Investigation of deep learning methods for classification and segmentation of chromosome and pulmonary images

Yulei Qin

N°d'ordre NNT : 2021LYSEI037

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
l'Institut National des Sciences Appliquées de Lyon
En cotutelle internationale avec
Shanghai Jiao Tong University

Ecole Doctorale N° ED160
Electronique, électrotechnique, automatique

Spécialité / discipline de doctorat :
Traitement du Signal et de l'Image

Soutenue publiquement/à huis clos le 30/06/2021, par :
Yulei QIN

Investigation of deep learning methods for classification and segmentation of chromosome and pulmonary images

---

Devant le jury composé de :

M. ZHENG Yuanjie, Professeur à Shandong Normal University	Rapporteur
MME RUAN Su, Professeure à Université de Rouen	Rapporteuse
M. DUPONT Florent, Professeur à Université Lyon 1	Examinateur
M. LIANG Dong, Professeur à Shenzhen Institute of Advanced Technology	Examinateur
M. ZHU Yue-Min, Directeur de Recherche CNRS à INSA de Lyon	Directeur de thèse
M. YANG Jie, Professeur à Shanghai Jiao Tong University	Co-directeur de thèse

Département FEDORA – INSA Lyon - Ecoles Doctorales

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
CHIMIE	**CHIMIE DE LYON** **https://www.edchimie-lyon.fr** Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage secretariat@edchimie-lyon.fr	**M. Stéphane DANIELE** C2P2-CPE LYON-UMR 5265 Bâtiment F308, BP 2077 43 Boulevard du 11 novembre 1918 69616 Villeurbanne **directeur@edchimie-lyon.fr**
E.E.A.	**ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE** **https://edeea.universite-lyon.fr** Sec. : Stéphanie CAUVIN Bâtiment Direction INSA Lyon Tél : 04.72.43.71.70 secretariat.edeea@insa-lyon.fr	**M. Philippe DELACHARTRE** INSA LYON Laboratoire CREATIS Bâtiment Blaise Pascal, 7 avenue Jean Capelle 69621 Villeurbanne CEDEX Tél : 04.72.43.88.63 **philippe.delachartre@insa-lyon.fr**
E2M2	**ÉVOLUTION, ÉCOSYSTÈME, MICROBIOLOGIE, MODÉLISATION** **http://e2m2.universite-lyon.fr** Sec. : Sylvie ROBERJOT Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 secretariat.e2m2@univ-lyon1.fr	**M. Philippe NORMAND** Université Claude Bernard Lyon 1 UMR 5557 Lab. d'Ecologie Microbienne Bâtiment Mendel 43, boulevard du 11 Novembre 1918 69 622 Villeurbanne CEDEX **philippe.normand@univ-lyon1.fr**
EDISS	**INTERDISCIPLINAIRE SCIENCES-SANTÉ** **http://ediss.universite-lyon.fr** Sec. : Sylvie ROBERJOT Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 secretariat.ediss@univ-lyon1.fr	**Mme Sylvie RICARD-BLUM** Institut de Chimie et Biochimie Moléculaires et Supramoléculaires (ICBMS) - UMR 5246 CNRS - Université Lyon 1 Bâtiment Raulin - 2ème étage Nord 43 Boulevard du 11 novembre 1918 69622 Villeurbanne Cedex Tél : +33(0)4 72 44 82 32 **sylvie.ricard-blum@univ-lyon1.fr**
INFOMATHS	**INFORMATIQUE ET MATHÉMATIQUES** **http://edinfomaths.universite-lyon.fr** Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage Tél : 04.72.43.80.46 infomaths@univ-lyon1.fr	**M. Hamamache KHEDDOUCI** Université Claude Bernard Lyon 1 Bât. Nautibus 43, Boulevard du 11 novembre 1918 69 622 Villeurbanne Cedex France Tél : 04.72.44.83.69 **hamamache.kheddouci@univ-lyon1.fr**
Matériaux	**MATÉRIAUX DE LYON** **http://ed34.universite-lyon.fr** Sec. : Yann DE ORDENANA Tél : 04.72.18.62.44 yann.de-ordenana@ec-lyon.fr	**M. Stéphane BENAYOUN** Ecole Centrale de Lyon Laboratoire LTDS 36 avenue Guy de Collongue 69134 Ecully CEDEX Tél : 04.72.18.64.37 **stephane.benayoun@ec-lyon.fr**
MEGA	**MÉCANIQUE, ÉNERGÉTIQUE, GÉNIE CIVIL, ACOUSTIQUE** **http://edmega.universite-lyon.fr** Sec. : Stéphanie CAUVIN Tél : 04.72.43.71.70 Bâtiment Direction INSA Lyon mega@insa-lyon.fr	**M. Jocelyn BONJOUR** INSA Lyon Laboratoire CETHIL Bâtiment Sadi-Carnot 9, rue de la Physique 69621 Villeurbanne CEDEX **jocelyn.bonjour@insa-lyon.fr**
ScSo	**ScSo*** **https://edsciencessociales.universite-lyon.fr** Sec. : Mélina FAVETON INSA : J.Y. TOUSSAINT Tél : 04.78.69.77.79 melina.faveton@univ-lyon2.fr	**M. Christian MONTES** Université Lumière Lyon 2 86 Rue Pasteur 69365 Lyon CEDEX 07 **christian.montes@univ-lyon2.fr**

*ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

Abstract

Pulmonary diseases can cause fatal damage to human health. Computed tomography (CT) helps display pulmonary structures and lesions for measurement and diagnosis. The advance of microscopy and karyotyping benefits pathogenesis study on the relationship between chromosomal abnormalities and lung diseases. In this thesis, to assist pulmonary disease analysis, we investigate deep learning methods for two purposes. The first is to classify Giemsa-stained chromosomes in microscopic imaging. The second is to segment pulmonary airways, arteries, veins, and nodules in CT.

We propose the Varifocal-Net for simultaneous classification of chromosome type and polarity via convolutional neural networks (CNNs). It performs robustly to different chromosome curvature, shape, and banding pattern.

For nodule segmentation, we propose a two-part CNNs-based method for all nodule textures and surroundings. The first part is to synthesize samples via generative adversarial network (GAN). The second part is to develop a segmentation model. For airways, their tree-like structure poses challenges to segmentation. We propose the AirwayNet to explicitly model connectivity between neighboring voxels. We further propose the AirwayNet-SE, more sophisticated than AirwayNet, by utilizing features of two context-scales. Finally, we propose a segmentation method for airways, arteries, and veins. To tackle sparse desired targets caused by severe class imbalance, we present the feature recalibration and attention distillation modules. Anatomy prior is incorporated for better artery-vein differentiation.

Keywords— Deep learning, Chromosome, Lung, Nodule, Airway, Artery-Vein, Computed tomography, Microscopy imaging, Classification, Segmentation

Résumé

Les maladies pulmonaires peuvent causer des dommages mortels à la santé humaine. La tomographie par rayons X (CT) permet d'obtenir les structures pulmonaires et les lésions pour la mesure et le diagnostic. L'avancée de la microscopie et du caryotypage profite à l'étude de la pathogenèse sur la relation entre les anomalies chromosomiques et les maladies pulmonaires. Dans cette thèse, pour aider à l'analyse des maladies pulmonaires, nous étudions des méthodes d'apprentissage en profondeur pour deux objectifs. Le premier est la classification des chromosomes colorés au Giemsa en imagerie microscopique. Le second est la segmentation des voies respiratoires pulmonaires, des artères, des veines et des nodules en CT.

Nous proposons le Varifocal-Net pour la classification simultanée du type et de la polarité des chromosomes via les réseaux de neurones convolutifs (CNN). Il fonctionne de manière robuste pour différentes courbures, formes et motifs de bandes chromosomiques.

Pour la segmentation des nodules, nous proposons une méthode de CNN composé de deux parties pour toutes les textures et tous les environnements des nodules. La première partie consiste à synthétiser des échantillons via un réseau antagoniste génératif (GAN). La deuxième partie vise à développer un modèle de segmentation. Pour les voies respiratoires, leur structure arborescente pose des problèmes de segmentation. Nous proposons AirwayNet pour modéliser explicitement la connectivité entre les voxels voisins. Nous proposons en outre AirwayNet-SE, plus sophistiqué que AirwayNet, en utilisant les caractéristiques des contextes à deux échelles. Enfin, nous proposons une méthode de segmentation des voies respiratoires, des artères et des veines. Pour faire face à des cibles désirées parcimonieux, causées par un sévère déséquilibre des classes, nous présentons les modules de recalibrage des caractéristiques et de distillation de l'attention. L'anatomie a priori est incorporée pour une meilleure différenciation artère-veine.

Mots-clés— Apprentissage profond, Chromosome, Poumon, Nodule, Bronche, Artère-Veine, Tomographie, Imagerie microscopique, Classification, Segmentation

Contents

Acknowledgement **v**

List of Figures **vii**

List of Tables **xiii**

Abbreviations **xvi**

Main Symbols **xviii**

Synthèse en Français de la thèse **1**

General Introduction **45**
0.1 Problem Statement and Objectives 45
0.2 Main Contributions 48
0.3 Organization of Thesis 50

1 Biomedical Context and Technical Background **53**
1.1 Chromosome Karyotyping 54
1.1.1 Giemsa Staining for Chromosome Imaging 54
1.1.2 Chromosome Separation and Classification 55
1.1.3 Enumeration and Abnormality Diagnosis 56
1.2 Pulmonary CT Image Segmentation 56
1.2.1 Anatomy of Human Pulmonary System 56
1.2.2 Lung Diseases Overview 60
1.2.3 Segmentation Methods Overview 61
1.3 State-of-the-art Deep Learning Methods 65
1.3.1 Artificial Neural Networks 65
1.3.2 Convolutional Neural Networks 67
1.3.3 CNNs Architectures Overview 69
1.3.4 Generative Adversarial Networks 72
1.3.5 Optimization Algorithms Overview 72
1.4 Summary 74

2 Development of a Chromosome Classification Approach Using Deep Convolutional Networks **77**
2.1 Introduction . . . 78
2.2 Methodology . . . 80
2.2.1 Stage 1: Global-Scale and Local-Scale Feature Learning . . . 81
2.2.2 Stage 2: Classification Based on the Fused Features . . . 87
2.2.3 Four-Step Training Strategy . . . 87
2.2.4 Stage 3: Type Assignment Using Dispatch Strategy . . . 88
2.3 Experiments and Results . . . 88
2.3.1 Materials . . . 88
2.3.2 Implementation Details . . . 89
2.3.3 Evaluation Metrics . . . 90
2.3.4 Results . . . 91
2.4 Discussion . . . 99
2.5 Conclusion . . . 105

3 Pulmonary Nodule Segmentation with CT Sample Synthesis Using Adversarial Networks **107**
3.1 Introduction . . . 108
3.2 Methodology . . . 111
3.2.1 Synthetic Image Generation . . . 111
3.2.2 Pulmonary Nodule Segmentation . . . 116
3.3 Experiments and Results . . . 120
3.3.1 Materials . . . 120
3.3.2 Implementation Details . . . 121
3.3.3 Evaluation Metrics . . . 121
3.3.4 Results . . . 122
3.4 Discussion . . . 125
3.5 Conclusion . . . 129

4 Development of a Voxel-Connectivity Aware Approach for Accurate Airway Segmentation Using Convolutional Neural Networks **131**
4.1 Introduction . . . 132
4.2 Methodology . . . 134
4.2.1 CT Volume Pre-processing . . . 134
4.2.2 Connectivity Modeling Using Binary Ground-Truth Labels . . . 135
4.2.3 Connectivity Prediction with AirwayNet . . . 137
4.2.4 Connectivity Prediction with AirwayNet-SE . . . 137
4.2.5 Airway Candidates Generation . . . 140
4.3 Experiments and Results . . . 140
4.3.1 Materials . . . 140
4.3.2 Implementation Details . . . 141
4.3.3 Evaluation Metrics . . . 141
4.3.4 Results . . . 141
4.4 Discussion . . . 144
4.5 Conclusion . . . 144

5 Learning Tubule-Sensitive Convolutional Neural Networks for Pulmonary Airway and Artery-Vein Segmentation in CT **145**
5.1 Introduction . . . 146
5.2 Methodology . . . 150
5.2.1 Feature Recalibration . . . 151
5.2.2 Attention Distillation . . . 153
5.2.3 Anatomy Prior for Artery-Vein Segmentation . . . 154
5.2.4 Model Design . . . 156
5.2.5 Training Loss . . . 157
5.3 Experiments and Results . . . 157
5.3.1 Materials . . . 157
5.3.2 Implementation Details . . . 158
5.3.3 Evaluation Metrics . . . 159
5.3.4 Results . . . 161
5.4 Discussion . . . 168
5.5 Conclusion . . . 175

6 General Conclusions and Perspectives **176**

List of Publications **179**

Bibliography **181**

Appendices **201**

Acknowledgement

Sticking to the cliché, I have to admit that the Ph.D. student days at INSA have gone in the blink of an eye. I still remember the first moment when I arrived in France and found everything brand-new and fascinating to me. But good times don't last forever, a severe pandemic suddenly spread over the world and changed the way we live. Immense gratitude is due to a lot of people who helped me go through this period with joy.

First and foremost, I would like to express sincere gratitude to my supervisor, Prof. Yue-Min Zhu, for his continuous support of my Ph.D. study since 2017. Without him, it would not be possible for me to know well how to conduct research in the field of medical imaging. He always provides me with an open research environment where I could freely choose interesting topics and do what I believe. His patient guidance, profound knowledge, and instructive comments helped me throughout the entire stage from problem definition to method development, experiment analysis, paper writing and revision. In addition, he often shares anecdotes about french culture and lifestyle with us, broadening my horizon. I will always be impelled and motivated by his rigorous research attitude, sense of responsibility, diligence, and shining personal characteristics.

My sincere thanks also go to my co-supervisor, Prof. Jie Yang, for his thoughtful and meticulous guidance. He encourages me to set high standard for academic discipline and publication. Besides, he offered me many opportunities to discuss with peer researchers, such as supporting me for attending international conferences and seminars. He concerns about the prospect of us students and provides me with a precious chance of exchange in France.

I would like to thank Prof. Guang-Zhong Yang, Prof. Xiaolin Huang, and Prof. Yun Gu for their collaboration. Their enlightening suggestions and comments helped me out of dilemmas whenever I felt confused and disappointed. I also acknowledge the support from my fellow colleagues and students including but not limited to Dr. Lei Zhou, Dr. Fanghui Liu, Dr. Donghao Shen, Hao Zheng, Xiang Wang, Mingjian Chen, Xingyu Chen, Sanli Tang, and Tianyi Zhang. I also want to thank Dr. Pei Niu, Dr. Bingqing Xie, Zexian Wang, Yunlong He, and Jiqing Huang for our friendship in Lyon.

Meanwhile, I would like to particularly thank my girlfriend Wen Yue for her accompany all the time. Thanks for her kindness, understanding, and sharing of sorrow and happiness. Her smile lights me up inside and dispels gloom away. Finally, I can't thank my parents and family enough for continuously supporting and encouraging me. I am grateful to their unconditional love and caring.

List of Figures

1 (a) Image microscopique des chromosomes mâles colorée au Giemsa pour un cas. (b) Le résultat du caryotype (alias caryogramme) de (a) est formé des chromosomes appariés et ordonnés (22 paires d'autosomes et 1 paire de chromosomes sexuels XY). 9
2 Le modèle de neurones artificiels (Figure inspirée de [Willems, 2019]). . . 12
3 Architecture du modèle LeNet-5 [LeCun et al., 1998]. 13
4 Architecture du modèle AlexNet [Krizhevsky et al., 2012]. 13
5 Différentes architectures CNN. (a) Architecture à chemin de liaison unique (b) Architecture à transmission par contournement (c) Architecture parallèle (d) Architecture à branchement multi-tâches 15
6 La focalisation varie de globale à locale. Étant donné les images de chromosomes (A, B, C), le sous-réseau de localisation détecte leurs régions les plus fines pour les recadrer et les agrandir. (a) Les images originales de chromosomes. (b) Les parties locales après un zoom avant. 16
7 Organigramme du réseau Varifocal proposé pour la classification des chromosomes. 18
8 Première étape du réseau varifocal proposé: extraction de caractéristiques à l'échelle globale et locale via le G-Net et le L-Net, respectivement. 19
9 La deuxième étape du Varifocal-Net proposé : la classification des chromosomes en utilisant des caractéristiques fusionnées à la fois à l'échelle globale et locale. 20
10 Vue d'ensemble du cadre proposé pour la segmentation des nodules pulmonaires. Des images synthétiques de nodules sont d'abord générées. Ensuite, les images originales et synthétiques sont utilisées pour entraîner le modèle de segmentation. Les résultats de la segmentation sont des masques binaires 3D de la VOI du nodule. 23
11 L'architecture réseau du cGAN proposé. 24
12 L'architecture réseau du cadre de segmentation proposé. 26
13 Organigramme de la proposition AirwayNet et AirwayNet-SE. 29
14 Illustration de l'étape de prétraitement de l'image CT. 30
15 Illustration de la modélisation de la 26-connectivité. 31
16 Illustration de l'AirwayNet. Le nombre de canaux est indiqué au-dessus de chaque carte de caractéristiques. 33

17 Illustration de l'AirwayNet-SE. Le nombre de canaux est indiqué sur chaque carte de caractéristiques. La première étape consiste à extraire des caractéristiques de deux échelles de contexte via DNN et SWN. La deuxième étape consiste à classifier la connectivité des voies aériennes en utilisant des caractéristiques fusionnées avec des contextes à grande et à petite échelle. 34
18 Vue d'ensemble de la méthode proposée pour la segmentation des voies aériennes pulmonaires et des veines artérielles. La normalisation des instances et l'activation ReLU sont effectuées après chaque couche de convolution, sauf la dernière. Le nombre de noyaux de convolution est indiqué au-dessus de chaque couche. 39
19 Illustration de la cartographie $\mathcal{Z}(\cdot)$ pour le recalibrage des caractéristiques. Son entrée est la caractéristique activée A_m de la m-ième couche de convolution. Tout d'abord, la carte spatiale qui met en évidence les régions importantes est intégrée par l'intermédiaire de $\mathcal{Z}_{spatial}(\cdot)$ selon trois axes : profondeur, hauteur et largeur. Ensuite, la recombinaison des canaux est effectuée sur la carte spatiale pour calculer le descripteur de canal U_m. La multiplication finale par éléments entre A_m et U_m produit la caractéristique recalibrée $\hat{A}_m$. Les notations r, C_m, D_m, H_m et W_m font référence au facteur de compression des canaux, au nombre de canaux, aux profondeurs, aux hauteurs et aux largeurs de A_m, respectivement. 40
20 Illustration de l'anatomie avant l'incorporation. La représentation visuelle des cartes de contexte pulmonaire et des cartes de transformation de distance générées, superposées aux CT images, est donnée en bas à gauche. . 42

1.1 (a) A Giemsa-stained microscopic image of male chromosomes for one case. (b) The karyotyping result (a.k.a. karyogram) of (a) is formed of the paired and ordered chromosomes (22 pairs of autosomes and 1 pair of sex chromosomes XY). 54
1.2 Anatomy of the human pulmonary system [Wikimedia, 2021]. 57
1.3 Generations of branches in a typical human airway tree [U. S. National Institutes of Health, 2021]. 58
1.4 Illustration of pulmonary circulation [Wikimedia, 2020]. 59
1.5 Segmentation of pulmonary nodule, airway, vessel (artery and vein) in CT. 61
1.6 The artificial neuron model (Figure inspired by [Willems, 2019]). 65
1.7 Illustration of back-propagation algorithm. 66
1.8 Architecture of the LeNet-5 model [LeCun et al., 1998]. 67
1.9 Hierarchical features of CNNs [Lee et al., 2011]. 67
1.10 Architecture of the AlexNet model [Krizhevsky et al., 2012]. 68
1.11 Convolution blocks in different CNNs. (a) VGG block (b) GoogLeNet Inception block (c) Residual block (d) Densely connected block 68
1.12 Different CNN architectures. (a) Single link path architecture (b) Bypass transmission architecture (c) Parallel architecture (d) Multi-task branch architecture 70
1.13 The training of GAN proceeds by alternatively training the generator and the discriminator. 73

2.1 The focus is varied from global to local. Given chromosome images (A, B, C), the localization subnet detects their finer regions to crop and magnify. (a) The original chromosome images. (b) The local parts after zooming in. 79
2.2 Flowchart of the proposed Varifocal-Net for chromosome classification. . 81
2.3 The first stage of the proposed Varifocal-Net: global-scale and local-scale feature extraction via the G-Net and the L-Net, respectively. 82
2.4 Wide residual unit. n_in and n_out stand for number of input and output feature channels, respectively. (a) if n_in $\neq$ n_out. (b) if n_in = n_out. . . . 83
2.5 The diagram of parameterizations for the sample x_i. (a) The red box is the predicted local region and the gray background square is the area where the box's center pixel (x_c, y_c) can be located. (b) The side length of the predicted box $(2t_l^i)$ is restricted, ranging from T_1 to $2T_1$. 85
2.6 The second stage of the proposed Varifocal-Net: chromosome classification using fused features from both global and local scales. 87
2.7 ROC analysis for the proposed Varifocal-Net and previous CNN models. Each ROC is averaged over all classes and its AUC is calculated. (a) ROC of type classification. (b) ROC of polarity classification. 100
2.8 Confusion matrix of the Varifocal-Net for type classification. The entry in the i-th row and j-th column denotes the percentage (%) of the testing samples from class i that were classified as class j. Best viewed magnified. 101
2.9 Feature embedding for chromosomes with t-SNE toolbox [Maaten and Hinton, 2008]. From the perspective of type classification, the global, local, and concatenated features are visualized in (a), (c), and (e), respectively. Similarly, these three features are visualized in (b), (d), and (f) correspondingly for polarity classification. The mixed regions of interest are marked with black circles. Best viewed in color. 102
2.10 Examples of correctly classified samples. Both global-scale and local-scale inputs are displayed to visually assess the varifocal mechanism. 103
2.11 Examples of misclassified samples. The probabilities of wrong predictions are displayed on the right of each image and each red rectangle encloses the predicted probability of the ground-truth label. 104

3.1 Typical cases for each nodule type. First row: Nodules are classified by internal texture. (a) GGO; (b) part-solid; (c) solid. Second row: Nodules are classified by external surroundings. (d) well-circumscribed; (e) juxta-vascular; (f) juxta-pleural. 108
3.2 An overview of the proposed pulmonary nodule segmentation framework. Synthetic nodule images are first generated. Then, both the original and synthesized images are used to train the segmentation model. The segmentation results are 3D binary masks of nodule VOI. 110
3.3 The process of generating ten-channel labels. (a) CT image of nodule; (b) Generated semantic label containing a nodule, pleural surfaces, and vascular structures; (c) Each attribute's scoring value is multiplied with a binary ground-truth label; (d) The semantic label and nine attribute labels are concatenated as an input image with ten channels. 112

3.4 The training of cGAN proceeds by alternatively training G and D. Given a label image and a noise vector, G is trained to obtain a realistic image. The synthetic pair and real pair refer to the ten-channel label concatenated with synthetic image and real image, respectively. D learns to distinguish real pairs from synthetic fake pairs. 114
3.5 The network architecture of the proposed cGAN. 115
3.6 The network architecture of the proposed segmentation framework. . . . 117
3.7 Residual unit. n_in and n_out denote the number of channels of input cube and output cube, respectively. 119
3.8 Post-block. 119
3.9 Examples of generated synthetic images. (a) Input labels; (b) Real images; (c) Generated images. Out of simplicity, ten-channel inputs are briefly displayed as semantic labels. 123
3.10 Comparison of five synthetic samples generated with and without the nine attributes labels. (a) Real CT images; (b) Semantic labels; (c) Images generated without nine attributes. (d), (e), and (f) stand for the images generated with nine attributes and their texture scores are set to 1, 3, and 5, respectively. 124
3.11 Qualitative segmentation results of validation samples. (a) Ground-truth labels are in green; (b) Predicted nodules are in red. The score beneath each pair is Dice coefficient of the result. Central slice of each VOI cube is displayed for simplicity. 126

4.1 The intensity distribution of trachea (a), primary (b) and secondary (c) bronchus, and peripheral bronchiole (d). The scale of contexts needed for airway segmentation on (a)-(d) is decreasing from large to small. 132
4.2 Flowchart of the proposed AirwayNet and AirwayNet-SE. 134
4.3 Illustration of the CT pre-processing step. 135
4.4 Illustration of 26-connectivity modeling. The binary ground-truth of airway (Dim: $1 \times Z \times H \times W$) is transformed into a connectivity label (Dim: $26 \times Z \times H \times W$). For each voxel P, we extract a $3 \times 3 \times 3$ neighborhood cube to check connected voxel pairs. Each pair $(P, Q_i), i \in \{1,2,...,26\}$ represents a connectivity orientation and is encoded with a binary label. For example, if an airway voxel P is connected to its neighbor Q_{20}, then the corresponding position "P" on the 20-th label is marked as 1. 136
4.5 Illustration of the AirwayNet. The number of channels is denoted above each feature map. 138
4.6 Illustration of the AirwayNet-SE. The number of channels is denoted on each feature map. The first stage is to extract features of two context scales via DNN and SWN. The second stage is to classify the connectivity of airways using fused features with both large-scale and small-scale contexts. . 139
4.7 Comparison of airway segmentation results between the AirwayNet-SE, DNN, SWN, and ground-truth. 143

5.1 Overview of the proposed method for pulmonary airway and artery-vein segmentation. Instance normalization and ReLU activation are performed after each convolution layer except the last one. The number of convolution kernels is denoted above each layer. 149
5.2 Illustration of the mapping $\mathcal{Z}(\cdot)$ for feature recalibration. Its input is the activated feature A_m of the m-th convolution layer. First, spatial map that highlights important regions is integrated through $\mathcal{Z}_{spatial}(\cdot)$ along three axes of depth, height, and width. Second, channel recombination is performed on the spatial map to compute the channel descriptor U_m. The final element-wise multiplication between A_m and U_m produces the recalibrated feature $\hat{A}_m$. The notations r, C_m, D_m, H_m, and W_m refer to the channel compression factor, the number of channels, depths, heights, and widths of A_m, respectively. 151
5.3 Difference among mapping functions $\mathcal{G}(\cdot)$ of computing the last attention map in decoder for airway and artery-vein segmentation tasks. 154
5.4 Illustration of anatomy prior incorporation. Visual display of the generated lung context maps and distance transform maps superimposed on CT scans is given in bottom left. 155
5.5 Rendering of pulmonary airway segmentation results on (a) easy and (b) hard testing cases. Best viewed magnified. 169
5.6 Rendering of pulmonary artery-vein segmentation results on (a) easy and (b) hard testing cases. Left: Wrongly segmented arteries and veins are zoomed in for better visual inspection. Right: Difference between prediction and label is categorized into 5 types. 170
5.7 Pseudo-color rendering of attention maps (decoder 1–4) before and after distillation process. These maps are min-max scaled and rendered with Jet colormap. Best viewed magnified. 173
5.8 Analysis of multi-class artery-vein segmentation results. (a) Normalized confusion matrix. (b) Percentage of different types of errors. (c) Typical examples of wrong predictions. Best viewed magnified. 174

B1 Percentage of different types of errors by (a) the proposed method, (b) dense CRFs (3 iterations), and (c) dense CRFs (10 iterations). 208

List of Tables

2.1 Statistics of the dataset. (H: Healthy Samples, U: Unhealthy Samples.) . . 90
2.2 The feature dimensions of the Varifocal-Net for the first stage. (T: Type, P: Polarity, Loc: Localization.) 90
2.3 The feature dimensions of the Varifocal-Net for the second stage. (T: Type, P: Polarity.) 90
2.4 Performance of the Varifocal-Net (mean±standard deviation). The results are presented in terms of four evaluation metrics: average F_1-score of all testing images (F_1), accuracy of all testing images (Acc.), average accuracy per patient case (Acc. per Case), and average accuracy per patient case using the proposed dispatch strategy (Acc. per Case-D). (T: Type, P: Polarity, PET: Per Epoch Time, TPI: Time Per Image.) 93
2.5 Performance of the Varifocal-Net for each chromosome type (mean±standard deviation). 94
2.6 Performance of the Varifocal-Net for each chromosome polarity (mean±standard deviation). 94
2.7 Performance of the Varifocal-Net for polarity classification within each type (mean±standard deviation). 95
2.8 Comparison results of the proposed method with state-of-the-art methods (mean±standard deviation). The results are presented in terms of four evaluation metrics: average F_1-score of all testing images (F_1), accuracy of all testing images (Acc.), average accuracy per patient case (Acc. per Case), and average accuracy per patient case using the proposed dispatch strategy (Acc. per Case-D). (T: Type, P: Polarity.) 96
2.9 Comparison results of the proposed method with state-of-the-art methods on unhealthy cases (mean±standard deviation). The results are presented in terms of four evaluation metrics: average F_1-score of all testing images (F_1), accuracy of all testing images (Acc.), average accuracy per patient case (Acc. per Case), and average accuracy per patient case using the proposed dispatch strategy (Acc. per Case-D). (T: Type, P: Polarity.) 97

3.1 Distributions of the 1182 pulmonary nodules from the LIDC-IDRI dataset. 111
3.2 Definition of scoring for each nodule attribute. 113
3.3 Quantitative results of synthetic image generation for different nodule categories. 123
3.4 The segmentation results of the proposed model for different nodule categories. 124

3.5 Comparison of segmentation results in DSC. 125
3.6 Quantitative comparison results of the control group. 125

4.1 Results of the proposed AirwayNet and AirwayNet-SE in comparison with state-of-the-art methods (mean±standard deviation). 142
4.2 Ablation study of the proposed AirwayNet and AirwayNet-SE (mean±standard deviation). The DNN and SWN stand for Deep-yet-Narrow Network, Shallow-yet-Wide Network, respectively. 143

5.1 Computational time of the proposed pulmonary airway and artery-vein segmentation method. 160
5.2 Comparison of pulmonary airway segmentation results. The results both under the same binarization threshold and under the same FPR are presented for each method. The FPR is controlled to be the same with 3-D U-Net (under threshold of 0.5) by respectively adjusting the binarization threshold on the probability outputs of each method. 162
5.3 Evaluation results on the EXACT'09 testing set. 163
5.4 Comparison of pulmonary artery-vein segmentation results. 164
5.5 Results of ablation study on pulmonary airway segmentation. The results both under the same binarization threshold and under the same FPR are presented for each method. The FPR is controlled to be the same with 3-D U-Net (under threshold of 0.5) by respectively adjusting the binarization threshold on the probability outputs of each method.
cSE = channel-Squeeze-Excitation, PE = Project-Excitation, FR = Feature Recalibration, AD = Attention Distillation, DS = Deep Supervision 166
5.6 Results of ablation study on pulmonary artery-vein segmentation.
cSE = channel-Squeeze-Excitation, PE = Project-Excitation, FR = Feature Recalibration, AD = Attention Distillation, DS = Deep Supervision, AP = Anatomy Prior 167

A1 Results of unpaired and paired t-tests between the proposed Varifocal-Net and other methods. Scientific notation is used to express numbers. (T: Type, P: Polarity.) 202
A2 Results of unpaired and paired t-tests between the proposed Varifocal-Net and other methods on unhealthy cases. Scientific notation is used to express numbers. (T: Type, P: Polarity.) 203
A3 Classification performance of the Leica's CytoVision System and the proposed Varifocal-Net on 10 patient cases (mean±standard deviation). . . . 203
B1 Computational time of the graph-based post-processing step for pulmonary artery-vein segmentation. 205

B2 Results of the graph-based post-processing after the proposed artery-vein segmentation method. For each 3-D CT scan, the original intensity image and CNNs' probability outputs of background, arteries, and veins are taken as inputs to Dense CRF for post-processing. **Union 1**: The union of artery voxels before and after post-processing is kept as final artery predictions. The union of vein voxels before and after post-processing intersects with the set of non-artery voxels to obtain the final vein predictions. The remaining voxels belong to the background. **Union 2**: The union of vein voxels before and after post-processing is kept as final vein predictions. The union of artery voxels before and after post-processing intersects with the set of non-vein voxels to obtain the final artery predictions. The remaining voxels belong to the background. 206
B3 Results of ablation study on pulmonary airway segmentation. Both the results under the same binarization threshold and under the same FPR are presented for each method. 209
B4 Results of ablation study on pulmonary artery-vein segmentation. 210

Abbreviations

Acc	Accuracy
Adam	Adaptive moment estimation
ANN	Artificial Neural Network
ARDS	Acute Respiratory Distress Syndrome
AUC	Area Under the Curve
AV	Artery-Vein
AVM	ArterioVenous Malformations
BD	Branches Detected
BN	Batch Normalization
BP	Back-Propagation
CAD	Computer-Aided Diagnosis
cGAN	conditional Generative Adversarial Network
CNN	Convolutional Neural Network
COPD	Chronic Obstructive Pulmonary Disease
CRF	Conditional Random Field
CT	Computed Tomography
CTPA	Computed Tomography Pulmonary Angiogram
DSC	Dice Similarity Coefficient
FC	Fully Connected
FCN	Fully Convolutional Network
FN	False Negative
FP	False Positive
FPR	False Positive Rate
GAN	Generative Adversarial Network

GGO Ground Glass Opacity

GNN Graph Neural Network

GPU Graphics Processing Unit

HU Hounsfield Unit

IN Instance Normalization

LBP Local Binary Pattern

MAT Medial Axis Transform

MLP Multi-Layer Perception

MSE Mean Squared Error

PE Pulmonary Embolism

PH Pulmonary Hypertension

PPV Positive Predictive Value

ReLU Rectified Linear Unit

ROC Receiver Operating Characteristic

ROI Region Of Interest

SGD Stochastic Gradient Descent

Std Standard deviation

SVM Support Vector Machine

TD Tree-length Detected

TN True Negative

TP True Positive

TPR True Positive Rate

VOI Volume Of Interest

Main Symbols

A_m	Activation output of the m-th convolution
$\mathcal{B}(\cdot)$	Broadcasting operation
D	Discriminator
E	Error between prediction and ground-truth
G	Generator
G_m	Attention map of A_m
$\mathcal{L}$	Loss
O	Output prediction
S_C	Cosine similarity
U_m	Channel descriptor of A_m
W_c	Parameters of convolution kernels
x_i	i-th sample
X	Sample
y_i	i-th sample's label
Y	Label
z	Random noise
$\|\cdot\|_F$	Frobenius norm
$\|\cdot\|_1$	L1 norm
$\|\cdot\|_2$	L2 norm
$\nabla \boldsymbol{X}$	Gradient matrix of $\boldsymbol{X}$
θ	Parameters of networks

Synthèse en Français de la thèse

Introduction Générale

Énoncé du Problème et Objectifs

Les maladies pulmonaires, dont la bronchopneumopathie chronique obstructive (BPCO) et le cancer du poumon, peuvent avoir des conséquences fatales pour la santé humaine. La BPCO se caractérise par une limitation persistante du débit d'air causée par des anomalies des voies respiratoires et des alvéoles pulmonaires [Vogelmeier et al., 2017, Halpin et al., 2021, Sethi and Rochester, 2000]. L'inflammation des voies aériennes et la destruction emphysémateuse du tissu pulmonaire sont souvent observées chez les patients qui sont exposés à des particules ou à des gaz toxiques pendant une longue période [Hogg, 2004]. Le cancer du poumon est l'un des principaux cancers chez les hommes et les femmes, causant 1,3 million de décès par an dans le monde [Torre et al., 2016]. Les nodules pulmonaires sont de petites excroissances de forme ronde ou ovale dans les poumons et sont souvent considérés comme une indication précoce de cancer. Les nodules sont généralement le résultat d'une inflammation du poumon. Plus de 90% des nodules solides sont bénins si leur diamètre est inférieur à 2 cm [Winer-Muram, 2006].

L'utilisation très répandue de la tomographie à rayons X (CT) permet de visualiser les structures pulmonaires pour un diagnostic précis des maladies. Pour l'analyse de l'imagerie tomographique pulmonaire, une étape préalable consiste à extraire les voies aériennes pulmonaires de la tomographie. La modélisation de l'arbre des voies respiratoires permet de quantifier ses changements morphologiques pour le diagnostic de la sténose bronchique, du syndrome de détresse respiratoire aiguë, de la fibrose pulmonaire idiopathique, de la BPCO, de la bronchiolite oblitérante et de la contusion pulmonaire [Howling et al., 1998, Shaw et al., 2002, Fetita et al., 2004, Li et al., 2019, Wu et al., 2019]. Combinées à un rendu et une projection photo-réalistes, les voies aériennes segmentées jouent un rôle important dans la bronchoscopie virtuelle et la navigation endobronchique pour la chirurgie; [Mori et al., 2000, Natori et al., 2005, Shen et al., 2015a, Shen et al., 2019]. Une autre étape essentielle consiste à extraire les artères et les veines pulmonaires du

scanner. Les maladies pulmonaires peuvent affecter les artères ou les veines, ou les deux, mais de manière différente [Melot and Naeije, 2011, Charbonnier et al., 2015]. Les modifications morphologiques des artères sont mesurées dans le diagnostic de l'embolie pulmonaire, des malformations artério-veineuses et de la BPCO [Zhou et al., 2007, Wittenberg et al., 2012, Cartin-Ceba et al., 2013, Estépar et al., 2013]. Les altérations artérielles servent également de biomarqueur d'imagerie dans l'hypertension pulmonaire thromboembolique chronique [Rahaghi et al., 2016]. Les caractéristiques d'imagerie des veines sont utiles pour le diagnostic des maladies veineuses [Porres et al., 2013]. Malgré les avantages de la segmentation des veines des voies respiratoires et des artères, elle nécessite de lourdes charges de travail pour la délimitation manuelle en raison de la complexité des structures tubulaires. Par conséquent, des méthodes de segmentation automatique ont été développées pour réduire la charge de travail et améliorer la précision. En particulier, si les artères et les veines peuvent être extraites à partir d'une coupe CT sans contraste (c'est-à-dire sans l'utilisation d'agents de contraste), l'angiographie pulmonaire par CT peut ne pas être nécessaire dans certains cas pour éviter les réactions indésirables aux agents de contraste [Cochran et al., 2001, Loh et al., 2010]. Néanmoins, l'extraction des voies respiratoires, des artères et des veines par des méthodes de segmentation automatique est sujette à la discontinuité, car il existe un déséquilibre de classe important entre l'avant-plan tubulaire et l'arrière-plan. Les voxels des voies respiratoires et des vaisseaux sont peu nombreux et dispersés par rapport à l'arrière-plan. En outre, il existe des différences entre les branches principales et épaisses et les branches périphériques et fines en termes d'intensité et de distribution spatiale. Les méthodes de segmentation des voies respiratoires, des artères et des veines sont nécessaires pour percevoir et traiter ces différences entre les échelles locale et globale.

Des systèmes de diagnostic assisté par ordinateur (DAO) ont été mis au point pour améliorer le diagnostic des nodules pulmonaires par CT: [Messay et al., 2010, Lopez Torres et al., 2015, Jacobs et al., 2014, Setio et al., 2016, Sakamoto and Nakano, 2016, Dou et al., 2017, Huang et al., 2017b]. Dans la conception des systèmes de CAO des nodules, l'étape préalable est la segmentation des nodules. Par rapport à la délimitation manuelle des nodules par les radiologues, ces systèmes fournissent efficacement des résultats de prédiction cohérents sans variance inter-observateur. La qualité de la segmentation affecte directement la mesure ultérieure des nodules pour la classification de la bénignité et de la malignité. La principale difficulté de la segmentation des nodules est de concevoir un algorithme qui s'adapte à la fois à la texture interne et à l'environnement externe des nodules pulmonaires. La plupart des méthodes de segmentation précédentes ont été développées pour les nodules solides. Peu de méthodes étaient applicables à la segmentation de tous les nodules solides, partiellement solides et à opacité en verre

dépoli (GGO). En outre, la similitude entre les nodules et le tissu pulmonaire en termes d'intensité et l'environnement compliqué des nodules posent des défis non négligeables à la généralisation des méthodes de segmentation. Pour les nodules qui sont reliés à la surface de la plèvre, aux vaisseaux et aux parois des voies respiratoires, les méthodes de segmentation ne parviennent souvent pas à générer des limites précises, ce qui entraîne une segmentation insuffisante ou excessive. La raison en est que les nodules partagent une intensité similaire à celle des tissus environnants. Dans ces circonstances, il est très important que la méthode de segmentation comprenne bien la forme, la texture et la distribution de la position des nodules.

Les progrès de l'imagerie microscopique permettent d'étudier la pathogenèse des maladies pulmonaires au niveau des chromosomes. Associées au cancer du poumon, les aberrations non aléatoires des chromosomes sont complexes, avec de multiples réarrangements numériques et structurels: [Balsara and Testa, 2002, Testa and Siegfried, 1992, Park et al., 2001, Masuda and Takahashi, 2002, Grigorova et al., 2005]. De plus, des anomalies en mosaïque des chromosomes somatiques ont été détectées dans les poumons de patients souffrant d'hypertension artérielle pulmonaire : "aldred2010somatic". Pour effectuer une analyse chromosomique dans le cadre de l'étude des maladies pulmonaires, une procédure importante est le caryotypage, au cours duquel les chromosomes en métaphase dans une cellule sont colorés, imagés, classés et triés dans l'ordre [Piper, 1990]. Selon la technique de coloration et le mécanisme d'imagerie, le caryotypage peut être divisé en caryotypage de Giemsa et caryotypage fluorescent. Le caryotype de Giemsa est préféré car il peut détecter presque toutes les anomalies avec un seul test peu coûteux. Cependant, les cytogénéticiens doivent déployer des efforts méticuleux pour classer les chromosomes colorés au Giemsa en fonction de leurs bandes. De nombreuses méthodes de classification automatique des chromosomes ont été proposées pour améliorer l'efficacité du caryotypage. La plupart d'entre elles reposent sur une extraction précise des axes médians et des centromères pour le calcul des caractéristiques: [Lerner et al., 1995, Ming and Tian, 2010, Markou et al., 2012, Stanley et al., 1996, Wang et al., 2008, Arachchige et al., 2013, Loganathan et al., 2013]. En raison de la difficulté de la squelettisation pour les chromosomes courbés et déformés, les méthodes précédentes n'ont souvent pas réussi à obtenir une classification robuste. Des méthodes capables de faire face à de grandes variations de forme et d'apparence des chromosomes sont nécessaires pour répondre aux normes cliniques.

Les principaux objectifs de recherche de cette thèse sont doubles: l'un consiste à développer une méthode de classification des chromosomes en imagerie microscopique pour le caryotypage au Giemsa ; l'autre consiste à développer des méthodes de segmentation des voies aériennes, des artères, des veines et des nodules pulmonaires en imagerie

CT pour la mesure et le diagnostic. Les études portant sur ces deux objectifs jettent les bases d'explications sur la pathogenèse et les conséquences des maladies pulmonaires, le premier objectif se situant à une micro-échelle et le second à une macro-échelle. Pour faire face aux limitations des méthodes de classification et de segmentation dans la littérature, les approches d'apprentissage profond sont étudiées dans le développement de méthodes. Par rapport aux méthodes traditionnelles qui dépendent de caractéristiques conçues manuellement, les méthodes basées sur l'apprentissage profond apprennent des caractéristiques efficaces qui devraient être plus robustes aux variations des objets. Ainsi, dans la présente étude, les approches d'apprentissage profond sont exploitées pour améliorer les performances de la classification et de la segmentation des images de chromosomes et de poumons.

Principales Contributions

Les principales contributions de cette thèse sont détaillées comme suit :

- Développement d'une Approche de Classification des Chromosomes à l'aide de Réseaux Convolutifs Profonds (Chapitre 2).

  La classification des chromosomes est essentielle pour le caryotypage dans le diagnostic des anomalies. Pour accélérer le diagnostic, nous présentons une nouvelle méthode appelée Varifocal-Net pour la classification simultanée du type et de la polarité des chromosomes en utilisant des réseaux convolutifs profonds. L'approche consiste en un réseau à échelle globale (G-Net) et un réseau à échelle locale (L-Net). Elle suit trois étapes. La première étape consiste à apprendre les caractéristiques globales et locales. Nous extrayons les caractéristiques globales et détectons les régions locales plus fines via le G-Net. En proposant un mécanisme Varifocal, nous zoomons sur les parties locales et extrayons les caractéristiques locales via le L-Net. Des stratégies d'apprentissage résiduel et d'apprentissage multi-tâches sont utilisées pour promouvoir l'extraction de caractéristiques de haut niveau. La détection des parties locales discriminantes est réalisée par un sous-réseau de localisation du G-Net, dont le processus d'apprentissage implique un apprentissage supervisé et faiblement supervisé. La deuxième étape consiste à construire deux classifieurs perceptron multicouches qui exploitent les caractéristiques des deux échelles pour améliorer les performances de classification. La troisième étape consiste à introduire une stratégie d'affectation de chaque chromosome à un type dans chaque cas de patient, en utilisant la connaissance du domaine du caryotypage. Les résultats de l'évaluation de 1909 cas de caryotypage ont montré que le réseau Varifocal proposé a atteint la plus haute précision par cas de patient de 99.2% pour les tâches de

type et de polarité. Il a surpassé les méthodes de pointe, démontrant l'efficacité de notre mécanisme Varifocal, de notre ensemble de caractéristiques multi-échelles et de notre stratégie de répartition. La méthode proposée a été appliquée pour aider au diagnostic pratique du caryotype.

- Segmentation de Nodules Pulmonaires par Synthèse d'échantillons CT à l'aide de Réseaux Adversariaux (Chapitre 3).

  La segmentation des nodules pulmonaires est essentielle pour l'analyse des nodules et le diagnostic du cancer du poumon. Nous présentons un nouveau cadre de segmentation pour différents types de nodules en utilisant des réseaux de neurones convolutifs (CNN). Le cadre proposé est composé de deux parties principales. La première partie consiste à augmenter la variété des échantillons et à construire un ensemble de données plus équilibré. Un réseau adversatif génératif conditionnel (cGAN) est utilisé pour produire des images CT synthétiques. Des étiquettes sémantiques sont générées pour transmettre au réseau des connaissances sur le contexte spatial. Neuf étiquettes de notation d'attributs sont également combinées pour préserver les caractéristiques des nodules. Pour affiner le réalisme des échantillons synthétisés, une perte d'erreur de reconstruction est introduite dans cGAN. La deuxième partie consiste à entraîner un réseau de segmentation des nodules sur le jeu de données étendu. Nous construisons un modèle CNN 3D qui exploite des cartes hétérogènes, notamment des cartes de bords et des cartes de motifs binaires locaux. L'incorporation de ces cartes informe le modèle des motifs de texture et des informations sur les limites des nodules, ce qui facilite l'apprentissage des caractéristiques de haut niveau pour la segmentation. L'unité résiduelle, qui apprend à réduire l'erreur résiduelle, est adoptée pour accélérer l'apprentissage et améliorer la précision. La validation sur le jeu de données LIDC-IDRI démontre que les échantillons générés sont réalistes. L'erreur quadratique moyenne et la similarité cosinus moyenne entre les échantillons réels et synthétisés sont respectivement de 1.55×10^{-2} et 0.9534. Le coefficient de Dice, la valeur prédite positive, la sensibilité et la précision sont respectivement de 0.8483, 0.8895, 0.8511 et 0.9904 pour les résultats de la segmentation. Le cadre de segmentation CNN 3D proposé, basé sur l'utilisation d'échantillons synthétisés et de cartes multiples avec apprentissage résiduel, permet une segmentation plus précise des nodules par rapport aux méthodes de pointe existantes. La méthode de synthèse d'images CT proposée peut non seulement produire des échantillons proches des images réelles, mais aussi permettre une variation stochastique de la diversité des images.

- Développement d'une Approche Tenant Compte de la Connectivité des Voxels

pour une Segmentation Précise des Voies Respiratoires à l'aide de Réseaux Neuronaux Convolutifs (Chapitre 4).

La segmentation précise des voies respiratoires à partir de tomographies thoraciques est cruciale pour le diagnostic des maladies pulmonaires et la navigation chirurgicale. Cependant, la variété intra-classe des voies respiratoires et leur structure arborescente intrinsèque posent des problèmes pour le développement de méthodes de segmentation automatique. Pour y remédier, nous proposons une approche basée sur la connectivité des voxels, appelée AirwayNet, pour une segmentation précise des voies respiratoires. Grâce à la modélisation de la connectivité, la tâche de segmentation binaire conventionnelle est transformée en 26 tâches de prédiction de la connectivité. Ainsi, notre AirwayNet apprend à la fois la structure des voies aériennes et la relation entre les voxels voisins. Nous proposons ensuite l'AirwayNet-SE en deux étapes, une approche simple mais efficace pour améliorer AirwayNet. La première étape de AirwayNet-SE consiste à adopter la modélisation de la connectivité pour transformer la tâche de segmentation binaire en une tâche de prédiction de 26 connectivités, facilitant ainsi la compréhension de l'anatomie des voies respiratoires par le modèle. La deuxième étape consiste à prédire la connectivité à l'aide d'une approche basée sur les CNN en deux étapes. Dans la première étape, un réseau Deep-yet-Narrow (DNN) et un Shallow-yet-Wide Network (SWN) sont respectivement utilisés pour apprendre des caractéristiques avec des connaissances contextuelles à grande et petite échelles. Ces deux caractéristiques sont fusionnées dans la deuxième étape pour prédire la probabilité que chaque voxel soit une voie aérienne et sa relation de connectivité entre voisins. Nous avons entraîné notre modèle sur 50 CT images provenant d'ensembles de données publics et l'avons testé sur 20 autres CT images. Par rapport aux méthodes de segmentation des voies respiratoires les plus récentes, la robustesse et la supériorité d'AirwayNet-SE ont confirmé l'efficacité de la fusion contextuelle à grande et petite échelles. En outre, nous avons publié nos annotations manuelles des voies respiratoires de 60 CT images provenant de jeux de données publics pour une étude supervisée de la segmentation des voies respiratoires.

- Apprentissage de Réseaux Neuronaux Convolutifs Sensibles aux Tubules pour la Segmentation des Voies Respiratoires et des Artères Pulmonaires dans le CT (Chapitre 5).

  L'entraînement des réseaux de neurones convolutifs (CNN) pour la segmentation des voies respiratoires, des artères et des veines pulmonaires est difficile en raison des cibles de supervision épars causés par le déséquilibre important entre les

cibles tubulaires et le fond. Nous présentons une méthode basée sur les CNN pour la segmentation précise des voies respiratoires et des artères et veines dans la tomographie par ordinateur sans contraste. Cette méthode présente une sensibilité supérieure aux bronchioles, artérioles et veinules périphériques ténues. La méthode utilise d'abord un module de recalibrage des caractéristiques pour utiliser au mieux les caractéristiques apprises par les réseaux neuronaux. Les informations spatiales des caractéristiques sont correctement intégrées pour conserver la priorité relative des régions activées, ce qui profite au recalibrage ultérieur par canal. Ensuite, le module de distillation de l'attention est introduit pour renforcer l'apprentissage de la représentation des objets tubulaires. Les détails les plus fins des cartes d'attention à haute résolution sont transmis d'une couche à la couche précédente de manière récursive pour enrichir le contexte. Les antécédents anatomiques de la carte contextuelle des poumons et de la carte de transformation de la distance sont conçus et incorporés pour améliorer la capacité de différenciation artère-veine. Des expériences approfondies ont démontré les gains de performance considérables apportés par ces composants. Par rapport aux méthodes de pointe, notre méthode a extrait beaucoup plus de branches tout en maintenant des performances de segmentation globales compétitives.

Organisation de la thèse

Le manuscrit de la thèse est organisé comme suit:

Au chapitre 1, intitulé "**Contexte Biomédical et Arrière-plan Technique**", le contexte biomédical du caryotypage des chromosomes et de la segmentation des images de tomographie pulmonaire est présenté. En outre, pour les techniques d'apprentissage profond, l'histoire du développement des réseaux neuronaux est présentée, ainsi que des examens des architectures CNN récentes. Nous expliquons également le mécanisme de l'apprentissage contradictoire via les GAN. Enfin, trois algorithmes d'optimisation fréquemment utilisés sont décrits.

Au chapitre 2, intitulé "**Développement d'une Approche de Classification des Chromosomes à l'aide de Réseaux Convolutifs Profonds**", nous avons proposé le VarifocalNet à trois étapes pour la classification des chromosomes, qui a été évalué sur un grand jeu de données construit manuellement. Chaque étape de la méthode proposée est décrite, y compris l'apprentissage des caractéristiques à l'échelle globale et à l'échelle locale, la classification basée sur les caractéristiques fusionnées et l'affectation des types à l'aide de la stratégie de répartition.

Au chapitre 3, intitulé "**Segmentation de Nodules Pulmonaires par Synthèse**

d'échantillons CT à l'aide de Réseaux Adversariaux", nous avons proposé un cadre en deux parties basé sur les CNN pour la segmentation des nodules pulmonaires. Dans la première partie, des réseaux adversaires sont introduits pour synthétiser des échantillons de nodules. Dans la deuxième partie, le modèle de segmentation 3D basé sur le CNN est proposé en utilisant des cartes de caractéristiques hétérogènes multiples et une stratégie d'apprentissage résiduel. La méthode de segmentation des nodules proposée a été évaluée sur le jeu de données LIDC-IDRI.

Au chapitre 4, intitulé "**Développement d'une Approche Tenant Compte de la Connectivité des Voxels pour une Segmentation Précise des Voies Respiratoires à l'aide de Réseaux Neuronaux Convolutifs**", nous avons proposé l'AirwayNet et sa variante AirwayNet-SE pour la segmentation des voies respiratoires. Les deux méthodes proposées apprennent explicitement la connectivité des voxels pour percevoir la structure inhérente des voies respiratoires. L'efficacité de la méthode proposée a été validée sur des jeux de données publics et privés.

Au chapitre 5, intitulé "**Apprentissage de Réseaux Neuronaux Convolutifs Sensibles aux Tubules pour la Segmentation des Voies Respiratoires et des Artères-Veines Pulmonaires dans le CT**", nous avons proposé une méthode sensible aux tubules pour la segmentation des voies respiratoires pulmonaires et des artères-veines. Des expériences approfondies ont été menées pour corroborer sa sensibilité supérieure aux méthodes de pointe et la validité de ses composants, notamment le module de recalibrage des caractéristiques, le module de distillation de l'attention et l'incorporation de l'anatomie préalable.

Au chapitre 6, intitulé "**Conclusions Générales et Perspectives**", un bref résumé des principales contributions, des conclusions et des perspectives futures potentielles est présenté.

Chapitre 1 Contexte Biomédical et Arrière-plan Technique

Caryotypage des Chromosomes

Les anomalies chromosomiques, notamment les anomalies numériques et structurelles, sont responsables de plusieurs maladies génétiques telles que la leucémie [Natarajan, 2002]. Les anomalies numériques résultent du gain ou de la perte d'un chromosome entier, ce qui constitue une grande proportion des anomalies [Theisen and Shaffer, 2010]. Les anomalies structurelles résultent de la perte, de la rupture et de la réunion de segments de chromosomes. Dans la pratique clinique, une procédure importante pour le diagnostic chromosomique est le caryotypage, qui est effectué sur des images microscopiques d'une seule cellule [Piper, 1990]. Le processus de caryotypage est réalisé par des cytogénéticiens cliniques expérimentés et comprend principalement trois étapes: 1) coloration et imagerie des chromosomes; 2) séparation et classification manuelles sur les images de chromosomes; 3) dénombrement et diagnostic des anomalies. Une image microscopique typique de chromosomes colorés au Giemsa et le caryogramme correspondant sont présentés dans la Fig. 1.

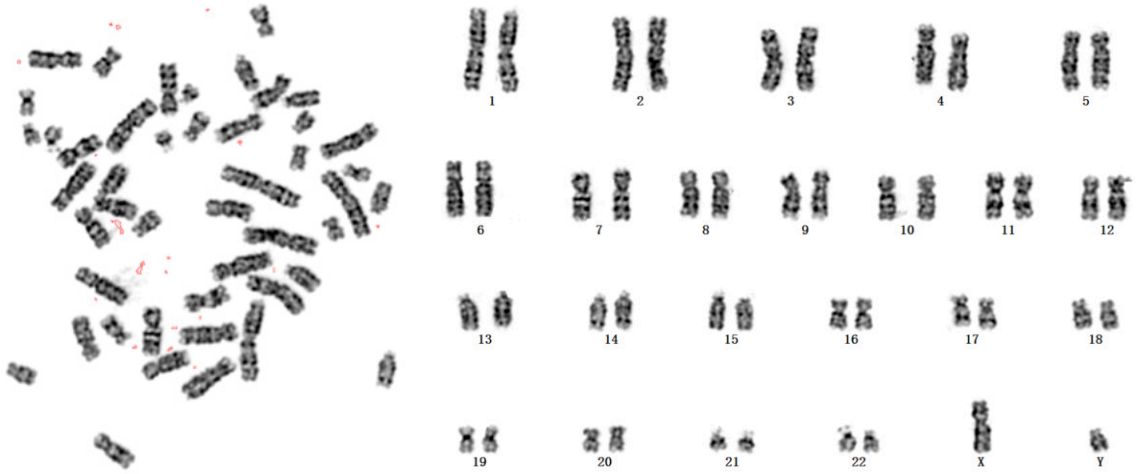


Figure 1: (a) Image microscopique des chromosomes mâles colorée au Giemsa pour un cas. (b) Le résultat du caryotype (alias caryogramme) de (a) est formé des chromosomes appariés et ordonnés (22 paires d'autosomes et 1 paire de chromosomes sexuels XY).

La première étape du caryotypage consiste à utiliser des techniques de coloration sur chaque cellule pour obtenir des chromosomes en métaphase colorés. Le caryotypage peut être classé en deux catégories principales selon la technique de coloration utilisée et le mécanisme d'imagerie : Le caryotypage Giemsa utilisant la coloration Giemsa et le

caryotypage fluorescent utilisant la coloration fluorescente (par exemple, SKY [Schröck et al., 1996] et M-FISH [Speicher et al., 1996]). Dans les applications cliniques, on préfère la coloration de Giemsa et le caryotypage plutôt que la coloration et le caryotypage fluorescents. Bien que le caryotypage fluorescent permette aux opérateurs de distinguer facilement les chromosomes par leur couleur, ses limites inhérentes (par exemple, la difficulté de détecter toutes les anomalies chromosomiques, la conservation impermanente des cibles de fluorescence, le coût prohibitif, la fiabilité controversée de l'hybridation des sondes, et la non-disponibilité de diverses sondes et d'échantillons cliniques) le rendent inapproprié comme outil de dépistage de premier niveau pour les examens [Lee et al., 2001,Huber et al., 2018,Gozzetti and Le Beau, 2000]. En revanche, le caryotype de Giemsa peut détecter presque toutes les anomalies avec un seul test peu coûteux.

La deuxième étape consiste à extraire et à classer manuellement chaque chromosome des clusters. Ces chromosomes classés sont ensuite triés et organisés en 22 paires d'autosomes et 1 paire de chromosomes sexuels (XX ou XY) dans la carte de caryotypage (aussi appelée caryogramme). Au cours de ce processus, une attention particulière est accordée à la longueur, à la position des centromères, au motif des bandes et à la courbure des contours des chromosomes. Par conséquent, le processus de caryotypage exige des efforts méticuleux de la part d'opérateurs bien formés. Afin de réduire la charge du caryotypage, de nombreuses méthodes de segmentation et de classification automatisées ont été développées pour analyser les chromosomes en métaphase [Ji, 1994,Minaee et al., 2014, Saleh et al., 2019, Cao et al., 2020, Lerner et al., 1995, Ming and Tian, 2010, Markou et al., 2012,Madian and Jayanthi, 2014,Biyani et al., 2005,Abid and Hamami, 2018,Sharma et al., 2017, Gupta et al., 2017, Wu et al., 2018b].

Enfin, les experts analysent le caryogramme afin de diagnostiquer d'éventuelles anomalies numériques et structurelles. Habituellement, le dénombrement des chromosomes est effectué sur au moins 20 caryogrammes par patient. Si une anomalie est détectée (par exemple, un mosaïcisme chromosomique) sur un caryogramme, 50 à 100 images microscopiques supplémentaires du même patient sont nécessaires pour confirmer le diagnostic. Étant donné que chaque cellule humaine contient normalement 46 chromosomes, l'ensemble du processus de diagnostic prend beaucoup de temps. Même un cytogénéticien sophistiqué doit consacrer 15 minutes ou plus à l'énumération des chromosomes pour un seul patient.

Segmentation d'images de Tomographie Pulmonaire

La segmentation des voies respiratoires est une étape clé dans l'analyse des maladies pulmonaires affectant les voies respiratoires. Elle permet de mesurer la taille, la forme et

l'épaisseur de la paroi des voies aériennes afin de quantifier le degré de rétrécissement des voies aériennes chez les patients atteints de BPCO [Wiemker et al., 2004]. En outre, il est nécessaire d'extraire des images CT des modèles de voies aériennes spécifiques au patient pour la navigation bronchoscopique [Mori et al., 2000].

Dans la tomographie thoracique, l'intensité de la lumière des voies respiratoires est généralement inférieure à celle de la paroi des voies respiratoires. Les méthodes basées sur la croissance des régions sont largement utilisées pour extraire la lumière des voies aériennes. Les méthodes de croissance de région basées sur un seuil [Kuhnigk et al., 2005, Zhou et al., 2006, Lassen et al., 2010, Ukil and Reinhardt, 2008] produisent des résultats satisfaisants pour l'extraction de la trachée et des bronches principales. Des règles heuristiques ont été étudiées pour prévenir les fuites, qui sont développées en fonction des caractéristiques géométriques des voies aériennes [Kiraly et al., 2002, Schlathoelter et al., 2002, van Ginneken et al., 2008, Mayer et al., 2004, Kitasaka et al., 2003, Graham et al., 2010, Tschirren et al., 2005]. Des caractéristiques d'image riches autres que l'intensité de l'image CT ont été explorées pour distinguer la lumière des voies aériennes des autres structures pulmonaires [Lo and de Bruijne, 2008, Lo et al., 2010b]. Par exemple, une technique de filtrage pour l'amélioration de la tubulure des voies respiratoires a été conçue [Lassen et al., 2012] pour renforcer le bord des voies respiratoires, améliorant ainsi les résultats de l'agrandissement de la région. En outre, le classifieur AdaBoost a été développé pour distinguer plusieurs échelles de voies respiratoires [Ochs et al., 2007].

L'extraction des vaisseaux pulmonaires est une étape importante de la quantification du volume des vaisseaux et du diagnostic des maladies pulmonaires. Compte tenu de la propriété de la structure tubulaire et de la forte intensité tomographique des vaisseaux, de nombreuses caractéristiques ont été conçues manuellement pour la segmentation des vaisseaux pulmonaires. Les méthodes existantes peuvent être généralement résumées en quatre catégories : seuillage [Fetita et al., 2009, Lassen et al., 2012], filtrage basé sur le Hessian [Frangi et al., 1998, Krissian et al., 2000, Aylward and Bullitt, 2002, Agam et al., 2005, Zhou et al., 2007], la croissance des régions [Metz et al., 2007, Bulow et al., 2004, Shikata et al., 2009, Zhou et al., 2012], et les méthodes basées sur l'apprentissage [Ochs et al., 2007, Korfiatis et al., 2011].

La segmentation des nodules pulmonaires est nécessaire pour le diagnostic assisté par ordinateur des nodules pulmonaires malins. La segmentation des nodules a toujours été une tâche difficile pour les raisons suivantes : 1) Le contraste entre le bord du nodule et le fond est souvent faible, en particulier pour les nodules non solides et partiellement solides. Le bruit, les artefacts et les différences d'équipement dégradent également la qualité de l'acquisition CT. 2) L'apparence des nodules pulmonaires varie beaucoup d'une personne à l'autre. La distribution des nodules est déséquilibrée en termes de taille,

de forme et d'intensité. 3) Les structures pulmonaires sont compliquées. La similitude d'intensité entre certaines structures adjacentes (par exemple, la plèvre, les vaisseaux, la paroi des voies respiratoires) et les nodules augmente la difficulté de délimiter avec précision les frontières des nodules. Pour relever ces défis, plusieurs méthodes de segmentation des nodules ont été proposées et peuvent être classées principalement en cinq catégories : seuillage [Reeves et al., 2006, Ye et al., 2009], croissance de région [Dehmeshki et al., 2008, Kubota et al., 2011, Gu et al., 2013], méthodes basées sur la morphologie [Kostis et al., 2003, Kuhnigk et al., 2006, Diciotti et al., 2011, Setio et al., 2015], modèles de contour actif [Awad et al., 2012, Farag et al., 2013, Farhangi et al., 2017, Alilou et al., 2017], et méthodes basées sur l'apprentissage [Ciompi et al., 2017, Wang et al., 2017, Wu et al., 2018a].

Méthodes d'apprentissage Profond de Pointe

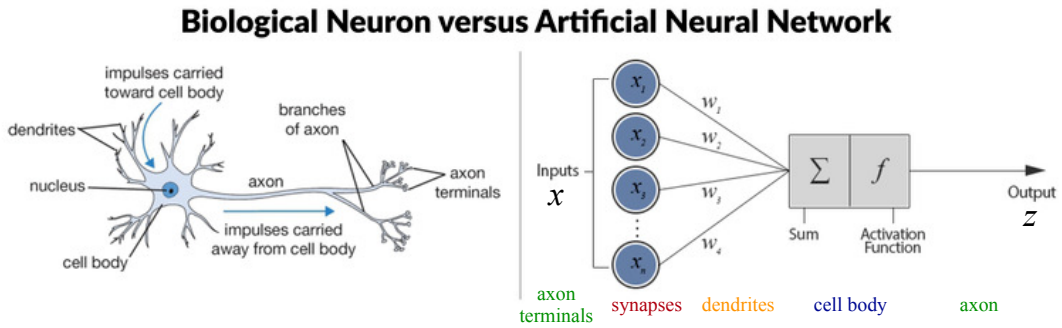


Figure 2: Le modèle de neurones artificiels (Figure inspirée de [Willems, 2019]).

En 1943, McCulloch et Pitts [McCulloch and Pitts, 1943] ont conçu le premier modèle de neurone artificiel en simulant la structure des neurones du cerveau (voir Fig. 2). Compte tenu d'un signal d'entrée x, la sortie finale z d'un neurone est exprimée comme suit :

$$z = f(\sum_i w_i x_i + b) \tag{1}$$

où x_i représente la i-ième entrée du neurone, w_i représente le poids attribué par le neurone à x_i, et b est le biais du neurone, f est la fonction d'activation du neurone. Étant donné que la fonction échelon est utilisée comme fonction d'activation à ce moment-là, il est difficile d'entraîner efficacement le réseau multicouche qui est construit sur la base de tels modèles de neurones artificiels. Par conséquent, les recherches ultérieures sur les réseaux de neurones artificiels ont stagné pendant un bon moment.

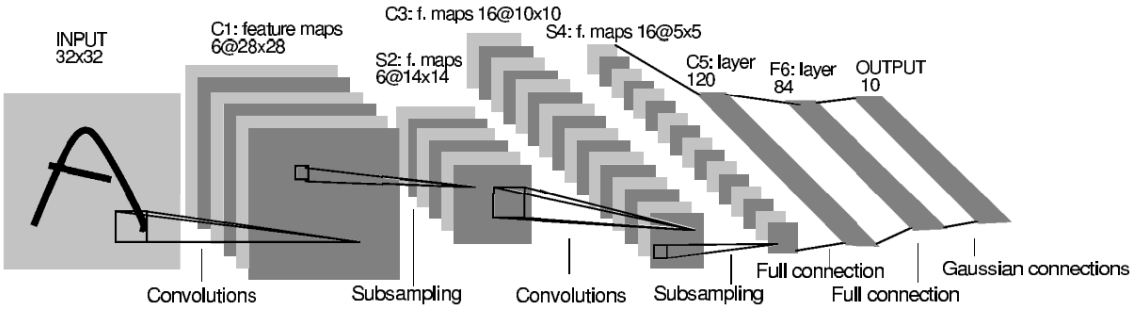


Figure 3: Architecture du modèle LeNet-5 [LeCun et al., 1998].

En 1998, LeCun et al. [LeCun et al., 1998] ont proposé le modèle LeNet-5 et l'ont appliqué avec succès à la reconnaissance de chiffres manuscrits. Ce modèle est composé de plusieurs types de couches de réseau (voir Fig. 3), notamment la couche convolutive, la couche de sous-échantillonnage (aussi appelée couche de mise en commun) et la couche entièrement connectée. Il s'agit de l'un des réseaux neuronaux convolutifs les plus représentatifs au stade initial. Sa caractéristique est d'utiliser 2 couches convolutionnelles et 3 couches entièrement connectées comme principales unités d'apprentissage du réseau. En outre, il applique avec succès les caractéristiques de partage et de réutilisation des poids spatiaux des opérations de convolution et de sous-échantillonnage afin de réduire la complexité de calcul globale du réseau.

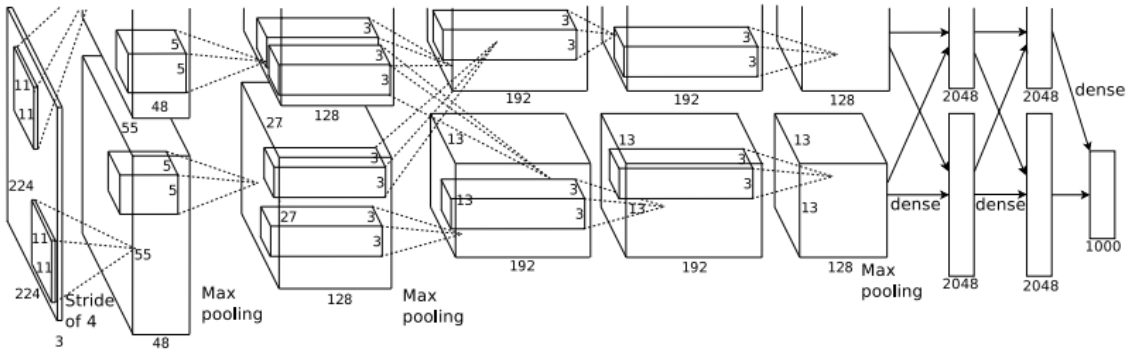


Figure 4: Architecture du modèle AlexNet [Krizhevsky et al., 2012].

En 2012, Krizhevsky et al. [Krizhevsky et al., 2012] ont développé l'AlexNet (voir Fig. 4), qui a largement dépassé les méthodes traditionnelles d'apprentissage automatique lors du défi de classification d'images ImageNet [Deng et al., 2009]. Il a officiellement ouvert le prélude au développement florissant des méthodes d'apprentissage profond dans le domaine de la vision par ordinateur. Comparé au modèle LeNet-5, AlexNet présente les caractéristiques exceptionnelles suivantes:

- Il possède des couches de réseau beaucoup plus profondes. Un noyau convolutif plus large (11×11) est utilisé dans les couches de convolution peu profondes pour apprendre les caractéristiques de bas niveau des images naturelles.
- L'unité linéaire rectifiée (ReLU) est utilisée comme fonction d'activation pour remplacer la fonction $tanh(\cdot)$ dans LeNet-5, ce qui améliore l'efficacité du calcul pendant l'apprentissage et atténue le problème du gradient évanescent dans les réseaux profonds.
- Il utilise une couche de normalisation de la réponse locale (LRN) pour établir un mécanisme de compétition de la réponse d'activation entre les neurones locaux et améliorer la capacité de généralisation du modèle.
- Il utilise le dropout [Srivastava et al., 2014] pour éviter le surajustement du modèle et obtenir un effet similaire à l'apprentissage d'ensemble.

Le développement de l'architecture réseau joue également un rôle important dans l'application des réseaux neuronaux convolutifs à diverses tâches de vision par ordinateur. Outre l'architecture à lien unique, les chercheurs ont mis au point plusieurs architectures de réseau importantes, telles que l'architecture de transmission en dérivation, l'architecture parallèle et l'architecture de branchements multitâches (voir Fig. 5).

Résumé

Nous présentons le contexte biomédical et les techniques d'apprentissage profond liées au sujet de la thèse :

- En ce qui concerne le contexte biomédical, l'objectif et le flux de travail du caryotypage sont bien expliqués. En outre, des méthodes de segmentation des voies respiratoires, des vaisseaux et des nodules pulmonaires sont présentées pour l'analyse des images CT.
- Pour les techniques d'apprentissage profond, l'histoire du développement des réseaux neuronaux est présentée. Ensuite, des études sur les architectures récentes des CNN sont présentées.

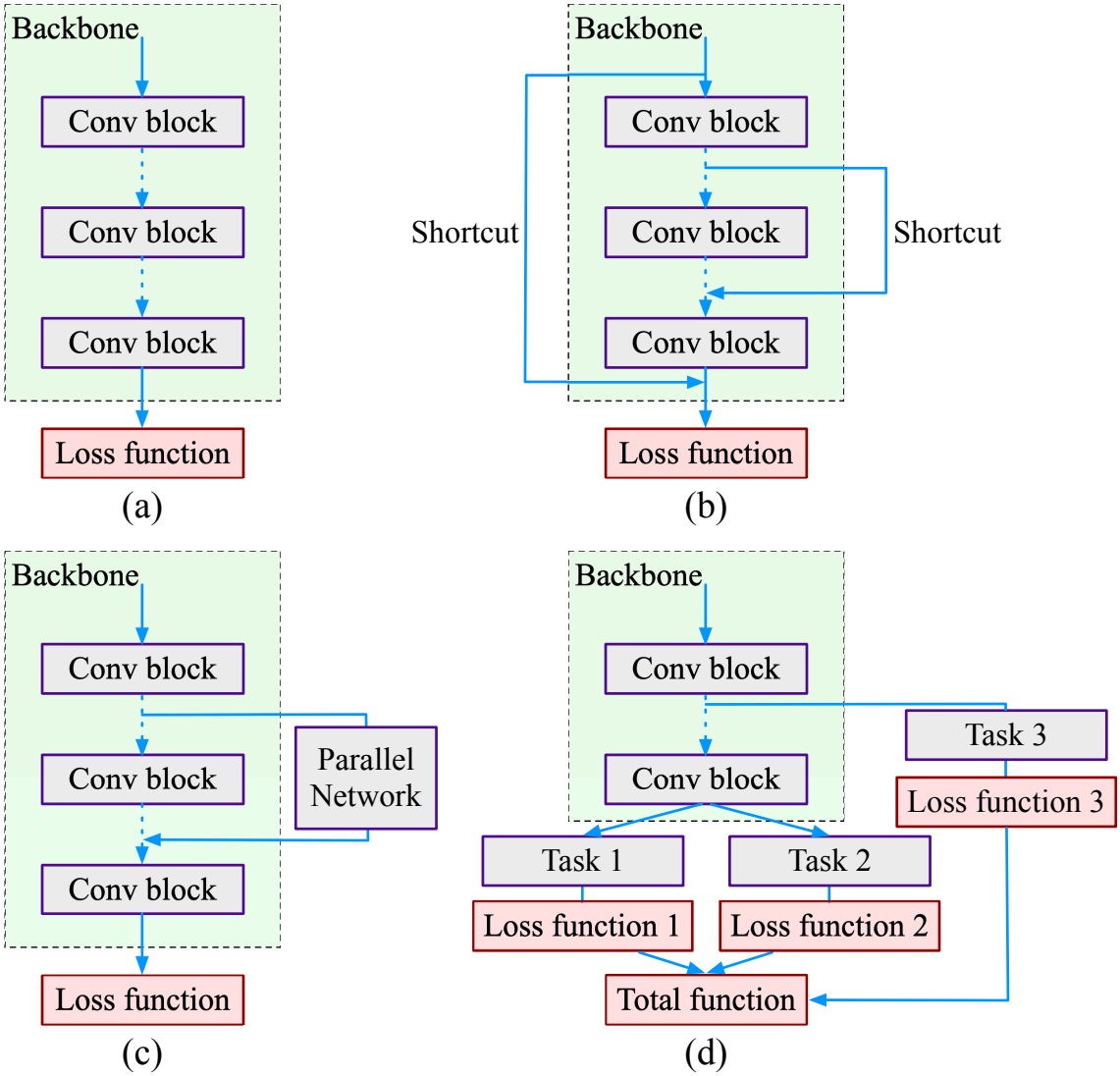


Figure 5: Différentes architectures CNN. (a) Architecture à chemin de liaison unique (b) Architecture à transmission par contournement (c) Architecture parallèle (d) Architecture à branchement multi-tâches

Chapitre 2 Développement d'une Approche de Classification des Chromosomes à l'aide de Réseaux Convolutifs Profonds

Introduction

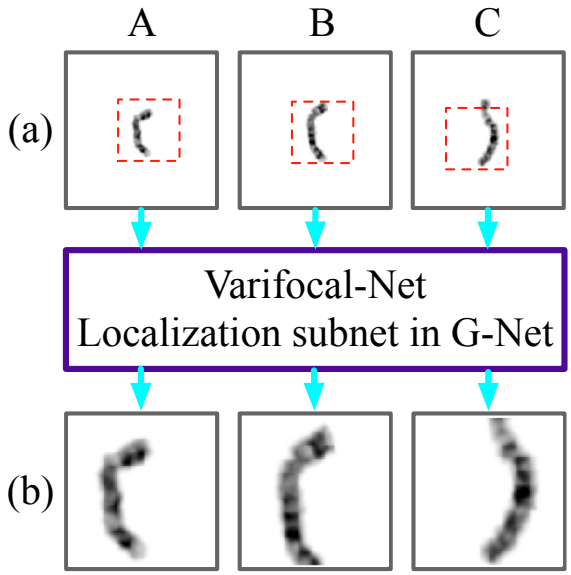


Figure 6: La focalisation varie de globale à locale. Étant donné les images de chromosomes (A, B, C), le sous-réseau de localisation détecte leurs régions les plus fines pour les recadrer et les agrandir. (a) Les images originales de chromosomes. (b) Les parties locales après un zoom avant.

