N°d'ordre NNT : xxx

**THESE de DOCTORAT DE L'UNIVERSITE DE LYON**
opérée au sein de
**L'Institut National des Sciences Appliquées de Lyon**

**Ecole Doctorale** N° 160
**Électronique, Électrotechnique, Automatique (EEA)**

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

Soutenue publiquement le 14/10/2021, par :
**Maryam HAMMAMI**

---

# A priori anatomique et augmentation de données pour la détection multi-organe en imagerie médicale

---

Devant le jury composé de :

| | | | |
|---|---|---|---|
| Caroline Petitjean | Professeur | Université de Rouen | Rapporteure |
| Michel Desvignes | Professeur | Grenoble INP | Rapporteur |
| Véronique Eglin | Professeur | INSA de Lyon | Examinatrice |
| Mireille Garreau | Professeur | Université de Rennes | Examinatrice |
| Denis Friboulet | Professeur | INSA de Lyon | Directeur de thèse |
| Razmig Kéchichian | Maître de conférences | INSA de Lyon | Co-encadrant de thèse |

# Département FEDORA – INSA Lyon - Ecoles Doctorales

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

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

# Remerciements

La réalisation de ce travail n'aurait pas été possible sans de nombreuses personnes à qui je suis reconnaissante.

Avant tout, je tiens à remercier mes superviseurs de thèse, Messieurs Denis Friboulet et Razmig Kéchichian pour leurs conseils et leur confiance durant ces années. Je les remercie pour leur assistance, encouragement et soutien. Leurs commentaires ainsi que leur généreux partage d'expérience tout au long du développement de ce traité ont agrémenté et consolidé son contenu scientifique et sa vigueur. De plus, je leur dois à tous les deux car ils m'ont soutenu moralement sans hésitation dans les moments difficiles de cette thèse.

Je tiens à remercier toutes les personnes du laboratoire CREATIS avec lesquelles j'ai collaboré. Une mention spéciale pour l'équipe "MYRIAD".

En dehors du cercle professionnel, mes sentiments de gratitude vont à mes parents, à qui je suis infiniment redevable : ma mère Leila, mon père Maher, qui croient continuellement en moi, inconditionnellement leur amour et leur soutien malgré la distance et qui m'encouragent à poursuivre cette aventure.

Enfin, je tiens à remercier tout particulièrement ma moitié Hamza, qui ne m'a jamais abandonné, il était toujours présent à mes côtés pendant les innombrables épreuves que j'ai traversées.

# Résumé

La détection d'objet, l'un des problèmes fondamentaux en vision par ordinateur, vise à localiser et à classer les instances d'objets. Elle peut constituer la première étape avant l'application d'autres méthodes de traitement d'images telles que la segmentation et le recalage. En imagerie médicale, elle est utile pour diverses applications, de la planification d'opérations chirurgicales à la recherche de pathologies.

Nous proposons une solution d'apprentissage profond au problème de la détection d'objets dans les images médicales. L'état de l'art nous a conduit à baser nos travaux sur le détecteur "You Only Look Once" (YOLO) qui fournit un bon compromis vitesse/-précision. Malheureusement cette méthode, comme toutes les méthodes d'apprentissage profond, s'avère être sensible à la dimension réduite de l'ensemble d'apprentissage, problème fréquemment rencontré en imagerie médicale car l'étiquetage manuel à réaliser par les experts pour chaque organe est long et coûteux en temps.

Dans ce cadre, notre première contribution a consisté à développer une approche d'augmentation des données basée sur un "Cycle Generative Adversarial Network" (CycleGAN). Nous montrons à partir des résultats expérimentaux obtenus sur des données TDM et IRM que cette augmentation de données permet de régulariser l'apprentissage du détecteur YOLO en conduisant à des performances de détection significativement meilleures. Ces résultats montrent cependant également que cette performance peut encore être améliorée, dans la mesure où ils comportent un certain nombre de détections anatomiquement aberrantes.

Notre deuxième contribution nous a donc conduit à intégrer un a priori dans le processus de détection afin de pénaliser les valeurs aberrantes. Cet a priori est basé sur les relations spatiales existantes entre les structures anatomiques et est intégré sous la forme d'un terme supplémentaire dans la fonction de perte du détecteur YOLO. Les résultats expérimentaux obtenus montrent clairement que cette contrainte joue pleinement son rôle en diminuant significativement les erreurs de détection.

# Table des matières

**Remerciements** iii

**Résumé** v

**Table des matières** vii

**Liste des figures** xi

**Liste des tableaux** xv

**Liste des algorithmes** xvii

**Liste d'abréviations et de symboles** xix

**1 Introduction** 1

**2 État de l'art** 5
2.1 Introduction . . . 5
2.2 Pourquoi la détection en imagerie médicale ? . . . 5
2.3 Imagerie médicale . . . 6
2.3.1 Imagerie par résonance magnétique . . . 6
2.3.2 Tomodensitométrie . . . 7
2.4 Détection d'objet . . . 8
2.5 Les approches de détection d'objet . . . 9
2.5.1 Méthodes traditionnelles . . . 9
2.5.1.1 Approches basées sur la classification . . . 10
2.5.1.2 Approches basées sur la régression . . . 11
2.5.2 Méthodes par apprentissage profond . . . 12
2.5.2.1 Généralités . . . 13
2.5.2.2 Méthodes basées sur les régions . . . 15
2.5.2.3 Méthodes utilisant un seul réseau . . . 17
2.6 Quelle est la méthode adéquate ? . . . 20
2.7 Conclusion . . . 22

**3 Détection multi-organe dans les images médicales** 24
3.1 Introduction . . . 24
3.2 Détection des organes par YOLO . . . 24
3.2.1 Architecture de YOLO . . . 25
3.2.1.1 Couches de convolution . . . 26
3.2.1.2 Couches de détection . . . 26

3.2.1.3 Hyperparamètres . . . 28
3.2.2 Fonction de perte . . . 29
3.3 Protocole expérimental . . . 33
3.3.1 Base de données . . . 33
3.3.2 Métriques d'évaluation . . . 34
3.3.3 Protocole d'entraînement . . . 36
3.4 Résultats . . . 36
3.4.1 Évaluation qualitative . . . 37
3.4.2 Évaluation quantitative . . . 38
3.5 Conclusion . . . 40

**4 Augmentation des données pour la détection multi-organe dans les images médicales 42**
4.1 Introduction . . . 42
4.2 Augmentation des données . . . 43
4.2.1 Méthodes par transformation . . . 43
4.2.2 Méthodes génératives . . . 43
4.3 Synthèse d'images multi-modalités . . . 44
4.3.1 Réseau Antagoniste Génératif . . . 44
4.3.2 CycleGAN . . . 46
4.4 Méthode proposée . . . 49
4.5 Protocole expérimental . . . 50
4.5.1 Métrique . . . 50
4.5.2 Implémentation du modèle . . . 51
4.5.3 Protocole d'entraînement . . . 52
4.6 Résultats . . . 52
4.6.1 Synthèse d'images multi-modalités . . . 52
4.6.1.1 Évaluation qualitative . . . 52
4.6.1.2 Évaluation quantitative . . . 53
4.6.2 Augmentation des données pour la détection multi-organe dans les images médicales . . . 57
4.7 Conclusion . . . 60

**5 A priori anatomique pour la détection multi-organe via YOLO 62**
5.1 Introduction . . . 62
5.2 Méthodes d'intégration d'a priori en apprentissage profond pour la détection 63
5.2.1 Méthodes d'intégration d'a priori implicites . . . 63
5.2.2 Méthodes d'intégration d'a priori explicites . . . 65
5.3 Contrainte d'orientation anatomique . . . 66
5.3.1 Définition de la contrainte . . . 66
5.3.2 Représentation des structures anatomiques . . . 67
5.4 Intégration de la contrainte anatomique . . . 69
5.5 Résultats . . . 71
5.5.1 Évaluation qualitative . . . 71
5.5.2 Évaluation quantitative . . . 72
5.5.2.1 Intégration des contraintes d'orientation anatomiques . . . 73
5.5.2.2 Intégration des contraintes sur les données augmentées . . . 77
5.5.2.3 Synthèse des résultats . . . 79
5.6 Conclusion . . . 80

**6 Conclusion** **82**

**Annexes** **85**

**A Matrice d'orientation anatomique** **85**
A.1 Matrice d'orientation anatomique Supérieur/Inférieur . . . . . 86
A.2 Matrice d'orientation anatomique Antérieur/Postérieur . . . . . 87
A.3 Matrice d'orientation anatomique Gauche/Droit . . . . . 88

**Bibliographie** **89**

# Liste des figures


2.1 Exemples d'images 2D de modalité IRM pour une coupe coronale. Source [Hanbury *et al.* (2012)] . . . 7
a Corps entier IRM T1 avec artefacts liés à l'inhomogénéité du champ magnétique global. . . . 7
b Thorax-Abdomen IRM T1. . . . 7
2.2 Exemples d'images 2D de modalité TDM pour une coupe coronale. Source [Hanbury *et al.* (2012)] . . . 7
a Corps entier TDM contrasté. . . . 7
b Thorax-Abdomen TDM contrasté. . . . 7
2.3 Classification, localisation et détection d'objets. Source [Thomas (2019)] . . . 8
2.4 Les différentes approches de détection d'objet. Source [Zou *et al.* (2019)] . . . 9
2.5 Détection des organes par des forêts aléatoires de classification. Source [Criminisi *et al.* (2009)] . . . 10
2.6 Détection des organes par des forêts aléatoires de régression. Les voxels de l'image votent pour la localisation des reins en utilisant des boîtes englobantes. Source [Gauriau *et al.* (2015)] . . . 12
2.7 Chronologie des méthodes profondes de détection d'objet. Source [Liu *et al.* (2018)] . . . 12
2.8 Représentation d'un perceptron multi-couche sous la forme d'un graphe. Source [Orbach (1962)] . . . 13
2.9 Architecture de Réseau de Neurones Convolutif. Source [LeCun *et al.* (2015)] 14
2.10 Architecture de la méthode R-CNN. Source [Girshick *et al.* (2014)] . . . 15
2.11 Architecture de la méthode Fast R-CNN. Source [Girshick (2015)] . . . 16
2.12 Architecture de la méthode Faster R-CNN. Source [Ren *et al.* (2015)] . . . 17
2.13 La couche de la détection de YOLO. Source [Kathuria (2018)] . . . 18
2.14 Architecture de la méthode SSD. Source [Liu *et al.* (2016)] . . . 19
2.15 Architecture de la méthode RetinaNet. Source [Lin *et al.* (2017)b] . . . 20
2.16 Étude comparative : Vitesse (ms) par rapport à la précision (Ap) sur la base COCO [Lin *et al.* (2014)]. Source [Redmon and Farhadi (2018)] . . . 20
2.17 Architecture de la méthode YOLOv3. Source [Redmon and Farhadi (2018)] 21

3.1 Architecture détaillée de la méthode YOLOv3. Source [De Palma (2020)] . 25
3.2 Détection et application de l'algorithme SNM. . . . 27
3.3 Les 3 échelles et les 3 rapports (hauteur / largeur) des boîtes d'ancrage. . . 28
3.4 Coordonnées de la boîte englobante. Source [Redmon and Farhadi (2017)] . 31
3.5 Courbe précision-rappel $mAp = \frac{1}{2}\sum_{i=1}^{C=2} Ap_i = 0.9535$. Source [Gad (2021)] 35
a Courbe précision-rappel $Ap_1$. . . . 35
b Courbe précision-rappel $Ap_2$. . . . 35
3.6 Détection multi-organe en 2D sur une image TDM axiale. . . . 37

a Vérité terrain. . . . . 37
b Prédiction. . . . . 37
3.7 Détection multi-organe en 2D sur une image IRM axiale. . . . . 37
a Prédiction 1. . . . . 37
b Prédiction 2. . . . . 37

4.1 Illustration d'un réseau antagoniste génératif. Source [Sadrach (2020)] . . . 45
4.2 Application d'un CycleGAN sur des images médicales. . . . . 46
4.3 Architecture du générateur et du discriminateur d'un CycleGAN. Source [Wolf (2018)] . . . . . 47
4.4 Perte antagoniste et perte de cohérence du cycle. Source [Zhu *et al.* (2017)] 48
4.5 Le modèle proposé : CycleGAN (synthèse d'image) + YOLO (détection multi-organe). . . . . 50
4.6 Résultats qualitatifs de la génération de inter-modalités (de l'IRM à l'image TDM). L'image IRM réelle (à gauche), l'image TDM générée (au centre) et l'image IRM reconstruite (à droite). . . . . 53
4.7 Résultats qualitatifs de la génération inter-modalités (de l'image TDM à l'image IRM). L'image TDM réelle (à gauche), l'image IRM générée (au centre) et l'image TDM reconstruite (à droite). . . . . 53
4.8 Premier mode d'évaluation : Calcul du SSIM entre l'image source et l'image reconstruite. . . . . 54
4.9 Mauvaise génération inter-modalités (de l'image IRM à l'image TDM). L'image IRM réelle x (à gauche) et l'image IRM reconstruite (à droite) sont similaires par contre l'image TDM générée (au centre) est de mauvaise qualité. . . . . 55
4.10 Histogramme de la différence d'altitude de coupe. . . . . 57
a Traduction IRM vers TDM. . . . . 57
b Traduction TDM vers IRM. . . . . 57
4.11 Exemples de détections aberrantes. . . . . 59
a Détection du grand psoas dans le thorax pour la modalité TDM. . . 59
b Détection de la vessie dans l'abdomen pour la modalité IRM. . . . . 59

5.1 Les différentes stratégies d'incorporation de connaissance du domaine dans l'apprentissage profond : (a) un extracteur de caractéristiques (b) réglage fin (fine-tuning) sur le jeu de données cible. Source [Xie *et al.* (2021)] . . . . 63
5.2 Structure globale du SRSCN. Source [Yue *et al.* (2019)]. . . . . 65
5.3 Illustration de la contrainte d'orientation anatomique : Supérieur/Inférieur, Gauche/Droit, Antérieur/Postérieur. . . . . 66
a Coupe coronale TDM contrastée. . . . . 66
b Coupe axiale TDM contrastée. . . . . 66
5.4 Illustration d'un graphe d'orientation et de la matrice d'orientation. . . . . 68
a Graphe d'orientation pour les relations Supérieur/Inférieur sur 6 organes présents sur une coupe coronale d'une image TDM. . . . . 68
b Graphe d'orientation pour les relations Antérieur/Postérieur sur 5 organes présents sur une coupe sagittale d'une image TDM. . . . . 68
c Graphe d'orientation pour les relations Gauche/Droit sur 5 organes présents sur une coupe axiale d'une image TDM. . . . . 68
5.5 Illustration du fonctionnement de l'entropie croisée binaire. . . . . 71
a Graphe d'orientation pour les relations Gauche/Droit sur 3 organes présents sur une coupe coronale d'une image TDM. . . . . 71

b Les cellules d'une image avec la cellule i qui se trouve à droite et la cellule j qui se trouve à gauche. . . . . 71
5.6 Exemple de correction de fausse détection après l'intégration des contraintes d'orientation pour la modalité TDM. . . . . 72
a Détection du poumon droit dans l'abdomen. . . . . 72
b Correction de la détection du poumon droit. . . . . 72
5.7 Exemple de correction de fausse détection après l'intégration des contraintes d'orientation pour la modalité IRM. . . . . 72
a Détection de la vessie dans l'abdomen. . . . . 72
b Correction de la détection de la vessie. . . . . 72

# Liste des tableaux


3.1 Détails de la base VISCERAL Gold : la modalité, l'anatomie, le contraste, le nombre de volumes, les annotations des structures anatomiques et les dimensions de volumes. . . . . 33
3.2 Distance moyenne pour la modalité TDM pour les boîtes englobantes obtenues à partir des coupes axiales et coronales. . . . . 39
3.3 Distance moyenne pour la modalité IRM pour les boîtes englobantes obtenues à partir des coupes axiales et coronales. . . . . 39

4.1 Choix du générateur G du CycleGAN (Traduction de IRM vers TDM). . . 51
4.2 Mode d'évaluation 1 : SSIM moyen obtenu pour chaque patient pour une traduction TDM vers IRM et IRM vers TDM. . . . . 54
a Traduction TDM vers IRM. . . . . 54
b Traduction IRM vers TDM. . . . . 54
4.3 Mode d'évaluation 2 : SSIM et différence d'altitude de coupe (DAC) moyens obtenus pour chaque patient pour une traduction TDM vers IRM et IRM vers TDM. . . . . 56
a Traduction TDM vers IRM. . . . . 56
b Traduction IRM vers TDM. . . . . 56
4.4 Comparaison des résultats de détection obtenus avec YOLO sur la base de données initiale et sur la base de données augmentée via le CycleGAN. Les résultats sont donnés en termes de distance moyenne pour chaque organe et chaque modalité. . . . . 58
a YOLO vs YOLO+CycleGAN pour la modalité TDM. . . . . 58
b YOLO vs YOLO+CycleGAN pour la modalité IRM. . . . . 58

5.1 Comparaison des résultats de détection obtenus à partir des coupes axiales et coronales sur la base de données VISCERAL Gold, pour YOLO sans contrainte, YOLO en intégrant une seule contrainte anatomique et YOLO en intégrant deux contraintes anatomiques. Les résultats sont donnés en termes de distance moyenne [mm] pour chaque organe pour la modalité TDM. . . . . 75
a YOLO vs YOLO+GD et YOLO+GD+AP, en coupes axiales. . . . . 75
b YOLO vs YOLO+GD et YOLO+GD+SI, en coupes coronales. . . . 75
5.2 Comparaison des résultats de détection obtenus à partir des coupes axiales et coronales sur la base de données VISCERAL Gold, pour YOLO sans contrainte, YOLO en intégrant une seule contrainte anatomique et YOLO en intégrant deux contraintes anatomiques. Les résultats sont donnés en termes de distance moyenne [mm] pour chaque organe pour la modalité IRM. 76
a YOLO vs YOLO+GD et YOLO+GD+AP, en coupes axiales. . . . . 76

b YOLO vs YOLO+GD et YOLO+GD+SI, en coupes coronales. . . . 76
5.3 Comparaison des résultats de détection obtenus à partir des coupes axiales sur la base de données VISCERAL Gold, YOLO sans contrainte, YOLO sur la base de données augmentée via le CycleGAN, YOLO+CycleGAN en intégrant une seule contrainte anatomique et YOLO+CycleGAN en intégrant deux contraintes anatomiques. Les résultats sont donnés en termes de distance moyenne [mm] pour chaque organe pour la modalité TDM. . . . . 78
5.4 Comparaison des résultats de détection obtenus à partir des coupes axiales sur la base de données VISCERAL Gold, YOLO sans contrainte, YOLO sur la base de données augmentée via le CycleGAN, YOLO+CycleGAN en intégrant une seule contrainte anatomique et YOLO+CycleGAN en intégrant deux contraintes anatomiques. Les résultats sont donnés en termes de distance moyenne [mm] pour chaque organe pour la modalité IRM. . . . . 79
5.5 Comparaison de la distance moyenne sur l'ensemble des organes pour les coupes axiales et les modalités IRM et TDM. . . . . 79

# Liste des algorithmes

1 Suppression de Non Maxima. . . . . 27

# Liste d'abréviations et de symboles

## Abréviations

| | |
|---|---|
| ACP | Analyse en Composantes Principales |
| AdaBoost | Adaptive Boosting |
| Ap | Précision moyenne (Average precision) |
| BCE | Entropie Croisée Binaire (Binary Cross Entropy) |
| COCO | Common Objects in COntext |
| DAC | Différence d'Altitude de Coupe |
| DPM | Modèle à Parties Déformables (Deformable Parts Model) |
| Fast R-CNN | Fast Region-based Convolutional Neural Network |
| Faster R-CNN | Faster Region-based Convolutional Neural Network |
| FCN | Fully Convolutional Network |
| FL | Erreur Focale |
| FN | Faux Négatif (False Negative) |
| FP | Faux Positif (False Positive) |
| GAN | Réseau Antagoniste Génératif (Generative Adverserial Networks) |
| GPU | Processus graphique (Graphical Processing Unit) |
| HOG | Histogramme de Gradient Orienté |
| IoU | Intersection sur Union (Intersection over Union) |
| IRM | Imagerie par Résonance Magnétique |
| k-means | Le partitionnement en k-moyennes |
| k-NN | k plus proches voisins (k-Nearest Neighbors) |
| LITS | LIver Tumor Segmentation challenge |
| LsGAN | Least Squares Generative Adversarial Networks |
| mAp | Précision moyenne globale (mean Average precision) |
| mSSIM | Moyenne SSIM |
| R-CNN | Region-based Convolutional Neural Network |
| ReLU | Unité de rectification linéaire |
| ResNET | Residual Networks |

| | |
|---|---|
| RMN | Résonance Magnétique Nucléaire . . . 6 |
| RNC | Réseau de Neurones Convolutif (Convolutional Neural Networks) . . . 13 |
| RoI | Region of Interest . . . 16 |
| RPN | Réseau de Proposition de Région (Region Proposal Network) . . . 15 |
| SC | Spatial Constraint . . . 66 |
| SCN | Spatial Constraint Network . . . 66 |
| Seg | Segmentation . . . 66 |
| SNM | Suppression de Non Maxima (Non maximum suppression) . . . 26 |
| SPP | Spatial Pyramid Pooling . . . 9 |
| SR | Shape Regularization . . . 66 |
| SRNN | Shape Reconstruction Neural Network . . . 66 |
| SRSCN | Shape Reconstruction Spatial Constraint Network . . . 66 |
| SSD | Single Shot object Detectors . . . 17 |
| SSIM | Structural Similarity Index Measure . . . 50 |
| SVM | Séparateur à Vaste Marge . . . 11 |
| T1 | Temps de relaxation longitudinale . . . 6 |
| T2 | Temps de relaxation transversal . . . 34 |
| TDM | Tomodensitométrie . . . 5 |
| TEP | Tomographie par émission de positons . . . 64 |
| TSV | Teinte Saturation Valeur . . . 36 |
| VGG | Oxford Vision Geometry Group . . . 19 |
| VISCERAL | VISual Concept Extraction challenge in RAdioLogy . . . 33 |
| VOC | Visual Object Classes Challenge . . . 11 |
| VP | Vrai Positif (True Positive) . . . 35 |
| YOLO | You Only Look Once . . . 17 |

# Lettre grecque

| | |
|---|---|
| $\lambda_{cont}$ | paramètre de pondération du terme de contrainte d'orientation anatomique . . . 70 |
| $\lambda_{coord}$ | paramètre de pondération de la fonction de perte de position prédite . . . 30 |
| $\lambda_{noobj}$ | paramètre de pondération de la fonction de perte associée à l'indice de confiance 30 |
| $\lambda_{pos}$ | paramètre de pondération de la fonction de perte de décalage de positionnement de la cible . . . 31 |
| $\mu_A$ (resp. $\mu_B$) | intensité moyenne de A (resp. B) . . . 51 |
| $\sigma(x)$ | fonction sigmoïde . . . 31 |
| $\sigma_A$ (resp. $\sigma_B$) | écart type de A (resp. B) . . . 51 |
| $\sigma_{AB}$ | covariance de A et B . . . 51 |

# Lettres latines

| | |
|---|---|
| $\mathbb{1}_{ij}^{noobj}$ | complément de $\mathbb{1}_{ij}^{obj}$ ..... 30 |
| $(c_x, c_y)$ | coordonnées de la cellule ..... 31 |
| $(p_w, p_h)$ | largeur et hauteur de la boîte d'ancrage ..... 31 |
| $\hat{p}_i(C)$ | représente la probabilité de classe $C$ dans la cellule $i$ ..... 30 |
| $\mathbb{1}_{ij}^{obj}$ | 1 si un objet est détecté dans la cellule $i$ et la $j^{\text{ème}}$ boîte englobante, et 0 sinon ..30 |
| Go | Gigaoctet ..... 38 |
| X | ondes électromagnétiques de hautes fréquences de l'ordre de $10^{16}$ Hz à $10^{20}$ Hz ... 6 |
| $^T$ | Transposé ..... 13 |
| $B$ | Boîtes englobantes ..... 17 |
| $b_x, b_y, b_w, b_h$ | coordonnées de la boîte englobante finale ..... 31 |
| $C$ | nombre de classes ..... 18 |
| $k_1; k_2$ | constante positive ..... 51 |
| $p_0$ | indice de confiance : l'objet est présent ou non dans une image ..... 17 |
| $p_1, p_2, \cdots, p_c$ | probabilité de chaque classe d'objets ..... 17 |
| $C_i$ | indice de confiance dans la cellule $i$ ..... 30 |
| 1D | Unidimensionnel ..... 14 |
| 2D | Bidimensionnel ..... 6 |
| 3D | Tridimensionnel ..... 6 |
| AP | Antérieur/Postérieur ..... 69 |
| GD | Gauche/Droit ..... 69 |
| min | minute ..... 21 |
| mm | millimètre ..... 38 |
| R | étendue de mesure ..... 51 |
| s | seconde ..... 17 |
| SI | Supérieur/Inférieur ..... 69 |
| D | Discriminateur ..... 44 |
| G | Générateur ..... 44 |

Chapitre 1

# Introduction

## Objectifs de la thèse

L'imagerie médicale joue un rôle important dans différentes applications cliniques notamment les procédures médicales utilisées pour la détection des structures anatomiques ou des lésions. La détection d'objet est un problème fondamental dans le domaine de la vision par ordinateur. Ces dernières années, elle a connu une croissance exponentielle avec le développement rapide de nouveaux outils et de nouvelles techniques [Zaidi *et al.* (2021)]. Souvent elle constitue le point de départ pour des algorithmes de plus haut niveau tels que la segmentation, la compréhension de scènes, le recalage, le suivi, reconstruction et la reconnaissance d'objets.

L'objectif de la détection d'objets est de déterminer la classe de chaque instance d'objet et son emplacement spatial. Elle est une condition préalable à plusieurs applications médicales telles que les procédures radiologiques et les interventions chirurgicales. Cependant, la détection des organes pour des images médicales est une tâche difficile pour plusieurs raisons. Par exemple, les artefacts liés à la modalité d'imagerie et aux conditions d'acquisition tels que le bruit et l'inhomogénéité du champ magnétique, dégradent la qualité des images et réduisent leur contenu informationnel. En parallèle, en imagerie médicale, il est difficile d'obtenir suffisamment de données. Cela est dû à la protection et la confidentialité des données médicales et au fait que l'étiquetage manuel par des experts d'un ensemble de données peut prendre un temps considérable. La difficulté d'obtenir des données annotées a limité leur application en imagerie médicale, en particulier la détection basée sur l'apprentissage profond.

L'imagerie médicale connaît actuellement une effervescence, grâce au développement des méthodes d'apprentissage profond basées sur les réseaux de neurones. La détection doit être efficace en termes de temps lors du traitement de grands ensembles de données.

C'est dans ce contexte que se situe mon travail de thèse qui porte sur la détection multi-organe pour des images médicales. Il traite les aspects méthodologiques suivants :

- l'augmentation des données à partir d'un modèle génératif pour la détection multi-organe afin de pallier à la rareté des données. Ce modèle génère des images synthétiques d'une modalité cible à partir d'une modalité source.
- l'intégration d'un a priori dans la fonction de perte du détecteur afin d'assurer la cohérence anatomique de la détection. Cet a priori utilise l'orientation, la relation spatiale entre les structures anatomiques.

## Organisation du manuscrit

Le manuscrit est composé de six chapitres. Le présent Chapitre 1 d'introduction donne les motivations et la méthodologie de la thèse, ainsi que l'organisation détaillée du manuscrit. Le Chapitre 2 est consacré à la présentation de l'état de l'art en détection d'images médicales. Nous présentons les principales propriétés des images médicales de modalité IRM et TDM, après avoir donné quelques exemples d'utilisations cliniques de la détection multi-organe. Ensuite, nous présentons une revue des méthodes de détection d'objet qui détaille les approches traditionnelles et profondes. Nous choisissons YOLOv3 [Redmon and Farhadi (2018)] comme détecteur. Il a été démontré qu'il offre un bon compromis entre précision et rapidité pour les images naturelles par rapport à d'autres détecteurs profonds. Les trois autres chapitres présentent nos principales contributions.

Dans le Chapitre 3, nous réalisons une détection multi-organe pour les images médicales en utilisant le détecteur YOLOv3. Nous commençons par détailler l'architecture et la fonction de perte du détecteur. Ensuite, nous présentons le protocole expérimental mis en place (la base de données, les métriques d'évaluation, etc.). Les performances de l'algorithme sont ensuite évaluées sur un jeu de données de 20 patients pour les deux modalités IRM et TDM. Enfin, nous commentons les résultats d'application de YOLOv3 sur les images médicales.

Dans le Chapitre 4, nous proposons une nouvelle approche d'augmentation de données pour la détection multi-organe afin d'améliorer les performances du détecteur. Dans un premier temps, nous dressons un état de l'art de techniques générales d'augmentation de données. Ensuite, nous nous intéressons à l'augmentation de données en utilisant un modèle génératif appelé CycleGAN [Zhu *et al.* (2017)]. Ce dernier a pour but de générer des images synthétiques d'une modalité cible (exemple la modalité IRM) à partir d'une modalité source (exemple la modalité TDM). Ensuite, nous évaluons ces images synthétiques à l'aide d'une métrique de similarité. Nous ajoutons par la suite ces images aux jeux de données d'entraînement du détecteur. Nous terminons par une évaluation de cette approche.

Dans le Chapitre 5, nous proposons de réduire les valeurs aberrantes de détection par l'intégration d'a priori dans le détecteur. Nous commençons par dresser un état de l'art des techniques d'intégration de l'a priori dans les réseaux profonds. Nous avons choisi la relation spatiale entre les structures anatomiques comme a priori, puisque elle est invariante d'un patient à un autre. Nous proposons d'intégrer cette contrainte d'orientation dans la fonction de perte du détecteur YOLOv3 et nous évaluons ensuite les résultats de cette approche.

Le Chapitre 6 est la conclusion générale et résume les principales réalisations et les perspectives de notre travail.

# Chapitre 2

# État de l'art

## 2.1 Introduction

La vision par ordinateur permet d'analyser, traiter et comprendre une ou plusieurs images obtenues par un système d'acquisition. Les applications de la vision par ordinateur incluent la détection d'objets, dans laquelle s'inscrit le présent sujet de thèse. Elle consiste à localiser et classer les régions de l'image numérique.

Dans ce chapitre, nous explorons l'utilisation de la détection des structures anatomiques en imagerie médicale multi-modalité. Nous commençons par présenter brièvement en Section 2.2 les applications les plus courantes de la détection des organes, puis les images sur lesquelles nous allons faire le traitement (Section 2.3). Ces images sont d'origine médicale et acquises à l'aide des technologies d'imagerie par résonance magnétique (IRM) et tomodensitométrie (TDM). Dans la Section 2.4, nous définissons la détection d'objet. Nous donnons ensuite une description des différentes méthodes de l'état de l'art (Section 2.5). Ces méthodes sont présentées selon deux axes, les méthodes traditionnelles (Section 2.5.1) et les méthodes basées sur l'apprentissage profond (Section 2.5.2).

## 2.2 Pourquoi la détection en imagerie médicale ?

La détection automatique de plusieurs organes dans les images médicales peut fournir des informations sémantiques importantes, qui peuvent être utilisées dans diverses applications hospitalières.

La détection des organes joue un rôle important dans la pratique clinique. Elle est une condition préalable à de nombreuses applications comme la planification, l'intervention

thérapeutique et les procédures radiologiques, telles que le dépistage et le diagnostic des patients par la localisation de structures ou de lésions anatomiques.

La détection des organes peut ainsi être utilisée comme initialisation pour de nombreuses tâches d'analyse automatique d'images médicales telles que la segmentation et le recalage des structures anatomiques. L'estimation préliminaire correcte de la position de l'organe peut grandement améliorer la précision des procédures de traitements ultérieurs. La détection d'objets est utilisée aussi comme une application plus directe pour localiser plusieurs organes dans une image 3D pour aider un médecin à naviguer dans un volume.

## 2.3 Imagerie médicale

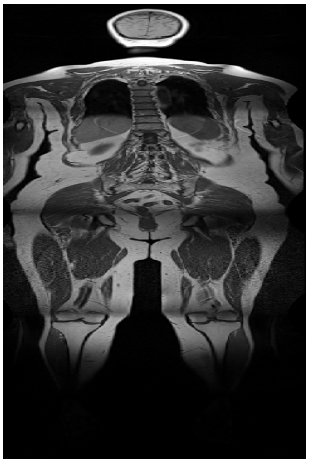
L'imagerie médicale apparue au 20$^{e}$ siècle est la base de la révolution de la médecine. Elle constitue un vaste domaine qui a émergé grâce aux progrès de l'instrumentation, des techniques d'acquisition, de la reconstruction d'image et du traitement du signal. Les avancées dans l'une ou l'autre de ces disciplines contribuent à améliorer la recherche ou la gestion clinique. Par exemple, l'évolution des techniques d'acquisition permet une meilleure visualisation de l'anatomie. Cela permet d'obtenir des images du corps humain à partir de phénomènes physiques comme la radioactivité, l'absorption des rayons X, les ondes ultra-sonores et la résonance magnétique. Grâce au développement que le domaine informatique a connu, ces phénomènes physiques ont pu être transformés en données exploitables constituant la base de l'imagerie médicale.

Dans cette section, nous allons présenter deux types d'acquisitions, l'IRM et la TDM pour lesquelles nous allons étudier la détection d'objets.

### 2.3.1 Imagerie par résonance magnétique

L'imagerie par résonance magnétique figure parmi les technologies les plus utilisées pour fournir des images en 2D ou en 3D des organes du corps humain (Figure 2.1). Elle permet d'obtenir une vue pour l'ensemble des organes du corps. Elle est basée sur le phénomène de la résonance magnétique. Ce phénomène est lié au comportement de plusieurs noyaux atomiques lors de l'application d'un champ magnétique externe, et a été présenté par Felix Bloch et Edward Mills Purcell en 1946. En se basant sur la capacité de la spectroscopie RMN de détecter les tumeurs, Raymond Vahan Damadian a proposé en 1969 d'intégrer la RMN dans le domaine médical. Depuis, l'imagerie par résonance magnétique a fait son apparition en 1977, date à laquelle la première image d'un corps humain a été obtenue *in vivo*.

L'avantage de la modalité IRM est le bon contraste, permettant de mieux différencier des tissus de compositions différentes et de bien visualiser les tissus mous. Par contre, elle présente des artefacts d'acquisition, tels que l'inhomogénéité du champ magnétique global comme le montre la Figure 2.1a.



(a) Corps entier IRM T1 avec artefacts liés à l'inhomogénéité du champ magnétique global.

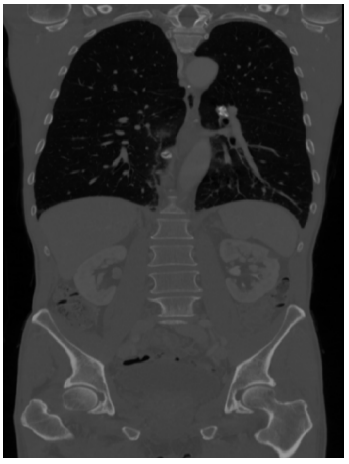
(b) Thorax-Abdomen IRM T1.

FIGURE 2.1 – Exemples d'images 2D de modalité IRM pour une coupe coronale. Source [Hanbury *et al.* (2012)]

### 2.3.2 Tomodensitométrie

La tomodensitométrie aussi fait partie des technologies les plus utilisées dans le domaine de l'imagerie médicale. Elle est basée sur les propriétés des rayons X découvertes par Wilhelm Röntgen en 1895. Cette technique repose sur l'absorption de ces rayons par les tissus. Elle fournit des images ciblées en coupes fines du corps. Les images qui en résultent sont alors traitées par ordinateur pour effectuer une reconstruction bidimensionnelle ou tridimensionnelle. Cette technique permet d'obtenir un contraste différent en fonction de la composition des objets. Cela permet de différencier les organes du corps humain et d'analyser ainsi leur état (identifier ou détecter des anomalies) comme le montre l'exemple de la Figure 2.2.

(a) Corps entier TDM contrasté.



(b) Thorax-Abdomen TDM contrasté.

FIGURE 2.2 – Exemples d'images 2D de modalité TDM pour une coupe coronale. Source [Hanbury *et al.* (2012)]

L'avantage de la modalité TDM est qu'elle offre un fort contraste des repères anatomiques tels que les os, les poumons, le foie et les vaisseaux sanguins et elle est robuste au bruit. Par contre, elle présente un contraste limité pour les tissus mous, contrairement à la modalité IRM.