Nous proposons une nouvelle approche basée sur les CNN pour la classification des chromosomes. Son nom, Varifocal-Net, souligne sa capacité à zoomer automatiquement sur des régions locales. Il est composé d'un réseau d'échelle globale (G-Net) et d'un réseau d'échelle locale (L-Net). Nous extrayons des caractéristiques globales et localisons des régions locales spécifiques via le G-Net. La vue est modifiée (voir Fig. 6) lorsque notre Varifocal-Net zoome sur la région discriminante d'un chromosome. Les caractéristiques locales sont extraites de ces parties locales via le L-Net. À première vue, cette idée de global-to-local ressemble au concept des CNN multi-échelles utilisés dans l'analyse d'images cellulaires [Godinez et al., 2017, Buyssens et al., 2012, Godinez et al., 2018, Pan et al., 2018] et d'autres tâches de vision [Shen et al., 2015b, Zeng et al., 2017, Lotter et al., 2017]. Cependant, contrairement aux méthodes multi-échelles précédentes, notre approche apprend les informations multi-échelles dans le mécanisme global-local. Elle localise la région locale discriminante et extrait les caractéristiques des deux échelles par le biais de deux réseaux indépendants. Le réseau varifocal proposé comprend trois étapes.

La première étape consiste à apprendre des représentations efficaces des caractéristiques à l'échelle globale et locale. Les représentations à l'échelle globale concernent principalement des informations générales telles que la longueur, la forme et la taille du chromosome, qui déterminent son type à un niveau de détail grossier. Les représentations à l'échelle locale décrivent des détails tels que les motifs de texture des parties locales, qui facilitent la discrimination entre les chromosomes à un niveau plus fin. La deuxième étape consiste à construire deux classifieurs MLP pour exploiter les caractéristiques des deux échelles afin de prédire le type et la polarité, respectivement. La troisième étape consiste à introduire une stratégie de répartition pour l'attribution du type dans chaque cas de patient. Pour valider l'efficacité et la généralisation de notre approche, nous construisons un grand ensemble de données contenant 1909 cas de caryotypage. Des expériences approfondies sur ce jeu de données corroborent le fait que le Varifocal-Net atteint de meilleures performances que les méthodes de pointe. Nos contributions peuvent être résumées comme suit:

- Inspirés par la capacité de zoom des appareils photo, nous proposons le réseau varifocal pour relever les défis de la classification des chromosomes. Nous extrayons les caractéristiques d'échelle globale de l'image entière et les caractéristiques d'échelle locale de la région locale sélectionnée par notre mécanisme varifocal. Des stratégies d'apprentissage résiduel et d'apprentissage multi-tâches sont utilisées pour promouvoir un apprentissage efficace des caractéristiques. La détection des parties locales discriminantes s'effectue par le biais d'un sous-réseau de localisation dont la formation implique un apprentissage supervisé et faiblement supervisé.

- Nous utilisons les caractéristiques concaténées des échelles globale et locale pour prédire le type et la polarité simultanément, combinant ainsi les connaissances acquises à deux échelles. À notre connaissance, il s'agit de la première tentative de prise en compte d'un ensemble de caractéristiques multi-échelles dans l'étude des chromosomes.

- Nous proposons une stratégie de répartition pour affecter chaque chromosome à un type en fonction de ses probabilités prédites. Le critère de vraisemblance maximale et les situations d'anomalies possibles sont pris en compte pour permettre à la stratégie d'être adaptée aux contextes cliniques.

- Nous évaluons l'approche proposée sur un grand ensemble de données. Elle démontre ses performances supérieures à celles des méthodes de pointe. La méthode de classification de bout en bout permet d'éviter le problème de l'extraction imprécise de l'axe médian et du redressement des chromosomes.

- Le Varifocal-Net a été mis en pratique clinique pour la classification des chromosomes. Pour chaque patient, il classe avec précision les chromosomes anormaux et sains et diagnostique les anomalies numériques si le nombre de chromosomes classés est irrégulier.

Méthodologie

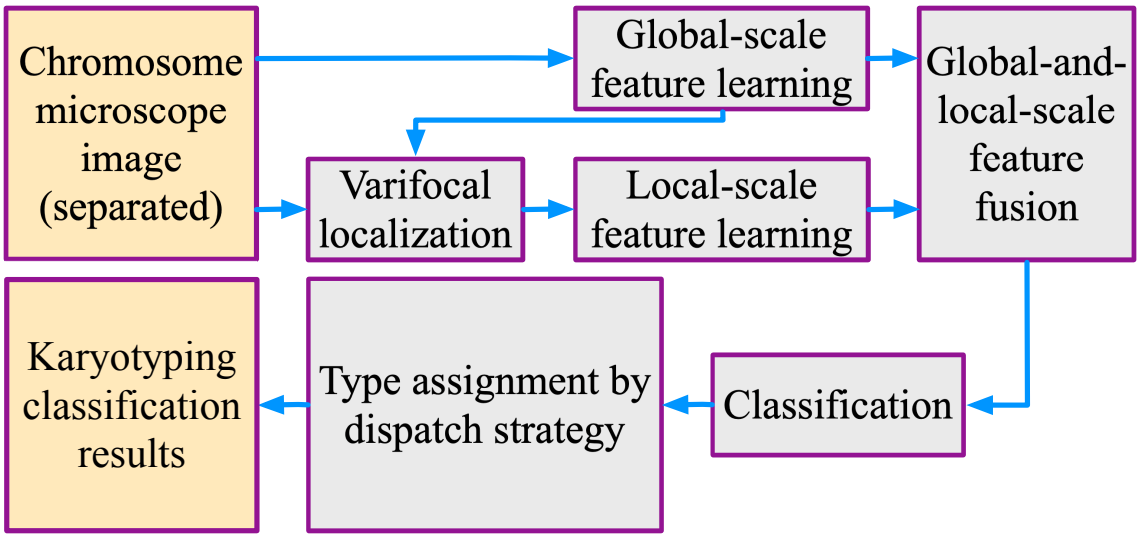


Figure 7: Organigramme du réseau Varifocal proposé pour la classification des chromosomes.

L'organigramme du Varifocal-Net proposé est représenté sur la Fig. 7. Il se compose de trois étapes : a) Apprentissage des caractéristiques à l'échelle globale et locale en optimisant le réseau Varifocal de manière alternative ; b) Classification du type et de la polarité via des classifieurs MLP utilisant les caractéristiques fusionnées ; c) Affectation des types de chromosomes avec la stratégie de répartition proposée. Les images originales de chromosomes sont séparées manuellement par des cytogénéticiens à partir d'images microscopiques capturées. Elles sont prétraitées pour être normalisées et prises comme entrées pour le G-Net dans la première étape. Le G-Net contient des CNN profonds, un sous-réseau de classification et un sous-réseau de localisation. Les caractéristiques d'échelle globale sont extraites par les CNN, qui sont optimisés par la fonction de perte du sous-réseau de classification. Une fois que les CNN et le sous-réseau de classification ont convergé, nous pré-entraînons le sous-réseau de localisation afin de produire des coordonnées initiales pour la détection des régions locales. Ensuite, avec les parties locales recadrées et redimensionnées, nous optimisons alternativement le L-Net et le sous-réseau de localisation du G-Net. Dans la deuxième étape, avec les caractéristiques fusionnées à

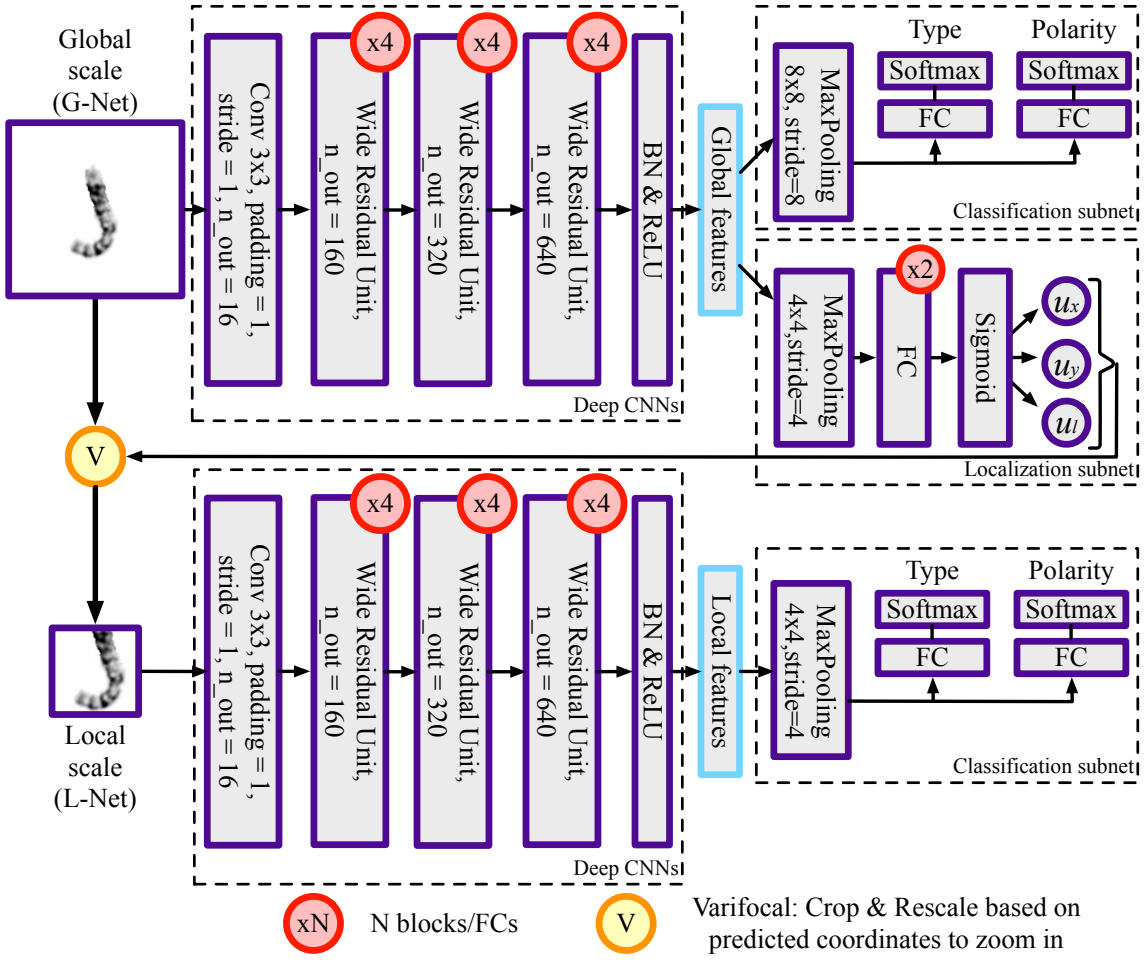


Figure 8: Première étape du réseau varifocal proposé: extraction de caractéristiques à l'échelle globale et locale via le G-Net et le L-Net, respectivement.

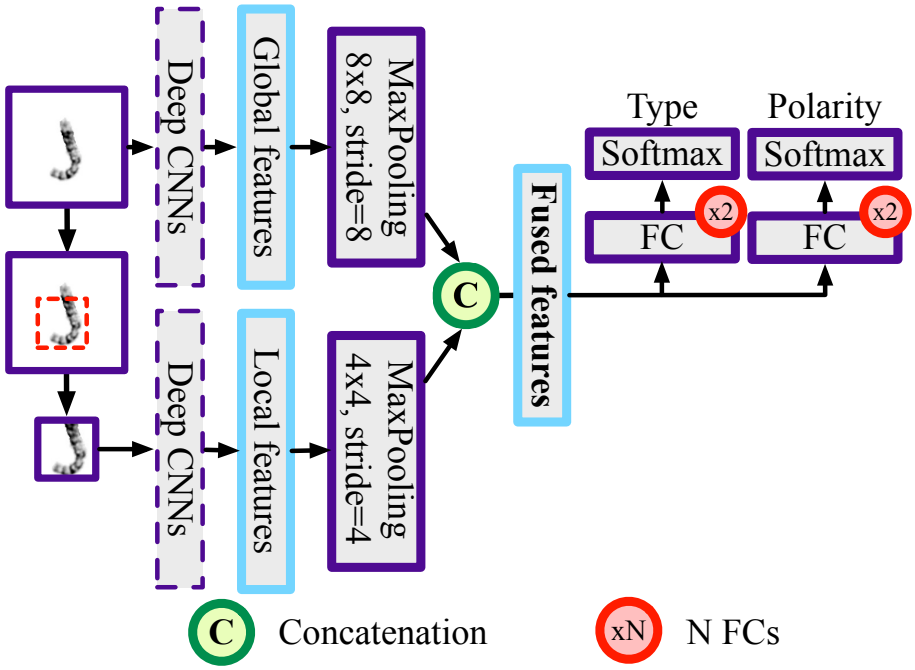


Figure 9: La deuxième étape du Varifocal-Net proposé : la classification des chromosomes en utilisant des caractéristiques fusionnées à la fois à l'échelle globale et locale.

deux échelles, nous construisons deux classifieurs MLP pour prédire le type et la polarité des chromosomes, respectivement. Les représentations schématiques de la première et de la deuxième étape de notre Varifocal-Net sont illustrées dans la Fig. 8 et la Fig. 9, respectivement. Pour chaque chromosome dans un cas de patient, une stratégie de répartition est employée dans la troisième étape pour l'affecter à un certain type en fonction de ses probabilités prédites.

Conclusion

En conclusion, nous avons proposé le Varifocal-Net pour la classification des chromosomes, qui a été évalué sur un grand jeu de données construit manuellement. Il s'agit d'une méthode basée sur le CNN en trois étapes. La première étape apprend efficacement les caractéristiques globales et locales par le biais du G-Net et du L-Net, respectivement. En prenant une image de chromosome à l'échelle globale comme entrée, il détecte précisément une région locale qui est discriminante et abondante en détail pour l'extraction de caractéristiques supplémentaires. La deuxième étape différencie de manière robuste les chromosomes en différents types et polarités via deux classifieurs MLP. Elle bénéficie d'un ensemble de caractéristiques multi-échelles, avec seulement quelques erreurs de classification. Dans la troisième étape, une stratégie de répartition est employée pour

affecter chaque chromosome à un type en fonction de ses probabilités prédites. Des résultats expérimentaux détaillés démontrent que notre approche surpasse les méthodes de l'état de l'art, corroborant sa grande précision et sa généralisation.

En ce qui concerne son rôle dans le flux de travail du caryotypage clinique, le Varifocal-Net peut effectuer une classification précise en moins d'une seconde après que les opérateurs aient segmenté manuellement les chromosomes d'une cellule pour chaque patient. Les cartes de résultats de caryotypage qu'il génère automatiquement offrent la possibilité aux experts humains de vérifier et de corriger les éventuelles erreurs de classification. En outre, les avertissements concernant d'éventuelles anomalies numériques permettent aux opérateurs d'accorder une attention particulière au diagnostic ultérieur. L'utilisation pratique du Varifocal-Net à l'hôpital Xiangya de l'université Central South suggère son potentiel prometteur pour alléger la charge de travail des médecins dans le processus de diagnostic.

Chapitre 3 Segmentation de Nodules Pulmonaires par Synthèse d'échantillons de CT à l'aide de Réseaux Adversariaux

Introduction

Nous proposons un cadre basé sur le CNN pour la segmentation des nodules pulmonaires. En adoptant des réseaux adversaires, des échantillons synthétiques sont générés pour obtenir un ensemble de données d'entraînement plus équilibré. Grâce à l'intégration de cartes de caractéristiques interprétables et à l'introduction d'une stratégie d'apprentissage résiduel, le modèle de segmentation fonctionne de manière robuste sur tous les types de nodules sans intervention manuelle des radiologues. Les principales contributions sont les suivantes : (1) Nous utilisons un GAN conditionnel qui génère des images CT de nodules pour étendre le jeu de données LIDC-IDRI. L'annotation originale ne portant que sur les limites de chaque nodule, nous concevons une méthode permettant d'obtenir des étiquettes sémantiques à dix canaux pour les plaques de nodules. Ces étiquettes contiennent non seulement des informations contextuelles mais représentent également les attributs sémantiques des nodules. Sur la base des étiquettes sémantiques, des échantillons synthétiques sont générés par des réseaux adversaires. La perte d'erreur de reconstruction $L2$ est introduite dans cGAN pour augmenter le réalisme des échantillons générés. Le problème de déséquilibre des données est atténué par cette expansion de l'ensemble de données, ce qui empêche le surajustement pour l'entraînement du modèle de segmentation. Par conséquent, la performance de notre méthode de segmentation est améliorée. (2) Nous proposons un modèle CNN 3D qui segmente avec précision les nodules pulmonaires. Pour générer des masques de segmentation, un réseau similaire à U-Net 3D est exploité. De multiples cartes hétérogènes, y compris des cartes d'arêtes et des cartes de caractéristiques de texture, sont introduites comme entrées et exploitées par le modèle CNN pour apprendre des caractéristiques de haut niveau. Pour les cartes de contours, nous appliquons l'opérateur de Canny [Canny, 1986] et l'opérateur de Sobel [Sobel, 1990] pour détecter les contours des images de nodules, ce qui constitue une base pour la tâche de segmentation. Les motifs binaires locaux (LBP) [Ojala et al., 2002] sont choisis pour capturer la structure spatiale des textures des nodules. Puisqu'il existe une grande différence de textures entre les nodules solides, partiellement solides et GGO, ces cartes de caractéristiques de texture sont considérées comme informatives pour le réseau afin de générer des résultats de segmentation précis pour chaque type de nodule. L'architecture 3D de notre modèle vise à mieux utiliser la connaissance volumétrique des images CT 3D. En outre, l'apprentissage résiduel est utilisé pour résoudre le problème du gradient évanescent. Il favorise un apprentissage efficace des caractéristiques et accélère

le processus de formation. (3) Le cadre de segmentation basé sur le CNN proposé est évalué sur le jeu de données public LIDC-IDRI.

Méthodologie

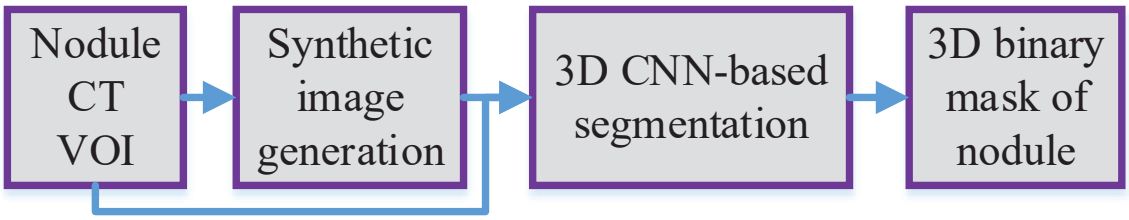


Figure 10: Vue d'ensemble du cadre proposé pour la segmentation des nodules pulmonaires. Des images synthétiques de nodules sont d'abord générées. Ensuite, les images originales et synthétiques sont utilisées pour entraîner le modèle de segmentation. Les résultats de la segmentation sont des masques binaires 3D de la VOI du nodule.

Le cadre de segmentation des nodules pulmonaires développé est composé de deux parties (voir Fig. 10) : (1) la génération d'images synthétiques et (2) la segmentation 3D basée sur CNN. Pour la première partie, des réseaux adversaires sont adoptés pour améliorer la diversité des échantillons de nodules et atténuer le problème des données déséquilibrées et limitées. La deuxième partie est conçue pour segmenter tous les types de nodules de la VOI en utilisant un modèle CNN 3D.

Dans le domaine de la segmentation d'images médicales, il est souvent inévitable que les échantillons collectés soient déséquilibrés et biaisés, ce qui pose des problèmes pour la généralisation des méthodes de segmentation. En particulier pour la segmentation des nodules pulmonaires, même le plus grand jeu de données public LIDC-IDRI [Armato et al., 2011] est déséquilibré en termes de texture et de taille des nodules. Le nombre de nodules solides est trois fois supérieur à celui des autres. Les gros nodules constituent une grande proportion de l'ensemble des nodules. Par ailleurs, les nodules GGO et les petits nodules sont si limités en quantité qu'ils sont facilement submergés par les autres nodules. Par conséquent, le modèle de segmentation peut souffrir de mauvaises performances sur les catégories minoritaires de nodules s'il est entraîné sur un tel ensemble de données. Pour résoudre ce problème, la génération d'images synthétiques apparaît alors comme une solution intéressante. Pour ce faire, les coupes qui contiennent des nodules sont d'abord sélectionnées parmi tous les cubes VOI recadrés. Une technique de transformation des étiquettes de vérité de terrain en étiquettes sémantiques à dix canaux est ensuite conçue pour introduire des informations contextuelles abondantes sur les nodules. Enfin, un modèle génératif conditionnel est utilisé pour traduire les étiquettes sé-

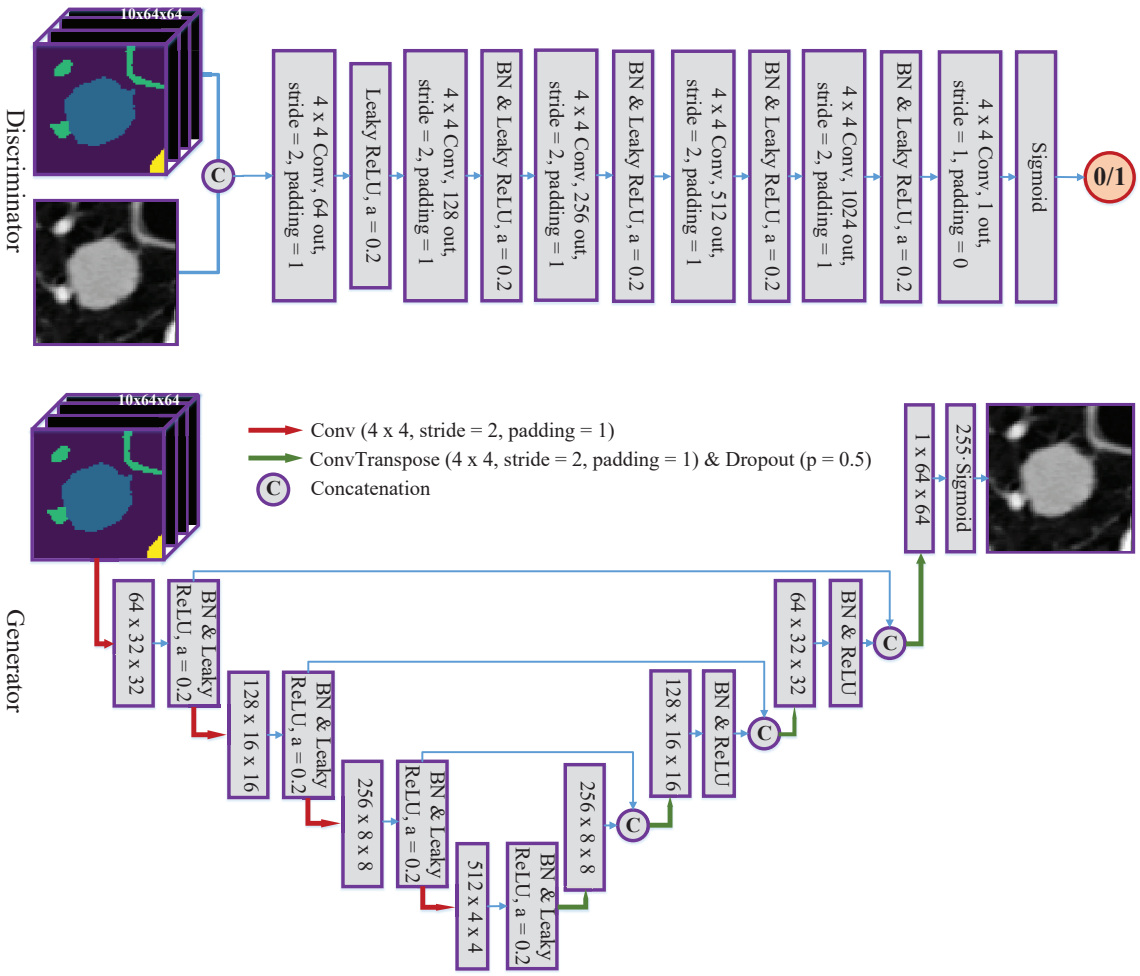


Figure 11: L'architecture réseau du cGAN proposé.

mantiques en images réalistes.

L'architecture de notre cGAN est décrite dans la Fig. 11. La structure U-Net 2D est utilisée comme colonne vertébrale pour construire un modèle génératif, qui génère des images synthétiques selon un mode codeur-décodeur avec des chemins de saut. Pour le chemin de contraction, au lieu de la couche de max-pooling utilisée dans le U-Net original, une couche de convolution en strides est adoptée pour sous-échantillonner l'image, suivie d'une couche de normalisation par lots (BN) et d'une couche d'unité linéaire rectifiée (ReLU) fuyante. Pour le chemin expansif, nous utilisons une convolution transposée pour ré-échantillonner les cartes de caractéristiques afin d'augmenter la résolution et les concaténer avec les caractéristiques du chemin de saut. La couche BN, la couche ReLU et la couche d'exclusion sont également présentées. Ensuite, un réseau entièrement convolutif (FCN) est conçu comme modèle discriminant. À l'exception de la première couche, toutes les couches de convolution stridentes sont suivies d'une couche BN et d'une couche ReLU fuyante. La couche de mise en commun dans le modèle générateur et le modèle discriminateur est remplacée par la convolution stridée parce que cette dernière apprend à résumer les pixels dans son noyau par une multiplication pondérée par éléments. Contrairement au max-pooling ou à l'avg-pooling, la façon dont la convolution stridée réduit la dimensionnalité des caractéristiques n'est pas déterminée à l'avance mais peut être apprise pendant l'apprentissage.

Comme le montre la Fig. 11, le bruit z est implicitement pris comme entrée du générateur. Nous utilisons une couche d'exclusion sur le chemin expansif pour introduire le bruit [Isola et al., 2017] en désactivant aléatoirement les neurones avec une probabilité de 0,5. Une étude précédente sur la couche d'exclusion [Park and Kwak, 2016] prouve que cette couche ajoute du bruit aux caractéristiques de sortie et améliore ainsi la robustesse à la variation des images d'entrée. En outre, la couche d'exclusion fournit une régularisation pour empêcher l'ajustement excessif en réduisant la co-dépendance entre les neurones. Elle désactive aléatoirement les neurones pendant le processus d'apprentissage, empêchant ainsi le modèle d'apprendre un ensemble interdépendant de poids de caractéristiques [Goodfellow et al., 2016].

L'architecture globale de segmentation des nodules est présentée dans la Fig. 12. Comme les nodules pulmonaires ont des textures internes différentes et que la méthode de segmentation doit s'adapter à cette variété, nous introduisons des cartes de texture pour donner implicitement au réseau la capacité d'appréhender si le nodule actuel est GGO, partiellement solide ou solide. En outre, les cartes d'arêtes sont concaténées comme entrées car elles fournissent des connaissances riches sur les marges et les limites des images de nodules, facilitant ainsi la tâche de segmentation. Le modèle de segmentation 3D CNN est un modèle de bout en bout qui exploite une structure similaire à un

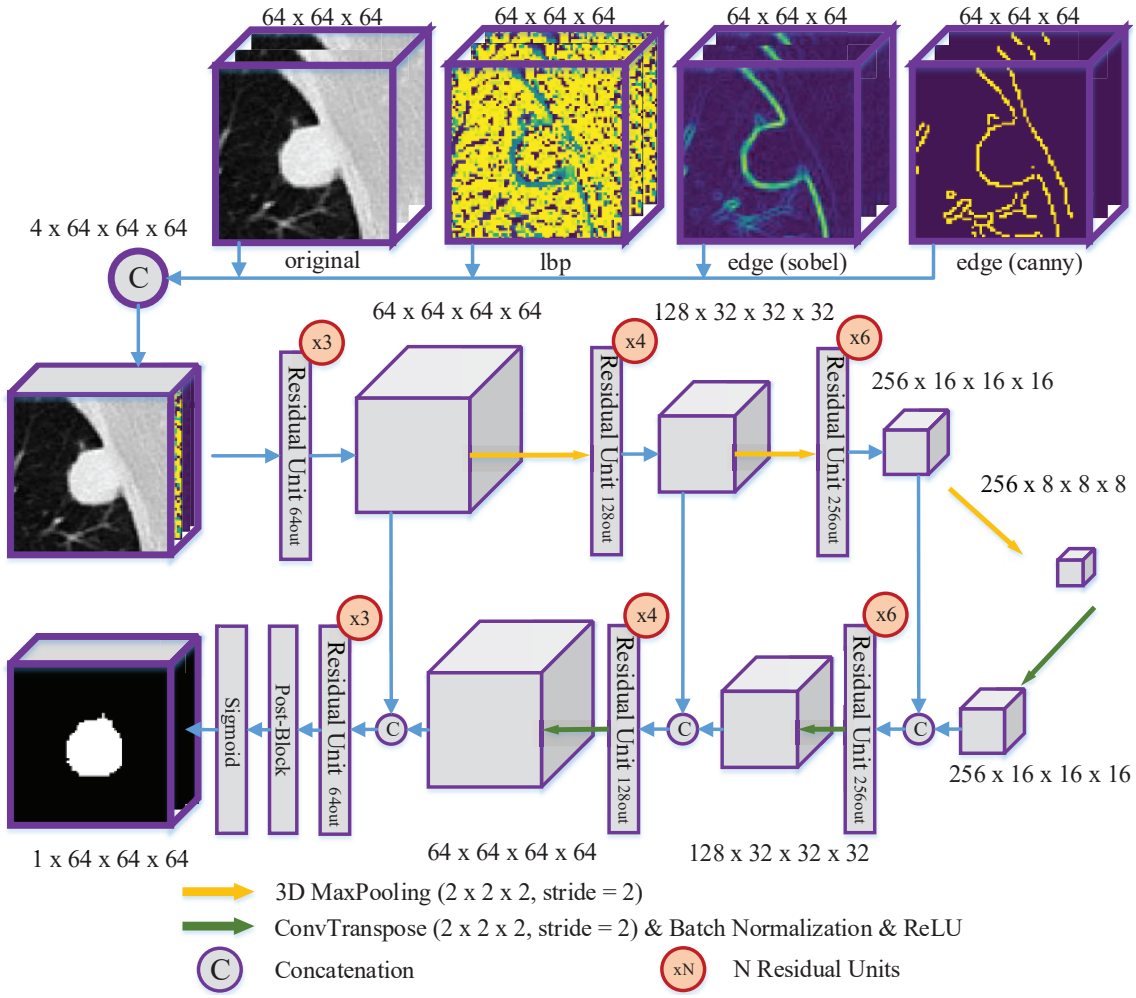


Figure 12: L'architecture réseau du cadre de segmentation proposé.

U-Net 3D [Çiçek et al., 2016]. L'apprentissage résiduel est introduit dans le réseau pour améliorer les performances de la segmentation.

Conclusion

Nous avons proposé un cadre en deux parties basé sur les réseaux CNN pour la segmentation des nodules pulmonaires. Dans la première partie, des réseaux adversaires sont utilisés pour synthétiser des échantillons de nodules. L'objectif est de créer un ensemble de données plus diversifié et plus équilibré pour la formation ultérieure du modèle. Les étiquettes sémantiques, ainsi que les étiquettes de notation de neuf attributs, sont exploitées pour fournir des connaissances sémantiques et contextuelles. La perte d'erreur de reconstruction est introduite pour améliorer le réalisme. Cette méthode d'extension du jeu de données présente plusieurs avantages. Les limites et les attributs sémantiques des nodules sont préservés pendant le processus de génération. De plus, le bruit aléatoire produit par la couche d'exclusion permet de faire varier l'environnement spatial et donc d'augmenter la diversité des images. Dans la deuxième partie, des cartes de caractéristiques multiples sont incorporées comme entrées dans le modèle CNN 3D. Grâce à la stratégie d'apprentissage résiduel, le modèle de segmentation entraîné sur le jeu de données étendu bénéficie d'un haut niveau de généralité. Les résultats obtenus sur le jeu de données LIDC-IDRI montrent que notre modèle CNN 3D permet une segmentation plus précise des nodules par rapport aux méthodes de l'état de l'art existantes, ce qui suggère sa valeur potentielle pour les applications cliniques.

Chapitre 4 Développement d'une Approche Tenant Compte de la Connectivité des Voxels pour une Segmentation Précise des Voies Respiratoires à l'aide de Réseaux Neuronaux Convolutifs

Introduction

Nous proposons AirwayNet, une approche basée sur les CNN pour une segmentation précise des voies respiratoires. Considérant que la structure arborescente des voies respiratoires est assez complexe et que la prédiction des candidats aux voies respiratoires est sujette à la discontinuité, nous mettons l'accent sur la connectivité des voxels des voies respiratoires. Contrairement aux méthodes précédentes, nous n'entraînons pas directement le réseau à classer les voxels d'avant-plan et d'arrière-plan. Au lieu de cela, la tâche de segmentation binaire est transformée en 26 tâches consistant à prédire si un voxel est connecté à ses voisins. Puisque les voxels des voies respiratoires s'étendent de la bronche principale vers l'extrémité de la bronchiole comme une région entière connectée, nous considérons que c'est une bonne solution pour permettre au modèle d'être conscient de la connectivité des voxels. Des travaux antérieurs sur la segmentation saillante [Kampffmeyer et al., 2018] ont démontré que la modélisation de la connectivité encode spontanément la relation entre deux pixels. Par conséquent, nous concevons une approche tenant compte de la connectivité des voxels pour mieux comprendre la structure inhérente des voies respiratoires.

De plus, nous allons plus loin en étendant le réseau AirwayNet à une étape au réseau AirwayNet-SE à deux étapes, une approche simple mais efficace qui incorpore deux échelles de contexte différentes pour comprendre les voies respiratoires de grande et de petite taille, respectivement. Avec la même modélisation de la connectivité 3D que lors de la première étape, les réseaux sont entraînés à prédire si un voxel est connecté à ses voisins au lieu de classer directement les voxels des voies aériennes. Le réseau AirwayNet-SE se compose d'un réseau **D**eep-yet-**N**arrow (DNN) et d'un réseau **S**hallow-yet-**W**ide (SWN). Le DNN, avec des couches plus profondes mais un plus petit nombre de canaux par couche, vise à extraire les caractéristiques des branches épaisses. Quatre opérations de mise en commun sont utilisées pour que le modèle soit conscient du contexte global de la cavité thoracique. Alors que pour le SWN, des couches moins profondes avec deux opérations de mise en commun sont adoptées pour éviter que les bronches fines ne disparaissent. Les canaux de caractéristiques du SWN sont élargis pour augmenter le pouvoir de représentation. La deuxième étape consiste à prédire la connectivité à l'aide de CNN à deux niveaux. Dans la première étape, nous entraînons respectivement nos DNN et SWN pour apprendre les caractéristiques efficaces des grandes et

petites bronches. Dans la deuxième étape, les caractéristiques du DNN et du SWN sont concaténées pour fusionner les connaissances contextuelles des deux échelles. Ces caractéristiques fusionnées sont utilisées pour la prédiction finale de la connectivité des voies respiratoires.

Nos contributions sont résumées comme suit: 1) La connectivité des voxels des voies respiratoires est modélisée à l'aide d'étiquettes binaires traditionnelles afin de mieux servir la tâche de segmentation des voies respiratoires. Le réseau AirwayNet proposé apprend automatiquement la relation entre les voxels adjacents et distingue les voies respiratoires du fond. Pour chaque voxel, le réseau prédit non seulement sa probabilité d'être une voie aérienne mais aussi sa connectivité avec ses voisins. 2) Le réseau AirwayNet-SE a proposé une solution au conflit causé par la différence entre les grandes et les petites voies respiratoires. Grâce à la modélisation de la connectivité, il a tiré parti de la fusion des connaissances contextuelles à deux échelles pour prédire si un voxel est une voie aérienne et s'il est connecté à ses voisins. 3) Nous avons publié les annotations manuelles de 60 CT images publics afin de promouvoir l'étude de la segmentation des voies respiratoires qui nécessite un apprentissage supervisé. À notre connaissance, il s'agit du plus grand ensemble de données publiques d'annotations des voies respiratoires. Les annotations sont disponibles sur `http://www.pami.sjtu.edu.cn/News/56`.

Méthodologie

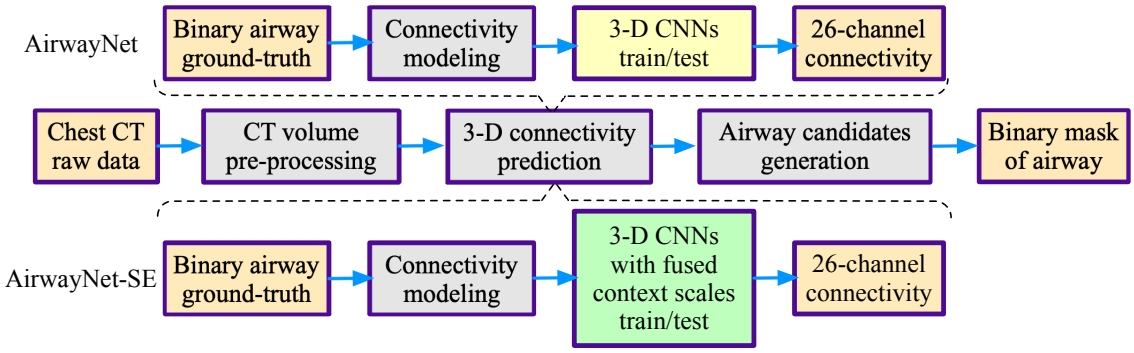


Figure 13: Organigramme de la proposition AirwayNet et AirwayNet-SE.

Dans cette section, nous présentons d'abord les détails du prétraitement de l'image CT et de la modélisation de la connectivité des voxels. Cette étape de modélisation est la condition préalable à la transformation du problème de segmentation en problème de prédiction de connectivité. Ensuite, nous décrivons la prédiction de la connectivité basée sur les CNN 3D. Ensuite, nous présentons comment étendre la prédiction de connectivité en une étape à sa contrepartie en deux étapes, où les caractéristiques des grandes et pe-

tites échelles de contexte sont fusionnées pour la prédiction de connectivité. Enfin, nous discutons du processus de génération des voies aériennes candidates. L'organigramme de l'AirwayNet et de l'AirwayNet-SE proposés est décrit dans la Fig. 13.

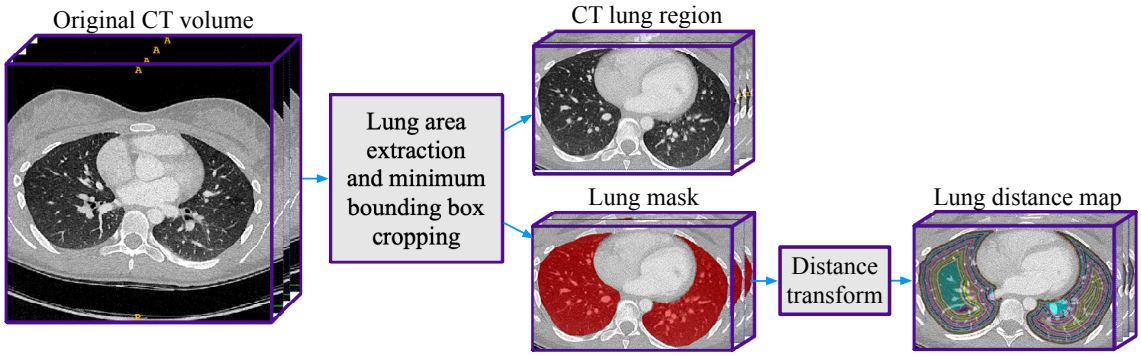


Figure 14: Illustration de l'étape de prétraitement de l'image CT.

L'un des défis de la segmentation des voies respiratoires est que les voxels de premier plan n'occupent qu'une petite proportion de tous les voxels de l'image CT. Pour éviter l'apprentissage de caractéristiques à partir de parties non pertinentes (par exemple, les côtes et la peau), nous limitons la région candidate valide des voies respiratoires à l'intérieur de la zone pulmonaire. Pour extraire le masque pulmonaire, chaque coupe de CT est d'abord filtrée avec un filtre gaussien ($\sigma = 1$) et binarisée avec un seuil (-600 unité Hounsfield). L'analyse en composantes connectées est appliquée pour éliminer les candidats peu sûrs et les deux plus grandes composantes sont choisies comme poumons gauche et droit, respectivement. Pour éviter une sous-segmentation, nous remplaçons la surface du poumon par sa coque convexe sur chaque coupe si la coque convexe a une surface supérieure de 50%. Nous effectuons également une transformation de distance euclidienne sur le masque pulmonaire pour calculer la carte de distance. Chaque voxel de la carte de distance enregistre sa distance minimale à la limite du poumon. Nous ajoutons cette carte au réseau car la position relative des voies respiratoires par rapport à la limite du poumon est considérée comme significative sur le plan anatomique. Pour préparer l'entraînement du réseau, l'intensité des voxels du CT est coupée par une fenêtre $[-1000, 600]$ (HU) et normalisée à $[0, 255]$. La Fig. 14 illustre l'étape de prétraitement du CT.

Dans un scanner tridimensionnel (3D), la 26-connectivité décrit bien la relation entre un voxel et ses 26 voisins (voir Fig. 15). Étant donné un voxel $P = (x, y, z)$ et son voisin $Q = (u, v, w)$, la distance entre P et Q est limitée par $d(P,Q) = max(|(x - u)|, |(y - v)|, |(z - w)|) \leq 1$, ce qui signifie que Q est situé dans un cube de $3 \times 3 \times 3$ centré sur P. Nous indexons les voisins Q de 1 à 26 et désignons chaque paire de voxels

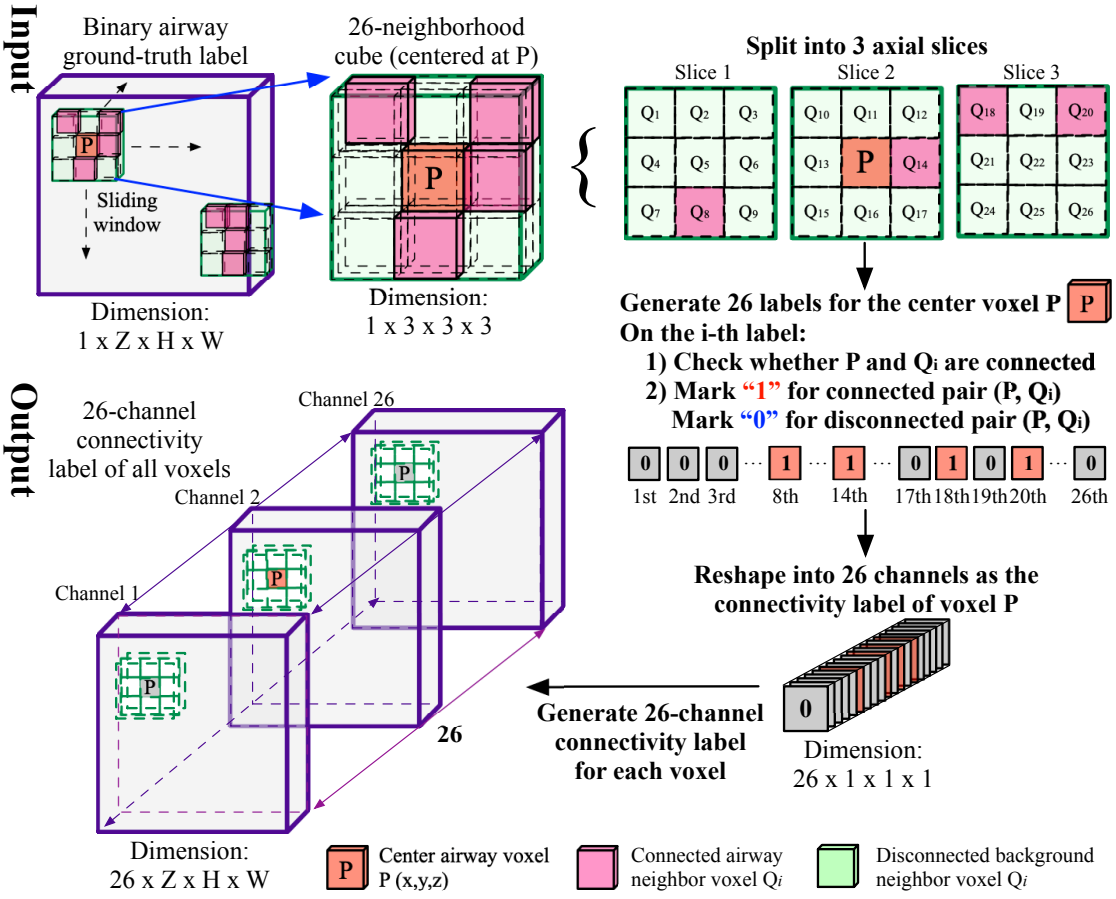


Figure 15: Illustration de la modélisation de la 26-connectivité.

$(P, Q_i), i \in \{1,2,...,26\}$ comme une orientation de connectivité. Chaque orientation est codée à l'aide d'une étiquette binaire à un canal. Si P et Q_i sont tous deux des voxels de voies aériennes, alors la paire (P, Q_i) est connectée et la position correspondante "P" sur la i-ième étiquette est marquée 1. Sinon, nous marquons 0 sur la i-ième étiquette pour représenter la paire déconnectée (P, Q_i). En glissant une telle fenêtre $3 \times 3 \times 3$ sur chaque voxel, nous obtenons 26 étiquettes binaires et les concaténons en une étiquette de connectivité à 26 canaux. La taille des étiquettes générées est maintenue inchangée grâce à l'ajout d'une marge nulle sur les limites du volume CT. Une telle étiquette de connectivité encode à la fois la position de la vérité du terrain et la relation de connectivité entre les voxels des voies respiratoires. Notez que toutes les opérations sont effectuées sur les étiquettes binaires conventionnelles de la vérité du terrain des voies aériennes. Nous n'avons pas besoin d'annotation manuelle supplémentaire pour les étiquettes de connectivité.

Le réseau AirwayNet proposé (voir Fig. 16) est basé sur l'ossature U-Net [Çiçek et al., 2016]. Notre architecture 3D complète capture plus d'informations spatiales que les CNN 2-D ou 2.5-D utilisés dans les [Charbonnier et al., 2017, Yun et al., 2019] et est plus adaptée à l'apprentissage de la continuité bronchique et des schémas de ramification. Le réseau AirwayNet se compose d'un chemin de contraction et d'un chemin d'expansion avec quatre échelles de résolution. À chaque échelle de résolution, la voie de contraction comporte deux couches de convolution (Conv) avec normalisation par lots (BN) et unité linéaire rectifiée (ReLU), suivies d'une couche de max-pooling. Dans la voie expansive, les caractéristiques plus fines de l'échelle de résolution inférieure sont d'abord suréchantillonnées linéairement, puis concaténées avec les caractéristiques grossières de la connexion de saut pour préserver les détails des bronches fines. Comme les voxels des voies respiratoires sont distribués dans la grande cavité thoracique, des informations sémantiques supplémentaires autres que l'intensité de l'échelle de gris sont considérées comme bénéfiques pour le modèle de classification des voxels des voies respiratoires. Nous utilisons ici les coordonnées des voxels et la carte de distance pulmonaire, et nous les concaténons avec les caractéristiques du chemin expansif à la dernière échelle. La fonction sigmoïde est appliquée sur le cube de connectivité prédite pour obtenir une distribution de probabilité.

Les principales différences entre AirwayNet et AirwayNet-SE résident dans l'architecture du réseau et la stratégie d'apprentissage des caractéristiques. AirwayNet adopte une approche à une étape, où un seul modèle basé sur CNNs est utilisé pour la prédiction. En revanche, AirwayNet-SE adopte une approche en deux étapes. Dans la première étape, deux modèles CNN (DNN et SWN) sont utilisés pour apprendre les caractéristiques des contextes à grande échelle et à petite échelle, respectivement. Dans la

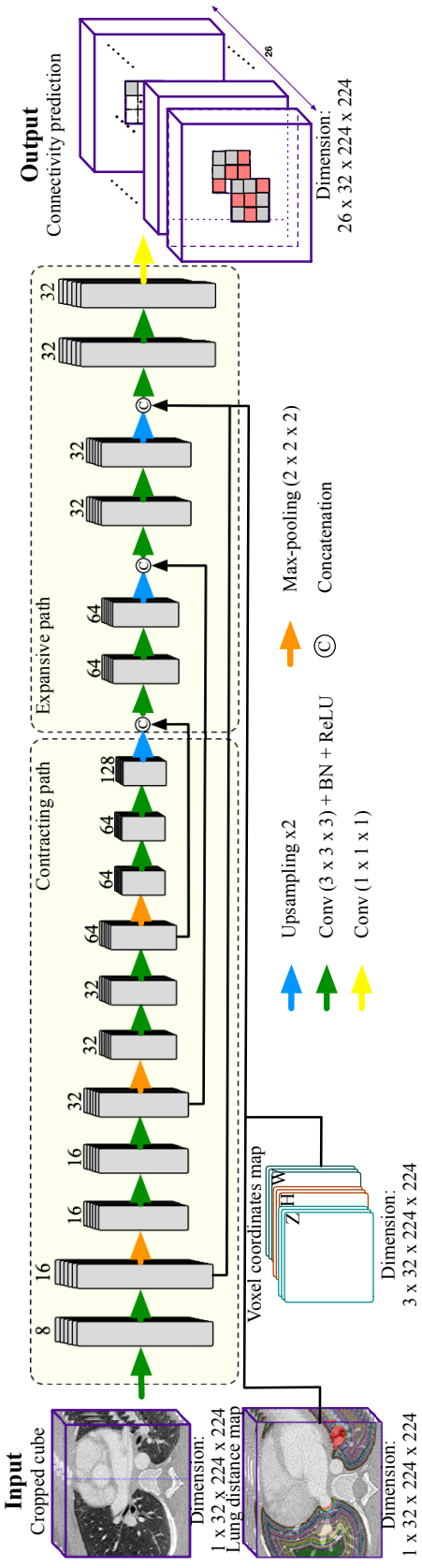


Figure 16: Illustration de l'AirwayNet. Le nombre de canaux est indiqué au-dessus de chaque carte de caractéristiques.

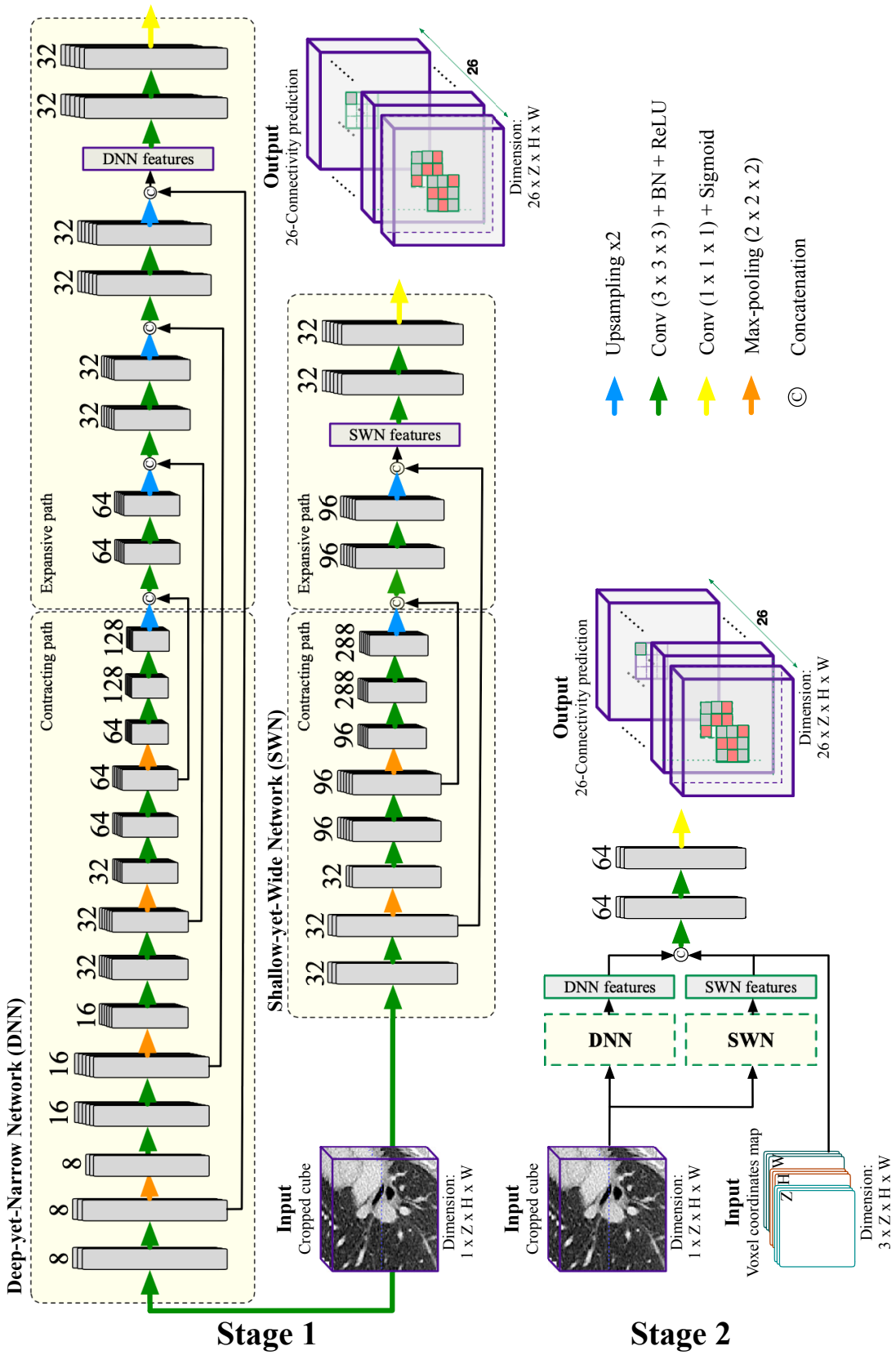


Figure 17: Illustration de l'AirwayNet-SE. Le nombre de canaux est indiqué sur chaque carte de caractéristiques. La première étape consiste à extraire des caractéristiques de deux échelles de contexte via DNN et SWN. La deuxième étape consiste à classifier la connectivité des voies aériennes en utilisant des caractéristiques fusionnées avec des contextes à grande et à petite échelle.

deuxième étape, ces caractéristiques des deux échelles sont fusionnées pour la prédiction finale de la connectivité. La Fig. 17 illustre le processus de prédiction du AirwayNet-SE.

Conclusion

Nous avons présenté l'AirwayNet et sa variante AirwayNet-SE pour la segmentation des voies respiratoires. Les deux méthodes proposées apprennent explicitement la connectivité des voxels pour percevoir la structure inhérente des voies respiratoires. Grâce à la modélisation de la connectivité, la tâche de segmentation classique est transformée en 26 tâches de prédiction de la connectivité, chaque tâche classant les voxels des voies aériennes selon une certaine orientation de connectivité. En outre, le AirwayNet-SE va un peu plus loin en fusionnant les caractéristiques de deux échelles de contexte. Les résultats expérimentaux ont prouvé que notre approche était efficace pour surmonter la différence de distribution entre les voies respiratoires de grande et de petite taille. Les annotations des voies respiratoires ont également été publiées pour stimuler la recherche sur l'extraction des voies respiratoires à l'aide de méthodes d'apprentissage supervisé.

À l'avenir, la méthode proposée pourrait être encore améliorée en travaillant sur (1) l'adoption de réseaux adversariaux génératifs pour produire divers échantillons d'entraînement afin d'améliorer la robustesse sur les scans de patients malsains et (2) l'exploration de mécanismes spécifiques d'amélioration des détails des bronches fines dans les scans CT de faible qualité afin d'améliorer les performances.

Chapitre 5 Apprentissage de Réseaux Neuronaux Convolutifs Sensibles aux Tubules pour la Segmentation des Voies Respiratoires et des Artères Pulmonaires dans le CT

Introduction

Nous présentons une méthode basée sur les CNN pour la segmentation des voies aériennes et des artères et veines pulmonaires. Comme les voies aériennes, les artères et les veines sont toutes des structures tubulaires, elles sont collectivement appelées tubules dans la présente étude. Grâce aux modules constitutifs soigneusement conçus, la méthode proposée apprend à comprendre la forme du contour, la distribution de l'intensité et la connectivité des bronches et des vaisseaux d'une manière guidée par les données. Elle relève les défis de l'application des CNN à la reconnaissance de tubules longs et fins et bénéficie d'une sensibilité élevée aux bronchioles, artérioles et veinules.

Tout d'abord, nous proposons un module de recalibrage des caractéristiques pour utiliser au maximum les caractéristiques apprises par les CNN. D'une part, pour augmenter le champ de vision pour la compréhension d'un contexte large, les architectures profondes avec de multiples couches de convolution et de mise en commun sont préférées. En conséquence, le nombre de paramètres apprenables augmente et le surajustement devient un problème. D'un autre côté, si le nombre de canaux de caractéristiques est simplement réduit pour éviter le sur-ajustement, on risque d'aller à l'autre extrême où le modèle ne parvient pas à apprendre des caractéristiques discriminantes. Par conséquent, le recalibrage des caractéristiques est envisagé car il intensifie les caractéristiques liées à la tâche pour une taille de modèle modérée. Dans la conception du module de recalibrage, nous supposons que les informations spatiales des caractéristiques sont indispensables pour le recalibrage par canal et qu'elles doivent être traitées différemment d'une position à l'autre et d'une couche à l'autre. Le moyen de regroupement utilisé dans [Rickmann et al., 2019, Zhu et al., 2019] pour la compression spatiale peut ne pas bien capturer l'emplacement des voies respiratoires et des vaisseaux dans différentes échelles de résolution. En revanche, nous visons à hiérarchiser les informations à des positions clés avec des poids apprenables, ce qui fournit des indications spatiales appropriées pour modéliser la dépendance entre les canaux et améliore ensuite le recalibrage.

Deuxièmement, nous introduisons un module de distillation de l'attention pour renforcer l'apprentissage de la représentation des voies respiratoires tubulaires, des artères et des veines. Les cartes d'attention de différentes échelles nous permettent de révéler potentiellement la morphologie et le modèle de distribution des voies respiratoires et

des vaisseaux. Inspirés par la distillation des connaissances [Zagoruyko and Komodakis, 2017, Hou et al., 2019], nous affinons les cartes d'attention de basse résolution en imitant celles de haute résolution. Les cartes d'attention plus fines (rôle de l'enseignant) avec un contexte plus riche peuvent engorger les cartes plus grossières (rôle de l'élève) avec des détails sur les voies respiratoires, les artères et les veines. La capacité du modèle à reconnaître les branches délicates est améliorée après une focalisation récursive sur l'anatomie cible. Face à des cibles de supervision insuffisants, la distillation elle-même agit comme une tâche d'apprentissage auxiliaire qui fournit des cibles supplémentaires pour aider à la formation.

Troisièmement, nous incorporons l'anatomie préalable dans la segmentation des artères et des veines en introduisant la carte du contexte pulmonaire et la carte de transformation de la distance. La carte du contexte pulmonaire, qui contient la lumière des voies aériennes, la paroi des voies aériennes et le poumon automatiquement segmentés, informe explicitement le modèle de la connaissance sémantique. La carte de transformation de la distance, calculée à l'aide des voies respiratoires extraites, enregistre la distance de chaque voxel à la paroi des voies respiratoires la plus proche.

Quatrièmement, la méthode de bout en bout proposée est applicable pour la segmentation à la fois des voies respiratoires pulmonaires et des artères et veines. Nous n'effectuons pas de segmentation indépendante des vaisseaux au préalable et nous n'avons pas besoin de post raffinement sur les sorties des CNNs. La segmentation basée sur la fenêtre glissante est utilisée et les coordonnées de chaque voxel dans la cavité thoracique sont introduites dans le modèle pour compenser la perte d'informations de position.

Enfin, bien que l'ensemble du cadre soit une solution intégrée à la segmentation des voies aériennes et des artères, les éléments qui le constituent peuvent être pris en compte pour concevoir des solutions à d'autres tâches. La méthode proposée peut également être facilement étendue en incorporant des techniques traditionnelles comme post-traitement (par exemple, les coupes de graphes), où la modélisation explicite des graphes et de la connectivité est introduite spécifiquement pour les structures tubulaires.

Nos contributions peuvent être brièvement résumées comme suit:

- Nous présentons une méthode basée sur des CNN sensibles aux tubules pour la segmentation des voies aériennes et des artères et veines pulmonaires. À notre connaissance, cette méthode représente la première tentative de segmentation simultanée des voies aériennes, des artères et des veines.

- Nous proposons un module de recalibrage des caractéristiques qui intègre des connaissances spatiales hiérarchisées pour un recalibrage par canal. Il encourage

l'apprentissage discriminant des caractéristiques.

- Nous introduisons un module de distillation de l'attention pour renforcer l'apprentissage de la représentation des voies respiratoires tubulaires, des artères et des veines. Aucun travail d'annotation supplémentaire n'est nécessaire.

- Nous incorporons un préalable anatomique explicite dans la segmentation des artères et des veines en utilisant la carte du contexte pulmonaire et la carte de transformation de la distance comme entrées supplémentaires.

- Nous validons respectivement la méthode proposée sur 110 et 55 scans CT cliniques sans contraste pour la segmentation des voies aériennes pulmonaires et des artères veineuses. Des expériences approfondies montrent que notre méthode a atteint une sensibilité supérieure pour les voies aériennes, les artères et les veines fines, tout en maintenant des performances de segmentation globales supérieures ou compétitives.

Méthodologie

La Fig. 18 présente une vue d'ensemble des méthodes proposées pour la segmentation des voies respiratoires et des artères. Pour réaliser un apprentissage efficace des caractéristiques des cibles tubulaires, des modules de recalibrage des caractéristiques et de distillation de l'attention sont introduits dans les CNN. Un antécédent d'anatomie est inclus pour fournir une connaissance sémantique de la tâche artère-veine.

Étant donné un volume CT d'entrée X, notre processus de segmentation peut être formulé comme suit : $P_{target} = \mathcal{F}(X)$, où $target$ peut être une voie aérienne, une artère ou une veine et P_{target} représente sa probabilité prédite correspondante. L'objectif est d'apprendre un mappage de bout en bout $\mathcal{F}$ via des CNN pour minimiser la différence entre P_{target} et son étiquette de vérité Y_{target}. En supposant que les CNN possèdent au total M couches de convolution, nous désignons la sortie d'activation de la m-ième convolution par $A_m \in R^{C_m \times D_m \times H_m \times W_m}$, $1 \leq m \leq M$. Les nombres de ses canaux, profondeurs, hauteurs et largeurs sont respectivement notés C_m, D_m, H_m et W_m.

Nous proposons la cartographie $\mathcal{Z}(\cdot)$ qui génère un descripteur de canal $U_m = \mathcal{Z}(A_m)$ pour recalibrer la caractéristique convolutive activée A_m. Un aperçu de $\mathcal{Z}(\cdot)$ pour le recalibrage de la caractéristique est donné dans la Fig. 19.

La distillation de l'attention est effectuée entre deux caractéristiques consécutives A_m et A_{m+1}. Tout d'abord, la carte d'attention est générée par $G_m = \mathcal{G}(A_m), G_m \in R^{1 \times D_m \times H_m \times W_m}$. La valeur absolue de chaque voxel dans G_m reflète la contribution de

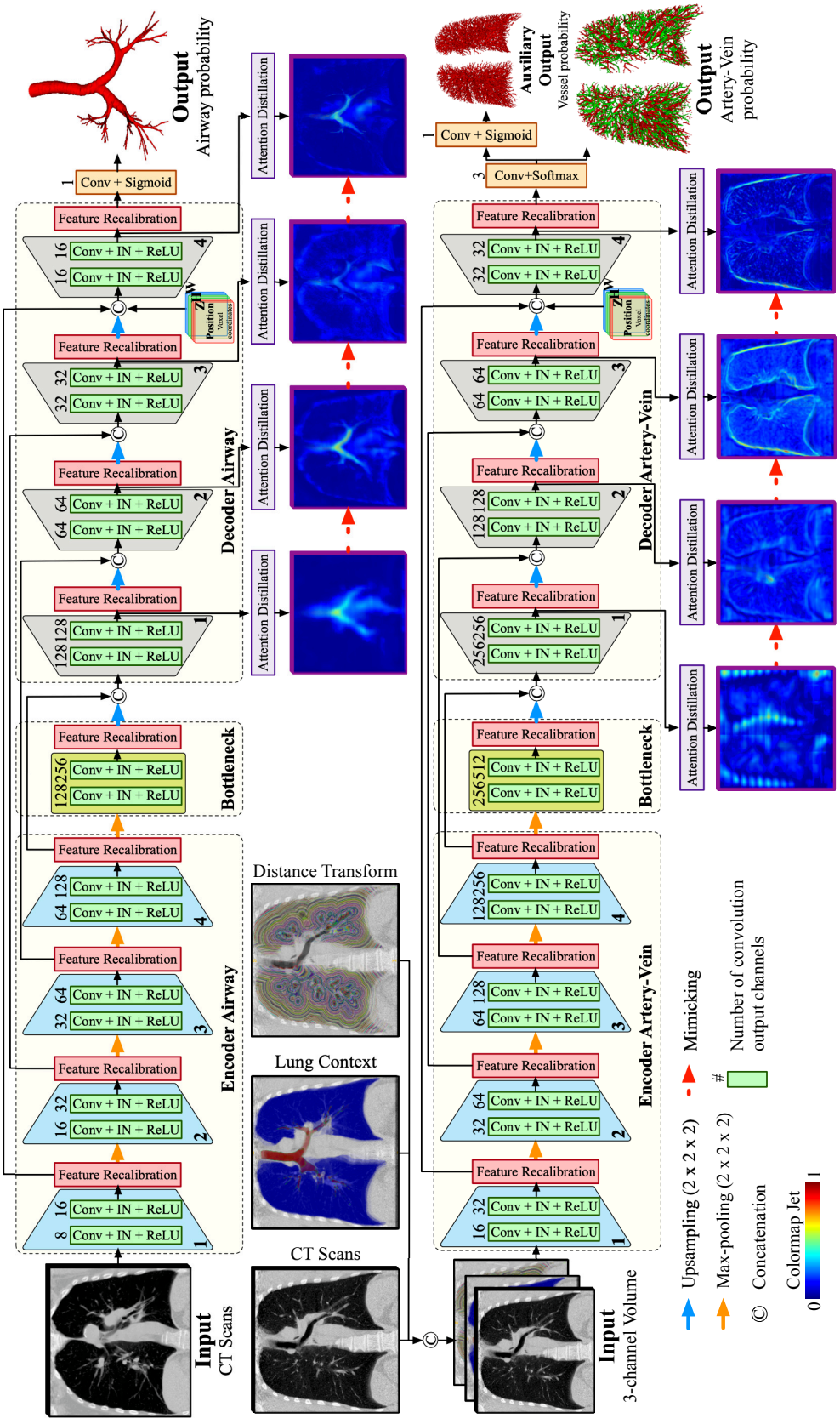


Figure 18: Vue d'ensemble de la méthode proposée pour la segmentation des voies aériennes pulmonaires et des veines artérielles. La normalisation des instances et l'activation ReLU sont effectuées après chaque couche de convolution, sauf la dernière. Le nombre de noyaux de convolution est indiqué au-dessus de chaque couche.

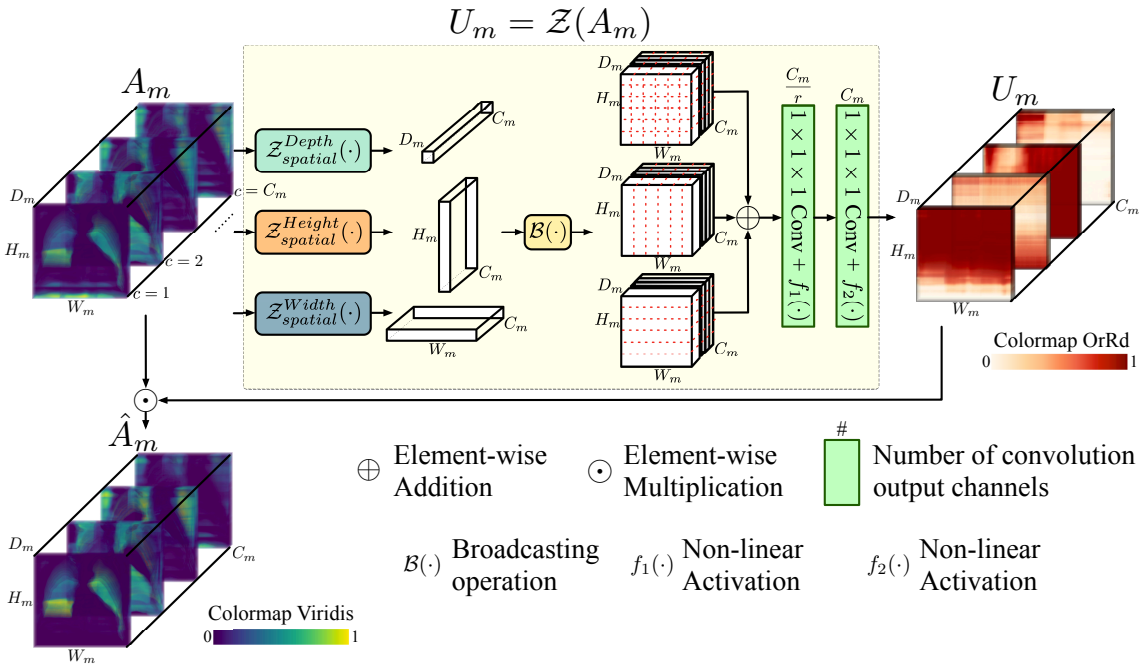


Figure 19: Illustration de la cartographie $\mathcal{Z}(\cdot)$ pour le recalibrage des caractéristiques. Son entrée est la caractéristique activée A_m de la m-ième couche de convolution. Tout d'abord, la carte spatiale qui met en évidence les régions importantes est intégrée par l'intermédiaire de $\mathcal{Z}_{spatial}(\cdot)$ selon trois axes : profondeur, hauteur et largeur. Ensuite, la recombinaison des canaux est effectuée sur la carte spatiale pour calculer le descripteur de canal U_m. La multiplication finale par éléments entre A_m et U_m produit la caractéristique recalibrée $\hat{A}_m$. Les notations r, C_m, D_m, H_m et W_m font référence au facteur de compression des canaux, au nombre de canaux, aux profondeurs, aux hauteurs et aux largeurs de A_m, respectivement.

sa correspondance dans A_m à l'ensemble du modèle de segmentation. Une façon de construire la fonction de mappage $\mathcal{G}(\cdot)$ est de calculer les statistiques des valeurs d'activation A_m à travers le canal:

$$G_m = \sum_{c=1}^{C_m} |A_m[c,:,:,:]|^p, \tag{2}$$

L'opération par éléments $|\cdot|^p$ désigne la valeur absolue élevée à la p-ième puissance. Les régions fortement activées font l'objet d'une plus grande attention si $p > 1$. Ici, nous adoptons la sommation par canal au lieu de maximiser $\max_c(\cdot)$ ou de calculer la moyenne $\frac{1}{C_m}\sum_{c=1}^{C_m}(\cdot)$ car elle est relativement moins biaisée. L'opération de sommation conserve toutes les informations d'activation saillantes implicites sans ignorer les éléments non maximaux ni affaiblir les éléments discriminants. Des expériences préliminaires montrent que la sommation avec $p > 1$ intensifie la plupart des régions sensibilisées liées à la tâche (par exemple, les bords du poumon, les bronches, les vaisseaux). Ensuite, une interpolation trilinéaire $\mathcal{I}(\cdot)$ est effectuée pour s'assurer que les cartes d'attention 3D traitées partagent la même dimension.

Ensuite, on applique spatialement la méthode Softmax $\mathcal{S}(\cdot)$ au voxel pour normaliser tous les éléments dans $[0,1]$. Enfin, nous rapprochons l'attention distillée $\hat{G}_m$ de $\hat{G}_{m+1}$ en minimisant la perte:

$$\mathcal{L}_{distill} = \sum_{m=1}^{M-1} \|\hat{G}_m - \hat{G}_{m+1}\|_F^2, \hat{G}_m = \mathcal{S}(\mathcal{I}(G_m)), \tag{3}$$

où $\|\cdot\|_F^2$ est la norme de Frobenius au carré. Avec $\hat{G}_m$ imitant récursivement son successeur $\hat{G}_{m+1}$, l'attention visuelle est transmise de la couche la plus profonde à la couche la moins profonde. Notez que ce processus de distillation ne nécessite pas de travail d'annotation supplémentaire et peut facilement fonctionner avec des CNNs arbitraires. Dans l'implémentation, pour éviter que la dernière attention $\hat{G}_{m+1}$ ne s'approche de la précédente $\hat{G}_m$, nous détachons $\hat{G}_{m+1}$ du graphe de calcul pour chaque m dans le calcul des pertes. Par conséquent, $\hat{G}_{m+1}$ ne sera pas modifié par les erreurs de rétropropagation. La raison pour laquelle nous ne sous-échantillonnons pas $\hat{G}_{m+1}$ à la taille de $\hat{G}_m$ est que $\hat{G}_{m+1}$ du côté décodeur a une résolution plus élevée que $\hat{G}_m$ par nature et que le sous-échantillonnage perd des informations riches qui n'existent que dans $\hat{G}_{m+1}$. Il est nécessaire de garder $\hat{G}_{m+1}$ inchangé afin que la perte de distillation résultante entre $\hat{G}_m$ et $\hat{G}_{m+1}$ puisse améliorer l'attention du modèle sur les détails fins des cibles.

Deux cartes sont introduites comme antécédents anatomiques pour la segmentation

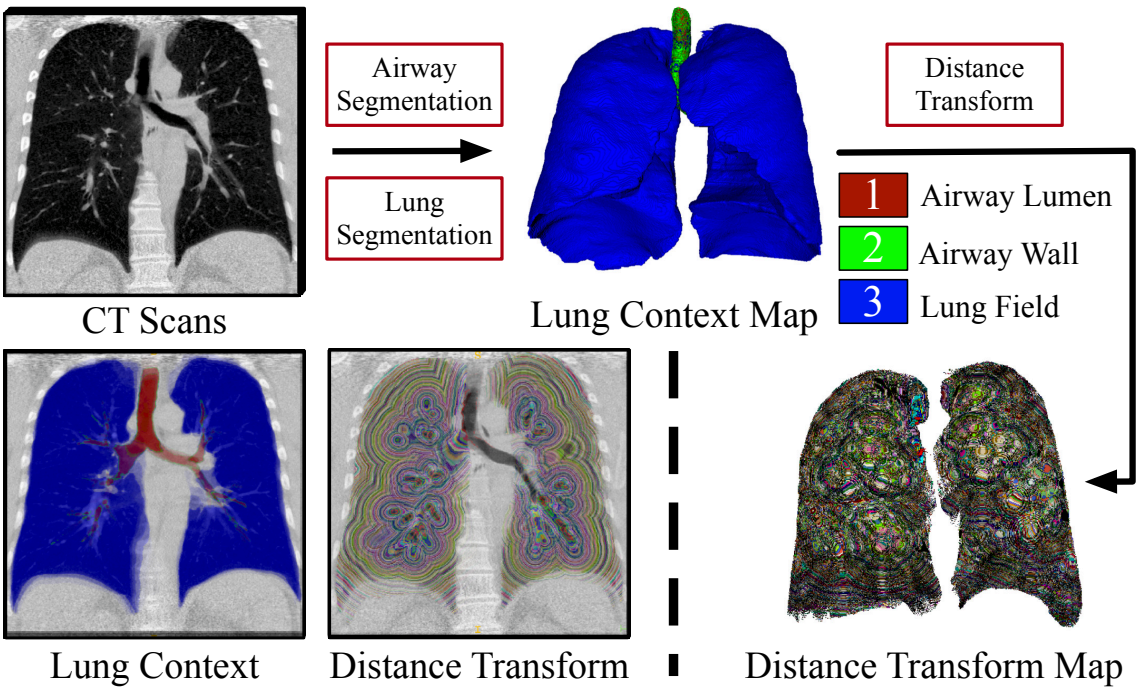


Figure 20: Illustration de l'anatomie avant l'incorporation. La représentation visuelle des cartes de contexte pulmonaire et des cartes de transformation de distance générées, superposées aux CT images, est donnée en bas à gauche.

des artères et des veines (voir Fig. 20) : la carte du contexte pulmonaire et la carte de transformation de la distance. La première carte offre une connaissance sémantique supplémentaire du poumon et la seconde reflète la proximité des voxels avec les voies respiratoires. Ces cartes sont concaténées avec le sous-volume du CT comme entrées du modèle de segmentation des artères et des veines.

Conclusion

Nous avons présenté une méthode sensible aux tubules pour la segmentation des voies aériennes pulmonaires et des veines artérielles. Elle utilise des CNN et ne nécessite aucun post-traitement. Avec le module proposé de recalibrage des caractéristiques en fonction de l'espace et le module de distillation de l'attention progressivement renforcé, l'apprentissage des caractéristiques de nos CNN devient plus efficace et plus pertinent pour la perception des tubules cibles. L'antériorité anatomique incorporée est également bénéfique pour la séparation artère-veine. Des expériences approfondies ont montré que notre méthode a détecté beaucoup plus de bronchioles, d'artérioles et de veinules tout en maintenant des performances globales de segmentation compétitives, ce qui corrobore sa sensibilité supérieure aux méthodes de l'état de l'art et la validité de ses constituants.

Chapitre 6 Conclusions Générales et Perspectives

Pour conclure, nous avons étudié plusieurs méthodes d'apprentissage profond pour la classification et la segmentation d'images de chromosomes et d'images pulmonaires. Les méthodes proposées présentent un grand potentiel clinique pour améliorer l'efficacité de l'analyse du caryotype, la mesure de la structure pulmonaire ainsi que le diagnostic des lésions. Compte tenu des limites constatées dans la présente étude, les travaux futurs sont discutés dans la section suivante. Les travaux futurs sont divisés en deux parties : 1) L'une consiste à résoudre de nouvelles tâches pour l'analyse des chromosomes et des images pulmonaires. 2) L'autre est d'améliorer encore les performances de la classification et de la segmentation en appliquant de nouvelles techniques d'apprentissage profond.

General Introduction

0.1 Problem Statement and Objectives

Pulmonary diseases including both chronic obstructive pulmonary disease (COPD) and lung cancer, could lead to fatal damage to human health. COPD is characterized by persistent airflow limitation caused by pulmonary airway and alveolar abnormalities [Vogelmeier et al., 2017, Halpin et al., 2021]. Airway inflammation and emphysematous destruction of lung tissue are often observed in patients that are exposed to toxic particles or gases for a long time [Hogg, 2004]. Emphysema, a typical type of COPD, is characterized by the difficulty of blowing air out. Air-filled cavities in the lung, which are connected to terminal bronchioles, are detected. All these structural deformities trigger off loss of lung elastic recoil and gas trapping, thereafter resulting in lung hyperinflation [Hogg, 2004]. Another type of COPD is chronic bronchitis, which is characterized by the inflammation and irritation of the bronchi tubes. The most common cause of chronic bronchitis is cigarette smoking [Sethi and Rochester, 2000]. Due to the mucus and swelling of airway bronchi, patients with such disease feel difficult to breathe.

Lung cancer is also a non-negligible death threat. It has been one of the leading cancers in both men and women, causing 1.3 million deaths worldwide per year [Torre et al., 2016]. On the basis of clinical behavior and histological appearance, lung cancer could be classified into two major categories: small cell cancer and non-small cell cancer [Minna et al., 2002]. Although the overall 5-year survival rate is only 18%, if early diagnosis and treatment are put into effect timely, the patients' chances of survival can be greatly increased [Siegel et al., 2016]. Pulmonary nodules are small round or oval-shaped growth in lung and often viewed as an early indication of cancer. Nodules are usually the result of inflammation in the lung. More than 90% of solid nodules are benign if their diameters are less than 2 cm [Winer-Muram, 2006]. It is therefore of great clinical value if nodules can be detected before deterioration.

Due to the development of biomedical imaging, especially the X-ray computed tomography (CT), pulmonary diseases can be revealed from tomographic features. The wide-spread use of conventional CT makes it possible to display pulmonary structures

for accurate diagnosis of diseases. For analysis of pulmonary CT imaging, one prerequisite step is to extract pulmonary airways from CT. The modeling of airway tree benefits the quantification of its morphological changes for diagnosis of bronchial stenosis, acute respiratory distress syndrome, idiopathic pulmonary fibrosis, COPD, obliterative bronchiolitis, and pulmonary contusion [Howling et al., 1998, Shaw et al., 2002, Fetita et al., 2004, Li et al., 2019, Wu et al., 2019]. Combined with photo-realistic rendering and projection, the segmented airways play an important role in virtual bronchoscopy and endobronchial navigation for surgery [Mori et al., 2000, Natori et al., 2005, Shen et al., 2015a, Shen et al., 2019]. Another essential step is to extract pulmonary arteries and veins from CT. Pulmonary diseases may affect artery or vein, or both but in different ways [Melot and Naeije, 2011, Charbonnier et al., 2015]. Morphological changes of arteries are measured in diagnosing pulmonary embolism, arteriovenous malformations, and COPD [Zhou et al., 2007, Wittenberg et al., 2012, Cartin-Ceba et al., 2013, Estépar et al., 2013]. The arterial alterations also serve as an imaging biomarker in chronic thromboembolic pulmonary hypertension [Rahaghi et al., 2016]. The imaging features of veins are found useful in diagnosis of vein diseases [Porres et al., 2013]. Despite the benefits of airway and artery-vein segmentation, it requires heavy workloads for manual delineation due to the complexity of tubular structures. Consequently, automatic segmentation methods were developed to reduce burden and improve accuracy. Especially if arteries and veins can be extracted from non-contrast CT (i.e. CT without the use of contrast agents), CT pulmonary angiogram may not be needed in certain cases to avoid adverse reactions to contrast agents [Cochran et al., 2001, Loh et al., 2010]. Nevertheless, extraction of airway, artery, and vein by automatic segmentation methods are prone to discontinuity since there exists severe class imbalance between tubular foreground and background. The airway and vessel voxels are sparse and scattered with respect to background. Besides, there exist differences between main, thick branches and peripheral, thin branches in intensity and spatial distribution. Segmentation methods on airway, artery, and vein are required to perceive and handle such differences between local and global scales.

Computer-aided diagnosis (CAD) systems have been developed to improve diagnosis of pulmonary nodules with CT [Messay et al., 2010, Lopez Torres et al., 2015, Jacobs et al., 2014, Setio et al., 2016, Sakamoto and Nakano, 2016, Dou et al., 2017, Huang et al., 2017b]. In the design of nodule CAD systems, the prerequisite step is nodule segmentation. Compared with manual delineation of nodule boundaries by radiologists, these systems efficiently provide consistent prediction results without inter-observer variance. The quality of segmentation directly affects the subsequent measurement of nodules for classification of benignity and malignancy. The main difficulty in nodule segmentation is to design an algorithm that adapts to both internal texture and external surround-

ings of pulmonary nodules. Most previous segmentation methods were developed for solid nodules. Few methods were applicable for segmentation of all solid, part-solid, and ground glass opacity (GGO) nodules. Furthermore, the similarity between nodules and lung tissue in intensity and the complicated surroundings of nodules pose non-negligible challenges to the generalizability of segmentation methods. For nodules that are connected to pleural surface, vessel, and airway wall, segmentation methods often fail to generate accurate boundaries, causing under or over segmentation. The reason behind is that nodules share similar intensity with surrounding tissues. Under such circumstance, it is of great importance for the segmentation method to comprehend well the shape, texture, and position distribution of nodules.

The advance of microscopy imaging makes it possible for pathogenesis study on pulmonary diseases at the chromosome level. Most human cancers display structural and numerical chromosome abnormalities [Mitelman, 2000, Rowley, 2001, Mitelman et al., 2004]. Associated with lung cancer, non-random aberrations of chromosomes are complex, with multiple numerical and structural rearrangements [Balsara and Testa, 2002, Testa and Siegfried, 1992, Park et al., 2001, Masuda and Takahashi, 2002, Grigorova et al., 2005]. Moreover, mosaic abnormalities of somatic chromosomes were detected in the lungs of patients with pulmonary arterial hypertension [Aldred et al., 2010]. To perform chromosome analysis for pulmonary disease study, one important procedure is karyotyping, where meta-phase chromosomes within one cell are stained, imaged, classified, and sorted in order [Piper, 1990]. According to the staining technique and imaging mechanism, the karyotyping can be divided into Giemsa karyotyping and florescent karyotyping. Giemsa karyotyping are preferred because it can detect nearly all abnormalities with a single low-cost test. However, it requires meticulous efforts for cytogeneticists to classify Giemsa-stained chromosomes by their banding patterns. Many automatic chromosome classification methods have been proposed to improve the efficiency of karyotyping. Most of them relied on accurate extraction of medial axes and centromeres for feature computation [Lerner et al., 1995, Ming and Tian, 2010, Markou et al., 2012, Stanley et al., 1996, Wang et al., 2008, Arachchige et al., 2013, Loganathan et al., 2013]. Due to the difficulty of skeletonization for bending, deformed chromosomes, previous methods often failed to perform robustly in classification. Methods, which can tackle large variations in shape and banding appearance of chromosomes, are needed to meet clinical standards.

The main research objectives of this thesis are two-fold: one is developing a chromosome classification method in microscopic imaging for Giemsa-karyotyping; the other is developing segmentation methods of pulmonary airway, artery, vein, and nodule in CT imaging for measurement and diagnosis. Studies of both the two objectives lay the

foundation for providing explanations on pathogenesis and consequences of pulmonary diseases, with the first objective at a micro scale and the second one at a macro scale. To cope with the limitations of classification and segmentation methods in literature, deep learning approaches are investigated in method development. Compared with traditional methods that depend on manually designed features, deep learning based methods learn effective features that are expected to be more robust to variations of objects. Thus, in the present study, deep learning approaches are exploited to improve performance in classification and segmentation of chromosome and pulmonary images.

0.2 Main Contributions

The main contributions of this thesis are detailed as follows:

- Development of a Chromosome Classification Approach Using Deep Convolutional Networks (Chapter 2).

  Chromosome classification is critical for karyotyping in abnormality diagnosis. To expedite the diagnosis, we present a novel method named Varifocal-Net for simultaneous classification of chromosome's type and polarity using deep convolutional networks. The approach consists of one global-scale network (G-Net) and one local-scale network (L-Net). It follows three stages. The first stage is to learn both global and local features. We extract global features and detect finer local regions via the G-Net. By proposing a Varifocal mechanism, we zoom into local parts and extract local features via the L-Net. Residual learning and multi-task learning strategies are utilized to promote high-level feature extraction. The detection of discriminative local parts is fulfilled by a localization subnet of the G-Net, whose training process involves both supervised and weakly-supervised learning. The second stage is to build two multi-layer perceptron classifiers that exploit features of both two scales to boost classification performance. The third stage is to introduce a dispatch strategy of assigning each chromosome to a type within each patient case, by utilizing the domain knowledge of karyotyping. Evaluation results from 1909 karyotyping cases showed that the proposed Varifocal-Net achieved the highest accuracy per patient case of 99.2% for both type and polarity tasks. It outperformed state-of-the-art methods, demonstrating the effectiveness of our Varifocal mechanism, multi-scale feature ensemble, and dispatch strategy. The proposed method has been applied to assist practical karyotype diagnosis.

- Pulmonary Nodule Segmentation with CT Sample Synthesis Using Adversarial Networks (Chapter 3).

Segmentation of pulmonary nodules is critical for the analysis of nodules and lung cancer diagnosis. We present a novel framework of segmentation for various types of nodules using convolutional neural networks (CNNs). The proposed framework is composed of two major parts. The first part is to increase the variety of samples and build a more balanced dataset. A conditional generative adversarial network (cGAN) is employed to produce synthetic CT images. Semantic labels are generated to impart spatial contextual knowledge to the network. Nine attribute scoring labels are combined as well to preserve nodule features. To refine the realism of synthesized samples, reconstruction error loss is introduced into cGAN. The second part is to train a nodule segmentation network on the extended dataset. We build a 3D CNN model that exploits heterogeneous maps including edge maps and local binary pattern maps. The incorporation of these maps informs the model of texture patterns and boundary information of nodules, which assists high-level feature learning for segmentation. Residual unit, which learns to reduce residual error, is adopted to accelerate training and improve accuracy. Validation on LIDC-IDRI dataset demonstrates that the generated samples are realistic. The mean squared error and average cosine similarity between real and synthesized samples are 1.55×10^{-2} and 0.9534, respectively. The Dice coefficient, positive predicted value, sensitivity, and accuracy are respectively 0.8483, 0.8895, 0.8511, 0.9904 for the segmentation results. The proposed 3D CNN segmentation framework, based on the use of synthesized samples and multiple maps with residual learning, achieves more accurate nodule segmentation compared to existing state-of-the-art methods. The proposed CT image synthesis method can not only output samples close to real images but also allow for stochastic variation in image diversity.

- Development of a Voxel-Connectivity Aware Approach for Accurate Airway Segmentation Using Convolutional Neural Networks (Chapter 4).

  Accurate segmentation of airways from chest CT scans is crucial for pulmonary disease diagnosis and surgical navigation. However, the intra-class variety of airways and their intrinsic tree-like structure pose challenges to the development of automatic segmentation methods. To address this, we propose a voxel-connectivity aware approach named AirwayNet for accurate airway segmentation. By connectivity modeling, conventional binary segmentation task is transformed into 26 tasks of connectivity prediction. Thus, our AirwayNet learns both airway structure and relationship between neighboring voxels. We further propose the two-step AirwayNet-SE, a Simple-yet-Effective approach to improve AirwayNet. The first step of AirwayNet-SE is to adopt connectivity modeling to transform the bi-

nary segmentation task into 26-connectivity prediction task, facilitating the model's comprehension of airway anatomy. The second step is to predict connectivity with a two-stage CNNs-based approach. In the first stage, a Deep-yet-Narrow Network (DNN) and a Shallow-yet-Wide Network (SWN) are respectively utilized to learn features with large-scale and small-scale context knowledge. These two features are fused in the second stage to predict each voxel's probability of being airway and its connectivity relationship between neighbors. We trained our model on 50 CT scans from public datasets and tested on another 20 scans. Compared with state-of-the-art airway segmentation methods, the robustness and superiority of the AirwayNet-SE confirmed the effectiveness of large-scale and small-scale context fusion. In addition, we released our manual airway annotations of 60 CT scans from public datasets for supervised airway segmentation study.

- Learning Tubule-Sensitive Convolutional Neural Networks for Pulmonary Airway and Artery-Vein Segmentation in CT (Chapter 5).

  Training convolutional neural networks (CNNs) for segmentation of pulmonary airway, artery, and vein is challenging due to sparse desired targets caused by the severe class imbalance between tubular targets and background. We present a CNNs-based method for accurate airway and artery-vein segmentation in non-contrast computed tomography. It enjoys superior sensitivity to tenuous peripheral bronchioles, arterioles, and venules. The method first uses a feature recalibration module to make the best use of features learned from the neural networks. Spatial information of features is properly integrated to retain relative priority of activated regions, which benefits the subsequent channel-wise recalibration. Then, attention distillation module is introduced to reinforce representation learning of tubular objects. Fine-grained details in high-resolution attention maps are passing down from one layer to its previous layer recursively to enrich context. Anatomy prior, consisting of lung context map and distance transform map, is designed and incorporated for better artery-vein differentiation capacity. Extensive experiments demonstrated considerable performance gains brought by these components. Compared with state-of-the-art methods, our method extracted much more branches while maintaining competitive overall segmentation performance.

0.3 Organization of Thesis

The thesis manuscript is organized as follows:

In Chapter 1, entitled "**Biomedical Context and Technical Background**", the biomed-

ical background about chromosome karyotyping and pulmonary CT image segmentation is introduced. Besides, for deep learning techniques, the history of neural network development is presented, together with reviews on recent CNNs architectures. We also explain the mechanism of adversarial learning via GANs. Finally, three frequently-used optimization algorithms are outlined.

In Chapter 2, entitled "**Development of a Chromosome Classification Approach Using Deep Convolutional Networks**", we have proposed the three-stage Varifocal-Net for chromosome classification, which has been evaluated on a large manually constructed dataset. Each stage of the proposed method is described, including global-scale and local-scale feature learning, classification based on the fused features, and type assignment using dispatch strategy.

In Chapter 3, entitled "**Pulmonary Nodule Segmentation with CT Sample Synthesis Using Adversarial Networks**", we have proposed a two-part CNNs-based framework for pulmonary nodule segmentation. In the first part, adversarial networks are introduced to synthesize nodule samples. In the second part, the 3D CNNs-based segmentation model is proposed using multiple heterogeneous feature maps and residual learning strategy. The proposed nodule segmentation method was evaluated on LIDC-IDRI dataset.

In Chapter 4, entitled "**Development of a Voxel-Connectivity Aware Approach for Accurate Airway Segmentation Using Convolutional Neural Networks**", we have proposed the AirwayNet and its variant AirwayNet-SE for airway segmentation. The proposed two methods explicitly learn voxel connectivity to perceive airway's inherent structure. The effectiveness of the proposed method was validated on both public and private datasets.

In Chapter 5, entitled "**Learning Tubule-Sensitive Convolutional Neural Networks for Pulmonary Airway and Artery-Vein Segmentation in CT**", we have proposed a tubule-sensitive method for both pulmonary airway and artery-vein segmentation. Extensive experiments were conducted to corroborate its superior sensitivity over state-of-the-art methods and the validity of its constituents including feature recalibration module, attention distillation module, and anatomy prior incorporation.

In Chapter 6, entitled "**General Conclusions and Perspectives**", a brief summary of the main contributions, conclusions, and potential future perspectives is presented.

Chapter 1

Biomedical Context and Technical Background

Contents

1.1 Chromosome Karyotyping 54
1.1.1 Giemsa Staining for Chromosome Imaging 54
1.1.2 Chromosome Separation and Classification 55
1.1.3 Enumeration and Abnormality Diagnosis 56
1.2 Pulmonary CT Image Segmentation 56
1.2.1 Anatomy of Human Pulmonary System 56
1.2.2 Lung Diseases Overview 60
1.2.3 Segmentation Methods Overview 61
1.3 State-of-the-art Deep Learning Methods 65
1.3.1 Artificial Neural Networks 65
1.3.2 Convolutional Neural Networks 67
1.3.3 CNNs Architectures Overview 69
1.3.4 Generative Adversarial Networks 72
1.3.5 Optimization Algorithms Overview 72
1.4 Summary 74

1.1 Chromosome Karyotyping

Chromosome anomalies, including numerical and structural abnormalities, are responsible for several genetic diseases such as leukemia [Natarajan, 2002]. Numerical abnormalities arise from the gain or loss of an entire chromosome, which constitute a great proportion of abnormalities [Theisen and Shaffer, 2010]. Structural abnormalities result from the loss, breakage, and reunion of chromosome segments. In clinical practice, an important procedure for chromosome diagnosis is karyotyping, which is carried out on microscopic images of a single cell [Piper, 1990]. The karyotyping process is performed by experienced clinical cytogeneticists and it mainly contains three steps: 1) chromosome staining and imaging; 2) manual separation and classification on chromosome images; 3) enumeration and abnormality diagnosis.

1.1.1 Giemsa Staining for Chromosome Imaging

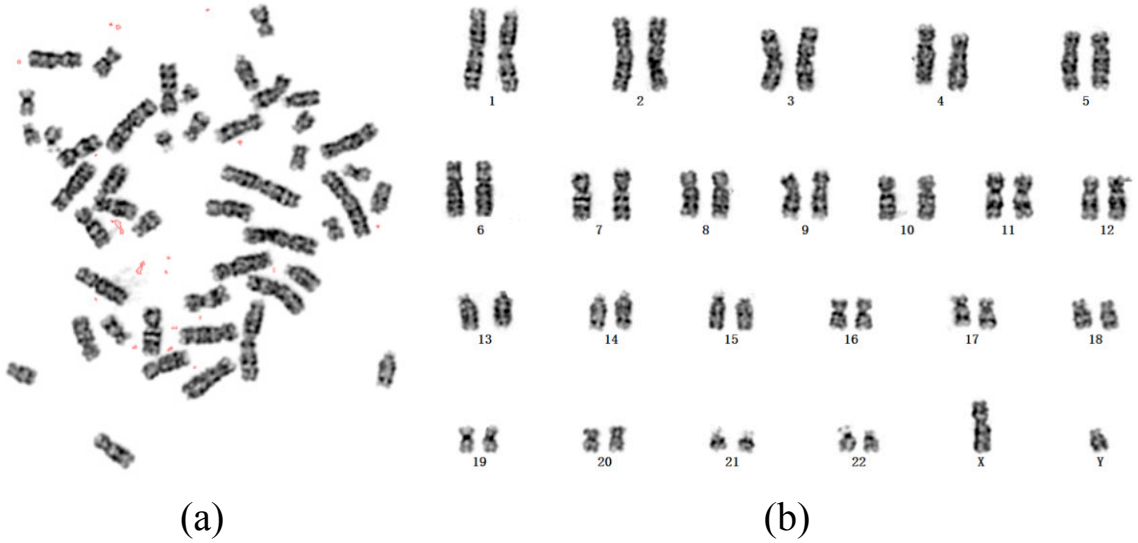


Figure 1.1: (a) A Giemsa-stained microscopic image of male chromosomes for one case. (b) The karyotyping result (a.k.a. karyogram) of (a) is formed of the paired and ordered chromosomes (22 pairs of autosomes and 1 pair of sex chromosomes XY).

The first step of karyotyping is to use staining techniques on each cell to obtain stained meta-phase chromosomes. The karyotyping can be classified into two main categories by the used staining technique and imaging mechanism: Giemsa karyotyping using Giemsa staining and fluorescent karyotyping using fluorescent staining. If fluorescent staining is employed together with deconvolution of fluorescence signals, chromosomes of different types will be dyed with different colors for fluorescent karyotyping (e.g., SKY [Schröck et al., 1996] and M-FISH [Speicher et al., 1996]). The Giemsa staining, on the other hand, refers to the staining technique that uses a visible light dye called Giemsa to stain meta-phase chromosomes. After staining, banding patterns that appear alternatively darker and lighter gray-levels (a.k.a. G-bands) will be produced for Giemsa karyotyping.

In clinical applications, Giemsa staining and karyotyping are preferred rather than fluorescent staining and karyotyping. Although fluorescent karyotyping is easy for operators to distinguish chromosomes by color, its inherent limitations (e.g., difficulty of detecting all chromosomal abnormalities, impermanent preservation of fluorescence signals, prohibitive cost, controversial reliability of probe hybridization, and unavailability of various probes and clinical samples) make it inappropriate as a first-tier screening tool for examinations [Lee et al., 2001, Huber et al., 2018, Gozzetti and Le Beau, 2000]. In contrast, Giemsa karyotyping can detect nearly all abnormalities with a single low-cost test. Besides, the Giemsa stained slides of chromosomes can be preserved permanently while the the fluorescence signals of fluorescent staining are difficult to be kept for a long term. Such advantages of Giemsa staining and karyotyping make it possible for fast, economic, and effective screening of chromosomal abnormalities.

Nowadays, there exist mainly three commercial microscope systems for chromosome imaging, including CytoVision System from Leica [Micci et al., 2001, Yang et al., 2010, Rødahl et al., 2005], Ikaros from MetaSystem [Gadhia et al., 2014], and HiBand from ASI [Fan et al., 2000]. These systems integrate both the microscope hardware and imaging software. They are targeted at the automation of the whole karyotyping workflow, which includes meta-phase scanning and capturing, data archive, and interactive karyotyping interface. A typical microscopic image of Giemsa-stained chromosomes and its corresponding karyogram are shown in Fig. 1.1.

1.1.2 Chromosome Separation and Classification

Given the captured chromosome image (see Fig. 1.1(a)), the second step is to manually extract and classify each chromosome from clusters. These classified chromosomes are further sorted and arranged into 22 pairs of autosomes and 1 pair of sex chromosomes (XX or XY) in the karyotyping map (a.k.a. karyogram). During this process, extra attention is paid to the length, centromere position, banding pattern, and contour curvature of chromosomes. Therefore, the process of karyotyping demands meticulous efforts from well-trained operators. To reduce the burden of karyotyping, many automated segmentation and classification methods have been developed for analyzing meta-phase chromosomes [Ji, 1994, Minaee et al., 2014, Saleh et al., 2019, Cao et al., 2020, Lerner et al., 1995, Ming and Tian, 2010, Markou et al., 2012, Madian and Jayanthi, 2014, Biyani et al., 2005, Abid and Hamami, 2018, Sharma et al., 2017, Gupta et al., 2017, Wu et al., 2018b]. In general, both segmentation and classification methods consist of four steps. The first is to preprocess the chromosome image, which usually involves thresholding, convex hull computation, and skeletonization algorithms to extract the medial axis of each chromosome in the image. The second is to locate end points of media axes and split touching chromosomes based on the identified cut points. After that, features along each chromosome's axis are computed. The final step is to build classifiers (e.g., multi-layer perceptron (MLP) and support vector machine (SVM)) to estimate chromosome's type based on the extracted features.

1.1.3 Enumeration and Abnormality Diagnosis

Finally, experts analyze the karyogram for possible numerical and structural abnormality diagnosis. Usually, the enumeration of chromosomes is performed on at least 20 karyograms per patient. If any abnormality is detected (e.g., chromosome mosaicism) on one karyogram, another 50–100 microscope images of the same patient are needed to confirm the diagnosis. Considering that each human cell normally contains 46 chromosomes, it is time-consuming to complete the entire diagnosis process. Even a sophisticated cytogeneticist has to spend 15 minutes or more in chromosome enumeration for one patient. To improve efficiency of chromosome enumeration and diagnosis, two computer-aided systems were developed. Gajendran et al. [Gajendran and Rodríguez, 2004] combined a variety of pre-processing methods and topological analysis in the counting algorithm. Li et al. [Xiao et al., 2020] developed a chromosome enumeration framework using convolution neural networks, where the object detection backbone–Faster R-CNN [Ren et al., 2015] was adopted for chromosome detection. Due to the variety of chromosome morphology and the complexity of overlapping and touching chromosomes, the performance of automatic enumeration and diagnosis methods remains to be further improved.

1.2 Pulmonary CT Image Segmentation

In X-ray computed tomography (CT), an image is a collection of measurements in three-dimensional (3-D) space and the location of each measurement is called a voxel [Pham et al., 2000]. Image segmentation refers to the process of partitioning an image into non-overlapping yet constituent regions with homogeneous characteristics such as intensity or texture. In CT imaging, the segmentation task can also be viewed as voxel-wise classification, where each voxel is classified into one category (e.g., background or foreground). Generally, segmentation methods can be grouped by the techniques in use: 1) thresholding; 2) region growing; 3) classifier; 4) clustering; 5) Markov random field models; 6) neural networks; 7) deformable models; 8) atlas models [Pham et al., 2000]. In view of the breadth of the segmentation topic, this section only introduces the background and techniques that are relevant to pulmonary CT image segmentation.

1.2.1 Anatomy of Human Pulmonary System

The function of human pulmonary system (a.k.a. respiratory system) is to enable gas exchange between the air and the blood, transporting oxygen from the atmosphere into the blood and releasing carbon dioxide back to the atmosphere [Selvan, 2018]. It includes organs like the nose, pharynx, larynx, trachea, bronchi, and lungs. The lungs are paired organs that take up most of the space within the thoracic cavity. They are divided into individual lobes by the bronchial tree and physically separated by pleural membranes (a.k.a. fissures). The left lung has two lobes: superior and inferior. In comparison, the right lung has three lobes: superior, middle, and inferior. Lungs consist of pulmonary vessels (arteries, veins, and capillaries), airways, and connective tissue [Miller, 1947, Netter, 2014]. The lung tissue is also called lung parenchyma. It consists of up to 700 million alveoli, which are the basic units of respiration. Fig. 1.2 presents an anatomical view of the human pulmonary system.

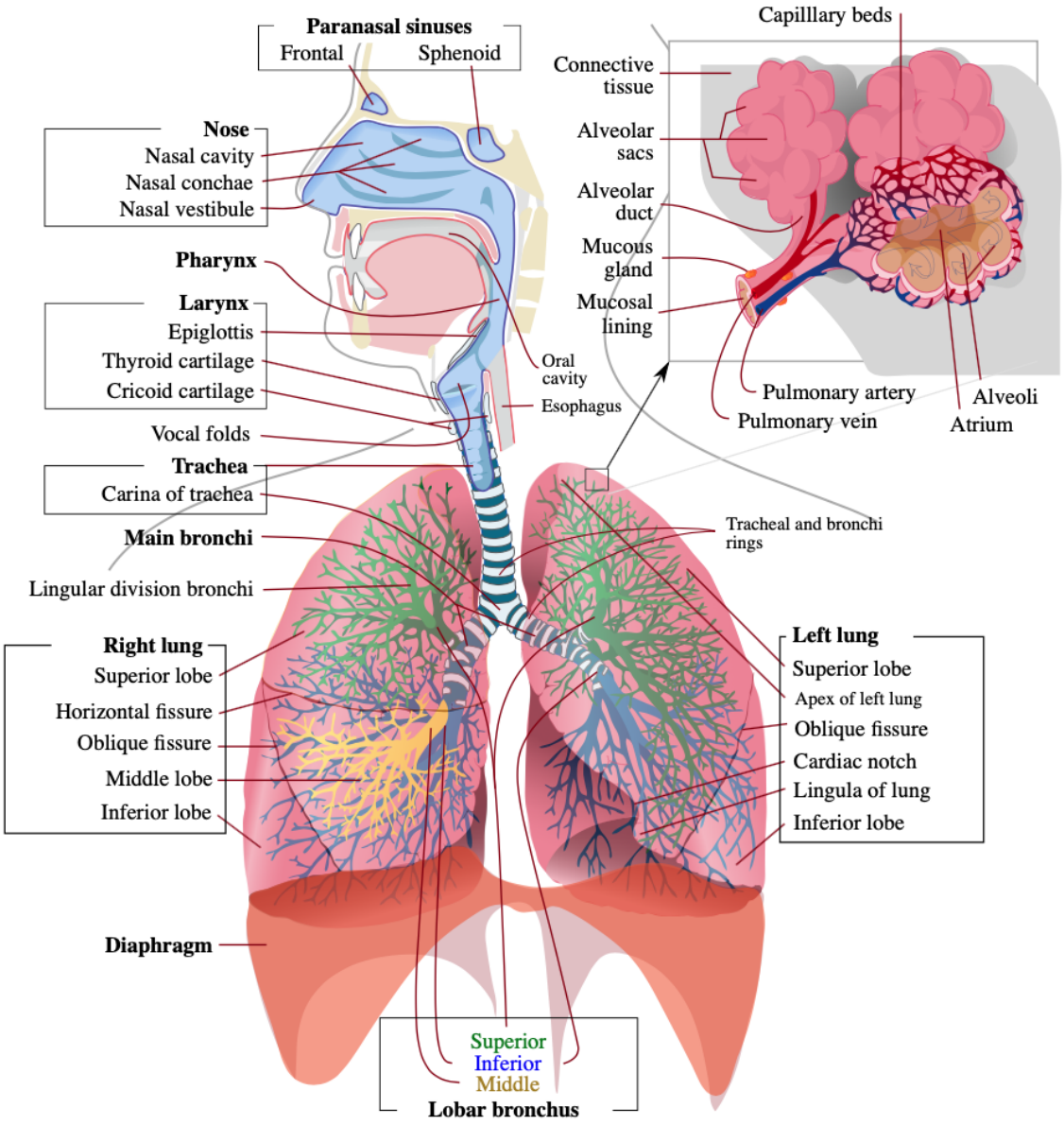


Figure 1.2: Anatomy of the human pulmonary system [Wikimedia, 2021].

Pulmonary Airway

The airway structure is composed of both upper and lower parts. The upper airway lies outside the thorax. It starts from the oral and nasal cavities, leading to the trachea. The lower airway consists of trachea, bronchi, and distal bronchioles. In the present study, the pulmonary airway refers to the lower airway that is of clinical interest in pulmonary CT images. Since the lower airway has a fractal like branching pattern for air supply [Weibel, 2009], it is also called the airway tree. As the airway stretches inside the lungs, it bifurcates into two or more smaller airway branches at each generation. According to the size and position, the hierarchy of airway includes trachea, primary bronchi, secondary (lobar) bronchi, tertiary (segmental) bronchi, bronchioles, and alveoli. The human airway has, on average, 23 generations of dichotomous branching [Weibel and Gomez, 1962]. Illustration of the generations of branches in an airway tree is given in Fig. 1.3.

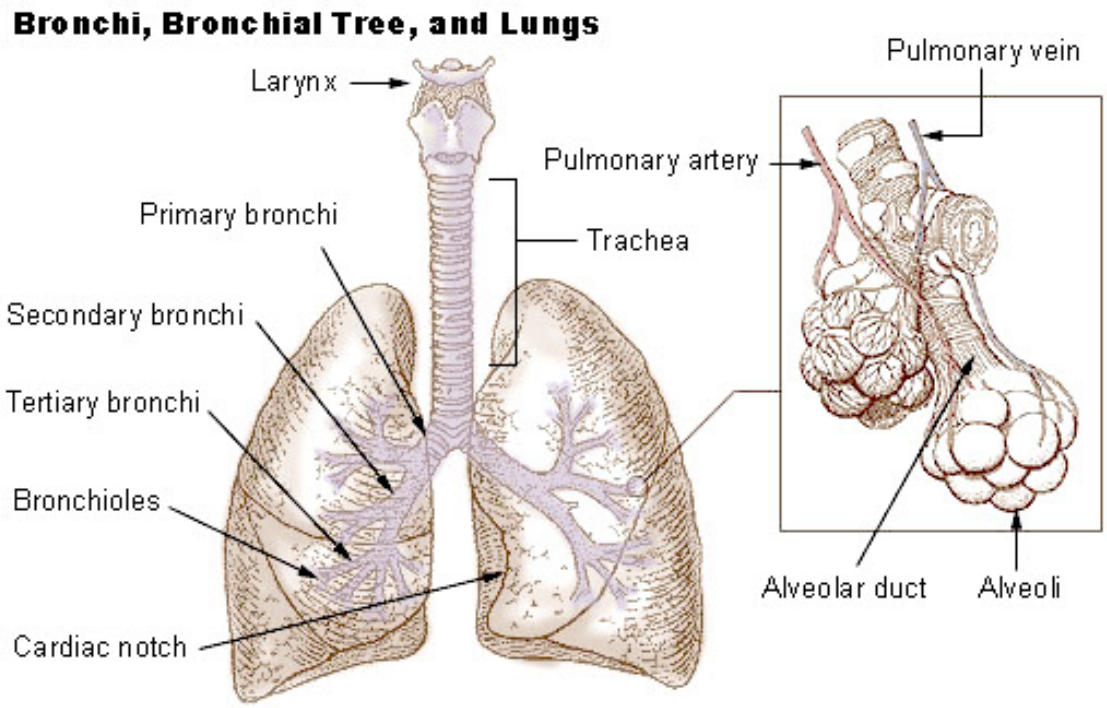


Figure 1.3: Generations of branches in a typical human airway tree [U. S. National Institutes of Health, 2021].

Pulmonary Vessel

Around the lungs, there exist two separate circulation systems: the high-pressure bronchial circulation and the low-pressure pulmonary circulation [Peacock et al., 2016, West, 2011]. The bronchial circulation supplies oxygenated blood to airway and lung parenchyma while the pulmonary circulation carries the deoxygenated blood from the right ventricle to lung and returns oxygenated blood to the left atrium. The vessels of pulmonary circulation are the pulmonary arteries and veins [Kandathil and Chamarthy,

2018]. Fig. 1.4 illustrates the pulmonary circulation system.

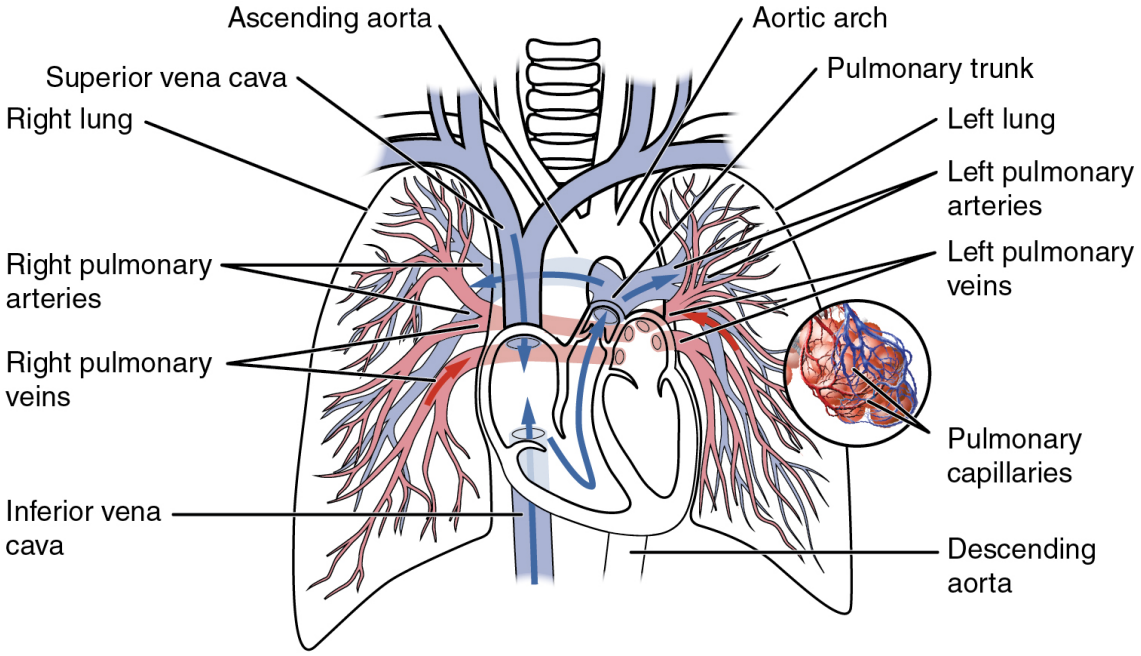


Figure 1.4: Illustration of pulmonary circulation [Wikimedia, 2020].

The pulmonary arteries start as the pulmonary trunk and split into the left and right main arteries. Once inside lungs, the left pulmonary artery divides into two lobar arteries, one for each lobe. In comparison, the right pulmonary artery courses longer and perpendicularly away from the trunk and left pulmonary artery. The right artery later divides into two branches: the upper lobar artery and the descending inter-lobar artery. The upper one supplies the right upper lobe and the descending one supplies the right middle and lower lobe. All pulmonary arteries bifurcate into smaller arterioles and eventually lead to the capillaries.

The pulmonary veins emerge from each lung hilum. In total, there are four main pulmonary veins, with one inferior vein and one superior vein for either side. The right superior vein drains the right upper and middle lobes, and the right inferior vein drains the right lower lobe. The left superior vein drains the left upper lobe and lingula, and the left inferior vein drains the left lower lobe. They transfer the oxygenated blood from lungs to the left atrium.

The close relationship of pulmonary vessels and airways is found throughout the lungs. Pulmonary arteries run in parallel with airways, developing close to each other and decreasing in diameter accordingly. Whereas pulmonary veins follow Miller's dictum, generally being as far away from the airways as possible [Miller, 1947, Hislop, 2002, Moreno et al., 2006]. The distance between pulmonary airway and artery is typically smaller than the distance between airway and vein. Besides, pulmonary arteries and veins have a roughly equivalent number of branches.

1.2.2 Lung Diseases Overview

The category and classification of lung diseases are various. Due to the complicated pathogenetic mechanisms, there exist many causes to certain lung lesions and diseases. Likewise, one lesion or disease may have multiple symptoms such as cough, chest pain, and shortness of breath. According to the affected pulmonary structures, lung lesions and diseases can be divided into those affecting pulmonary airway, air sac, interstitium, vessel, pleura, and chest wall [Mason et al., 2010]. It is noted that some diseases may affect multiple structures as well. In this section, we mainly introduce diseases that affect airway, air sac, and vessel.

Lung disease affecting airway

- Asthma: Chronic inflammation occurs around the conducting zone of airways, resulting in increased contractability of muscles and the narrowing of airways. Usually the airway obstruction in asthma is reversible, however, if left untreated, irreversible damage may occur due to airway remodeling. Allergies, infections, pollution, and genetic interactions can trigger the symptoms.
- Chronic obstructive pulmonary disease (COPD): COPD is characterized by persistent airflow limitation caused by airway inflammation and emphysematous destruction of lung tissue. COPD is a progressive disease where small airways become narrow and lung tissue break down over time.
- Emphysema: One typical type of COPD where Air-filled cavities in the lung are detected. It is characterized by the difficulty of blowing air out.
- Chronic bronchitis: Another typical type of COPD which is characterized by productive cough and inflammation of bronchi.
- Acute bronchitis: The short-term inflammation of the bronchi with symptom of cough. Most of the cases is caused by viral infection.
- Cystic fibrosis: A genetic disorder that causes poor clearance of mucus from bronchi, clogging small airways.

Lung disease affecting air sac

- Pneumonia: An inflammatory condition of the lung that primarily affects small air sacs.
- Tuberculosis: A slow progressive, infectious disease caused by Mycobacterium tuberculosis (MTB) bacteria.
- Pulmonary edema: Fluid leaks out of small vessels into the air sacs and air spaces of lungs.
- Lung cancer/nodule: Lung cancer may develop in any part of lungs and is most often found near air sacs. Pulmonary nodules are small focal lesions inside lungs. Some malignant nodules may grow into caner.

- Acute respiratory distress syndrome (ARDS): Severe injury to lungs caused by illnesses. It is a type of respiratory failure characterized by widespread and rapid inflammation of lungs.

Lung disease affecting vessel

- Pulmonary embolism (PE): A blockage of an artery in lungs. The blood clots in the leg or other body part travel to the lungs, causing shortness of breath and low blood oxygen levels.
- Pulmonary hypertension (PH): A pathophysiologic condition of increased blood pressure within pulmonary arteries. Its causes are unclear and no cure is available now.
- Arteriovenous malformations (AVM): An abnormal structural connection between arteries and veins, bypassing the capillary system. It leads to a right-to-left blood shunt and thereafter gives rise to serious complications such as hemorrhage and infection.

1.2.3 Segmentation Methods Overview

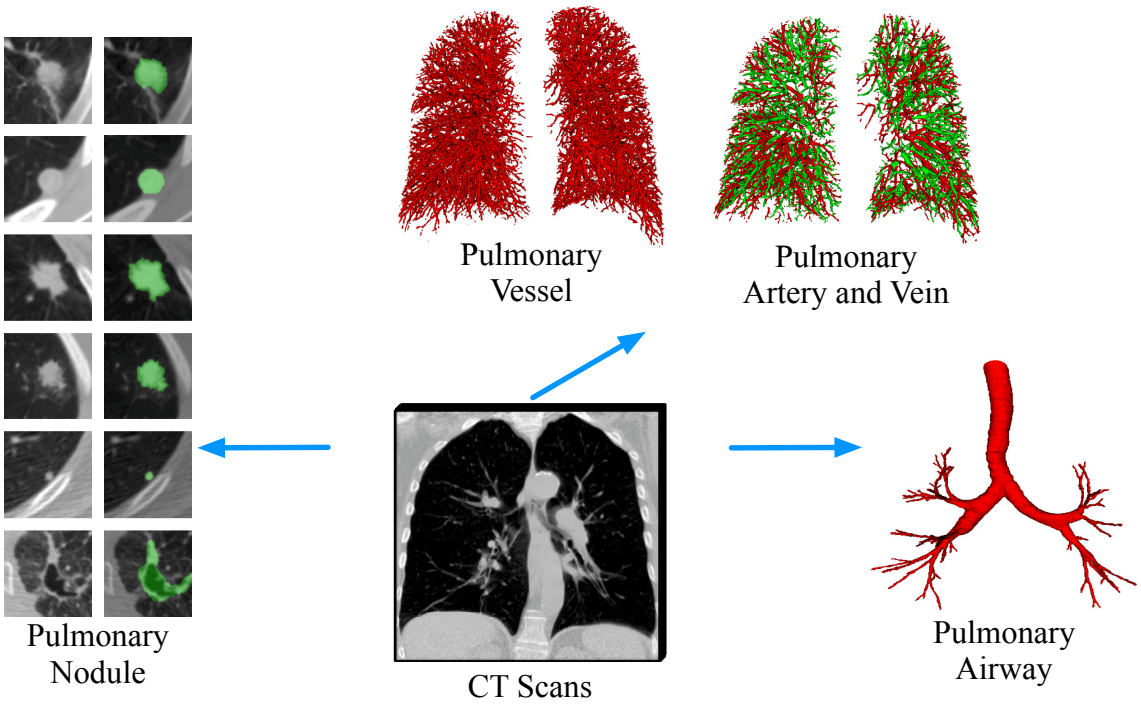


Figure 1.5: Segmentation of pulmonary nodule, airway, vessel (artery and vein) in CT.

Due to the development of CT hardware and software (e.g., high-resolution CT, low-dose CT), chest CT has been widely adopted as the “de facto" pulmonary imaging modality [Beutel et al., 2000]. It provides multi-planar images of the entire chest, where lung

parenchyma, airway, vessel, as well as abnormal findings such as nodules are clearly visible. To help radiologists better analyze pulmonary CT images, a variety of segmentation methods have been developed for key lung structures and lesions of interest. Automatic segmentation of these objects in CT makes it accurate and efficient for quantitative measurement. Subsequently, features computed on the segmented objects benefit disease diagnosis. In this section, segmentation methods for pulmonary airway, vessel, and nodule (see Fig. 1.5) are presented as follows.

Pulmonary Airway Segmentation

Airway segmentation is a key step in the analysis of lung diseases affecting airway. It helps to measure airway size, shape, and wall thickness to quantify the degree of airway narrowing in COPD patients [Wiemker et al., 2004]. In addition, it is required to extract patient-specific airway models from CT for bronchoscopic navigation [Mori et al., 2000].

In chest CT, the intensity of airway lumen is usually lower than that of airway wall. The methods based on region-growing are widely used to extract airway lumen. Threshold-based region-growing methods [Kuhnigk et al., 2005, Zhou et al., 2006, Lassen et al., 2010, Ukil and Reinhardt, 2008] produce satisfactory results for extracting the trachea and main bronchi. However, the intensity between airway lumen and wall becomes smaller as the airway extends towards peripheral branches. Image artifacts such as partial volume effect (PVE) and noise can cause breakage or holes in airway wall, leading to under-segmentation or leakage into lung parenchyma. In order to extract more smaller airways while preventing leakage, an improved region-growing method for airway segmentation has been developed. There are mainly two strategies for alleviating the problem: leakage removal processing and incorporation of rich image features.

Heuristic rules have been studied for preventing leakage, which are developed according to the geometric characteristics of airway. Kiraly et al. [Kiraly et al., 2002] combined adaptive region growing with morphological operations, and evaluated the effectiveness of airway segmentation by the volume of segmentation results. In addition, a front-propagation method was proposed [Schlathoelter et al., 2002] to detect any leakage. Van Ginneken et al. [van Ginneken et al., 2008] adopted wavefront propagation and region-growing to extract the centerline and skeleton of airway at the same time. Mayer et al. [Mayer et al., 2004] proposed a multi-stage strategy for segmenting airways, in which region growth, two-dimensional (2-D) wave propagation, and template matching were used according to different diameter ranges of airway. Moreover, Kitasaka et al. [Kitasaka et al., 2003] detected leakage by analyzing the cross-sectional area segmented airway branches. Similarly, Graham et al. [Graham et al., 2010] generated initial airway segmentation by region-growing and analyzed the cross-sectional area to determine the location of true positive airway branches. Finally, they obtained the best airway tree by using a graph partitioning method to minimize the cost of linking all disconnected airway components. Tschirren et al. [Tschirren et al., 2005] refined the segmented airway tree by first calculating the skeleton with thinning algorithm and then extracting topological features.

Rich image features other than CT intensity have been explored to distinguish between airway lumen and other pulmonary structures. For example, an airway tubular enhancement filtering technique has been designed [Lassen et al., 2012] to enhance

the edge of airway, thereby improving the results of region-growing. In addition, the AdaBoost classifier was developed to distinguish multiple scales of airway [Ochs et al., 2007]. Lo et al. [Lo and de Bruijne, 2008] used the k-nearest neighbor (k-NN) algorithm with multi-scale local image descriptors to distinguish airway from surrounding tissues. Lo et al. [Lo et al., 2010b] used the positional relationship between pulmonary vessels and bronchi as anatomical guidance for airway segmentation. They assumed that airway branches should have a similar orientation to their adjacent vessels. A combination of techniques including appearance model of airway, vessel orientation constraints, and region-growing is exploited to improve the performance.

Pulmonary Vessel Segmentation

Pulmonary vessel extraction is an important step in vessel volume quantification and lung disease diagnosis. In consideration of the tubular structure property and high CT intensity of vessels, many features were manually designed for pulmonary vessel segmentation. The existing methods can be generally summarized into four categories: thresholding [Fetita et al., 2009, Lassen et al., 2012], Hessian-based filtering [Frangi et al., 1998, Krissian et al., 2000, Aylward and Bullitt, 2002, Agam et al., 2005, Zhou et al., 2007], region-growing [Metz et al., 2007, Bulow et al., 2004, Shikata et al., 2009, Zhou et al., 2012], and learning-based methods [Ochs et al., 2007, Korfiatis et al., 2011].

Given chest CT of healthy patients, thresholding is a simple yet effective method to extract pulmonary vessels [Lassen et al., 2012]. Frangi's vessel filtering [Frangi et al., 1998] enhances tubular structures in 3-D CT. By changing the judging criterion, the same filtering method can be used for airway segmentation. Morphological operations [Agam et al., 2005] or connected component analysis [Zhou et al., 2007] are employed on the results of vessel filtering to reconstruct the vessel tree. However, the detected vessels are prone to discontinuity due to weak intensity. Besides, lung lesions such as mucus-filled bronchial tubes, nodules may share similar intensity and shape with vessels. False positives are often observed in these methods. In order to alleviate the shortcomings of Hessian-based filtering, post-processing methods have been proposed to refine the initial vessel segmentation results. van Dongen et al. [van Dongen and van Ginneken, 2010] applied a thinning algorithm to obtain the centerline of segmented vessels and applied local thresholding along the skeleton for refinement. Zhou et al. [Zhou et al., 2007] designed a response function that enhances bifurcation and suppresses non-vascular structures. Ochs et al. [Ochs et al., 2007] designed eigenvalue-based features to perform voxel-wise classification with Adaboost. Korfiatis et al. [Korfiatis et al., 2011] first generated vessel candidates by Hessian-based filtering, and then refined the candidates using an SVM classifier with 3D co-occurring texture features. Region-growing based methods were also popular in vessel segmentation. Results of vessel filtering were used to initialize the seed points for region-growing. Bulow et al. [Bulow et al., 2004] employed a fast marching wave propagation approach where bifurcations and termination of vessels are checked based on connectivity. Leakage was detected by measuring the diameter of the wave front. Shikata et al. [Shikata et al., 2009] proposed a vessel traversal approach where the vessel trajectories are tracked and linked at the nearest bifurcation for complete vessel reconstruction. Lo et al. and Zhou et al. [Lo et al., 2010b, Zhou et al., 2012] both first generated initial vessel candidates with filtering and then linked independent vessel

components with Dijkstra algorithm [Dijkstra, 1959].

Pulmonary Nodule Segmentation

Pulmonary nodule segmentation is required for computer-aided diagnosis of malignant lung nodules. The segmentation of nodules has always been a challenging task for the following reasons: 1) Low contrast between nodule border and background often exists, especially for non-solid and part-solid nodules. Noise, artifact, and equipment differences also degrade the quality of CT acquisition. 2) The appearance of pulmonary nodules varies greatly from person to person. The distribution of nodules is imbalanced in terms of size, shape, and intensity. 3) The pulmonary structures are complicated. The intensity similarity between certain adjacent structures (e.g., pleura, vessel, airway wall) and nodules increase the difficulty of accurately delineating boundaries of nodules. To address these challenges, several nodule segmentation methods have been proposed and can be mainly classified into five categories: thresholding [Reeves et al., 2006, Ye et al., 2009], region-growing [Dehmeshki et al., 2008, Kubota et al., 2011, Gu et al., 2013], morphology-based methods [Kostis et al., 2003, Kuhnigk et al., 2006, Diciotti et al., 2011, Setio et al., 2015], active contour models [Awad et al., 2012, Farag et al., 2013, Farhangi et al., 2017, Alilou et al., 2017], and learning-based methods [Ciompi et al., 2017, Wang et al., 2017, Wu et al., 2018a].

For thresholding-based methods, Reeves et al. [Reeves et al., 2006] proposed an adaptive thresholding method to segment pulmonary nodules from the lung parenchyma. They used the histogram analysis method to calculate the average intensity values of nodules and lung parenchyma, respectively, and then obtained the adaptive threshold. For solid nodules with clear boundaries and uniform intensity distribution, the thresholding-based methods may achieve good results. However, their performance dropped largely for other intricate situations. Therefore, thresholding is generally only used as initial segmentation method, or incorporated with other segmentation algorithms. Region-growing is another widely used method. Dehmeshki et al. [Dehmeshki et al., 2008] proposed a region-growing algorithm based on contrast and fuzzy connectivity. Kubota et al. [Kubota et al., 2011] used the Euclidean distance transformation to calculate the distance map, and then applied region-growing on the Euclidean distance map to separate nodules from their surrounding pulmonary structures. Gu et al. [Gu et al., 2013] proposed a "click and grow" region-growing method for interactive nodule segmentation. Kostis et al. [Kostis et al., 2003] used a morphological operation with fixed-size structuring elements to distinguish nodules from vessels. Kuhnigk et al. [Kuhnigk et al., 2006] developed an adaptive structuring element for morphological operation to segment nodules. Diciotti et al. [Diciotti et al., 2011] applied morphological erosion and dilation to refine the segmented pulmonary nodules. Setio et al. [Setio et al., 2015] combined both thresholding and morphological operation for detection and segmentation of large nodules. For active contour models, Farag et al. [Farag et al., 2013] proposed a level-set based method to adaptively segment nodules. Farhangi et al. [Farhangi et al., 2017] exploited shape prior knowledge and captured a sparse representation of predefined shapes for given nodules. They updated shape prior over level set evolution to handle various types of nodules. With the development of machine learning, especially deep learning, learning-based methods are designed for nodule segmentation. Ciompi et

al. [Ciompi et al., 2017] proposed a deep learning method using multi-stream and multi-scale convolutional neural networks. Wang et al. [Wang et al., 2017] proposed a central focused convolutional neural network where the central pooling layer can retain a large amount of information around center voxels. Furthermore, a weighted sampling method was developed for efficient model training.

1.3 State-of-the-art Deep Learning Methods

The goal of the present thesis is to develop accurate and reliable classification and segmentation methods for chromosome and pulmonary images with the help of deep learning technology. Convolutional Neural Network (CNN), as one of the most important implementation forms of deep learning [LeCun et al., 2015], is the technical approach that the present work mainly relies on. In the light of this statement, it is indispensable to introduce related background knowledge in this section.

1.3.1 Artificial Neural Networks

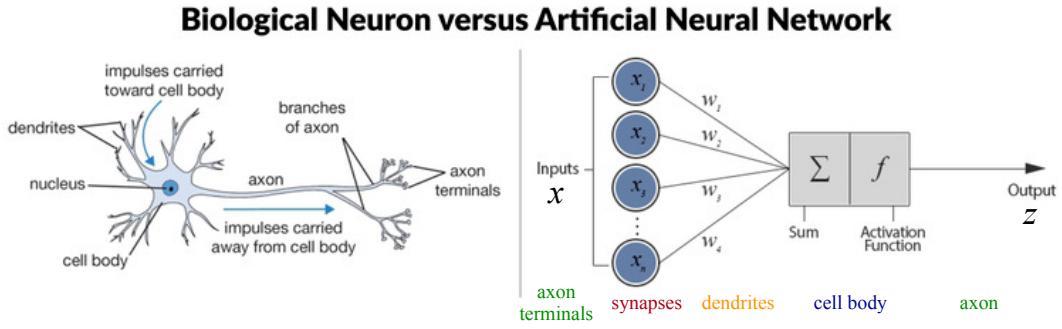


Figure 1.6: The artificial neuron model (Figure inspired by [Willems, 2019]).

In 1943, McCulloch and Pitts [McCulloch and Pitts, 1943] designed the earliest artificial neuron model by simulating the structure of brain neurons (see Fig. 1.6). Given an input signal x, the final output z of a neuron is expressed as:

$$z = f(\sum_i w_i x_i + b) \tag{1.1}$$

where x_i represents the i-th input of the neuron, w_i represents the weight assigned by the neuron to x_i, and b is the bias of the neuron, f is the activation function of the neuron. Since the step function is used as activation function at that time, it is difficult to effectively train the multi-layer network that is built based on such artificial neuron models. Therefore, the subsequent research on artificial neural networks stagnated for quite a long while.

In 1986, Rumelhart et al. [Rumelhart et al., 1986] replaced the original step function by the non-linear differentiable sigmoid function ($f(x) = \frac{1}{1+e^{-x}}$) as activation function.

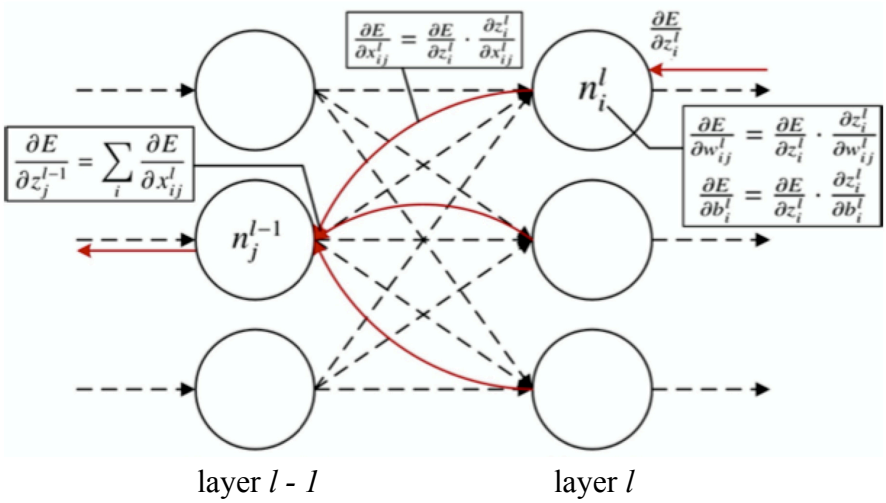


Figure 1.7: Illustration of back-propagation algorithm.

Besides, a back-propagation (BP) algorithm was proposed for supervised learning of networks. This not only laid the foundation for multi-layer network optimization, but also opened a new era of artificial neural network (ANN) research. The back-propagation algorithm applies the chain derivation rule, and the detailed process is shown in Fig. 1.7. For the i-th neuron n_i^l in the l-th layer, if the derivative of prediction error E of the entire network with respect to its output z_i^l is known as $\frac{\partial E}{\partial z_i^l}$, then the derivative of E with respect to the internal parameters of the artificial neuron can be expressed as:

$$\frac{\partial E}{\partial w_{ij}^l} = \frac{\partial E}{\partial z_i^l} \frac{\partial z_i^l}{\partial w_{ij}^l}, \tag{1.2}$$

$$\frac{\partial E}{\partial b_i^l} = \frac{\partial E}{\partial z_i^l} \frac{\partial z_i^l}{\partial b_i^l}, \tag{1.3}$$

where w_{ij}^l is the weight between the neuron n_i^l and the j-th neuron n_j^{l-1} of the previous layer $l-1$, b_i^l is the bias. For the neuron n_j^{l-1}, the derivative of E with respect to its output z_j^{l-1} is the sum of feedback of all neurons in the l-th layer that are connected to n_j^{l-1}:

$$\frac{\partial E}{\partial x_{ij}^l} = \frac{\partial E}{\partial z_i^l} \frac{\partial z_i^l}{\partial x_{ij}^l}, \tag{1.4}$$

$$\frac{\partial E}{\partial z_j^{l-1}} = \sum_i \frac{\partial E}{\partial x_{ij}^l}, \tag{1.5}$$

where the output of the neuron n_j^{l-1} is exactly the input to the neuron n_i^l: $x_{ij}^l = z^{l-1}j$. By recursively using the chain rule of derivation from the end of network layer to the front, the derivative of the error E with respect to the parameters of each layer can be obtained for any given network. Finally, the parameters of the network can be updated by gradient descent to minimize E, which fulfills the optimization of the entire network.

1.3.2 Convolutional Neural Networks

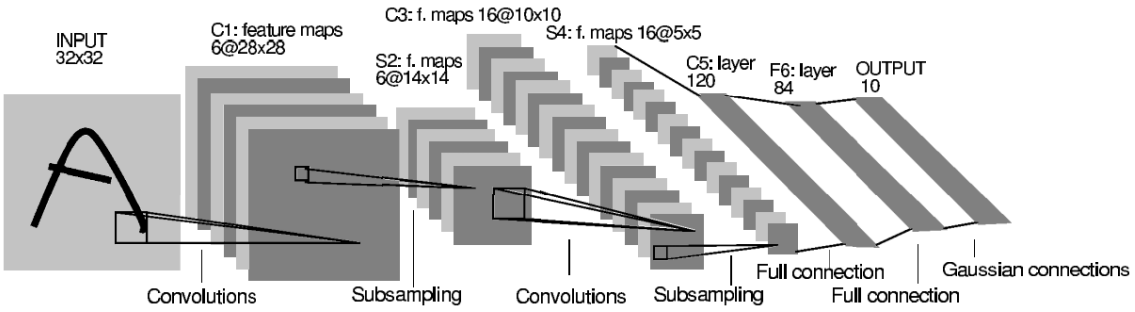


Figure 1.8: Architecture of the LeNet-5 model [LeCun et al., 1998].

In 1998, LeCun et al. [LeCun et al., 1998] proposed the LeNet-5 model and successfully applied it to the recognition of handwritten digits. This model is composed of many kinds of network layers (see Fig. 1.8), including convolutional layer, down-sampling layer (a.k.a. pooling layer) and fully connected layer. It is one of the most representative convolutional neural networks in early stage. Its characteristic is to use 2 convolutional layers and 3 fully connected layers as the main learning units of the network. Besides, it successfully applies the spatial weights sharing and reusing characteristics of convolution and down-sampling operations to reduce the overall computational complexity of the network.

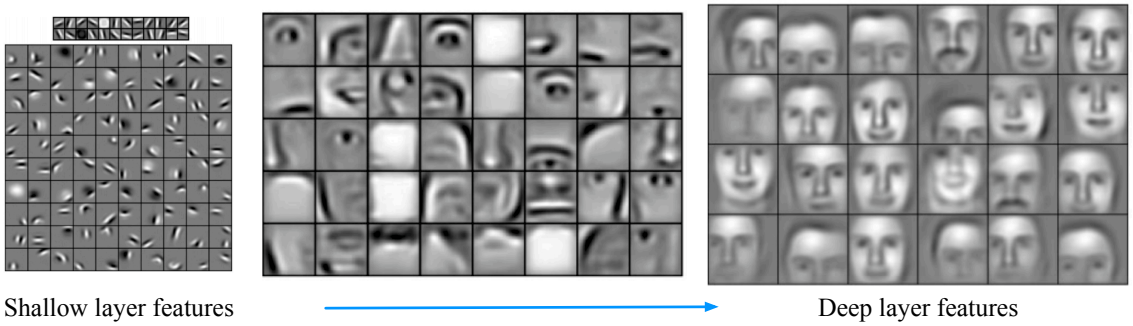


Figure 1.9: Hierarchical features of CNNs [Lee et al., 2011].

In 2011, Lee et al. [Lee et al., 2011] used the generative network — restricted Boltzmann machine (RBM) to visualize the stacked convolutional layers. It provides a more intuitive understanding of the hierarchical feature extraction process of the convolutional neural network (see Fig. 1.9).

In 2012, Krizhevsky et al. [Krizhevsky et al., 2012] developed the AlexNet (see Fig. 1.10), which significantly surpassed traditional machine learning methods in the ImageNet [Deng et al., 2009] image classification challenge. It officially opened the prelude to the booming development of deep learning methods in the field of computer vision. Compared with the LeNet-5 model, AlexNet has the following outstanding features:

- It has much deeper network layers. A larger convolutional kernel (11×11) is used

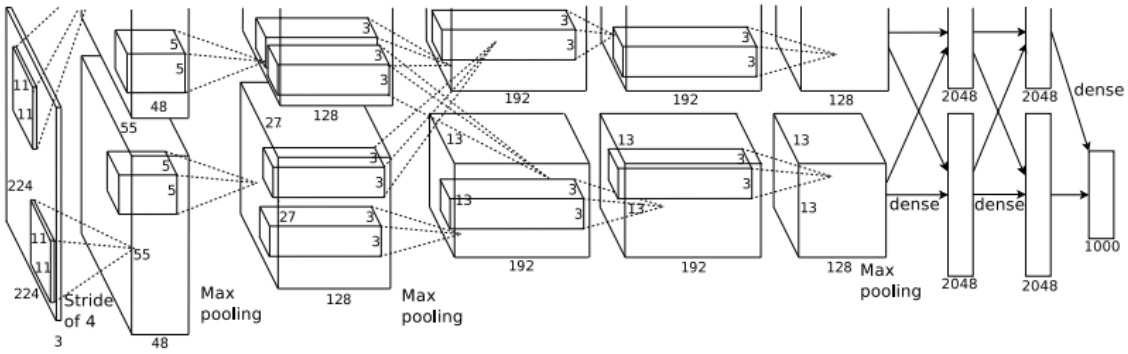


Figure 1.10: Architecture of the AlexNet model [Krizhevsky et al., 2012].

in shallow convolution layers to learn low-level features of natural images.

- The rectified linear unit (ReLU) is used as activation function to replace the $tanh(\cdot)$ function in LeNet-5, which improves the computational efficiency during training and at the same time alleviates the vanishing gradient problem in deep networks.
- It uses local response normalization (LRN) layer to establish a competition mechanism of activation response between local neurons and enhance the generalization ability of the model.
- It uses dropout [Srivastava et al., 2014] to avoid overfitting of the model and achieve an effect similar to ensemble learning.

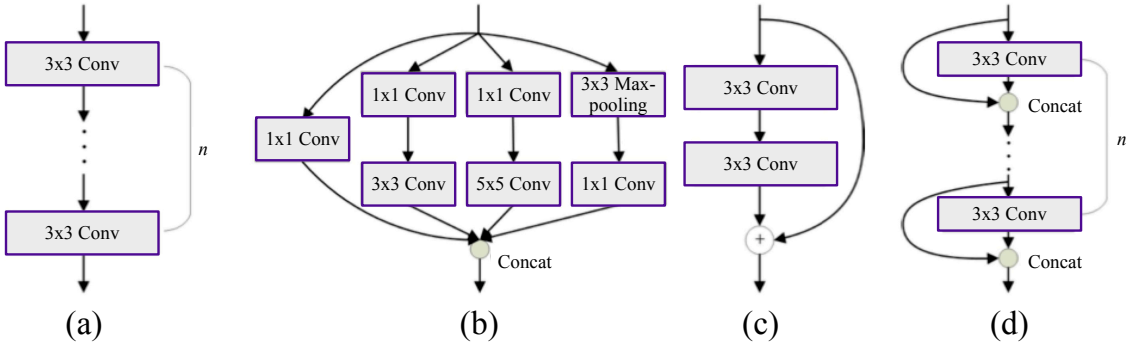


Figure 1.11: Convolution blocks in different CNNs. (a) VGG block (b) GoogLeNet Inception block (c) Residual block (d) Densely connected block

In 2014, Simonyan and Zisserman [Simonyan and Zisserman, 2014] proposed the VGG network with a depth of up to 19 layers. They used convolution blocks in VGG to obtain the same receptive field as the single convolution kernel in LeNet-5 or AlexNet. Each convolutional block contains multiple convolutional layers with kernel size of 3×3 (see Fig. 1.11(a)). The depth of VGG is increased with such blocks. Later, Szegedy et al. [Szegedy et al., 2016] proposed the GoogLeNet with a depth of 22 layers. They used

the inception structure (see Fig. 1.11(b)), which includes multi-path convolution and max-pooling for diversity and sparsity of feature extraction. The concatenation operation is performed to fuse features from different paths. In order to reduce computational complexity, the inception structure further adopts a 1×1 convolution to reduce the number of feature channels. The GoogLeNet increases the network depth while restricting the number of total parameters, reducing the risk of overfitting.

As the depth of network increases, the difficulty of network training also increases. In 2015, Ioffe and Szegedy [Ioffe and Szegedy, 2015] proposed batch normalization (BN), which is used to solve the problem of large deviations between the input and output distribution of each layer in deep networks. Usually, the normalization layer is placed before the activation function, immediately following the convolution layer. Its effect is similar to a regularizer that corrects the output distribution of convolution, allowing deep networks to be trained with a large learning rate.

In 2016, He et al. [He et al., 2016a] developed the method of residual learning and proposed the ResNet with a depth of up to 152 layers. The ResNet employs a direct shortcut connection in each residual block to achieve the identity transformation of features. It also uses parallel convolution layers for non-linear transformation of the residuals (see Fig. 1.11(c)). Such architecture can better handle the subtle changes of features between different layers of the network, and therefore achieved better performance. Subsequently, modifications on the residual architecture have been conducted to further increase the depth and width of ResNet [He et al., 2016b, Zagoruyko and Komodakis, 2016].

Recently, Huang [Huang et al., 2017a] found that there still exists a large number of redundant features in ResNet. It may cause the vanishing gradient problem in network training. Therefore, based on the idea of feature reusing, they proposed the DenseNet model in 2017. The DenseNet makes extensive use of dense blocks that are composed of multiple convolution layers. In each dense block, the input of one convolution layer is the concatenation of output features from all previous convolution layers (see Fig. 1.11(d)). Such design not only reutilizes all previous features, but also realizes multi-layer feature fusion. In addition, with the benefit of feature reusing, the number of parameters of each dense block can be reduced to avoid overfitting. With only 100 layers of depth, the DenseNet surpassed ResNet-1001 in classification performance. However, due to the huge number of dense connections within dense blocks, special optimization techniques are needed to reduce memory consumption during the training process. So far, many deep learning frameworks are still in lack of solutions to memory-efficient and light-weight implementation.

1.3.3 CNNs Architectures Overview

Similar to the development of convolution blocks, the development of network architecture also plays an important role in applying convolutional neural networks for various computer vision tasks. In addition to the single link path architecture, researchers have developed several important network architectures such as bypass transmission architecture, parallel architecture, and multi-task branch architecture (see Fig. 1.12).

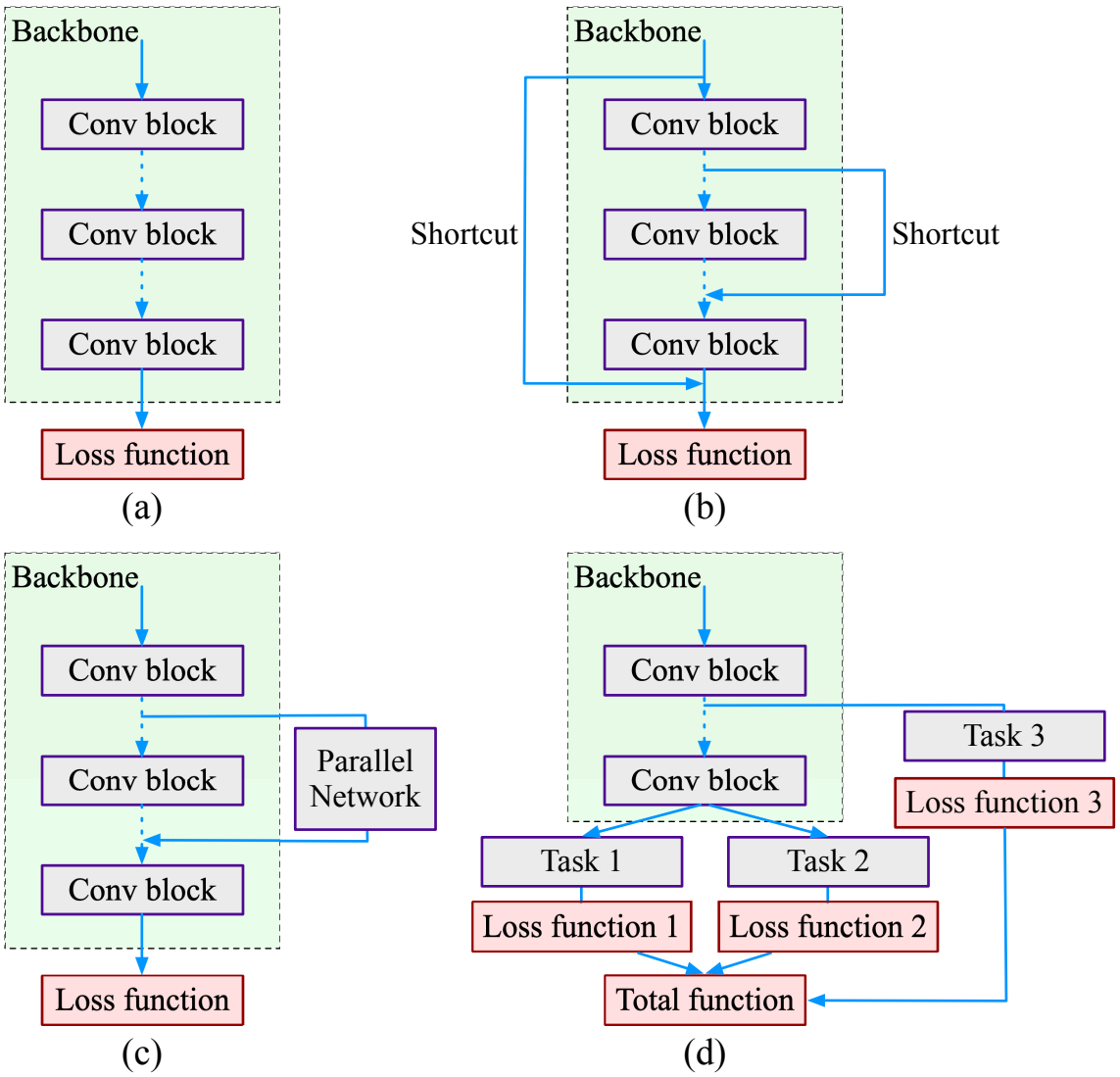


Figure 1.12: Different CNN architectures. (a) Single link path architecture (b) Bypass transmission architecture (c) Parallel architecture (d) Multi-task branch architecture

Single Link Path Architecture

The characteristic of the single link path architecture is that each component block of the network is connected sequentially by a unique link as a backbone. Such architecture is consistent with the current hardware stream processing pipeline. Therefore, it can be trained efficiently and deployed easily. Several typical representative networks with single link path architecture are LeNet-5, AlexNet, and VGG for image classification.

Bypass Transmission Architecture

The characteristic of the bypass transmission architecture is to use long-distance skip connections to effectively transfer information from lower layers to higher layers of the network, or even to the loss function layer. Such skip connection not only merges low-level and high-level features, but also alleviates the problem of vanishing gradient during training. Typical representatives of this architecture are the HED [Xie and Tu, 2015] for edge detection, FCN [Long et al., 2015] and U-Net [Ronneberger et al., 2015] for semantic segmentation. Among them, HED uses skip connections to directly connect different levels of features to the edge detection classifier for loss computation. During back-propagation, errors can be directly transmitted to lower layers of the network through skip connections for parameter update. FCN also uses the same strategy for image segmentation. U-Net further exploits the role of skip connection, fusing low-level and high-level features at each resolution scale for accurate pixel-wise segmentation.

Parallel Architecture

The characteristic of parallel architecture is to attach sub-networks in parallel with the backbone network. Such sub-networks control the information flow of the backbone with specific purpose. One typical example is the spatial transformer network (STN) [Jaderberg et al., 2015]. It enables the entire network to respond according to the variations of targets in posture and position. Usually, the ability of CNNs to learn spatial-invariant and posture-invariant features is limited. To alleviate this problem, data augmentation methods such as random shifting, scaling, rotation, and elastic wrapping transform are often applied on input images. The STN provides another solution by adding a localization sub-network in parallel with the backbone. It estimates the deformation parameters of the input object and enables the model to recognize various spatial transformations of the object (e.g., shifting, cropping, rotation, shearing, and scaling). Furthermore, based on the estimated deformation parameters, it uses a differentiable spatial sampling algorithm to correct features of the backbone network. The localization sub-network can be optimized simultaneously with the backbone network.

Multi-task Branch Architecture

The characteristic of the multi-task branch architecture is to introduce multiple prediction heads or sub-networks to the backbone network for different tasks. Such architecture is designed for the joint feature learning of closely-related tasks. By sharing the common backbone network as much as possible between different tasks, it saves computing resources and reduces the labor of designing networks independently for each task. Be-

sides, multi-task learning helps extract effective and discriminative features because it introduces information from different tasks to the backbone network [Caruana, 1997]. Complementary knowledge learned from multiple tasks may improve the model's performance on each task. Several representative architectures are Fast R-CNN [Girshick, 2015], Faster R-CNN [Ren et al., 2015] for object detection and recognition, and Mask R-CNN [He et al., 2017] for instance segmentation.

1.3.4 Generative Adversarial Networks

The generative adversarial networks (GANs) [Goodfellow et al., 2014] are differentiable generative models with adversarial training strategies. A typical GAN has two component networks:

- Generator. The generator learns to produce plausible samples that share similar distribution with target data: $\boldsymbol{x} = G(\boldsymbol{z};\boldsymbol{\theta}^{(G)})$, where $\boldsymbol{z}$ is the random noise vector sampled from a prior distribution, and $\boldsymbol{x}$ is the generated sample.
- Discriminator. The discriminator attempts to distinguish between samples drawn from real training data and samples produced by the generator. It emits a probability value $d(\boldsymbol{x};\boldsymbol{\theta}^{(D)})$ which indicates the probability that $\boldsymbol{x}$ is a real training sample instead of a fake synthetic sample. The discriminator penalizes the generator for producing implausible results.

The optimization of GANs adopts the idea of zero-sum game, where the generator and discriminator evolve in an adversarial way. A function $v(\boldsymbol{\theta}^{(G)},\boldsymbol{\theta}^{(D)})$ determines the payoff of the discriminator. The generator receives $-v(\boldsymbol{\theta}^{(G)},\boldsymbol{\theta}^{(D)})$ as its own payoff. During adversarial learning, each player tries to maximize its own payoff until convergence:

$$G^*, D^* = \arg\min_G \max_D v(\boldsymbol{\theta}^{(G)},\boldsymbol{\theta}^{(D)}) \tag{1.6}$$

The default choice for v is formulated as binary cross entropy:

$$v(\boldsymbol{\theta}^{(G)},\boldsymbol{\theta}^{(D)}) = \mathbb{E}_{\boldsymbol{x}\sim p_{data}} \log d(\boldsymbol{x}) + \mathbb{E}_{\boldsymbol{x}\sim p_{model}} \log(1 - d(\boldsymbol{x})) \tag{1.7}$$

The loss above drives the discriminator to learn to classify generated samples as fake and real training samples as real. On the contrary, the generator learns to produce samples as real as possible, fooling the discriminator into believing its samples are real. After convergence, samples from the generator are indistinguishable from real data, and the discriminator outputs the probability of $\frac{1}{2}$ for both real and fake samples. The training process of GAN is illustrated in Fig. 1.13.

1.3.5 Optimization Algorithms Overview

With the increasing number of training data and learnable network parameters, it becomes more and more difficult to train convolutional neural networks. In order to improve the efficiency of network training and performance of deep networks, researchers have proposed several optimization algorithms including the stochastic gradient descent (SGD), SGD with momentum, SGD with Nesterov momentum, AdaGrad, RMSProp, and Adam [Goodfellow et al., 2016]. Among them, SGD, SGD with momentum, and Adam algorithms are most widely adopted in previous studies.