## 2.4 Détection d'objet

Face à l'augmentation du nombre d'examens en imagerie médicale, la détection d'objets joue un rôle essentiel dans l'accompagnement des fonctions des praticiens. Elle se décompose en deux étapes, la classification et la localisation des objets dans l'image :

- Localisation : elle consiste à déterminer la position spatiale d'un objet détecté.
- Classification : elle consiste à identifier la présence d'une instance d'une classe (e.g. d'un organe ou d'une pathologie) dans une image numérique.

La classification et la localisation sont les deux principaux axes sur lesquels la détection d'objets repose. La détection d'objets est appliquée dans différents domaines comme l'automobile (voitures autonomes...) et la sécurité (détection de visage... ). Elle est utilisée lorsqu'on a des images contenant de nombreux objets de différentes classes comme le montre la Figure 2.3.

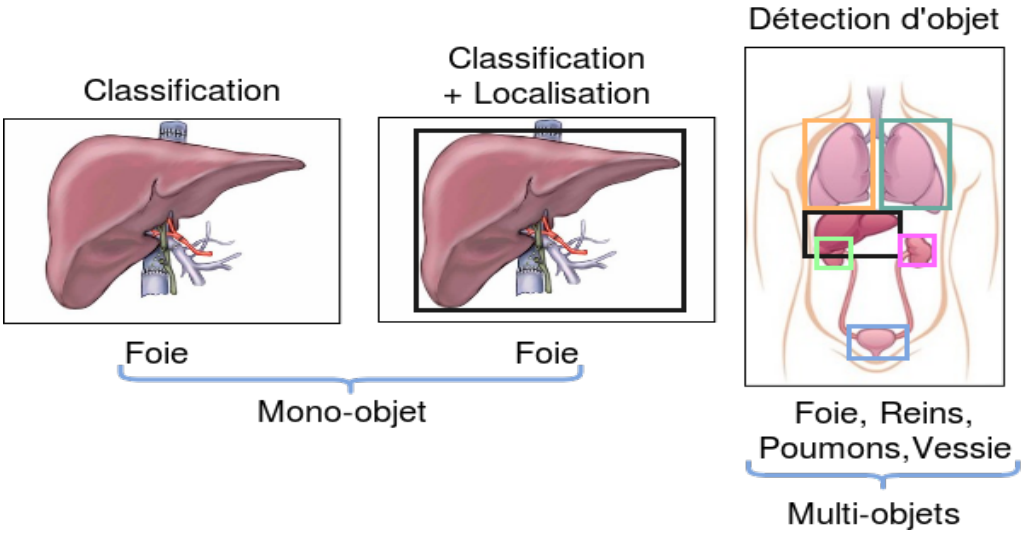



FIGURE 2.3 – Classification, localisation et détection d'objets. Source [Thomas (2019)]

Le principe de la détection d'objets est le suivant : pour une image donnée, on cherche les régions de celle-ci qui pourraient contenir un objet puis pour chacune des régions découvertes, on extrait l'objet et on le classe à l'aide d'un classifieur et on le localise.

## 2.5 Les approches de détection d'objet

La détection d'objet a beaucoup progressé pendant les 20 dernières années [Zou *et al.* (2019), Zaidi *et al.* (2021)]. La Figure 2.4 montre les différentes approches de détection d'objets ; la première branche présente les méthodes traditionnelles (non profondes) (Section 2.5.1) et la deuxième branche présente les méthodes basées sur l'apprentissage profond (Section 2.5.2).

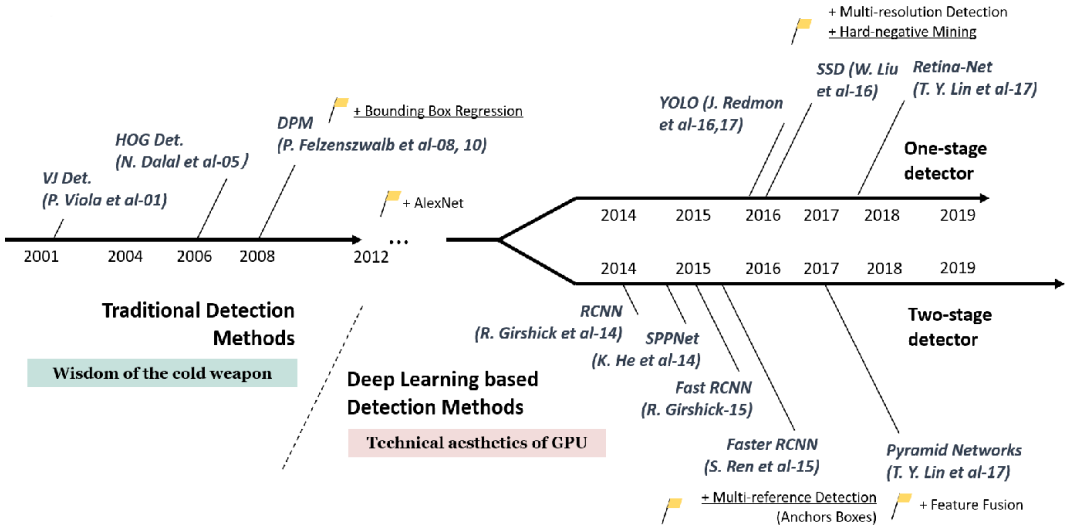



FIGURE 2.4 – Les différentes approches de détection d'objet. Source [Zou *et al.* (2019)]

### 2.5.1 Méthodes traditionnelles

Dans cette section, nous présentons les méthodes traditionnelles (non profondes), qui sont basées sur l'apprentissage automatique. L'apprentissage automatique est un vaste domaine à l'intersection des statistiques, des probabilités et de l'informatique. Il y a deux types d'apprentissage : l'apprentissage supervisé et non supervisé.

L'apprentissage supervisé comprend toutes les méthodes de prédiction utilisant des données annotées. Les arbres de décision [Breiman (2001)] et les séparateurs à vaste marge (SVM) [Hasan and Boris (2006)] figurent parmi les classifieurs les plus employés pour la détection des organes dans les images médicales.

L'apprentissage non supervisé comprend toutes les méthodes qui n'utilisent pas d'annotations dans le but de trouver des similitudes (clustering) dans les données ou de réduire la dimensionnalité de l'espace de représentation. En traitement d'image médicale, les méthodes les plus connues incluent la méthode des k plus proches voisins (k-NN) [Altman (1992)] et la méthode de l'analyse en composantes principales (ACP) [Wold *et al.* (1987)].

Dans cette section, nous présentons les différentes approches traditionnelles de la détection d'objet selon deux grands types : les approches basées sur la classification et les approches basées sur la régression.

#### 2.5.1.1 Approches basées sur la classification

La plupart des algorithmes de détection basés sur la classification impliquent la mise en place d'un classifieur dont le rôle est de prédire à quel organe appartient le voxel en fonction des caractéristiques locales. Dans [Criminisi *et al.* (2009)], Criminisi et al. parviennent à localiser des organes dans les volumes de TDM en utilisant les forêts décisionnelles (Figure 2.5) avec un contexte spatial à long terme. Les auteurs [Dolejsi *et al.* (2008)] utilisent un AdaBoost asymétrique [Viola and Jones (2001)a] pour détecter les nodules du poumon pour des images de modalité TDM.

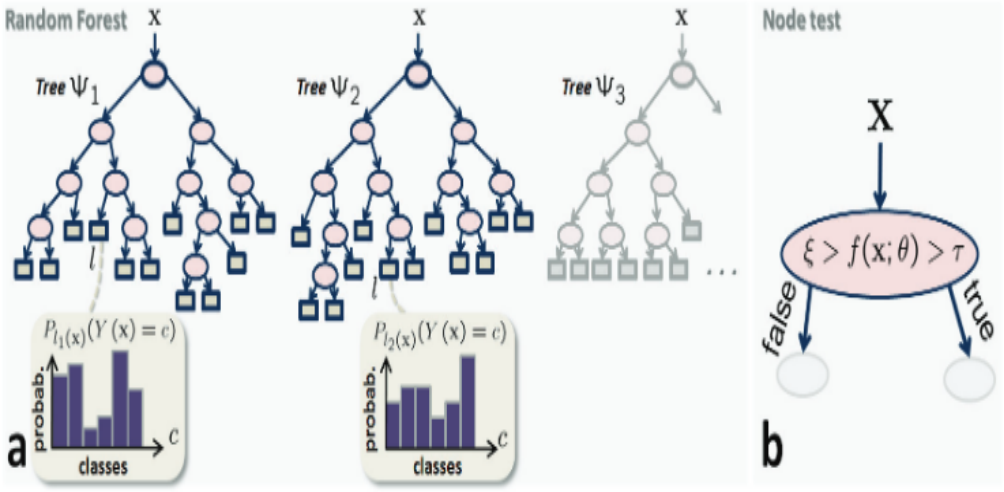



FIGURE 2.5 – Détection des organes par des forêts aléatoires de classification. Source [Criminisi *et al.* (2009)]

Par la suite, nous présentons les principaux descripteurs utilisés dans les approches basées sur la classification.

**Méthode de Viola et Jones** La méthode de Viola et Jones [Viola and Jones (2001)b] a été proposée par Paul Viola et Michael Jones en 2001. Elle est l'une des plus anciennes méthodes. Elle permet la détection et localisation des objets dans une image en temps réel. Elle permet non seulement la détection des visages, son intérêt d'origine, mais aussi la détection d'objets comme par exemple les voitures.

Cette méthode a permis d'introduire des notions comme la classification construite comme une cascade de classifieurs boostés qui ont été très utilisés par la suite en vision par ordinateur. Pour entraîner un classifieur à l'aide du détecteur de Viola et Jones, il est nécessaire de disposer d'un grand nombre d'exemples d'objets (de centaines à des milliers d'objets). Une fois son apprentissage réalisé, ce classifieur est utilisé pour détecter la présence de l'objet dans l'image en la parcourant de manière exhaustive à toutes les positions possibles et à toutes les échelles possibles.

**Histogramme de Gradient Orienté** L'histogramme de Gradient Orienté (HOG) a été proposé en 2005 par N. Dalal et B.Triggs [Dalal and Triggs (2005)]. C'est un descripteur de caractéristique largement déployé dans le domaine de la détection d'objet. HOG est basé sur l'analyse de l'orientation des gradients d'intensités locales et de leurs distributions afin de décrire la forme d'un objet local. Comme pour la plupart des méthodes de détection, l'image d'entrée est divisée en plusieurs cellules. Dans chaque cellule on calcule le vecteur gradient de chaque pixel, ainsi que sa magnitude et sa direction. Le vecteur caractéristique HOG final est la concaténation de tous les vecteurs de cellule. La phase finale est la classification qui utilise ensuite un algorithme d'apprentissage classique (par exemple un séparateur à vaste marge (SVM)) [Hasan and Boris (2006)].

**Modèle basé sur les parties déformables** Remportant les compétitions de détection VOC-07, -08 et -09, le modèle à parties déformables (DPM) se présente comme l'un des meilleurs modèles de détection. Le détecteur DPM a été proposé à l'origine par P. Felzenszwalb [Felzenszwalb *et al.* (2008)] en 2008 comme une extension du détecteur HOG. Le DPM est composé d'un modèle racine qui représente le filtre à faible résolution et de filtres partiels qui représentent la haute résolution. Pour la détection, ces filtres sont appliqués sur toute l'image.

Ensuite, R. Girshick [Felzenszwalb *et al.* (2010)] a apporté diverses améliorations. Les améliorations principales consistent à configurer automatiquement les filtres partiels par une méthode d'apprentissage.

#### 2.5.1.2 Approches basées sur la régression

Dans le domaine des applications médicales, les chercheurs utilisent aussi des solutions de détection d'organes basées sur la régression. Les méthodes basées sur la classification prennent souvent le contexte local au prix d'une analyse complète des images. Contrairement à ces méthodes, les méthodes basées sur la régression reposent davantage sur le contexte global, afin d'atteindre une vitesse plus rapide (l'analyse complète de l'image n'est pas nécessaire) [Gauriau *et al.* (2015)].

Ainsi dans [Zhou *et al.* (2007)], les auteurs ont utilisé une méthode basée sur la régression dite "ridge". Cette méthode est introduite pour détecter et localiser le ventricule gauche dans des images ultra-sonores cardiaques 2D. Cette approche prédit la position relative, la taille et la direction du ventricule gauche sur la base de caractéristiques basées sur les ondelettes de Haar. Elle a montré des résultats impressionnants de détection sur les séquences échocardiographiques. Criminisi et al. [Criminisi *et al.* (2010), Criminisi *et al.* (2013)] ont proposé une méthode de régression basée sur la forêt aléatoire pour localiser les organes en TDM 3D. Dans [Cuingnet *et al.* (2012), Gauriau *et al.* (2015)], les auteurs ont montré que l'utilisation des forêts de régression qui sont appliquées en cascade de l'échelle globale à l'échelle locale améliore la détection des organes en TDM 3D (voir Figure 2.6).

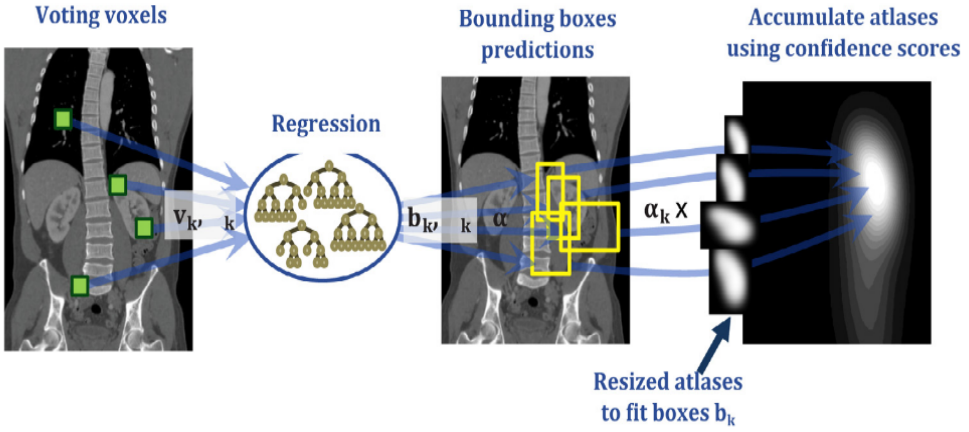



FIGURE 2.6 – Détection des organes par des forêts aléatoires de régression. Les voxels de l'image votent pour la localisation des reins en utilisant des boîtes englobantes. Source [Gauriau *et al.* (2015)]

### 2.5.2 Méthodes par apprentissage profond

Ces dernières années, les progrès impressionnants réalisés dans le domaine de l'apprentissage profond ont favorisé le développement de la plupart des méthodes de reconnaissance visuelle (Figure 2.7). L'apprentissage profond a été largement utilisé dans tout le domaine de la vision par ordinateur, y compris la détection d'objets.

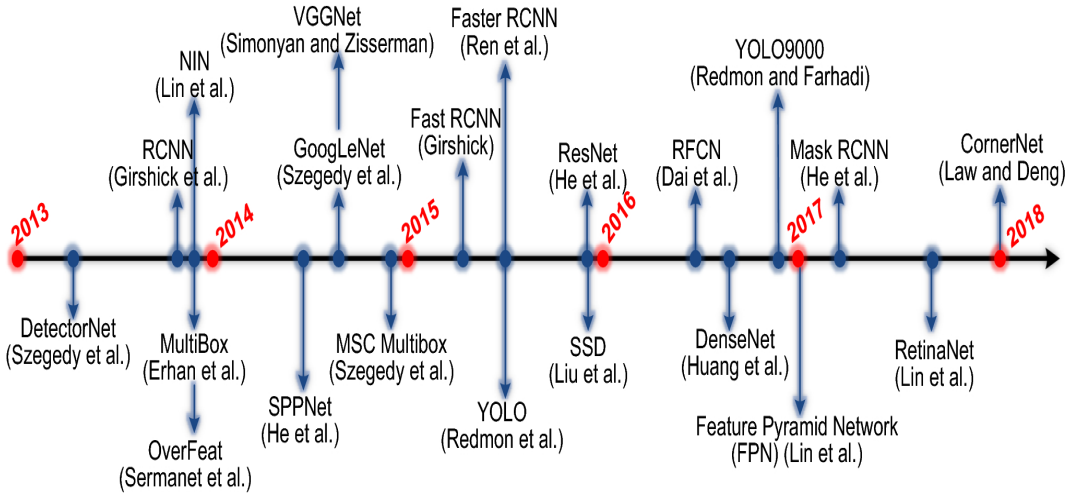



FIGURE 2.7 – Chronologie des méthodes profondes de détection d'objet. Source [Liu *et al.* (2018)]

Dans cette section nous détaillons les principales approches proposées récemment exploitant les réseaux de neurones convolutifs pour la détection d'objets. En premier lieu, les généralités sont abordées pour introduire les concepts de base d'un réseau neurone. En deuxième lieu, nous détaillons les approches profondes pour la détection d'objet qui se décomposent en deux familles : les méthodes basées sur des régions présentées dans la Section 2.5.2.2, et les méthodes basées sur un seul réseau présentées dans la Section 2.5.2.3.

#### 2.5.2.1 Généralités

Un bref aperçu de l'histoire de l'apprentissage profond [Goodfellow *et al.* (2016)] révèle trois grandes vagues ; la première dénomination connue sous le nom *Cybernétique* durant les années 1940 − 1960, la deuxième dénomination connue sous le nom *Connectionisme* durant les années 1980 − 1990 et la tendance actuelle appelée *apprentissage profond* à partir de 2006.

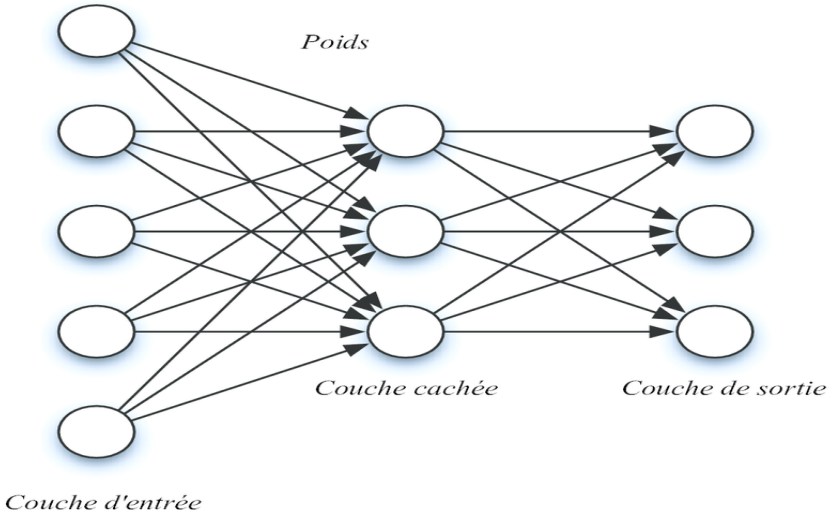



FIGURE 2.8 – Représentation d'un perceptron multi-couche sous la forme d'un graphe. Source [Orbach (1962)]

Les réseaux de neurones sont inspirés de la modélisation du système nerveux. Un exemple typique de ce type de réflexion est le Perceptron (Figure 2.8), proposé dans les années 60 dans l'article [Orbach (1962)], qui est inspiré du fonctionnement d'un neurone. En biologie, le signal d'entrée transmis par les dendrites est accumulé à l'intérieur du neurone, puis le signal de sortie est généré lorsqu'un certain seuil est atteint.

Par conséquent, le but du Perceptron est d'approcher une fonction $f$ afin de produire une sortie $y$ (catégorie ou valeur réelle) de la forme $y = f(x; w)$ à partir du vecteur d'entrée $x$ et du vecteur de paramètres $w$. Le perceptron est donc de la forme suivante :

$$f(x, w) = g(x^T w + b) \tag{2.1}$$

avec $g$ une fonction non linéaire dite d'activation, destinée à simuler le phénomène de seuil neuronal, et $b$ le paramètre de biais. Les paramètres $w$ et $b$ sont déterminés de manière itérative pendant le processus d'apprentissage du modèle.

Dans cette section, nous présentons par la suite le type de réseaux de neurones le plus fréquemment mis en oeuvre : le Réseau de Neurones Convolutif (RNC) [LeCun *et al.* (2015)].

**Réseau de Neurones Convolutif** Les réseaux de neurones convolutifs [LeCun *et al.* (2015)] sont une classe de modèles basés sur l'apprentissage profond. Un réseau de neurones convolutif est dédié aux tâches de vision par ordinateur et se caractérise par l'utilisation de couches de convolution et de sous-échantillonnage pour apprendre des représentations visuelles efficaces.

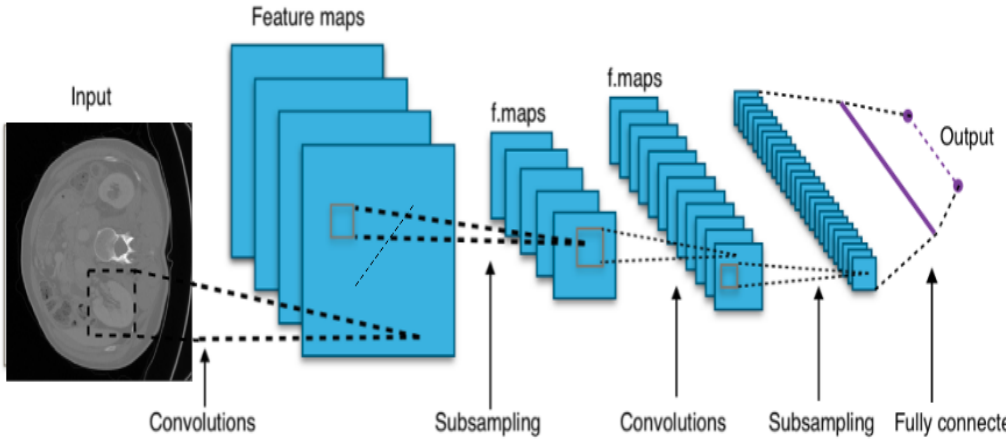



FIGURE 2.9 – Architecture de Réseau de Neurones Convolutif. Source [LeCun *et al.* (2015)]

Les RNCs sont adaptés pour les problèmes d'apprentissage sur des données structurées, organisées en grilles grille 1D, 2D, et 3D (signal audio, les images ou encore la vidéo). Un réseau de neurones convolutif est généralement basé sur trois éléments principaux (Figure 2.9) :

- *Une couche de convolution* : cette couche consiste à appliquer un filtre convolutif à l'entrée de la couche, ce qui donne en sortie une carte de caractéristiques. Les poids du filtre de convolution sont l'objet de l'apprentissage.
- *Une fonction d'activation* : c'est une fonction mathématique non-linéaire appliquée à la carte de caractéristiques en sortie de la couche de convolution.
- *Sous-échantillonnage* : cette couche appelée aussi "couche avec pooling" consiste à réduire les dimensions de la sortie de la fonction d'activation. Cette action est faite soit par une fonction de maximum ou une fonction de moyenne.

Les réseaux de neurones convolutifs consistent à trouver une fonction reliant la sortie à l'entrée au travers d'une succession de convolutions dont les paramètres sont appris par rétropropagation [Rumelhart *et al.* (1995)] qui permet d'entraîner efficacement des réseaux de neurones à plusieurs couches. Cette démarche consiste à définir une fonction de coût, calculé à partir de l'écart entre les données d'apprentissage (vérité terrain) et la sortie du RNC. L'erreur sur la fonction de coût est alors propagée à chaque paramètre depuis la sortie vers l'entrée selon le théorème de dérivation des fonctions composées. Le modèle est adapté grâce à une série de mises à jour des paramètres de filtre afin que la sortie

du réseau optimise la fonction de coût. Cette fonction est généralement réalisée par une méthode d'optimisation de descente de gradient particulière appelée descente de gradient stochastique (SGD) [LeCun *et al.* (2015)].

Les exemples d'applications d'un RNC sur des images médicales sont nombreux. Parmi ceux-ci, les auteurs de [De Vos *et al.* (2016)] utilisent un RNC pour détecter les organes thoraciques sur des images TDM, alors que les auteurs de [Xu *et al.* (2019)] proposent un modèle basé sur un RNC 3D pour la classification et par un réseau de proposition de région (RPN) 3D pour la localisation des organes sur des images TDM.

#### 2.5.2.2 Méthodes basées sur les régions

Les méthodes basées sur les régions correspondent au mécanisme attentionnel de la perception humaine. Elles se décomposent en deux étapes. La première consiste en un balayage grossier de l'ensemble des données, et la seconde se focalise sur des régions d'intérêt. Ces régions d'intérêt font ressortir les régions intéressantes dans l'image, c'est-à-dire présentant des caractéristiques locales importantes. Ces régions peuvent se présenter sous forme de points, de courbes continues ou de zones connectées. Sur nos données, nous utilisons des boîtes rectangulaires pour la localisation appelée "boîtes englobantes".

Parmi les travaux basés sur les régions avec des réseaux de neurones convolutifs (R-CNN) nous présentons dans la suite les méthodes R-CNN [Girshick *et al.* (2014)], Fast R-CNN [Girshick (2015)] et Faster R-CNN [Ren *et al.* (2015)].

**R-CNN** L'algorithme de région avec des réseaux de neurones convolutifs [Girshick *et al.* (2014)] proposé par Ross Girshick en 2014 est un réseau profond pour la détection d'objet dans une image.

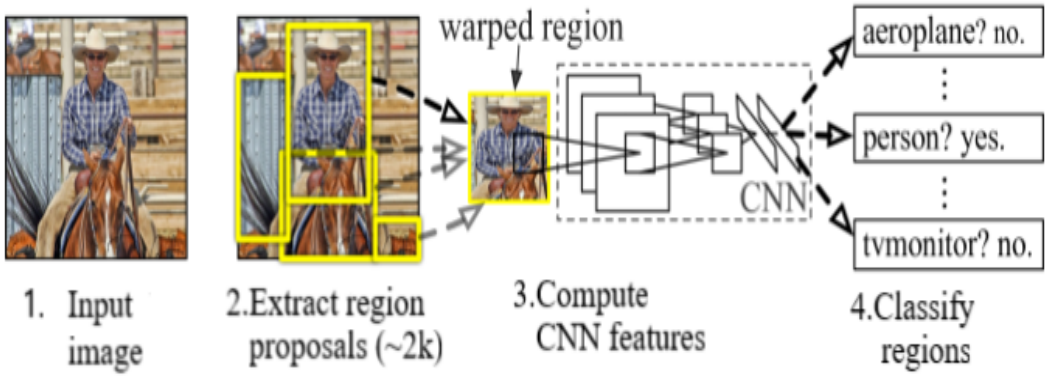



FIGURE 2.10 – Architecture de la méthode R-CNN. Source [Girshick *et al.* (2014)]

La méthode R-CNN, comme le montre la Figure 2.10, est répartie en trois étapes. Dans la première étape, on commence par extraire des régions de l'image à l'aide d'un algorithme de recherche sélective [Uijlings *et al.* (2013)]. Ensuite, on considère chaque région sélectionnée comme une entrée d'un CNN pour créer des vecteurs de caractéristiques représentant

l'image d'entrée en dimensions réduites. En dernière étape, on classifie ces vecteurs de caractéristiques en utilisant un algorithme d'apprentissage tel que le SVM. L'algorithme de recherche sélective de la méthode R-CNN est un algorithme externe au réseau. Cela présente un inconvénient pour cette méthode en augmentant le temps d'exécution et en générant de mauvaises régions d'intérêt. Un autre inconvénient de l'algorithme R-CNN est le nombre important des régions d'intérêt ce qui implique un grand nombre de CNN. Cela rend l'opération très coûteuse.

**Fast R-CNN** Fast R-CNN [Girshick (2015)] est une amélioration de la rapidité de la méthode R-CNN (R-CNN prend beaucoup de temps pour extraire les régions). Le principe de Fast R-CNN comme le montre la Figure 2.11, consiste à prendre toute l'image numérique comme une entrée d'un seul réseau CNN pour extraire les vecteurs de caractéristiques à la différence de R-CNN qui prend chaque vecteur de caractéristiques de chaque région comme une entrée d'un seul CNN. Ensuite, à partir de la carte des caractéristiques, deux étapes sont appliquées : la localisation des régions en utilisant la couche "ROI pooling" et la classification de ces régions en utilisant un softmax. Fast R-CNN extrait les vecteurs de caractéristiques avec un traitement comprenant une seule étape, ce qui le rend plus efficace que R-CNN qui est un traitement à plusieurs étapes. Par ailleurs, le Fast R-CNN repose encore sur la recherche sélective pour extraire les propositions de régions, ce qui limite sa vitesse de test et d'entraînement.

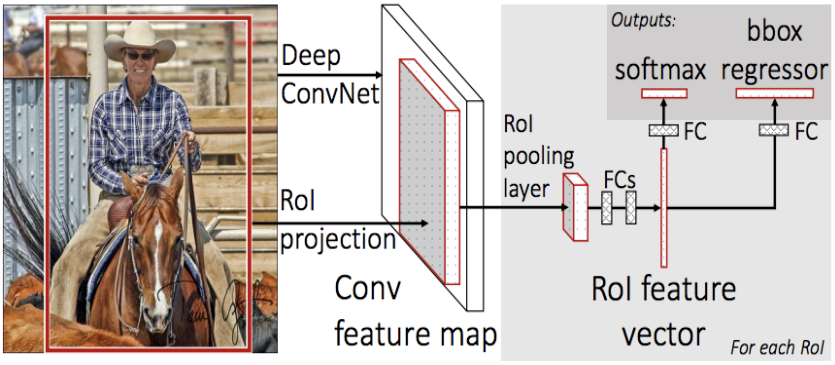



FIGURE 2.11 – Architecture de la méthode Fast R-CNN. Source [Girshick (2015)]

**Faster R-CNN** Faster R-CNN [Ren *et al.* (2015)] est l'évolution des méthodes Fast R-CNN et R-CNN. Ces dernières s'appuient sur un algorithme fixe pour extraire les propositions de régions, ce qui demeure lent et coûteux. Faster R-CNN est basé sur deux réseaux de neurones qui partagent la partie convolution comme l'illustre la Figure 2.12. Le premier réseau remplace l'algorithme de recherche sélective, ce réseau est appelé réseau de proposition de région. Le deuxième réseau prend en entrée les régions proposées par le premier réseau et recherche si elles contiennent l'objet à détecter. La détection est ensuite faite par un ROI pooling et un classifieur.

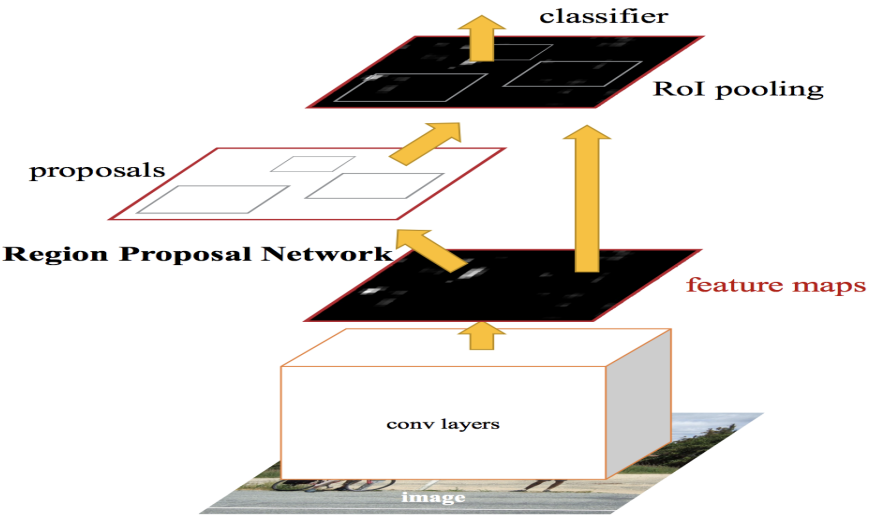



FIGURE 2.12 – Architecture de la méthode Faster R-CNN. Source [Ren *et al.* (2015)]

Des travaux ont ainsi proposé d'utiliser Faster R-CNN pour détecter le disque intervertébral [Sa *et al.* (2017)] sur des images TDM où une précision moyenne de 0.905 a été obtenue avec un temps d'exécution de 3 s à comparer à une méthode traditionnelle (HOG + SVM) conduisant à une précision moyenne de 0.091 et un temps d'exécution de 82 s.

#### 2.5.2.3 Méthodes utilisant un seul réseau

Contrairement aux méthodes basées sur des régions, le principe de la méthode de réseau unique est de prédire à la fois les classes et les boîtes englobantes en les passant à travers un seul réseau de neurones profond. L'avantage de ces méthodes est qu'elles sont plus rapides que les détecteurs basés sur les régions. Ces méthodes incluent You Only Look Once (YOLO) [Redmon *et al.* (2016)], SSD [Liu *et al.* (2016)] et RetinaNet [Lin *et al.* (2017)b].

**You Only Look Once** YOLO [Redmon *et al.* (2016)] est une méthode de détection d'objets proposée par Redmon et al. en 2016. L'approche de YOLO utilise un seul réseau profond. Elle consiste à prendre une image numérique en entrée et à la diviser en $S \times S$ cellules. Chaque cellule prédit trois vecteurs :

- L'indice de confiance d'objet qui représente la probabilité qu'un objet soit contenu dans une boîte englobante.
- Un tuple de 4 cordonnées (centre $x$, centre $y$, largeur $w$, hauteur $h$) qui représente l'emplacement et les dimensions des boîtes englobantes $B$.
- Une probabilité de classe d'objet qui représente la probabilité d'appartenance de l'objet détecté à une classe particulière.

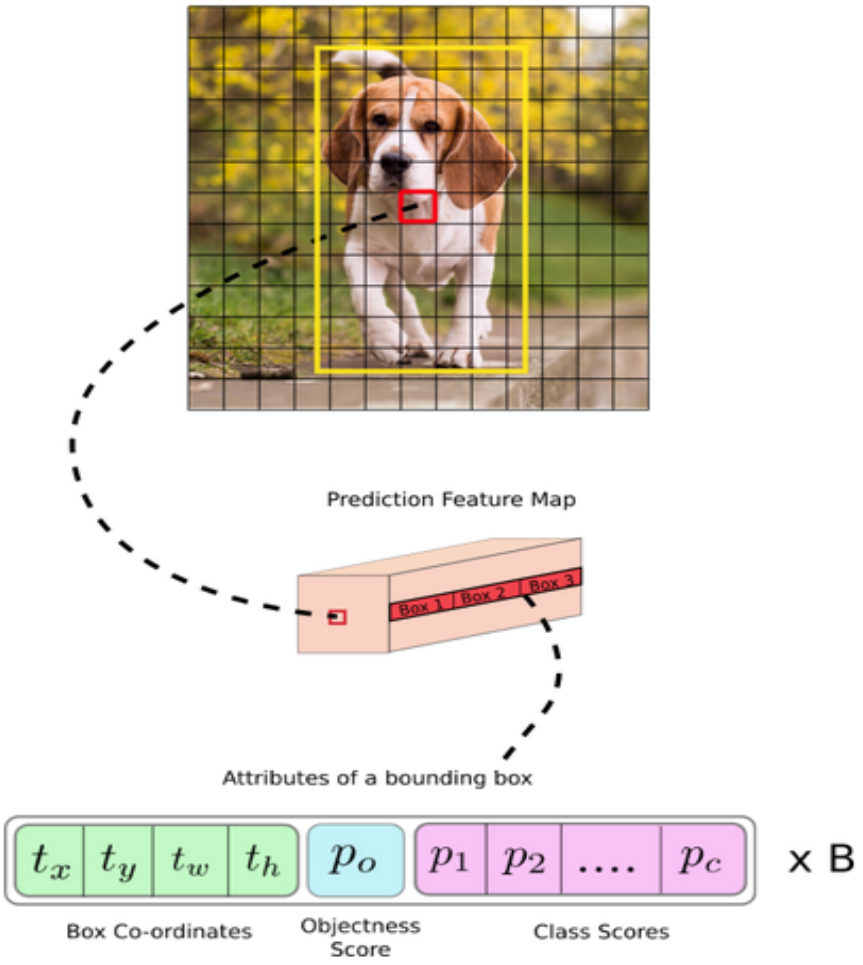



FIGURE 2.13 – La couche de la détection de YOLO. Source [Kathuria (2018)]

À la sortie de YOLO, comme le montre la Figure 2.13, l'image contient un total de $S \times S \times B$ boîtes englobantes. Chaque boîte englobante est associée à 4 coordonnées, 1 indice de confiance et $C$ probabilités de classes. Nous présenterons en détail le fonctionnement de la méthode YOLO dans le Chapitre 3.

En imagerie médicale, YOLO a été utilisé pour la détection des nodules pulmonaires en TDM avec une précision de 0.93 [Sindhu *et al.* (2018)], et YOLOv3 pour la détection des reins en TDM [Lemay (2019)] avec une précision de 0.85 en 2D et de 0.74 en 3D.

**Single shot object detectors** La méthode Single Shot Detector (SSD) a été proposée par Liu et al. en 2016 [Liu *et al.* (2016)]. L'architecture de SSD correspond à une pyramide de différentes échelles comme le montre la Figure 2.14. Les cartes de caractéristiques d'une image extraites par le modèle VGG-16 [Simonyan and Zisserman (2014)] sont représentées à des différents niveaux et différentes échelles. Au contraire de YOLO, la détection avec SSD se fait dans chaque couche pyramidale, en ciblant des objets de tailles différentes.

En imagerie médicale, le détecteur SSD a été utilisé dans [Lee *et al.* (2018)] pour

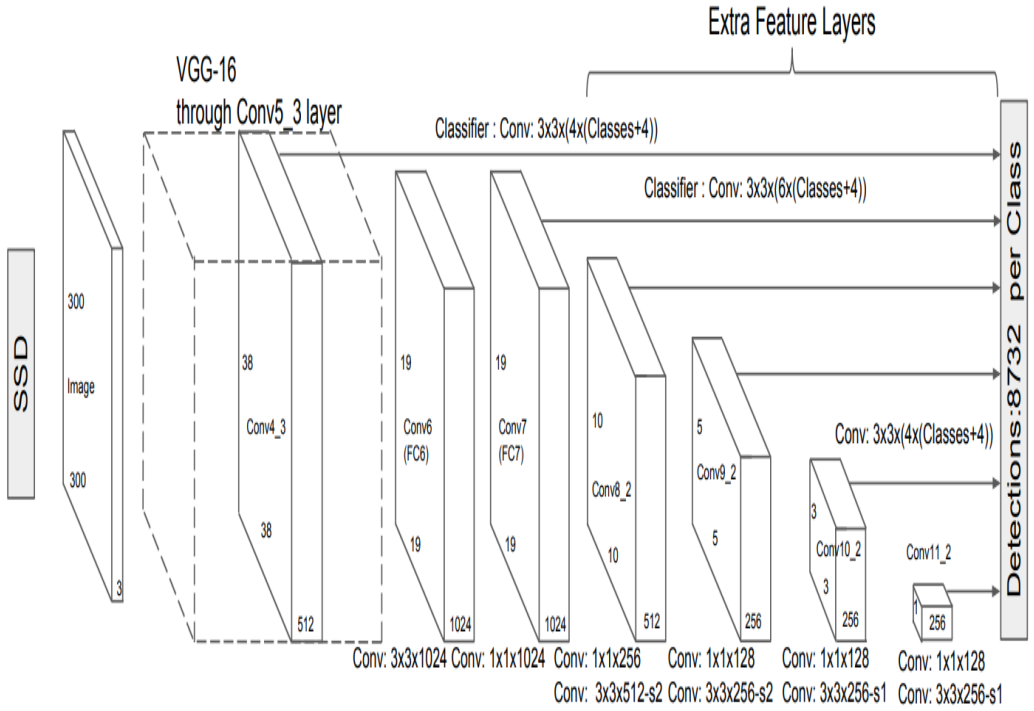



FIGURE 2.14 – Architecture de la méthode SSD. Source [Liu *et al.* (2016)]

détecter des lésions hépatiques dans des volumes TDM avec une précision moyenne de 0.53.

**RetinaNet** RetinaNet [Lin *et al.* (2017)b] proposé par Lin et al. est un détecteur à un seul réseau. Ce détecteur utilise l'erreur focale (FL) comme fonction de perte de classification. Cette fonction de perte FL consiste à résoudre le problème de déséquilibre de classe. L'architecture de "RetinaNet" est composée de quatre parties (voir Figure 2.15). La première partie a pour rôle d'extraire des cartes de caractéristiques à différentes échelles. La deuxième échantillonne ces cartes à l'aide d'un réseau sous la forme d'une pyramide commençant par l'échelle la plus faible. Elle présente une connexion latérale qui fusionne les couches descendantes et ascendantes ayant la même taille spatiale pour résoudre le problème de l'atténuation des signaux importants au cours du passage à travers les couches. Ensuite, la troisième partie réalise la classification. La dernière partie consiste à localiser les régions d'intérêt.

En imagerie médicale, RetinaNet a été utilisé dans [Yang *et al.* (2020)] pour détecter le ventricule gauche dans des images d'échocardiographie multi-vues. Les précisions obtenues pour les orientations d'acquisition deux chambres apicales (A2C), trois chambres apicales (A3C) et quatre chambres apicales (A4C) sont respectivement de 0.86, 0.79 et 0.84.

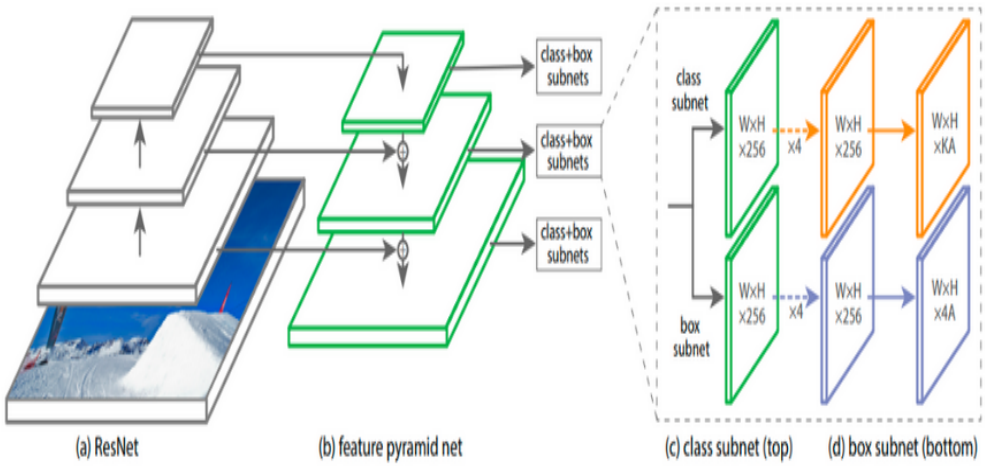



FIGURE 2.15 – Architecture de la méthode RetinaNet. Source [Lin *et al.* (2017)b]

## 2.6 Quelle est la méthode adéquate ?

Le but de notre approche est de détecter les organes rapidement et avec une bonne précision. Pour cela, nous avons détaillé précédemment les différentes approches de détection d'objet pour les images médicales. Dans cette section nous allons utiliser une comparaison de ces approches afin d'identifier un compromis entre la précision et la rapidité. Dans cette comparaison nous allons adapter l'utilisation de l'apprentissage profond aux exigences spécifiques de l'imagerie médicale.

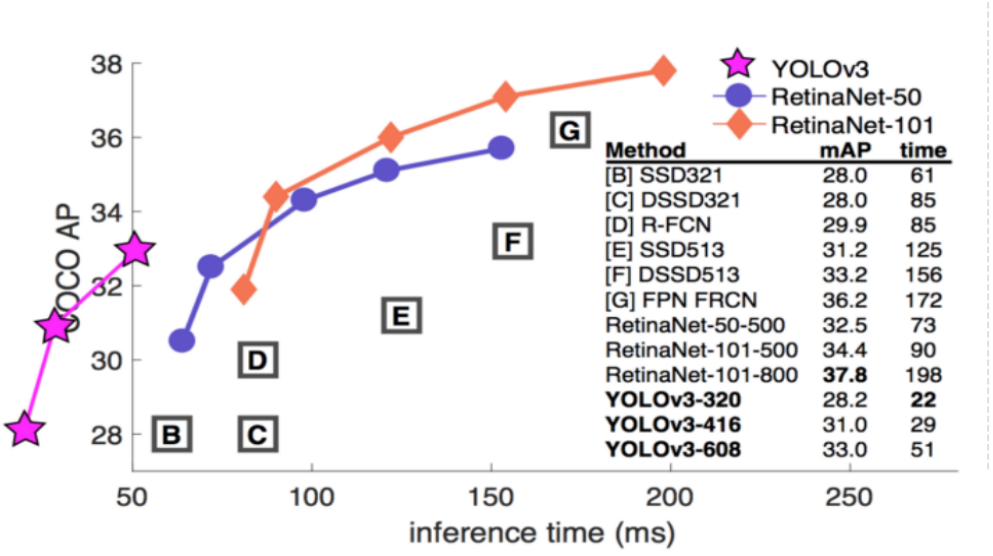



FIGURE 2.16 – Étude comparative : Vitesse (ms) par rapport à la précision (Ap) sur la base COCO [Lin *et al.* (2014)]. Source [Redmon and Farhadi (2018)]

La Figure 2.16 ci-dessous montre la comparaison entre les différentes approches profondes faite par [Redmon and Farhadi (2018)] sur la base COCO [Lin *et al.* (2014)]. Les différents nombres (320/416/608) sont les dimensions d'entrée des images. Celle-ci influe sur les performances du modèle. Nous avons choisi la méthode YOLO (la courbe en violet) comme méthode de détection, pour sa précision comparable aux autres méthodes et sa rapidité. YOLO fonctionne beaucoup plus rapidement avec un temps d'inférence de 22 *s* par rapport aux autres méthodes qui dépassent 1 min. Ce temps d'exécution est calculé sur des GPU similaires comme M40 ou Titan X.

Une fois le choix de travailler avec YOLO effectué, il a fallu décider quelle version à utiliser. YOLO présente quatre variantes principales : YOLOv1 est la première version qui propose l'architecture globale, YOLOv2 [Redmon and Farhadi (2017)] améliore l'entraînement et utilise des boîtes d'ancrage prédéfinies pour améliorer les suggestions de boîtes englobantes. YOLOv3 [Redmon and Farhadi (2018)] améliore de plus l'architecture du modèle en utilisant 3 couches de détection comme le montre la Figure 2.17. La version la plus récente YOLOv4 [Bochkovskiy *et al.* (2020)] améliore l'entraînement en ajoutant de nouvelles fonctionnalités comme l'augmentation des données (Mosaic) et la régularisation DropBlock.

La troisième version de YOLO sera appliqué dans la suite de thèse, dans la mesure où cette la version était la plus récente et performante lorsque nous avons commencé ce travail de recherche.

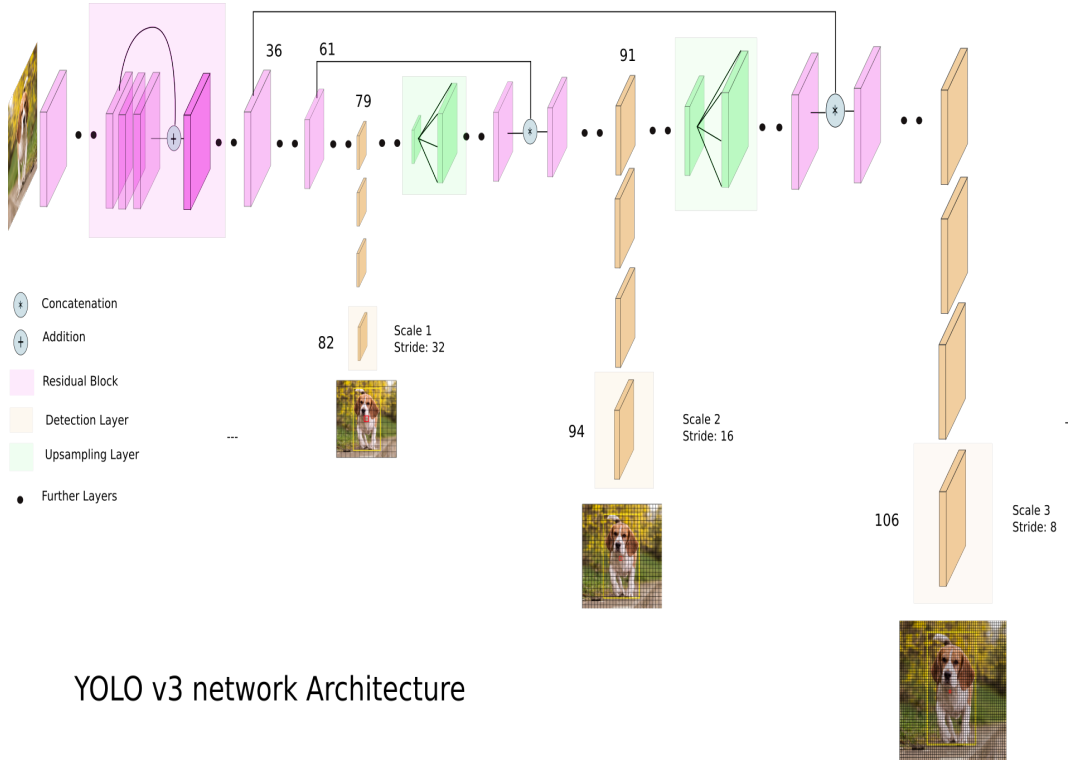



FIGURE 2.17 – Architecture de la méthode YOLOv3. Source [Redmon and Farhadi (2018)]

## 2.7 Conclusion

Au cours de ce chapitre, nous avons présenté l'importance de la détection des structures anatomiques dans le corps humain pour l'imagerie médicale. Ensuite, nous avons défini la détection d'objet dans le domaine de l'imagerie. Puis, nous avons présenté les images sur laquelle nous allons appliquer les méthodes développées. Ces images sont des images médicales des modalités TDM et IRM. En outre, nous avons détaillé l'état de l'art afin de mettre l'accent sur l'importance de la détection en imagerie médicale.

Au cours de ces dernières années, le développement des approches de la détection a évolué très rapidement, notamment en apprentissage automatique, où la progression rapide des réseaux de neurones profonds révolutionne les technologies existantes. Dans ce contexte, nous avons présenté les spécificités des réseaux de neurones profonds, qui sont aujourd'hui utilisés dans de nombreuses études ayant pour objectif la détection d'objet en imagerie médicale, et qui sont à la base des travaux réalisés au cours de cette thèse.

La comparaison des approches profondes nous a conduits à sélectionner comme détecteur la version 3 de YOLO pour son bon compromis entre précision et rapidité.

Dans le chapitre suivant, nous détaillons le modèle YOLOv3, son architecture et sa fonction de perte. Nous présentons ensuite son application à la détection d'organes en imagerie médicale multi-modalité.

Chapitre 3

# Détection multi-organe dans les images médicales

## 3.1 Introduction

La détection des organes dans des images médicales est utilisée dans plusieurs applications hospitalières pour planifier des opérations chirurgicales ou rechercher une pathologie [Lee *et al.* (2018), Lemay (2019), Sa *et al.* (2017)]. En raison de la variabilité entre les patients et la présence des artefacts d'acquisition, tels que le bruit et l'inhomogénéité du champ magnétique en IRM, la détection est une tâche difficile.

Dans ce chapitre, nous présentons l'approche choisie adoptée pour la détection des organes via un détecteur profond que nous avons sélectionné "YOLO". Dans la première partie de ce chapitre (Section 3.2), nous détaillons l'architecture et la fonction de perte de YOLO. Ensuite, nous introduisons le protocole expérimental en présentant la base de données, les métriques d'évaluation et le protocole d'entraînement (Section 3.3). Dans la Section 3.4, nous présentons les résultats qualitatifs et quantitatifs de l'algorithme de YOLO. Cette expérience est réalisée sur des images médicales de différentes modalités TDM, IRM et différentes orientations de coupe, axiales et coronales. Enfin, nous commentons les résultats obtenus pour jauger l'intérêt de la méthode (Section 3.4).

## 3.2 Détection des organes par YOLO

Dans le chapitre précédent, nous avons vu de nombreux algorithmes basés sur des réseaux profonds et nous avons sélectionné la version 3 de YOLO [Redmon and Farhadi

(2018)] comme détecteur pour sa rapidité et sa précision par rapport aux autres méthodes profondes. Dans cette section, nous commençons par détailler l'architecture du réseau YOLOv3 : les différentes couches du réseau et les hyperparamètres. Ensuite, nous présentons la fonction de coût utilisée pour optimiser les paramètres du réseau.

### 3.2.1 Architecture de YOLO

Le principe du modèle YOLO est de ne parcourir l'image qu'une seule fois à travers un réseau neuronal profond, ce qui est le contraire des méthodes basées région (Section 2.5.2.2). L'architecture globale de YOLOv3 est présentée dans la Figure 3.1.

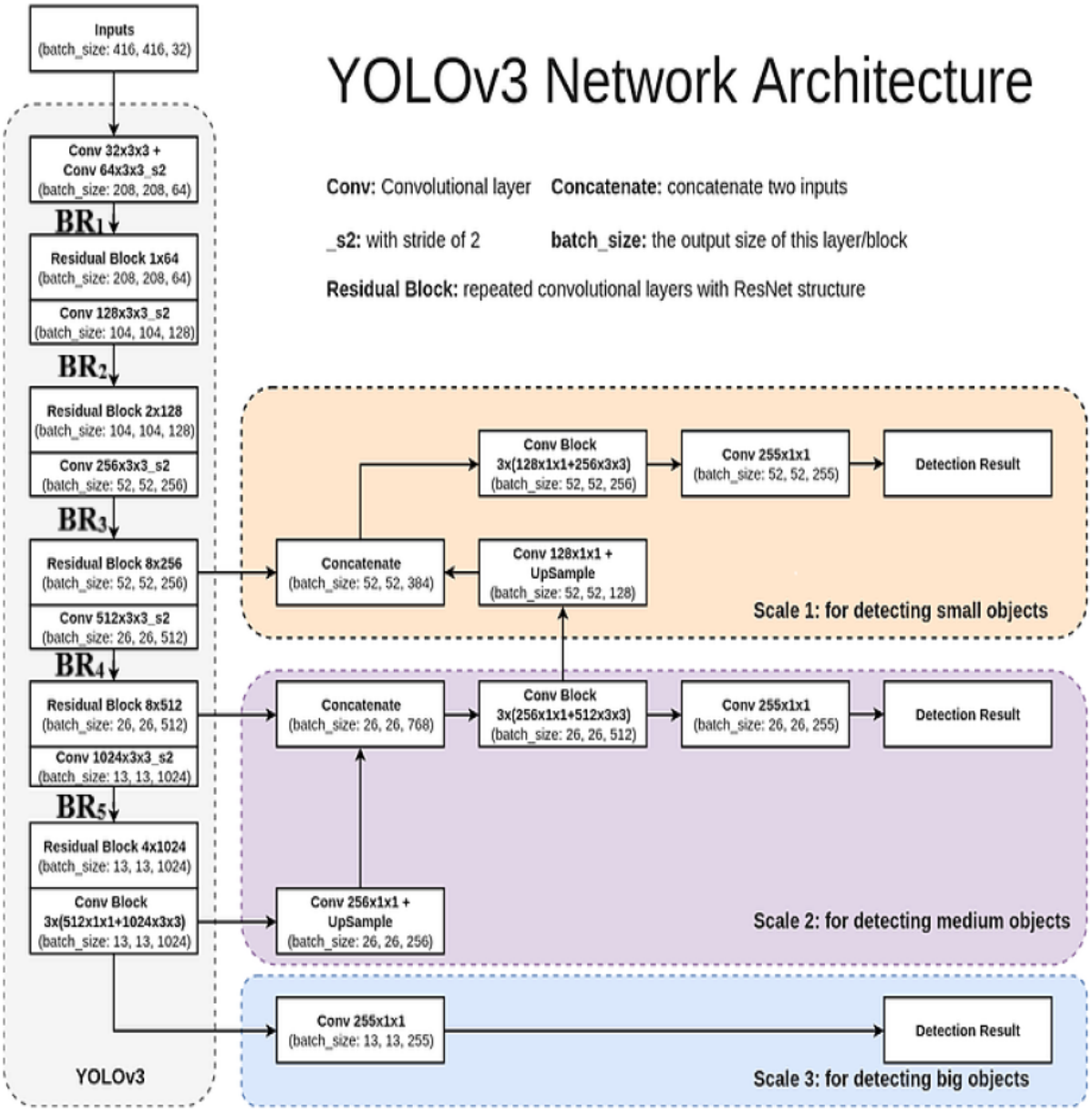



FIGURE 3.1 – Architecture détaillée de la méthode YOLOv3. Source [De Palma (2020)]

Son architecture est basée sur les couches de convolutions (Section 3.2.1.1) et trois couches de détection (Section 3.2.1.2).

#### 3.2.1.1 Couches de convolution

YOLOv3 est basé sur l'architecture d'un réseau de neurones de convolution (RNC) à 53 couches de convolutions nommé "Darknet53" (Figure 3.1). Le modèle YOLOv3 est inspiré de l'idée du réseau pyramidal de caractéristiques (feature pyramid network) [Lin *et al.* (2017)a]. Afin d'extraire des caractéristiques plus profondes, YOLOv3 ajoute cinq blocs résiduels (BR) au réseau comme le réseau pyramidal de caractéristiques. Chaque bloc contient une couche résiduelle suivie d'une couche de convolution. La couche résiduelle inclut l'ajout de zéros complémentaires (zéro padding) sur les bords de la convolution et l'unité résiduelle. Chaque couche de convolution est suivie d'une normalisation par lot (Batch Normalization) [Ioffe and Szegedy (2015)] et d'un ReLU avec fuite (leaky ReLU) [Nair and Hinton (2010)]. L'algorithme de normalisation par lot est utilisé pour accélérer la vitesse de convergence pendant l'entraînement du modèle, rendre le processus d'entraînement du modèle plus stable et éviter l'explosion du gradient ou la disparition du gradient.

Les sorties des couches résiduelles des blocs BR 3 et BR 4 (la partie gauche de la Figure 3.1), ainsi que la sortie de la couche de convolution du bloc BR 5 sont utilisées pour réaliser la détection des objets à 3 échelles différentes. Pour cela, le troisième bloc produit des cartes de caractéristiques $52 \times 52$ afin de détecter les objets de petite taille. De même, le quatrième bloc produit des cartes de caractéristiques $26 \times 26$ pour la détection de cibles de taille moyenne. YOLOv3 dispose d'une couche de convolution $255 \times 1 \times 1$ pour produire des cartes de caractéristiques $13 \times 13$ avec 255 canaux pour la détection de gros objets. Enfin, YOLOv3 détecte des images à trois échelles différentes avec des cartes de caractéristiques $32 \times 32$, $16 \times 16$ et $8 \times 8$ à l'aide de la couche de la détection.

#### 3.2.1.2 Couches de détection

Comme le montre la Figure 3.1, la détection est réalisée sur 3 différentes échelles. Cette couche est du type densément connecté (Fully Connected). Elle effectue une somme pondérée de tous les éléments de la matrice d'entrée pour obtenir un seul nombre en résultat. Cette couche du réseau est de dimension $S \times S \times 3 \times$ [($C$ classes) + 1 indice de confiance + 4 coordonnées de la boîte englobante] où $S$ représente le nombre de cellules (Figure 2.13).

La couche de détection produit plusieurs boîtes englobantes qui se chevauchent pour un même objet comme le montre la Figure 3.2. Ces boîtes englobantes sont sélectionnées pour la détection finale à l'aide de l'algorithme de suppression de non maxima (SNM) détaillé ci-dessous afin de ne garder que les boîtes englobantes de prédiction les plus fiables. Auparavant, nous détaillons ci-dessous la métrique Intersection sur Union qui est utilisée dans l'algorithme SNM.

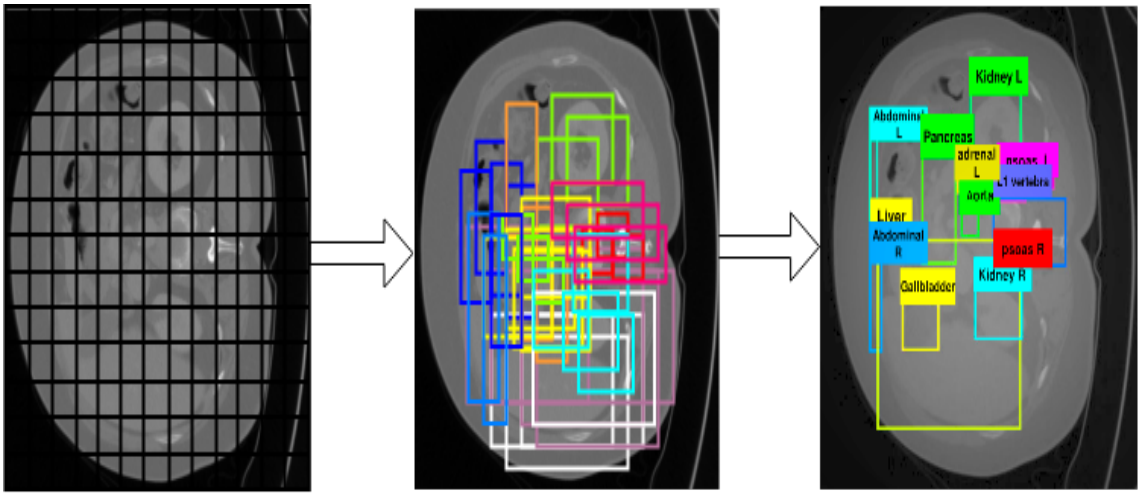



FIGURE 3.2 – Détection et application de l'algorithme SNM.

**Intersection sur Union (Intersection over Union)** L'intersection sur l'union (IoU) est une métrique d'évaluation utilisée pour mesurer la précision du détecteur. IoU est l'aire de l'intersection entre la boîte englobante de la prédiction de la détection $X$ et la boîte englobante de vérité terrain $Y$ divisé par l'aire de l'union comme le montre l'Équation (3.1).

$$IoU(X,Y) = \frac{|\, X \cap Y \,|}{|\, X \cup Y \,|} \tag{3.1}$$

Cette métrique mesure le chevauchement entre deux boîtes englobantes. Elle va de 0 à 1, 0 signifiant qu'il n'y a pas de chevauchement et 1 signifiant que la détection et la vérité terrain se chevauchent parfaitement.

**Suppression de non maxima** L'algorithme SNM (Algorithme 1) traite les prédictions classe par classe. La prédiction avec la plus grande probabilité d'être l'objet souhaité est conservée et les boîtes englobantes trop proches de cette dernière ne le sont pas. En supprimant d'abord toutes les prédictions avec des indices de confiance trop faibles, seul un nombre pertinent de prédictions est conservé comme le montre la Figure 3.2.

**Algorithme 1** Suppression de Non Maxima.

procédure : SNM
définir le seuil de IoU $\beta$
définir le seuil de l'indice de confiance $\tau$
supprimer toutes les boîtes englobantes avec un indice de confiance $< \tau$
**pour** chaque classe $c \in C$ **faire**
    **tant que** il reste des boîtes avec la classe $c$ **faire**
        **sélectionner** la boîte englobante $b$ avec le plus grand indice de confiance
        **supprimer** les boîtes englobantes dont IoU avec $b >= \beta$
    **fin tant que**
**fin pour**

#### 3.2.1.3 Hyperparamètres

Les hyperparamètres du modèle sont des paramètres qui doivent être définis manuellement avant le début de l'entraînement. Il existe plusieurs hyperparamètres à définir :

- Le nombre de classes (pour savoir le nombre de sorties à donner à la couche)
- Le seuil de l'algorithme suppression de non maxima.
- La taille de la cellule (qui divise l'image).
- Les boîtes d'ancrage et leurs tailles qui sont définies en fonction des annotations des données d'entraînement.

**Les boîtes d'ancrage** L'algorithme de YOLOv3 divise l'image d'entrée en plusieurs cellules. Afin de détecter de nombreux objets de différentes échelles, il utilise des boîtes englobantes prédéfinies avec des tailles et des rapports d'aspects différents, qui se concentrent sur chaque cellule. Ces boîtes englobantes sont appelées *boîtes d'ancrage*. L'intuition est d'utiliser des boîtes d'ancrage pour représenter les dimensions intrinsèques des objets.

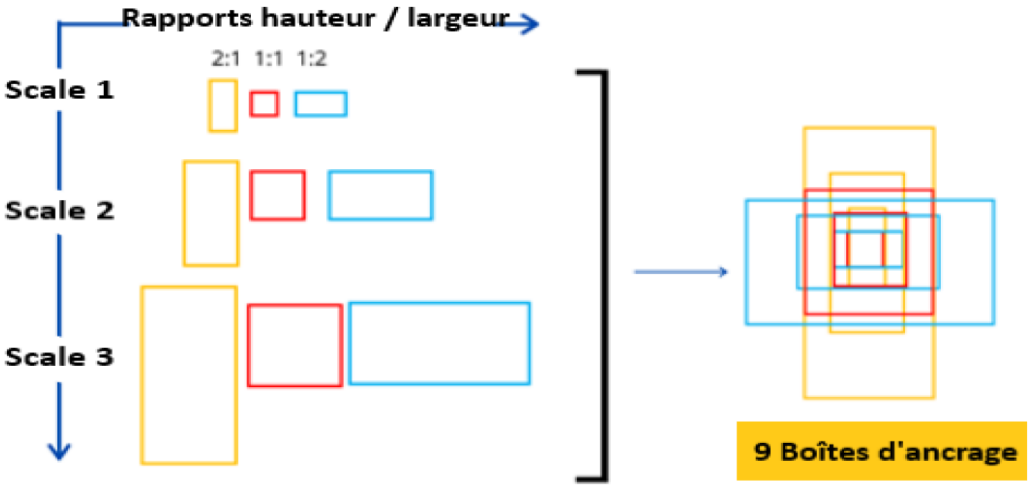



FIGURE 3.3 – Les 3 échelles et les 3 rapports (hauteur / largeur) des boîtes d'ancrage.

L'ensemble des boîtes d'ancrage ($B$) est prédéfini d'une certaine hauteur et largeur. Comme le montre la Figure 3.3, les boîtes d'ancrage sont utilisées pour représenter l'échelle et le rapport hauteur / largeur de la catégorie d'objets spécifique à détecter, et sont généralement sélectionnées en fonction de la taille de l'objet dans l'ensemble de données d'apprentissage.

YOLOv3 utilise 9 boîtes d'ancrage c'est-à-dire 3 pour chaque échelle. Chaque boîte d'ancrage est utilisée pour une taille d'objet différente. Les boîtes d'ancrage sont générées à l'aide de l'algorithme de clustering "k-means" [MacQueen (1967)] qui vérifie toutes les boîtes englobantes de l'ensemble des données afin de ne garder que les boîtes englobantes les plus représentatives.

### 3.2.2 Fonction de perte

La fonction de perte de la version initiale de YOLO est une somme de trois termes : le premier traite les 4 coordonnées de la boîte englobante, le second traite l'indice de confiance d'objet et le troisième traite la prédiction du score de classe. La fonction de perte est donnée par l'Équation (3.2) [Redmon *et al.* (2016)] :

$$\overbrace{\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{obj}\left[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2\right]+\lambda_{coord}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{obj}\left[\left(\sqrt{w_i}-\sqrt{\hat{w}_i}\right)^2+\left(\sqrt{h_i}-\sqrt{\hat{h}_i}\right)^2\right]}^{L_{coord}}$$
$$\underbrace{+\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{obj}\left(C_i-\hat{C}_i\right)^2+\lambda_{noobj}\sum_{i=0}^{S^2}\sum_{j=0}^{B}\mathbb{1}_{ij}^{noobj}\left(C_i-\hat{C}_i\right)^2}_{L_{obj}}+\underbrace{\sum_{i=0}^{S^2}\mathbb{1}_{i}^{obj}\sum_{c\,\in\,classes}(p_i(c)-\hat{p}_i(c))^2}_{L_{prob}} \quad (3.2)$$

Avec

$S$ le nombre de cellules et $B$ le nombre de boîtes englobantes ;

$x_i$, $y_i$, $w_i$, $h_i$ les coordonnées de la boîte englobante ;

$C_i$ l'indice de confiance ;

$p_i$ la probabilité de classe ;

$\mathbb{1}_i^{obj}$ vaut 1 si un objet est détecté dans la cellule $i$, et vaut 0 sinon ;

$\mathbb{1}_{ij}^{obj}$ vaut 1 si un objet est détecté dans la cellule $i$ et la $j^{\text{ème}}$ boîte englobante, et vaut 0 sinon ;

$\mathbb{1}_{ij}^{noobj}$ vaut 1 si un objet n'est pas détecté dans la cellule $i$ et la $j^{\text{ème}}$ boîte englobante, et vaut 0 sinon ;

$\lambda_{coord}$ et $\lambda_{noobj}$ sont des coefficients de pondération ;

Le modèle divise chaque image d'entrée en une grille $S^2 = S \times S$ de cellules et chaque cellule prédit $B$ boîtes englobantes. Les $B$ boîtes englobantes sont associées au nombre des boîtes d'ancrage utilisées. Chaque boîte possède $5 + C$ attributs, où 5 fait référence aux cinq attributs de la boîte englobante (les coordonnées du centre $(x_i, y_i)$, la hauteur $(h_i)$, la largeur $(w_i)$ et l'indice de confiance $C_i$) et $C$ est le nombre de classes. Les fonctions des trois termes $L_{coord}$, $L_{obj}$ et $L_{prob}$ sont les suivantes :

$L_{coord}$ **:** Le premier terme calcule la perte liée à la position prédite de la boîte englobante : coordonnées $(x, y)$, largeur $w$ et hauteur $h$. Dans ce terme, $(\hat{x}, \hat{y}, \hat{w}, \hat{h})$ correspondent à la référence, fournie par les données d'entraînement.

$L_{obj}$ **:** Le deuxième terme calcule la perte associée à l'indice de confiance pour chaque prédiction de la boîte englobante, $C_i$ est l'indice de confiance et $\hat{C}_i$ est l'intersection sur l'union de la boîte englobante prédite avec la vérité terrain.

$L_{prob}$ **:** Le troisième terme calcule la perte de classification.

Par choix de performance, nous avons choisi la version 3 de YOLO comme détecteur. La fonction de perte n'est pas présentée en détail dans la publication relative à YOLOv3 [Redmon and Farhadi (2018)], mais celle-ci peut être déduite du code source.

La fonction de perte de YOLOv3 est présentée dans l'Équation (3.3). Cette version est aussi la somme de trois termes. La première perte $L_{pos}$ est la perte de décalage de positionnement de la cible, la deuxième perte $L_{conf}$ est la perte de confiance de la cible et la dernière perte $L_{class}$ est la perte de classification de la cible, et $\lambda_{pos}, \lambda_{obj}, \lambda_{noobj}, \lambda_{class}$ sont des coefficients de pondération associés.

La première perte utilise l'erreur quadratique moyenne (MSE). La deuxième et la troisième fonction de perte utilisent l'entropie croisée binaire (BCE) [De Boer *et al.* (2005)] à l'opposé de YOLOv1 qui utilise des pertes quadratiques (Équation (3.2)).

$$\begin{aligned}
&\underbrace{\lambda_{pos} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} \left[ \left(\sigma(t_x) - \sigma(\hat{t}_x)\right)^2 + \left(\sigma(t_y) - \sigma(\hat{t}_y)\right)^2 + \left(t_w - \hat{t}_w\right)^2 + \left(t_h - \hat{t}_h\right)^2 \right]}_{L_{pos}} \\
&\quad + \underbrace{\lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} BCE(\hat{C}_i, C_i) + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{noobj} BCE(\hat{C}_i, C_i)}_{L_{conf}} \\
&\quad + \underbrace{\lambda_{class} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} \left[ \sum_{k=1}^{C} BCE\left(\hat{q}_k, \sigma(s_k)\right) \right]}_{L_{class}}
\end{aligned} \tag{3.3}$$

$L_{pos}$ **:** Perte liée au positionnement de la cible. Ce terme correspond à l'écart quadratique entre les 4 coordonnées de la boîte de référence $(t_x, t_y, t_w, t_h)$ et celles prédites par le réseau $(\hat{t}_x, \hat{t}_y, \hat{t}_w, \hat{t}_h)$. Le détail de ces valeurs est donné en Figure 3.4. Dans la Figure 3.4, la boîte rectangulaire en pointillés est la boîte englobante de vérité terrain $\hat{b} = (\hat{b}_x, \hat{b}_y, \hat{b}_w, \hat{b}_h)$ et la boîte rectangulaire pleine est la boîte englobante prédite obtenue en calculant le décalage prédit par le réseau. Par conséquent, la boîte englobante prédite finale a comme coordonnées $b = (b_x, b_y, b_w, b_h)$ qui sont comme le montre l'Équation (3.4) :

$$\begin{aligned} b_x &= \sigma(t_x) + c_x \quad \Rightarrow \quad \sigma(t_x) = b_x - c_x \\ b_y &= \sigma(t_y) + c_y \quad \Rightarrow \quad \sigma(t_y) = b_y - c_y \\ b_w &= p_w \exp(t_w) \quad \Rightarrow \quad t_w = \log\left(\frac{b_w}{p_w}\right) \\ b_h &= p_h \exp(t_h) \quad \Rightarrow \quad t_h = \log\left(\frac{b_h}{p_h}\right) \end{aligned} \tag{3.4}$$

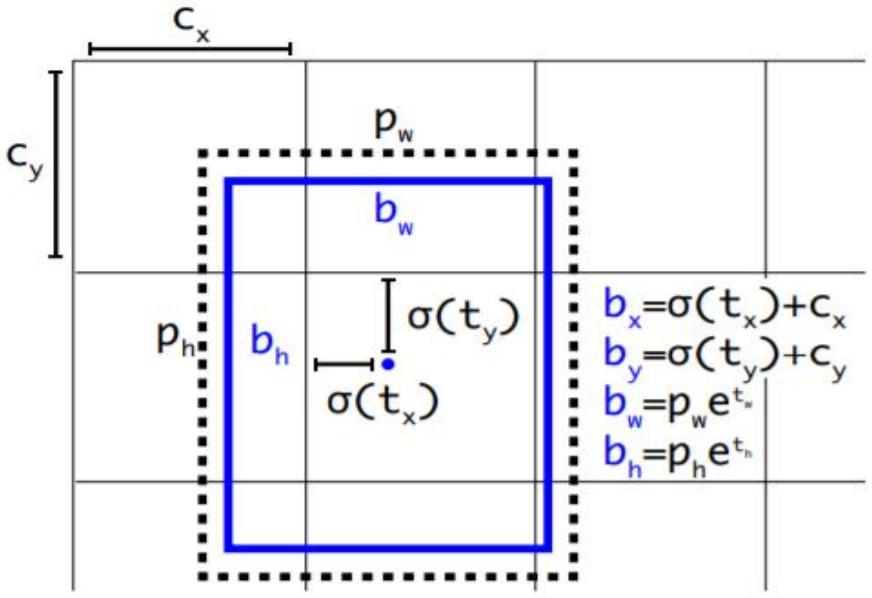



FIGURE 3.4 – Coordonnées de la boîte englobante. Source [Redmon and Farhadi (2017)]

Avec :

$\sigma(x)$ la fonction sigmoïde, qui a été utilisée pour borner la position des boîtes englobantes prédites, afin d'empêcher la croissance infinie des dimensions de la boîte englobante.