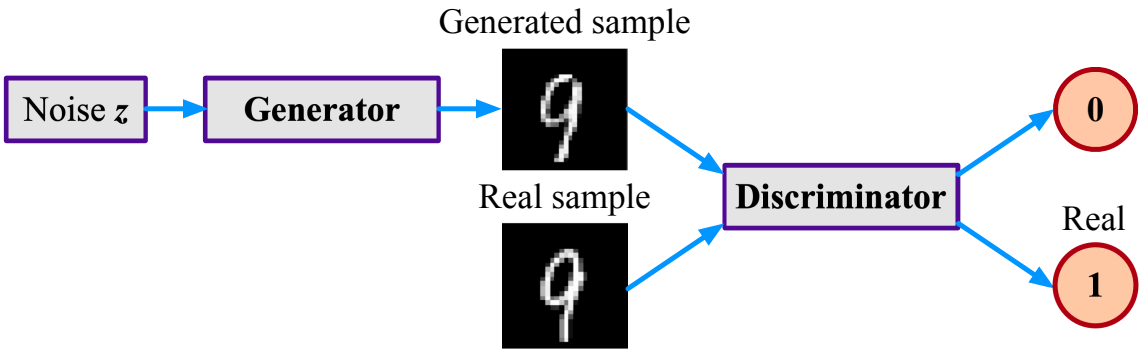


Figure 1.13: The training of GAN proceeds by alternatively training the generator and the discriminator.

Stochastic Gradient Descent

The stochastic gradient descent algorithm randomly selects part of the training samples in each iteration for loss computation, and updates the network parameters by the average gradients calculated from the loss. This algorithm avoids traversal of all training samples, reducing the memory consumption. In practice, it is noted that the statistics of samples in each training batch should be kept similar to those of the entire training set. Otherwise, the oscillation caused by inconsistent gradient orientation may exert negative effects on convergence. The optimization process of SGD is given in Alg. 1.

Algorithm 1 Stochastic gradient descent (SGD)

Input: Learning rate η.
Input: Initial parameter $\boldsymbol{\theta}$.
1: **while** Stopping criterion is not met **do**
2: Sample a mini-batch of m examples from all training samples $\{x^{(1)},...,x^{(m)}\}$.
3: Compute gradient estimate: $\boldsymbol{g} = \frac{1}{m}\nabla_{\boldsymbol{\theta}}\sum_i L(f(\boldsymbol{x}^{(i)};\boldsymbol{\theta}),\boldsymbol{y}^{(i)})$
4: Apply update: $\boldsymbol{\theta} = \boldsymbol{\theta} - \eta\boldsymbol{g}$
5: **end while**
6: **return** Updated parameter $\boldsymbol{\theta}$.

SGD with Momentum

One disadvantage of SGD is that when the mini-batch size m is small, the distribution of samples between different batches may vary greatly and the statistics of each batch are not in accord with those of all training samples. Such inconsistency will cause a sharp change in the orientation of gradients. In order to stabilize and accelerate the optimization process, the momentum algorithm was proposed on the basis of SGD. It simulates the physical law of moving objects (inertia) by accumulating gradients in the past in an exponential decay way. At each iteration, both the current gradient and the accumulated gradients of all previous iterations are used for parameter update. Compared with SGD, the SGD with momentum algorithm changed smoothly in gradient orientation. The optimization process of momentum is given in Alg. 2.

Algorithm 2 Stochastic gradient descent (SGD) with momentum

Input: Learning rate η, momentum hyper-parameter α.
Input: Initial parameter $\boldsymbol{\theta}$, initial velocity $\boldsymbol{v}$.
while Stopping criterion is not met **do**
  Sample a mini-batch of m examples from all training samples $\{x^{(1)},...,x^{(m)}\}$.
  Compute gradient estimate: $\boldsymbol{g} = \frac{1}{m}\nabla_{\boldsymbol{\theta}} \sum_i L(f(\boldsymbol{x}^{(i)};\boldsymbol{\theta}),\boldsymbol{y}^{(i)})$
  Compute velocity update: $\boldsymbol{v} = \alpha\boldsymbol{v} - \eta\boldsymbol{g}$
  Apply update: $\boldsymbol{\theta} = \boldsymbol{\theta} + \boldsymbol{v}$
end while
return Updated parameter $\boldsymbol{\theta}$.

Adam

The Adam algorithm [Kingma and Ba, 2014] is inspired by the SGD with momentum algorithm. It goes one step further by introducing the estimation of the first-order and second-order moment of the gradient to adaptively adjust the learning rate of each parameter. The Adan optimizer is particularly suitable for training networks with complex architectures (e.g., U-Net [Ronneberger et al., 2015]) or networks that are hard to train (e.g., generative adversarial networks [Goodfellow et al., 2014]). The optimization process of Adam is given in Alg. 3.

Algorithm 3 Adam

Input: Learning rate η, first-order momentum hyper-parameter ρ_1, second-order momentum hyper-parameter ρ_2, constant δ, iteration time t.
Input: Initial parameter $\boldsymbol{\theta}$, initial first-order momentum $\boldsymbol{s}$, initial second-order momentum $\boldsymbol{r}$.
while Stopping criterion is not met **do**
  Sample a mini-batch of m examples from all training samples $\{x^{(1)},...,x^{(m)}\}$.
  Compute gradient estimate: $\boldsymbol{g} = \frac{1}{m}\nabla_{\boldsymbol{\theta}} \sum_i L(f(\boldsymbol{x}^{(i)};\boldsymbol{\theta}),\boldsymbol{y}^{(i)})$
  Estimate first-order momentum: $\boldsymbol{s} = \rho_1\boldsymbol{s} + (1-\rho_1)\boldsymbol{g}$
  Estimate second-order momentum: $\boldsymbol{r} = \rho_2\boldsymbol{r} + (1-\rho_2)\boldsymbol{g} \odot \boldsymbol{g}$
  Calibrate first-order momentum: $\hat{\boldsymbol{s}} = \frac{s}{1-\rho_1^t}$
  Calibrate second-order momentum: $\hat{\boldsymbol{r}} = \frac{r}{1-\rho_2^t}$
  Apply update: $\boldsymbol{\theta} = \boldsymbol{\theta} - \eta\frac{\hat{\boldsymbol{s}}}{\sqrt{\hat{r}}+\delta}$
  $t = t + 1$
end while
return Updated parameter $\boldsymbol{\theta}$.

1.4 Summary

This chapter presents the biomedical background and deep learning techniques related to the thesis topic:

- For the biomedical background, the objective and workflow of karyotyping are well

explained, together with the involved operations such as chromosome separation, classification, enumeration, and diagnosis. Besides, the anatomy of human pulmonary system is illustrated. Segmentation methods of pulmonary airway, vessel, and nodule are introduced for analysis of CT images.

- For the deep learning techniques, the history of neural network development is introduced. Then, reviews on recent CNNs architectures are presented. In addition, we explain the mechanism of adversarial learning via GANs. Finally, three frequently-used optimization algorithms are outlined.

Chapter 2

Development of a Chromosome Classification Approach Using Deep Convolutional Networks

Contents

2.1 Introduction 78
2.2 Methodology 80
2.2.1 Stage 1: Global-Scale and Local-Scale Feature Learning 81
2.2.2 Stage 2: Classification Based on the Fused Features 87
2.2.3 Four-Step Training Strategy 87
2.2.4 Stage 3: Type Assignment Using Dispatch Strategy 88
2.3 Experiments and Results 88
2.3.1 Materials 88
2.3.2 Implementation Details 89
2.3.3 Evaluation Metrics 90
2.3.4 Results 91
2.4 Discussion 99
2.5 Conclusion 105

2.1 Introduction

For classification of Giemsa-stained chromosomes, traditional automatic methods mainly rely on geometrical features (e.g., a chromosome's length, centromere position, and banding pattern features). Lerner et al. [Lerner et al., 1995] first proposed two approaches of computing medial axis transform (MAT) to detect medial axes of chromosomes. Then, intensity-based features and centromeric indexes were fed into an MLP network for classification. Ming et al. [Ming and Tian, 2010] computed medial axes using a middle point algorithm. They extracted banding patterns by average intensity, gradient, and shape profiles and adopted an MLP classifier. Markou et al. [Markou et al., 2012] proposed a robust method to first extract medial axes using a thinning algorithm. Bifurcations of the axis were removed iteratively via a pixel-neighborhood-based pruning algorithm. Then, the axis was smoothed and extended, with the band-profile features extracted along it. An SVM classifier was finally adopted for type classification. Several other methods targeted at precise detection of the medial axis and centromere location [Stanley et al., 1996, Wang et al., 2008, Arachchige et al., 2013, Loganathan et al., 2013], providing a foundation for accurate chromosome classification.

With the advent of deep learning, researchers tended to employ convolutional neural networks (CNNs) for feature extraction in classification tasks [LeCun et al., 1998, Krizhevsky et al., 2012, Szegedy et al., 2016, Simonyan and Zisserman, 2014, He et al., 2016a, Huang et al., 2017a, Lin et al., 2015, Fu et al., 2017, Jaderberg et al., 2015]. Three methods were reported on using deep learning techniques in chromosome studies. Sharma et al. [Sharma et al., 2017] proposed a CNN-based method for classification. Bent chromosomes were first straightened by cropping and stitching, and then normalized by length. The accuracy of classification was 86.7% for such preprocessed chromosomes. Gupta et al. [Gupta et al., 2017] developed a classification method based on the Siamese Network. Chromosomes were first straightened using two proposed approaches and then fed into the Siamese Network for high-level feature embeddings. An MLP classifier exploited such embeddings for classification and an average accuracy of 85.6% was achieved. Very recently, Wu et al. [Wu et al., 2018b] proposed a VGG-Net-D-based approach for category classification. Due to inadequate labeled data, they adopted generative adversarial network (GAN) to generate samples as data augmentation. Their performance was far below requirement for clinical application, with an average precision of 63.5% achieved.

Although many microscopes are nowadays equipped with chromosome classification systems (e.g., CytoVision [Micci et al., 2001, Yang et al., 2010, Rødahl et al., 2005], Ikaros [Gadhia et al., 2014], and ASI HiBand [Fan et al., 2000]), users still have to manually drag each chromosome image and drop it to the target position in the practical karyotyping process due to their poor performance. Research studies reveal that the challenges of chromosome classification mainly lie in the following aspects: 1) Chromosomes are often curved and bent due to their non-rigid nature, making it difficult to accurately extract their medial axes. Hence, errors accumulate in the process of straightening and feature computation along such axes, leading to an accuracy drop. 2) Even for the chromosomes of the same class, they vary slightly from person to person in terms of local details. The generalizability and performance of traditional methods, which depend on manually designed features, may degrade for clinical applications. 3) The chromosome

polarity, which reflects whether a chromosome's q-arm (long arm) is downward or upward, is often not considered in previous work. However, it is important to decide the chromosome's orientation in the process of repositioning for the karyotyping map generation. All the q-arms should stay downward.

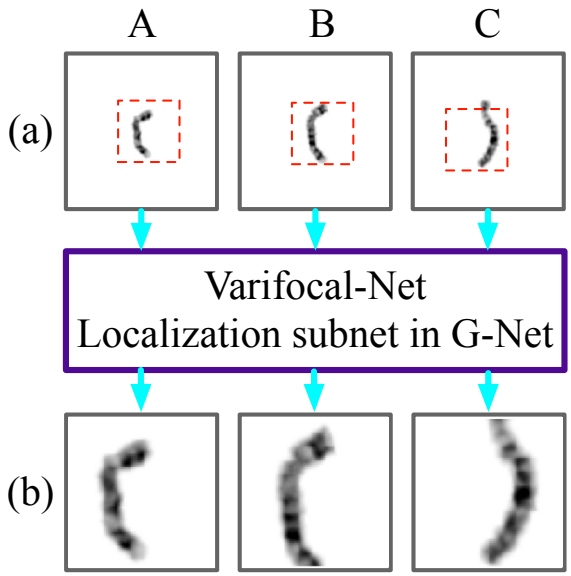


Figure 2.1: The focus is varied from global to local. Given chromosome images (A, B, C), the localization subnet detects their finer regions to crop and magnify. (a) The original chromosome images. (b) The local parts after zooming in.

To tackle the above challenges, we propose a novel CNN-based approach for chromosome classification. Its name, Varifocal-Net, highlights the capacity to zoom into local regions automatically. It has one global-scale network (G-Net) and one local-scale network (L-Net). We extract global features and pinpoint specific local regions via the G-Net. The view is changed (see Fig. 2.1) as our Varifocal-Net zooms into the discriminative region of a chromosome. Local features are extracted from such local parts via the L-Net. At first glance, such a global-to-local idea resembles the concept of multi-scale CNNs used in cellular image analysis [Godinez et al., 2017, Buyssens et al., 2012, Godinez et al., 2018, Pan et al., 2018] and other vision tasks [Shen et al., 2015b, Zeng et al., 2017, Lotter et al., 2017]. However, unlike previous multi-scale methods, our approach learns multi-scale information in the global-to-local mechanism. It locates the discriminative local region and extracts the features of the two scales through two independent networks. The proposed Varifocal-Net comprises three stages. The first stage is to learn effective feature representations at both global and local scales. The global-scale representations mainly concern overall information such as the chromosome's length, shape, and size, which determines its type on a coarse-grained level. The local-scale representations depict details such as texture patterns of local parts, which facilitate discrimination among chromosomes on a fine-grained level. The second stage is to build two MLP classifiers to leverage features of both two scales for prediction of type and polarity, respectively. The third stage is to introduce a dispatch strategy for type assignment within each patient case. To validate

the effectiveness and generalizability of our approach, we construct a large dataset containing 1909 karyotyping cases. Extensive experiments on the dataset corroborate that the Varifocal-Net achieved better performance than state-of-the-art methods. Our contributions can be summarized as follows:

- Inspired by the zoom capability of cameras, we propose the Varifocal-Net to address the challenges of chromosome classification. We extract global-scale features from the whole image and local-scale features from the local region selected by our varifocal mechanism. Residual learning and multi-task learning strategies are utilized to promote effective feature learning. The detection of discriminative local parts is fulfilled via a localization subnet whose training involves both supervised and weakly-supervised learning.
- We utilize the concatenated features from both global and local scales to predict type and polarity simultaneously, thereby combining the knowledge acquired at two scales. To our best knowledge, this represents the first attempt to take multi-scale feature ensemble into account in chromosome studies.
- We propose a dispatch strategy to assign each chromosome to a type based on its predicted probabilities. Both the maximum likelihood criterion and possible abnormality situations are taken into account to enable the strategy suitable for clinical settings.
- We evaluate the proposed approach on a large dataset. It demonstrates its superior performance compared with state-of-the-art methods. The end-to-end manner of classification sidesteps the problem of inaccurate medial axis extraction and chromosome straightening.
- The Varifocal-Net has been put into clinical practice for chromosome classification. For each patient, it accurately classifies both abnormal and healthy chromosomes and diagnoses numerical abnormalities if the number of classified chromosomes is irregular.

The rest of the chapter is structured as follows: In Section 2.2, we describe the proposed method. In Section 2.3, we provide experiments and results. Section 2.4 discusses our findings, followed by the conclusion in Section 2.5.

2.2 Methodology

The flowchart of the proposed Varifocal-Net is depicted in Fig. 2.2. It is composed of three stages: a) Global-scale and local-scale feature learning by optimizing the Varifocal-Net in an alternative way; b) Classification of type and polarity via MLP classifiers utilizing the fused features; c) Assignment of chromosomes' types with the proposed dispatch strategy. Original chromosome images are separated manually by cytogeneticists from captured microscopic images. They are preprocessed as discussed in Sec. 2.3.2 and taken as inputs to the G-Net in the first stage. The G-Net contains deep CNNs, one classification subnet, and one localization subnet, as shown in Fig. 2.3. Global-scale features are extracted via the CNNs, which are optimized by the loss function of the classification

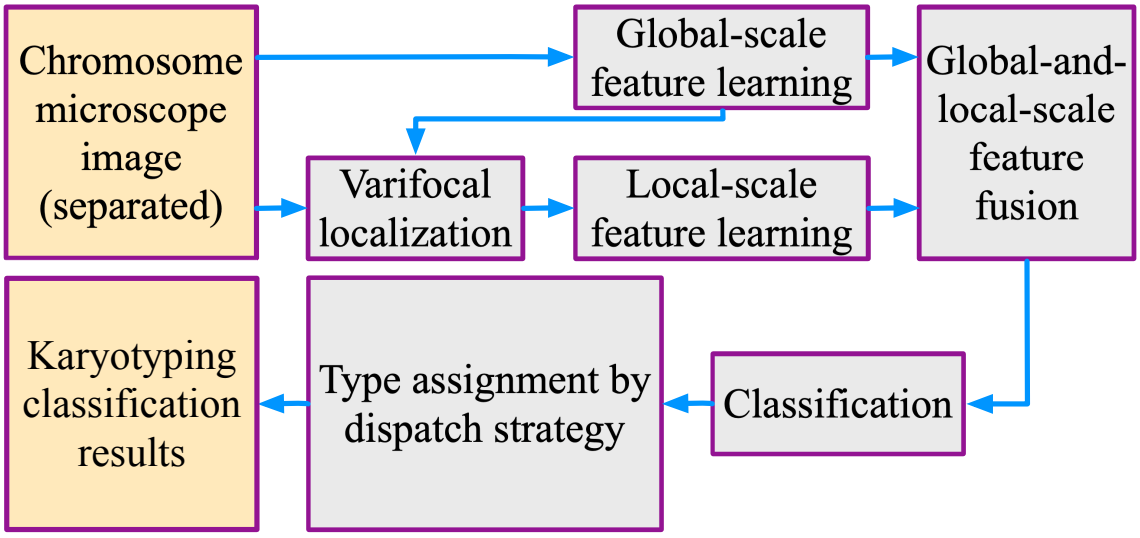


Figure 2.2: Flowchart of the proposed Varifocal-Net for chromosome classification.

subnet. After the CNNs and classification subnet converge, we pre-train the localization subnet to output initial coordinates for local region detection. Then, with local parts cropped and rescaled, we optimize the L-Net and the localization subnet of the G-Net alternatively. In the second stage, with the fused two-scale features, we build two MLP classifiers to predict chromosome's type and polarity, respectively. The schematic representations of the first stage and the second stage of our Varifocal-Net are illustrated in Fig. 2.3 and Fig. 2.6, respectively. For each chromosome within one patient case, a dispatch strategy is employed in the third stage to assign it to a certain type based on its predicted probabilities.

2.2.1 Stage 1: Global-Scale and Local-Scale Feature Learning

Feature Extraction with Residual Learning

The architecture of deep CNNs for feature extraction is the same for both the G-Net and the L-Net. Inspired by the success of ResNet [He et al., 2016a,Zagoruyko and Komodakis, 2016], we adopt wide residual blocks to introduce residual learning. Such CNNs consist of one convolution layer (Conv), three residual blocks, one batch normalization layer (BN), and one rectified linear unit (ReLU). Each residual block has four residual units as illustrated in Fig. 2.4, with the first unit increasing the number of channels and down-sampling features through strided convolution.

Multi-task Learning with Weighted Classification Loss

Since the tasks of type and polarity classification are correlated, we adopt multi-task learning to take inner relation between these tasks into consideration. It improves the effectiveness of feature extraction through a shared representation of CNNs [Caruana, 1997]. In the classification subnet, a max-pooling layer is followed by two fully-connected

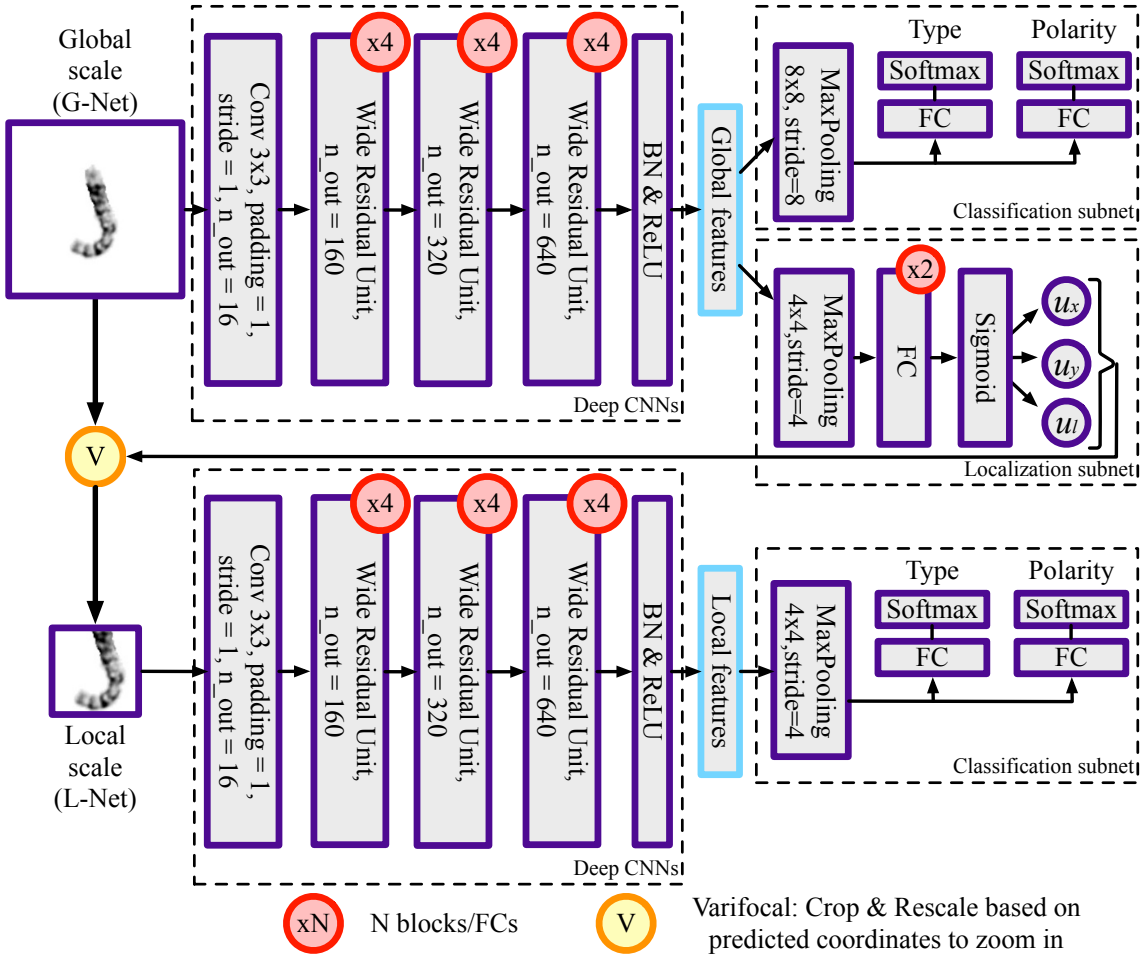


Figure 2.3: The first stage of the proposed Varifocal-Net: global-scale and local-scale feature extraction via the G-Net and the L-Net, respectively.

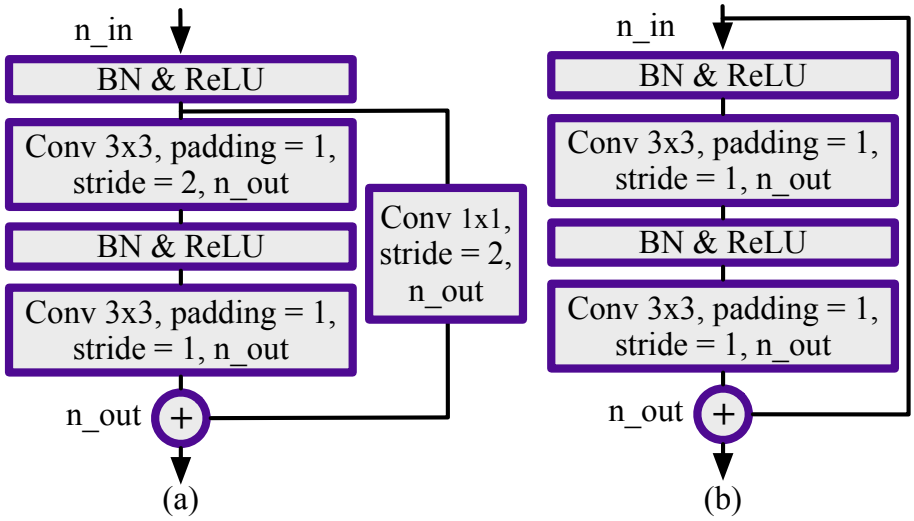


Figure 2.4: Wide residual unit. n_in and n_out stand for number of input and output feature channels, respectively. (a) if n_in $\neq$ n_out. (b) if n_in = n_out.

(FC) layers respectively to predict type and polarity. The FC layers map the feature vector to the probability vectors of 24 dimensions (for the type task) and 2 dimensions (for the polarity task). We train the deep CNNs in the G-Net and the L-Net independently by minimizing a weighted loss of the classification subnet. For the type task, given a set of N training triplets $\{(x_i, y_i^t, y_i^p)\}_{i=1,2,...,N}$, the cross-entropy loss between the output vector $\mathbf{O^t}$ and the target vector $\mathbf{Y^t}$ is given by:

$$\mathcal{L}_t(\mathbf{O^t}, \mathbf{Y^t}) = \sum_{i=1}^{N} -log(\frac{exp(o_i^t[y_i^t])}{\sum_{j=1}^{24} exp(o_i^t[j])}), \tag{2.1}$$

where o_i^t and y_i^t denote the output probability vector and the target type for the sample x_i, respectively. Note that here we combine the softmax function and the standard cross-entropy function into one formula. Similarly, the polarity classification loss between the predicted vector $\mathbf{O^p}$ and the target vector $\mathbf{Y^p}$ is defined as:

$$\mathcal{L}_p(\mathbf{O^p}, \mathbf{Y^p}) = \sum_{i=1}^{N} -log(\frac{exp(o_i^p[y_i^p])}{\sum_{j=1}^{2} exp(o_i^p[j])}), \tag{2.2}$$

where o_i^p and y_i^p stand for the probability vector and the target polarity, respectively. The total multi-task loss is given by:

$$\mathcal{L}_{cls}(\mathbf{O^t}, \mathbf{Y^t}, \mathbf{O^p}, \mathbf{Y^p}) = \mathcal{L}_t(\mathbf{O^t}, \mathbf{Y^t}) + \lambda \mathcal{L}_p(\mathbf{O^p}, \mathbf{Y^p}), \tag{2.3}$$

in which λ is a weight controlling the balance between the two loss terms. We place more emphasis on the type task, thus setting $\lambda = 0.5$ in our experiments.

Varifocal Mechanism

Previous work on chromosome classification takes no advantage of multi-scale feature learning and fusing. These methods do not detect specific finer parts for detail description (e.g., nuance of banding's number, width, and intensity among similar chromosomes). Motivated by the success of region proposal network (RPN) [Girshick, 2015, Ren et al., 2015] and attention proposal network (APN) [Fu et al., 2017], we propose a varifocal mechanism that zooms into local regions of chromosomes automatically for finer feature extraction. Given a chromosome sample x_i, it first predicts the position and size of a local region box via the localization subnet, which is sequentially composed of a max-pooling layer, two FC layers, and a sigmoid layer. The square box prediction is expressed as:

$$(u_x^i, u_y^i, u_l^i) = f(\mathbf{W_c} * x_i), \tag{2.4}$$

where $\mathbf{W_c}$ and $*$ denote all parameters of deep CNNs and their related operations (e.g., Conv, BN, and ReLU), respectively. $\mathbf{W_c} * x_i$ gives the global feature of x_i and $f(\cdot)$ represents the proposed localization subnet. The variables u_x^i and u_y^i denote the relative coordinates of the box's center (x_c, y_c) and u_l^i is the relative length of the half of its side. All these variables range from 0 to 1. Assuming the top-left corner of x_i as the origin of the global pixel coordinate system where x-axis starts from left to right and y-axis from top to bottom, we adopt the parameterizations of the top-left (tl) and bottom-right (br) pixels of the region box as follows:

$$\begin{aligned} t_{x(tl)}^i = T_1 + u_x^i \cdot T_2 - t_l^i,\ t_{x(br)}^i = T_1 + u_x^i \cdot T_2 + t_l^i, \\ t_{y(tl)}^i = T_1 + u_y^i \cdot T_2 - t_l^i,\ t_{y(br)}^i = T_1 + u_y^i \cdot T_2 + t_l^i, \\ t_l^i = u_l^i \cdot T_1/2 + T_1/2, \end{aligned} \tag{2.5}$$

where T_1, T_2, and t_l^i denote the minimum margin, maximum shift, and half of the side length, respectively. Fig. 2.5 illustrates these parameterizations. Note that here we restrict the position and size of the predicted local region for two reasons. First, the predicted region should focus on a discriminative part of the chromosome, which is in the center of the image. Second, the region cannot exceed the image boundary and its size should be moderate to effectively capture local features. In our implementation, we set $T_2 = 2T_1$ empirically because it forces the localization subnet to focus on the central region.

Once a local region is predicted, the focus is moved onto it by cropping and rescaling. The cropping operation is implemented using a variant of two-dimensional (2-D) boxcar function [Fu et al., 2017] as an approximation. Given the coordinate tuple $(t_{x(tl)}^i, t_{y(tl)}^i, t_{x(br)}^i, t_{y(br)}^i)$, we use the boxcar function to generate a region mask and multiply it with the original image in an element-wise manner. It is mathematically expressed as:

$$\begin{aligned} x_i^{loc} &= x_i \odot boxcar(t_x^i, t_y^i, t_l^i), \\ boxcar(t_x^i, t_y^i, t_l^i) &= (H(x - t_{x(tl)}^i) - H(x - t_{x(br)}^i)) \\ &\quad \cdot (H(y - t_{y(tl)}^i) - H(y - t_{y(br)}^i)) \end{aligned} \tag{2.6}$$

where $\odot$ denotes element-wise multiplication and x_i^{loc} stands for the cropped local part. The 2-D $boxcar(t_x^i, t_y^i, t_l^i)$ function serves as a mask and $H(x)$ is the Heaviside step function. Note that the derivative of $H(x)$ is infinite at $x = 0$. Since its derivative is required

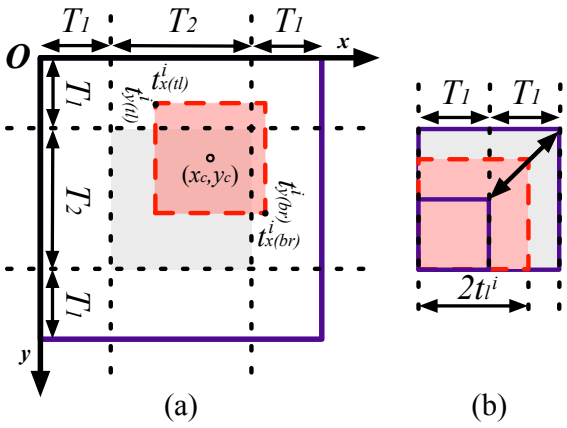


Figure 2.5: The diagram of parameterizations for the sample x_i. (a) The red box is the predicted local region and the gray background square is the area where the box's center pixel (x_c, y_c) can be located. (b) The side length of the predicted box $(2t_l^i)$ is restricted, ranging from T_1 to $2T_1$.

in back-propagation, we use the logistic function as a smooth analytic approximation for $H(x)$ in experiments, which is computed by:

$$H(x) = \frac{1}{1 + e^{-kx}},\ k > 0 \tag{2.7}$$

in which a larger k (e.g., $k = 10$) leads to a sharper change at $x = 0$. The multiplication with $boxcar(t_x^i, t_y^i, t_l^i)$ will mask out the target local region by keeping the value of pixels inside the region almost unchanged and that of others close to zero. Then, we crop the target region in x_i^{loc} and rescale it to a unified size via bilinear interpolation, which makes it easier for both algorithm implementation and finer feature extraction in the L-Net. So far, the Varifocal-Net has zoomed into a particular local part. Note that in the forward process, the local region is cropped directly by indexed slicing. In the backward propagation process, since the cropping operation is not derivative, the boxcar function is used to approximate it and provide necessary gradient for proper parameter optimization. Detailed analytical derivations are presented in Sec. 2.2.1.

Loss Function of the Localization Subnet

With definitions of the localization subnet $f(\cdot)$, we adopt both supervised and weakly-supervised learning to optimize it. The supervised method is employed in pre-training to initialize the parameters of $f(\cdot)$. For such pre-training, we assign the ground-truth coordinates $(u_x^{i*}, u_y^{i*}, u_l^{i*})$ for the sample x_i as follows: 1) The locations u_x^{i*} and u_y^{i*} are set to 0.5 since a chromosome is centered in the image. 2) Based on u_x^{i*} and u_y^{i*}, the smallest region that covers the whole chromosome is calculated and u_l^{i*} is computed accordingly. The lower bound of u_l^{i*} is 0 and if the width or height of chromosome exceeds $2T_1$, u_l^{i*} will be set to 1. Given a set of N sample pairs $\{(x_i, u_x^{i*}, u_y^{i*}, u_l^{i*})\}_{i=1,2,...,N}$, our loss function

for supervised learning is defined as:

$$\begin{aligned} \mathcal{L}_u(\mathbf{U}, \mathbf{U}^*) &= \sum_{i=1}^{N} \sum_{\gamma \in \{x,y,l\}} \text{smooth}_{L_1}(u_\gamma^i - u_\gamma^{i*}), \\ \text{smooth}_{L_1}(x) &= \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases} \end{aligned} \tag{2.8}$$

where $\mathbf{U}$ and $\mathbf{U}^*$ denote the vector of the predicted coordinates and their ground-truth labels, respectively. The robust smooth_{L_1} loss [Girshick, 2015] is used to directly train the localization subnet to output initial local region coordinates. It is less sensitive to outliers than L_2 loss and smoother near zero compared to the standard L_1 norm. Its gradient is uniquely defined at zero point.

While the weakly-supervised method is aimed at improving the classification performance of the L-Net by optimizing the $f(\cdot)$ for finer part localization, we keep all parameters of the L-Net unchanged and only fine-tune the localization subnet by minimizing the multi-task loss (2.3) of the L-Net. Without ground-truth coordinates provided, the subnet $f(\cdot)$ autonomously learns to locate discriminative parts, making the extracted features meaningful. Thus, the total loss is given by:

$$\begin{aligned} \mathcal{L}_{loc}(\mathbf{U}, \mathbf{U}^*, \mathbf{O^t}, \mathbf{Y^t}, \mathbf{O^p}, \mathbf{Y^p}) = \mathcal{L}_u(\mathbf{U}, \mathbf{U}^*) + \\ \mathcal{L}_{cls}(\mathbf{O^t}, \mathbf{Y^t}, \mathbf{O^p}, \mathbf{Y^p}). \end{aligned} \tag{2.9}$$

Here, the subnet is only pre-trained once by minimizing $\mathcal{L}_u(\mathbf{U}, \mathbf{U}^*)$. Then, its optimization process is dominant by weakly-supervised learning. The training details of our proposed Varifocal-Net will be introduced in Sec. 2.2.3.

Back-propagation through Boxcar Function

We adopt the boxcar function for localization because it provides analytical representations between region cropping and the predicted relative coordinates (u_x^i, u_y^i, u_l^i), which is indispensable for parameter update in back-propagation. When optimizing $\mathcal{L}_{cls}(\mathbf{O^t}, \mathbf{Y^t}, \mathbf{O^p}, \mathbf{Y^p})$ to train the localization subnet, gradients back-propagate through the boxcar function. For one single image x_i, we designate the gradients that back-propagate to the input layer of the L-Net as $\mathbf{G}_{top}$. The partial derivatives of the loss to coordinates are then given by:

$$\begin{aligned} \frac{\partial \mathcal{L}_{cls}(\mathbf{O^t}, \mathbf{Y^t}, \mathbf{O^p}, \mathbf{Y^p})}{\partial u_\gamma^i} &\propto \mathbf{G}_{top} \odot \frac{\partial boxcar(t_x^i, t_y^i, t_l^i)}{\partial t_\gamma^i} \cdot \frac{\partial t_\gamma^i}{\partial u_\gamma^i}, \\ \frac{\partial t_x^i}{\partial u_x^i} = \frac{\partial t_y^i}{\partial u_y^i} &= T_2, \ \frac{\partial t_l^i}{\partial u_l^i} = T_1/2, \ \gamma \in \{x,y,l\}, \end{aligned} \tag{2.10}$$

where $\odot$ denotes element-wise multiplication. Hence, the derivatives of $boxcar(t_x^i, t_y^i, t_l^i)$ with respect to t_x^i, t_y^i and t_l^i largely influence the moving direction and size of the local region box. Note that in the context of minimizing our loss, it holds true for $\forall \gamma \in \{x,y,l\}$ that t_γ^i increases when $\frac{\partial \mathcal{L}_{cls}(\mathbf{O^t}, \mathbf{Y^t}, \mathbf{O^p}, \mathbf{Y^p})}{\partial u_\gamma^i} < 0$ and decreases otherwise. To achieve a consistent optimization direction, we follow [Fu et al., 2017] to calculate the negative squared

norm of derivatives $\mathbf{G}_{top}$ and compute the $boxcar(t_x^i, t_y^i, t_l^i)$'s partial derivatives explicitly in the back-propagation process.

2.2.2 Stage 2: Classification Based on the Fused Features

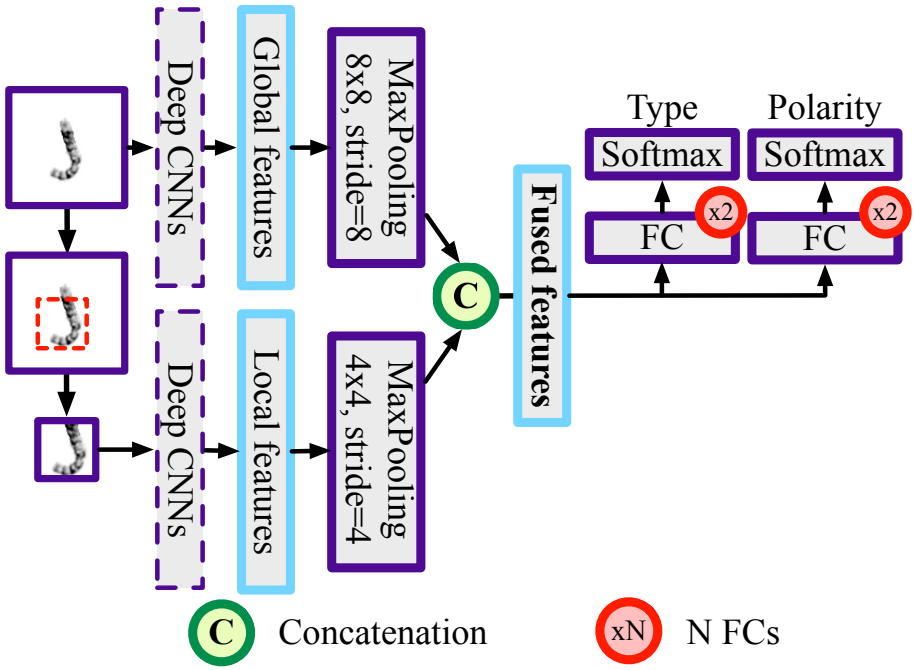


Figure 2.6: The second stage of the proposed Varifocal-Net: chromosome classification using fused features from both global and local scales.

Once both the G-Net and the L-Net are optimized, global-scale and local-scale features can be extracted via deep CNNs. To make full use of these two representations, it is reasonable to concatenate them into a feature ensemble. We build two MLP classifiers (see Fig. 2.6) to learn the mapping from the fused features to classification probabilities of type and polarity, respectively. Each classifier consists of two FC layers and one Softmax layer. With the trained classifiers, the proposed Varifocal-Net simultaneously predicts chromosome's type and polarity in an end-to-end manner.

2.2.3 Four-Step Training Strategy

In the present study, we adopt a four-step optimization technique to alternatively train the network. In the first step, we initialize deep CNNs of the G-Net and L-Net via He's method [He et al., 2015]. In the second step, we train deep CNNs and the classification subnet in the G-Net until convergence. At this point, the localization subnet and the L-Net are not optimized. In the third step, we prepare all the ground-truth coordinates of local region boxes and only pre-train the localization subnet once. Finally, we train the L-Net and the localization subnet alternatively in the fourth step. Keeping the parameters of the localization subnet fixed, we optimize the L-Net by minimizing our multi-task loss.

Then we fix the parameters of the L-Net and fine-tune the localization subnet alone. Such alternative training can be run for iterations until there is no further error loss decrease.

2.2.4 Stage 3: Type Assignment Using Dispatch Strategy

In karyotyping practice, the classification of chromosome's type is conducted within each patient case. Therefore the classification can also be viewed as dispatching each chromosome to a certain type. This led us to propose a dispatch strategy for type assignment in the third stage. The design of the dispatch strategy follows two simple rules about karyotyping's domain knowledge [McGowan-Jordan et al., 2016]:

- Each healthy patient has 46 chromosomes for 23 classes (female) or 24 classes (male).
- For unhealthy patient, the number of each type falls between 1 and 3 (e.g., monosomy 21 and trisomy 21) except extremely rare cases. Type Y has less than 3 chromosomes.

Considering both the maximum likelihood criterion and possible abnormality situations, we dispatch chromosomes twice. Given the predicted probabilities from the second stage of the Varifocal-Net, the first-time dispatch is to assign each chromosome to the type having the highest probability. The second-time dispatch is to check and compare the probabilities of different chromosomes that are assigned to the same type. The confidence threshold th is designed to filter out uncertain assignments. The dispatch strategy is described in details in Alg. 4. Note that it is not used for polarity prediction because polarity only involves 2 classes (q-arm upward or downward).

2.3 Experiments and Results

2.3.1 Materials

For the experiments conducted in this section, we collected 1909 different patients' karyotyping cases from the Xiangya Hospital of Central South University, China. Each patient case contains one Giemsa stained microscopic image of meta-phase chromosomes. All images are grayscale and sampled with the same resolution, using the Leica's CytoVision System (GSL-120). Each chromosome is of approximate 300-band levels. The datasets contain 1784 karyotyping cases from healthy patients (1061 male and 723 female) and 125 cases from unhealthy patients (73 male and 52 female). The unhealthy cases contain both numerical and structural abnormalities. Each chromosome's type is manually annotated by cytogeneticists in real-world clinical environments. The type of autosomes is labeled from 0 to 21 and the type of sex chromosomes X and Y are denoted as 22 and 23, respectively. The polarity of a chromosome is labeled as 1 if its q-arm is downward and 0 otherwise.

We obtain each individual chromosome image by manually segmenting it from microscopic images. In total, there exist 87831 separated chromosomes. We randomly split both healthy and unhealthy samples into five subsets to perform five-fold cross validation. Each time, four subsets are used for training the model and fine-tuning the hyper-parameters. The remaining one subset is left for testing. Note that the chromosome

Algorithm 4 Dispatch strategy for chromosome's type.

Input: N chromosomes; the probabilities of 24 types P_i for the i-th chromosome (P_{ij} stands for its probability of being type j, $i = 1,2,...,N, j = 1,2,...,24$); confidence threshold th.

Output: The set of chromosomes assigned to type k ($O_k, k = 1,2,...,24$); possible abnormal warnings.

$T_k = \varnothing, O_k = \varnothing, \forall k \in \{1,2,...,24\}$.
for each $i \in \{1,2,...,N\}$ **do**
    Compute the most probable type $j^* = \arg\max_j P_{ij}$ and dispatch the i-th chromosome to type j^* by $T_{j^*} = T_{j^*} \cup \{i\}$;
end for
for each $k \in \{1,2,...,24\}$ **do**
    $S = 1$ if $k = 24$, otherwise $S = 2$;
    if $|T_k| > S$ **then**
        Sort each element in T_k based on its probability. From T_k, choose $S+1$ elements ($Q_k = \{i^1,...,i^{S+1}\}$) with the highest probability if $P_{ik} > th, \forall i \in Q_k$, otherwise choose only S elements ($Q_k = \{i^1,...,i^S\}$);
        $O_k = O_k \cup Q_k$;
        for $i \in T_k \setminus Q_k$ **do**
            Compute the second probable type $j^* = \arg\max_{j,j\neq k} P_{ij}$ and dispatch it to type j^* by $O_{j^*} = O_{j^*} \cup \{i\}$;
        end for
    else
        $O_k = O_k \cup T_k$;
    end if
end for
Print abnormal warnings if $|O_k| \neq 2, \forall k \in \{1,2,...,22\}$ or $|O_{23}| + |O_{24}| \neq 2$;
return $O_k, k = 1,2,...,24$.

samples are divided by patient case. All chromosomes of the same case stay in the same subset. Table 2.1 provides the details of our datasets.

2.3.2 Implementation Details

The size of images differs from each other and we first padded them with pixels into square images of the same size. The padding value is set as 255 to imitate the background of the original Giemsa stained images. And the size of padded image is 320×320 pixels. Then, we resized the image to 256×256 pixels and normalized all N images as follows:

$$x_i' = (x_i - \mu_i)/\sigma_i,\ i = 1,2,...,N \tag{2.11}$$

where μ_i and σ_i are the mean value and the standard deviation of the sample x_i, respectively. x_i' denotes the normalized input, which has a zero mean and a unit variance. For local region prediction, the margin T_1 is 64 and the shift range T_2 is 128. The cropped target region was then upsampled to 128×128 pixels as the input to the L-Net. In Table 2.2 and Table 2.3, we describe feature dimensions of the proposed Varifocal-Net for the

Table 2.1: Statistics of the dataset. (H: Healthy Samples, U: Unhealthy Samples.)

Dataset		Case #		Image #		Total image #
		Male	Female	Male	Female	
Total samples	H	1061	723	48806	33258	87831
	U	73	52	3384	2383	

first and the second stages, respectively. For the dispatch strategy, the confidence threshold th was set to 0.9 because we only keep highly-confident chromosomes when possible numerical abnormalities happen.

Table 2.2: The feature dimensions of the Varifocal-Net for the first stage. (T: Type, P: Polarity, Loc: Localization.)

Layer	Dimension				
	G-Net			L-Net	
Input	256×256			128×128	
Deep CNNs	$640 \times 32 \times 32$			$640 \times 16 \times 16$	
Max-pooling	$640 \times 4 \times 4$		$640 \times 8 \times 8$	$640 \times 4 \times 4$	
FC1	24 (T)	2 (P)	1024	24 (T)	2 (P)
FC2	–	–	3 (Loc)	–	–

Table 2.3: The feature dimensions of the Varifocal-Net for the second stage. (T: Type, P: Polarity.)

Layer	Dimension	
Input	256×256 (G-Net)	128×128 (L-Net)
Deep CNNs	$640 \times 32 \times 32$	$640 \times 16 \times 16$
Max-pooling	$640 \times 4 \times 4$	$640 \times 4 \times 4$
Concatenation	$640 \times 4 \times 4 \times 2$	
FC1	512	512
FC2	24 (T)	2 (P)

In the training process, we adopted horizontal flipping and random rotation between $[0°, 45°]$ for data augmentation. The vertical flipping operation was performed to change the polarity label of a chromosome. All modules of the Varifocal-Net were trained from scratch using Adam optimizer [Kingma and Ba, 2014] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The initial learning rate was set to 0.0001 and it decreased by nine-tenth every 10 epochs. We implemented the proposed Varifocal-Net and other CNN-based methods in Python, with PyTorch framework [Paszke et al., 2019]. All experiments were conducted under a Ubuntu OS workstation with Intel Xeon(R) CPU E5-2620 v4 @ 2.10GHz, 128 GB of RAM, and 4 NVIDIA GTX Titan X GPUs.

2.3.3 Evaluation Metrics

The performance of the Varifocal-Net was evaluated by four metrics: the accuracy of all

the testing images (Acc.), the average F_1-score over classes of all the testing images (F_1), the average accuracy of the complete karyotyping per patient case (Acc. per Case), and the average accuracy of the complete karyotyping per patient case using the proposed dispatch strategy (Acc. per Case-D). The Acc. is an intuitive measurement defined as the fraction of the testing samples which are correctly classified.

For the computation of F_1-score, we first define the following four criteria to fit the context of multi-class classification:

- True positives (TP_j): images predicted as class j which actually belong to class j
- False positives (FP_j): images predicted as class j which actually do not belong to class j
- False negatives (FN_j): images predicted as class k ($\forall k \neq j$) which actually belong to class j
- True negatives (TN_j): images predicted as class k ($\forall k \neq j$) which actually do not belong to class j

Then, the F_1-score is computed as:

$$\begin{aligned} F_1 &= \frac{1}{N_{cls}} \sum_{j=1}^{N_{cls}} \frac{2 \cdot Precision_j \cdot Recall_j}{Precision_j + Recall_j}, \\ Precision_j &= \frac{TP_j}{TP_j + FP_j}, \\ Recall_j &= \frac{TP_j}{TP_j + FN_j}, \end{aligned} \tag{2.12}$$

where N_{cls} equals 24 and 2 for type and polarity recognition, respectively.

The accuracy per patient case was adopted to evaluate the performance in clinical settings. It is computed by checking the fraction of the correctly classified samples within each patient case. No dispatch strategy is used for computing Acc. per Case. We only assign each chromosome to the type having the highest predicted probability. For the computation of Acc. per Case-D, the proposed dispatch strategy is employed and accuracy within each case is recalculated for all samples.

The mean value and the standard deviation of these four metrics are provided to assess performance stability. They were calculated based on the results of five-fold cross validation and displayed in percentage.

Furthermore, we also adopted a receiver operating characteristic (ROC) analysis for performance comparison. The ROC curves averaged over all classes were plotted and the area under each curve (AUC) was calculated as well.

2.3.4 Results

This section presents experimental results in three parts. We first provide detailed evaluation results of the proposed Varifocal-Net. Then, a comparison of the proposed method with state-of-the-art methods is given. Finally, we present additional results for analyzing our performance.

Evaluation Results

Table 2.4 gives the classification results of the G-Net, L-Net, and the entire Varifocal-Net. The global-scale G-Net achieved the accuracy (%) of 97.8 and 99.0 for type and polarity recognition, respectively. With the localization subnet for finer region detection, the local-scale L-Net reduced classification errors. By utilizing the knowledge learned at two scales, the proposed Varifocal-Net yielded the best performance. The accuracy (%) of type and polarity tasks were boosted to 98.9 and 99.2, respectively. Due to the proposed dispatch strategy, the accuracy of type classification per case is further improved for each method. The proposed Varifocal-Net achieved the averaged Acc. per Case-D (%) of 99.2. Though the total training time is relatively long, the testing time of the Varifocal-Net is only 5.9ms per sample.

To observe the performance of the Varifocal-Net on each class of chromosomes, Table 2.5 and Table 2.6 provide the F_1-score, precision, and recall, which were computed within each category. For type recognition, the proposed method performed worst on Y chromosomes, with only a F_1-score (%) of 94.3 achieved. The evaluation results of classes No. 4, No. 5, No.15, No. 16, No. 20–No. 22, X, and Y are below average. For polarity recognition, the orientation of q-arm was accurately predicted, with the F_1-score of each class above 99%.

Besides, for polarity classification, we also computed the accuracy within each type category to learn the performance difference among chromosome types. Table 2.7 indicates that our prediction is relatively inaccurate for two long types (classes No. 2 and No. 5) and four short types (classes No. 15, No.16, No. 20 and Y).

Comparison with the State-Of-The-Art Methods

Table 2.8 provides a comparison of the proposed Varifocal-Net with state-of-the-art methods. The first two methods [Sharma et al., 2017, Gupta et al., 2017] were proposed specifically for classifying Giemsa stained chromosomes. Both the two existing methods employed CNNs for feature extraction, and they relied on straightening chromosomes for normalization and used small datasets. In contrast, we adopted an end-to-end fashion for prediction. We implemented the two methods and evaluated them using five-fold cross validation. Their performance of type recognition on our large testing set proves the superiority of our method, which surpasses [Sharma et al., 2017] and [Gupta et al., 2017] by nearly 6.7% and 7.5% in average F_1-score, respectively.

To test the usefulness of the varifocal mechanism, we replaced the localization subnet by a simple preprocessing method. The input of the L-Net is not the cropped local region of the original image. Instead, we directly rescaled and padded the minimum bounding box of each chromosome image into the same size (256×256). The processed image contains the whole chromosome part and consequently the extracted features are no longer local. After the L-Net converges, the features learned from the G-Net and the L-Net are concatenated as well for the training of the second stage. We named the L-Net and the Varifocal-Net using such a simple preprocessing step as L-Net (Simple) and Varifocal-Net (Simple), respectively.

Table 2.8 shows that the simple preprocessing method does not facilitate feature learning of fine-grained details. Our method outperforms the L-Net (Simple) and the

Table 2.4: Performance of the Varifocal-Net (mean±standard deviation). The results are presented in terms of four evaluation metrics: average F_1-score of all testing images (F_1), accuracy of all testing images (Acc.), average accuracy per patient case (Acc. per Case), and average accuracy per patient case using the proposed dispatch strategy (Acc. per Case-D). (T: Type, P: Polarity, PET: Per Epoch Time, TPI: Time Per Image.)

Stage	Method	F_1 (%)		Acc. (%)		Acc. per Case (%)		Acc. per Case-D (%)	# Epoch ×PET (s)	Testing TPI (ms)
		T	P	T	P	T	P	T		
1	G-Net	97.5±0.4	99.0±0.1	97.8±0.4	99.0±0.1	97.8±3.8	99.0±1.9	98.2±3.3	30×956.3±1.5	5.7±0.1
	L-Net	98.2±0.5	99.2±0.1	98.4±0.5	99.2±0.1	98.4±2.9	99.2±1.6	98.9±2.5	30×1142.3±2.3	6.8±0.1
2	**Varifocal-Net**	**98.7±0.7**	**99.2±0.3**	**98.9±0.7**	**99.2±0.3**	**98.9±2.3**	**99.2±1.5**	**99.2±2.1**	**20×1150.8±11.3**	**5.9±0.1**

Table 2.5: Performance of the Varifocal-Net for each chromosome type (mean±standard deviation).

Class (No.)	F_1 (%)	Precision (%)	Recall (%)
1	99.6±0.6	99.5±0.7	99.7±0.5
2	99.3±0.7	98.8±0.9	99.7±0.5
3	99.5±0.6	99.4±0.8	99.6±0.5
4	98.6±1.1	98.4±1.1	98.7±1.1
5	98.6±0.7	98.7±0.7	98.6±0.7
6	99.4±0.6	99.7±0.3	99.2±0.8
7	99.7±0.2	99.7±0.3	99.6±0.3
8	98.9±0.8	98.9±0.9	98.9±0.8
9	98.7±0.4	98.8±0.8	98.6±0.5
10	98.7±0.7	98.7±0.8	98.7±0.7
11	99.6±0.2	99.6±0.3	99.6±0.3
12	99.7±0.2	99.8±0.1	99.6±0.4
13	98.7±0.7	98.8±0.5	98.7±1.0
14	99.0±0.5	99.2±0.6	98.9±0.5
15	98.5±0.8	98.6±0.9	98.4±0.7
16	97.9±1.3	97.9±1.2	97.9±1.4
17	99.3±0.6	99.1±0.8	99.4±0.4
18	98.8±1.1	98.9±0.8	98.7±1.5
19	98.7±0.8	98.6±0.9	98.8±0.8
20	98.4±1.0	98.5±1.1	98.4±0.9
21	98.5±0.5	98.5±0.4	98.6±0.7
22	98.4±0.8	98.3±0.8	98.6±0.9
X	98.3±1.1	98.6±0.9	98.1±1.4
Y	94.3±3.6	95.0±3.5	93.6±3.8

Table 2.6: Performance of the Varifocal-Net for each chromosome polarity (mean±standard deviation).

Class	F_1 (%)	Precision (%)	Recall (%)
q-arm upward	99.2±0.3	99.1±0.4	99.4±0.1
q-arm downward	99.3±0.3	99.4±0.1	99.1±0.4

Varifocal-Net (Simple), which validates the effectiveness of the localization subnet.

Table 2.8 also provides the results of comparison with other CNN models. To assess our multi-scale feature ensemble strategy, we evaluated the performance of the well-known models that have been proved powerful on the ImageNet dataset, including AlexNet [Krizhevsky et al., 2012], GoogLeNet [Szegedy et al., 2016], VGG-Net-D [Simonyan and Zisserman, 2014], ResNet-101 [He et al., 2016a], and DenseNet-121 [Huang et al., 2017a]. The number of convolution layers in these five models and our Varifocal-Net (feature extractor part) are respectively 5, 22, 13, 100, 120, and 28, which are much deeper than previous work in chromosome classification [Sharma et al., 2017,Gupta et al., 2017]. Besides, we also evaluated Spatial Transformer Network (STN) [Jaderberg et al., 2015] for performance comparison. It contains 4 Conv layers, 4 Max-Pooling layers, and 2 FC layers. We inserted STN into the first layer of each model and retrained it. Since the parameters of these popular models were compatible with the 3-channel 224×224 natural images (ImageNet), we rescaled our 256×256 grayscale images into 224×224 pixels and then generated 3 channels by directly stacking the original grayscale channel. The preprocessing step was also adopted to normalize all the inputs as mentioned in Sec. 2.3.2. To introduce multi-task learning, we duplicated the classifier settings in each model so that both type and polarity could be predicted at the same time. The loss function is defined as (2.3) with $\lambda = 0.5$. We trained all models from scratch because the collected samples are sufficient. The results show that all models have acceptable performance. Even the shallowest AlexNet achieved the accuracy (%) of 90.8 and 97.1 for type and polarity classifications, respectively. Among these single-scale CNN models, the highest accuracy and F_1-score were achieved by DenseNet-STN for both type and polarity tasks. However, its result is still inferior to ours, where the error rates of type classification are reduced by half. For the polarity task, our method also outperformed other CNN models. Note that the use of STN does not necessarily improve the performance. Its introduction in GoogLeNet and G-Net brings about obvious decrease.

In the real clinical environment, it is imperative to correctly classify chromosomes having numerical and structural anomalies. To test the robustness of different methods under abnormal circumstance, we specially provide the evaluation results only on unhealthy cases in Table 2.9. For most CNN-based methods, the performance degraded

Table 2.7: Performance of the Varifocal-Net for polarity classification within each type (mean±standard deviation).

Class (No.)	Acc. (%)	Class (No.)	Acc. (%)	Class (No.)	Acc. (%)
1	99.3±0.5	9	99.6±0.1	17	99.3±0.5
2	99.1±0.5	10	99.5±0.2	18	99.5±0.1
3	99.5±0.4	11	99.8±0.2	19	99.3±0.4
4	99.2±0.4	12	99.6±0.2	20	96.2±1.1
5	99.1±0.3	13	99.6±0.4	21	99.3±0.3
6	99.5±0.2	14	99.8±0.2	22	99.5±0.1
7	99.8±0.2	15	98.9±0.5	X	99.2±0.3
8	99.5±0.2	16	99.1±0.4	Y	98.0±0.8

Table 2.8: Comparison results of the proposed method with state-of-the-art methods (mean±standard deviation). The results are presented in terms of four evaluation metrics: average F_1-score of all testing images (F_1), accuracy of all testing images (Acc.), average accuracy per patient case (Acc. per Case), and average accuracy per patient case using the proposed dispatch strategy (Acc. per Case-D). (T: Type, P: Polarity.)

Method	F_1 (%)		Acc. (%)		Acc. per Case (%)		Acc. per Case-D (%)
	T	P	T	P	T	P	T
Sharma et al. [Sharma et al., 2017]	92.0±1.6	–	92.6±1.5	–	92.6±7.9	–	93.6±7.4
Gupta et al. [Gupta et al., 2017]	91.2±2.3	–	91.8±2.2	–	91.8±9.9	–	92.6±9.5
AlexNet [Krizhevsky et al., 2012]	90.2±1.9	97.1±0.5	90.8±1.8	97.1±0.5	90.8±9.5	97.1±3.9	92.4±9.1
GoogLeNet [Szegedy et al., 2016]	95.6±1.6	98.6±0.5	96.0±1.5	98.6±0.5	96.0±6.5	98.6±2.7	96.8±6.2
VGG-Net [Simonyan and Zisserman, 2014]	96.0±0.7	98.8±0.2	96.3±0.6	98.8±0.2	96.3±5.3	98.8±2.2	97.1±4.9
ResNet [He et al., 2016a]	96.6±0.9	98.9±0.2	96.9±0.9	98.9±0.2	96.9±4.7	98.9±2.1	97.5±4.2
DenseNet [Huang et al., 2017a]	96.2±1.3	98.8±0.4	96.5±1.2	98.8±0.4	96.5±5.4	98.8±2.2	97.3±4.9
AlexNet-STN [Krizhevsky et al., 2012] [Jaderberg et al., 2015]	92.9±2.1	97.8±0.5	93.4±2.0	97.8±0.5	93.4±7.4	97.8±3.3	94.7±6.9
GoogLeNet-STN [Szegedy et al., 2016] [Jaderberg et al., 2015]	90.7±1.8	97.4±0.3	91.2±1.8	97.4±0.3	91.2±9.8	97.4±3.8	93.1±9.5
VGG-Net-STN [Simonyan and Zisserman, 2014] [Jaderberg et al., 2015]	96.8±0.8	99.0±0.3	97.1±0.8	99.0±0.3	97.0±4.4	99.0±1.9	97.7±4.1
ResNet-STN [He et al., 2016a] [Jaderberg et al., 2015]	96.9±0.9	98.9±0.2	97.2±0.9	98.9±0.2	97.2±3.7	98.9±1.8	97.8±3.3
DenseNet-STN [Huang et al., 2017a] [Jaderberg et al., 2015]	97.0±1.5	99.0±0.4	97.3±1.4	99.0±0.3	97.3±4.4	99.0±1.8	97.9±3.9
G-Net-STN [Jaderberg et al., 2015]	95.9±1.8	98.7±0.4	96.2±1.6	98.7±0.4	96.2±5.6	98.7±2.2	97.2±5.0
L-Net (Simple)	95.3±0.8	98.3±0.4	95.8±0.7	98.3±0.4	95.8±4.9	98.3±2.5	97.0±4.1
Varifocal-Net (Simple)	96.1±0.6	98.3±0.2	96.5±0.5	98.3±0.2	96.5±4.7	98.3±2.4	96.6±4.7
Varifocal-Net	**98.7±0.7**	**99.2±0.3**	**98.9±0.7**	**99.2±0.3**	**98.9±2.3**	**99.2±1.5**	**99.2±2.1**

Table 2.9: Comparison results of the proposed method with state-of-the-art methods on unhealthy cases (mean±standard deviation). The results are presented in terms of four evaluation metrics: average F_1-score of all testing images (F_1), accuracy of all testing images (Acc.), average accuracy per patient case (Acc. per Case), and average accuracy per patient case using the proposed dispatch strategy (Acc. per Case-D). (T: Type, P: Polarity.)

Method	F_1 (%)		Acc. (%)		Acc. per Case (%)		Acc. per Case-D (%)
	T	P	T	P	T	P	T
Sharma et al. [Sharma et al., 2017]	88.4±3.5	–	88.9±3.5	–	88.9±12.9	–	90.3±12.5
Gupta et al. [Gupta et al., 2017]	90.6±1.0	–	90.8±1.0	–	90.8±14.7	–	92.3±14.1
AlexNet [Krizhevsky et al., 2012]	80.7±5.2	93.8±2.2	81.1±5.3	94.0±2.1	81.2±18.6	94.0±7.1	83.6±18.2
GoogLeNet [Szegedy et al., 2016]	89.0±4.8	96.1±2.2	89.3±4.7	96.1±2.2	89.2±16.6	96.1±7.5	90.5±16.2
VGG-Net [Simonyan and Zisserman, 2014]	92.1±2.4	97.6±1.1	92.5±2.5	97.6±1.1	92.5±10.8	97.6±4.0	93.7±10.0
ResNet [He et al., 2016a]	93.5±1.7	98.0±1.2	93.8±1.8	98.0±1.2	93.8±9.7	98.0±3.8	94.7±9.0
DenseNet [Huang et al., 2017a]	92.3±2.8	97.7±1.0	92.7±2.7	97.8±1.0	92.6±11.4	97.8±4.0	93.7±10.9
AlexNet-STN [Krizhevsky et al., 2012] [Jaderberg et al., 2015]	87.2±5.5	95.7±2.1	87.6±5.6	95.8±2.1	87.5±15.1	95.8±6.1	89.2±15.1
GoogLeNet-STN [Szegedy et al., 2016] [Jaderberg et al., 2015]	81.0±4.4	94.2±1.8	81.5±4.6	94.3±1.8	81.5±18.7	94.2±8.6	83.4±19.6
VGG-Net-STN [Simonyan and Zisserman, 2014] [Jaderberg et al., 2015]	94.4±2.6	98.4±0.8	94.7±2.5	98.4±0.8	94.7±8.3	98.5±2.9	95.7±8.2
ResNet-STN [He et al., 2016a] [Jaderberg et al., 2015]	96.0±0.5	98.6±0.7	96.2±0.5	98.6±0.6	96.2±5.6	98.6±2.7	96.8±5.0
DenseNet-STN [Huang et al., 2017a] [Jaderberg et al., 2015]	95.8±2.9	98.5±0.9	96.0±2.8	98.5±0.9	96.0±6.7	98.5±2.4	96.7±6.0
G-Net	95.4±1.4	98.1±0.8	95.6±1.4	98.1±0.7	95.6±7.3	98.1±3.5	96.3±6.6
L-Net	96.6±1.3	98.5±0.5	96.8±1.3	98.6±0.5	96.8±5.8	98.6±2.4	97.5±5.2
G-Net-STN [Jaderberg et al., 2015]	93.4±3.2	97.8±1.2	93.6±3.0	97.8±1.2	93.6±9.3	97.8±3.5	94.7±8.8
L-Net (Simple)	94.5±1.7	97.7±0.8	94.8±1.6	97.8±0.8	94.8±6.7	97.8±3.0	95.8±6.6
Varifocal-Net (Simple)	95.1±1.1	97.5±0.4	95.3±0.9	97.5±0.4	95.3±5.3	97.5±3.1	95.4±5.3
Varifocal-Net	**97.7±1.6**	**98.6±0.6**	**97.8±1.7**	**98.6±0.6**	**97.8±4.3**	**98.6±2.5**	**98.4±3.9**

dramatically on abnormal cases. The AlexNet and GoogLeNet-STN even suffered over 9% loss of accuracy and F_1-score. In contrast, our Varifocal-Net had only a slight performance drop around 1.1% and 0.6% in Acc. per Case of the type and polarity task, respectively. We remarkably outperformed state-of-the-art methods on abnormal chromosome classification.

In Fig. 2.7, the results of ROC analysis are illustrated for both type and polarity classifications. We first performed ROC analysis per class using a one-vs-all scheme. Then, we averaged all ROC curves over classes and calculated the AUC for each method. It is observed that the proposed Varifocal-Net outperformed other methods with the least false positive predictions and the highest true positive rates. We achieved the highest AUC for both the type and polarity tasks. It demonstrates that in the case of not redesigning a completely brand-new feature extraction architecture, our Varifocal-Net, which benefited from the global and local feature ensemble, could further boost the overall classification performance. The lowest three AUCs of the type task were observed for [Krizhevsky et al., 2012], [Sharma et al., 2017], [Gupta et al., 2017], and simple processing methods, which is consistent with Table 2.8. Furthermore, statistical tests were performed using both unpaired and paired t-tests [Samuels et al., 2003, Hsu and Lachenbruch, 2014]. The Acc. per Case of all five fold testing samples were tested and the results of two t-tests confirm the significant superiority of the proposed Varifocal-Net against all other methods (p-value $\ll 0.05$) for both type and polarity tasks.

Performance Analysis Results

In this section, we present further experiment results of performance analysis. We computed the confusion matrix to get explanatory insights into the results of type prediction. As shown in Fig. 2.8, the confusion between class Y and classes No. 13, No. 15, No. 18, No. 21, and No. 22 mainly contributes to the performance drop.

We probed the embedded representations, including the global, local, and concatenated features, in order to illustrate their discrimination capability. We applied the t-SNE [Maaten and Hinton, 2008] approach on testing samples' features to reduce their dimensionality for 2-D visualization. As shown in Fig. 2.9, the testing samples were clustered by categories and separately dispersed for the concatenated features, with only a small set of samples mixed together. In contrast, for the single-scale global or local features, there exist many large regions where samples of different classes blend together. Compared to Fig. 2.9(e) and (f), the distance between adjacent clusters in Fig. 2.9(a)–(d) is smaller. The clusters of global-scale or local-scale features are less compact than that of multi-scale features, making it hard to find a clear boundary for differentiation.

Figs. 2.10 and 2.11 illustrate typical examples of correctly and incorrectly classified chromosomes, respectively. Fig. 2.10 shows that our varifocal mechanism can precisely locate the target region and capture the most discriminative local part with appropriate position and size. For small chromosomes, the predicted box can cover the whole body, while for larger chromosomes, the localization subnet selects partial segments of interest to facilitate accurate recognition.

In Fig. 2.11, misclassified samples are accompanied with their top 5 probabilities for wrong type predictions and 2 probabilities for wrong polarity predictions. It is observed that for most incorrect predictions, the probability of the true label ranks just the sec-

ond highest in order. Besides, some chromosomes are grossly distorted or have unusual shapes of their kinds, increasing the difficulty of accurate classification.

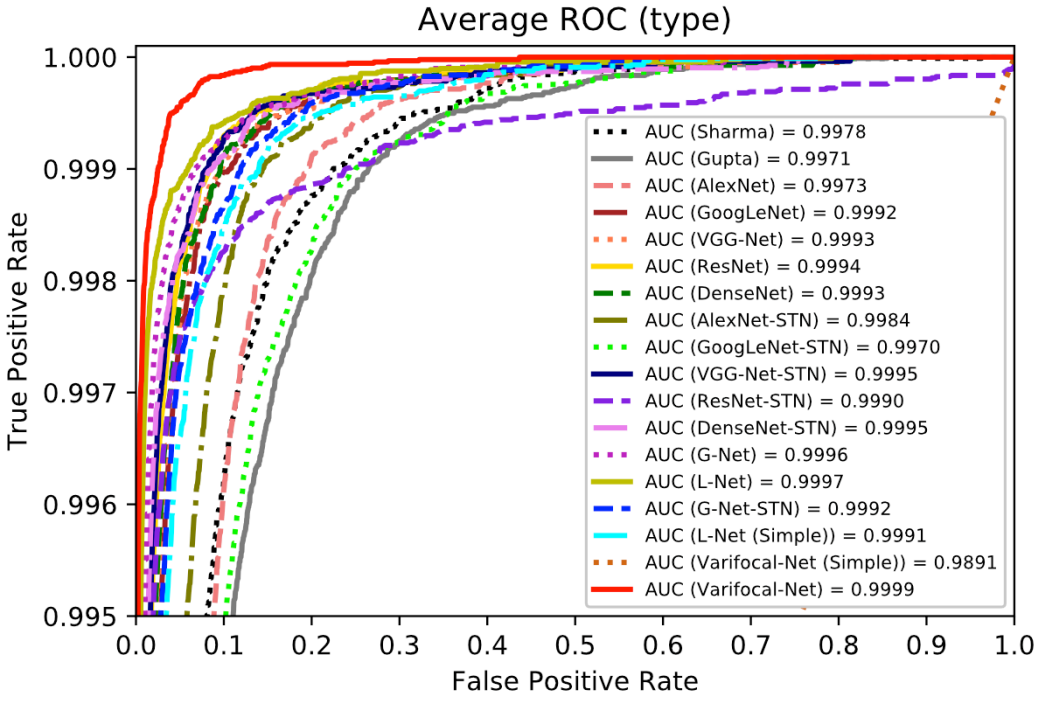
2.4 Discussion

In this chapter, a three-stage CNN method was proposed for chromosome classification. Its most distinctive characteristics include: 1) the adoption of varifocal mechanism to detect local discriminative regions; 2) the introduction of residual learning and multi-task learning to facilitate feature extraction; 3) the ensemble of global and local features to boost performance; 4) the use of a dispatch strategy for type assignment in practical karyotyping per case.

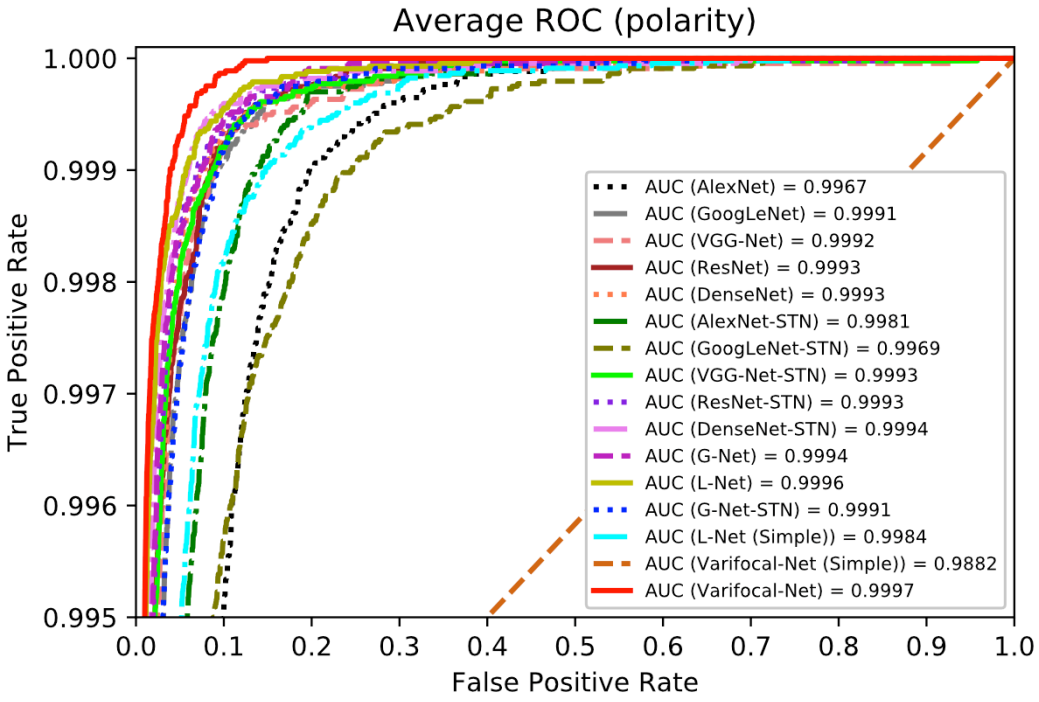
There are mainly two reasons contributing to the inferior performance of the previous CNN-based methods [Sharma et al., 2017, Gupta et al., 2017]. One is the loss of fidelity caused by the straightening step in their pipelines. Although this step is designed to rectify the shape of chromosomes for normalization, it damages the chromosome's morphological consistency and structural information due to inaccurate medial axis extraction and pixel interpolation. In contrast, the proposed Varifocal-Net is an end-to-end method without any shape correction in advance. The other is the lack of large labeled dataset. Their CNNs, which are designed on small datasets, cannot effectively describe the diversity and variety of chromosomes. Hence, these methods lack generality when evaluated on a large testing set.

As observed from the comparison results in Table 2.8, the potent CNN models [Krizhevsky et al., 2012, Szegedy et al., 2016, Simonyan and Zisserman, 2014, He et al., 2016a, Huang et al., 2017a, Jaderberg et al., 2015] performed well because we adapted them into the same settings as ours. In experiments, we adopted multi-task learning and applied necessary normalization on images. Hence, the performance difference among these models, to a certain extent, reflects the difference of their capabilities of global feature extraction. With respect to their accuracy, there exists a bottleneck of improvement for such single-scale models, which inspired us to resort to multi-scale feature ensemble. With the design of the proposed Varifocal-Net, we keep two aims in mind: the excellent feature extraction ability for classification and the strong discrimination of finer regions detected by the localization subnet. Since residual units are employed as the backbone of feature extraction CNNs, the proposed method benefits from the introduction of residual learning. Besides, the multi-task learning strategy also contributes to training the network. For the localization, the varifocal mechanism autonomously focuses on the local part which boosts local feature learning. We can see from Fig. 2.9 that the integration of both global and local features makes samples of the same category gather closely. It increases between-class distance and reduces chaotic outliers, which explains why our method is superior to the models that only count upon global-scale features.

Additional comparison on unhealthy cases (see Table 2.9) demonstrated the superior robustness of our method on abnormal chromosome classification. Compared with those single-scale models, our Varifocal-Net utilizes local-scale detail depiction to make up the deficiency of mere consideration of coarse-grained features. Compared to the Varifocal-Net, there does exist a larger performance decrease for the only G-Net and the only L-Net, which confirms the importance of multi-scale feature ensemble strategy. Hence, the



(a)



(b)

Figure 2.7: ROC analysis for the proposed Varifocal-Net and previous CNN models. Each ROC is averaged over all classes and its AUC is calculated. (a) ROC of type classification. (b) ROC of polarity classification.