$(t_x, t_y, t_w, t_h)$ sont les coordonnées prédites de la boîte englobante par le réseau, $(c_x, c_y)$ sont les dimensions d'une cellule et $(p_w, p_h)$ sont respectivement la largeur et la hauteur de la boîte d'ancrage.

$L_{conf}$ **:** Perte liée à la mesure de confiance de l'existence d'un objet dans la boîte englobante prédite. Ce terme est basé sur la fonction d'entropie croisée et est donné par l'équation suivante (Équation (3.5)) :

$$\sum_{i=0}^{S^2} BCE(\hat{C}_i, C_i) = \sum_{i=0}^{S^2} \left[ -\hat{C}_i \log(C_i) - (1 - \hat{C}_i) \log(1 - C_i) \right] \tag{3.5}$$

Avec :

$C_i$ : correspond à la mesure de confiance de la présence d'un objet, évaluée à partir de la probabilité qu'un objet soit présent dans la cellule ($obj_i$) et de l'intersection sur l'union (IoU) de la boîte englobante de la prédiction (b) et de la vérité terrain ($\hat{b}$) (Équation (3.6)) :

$$C_i = Pr(obj_i) * IoU(b, \hat{b}) \tag{3.6}$$

$\hat{C}_i$ : indique la présence ou non d'un objet dans la boîte. Ce terme prend la valeur 1 s'il existe un objet dans la boîte englobante $i$, et 0 dans le cas contraire ;

$\mathbb{1}_{ij}^{obj}$ : vaut 1 si un objet est détecté dans la cellule $i$ et la $j^{\text{ème}}$ boîte englobante, et vaut 0 sinon ;

$\mathbb{1}_{ij}^{noobj}$ : vaut 1 si un objet n'est pas détecté dans la cellule $i$ et la $j^{\text{ème}}$ boîte englobante, et vaut 0 sinon ;

$L_{class}$ **:** Perte liée à l'erreur de classification. Ce terme utilise également la fonction d'entropie croisée et est donné par l'équation suivante (Équation (3.7)) :

$$\sum_{k=1}^{C} \left[BCE\left(\hat{q_k}, \sigma(s_k)\right)\right] = \sum_{k=1}^{C} \left[-\hat{q_k} \log\left(\sigma(s_k)\right) - (1 - \hat{q_k}) \log\left(1 - \sigma(s_k)\right)\right] \tag{3.7}$$

Avec :

$\sigma(s_k)$ : correspond, via une sigmoïde, à la probabilité de présence de l'objet de la classe $k$ prédit par le réseau dans la $i^{\text{ème}}$ boîte englobante de la $j^{\text{ème}}$ cellule ;

$\hat{q_k}$ : indique si l'objet de la classe $k$ existe effectivement. Ce terme vaut 1 si l'objet est présent et 0 sinon ;

La fonction de perte de YOLOv3 (Équations (3.3), (3.5) et (3.7)) s'écrit comme suit (Équation (3.8)) :

$$\begin{aligned}
&\lambda_{pos} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} \left[ \left(\sigma(t_x) - \sigma(\hat{t}_x)\right)^2 + \left(\sigma(t_y) - \sigma(\hat{t}_y)\right)^2 + \left(t_w - \hat{t}_w\right)^2 + \left(t_h - \hat{t}_h\right)^2 \right] \\
&+\lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} \left[ -\hat{C}_i \log\left(C_i\right) - (1 - \hat{C}_i) \log\left(1 - C_i\right) \right] \\
&+\lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{noobj} \left[ -\hat{C}_i \log\left(C_i\right) - (1 - \hat{C}_i) \log\left(1 - C_i\right) \right] \\
&+\lambda_{class} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} \sum_{k=1}^{C} \left[ -\hat{q}_k \log\left(\sigma(s_k)\right) - (1 - \hat{q}_k) \log\left(1 - \sigma(s_k)\right) \right]
\end{aligned} \tag{3.8}$$

## 3.3 Protocole expérimental

Dans cette section nous détaillons le protocole expérimental mis en place au cours des expériences d'évaluation du détecteur YOLOv3 sur les images médicales de modalité TDM et IRM. Nous présentons dans la Section 3.3.1, la base de données utilisée dans les expériences. Dans la Section 3.3.2, nous exposons les détails des métriques d'évaluation. Enfin dans la Section 3.3.3, nous présentons le protocole d'entraînement de l'expérience.

### 3.3.1 Base de données

Les données médicales sont le point central de tout projet de recherche clinique. Elles peuvent être utilisées pour déterminer l'effet du traitement sur un groupe de patients, et elles peuvent également être utilisées pour suivre la progression de la maladie. L'annotation constitue une partie cruciale et délicate du processus d'acquisition de la base de données. En effet, selon la complexité du travail d'identification ou de représentation requis, l'étude de chaque image peut prendre de quelques secondes à plusieurs heures.

Dans la section suivante nous présentons la base de données de challenge VISual Concept Extraction challenge in RAdioLogy (VISCERAL) [Hanbury *et al.* (2012)].

#### VISCERAL Anatomy Benchmark

Le challenge VISCERAL [Hanbury *et al.* (2012)] a mis en place plusieurs compétitions en segmentation d'images, détection de lésions, points de repères et de recherche d'images par le contenu [Jimenez-del Toro *et al.* (2016)]. Les données utilisées dans cette étude proviennent de ce challenge. Ils comportent deux jeux de données : (1) un jeu de données VISCERAL *Gold* et (2) un jeu de données VISCERAL *Silver*.

**VISCERAL Gold** Le projet VISCERAL a collecté, anonymisé et a mis à disposition pour les challenges [Jimenez-del Toro *et al.* (2016)] un grand nombre d'images annotées manuellement appelées «VISCERAL Gold». Les annotations de cette base ont été crées par des experts médicaux. Le Tableau 3.1 donne les détails sur cette base.

TABLEAU 3.1 – Détails de la base VISCERAL Gold : la modalité, l'anatomie, le contraste, le nombre de volumes, les annotations des structures anatomiques et les dimensions de volumes.

| Modalité | Anatomie | Contraste | Volumes | Annotations | Dimensions |
|---|---|---|---|---|---|
| TDM | corps-entier | non-contrasté | 20 | 384 | 512 × 512 × 880 |
| | thorax-abdomen | contrasté | 20 | 387 | 512 × 512 × 438 |
| IRM T1w | corps-entier | non-contrasté | 20 | 305 | 387 × 25 × 1506 |
| IRM T2w | abdomen | contrasté | 20 | 219 | 312 × 72 × 384 |

**VISCERAL Silver** En plus de la base de données d'image VISCERAL Gold avec des annotations d'experts décrite dans la section précédente, le challenge VISCERAL a

également produit une base de données plus grande appelée "VISCERAL Silver". Les annotations de cette base ont été créés en fusionnant les résultats des algorithmes de segmentation conçus par les participants dans les challenges. Bien que les annotations de la base de données Silver ne soient pas aussi précises que les annotations des experts, le résultat combiné des algorithmes est plus précis que celui des algorithmes individuels.

Par la suite, pour les expériences d'application YOLO sur les images médicales, nous allons utiliser l'ensemble des données VISCERAL Gold (20 patients par modalité). Dans cette base, nous avons sélectionné les modalités TDM thorax-abdomen et IRM abdomen fournissant respectivement 20 et 15 annotations de structure.

### 3.3.2 Métriques d'évaluation

L'évaluation expérimentale du détecteur est la principale étape pour comprendre ses avantages et ses inconvénients.

Nous allons utiliser la distance moyenne pour évaluer les résultats de l'application du détecteur YOLO. La distance moyenne est utilisée pour quantifier l'écart millimétrique des boîtes englobantes de détection par rapport aux boîtes englobantes de la vérité terrain.

Par ailleurs, nous utiliserons la précision moyenne globale (mAp) pour sélectionner le meilleur modèle en utilisant une procédure de validation croisée à $k$ plis.

**La distance moyenne :** La distance moyenne (Avg Dist) est utilisée pour évaluer la distance des boîtes englobantes de détection par rapport aux boîtes englobantes de la vérité terrain. Elle est déterminée en calculant la moyenne des 6 distances séparant les faces des boîtes englobantes détectées et les boîtes correspondant à la vérité terrain.

**La précision moyenne globale :** La précision moyenne globale (mAp, de l'anglais "mean Agerage precision") est couramment utilisée pour mesurer la précision des détecteurs d'objets. Cette métrique est la moyenne de la précision moyenne (Ap) sur toutes les classes, définie ci-dessous. Plus le score est élevé, plus la détection est précise. Cette métrique est définie à partir des notions de précision et de rappel, ces derniers étant calculés à partir du nombre de : Vrai positif (VP), Faux positif (FP) et Faux négatif (FN). Ces derniers éléments ont obtenu en seuillant l'Intersection sur Union (IoU) : pour une classe donnée, la détection est marquée comme correcte lorsque l'IoU entre la boîte de détection et la vérité terrain est supérieure à un seuil prédéfini, $\beta$.

**Vrai Positif, Faux Positif, Faux Négatif et Vrai Négatif :**

**Vrai positif (VP) :** Une détection est définie comme correcte lorsque IoU $\geq \beta$.
VP correspond alors au nombre de détections correctes.

**Faux positif (FP) :** Une détection est définie comme erronée lorsque IoU $< \beta$.
FP correspond alors au nombre de détection erronées.

**Faux négatif (FN) :** Cet élément correspond au cas d'un objet présent non détecté. FN correspond au nombre de ces cas.

**Vrai négatif (VN) :** Cet élément correspond au cas de la prédiction correcte d'absence d'objet. VN correspond au nombre de ces cas.

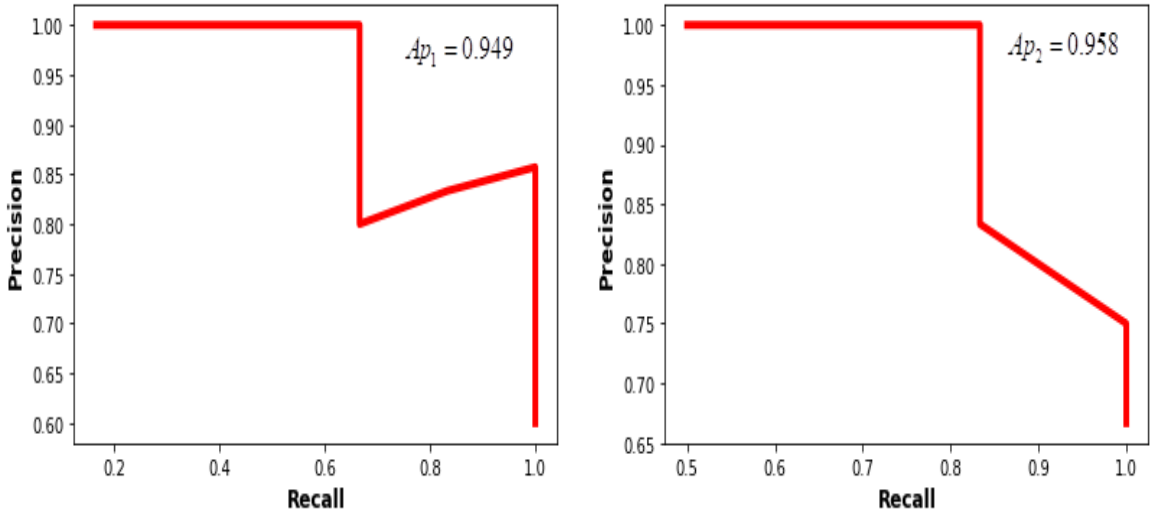
**Précision :** La précision est définie par la proportion de prédictions positives correctes donnée par (Équation (3.9)) :

$$\text{Précision} = \frac{VP}{VP + FP} = \frac{VP}{\text{toutes les détections}} \tag{3.9}$$

**Rappel :** Le rappel est la capacité d'un modèle à trouver tous les cas pertinents. Pour une classe donnée, il s'agit de la proportion de vrais positifs détectés par rapport à l'ensemble des objets à détecter de cette classe. Le rappel est donc donné par (Équation (3.10)) :

$$\text{Rappel} = \frac{VP}{VP + FN} = \frac{VP}{\text{l'ensemble des objets à détecter}} \tag{3.10}$$

En faisant varier la valeur du seuil d'IoU $\beta$, il est possible de tracer la courbe rappel-précision pour une classe donnée. Cette courbe (Figure 3.5) permet d'évaluer les performances d'un modèle de détection d'objets : le modèle est considéré comme un bon modèle prédictif si la précision reste élevée alors que le rappel augmente. La précision moyenne Ap est alors définie comme l'aire sous cette courbe . La précision moyenne globale (mAp) correspond à la moyenne des Ap de toutes les classes.




(a) Courbe précision-rappel $Ap_1$.

(b) Courbe précision-rappel $Ap_2$.

FIGURE 3.5 – Courbe précision-rappel $mAp = \frac{1}{2}\sum_{i=1}^{C=2} Ap_i = 0.9535$. Source [Gad (2021)]

### 3.3.3 Protocole d'entraînement

Nous avons appliqué une approche d'apprentissage par transfert [Pan and Yang (2009)]. Cette méthode s'appuie sur des modèles pré-entraînés pour réaliser des tâches similaires aux nôtres afin de pouvoir effectuer un apprentissage de convergence plus rapide. Dans notre cas, nous avons utilisé une base de données d'images naturelles. Le réseau YOLOv3 a été pré-entraîné sur les données de la base ImageNet [Deng *et al.* (2009)] (tâche de classification avec 1000 classes). Nous avons utilisé ce réseau pré-entraîné sur nos données à l'exception de la dernière couche.

La détection d'organes multiples a été évaluée sur 20 patients en utilisant la validation croisée avec 10 plis. La validation croisée consiste à diviser les images d'entrée en $K = 10$ plis et chaque pli est utilisé comme un ensemble de test à un moment donné. À chaque itération de la validation croisée, nous avons donc utilisé 18 patients pour l'apprentissage et 2 patients pour le test. Dans la première expérience, le premier pli est utilisé pour tester le modèle et le reste est utilisé pour entraîner le modèle. Dans la deuxième expérience, le deuxième pli est utilisé comme un ensemble de test, et le reste est utilisé comme un ensemble d'apprentissage. Ce processus est répété jusqu'à ce que chacune des 10 plis ait été utilisé comme un ensemble de test. La division du jeu de données est identique dans les deux modalités IRM et TDM. Pour chaque pli, le modèle est sélectionné suivant la meilleure performance de détection mesurée par la précision moyenne globale mAp.

Dans l'étude expérimentale, nous avons choisi d'appliquer le modèle YOLO sur des coupes axiales et coronales 2D extraites des images 3D en raison des restrictions de mémoire du GPU et de la complexité du modèle de réseau si celui-ci devait être défini sur des images 3D. Nous utilisons les annotations de segmentation fournies pour définir les annotations des boîtes englobantes 2D afin d'entraîner les détecteurs YOLO. Par la suite, nous avons normalisé les dimensions de l'image d'entrée, ce qui donne lieu à la normalisation des coordonnées des cases englobantes.

Nous entraînons YOLO avec des époques de 450 et un taux d'apprentissage décroissant. En outre, YOLO applique une modification photométrique dans l'espace Teinte Saturation Valeur (TSV) afin d'augmenter les données. Nous n'avons pas appliqué de rotation ou de retournement afin de pouvoir distinguer les organes gauches et droits en tant que classes distinctes (par exemple poumon droit et gauche).

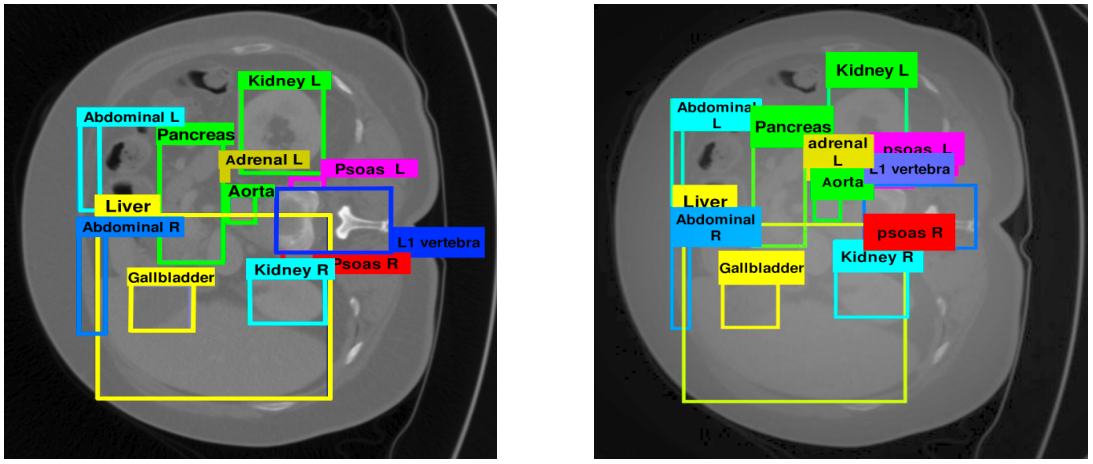
Pendant toutes les expériences, nous avons fixé les coefficients de pondération $\lambda_{pos}$, $\lambda_{obj}$, $\lambda_{noobj}$, $\lambda_{class}$ à 1.

## 3.4 Résultats

Dans cette section, nous commentons les résultats de l'application du détecteur YOLOv3 sur les images médicales de modalités TDM et IRM (pour les coupes axiales et coronales). Pour cela, nous présentons les résultats des évaluations qualitatives et quantitatives du détecteur YOLOv3.

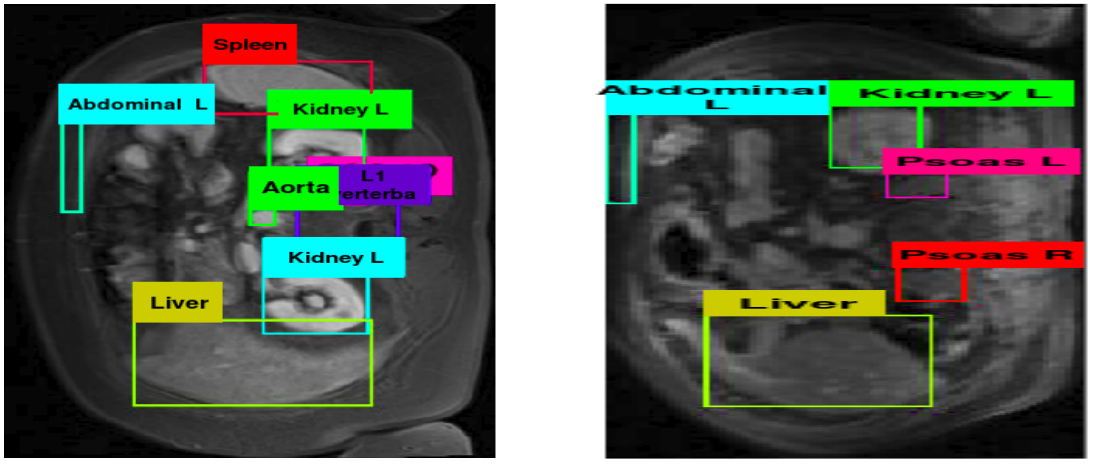
### 3.4.1 Évaluation qualitative

L'évaluation qualitative pour la détection des organes sur des images médicales nous permet de mieux visualiser le comportement du détecteur. Les Figures 3.6 et 3.7 montrent un exemple de détection multi-organe utilisant YOLO respectivement pour une image TDM axiale et une image IRM axiale.




(a) Vérité terrain. (b) Prédiction.

FIGURE 3.6 – Détection multi-organe en 2D sur une image TDM axiale.




(a) Prédiction 1. (b) Prédiction 2.

FIGURE 3.7 – Détection multi-organe en 2D sur une image IRM axiale.

La Figure 3.6 est organisée en deux vues ; la vérité terrain (Figure 3.6a) et la prédiction de détection (Figure 3.6b). Elle montre la similarité des boîtes englobantes prédites avec les boîtes de la vérité terrain. La Figure 3.7 illustre deux images IRM de prédiction 2D. Elle montre que les boîtes englobantes sont centrées sur les organes, même pour les plus petits. De plus, nous pouvons noter que le détecteur se comporte bien en présence d'une anomalie, telle que celle présente dans la Figure 3.7b, où le patient n'a qu'un seul rein.

### 3.4.2 Évaluation quantitative

Dans cette section, nous présentons l'évaluation quantitative pour les modalités TDM et IRM pour les coupes axiales et coronales. Nous évaluons la détection pour ces deux coupes 2D, du fait de l'impossibilité d'une étude directe en 3D, liée aux limitations de l'espace mémoire GPU.

Dans les Tableaux 3.2 et 3.3, nous donnons les mesures de la distance moyenne pour les modalité IRM et TDM. Ces mesures sont réalisées sur les boîtes englobantes 3D de chaque organe. Ces boîtes 3D sont obtenues en prenant les coordonnées maximales de la série de boîtes 2D fournies par le réseau.

- *Résultats obtenus à partir des coupes axiales* : Dans le Tableau 3.2a, nous observons que les distances correspondant aux organes de grande taille dans les deux modalités sont satisfaisantes, tels que la rate $6.90 \pm 7.04$ mm pour le TDM ($11.75 \pm 17.75$ mm pour l'IRM (Tableau 3.3a) et le rein gauche $5.61 \pm 12.93$ mm pour le TDM, contrairement aux organes difficiles à détecter tels que le pancréas $14.34 \pm 10.42$ mm pour le TDM et le corps musculaire du rectus abdominis gauche $55.99 \pm 5.08$ mm pour l'IRM (Tableau 3.3a).

  La distance moyenne est de $8.00 \pm 7.65$ mm pour la modalité TDM et de $22.63 \pm 9.34$ mm pour la modalité IRM.

- *Résultats obtenus à partir des coupes coronales* : De la même façon que pour les coupes axiales, la détection des grands organes est meilleure que la détection des petits organes (Tableau 3.3b). La distance pour le rein droit est de $3.39 \pm 1.71$ mm (Tableau 3.2b) pour la modalité TDM ($5.51 \pm 4.88$ mm (Tableau 3.3b) pour la modalité IRM ) contrairement au muscle droit qui est une partie difficile à détecter car il est différent d'un patient à un autre où la distance est de $10.69 \pm 5.34$ mm (Tableau 3.2b) pour la modalité TDM ($41.31 \pm 0.00$ mm (Tableau 3.3b) pour la modalité IRM).

  La distance moyenne est de $7.07 \pm 5.32$ mm pour la modalité TDM et de $17.37 \pm 8.50$ mm pour la modalité IRM.

Nous observons que les résultats globaux (moyenne des métriques sur l'ensemble des organes) sont meilleurs pour la détection obtenue à partir des coupes coronales que celle obtenue à partir des coupes axiales. Cela est dû au fait que les organes présentent plus d'informations (bien distribués) en coupe coronale.

Par ailleurs, en ce qui concerne le temps d'exécution de YOLO, nous traitons un volume TDM entier en 8 s, et un volume IRM entier en 3 s pour une inférence réalisée sur des GPU NVIDIA Tesla V100 avec 32 Go de mémoire.

TABLEAU 3.2 – Distance moyenne pour la modalité TDM pour les boîtes englobantes obtenues à partir des coupes axiales et coronales.

| | (a) Axiales. | (b) Coronales. |
|---|---|---|
| Trachée | 9.81 ± 23.97 | 3.83 ± 4.31 |
| Poumon droit | 3.80 ± 1.32 | 3.39 ± 1.71 |
| Poumon gauche | 3.52 ± 1.51 | 4.59 ± 7.58 |
| Pancréas | 14.34 ± 10.42 | 12.79 ± 7.61 |
| Vésicule biliaire | 6.95 ± 10.99 | 6.21 ± 3.30 |
| Vessie | 4.04 ± 1.21 | 5.85 ± 7.27 |
| Sternum | 7.05 ± 8.33 | 8.13 ± 9.31 |
| Vertèbre L1 | 6.24 ± 3.52 | 9.31 ± 4.98 |
| Rein droit | 4.71 ± 5.81 | 3.39 ± 5.08 |
| Rein gauche | 5.61 ± 12.93 | 3.38 ± 4.61 |
| Surrénale droite | 6.63 ± 6.54 | 7.87 ± 4.20 |
| Surrénale gauche | 8.14 ± 8.48 | 8.67 ± 7.76 |
| Psoas droit | 16.67 ± 13.48 | 10.70 ± 6.80 |
| Psoas gauche | 12.70 ± 7.10 | 11.41 ± 6.53 |
| Abdominale D. | 13.63 ± 12.18 | 10.69 ± 5.34 |
| Abdominale G. | 11.92 ± 7.33 | 9.14 ± 4.81 |
| Aorte | 4.03 ± 3.06 | 7.20 ± 4.34 |
| Foie | 7.45 ± 4.47 | 5.81 ± 2.78 |
| Thyroïde | 5.91 ± 3.22 | 5.33 ± 3.79 |
| Rate | 6.90 ± 7.04 | 3.72 ± 4.38 |
| Moyenne | 8.00 ± 7.65 | 7.07 ± 5.32 |

TABLEAU 3.3 – Distance moyenne pour la modalité IRM pour les boîtes englobantes obtenues à partir des coupes axiales et coronales.

| | (a) Axiales. | (b) Coronales. |
|---|---|---|
| Pancréas | 17.70 ± 8.33 | 16.63 ± 7.55 |
| Vésicule biliaire | 17.78 ± 10.49 | 8.26 ± 3.22 |
| Vessie | 15.12 ± 17.73 | 10.83 ± 10.08 |
| Vertèbre L1 | 13.59 ± 5.10 | 11.39 ± 4.57 |
| Rein droit | 8.84 ± 11.93 | 5.51 ± 4.88 |
| Rein gauche | 11.49 ± 15.71 | 9.88 ± 16.46 |
| Psoas droit | 12.03 ± 5.29 | 10.94 ± 5.80 |
| Psoas gauche | 13.80 ± 7.19 | 11.54 ± 7.69 |
| Abdominale D. | | 41.31 ± 0.00 |
| Abdominale G. | 55.99 ± 5.08 | 40.21 ± 7.77 |
| Aorte | 81.20 ± 0.00 | 36.54 ± 0.00 |
| Foie | 12.26 ± 7.59 | 13.44 ± 13.00 |
| Rate | 11.75 ± 17.75 | 9.31 ± 12.97 |
| Moyenne | 22.63 ± 9.34 | 17.37 ± 8.50 |

## 3.5 Conclusion

Dans ce chapitre, nous avons détaillé le fonctionnement et l'architecture du détecteur YOLO et nous sommes entrés dans les détails de sa fonction de perte. Ensuite, nous avons décrit les métriques d'évaluation de détection et la base de données utilisée dans les expériences. Puis, nous avons présenté les résultats de détection obtenus avec YOLO pour la détection multi-organe sur des images médicales de modalité IRM et TDM. Nous avons remarqué que la détection des grands organes est très satisfaisante pour les deux modalités. De plus, le temps d'inférence est faible, que ce soit pour la modalité TDM (8 s) ou pour la modalité IRM (3 s). En revanche, la limitation de cette méthode est que la détection de petits organes n'est pas satisfaisante. Cela est dû au manque d'informations sur ces organes, c'est-à-dire au manque de données d'entraînement. En imagerie médicale, le manque de données est l'une des principales limites de l'apprentissage profond.

Dans le chapitre suivant, nous développons une approche pour réduire ces limitations de détection par YOLOv3, en augmentant par des données synthétiques la base de données d'entraînement du détecteur.

Chapitre 4

# Augmentation des données pour la détection multi-organe dans les images médicales

## 4.1 Introduction

En imagerie médicale, il est difficile d'acquérir une quantité suffisante de données d'entraînement pour former un réseau neuronal profond. Le manque de données d'entraînement est l'une des principales limites de la détection basée sur l'apprentissage profond. Cela est dû au fait que l'étiquetage manuel que les experts doivent réaliser pour chaque organe s'avère long et coûteux en temps. Cette rareté de données d'entraînement rend primordial le fait de pouvoir enrichir les données initiales, ce qui permet de régulariser l'apprentissage et de minimiser l'influence de la pénurie de données. Nous appelons cette technique l'augmentation de données. L'augmentation de données peut ainsi être vue comme une forme de régularisation conçue pour améliorer les performances d'un modèle en évitant le surapprentissage.

Dans ce chapitre, nous présentons l'augmentation des données par transformation des données initiales (transformations géométriques et photométriques). Ensuite, nous détaillons les avancées récentes basées sur des modèles génératifs pour augmenter les données en générant des données synthétiques. Dans la Section 4.4, nous présentons la méthode que nous avons utilisée pour générer des images synthétiques et nous montrons comment ajouter ces images pour entraîner notre détecteur. Par la suite, nous détaillons le protocole

mis en place lors des expériences. Dans la Section 4.6, nous détaillons en premier lieu les résultats quantitatifs et qualitatifs pour les expériences de générations des images synthétiques et ensuite les résultats sur l'augmentation des données. Nous dressons finalement les conclusions et les perspectives de ce travail.

## 4.2 Augmentation des données

L'augmentation des données est une technique qui a pour but d'améliorer les performances du modèle en simulant des versions similaires des données. Dans cette section, nous présentons deux grandes familles de méthodes d'augmentation des données : méthode par transformation et méthode générative.

### 4.2.1 Méthodes par transformation

La technique d'augmentation des données par transformation est la plus utilisée dans le domaine de l'imagerie. Elle consiste à générer des données d'entraînement supplémentaires en appliquant des transformations aux données d'entraînement initiales. Les transformations utilisées sont des transformations géométriques et photométriques [Shorten and Khoshgoftaar (2019)]. Les transformations géométriques incluent le retournement (flip), la rotation, la translation, le recadrage aléatoire, le redimensionnement. Les transformations photométriques incluent la perturbation des couleurs (color jittering), le renforcement des bords, l'Analyse en Composantes Principales (PCA) [Bargoti and Underwood (2017)] et l'ajout de bruit.

Les références [Dvornik *et al.* (2017), Karsch *et al.* (2011), Su *et al.* (2015)] fournissent des exemples de détection d'objet qui utilisent les techniques d'augmentation des données standard afin d'améliorer les performances du modèle.

### 4.2.2 Méthodes génératives

L'augmentation des données avec des méthodes génératives est une approche complètement différente de l'augmentation des données par transformation. Dans cette méthode, les images d'entrée ne sont pas transformées, mais un modèle génératif est entraîné pour générer des données supplémentaires, synthétiques, similaires aux données réelles. Les auteurs [Rozantsev *et al.* (2015), Riegler *et al.* (2015), Shrivastava *et al.* (2017)] utilisent des méthodes génératives afin d'améliorer les performances de détecteurs.

L'augmentation des données par des méthodes génératives permet de régler le manque des données en générant des images semblables à des données réelles. Par contre, elle ne résout pas le manque d'étiquetage. Cette opération d'étiquetage qui est lente et coûteuse peut être allégée si l'on peut réutiliser les étiquettes obtenues sur une autre modalité (par exemple, l'IRM) pour former des données étiquetées dans une nouvelle modalité (par exemple, le TDM). Par conséquent, pour améliorer les performances de détection YOLO sur notre base de données, nous allons générer des images synthétiques d'une modalité

cible à partir des images d'une modalité source, tout en conservant les étiquettes connues des images source. Cette approche est détaillée dans la suite.

## 4.3 Synthèse d'images multi-modalités

Plusieurs travaux utilisant des réseaux antagonistes génératifs (par la suite nous allons utiliser la nomenclature en anglais Generative Adversarial Network (GAN)) pour synthétiser des images dans une modalité à partir de celles d'une autre ont été proposés [Yi *et al.* (2019)]. Ainsi l'approche dite "CycleGAN" a été proposée par [Zhu *et al.* (2017)] et est devenue l'une des approches couramment utilisée [Welander *et al.* (2018), Wolterink *et al.* (2017), Jiang *et al.* (2018), Huo *et al.* (2018)] pour la génération d'images médicales synthétiques. Il est important de noter qu'un CycleGAN peut être utilisé dans le cas de données non appariées, ce qui est particulièrement utile dans notre application, car il est rare d'avoir des images de différentes modalités pour le même patient dans les mêmes conditions. En d'autres termes, les instances ne sont pas mutuellement mises en correspondance entre les domaines source et cible.

Un CycleGAN a été utilisé dans [Wolterink *et al.* (2017)] pour générer des images TDM du cerveau à partir d'images IRM, et dans [Jiang *et al.* (2018)] pour générer des images IRM du poumon à partir d'images TDM afin de segmenter des tumeurs du poumon. Notre travail s'inspire de [Huo *et al.* (2018)] où l'objectif était de segmenter un seul organe (le foie) sans disposer d'annotations de vérité du terrain pour la modalité cible. Un CycleGAN a été utilisé pour générer des images de la modalité cible à partir des images sources étiquetées. Les étiquettes sources sont ensuite transférées à la cible. Comme nous l'avons déjà évoqué, notre travail vise la détection de plusieurs organes.

Dans la suite de cette section, nous présentons en détail le modèle génératif le plus utilisé le Réseau Antagoniste Génératif [Goodfellow *et al.* (2014)]. Ensuite, nous présentons le modèle CycleGAN [Zhu *et al.* (2017)] utilisé lors des expérimentations au cours de cette thèse pour l'augmentation des données médicales.

### 4.3.1 Réseau Antagoniste Génératif

Un réseau antagoniste génératif est un modèle génératif proposé par Ian Goodfellow en 2014 [Goodfellow *et al.* (2014)]. Un GAN est une méthode de réseau profond pour générer des images synthétiques similaires à des images réelles. Ce modèle repose sur la mise en compétition de deux réseaux : un générateur (G) et un discriminateur (D) (Figure 4.1) :

- **G** : Un réseau générateur qui prend en entrée un vecteur aléatoire et doit générer en sortie une image. Au fur et à mesure que le réseau va apprendre, les images en sortie seront de meilleure qualité. Les images générées sont soumises au discriminateur qui va les évaluer et essayer de deviner si elles sont réelles ou pas. L'entraînement du générateur est supervisé : les poids doivent être modifiés pour faire correspondre le vecteur aléatoire en entrée à une perte minimale en sortie du discriminateur.

Les prédictions faites sur un même vecteur aléatoire en entrée du générateur vont s'affiner au fil des générations à la manière d'une image dont la qualité s'améliore avec le temps.

- **D** : Un réseau discriminateur qui estime la probabilité qu'un échantillon provienne des données d'entraînement plutôt que de G. L'entraînement est supervisé et réalise cette classification binaire supervisée par régression logistique.

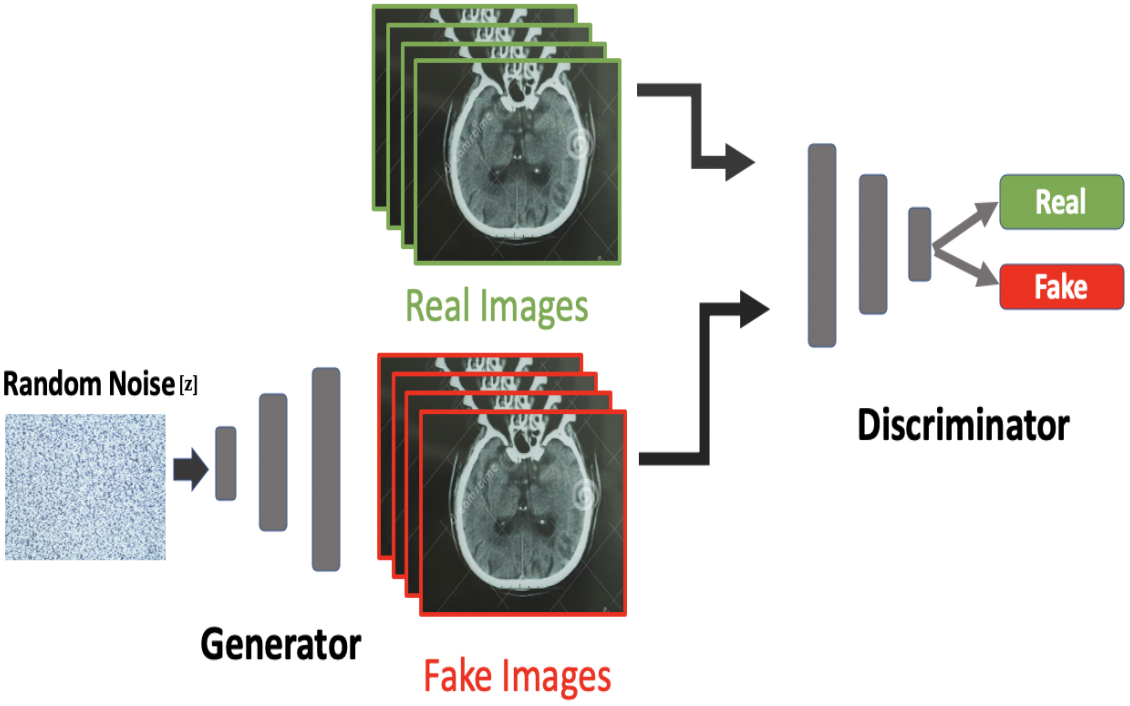



FIGURE 4.1 – Illustration d'un réseau antagoniste génératif. Source [Sadrach (2020)]

**Entraînement d'un GAN :** Un GAN est entraîné en résolvant le problème de maximisation/minimisation suivant (Équation (4.1)) [Goodfellow *et al.* (2014)] :

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}\left[\log(D(x))\right] + \mathbb{E}_{z\sim p_z(z)}\left[\log(1-D(G(z)))\right] \tag{4.1}$$

Dans cette expression, $D(x)$ est l'estimation par le discriminateur de la probabilité que l'instance de données réelles $x$ soit réelle, $\mathbb{E}_x$ est l'espérance évaluée sur l'ensemble des données réelles $x$, $G(z)$ est la sortie du générateur en fonction du bruit $z$, $D(G(z))$ est l'estimation par le discriminateur du fait que l'instance $G(z)$ soit réelle ou fausse et $\mathbb{E}_z$ est l'espérance évaluée sur l'entrée aléatoire $z$ du générateur.

Ainsi que cela a été montré dans [Goodfellow *et al.* (2014)], l'entraînement du GAN peut alors être réalisé en alternant :

– l'entraînement du discriminateur D pour un générateur G fixé, ce qui revient à réaliser l'optimisation (Équation (4.2)) :

$$\max_{D} V(D,G) \tag{4.2}$$

– l'entraînement du générateur G pour un discriminateur fixé, à savoir optimiser (Équation (4.3)) :

$$\min_{G} V(D,G) = \min_{G} \mathbb{E}_{z\sim p_z(z)} \left[\log(1 - D(G(z)))\right] \tag{4.3}$$

### 4.3.2 CycleGAN

Un CycleGAN [Zhu *et al.* (2017)] est une méthode d'apprentissage profond non supervisée qui permet une traduction bidirectionnelle entre le domaine source $X$, $\{x_i\}_{i=1}^{M}$ $(x_i \in X)$ et le domaine cible $Y$, $\{y_i\}_{i=1}^{N}$ $(y_i \in Y)$. Nous désignons la distribution des données par $x \sim p_X(x)$ et $y \sim p_Y(y)$. Elle utilise deux réseaux générateurs $G_1$, $G_2$ tels que $G_1 : X \to Y$ et $G_2 : Y \to X$, associés chacun à un réseau discriminant, $D_1$ et $D_2$ comme le montre la Figure 4.2. Les réseaux $G$ et $D$ sont en concurrence les uns avec les autres. Par la suite, nous détaillons l'architecture et la fonction de perte d'un CycleGAN.

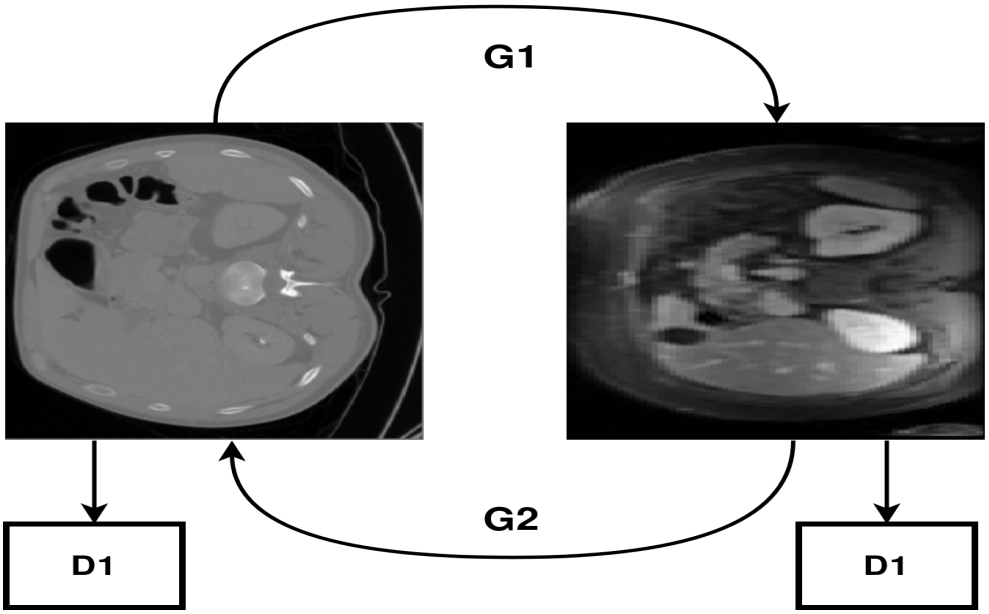



FIGURE 4.2 – Application d'un CycleGAN sur des images médicales.

**Architecture d'un CycleGAN** Les architectures des générateurs G et discriminateurs D sont présentées dans la Figure 4.3 :

– **G** : Le générateur du CycleGAN comporte trois parties : un encodeur, un transformateur et un décodeur. L'image d'entrée est introduite directement dans l'encodeur pour extraire les caractéristiques ce qui réduit la taille de la représentation tout en

augmentant le nombre de canaux. Ensuite, la sortie de l'encodeur est passée dans le transformateur qui utilise ces caractéristiques et les combine pour les transformer d'un domaine à un autre. La sortie du transformateur est ensuite passée dans le décodeur qui utilise des blocs de convolution transposée pour ramener la taille de la représentation à la taille originale.

– **D** : Le discriminateur est implémenté comme un modèle PatchGAN [Zhu *et al.* (2017)] qui vise à classer les images comme réelles ou synthétiques. Ce réseau examine une sous-image (patch) de l'image d'entrée et donne en sortie la probabilité que ce patch soit "réel" ou pas. Cette méthode est efficace car elle permet au discriminateur de se concentrer sur les caractéristiques locales (telles que la texture), qui sont généralement les éléments modifiés dans les tâches de synthèse d'images.

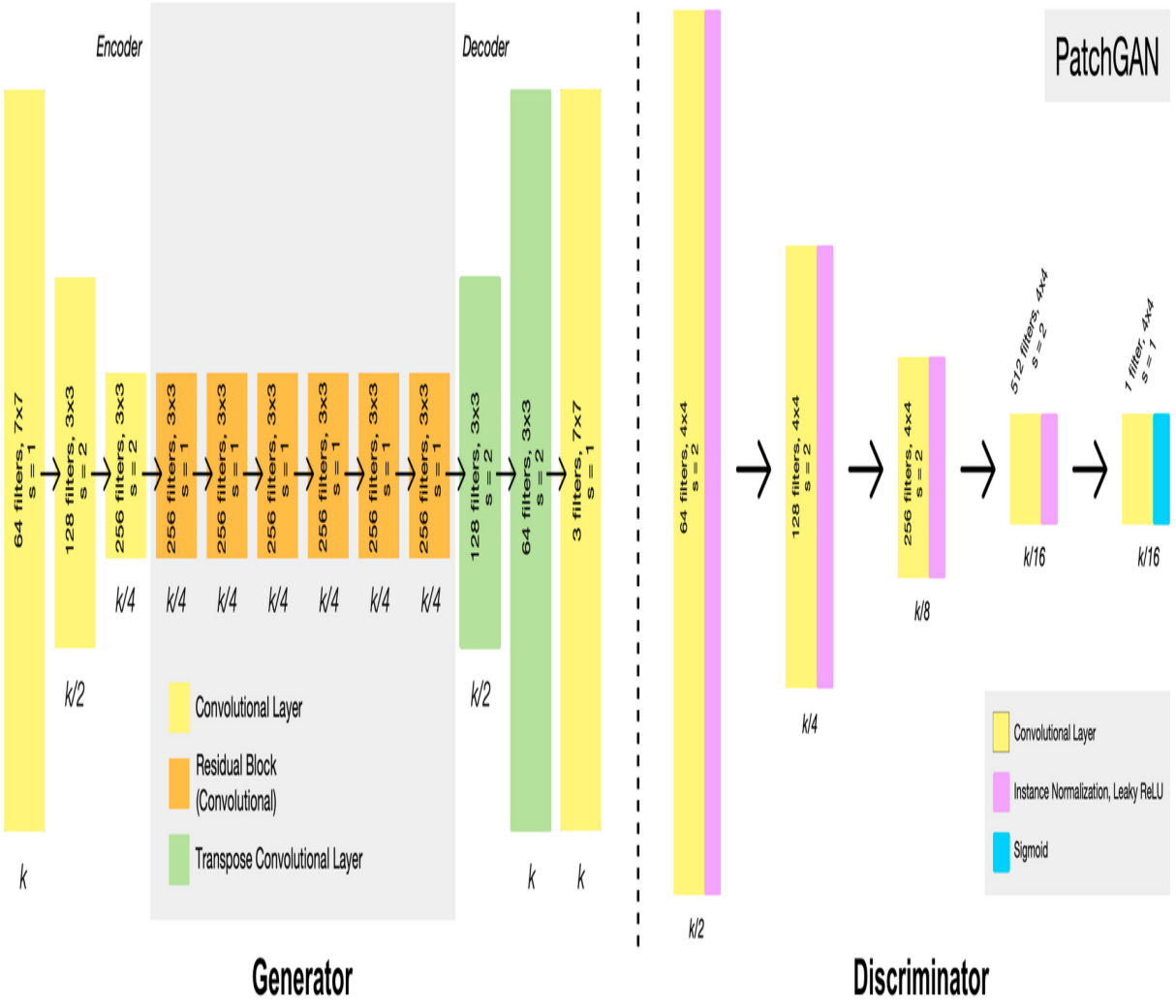



FIGURE 4.3 – Architecture du générateur et du discriminateur d'un CycleGAN. Source [Wolf (2018)]

**Fonction de perte** La fonction de perte de CycleGAN comporte deux termes : *une perte antagoniste* et *une perte de cohérence de cycle* :

– **Perte antagoniste :** La perte antagoniste d'un réseau antagoniste génératif $L_{GAN}$ est appliquée aux deux générateurs et discriminateurs (Figure 4.4a). La fonction de perte classiquement associée utilise la perte d'entropie croisée (Cross-Entropy) [Goodfellow *et al.* (2014)]. Elle a pour but de rapprocher (matching) la distribution des images générées de la distribution des données dans le domaine cible. Cette perte s'écrit comme suit (Équation (4.4) et Équation (4.5)) :

$$L_{GAN}(G_1, D_2, X, Y) = \mathbb{E}_{y \sim p_Y(y)}[\log(D_2(y))] + \mathbb{E}_{x \sim p_X(x)}[\log(1 - D_2(G_1(x)))] \quad (4.4)$$

$$L_{GAN}(G_2, D_1, Y, X) = \mathbb{E}_{x \sim p_X(x)}[\log(D_1(x))] + \mathbb{E}_{y \sim p_Y(y)}[\log(1 - D_1(G_2(y)))] \quad (4.5)$$

Ainsi $G_1$ (resp. $G_2$) génère des images $G_1(x)$ (resp. $G_2(x)$) similaire à des images du domaine cible $Y$ (resp. $X$), tandis que $D_2$ (resp. $D_1$) a pour but de distinguer entre les images synthétiques $G_1(x)$ (resp. $G_2(x)$) et les images réelles $y$ (resp. $x$).

Certains auteurs ont proposé d'utiliser une autre fonction de perte basée sur l'erreur quadratique [Mao *et al.* (2017)]. Cette fonction est plus stable pendant l'entraînement et génère des résultats de meilleure qualité [Zhu *et al.* (2017)]. Les fonctions de perte définies en (Équation (4.4)) et (Équation (4.5)) deviennent alors (Équation (4.6)) et (Équation (4.7)) :

$$L_{LsGAN}(G_1, D_2, X, Y) = \mathbb{E}_{y \sim p_Y(y)}\left[(D_2(y) - 1)^2\right] + \mathbb{E}_{x \sim p_X(x)}\left[D_2(G_1(x))^2\right] \quad (4.6)$$

$$L_{LsGAN}(G_2, D_1, Y, X) = \mathbb{E}_{x \sim p_X(x)}\left[(D_1(x) - 1)^2\right] + \mathbb{E}_{y \sim p_Y(y)}\left[D_1(G_2(y))^2\right] \quad (4.7)$$

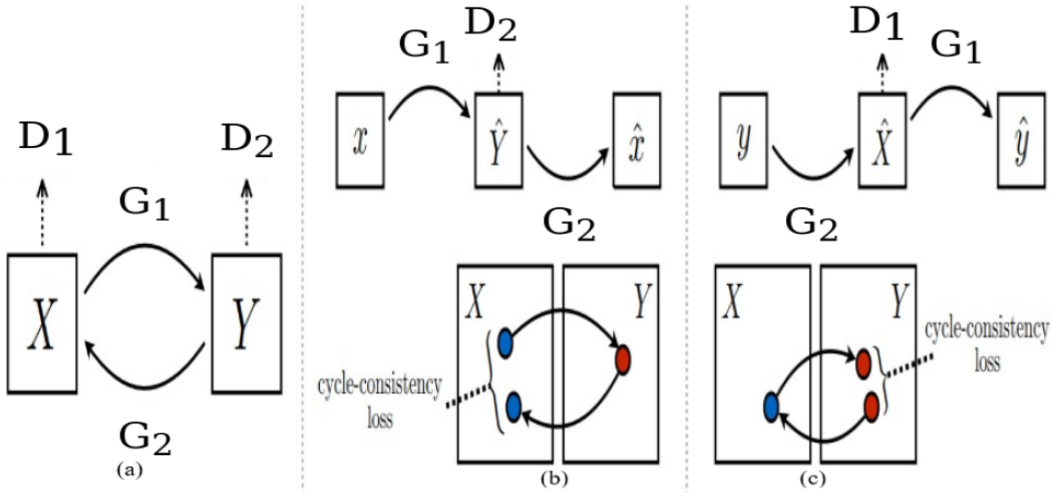



FIGURE 4.4 – Perte antagoniste et perte de cohérence du cycle. Source [Zhu *et al.* (2017)]

– **Perte de cohérence de cycle :** Cette fonction de perte $L_{cyc}$ est utilisée dans le CycleGAN pour comparer les images reconstruites avec les images réelles. L'objectif de cette fonction de perte est d'imposer une similarité entre l'image réelle ($X$) et l'image reconstruite $G_1(G_2(X))$. Cette perte utilise la norme $L_1$.

Comme le montre la Figure 4.4b, l'image $X$ est transformée via le générateur $G_1$ qui donne l'image générée $\hat{Y}$. Cette image générée $\hat{Y}$ est ensuite transformée via le générateur $G_2$ qui donne l'image générée $\hat{X}$ (l'image reconstruite).

La perte de cohérence de cycle est définie comme suit :
Perte de cohérence de cycle "direct" (Figure 4.4b) : $X \rightarrow G_1(X) \rightarrow G_2(G_1(X)) \sim \hat{X}$.
Perte de cohérence de cycle "inverse" (Figure 4.4c) : $Y \rightarrow G_2(Y) \rightarrow G_1(G_2(Y)) \sim \hat{Y}$.

$$L_{cyc}(G_1, G_2) = \mathbb{E}_{x \sim p_X(x)}[\|G_2(G_1(x)) - x\|_1] + \mathbb{E}_{y \sim p_Y(y)}[\|G_1(G_2(y)) - y\|_1] \quad (4.8)$$

La fonction de coût du CycleGAN est la somme des pertes antagonistes (Équations (4.4), (4.5)) et de la perte de cohérence de cycle (Équation (4.8)) :

$$L\left(G_1, G_2, D_1, D_2\right) = L_{GAN}\left(G_1, D_2, X, Y\right) + L_{GAN}\left(G_2, D_1, Y, X\right) + \lambda L_{cyc}\left(G_1, G_2\right) \quad (4.9)$$

où $\lambda$ contrôle l'importance relative des deux types de termes (Équation (4.9)).

Comme pour le GAN, l'optimisation du CycleGAN est alors obtenue par la formulation min-max suivante (Équation (4.10)) :

$$G_1^*, G_2^* = \arg \min_{G_1, G_2} \max_{D_1, D_2} L\left(G_1, G_2, D_1, D_2\right) \quad (4.10)$$

## 4.4 Méthode proposée

Dans cette section, nous présentons l'augmentation des données pour la détection multi-organe dans les images médicales. Comme le montre la Figure 4.5, notre approche s'articule en deux étapes :

– *La première étape* est la synthèse intermodalité par CycleGAN [Zhu *et al.* (2017)]. Celle-ci consiste à générer des images synthétiques à partir d'une modalité source (par exemple la modalité IRM) vers une autre modalité (par exemple la modalité TDM). Lors de cette génération les annotations de la modalité source sont également transférées vers la modalité cible.

– *La deuxième étape* est la détection d'organes multiples avec l'algorithme YOLOv3 [Redmon and Farhadi (2018)]. YOLOv3 a été choisi comme le détecteur en raison de sa rapidité et de sa précision, comme mentionné dans le chapitre précédent. Les images synthétiques générées par le CycleGAN sont alors utilisées, avec les annotations des images sources, pour augmenter les jeux de données d'entraînement du détecteur YOLOv3.

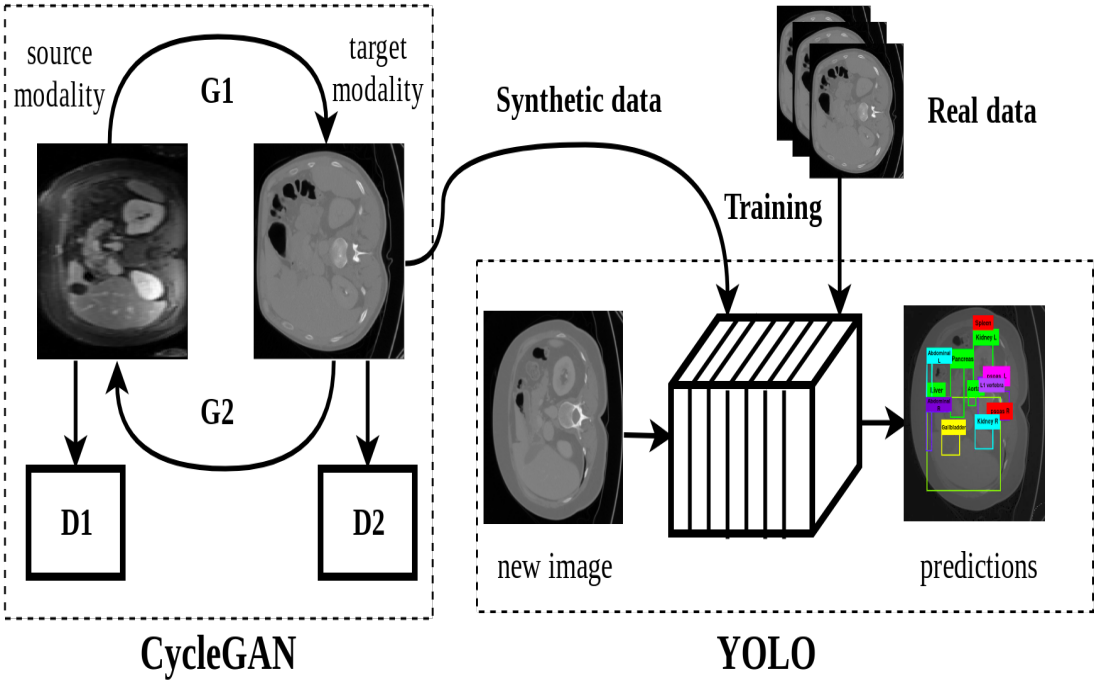



FIGURE 4.5 – Le modèle proposé : CycleGAN (synthèse d'image) + YOLO (détection multi-organe).

## 4.5 Protocole expérimental

Dans cette section, nous présentons en détail le protocole mis en place pour l'expérience de l'augmentation des données pour la détection des organes multiples. Tout d'abord (Section 4.5.1), nous exposons la métrique utilisée pour mesurer la qualité des images médicales synthétiques générées par CycleGAN. Ensuite nous détaillons comment nous avons implémenté le CycleGAN (Section 4.5.2). Dans la Section 4.5.3, nous présentons les détails du protocole d'entraînement et la sélection des hyperparamètres de l'approche CycleGAN+YOLO.

### 4.5.1 Métrique

Dans cette section, nous présentons la métrique de similarité "Mesure de l'indice de similarité structurelle" (par la suite nous allons utiliser la nomenclature en anglais Structural Similarity Index Measure (SSIM)) [Wang *et al.* (2004)] pour mesurer la qualité des images synthétiques générées par CycleGAN.

**Structural Similarity Index Measure** La métrique SSIM a été proposée par Wang et al. en 2004 [Wang *et al.* (2004)] et s'inspire du système visuel humain. La mesure de similarité fournie par SSIM est donnée par l'Équation (4.11) suivante :

$$SSIM(A,B) = \frac{(2\mu_A\mu_B + C_1)(2\sigma_{AB} + C_2)}{(\mu_A^2 + \mu_B^2 + C_1)(\sigma_A^2 + \sigma_B^2 + C_2)} \tag{4.11}$$

où $\mu_A$ (resp. $\mu_B$) est l'intensité moyenne de A (resp. B), $\sigma_A$ (resp. $\sigma_B$) est l'écart-type des intensités de A (resp. B) et $\sigma_{AB}$ est la covariance entre les intensités. $C_1 = (k_1 R)^2$, $C_2 = (k_2 R)^2$ sont deux petites constantes positives nécessaires pour stabiliser la division. R est l'étendue des intensités.

### 4.5.2 Implémentation du modèle

Comme nous l'avons expliqué auparavant, le CycleGAN est composé de deux générateurs et deux discriminateurs. Dans notre implémentation, notre discriminateur D utilise un PatchGAN de dimension 70x70, qui a pour but de déterminer si les sous-images (patches) d'entrée $70 \times 70$ qui se chevauchent sont réelles ou fausses. Le générateur G est soit un ResNet [He *et al.* (2016)] soit un U-Net [Ronneberger *et al.* (2015)]. De plus, comme nous l'avons mentionné dans la Section 4.3.2, la fonction de perte associée au CycleGAN correspond soit à un terme d'entropie croisée (Équations 4.4 et 4.5) soit à un terme quadratique (Équations 4.6 et 4.7). Dans cette expérience, chacun des deux générateurs a été implanté avec ces deux fonctions de perte. Nous comparons la qualité des résultats fournis dans les quatre cas de figure en termes de SSIM dans le Tableau 4.1. La mesure de cette qualité est réalisée en comparant, pour chaque cas, les images sources et les images reconstruites (voir Section 4.6.1.2).

Tableau 4.1 – Choix du générateur G du CycleGAN (Traduction de IRM vers TDM).

| Patient | U–Net Entropie croisée | U–Net Quadratique | ResNet Entropie croisée | ResNet Quadratique |
|---|---|---|---|---|
| 130 | 0.43 | 0.58 | 0.68 | 0.95 |
| 300 | 0.38 | 0.51 | 0.65 | 0.94 |
| 323 | 0.70 | 0.72 | 0.94 | 0.98 |
| 324 | 0.73 | 0.74 | 0.94 | 0.98 |
| 326 | 0.70 | 0.73 | 0.93 | 0.98 |
| 327 | 0.70 | 0.75 | 0.93 | 0.98 |
| 329 | 0.72 | 0.76 | 0.95 | 0.98 |
| 330 | 0.69 | 0.73 | 0.93 | 0.98 |
| 331 | 0.71 | 0.75 | 0.94 | 0.98 |
| 334 | 0.68 | 0.74 | 0.79 | 0.98 |
| 335 | 0.71 | 0.76 | 0.94 | 0.98 |
| 336 | 0.71 | 0.75 | 0.94 | 0.98 |
| 337 | 0.67 | 0.72 | 0.92 | 0.97 |
| 339 | 0.72 | 0.75 | 0.94 | 0.98 |
| 340 | 0.72 | 0.76 | 0.94 | 0.98 |
| 341 | 0.70 | 0.73 | 0.80 | 0.98 |
| 342 | 0.71 | 0.76 | 0.94 | 0.98 |
| 359 | 0.71 | 0.75 | 0.93 | 0.98 |
| 365 | 0.74 | 0.77 | 0.94 | 0.97 |
| 381 | 0.53 | 0.64 | 0.77 | 0.96 |
| Moyenne | 0.67 | 0.72 | 0.89 | **0.97** |

Les résultats montrent clairement que c'est le ResNet associé à la perte quadratique qui conduit aux meilleurs résultats. C'est donc cette configuration que nous avons sélectionnée pour la suite de ce travail.

### 4.5.3 Protocole d'entraînement

Le CycleGAN a été entraîné en utilisant 200 époques. Nous avons fixé le taux d'apprentissage à 0.0002 pour les premières 100 époques, puis nous l'avons diminué linéairement jusqu'à atteindre zéro. L'évaluation a été réalisée sur les images du jeu de données VISCERAL Gold (Section 3.3.1) avec un modèle entraîné sur celles du jeu de données VISCERAL Silver (Section 3.3.1). Pour l'évaluation expérimentale de synthèse d'images multimodales, nous recadrons les images de la modalité TDM autour de l'abdomen. Cela est dû au fait que le thorax est absent dans les images de modalité IRM. De plus, nous redimensionnons les images des modalités IRM et TDM à une résolution de $320 \times 320$ pixels.

Concernant les détails d'implémentation du YOLOv3, nous avons utilisé le protocole d'entraînement précédemment décrit dans la Section 3.3.3.

## 4.6 Résultats

Dans cette section, nous présentons les résultats de l'augmentation des données des images d'apprentissage du détecteur. Tout d'abord, nous évaluons qualitativement et quantitativement la qualité des images synthétiques générées par CycleGAN (Section 4.6.1). Nous commentons ensuite les résultats de l'intégration de ces images synthétiques avec les images d'entrée de base du détecteur YOLO (Section 4.6.2).

### 4.6.1 Synthèse d'images multi-modalités

Nous effectuons une traduction d'images d'une modalité à l'autre en utilisant un CycleGAN. Nous générons donc des images de modalité IRM à partir de la modalité TDM et nous générons des images de modalité TDM à partir de la modalité IRM.

#### 4.6.1.1 Évaluation qualitative

L'évaluation qualitative nous permet de visualiser le comportement du générateur CycleGAN sur nos données en comparant les images synthétiques aux images réelles.

Un exemple de synthèse de la modalité IRM vers la modalité TDM est présenté dans la Figure 4.6. Cette figure montre la cohérence des structures transférées avec les structures réelles de la modalité TDM. Ainsi, par exemple, les vertèbres et les reins sont plus clairs que les muscles. Inversement, la Figure 4.7 fournit un exemple de synthèse de la modalité TDM vers la modalité IRM. Là encore, nous notons que les structures de limage IRM synthétique sont similaires à l'IRM réelle. Ainsi, par exemple, le foie est clair contrairement à son apparence dans la modalité TDM.

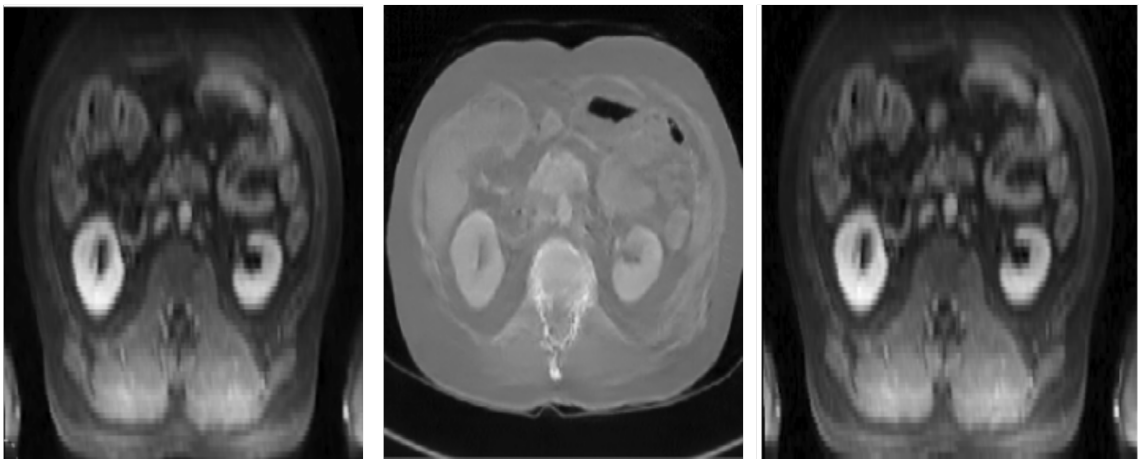


FIGURE 4.6 – Résultats qualitatifs de la génération de inter-modalités (de l'IRM à l'image TDM). L'image IRM réelle (à gauche), l'image TDM générée (au centre) et l'image IRM reconstruite (à droite).

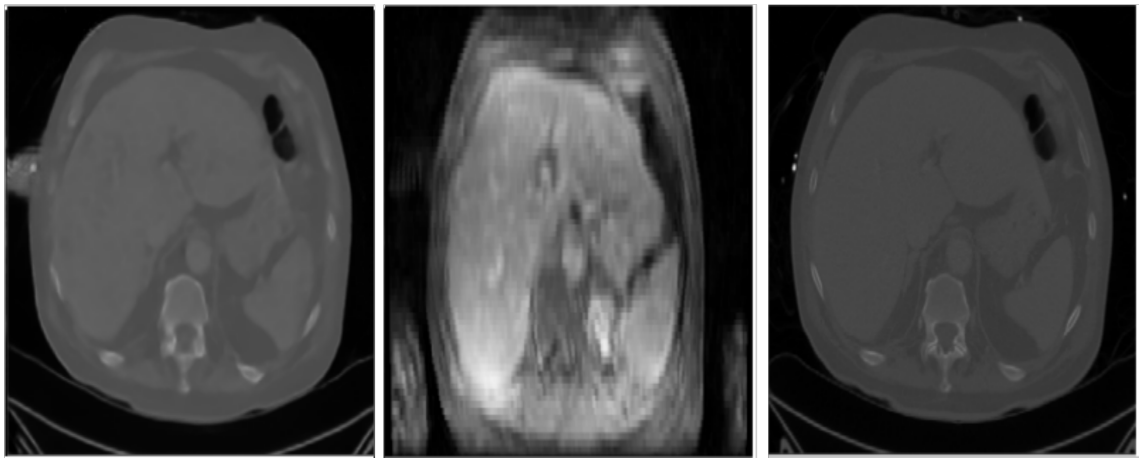


FIGURE 4.7 – Résultats qualitatifs de la génération inter-modalités (de l'image TDM à l'image IRM). L'image TDM réelle (à gauche), l'image IRM générée (au centre) et l'image TDM reconstruite (à droite).

#### 4.6.1.2 Évaluation quantitative

Pour mesurer quantitativement les performances du CycleGAN, nous utilisons l'indice de similarité Structural Similarity Index Measure (SSIM). Le calcul de cet indice pose le problème de définition des images de référence auxquelles comparer les images synthétiques. Nous proposons deux modes d'évaluation, définis ci-dessous.

**Mode d'évaluation 1 :** Cette évaluation consiste à calculer le SSIM entre une image source $X$ et sa reconstruction $G_2\left(G_1(X)\right)$ par le CycleGAN, comme le montre la Figure 4.8. Ce type d'évaluation est celui communément utilisé dans la littérature traitant de la synthèse d'images par CycleGAN [Jin *et al.* (2017)].

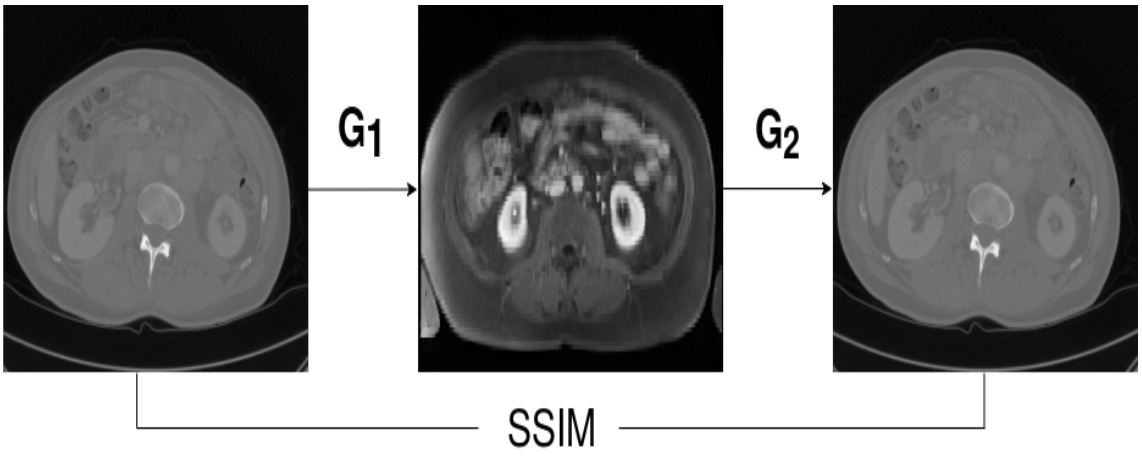



FIGURE 4.8 – Premier mode d'évaluation : Calcul du SSIM entre l'image source et l'image reconstruite.

TABLEAU 4.2 – Mode d'évaluation 1 : SSIM moyen obtenu pour chaque patient pour une traduction TDM vers IRM et IRM vers TDM.

(a) Traduction TDM vers IRM.

| Patient | mSSIM |
|---|---|
| 100 | 0.82 |
| 104 | 0.86 |
| 105 | 0.84 |
| 106 | 0.83 |
| 108 | 0.86 |
| 109 | 0.86 |
| 110 | 0.84 |
| 111 | 0.82 |
| 112 | 0.86 |
| 113 | 0.88 |
| 127 | 0.86 |
| 128 | 0.85 |
| 129 | 0.85 |
| 130 | 0.87 |
| 131 | 0.86 |
| 132 | 0.86 |
| 133 | 0.87 |
| 134 | 0.83 |
| 135 | 0.85 |
| 136 | 0.85 |
| Moyenne | 0.85 |

(b) Traduction IRM vers TDM.

| Patient | mSSIM |
|---|---|
| 130 | 0.80 |
| 300 | 0.76 |
| 323 | 0.97 |
| 324 | 0.98 |
| 326 | 0.97 |
| 327 | 0.97 |
| 329 | 0.98 |
| 330 | 0.98 |
| 331 | 0.98 |
| 334 | 0.92 |
| 335 | 0.98 |
| 336 | 0.97 |
| 337 | 0.97 |
| 339 | 0.98 |
| 340 | 0.98 |
| 341 | 0.93 |
| 342 | 0.97 |
| 359 | 0.97 |
| 365 | 0.96 |
| 381 | 0.91 |
| Moyenne | 0.95 |

Le Tableau 4.2 présente la moyenne du SSIM (mSSIM) obtenu pour chacune des coupes de chacun des cas étudiés, ainsi que la moyenne globale correspondant à l'ensemble des cas. Nous obtenons pour la génération des images de modalité IRM à partir des images de

modalité TDM (Tableau 4.2a) un SSIM global moyen de 0.85. Le Tableau 4.2b présente les résultats pour la génération des images de modalité TDM à partir des images de modalité IRM. Dans ce tableau nous obtenons un SSIM global moyen de 0.94.