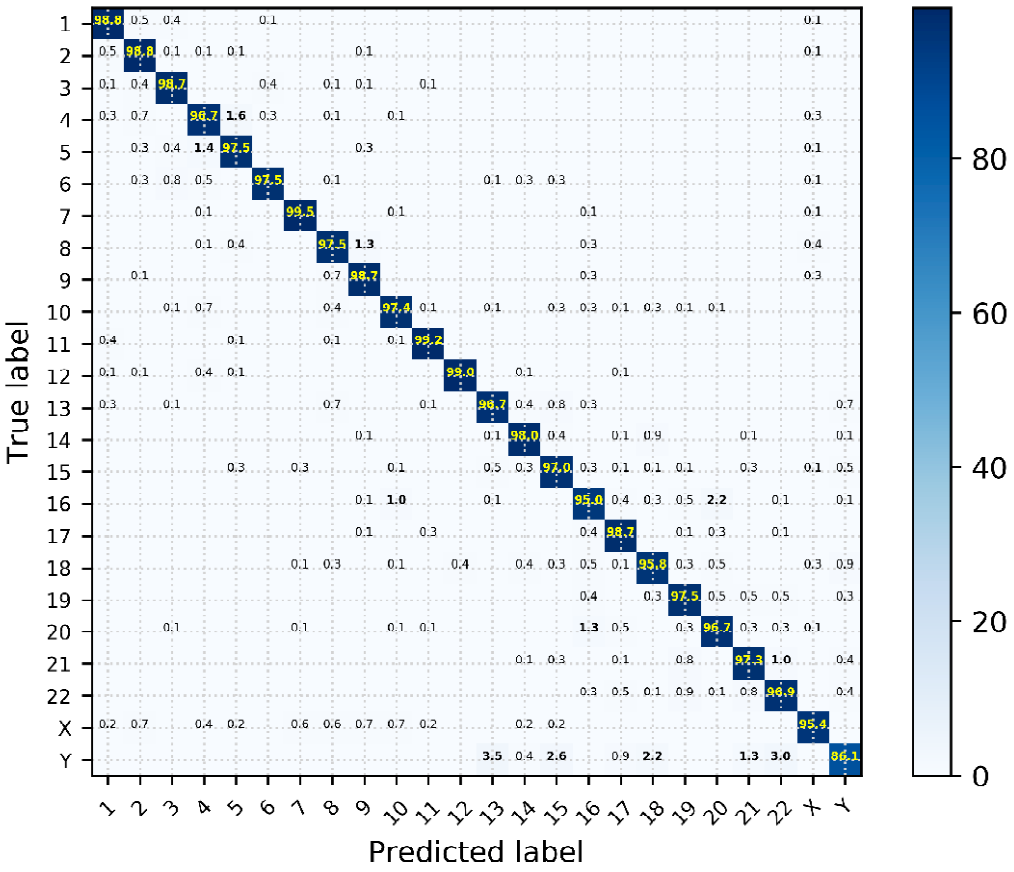


Figure 2.8: Confusion matrix of the Varifocal-Net for type classification. The entry in the i-th row and j-th column denotes the percentage (%) of the testing samples from class i that were classified as class j. Best viewed magnified.

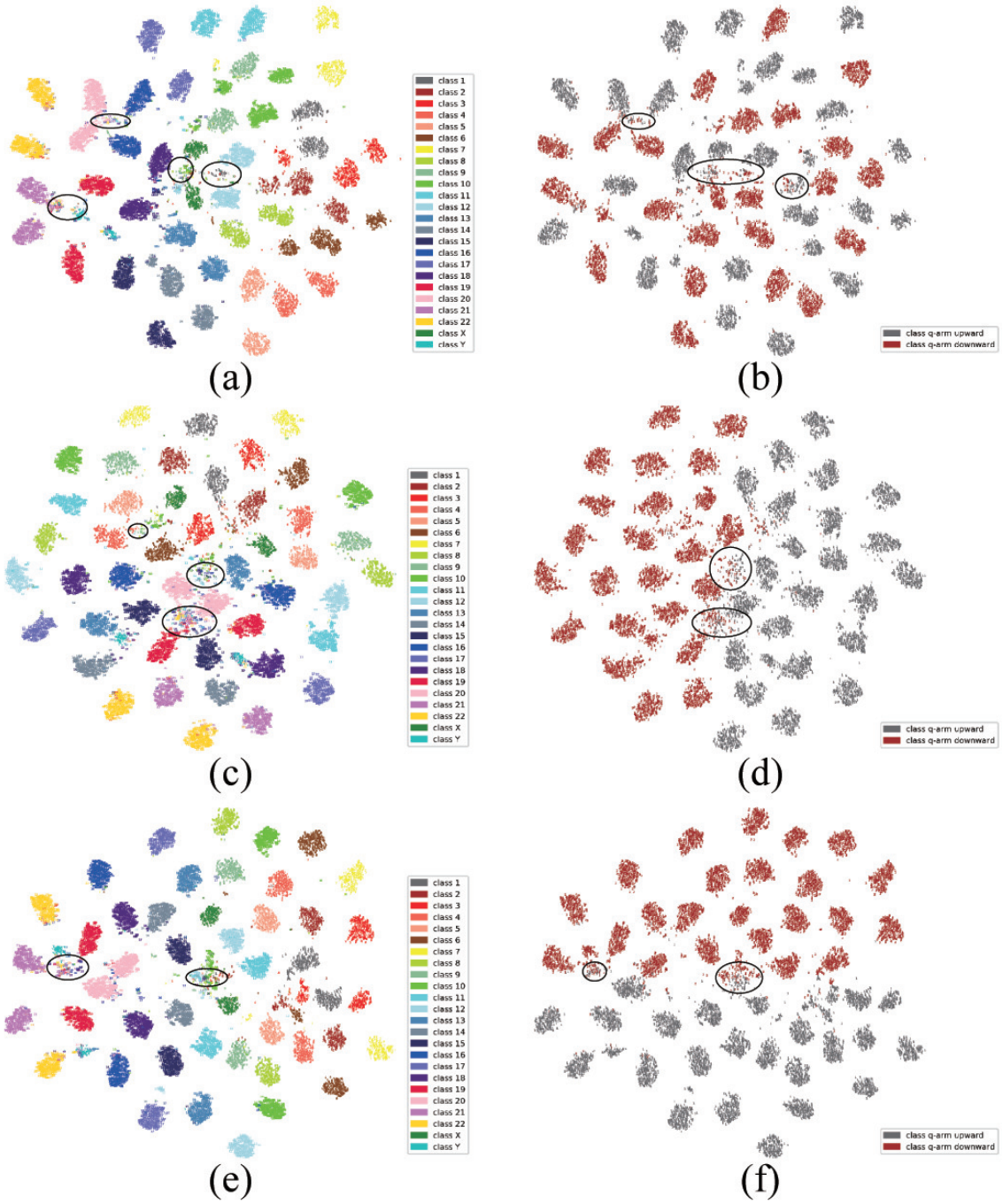


Figure 2.9: Feature embedding for chromosomes with t-SNE toolbox [Maaten and Hinton, 2008]. From the perspective of type classification, the global, local, and concatenated features are visualized in (a), (c), and (e), respectively. Similarly, these three features are visualized in (b), (d), and (f) correspondingly for polarity classification. The mixed regions of interest are marked with black circles. Best viewed in color.

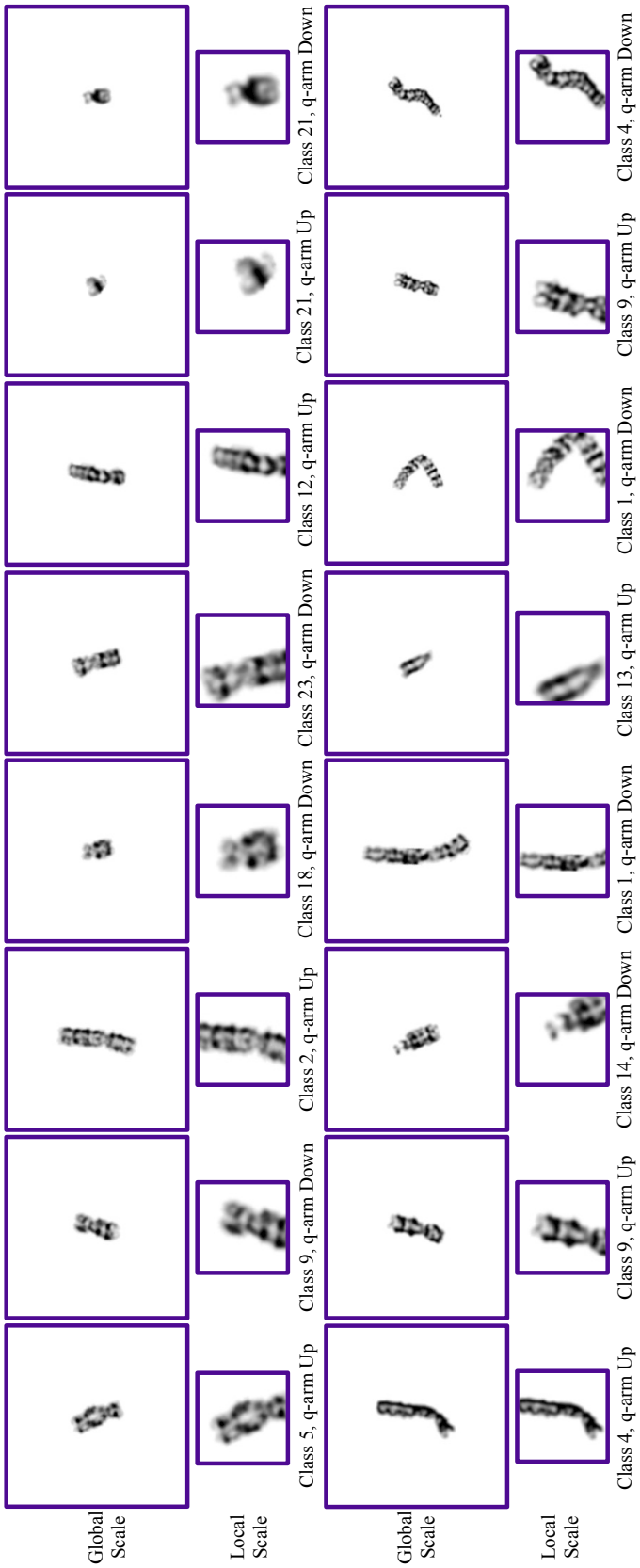


Figure 2.10: Examples of correctly classified samples. Both global-scale and local-scale inputs are displayed to visually assess the varifocal mechanism.

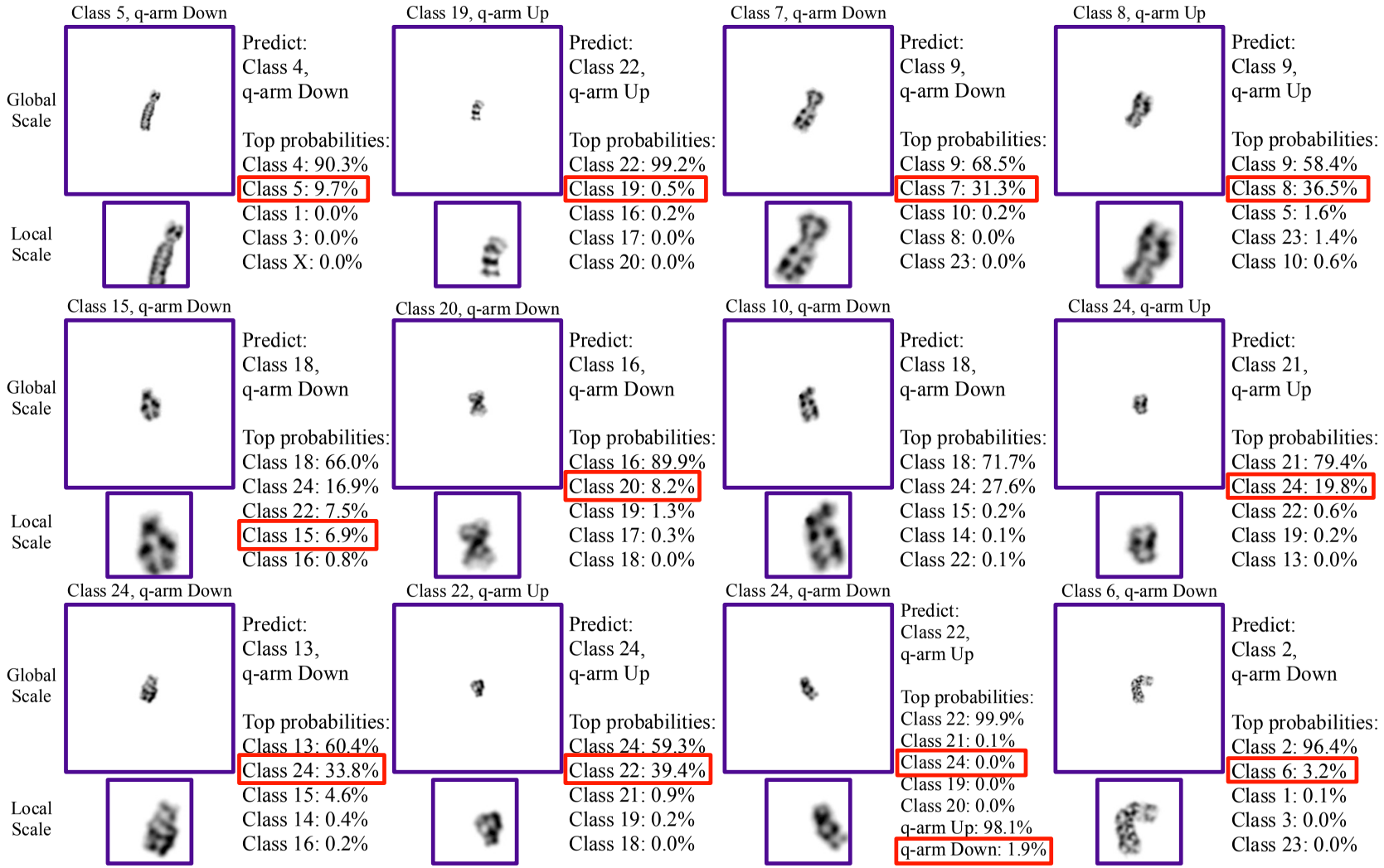


Figure 2.11: Examples of misclassified samples. The probabilities of wrong predictions are displayed on the right of each image and each red rectangle encloses the predicted probability of the ground-truth label.

proposed method, which possesses excellent generalization abilities, can assist doctors in the clinical karyotyping process where abnormal cases occur from time to time.

The performance improvement of Acc. per Case-D with respect to Acc. per Case in type classification substantiates that the proposed dispatch strategy is effective and suitable for karyotyping within each case. For each method in Table 2.8 and Table 2.9, the adoption of the dispatch strategy improves the average accuracy and diminishes the standard deviation for both healthy and unhealthy cases. The generalizability of such strategy lies in the consideration of both maximum likelihood criterion and chromosomal numerical abnormalities.

The proposed method performed less well on chromosomes No. 15, No. 21, and No. 22 (see Table 2.5 and Fig. 2.8) with respect to other classes. Such three kinds of chromosomes are acrocentric and contain a segment called satellite, which is separated from the main body. The shape, size, and orientation of satellites differ from one person to another, thus making it difficult for our model to handle all possible situations. Fig. 2.8 also shows that chromosome Y is often confused with No. 21 and No. 22. It is because the size and texture of class Y are similar to that of No. 21 and No. 22. Furthermore, the comparatively imbalanced Y samples are not processed with additional data augmentation method, which triggers off poorer recognition of Y. It is noted that although we collected a much larger dataset than previous work, the dataset is still insufficient to cover all possible shapes and sizes of chromosomes. Samples of sex chromosome Y and diversified satellite chromosomes are still in shortage. Therefore for better performance, more data should be collected and generative adversarial networks could be used for sample synthesis in the future.

From the results of Table 2.7 and examples in Fig. 2.11, it is observed that some long chromosomes (e.g., No. 2 and No. 5) may be misclassified because their long arms tend to bend or distort greatly during the sampling process. Since the proposed method cannot accurately recognize greatly bent chromosomes, future work may involve particular strategies to cope with this situation. Instead of straightening the chromosomes, we might inform the network of the degree of bending deformation by detecting the rotation pivot (e.g., the centromere) and its angle between two arms. Furthermore, for the G-Net and the L-Net, current feature extractor employs the residual block as a backbone. To further improve performance, we may meticulously redesign the network architecture.

2.5 Conclusion

In conclusion, we have proposed the Varifocal-Net for chromosome classification, which has been evaluated on a large manually constructed dataset. It is a three-stage CNN-based method. The first stage effectively learns global and local features through the G-Net and the L-Net, respectively. Taking a global-scale chromosome image as the input, it precisely detects a local region that is discriminative and abundant in details for further feature extraction. The second stage robustly differentiates chromosomes into various types and polarities via two MLP classifiers. It benefits from multi-scale feature ensemble, with only a few misclassifications. In the third stage, a dispatch strategy was employed to assign each chromosome to a type based on its predicted probabilities. Extensive experimental results demonstrate that our approach outperforms state-of-the-art

methods, corroborating its high accuracy and generalizability.

Concerning its role in clinical karyotyping workflow, the Varifocal-Net can accurately perform classification within 1 second after operators manually segment chromosomes of a cell for each patient. The karyotyping result maps it automatically generates offer the possibility for human experts to further check and correct possible misclassifications. Moreover, warnings about possible numerical abnormalities allow operators to pay extra attention to the subsequent diagnosis. The practical use of the Varifocal-Net in the Xiangya Hospital of Central South University suggests its promising potential for alleviating doctors' workload in the diagnosis process.

Chapter 3

Pulmonary Nodule Segmentation with CT Sample Synthesis Using Adversarial Networks

Contents

3.1 Introduction . . . 108
3.2 Methodology . . . 111
3.2.1 Synthetic Image Generation . . . 111
3.2.2 Pulmonary Nodule Segmentation . . . 116
3.3 Experiments and Results . . . 120
3.3.1 Materials . . . 120
3.3.2 Implementation Details . . . 121
3.3.3 Evaluation Metrics . . . 121
3.3.4 Results . . . 122
3.4 Discussion . . . 125
3.5 Conclusion . . . 129

3.1 Introduction

Pulmonary cancer has been one of the leading cancers in both men and women and annually causes 1.3 million deaths worldwide [Torre et al., 2016]. Although the overall 5-year survival rate is only 18%, if early diagnosis and treatment are put into effect timely, the patients' chances of survival can be greatly increased [Siegel et al., 2016]. Pulmonary nodules are small masses in lung and often viewed as an early indication of cancer. The wide-spread use of computer tomography (CT) helps radiologists make accurate diagnosis of nodules. However, due to the high demand for CT scanning and similarity of nodules to lung tissue (e.g., blood vessels and bronchi), it may take radiologists long reading time to analyze suspicious lesions. Therefore, computer-aided diagnosis (CAD) systems are developed to improve doctors' reading efficiency.

Many current CAD systems focus on the detection of pulmonary nodules in CT [Messay et al., 2010, Lopez Torres et al., 2015, Jacobs et al., 2014, Setio et al., 2016, Sakamoto and Nakano, 2016, Dou et al., 2017, Huang et al., 2017b]. These CAD systems process CT images and predict the coordinates of bounding boxes that contain suspicious nodules. However, bounding box alone is not sufficient. In clinical practice, radiologists need to measure volumetric changes of nodules to estimate their malignancy likelihood effectively [Yankelevitz et al., 2000, de Hoop et al., 2012, Wilson et al., 2012, Goodman et al., 2006], which requires manual delineation of nodules' boundaries. The pixel-level manual segmentation by radiologists is time-consuming since nodules differ in size (diameter ranging from 3 to 30 mm), shape, brightness, and compactness [Setio et al., 2016]. Therefore, it is imperative to develop CAD systems for accurate and robust nodule segmentation.

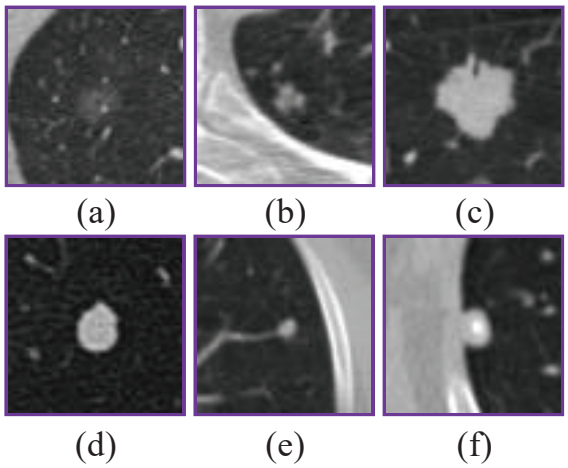


Figure 3.1: Typical cases for each nodule type. First row: Nodules are classified by internal texture. (a) GGO; (b) part-solid; (c) solid. Second row: Nodules are classified by external surroundings. (d) well-circumscribed; (e) juxta-vascular; (f) juxta-pleural.

The main difficulty in nodule segmentation is to design an algorithm that adapts to both internal texture and external surroundings of pulmonary nodules. According to the variation in internal texture characteristics, lung nodules can be classified into the categories: solid, part-solid, and ground glass opacity (GGO). The solid nodules exhibit

explicit shapes and margins while GGO nodules are of low contrast and have fuzzy boundaries. The part-solid nodules fall in between. Pulmonary nodules can also be classified into the categories: well-circumscribed, juxta-vascular, and juxta-pleural. The well-circumscribed nodules stay inside the lung alone. The juxta-vascular nodules and the juxta-pleural nodules connect vascular structures and pleural surfaces, respectively. Typical cases for each category are shown in Fig. 3.1.

In the past, several methods have been proposed to mainly segment on solid nodules [Dehmeshki et al., 2008, Diciotti et al., 2011, Reeves et al., 2006, Wang et al., 2007]. Dehmeshki et al. [Dehmeshki et al., 2008] employed a 3D region growing method for user-interactive segmentation. Their method performs a sphericity-oriented contrast region growing on the fuzzy connectivity map of the target object. It combines distance and intensity information as growing conditions. Diciotti et al. [Diciotti et al., 2011] developed an automated method to refine initial rough segmentation results of small juxta-vascular solid nodules. The rough segmentation is corrected by 3D local shape analysis, which removes vessel attachments with nodule boundaries preserved. GGO nodules are not considered in their work. Reeves et al. [Reeves et al., 2006] designed an iterative method to separate a nodule from the pleural surface using plane fitting technique. Adaptive thresholding is then applied to adjust segmentation. Wang, Engelmann, and Li [Wang et al., 2007] proposed a segmentation method that transforms 3D volume of interest (VOI) into 2D images using a spiral-scanning technique. The optimal outlines of nodules in 2D images are delineated by dynamic programming method. Then, they are transformed back to 3D images for surface reconstruction.

Few methods were developed for segmentation of all solid, part-solid, and GGO nodules [Kubota et al., 2011, Qiang et al., 2014, Mukhopadhyay, 2016]. Kubota et al. [Kubota et al., 2011] proposed a general segmentation method. It combines morphological operation and convexity models to segment on juxta-vascular and juxta-pleural nodules without separating lung walls. Qiang et al. [Qiang et al., 2014] employed a scheme that utilizes freehand sketch analysis. Nodules are automatically segmented with an improved shape break-and-repair strategy. Mukhopadhyay [Mukhopadhyay, 2016] adopted a two-step segmentation method. It first categorizes nodules by internal texture. Then, vascular structures and pleural surfaces are removed. The method was evaluated on LIDC-IDRI public database [Armato et al., 2011].

With the development of convolutional neural networks (CNNs) [LeCun et al., 1998, Krizhevsky et al., 2012, He et al., 2016a], researchers tended to employ CNNs for segmentation in an end-to-end manner [Long et al., 2015, Ronneberger et al., 2015, Milletari et al., 2016]. However, at the moment, only one method is reported on adopting CNNs for nodule segmentation. Wu et al. [Wu et al., 2018a] developed a 3D CNN model for segmentation of pulmonary nodules from VOI. They evaluated the method on LIDC-IDRI dataset and achieved Dice coefficient of 0.7405.

Despite the fact that there exists an interest in designing CAD systems based on deep learning techniques, the performance of these systems is limited by the availability of large labeled datasets. Medical data are not easy to access due to privacy issues. In addition, it is laboursome for doctors to collect, organize, and annotate them, making the size of dataset restricted. Motivated by recent development of generative adversarial networks (GAN) [Goodfellow et al., 2014, Mirza and Osindero, 2014, Radford et al., 2015, Shrivastava et al., 2017, Isola et al., 2017, Guibas et al., 2017], we believe synthetic image

generation may be a good choice in the face of the underlying problem of imbalanced and limited data. In order to build a more balanced and diverse dataset, we capitalize on generating nodule CT images through adversarial networks, which is not considered in previously reported works.

In this chapter, we propose a CNN-based framework for pulmonary nodule segmentation. By adopting adversarial networks, synthetic samples are generated to achieve a more balanced training dataset. With interpretable feature maps incorporated and residual learning strategy introduced, the segmentation model performs robustly on all kinds of nodules without radiologists' manual intervention. The main contributions are as follows: (1) We employ a conditional GAN that generates nodule CT images to extend the LIDC-IDRI dataset. Since original annotation is only the boundary of each nodule, we design a method to obtain ten-channel semantic labels of nodule patches. These labels not only contain contextual information but also represent nodules' semantic attributes. Based on semantic labels, synthetic samples are generated through adversarial networks. The $L2$ reconstruction error loss is introduced into cGAN to increase the realism of generated samples. The imbalanced data problem is alleviated by such expansion of dataset, which prevents overfitting for the training of segmentation model. Hence, the performance of our segmentation method gets improved. (2) We propose a 3D CNN model that accurately segments pulmonary nodules. To generate segmentation masks, a 3D U-Net [Çiçek et al., 2016] similar network is exploited. Multiple heterogeneous maps, including edge maps and texture feature maps, are introduced as inputs and leveraged by the CNN model to learn high-level features. For edge maps, we apply Canny operator [Canny, 1986] and Sobel operator [Sobel, 1990] to detect the edges of nodule images, which lay a foundation for the task of segmentation. Local binary patterns (LBP) [Ojala et al., 2002] are chosen to capture spatial structure of nodules' textures. Since there exists a great difference in textures between solid, part-solid, and GGO nodules, these texture feature maps are considered informative for the network to generate accurate segmentation results for each kind of nodule. The 3D architecture of our model aims at better utilizing volumetric knowledge of 3D CT images. Besides, residual learning is employed to resolve vanishing gradient problem. It promotes effective feature learning and accelerates training process. (3) The proposed CNN-based segmentation framework is evaluated on the public LIDC-IDRI dataset.

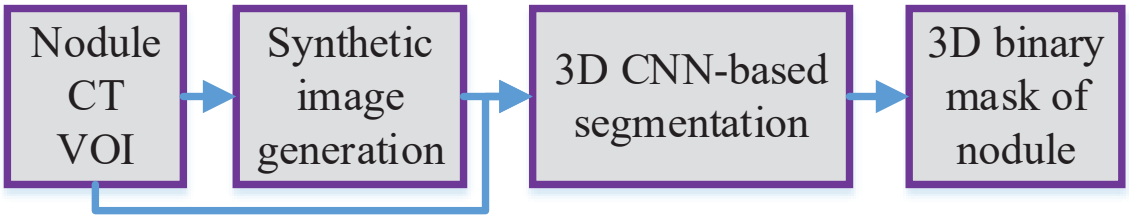


Figure 3.2: An overview of the proposed pulmonary nodule segmentation framework. Synthetic nodule images are first generated. Then, both the original and synthesized images are used to train the segmentation model. The segmentation results are 3D binary masks of nodule VOI.

3.2 Methodology

The developed pulmonary nodule segmentation framework is composed of two parts (see Fig. 3.2): (1) Synthetic image generation and (2) 3D CNN-based segmentation. For the first part, adversarial networks are adopted to enhance the diversity of nodule samples and mitigate the problem of imbalanced and limited data. The second part is designed to segment all kinds of nodules from VOI using a 3D CNN model. The details of the proposed framework is presented as follows.

3.2.1 Synthetic Image Generation

Table 3.1: Distributions of the 1182 pulmonary nodules from the LIDC-IDRI dataset.

	Category	No. of nodules
Texture	Solid	927
	Part-solid	188
	GGO	67
Diameter	$< 6\ mm$	38
	$6 \sim 10\ mm$	424
	$> 10\ mm$	720
In total		1182

In the field of medical image segmentation, it is often inevitable that collected samples are imbalanced and biased, posing challenges to the generalization of segmentation methods. Especially for pulmonary nodule segmentation, even the largest public dataset LIDC-IDRI [Armato et al., 2011] is imbalanced in terms of nodule's texture and size (see Table 3.1). The number of solid nodules is three times as many as that of the rest. Large nodules constitute a great proportion of all nodules. Besides, GGO nodules and small nodules are so limited in quantity that they are easily overwhelmed by other nodules. Consequently, the segmentation model may suffer poor performance on the minority categories of nodules if trained on such dataset. To tackle this problem, synthetic image generation then appears as an interesting solution. To do that, slices that contain nodules are first selected from all cropped VOI cubes. A technique of transforming ground-truth labels into ten-channel semantic labels is then designed to introduce abundant contextual information about nodules. Finally, a conditional generative model is employed to translate semantic labels into realistic images.

Semantic Label Generation

The ground-truth labels from LIDC-IDRI dataset only describe shape, size, and attributes of nodules. These labels are sufficient for the task of nodule segmentation, but not for sample synthesis. It is difficult for generative adversarial networks to produce authentic images if only information about nodules is provided. The semantic knowledge of nodules' surroundings is of great importance since it depicts external attachments and nodules' relative position in thoracic cavity. For example, two nodules may have similar boundaries but one is attached to pleural surface and the other stays alone. Hence,

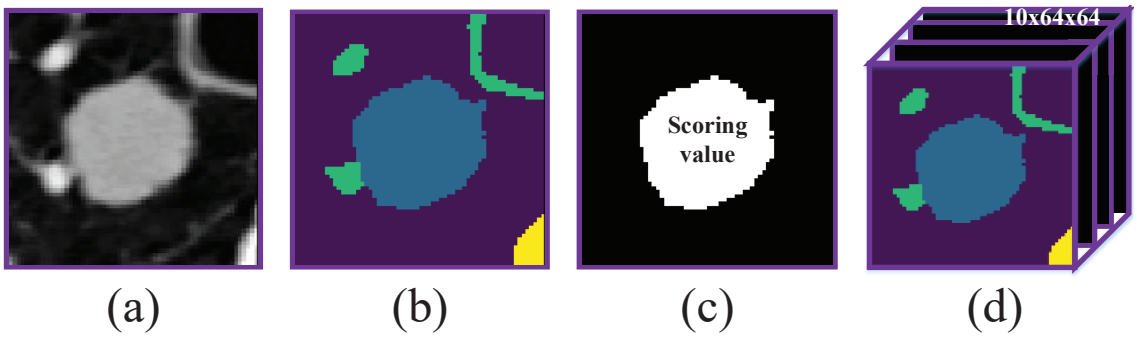


Figure 3.3: The process of generating ten-channel labels. (a) CT image of nodule; (b) Generated semantic label containing a nodule, pleural surfaces, and vascular structures; (c) Each attribute's scoring value is multiplied with a binary ground-truth label; (d) The semantic label and nine attribute labels are concatenated as an input image with ten channels.

in order to enable the network to learn from nodules' contextual information, semantic labels are generated as stated in the following. First, slices are thresholded to extract all components with high intensity. The thresholding value is determined by the category of nodule's internal texture. For GGO nodules whose scoring of texture is lower than three, the grayscale value of 60 is chosen. For part-solid and solid nodules, 70 is set as threshold value. These two values are calculated based on the studies of nodule's density distribution in CT [Kauczor et al., 2000, Zhao et al., 2003] and our clipping window of Hounsfield unit. Secondly, a disk-shape structuring element with radius of one is adopted for morphological opening. Each slice is opened to remove tiny objects and smooth image since only obvious parenchymal structures, including large vessels and pleural surfaces, are considered for labeling. Then, connect component analysis is employed to differentiate between vascular structures and pleural surfaces. For each connected component, if its area is larger than 640 or if it intersects at least two borders of image with a minimum area of 32, it is labeled as pleural surface with a value of 3. Other components are labeled as vascular structures with a value of 2. The ground-truth label of nodule is set as 1. This generated label is not accurate enough to directly train a semantic segmentation model, but it provides adequate nodule's surrounding knowledge for image synthesis.

In addition, nine semantic attributes (see Table 3.2) are introduced to describe nodules in more details. These features represent nodule's internal variation of intensity and shape, and are closely related to diagnosis. For each nodule, nine original binary ground-truth labels are multiplied with its nine attribute scorings respectively. Each label corresponds to one attribute scoring. Then, all nine attribute labels and one semantic label are concatenated together to form a ten-channel image as the input of cGAN. The input size is $10 \times 64 \times 64$ and the process of label generation is shown in Fig. 3.3.

Conditional Generative Adversarial Network

The initially proposed GAN [Goodfellow et al., 2014] learns to generate samples from the random noise vector. The noise z is passed explicitly into the generator as input. Different from the original GAN, the random noise z of cGAN [Mirza and Osindero, 2014] is

Table 3.2: Definition of scoring for each nodule attribute.

Attribute	Scoring					
	1	2	3	4	5	6
Subtlety	Extreme subtlety				Obvious	\
Internal Structure	Soft	Fluid	Fat	Air	\	\
Calcification	Popcorn	Luminated	Solid	Non-central	Central	Absent
Sphericity	Linear		Ovoid		Round	\
Margin	Poorly defined				Sharp	\
Lobulation	None				Marked	\
Spiculation	None				Marked	\
Texture	GGO		Part-solid		Solid	\
Malignancy	Highly unlikely	Moderately unlikely	Indeter-minate	Moderately suspicious	Highly suspicious	\

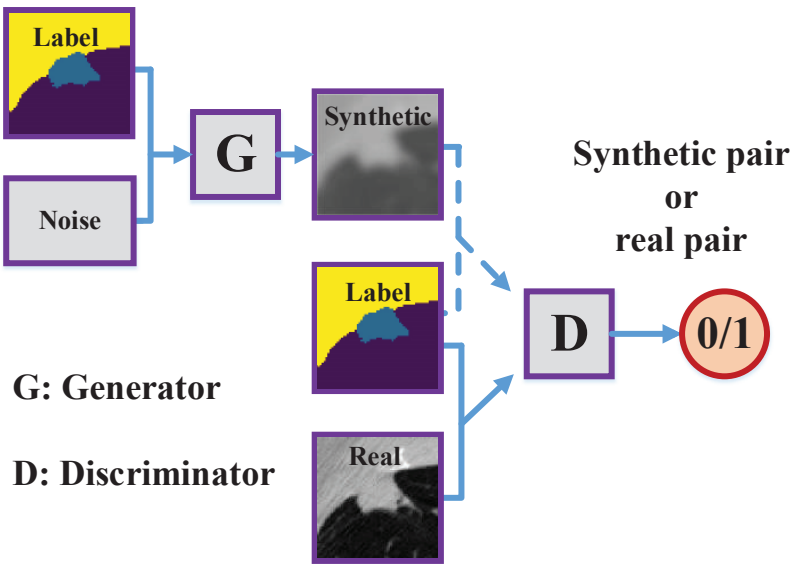


Figure 3.4: The training of cGAN proceeds by alternatively training G and D. Given a label image and a noise vector, G is trained to obtain a realistic image. The synthetic pair and real pair refer to the ten-channel label concatenated with synthetic image and real image, respectively. D learns to distinguish real pairs from synthetic fake pairs.

introduced into the generator during the process of generating samples. The cGAN maps z to the realistic CT image y in the conditional setting of a ten-channel semantic label x. The training of cGAN involves gaming between the generator model G and the discriminator model D. The objective function of the original cGAN [Mirza and Osindero, 2014] is defined as:

$$\begin{aligned} \mathcal{L}_{cGAN} &= \mathbb{E}_{x,y\sim p_{data}(x,y)}[\log D(y|x)] + \mathbb{E}_{x\sim p_{data}(x),z\sim p_z(z)}[\log(1 - D(G(z|x)|x))], \\ G,D &= \arg\min_G \max_D \mathcal{L}_{cGAN}, \end{aligned} \tag{3.1}$$

where p_z and p_{data} here denote the prior noise distribution and the real nodule data distribution, respectively. G tries to capture the nodule images' distribution with the condition of label x and its generated sample is $G(z|x)$. D estimates the probability that the current pair is real nodule data pair (x,y) rather than synthetic data pair $(x,G(z|x))$. G is trained by minimizing such adversarial loss while D by maximizing it. It is noted that G is optimized to output images that are difficult for D to distinguish from real ones. To directly guide G to produce samples that are similar to realistic images, we introduce $L2$ reconstruction error loss to the training of generator as the following:

$$\mathcal{L}_G = \mathbb{E}_{x,y\sim p_{data}(x,y),z\sim p_z(z)}[\|y - G(z|x)\|_2^2], \tag{3.2}$$

where the real nodule data y serve as the ground-truth for $G(z|x)$. Such $L2$ loss function penalizes the model to explicitly reduce the difference between real CT images and synthetic images. The modified objective function is given by:

$$G, D = \arg\min_{G}\max_{D} \mathcal{L}_{cGAN} + \lambda \mathcal{L}_{G}, \tag{3.3}$$

where λ is a weight balancing these two terms. We set $\lambda = 100$ in the present study. The adversarial training process is illustrated in Fig. 3.4. Note that the noise here, to a certain degree, can be viewed as an implicit input.

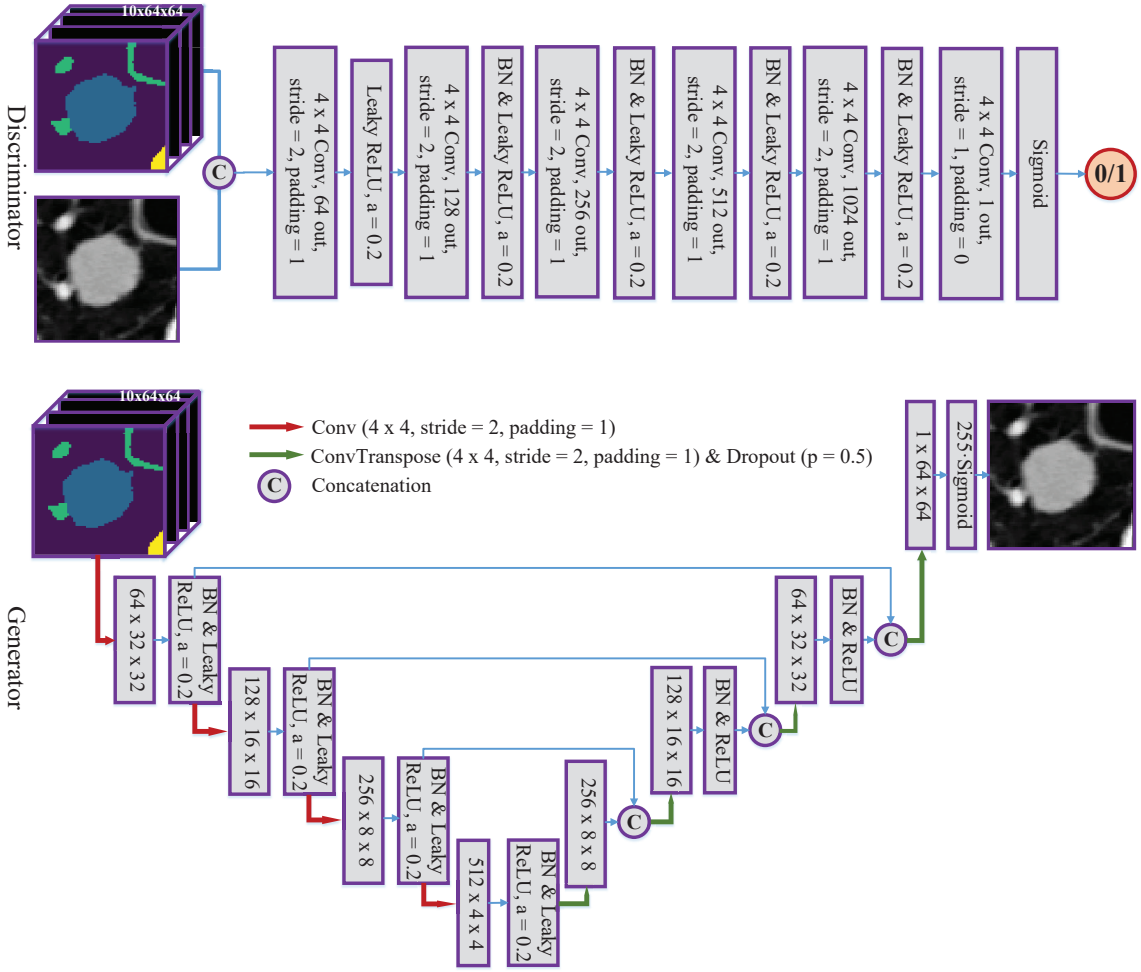


Figure 3.5: The network architecture of the proposed cGAN.

The architecture of our cGAN is depicted in Fig. 3.5. The 2D U-Net structure is used as a backbone to build a generative model, which generates synthetic images in an encoder-decoder fashion with skip paths. For the contracting path, instead of max-pooling layer used in the original U-Net [Ronneberger et al., 2015], strided convolution layer is adopted to downsample the image, followed by a batch normalization (BN) layer and a leaky rectified linear unit (ReLU) layer. For the expansive path, we employ transposed convolution to upsample feature maps to increase resolution and concatenate them with features from skip path. The BN layer, ReLU layer and dropout layer are also introduced. Then, a fully convolutional net (FCN) is designed as a discriminator model. Except the first layer, all strided convolution layers are followed by a BN layer and a leaky ReLU

layer. The pooling layer in both generator model and discriminator model is replaced by strided convolution because the latter learns to summarize the pixels within its kernel by a weighted element-wise multiplication. Different from max-pooling or avg-pooling, the way that strided convolution reduces feature dimensionality is not determined in advance but learnable during training.

As shown in Fig. 3.5, the noise z is implicitly taken as input to the generator. We use dropout layer on the expansive path to introduce noise [Isola et al., 2017] by randomly deactivating neurons with a probability of 0.5. Previous study on dropout layer [Park and Kwak, 2016] proves that such layer adds noise to the output features and thus improves robustness to the variation of input images. Furthermore, the dropout layer provides regularization to prevent over-fitting by reducing co-dependency among neurons. It randomly deactivates neurons during the training process, thereby preventing the model from learning interdependent set of feature weights [Goodfellow et al., 2016].

3.2.2 Pulmonary Nodule Segmentation

The overall nodule segmentation architecture is given in Fig. 3.6. As pulmonary nodules have different internal textures and segmentation method should adapt to such variety, we introduce texture maps to implicitly impart to the network the ability of apprehending whether current nodule is GGO, part-solid, or solid. In addition, edge maps are concatenated as inputs since they provide rich knowledge about margins and boundaries of nodule images, thereby assisting the task of segmentation. The 3D CNN segmentation model is an end-to-end model that exploits a 3D U-Net [Çiçek et al., 2016] similar structure. Residual learning is brought into the network to improve the performance of segmentation.

Heterogeneous Maps

Local Binary Pattern (LBP) [Ojala et al., 2002] characterizes the spatial structure of local image texture by encoding the difference between a center pixel and its neighboring pixels. We use LBP maps as the representation of nodule's texture to describe different types of nodules for the network. For each pixel in the original image, its LBP encoding is computed by thresholding neighboring pixels with its intensity:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^P, s(x) = \begin{cases} 1, \text{if } x \geq 0 \\ 0, \text{if } x < 0 \end{cases}, \tag{3.4}$$

where g_c and g_p are grayscale values of the center pixel and its surrounding pixels inside a circle with radius of R, respectively. The total number of neighboring pixels is P.

The LBP operator only considers the relative intensity of neighboring pixels with respect to the center pixel. Its value changes if rotation operation is implemented on the image. Since rotation is used for data augmentation, rotation-invariant LBP is preferred in order to extract essential characteristics of nodule's texture. Hence, we use the new type of LBP:

$$LBP^{ri}_{P,R} = \min\{ROR(LBP_{P,R}, i) | i = 0, 1, .., P-1\}, \tag{3.5}$$

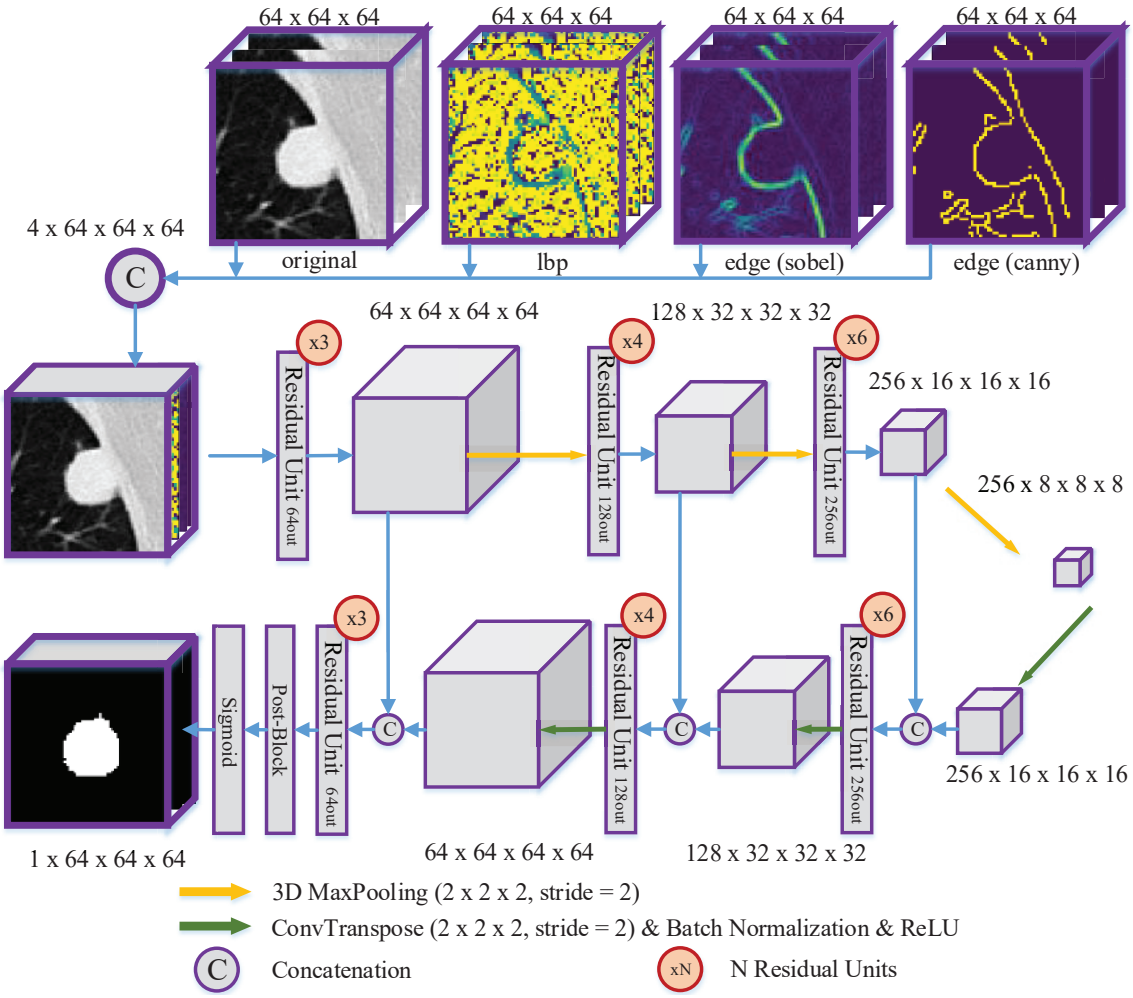


Figure 3.6: The network architecture of the proposed segmentation framework.

where $ROR(x,i)$ performs a circular bit-wise right shift on the encoded value x, i times. It can be viewed as texture feature detector to capture micro-features that are invariant to rotation. Furthermore, with rotation-invariant LBP texture maps fed into the network, the learned high-level CNN features are rotation-invariant as well. In the experiments, we set $P = 24$ and $R = 3$ and compute LBP maps slice by slice.

In the segmentation task, accurate detection of meaningful edges is fundamental. The edge map reflects the discontinuity of an image. Especially for the solid nodule that has a clear margin, there exists an abrupt change in intensity around nodule's border. Edge maps are used to filter out useless information and only preserve structure properties of images. For well-circumscribed nodules, the edge maps directly detect their boundaries. For nodules that connect pleural surface or vascular structures, their edge maps also provide the outlines of their attachment. These maps can be viewed as initial segmentation results, which are then polished up by our network for final precise results.

There are many methods for edge detection. In our experiments, two most widely used methods, Sobel [Sobel, 1990] and Canny [Canny, 1986] edge detectors, are employed together for each slice since their performance varies depending on the categories of nodules and the integrated use of both the methods is better than using one. For Sobel edge detection, two 3×3 kernels are convoluted with images to estimate the gradient in x and y directions. After convolution with horizontal and vertical kernels, two images of the approximated gradient of intensity are obtained as G_x and G_y. Then, the magnitude of gradient is computed as edge map.

For Canny edge detection, we first smooth the image using a 3×3 Gaussian filter to reduce noise. Gaussian filter is adopted because it is faster than other non-linear filters such as Median filter. Then, horizontal and vertical Sobel operators are applied to compute the magnitude and orientation of the gradient. After that, non-maximum suppression is performed on the magnitude map to suppress all gradient values except local maxima. Finally, two thresholding values t_1 and t_2, determined respectively as 10% and 20% of the maximum magnitude's value, are applied to threshold the edge map. All pixels with magnitude value higher than t_2 are labeled as edges. Pixels with value higher than t_1, which are also 8-connected to the labeled edge pixels, are recursively labeled as edges.

Segmentation Network

The input of the network is a four-channel 3D cube, which consists of four different cubes: cropped CT volume cubes, LBP maps, and two edge maps. The full 3D CNN architecture is developed to exploit spatial contextual knowledge for high-level feature extraction. Residual units, which consist of a few stacked layers, are introduced into the network. Given the input x of the residual unit, the underlying mapping to be fit by the layers is denoted as $H(x)$. Rather than directly approximating $H(x)$, these layers approximate a residual function $F(x) = H(x) - x$. By reducing such residual, it is easier to learn the underlying mapping. This learning strategy is known as residual learning. Specifically, we define a residual unit as:

$$y = \mathcal{F}(x, \{W_i\}) + x, \tag{3.6}$$

where x and y are the input and output of residual unit, respectively. $\mathcal{F}(x, \{W_i\})$ is a 3D

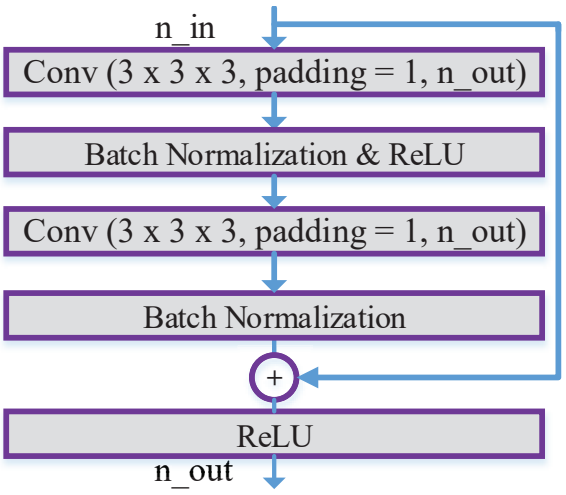


Figure 3.7: Residual unit. n_in and n_out denote the number of channels of input cube and output cube, respectively.

mapping to high-level features, which includes two convolution layers, two BN layers, and one ReLU layer. $\{W_i\}$ contains all learned parameters. Such residual unit allows gradient to propagate directly through a shortcut and thus avoids vanishing gradient problem. The introduction of residual learning benefits optimization process of deep network and improves the accuracy of segmentation. The schematic representation of residual unit is illustrated in Fig. 3.7.

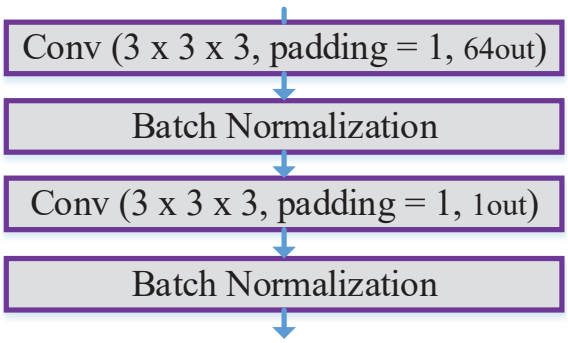


Figure 3.8: Post-block.

For the contracting (forward) path, three blocks of residual units are adopted and each block is followed by a max-pooling layer to reduce the dimension of cube. The residual block contains multiple residual units and only the first unit increases the feature channel to the desired size. For the expansive (backward) path, we first use transposed convolution, BN, and ReLU to upsample cube. Secondly, we concatenate it with the corresponding shallow features that propagate via the skip path. Then, the concatenated features are fed into a residual block. At the end of the last residual block, a post block is attached in order to map the 64-channel feature cube to the size of $1 \times 64 \times 64 \times 64$. It is composed of two convolution layers and two BN layers as shown in Fig. 3.8. In total,

the network has 57 convolution layers. For each voxel in the final cube, the probability of being nodule is calculated via a sigmoid function.

The loss function of our segmentation network is based on Dice coefficient, which measures the similarity between segmentation results and ground-truth labels. Given two binary volumes P and T, the Dice similarity coefficient (DSC) is defined as:

$$DSC = \frac{2\sum_i^N p_i t_i}{\sum_i^N p_i^2 + \sum_i^N t_i^2}, \ \ p_i \in P, t_i \in T, \tag{3.7}$$

where p_i and t_i are voxels in the predicted segmentation result and ground-truth target, respectively. N is the total number of voxels. The value of DSC ranges from 0 to 1 and if P is exactly equivalent to T, the DSC achieves the maximum value of 1. In our implementation, the goal being to minimize the loss function, we define the Dice loss as:

$$\mathcal{L}_{seg} = 1 - \frac{2\sum_i^N p_i t_i + \epsilon}{\sum_i^N p_i^2 + \sum_i^N t_i^2 + \epsilon}, \ \ p_i \in P, t_i \in T, \tag{3.8}$$

where ϵ is a smoothing coefficient that not only prevents division by zero but also avoids overfitting. We set $\epsilon = 1$ here in consideration of Laplace's rule of succession [Jurafsky and Martin, 2014, Russell and Norvig, 2016].

3.3 Experiments and Results

3.3.1 Materials

The public LIDC-IDRI dataset is used to generate synthetic nodule images and validate the proposed segmentation method. The dataset contains 1010 patients' CT scans. Each CT scan was reviewed by four experienced radiologists through a two-stage process: blinded and unblinded reading. In the blinded phase, each radiologist reviewed and marked each CT scan independently. In the unblinded phase, with three other radiologists' marks provided, each radiologist modified the original annotations to improve the quality of ground-truth labels. Nodules, of which the diameters are larger than 3 mm, are annotated with the boundaries and nine semantic attributes of subtlety, internal structure, margin, calcification, sphericity, lobulation, spiculation, texture, and malignancy. In our experiments, we exclude CT scans whose slice thickness is greater than 2.5 mm in consideration of image quality. Hence, there are 888 CT scans with 1182 nodules in total. The distributions of these nodules are listed in Table 3.1 in terms of texture and size. For each nodule, the rating scores of its attributes are computed as the average of ratings from the four radiologists. The definition of attributes' scoring can be found in Table 3.2. The scores of internal structure and calcification reflect corresponding classes while other feature scores represent sequential degrees. First, the internal area inside each nodule's boundary is filled to obtain its ground-truth label. For each CT slice, the Hounsfield unit (HU) is clipped in the range of [-1200 HU, 600 HU]. Then, all slices are normalized to [0, 255] and resampled to the same spacing of $1 \times 1 \times 1$ mm. VOI cubes containing nodules are cropped from slices based on their coordinates and the cropped size is $64 \times 64 \times 64$ pixels.

3.3.2 Implementation Details

Synthetic Image Generation

We use the cropped VOI cubes from LIDC dataset to train our cGAN. All slices containing nodules are chosen to generate their semantic labels and the total number of real CT pairs is 4694. Then, we split the dataset into ten subsets and perform ten-fold cross-validation. Each time, nine subsets are used for training. The remaining subset is left for validation, which generates new synthetic images. Thus, a new synthetic dataset of 4694 slices is obtained.

The cGAN model is initialized from a Gaussian distribution $\mathcal{N}(0, 0.02)$ and optimized using Adam [Kingma and Ba, 2014] with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The initial learning rate for the first 200 epochs is set to 0.0002 and then decreases to 0 linearly after 200 epochs. The model is implemented in PyTorch [Paszke et al., 2019] using 4 NVIDIA GTX 1080Ti GPUs.

Pulmonary Nodule Segmentation

In segmentation experiments, we first replace the original slices in the 1182 nodule cubes with the generated slices to form new VOI. In total, 2364 nodule CT cubes are used, with half from the LIDC-IDRI dataset and half from our generated images. All cubes are of the same size: $64 \times 64 \times 64$. Ten-fold cross validation is adopted to evaluate our model. It should be noted that each time we use nine subsets of LIDC-IDRI dataset and their corresponding synthesized samples to train our model. Then we evaluate the model on the remaining one LIDC-IDRI subset. The validation set has no overlap with the training set.

The segmentation model is initialized from a Gaussian distribution $\mathcal{N}(0, 0.01)$ and trained using Adam [Kingma and Ba, 2014] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ for 150 epochs. The data augmentation method includes random rotation between $[0°, 180°]$, random flipping, and random axis swapping. The initial learning rate is set to 0.05 and decreases by half after every 20 epochs. The validation time for each nodule VOI is within 0.1 second and the segmentation model is also implemented in PyTorch.

3.3.3 Evaluation Metrics

Synthetic Image Generation

It is an open and difficult problem to find suitable metrics for evaluating the quality of synthesized images [Isola et al., 2017]. In the loss function of our cGAN model, $\mathcal{L}_G$ is explicitly optimized. Hence, it is reasonable and natural to use mean squared error (MSE) and cosine similarity (S_C) to evaluate our model. The two metrics are aimed at measuring the similarity between real nodule images and synthetic images. Given a trained generator G and a set of ten-channel semantic labels $\{x^i | i = 1, 2, ..., m\}$, the metrics are defined as:

$$
\begin{aligned}
MSE &= \frac{1}{m}\sum_{i=1}^{m}\|y^i - G(z|x^i)\|_2^2, \\
S_C &= \frac{1}{m}\sum_{i=1}^{m}\frac{y^i \cdot G(z|x^i)}{\|y^i\|_2\|G(z|x^i)\|_2},
\end{aligned}
\tag{3.9}
$$

where y^i and $G(z|x^i)$ here are the vectorized real nodule image and synthetic sample, respectively.

Pulmonary Nodule Segmentation

The performance of the proposed segmentation model is measured by four metrics: DSC, positive predicted value (PPV), sensitivity and accuracy. The DSC, defined in Eq. 3.7, is one of the most commonly used evaluation criteria. The PPV and sensitivity are respectively defined by:

$$
\begin{aligned}
PPV &= \frac{\sum_i^N p_i t_i}{\sum_i^N p_i^2}, \;\; p_i \in P, t_i \in T, \\
Sensitivity &= \frac{\sum_i^N p_i t_i}{\sum_i^N t_i^2}, \;\; p_i \in P, t_i \in T,
\end{aligned}
\tag{3.10}
$$

where P is the predicted result and T is the ground-truth label. N is the total number of voxels of VOI cubes. All numerators of DSC, PPV and sensitivity are the intersection voxels between P and T. For DSC, its denominator is the average union voxels of P and T while for PPV and sensitivity, their denominators are the voxels predicted as positive for nodule region and true lesion voxels, respectively.

In addition, hard thresholding is applied on the probability map to obtain the binary segmentation result. Voxels having probability higher than 0.5 are considered as foreground objects. Then, accuracy is computed as:

$$
\begin{aligned}
Accuracy &= \frac{\sum_i^N \mathbb{1}(p_i == t_i)}{N}, \;\; p_i \in P, t_i \in T, \\
\mathbb{1}(\text{statement}) &= \begin{cases} 1, \text{if statement is True} \\ 0, \text{otherwise} \end{cases},
\end{aligned}
\tag{3.11}
$$

where $\mathbb{1}(\cdot)$ is an indicator function.

3.3.4 Results

Synthetic Image Generation

The evaluation results inside different categories are given in Table 3.3. The MSE and cosine similarity for all nodules are 1.55×10^{-2} and 0.9534, respectively. The MSE of GGO nodules is 1.70×10^{-2}, which exceeds solid and part-solid nodules. The cosine

Table 3.3: Quantitative results of synthetic image generation for different nodule categories.

	Category	MSE ($\times 10^{-2}$)	S_C
Texture	Solid	1.55	0.9538
	Part-solid	1.47	0.9529
	GGO	1.70	0.9491
Diameter	<6 *mm*	1.65	0.9556
	6∼10 *mm*	1.52	0.9524
	>10 *mm*	1.55	0.9539
All nodules		1.55	0.9534

similarity of solid nodules is higher than that of part-solid and GGO nodules. In terms of nodule's size, small nodules achieve the highest cosine similarity of 0.9556 and medium-sized nodules have the lowest MSE of 1.52×10^{-2}. All MSE and cosine similarity results are computed on 4694 nodule images of size 64×64.

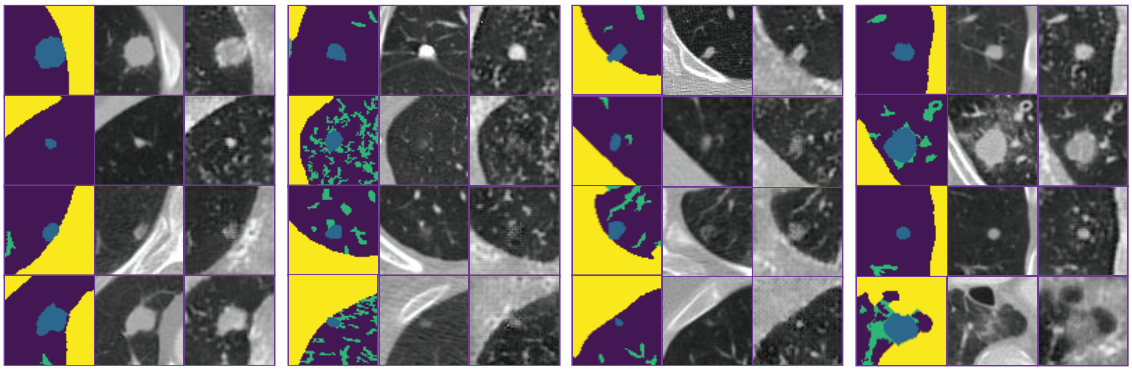


Figure 3.9: Examples of generated synthetic images. (a) Input labels; (b) Real images; (c) Generated images. Out of simplicity, ten-channel inputs are briefly displayed as semantic labels.

Besides, visual examination of generated images is also employed to evaluate our cGAN model. Such evaluation metric is one of the most simple, intuitive yet effective methods to estimate sample's quality. Fig. 3.9 offers qualitative results of some generated samples. It shows that nodules and their surroundings are well reconstructed through our cGAN model.

To corroborate the effectiveness of nine attributes' labels in sample synthesis, we provide a comparison of samples generated with and without the nine attributes in Fig. 3.10. It shows that if only semantic labels are provided, the synthetic samples resemble real CT images with limited variety. In contrast, with additional nine attributes' labels incorporated as inputs, the cGAN can produce various images according to different configurations of attributes' scorings. By setting the value of texture as 1, 3, and 5, the output nodules indeed exhibit the characteristics of GGO, part-solid, and solid nodules, respectively. The 10-channel inputs (see Fig. 3.3) allow the cGAN to generate a large variety of

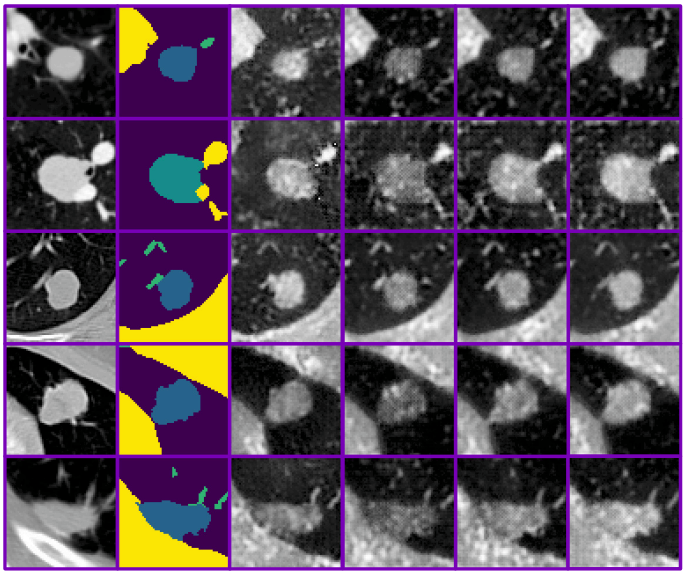


Figure 3.10: Comparison of five synthetic samples generated with and without the nine attributes labels. (a) Real CT images; (b) Semantic labels; (c) Images generated without nine attributes. (d), (e), and (f) stand for the images generated with nine attributes and their texture scores are set to 1, 3, and 5, respectively.

nodule images that do not exist in the original LIDC-IDRI dataset, thereby enriching the training data greatly.

Pulmonary Nodule Segmentation

Table 3.4: The segmentation results of the proposed model for different nodule categories.

	Category	DSC	PPV	Sensitivity	Accuracy
Texture	Solid	0.8605	0.8927	0.8681	0.9909
	Part-solid	0.8096	0.8755	0.8023	0.9891
	GGO	0.7865	0.8850	0.7515	0.9871
Diameter	$< 6mm$	0.7776	0.8748	0.7719	0.9849
	$6 \sim 10mm$	0.8382	0.8788	0.8494	0.9897
	$> 10mm$	0.8578	0.8966	0.8560	0.9911
All nodules		0.8483	0.8895	0.8511	0.9904

Table 3.4 summarizes the segmentation results in terms of four metrics. The average DSC, PPV, sensitivity and accuracy of all nodules are 0.8483, 0.8895, 0.8511, 0.9904, respectively. The performance of the proposed method on GGO nodules is the worst in terms of DSC, sensitivity, and accuracy. It is noted that nodules with larger diameter or solid texture have the highest segmentation scores in any evaluation metric.

Table 3.5: Comparison of segmentation results in DSC.

Approach	DSC
Mukhopadhyay [Mukhopadhyay, 2016]	0.3900
Çiçek et al. [Çiçek et al., 2016]	0.7197
Wu et al. [Wu et al., 2018a]	0.7405
Proposed method	**0.8483**

The comparison of segmentation results with state-of-the-art methods [Mukhopadhyay, 2016, Çiçek et al., 2016, Wu et al., 2018a] is given in Table 3.5. All the methods are evaluated on LIDC-IDRI dataset and the commonly used metric is Dice coefficient. Our model achieves the highest DSC score of **0.8483** and it outperforms existing methods. The traditional segmentation techniques by Mukhopadhyay [Mukhopadhyay, 2016] can not adapt to large variation of nodules such as size, shape and texture. Although both methods by Çiçek et al. [Çiçek et al., 2016] and Wu et al. [Wu et al., 2018a] adopt 3D CNNs for segmentation task, our method surpasses them over 10% on average.

Table 3.6: Quantitative comparison results of the control group.

Approach	DSC	PPV	Sensitivity	Accuracy
Seg-NMaps	0.7993	0.8523	0.8121	0.9881
Seg-NEdge	0.8176	0.8233	**0.8610**	0.9891
Seg-NLBP	0.8101	0.8559	0.8261	0.9890
Seg-NSynthetic	0.8104	0.8431	0.8596	0.9876
Proposed method	**0.8483**	**0.8895**	0.8511	**0.9904**

Table 3.6 summarizes the results of quantitative comparison between different configurations having different inputs. A control group of four methods is constituted to evaluate our proposed method. Seg-NMaps refers to the proposed method without taking any map as input to the segmentation network. Seg-NEdge and Seg-NLBP denote the proposed method that does not use edge maps and LBP maps, respectively. For Seg-NSynthetic, generated synthetic samples are not added into the dataset to train our model. The Seg-NMaps method has the lowest DSC of 0.7993. Both LBP and edge maps contribute to better results, increasing DSC to 0.8176 and 0.8101, respectively. Without the extension of dataset, the Seg-NSynthetic method achieves the lowest accuracy of 0.9876. Except sensitivity, the proposed method enjoys the highest scores on other three metrics, which demonstrates the pertinence of each component of the proposed method. The accuracy of all methods is over 0.98.

More visually, qualitative results of different validation samples are shown in Fig. 3.11.

3.4 Discussion

We have shown that the proposed method can achieve accurate segmentation on pulmonary nodules. Its most distinctive characteristics include (1) the adoption of adversarial networks to promote samples' diversity for a more balanced training dataset and (2)

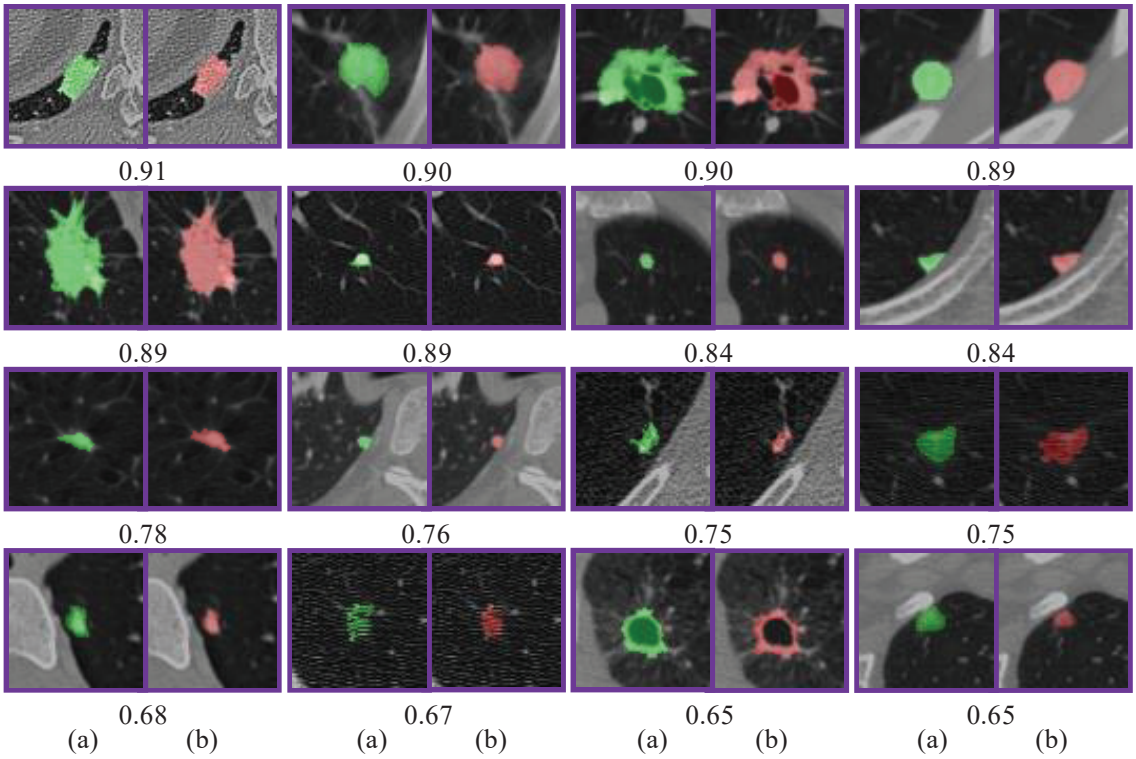


Figure 3.11: Qualitative segmentation results of validation samples. (a) Ground-truth labels are in green; (b) Predicted nodules are in red. The score beneath each pair is Dice coefficient of the result. Central slice of each VOI cube is displayed for simplicity.

the 3D segmentation network that takes advantage of interpretable heterogenous maps and residual learning. The results on synthetic image generation show that the cGAN simulates well the real nodule images to generate satisfactory samples and that each component of the segmentation network is instrumental in improving accuracy.

There are mainly two elements contributing to the realistic synthetic image generation. One is the preprocessing technique designed to obtain the ten-channel label that is rich in semantic information about nodule's attributes and surroundings. In conventional synthetic image generation [Shrivastava et al., 2017, Isola et al., 2017, Guibas et al., 2017], only ordinary geometry images are viewed as conditions of adversarial networks. Such type of images is in lack of depiction of context and characteristics of nodules. The other is the modification on the objective function of the original cGAN. The objective function defined in the original cGAN [Mirza and Osindero, 2014] only considers producing images that can deceive the discriminator, which is not sufficient in our medical setting. In contrast, the introduced $L2$ reconstruction error loss [Eq. (3.2)] explicitly minimizes the difference between generated images and real nodule images.

During the semantic label generation process, all nine channels were employed to represent nodule's attributes. No selection or weighted combination of the nine attributes was conducted in advance because all these attributes are important for describing nodules. Each attribute is annotated and corrected meticulously by radiologists. If without nine channels, the diversity of synthetic images is substantially reduced. Given a one-channel label of a nodule such as a disk mask, the cGAN will produce a geometrically similar sample that only differs from the original data in grayscale value. While with the attributes provided, various nodule images can be generated by changing the rating scores of each attribute.

Although the patterns or styles of synthetic samples are kept similar to real ones, the generated images differ from the existing dataset in specific details. Firstly, in fact, given any artificial 10-channel semantic label, the cGAN can generate realistic CT samples that are missing in the original dataset. In accordance with different scorings of attribute labels, diverse types of synthetic images can be generated to improve the variety of training samples. Secondly, random noise is introduced into the generating process by dropout layer. Even with the same 10-channel semantic labels as inputs, the generated samples are different from their corresponding real CT images. Thirdly, tiny objects, such as small vessels and parenchymal structures, are removed in the generation process of semantic labels. Hence, conditioned on the resulting coarse-grained semantic labels, the synthetic images do possess a high level of variety.

The MSEs of solid and part-solid nodules are smaller than those of GGO nodules (Table 3.3). This can be explained as follows. First, the boundaries of solid and part-solid nodules are clearer and their intensity varies abruptly, which is easier for cGAN to learn the discrepancy between nodules and their surroundings. The second reason is that the internal distribution of GGO nodules is comparatively complex and scattered. The intensity inside GGO is relatively low and not as constant as that of solid nodules. As shown in Fig. 3.9, compared to real images, nodules on synthetic images tend to be more distinct because they are generated from labels which have sharp margins and specific borders. Due to the introduction of stochastic noise, the background of generated samples has more vascular-like structures than that of real nodule images.

Concerning the segmentation (Table 3.4), our method performs better on solid and

part-solid nodules than on GGO nodules. This is because the boundaries of GGO nodules are fuzzier than other nodules, especially if there exist vascular structures in their vicinity. Furthermore, the number of GGO nodules in LIDC-IDRI dataset is far smaller than that of solid and part-solid nodules and thus the diversity and quality of generated samples are limited. Training on such dataset, the proposed model is difficult to capture strong feature representations for segmentation of GGO nodules. In terms of nodule's size, the larger the nodule is, the better the result is due to the fact that for larger nodules, it is easier to detect their position inside VOI and determine accurate margins.

Table 3.5 provided comparison with state-of-the-art methods [Mukhopadhyay, 2016, Çiçek et al., 2016, Wu et al., 2018a]. The performance of the traditional method by Mukhopadhyay [Mukhopadhyay, 2016] is poor because it requires careful tuning of hyper-parameters (e.g., thresholding value of density for different nodules), which triggers off weak generalization ability on large dataset. Although Çiçek et al. [Çiçek et al., 2016] and Wu et al. [Wu et al., 2018a] employed deep learning techniques as well, they did not regard the effect of multiple interpretable maps on conveying useful information (e.g. portrayal of nodule's texture by LBP maps and emphasis on nodule's border by edge maps) to the network. Besides, they did not take residual learning into consideration, which is crucial for developing a deep model. Examples in Fig. 3.11 demonstrate the performance on different kinds of nodules. Compared with well-circumscribed and solid nodules, juxta-vascular and GGO nodules are relatively harder to segment accurately due to their complex outer attachments and internal texture patterns, respectively. The results predicted by our model tend to provide conservative boundaries if the intensity drops sharply at margins. It may be because the inclusion of edge maps makes the model sensitive to borders. In Table 3.6, a possible reason that Seg-NEdge has the highest sensitivity is that without edge maps, the segmentation is not sensitive to the contours of nodules. It may tend to predict more pixels outside the contour as nodules than true nodule pixels. According to Eq. (3.10), the sensitivity becomes high when the numerator increases. If neither the maps nor the synthetic data are used, the proposed framework degenerates back to a normal 3D CNN-based segmentation model, which differs from the existing 3D U-Net in two aspects: (1) the number of feature channels and (2) the introduction of residual learning strategy. Since in this case only real VOIs are fed into the 3D CNN without their features included, the segmentation performance is worse than the proposed framework.

There exist some limitations associated with the proposed framework. First, for the semantic labels in synthetic image generation, we only consider vessels and pleural surfaces and omit other structures such as bones and bronchi. To further improve the realism of generated samples, all structures would need to be labeled, which requires more complicated preprocessing techniques and parameter tuning. Since the quality of semantic labeling is heavily dependent on prior knowledge, it is challenging to develop a method of automatic labeling at an expert level. Second, it is difficult to find the optimal form of introducing random noises into cGAN. In the present study, dropout layer is applied as noise z to generate stochastic output, which is consistent with Isola et al. [Isola et al., 2017]. It needs a different study to determine the impact of noise on the generated samples. Third, the distribution of training dataset is still not even. Although we extend LIDC-IDRI dataset via the cGAN model, the quantity and diversity of some nodules (e.g., GGO and juxta-vascular nodules) are still in shortage. Future work may include

designing new schemes to solve imbalanced dataset problem. Finally, it is noted that the segmentation labels of nodules are obtained from radiologists in LIDC. However, the annotation process is decided by each radiologist's subjective judgment [Mukhopadhyay, 2016,Qiang et al., 2014,Armato et al., 2011], leading to different ground-truth labels. Hence, the performance of the proposed method may be affected by such variation.

3.5 Conclusion

In this chapter, we have proposed a two-part CNN-based framework for pulmonary nodule segmentation. In the first part, adversarial networks are employed to synthesize nodule samples. It targets at building a more diverse and balanced dataset for the subsequent model training. Semantic labels, together with nine attribute scoring labels, are exploited to provide semantic and contextual knowledge. Reconstruction error loss is introduced to improve realism. Such method of extending dataset presents several advantages. The boundaries and semantic attributes of nodules are preserved during generation process. Moreover, the random noise produced by dropout layer allows for the variation of spatial surroundings and thus boosts image diversity. In the second part, multiple feature maps are incorporated as inputs into the 3D CNN model. With residual learning strategy, the segmentation model trained on the extended dataset enjoys a high level of generality. The results on LIDC-IDRI dataset demonstrate that our 3D CNN model achieves more accurate nodule segmentation compared to existing state-of-the-art methods, which suggests its potential value for clinical applications.