Ces résultats montrent que les valeurs de SSIM obtenues sont assez élevées, mais ils indiquent cependant que la qualité des images de modalité TDM synthétiques est meilleure que la qualité des images de modalités IRM synthétiques. Cela dû au fait que la modalité IRM présente un meilleur contraste, permettant de mieux différencier des tissus de compositions différentes. De plus, les images de modalité IRM sont plus difficiles à générer aussi par une approche qui ne tient pas compte de l'inhomogénéité du champ magnétique.

Ce mode d'évaluation nous permet de mesurer la qualité des images reconstruites et d'évaluer si le cycle de génération est bien réalisé. En revanche, elle ne garantit malheureusement pas la qualité de l'image synthétique générée, comme le montrent les exemples présentés dans la Figure 4.9. De ce fait, nous proposons un deuxième mode d'évaluation, centré sur l'évaluation réelle de la qualité des images synthétiques.

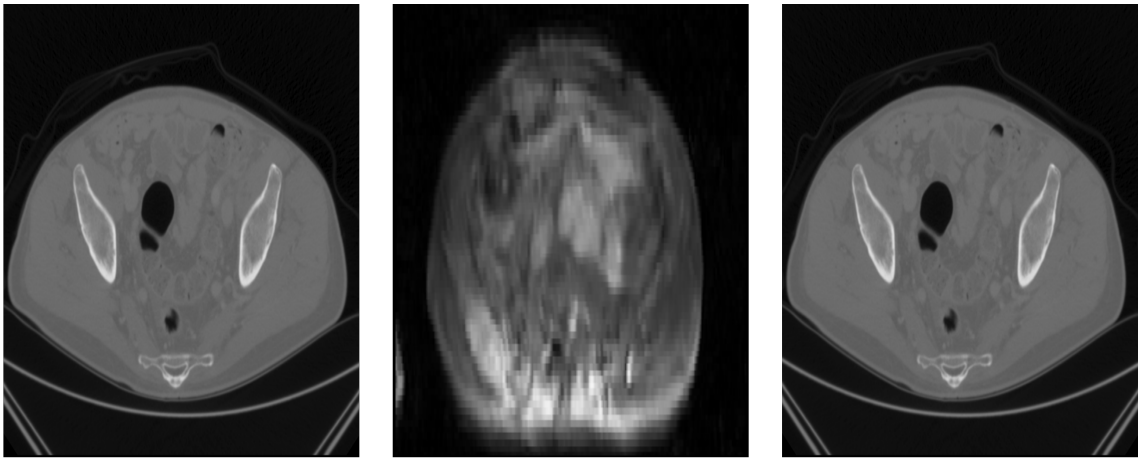


FIGURE 4.9 – Mauvaise génération inter-modalités (de l'image IRM à l'image TDM). L'image IRM réelle x (à gauche) et l'image IRM reconstruite (à droite) sont similaires par contre l'image TDM générée (au centre) est de mauvaise qualité.

**Mode d'évaluation 2 :** Puisque nous ne possédons pas de référence directe à laquelle nous pouvons comparer une image synthétique, nous proposons d'évaluer son degré de réalisme. Pour une modalité donnée, l'idée est de déterminer dans quelle mesure une image synthétique (TDM par exemple) ressemble à une image réelle, en utilisant comme référence de réalité l'ensemble des images TDM réelles de la base de données. Cette évaluation est réalisée de la façon suivante : en premier lieu, nous extrayons des images 2D réelles des volumes 3D réels disponibles. En second lieu, nous appliquons un recalage affine entre les images réelles et les images synthétiques pour les orienter toutes dans le même espace.

Chaque image synthétique d'une modalité donnée est alors comparée via la métrique de SSIM à l'ensemble des images réelles de la même modalité dans la base. L'image réelle correspondant à la valeur maximum du SSIM est retenue et deux mesures sont alors déterminées :

- Le SSIM maximum obtenu, considéré donc comme la similarité de l'image synthétique à la réalité fournie par la base de données ;
- La différence d'altitude de coupe en valeur absolue entre l'image synthétique et l'image réelle associée au SSIM maximum, afin de vérifier si les deux images correspondent aux mêmes structures anatomiques. Cette mesure sera notée DAC dans la suite.

TABLEAU 4.3 – Mode d'évaluation 2 : SSIM et différence d'altitude de coupe (DAC) moyens obtenus pour chaque patient pour une traduction TDM vers IRM et IRM vers TDM.

(a) Traduction TDM vers IRM.

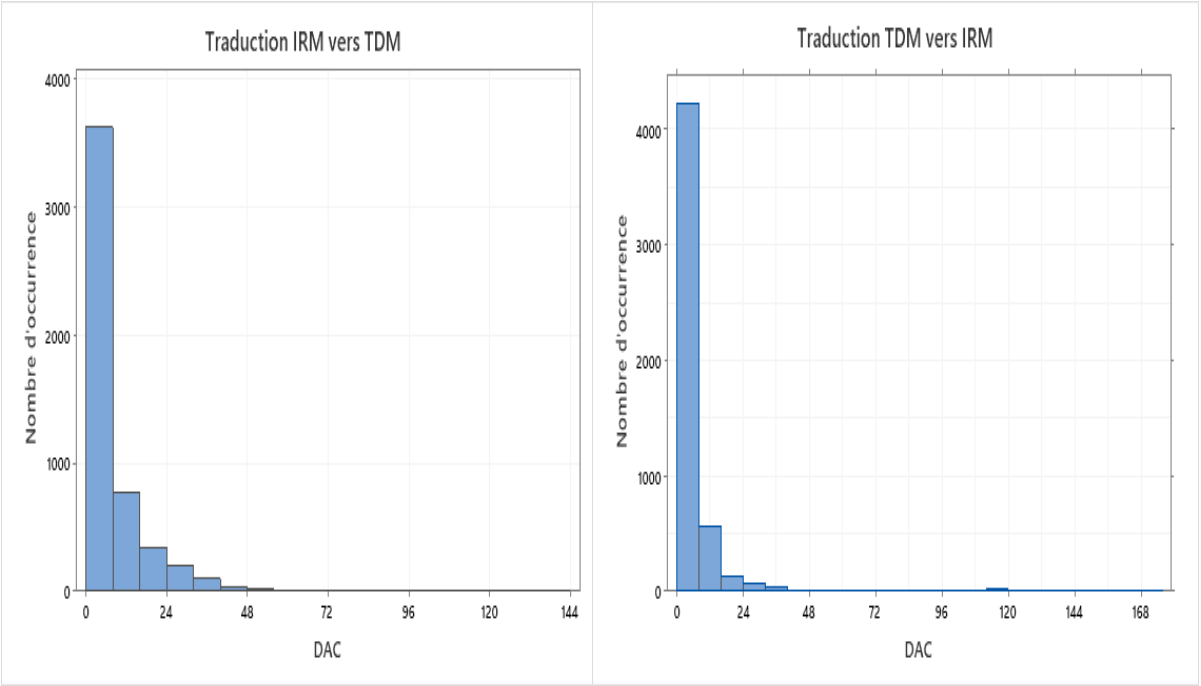
| Patient | SSIM | DAC |
|---|---|---|
| 100 | 0.71 | 7.23 |
| 104 | 0.73 | 2.94 |
| 105 | 0.74 | 5.96 |
| 106 | 0.63 | 4.18 |
| 108 | 0.70 | 3.59 |
| 109 | 0.72 | 4.45 |
| 110 | 0.70 | 6.14 |
| 111 | 0.70 | 2.69 |
| 112 | 0.71 | 3.82 |
| 113 | 0.68 | 5.70 |
| 127 | 0.71 | 3.41 |
| 128 | 0.71 | 3.14 |
| 129 | 0.72 | 4.54 |
| 130 | 0.72 | 2.61 |
| 131 | 0.72 | 4.52 |
| 132 | 0.69 | 3.08 |
| 133 | 0.71 | 3.93 |
| 134 | 0.74 | 5.89 |
| 135 | 0.77 | 7.30 |
| 136 | 0.75 | 5.77 |
| Moyenne | 0.71 | 4.55 |

(b) Traduction IRM vers TDM.

| Patient | SSIM | DAC |
|---|---|---|
| 130 | 0.75 | 10.91 |
| 300 | 0.76 | 15.10 |
| 323 | 0.80 | 4.52 |
| 324 | 0.81 | 4.08 |
| 326 | 0.80 | 6.15 |
| 327 | 0.81 | 9.78 |
| 329 | 0.82 | 5.37 |
| 330 | 0.80 | 9.86 |
| 331 | 0.78 | 5.60 |
| 334 | 0.79 | 5.22 |
| 335 | 0.81 | 6.84 |
| 336 | 0.79 | 5.63 |
| 337 | 0.80 | 6.17 |
| 339 | 0.81 | 11.56 |
| 340 | 0.81 | 7.43 |
| 341 | 0.82 | 8.72 |
| 342 | 0.79 | 7.35 |
| 359 | 0.79 | 7.99 |
| 365 | 0.76 | 9.50 |
| 381 | 0.83 | 8.32 |
| Moyenne | 0.80 | 7.80 |

Le Tableau 4.3 présente une synthèse des résultats obtenus, à savoir la moyenne du SSIM et du DAC obtenus pour l'ensemble des coupes de chacun des cas étudiés, ainsi que la moyenne globale correspondant à l'ensemble des cas.

En ce qui concerne la traduction TDM vers IRM, nous observons un SSIM global moyen de 0.71 et un DAC global moyen de 4.55 (Tableau 4.3a). Pour la traduction IRM vers TDM (Tableau 4.3b), nous obtenons un SSIM global moyen de 0.80 et un DAC global moyen de 7.80. Ces valeurs indiquent que les images synthétiques sont similaires aux images réelles. De la même façon que pour le mode d'évaluation 1, les images TDM synthétisées à partir de IRM sont meilleures en qualité relativement à la traduction TDM vers IRM.




(a) Traduction IRM vers TDM. (b) Traduction TDM vers IRM.

FIGURE 4.10 – Histogramme de la différence d'altitude de coupe.

La Figure 4.10 présente les histogrammes de la différence d'altitude de coupe de la génération des images synthétiques. Nous remarquons que 70% de DAC se concentre dans l'intervalle de $[0; 8]$ dans la Figure 4.10a de la traduction IRM vers TDM et 82% de DAC se concentre dans l'intervalle de $[0; 8]$ dans la Figure 4.10b de la traduction TDM vers IRM. Ces valeurs montrent que les images synthétiques et les images réelles correspondent aux mêmes structures.

### 4.6.2 Augmentation des données pour la détection multi-organe dans les images médicales

La détection multi-organe a été évaluée sur le jeu de données VISCERAL Gold en utilisant la validation croisée sur 10 plis dans deux scénarios, sans augmentation des données (scénario "YOLO") et avec augmentation des données (scénario "YOLO+CycleGAN"). Pour ce dernier scénario, un CycleGAN entraîné sur le jeu de données VISCERAL Silver a été utilisé pour transférer les images TDM du jeu de données Gold en images IRM, qui ont été utilisées pour augmenter les données d'entraînement de YOLO dans chacun des 10 plis. Le même scénario a été utilisé pour augmenter les données d'images TDM à partir d'images IRM. Les données de test sont identiques dans les deux scénarios. Pour chaque pli, le modèle qui a donné la meilleure performance de détection (mesurée par la métrique mAp décrite dans la Section 3.3.2) est sélectionné.

TABLEAU 4.4 – Comparaison des résultats de détection obtenus avec YOLO sur la base de données initiale et sur la base de données augmentée via le CycleGAN. Les résultats sont donnés en termes de distance moyenne pour chaque organe et chaque modalité.

(a) YOLO vs YOLO+CycleGAN pour la modalité TDM.

| | YOLO | YOLO + CycleGAN |
|---|---|---|
| Pancréas | 14.34 ± 10.42 | **10.60 ± 5.28** |
| Vésicule biliaire | **6.95 ± 10.99** | 7.47 ± 11.29 |
| Vessie | **4.04 ± 1.21** | 4.56 ± 1.65 |
| Vertèbre L1 | 6.24 ± 3.52 | **5.87 ± 3.39** |
| Rein droit | **5.61 ± 12.93** | 5.98 ± 12.42 |
| Rein gauche | 4.71 ± 5.81 | **4.39 ± 4.82** |
| Surrénale D. | 6.63 ± 6.54 | **6.37 ± 5.93** |
| Surrénale G. | 8.14 ± 8.48 | **7.86 ± 8.71** |
| Psoas droit | 16.67 ± 13.48 | **11.81 ± 6.96** |
| Psoas gauche | **12.70 ± 7.10** | 12.87 ± 5.77 |
| Abdominale D. | 13.63 ± 12.18 | **11.92 ± 6.77** |
| Abdominale G. | **11.92 ± 7.33** | 12.23 ± 7.77 |
| Aorte | 4.03 ± 3.06 | **3.93 ± 2.67** |
| Foie | 7.45 ± 4.47 | **6.92 ± 3.41** |
| Rate | 6.90 ± 7.03 | **6.54 ± 6.24** |
| Moyenne | 8.66 ± 7.63 | **7.95 ± 6.20** |

(b) YOLO vs YOLO+CycleGAN pour la modalité IRM.

| | YOLO | YOLO+ CycleGAN |
|---|---|---|
| Pancréas | 17.70 ± 8.33 | **14.93 ± 5.68** |
| Vésicule biliaire | 17.78 ± 10.49 | **13.90 ± 5.96** |
| Vessie | 15.12 ± 17.73 | **11.35 ± 12.04** |
| Vertèbre L1 | 13.59 ± 5.10 | **9.70 ± 3.12** |
| Rein droit | 11.49 ± 15.71 | **10.01 ± 15.64** |
| Rein gauche | **8.84 ± 11.93** | 10.20 ± 13.42 |
| Psoas droit | **12.03 ± 5.29** | 12.95 ± 5.69 |
| Psoas gauche | 13.80 ± 7.19 | **12.59 ± 6.31** |
| Abdominale G. | 55.99 ± 5.08 | **35.88 ± 34.32** |
| Aorte | 81.20 ± 0.00 | **37.62 ± 20.68** |
| Foie | **12.26 ± 7.59** | 14.08 ± 10.57 |
| Rate | 11.75 ± 17.75 | **10.95 ± 6.90** |
| Moyenne | 22.63 ± 9.34 | **16.18 ± 11.69** |

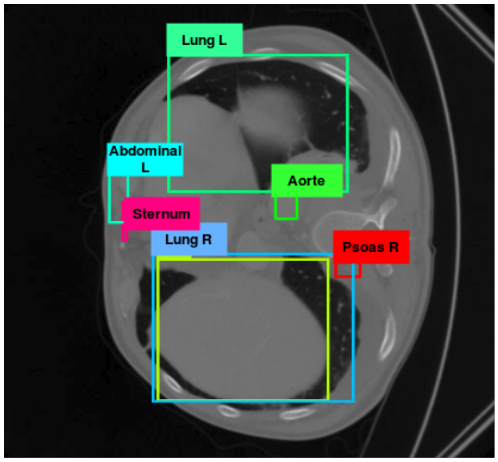
### Moyenne des distances obtenues (précision)

Comme indiqué précédemment (Section 3.4.2), la précision de la détection est mesurée en 3D sur les boîtes englobantes reconstruites en utilisant la métrique de distance moyenne. Le Tableau 4.4 donne les résultats moyens obtenus pour chaque organe (moyenne sur l'ensemble des images traitées pour un organe donné). Nous observons que les distances correspondant aux organes de grande taille dans les deux modalités, tels que la vessie (4.04 mm pour le TDM (Tableau 4.4a) et 15.12 mm pour l'IRM (Tableau 4.4b) et le rein droit (5.61 mm pour le TDM et 11.49 mm pour l'IRM) sont satisfaisantes, contrairement aux organes difficiles à détecter tels que le pancréas (14.34 mm pour le TDM) et l'aorte (81.20 mm pour l'IRM).

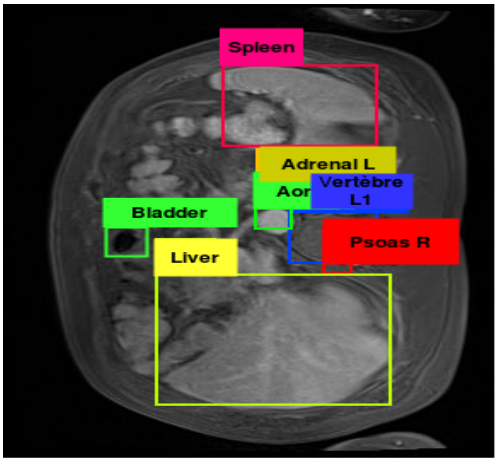
Le scénario YOLO+CycleGAN [Hammami *et al.* (2020)b] donne de bien meilleurs résultats pour la plupart des organes dans les deux modalités. Pour la modalité TDM, la distance moyenne est de 7.95 mm pour le scénario YOLO+CycleGAN par rapport à 8.66 mm pour le scénario YOLO seul. Cette amélioration est statistiquement significative ($p = 0.046$ sur un test $t$ unilatéral apparié). Pour la modalité IRM, la distance moyenne est de 16.18 mm pour le scénario YOLO+CycleGAN par rapport à 22.63 mm pour le scénario YOLO seul. Là encore, cette amélioration est statistiquement significative ($p = 0.050$ sur un test $t$ unilatéral apparié).

### Écart-type des distances obtenues (stabilité)

En contraste avec ces bons résultats globaux, le Tableau 4.4 indique également que l'écart-type des résultats est élevé pour plusieurs organes. Ainsi, par exemple, pour le rein droit cet écart-type est de 12.42 mm pour la modalité TDM et de 15.64 mm pour la modalité IRM.




(a) Détection du grand psoas dans le thorax pour la modalité TDM.




(b) Détection de la vessie dans l'abdomen pour la modalité IRM.

FIGURE 4.11 – Exemples de détections aberrantes.

Un examen attentif des prédictions obtenues a permis de confirmer que ce phénomène est principalement dû à des valeurs aberrantes dans la détection, comme l'illustre la Figure 4.11 sur deux exemples. La Figure 4.11a montre pour la modalité TDM, la détection du grand psoas qui est un organe de l'abdomen dans le thorax et la Figure 4.11b pour la modalité IRM montre la détection de la vessie (bladder) dans l'abdomen.

## 4.7 Conclusion

Dans ce chapitre, nous avons présenté notre contribution basée sur l'augmentation des données pour la détection multi-organe dans les images médicales. Nous avons ainsi proposé une combinaison YOLO+CycleGAN pour augmenter les données afin d'entraîner un détecteur multi-organe pour des images multi-modalités. Nous avons montré que cette approche permet d'améliorer les résultats de détection et conduit à une différence de distance moyenne de 0.7 mm pour la modalité TDM et à une différence de distance moyenne de 6.5 mm pour la modalité IRM. Ces résultats ont fait l'objet de communications dans deux conférences internationales [Hammami *et al.* (2020)a], [Hammami *et al.* (2020)b].

Nous avons également noté une limitation de notre approche, en cela qu'elle ne permet pas de réduire l'apparition de détections aberrantes, ce qui s'est traduit par l'observation d'écart-types élevés dans les résultats. Le chapitre suivant est consacré à une nouvelle contribution, où nous développons une approche pour réduire cette limitation en intégrant de l'a priori dans le détecteur.

Chapitre 5

# A priori anatomique pour la détection multi-organe via YOLO

## 5.1 Introduction

L'apprentissage profond a prouvé une efficacité remarquable dans le domaine de l'imagerie médicale. Cependant un volume de données d'apprentissage trop faible peut conduire à un manque de robustesse : dans ce cas, le réseau n'est en effet pas en mesure d'apprendre les contraintes anatomiques de haut niveau du corps humain (formes, adjacences, orientations ...), ce qui peut se traduire par des détections erronées. C'est pourquoi nous proposons une nouvelle méthodologie qui permet de réduire le nombre de détections erronées en intégrant des contraintes d'orientation anatomiques. Ces contraintes sont basées sur la connaissance a priori de l'orientation relative des structures anatomiques, par exemple le poumon droit se trouve au-dessus du foie, la rate se trouve à gauche du rein droit.

Dans ce chapitre, nous détaillons en premier lieu l'état de l'art de l'intégration d'a priori dans les réseaux profonds. Ensuite, nous définissons les contraintes d'orientation anatomique formellement (Section 5.3), puis dans (Section 5.4) nous détaillons de nouveaux termes pour la fonction de perte de YOLO (Équation (3.3)) afin de permettre à ce réseau de tenir compte des connaissances de l'a priori. Dans la dernière section de ce chapitre, nous exposons les résultats de détection par YOLO avec les contraintes anatomiques sur les modalités TDM et IRM.

## 5.2 Méthodes d'intégration d'a priori en apprentissage profond pour la détection

Un certain nombre de travaux ont proposé d'incorporer de l'a priori dans le but d'améliorer les performances de l'apprentissage profond. Ces travaux peuvent être classés en deux grandes familles [Xie *et al.* (2021)]. La première famille (Section 5.2.1), que nous appellerons implicites dans la suite, consiste à utiliser des connaissances extraites d'un autre jeu de données d'images naturelles ou médicales et de l'affiner sur l'ensemble de données médicales cibles : c'est l'apprentissage par transfert. La deuxième famille (Section 5.2.2), que nous qualifierons d'explicite dans la suite, consiste à intégrer l'a priori directement dans la fonction de perte du réseau.

### 5.2.1 Méthodes d'intégration d'a priori implicites

Comme nous l'avons dit plus haut, cette approche consiste à incorporer de la connaissance issue d'autres jeux de données. L'idée est d'améliorer les performances de détection en utilisant une représentation latente des structures anatomiques obtenue à partir d'un jeu de données initial et de l'incorporer pour régulariser l'apprentissage profond sur le jeu de données médicales cible.

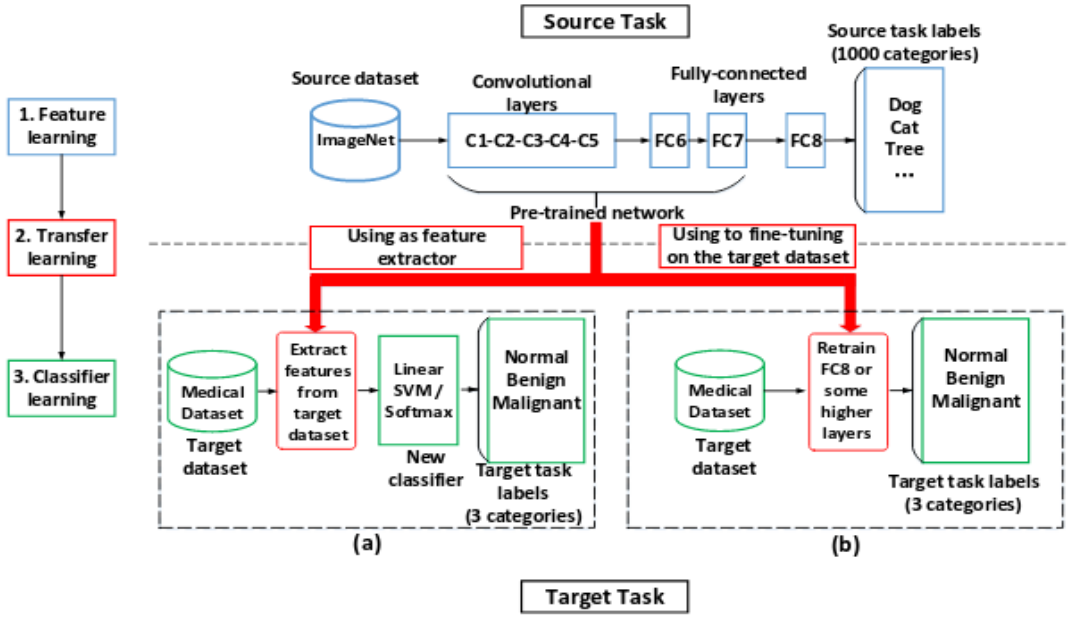



FIGURE 5.1 – Les différentes stratégies d'incorporation de connaissance du domaine dans l'apprentissage profond : (a) un extracteur de caractéristiques (b) réglage fin (fine-tuning) sur le jeu de données cible. Source [Xie *et al.* (2021)]

La première forme de ce type d'approche consiste à utiliser un jeu de données initiales non médicales. Ainsi, il est assez courant d'effectuer un pré-entraînement sur un grand ensemble de données d'images naturelles (généralement ImageNet [Deng *et al.* (2009)])

afin d'introduire des informations pour la détection d'objets dans le domaine médical. Comme l'illustre la Figure 5.1 cette forme d'approche est utilisée selon deux stratégies : (a) comme un extracteur de caractéristiques et (b) comme une initialisation qui sera affinée sur le jeu de données cible. La première approche consiste à supprimer la dernière couche de classification - entièrement connectée (fully connected) - d'un réseau pré-entraîné sur les données initiales et à ne conserver que les autres couches du réseau, qui constituent ainsi un extracteur de caractéristiques fixe. Le transfert est alors réalisé en entraînant un classifieur sur la représentation latente des données cibles fournie par cet extracteur. La deuxième approche consiste à partir du réseau pré-entraîné sur les données initiales et à affiner ensuite les poids de toutes ou de certaines couches (fine-tuning) du réseau en l'entraînant sur les données cibles. Des exemples peuvent être trouvés dans la détection des ganglions lymphatiques [Shin *et al.* (2016)], la détection des polypes et des embolies pulmonaires [Tajbakhsh *et al.* (2016)], la détection des tumeurs du sein [Yap *et al.* (2018)], la détection des polypes colorectaux [Näppi *et al.* (2016)] et [Zhang *et al.* (2016)].

La deuxième forme des approches implicites consiste à utiliser un jeu de données médicales d'une même modalité ou d'une modalité similaire. L'avantage de cette forme d'approche est que les ensembles de données médicales ont une distribution similaire. Contrairement à la première forme d'approche, les travaux relevant de cette deuxième forme utilisent des stratégies très diverses, qui sont difficiles à regrouper synthétiquement.

Ainsi par exemple, [Ben-Cohen *et al.* (2019)] utilise des images TEP pour faciliter la détection des lésions du foie pour la modalité TDM. Plus précisément, les images TEP sont d'abord générées à partir des TDM à l'aide d'une structure combinée de FCN et de GAN, puis les images TEP synthétisées sont utilisées dans une couche de réduction des faux positifs pour détecter les lésions malignes. Les résultats quantitatifs montrent une réduction de 28% du nombre moyen de faux positifs par cas. Dans une autre étude, [Zhang *et al.* (2018)] développe une stratégie de détection de masses mammaires à partir de la tomosynthèse numérique en affinant le modèle pré-entraîné sur des jeux de données de mammographie. Un autre exemple d'utilisation d'images médicales multi-modalités peut être trouvé dans la détection de tumeurs du foie [Zhao *et al.* (2020)]. Dans [Oktay *et al.* (2017)], les auteurs proposent un modèle d'auto-encodeur entraîné sur les cartes de segmentation cardiaque afin de déterminer une représentation latente des structures. Ce modèle est destiné à la tâche de segmentation, afin de minimiser le terme d'entropie croisée classique et la distance euclidienne entre la sortie du réseau et la vérité terrain dans l'espace latent fourni par l'auto-encodeur. L'approche proposée dans [Ravishankar *et al.* (2017)] incite le modèle à incorporer un a priori de forme lors de l'apprentissage du RNC, avec un premier réseau convolutif entraîné à l'aide de la stratégie d'augmentation de données. La fonction de coût de cette approche vise à minimiser la similitude entre la segmentation prédite et la vérité-terrain, ainsi que leur distance dans l'espace latent. Dans [Baumgartner *et al.* (2019)], les auteurs suggèrent une approche de segmentation basée sur un auto-encodeur. Cette approche modélise la distribution de probabilité conditionnelle des segmentations à l'aide de l'image d'entrée.

### 5.2.2 Méthodes d'intégration d'a priori explicites

La base commune de ces approches est d'intégrer un a priori (contrainte de forme, d'adjacence, de topologie, de géométrie, etc.) au travers de la modification de la fonction de perte du RNC associé au détecteur.

Ainsi, dans [BenTaieb and Hamarneh (2016)], les auteurs proposent de modéliser des descripteurs d'image de haut niveau, tels que la continuité des contours et l'interaction entre les régions (inclusion et exclusion) dans le but de les intégrer dans l'apprentissage d'un réseau. Ces deux contraintes topologiques sont directement intégrées dans la fonction de coût et optimisées conjointement à l'apprentissage. Les auteurs de [Tofighi *et al.* (2018)] proposent des contraintes de forme pour la détection de nucléus dans un RNC. Ces contraintes de forme sont créées par des experts de domaine. L'apprentissage est fait en deux parties : la première correspond à des couches dont les poids sont mis à jour par l'apprentissage et qui effectuent la détection des noyaux et la deuxième correspond à des couches de poids fixes qui guident l'apprentissage avec des informations préalables.

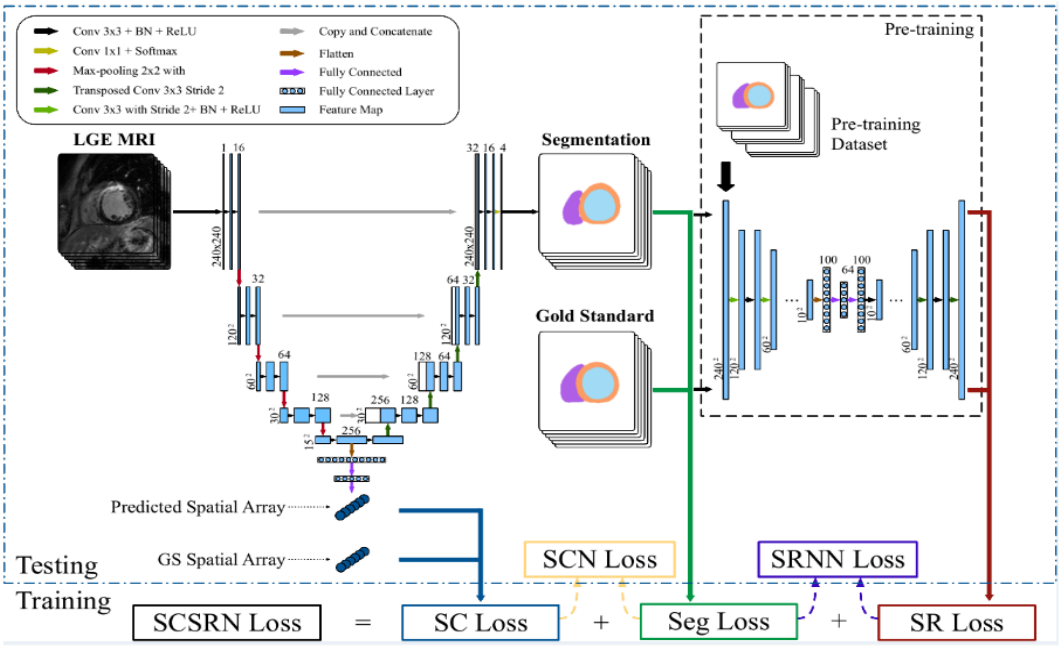



FIGURE 5.2 – Structure globale du SRSCN. Source [Yue *et al.* (2019)].

Dans [Yue *et al.* (2019)] les auteurs utilisent des contraintes spatiales et de forme dans un réseau de neurones profond pour la segmentation cardiaque. Ces contraintes sont intégrées dans la fonction de perte (Figure 5.2) avec un terme de contrainte spatiale (SC) pour aider à la segmentation et un terme de reconstruction de forme (SR) pour la régularisation de la forme. Dans [Ganaye *et al.* (2019)] les auteurs proposent d'intégrer une contrainte d'adjacence dans la fonction de coût d'un RNC afin de réduire les anomalies de la segmentation. L'objectif de cette contrainte est de pénaliser les adjacences non cohérentes avec l'anatomie humaine.

Dans la suite, l'ajout de l'a priori dans le détecteur YOLOv3 que nous proposons appartient à cette famille de méthodes explicites.

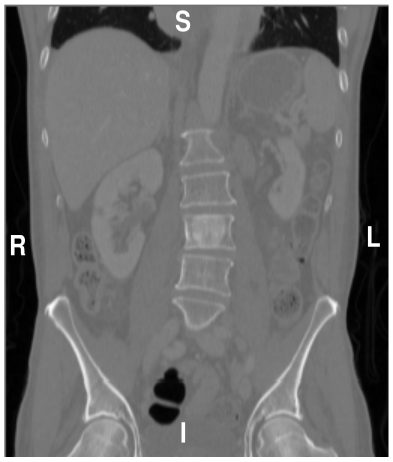
## 5.3 Contrainte d'orientation anatomique

Dans cette section, nous définissons l'a priori que nous allons intégrer dans le détecteur YOLOv3 : la contrainte d'orientation anatomique. Ensuite, nous détaillons comment nous avons formulé et représenté cette contrainte.

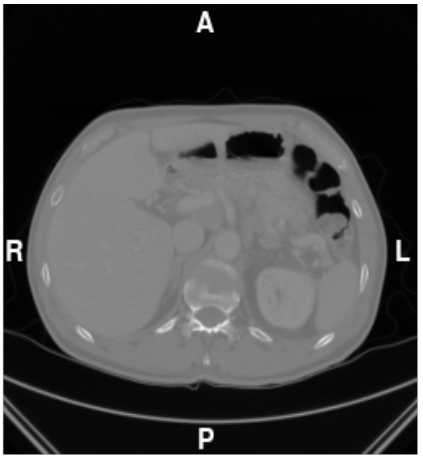
### 5.3.1 Définition de la contrainte

Tous les patients présentent la même anatomie et les mêmes relations inter-organes même si leurs géométries (forme, volume) peuvent varier. L'orientation joue un rôle important dans l'ensemble des informations disponibles dans les annotations des régions anatomiques. En effet, elle est invariable d'un patient à un autre. Les organes ne changent pas de positions relatives : ainsi par exemple le foie est toujours à droite du rein gauche et en dessous du poumon droit. Par conséquent, nous définissons un nouvel a priori appelé : contrainte d'orientation anatomique. Cette contrainte est définie par un ensemble d'orientation $O \in \{SI, AP, GD\}$, avec SI : Supérieur/Inférieur, AP : Antérieur/Postérieur, GD : Gauche/Droit.

La Figure 5.3 illustre ces relations d'orientation selon deux coupes : la coupe coronale (Figure 5.3a) implique les relations supérieur/inférieur et gauche/droit et la coupe axiale (Figure 5.3b) implique les relations antérieur/postérieur et gauche/droit.




(a) Coupe coronale TDM contrastée.




(b) Coupe axiale TDM contrastée.

FIGURE 5.3 – Illustration de la contrainte d'orientation anatomique : Supérieur/Inférieur, Gauche/Droit, Antérieur/Postérieur.

### 5.3.2 Représentation des structures anatomiques

Les relations d'orientation entre les organes peuvent être représentées par un graphe orienté. Nous en donnons ci-dessous quelques exemples pour chaque orientation :

**Supérieur/Inférieur (SI) :** La Figure 5.4a représente le graphe d'orientation supérieur/inférieur dans une coupe coronale 2D présentant 6 organes : le poumon droit, le poumon gauche, la vessie, le rein droit, le foie et la rate. Nous remarquons que le poumon droit est spatialement supérieur au foie, le foie est supérieur à la vessie. À l'inverse, le poumon droit et le poumon gauche ne peuvent être reliés par une relation d'orientation supérieure/inférieure, dans la mesure où ils ne peuvent être séparés selon un plan perpendiculaire à cette orientation.

**Antérieur/Postérieur (AP) :** La Figure 5.4b présente le graphe d'orientation antérieur/postérieur dans une coupe sagittale 2D présentant 5 organes : le sternum, la vertèbre L1, le muscle abdominal droit, l'aorte et le foie. Nous remarquons que le muscle abdominal droit est antérieur relativement à la vertèbre L1 et au même niveau que le sternum. Par contre, l'aorte ne présente aucune information d'orientation antérieure/postérieure sauf pour la vertèbre qui est en position d'antériorité.