Chapter 4

Development of a Voxel-Connectivity Aware Approach for Accurate Airway Segmentation Using Convolutional Neural Networks

Contents

4.1 Introduction 132
4.2 Methodology 134
4.2.1 CT Volume Pre-processing 134
4.2.2 Connectivity Modeling Using Binary Ground-Truth Labels 135
4.2.3 Connectivity Prediction with AirwayNet 137
4.2.4 Connectivity Prediction with AirwayNet-SE 137
4.2.5 Airway Candidates Generation 140
4.3 Experiments and Results 140
4.3.1 Materials 140
4.3.2 Implementation Details 141
4.3.3 Evaluation Metrics 141
4.3.4 Results 141
4.4 Discussion 144
4.5 Conclusion 144

4.1 Introduction

Pulmonary diseases, including chronic obstructive pulmonary diseases (COPD) and lung cancer, pose high risks to human health. The standard computed tomography (CT) helps radiologists detect pathological changes. For tracheal and bronchial surgery, airway tree modeling on CT scans is often considered a prerequisite. Meticulous efforts are required to manually segment airway due to its tree-like structure and variety in size, shape, and intensity.

Several methods have been proposed for airway segmentation on CT images. Van Rikxoort et al. [Van Rikxoort et al., 2009] proposed a region growing method with adaptive thresholding. Xu et al. [Xu et al., 2015] combined two tubular structure enhancement techniques within the fuzzy connectedness segmentation framework. Lo et al. [Lo et al., 2010a] designed a learning-based approach that models airway appearance and utilized vessel orientation similarity. In the EXACT'09 challenge, fifteen airway extraction algorithms were summarized by Lo et al. [Lo et al., 2012]. Most algorithms adopted region growing with additional constraints such as tube likeness. Although successfully segmenting bronchi of large size, these conventional methods performed worse on peripheral bronchi.

Recently, convolutional neural networks (CNNs) were increasingly used in segmentation tasks [Ronneberger et al., 2015,Çiçek et al., 2016]. For airway extraction, CNNs-based methods [Charbonnier et al., 2017, Yun et al., 2019, Meng et al., 2017, Jin et al., 2017, Juarez et al., 2018] were developed and proved superior to previous methods in [Lo et al., 2012]. Charbonnier et al. [Charbonnier et al., 2017] and Yun et al. [Yun et al., 2019] respectively used two-dimensional (2-D) and 2.5-D CNNs on already coarsely segmented bronchi to reduce false positives and increase detected tree length. Meng et al. [Meng et al., 2017] embedded CNNs-based segmentation into the airway volume of interest (VOI) tracking framework. Jin et al. [Jin et al., 2017] employed graph-based refinement on the probability output of CNNs. Juarez et al. [Juarez et al., 2018] designed an end-to-end CNN model with simple pre- and post-processing. Graph neural networks (GNNs) [Selvan et al., 2020] have also been studied for airway extraction.

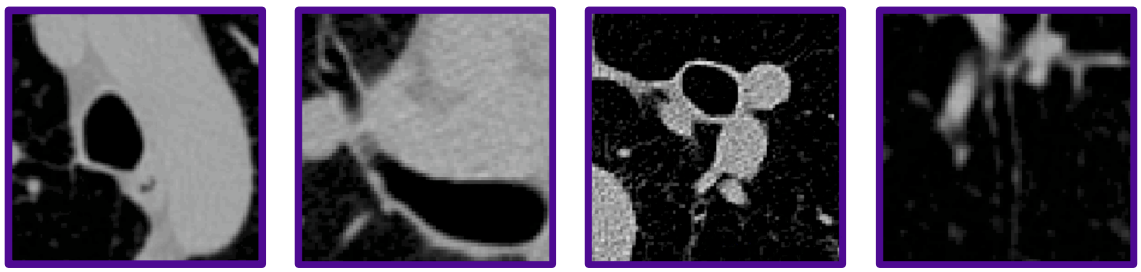


Figure 4.1: The intensity distribution of trachea (a), primary (b) and secondary (c) bronchus, and peripheral bronchiole (d). The scale of contexts needed for airway segmentation on (a)-(d) is decreasing from large to small.

Although deep learning approaches achieved superior performance, there still remain challenges to be solved. First, the intensity distribution of airways is quite different

among trachea, primary bronchus, secondary bronchus, and peripheral bronchiole (see Fig. 4.1). The intensity contrast between lumen and wall is clear at trachea regions, but becomes weaker as the airway bifurcates into smaller branches. The airway wall is much thinner and darker at bronchiole regions. Second, the scale of context is dissimilar for segmentation on large and small bronchi. The context refers to the feature information that describes the mutual relationship between airways and background. To extract trachea and main bronchus, large-scale context is preferred for the model to perceive the main body with large field-of-view. On the contrary, for segmenting bronchiole, context of close neighborhood is enough. Third, for CNNs architecture, the number of pooling layers requires careful design. Features of thin bronchi, whose diameters are usually only 2-3 voxels, are prone to vanish after three times of pooling, making it difficult to reconstruct and recover. However, for large bronchi, multiple pooling layers are necessary to extract effective context. Furthermore, public datasets with airway annotations are unavailable for model training and fair comparison between different methods.

To address these gaps, we propose AirwayNet, a CNNs-based approach for accurate airway segmentation. Considering that the tree-like structure of airway is rather complex and the prediction of airway candidates is prone to discontinuity, we put emphasis on the connectivity of airway voxels. Unlike previous methods, we do not directly train the network to classify foreground and background voxels. Instead, binary segmentation task is transformed into 26 tasks of predicting whether a voxel is connected to its neighbors. Since airway voxels are stretching from the main bronchus towards bronchiole end as a whole connected region, we consider it a good solution to enable the model being aware of voxel connectivity. Previous work on salient segmentation [Kampffmeyer et al., 2018] demonstrated that connectivity modeling spontaneously encodes the relation between two pixels. Therefore, we design a voxel connectivity-aware approach to better comprehend the inherent structure of airway.

Moreover, we go one step further by extending the one-step AirwayNet into the two-step AirwayNet-SE, a **S**imple-yet-**E**ffective approach that incorporates two different scales of context to comprehend large and small airways, respectively. With the same 3-D connectivity modeling as the first step, the networks are trained to predict whether a voxel is connected to its neighbors instead of directly classifying airway voxels. The AirwayNet-SE consists of one **d**eep-yet-**n**arrow **n**etwork (DNN) and one **s**hallow-yet-**w**ide **n**etwork (SWN). The DNN, with deeper layers yet smaller number of channels per layer, aims at extracting features of thick branches. Four pooling operations are used for the model to be aware of the overall context of thoracic cavity. While for the SWN, shallower layers with two pooling operations are adopted to prevent thin bronchi from vanishing. The feature channels of SWN are widened to increase representation power. The second step is to predict connectivity using two-stage CNNs. In the first stage, we respectively train our DNN and SWN to learn effective features of large and small bronchi. In the second stage, features from both DNN and SWN are concatenated as the fusion of context knowledge from two scales. Such fused features are utilized for the final airway connectivity prediction.

Our contributions are summarized as follows: 1) The voxel connectivity of airway is modeled using conventional binary ground-truth labels to better serve the airway segmentation task. The proposed AirwayNet automatically learns relationship between adjacent voxels and discriminates airway from the background. For each voxel, the network

predicts not only its probability of being airway but also its connectivity to neighbors. 2) The AirwayNet-SE proposed a solution to the conflict caused by the difference between large and small airways. With connectivity modeling, it leveraged the fusion of context knowledge from two scales to predict whether a voxel is airway and connects to its neighbors. 3) We released the manual annotations of 60 public CT scans to promote airway segmentation study that requires supervised learning. To the best of our knowledge, this is the largest publicly available dataset of airway annotations[1].

4.2 Methodology

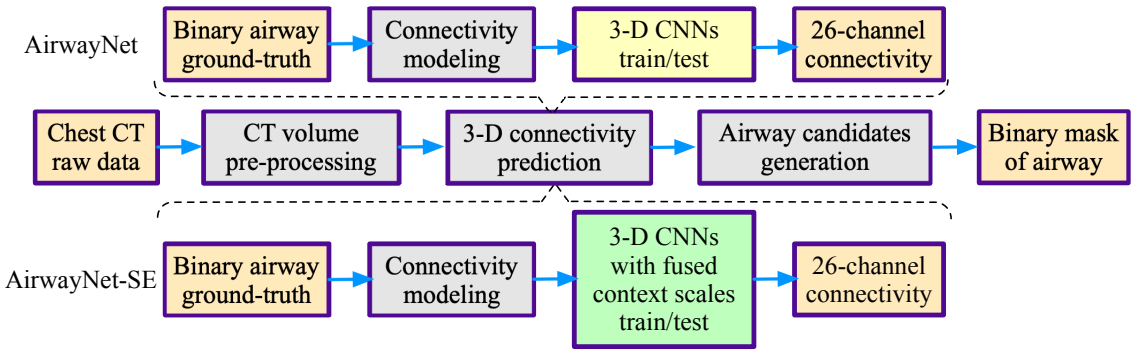


Figure 4.2: Flowchart of the proposed AirwayNet and AirwayNet-SE.

In this section, we first introduce the details about CT pre-processing and voxel connectivity modeling. Such modeling step is the prerequisite for transforming the segmentation problem into connectivity prediction problem. Then, the 3-D CNNs-based connectivity prediction is described. Subsequently, we introduce how to extend the one-stage connectivity prediction into its two-stage counterpart, where features of large and small context scales are fused for connectivity prediction. Finally, we discuss the generation process of airway candidates. The flowchart of the proposed AirwayNet and AirwayNet-SE is depicted in Fig. 4.2.

4.2.1 CT Volume Pre-processing

One challenge in airway segmentation is that the foreground voxels only occupy a small proportion of all CT voxels. To avoid feature learning from irrelevant parts (e.g., ribs and skin), we restrict the valid airway candidate region inside the lung area. To extract lung mask, each CT slice is first filtered with a Gaussian filter ($\sigma = 1$) and binarized with a threshold (-600 Hounsfield unit). The connected component analysis is applied to remove unconfident candidates and the largest two components are chosen as left and right lungs, respectively. To avoid under-segmentation, we replace the lung area by its convex hull on each slice if the convex hull has 50% more area. We also perform Euclidean distance transform on the lung mask to calculate the distance map. Each voxel on the distance map records its minimum distance to the lung border. We add such

[1] Annotations are available at `http://www.pami.sjtu.edu.cn/News/56`

map into the network because the airway's relative position to lung border is considered anatomically meaningful. To prepare for network training, CT voxel intensity is clipped by a window $[-1000, 600]$ (HU) and normalized to $[0, 255]$. Fig. 4.3 illustrates the CT pre-processing step.

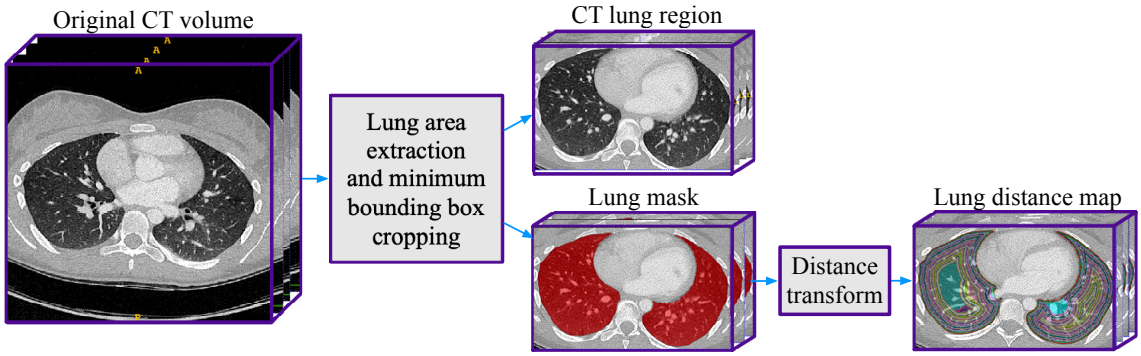


Figure 4.3: Illustration of the CT pre-processing step.

4.2.2 Connectivity Modeling Using Binary Ground-Truth Labels

In a three-dimensional (3-D) CT, 26-connectivity describes well the relation between one voxel and its 26 neighbors (see Fig. 4.4). Given a voxel $P = (x, y, z)$ and its neighbor $Q = (u, v, w)$, the distance between P and Q is restricted by $d(P, Q) = max(|(x-u)|, |(y-v)|, |(z-w)|) \leq 1$, which means that Q is located within a $3 \times 3 \times 3$ cube centered at P. We index neighbors Q from 1 to 26 and denote each voxel pair $(P, Q_i), i \in \{1, 2, ..., 26\}$ as a connectivity orientation. Each orientation is encoded using a 1-channel binary label. If both P and Q_i are airway voxels, then the pair (P, Q_i) is connected and the corresponding position "P" on the i-th label is marked as 1. Otherwise, we mark 0 on the i-th label to represent disconnected pair (P, Q_i). By sliding such a $3 \times 3 \times 3$ window over each voxel, we obtain 26 binary labels and concatenate them into a 26-channel connectivity label. Zero padding is performed on CT volume borders to keep the size of generated labels unchanged. Such connectivity label encodes both ground-truth position and 26-connectivity relation between airway voxels. Note that all operations are performed on conventional binary labels of airway ground-truth. We do not require extra manual annotation for connectivity labels.

After modeling the 3-D connectivity, the original airway ground-truth label is reformed, and the conventional segmentation task is then transformed into 26 connectivity prediction tasks. The objective here is to classify and merge connected airway voxels along each connectivity orientation. An advantage of decomposing one task into 26 different tasks is that multi-task learning strategy helps the network learn more generalized features. These 26 tasks are correlated in depicting voxel connectivity, so that our AirwayNet can extract essential and robust features. Furthermore, the connectivity label emphasizes airway's structure attribute. The network trained using such label is encouraged to grasp the tree-like pattern of bronchial airway.

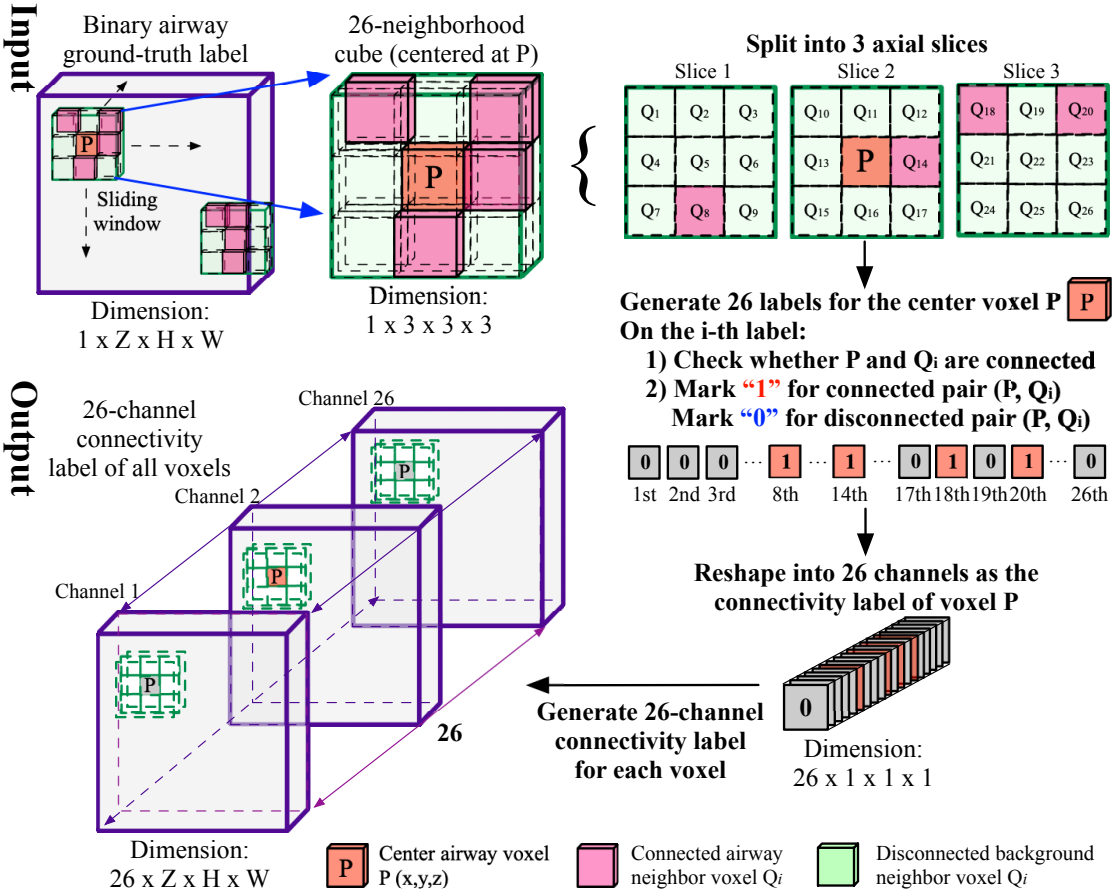


Figure 4.4: Illustration of 26-connectivity modeling. The binary ground-truth of airway (Dim: $1 \times Z \times H \times W$) is transformed into a connectivity label (Dim: $26 \times Z \times H \times W$). For each voxel P, we extract a $3 \times 3 \times 3$ neighborhood cube to check connected voxel pairs. Each pair $(P, Q_i), i \in \{1,2,...,26\}$ represents a connectivity orientation and is encoded with a binary label. For example, if an airway voxel P is connected to its neighbor Q_{20}, then the corresponding position "P" on the 20-th label is marked as 1.

4.2.3 Connectivity Prediction with AirwayNet

Given a cropped CT cube, the objective is to use 3-D CNNs to predict voxel connectivity of the cube. The proposed AirwayNet (see Fig. 4.5) is based on the U-Net [Çiçek et al., 2016] backbone. Our full 3-D architecture captures more spatial information than the 2-D or 2.5-D CNNs used in [Charbonnier et al., 2017, Yun et al., 2019] and is more suitable for learning the bronchial continuity and branching patterns. The AirwayNet consists of a contracting path and an expansive path with four resolution scales. At each resolution scale, the contracting path has two convolution layers (Conv) with batch normalization (BN) and rectified linear unit (ReLU), followed by a max-pooling layer. In the expansive path, finer features from lower resolution scale are linearly upsampled first and then concatenated with coarse features from skip connection to preserve details of thin bronchi. Since airway voxels are distributed within the large thoracic cavity, extra semantic information other than grayscale intensity is considered beneficial for the model to classify airway voxels. Here we use voxel coordinates and lung distance map, and concatenate them with features on the expansive path at the last scale. The sigmoid function is applied on the predicted connectivity cube to obtain probability distribution.

We use the Dice similarity coefficient (DSC) loss to optimize our AirwayNet. For each voxel x in the cropped cube X, given its label $y_i(x)$ and prediction probability $p_i(x), i \in \{1,2,...,26\}$, the total connectivity loss is defined as the averaged DSC loss over all channels:

$$\mathcal{L} = 1 - \frac{1}{26}\sum_{i=1}^{26}\frac{2\sum_{x\in X} p_i(x)y_i(x)}{\sum_{x\in X}(p_i(x)+y_i(x))+\epsilon}, \tag{4.1}$$

where ϵ is used to avoid division by zero.

4.2.4 Connectivity Prediction with AirwayNet-SE

The major differences between AirwayNet and AirwayNet-SE lie in the network architecture and feature learning strategy. The AirwayNet adopts a one-stage approach, where only one CNNs-based model is used for prediction. In contrast, the AirwayNet-SE adopts a two-stage approach. In the first stage, two CNNs-based models (DNN and SWN) are used to learn features of large-scale and small-scale contexts, respectively. In the second stage, such features of two scales are fused for final connectivity prediction. Fig. 4.6 illustrates the prediction process of the AirwayNet-SE.

Stage 1: Feature Learning with Large-scale and Small-scale Contexts

In this stage, we respectively employ the DNN and SWN to learn features of airway connectivity with different context scales. The 3-D U-Net [Çiçek et al., 2016], containing a contracting path and an expansive path, is used in both DNN and SWN as backbone. To enlarge the receptive field of DNN, four pooling layers are used and accordingly ten convolution layers (Conv) with batch normalization (BN) and rectified linear unit (ReLU) are set on the contracting path. For SWN, only two pooling layers are kept to preserve the details of "delicate" and thin bronchi. The number of feature channels in DNN and SWN are separately designed to fit for such architecture difference. The trade-off between feature extraction ability and GPU memory limit is considered as well.

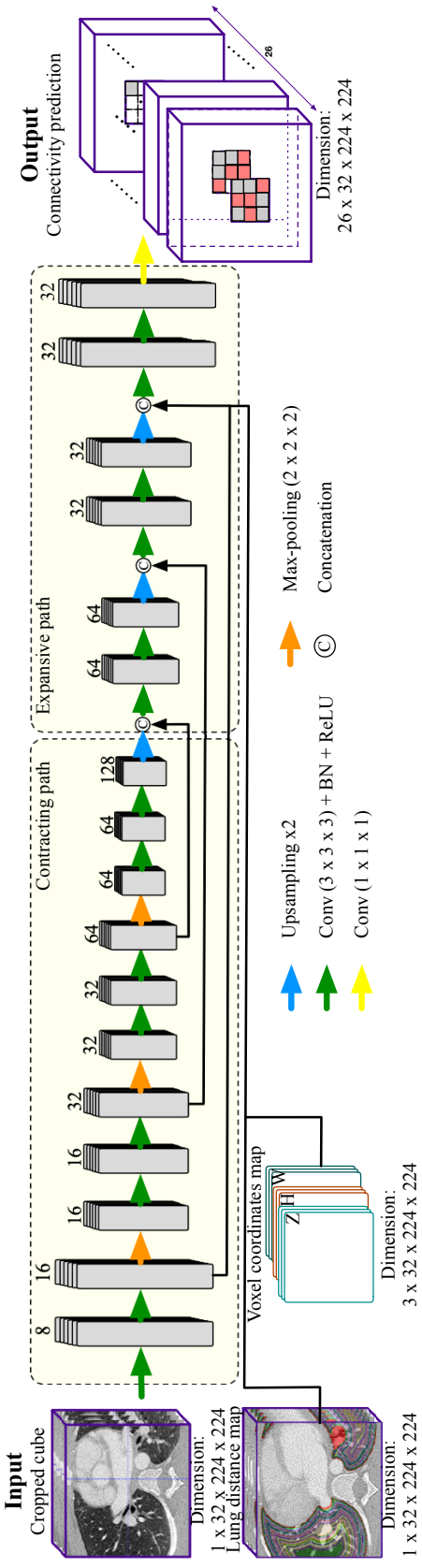


Figure 4.5: Illustration of the AirwayNet. The number of channels is denoted above each feature map.

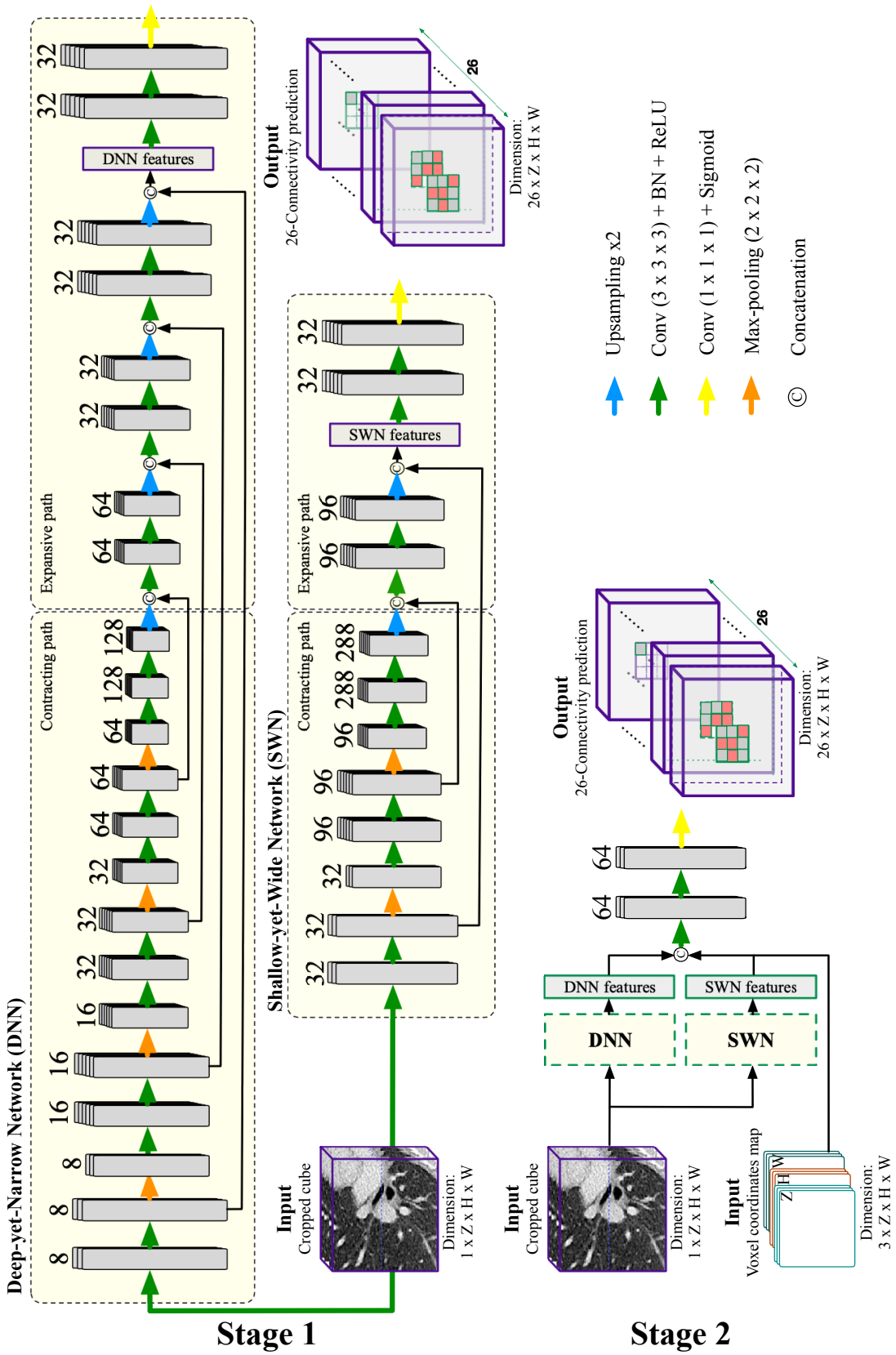


Figure 4.6: Illustration of the AirwayNet-SE. The number of channels is denoted on each feature map. The first stage is to extract features of two context scales via DNN and SWN. The second stage is to classify the connectivity of airways using fused features with both large-scale and small-scale contexts.

Stage 2: Connectivity Prediction Based on Fused Context Knowledge

In this stage, feature representations from DNN and SWN are concatenated as context knowledge fusion. The voxel coordinates are also included as inputs. They are considered beneficial for the model to comprehend the anatomical structure of airways because the position of airways within the thoracic cavity is not randomly distributed. We build a simple three-layer CNN to learn the mapping from the fused features to the 26-channel connectivity prediction. As introduced in Sec. 4.2.3, the Dice coefficient loss is used in both the two stages for optimization.

4.2.5 Airway Candidates Generation

The final step is to generate airway candidates based on the predicted connectivity cube. First, a threshold $t = 0.5$ is used to binarize prediction results. Here we consider that pairwise voxels should agree with each other in connectivity. For example, if voxel P is connected to its neighbor Q_{14} (see Fig. 4.4), then voxel P on the 14-th connectivity channel is marked as 1. Meanwhile, on the 13-th channel, voxel Q_{14} is marked as 1 because P is also at the position "Q_{13}" of the $3 \times 3 \times 3$ neighborhood of Q_{14}. The connectivity between P and Q_{14} is coded in both the 13-th and the 14-th channels. Therefore, we only keep voxels that comply with such pairwise agreement. Then, channel-wise summation is performed on the connectivity cube and those non-zero voxels are marked as airway candidates. The segmentation results are multiplied with the lung field mask to filter out false positives. No post-processing method is employed for refinement.

4.3 Experiments and Results

4.3.1 Materials

In this chapter, the experiment dataset contains 70 clinical chest CT scans in total, with 60 public CT scans and 10 privately collected CT scans. The acquisition and investigation of data were conformed to the principles outlined in the declaration of Helsinki [Association et al., 2001]. We used 20 scans from the training set of EXACT'09 [Lo et al., 2012] and 40 scans from LIDC-IDRI [Armato et al., 2011]. The EXACT'09 only provides raw CT images without airway annotation. The LIDC-IDRI (under Creative Commons Attribution 3.0 Unported License) includes 1018 scans with pulmonary nodule annotations. In view of image quality, 40 scans whose slice thickness ≤ 0.625 mm are randomly chosen. The 10 private CT scans were obtained from patients with severe lung diseases such as emphysema and pneumonia. The axial slices of all scans have the same size of 512×512, with their spatial resolution in the range of 0.5–0.781 mm. Their slice thickness varies from 0.45 to 1.0 mm.

For each CT scan, the ground-truth annotation of airway lumen was obtained by: 1) using an interactive segmentation method to generate an initial rough airway tree via ITK-SNAP [Yushkevich et al., 2006]; 2) manual correction and delineation by well-trained experts. The annotations of 60 public CT scans are released to promote further study of airway extraction using supervised learning methods. However, the 10 private CT scans and annotations will not be made available online at the moment.

In view of the dataset under investigation, we randomly chose 50 public CT scans for training and hyper-parameter fine-tuning. The remaining 10 public and 10 private scans were used for evaluation. Performance comparison and ablation studies are conducted to confirm the validity of our method.

4.3.2 Implementation Details

To improve the model's generalizability on diverse CT scans, data augmentation is performed on-the-fly during training via random horizontal flipping and Gaussian smoothing. For the AirwayNet and the second stage of AirwayNet-SE, all cropped cubes near airways are used for network training, resulting in around 6000 samples. For the first stage of AirwayNet-SE, we densely sampled cubes near trachea and main bronchus regions for training DNN. The cubes containing thin peripheral bronchiole are mostly chosen for SWN to learn their complicated branching patterns. We implemented our method in Keras [Chollet et al., 2015] and the training was conducted on 4 NVIDIA Titan Xp GPUs. The Adam optimizer [Kingma and Ba, 2014] ($\beta_1 = 0.9, \beta_2 = 0.999$) is used with the initial learning rate set as 10^{-4}. The training converged after 30 epochs for all models. Due to the limit of GPU memory, patch-based training and testing are adopted, where pre-processed CT images are cropped into smaller cubes using a sliding window technique. The patch size is $32 \times 224 \times 224$ and the sliding stride is $[8,56,56]$. Such cubes include abundant context knowledge for our model to distinguish between the airway and the background. During testing, the stride of the sliding window is $[16,128,128]$ and the prediction results are averaged on overlapping margins.

4.3.3 Evaluation Metrics

The performance of the proposed AirwayNet and AirwayNet-SE were evaluated in terms of three metrics: 1) Dice similarity coefficient (DSC), 2) True positive rate (TPR), and 3) False positive rate (FPR). Given two binary volumes P and T, the DSC, TPR, and FPR are respectively defined as:

$$\begin{aligned} DSC &= \frac{2\sum_i^N p_i t_i}{\sum_i^N p_i^2 + \sum_i^N t_i^2}, \ \ p_i \in P, t_i \in T, \\ TPR &= \frac{\sum_i^N p_i t_i}{\sum_i^N t_i^2}, \ \ p_i \in P, t_i \in T, \\ FPR &= \frac{\sum_i^N p_i - p_i t_i}{\sum_i^N 1 - t_i^2}, \ \ p_i \in P, t_i \in T, \end{aligned} \tag{4.2}$$

4.3.4 Results

Comparison with the State-Of-The-Art Methods

We compare the proposed AirwayNet and AirwayNet-SE with two state-of-the-art methods: Jin's method [Jin et al., 2017] and Juarez's method [Juarez et al., 2018]. They both employ 3-D CNNs with a sliding window technique for airway extraction. We

re-implemented their methods in Keras [Chollet et al., 2015] by ourselves. They were trained from scratch and evaluated on the same dataset.

Table 4.1: Results of the proposed AirwayNet and AirwayNet-SE in comparison with state-of-the-art methods (mean±standard deviation).

Public testing set			
Method	DSC (%)	TPR (%)	FPR (%)
Jin et al. [Jin et al., 2017]	90.5±4.0	**94.7±2.6**	0.044±0.029
Juarez et al. [Juarez et al., 2018]	92.8±3.5	89.2±6.5	**0.008±0.004**
AirwayNet	90.9±4.3	92.7±3.5	0.033±0.027
AirwayNet-SE	**93.0±3.1**	92.4±4.0	0.018±0.012
Private testing set			
Method	DSC (%)	TPR (%)	FPR (%)
Jin et al. [Jin et al., 2017]	86.3±5.4	81.7±8.9	0.027±0.022
Juarez et al. [Juarez et al., 2018]	87.2±5.3	81.0±8.8	**0.016±0.021**
AirwayNet	88.5±4.0	**86.5±6.0**	0.033±0.029
AirwayNet-SE	**88.7±5.3**	84.6±8.7	0.020±0.017

In Table 4.1, it is observed that the proposed AirwayNet-SE achieved the highest DSC of 93.0% and 88.7% on both the public and private testing sets, respectively. Compared with AirwayNet, the two-stage AirwayNet-SE achieved a higher DSC, a lower FPR, and a comparable TPR. On the public testing set, Juarez et al. [Juarez et al., 2018] segmented more conservatively and they had the lowest FPR with a competitive DSC of 92.8%. Jin et al. [Jin et al., 2017], on the other hand, performed more aggressively than others. Their sensitivity to airway voxels also comes with a side effect of a much higher FPR. For the independent private testing set, the two state-of-the-art methods [Juarez et al., 2018, Jin et al., 2017] obtained the lowest DSC and TPR. Compared with AirwayNet-SE, AirwayNet achieved a higher TPR of 86.5% and a comparable DSC of 88.5%.

Ablation Study

The ablation study was performed to evaluate each constituting component of the proposed method. Compared with DNN and SWN that only rely on context of one scale for airway extraction, the AirwayNet-SE increased the DSC by over 2% on average. The SWN extracted more airway voxels than DNN on both the public and private testing sets. To measure the performance of airway segmentation without connectivity modeling, we also trained the AirwayNet-SE using the conventional binary airway labels as targets. The same network architecture and experimental settings are maintained except that the current objective is to directly predict airway voxels via CNNs. Results in Table 4.2 show that without connectivity modeling, the AirwayNet-SE had lower DSC, TPR, and FPR on both the two testing sets.

Qualitative Results

Qualitative comparison of the segmentation results is visualized in Fig. 4.7. Compared with SWN and DNN, more peripheral branches were successfully detected by

Table 4.2: Ablation study of the proposed AirwayNet and AirwayNet-SE (mean±standard deviation). The DNN and SWN stand for Deep-yet-Narrow Network, Shallow-yet-Wide Network, respectively.

Public testing set			
Method	DSC (%)	TPR (%)	FPR (%)
AirwayNet	90.9±4.3	92.7±3.5	0.033±0.027
DNN	90.1±3.8	93.3±3.3	0.041±0.024
SWN	89.5±5.4	**95.4±2.4**	0.055±0.043
AirwayNet-SE w/o Connectivity modeling	92.7±3.2	89.2±5.1	**0.009±0.008**
AirwayNet-SE	**93.0±3.1**	92.4±4.0	0.018±0.012
Private testing set			
Method	DSC (%)	TPR (%)	FPR (%)
AirwayNet	88.5±4.0	**86.5±6.0**	0.033±0.029
DNN	86.9±6.8	82.0±11.2	0.022±0.021
SWN	87.4±4.7	83.4±7.6	0.027±0.023
AirwayNet-SE w/o Connectivity modeling	87.2±6.5	80.7±10.5	**0.013±0.014**
AirwayNet-SE	**88.7±5.3**	84.6±8.7	0.020±0.017

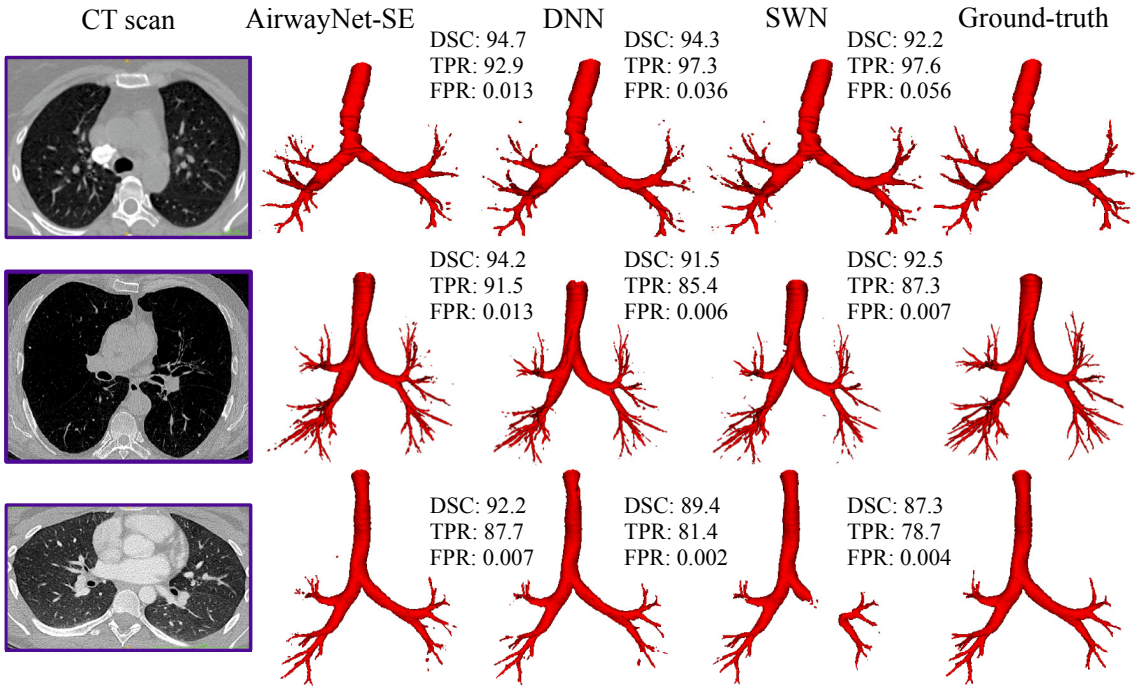


Figure 4.7: Comparison of airway segmentation results between the AirwayNet-SE, DNN, SWN, and ground-truth.

AirwayNet-SE and there exists less discontinuity in the predicted airway regions. Furthermore, SWN performed worse on thick bronchi. In contrast, the prediction of DNN was prone to missing thin bronchiole and produce ruptures.

4.4 Discussion

Table 4.1 demonstrated that the two state-of-the-art methods did not generalize well on the independent private testing set. Although Jin et al. [Jin et al., 2017] achieved the highest TPR of 94.7% on the public testing set, performance of their method on the private set dropped by over 4.2% and 13% in DSC and TPR, respectively. Since the intensity distribution of airway lumen and wall may differ in different CT scans, it is important for CNNs to handle such variation in airway segmentation. The relative robustness of AirwayNet and AirwayNet-SE confirmed the effectiveness of connectivity modeling. Besides, considering that DSC is a comprehensive metric that weighs up both sensitivity and specificity, we believe the highest DSC of AirwayNet-SE with comparable TPR and FPR verified its superiority over other methods.

Table 4.2 reflected that the feature fusion from two context scales did improve the performance of airway segmentation. Due to the differences between DNN and SWN in model design, their performance varied accordingly. Compared with DNN, SWN used less pooling layers and wider convolution features. More small, peripheral airway voxels are preserved with SWN and therefore its TPR is relatively higher. Meanwhile, due to the small receptive field of SWN, it did not perceive well the large, thick branches. The AirwayNet-SE combined the advantages of both DNN and SWN by fusing the features of global-scale and local-scale contexts. Moreover, the performance gains in DSC and TPR brought by connectivity modeling confirmed its effectiveness to AirwayNet-SE. Owing to the explicit modeling of airway connectivity, the proposed AirwayNet-SE enriched details of segmented peripheral bronchi and therefore achieved more accurate segmentation results.

4.5 Conclusion

This chapter introduced the AirwayNet and its variant AirwayNet-SE for airway segmentation. The proposed two methods explicitly learn voxel connectivity to perceive airway's inherent structure. By connectivity modeling, conventional segmentation task is transformed into 26 tasks of connectivity prediction, with each task classifying airway voxels along a certain connectivity orientation. Besides, the AirwayNet-SE goes one-step further by fusing features of two context scales. Experimental results proved that our approach was effective at overcoming the distribution differences between large and small airways. The airway annotations were also released to boost research on airway extraction using supervised learning methods.

In the future, the proposed method could further be improved by: (1) the adoption of generative adversarial networks to produce various training samples to improve robustness on CT scans of unhealthy patients; (2) the exploration of specific techniques (e.g., Hessian-based filtering) for enhancing thin bronchi details in low-quality CT scans.

Chapter 5

Learning Tubule-Sensitive Convolutional Neural Networks for Pulmonary Airway and Artery-Vein Segmentation in CT

Contents

5.1 Introduction 146
5.2 Methodology 150
5.2.1 Feature Recalibration 151
5.2.2 Attention Distillation 153
5.2.3 Anatomy Prior for Artery-Vein Segmentation 154
5.2.4 Model Design 156
5.2.5 Training Loss 157
5.3 Experiments and Results 157
5.3.1 Materials 157
5.3.2 Implementation Details 158
5.3.3 Evaluation Metrics 159
5.3.4 Results 161
5.4 Discussion 168
5.5 Conclusion 175

5.1 Introduction

Pulmonary diseases pose high risks to human health. As a diagnostic tool, computed tomography (CT) has been widely adopted to reveal tomographic patterns of pulmonary diseases. It is of significant clinical interest to study pulmonary structures in volume-of-interest (VOI). One prerequisite step is to extract pulmonary airways from CT. The modeling of airway tree benefits the quantification of its morphological changes for diagnosis of bronchial stenosis, acute respiratory distress syndrome, idiopathic pulmonary fibrosis, chronic obstructive pulmonary diseases (COPD), obliterative bronchiolitis, and pulmonary contusion [Howling et al., 1998, Shaw et al., 2002, Fetita et al., 2004, Li et al., 2019, Wu et al., 2019]. Combined with photo-realistic rendering and projection, the segmented airways play an important role in virtual bronchoscopy and endobronchial navigation for surgery [Mori et al., 2000, Natori et al., 2005, Shen et al., 2015a, Shen et al., 2019]. Another essential step is to extract pulmonary arteries and veins from CT. Pulmonary diseases may affect artery or vein, or both but in different ways [Melot and Naeije, 2011,Charbonnier et al., 2015]. Morphological changes of arteries are measured in diagnosing pulmonary embolism, arteriovenous malformations, and COPD [Zhou et al., 2007, Wittenberg et al., 2012, Cartin-Ceba et al., 2013, Estépar et al., 2013]. The arterial alterations also serve as an imaging biomarker in chronic thromboembolic pulmonary hypertension [Rahaghi et al., 2016]. Accurate separation of veins from arteries may improve computer-aided diagnosis of embolism because most false positive lesions were found in veins [Wittenberg et al., 2012]. The imaging features of veins are found useful in diagnosis of vein diseases [Porres et al., 2013]. Despite the benefits of airway and artery-vein segmentation, it requires heavy workloads for manual delineation due to the complexity of tubular structures. Consequently, automatic segmentation methods were developed to reduce burden and improve accuracy. Especially if arteries and veins can be extracted from non-contrast CT (i.e. CT without the use of contrast agents), CT pulmonary angiogram may not be needed in certain cases to avoid adverse reactions to contrast agents [Cochran et al., 2001, Loh et al., 2010].

Pulmonary Airway Segmentation Over the past decades, several methods have been proposed for airway segmentation [Mori et al., 2000, Van Rikxoort et al., 2009, Lo et al., 2012]. Most of them employed techniques such as adaptive thresholding, region growing and filtering-based enhancement. These methods successfully segmented thick bronchi, but often failed to extract peripheral bronchioles due to the fact that the intensity contrast between airway lumen and wall weakens as airways bifurcate into thinner branches. Recent progress of convolutional neural networks (CNNs) has spawned research on airway segmentation using CNNs [Selvan et al., 2020, Juarez et al., 2019, Wang et al., 2019, Qin et al., 2019, Yun et al., 2019, Charbonnier et al., 2017, Meng et al., 2017, Jin et al., 2017, Juarez et al., 2018, Zhao et al., 2019]. Two-dimensional (2-D) and 2.5-D CNNs [Yun et al., 2019, Charbonnier et al., 2017] were respectively applied on the initial coarsely segmented bronchi to reduce false positives and increase length of the detected airway tree. Three-dimensional (3-D) CNNs were developed for direct airway segmentation in either a dynamic VOI-based tracking way [Meng et al., 2017] or a fixed-stride sliding window way [Juarez et al., 2018]. The spatial recurrent convolution layer and radial distance loss were proposed by [Wang et al., 2019] for tubular topology perception. In [Qin et al., 2019],

the airway segmentation task was transformed into 26-neighbor connectivity prediction task for inherent structure comprehension. Both 2-D and 3-D CNNs were combined with linear programming-based tracking in [Zhao et al., 2019]. Graph neural networks [Selvan et al., 2020, Juarez et al., 2019] were explored to incorporate neighborhood knowledge in feature utilization.

Pulmonary Artery-Vein Segmentation Previous methods on artery-vein separation relied on the enhanced or segmented vessels as premise [Buelow et al., 2005, Mekada et al., 2006, Saha et al., 2010, Gao et al., 2012, Park et al., 2013, Kitamura et al., 2016, Payer et al., 2016, Charbonnier et al., 2015, Nardelli et al., 2018]. To tackle the variety of vessels, combination of techniques such as local filtering and anatomical guidance is employed in the literature. Specifically, they utilized the proximity of airways to arteries for differentiation and suppressed airway walls to reduce false positives. Buelow *et al.* [Buelow et al., 2005] proposed a measure of "arterialness" by identifying airway candidates in the vicinity of given vessels and assigning high value to vessels that run in parallel with bronchi. Mekada *et al.* [Mekada et al., 2006] calculated the distance from vessels to airways and to inter-lobar fissures. Vessels closer to airways are arteries and those closer to fissures are veins. Both Saha *et al.* [Saha et al., 2010] and Gao *et al.* [Gao et al., 2012] combined distance transform and fuzzy connectivity with morphological opening for separation. Recently, three methods were developed to improve artery-vein segmentation [Charbonnier et al., 2015, Payer et al., 2016, Nardelli et al., 2018]. Charbonnier *et al.* [Charbonnier et al., 2015] first constructed a graph representation of the segmented vessels to extract sub-trees. These trees were grouped iteratively and the final classification was performed by comparing the volume size of the linked trees. Payer *et al.* [Payer et al., 2016] extracted vessel sub-trees and classified each sub-tree via integer programming. Two anatomy properties were used: 1) proximity of arteries to bronchi; 2) uniform distribution of arteries and veins. CNNs were at the first time introduced to artery-vein classification by Nardelli *et al.* [Nardelli et al., 2018]. Graph-cut was adopted as post-processing to remove spatial inconsistency.

Limitations and Challenges Despite the improved performance of pulmonary airway and artery-vein segmentation by deep learning, there still remain limitations and challenges to be overcome.

First, for both airway and artery-vein segmentation, the severe class imbalance between tubular foreground and background poses a threat to the training of 3-D CNNs. Most CNNs heavily rely on airway and vessel ground-truth as supervisory targets. Unlike bulky or spheroid-like organs (e.g., liver and kidney), tree-like airways, arteries, and veins are thin, tenuous and divergent. The number of annotated voxels are far fewer than that of background voxels in the thoracic cavity. It is difficult to train deep models using such sparse and scattered targets. Although weighted cross-entropy loss and data sampling strategies were proposed to focus on the minority, single source of sparse desired targets from deficient airway and artery-vein labels still makes optimization ineffective.

Second, the spatial distribution and branching pattern of airways and vessels require the model to utilize both global-scale and local-scale context to perceive the main body (e.g., trachea, main branches) and limbs (e.g., peripheral bronchi and vessels). Previous

deep learning models used 2 or 3 pooling layers and the coarsest resolution features provide limited long-range context. If more layers are simply piled up, the increased parameters may cause over-fitting due to inadequate training data. If the width of CNNs (a.k.a. number of feature channels) is sacrificed for the depth (a.k.a. number of convolution layers) to avoid such parameter "explosion", the model's learning and fitting capacity may get restricted.

Third, it is more rigorous to deem pulmonary artery-vein separation methods in the literature as classification rather than segmentation. They used two-stage strategy and counted on vessel segmentation in the first stage. The subsequent artery-vein separation was treated as an independent classification task in the second stage. Different techniques were deployed in two stages and such isolation has two drawbacks: 1) It blocks the path for the second model to exploit rich context from the first one, especially when CNNs are applied as backbone. CNNs cannot take advantage of high correlation between the two tasks and have to learn from scratch. 2) The performance of artery-vein segmentation is largely affected by that of vessel extraction and errors accumulate along the whole pipeline.

Last but not least, for artery-vein separation, previous CNNs-based method [Nardelli et al., 2018] did not consider the relationship between airways, arteries, and veins. Auxiliary anatomy prior (e.g., close proximity of arteries to airways, intensity similarity between airway walls and vessels) was not involved in algorithm design, leaving room for further improvement.

Contributions To address these concerns, we present a CNNs-based method for pulmonary airway and artery-vein segmentation. Since airways, arteries, and veins are all tubular structures, they are collectively referred to as tubules in the present study. With the carefully designed constituent modules, the proposed method learns to comprehend the contour shape, intensity distribution, and connectivity of bronchi and vessels in a data-driven way. It tackles the challenges of applying CNNs to recognition of long, thin tubules and enjoys high sensitivity to bronchioles, arterioles, and venules.

First, we propose a feature recalibration module to maximally utilize the features learned from CNNs. On one hand, to increase the field-of-view for large context comprehension, deep architectures with multiple convolution and pooling layers are preferred. Accordingly, the number of learnable parameters increases and then overfitting becomes a problem. On the other hand, if the number of feature channels is simply reduced to avoid over-fitting, it might go to the other extreme where the model fails to learn discriminative features. Therefore, feature recalibration is considered because it intensifies task-related features given a moderate model size. In the design of recalibration module, we hypothesize that spatial information of features is indispensable for channel-wise recalibration and should be treated differently from position to position and layer to layer. The average pooling used in [Rickmann et al., 2019, Zhu et al., 2019] for spatial compression may not well capture the location of airways and vessels in different resolution scales. In contrast, we aim at prioritizing information at key positions with learnable weights, which provides appropriate spatial hints to model inter-channel dependency and thereafter improves recalibration.

Second, we introduce an attention distillation module to reinforce representation learning of tubular airway, artery, and vein. Attention maps of different scales enable

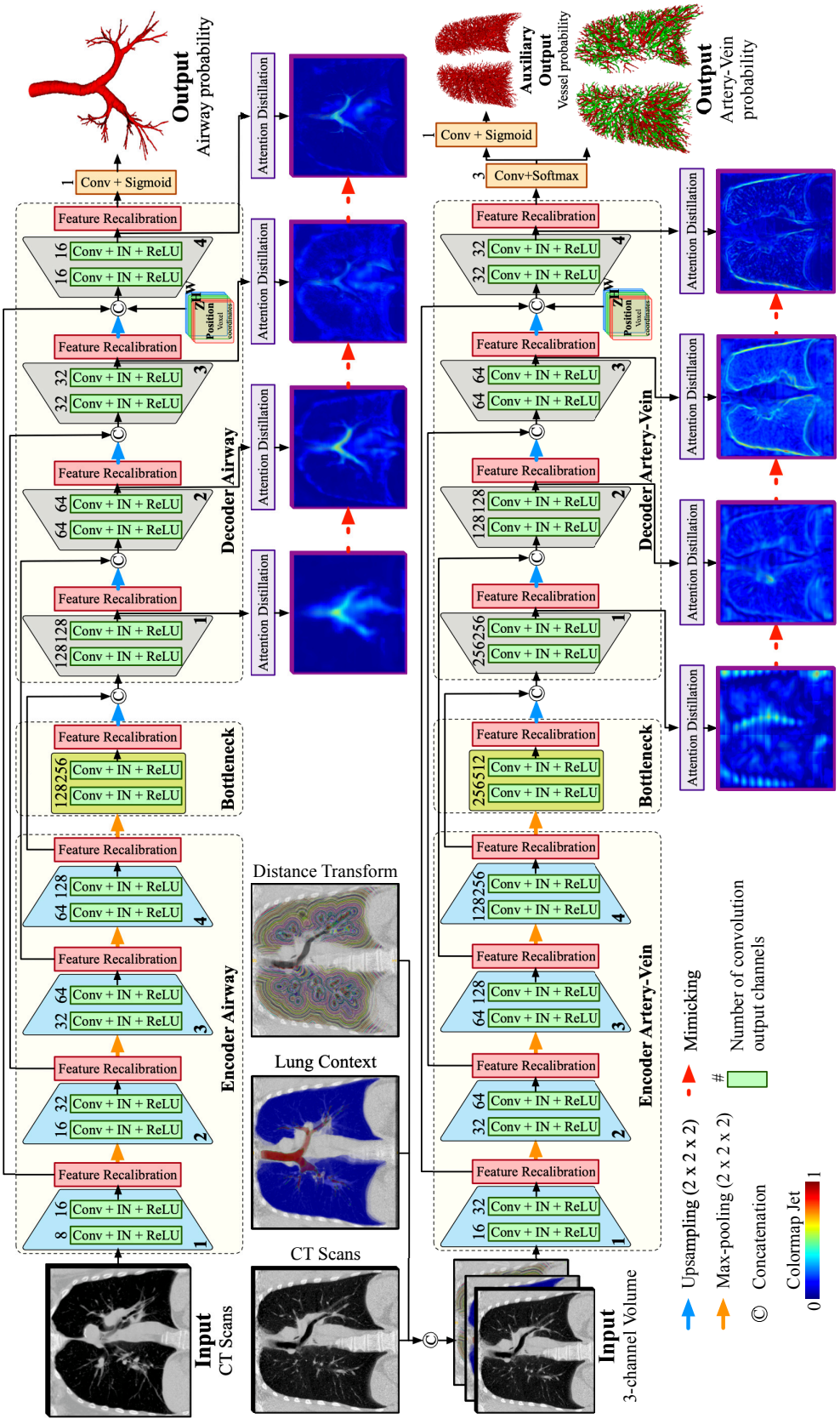


Figure 5.1: Overview of the proposed method for pulmonary airway and artery-vein segmentation. Instance normalization and ReLU activation are performed after each convolution layer except the last one. The number of convolution kernels is denoted above each layer.

us to potentially reveal the morphology and distribution pattern of airways and vessels. Inspired by knowledge distillation [Zagoruyko and Komodakis, 2017, Hou et al., 2019], we refine the attention maps of lower resolution by mimicking those of higher resolution. Finer attention maps (teacher's role) with richer context can cram coarser ones (student's role) with details about airways, arteries, and veins. The model's ability to recognize delicate branches is ameliorated after recursively focusing on the target anatomy. Dealing with insufficient supervisory desired targets, the distillation itself acts as an auxiliary learning task that provides extra gradients to assist training.

Third, we incorporate anatomy prior into artery-vein segmentation by introducing lung context map and distance transform map. The lung context map, containing automatically segmented airway lumen, airway wall, and lung, explicitly informs the model of semantic knowledge. The distance transform map, computed using extracted airways, records the distance of each voxel to its nearest airway wall.

Fourth, the proposed end-to-end method is applicable for both pulmonary airway and artery-vein segmentation. We do not perform independent vessel segmentation beforehand and require no post-refinement on the outputs of CNNs. The sliding window-based segmentation is used and each voxel's coordinates within the thoracic cavity are fed into the model to make up for the loss of position information.

Finally, although the entire framework is an integrated solution to airway and artery-vein segmentation, its constituting components can be considered for designing solutions to other tasks. The proposed method may also be readily extended by incorporating traditional techniques as post-processing (e.g., graph-cuts [Boykov et al., 2001]), where explicit graph and connectivity modeling are introduced specifically for tubular structures.

Our contributions can be briefly summarized as follows:

- We present a tubule-sensitive CNNs-based method for pulmonary airway and artery-vein segmentation. To our best knowledge, this method represents the first attempt to segment airways, arteries, and veins simultaneously.
- We propose a feature recalibration module that integrates prioritized spatial knowledge for channel-wise recalibration. It encourages discriminative feature learning.
- We introduce an attention distillation module to reinforce representation learning of tubular airway, artery, and vein. No extra annotation labor is required.
- We incorporate explicit anatomy prior into artery-vein segmentation by utilizing the lung context map and distance transform map as additional inputs.
- We respectively validate the proposed method on 110 and 55 non-contrast clinical CT scans for pulmonary airway and artery-vein segmentation. Extensive experiments show that our method achieved superior sensitivity to thin airways, arteries, and veins, with surpassing or competitive overall segmentation performance maintained.

5.2 Methodology

Overview of the proposed airway and artery-vein segmentation methods is illustrated in Fig. 5.1. To fulfill effective feature learning of tubular targets, feature recalibration and

attention distillation modules are introduced into CNNs. Anatomy prior is included to provide semantic knowledge for artery-vein task.

Given an input CT volume X, our segmentation process can be formulated as $P_{Target} = \mathcal{F}(X)$, where *Target* can be airway, artery, or vein and P_{Target} denotes its corresponding predicted probability. The objective is to learn an end-to-end mapping $\mathcal{F}$ via CNNs to minimize the difference between P_{Target} and its ground-truth label Y_{Target}. Assuming CNNs have M convolution layers in total, we denote the activation output of the m-th convolution as $A_m \in R^{C_m \times D_m \times H_m \times W_m}$, $1 \leq m \leq M$. The number of its channels, depths, heights, and widths are respectively denoted as C_m, D_m, H_m and W_m.

5.2.1 Feature Recalibration

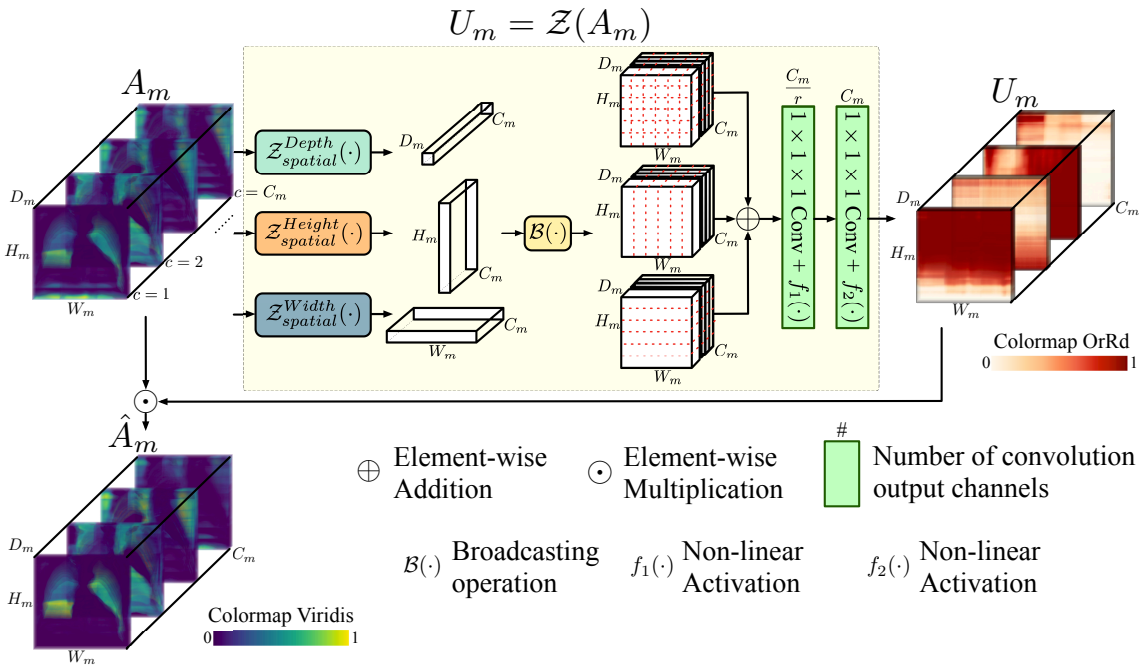


Figure 5.2: Illustration of the mapping $\mathcal{Z}(\cdot)$ for feature recalibration. Its input is the activated feature A_m of the m-th convolution layer. First, spatial map that highlights important regions is integrated through $\mathcal{Z}_{spatial}(\cdot)$ along three axes of depth, height, and width. Second, channel recombination is performed on the spatial map to compute the channel descriptor U_m. The final element-wise multiplication between A_m and U_m produces the recalibrated feature $\hat{A}_m$. The notations r, C_m, D_m, H_m, and W_m refer to the channel compression factor, the number of channels, depths, heights, and widths of A_m, respectively.

We propose the mapping $\mathcal{Z}(\cdot)$ that generates a channel descriptor $U_m = \mathcal{Z}(A_m)$ to recalibrate the activated convolutional feature A_m. An overview of $\mathcal{Z}(\cdot)$ for feature recalibration is given in Fig. 5.2. The channel-wise weight map U_m is learned to not only unearth crucial spatial locations of A_m, but also strengthen basis channels that affect most the output decision. The mapping $\mathcal{Z}(\cdot)$ is composed of two steps: 1) spatial knowledge

integration to obtain the compressed spatial map; 2) channel-wise recombination of such spatial map. For the first step, previously proposed feature recalibration methods [Rickmann et al., 2019, Zhu et al., 2019] treated all positions equally by condensing spatial information of features into vector or scalar value with average pooling, which might not be appropriate for processing large volumetric features. Instead, we integrate spatial information using weighted combination of features along each spatial dimension. Our hypothesis is that different positions may hold different degrees of importance both within the current A_m and across resolution scales (e.g., the shallower A_{m-1} and the deeper A_{m+1}). An operation like adaptive or global pooling is not spatially discriminating between the finest features (containing thin bronchioles, arterioles, and venules which are easily "erased" by averaging) and the coarsest features (containing mostly thick bronchi and vessels). Therefore, we introduce the following spatial integration method $\mathcal{Z}_{spatial}(\cdot)$ that preserves relatively important regions. It can be formulated as:

$$\begin{aligned}\mathcal{Z}_{spatial}(A_m) = \mathcal{B}(\mathcal{Z}^{Depth}_{spatial}(A_m)) + \mathcal{B}(\mathcal{Z}^{Height}_{spatial}(A_m)) \\ + \mathcal{B}(\mathcal{Z}^{Width}_{spatial}(A_m)),\end{aligned} \tag{5.1}$$

$$\begin{aligned}&\mathcal{Z}^{Depth}_{spatial}(A_m) = \sum_{j=1}^{H_m} h_j \sum_{k=1}^{W_m} w_k A_m[:,:,j,k], \\ &\mathcal{Z}^{Depth}_{spatial}(A_m) \in R^{C_m \times D_m \times 1 \times 1},\end{aligned} \tag{5.2}$$

$$\begin{aligned}&\mathcal{Z}^{Height}_{spatial}(A_m) = \sum_{i=1}^{D_m} d_i \sum_{k=1}^{W_m} w_k A_m[:,i,:,k], \\ &\mathcal{Z}^{Height}_{spatial}(A_m) \in R^{C_m \times 1 \times H_m \times 1},\end{aligned} \tag{5.3}$$

$$\begin{aligned}&\mathcal{Z}^{Width}_{spatial}(A_m) = \sum_{i=1}^{D_m} d_i \sum_{j=1}^{H_m} h_j A_m[:,i,j,:], \\ &\mathcal{Z}^{Width}_{spatial}(A_m) \in R^{C_m \times 1 \times 1 \times W_m},\end{aligned} \tag{5.4}$$

where indexed slicing (using Python notation) and broadcasting $\mathcal{B}(\cdot)$ are performed. Notations C_m, D_m, H_m, and W_m are referring to the number of channels, depths, heights, and widths of the m-th layer feature A_m. The learnable parameters d_i, h_j, w_k denote the combination weights for each feature slice in depth, height, and width dimension, respectively. During training, crucial airway and artery-vein regions are gradually preferred with higher weights while uninformative corner regions are neglected with lower weights. For the second step, we apply the excitation technique [Rickmann et al., 2019] on the compressed spatial map to model inter-channel dependency. Specifically, the channel descriptor U_m is obtained by:

$$U_m = \mathcal{Z}(A_m) = f_2(K_2 * f_1(K_1 * \mathcal{Z}_{spatial}(A_m))), \tag{5.5}$$

where K_1, K_2 are 3-D kernels of size $1\times1\times1$ and "$*$" denotes convolution. Convolving with K_1 decreases the channel number to C_m/r and that with K_2 recovers back to C_m. The ratio r is the compression factor that determines reduction extent. $f_1(\cdot)$ and $f_2(\cdot)$ are nonlinear activation functions. We choose Rectified Linear Unit (ReLU) as $f_1(\cdot)$ and Sigmoid

as $f_2(\cdot)$ in the present study. Multiple channels are recombined through such channel reduction and increment, with informative ones emphasized and redundant ones suppressed. Given the activated convolutional feature A_m and its channel descriptor U_m, the recalibrated feature $\hat{A}_m$ is defined as:

$$\hat{A}_m = U_m \odot A_m, \tag{5.6}$$

where $\odot$ denotes element-wise multiplication.

5.2.2 Attention Distillation

In both airway and artery-vein segmentation tasks, the segmentation model is required to identify thin tubules like distal bronchi, arteries, and veins. It could be expected that reinforced attention on such objects during feature learning may conduce to improved performance. Recent studies [Zagoruyko and Komodakis, 2017, Hou et al., 2019] on knowledge distillation showed that attention maps serve as valuable knowledge and can be transferred layer-by-layer from teacher networks to student networks. Motivated by knowledge transferability and self-attention mechanism, we introduce the attention distillation module into our 3-D CNNs for recognition of narrow, thin objects. The activation-based attention maps, which guide where to look at, are distilled and exploited during backward transfer process. Without separately setting two different models, later layers play the role of teacher and "impart" such attention to earlier layers within the same model. Besides, to tackle insufficient supervisory targets caused by the severe class imbalance, the distillation can be viewed as another source of supervision. It produces additional gradients by forcing low-resolution attention maps to resemble high-resolution ones, aiding the training of deep CNNs. Specifically, the attention distillation is performed between two consecutive features A_m and A_{m+1}.

Firstly, the attention map is generated by $G_m = \mathcal{G}(A_m), G_m \in R^{1\times D_m \times H_m \times W_m}$. Each voxel's absolute value in G_m reflects the contribution of its correspondence in A_m to the entire segmentation model. One way of constructing the mapping function $\mathcal{G}(\cdot)$ is to compute the statistics of activation values A_m across channel:

$$G_m = \sum_{c=1}^{C_m} |A_m[c,:,:,:]|^p, \tag{5.7}$$

The element-wise operation $|\cdot|^p$ denotes the absolute value raised to the p-th power. More attention is addressed to highly activated regions if $p > 1$. Here, we adopt channel-wise summation instead of maximizing $\max_c(\cdot)$ or averaging $\frac{1}{C_m}\sum_{c=1}^{C_m}(\cdot)$ because it is relatively less biased. The sum operation retains all implied salient activation information without ignoring non-maximum elements or weakening discriminative elements. For intuitive comparison of different $\mathcal{G}(\cdot)$, visualization of 3-D attention maps on 2-D plane is presented in Fig. 5.3 by first choosing multiple 2-D slices that contain airways and vessels and then super-imposing them together with opacity of 30%. Visual comparison exhibits that summation with $p > 1$ intensifies most the sensitized task-related regions (e.g., lung borders, bronchi, vessels).

Secondly, trilinear interpolation $\mathcal{I}(\cdot)$ is performed to ensure that processed 3-D attention maps share the same dimension.

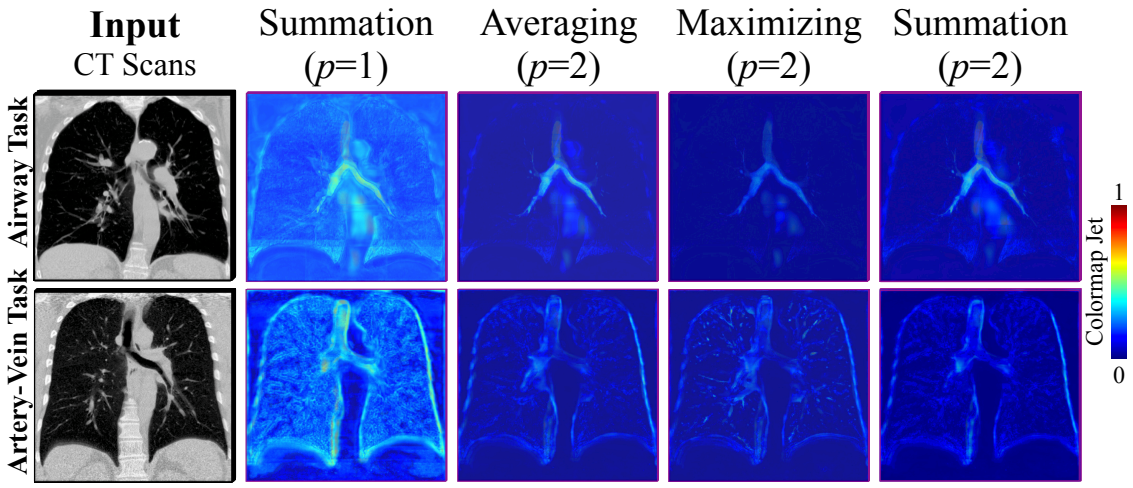


Figure 5.3: Difference among mapping functions $\mathcal{G}(\cdot)$ of computing the last attention map in decoder for airway and artery-vein segmentation tasks.

Then, voxel-wise Softmax $\mathcal{S}(\cdot)$ is spatially applied to normalize all elements in $[0,1]$. Finally, we drive the distilled attention $\hat{G}_m$ closer to $\hat{G}_{m+1}$ by minimizing the loss:

$$\mathcal{L}_{distill} = \sum_{m=1}^{M-1} \|\hat{G}_m - \hat{G}_{m+1}\|_F^2, \hat{G}_m = \mathcal{S}(\mathcal{I}(G_m)), \tag{5.8}$$

where $\|\cdot\|_F^2$ is the squared Frobenius norm. With $\hat{G}_m$ recursively mimicking its successor $\hat{G}_{m+1}$, visual attention is transmitted from the deepest to the shallowest layer. Note that such distillation process does not require extra annotation labor and can work with arbitrary CNNs readily. In implementation, to prevent the latter attention $\hat{G}_{m+1}$ from approximating the previous $\hat{G}_m$, we detach $\hat{G}_{m+1}$ from the computation graph for each m in loss calculation. Consequently, $\hat{G}_{m+1}$ will not be changed by back-propagating errors. The reasons why we do not down-sample $\hat{G}_{m+1}$ to the size of $\hat{G}_m$ is that $\hat{G}_{m+1}$ at decoder side has higher resolution than $\hat{G}_m$ by nature and down-sampling loses rich information that only exists in $\hat{G}_{m+1}$. It is necessary to keep $\hat{G}_{m+1}$ unchanged so that the resultant distillation loss between $\hat{G}_m$ and $\hat{G}_{m+1}$ can improve model's attention on fine details about targets.

5.2.3 Anatomy Prior for Artery-Vein Segmentation

In the present study, artery-vein segmentation in the lung hilum (e.g., pulmonary trunk, left and right main pulmonary veins) is excluded since recognition of these vessels in non-contrast CT is extremely difficult for both computers and medical experts [Charbonnier et al., 2015]. Considering that the valid artery and vein targets are mainly restricted inside the two lungs, we believe it is reasonable to provide segmented lung masks as VOI hint. The lung segmentation is performed by: 1) binarization using OTSU thresholding [Otsu, 1979]; 2) hole filling using morphological operations; 3) selection of the two largest connected components as left and right lungs; 4) convex hull computation

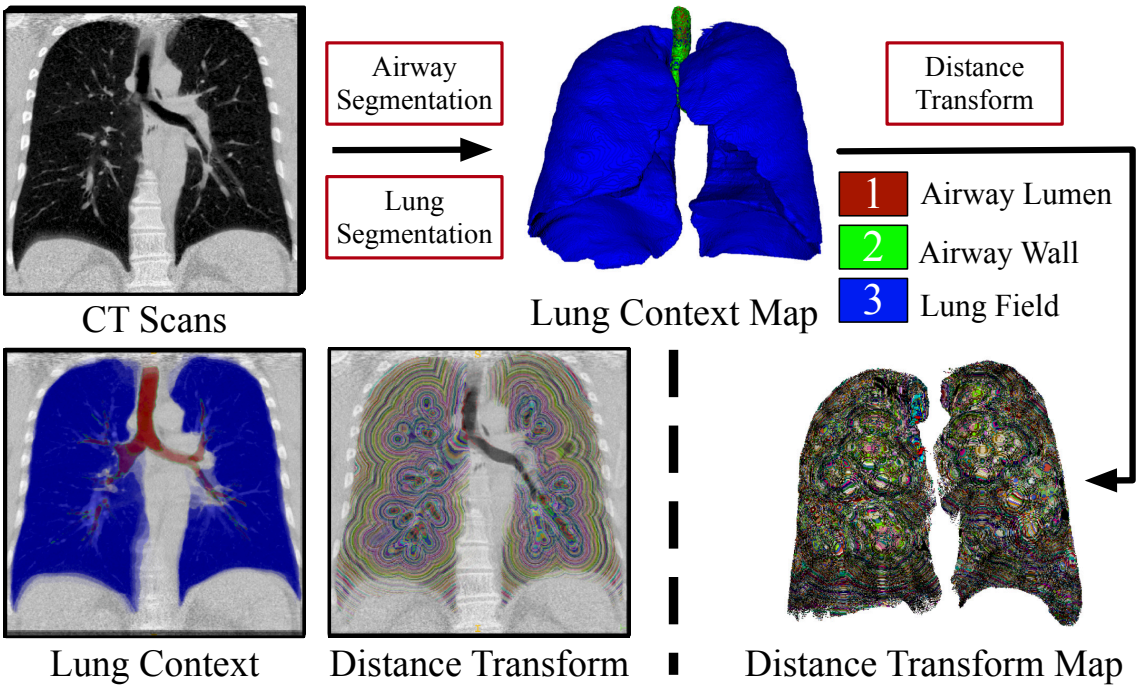


Figure 5.4: Illustration of anatomy prior incorporation. Visual display of the generated lung context maps and distance transform maps superimposed on CT scans is given in bottom left.

to prevent over-segmentation. Besides, we hypothesize that proper representation of airway anatomy is beneficial for the model to distinguish between vessels and airway walls, where similar intensity distribution is shared. Therefore, automatic airway segmentation is first performed using the proposed method to obtain airway lumen. Then, assuming the thickness of airway wall is less than 2 mm [van Dongen and van Ginneken, 2010], we extract airway wall by subtracting airway lumen from its morphological dilation result. The structuring element for dilation is a sphere with diameter of 3 voxels. Given the segmented airway lumen, wall, and lung field, we respectively label them as 1, 2 and 3 to generate the lung context map.

Since pulmonary arteries inside lungs often accompany with airways in parallel, we believe the proximity of arteries to airways might be informative for the segmentation model to discriminate arteries from veins [Hislop, 2002, Miller, 1947, Berend et al., 1979, Kandathil and Chamarthy, 2018]. Consequently, Euclidean distance transform is performed on the segmented airways to calculate the distance of each voxel to its nearest airway wall. The computed distance transform map is multiplied with lung mask to keep valid regions.

To summarize, two maps are introduced as anatomy prior for artery-vein segmentation (see Fig. 5.4): lung context map and distance transform map. The first map offers extra semantic knowledge of lung and the second map reflects voxels' closeness to airway. These maps are concatenated with CT sub-volume as inputs to the artery-vein segmentation model.

5.2.4 Model Design

The proposed method employs 3-D U-Net [Çiçek et al., 2016] as network backbone. Such encoder-decoder CNNs first extract a condensed representation of input image and then reconstruct it in response to different tasks. To enlarge the receptive field of CNNs and facilitate feature learning of long-range relationship, four pooling layers are used with five resolution scales involved in total. At each scale, both encoders and decoders have two convolution layers (kernel size $3 \times 3 \times 3$) followed by instance normalization and ReLU. The feature recalibration module is inserted at the end of each resolution scale. Since high-level features in decoders are also of high-resolution and high-relevance to segmented targets, we perform the decoder-side attention distillation to pass down the fine-grained details that are missing in previous low-resolution attention maps. The encoder-side distillation is not favored because low-level features are more local-scale and general. Furthermore, voxel coordinate map, which records voxels' global position inside the thoracic cavity, is concatenated with features at decoder 4 to make the model explicitly consider location. In view of the patch-wise training, such coordinate map is used to offset the loss of position information. Since both arteries and veins are vessels, we introduce an auxiliary task of vessel segmentation by adding another convolution layer with sigmoid activation to the artery-vein output. Such multi-head design takes advantage of: 1) the inclusion relationship between vessel and artery-vein; 2) the reduced difficulty of learning to recognize vessels. Preliminary ablation study on the auxiliary vessel segmentation output has confirmed its effectiveness. The probability outputs of airway and artery-vein are respectively obtained using sigmoid and softmax activation. Preliminary experiments confirmed that such model design and feature number choice

are optimum for our tasks.