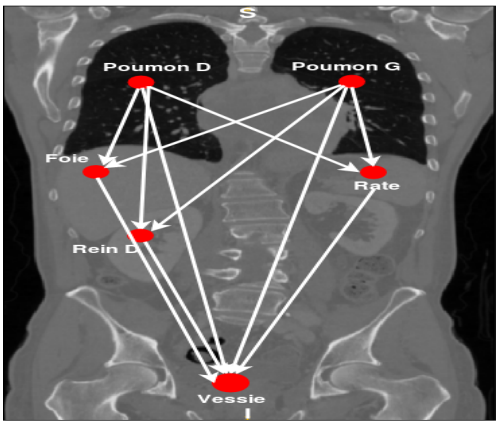
**Gauche/Droit (GD) :** La Figure 5.4c représente le graphe d'orientation gauche/droit dans une coupe axiale 2D présentant 5 organes : le pancréas, la vertèbre L1, le rein gauche, le foie et la rate. Nous remarquons que le pancréas se situe à la gauche du foie et la rate se trouve à gauche de tous les organes. Le pancréas est un organe de forme très variable et ne se trouve ni à droite ni à gauche par rapport au rein gauche.

Nous représentons dans la suite ces graphes sous la forme d'une matrice d'orientation $R^O$ avec $O \in \{SI, AP, GD\}$. Les termes de la matrice sont déterminés comme suit (Équation (5.1)) :

$$r^O_{i,j} = \begin{cases} 1 & \text{si la relation selon l'orientation } O \text{ entre les structures } i \text{ et } j \text{ est vérifiée} \\ 0 & \text{sinon} \end{cases} \tag{5.1}$$

Les graphes précédents (Figures 5.4a, 5.4b et 5.4c) sont alors représentés par les matrices d'orientation suivantes :

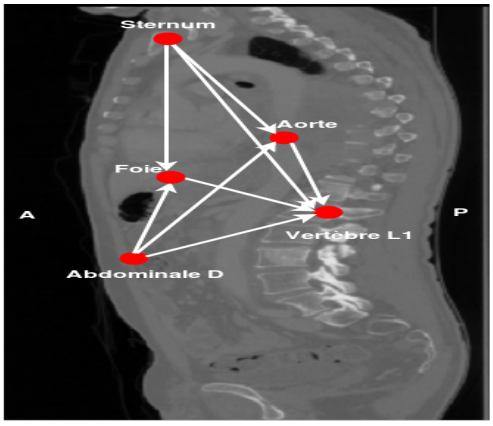
**SI :** L'Équation (5.2) représente la matrice d'orientation supérieur/inférieur pour le graphe donné en Figure 5.4a.
Les termes $r^{SI}_{\text{poumon droit,vessie}} = 1$ et $r^{SI}_{\text{poumon droit,poumon gauche}} = 0$ montrent que le poumon droit est supérieur à la vessie, et qu'il est ni supérieur ni inférieur par rapport au poumon gauche.




(a) Graphe d'orientation pour les relations Supérieur/Inférieur sur 6 organes présents sur une coupe coronale d'une image TDM.

$$\begin{array}{l} \\ \text{Poumon D.} \\ \text{Poumon G.} \\ \text{Vessie} \\ \text{Rein D.} \\ \text{Foie} \\ \text{Rate} \end{array} \begin{array}{c} \begin{array}{cccccc} \text{P. D.} & \text{P. G.} & \text{Vessie} & \text{Rein D.} & \text{Foie} & \text{Rate} \end{array} \\ \begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} \end{array} \tag{5.2}$$

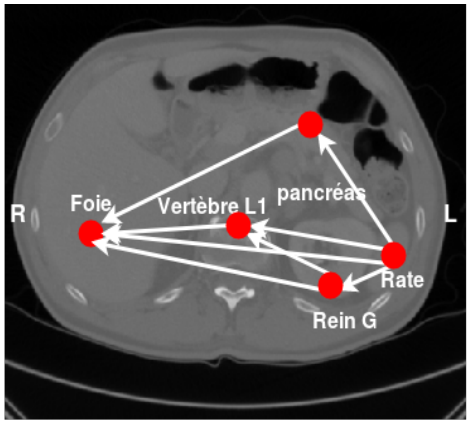
(a) Matrice d'orientation pour les relations Supérieur/Inférieur sur 6 organes.




(b) Graphe d'orientation pour les relations Antérieur/Postérieur sur 5 organes présents sur une coupe sagittale d'une image TDM.

$$\begin{array}{l} \\ \text{Sternum} \\ \text{Vertèbre L1} \\ \text{Abdominal D.} \\ \text{Aorte} \\ \text{Foie} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Sternum} & \text{Vertèbre L1} & \text{A. D.} & \text{Aorte} & \text{Foie} \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \end{array} \tag{5.3}$$

(b) Matrice d'orientation pour les relations Antérieur/Postérieur sur 5 organes.




(c) Graphe d'orientation pour les relations Gauche/Droit sur 5 organes présents sur une coupe axiale d'une image TDM.

$$\begin{array}{l} \\ \text{Pancréas} \\ \text{Vertèbre L1} \\ \text{Rein G.} \\ \text{Foie} \\ \text{Rate} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{Pancréas} & \text{Vertèbre L1} & \text{Rein G.} & \text{Foie} & \text{Rate} \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \end{array} \tag{5.4}$$

(c) Matrice d'orientation pour les relations Gauche/Droit sur 5 organes.

FIGURE 5.4 – Illustration d'un graphe d'orientation et de la matrice d'orientation.

**AP :** L'Équation (5.3) représente la matrice d'orientation antérieur/postérieur pour le graphe donné en Figure 5.4b.
Les termes $r^{AP}_{\text{muscle abdominale,aorte}} = 1$ et $r^{AP}_{\text{muscle abdominale,sternum}} = 0$ montrent que le muscle abdominal est antérieur par rapport à l'aorte et qu'il est ni antérieur ni postérieur par rapport au sternum.

**GD :** L'Équation (5.4) représente la matrice d'orientation gauche/droit pour le graphe donné en Figure 5.4c.
Le terme $r^{GD}_{\text{rate,pancréas}} = 1$ montre que la rate est à gauche du pancréas et le terme $r^{GD}_{\text{pancréas,foie}} = 1$ montre que le pancréas est à gauche du foie.

Les matrices d'orientation utilisées tout au long de cette approche pour les 20 organes sont présentées dans l'annexe pour l'orientation supérieur/inférieur (Équation (A.1)), antérieur/postérieur (Équation (A.2)) et gauche/droite (Équation (A.3).

## 5.4 Intégration de la contrainte anatomique

Nous désirons entraîner le détecteur YOLOv3 en tenant compte des contraintes d'orientation anatomiques présentées dans la section précédente. Cette contrainte est basée sur la matrice d'orientation dont les termes sont binaires. De ce fait, nous avons choisi l'entropie croisée binaire comme fonction de perte pour cette contrainte. Cette fonction de perte est une fonction différentiable et pourra donc être intégrée sans difficulté dans le processus de rétropropagation du gradient de la fonction de perte. Dans l'Équation (5.5) ci-dessous, nous avons intégré la perte de contrainte $L_{cont}$ dans la fonction de perte de YOLOv3 (Équation (3.3) détaillée dans la Section 3.2.2).

$$
\begin{aligned}
&\underbrace{\lambda_{pos} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} \left[ \left(\sigma(t_x) - \sigma(\hat{t}_x)\right)^2 + \left(\sigma(t_y) - \sigma(\hat{t}_y)\right)^2 + \left(t_w - \hat{t}_w\right)^2 + \left(t_h - \hat{t}_h\right)^2 \right]}_{L_{pos}} \\
&\quad + \underbrace{\lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} BCE(\hat{C}_i, C_i) + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{noobj} BCE(\hat{C}_i, C_i)}_{L_{conf}} \\
&\quad + \underbrace{\lambda_{class} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{obj} \left[ \sum_{k=1}^{C} BCE\left(\hat{q}_k, \sigma(s_k)\right) \right]}_{L_{class}} \\
&\quad + \underbrace{\lambda_{cont} \sum_{i \in S} \sum_{j \in B} \sum_{i' \in S^O} \sum_{j' \in B} \mathbb{1}_{i,j}^{obj} \mathbb{1}_{i',j'}^{obj} BCE\left(\hat{r}^O_{ij,i'j'}, r^O_{ij,i'j'}\right)}_{L_{cont}}
\end{aligned}
\tag{5.5}
$$

Le terme supplémentaire de perte de contrainte anatomique, $L_{cont}$, utilise l'entropie croisée binaire qui s'écrit sous la forme détaillée suivante (Équation (5.6)) :

$$BCE\left(\hat{r}^{O}_{ij,i'j'}, r^{O}_{ij,i'j'}\right) = -\hat{r}^{O}_{ij,i'j'} \log\left(r^{O}_{ij,i'j'}\right) - (1 - \hat{r}^{O}_{ij,i'j'}) \log\left(1 - r^{O}_{ij,i'j'}\right) \tag{5.6}$$

Les différents éléments de l'expression globale de $L_{cont}$ sont les suivants :

$\lambda_{cont}$ est le paramètre de pondération du terme de contrainte d'orientation anatomique ;

$S$ est le nombre de cellules de l'image ;

$B$ est le nombre de boîtes englobantes ;

$S^O$ est le nombre de cellules dans la direction $O$ ;

$\mathbb{1}^{obj}_{i,j}$ vaut 1 si un objet est détecté dans la boîte englobante $j$ associé à la $i^{\text{ème}}$ cellule, et vaut 0 dans le cas contraire ;

$\mathbb{1}^{obj}_{i',j'}$ vaut 1 si un objet est détecté dans la boîte englobante $j'$ associé à la $i'^{\text{ème}}$ cellule, et vaut 0 dans le cas contraire ;

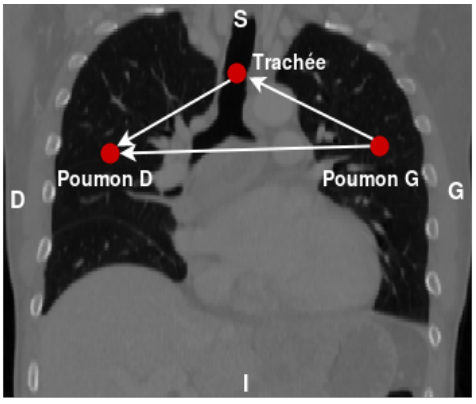
$r^{O}_{ij,i'j'}$ Ce terme représente la valeur de l'orientation prédite par le réseau. Il est calculé à partir de la prédiction entre l'objet qui se trouve dans la boîte englobante $j$ associé à la $i^{\text{ème}}$ cellule et l'objet dans l'orientation O qui se trouve dans la boîte englobante $j'$ associée à la $i'^{\text{ème}}$ cellule ;

$\hat{r}^{O}_{ij,i'j'}$ Ce terme représente la valeur de l'orientation fournie par la vérité terrain. Il est donc associé à l'objet cible qui se trouve dans la boîte englobante $j$ associé à la $i^{\text{ème}}$ cellule et l'objet cible dans l'orientation O qui se trouve dans la boîte englobante $j'$ associée à la $i'^{\text{ème}}$ cellule ;

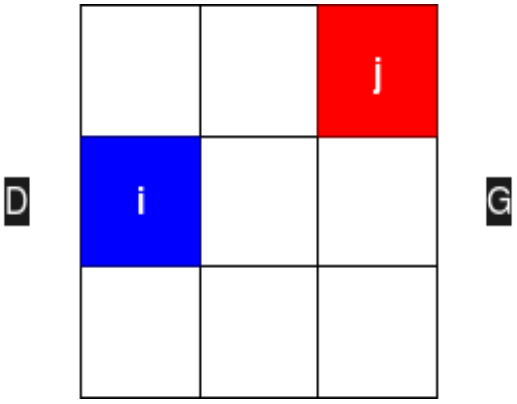
Nous présentons dans la Figure 5.5 le fonctionnement de la contrainte d'orientation anatomique en utilisant l'entropie croisée binaire. La Figure 5.5a représente le graphe d'orientation gauche/droit dans une coupe coronale 2D présentant 3 organes : la trachée, le poumon gauche et le poumon droit. Le poumon gauche se trouve à gauche de la trachée et du poumon droit et la trachée se trouve à gauche du poumon droit.

La Figure 5.5b présente les cellules d'une image avec la cellule $i$ qui se trouve à droite et la cellule $j$ qui se trouve à gauche. Nous illustrons deux différent cas :

- 1$^{\text{er}}$ Cas : Les relations sont vérifiées, par exemple $i$ = poumon droit et $j$ = poumon gauche. Dans ce cas, nous avons pour le terme de la matrice d'orientation de la vérité terrain $\hat{r}^{GD}_{i,j} = 1$. L'entropie croisée binaire se réduit alors au terme $-\log\left(r^{GD}_{i,j}\right)$. La minimisation de ce terme implique donc que le terme de la matrice d'orientation prédit $r^{GD}_{i,j}$ tende vers 1.




(a) Graphe d'orientation pour les relations Gauche/Droit sur 3 organes présents sur une coupe coronale d'une image TDM.




(b) Les cellules d'une image avec la cellule i qui se trouve à droite et la cellule j qui se trouve à gauche.

FIGURE 5.5 – Illustration du fonctionnement de l'entropie croisée binaire.

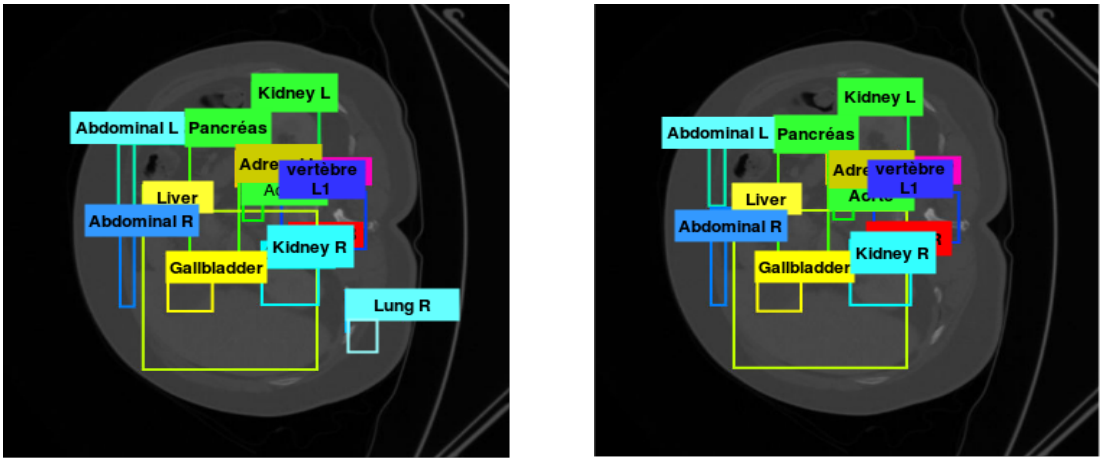
- $2^{\text{ème}}$ Cas : Les relations ne sont pas vérifiés, exemple $i$ = poumon gauche et $j$ = trachée. Dans ce cas, nous avons pour le terme de la matrice d'orientation de la vérité terrain $\hat{r}_{i,j}^{GD} = 0$. L'entropie croisée binaire se réduit alors au terme $-\log\left(1 - r_{i,j}^{GD}\right)$. La minimisation de ce terme implique donc que le terme de la matrice d'orientation prédit $r_{i,j}^{GD}$ tende vers 0.

## 5.5 Résultats

Dans cette section, nous présentons les résultats obtenus en intégrant la contrainte d'orientation anatomique dans le détecteur YOLOv3. Tout d'abord, nous évaluons qualitativement l'influence de la contrainte anatomique sur la détection de valeurs aberrantes (outliers) pour les modalités IRM et TDM. Ensuite, nous évaluons quantitativement l'intérêt de cette contrainte en comparant les résultats obtenus avec YOLOv3 seul (à savoir sans contrainte). Dans la Section 5.5.2.2, nous évaluons l'apport de la contrainte anatomique lorsque celle-ci est appliquée après l'augmentation des données décrite au chapitre précédent. Tout au long de ces expériences, nous avons fixé le paramètre pondérant le terme de contrainte d'orientation anatomique $\lambda_{cont}$ à 1.

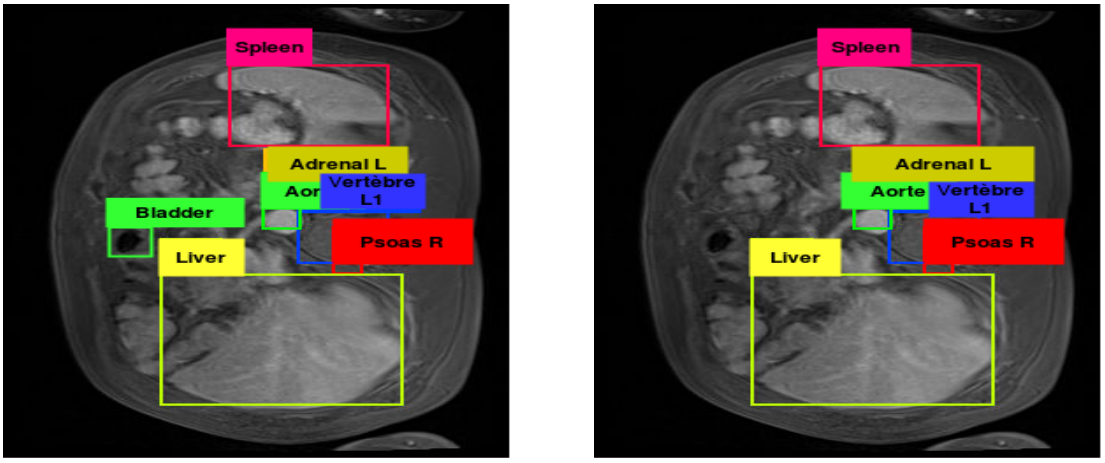
### 5.5.1 Évaluation qualitative

Dans le Chapitre 4, nous avons vu que l'approche de l'augmentation des données avec un CycleGAN pour la détection multi-organes avec YOLOv3 présentait une limitation, liée à l'apparition de détections aberrantes, qui conduisait dans certains cas à des écarts-types élevés sur la distance moyenne. Dans cette section, nous évaluons qualitativement l'influence de la contrainte anatomique sur la détection des structures multiples pour les modalités IRM et TDM.




(a) Détection du poumon droit dans l'abdomen. (b) Correction de la détection du poumon droit.

FIGURE 5.6 – Exemple de correction de fausse détection après l'intégration des contraintes d'orientation pour la modalité TDM.




(a) Détection de la vessie dans l'abdomen. (b) Correction de la détection de la vessie.

FIGURE 5.7 – Exemple de correction de fausse détection après l'intégration des contraintes d'orientation pour la modalité IRM.

Les Figures 5.6a et 5.7a illustrent respectivement une mauvaise détection du poumon droit dans l'abdomen pour la modalité TDM et une mauvaise détection de la vessie dans le haut de l'abdomen. Ces fausses détections ont été corrigées en intégrant des contraintes d'orientation de relation supérieur/inférieur et gauche/droit comme le montre la Figure 5.6b et la Figure 5.7b.

### 5.5.2 Évaluation quantitative

Pour étudier toutes les relations d'orientation antérieur/postérieur, supérieur/inférieur et gauche/droit, nous allons appliquer les contraintes sur deux orientations de coupe, à savoir les coupes coronales et axiales. En effet, une orientation de coupe donnée nous permet d'évaluer deux contraintes projetées dans le plan de ces coupes. Les relations

supérieur/inférieur et gauche/droit seront étudiées pour les coupes coronales et les relations antérieur/postérieur et gauche/droit pour les coupes axiales.

#### 5.5.2.1 Intégration des contraintes d'orientation anatomiques

Les résultats de l'application de la contrainte d'orientation anatomique aux coupes coronales et axiales pour la modalité TDM sont présentés respectivement dans les Tableaux 5.1a et 5.1b. Celles de la modalité IRM sont présentés dans les Tableaux 5.2a et 5.2b. Nous intégrons d'abord une seule contrainte ($2^{\text{ème}}$ colonne) puis deux contraintes ($3^{\text{ème}}$ colonne).

*Résultats obtenus à partir des coupes axiales* : Le Tableau 5.1a (Tableau 5.2a pour la modalité IRM) montre les résultats d'intégration de la contrainte d'orientation anatomique de la modalité TDM pour les coupes axiales. Avec l'intégration d'une seule contrainte gauche/droit, nous observons que la détection est améliorée pour la plupart des organes. Ainsi par exemple, le rein gauche présente une distance de $2.21 \pm 1.06$ mm ($7.31 \pm 9.1$ mm pour l'IRM) contre une distance de $5.61 \pm 12.93$ mm ($8.84 \pm 11.93$ mm pour l'IRM) pour YOLO seul. L'intégration de deux contraintes d'orientation anatomique gauche/droit et antérieur/postérieur permet d'améliorer les résultats de façon encore plus importante, en particulier pour les petits organes. Ainsi par exemple, la vésicule biliaire présente une distance de $4.87 \pm 1.99$ mm ($12.04 \pm 7.16$ mm pour l'IRM) contre une distance de $6.95 \pm 10.99$ mm ($17.78 \pm 10.49$ mm pour l'IRM) pour YOLO seul.

Pour les coupes axiales, nous avons une amélioration significative de la distance moyenne lorsqu'une seule contrainte est intégrée avec un $p$-value $p = 0.0003$ ($p = 0.028$ pour l'IRM) et lorsque deux contraintes sont intégrées avec $p = 7.82 * 10^{-5}$ ($p = 0.018$ pour l'IRM).

*Résultats obtenus à partir des coupes coronales* : Les résultats obtenus pour les coupes coronales sont similaires à ceux observés pour les coupes axiales. Le Tableau 5.1b (Tableau 5.2b pour la modalité IRM) montre les résultats d'intégration de la contrainte d'orientation anatomique de la modalité TDM. Avec l'ajout d'une seule contrainte, nous remarquons une amélioration pour la plupart des structures anatomique exemple pour la vessie qui a une distance moyenne de $3.47 \pm 1.58$ mm ($7.37 \pm 7.74$ mm pour l'IRM) contre $5.85 \pm 7.27$ mm avec YOLO seul ($10.83 \pm 10.08$ mm pour l'IRM). Le foie qui a une distance moyenne de $4.76 \pm 2.66$ ($9.90 \pm 10.52$ pour l'IRM) contre $5.81 \pm 2.78$ mm ($13.44 \pm 13.00$ mm pour l'IRM) pour YOLO seul. Avec l'intégration des deux contraintes d'orientation anatomiques gauche/droit et supérieur/inférieur, nous remarquons une amélioration pour tous les organes même pour les organes difficiles à détecter exemple le pancréas qui est un organe très variable en forme. Il présente une amélioration de distance de $8.05 \pm 3.17$ mm ($10.45 \pm 4.64$ mm pour l'IRM) contre $12.79 \pm 7.61$ mm ($16.63 \pm 7.55$ mm pour l'IRM) pour YOLO seul. Le muscle grand psoas droit présente aussi une amélioration de distance de $6.63 \pm 3.83$ mm ($7.28 \pm 2.12$ mm pour l'IRM) contre $10.70 \pm 6.80$ mm ($10.94 \pm 5.80$ mm pour l'IRM) pour YOLO seul.

Pour les coupes coronales, nous obtenons des améliorations très significatives de la distance moyenne pour la modalité TDM lors de l'intégration des contraintes anatomique. En effet, nous avons un $p$-value de $p = 1.06 * 10^{-5}$ (0.005 pour l'IRM) lors de l'intégration d'une seule contrainte anatomique et un $p$-value de $p = 8.39 * \ 10^{-7}$ (0.004 pour l'IRM) lors de l'intégration de deux contraintes anatomiques.

**Évolution de l'écart type des distances**

Dans le Chapitre 4, nous avons vu que l'approche de l'augmentation des données avec un CycleGAN pour la détection multi-organes avec YOLOv3 présentait une limitation, liée à l'apparition des fausses détections, qui conduisait dans certains cas à des écarts-types élevés sur la distance moyenne.

*Résultats obtenus à partir des coupes axiales* : Concernant l'écart type, nous remarquons que l'ajout de la contrainte d'orientation anatomique pour la coupe axiale entraîne une amélioration pour la plupart des structures anatomiques. Ainsi par exemple, pour la modalité TDM (Tableau 5.1a), la trachée est associée à un écart type de 23.97 mm pour YOLO seul relativement à une valeur de 1.96 mm avec une seule contrainte et de 1.25 mm avec deux contraintes. Pour la modalité IRM (Tableau 5.2a), la rate est associée à un écart type de 17.75 mm pour YOLO seul relativement à une valeur de 11.72 mm avec une seule contrainte et de 9.70 mm avec deux contraintes anatomiques ce qui amène à une stabilisation dans la détection.

Pour les coupes axiales, l'intégration de deux contraintes d'orientation anatomiques donne de bons résultats pour la plupart des organes dans les deux modalités. Pour la modalité TDM, l'écart-type moyen est de 2.31 mm pour le scénario YOLO+GD+AP par rapport à 7.65 mm pour le scénario YOLO. Pour la modalité IRM, l'écart-type moyen est de 7.07 mm pour le scénario YOLO+GD+AP par rapport à 9.34 mm pour le scénario YOLO.

*Résultats obtenus à partir des coupes coronales* : Nous remarquons aussi une bonne amélioration de l'écart type pour les coupes coronales ce qui amène à une stabilisation pour la détection. Ainsi par exemple, pour la modalité TDM (Tableau 5.1b), le sternum est associé à un écart type de 9.31 mm pour YOLO seul relativement à une valeur de 3.99 mm avec une seule contrainte et de 2.75 mm avec deux contraintes. Pour la modalité IRM (Tableau 5.2b), le rein droit est associé à un écart type de 16.46 mm pour YOLO seul relativement à une valeur de 7.40 mm avec une seule contrainte et de 5.13 mm avec deux contraintes anatomiques.

Pour les coupes coronales, l'intégration de deux contraintes d'orientation anatomiques donne de bons résultats pour la plupart des organes dans les deux modalités. Pour la modalité TDM, l'écart-type moyen est de 2.75 mm pour le scénario YOLO+GD+SI par rapport à 5.32 mm pour le scénario YOLO. Pour la modalité IRM, l'écart-type moyen est de 5.12 mm pour le scénario YOLO+GD+SI par rapport à 8.54 mm pour le scénario YOLO.

TABLEAU 5.1 – Comparaison des résultats de détection obtenus à partir des coupes axiales et coronales sur la base de données VISCERAL Gold, pour YOLO sans contrainte, YOLO en intégrant une seule contrainte anatomique et YOLO en intégrant deux contraintes anatomiques. Les résultats sont donnés en termes de distance moyenne [mm] pour chaque organe pour la modalité TDM.

(a) YOLO vs YOLO+GD et YOLO+GD+AP, en coupes axiales.

| | YOLO | YOLO+GD | YOLO+GD+AP |
|---|---|---|---|
| Trachée | 9.81 ± 23.97 | 2.60 ± 1.96 | **1.96 ± 1.25** |
| Poumon droit | 3.80 ± 1.32 | 3.76 ± 0.85 | **3.21 ± 0.88** |
| Poumon gauche | 3.52 ± 1.51 | 3.50 ± 1.14 | **2.65 ± 1.20** |
| Pancréas | 14.34 ± 10.42 | 10.91 ± 4.73 | **8.19 ± 3.30** |
| Vésicule biliaire | 6.95 ± 10.99 | 6.76 ± 3.14 | **4.87 ± 1.99** |
| Vessie | 4.04 ± 1.21 | 4.03 ± 1.18 | **3.72 ± 1.36** |
| Sternum | 7.05 ± 8.33 | 3.08 ± 1.65 | **2.55 ± 1.17** |
| Vertèbre L1 | 6.24 ± 3.52 | 4.80 ± 2.57 | **4.28 ± 2.12** |
| Rein droit | 4.71 ± 5.81 | **3.24 ± 2.19** | 3.42 ± 3.69 |
| Rein gauche | 5.61 ± 12.93 | **2.21 ± 1.06** | 2.56 ± 1.82 |
| Surrénale droite | 6.63 ± 6.54 | 6.00 ± 2.88 | **4.68 ± 1.64** |
| Surrénale gauche | 8.14 ± 8.48 | 6.85 ± 4.32 | **4.21 ± 1.81** |
| Psoas droit | 16.67 ± 13.48 | 9.70 ± 5.75 | **5.50 ± 2.77** |
| Psoas gauche | 12.70 ± 7.10 | 9.48 ± 5.30 | **6.75 ± 4.17** |
| Abdominale D. | 13.63 ± 12.18 | 10.51 ± 5.58 | **6.26 ± 2.25** |
| Abdominale G. | 11.92 ± 7.33 | 10.20 ± 5.14 | **10.16 ± 4.69** |
| Aorte | 4.03 ± 3.06 | 3.43 ± 1.75 | **3.41 ± 1.33** |
| Foie | 7.45 ± 4.47 | 7.42 ± 2.75 | **7.37 ± 3.41** |
| Thyroïde | **5.91 ± 3.22** | 6.60 ± 3.37 | 6.01 ± 2.68 |
| Rate | 6.90 ± 7.03 | 5.56 ± 3.34 | **5.10 ± 2.78** |
| Moyenne | 8.00 ± 7.65 | 6.03 ± 3.03 | **4.84 ± 2.31** |

(b) YOLO vs YOLO+GD et YOLO+GD+SI, en coupes coronales.

| | YOLO | YOLO+GD | YOLO+GD+SI |
|---|---|---|---|
| Trachée | 3.83 ± 4.31 | 2.88 ± 3.20 | **2.55 ± 2.52** |
| Poumon droit | 3.39 ± 1.71 | 3.37 ± 2.88 | **2.78 ± 0.93** |
| Poumon gauche | 4.59 ± 7.58 | 2.37 ± 1.07 | **2.25 ± 0.95** |
| Pancréas | 12.79 ± 7.61 | 8.90 ± 4.26 | **8.05 ± 3.17** |
| Vésicule biliaire | 6.21 ± 3.30 | 4.81 ± 2.30 | **4.79 ± 2.35** |
| Vessie | 5.85 ± 7.27 | 3.47 ± 1.58 | **3.35 ± 1.60** |
| Sternum | 8.13 ± 9.31 | 5.26 ± 3.99 | **4.67 ± 2.75** |
| Vertèbre L1 | 9.31 ± 4.98 | 8.27 ± 4.05 | **7.84 ± 4.15** |
| Rein droit | 3.39 ± 5.08 | 2.80 ± 2.98 | **2.69 ± 3.02** |
| Rein gauche | 3.37 ± 4.61 | 2.83 ± 4.15 | **2.34 ± 2.79** |
| Surrénale droite | 7.87 ± 4.20 | 6.45 ± 3.85 | **4.73 ± 1.56** |
| Surrénale gauche | **8.67 ± 7.76** | 9.39 ± 6.00 | 8.98 ± 4.59 |
| Psoas droit | 10.70 ± 6.80 | 7.75 ± 4.90 | **6.63 ± 3.83** |
| Psoas gauche | 11.41 ± 6.52 | 9.20 ± 4.75 | **7.91 ± 3.23** |
| Abdominale D. | 10.69 ± 5.34 | 9.53 ± 5.03 | **7.10 ± 3.06** |
| Abdominale G. | 9.14 ± 4.81 | 7.93 ± 4.17 | **7.27 ± 3.61** |
| Aorte | 7.20 ± 4.34 | 6.12 ± 3.89 | **5.51 ± 2.94** |
| Foie | 5.81 ± 2.78 | 4.76 ± 2.66 | **4.48 ± 2.44** |
| Thyroïde | 5.33 ± 3.79 | 4.23 ± 3.19 | **3.80 ± 2.41** |
| Rate | 3.71 ± 4.38 | 3.80 ± 4.16 | **3.39 ± 3.03** |
| Moyenne | 7.07 ± 5.32 | 5.71 ± 3.65 | **5.06 ± 2.75** |

TABLEAU 5.2 – Comparaison des résultats de détection obtenus à partir des coupes axiales et coronales sur la base de données VISCERAL Gold, pour YOLO sans contrainte, YOLO en intégrant une seule contrainte anatomique et YOLO en intégrant deux contraintes anatomiques. Les résultats sont donnés en termes de distance moyenne [mm] pour chaque organe pour la modalité IRM.

(a) YOLO vs YOLO+GD et YOLO+GD+AP, en coupes axiales.

| | YOLO | YOLO+GD | YOLO+GD+AP |
|---|---|---|---|
| Pancréas | 17.70 ± 8.33 | 15.45 ± 6.47 | **12.65 ± 5.11** |
| Vésicule biliaire | 17.78 ± 10.49 | 13.23 ± 9.55 | **12.04 ± 7.16** |
| Vessie | 15.12 ± 17.7 | 13.63 ± 14.89 | **11.21 ± 11.10** |
| Vertèbre L1 | 13.59 ± 5.10 | 11.69 ± 4.85 | **11.31 ± 4.09** |
| Rein droit | 11.49 ± 15.71 | 8.53 ± 11.24 | **6.31 ± 8.28** |
| Rein gauche | 8.84 ± 11.93 | 7.31 ± 9.1 | **6.25 ± 6.59** |
| Psoas droit | 12.03 ± 5.29 | 9.67 ± 4.65 | **8.54 ± 4.26** |
| Psoas gauche | 13.80 ± 7.19 | 10.04 ± 6.69 | **8.19 ± 5.54** |
| Abdominale D. | | 51.25 ± 0.00 | **23.39 ± 10.65** |
| Abdominale G. | 55.99 ± 5.08 | 34.34 ± 4.87 | **29.37 ± 3.77** |
| Aorte | 81.20 ± 0.00 | 35.73 ± 13.53 | **27.16 ± 9.43** |
| Foie | 12.26 ± 7.59 | 10.17 ± 6.93 | **9.08 ± 6.17** |
| Rate | 11.75 ± 17.75 | 9.30 ± 11.72 | **8.52 ± 9.70** |
| Moyenne | 22.63 ± 9.34 | 17.72 ± 8.71 | **13.39 ± 7.07** |

(b) YOLO vs YOLO+GD et YOLO+GD+SI, en coupes coronales.

| | YOLO | YOLO+GD | YOLO+GD+SI |
|---|---|---|---|
| Pancréas | 16.63 ± 7.55 | 14.45 ± 5.12 | **10.45 ± 4.64** |
| Vésicule biliaire | 8.26 ± 3.22 | 7.66 ± 3.10 | **7.66 ± 2.63** |
| Vessie | 10.83 ± 10.08 | 7.37 ± 7.74 | **6.70 ± 5.03** |
| Vertèbre L1 | 11.39 ± 4.57 | 10.18 ± 4.08 | **9.58 ± 3.73** |
| Rein droit | 9.88 ± 16.46 | 7.79 ± 7.40 | **7.03 ± 5.13** |
| Rein gauche | **5.51 ± 4.88** | 5.54 ± 4.16 | 5.54 ± 4.16 |
| Psoas droit | 10.94 ± 5.80 | 9.72 ± 4.78 | **7.28 ± 2.12** |
| Psoas gauche | 11.54 ± 7.69 | 10.22 ± 5.83 | **8.98 ± 3.51** |
| Abdominale D. | 41.31 ± 0.00 | 25.89 ± 0.00 | **18.28 ± 9.81** |
| Abdominale G. | 40.21 ± 7.77 | 23.02 ± 0.00 | **11.65 ± 7.65** |
| Aorte | 36.54 ± 0.00 | 27.13 ± 0.00 | **11.98 ± 0.00** |
| Foie | 13.44 ± 13.00 | 9.90 ± 10.52 | **7.49 ± 7.88** |
| Rate | 9.31 ± 12.97 | 7.30 ± 7.99 | **6.51 ± 3.13** |
| Moyenne | 17.37 ± 8.54 | 13.24 ± 5.86 | **9.38 ± 5.12** |

#### 5.5.2.2 Intégration des contraintes sur les données augmentées

Dans cette section nous présentons les résultats des deux contributions réunies, à savoir l'application de la contrainte d'orientation anatomique aux données augmentées par CycleGAN, présentées dans le Chapitre 4. Ces résultats pour les modalités TDM et IRM aux coupes axiales sont présentés respectivement dans les Tableaux 5.3 et 5.4.