5.2.5 Training Loss

To deal with hard samples, we use both the Dice [Milletari et al., 2016] and Focal loss [Lin et al., 2017] for training CNNs. For airway segmentation, given the binary label $y^a(x)$ and prediction $p^a(x)$ of each voxel x in the volume set X, the combined loss is defined as:

$$\begin{aligned}\mathcal{L}_{Airway} = -(&\frac{2\sum_{x\in X} p^a(x)y^a(x)}{\sum_{x\in X}(p^a(x)+y^a(x))+\epsilon} \\ &+ \frac{1}{|X|}\sum_{x\in X}(1-p_t^a(x))^2\log(p_t^a(x))),\end{aligned} \tag{5.9}$$

where $p_t^a(x) = p^a(x)$ if $y^a(x) = 1$. Otherwise, $p_t^a(x) = 1 - p^a(x)$. Parameter ϵ is used to avoid division by zero. For multi-class artery-vein and binary vessel segmentation tasks, the losses are defined as the following:

$$\begin{aligned}\mathcal{L}_{A-V} = -\frac{1}{3}\sum_{i=0}^{2}(&\frac{2\sum_{x\in X} p_i^{av}(x)y_i^{av}(x)}{\sum_{x\in X}(p_i^{av}(x)+y_i^{av}(x))+\epsilon} \\ &+ \frac{1}{|X|}\sum_{x\in X}(1-p_{it}^{av}(x))^2\log(p_{it}^{av}(x))), \\ \mathcal{L}_{Vessel} = -(&\frac{2\sum_{x\in X} p^v(x)y^v(x)}{\sum_{x\in X}(p^v(x)+y^v(x))+\epsilon} \\ &+ \frac{1}{|X|}\sum_{x\in X}(1-p_t^v(x))^2\log(p_t^v(x))),\end{aligned} \tag{5.10}$$

where $p_i^{av}(x)$ and $y_i^{av}(x)$ are respectively the artery-vein prediction and label of the i-th class ($i = 0$ for background, 1 for artery, and 2 for vein), $p_{it}^{av}(x) = p_i^{av}(x)$ if $y_i^{av}(x) = 1$ and otherwise $p_{it}^{av}(x) = 1 - p_i^{av}(x)$. The vessel prediction and label are respectively denoted as $p^v(x)$ and $y^v(x)$, with $y^v(x) = 1$ if x belongs to artery or vein. Similarly, we have $p_t^v(x) = p^v(x)$ if $y^v(x) = 1$. Otherwise, $p_t^v(x) = 1 - p^v(x)$. The total losses for training the airway segmentation model and artery-vein segmentation model are respectively:

$$\mathcal{L}_{Airway}^{total} = \mathcal{L}_{Airway} + \alpha\mathcal{L}_{distill} \tag{5.11}$$

$$\mathcal{L}_{A-V}^{total} = \mathcal{L}_{A-V} + \mathcal{L}_{Vessel} + \alpha\mathcal{L}_{distill}, \tag{5.12}$$

where α balances these terms.

5.3 Experiments and Results

5.3.1 Materials

In total, 110 and 55 non-contrast CT scans from multiple sources were respectively used for airway and artery-vein segmentation. We respectively conducted experiments for these two tasks with non-overlapping datasets. The acquisition and investigation of data were conformed to the principles outlined in the declaration of Helsinki [Association et al., 2001].

Pulmonary Airway Segmentation

We used 110 chest CT scans for airway segmentation. They were collected from two public datasets: a) 70 scans from LIDC-IDRI [Armato et al., 2011]; b) 40 scans from the EXACT'09 [Lo et al., 2012]. The LIDC-IDRI dataset contains 1018 CT scans with pulmonary nodule annotations. In view of image quality, 70 scans with slice spacing less than 0.625 mm were randomly chosen. The EXACT'09 Challenge provides 20 scans in the training set and 20 scans in the testing set. No airway annotation is openly available. The axial size of all CT scans is 512×512 pixels with spatial resolution of 0.5–0.781 mm. The number of slices in each scan varies from 157 to 764. Their slice spacing ranges between 0.45–1.0 mm. The airway ground-truth of each CT scan was acquired by: a) performing an interactive segmentation via ITK-SNAP [Yushkevich et al., 2006] to generate a rough airway tree; b) manual delineation and correction by well-trained experts. All 70 scans from LIDC-IDRI and 20 scans of the EXACT'09 training set were used in model training and evaluation. These 90 scans were randomly split into the training set (63 scans), validation set (9 scans) and testing set (18 scans). In addition, all 20 scans of the EXACT'09 testing set were reserved as an independent evaluation set. Segmentation results on this extra dataset were evaluated by EXACT'09 organizers for fair comparison experiments.

Pulmonary Artery-Vein Segmentation

We used all 55 non-contrast chest CT scans from CARVE14 [Charbonnier et al., 2015] for artery-vein segmentation. These scans share the same axial size of 512×512 pixels. They have the same slice spacing of 0.7 mm and the spatial resolution is 0.59–0.83 mm. The number of axial slices ranges from 349 to 498. Two kinds of artery-vein reference were available: 1) full annotations of 10 CT scans; 2) partial annotations of a small portion of vessel segments for the remaining 45 CT scans. We randomly split these 45 CT scans into the training set (40 scans) and validation set (5 scans). The 10 CT scans with full artery-vein labels were kept as the testing set. Since the number of labeled vessel segments of the 45 scans was too small to train and validate CNNs, we used semi-automatic segmentation results released by [Charbonnier et al., 2015] as complement target labels. Due to annotation difficulty, some voxels were marked as non-determined by two observers and we excluded these voxels in training and evaluation. Vessel roots, which are large main vessels entering from the lung hilum into the lung field, were not marked as they are too difficult to delineate in non-contrast CT [Charbonnier et al., 2015]. Disagreement between observers exists and vessel annotations in this region are unavailable. Therefore, the segmentation targets of interest are limited inside lungs.

5.3.2 Implementation Details

Since CT scans were acquired from different scanners using different parameter settings, data pre-processing is imperative before model training. The voxel intensity of all scans was truncated within the Hounsfield Unit (HU) window of $[-1000, 400]$ and normalized to $[0, 1]$. To avoid learning irrelevant marginal area outside lung, lung mask was extracted using the method mentioned in Sec. 5.2.3. The minimum bounding box of lung was cropped as valid input region. Here, isotropic resampling is not used because

it triggers off mismatch between CT images and ground-truth labels during voxel interpolation. The resampled annotations are discontinuous and incomplete, which are detrimental for CNNs to learn effective representation of long, thin, tubular targets. Due to GPU memory limit, CT scans were respectively cropped into sub-volume cubes of the size $80 \times 192 \times 304$ and $64 \times 176 \times 176$ for airway and artery-vein tasks. Such size was chosen specifically to maximally utilize the available GPU resource but is not related to the shape of the lung. Given the same GPU memory, the size of cropped cubes for artery-vein task is smaller than that for airway task because the proposed artery-vein model has around 4 times as many parameters as the proposed airway model. The size of the cropped cubes is kept the same for all phases of training, validation, and testing. In training phase, random horizontal flipping, shifting, Gaussian smoothing ($\sigma = 1$), and voxel intensity jittering were applied as on-the-fly data augmentation. The Adam optimizer was used with an initial learning rate of 3×10^{-3}. If the training loss stayed at a plateau for over 10 epochs, the learning rate was reduced by a factor of 10. With batch size of 1, training converged after 60 epochs for each model. In validation and testing phases, we performed sliding window prediction with axial stride of 64. Results were averaged on overlapping margins. Each voxel's category of artery-vein or airway was assigned by respectively performing channel-wise arg max or thresholding ($th = 0.5$) on the probability outputs. No post-processing was involved. All models were implemented in Python with PyTorch or Keras. Model training and hyper-parameter tuning were performed only on the training set. The model that achieved the best validation results was chosen and tested on the testing set for objectivity. In experiments, model training was executed on a Linux workstation with Intel Xeon Silver 4114 CPU, 192 GB RAM, and NVIDIA Tesla V100 GPU. Model inference and anatomy prior computation were carried out on a Linux PC with Intel Core i7-8700 CPU, 64 GB RAM, and NVIDIA Quadro P4000 GPU. The computational time of the proposed airway and artery-vein segmentation method is reported in Table 5.1 under the current hardware configuration. For hyper-parameter settings, we empirically chose $\alpha = 0.1$, $\epsilon = 10^{-7}$, $p = 2$ and $r = 2$. Note that current settings worked well for our tasks but are not necessarily optimum. Elaborate tuning may be conducted in specific tasks.

5.3.3 Evaluation Metrics

Pulmonary Airway Segmentation

For airway task, only the largest connected component of the binarized segmentation output was kept in view of clinical practice. Five metrics were used: (a) Branches detected (BD); (b) Tree-length detected (TD); (c) True positive rate (TPR); (d) False positive rate (FPR); (e) Dice similarity coefficient (DSC). We referred to [Lo et al., 2012] for definitions of metrics (a)–(d). The first two metrics are centerline-based measurements. We computed the centerlines of reference annotations using the algorithm described in [Lee et al., 1994]. Then, the centerlines were multiplied with segmentation results to compute the length of the overlapped centerlines L_{seg}. The fraction of the correctly segmented tree's length relative to the total tree length of the reference centerlines L_{ref} is defined as TD, $TD = \frac{L_{seg}}{L_{ref}} \times 100\%$. For any branch segment between two nodes (bifurcation node or terminal node) on the reference centerlines, if the segmentation results and this seg-

Table 5.1: Computational time of the proposed pulmonary airway and artery-vein segmentation method.

Item Name		Time (seconds)
Pulmonary Airway Segmentation	Training (# Epoch×Time Per Epoch)	60×3294.0±31.1
	Inference (Per CT Volume)	48.4±18.5
Anatomy Prior (Per CT Volume)	Lung Segmentation	115.9±95.8
	Airway Segmentation (Inference)	45.4±8.9
	Lung Context Map	3.7±0.4
	Distance Transform Map	9.2±1.1
Pulmonary Artery-Vein Segmentation	Training (# Epoch×Time Per Epoch)	60×2856.7±153.9
	Inference (Per CT Volume)	75.9±13.9

ment overlap with over 1 voxel, then this branch is counted as "detected". The number of branches that are successfully detected N_{seg} with respect to the total number of branches in reference N_{ref} is defined as BD, $BD = \frac{N_{seg}}{N_{ref}} \times 100\%$. The metrics of TPR, FPR, and DSC are voxel-based measurements. TPR is defined as the number of true airway voxels in segmentation results N_{TP} divided by the total number of airway voxels in reference N_P, $TPR = \frac{N_{TP}}{N_P} \times 100\%$. FPR is defined as the number of false airway voxels in segmentation results N_{FP} divided by the total number of background voxels in reference N_N, $FPR = \frac{N_{FP}}{N_N} \times 100\%$. With the categorized N_{TP}, N_{FP}, and N_P, DSC is given by: $DSC = \frac{2 \times N_{TP}}{N_{TP}+N_{FP}+N_P} \times 100\%$. Note that trachea region is excluded in calculating BD, TD, TPR, and FPR to reflect model's ability to extract peripheral airways. However, trachea is included in DSC computation as it measures overall segmentation quality.

Pulmonary Artery-Vein Segmentation

For artery-vein task, six metrics were used: (a) Accuracy (ACC); (b) TPR; (c) FPR; (d) DSC; (e) BD; (f) TD. All connected components of artery and vein subtrees were involved in measurements. We followed [Charbonnier et al., 2015] to report both mean and median ACC of artery-vein separation, with 95% confidence interval (CI) estimated. Other metrics were reported in mean ± standard deviation. The definitions of BD and TD are the same as those in airway tasks except that arteries and veins are first measured respectively to obtain the number of detected branches (N_{seg}^{artery}, N_{seg}^{vein}) and the total length of segmented subtrees (L_{seg}^{artery}, L_{seg}^{vein}). Then, BD and TD are given as the averaged artery and vein results over their corresponding ground-truth: $BD = \frac{1}{2} \times (\frac{N_{seg}^{artery}}{N_{ref}^{artery}} + \frac{N_{seg}^{vein}}{N_{ref}^{vein}}) \times 100\%$, $TD = \frac{1}{2} \times (\frac{L_{seg}^{artery}}{L_{ref}^{artery}} + \frac{L_{seg}^{vein}}{L_{ref}^{vein}}) \times 100\%$.

5.3.4 Results

Evaluation of the proposed method is structured as follows. First, we provide quantitative results of pulmonary airway and artery-vein segmentation in comparison with state-of-the-art methods. Second, ablation study is conducted to validate each constituting component of our method. Third, qualitative segmentation results are presented for visual analysis.

Comparison with the State-Of-The-Art Methods

Pulmonary Airway Segmentation Table 5.2 reports comparison results with state-of-the-art pulmonary airway segmentation methods. Since we adopted U-Net as network backbone, comparison experiments were performed with other encoder-decoder CNNs: the original 3-D U-Net [Çiçek et al., 2016], its variants V-Net [Milletari et al., 2016], VoxResNet [Chen et al., 2018], and Attention-Gated (AG) U-Net [Schlemper et al., 2019]. The network architecture of these methods has similar encoding path but varied decoding path. We also compared our method with five state-of-the-art methods: Wang et al. [Wang et al., 2019], Juarez et al. [Juarez et al., 2019], Qin et al. [Qin et al., 2019], Juarez et al. [Juarez et al., 2018] and Jin et al. [Jin et al., 2017]. These methods were re-implemented by ourselves and fine-tuned on the same dataset. Only methods in [Wang et al., 2019, Qin et al., 2019, Juarez et al., 2018, Jin et al., 2017] were reproduced with Keras. Other pulmonary airway or artery-vein segmentation methods were implemented with PyTorch. Furthermore, we evaluated our method on the independent testing set of EXACT'09 [Lo et al., 2012]. These 20 testing cases were not used for training or fine-tuning. For a fair comparison, results of three available metrics (BD, TD, and FPR) were given by EXACT'09 organizers and shown in Table 5.3.

Table 5.2 shows that under the same thresholding value ($th = 0.5$), the proposed method achieved the highest BD of 96.2%, TD of 90.7%, and TPR of 93.6% with a compelling DSC of 92.5%. Such high sensitivity was accompanied with an inferior FPR of 0.035%. Since the threshold th directly affects airway segmentation results, we adjusted th to enforce the same FPR for all methods. The FPR of 3-D U-Net [Çiçek et al., 2016] under $th = 0.5$ was chosen as the "anchor" FPR for alignment. Except V-Net [Milletari et al., 2016], AG U-Net [Schlemper et al., 2019], and the proposed method, all methods have to be thresholded with a rather low $th < 0.5$ to control FPR. Under the same FPR, results of state-of-the-art methods are closer to each other than those under the same $th = 0.5$. In that case, the proposed method still achieved the highest BD of 94.3%, TPR of 90.6%, and DSC of 93.5% with a competitive TD of 86.7%.

In Table 5.3, results of recent participants are reported by EXACT'09 organizers and are not publicly accessible. It is impossible to control all FPRs to be the same. Instead, we binarized our probability results with three thresholding values ($th = 0.1, 0.5, 0.8$) and submitted them for official evaluation. Different FPR levels are presented as reference. Under $th = 0.1$, the proposed method (FPR: 9.71%) achieved a 2.4% higher BD and a comparable TD with respect to team FF_ITC (FPR: 11.92%). Under $th = 0.5$, compared with teams HybAir (FPR: 6.78%) and NTNU (FPR: 3.60%), our method (FPR: 3.65%) achieved an over 25% higher BD and an over 28% higher TD. Under $th = 0.8$, we (FPR: 1.28%) obtained over 1.6 times higher BD and TD than teams Neko (FPR: 0.89%), UCCTeam (FPR:

Table 5.2: Comparison of pulmonary airway segmentation results. The results both under the same binarization threshold and under the same FPR are presented for each method. The FPR is controlled to be the same with 3-D U-Net (under threshold of 0.5) by respectively adjusting the binarization threshold on the probability outputs of each method.

Method	Params ($\times 10^4$)	*th*	BD (%)	TD (%)	TPR (%)	FPR (%)	DSC (%)
3-D U-Net [Çiçek et al., 2016]*	477.1	0.5	87.2±13.7	73.8±18.7	85.3±10.4	0.021±0.015	91.5±2.9
V-Net [Milletari et al., 2016]	1047.1		91.0±16.2	81.6±19.5	87.1±13.6	0.024±0.017	92.1±3.6
VoxResNet [Chen et al., 2018]	170.9		88.2±12.6	76.4±13.7	84.3±10.4	0.012±0.009	92.7±3.0
AG U-Net [Schlemper et al., 2019]	621.3		93.8±7.9	88.2±9.4	91.7±6.6	0.031±0.015	92.5±2.0
Wang et al. [Wang et al., 2019]	549.7		93.4±8.0	85.6±9.9	88.6±8.8	0.018±0.012	93.5±2.2
Juarez et al. [Juarez et al., 2019]	5.3		77.5±20.9	66.0±20.4	77.5±15.5	**0.009±0.009**	87.5±13.2
Qin et al. [Qin et al., 2019]	106.8		91.6±8.3	82.1±10.9	87.2±8.9	0.014±0.009	**93.7±1.9**
Juarez et al. [Juarez et al., 2018]	352.7		91.9±9.2	80.7±11.3	86.7±9.1	0.014±0.009	93.6±2.2
Jin et al. [Jin et al., 2017]	473.4		93.1±7.9	84.8±9.9	88.1±8.5	0.017±0.010	93.6±2.0
Our proposed	423.1		**96.2±5.8**	**90.7±6.9**	**93.6±5.0**	0.035±0.014	92.5±2.0
3-D U-Net [Çiçek et al., 2016]*	477.1	0.5	87.2±13.7	73.8±18.7	85.3±10.4	0.021±0.015	91.5±2.9
V-Net [Milletari et al., 2016]	1047.1	0.6	88.2±21.3	79.1±21.6	85.4±14.9	0.021±0.014	92.0±3.9
VoxResNet [Chen et al., 2018]	170.9	0.001	91.6±10.4	81.6±11.2	88.3±8.6	0.021±0.012	92.9±2.3
AG U-Net [Schlemper et al., 2019]	621.3	0.7	90.7±12.1	83.1±15.8	87.3±11.9	0.021±0.012	92.7±2.6
Wang et al. [Wang et al., 2019]	549.7	0.2	94.1±7.7	86.7±9.5	89.5±8.4	0.021±0.012	93.3±2.2
Juarez et al. [Juarez et al., 2019]	5.3	0.00001	86.5±16.3	76.3±17.8	84.7±12.5	0.021±0.012	89.7±9.2
Qin et al. [Qin et al., 2019]	106.8	0.005	93.7±6.3	85.7±9.5	89.6±8.0	0.021±0.011	93.3±1.7
Juarez et al. [Juarez et al., 2018]	352.7	0.05	93.7±7.8	83.4±10.1	89.3±7.9	0.021±0.011	93.4±1.9
Jin et al. [Jin et al., 2017]	473.4	0.005	94.3±7.3	**87.3±9.1**	89.6±8.0	0.021±0.012	93.5±1.9
Our proposed	423.1	0.77	**94.3±6.6**	86.7±8.5	**90.6±6.7**	**0.021±0.011**	**93.5±1.6**

* Feature channels were halved to have similar number of parameters with the proposed method.

Table 5.3: Evaluation results on the EXACT'09 testing set.

Participants†	BD (%)	TD (%)	FPR (%)
Neko	35.5±8.2	30.4±7.4	0.89±1.78
UCCTeam	41.6±9.0	36.5±7.6	**0.71±1.67**
FF_ITC	79.6±13.5	**79.9±12.1**	11.92±13.16
HybAir	51.1±10.9	43.9±9.6	6.78±26.60
MISLAB	42.9±9.6	37.5±7.1	0.89±1.64
NTNU	31.3±10.4	27.4±9.6	3.60±3.37
Our proposed ($th = 0.1$)	**82.0±9.9**	79.4±10.0	9.71±5.59
Our proposed ($th = 0.5$)	76.7±11.5	72.7±11.6	3.65±2.86
Our proposed ($th = 0.8$)	68.8±13.4	62.6±12.7	1.28±1.29

† Results of recent participants were directly quoted here.

0.71%), and MISLAB (FPR: 0.89%).

Pulmonary Artery-Vein Segmentation Table 5.4 gives comparison results with state-of-the-art pulmonary artery-vein segmentation methods. Apart from the well-known medical image segmentation models mentioned in Sec. 5.3.4, two recently proposed artery-vein classification methods were evaluated: Charbonnier et al. [Charbonnier et al., 2015] and Nardelli et al. [Nardelli et al., 2018]. Both two methods were developed to recognize arteries and veins from the already segmented vessels, where vessel segmentation was performed independently in advance. The comparison of the proposed method against labels yielded a mean ACC of 90.3%, a medium ACC of 90.9%, a TPR of 90.3%, a FPR of 0.151%, a DSC of 82.4%, a BD of 85.4%, and a TD of 90.9%. It outperformed state-of-the-art segmentation CNNs by a large margin in ACC, TPR, BD, and TD with comparable FPR and DSC. Admittedly, compared with methods that adopted graph-based representation for artery-vein separation [Charbonnier et al., 2015, Nardelli et al., 2018], the proposed method has room for improvement.

Ablation Study

We investigated the validity of key constituents of the proposed method: 1) feature recalibration (FR); 2) attention distillation (AD); 3) anatomy prior (AP). FR and AD were employed in both airway and artery-vein segmentation while AP was only used in artery-vein task. The model trained without FR, AD, and AP was indicated as baseline. Two very recently proposed feature recalibration modules (cSE [Zhu et al., 2019] and PE [Rickmann et al., 2019]) were introduced into our baseline for comparison. They were both adapted from the 2-D squeeze-and-excitation [Hu et al., 2018] technique for 3-D channel-wise feature recalibration. We replaced all FR with these two modules and trained models from scratch. For assessing AD, deep supervision (DS) [Zhu et al., 2017] was introduced for comparison. DS allows features of lower resolution to be supervised directly by targets. Specifically, we respectively added one convolution layer (kernel size $3 \times 3 \times 3$) and one trilinear upsampling layer to features of decoder 1–3. After sigmoid or softmax activation, these outputs were involved in loss computation for airway or artery-vein seg-

Table 5.4: Comparison of pulmonary artery-vein segmentation results.

Method	Params ($\times 10^4$)	ACC-mean [95%-CI] (%)	ACC-median [95%-CI] (%)	TPR (%)	FPR (%)	DSC (%)	BD (%)	TD (%)
3-D U-Net [Çiçek et al., 2016]*	1907.4	88.3 [85.4,91.1]	88.7 [85.3,93.6]	88.2±3.9	0.117±0.043	83.8±2.9	82.8±6.6	89.1±4.5
V-Net [Milletari et al., 2016]*	1047.1	84.4 [80.7,88.1]	84.9 [80.3,90.6]	84.4±4.9	0.104±0.047	82.7±4.5	80.0±7.5	85.9±5.9
VoxResNet [Chen et al., 2018]*	170.9	87.0 [83.9,90.0]	86.0 [84.3,93.7]	87.2±4.1	0.116±0.054	83.2±3.4	82.4±8.0	88.7±5.0
AG U-Net [Schlemper et al., 2019]*	621.4	84.6 [80.8,88.5]	85.9 [80.9,90.3]	84.7±5.2	0.098±0.037	83.2±3.7	77.3±7.5	85.4±6.1
Charbonnier et al. [Charbonnier et al., 2015]†	–	92 [88,95]	94 [84,96]	–	–	–	–	–
Nardelli et al. [Nardelli et al., 2018]†	504.3	94 [91, 96]	95 [93,97]	–	–	–	–	–
Our proposed	1691.0	90.3 [87.7,92.9]	90.9 [87.4,94.6]	90.3±3.5	0.151±0.043	82.4±3.0	85.4±5.3	90.9±3.8
Our proposed + Graph-cuts (a)‡		92.3 [90.1,94.4]	92.7 [89.9,95.9]	92.2±2.9	0.141±0.043	84.2±2.7	87.3±4.8	92.9±3.1
Our proposed + Graph-cuts (b)‡		**97.2 [96.2,98.2]**	**97.5 [96.8,98.6]**	**97.1±1.4**	**0.015±0.008**	**97.2±1.3**	**95.6±1.9**	**96.8±1.5**

* The same auxiliary vessel segmentation task was introduced as the proposed method.

† Results on the same testing set were directly quoted here.

‡ The graph-cuts post-processing was introduced to classify the segmented vessels into arteries and veins. Two possible ways of building vessel graphs are considered: (a) combining both the predicted arteries and veins from the proposed method as vessels; (b) combining both the ground-truth arteries and veins from labels as vessels. Each vessel voxel is regarded as a non-terminal node and is connected to its neighbor vessel nodes, source node (artery), and sink node (vein). For the regional term of graph-cuts, the CNNs' predicted probabilities of being background (p_0), artery (p_1) and vein (p_2) are re-normalized for each vessel voxel: $p'_1 = p_1/(p_1+p_2)$ and $p'_2 = p_2/(p_1+p_2)$, where $p_0+p_1+p_2=1$. The two probabilities are respectively used as the weights of edges between each vessel node and the source node or sink node: $w_{source} = p'_1$ and $w_{sink} = p'_2$. For the boundary term of graph-cuts, the weight of edge between two connected vessel nodes is calculated by the Gaussian kernel of their CT intensity difference: $w_{AB} = \kappa * exp(-(I_A - I_B)^2/\sigma)$, where I_A and I_B are respectively the intensity of vessel node A and node B. The κ balances between the regional term and boundary term. The σ determines how fast the values decay towards zero with increasing intensity dissimilarity. The final cut is obtained by min-cut/max-flow algorithm [Boykov et al., 2001], assigning each vessel node to artery or vein. In the present study, hyper-parameters $\kappa = 8$ and $\sigma = 100$ were found as optimal by means of grid search.

mentation. The key difference between AD and DS is that DS uses segmentation targets as “hard" supervision. In contrast, AD acts as “soft" supervision. It guides preceding layers to pay attention to “hot areas" in latter layers, where voxel-wise supervision of segmentation is not enforced. To evaluate AP, we calculated AP using airway prediction results from 3-D U-Net [Çiçek et al., 2016], V-Net [Milletari et al., 2016], VoxResNet [Chen et al., 2018], AG U-Net [Schlemper et al., 2019], and the proposed airway segmentation method. AP methods with different airway sources are respectively referred to as AP (3-D U-Net [Çiçek et al., 2016]), AP (V-Net [Milletari et al., 2016]), AP (VoxResNet [Chen et al., 2018]), AP (AG U-Net [Schlemper et al., 2019]), and AP (proposed).

Feature Recalibration Table 5.5 shows that under the same threshold $th = 0.5$, all three recalibration modules (cSE [Zhu et al., 2019], PE [Rickmann et al., 2019], and FR) bring performance gains to baseline in BD, TD, and TPR. Specifically, the proposed FR leads to the highest increase of 4.5% in BD, 9.5% in TD, and 5.7% in TPR. Meanwhile, all these modules more or less worsen FPR and DSC. Under the same FPR, results of different methods become closer than those under $th = 0.5$. All methods except baseline are binarized with $th > 0.5$ to reduce FPR. Although FPR of baseline is relaxed to be higher, it only achieved the same BD, 0.3% higher TD, 1.1% higher TPR, and 0.8% lower DSC. The proposed FR boosted performance to a BD of 94.2%, a TD of 87.5%, a TPR of 90.1%, and a DSC of 93.2%. FR outperformed cSE [Zhu et al., 2019] and PE [Rickmann et al., 2019] in BD, TD, TPR, and DSC, which is in line with results under $th = 0.5$.

Table 5.6 reveals that compared with baseline (AP), all recalibration modules increase mean ACC, TPR by over 0.7%, and DSC by over 0.3%. The baseline with FR obtained a mean ACC of 89.4%, a median ACC of 90.2%, a TPR of 89.4%, a FPR of 0.150%, a DSC of 82.0%, a BD of 83.8%, and a TD of 89.9%. Both PE [Rickmann et al., 2019] and FR share similar results, surpassing cSE [Zhu et al., 2019] in all metrics but FPR and DSC.

Attention Distillation In Table 5.5, under the same threshold $th = 0.5$, AD respectively improved baseline in BD, TD, and TPR by 3.3%, 7.0%, and 4.6%. It exceeds DS [Zhu et al., 2017] in BD, TD, and TPR no matter whether FR is introduced or not. Under the same FPR, DS [Zhu et al., 2017] alone performed slightly better than AD. When FR was combined, AD gained a slim advantage over DS [Zhu et al., 2017] in BD and TPR.

In Table 5.6, consistent improvements with 1.7% of mean ACC, 1.1% of median ACC, 1.6% of TPR, 0.003% of FPR, 1.2% of DSC, 1.2% of BD, and 0.9% of TD were observed in AD with regard to baseline + AP (proposed). DS [Zhu et al., 2017] also boosted performance of baseline + AP (proposed) with 0.9% of mean ACC, 0.4% of median ACC, 0.9% of TPR, 0.6% of DSC, 0.3% of BD, and 0.2% of TD. Moreover, AD surpassed DS [Zhu et al., 2017] in all metrics regardless of the presence of FR.

Anatomy Prior Table 5.6 shows that AP (proposed) improved baseline by 1.1% of mean ACC, 2.1% of median ACC, 1.4% of TPR, 2.3% of BD, and 1.8% of TD. AP methods that were calculated using other airway results also performed better than baseline in mean and median ACC, TPR, BD, and TD. Among the baseline + AP methods that were computed from different airway segmentation sources, although the performance variation is small, the highest increments of mean ACC, TPR, and TD were achieved by AP (pro-

Table 5.5: Results of ablation study on pulmonary airway segmentation. The results both under the same binarization threshold and under the same FPR are presented for each method. The FPR is controlled to be the same with 3-D U-Net (under threshold of 0.5) by respectively adjusting the binarization threshold on the probability outputs of each method.
cSE = channel-Squeeze-Excitation, PE = Project-Excitation, FR = Feature Recalibration, AD = Attention Distillation, DS = Deep Supervision

Method	Params ($\times 10^4$)	*th*	BD (%)	TD (%)	TPR (%)	FPR (%)	DSC (%)
Baseline	411.8	0.5	91.6±9.2	81.3±11.5	87.2±8.6	**0.014±0.008**	**93.7±1.7**
+ cSE [Zhu et al., 2019]	422.8		95.1±6.2	88.5±8.3	92.4±5.5	0.033±0.015	92.3±1.9
+ PE [Rickmann et al., 2019]	422.8		95.7±5.1	88.4±7.9	92.3±5.9	0.037±0.019	91.8±2.8
+ FR	423.1		96.1±5.9	**90.8±7.5**	92.9±5.9	0.034±0.016	92.3±2.3
+ AD	411.8		94.9±6.9	88.3±8.2	91.8±6.2	0.029±0.014	92.8±1.4
+ DS [Zhu et al., 2017]	412.4		94.8±7.1	87.6±8.8	91.7±6.2	0.027±0.013	93.1±1.7
+ DS [Zhu et al., 2017] + FR	423.7		96.0±5.3	89.9±7.3	92.7±5.7	0.031±0.013	92.9±1.6
Our proposed	423.1		**96.2±5.8**	90.7±6.9	**93.6±5.0**	0.035±0.014	92.5±2.0
Baseline	411.8	0.001	91.6±10.4	81.6±11.2	88.3±8.6	0.021±0.012	92.9±2.3
+ cSE [Zhu et al., 2019]	422.8	0.83	92.1±7.9	83.0±10.3	89.2±6.9	0.021±0.012	92.8±1.7
+ PE [Rickmann et al., 2019]	422.8	0.71	90.6±8.9	81.5±11.4	86.6±8.8	0.021±0.013	92.1±2.1
+ FR	423.1	0.76	94.2±7.2	**87.5±8.8**	90.1±7.3	0.021±0.012	93.2±1.9
+ AD	411.8	0.66	93.8±7.9	85.7±8.9	89.9±7.1	0.021±0.012	93.3±1.4
+ DS [Zhu et al., 2017]	412.4	0.65	93.9±7.6	85.9±9.4	90.3±6.9	0.021±0.011	93.5±1.6
+ DS [Zhu et al., 2017] + FR	423.7	0.72	94.2±6.8	86.8±8.7	90.4±7.2	0.021±0.011	**93.5±1.5**
Our proposed	423.1	0.77	**94.3±6.6**	86.7±8.5	**90.6±6.7**	**0.021±0.011**	93.5±1.6

Table 5.6: Results of ablation study on pulmonary artery-vein segmentation.
cSE = channel-Squeeze-Excitation, PE = Project-Excitation, FR = Feature Recalibration, AD = Attention Distillation, DS = Deep Supervision, AP = Anatomy Prior

Method	Params ($\times 10^4$)	ACC-mean [95%-CI] (%)	ACC-median [95%-CI] (%)	TPR (%)	FPR (%)	DSC (%)	BD (%)	TD (%)
Baseline	1646.9	87.0 [84.1,89.8]	87.7 [84.4,92.4]	86.7±3.8	**0.130±0.037**	82.1±2.9	81.3±6.1	87.6±4.5
+ AP (proposed)		88.1 [84.5,91.7]	89.8 [84.9,93.1]	88.1±4.8	0.147±0.026	81.4±2.8	83.6±6.1	89.4±4.6
+ AP (3-D U-Net) [Çiçek et al., 2016]		87.9 [84.1,91.7]	90.1 [84.5,92.7]	87.9±5.1	0.151±0.026	81.0±3.0	83.6±5.9	89.2±4.6
+ AP (V-Net) [Milletari et al., 2016]	1647.1	88.0 [84.3,91.6]	90.1 [84.6,92.7]	88.0±4.9	0.149±0.026	81.1±2.8	83.7±5.9	89.3±4.5
+ AP (VoxResNet) [Chen et al., 2018]		88.0 [84.4,91.6]	90.0 [84.6,92.8]	88.0±4.9	0.152±0.027	81.0±2.7	83.7±5.8	89.3±4.5
+ AP (AG U-Net) [Schlemper et al., 2019]		88.0 [84.3,91.7]	90.0 [84.7,92.7]	88.0±4.9	0.148±0.027	81.2±2.8	83.7±5.9	89.3±4.6
+ AP + cSE [Zhu et al., 2019]	1690.8	88.8 [85.6,92.0]	89.1 [85.4,94.2]	88.8±4.3	0.139±0.039	82.4±3.2	83.3±6.6	89.4±4.8
+ AP + PE [Rickmann et al., 2019]	1690.8	89.5 [86.4,92.7]	90.1 [86.4,94.1]	89.5±4.2	0.155±0.049	81.7±3.5	84.3±6.5	90.1±4.7
+ AP + FR	1691.0	89.4 [86.5,92.3]	90.2 [86.1,94.3]	89.4±3.8	0.150±0.044	82.0±3.5	83.8±5.9	89.9±4.2
+ AP + AD	1647.1	89.8 [86.9,92.7]	90.9 [86.6,94.4]	89.7±3.9	0.144±0.043	**82.6±3.9**	84.8±5.4	90.3±4.2
+ AP + DS [Zhu et al., 2017]	1647.2	89.0 [86.0,92.1]	90.2 [87.0,93.9]	89.0±4.1	0.146±0.046	82.0±3.8	83.9±5.9	89.6±4.4
+ AP + FR + DS [Zhu et al., 2017]	1691.2	89.7 [86.9,92.5]	90.5 [86.9,94.2]	89.7±3.7	0.153±0.044	82.0±3.2	84.0±5.7	90.1±4.2
Our proposed	1691.0	**90.3 [87.7,92.9]**	**90.9 [87.4,94.6]**	**90.3±3.5**	0.151±0.043	82.4±3.0	**85.4±5.3**	**90.9±3.8**

posed) whereas AP (3-D U-Net [Çiçek et al., 2016])) yielded the worst results in mean ACC, TPR, DSC, BD, and TD.

Qualitative Results

Results of pulmonary airway and artery-vein segmentation are 3-D rendered in Fig. 5.5, illustrating the robustness of our airway segmentation method on both easy and hard cases. In line with Table 5.2, all methods performed well on extracting thick bronchi. Compared with state-of-the-art methods, more visible tiny branches were reconstructed by the proposed method with high overall segmentation performance maintained. Some false positives were actually true airway branches. Fig. 5.6 reveals that the proposed artery-vein segmentation method successfully extracted multiple arteries and veins. After close inspection of wrong predictions, we noticed that our method may fail to correctly classify some isolated vessel segments. Spatial inconsistency was also observed at terminal ends of arteries and veins (e.g., top and bottom area).

5.4 Discussion

From results in Table 5.2, it is conclusive that our method outperformed the others in airway segmentation, especially distal thin branches. This can be ascribed to the recalibrated features and reinforced attention on hard, tiny, peripheral branches. Although CNNs possess strong fitting ability, it is necessary to suppress redundant, irrelevant features and strengthen task-related ones. Under the same threshold, two reasons are responsible for our relatively inferior FPR and DSC: 1) Our model successfully detected some true thin airways that were too indistinct to be annotated properly by experts. After careful examination of segmented airways and retrospective evaluation of labels, some branches were unintentionally neglected due to annotation difficulty. When calculating the evaluation metrics, these actually existing branches were counted as false positives and caused higher FPR with lower DSC. 2) A little leakage was produced at bifurcations when the contrast between airway lumen and wall was fairly low. In this situation, the proposed method was inclined to predict voxels as airway while other methods were relatively conservative. Some leakage regions do resemble airway in appearance, where tubular parenchyma with high-intensity circular boundary and low-intensity hollow was observed. Under the same FPR, the superior BD, TPR, and DSC of the proposed method demonstrated its sensitivity and robustness on extracting small airways. Since our threshold th was increased to 0.77, airway predictions in low confidence were excluded and false positives were suppressed. The overall performance indicator DSC was consequently improved. By considering results both under the same threshold and FPR, we believe the superiority of the proposed method is well revealed.

An additional evaluation on the EXACT'09 testing set verified that under different FPR levels, our method did extract much more branches than previous methods. Besides, there exists a gap between results on our testing set and results on EXACT'09 testing set. The reasons behind are 3-fold: 1) difference in quality between our labels and EXACT'09 labels (e.g., inter-observer variation in labeling the 5-th and 6-th order bronchi); 2) difference in implementation of metrics calculation (e.g., centerline extraction); 3) difference

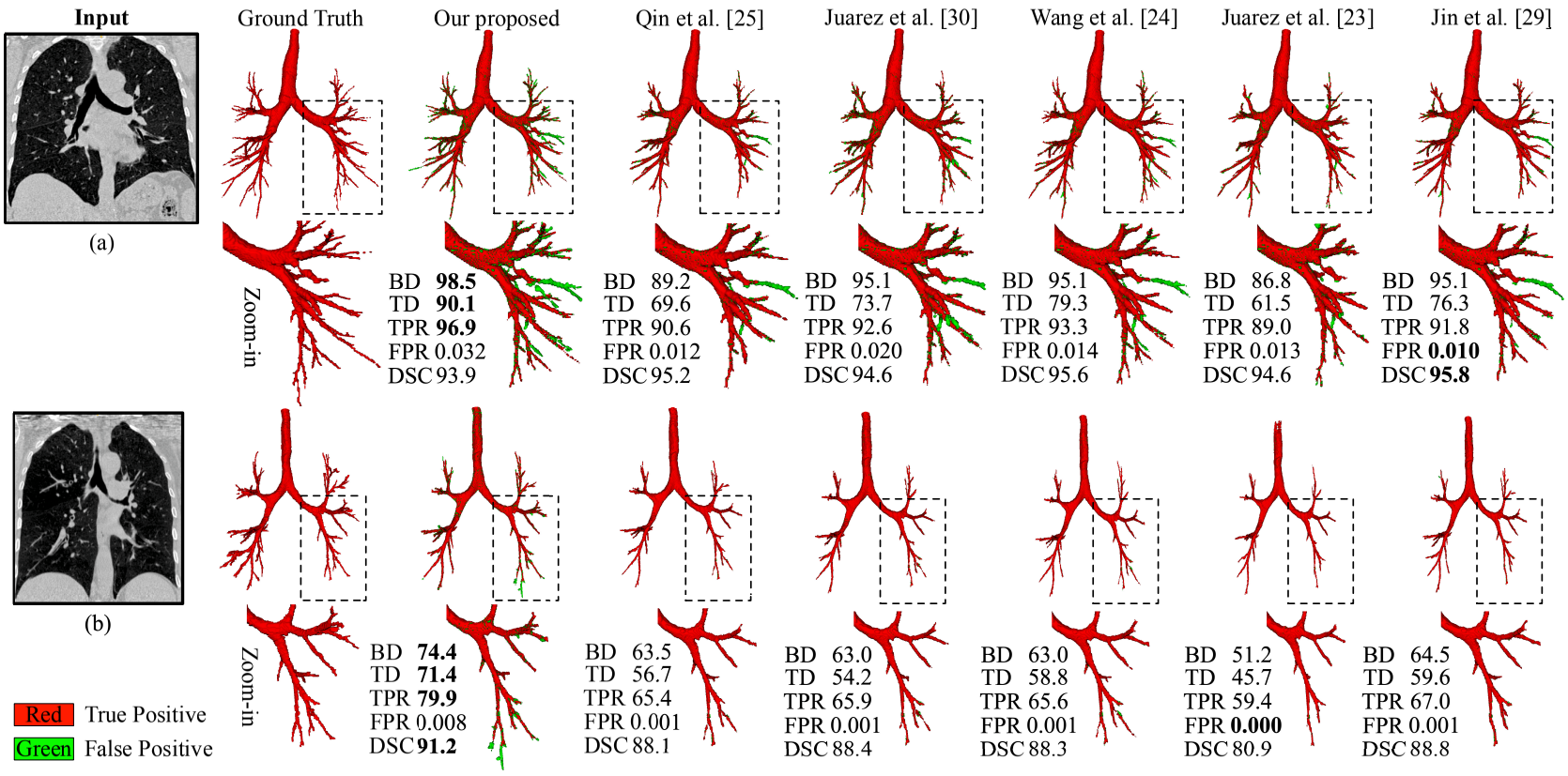


Figure 5.5: Rendering of pulmonary airway segmentation results on (a) easy and (b) hard testing cases. Best viewed magnified.

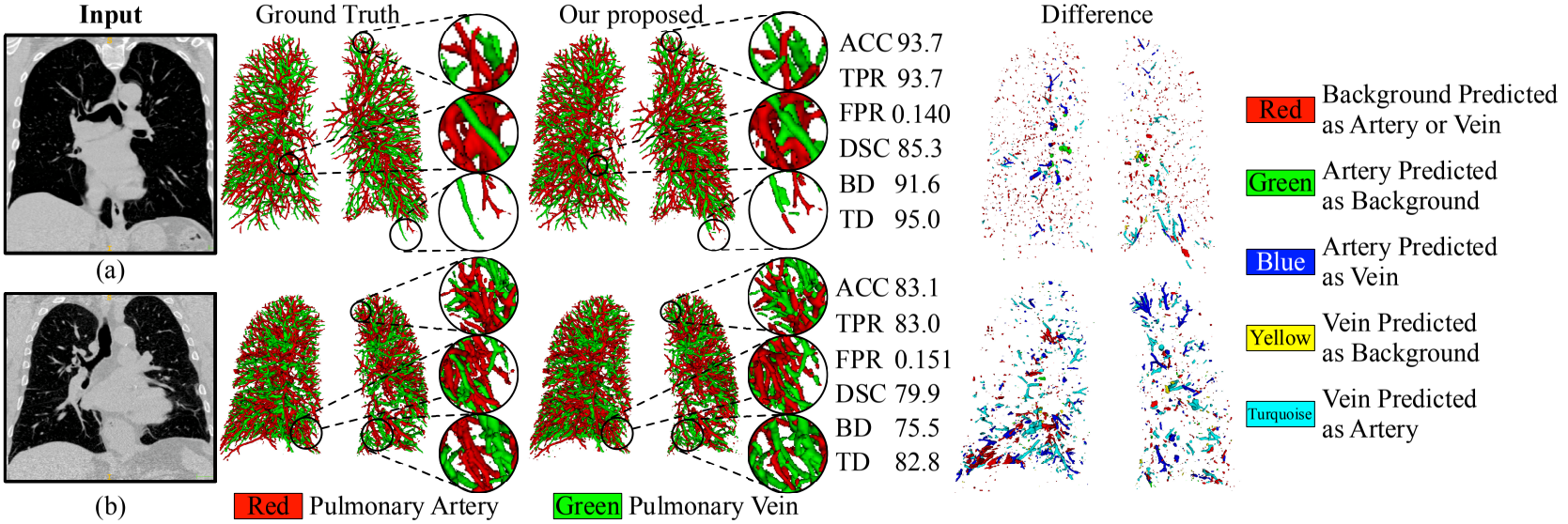


Figure 5.6: Rendering of pulmonary artery-vein segmentation results on (a) easy and (b) hard testing cases. Left: Wrongly segmented arteries and veins are zoomed in for better visual inspection. Right: Difference between prediction and label is categorized into 5 types.

in data distribution between the training set and EXACT'09 testing set (e.g., multi-center CT scans, dissimilar lung diseases).

Table 5.4 shows that state-of-the-art CNNs may not be effective to segment arteries and veins if no "customization" was involved for this task. It is undeniable that two state-of-the-art methods [Charbonnier et al., 2015,Nardelli et al., 2018] performed well in this task. In [Charbonnier et al., 2015], graph representation was computed and the label of artery or vein was assigned to the entire linked sub-trees. It eliminated the possibility of label inconsistency for each branch and therefore performed better than ours in terms of ACC. The two-stage CNNs-based classification method [Nardelli et al., 2018] achieved the highest ACC. It highly relied on graph-cuts post-processing to refine the predictions of CNNs, where over 10% performance gains were brought by graph-cuts. In contrast, our method is end-to-end and segmentation was directly fulfilled by CNNs. No vessel segmentation beforehand or post-processing afterwards was designed in the pipeline, which avoids error accumulation. In addition, the proposed CNNs-based method is effective for both pulmonary airway and artery-vein segmentation, which is the first of its kind in literature.

Furthermore, post-processing experiments were performed to assess the performance gains brought by the graph-cuts [Boykov et al., 2001] method. Due to the fundamental differences between the proposed method and Ref. [Nardelli et al., 2018], the same graph-cuts technique in [Nardelli et al., 2018] is not directly applicable. Modifications were made as described in Table 5.4. The Graph-cuts (a) and (b) differ in vessel graph construction: (a) combining both the predicted arteries and veins from our CNNs as vessels; (b) combining both the ground-truth arteries and veins from labels as vessels. We observe that both Graph-cuts (a) and (b) improved artery-vein separation. Given segmented vessels, the connectivity of each voxel node with respect to its neighbor nodes is explicitly modeled in each graph. After graph construction, adjacent voxel nodes that share similar intensity tend to be in the same category. Consequently, spatial inconsistency of label assignment could be reduced for connected vessels. The comparison between Graph-cuts (a) and (b) reveals that the performance of artery-vein classification is positively associated with that of vessel segmentation. Results of Graph-cuts (b) demonstrate our superior artery-vein classification competence given ground-truth vessel branches. It is noted that the capability of graph-cuts may not be fully exploited under the current segmentation framework. Elaborate design on the incorporation of graph-cuts is needed but beyond the scope of the present study.

To study the effectiveness of the proposed FR, we conducted extensive ablation experiments. As shown in Table 5.5, cSE [Zhu et al., 2019], PE [Rickmann et al., 2019], and FR all contributed to model's capacity of peripheral bronchiole extraction. Interestingly, these modules more or less worsen FPR and DSC under the same threshold. We believe the recalibration mechanism by element-wise weighting prefers airway voxels and the model tends to identify as many branches as possible, causing an increase of FPR and a decrease of DSC. Under the same FPR, the advantage of FR over baseline and other recalibration modules is clearly revealed in BD, TD, and TPR, substantiating the importance of reasonably integrating spatial knowledge for channel-wise recalibration. Table 5.6 reveals that all recalibration modules improved sensitivity of baseline (AP) to arteries and veins. The similar performance of PE [Rickmann et al., 2019] and FR might be explained by the spatial distribution of artery-vein targets. Arteries and veins spread all over the

lung and their difference in position may not be highly informative for artery-vein separation.

For completeness, we also conducted experiments to investigate the efficacy of the proposed AD. In Table 5.5, both AD and DS [Zhu et al., 2017] improved baseline, confirming our hypothesis that it is difficult for CNNs to learn effective representation of small, thin targets only with supervision from the last output layer. In Table 5.6, AD outperformed baseline (AP) in all metrics. Such improved sensitivity was attributed to the mechanism that features of shallower layers learned to focus on fine-grained details in features of deeper layers. To intuitively assess AD, attention maps from decoder 1–4 are visualized in Fig. 5.7. After distillation, activated regions become more distinct and the target tubules are enhanced. The improved attention on airway, vessel, and lung border explains that our model comprehended more context and therefore achieved higher sensitivity to intricate tubules. Another interesting finding is that although the last attention map is not refined in distillation, it still gets polished up because better representation learned at previous layers in turn affects late-layer features. Moreover, it is noted that AD surpassed DS [Zhu et al., 2017] if their performance in both airway and artery-vein segmentation was comprehensively considered. It may not be optimum for earlier layers to be directly supervised by scattered and sparse targets. As shown in Fig. 5.7, not only segmentation targets but also lung contours are enhanced. DS [Zhu et al., 2017] may hamper shallow layers from learning rich context for later comprehension.

Ablation study on AP in Table 5.6 suggested that: 1) The lung context and distance transform maps do contribute to recognizing arteries and veins. 2) The accuracy of segmented airways was positively associated with the artery-vein segmentation performance. The more complete and precise the airway is, the more informative the calculated anatomy prior is. 3) The proposed model performed robustly to AP using different airway segmentation methods. No drastic decline was observed due to poor airway prediction results. Note that the loss of DSC by introducing AP was later offset by FR and AD.

From visual analysis, we find that some true airways were neglected unintentionally in labels (see Fig. 5.5). Such mistake was due to the weak intensity contrast and limitation of annotating 3-D objects in 2-D planes. In Fig. 5.6, the reason why wrong classification was observed on isolated vessel segments and terminal ends might be explained as follows. In these regions, CNNs may not capture enough context knowledge for proper inference and no label propagation was enforced to make consistent decisions on the same vessel segment.

We further provide analysis for the artery-vein task. The confusion matrix in Fig. 5.8(a) shows that the proposed method performed worst on recognizing veins. Lacking anatomic relationship such as the proximity of arteries to airways, veins are a bit more difficult to segment. By dividing errors into 5 types (see Fig. 5.8(b)), we find that the Type 1 error of predicting background as artery or vein makes up a large proportion of all errors. In accord with Fig. 5.6 and Fig. 5.8(c), Type 1 error voxels distribute all over the lung. Besides, most errors of Type 1, 2 and 4 appear at vessel boundaries which are ambiguous for accurate and unified definition. In that case, the proposed method behaved a bit aggressively due to the reinforced learning of tubular targets.

Although the proposed method solved both pulmonary airway and artery-vein segmentation with higher sensitivity to peripheral tubular branches, there exist some limi-

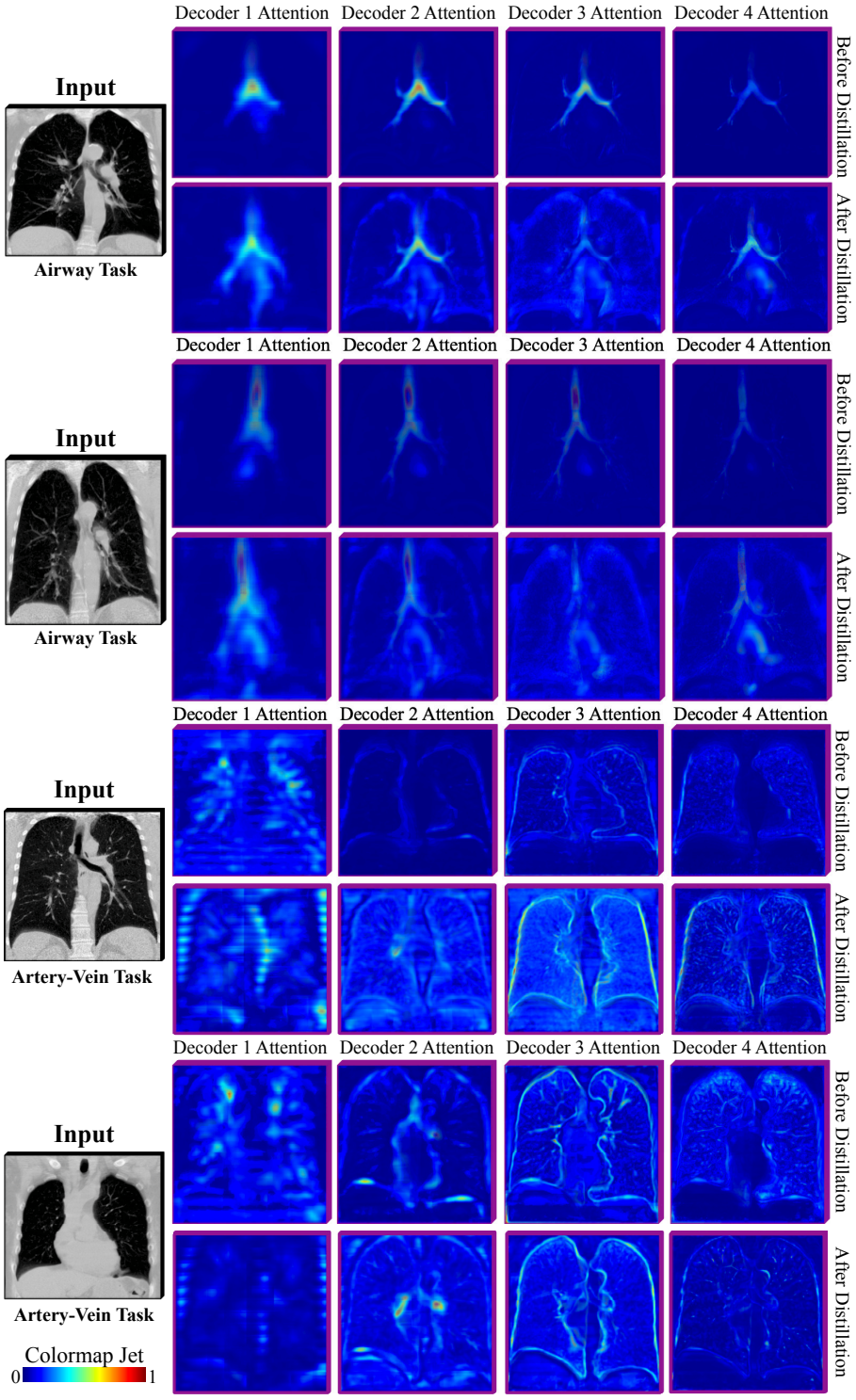


Figure 5.7: Pseudo-color rendering of attention maps (decoder 1–4) before and after distillation process. These maps are min-max scaled and rendered with Jet colormap. Best viewed magnified.

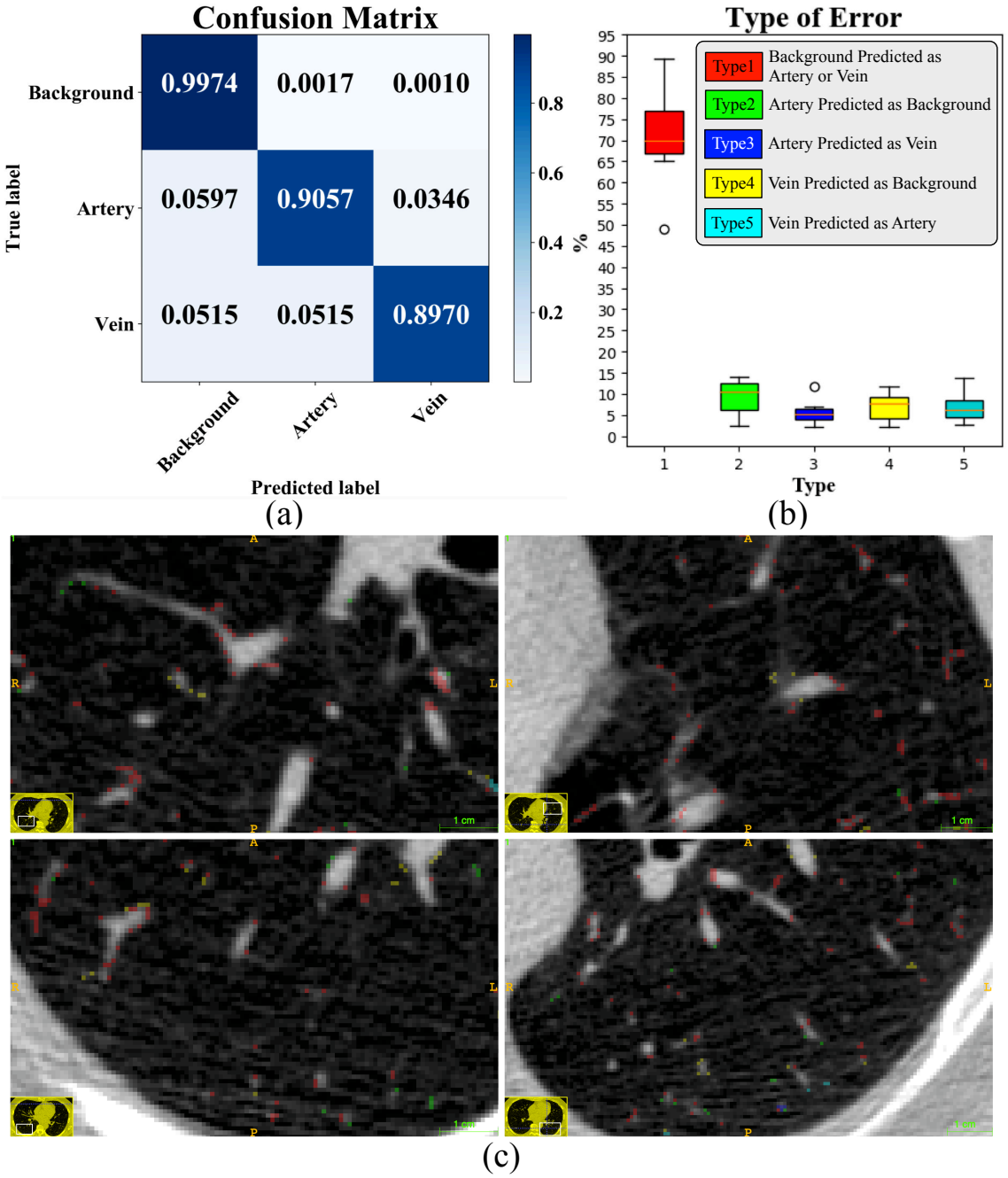


Figure 5.8: Analysis of multi-class artery-vein segmentation results. (a) Normalized confusion matrix. (b) Percentage of different types of errors. (c) Typical examples of wrong predictions. Best viewed magnified.

tations. First, for the current artery-vein segmentation task, only vessels inside lungs are considered as targets. The main pulmonary artery and vein vessels, including the trunk, are too difficult for human observers to delineate their boundaries in non-contrast CT. To have a broader application, CT pulmonary angiogram would be investigated in the future for segmentation of vessels in the lung hilum. Second, refinement of airway annotations might be carefully conducted to incorporate fuzzy and unclear airway branches. One efficient way of correction would be to first apply the proposed method on CT and then review both the predicted and manually labeled branches. Third, the proposed method did not enforce label compatibility in artery-vein segmentation. Spatial inconsistency was therefore observed in distal vessels. Future work includes the adoption of strategies like label propagation and majority voting. It could explicitly remove conflicting predictions that mainly caused Type 3 and 5 errors. Finally, for both airway and artery-vein tasks, more diverse clinical data might be collected to improve and examine the generalizability of our method.

5.5 Conclusion

This chapter presented a tubule-sensitive method for both pulmonary airway and artery-vein segmentation. It utilizes CNNs and requires no post-processing. With the proposed spatial-aware feature recalibration module and the gradually reinforced attention distillation module, feature learning of our CNNs becomes more effective and relevant to target tubule perception. The incorporated anatomy prior is also beneficial for artery-vein separation. Extensive experiments showed that our method detected much more bronchioles, arterioles, and venules while maintaining competitive overall segmentation performance, which corroborates its superior sensitivity over state-of-the-art methods and the validity of its constituents.

Chapter 6

General Conclusions and Perspectives

Conclusions

In this thesis, we present the original work of developing deep learning methods for chromosome image classification and pulmonary CT image segmentation. Before presenting specific achievements, introduction about the relevant biomedical context and technical background is presented in Chapter 1, including chromosome karyotyping, pulmonary CT image segmentation, and trends in development of deep learning. The main contributions in Chapter 2 — Chapter 5 are summarized as follows:

- In Chapter 2, we proposed a chromosome classification method called VarifocalNet. It is a three-stage CNN-based method. The first stage effectively learns global and local features through the G-Net and the L-Net, respectively. The second stage robustly differentiates chromosomes into various types and polarities via two MLP classifiers. In the third stage, a dispatch post-processing strategy was employed to assign each chromosome to a type based on its predicted probabilities. Extensive experimental results demonstrate that our approach outperforms state-of-the-art methods, corroborating its high accuracy and generalizability. Furthermore, its practical value has been validated in Xiangya Hospital of Central South University.

- In Chapter 3, we proposed a two-part CNN-based framework for pulmonary nodule segmentation. In the first part, adversarial networks are employed to synthesize nodule samples that do not exist in the collected dataset. It targets at building a more diverse and balanced dataset for the subsequent model training. In the second part, multiple feature maps are incorporated as inputs into the 3D CNN model. With residual learning, the segmentation model trained on the extended dataset enjoys a high level of generality. Results on LIDC-IDRI dataset demonstrate that our method achieved more accurate nodule segmentation compared with state-of-the-art methods, which suggests its potential value for clinical applications.

- In Chapter 4, we proposed the AirwayNet and AirwayNet-SE for pulmonary airway segmentation. These two methods explicitly learn voxel connectivity to perceive airway's inherent structure. By connectivity modeling, the segmentation task is transformed into 26 tasks of connectivity prediction. Besides, the AirwayNet-SE extends the AirwayNet by fusing features of two context scales for recognition of both large and small airways. Experimental results proved that our approach was effective at overcoming the distribution difference between large and small airways. The airway annotations were also released to boost research on airway extraction using supervised learning methods.

- In Chapter 5, we presented a tubule-sensitive method using CNNs for both pulmonary airway and artery-vein segmentation. With the proposed spatial-aware feature recalibration module and the gradually reinforced attention distillation module, feature learning of our CNNs becomes more effective and relevant to target tubule perception. The incorporated anatomy prior is also beneficial for artery-vein separation. Extensive experiments showed that our method detected much more bronchioles, arterioles, and venules while maintaining competitive overall segmentation performance, which corroborates its superior sensitivity over state-of-the-art methods and the validity of its constituents.

To conclude, we investigated several deep learning methods for classification and segmentation of chromosome and pulmonary images. The proposed methods manifest great clinical potentials for improving the efficiency of karyotyping analysis, pulmonary structure measurement as well as lesion diagnosis. Regarding the limitations found in the present study, future work is discussed in the following section.

Perspectives

Future work is divided into two parts: 1) One is to solve new tasks for chromosome and pulmonary image analysis; 2) The other is to further improve performance of classification and segmentation by applying new deep learning techniques. Concerning the first part, future work includes:

- For chromosome image analysis, it is still difficult and time-consuming for manual detection and segmentation of chromosomes. To fully automate the entire karyotyping process, automatic detection and segmentation methods are yet to be developed. We believe the current instance segmentation framework (Mask R-CNN [He et al., 2017]) might be a good starting point. Modifications on this backbone are needed to adapt to chromosome microscopic images.
- For pulmonary image analysis, it would be comprehensive to design methods for segmentation of pulmonary lobes, pulmonary segments, main pulmonary arteries and veins, and pulmonary trunk. To fulfill these tasks, more data annotations are needed for supervised learning. Besides, other image modalities such as computed tomography pulmonary angiogram (CTPA) are prioritized for full artery-vein segmentation and analysis.

For the second part, future work includes:

- For chromosome classification, attention mechanism might be introduced for discrimination between short, curve chromosomes such as sex chromosome Y and chromosome Nos. 15, 21, 22. In addition, methods such as spatial transform and deformation may benefit the classification model in handling bending chromosomes.
- For pulmonary nodule segmentation, the realness of generated samples can be improved by exploring newly developed generative adversarial networks. To enrich the context of synthetic samples, the semantic labels used in the proposed GAN might be labeled with more structures such as airway wall, airway lumen, and ribs.
- For pulmonary airway segmentation, graph representation and graph neural networks (GNNs) are expected to explicitly model the relationship between branches. Graph-based methods may improve segmentation for patients with severe pathological changes, where airway structures appear discontinuous in CT.
- For pulmonary artery-vein segmentation, there still remains a challenge to remove spatial inconsistency within one vessel segment. An adversarial training strategy might be useful when a discriminator is trained to identify the misclassified arteries and veins. The adoption of GNNs is also a solution to improving the graph integrity of arteries and veins.

List of Publications

Journal Papers:

- **Qin, Y.**, Zheng, H., Gu, Y., Huang, X., Yang, J., Wang, L., Yao, F., Zhu, Y.M. and Yang, G.Z., 2021. Learning Tubule-Sensitive CNNs for Pulmonary Airway and Artery-Vein Segmentation in CT. **IEEE Transactions on Medical Imaging**, 40(6), pp.1603-1617.
- **Qin, Y.**, Wen, J., Zheng, H., Huang, X., Yang, J., Song, N., Zhu, Y.M., Wu, L. and Yang, G.Z., 2019. Varifocal-Net: A Chromosome Classification Approach Using Deep Convolutional Networks. **IEEE Transactions on Medical Imaging**, 38(11), pp.2569-2581.
- **Qin, Y.**, Zheng, H., Huang, X., Yang, J. and Zhu, Y.M., 2019. Pulmonary nodule segmentation with CT sample synthesis using adversarial networks. **Medical Physics**, 46(3), pp.1218-1229.
- Zheng, H., **Qin, Y.**, Gu, Y., Xie, F., Yang, J., Sun, J. and Yang, G.Z., 2021. Alleviating Class-wise Gradient Imbalance for Pulmonary Airway Segmentation. **IEEE Transactions on Medical Imaging**, Early Access, pp. 1-17.
- Zheng, H., Qian, L., **Qin, Y.**, Gu, Y. and Yang, J., 2020. Improving the slice interaction of 2.5D CNN for automatic pancreas segmentation. **Medical Physics**, 47(11), pp.5543-5554.

Conference Papers:

- **Qin, Y.**, Zheng, H., Gu, Y., Huang, X., Yang, J., Wang, L. and Zhu, Y.M., 2020. Learning Bronchiole-Sensitive Airway Segmentation CNNs by Feature Recalibration and Attention Distillation. International Conference on Medical Image Computing and Computer Assisted Intervention – **MICCAI 2020**, pp.221-231.
- **Qin, Y.**, Gu, Y., Zheng, H., Chen, M., Yang, J. and Zhu, Y.M., 2020. AirwayNet-SE: A Simple-Yet-Effective Approach to Improve Airway Segmentation Using Context Scale Fusion. 2020 IEEE 17th International Symposium on Biomedical Imaging – **ISBI 2020**, pp.809-813.

- **Qin, Y.**, Chen, M., Zheng, H., Gu, Y., Shen, M., Yang, J., Huang, X., Zhu, Y.M. and Yang, G.Z., 2019. AirwayNet: A Voxel-Connectivity Aware Approach for Accurate Airway Segmentation Using Convolutional Neural Networks. International Conference on Medical Image Computing and Computer Assisted Intervention – **MICCAI 2019**, pp.212-220.
- **Qin, Y.**, Zheng, H., Zhu, Y.M. and Yang, J., 2018. Simultaneous Accurate Detection of Pulmonary Nodules and False Positive Reduction Using 3D CNNs. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing – **ICASSP 2018**, pp.1005-1009.
- Zhang, H., Gu, Y., **Qin, Y.**, Yao, F. and Yang, G.Z., 2020. Learning with Sure Data for Nodule-Level Lung Cancer Prediction. International Conference on Medical Image Computing and Computer Assisted Intervention – **MICCAI 2020**, pp.570-578.
- Zheng, H., Zhuang, Z., **Qin, Y.**, Gu, Y., Yang, J. and Yang, G.Z., 2020. Weakly Supervised Deep Learning for Breast Cancer Segmentation with Coarse Annotations. International Conference on Medical Image Computing and Computer Assisted Intervention – **MICCAI 2020**, pp.450-459.
- Zheng, H., Gu, Y., **Qin, Y.**, Huang, X., Yang, J. and Yang, G.Z., 2018. Small Lesion Classification in Dynamic Contrast Enhancement MRI for Breast Cancer Early Detection. International Conference on Medical Image Computing and Computer Assisted Intervention – **MICCAI 2018**, pp.876-884.

Bibliography

[Abid and Hamami, 2018] Abid, F. and Hamami, L. (2018). A survey of neural network based automated systems for human chromosome classification. *Artificial Intelligence Review*, 49(1):41–56.

[Agam et al., 2005] Agam, G., Armato, S. G., and Wu, C. (2005). Vessel tree reconstruction in thoracic ct scans with application to nodule detection. *IEEE Transactions on Medical Imaging*, 24(4):486–499.

[Aldred et al., 2010] Aldred, M. A., Comhair, S. A., Varella-Garcia, M., Asosingh, K., Xu, W., Noon, G. P., Thistlethwaite, P. A., Tuder, R. M., Erzurum, S. C., Geraci, M. W., et al. (2010). Somatic chromosome abnormalities in the lungs of patients with pulmonary arterial hypertension. *American Journal of Respiratory and Critical Care Medicine*, 182(9):1153–1160.

[Alilou et al., 2017] Alilou, M., Beig, N., Orooji, M., Rajiah, P., Velcheti, V., Rakshit, S., Reddy, N., Yang, M., Jacono, F., Gilkeson, R. C., et al. (2017). An integrated segmentation and shape-based classification scheme for distinguishing adenocarcinomas from granulomas on lung ct. *Medical Physics*, 44(7):3556–3569.

[Arachchige et al., 2013] Arachchige, A. S., Samarabandu, J., Knoll, J. H., and Rogan, P. K. (2013). Intensity integrated laplacian-based thickness measurement for detecting human metaphase chromosome centromere location. *IEEE Transactions on Biomedical Engineering*, 60(7):2005–2013.

[Armato et al., 2011] Armato, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., et al. (2011). The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans. *Medical Physics*, 38(2):915–931.

[Association et al., 2001] Association, W. M. et al. (2001). World medical association declaration of helsinki. ethical principles for medical research involving human subjects. *Bulletin of the World Health Organization*, 79(4):373.

[Awad et al., 2012] Awad, J., Owrangi, A., Villemaire, L., O'Riordan, E., Parraga, G., and Fenster, A. (2012). Three-dimensional lung tumor segmentation from x-ray computed tomography using sparse field active models. *Medical Physics*, 39(2):851–865.

[Aylward and Bullitt, 2002] Aylward, S. R. and Bullitt, E. (2002). Initialization, noise, singularities, and scale in height ridge traversal for tubular object centerline extraction. *IEEE Transactions on Medical Imaging*, 21(2):61–75.

[Balsara and Testa, 2002] Balsara, B. R. and Testa, J. R. (2002). Chromosomal imbalances in human lung cancer. *Oncogene*, 21(45):6877–6883.

[Berend et al., 1979] Berend, N., Woolcock, A., and Marlin, G. (1979). Relationship between bronchial and arterial diameters in normal human lungs. *Thorax*, 34(3):354–358.

[Beutel et al., 2000] Beutel, J., Kundel, H. L., and Van Metter, R. L. (2000). *Handbook of Medical Imaging*, volume 1. SPIE Press.

[Biyani et al., 2005] Biyani, P., Wu, X., and Sinha, A. (2005). Joint classification and pairing of human chromosomes. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2(2):102–109.

[Boykov et al., 2001] Boykov, Y., Veksler, O., and Zabih, R. (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239.

[Buelow et al., 2005] Buelow, T., Wiemker, R., Blaffert, T., Lorenz, C., and Renisch, S. (2005). Automatic extraction of the pulmonary artery tree from multi-slice CT data. In *Medical Imaging 2005: Physiology, Function, and Structure from Medical Images*, volume 5746, pages 730–740. International Society for Optics and Photonics.

[Bulow et al., 2004] Bulow, T., Lorenz, C., and Renisch, S. (2004). A general framework for tree segmentation and reconstruction from medical volume data. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 533–540. Springer.

[Buyssens et al., 2012] Buyssens, P., Elmoataz, A., and Lézoray, O. (2012). Multiscale convolutional neural networks for vision–based classification of cells. In *Asian Conference on Computer Vision*, pages 342–352. Springer.

[Canny, 1986] Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698.

[Cao et al., 2020] Cao, X., Lan, F., Liu, C.-M., Lam, T.-W., and Luo, R. (2020). Chromseg: Two-stage framework for overlapping chromosome segmentation and reconstruction. In *2020 IEEE International Conference on Bioinformatics and Biomedicine*, pages 2335–2342. IEEE.

[Cartin-Ceba et al., 2013] Cartin-Ceba, R., Swanson, K. L., and Krowka, M. J. (2013). Pulmonary arteriovenous malformations. *Chest*, 144(3):1033–1044.

[Caruana, 1997] Caruana, R. (1997). Multitask learning. *Machine Learning*, 28(1):41–75.

[Charbonnier et al., 2015] Charbonnier, J.-P., Brink, M., Ciompi, F., Scholten, E. T., Schaefer-Prokop, C. M., and Van Rikxoort, E. M. (2015). Automatic pulmonary artery-vein separation and classification in computed tomography using tree partitioning and peripheral vessel matching. *IEEE Transactions on Medical Imaging*, 35(3):882–892.

[Charbonnier et al., 2017] Charbonnier, J.-P., Van Rikxoort, E. M., Setio, A. A., Schaefer-Prokop, C. M., van Ginneken, B., and Ciompi, F. (2017). Improving airway segmentation in computed tomography using leak detection with convolutional networks. *Medical Image Analysis*, 36:52–60.

[Chen et al., 2018] Chen, H., Dou, Q., Yu, L., Qin, J., and Heng, P.-A. (2018). VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage*, 170:446–455.

[Chollet et al., 2015] Chollet, F. et al. (2015). Keras. `https://keras.io`.

[Çiçek et al., 2016] Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., and Ronneberger, O. (2016). 3D U-Net: learning dense volumetric segmentation from sparse annotation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 424–432.

[Ciompi et al., 2017] Ciompi, F., Chung, K., Van Riel, S. J., Setio, A. A. A., Gerke, P. K., Jacobs, C., Scholten, E. T., Schaefer-Prokop, C., Wille, M. M., Marchiano, A., et al. (2017). Towards automatic pulmonary nodule management in lung cancer screening with deep learning. *Scientific Reports*, 7(1):1–11.

[Cochran et al., 2001] Cochran, S. T., Bomyea, K., and Sayre, J. W. (2001). Trends in adverse events after iv administration of contrast media. *American Journal of Roentgenology*, 176(6):1385–1388.

[CytoVision, 2021] CytoVision (2021). Cytovision automated cytogenetics platform.

[de Hoop et al., 2012] de Hoop, B., van Ginneken, B., Gietema, H., and Prokop, M. (2012). Pulmonary perifissural nodules on CT scans: rapid growth is not a predictor of malignancy. *Radiology*, 265(2):611–616.

[Dehmeshki et al., 2008] Dehmeshki, J., Amin, H., Valdivieso, M., and Ye, X. (2008). Segmentation of pulmonary nodules in thoracic CT scans: a region growing approach. *IEEE Transactions on Medical Imaging*, 27(4):467–480.

[Deng et al., 2009] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. IEEE.

[Diciotti et al., 2011] Diciotti, S., Lombardo, S., Falchini, M., Picozzi, G., and Mascalchi, M. (2011). Automated segmentation refinement of small lung nodules in CT scans by local shape analysis. *IEEE Transactions on Biomedical Engineering*, 58(12):3418–3428.

[Dijkstra, 1959] Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271.

[Dou et al., 2017] Dou, Q., Chen, H., Yu, L., Qin, J., and Heng, P.-A. (2017). Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection. *IEEE Transactions on Biomedical Engineering*, 64(7):1558–1567.

[Estépar et al., 2013] Estépar, R. S. J., Kinney, G. L., Black-Shinn, J. L., Bowler, R. P., Kindlmann, G. L., Ross, J. C., Kikinis, R., Han, M. K., Come, C. E., Diaz, A. A., et al. (2013). Computed tomographic measures of pulmonary vascular morphology in smokers and their clinical implications. *American Journal of Respiratory and Critical Care Medicine*, 188(2):231–239.

[Fan et al., 2000] Fan, Y.-S., Siu, V. M., Jung, J. H., and Xu, J. (2000). Sensitivity of multiple color spectral karyotyping in detecting small interchromosomal rearrangements. *Genetic Testing*, 4(1):9–14.

[Farag et al., 2013] Farag, A. A., Abd El Munim, H. E., Graham, J. H., and Farag, A. A. (2013). A novel approach for lung nodules segmentation in chest CT using level sets. *IEEE Transactions on Image Processing*, 22(12):5202–5213.

[Farhangi et al., 2017] Farhangi, M. M., Frigui, H., Seow, A., and Amini, A. A. (2017). 3-D active contour segmentation based on sparse linear combination of training shapes (scots). *IEEE Transactions on Medical Imaging*, 36(11):2239–2249.

[Fetita et al., 2009] Fetita, C., Brillet, P.-Y., and Preteux, F. J. (2009). Morpho-geometrical approach for 3d segmentation of pulmonary vascular tree in multi-slice CT. In *Medical Imaging 2009: Image Processing*, volume 7259, page 72594F. International Society for Optics and Photonics.

[Fetita et al., 2004] Fetita, C. I., Prêteux, F., Beigelman-Aubry, C., and Grenier, P. (2004). Pulmonary airways: 3-D reconstruction from multislice CT and clinical investigation. *IEEE Transactions on Medical Imaging*, 23(11):1353–1364.

[Frangi et al., 1998] Frangi, A. F., Niessen, W. J., Vincken, K. L., and Viergever, M. A. (1998). Multiscale vessel enhancement filtering. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 130–137. Springer.

[Fu et al., 2017] Fu, J., Zheng, H., and Mei, T. (2017). Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4438–4446.

[Gadhia et al., 2014] Gadhia, P. K., Vaniawala, S. N., et al. (2014). A rare double aneuploidy with 48, xxy,+ 21 karyotype in down syndrome from gujarat, india. *International Journal of Molecular Medical Science*, 4(4).

[Gajendran and Rodríguez, 2004] Gajendran, V. and Rodríguez, J. J. (2004). Chromosome counting via digital image analysis. In *IEEE International Conference on Image Processing*, volume 5, pages 2929–2932. IEEE.

[Gao et al., 2012] Gao, Z., Grout, R. W., Holtze, C., Hoffman, E. A., and Saha, P. (2012). A new paradigm of interactive artery/vein separation in noncontrast pulmonary CT imaging using multiscale topomorphologic opening. *IEEE Transactions on Biomedical Engineering*, 59(11):3016–3027.

[Girshick, 2015] Girshick, R. (2015). Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1440–1448.

[Godinez et al., 2017] Godinez, W. J., Hossain, I., Lazic, S. E., Davies, J. W., and Zhang, X. (2017). A multi-scale convolutional neural network for phenotyping high-content cellular images. *Bioinformatics*, 33(13):2010–2019.

[Godinez et al., 2018] Godinez, W. J., Hossain, I., and Zhang, X. (2018). Unsupervised phenotypic analysis of cellular images with multi-scale convolutional neural networks. *bioRxiv*, page 361410.

[Goodfellow et al., 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*, volume 1. Cambridge: MIT press.

[Goodfellow et al., 2014] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Proceedings of the Advances in Neural Information Processing Systems*, pages 2672–2680.

[Goodman et al., 2006] Goodman, L. R., Gulsun, M., Washington, L., Nagy, P. G., and Piacsek, K. L. (2006). Inherent variability of CT lung nodule measurements in vivo using semiautomated volumetric measurements. *American Journal of Roentgenology*, 186(4):989–994.

[Gozzetti and Le Beau, 2000] Gozzetti, A. and Le Beau, M. M. (2000). Fluorescence in situ hybridization: uses and limitations. *Seminars in Hematology*, 37(4):320–333.

[Graham et al., 2010] Graham, M. W., Gibbs, J. D., Cornish, D. C., and Higgins, W. E. (2010). Robust 3-D airway tree segmentation for image-guided peripheral bronchoscopy. *IEEE Transactions on Medical Imaging*, 29(4):982–997.

[Grigorova et al., 2005] Grigorova, M., Lyman, R. C., Caldas, C., and Edwards, P. A. (2005). Chromosome abnormalities in 10 lung cancer cell lines of the nci-h series analyzed with spectral karyotyping. *Cancer Genetics and Cytogenetics*, 162(1):1–9.

[Gu et al., 2013] Gu, Y., Kumar, V., Hall, L. O., Goldgof, D. B., Li, C.-Y., Korn, R., Bendtsen, C., Velazquez, E. R., Dekker, A., Aerts, H., et al. (2013). Automated delineation of lung tumors from CT images using a single click ensemble segmentation approach. *Pattern Recognition*, 46(3):692–702.

[Guibas et al., 2017] Guibas, J. T., Virdi, T. S., and Li, P. S. (2017). Synthetic medical images from dual generative adversarial networks. *arXiv preprint arXiv:1709.01872*.

[Gupta et al., 2017] Gupta, G., Yadav, M., Sharma, M., Vig, L., et al. (2017). Siamese networks for chromosome classification. In *2017 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pages 72–81. IEEE.

[Halpin et al., 2021] Halpin, D. M., Criner, G. J., Papi, A., Singh, D., Anzueto, A., Martinez, F. J., Agusti, A. A., and Vogelmeier, C. F. (2021). Global initiative for the diagnosis, management, and prevention of chronic obstructive lung disease. the 2020 gold science committee report on covid-19 and chronic obstructive pulmonary disease. *American Journal of Respiratory and Critical Care Medicine*, 203(1):24–36.

[He et al., 2017] He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2961–2969.

[He et al., 2015] He, K., Zhang, X., Ren, S., and Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1026–1034.

[He et al., 2016a] He, K., Zhang, X., Ren, S., and Sun, J. (2016a). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778.

[He et al., 2016b] He, K., Zhang, X., Ren, S., and Sun, J. (2016b). Identity mappings in deep residual networks. In *European Conference on Computer Vision*, pages 630–645. Springer.

[HiBand, 2021] HiBand (2021). Hiband advanced chromosome analysis.

[Hislop, 2002] Hislop, A. A. (2002). Airway and blood vessel interaction during lung development. *Journal of Anatomy*, 201(4):325–334.

[Hogg, 2004] Hogg, J. C. (2004). Pathophysiology of airflow limitation in chronic obstructive pulmonary disease. *The Lancet*, 364(9435):709–721.

[Hou et al., 2019] Hou, Y., Ma, Z., Liu, C., and Loy, C. C. (2019). Learning lightweight lane detection CNNs by self attention distillation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1013–1021. IEEE.

[Howling et al., 1998] Howling, S., Evans, T., and Hansell, D. (1998). The significance of bronchial dilatation on ct in patients with adult respiratory distress syndrome. *Clinical Radiology*, 53(2):105–109.

[Hsu and Lachenbruch, 2014] Hsu, H. and Lachenbruch, P. A. (2014). Paired t test. *Wiley StatsRef: Statistics Reference Online*.

[Hu et al., 2018] Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7132–7141.

[Huang et al., 2017a] Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017a). Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708.

[Huang et al., 2017b] Huang, X., Shan, J., and Vaidya, V. (2017b). Lung nodule detection in ct using 3d convolutional neural networks. In *IEEE International Symposium on Biomedical Imaging*, pages 379–383.

[Huber et al., 2018] Huber, D., von Voithenberg, L. V., and Kaigala, G. (2018). Fluorescence in situ hybridization (FISH): history, limitations and what to expect from micro-scale FISH? *Micro and Nano Engineering*, 1:15–24.

[Ikaros, 2021] Ikaros (2021). Ikaros karyotyping system.

[Ioffe and Szegedy, 2015] Ioffe, S. and Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning*, pages 448–456. PMLR.

[Isola et al., 2017] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *arXiv preprint arXiv: 1611.07004*.

[Jacobs et al., 2014] Jacobs, C., van Rikxoort, E. M., Twellmann, T., Scholten, E. T., de Jong, P. A., Kuhnigk, J.-M., Oudkerk, M., de Koning, H. J., Prokop, M., Schaefer-Prokop, C., et al. (2014). Automatic detection of subsolid pulmonary nodules in thoracic computed tomography images. *Medical Image Analysis*, 18(2):374–384.

[Jaderberg et al., 2015] Jaderberg, M., Simonyan, K., Zisserman, A., et al. (2015). Spatial transformer networks. In *Advances in Neural Information Processing Systems*, pages 2017–2025.

[Ji, 1994] Ji, L. (1994). Fully automatic chromosome segmentation. *Cytometry: The Journal of the International Society for Analytical Cytology*, 17(3):196–208.

[Jin et al., 2017] Jin, D., Xu, Z., Harrison, A. P., George, K., and Mollura, D. J. (2017). 3D convolutional neural networks with graph refinement for airway segmentation using incomplete data labels. In *International Workshop on Machine Learning in Medical Imaging*, pages 141–149. Springer.

[Juarez et al., 2019] Juarez, A. G.-U., Selvan, R., Saghir, Z., and de Bruijne, M. (2019). A joint 3D UNet-graph neural network-based method for airway segmentation from chest CTs. In *International Workshop on Machine Learning in Medical Imaging*, pages 583–591. Springer.

[Juarez et al., 2018] Juarez, A. G.-U., Tiddens, H., and de Bruijne, M. (2018). Automatic airway segmentation in chest CT using convolutional neural networks. In *Image Analysis for Moving Organ, Breast, and Thoracic Images*, pages 238–250. Springer.

[Jurafsky and Martin, 2014] Jurafsky, D. and Martin, J. H. (2014). *Speech and Language Processing*, volume 3. London: Pearson.

[Kamnitsas et al., 2017] Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., and Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis*, 36:61–78.

[Kampffmeyer et al., 2018] Kampffmeyer, M., Dong, N., Liang, X., Zhang, Y., and Xing, E. P. (2018). Connnet: A long-range relation-aware pixel-connectivity network for salient segmentation. *IEEE Transactions on Image Processing*, 28(5):2518–2529.

[Kandathil and Chamarthy, 2018] Kandathil, A. and Chamarthy, M. (2018). Pulmonary vascular anatomy & anatomical variants. *Cardiovascular Diagnosis and Therapy*, 8(3):201.

[Kauczor et al., 2000] Kauczor, H.-U., Heitmann, K., Heussel, C. P., Marwede, D., Uthmann, T., and Thelen, M. (2000). Automatic detection and quantification of ground-glass opacities on high-resolution CT using multiple neural networks: comparison with a density mask. *American Journal of Roentgenology*, 175(5):1329–1334.

[Kingma and Ba, 2014] Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

[Kiraly et al., 2002] Kiraly, A. P., Higgins, W. E., McLennan, G., Hoffman, E. A., and Reinhardt, J. M. (2002). Three-dimensional human airway segmentation methods for clinical virtual bronchoscopy. *Academic Radiology*, 9(10):1153–1168.

[Kitamura et al., 2016] Kitamura, Y., Li, Y., Ito, W., and Ishikawa, H. (2016). Data-dependent higher-order clique selection for artery–vein segmentation by energy minimization. *International Journal of Computer Vision*, 117(2):142–158.

[Kitasaka et al., 2003] Kitasaka, T., Mori, K., Suenaga, Y., Hasegawa, J.-i., and Toriwaki, J.-i. (2003). A method for segmenting bronchial trees from 3D chest x-ray CT images. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 603–610. Springer.

[Korfiatis et al., 2011] Korfiatis, P. D., Kalogeropoulou, C., Karahaliou, A. N., Kazantzi, A. D., and Costaridou, L. I. (2011). Vessel tree segmentation in presence of interstitial lung disease in MDCT. *IEEE Transactions on Information Technology in Biomedicine*, 15(2):214–220.

[Kostis et al., 2003] Kostis, W. J., Reeves, A. P., Yankelevitz, D. F., and Henschke, C. I. (2003). Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical CT images. *IEEE Transactions on Medical Imaging*, 22(10):1259–1274.

[Krähenbühl and Koltun, 2011] Krähenbühl, P. and Koltun, V. (2011). Efficient inference in fully connected CRFs with gaussian edge potentials. In *Advances in Neural Information Processing Systems*, pages 109–117.

[Krissian et al., 2000] Krissian, K., Malandain, G., Ayache, N., Vaillant, R., and Trousset, Y. (2000). Model-based detection of tubular structures in 3D images. *Computer Vision and Image Understanding*, 80(2):130–171.

[Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105.

[Kubota et al., 2011] Kubota, T., Jerebko, A. K., Dewan, M., Salganicoff, M., and Krishnan, A. (2011). Segmentation of pulmonary nodules of various densities with morphological approaches and convexity models. *Medical Image Analysis*, 15(1):133–154.

[Kuhnigk et al., 2006] Kuhnigk, J.-M., Dicken, V., Bornemann, L., Bakai, A., Wormanns, D., Krass, S., and Peitgen, H.-O. (2006). Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic ct scans. *IEEE Transactions on Medical Imaging*, 25(4):417–434.

[Kuhnigk et al., 2005] Kuhnigk, J.-M., Dicken, V., Zidowitz, S., Bornemann, L., Kuemmerlen, B., Krass, S., Peitgen, H.-O., Yuval, S., Jend, H.-H., Rau, W. S., et al. (2005). New tools for computer assistance in thoracic ct. part 1. functional analysis of lungs, lung lobes, and bronchopulmonary segments. *Radiographics*, 25(2):525–536.

[Lassen et al., 2010] Lassen, B., Kuhnigk, J.-M., Friman, O., Krass, S., and Peitgen, H.-O. (2010). Automatic segmentation of lung lobes in ct images based on fissures, vessels, and bronchi. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 560–563. IEEE.

[Lassen et al., 2012] Lassen, B., van Rikxoort, E. M., Schmidt, M., Kerkstra, S., van Ginneken, B., and Kuhnigk, J.-M. (2012). Automatic segmentation of the pulmonary lobes from chest CT scans based on fissures, vessels, and bronchi. *IEEE Transactions on Medical Imaging*, 32(2):210–222.

[LeCun et al., 2015] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

[LeCun et al., 1998] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

[Lee et al., 2001] Lee, C., Gisselsson, D., Jin, C., Nordgren, A., Ferguson, D. O., Blennow, E., Fletcher, J. A., and Morton, C. C. (2001). Limitations of chromosome classification by multicolor karyotyping. *The American Journal of Human Genetics*, 68(4):1043–1047.

[Lee et al., 2011] Lee, H., Grosse, R., Ranganath, R., and Ng, A. Y. (2011). Unsupervised learning of hierarchical representations with convolutional deep belief networks. *Communications of the ACM*, 54(10):95–103.

[Lee et al., 1994] Lee, T.-C., Kashyap, R. L., and Chu, C.-N. (1994). Building skeleton models via 3-D medial surface axis thinning algorithms. *CVGIP: Graphical Models and Image Processing*, 56(6):462–478.

[Lerner et al., 1995] Lerner, B., Guterman, H., Dinstein, I., and Romem, Y. (1995). Medial axis transform-based features and a neural network for human chromosome classification. *Pattern Recognition*, 28(11):1673–1683.

[Li et al., 2019] Li, Y., Dai, Y., Duan, X., Zhang, W., Guo, Y., and Wang, J. (2019). Application of automated bronchial 3D-CT measurement in pulmonary contusion complicated with acute respiratory distress syndrome. *Journal of X-ray Science and Technology*, 27(4):641–654.

[Lin et al., 2017] Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2980–2988. IEEE.

[Lin et al., 2015] Lin, T.-Y., RoyChowdhury, A., and Maji, S. (2015). Bilinear cnn models for fine-grained visual recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1449–1457.

[Lo and de Bruijne, 2008] Lo, P. and de Bruijne, M. (2008). Voxel classification based airway tree segmentation. In *Medical Imaging 2008: Image Processing*, volume 6914, page 69141K. International Society for Optics and Photonics.

[Lo et al., 2010a] Lo, P., Sporring, J., Ashraf, H., Pedersen, J. J., and de Bruijne, M. (2010a). Vessel-guided airway tree segmentation: a voxel classification approach. *Medical Image Analysis*, 14(4):527–538.

[Lo et al., 2010b] Lo, P., van Ginneken, B., and de Bruijne, M. (2010b). Vessel tree extraction using locally optimal paths. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 680–683. IEEE.

[Lo et al., 2012] Lo, P., Van Ginneken, B., Reinhardt, J. M., Yavarna, T., De Jong, P. A., Irving, B., Fetita, C., Ortner, M., Pinho, R., Sijbers, J., et al. (2012). Extraction of airways from CT (EXACT'09). *IEEE Transactions on Medical Imaging*, 31(11):2093–2107.

[Loganathan et al., 2013] Loganathan, E., Anuja, M., and Madian, N. (2013). Analysis of human chromosome images for the identification of centromere position and length. In *Point-of-Care Healthcare Technologies (PHT), 2013 IEEE*, pages 314–317. IEEE.

[Loh et al., 2010] Loh, S., Bagheri, S., Katzberg, R. W., Fung, M. A., and Li, C.-S. (2010). Delayed adverse reaction to contrast-enhanced ct: a prospective single-center study comparison to control group without enhancement. *Radiology*, 255(3):764–771.

[Long et al., 2015] Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440.

[Lopez Torres et al., 2015] Lopez Torres, E., Fiorina, E., Pennazio, F., Peroni, C., Saletta, M., Camarlinghi, N., Fantacci, M., and Cerello, P. (2015). Large scale validation of the M5L lung CAD on heterogeneous CT datasets. *Medical Physics*, 42(4):1477–1489.

[Lotter et al., 2017] Lotter, W., Sorensen, G., and Cox, D. (2017). A multi-scale cnn and curriculum learning strategy for mammogram classification. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 169–177. Springer.

[Maaten and Hinton, 2008] Maaten, L. v. d. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605.

[Madian and Jayanthi, 2014] Madian, N. and Jayanthi, K. (2014). Analysis of human chromosome classification using centromere position. *Measurement*, 47:287–295.

[Markou et al., 2012] Markou, C., Maramis, C., Delopoulos, A., Daiou, C., and Lambropoulos, A. (2012). Automatic chromosome classification using support vector machines. In *Pattern Recognition: Methods and Applications*, pages 1–24. iConcept Press.

[Mason et al., 2010] Mason, R. J., Broaddus, V. C., Martin, T. R., King, T. E., Schraufnagel, D., Murray, J. F., and Nadel, J. A. (2010). *Murray and Nadel's Textbook of Respiratory Medicine E-Book: 2-Volume Set*. Elsevier Health Sciences.

[Masuda and Takahashi, 2002] Masuda, A. and Takahashi, T. (2002). Chromosome instability in human lung cancers: possible underlying mechanisms and potential consequences in the pathogenesis. *Oncogene*, 21(45):6884–6897.

[Mayer et al., 2004] Mayer, D., Bartz, D., Fischer, J., Ley, S., del Rıo, A., Thust, S., Kauczor, H.-U., and Heussel, C. P. (2004). Hybrid segmentation and virtual bronchoscopy based on CT images. *Academic Radiology*, 11(5):551–565.

[McCulloch and Pitts, 1943] McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133.

[McGowan-Jordan et al., 2016] McGowan-Jordan, J., Simons, A., and Schmid, M. (2016). An international system for human cytogenomic nomenclature (ISCN). *Reprint of: Cytogenetic and Genome Research 2016*, 149(1–2).

[Mekada et al., 2006] Mekada, Y., Nakamura, S., Ide, I., Murase, H., and Otsuji, H. (2006). Pulmonary artery and vein classification using spatial arrangement features from X-ray CT images. In *Proceedings of the 7th Asia-pacific Conference on Control and Measurement*, pages 232–235.

[Melot and Naeije, 2011] Melot, C. and Naeije, R. (2011). Pulmonary vascular diseases. *Comprehensive Physiology*, 1(2):593–619.

[Meng et al., 2017] Meng, Q., Roth, H. R., Kitasaka, T., Oda, M., Ueno, J., and Mori, K. (2017). Tracking and segmentation of the airways in chest CT using a fully convolutional network. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 198–207. Springer.

[Messay et al., 2010] Messay, T., Hardie, R. C., and Rogers, S. K. (2010). A new computationally efficient CAD system for pulmonary nodule detection in CT imagery. *Medical Image Analysis*, 14(3):390–406.

[Metz et al., 2007] Metz, C., Schaap, M., and Niessen, W. (2007). Semi-automatic coronary artery centerline extraction in computed tomography angiography data. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 856–859. IEEE.

[Micci et al., 2001] Micci, F., Teixeira, M. R., and Heim, S. (2001). Complete cytogenetic characterization of the human breast cancer cell line ma11 combining g-banding, comparative genomic hybridization, multicolor fluorescence in situ hybridization, rxfish, and chromosome-specific painting. *Cancer Genetics and Cytogenetics*, 131(1):25–30.

[Miller, 1947] Miller, W. S. (1947). *The lung*. CC Thomas.

[Milletari et al., 2016] Milletari, F., Navab, N., and Ahmadi, S.-A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *Proceedings of the 2016 International Conference on 3D Vision*, pages 565–571.

[Minaee et al., 2014] Minaee, S., Fotouhi, M., and Khalaj, B. H. (2014). A geometric approach to fully automatic chromosome segmentation. In *2014 IEEE Signal Processing in Medicine and Biology Symposium*, pages 1–6. IEEE.

[Ming and Tian, 2010] Ming, D. and Tian, J. (2010). Automatic pattern extraction and classification for chromosome images. *Journal of Infrared, Millimeter, and Terahertz Waves*, 31(7):866–877.

[Minna et al., 2002] Minna, J. D., Roth, J. A., and Gazdar, A. F. (2002). Focus on lung cancer. *Cancer Cell*, 1(1):49–52.

[Mirza and Osindero, 2014] Mirza, M. and Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

[Mitelman, 2000] Mitelman, F. (2000). Recurrent chromosome aberrations in cancer. *Mutation Research/Reviews in Mutation Research*, 462(2-3):247–253.

[Mitelman et al., 2004] Mitelman, F., Johansson, B., and Mertens, F. (2004). Fusion genes and rearranged genes as a linear function of chromosome aberrations in cancer. *Nature Genetics*, 36(4):331–334.

[Moreno et al., 2006] Moreno, L., Perez-Vizcaino, F., Harrington, L., Faro, R., Sturton, G., Barnes, P. J., and Mitchell, J. A. (2006). Pharmacology of airways and vessels in lung slices in situ: role of endogenous dilator hormones. *Respiratory Research*, 7(1):1–7.

[Mori et al., 2000] Mori, K., Hasegawa, J.-I., Suenaga, Y., and Toriwaki, J.-I. (2000). Automated anatomical labeling of the bronchial branch and its application to the virtual bronchoscopy system. *IEEE Transactions on Medical Imaging*, 19(2):103–114.

[Mukhopadhyay, 2016] Mukhopadhyay, S. (2016). A segmentation framework of pulmonary nodules in lung CT images. *Journal of Digital Imaging*, 29(1):86–103.

[Nardelli et al., 2018] Nardelli, P., Jimenez-Carretero, D., Bermejo-Pelaez, D., Washko, G. R., Rahaghi, F. N., Ledesma-Carbayo, M. J., and Estépar, R. S. J. (2018). Pulmonary artery–vein classification in CT images using deep learning. *IEEE Transactions on Medical Imaging*, 37(11):2428–2440.

[Natarajan, 2002] Natarajan, A. T. (2002). Chromosome aberrations: past, present and future. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 504(1):3–16.

[Natori et al., 2005] Natori, H., Takabatake, H., Mori, M., Koba, H., Mori, K., Kitasaka, T., and Suenaga, Y. (2005). Virtual navigation of central and peripheral airways. *Chest*, 128(4):327S.

[Netter, 2014] Netter, F. H. (2014). *Atlas of Human Anatomy*. Elsevier.

[Ochs et al., 2007] Ochs, R. A., Goldin, J. G., Abtin, F., Kim, H. J., Brown, K., Batra, P., Roback, D., McNitt-Gray, M. F., and Brown, M. S. (2007). Automated classification of lung bronchovascular anatomy in CT using adaboost. *Medical Image Analysis*, 11(3):315–324.

[Ojala et al., 2002] Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987.

[Otsu, 1979] Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66.

[Pan et al., 2018] Pan, X., Yang, D., Li, L., Liu, Z., Yang, H., Cao, Z., He, Y., Ma, Z., and Chen, Y. (2018). Cell detection in pathology and microscopy images with multi-scale fully convolutional neural networks. *World Wide Web*, pages 1–23.

[Park and Kwak, 2016] Park, S. and Kwak, N. (2016). Analysis on the dropout effect in convolutional neural networks. In *Proceedings of the Asian Conference on Computer Vision*, pages 189–204.

[Park et al., 2013] Park, S., Min Lee, S., Kim, N., Beom Seo, J., and Shin, H. (2013). Automatic reconstruction of the arterial and venous trees on volumetric chest CT. *Medical Physics*, 40(7):071906.

[Park et al., 2001] Park, S.-Y., Choi, H.-C., Chun, Y.-H., Kim, H., and Park, S.-H. (2001). Characterization of chromosomal aberrations in lung cancer cell lines by cross-species color banding. *Cancer Genetics and Cytogenetics*, 124(1):62–70.

[Paszke et al., 2019] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703*.

[Payer et al., 2016] Payer, C., Pienn, M., Bálint, Z., Shekhovtsov, A., Talakic, E., Nagy, E., Olschewski, A., Olschewski, H., and Urschler, M. (2016). Automated integer programming based separation of arteries and veins from thoracic CT images. *Medical Image Analysis*, 34:109–122.

[Peacock et al., 2016] Peacock, A. J., Naeije, R., and Rubin, L. J. (2016). *Pulmonary circulation: diseases and their treatment*. CRC Press.

[Pham et al., 2000] Pham, D. L., Xu, C., and Prince, J. L. (2000). Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2(1):315–337.

[Piper, 1990] Piper, J. (1990). Automated cytogenetics in the study of mutagenesis and cancer. In *Advances in Mutagenesis Research*, pages 127–153. Springer.

[Porres et al., 2013] Porres, D. V., Morenza, Ó. P., Pallisa, E., Roque, A., Andreu, J., and Martínez, M. (2013). Learning from the pulmonary veins. *Radiographics*, 33(4):999–1022.

[Qiang et al., 2014] Qiang, Y., Wang, Q., Xu, G., Ma, H., Deng, L., Zhang, L., Pu, J., and Guo, Y. (2014). Computerized segmentation of pulmonary nodules depicted in CT examinations using freehand sketches. *Medical Physics*, 41(4):041917.

[Qin et al., 2019] Qin, Y., Chen, M., Zheng, H., Gu, Y., Shen, M., Yang, J., Huang, X., Zhu, Y.-M., and Yang, G.-Z. (2019). AirwayNet: A voxel-connectivity aware approach for accurate airway segmentation using convolutional neural networks. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 212–220. Springer.

[Radford et al., 2015] Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

[Rahaghi et al., 2016] Rahaghi, F., Ross, J., Agarwal, M., González, G., Come, C., Diaz, A., Vegas-Sánchez-Ferrero, G., Hunsaker, A., Estépar, R. S. J., Waxman, A., et al. (2016). Pulmonary vascular morphology as an imaging biomarker in chronic thromboembolic pulmonary hypertension. *Pulmonary Circulation*, 6(1):70–81.

[Reeves et al., 2006] Reeves, A. P., Chan, A. B., Yankelevitz, D. F., Henschke, C. I., Kressler, B., and Kostis, W. J. (2006). On measuring the change in size of pulmonary nodules. *IEEE Transactions on Medical Imaging*, 25(4):435–450.

[Ren et al., 2015] Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, pages 91–99.

[Rickmann et al., 2019] Rickmann, A.-M., Roy, A. G., Sarasua, I., Navab, N., and Wachinger, C. (2019). 'Project & Excite' modules for segmentation of volumetric medical scans. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 39–47. Springer.

[Rødahl et al., 2005] Rødahl, E., Lybæk, H., Arnes, J., and Ness, G. O. (2005). Chromosomal imbalances in some benign orbital tumours. *Acta Ophthalmologica Scandinavica*, 83(3):385–391.

[Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241.

[Rowley, 2001] Rowley, J. D. (2001). Chromosome translocations: dangerous liaisons revisited. *Nature Reviews Cancer*, 1(3):245–250.

[Rumelhart et al., 1986] Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088):533–536.

[Russell and Norvig, 2016] Russell, S. J. and Norvig, P. (2016). *Artificial Intelligence: a Modern Approach*. Malaysia: Pearson Education Limited.

[Saha et al., 2010] Saha, P. K., Gao, Z., Alford, S. K., Sonka, M., and Hoffman, E. A. (2010). Topomorphologic separation of fused isointensity objects via multiscale opening: Separating arteries and veins in 3-D pulmonary CT. *IEEE Transactions on Medical Imaging*, 29(3):840–851.

[Sakamoto and Nakano, 2016] Sakamoto, M. and Nakano, H. (2016). Cascaded neural networks with selective classifiers and its evaluation using lung X-ray CT images. *arXiv preprint arXiv:1611.07136*.

[Saleh et al., 2019] Saleh, H. M., Saad, N. H., and Isa, N. A. M. (2019). Overlapping chromosome segmentation using u-net: Convolutional networks with test time augmentation. *Procedia Computer Science*, 159:524–533.

[Samuels et al., 2003] Samuels, M. L., Witmer, J. A., and Schaffner, A. A. (2003). *Statistics For the Life Sciences*, volume 4. Pearson Prentice Hall, Upper Saddle River.

[Schlathoelter et al., 2002] Schlathoelter, T., Lorenz, C., Carlsen, I. C., Renisch, S., and Deschamps, T. (2002). Simultaneous segmentation and tree reconstruction of the airways for virtual bronchoscopy. In *Medical Imaging 2002: Image Processing*, volume 4684, pages 103–113. International Society for Optics and Photonics.

[Schlemper et al., 2019] Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., and Rueckert, D. (2019). Attention gated networks: Learning to leverage salient regions in medical images. *Medical Image Analysis*, 53:197–207.

[Schröck et al., 1996] Schröck, E., Du Manoir, S., Veldman, T., Schoell, B., Wienberg, J., Ferguson-Smith, M., Ning, Y., Ledbetter, D., Bar-Am, I., Soenksen, D., et al. (1996). Multicolor spectral karyotyping of human chromosomes. *Science*, 273(5274):494–497.

[Selvan, 2018] Selvan, R. (2018). *Extraction of Airways from Volumetric Data*. PhD thesis, University of Copenhagen.

[Selvan et al., 2020] Selvan, R., Kipf, T., Welling, M., Juarez, A. G.-U., Pedersen, J. H., Petersen, J., and de Bruijne, M. (2020). Graph refinement based airway extraction using mean-field networks and graph neural networks. *Medical Image Analysis*, 64:101751.

[Sethi and Rochester, 2000] Sethi, J. M. and Rochester, C. L. (2000). Smoking and chronic obstructive pulmonary disease. *Clinics in Chest Medicine*, 21(1):67–86.

[Setio et al., 2015] Setio, A. A., Jacobs, C., Gelderblom, J., and van Ginneken, B. (2015). Automatic detection of large pulmonary solid nodules in thoracic CT images. *Medical Physics*, 42(10):5642–5653.

[Setio et al., 2016] Setio, A. A. A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., van Riel, S. J., Wille, M. M. W., Naqibullah, M., Sánchez, C. I., and van Ginneken, B. (2016). Pulmonary nodule detection in CT images: false positive reduction using multi-view convolutional networks. *IEEE Transactions on Medical Imaging*, 35(5):1160–1169.

[Sharma et al., 2017] Sharma, M., Saha, O., Sriraman, A., Hebbalaguppe, R., Vig, L., and Karande, S. (2017). Crowdsourcing for chromosome segmentation and deep classification. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 786–793. IEEE.

[Shaw et al., 2002] Shaw, R., Djukanovic, R., Tashkin, D., Millar, A., Du Bois, R., and Corris, P. (2002). The role of small airways in lung disease. *Respiratory Medicine*, 96(2):67–80.

[Shen et al., 2015a] Shen, M., Giannarou, S., and Yang, G.-Z. (2015a). Robust camera localisation with depth reconstruction for bronchoscopic navigation. *International Journal of Computer Assisted Radiology and Surgery*, 10(6):801–813.

[Shen et al., 2019] Shen, M., Gu, Y., Liu, N., and Yang, G.-Z. (2019). Context-aware depth and pose estimation for bronchoscopic navigation. *IEEE Robotics and Automation Letters*, 4(2):732–739.

[Shen et al., 2015b] Shen, W., Zhou, M., Yang, F., Yang, C., and Tian, J. (2015b). Multi-scale convolutional neural networks for lung nodule classification. In *International Conference on Information Processing in Medical Imaging*, pages 588–599. Springer.

[Shikata et al., 2009] Shikata, H., McLennan, G., Hoffman, E. A., and Sonka, M. (2009). Segmentation of pulmonary vascular trees from thoracic 3D CT images. *International Journal of Biomedical Imaging*, 2009.

[Shrivastava et al., 2017] Shrivastava, A., Pfister, T., Tuzel, O., Susskind, J., Wang, W., and Webb, R. (2017). Learning from simulated and unsupervised images through adversarial training. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, volume 3, page 6.

[Siegel et al., 2016] Siegel, R. L., Miller, K. D., and Jemal, A. (2016). Cancer statistics, 2016. *CA: A Cancer Journal for Clinicians*, 66(1):7–30.

[Simonyan and Zisserman, 2014] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[Sobel, 1990] Sobel, I. (1990). An isotropic $3\times$ 3 image gradient operator. *Machine Vision for Three-Dimensional Scenes*, pages 376–379.

[Speicher et al., 1996] Speicher, M. R., Ballard, S. G., and Ward, D. C. (1996). Karyotyping human chromosomes by combinatorial multi-fluor fish. *Nature Genetics*, 12(4):368.

[Srivastava et al., 2014] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958.

[Stanley et al., 1996] Stanley, R. J., Keller, J., Caldwell, C. W., and Gader, P. (1996). Centromere attribute integration based chromosome polarity assignment. In *Proceedings of the AMIA Annual Fall Symposium*, page 284. American Medical Informatics Association.

[Szegedy et al., 2016] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826.

[Testa and Siegfried, 1992] Testa, J. R. and Siegfried, J. M. (1992). Chromosome abnormalities in human non-small cell lung cancer. *Cancer Research*, 52(9):2702–2706.

[Theisen and Shaffer, 2010] Theisen, A. and Shaffer, L. G. (2010). Disorders caused by chromosome abnormalities. *The Application of Clinical Genetics*, 3:159–174.

[Torre et al., 2016] Torre, L. A., Siegel, R. L., and Jemal, A. (2016). Lung cancer statistics. In *Lung Cancer and Personalized Medicine*, pages 1–19. Springer.

[Tschirren et al., 2005] Tschirren, J., Hoffman, E. A., McLennan, G., and Sonka, M. (2005). Intrathoracic airway trees: segmentation and airway morphology analysis from low-dose CT scans. *IEEE Transactions on Medical Imaging*, 24(12):1529–1539.

[U. S. National Institutes of Health, 2021] U. S. National Institutes of Health, N. C. I. (2021). Seer training modules, cancer registration & surveillance modules. [Online; accessed 10-March-2021].

[Ukil and Reinhardt, 2008] Ukil, S. and Reinhardt, J. M. (2008). Anatomy-guided lung lobe segmentation in x-ray CT images. *IEEE Transactions on Medical Imaging*, 28(2):202–214.

[van Dongen and van Ginneken, 2010] van Dongen, E. and van Ginneken, B. (2010). Automatic segmentation of pulmonary vasculature in thoracic CT scans with local thresholding and airway wall removal. In *IEEE International Symposium on Biomedical Imaging*, pages 668–671. IEEE.

[van Ginneken et al., 2008] van Ginneken, B., Baggerman, W., and van Rikxoort, E. M. (2008). Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 219–226. Springer.

[Van Rikxoort et al., 2009] Van Rikxoort, E. M., Baggerman, W., and van Ginneken, B. (2009). Automatic segmentation of the airway tree from thoracic CT scans using a multi-threshold approach. In *Proceedings of the Second International Workshop on Pulmonary Image Analysis*, pages 341–349.

[Vogelmeier et al., 2017] Vogelmeier, C. F., Criner, G. J., Martinez, F. J., Anzueto, A., Barnes, P. J., Bourbeau, J., Celli, B. R., Chen, R., Decramer, M., Fabbri, L. M., et al. (2017). Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report. gold executive summary. *American Journal of Respiratory and Critical Care Medicine*, 195(5):557–582.

[Wang et al., 2019] Wang, C., Hayashi, Y., Oda, M., Itoh, H., Kitasaka, T., Frangi, A. F., and Mori, K. (2019). Tubular structure segmentation using spatial fully connected network with radial distance loss for 3D medical images. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 348–356. Springer.

[Wang et al., 2007] Wang, J., Engelmann, R., and Li, Q. (2007). Segmentation of pulmonary nodules in three-dimensional CT images by use of a spiral-scanning technique. *Medical Physics*, 34(12):4678–4689.

[Wang et al., 2017] Wang, S., Zhou, M., Liu, Z., Liu, Z., Gu, D., Zang, Y., Dong, D., Gevaert, O., and Tian, J. (2017). Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation. *Medical Image Analysis*, 40:172–183.

[Wang et al., 2008] Wang, X., Zheng, B., Li, S., Mulvihill, J. J., and Liu, H. (2008). A rule-based computer scheme for centromere identification and polarity assignment of metaphase chromosomes. *Computer Methods and Programs in Biomedicine*, 89(1):33–42.

[Weibel, 2009] Weibel, E. R. (2009). What makes a good lung? *Swiss Medical Weekly*, 139(2728).

[Weibel and Gomez, 1962] Weibel, E. R. and Gomez, D. M. (1962). Architecture of the human lung: Use of quantitative methods establishes fundamental relations between size and number of lung structures. *Science*, 137(3530):577–585.

[West, 2011] West, J. B. (2011). *Pulmonary pathophysiology: the essentials*. Lippincott Williams & Wilkins.

[Wiemker et al., 2004] Wiemker, R., Blaffert, T., Bülow, T., Renisch, S., and Lorenz, C. (2004). Automated assessment of bronchial lumen, wall thickness and bronchoarterial diameter ratio of the tracheobronchial tree using high-resolution CT. In *International Congress Series*, volume 1268, pages 967–972. Elsevier.

[Wikimedia, 2020] Wikimedia, C. (2020). File:2119 pulmonary circuit.jpg — wikimedia commons, the free media repository. [Online; accessed 11-March-2021].

[Wikimedia, 2021] Wikimedia, C. (2021). File:respiratory_system_complete_en.svg — wikimedia commons, the free media repository. [Online; accessed 10-March-2021].

[Willems, 2019] Willems, K. (2019). Keras tutorial: Deep learning in python. [Online; accessed March 6, 2021].

[Wilson et al., 2012] Wilson, D. O., Ryan, A., Fuhrman, C., Schuchert, M., Shapiro, S., Siegfried, J. M., and Weissfeld, J. (2012). Doubling times and CT screen–detected lung cancers in the pittsburgh lung screening study. *American Journal of Respiratory and Critical Care Medicine*, 185(1):85–89.

[Winer-Muram, 2006] Winer-Muram, H. T. (2006). The solitary pulmonary nodule. *Radiology*, 239(1):34–49.

[Wittenberg et al., 2012] Wittenberg, R., Berger, F. H., Peters, J. F., Weber, M., van Hoorn, F., Beenen, L. F., van Doorn, M. M., van Schuppen, J., Zijlstra, I. A., Prokop, M., et al. (2012). Acute pulmonary embolism: effect of a computer-assisted detection prototype on diagnosis—an observer study. *Radiology*, 262(1):305–313.

[Wu et al., 2018a] Wu, B., Zhou, Z., Wang, J., and Wang, Y. (2018a). Joint learning for pulmonary nodule segmentation, attributes and malignancy prediction. In *IEEE International Symposium on Biomedical Imaging*, pages 1109–1113. IEEE.

[Wu et al., 2019] Wu, X., Kim, G. H., Salisbury, M. L., Barber, D., Bartholmai, B. J., Brown, K. K., Conoscenti, C. S., De Backer, J., Flaherty, K. R., Gruden, J. F., et al. (2019). Computed tomographic biomarkers in idiopathic pulmonary fibrosis. the future of quantitative analysis. *American Journal of Respiratory and Critical Care Medicine*, 199(1):12–21.

[Wu et al., 2018b] Wu, Y., Yue, Y., Tan, X., Wang, W., and Lu, T. (2018b). End-to-end chromosome karyotyping with data augmentation using gan. In *IEEE International Conference on Image Processing*, pages 2456–2460. IEEE.

[Xiao et al., 2020] Xiao, L., Luo, C., Yu, T., Luo, Y., Wang, M., Yu, F., Li, Y., Tian, C., and Qiao, J. (2020). Deepacev2: Automated chromosome enumeration in metaphase cell images using deep convolutional neural networks. *IEEE Transactions on Medical Imaging*, 39(12):3920–3932.

[Xie and Tu, 2015] Xie, S. and Tu, Z. (2015). Holistically-nested edge detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1395–1403.

[Xu et al., 2015] Xu, Z., Bagci, U., Foster, B., Mansoor, A., Udupa, J. K., and Mollura, D. J. (2015). A hybrid method for airway segmentation and automated measurement of bronchial wall thickness on CT. *Medical Image Analysis*, 24(1):1–17.

[Yang et al., 2010] Yang, W., Stotler, B., Sevilla, D. W., Emmons, F. N., Murty, V. V., Alobeid, B., and Bhagat, G. (2010). Fish analysis in addition to g-band karyotyping: utility in evaluation of myelodysplastic syndromes? *Leukemia Research*, 34(4):420–425.

[Yankelevitz et al., 2000] Yankelevitz, D. F., Reeves, A. P., Kostis, W. J., Zhao, B., and Henschke, C. I. (2000). Small pulmonary nodules: volumetrically determined growth rates based on CT evaluation. *Radiology*, 217(1):251–256.

[Ye et al., 2009] Ye, X., Lin, X., Dehmeshki, J., Slabaugh, G., and Beddoe, G. (2009). Shape-based computer-aided detection of lung nodules in thoracic CT images. *IEEE Transactions on Biomedical Engineering*, 56(7):1810–1820.

[Yun et al., 2019] Yun, J., Park, J., Yu, D., Yi, J., Lee, M., Park, H. J., Lee, J.-G., Seo, J. B., and Kim, N. (2019). Improvement of fully automated airway segmentation on volumetric computed tomographic images using a 2.5 dimensional convolutional neural net. *Medical Image Analysis*, 51:13–20.

[Yushkevich et al., 2006] Yushkevich, P. A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J. C., and Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*, 31(3):1116–1128.

[Zagoruyko and Komodakis, 2016] Zagoruyko, S. and Komodakis, N. (2016). Wide residual networks. *arXiv preprint arXiv:1605.07146*.

[Zagoruyko and Komodakis, 2017] Zagoruyko, S. and Komodakis, N. (2017). Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. In *Proceedings of the International Conference on Learning Representations*. OpenReview.net.

[Zeng et al., 2017] Zeng, L., Xu, X., Cai, B., Qiu, S., and Zhang, T. (2017). Multi-scale convolutional neural networks for crowd counting. In *IEEE International Conference on Image Processing*, pages 465–469. IEEE.

[Zhao et al., 2003] Zhao, B., Gamsu, G., Ginsberg, M. S., Jiang, L., and Schwartz, L. H. (2003). Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm. *Journal of Applied Clinical Medical Physics*, 4(3):248–260.

[Zhao et al., 2019] Zhao, T., Yin, Z., Wang, J., Gao, D., Chen, Y., and Mao, Y. (2019). Bronchus segmentation and classification by neural networks and linear programming. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 230–239. Springer.

[Zhou et al., 2012] Zhou, C., Chan, H.-P., Kuriakose, J. W., Chughtai, A., Wei, J., Hadjiiski, L. M., Guo, Y., Patel, S., and Kazerooni, E. A. (2012). Pulmonary vessel segmentation utilizing curved planar reformation and optimal path finding (CROP) in computed tomographic pulmonary angiography (CTPA) for CAD applications. In *Medical Imaging 2012: Computer-Aided Diagnosis*, volume 8315, page 83150N. International Society for Optics and Photonics.

[Zhou et al., 2007] Zhou, C., Chan, H.-P., Sahiner, B., Hadjiiski, L. M., Chughtai, A., Patel, S., Wei, J., Ge, J., Cascade, P. N., and Kazerooni, E. A. (2007). Automatic multiscale enhancement and segmentation of pulmonary vessels in CT pulmonary angiography images for CAD applications. *Medical Physics*, 34(12):4567–4577.

[Zhou et al., 2006] Zhou, X., Hayashi, T., Hara, T., Fujita, H., Yokoyama, R., Kiryu, T., and Hoshi, H. (2006). Automatic segmentation and recognition of anatomical lung structures from high-resolution chest ct images. *Computerized Medical Imaging and Graphics*, 30(5):299–313.

[Zhu et al., 2017] Zhu, Q., Du, B., Turkbey, B., Choyke, P. L., and Yan, P. (2017). Deeply-supervised CNN for prostate segmentation. In *International Joint Conference on Neural Networks*, pages 178–184. IEEE.

[Zhu et al., 2019] Zhu, W., Huang, Y., Zeng, L., Chen, X., Liu, Y., Qian, Z., Du, N., Fan, W., and Xie, X. (2019). AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Medical Physics*, 46(2):576–589.

Appendices

A. Supplementary Materials of Chromosome Classification Approach in Chapter 2

A.1 Comparison with the State-Of-The-Art Methods

Statistical tests analysis

We conducted statistical tests between the proposed Varifocal-Net and other state-of-the-art methods. Both the two-tailed unpaired (independent) [Samuels et al., 2003] and paired (dependent) t-tests [Hsu and Lachenbruch, 2014] were performed on the accuracy of karyotyping per case (Acc. per Case). The dispatch strategy was not used here for statistical tests since we only care about the classification performance of each network itself. The two t-tests are used as follows:

- The unpaired t-test was used to check whether the two sets of independent and identically distributed samples have the same mean value. The null hypothesis for the unpaired t-test is that the means of the testing results from two different methods are equal. When using the unpaired t-test, we assume that the testing cases for two methods are not related. We just aim to compare the means of these two sets of results and ignore that the testing cases are actually the same.
- The paired t-test was used because all methods were tested on the same testing cases. For comparison between the Varifocal-Net and the other method, each case was tested twice (one for each method). Therefore, the samples of two tests are dependent and can be paired by case. The null hypothesis for the paired t-test is that the pairwise differences between two testing samples are equal. We aim to know whether the performance of the Varifocal-Net is similar to that of the other method.

For both the two t-tests, if the calculated p-value is smaller than 0.05, then the null hypothesis is rejected and the difference between two methods is proved significant. The results of p-value are presented in Table A1 and Table A2 for demonstrating the performance difference on total cases and unhealthy cases, respectively. The results of both two t-tests from Table A1 confirm that the superiority of the proposed Varifocal-Net against all other methods is statistically significant (p-value $\ll 0.05$). The only exception is the performance of L-Net on polarity classification, which is consistent with the results in Table 2.4. For the performance on unhealthy cases, Table A2 demonstrates that except L-Net, the difference between the proposed Varifocal-Net and other methods is statistically

Table A1: Results of unpaired and paired t-tests between the proposed Varifocal-Net and other methods. Scientific notation is used to express numbers. (T: Type, P: Polarity.)

Method	p-value (unpaired)		p-value (paired)	
	T	P	T	P
Sharma *et al.*	2.6×10^{-210}	–	4.3×10^{-243}	–
Gupta *et al.*	1.9×10^{-181}	–	1.5×10^{-192}	–
AlexNet	3.2×10^{-242}	1.9×10^{-94}	4.3×10^{-260}	5.0×10^{-135}
GoogLeNet	4.0×10^{-69}	7.2×10^{-15}	9.1×10^{-101}	6.9×10^{-27}
VGG-Net	9.2×10^{-75}	3.9×10^{-11}	2.8×10^{-99}	1.5×10^{-14}
ResNet	1.5×10^{-56}	2.0×10^{-7}	2.8×10^{-99}	1.5×10^{-14}
DenseNet	1.1×10^{-62}	4.1×10^{-10}	2.4×10^{-105}	5.7×10^{-20}
AlexNet-STN	3.6×10^{-182}	5.1×10^{-57}	6.4×10^{-215}	4.5×10^{-92}
GoogLeNet-STN	5.4×10^{-212}	1.9×10^{-75}	1.6×10^{-223}	3.2×10^{-106}
VGG-Net-STN	6.9×10^{-51}	2.6×10^{-5}	3.8×10^{-97}	6.5×10^{-12}
ResNet-STN	3.5×10^{-57}	2.7×10^{-8}	3.3×10^{-108}	4.1×10^{-18}
DenseNet-STN	1.9×10^{-41}	3.4×10^{-4}	9.4×10^{-86}	1.2×10^{-8}
G-Net-STN	9.0×10^{-76}	5.1×10^{-17}	1.7×10^{-110}	1.7×10^{-34}
L-Net (Simple)	2.2×10^{-123}	1.0×10^{-46}	2.8×10^{-177}	3.4×10^{-80}
Varifocal-Net (Simple)	5.2×10^{-80}	3.2×10^{-41}	6.1×10^{-127}	1.5×10^{-76}
G-Net	1.8×10^{-26}	7.7×10^{-5}	5.3×10^{-59}	2.0×10^{-10}
L-Net	7.2×10^{-8}	7.6×10^{-1}	1.1×10^{-25}	5.6×10^{-1}
Varifocal-Net	–	–	–	–

significant (p-value < 0.05) for type classification. While for polarity classification, the difference between the Varifocal-Net and VGG-Net-STN, ResNet-STN, DenseNet-STN, and L-Net is not considered significant via both of the t-tests. The results in Table A1 and Table A2 are in line with the corresponding results in Table 2.8 and Table 2.9, respectively. In view of the results of the two t-tests, it is reasonable to believe that the proposed Varifocal-Net outperforms state-of-the-art methods statistically.

Performance of chromosome classification in commercial microscope systems

There are mainly three microscope systems that are equipped with automated chromosome classification function: CytoVision System from Leica [CytoVision, 2021], Ikaros from MetaSystem [Ikaros, 2021], and HiBand from ASI [HiBand, 2021]. The classification accuracy of these microscope systems is not provided in their manual, product introduction website or literature. Besides, these systems also control the microscope hardware. They are targeted at the automation of the whole karyotyping workflow, which includes meta-phase scanning and capturing, data archive, and interactive karyotyping. The process of chromosome classification is not specifically optimized. According to our medical colleagues, operators still manually drag each chromosome image for karyotyping in clinical practice.

To quantitatively assess the performance of the classification function in current commercial microscope systems, we have performed additional clinical experiments on 10 new cases from different healthy people (not included in the 1909 cases). The experi-

Table A2: Results of unpaired and paired t-tests between the proposed Varifocal-Net and other methods on unhealthy cases. Scientific notation is used to express numbers. (T: Type, P: Polarity.)

Method	p-value (unpaired)		p-value (paired)	
	T	P	T	P
Sharma *et al.*	5.1×10^{-12}	-	2.1×10^{-14}	-
Gupta *et al.*	1.4×10^{-5}	-	2.7×10^{-6}	-
AlexNet	5.5×10^{-19}	9.4×10^{-11}	5.1×10^{-20}	2.3×10^{-15}
GoogLeNet	7.8×10^{-8}	5.4×10^{-4}	9.3×10^{-10}	8.1×10^{-6}
VGG-Net	8.5×10^{-7}	2.3×10^{-2}	7.9×10^{-10}	3.2×10^{-5}
ResNet	5.6×10^{-5}	1.3×10^{-1}	1.2×10^{-7}	9.7×10^{-3}
DenseNet	4.8×10^{-6}	4.5×10^{-2}	8.8×10^{-9}	1.3×10^{-3}
AlexNet-STN	5.8×10^{-12}	2.2×10^{-6}	1.2×10^{-13}	1.2×10^{-9}
GoogLeNet-STN	3.3×10^{-18}	1.4×10^{-7}	4.1×10^{-19}	1.5×10^{-9}
VGG-Net-STN	2.9×10^{-4}	6.3×10^{-1}	2.9×10^{-8}	3.5×10^{-1}
ResNet-STN	1.3×10^{-2}	9.9×10^{-1}	1.7×10^{-5}	9.9×10^{-1}
DenseNet-STN	1.5×10^{-2}	7.3×10^{-1}	6.0×10^{-7}	5×10^{-1}
G-Net-STN	1.1×10^{-5}	3.9×10^{-2}	3.8×10^{-9}	3.3×10^{-3}
L-Net (Simple)	5.3×10^{-5}	1.6×10^{-2}	1.6×10^{-6}	5.7×10^{-4}
Varifocal-Net (Simple)	6.4×10^{-5}	1.5×10^{-3}	2.6×10^{-10}	6.5×10^{-7}
G-Net	3.9×10^{-3}	1.9×10^{-1}	1.1×10^{-5}	1.9×10^{-2}
L-Net	1.2×10^{-1}	9.4×10^{-1}	2.3×10^{-3}	8.7×10^{-1}
Varifocal-Net	–	–	–	–

Table A3: Classification performance of the Leica's CytoVision System and the proposed Varifocal-Net on 10 patient cases (mean±standard deviation).

Case	# total image	# Misclassified image (Leica)	Acc. per Case (%) (Leica)	Acc. per Case-D (%) (Varifocal-Net)
1	46	10	78.2	100.0
2	46	20	56.5	100.0
3	46	12	73.9	100.0
4	46	12	73.9	100.0
5	46	8	82.6	100.0
6	46	17	63.0	100.0
7	46	21	54.3	100.0
8	46	15	67.3	100.0
9	46	9	80.4	100.0
10	46	18	60.9	100.0
Avg.	46	14.2±4.7	69.1±10.1	100.0±0.0

ments were conducted on the Leica's CytoVision System (GSL-120), which is the leading karyotyping solution provider in the world. For each patient case, the chromosomes were manually segmented and separated by doctors. Then, the classification of each chromosome was automatically accomplished by CytoVision and the final karyotyping result map was generated. Meanwhile, the Varifocal-Net was also employed to test its performance on these new cases. We calculated the accuracy per patient case and presented the results in Table A3. The average accuracy of all 10 cases is 69.1% for Leica's CytoVision System, while the proposed Varifocal-Net achieved 100% accuracy per patient case using the proposed dispatch strategy (Acc. per Case-D), which reflects that there does exist a gap between the current classification performance of microscope systems and the satisfactory clinical performance.

Conclusion

In this supplementary material, we presented statistical test analysis of the proposed Varifocal-Net in comparison with state-of-the-art methods. Results confirmed the statistical significance in superiority of the proposed method against others. Besides, the classification performance of current commercial chromosome karyotyping systems was measured. The proposed method does achieve much better accuracy in chromosome classification, suggesting its great potential in clinical practice.

B. Supplementary Materials of Pulmonary Airway and Artery-Vein Segmentation Approach in Chapter 5

B.1 Graph-based post-processing

This section presents a detailed comparison of artery-vein segmentation before and after graph-based post-processing. Since the graph-cuts in [Nardelli et al., 2018] cannot be applied to the proposed method, we consider another popular graph-based post-processing: fully connected conditional random fields (dense CRFs) [Krähenbühl and Koltun, 2011]. The three-dimensional (3-D) dense CRFs [Kamnitsas et al., 2017] was adopted to model arbitrarily large voxel-neighborhoods. The Gibbs energy function defined in dense CRFs includes the unary potential term and the pairwise potential term. For each voxel, the CNNs' probability outputs of three categories are used with negative log-likelihood in the unary term. The local smoothness, the similarity of CT intensity, and the distance between the current voxel and each one of the other voxels are represented as a linear combination of Gaussian kernels in the pairwise term. Combined with message passing using Gaussian kernel-based convolutions, the CRFs model lends itself for efficient inference with mean field approximation. The number of iterations for CRF regularization is commonly chosen as 5-10 [Kamnitsas et al., 2017]. However, the trade-off between inference time and accuracy needs to be considered.

In the present study, we conducted CRFs under 3 and 10 iterations. Hyper-parameters are manually tuned to optimal using 6 CT scans randomly chosen from the training set. Detailed hyper-parameter settings [Kamnitsas et al., 2017] are given as follows: $pR = 12.0$, $pC = 12.0$, $pZ = 12.0$, $pW = 1.0$, $bR = 12.0$, $bC = 12.0$, $bZ = 12.0$, $bW = 3.0$, and $bMods = 10.0$. The computational time of dense CRFs is reported in Table B1.

Table B1: Computational time of the graph-based post-processing step for pulmonary artery-vein segmentation.

Item Name		Time (seconds)
Pulmonary Artery-Vein Segmentation Using 55 CT Scans	Graph-based Dense CRFs Post-processing 3 Iterations (Per CT Volume)	515.7±79.9
	Graph-based Dense CRFs Post-processing 10 Iterations (Per CT Volume)	1261.4±150.9

Table B2 demonstrates that after dense CRFs, the performance in mean and median ACC, TPR, BD, and TD all decreased around 3-5%. The reasons behind are two-fold: 1) The Gibbs energy defined in CRFs imposed regularization constraints. The smoothness kernel enforced local smoothness, meaning that thin arteries and veins could be easily smoothed out by their surrounding background. Especially for peripheral branches, these scattered voxels are overwhelmed by the background majority. 2) The appearance similarity kernel enforced homogeneous appearance in CT intensity when voxels in nearby neighborhoods were identically labelled. However, there exists intraclass vari-

Table B2: Results of the graph-based post-processing after the proposed artery-vein segmentation method. For each 3-D CT scan, the original intensity image and CNNs' probability outputs of background, arteries, and veins are taken as inputs to Dense CRF for post-processing. **Union 1**: The union of artery voxels before and after post-processing is kept as final artery predictions. The union of vein voxels before and after post-processing intersects with the set of non-artery voxels to obtain the final vein predictions. The remaining voxels belong to the background. **Union 2**: The union of vein voxels before and after post-processing is kept as final vein predictions. The union of artery voxels before and after post-processing intersects with the set of non-vein voxels to obtain the final artery predictions. The remaining voxels belong to the background.

Method	Params ($\times 10^4$)	ACC-mean [95%-CI] (%)	ACC-median [95%-CI] (%)	TPR (%)	FPR (%)	DSC (%)	BD (%)	TD (%)
Our proposed	1691.0	90.3 [87.7,92.9]	90.9 [87.4,94.6]	90.3±3.5	0.151±0.043	82.4±3.0	85.4±5.3	90.9±3.8
+ Dense CRFs (3 iterations)	–	85.1 [81.7,88.4]	85.8 [81.2,91.5]	85.1±4.5	0.077±0.024	85.2±2.6	79.6±6.5	87.6±4.8
+ Dense CRFs (10 iterations)	–	85.0 [81.7,88.4]	85.7 [81.2,91.5]	85.1±4.5	0.077±0.024	85.2±2.6	79.6±6.5	87.6±4.8
+ Dense CRFs (3 iterations) + Union 1	–	90.5 [87.9,93.1]	90.9 [87.6,94.6]	90.4±3.5	0.151±0.043	82.6±2.9	85.4±5.3	90.9±3.8
+ Dense CRFs (3 iterations) + Union 2	–	90.5 [87.9,93.2]	91.2 [87.7,94.9]	90.6±3.5	0.151±0.043	82.6±2.9	85.7±5.3	91.1±3.8
+ Dense CRFs (10 iterations) + Union 1	–	90.5 [87.9,93.1]	90.9 [87.6,94.6]	90.4±3.5	0.151±0.043	82.6±2.9	85.4±5.3	90.9±3.8
+ Dense CRFs (10 iterations) + Union 2	–	90.5 [87.9,93.2]	91.2 [87.7,94.9]	90.6±3.5	0.151±0.043	82.6±2.9	85.7±5.3	91.1±3.8

ance in the appearance of arteries and veins. The intensity contrast between vessel and background becomes weak and implicit around lung borders. Besides, even for human observers, it is difficult to distinguish between arteries and veins only from intensity values. The feature of CT intensity alone is unreliable for accurate inference.

The reason why FPR was reduced by half and DSC was improved by 2.8% can also be explained by the regularization mechanism of dense CRFs. Since local smoothness and homogeneous appearance were enforced, many isolated false positives were removed and the FPR was reduced. Meanwhile, CRFs enforced connectivity within neighborhoods for thick, large vessel segments whose intensity values are similar and distinct from background. Consequently, major artery and vein segments were correctly segmented and the overall performance indicator DSC increased.

Fig. B1 shows that compared with the proposed method, dense CRFs increased the percentage of both type 2 and type 4 errors by about 4%. Many arteries and veins were incorrectly predicted as background, suggesting that the model's sensitivity to small, peripheral vessels was restricted. It also partly explains why performance in ACC, TPR, BD, and TD declined after dense CRFs.

The comparison of dense CRFs between 3 and 10 iterations shows that no performance gains were achieved when the number of CRFs iterations was increased.

To check if dense CRFs removed many peripheral vessels, we fused the results before and after post-processing into a union set. Arteries and veins were separately processed. To avoid overlapping between the fused arteries and veins, we used one of the following operations: 1) Union 1: intersection between the fused vein voxels and non-artery voxels; 2) Union 2: intersection between the fused artery voxels and non-vein voxels. Table B2 shows that the union did "make up" the loss of peripheral arteries and veins, recovering performance in ACC, TPR, BD, and TD. The fusing trick even improved segmentation a bit, suggesting that CRFs corrected some false predictions on thick branches (e.g., spatial inconsistency).

In summary, considering the performance and computational time, the graph-based post-processing by dense CRFs may not be suitable for the current task.

B.2 Ablation study

This section presents additional ablation study experiments of the proposed method.

Effectiveness of Auxiliary Vessel Segmentation

To validate the effectiveness of auxiliary vessel segmentation, we removed this prediction head from CNNs' architecture and re-trained the proposed method from scratch for 70 epochs. All hyper-parameter settings remained the same.

Table B4 shows that without auxiliary vessel output, the performance declined by around 0.6% in mean ACC, 0.6% in TPR, 0.007% in FPR, 0.8% in DSC, 0.2% in BD, and 0.4% in TD. It demonstrates that the model's ability to differentiate between vessels and background is beneficial to separation of arteries and veins. The auxiliary vessel task provides additional gradients that may assist the optimization of CNNs.

Although the performance gains are minor, the auxiliary vessel segmentation is effective in artery-vein segmentation.

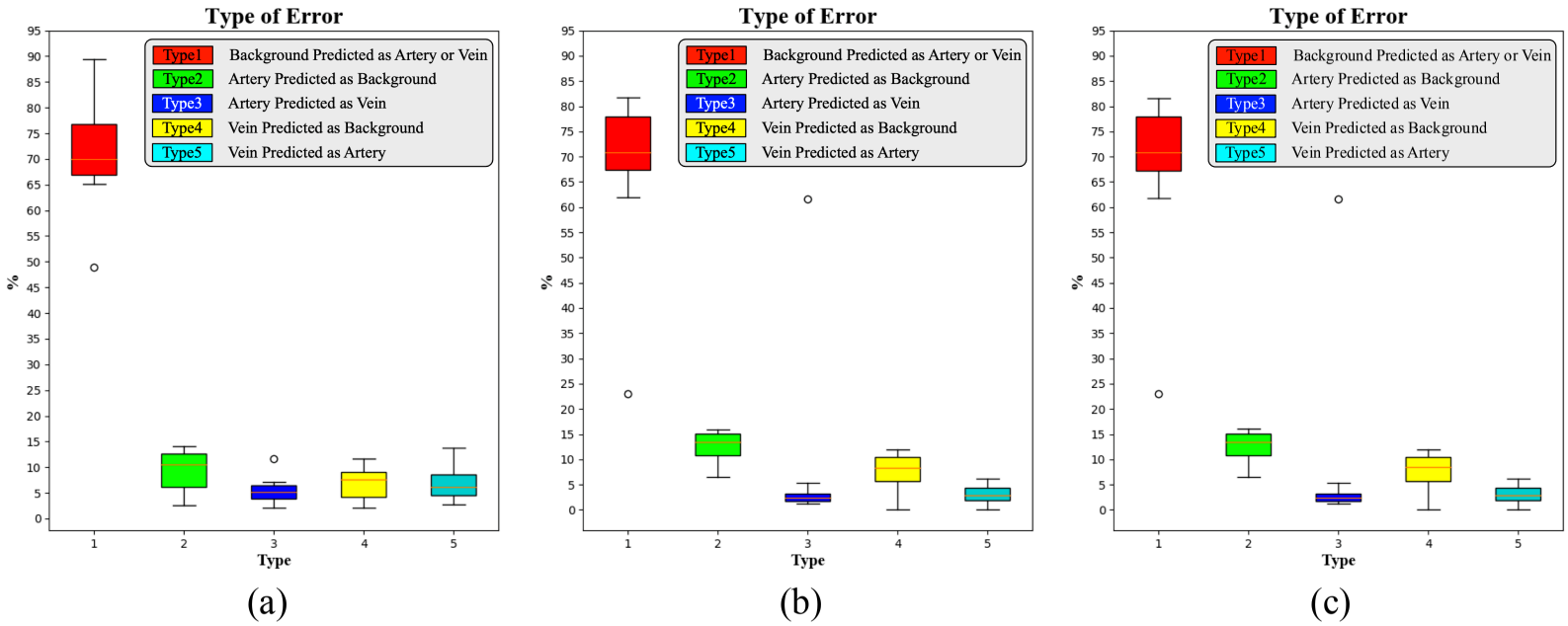


Figure B1: Percentage of different types of errors by (a) the proposed method, (b) dense CRFs (3 iterations), and (c) dense CRFs (10 iterations).

Table B3: Results of ablation study on pulmonary airway segmentation. Both the results under the same binarization threshold and under the same FPR are presented for each method.

Method	Params ($\times 10^4$)	th	BD (%)	TD (%)	TPR (%)	FPR (%)	DSC (%)
Our proposed (w/o coordinate)	422.9	0.5	96.1±4.9	90.3±7.1	93.4±5.0	0.033±0.013	92.7±1.7
Our proposed (w/ coordinate)	423.1		96.2±5.8	90.7±6.9	93.6±5.0	0.035±0.014	92.5±2.0
Our proposed (w/o coordinate)	422.9	0.75	94.0±6.3	86.8±8.7	90.7±6.5	0.021±0.011	93.6±1.5
Our proposed (w/ coordinate)	423.1	0.77	94.3±6.6	86.7±8.5	90.6±6.7	0.021±0.011	93.5±1.6
Baseline (w/o resampling)	411.8	0.5	91.6±9.2	81.3±11.5	87.2±8.6	0.014±0.008	93.7±1.7
Baseline (w/ resampling)			87.1±10.9	75.2±12.7	86.2±7.5	0.017±0.008	92.5±1.7
Baseline (w/o resampling)		0.001	91.6±10.4	81.6±11.2	88.3±8.6	0.021±0.012	92.9±2.3
Baseline (w/ resampling)		0.3	88.9±10.2	77.9±12.2	87.7±7.0	0.021±0.009	92.5±1.6
Our proposed (w/o resampling)	423.1	0.5	96.2±5.8	90.7±6.9	93.6±5.0	0.035±0.014	92.5±2.0
Our proposed (w/ resampling)			93.6±7.3	87.4±9.1	92.3±5.5	0.038±0.015	91.4±1.9
Our proposed (w/o resampling)		0.77	94.3±6.6	86.7±8.5	90.6±6.7	0.021±0.011	93.5±1.6
Our proposed (w/ resampling)		0.81	90.3±9.2	80.5±11.4	87.7±7.1	0.021±0.009	92.3±1.6
Our proposed ($\alpha = 1.0$)	423.1	0.5	95.3±6.3	88.9±7.8	91.9±6.1	0.026±0.012	93.3±1.6
Our proposed ($\alpha = 0.5$)			95.3±5.3	88.4±7.5	92.3±5.6	0.028±0.013	93.1±1.7
Our proposed ($\alpha = 0.1$)			96.2±5.8	90.7±6.9	93.6±5.0	0.035±0.014	92.5±2.0
Our proposed ($\alpha = 0.01$)			94.8±5.9	88.8±8.1	91.9±6.2	0.027±0.013	93.2±1.6
Our proposed ($\alpha = 1.0$)		0.65	93.6±6.7	86.6±8.2	90.4±6.7	0.021±0.011	93.6±1.5
Our proposed ($\alpha = 0.5$)		0.66	93.3±6.4	85.9±8.4	90.5±6.4	0.021±0.011	93.5±1.5
Our proposed ($\alpha = 0.1$)		0.77	94.3±6.6	86.7±8.5	90.6±6.7	0.021±0.011	93.5±1.6
Our proposed ($\alpha = 0.01$)		0.65	92.8±7.6	86.7±8.9	90.5±6.9	0.021±0.011	93.5±1.5
Our proposed ($p = 10$)	423.1	0.5	95.1±5.4	88.0±7.9	91.8±5.9	0.025±0.011	93.4±1.5
Our proposed ($p = 4$)			95.0±6.3	88.1±7.8	91.6±6.3	0.025±0.012	93.3±1.6
Our proposed ($p = 2$)			96.2±5.8	90.7±6.9	93.6±5.0	0.035±0.014	92.5±2.0
Our proposed ($p = 1$)			95.4±5.9	89.2±7.7	92.2±5.9	0.028±0.012	93.2±1.5
Our proposed ($p = 10$)		0.62	94.2±5.9	86.2±8.6	90.5±6.5	0.021±0.010	93.6±1.5
Our proposed ($p = 4$)		0.62	94.1±7.1	86.3±8.5	90.3±6.9	0.021±0.011	93.6±1.6
Our proposed ($p = 2$)		0.77	94.3±6.6	86.7±8.5	90.6±6.7	0.021±0.011	93.5±1.6
Our proposed ($p = 1$)		0.67	94.1±6.5	86.9±8.5	90.6±6.8	0.021±0.010	93.6±1.5

Table B4: Results of ablation study on pulmonary artery-vein segmentation.

Method	Params ($\times 10^4$)	ACC-mean [95%-CI] (%)	ACC-median [95%-CI] (%)	TPR (%)	FPR (%)	DSC (%)	BD (%)	TD (%)
Our proposed (w/o auxiliary task)	1691.0	89.7 [86.2,93.2]	91.1 [87.3,94.8]	89.7±4.7	0.158±0.048	81.6±4.2	85.2±5.9	90.5±4.5
Our proposed (w/ auxiliary task)		90.3 [87.7,92.9]	90.9 [87.4,94.6]	90.3±3.5	0.151±0.043	82.4±3.0	85.4±5.3	90.9±3.8
Our proposed (w/o coordinate)	1690.8	89.0 [85.8,92.3]	90.2 [85.9,93.9]	89.0±4.3	0.151±0.042	81.7±3.9	83.7±6.1	89.5±4.7
Our proposed (w/ coordinate)	1691.0	90.3 [87.7,92.9]	90.9 [87.4,94.6]	90.3±3.5	0.151±0.043	82.4±3.0	85.4±5.3	90.9±3.8
Our proposed ($\alpha = 1.0$)	1691.0	89.2 [86.2,92.1]	90.1 [86.3,94.2]	89.2±3.9	0.148±0.043	81.9±3.5	83.9±5.8	89.8±4.2
Our proposed ($\alpha = 0.5$)		89.8 [86.9,92.7]	90.7 [86.7,94.4]	89.8±3.9	0.158±0.043	81.6±3.5	84.4±5.8	90.2±4.2
Our proposed ($\alpha = 0.1$)		90.3 [87.7,92.9]	90.9 [87.4,94.6]	90.3±3.5	0.151±0.043	82.4±3.0	85.4±5.3	90.9±3.8
Our proposed ($\alpha = 0.01$)		88.7 [85.6,91.7]	89.8 [85.7,93.9]	88.7±4.1	0.144±0.043	81.9±3.8	83.3±5.8	89.5±4.2
Our proposed ($p = 10$)	1691.0	88.8 [85.9,91.8]	89.7 [85.8,93.8]	88.8±3.9	0.141±0.041	82.2±3.5	83.1±5.8	89.4±4.3
Our proposed ($p = 4$)		89.3 [86.4,92.2]	90.3 [85.9,94.2]	89.3±3.9	0.148±0.042	82.1±3.5	83.9±5.8	89.9±4.2
Our proposed ($p = 2$)		90.3 [87.7,92.9]	90.9 [87.4,94.6]	90.3±3.5	0.151±0.043	82.4±3.0	85.4±5.3	90.9±3.8
Our proposed ($p = 1$)		88.8 [85.8,91.8]	89.8 [85.9,93.9]	88.8±4.0	0.140±0.042	82.3±3.6	83.3±5.9	89.5±4.2

Effectiveness of Voxel Coordinate Map

To check effectivity of voxel coordinate map, the proposed method without coordinate information was trained from scratch and evaluated on the same dataset.

Table B3 shows that results of airway segmentation with and without coordinate information were similar. In contrast, performance of artery-vein segmentation without coordinate information degraded in almost all metrics by 0.7%–1.7% (Table B4). Given limited GPU memory, since the number of parameters of the artery-vein segmentation model is much larger than that of the airway segmentation model, the input patch size for the artery-vein task ($64 \times 176 \times 176$) is smaller than that for the airway task ($80 \times 192 \times 304$). In that case, one CT sub-volume patch covers limited context and position information. The coordinates provide supplementary information about each voxel's position relative to the entire CT volume. If such coordinate map is not explicitly used, the model may not learn well the location relationship.

The voxel coordinate map did not affect much airway segmentation. Yet it did improve artery-vein segmentation.

Negative Effects of Isometric Resampling

Isometric resampling was performed to demonstrate its negative effects on airway segmentation. We performed trilinear interpolation in CT and nearest neighbor interpolation in annotations. The resampled data share the same isotropic resolution and slice thickness of 0.625 mm. We re-trained the proposed CNNs on resampled data from scratch.

Results in Table B3 show that performance on resampled data degraded for both baseline and the proposed method. Two reasons are responsible: 1) The model was trained with annotations that mismatched their corresponding CT scans. Some voxels were not labelled correctly due to interpolation. 2) The evaluation metrics calculated with the resampled labels may also be inaccurate. Furthermore, extensive experiments demonstrated that without isometric resampling, CNNs can still learn effective representation of airways. Therefore, it is recommended not to perform isometric resampling.

Hyper-parameter Tuning on α

The hyper-parameter α typically ranges between 0 and 1. If $\alpha = 1$, the attention distillation loss and segmentation loss are weighted equally in the total loss function. If $\alpha = 0$, the distillation loss is set to 0 and the proposed method degenerates to the one without attention distillation. The higher the α is, the more emphasis the model will put on the attention distillation task instead of the segmentation task. Three new α values were tested: $\alpha = 1.0$, $\alpha = 0.5$, and $\alpha = 0.01$. For each α, the proposed airway and artery-vein segmentation method was trained from scratch and other hyper-parameters were kept the same with the original settings.

Table B3 shows that when α was increased or decreased from 0.1, BD, TD, and TPR of the proposed method declined for results both under the same threshold and the same FPR. Table B4 shows that when α was respectively increased or decreased from 0.1, mean and median ACC, TPR, DSC, BD, and TD all decreased but FPR remained similar.

In summary, we believe α should be tuned around 0.1 to achieve a balance between the segmentation loss and the attention distillation loss.

Hyper-parameter Tuning on p

The value of p is typically set greater than or equal to 1. The higher the p is, the more attention is addressed to highly activated regions in feature A_m whose voxels' absolute values are greater than 1. Compared to $p = 1$, the p-th power ($p > 1$) magnifies such differentiation between targets and background. Three new values were tested: $p = 1$, $p = 4$, and $p = 10$. For each p, the proposed airway and artery-vein segmentation method was trained from scratch and other hyper-parameters were kept the same with the original settings.

Table B3 shows that when p was increased or decreased from 2, BD, TD, and TPR of the proposed method declined. If FPR was controlled to be the same, such performance drop was relatively slight. Table B4 shows that when p was respectively increased or decreased from 2, mean and median ACC, TPR, DSC, BD, and TD all decreased but FPR got improved marginally.

In summary, p should not be set too high or too low. The current $p = 2$ worked well because: 1) The task-related regions are intensified properly over their surrounding background. 2) The difference in activation values between thick and thin branches is not that significant. Attention was focused on both large or small airways and vessels to avoid imbalance.

B.3 Conclusion

In this supplementary material, we presented a detailed comparison of graph-based post-processing. The dense CRFs did not perform well on artery-vein segmentation. Ablation study substantiated the effectiveness of auxiliary vessel segmentation and voxel coordinate map. Besides, isometric resampling is not encouraged. In hyper-parameter tuning experiments, the segmentation performance did vary when α and p were respectively changed. But the variation is minor.

FOLIO ADMINISTRATIF

THESE DE L'UNIVERSITE DE LYON OPEREE AU SEIN DE L'INSA LYON

NOM : QIN
(avec précision du nom de jeune fille, le cas échéant)

DATE de SOUTENANCE : 30/06/2021

Prénoms : Yulei

TITRE : Investigation of deep learning methods for classification and segmentation of chromosome and pulmonary images

NATURE : Doctorat

Numéro d'ordre : 2021LYSEI037

Ecole doctorale : Electronique, Électrotechnique, Automatique (EEA) – ED160

Spécialité : Traitement du Signal et de l'Image

RESUME :
Les maladies pulmonaires peuvent causer des dommages mortels à la santé humaine. La tomographie par rayons X (CT) permet d'obtenir les structures pulmonaires et les lésions pour la mesure et le diagnostic. L'avancée de la microscopie et du caryotypage profite à l'étude de la pathogenèse sur la relation entre les anomalies chromosomiques et les maladies pulmonaires. Dans cette thèse, pour aider à l'analyse des maladies pulmonaires, nous étudions des méthodes d'apprentissage en profondeur pour deux objectifs. Le premier est la classification des chromosomes colorés au Giemsa en imagerie microscopique. Le second est la segmentation des voies respiratoires pulmonaires, des artères, des veines et des nodules en CT.

Nous proposons le Varifocal-Net pour la classification simultanée du type et de la polarité des chromosomes via les réseaux de neurones convolutifs (CNN). Il fonctionne de manière robuste pour différentes courbures, formes et motifs de bandes chromosomiques.

Pour la segmentation des nodules, nous proposons une méthode de CNN composé de deux parties pour toutes les textures et tous les environnements des nodules. La première partie consiste à synthétiser des échantillons via un réseau antagoniste génératif (GAN). La deuxième partie vise à développer un modèle de segmentation. Pour les voies respiratoires, leur structure arborescente pose des problèmes de segmentation. Nous proposons AirwayNet pour modéliser explicitement la connectivité entre les voxels voisins. Nous proposons en outre AirwayNet-SE, plus sophistiqué que AirwayNet, en utilisant les caractéristiques des contextes à deux échelles. Enfin, nous proposons une méthode de segmentation des voies respiratoires, des artères et des veines. Pour faire face à des cibles désirées parcimonieux, causées par un sévère déséquilibre des classes, nous présentons les modules de recalibrage des caractéristiques et de distillation de l'attention. L'anatomie a priori est incorporée pour une meilleure différenciation artère-veine.

MOTS-CLÉS : Apprentissage profond, Chromosome, Poumon, Nodule, Bronche, Artère-Veine, Tomographie, Imagerie microscopique, Classification, Segmentation

Laboratoire (s) de recherche : CREATIS

Directeur de thèse: ZHU Yue-Min

Président de jury:

Composition du jury: ZHENG Yuanjie, RUAN Su, DUPONT Florent, LIANG Dong, ZHU Yue-Min, YANG Jie