La Tableau 5.3 illustre l'évaluation de l'intégration de la contrainte avec les données augmentées dans la base d'entraînement du détecteur YOLO pour la modalité TDM. Avec l'intégration d'une seule contrainte anatomique gauche/droit, nous notons une amélioration de la distance moyenne pour la plupart des structures anatomiques. Ainsi par exemple, la glande surrénale droite est un organe difficile à détecter puisqu'elle est de petite taille. Nous notons qu'elle présente une distance moyenne de 4.61 mm contre une distance de 6.37 mm pour YOLO+CycleGAN. La trachée présente une distance de 2.64 mm contre YOLO+CycleGAN 6.23 mm. Avec l'intégration de deux contraintes anatomiques d'orientation gauche/droit et antérieur/postérieur, nous présentons une amélioration pour tous les organes par rapport à YOLO seul et YOLO+CycleGAN. Par exemple, le grand muscle psoas droit présente une distance moyenne de 8.60 mm contre YOLO seul 16.67 mm et YOLO+CycleGAN 11.81 mm et le sternum de distance moyenne 4.13 mm contre YOLO seul 7.05 mm et 8.96 mm YOLO+CycleGAN.

Pour la modalité TDM nous avons une amélioration très significative de la distance moyenne lorsqu'une seule contrainte est intégrée avec un $p$-value $p = 0.005$ et lorsque deux contraintes sont intégrées avec $p = 1.97 * 10^{-7}$.

Dans le chapitre précédent, nous avons indiqué que la détection n'était pas satisfaisante pour la modalité IRM plus précisément pour les petits organes. Le Tableau 5.4 illustre les résultats de l'intégration de la contrainte d'orientation anatomique pour la modalité IRM. Nous observons une amélioration pour ces petits organes tels que la vésicule biliaire qui présente une distance de 13.90 mm en intégrant une seule contrainte et une distance de 11.14 mm en intégrant deux contraintes anatomiques contre une distance de 17.78 mm pour YOLO seul et une distance de 13.90 mm pour YOLO+CycleGAN. Nous notons de plus une amélioration significative pour les organes difficiles à détecter. Ainsi par exemple, le pancréas présente une distance de 13.31 mm en intégrant une seule contrainte et une distance de 11.96 mm en intégrant deux contraintes anatomiques contre une distance de 17.70 mm pour YOLO seul et une distance de 14.93 mm pour YOLO+CycleGAN.

Pour la modalité IRM nous avons une amélioration très significative de la distance moyenne lorsqu'une seule contrainte est intégrée avec un $p$-value $p = 0.007$ et lorsque deux contraintes sont intégrées avec $p = 0.003$.

**Évolution de l'écart type des distances**

Nous remarquons que l'ajout de la contrainte d'orientation anatomique entraîne une amélioration d'écart type pour la plupart des structures anatomiques. Pour la modalité

TDM (Tableau 5.3), la vésicule biliaire est associée à un écart type de 11.29 mm pour YOLO+CycleGAN relativement à une valeur de 2.88 mm avec une seule contrainte et de 2.14 mm avec deux contraintes. Pour la modalité IRM (Tableau 5.4), la vessie est associée à un écart type de 12.04 mm pour YOLO+CycleGAN relativement à une valeur de 9.27 mm avec une seule contrainte et de 7.93 mm avec deux contraintes anatomiques ce qui amène à une stabilisation dans la détection.

L'intégration de deux contraintes d'orientation anatomiques donne de bons résultats pour la plupart des organes dans les deux modalités. Pour la modalité TDM, l'écart-type moyen est de 2.56 mm pour le scénario YOLO+CycleGAN+GD+AP par rapport à 6.60 mm pour le scénario YOLO+CycleGAN. Pour la modalité IRM, l'écart-type moyen est de 6.84 mm pour le scénario YOLO+CycleGAN+GD+AP par rapport à 11.69 mm pour le scénario YOLO+CycleGAN.

TABLEAU 5.3 – Comparaison des résultats de détection obtenus à partir des coupes axiales sur la base de données VISCERAL Gold, YOLO sans contrainte, YOLO sur la base de données augmentée via le CycleGAN, YOLO+CycleGAN en intégrant une seule contrainte anatomique et YOLO+CycleGAN en intégrant deux contraintes anatomiques. Les résultats sont donnés en termes de distance moyenne [mm] pour chaque organe pour la modalité TDM.

| | YOLO | YOLO +CycleGAN | YOLO +CycleGAN +GD | YOLO +CycleGAN +GD+AP |
|---|---|---|---|---|
| Trachée | 9.81 ± 23.97 | 6.23 ± 17.83 | 2.64 ± 4.43 | **1.63 ± 1.24** |
| Poumon droit | 3.80 ± 1.32 | 6.38 ± 4.96 | 3.54 ± 3.08 | **2.63 ± 0.85** |
| Poumon gauche | 3.52 ± 1.51 | 4.35 ± 1.90 | 2.96 ± 1.95 | **2.43 ± 1.17** |
| Pancréas | 14.34 ± 10.42 | 10.60 ± 5.28 | 8.95 ± 3.51 | **4.61 ± 1.49** |
| Vésicule biliaire | 6.95 ± 10.99 | 7.47 ± 11.29 | 5.23 ± 2.88 | **4.13 ± 2.14** |
| Vessie | 4.04 ± 1.21 | 4.56 ± 1.65 | 3.92 ± 1.54 | **3.48 ± 1.44** |
| Sternum | 7.05 ± 8.33 | 8.96 ± 11.10 | 6.84 ± 8.84 | **4.13 ± 2.77** |
| Vertèbre L1 | 6.24 ± 3.52 | 5.87 ± 3.39 | 5.49 ± 2.74 | **4.13 ± 1.57** |
| Rein droit | 5.61 ± 12.93 | 5.98 ± 12.42 | 3.49 ± 4.09 | **2.49 ± 2.56** |
| Rein gauche | 4.71 ± 5.81 | 4.39 ± 4.82 | 2.32 ± 2.88 | **1.87 ± 1.47** |
| Surrénale droite | 6.63 ± 6.54 | 6.37 ± 5.93 | 4.61 ± 3.15 | **3.94 ± 2.22** |
| Surrénale gauche | 8.14 ± 8.48 | 7.86 ± 8.71 | 7.51 ± 7.39 | **6.74 ± 7.06** |
| Psoas droit | 16.67 ± 13.48 | 11.81 ± 6.96 | 11.73 ± 6.34 | **8.60 ± 2.70** |
| Psoas gauche | 12.70 ± 7.10 | 12.87 ± 5.77 | 14.33 ± 14.12 | **10.13 ± 3.66** |
| Abdominale D. | 13.63 ± 12.18 | 11.92 ± 6.77 | 12.79 ± 5.60 | **10.62 ± 2.95** |
| Abdominale G | 11.92 ± 7.33 | 12.23 ± 7.77 | 11.41 ± 6.83 | **9.89 ± 4.91** |
| Aorte | 4.03 ± 3.06 | 3.93 ± 2.67 | 3.99 ± 2.81 | **3.46 ± 1.51** |
| Foie | 7.45 ± 4.47 | 6.92 ± 3.41 | 7.18 ± 6.87 | **5.36 ± 2.60** |
| Thyroïde | 5.91 ± 3.22 | 5.35 ± 3.06 | 5.23 ± 2.43 | **4.93 ± 2.01** |
| Rate | 6.90 ± 7.03 | 6.54 ± 6.24 | 8.35 ± 13.55 | **5.52 ± 4.85** |
| Moyenne | 8.00 ± 7.65 | 7.53 ± 6.60 | 6.63 ± 5.25 | **5.04 ± 2.56** |

TABLEAU 5.4 – Comparaison des résultats de détection obtenus à partir des coupes axiales sur la base de données VISCERAL Gold, YOLO sans contrainte, YOLO sur la base de données augmentée via le CycleGAN, YOLO+CycleGAN en intégrant une seule contrainte anatomique et YOLO+CycleGAN en intégrant deux contraintes anatomiques. Les résultats sont donnés en termes de distance moyenne [mm] pour chaque organe pour la modalité IRM.

| | YOLO | YOLO +CycleGAN | YOLO +CycleGAN +GD | YOLO +CycleGAN +GD+AP |
|---|---|---|---|---|
| Pancréas | 17.70 ± 8.33 | 14.93 ± 5.68 | 13.31 ± 4.30 | **11.96 ± 4.09** |
| Vésicule biliaire | 17.78 ± 10.49 | 13.90 ± 5.96 | 12.84 ± 5.03 | **11.14 ± 4.22** |
| Vessie | 15.12 ± 17.73 | 11.35 ± 12.04 | 9.21 ± 9.27 | **8.42 ± 7.93** |
| Vertèbre L1 | 13.59 ± 5.10 | 9.70 ± 3.12 | 8.24 ± 2.92 | **7.11 ± 2.84** |
| Rein droit | 11.49 ± 15.71 | 10.01 ± 15.64 | 8.65 ± 10.87 | **7.44 ± 8.48** |
| Rein gauche | 8.84 ± 11.93 | 10.20 ± 13.42 | 9.06 ± 10.19 | **8.20 ± 7.94** |
| Psoas droit | 12.03 ± 5.29 | 12.95 ± 5.69 | 11.32 ± 4.89 | **10.91 ± 4.62** |
| Psoas gauche | 13.80 ± 7.19 | 12.59 ± 6.31 | 11.31 ± 5.94 | **10.10 ± 4.39** |
| Abdominale G. | 55.99 ± 5.08 | 35.88 ± 34.32 | 24.80 ± 14.30 | **21.93 ± 13.39** |
| Aorte | 81.20 ± 0.00 | 37.62 ± 20.68 | 27.95 ± 9.79 | **24.22 ± 9.61** |
| Foie | **12.26 ± 7.59** | 14.08 ± 10.57 | 13.87 ± 9.57 | 13.73 ± 9.70 |
| Rate | 11.75 ± 17.75 | 10.95 ± 6.90 | 9.57 ± 5.19 | **8.95 ± 4.91** |
| Moyenne | 22.63 ± 9.34 | 16.18 ± 11.69 | 13.34 ± 7.69 | **12.01 ± 6.84** |

#### 5.5.2.3 Synthèse des résultats

Le Tableau 5.5 montre les résultats globaux de détection à partir de la distance moyenne obtenue sur l'ensemble des organes pour les coupes axiales et les modalités IRM et TDM. Ce tableau permet une comparaison entre les différentes configurations testées : YOLO seul et avec intégration d'une ou deux contraintes d'orientation ainsi que YOLO avec les données augmentées par CycleGAN et avec intégration d'une ou deux contraintes d'orientation. Nous remarquons que les résultats sont meilleurs lorsque les deux contributions sont réunies pour la modalité IRM par rapport à celle de la modalité TDM. Cela est dû au fait que les organes thoraciques dans la modalité TDM ne bénéficient pas de l'augmentation de données puisque le thorax est absent dans les images de la modalité IRM.

TABLEAU 5.5 – Comparaison de la distance moyenne sur l'ensemble des organes pour les coupes axiales et les modalités IRM et TDM.

| | YOLO | YOLO +GD | YOLO +GD+AP | YOLO +CycleGAN | YOLO +CycleGAN +GD | YOLO +CycleGAN +GD+AP |
|---|---|---|---|---|---|---|
| Modalité TDM | 8.00 ± 7.65 | 6.03 ± 3.03 | **4.84 ± 2.31** | 7.53 ± 6.60 | 6.63 ± 5.25 | 5.04 ± 2.56 |
| Modalité IRM | 22.63 ± 9.34 | 17.72 ± 8.71 | 13.39 ± 7.07 | 16.18 ± 11.69 | 13.34 ± 7.69 | **12.01 ± 6.84** |

## 5.6 Conclusion

Dans ce chapitre, nous avons intégré un a priori anatomique dans la fonction de perte du détecteur YOLO. Cet a priori est basé sur les relations d'orientation spatiales (supérieur/inférieur, gauche/droit et antérieur/postérieur) entre les structures anatomiques. La contrainte résultant de cet a priori est exprimée sous la forme de matrices d'orientation et intégrée à la fonction de perte de YOLOv3 via l'entropie croisée binaire.

L'influence de la contrainte sur la performance de détection a été étudiée sur les coupes axiales et coronales pour les modalités TDM et IRM. Les résultats obtenus ont montré que la contrainte conduisait à une nette amélioration de la détection en termes de distance moyenne. Cette amélioration est de 44% en coupe axiale et de 28% en coupe coronale pour la modalité TDM. Cette amélioration est de 40% en coupe axiale et de 45% en coupe coronale pour la modalité IRM.

Nous avons intégré cette contrainte pour des images de modalité IRM et TDM dans le cas particulier de la détection des organes thoraciques et abdominaux. Cependant le principe de cette contrainte est généralisable à d'autres modalités et d'autres organes, telle que des images cérébrales. De même, au-delà de la détection, cette contrainte d'orientation a le potentiel d'être intégrée dans d'autres applications telles que la segmentation, la reconstruction et la génération des images médicales.

# Chapitre 6

# Conclusion

Au cours de cette thèse, nous avons proposé une méthode de détection multi-organe pour les images médicales multi-modalités. Les contributions principales de cette thèse sont résumées dans la suite.

## Contributions principales

Tout d'abord, nous avons réalisé la détection des organes des images médicales pour les modalités TDM et IRM en utilisant le détecteur YOLOv3. Nous avons obtenu une bonne détection des grands organes et moins bonne des petits. Par conséquent, nous avons proposé une nouvelle approche d'augmentation de données permettant d'améliorer les performances du détecteur. Cette approche présente notre première contribution. L'augmentation de données est réalisée par une génération de données synthétiques à partir d'un modèle générateur CycleGAN. CycleGAN est une méthode d'apprentissage profond non supervisé qui traduit des images d'une modalité cible (exemple modalité IRM) à partir d'une modalité source (exemple modalité TDM). Une fois générées, les données synthétiques sont ajoutées dans le jeu de données d'entraînement du détecteur YOLOv3 avec les annotations de la modalité source. Le CycleGAN+YOLO donne des résultats statistiquement meilleurs relativement à YOLO seul pour la plupart des organes. Par contre, l'écart-type est élevé pour certains organes et cela est dû aux valeurs aberrantes de détection. Pour résoudre ce problème nous avons proposé d'intégrer une contrainte anatomique dans la fonction de perte de YOLOv3. Cette contrainte utilise la connaissance a priori de l'orientation relative des structures anatomiques (inférieur/supérieur, antérieur/postérieur, gauche/droit). Cette deuxième contribution donne de façon statistiquement significative de meilleurs résultats par rapport à YOLO seul.

## Perspectives

Ce travail de recherche sur l'augmentation de données et l'intégration d'a priori pour la détection multi-organe en imagerie médicale ouvre à plusieurs perspectives.

Des perspectives concernent l'élargissement du champ d'application et l'amélioration de résultats peuvent être réalisées dans un futur proche. La première consiste à améliorer encore les résultats en optimisant certains aspects des méthodes. Tout d'abord, il serait intéressant de tester si une évolution de l'architecture de YOLO pourrait permettre de mieux détecter les petits objets, en ajoutant des couches supplémentaires. Tout au long de la thèse, nous avons en effet utilisé le détecteur YOLO avec 3 couches de détection, et il serait tout à fait possible d'envisager d'accroître ce nombre avec des couches travaillant à des échelles plus réduites pour une meilleure détection des petits organes. Un autre axe concerne l'étude plus approfondie du choix des hyperparamètres et de leur influence, afin de caractériser le comportement de l'algorithme à des changements de pondération. Cette étude serait en particulier intéressante pour le paramètre pondérant le terme de contrainte d'orientation anatomique. Une deuxième perspective concerne la comparaison avec l'état de l'art. C'est un travail qui est en cours sur la base de données du challenge LITS [Xu *et al.* (2019)], qui permettra de plus de tester la capacité de généralisation de l'approche proposée. Une troisième perspective qui peut être envisagée dans le futur consiste en l'application de l'approche proposée en 3D, qui est pour le moment encore difficile à envisager pour des raisons de limitation mémoire. Cependant, il convient de noter que cette généralisation est immédiate, puisqu'il suffirait dans le terme de contrainte de considérer trois orientations au lieu des deux utilisées en 2D dans ce travail.

A plus long terme, il est à noter que la formulation que nous avons proposée peut facilement être étendue à d'autres applications. L'approche consistant à intégrer une contrainte anatomique d'orientation présente en effet l'intérêt d'être facilement extensible, dans la mesure où elle repose sur un a priori très général. Ainsi, en imagerie médicale, celle-ci pourrait facilement être mise à profit pour traiter d'autres régions du corps humain (et même le corps entier si l'ensemble des relations inter-organes est encodée) et être intégrée à d'autres applications que la détection, en particulier en segmentation ou recalage. Au-delà de l'imagerie médicale, (par exemple en imagerie microscopique, satellitaire, etc.), cette approche est là encore applicable à partir du moment où un ensemble stable de relations d'orientation peut être défini pour les objets à détecter.

# Annexes

Annexe A

# Matrice d'orientation anatomique

## A.1 Matrice d'orientation anatomique Supérieur/Inférieur

| | *Tra* | *P.D.* | *P.G.* | *Panc* | *Ves.bil* | *Ves.* | *Ster* | *Ver.L*1 | *ReinD* | *ReinG* | *Su.D* | *Su.G.* | *Ps.D.* | *Ps.G.* | *Ab.D.* | *Ab.G.* | *Aor* | *Foie* | *Thy* | *Rate* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Trachée | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| Poumon droit | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Poumon gauche | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| Pancréas | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Vésicule biliaire | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Vessie | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| Sternum | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| Vertèbre L1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| Rein droit | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Rein gauche | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| Surrénale droite | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Surrénale gauche | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| Psoas droit | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Psoas gauche | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| Abdominale droit | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Abdominale gauche | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| Aorte | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| Foie | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Thyroïde | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| Rate | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |

(A.1)

## A.2 Matrice d'orientation anatomique Antérieur/Postérieur

| | *Tra* | *P.D.* | *P.G.* | *Panc* | *Ves.B.* | *Ves.* | *Ster* | *VerL*1 | *ReinD.* | *ReinG.* | *Su.D.* | *Su.G.* | *Ps.D.* | *Ps.G.* | *Abd.D.* | *Abd.G.* | *Aor* | *Foie* | *Thy* | *Rate* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Trachée | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Poumon droit | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Poumon gauche | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Pancréas | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Vésicule biliaire | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 |
| Vessie | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Sternum | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 |
| Vertèbre L1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Rein droit | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Rein gauche | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Surrénale D. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Surrénale G. | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Psoas droit | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Psoas gauche | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Abdominale D. | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 |
| Abdominale G. | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 |
| Aorte | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Foie | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Thyroïde | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Rate | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

(A.2)

## A.3 Matrice d'orientation anatomique Gauche/Droit

| | | *Tra* | *Po.D* | *Po.G* | *Pan* | *Ves.bil* | *Vess.* | *Ste* | *Ver.L1* | *Re.D.* | *Re.G* | *Sur.D* | *Sur.G* | *Ps.D* | *Ps.G* | *Ab.D* | *Ab.G* | *Aor* | *Foie* | *Thy* | *Rate* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Trachée | 1247 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| Poumon D. | 1302 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Poumon G. | 1326 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| Pancréas | 170 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Vésicule biliaire | 187 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Vessie | 237 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| Sternum | 2473 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| Vertèbre L1 | 29193 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| Rein droit | 29662 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Rein gauche | 29663 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| Surrénale D. | 30324 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Surrénale G. | 30325 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| Psoas droit | 32248 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Psoas gauche | 32249 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| Abdominale D. | 40357 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Abdominale G. | 40358 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 |
| Aorte | 480 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| Foie | 58 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Thyroïde | 7578 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| Rate | 86 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |

(A.3)

# Bibliographie


[Altman (1992)] Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3) :175–185.

[Bargoti and Underwood (2017)] Bargoti, S. et Underwood, J. P. (2017). Deep Fruit Detection in Orchards. In *2017 International Conference on Robotics and Automation (ICRA)*, pages 3626–3633. http://arxiv.org/abs/1610.03677.

[Baumgartner *et al.* (2019)] Baumgartner, C. F. , Tezcan, K. C. , Chaitanya, K. , Hötker, A. M. , Muehlematter, U. J. , Schawkat, K. , Becker, A. S. , Donati, O., et Konukoglu, E. (2019). PHiSeg : Capturing Uncertainty in Medical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 119–127. Springer. arXiv:1906.04045v2.

[Ben-Cohen *et al.* (2019)] Ben-Cohen, A. , Klang, E. , Raskin, S. , Soffer, S. , Ben-Haim, S. , Konen, E. , Amitai, M., et Greenspan, H. (2019). Cross-Modality Synthesis from CT to PET using FCN and GAN Networks for Improved Automated Lesion Detection. *Eng. Appl. Artif. Intell.*, 78 :186–194. arXiv:1802.07846v2.

[BenTaieb and Hamarneh (2016)] BenTaieb, A. et Hamarneh, G. (2016). Topology Aware Fully Convolutional Networks for Histology Gland Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 460–468. Springer. https://doi.org/10.1007/978-3-319-46723-8_53.

[Bochkovskiy *et al.* (2020)] Bochkovskiy, A. , Wang, C.-Y., et Liao, H.-Y. M. (2020). YOLOv4 : Optimal Speed and Accuracy of Object Detection. https://arxiv.org/abs/2004.10934.

[Breiman (2001)] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1) :5–32. https://doi.org/10.1023/A:1010933404324.

[Criminisi *et al.* (2009)] Criminisi, A. , Shotton, J., et Bucciarelli, S. (2009). Decision Forests with Long-Range Spatial Context for Organ Localization in CT Volumes. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 69–80. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.646.3583&rep=rep1&type=pdf.

[Criminisi *et al.* (2010)] Criminisi, A. , Shotton, J. , Robertson, D., et Konukoglu, E. (2010). Regression Forests for Efficient Anatomy Detection and Localization in CT Studies. In *International MICCAI Workshop on Medical Computer Vision*, pages 106–117. Springer. https://doi.org/10.1007/978-3-642-18421-5_11.

[Criminisi *et al.* (2013)] Criminisi, A. , Robertson, D. , Konukoglu, E. , Shotton, J. , Pathak, S. , White, S., et Siddiqui, K. (2013). Regression Forests for Efficient Anatomy

Detection and Localization in Computed Tomography Scans. *Medical Image Analysis*, 17(8) :1293–1303. DOI:10.1016/j.media.2013.01.001.

[Cuingnet *et al.* (2012)] Cuingnet, R. , Prevost, R. , Lesage, D. , Cohen, L. , Mory, B., et Ardon, R. (2012). Automatic Detection and Segmentation of Kidneys in 3D CT images using Random Forests. In *International Conference on Medical Image Computing and Computer–Assisted Intervention*, pages 66–74. Springer. DOI:10.1007/978-3-642-33454-2_9.

[Dalal and Triggs (2005)] Dalal, N. et Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893. doi:10.1109/CVPR.2005.177.

[Deng *et al.* (2009)] Deng, J. , Dong, W. , Socher, R. , Li, L.-J. , Li, K., et Fei-Fei, L. (2009). ImageNet : A Large–scale Hierarchical Image Database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. DOI:10.1109/CVPR.2009.5206848.

[De Boer *et al.* (2005)] De Boer, P.-T. , Kroese, D. P. , Mannor, S., et Rubinstein, R. Y. (2005). A Tutorial on the Cross–Entropy Method. *Annals of Operations Research*, 134(1) :19–67. doi=10.1007/s10479-005-5724-z.

[De Palma (2020)] De Palma, R. (2020). YOLOv3 Architecture : Best Model in Object Detection. In *Consulté le 01/06/2020.* https://d33wubrfki0l68.cloudfront.net/c6fd049f28b66dbd35faed6965905ec6281f7d7d/c0399/assets/images/yolo/yolo-architecture.webp.

[De Vos *et al.* (2016)] De Vos, B. D. , Wolterink, J. M. , De Jong, P. A. , Viergever, M. A., et Išgum, I. (2016). 2D Image Classification for 3D Anatomy Localization : Employing Deep Convolutional Neural Networks. In *Medical Imaging 2016 : Image processing (Progress in Biomedical Optics and Imaging - Proceedings of SPIE)*, volume 9784, page 97841Y. https://doi.org/10.1117/12.2216971.

[Dolejsi *et al.* (2008)] Dolejsi, M. , Kybic, J. , Tuma, S., et Polovincák, M. (2008). Reducing False Positive Responses in Lung Nodule Detector System by Asymmetric AdaBoost. In *5th IEEE International Symposium on Biomedical Imaging : From Nano to Macro*, pages 656–659. DOI:10.1109/ISBI.2008.4541081.

[Dvornik *et al.* (2017)] Dvornik, N. , Shmelkov, K. , Mairal, J., et Schmid, C. (2017). Blitznet : A Real-Time Deep Network for Scene Understanding. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4174–4182. DOI:10.1109/ICCV.2017.447.

[Felzenszwalb *et al.* (2008)] Felzenszwalb, P. , McAllester, D., et Ramanan, D. (2008). A Discriminatively Trained, Multiscale, Deformable part Model. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. doi:10.1109/CVPR.2008.4587597.

[Felzenszwalb *et al.* (2010)] Felzenszwalb, P. F. , Girshick, R. B., et McAllester, D. (2010). Cascade object Detection with Deformable part Models. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2241–2248. DOI:10.1109/CVPR.2010.5539906.

[Gad (2021)] Gad, A. (2021). Evaluating Object Detection Models Using Mean Average Precision. https://www.kdnuggets.com/2021/03/evaluating-object-detection-models-using-mean-average-precision.html.

[Ganaye *et al.* (2019)] Ganaye, P.-A. , Sdika, M. , Triggs, B., et Benoit-Cattin, H. (2019). Removing Segmentation Inconsistencies with Semi-Supervised Non-Adjacency Constraint. *Medical Image Analysis*, 58 :101551. https://doi.org/10.1016/j.media.2019.101551.

[Gauriau *et al.* (2015)] Gauriau, R. , Cuingnet, R. , Lesage, D., et Bloch, I. (2015). Multi–Organ Localization with Cascaded Global–to–Local Regression and Shape Prior. *Medical Image Analysis*, 23(1) :70–83. DOI:10.1016/j.media.2015.04.007.

[Girshick (2015)] Girshick, R. (2015). Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1440–1448. doi:10.1109/ICCV.2015.169.

[Girshick *et al.* (2014)] Girshick, R. , Donahue, J. , Darrell, T., et Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587. doi:10.1109/CVPR.2014.81.

[Goodfellow *et al.* (2014)] Goodfellow, I. J. , Pouget-Abadie, J. , Mirza, M. , Xu, B. , Warde-Farley, D. , Ozair, S. , Courville, A., et Bengio, Y. (2014). Generative Adversarial Networks. *stat.ML*. arXiv:1406.2661v1.

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

[Hammami *et al.* (2020)a] Hammami, M. , Friboulet, D., et Kechichian, R. (2020a). Cycle GAN-Based Data Augmentation for Multi-Organ Detection in CT Images via YOLO. In *IEEE International Conference on Image Processing (ICIP)*, pages 390–393. doi:10.1109/ICIP40778.2020.9191127.

[Hammami *et al.* (2020)b] Hammami, M. , Friboulet, D., et Kechichian, R. (2020b). Data Augmentation for Multi-Organ Detection in Medical Images. In *IEEE Tenth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pages 1–6. DOI:10.1109/IPTA50016.2020.9286712.

[Hanbury *et al.* (2012)] Hanbury, A. , Müller, H. , Langs, G. , Weber, M. A. , Menze, B. H., et Fernandez, T. S. (2012). Bringing the Algorithms to the Data : Cloud–based Benchmarking for Medical Image Analysis. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pages 24–29. Springer. https://doi.org/10.1007/978-3-642-33247-0_3.

[Hasan and Boris (2006)] Hasan, M. et Boris, F. (2006). SVM : Machines à Vecteurs de Support ou Séparateurs à Vastes Marges. *BD Web, ISTY3, Rapport technique, Versailles St Quentin, France. Cité*, 64. http://georges.gardarin.free.fr/Surveys_DM/Survey_SVM.pdf.

[He *et al.* (2016)] He, K. , Zhang, X. , Ren, S., et Sun, J. (2016). Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778. DOI:10.1109/CVPR.2016.90.

[Huo *et al.* (2018)] Huo, Y. , Xu, Z. , Bao, S. , Assad, A. , Abramson, R., et Landman, B. (2018). Adversarial Synthesis Learning Enables Segmentation Without Target Modality Ground Truth. In *IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1217–1220. https://doi.org/10.1109/isbi.2018.8363790.

[Ioffe and Szegedy (2015)] Ioffe, S. et Szegedy, C. (2015). Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *International Conference on Machine Learning*, pages 448–456. PMLR. http://arxiv.org/abs/1502.03167.

[Jiang *et al.* (2018)] Jiang, J. , Hu, Y.-C. , Tyagi, N. , Zhang, P. , Rimner, A. , Mageras, G. , Deasy, J., et Veeraraghavan, H. (2018). Tumor–aware, Adversarial Domain Adaptation from CT to MRI for Lung Cancer Segmentation. In *International Conference on Medical Image Computing and Computer–Assisted Intervention*, pages 777–785. Springer. DOI:10.1007/978-3-030-00934-2_86.

[Jimenez-del Toro *et al.* (2016)] Jimenez-del Toro, O. *et al.* (2016). Cloud-based Evaluation of Anatomical Structure Segmentation and Landmark Detection Algorithms : VISCERAL Anatomy Benchmarks. *IEEE Trans. Med. Imag.*, 35(11) :2459–2475. DOI:10.1109/TMI.2016.2578680.

[Jin *et al.* (2017)] Jin, X. , Qi, Y., et Wu, S. (2017). CycleGAN Face-off. *CoRR*, abs/1712.03451. http://arxiv.org/abs/1712.03451.

[Karsch *et al.* (2011)] Karsch, K. , Hedau, V. , Forsyth, D., et Hoiem, D. (2011). Rendering Synthetic Objects Into Legacy Photographs. *ACM Transactions on Graphics (TOG)*, 30(6) :1–12. https://doi.org/10.1145/2070781.2024191.

[Kathuria (2018)] Kathuria, A. (2018). How to Implement a YOLO(v3) Object Detector from Scratch in Pytorch : Part 1. Mis en ligne le 17 Mai 2018, consulté le 20 septembre 2020. https://www.kdnuggets.com/2018/05/implement-yolo-v3-object-detector-pytorch-part-1.html.

[LeCun *et al.* (2015)] LeCun, Y. *et al.* (2015). LeNet-5, Convolutional Neural Networks. 20(5) :14. URL : http://yann.lecun.com/exdb/lenet.

[Lee *et al.* (2018)] Lee, S.-g. , Bae, J. S. , Kim, H. , Kim, J. H., et Yoon, S. (2018). Liver Lesion Detection from Weakly-Labeled multi-phase CT Volumes with a Grouped Single Shot Multibox Detector. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 693–701. Springer. https://doi.org/10.1007/978-3-030-00934-2_77.

[Lemay (2019)] Lemay, A. (2019). Kidney Recognition in CT using YOLOv3. *arXiv preprint arXiv: 1910. 01268*.

[Lin *et al.* (2014)] Lin, T.-Y. , Maire, M. , Belongie, S. , Hays, J. , Perona, P. , Ramanan, D. , Dollár, P., et Zitnick, C. L. (2014). Microsoft COCO : Common Objects in Context. In *European Conference on Computer Vision*, volume 8693, pages 740–755. Springer. https://doi.org/10.1007/978-3-319-10602-1_48.

[Lin *et al.* (2017)a] Lin, T.-Y. , Dollár, P. , Girshick, R. , He, K. , Hariharan, B., et Belongie, S. J. (2017a). Feature Pyramid Networks for Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2117–2125. http://arxiv.org/abs/1612.03144.

[Lin *et al.* (2017)b] Lin, T.-Y. , Goyal, P. , Girshick, R. , He, K., et Dollár, P. (2017b). Focal Loss for Dense Object Detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2980–2988. doi:10.1109/ICCV.2017.324.

[Liu *et al.* (2016)] Liu, W. , Anguelov, D. , Erhan, D. , Szegedy, C. , Reed, S. , Fu, C.-Y., et Berg, A. C. (2016). SSD : Single Shot Multibox Detector. In *European Conference on Computer Vision*, pages 21–37. Springer. https://doi.org/10.1007/978-3-319-46448-0_2.

[Liu *et al.* (2018)] Liu, L. , Ouyang, W. , Wang, X. , Fieguth, P. , Chen, J. , Liu, X., et Pietikäinen, M. (2018). Deep Learning for Generic Object Detection : A Survey. *International Journal of Computer Vision*, pages 1–58. http://arxiv.org/abs/1809.02165.

[MacQueen (1967)] MacQueen, J. (1967). Some Methods for lassification and Analysis of Multivariate Observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 281–297. Oakland, CA, USA. doi=10.1.1.308.8619.

[Mao *et al.* (2017)] Mao, X. , Li, Q. , Xie, H. , Lau, R. Y. K. , Wang, Z., et Smolley, S. P. (2017). Least Squares Generative Adversarial Networks. *IEEE International Conference on Computer Vision (ICCV)*, pages 2813–2821. doi:10.1109/ICCV.2017.304.

[Nair and Hinton (2010)] Nair, V. et Hinton, E. G. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 807–814. https://icml.cc/Conferences/2010/papers/432.pdf.

[Näppi *et al.* (2016)] Näppi, J. , Hironaka, T. , Regge, D., et Yoshida, H. (2016). Deep Transfer Learning of Virtual Endoluminal views for the Detection of Polyps in CT Colonography. *International Society for Optics and Photonics*, 9785 :97852B. https://doi.org/10.1117/12.2217260.

[Oktay *et al.* (2017)] Oktay, O. , Ferrante, E. , Kamnitsas, K. , Heinrich, M. , Bai, W. , Caballero, J. , Cook, S. A. , De Marvao, A. , Dawes, T. , O'Regan, D. P., *et al.* (2017). Anatomically Constrained Neural Networks (ACNNs) : Application to Cardiac Image Enhancement and Segmentation. *IEEE Transactions on Medical Imaging*, 37(2) :384–395. DOI:10.1109/TMI.2017.2743464.

[Orbach (1962)] Orbach, J. (1962). Principles of Neurodynamics. Perceptrons and the Theory of Brain Mechanisms. Technical report, Cornell Aeronautical Lab Inc Buffalo NY. doi:10.1001/archpsyc.1962.01720030064010.

[Pan and Yang (2009)] Pan, S. J. et Yang, Q. (2009). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10) :1345–1359. doi:10.1109/TKDE.2009.191.

[Ravishankar *et al.* (2017)] Ravishankar, H. , Venkataramani, R. , Thiruvenkadam, S. , Sudhakar, P., et Vaidya, V. (2017). Learning and Incorporating Shape Models for Semantic Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 203–211. Springer. https://doi.org/10.1007/978-3-319-66182-7_24.

[Redmon *et al.* (2016)] Redmon, J. , Divvala, S. , Girshick, R., et Farhadi, A. (2016). You Only Look Once : Unified, Real-Time Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 779–788. (cs.CV) arXiv:1506.02640v5.

[Redmon and Farhadi (2017)] Redmon, J. et Farhadi, A. (2017). YOLO9000 : Better, Faster, Stronger. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6517–6525. DOI : 10.1109/CVPR.2017.690.

[Redmon and Farhadi (2018)] Redmon, J. et Farhadi, A. (2018). YOLOv3 : An Incremental Improvement. arXiv preprint http://arxiv.org/abs/1804.02767.

[Ren *et al.* (2015)] Ren, S. , He, K. , Girshick, R., et Sun, J. (2015). Faster R–CNN : Towards Real–Time Object Detection with Region Proposal Networks. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 39(6), pages 91–99. doi:10.1109/TPAMI.2016.2577031.

[Riegler *et al.* (2015)] Riegler, G. , Urschler, M. , Ruther, M. , Bischof, H., et Stern, D. (2015). Anatomical Landmark Detection in Medical Applications Driven by Synthetic Data. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 12–16. DOI:10.1109/ICCVW.2015.21.

[Ronneberger *et al.* (2015)] Ronneberger, O. , Fischer, P., et Brox, T. (2015). U-Net : Convolutional Networks for Biomedical Image Segmentation. In *International Conference on Medical Image Computing and Computer–Assisted Intervention*, pages 234–241. Springer. http://arxiv.org/abs/1505.04597.

[Rozantsev *et al.* (2015)] Rozantsev, A. , Lepetit, V., et Fua, P. (2015). On Rendering Synthetic Images for Training an Object Detector. *Computer Vision and Image Understanding*, 137 :24–37. DOI:10.1016/j.cviu.2014.12.006.

[Rumelhart *et al.* (1995)] Rumelhart, D. E. , Durbin, R. , Golden, R., et Chauvin, Y. (1995). Backpropagation : The Basic Theory. *Backpropagation : Theory, architectures and applications, Chapter 1*, pages 1–34.

[Sa *et al.* (2017)] Sa, R. , Owens, W. , Wiegand, R. , Studin, M. , Capoferri, D. , Barooha, K. , Greaux, A. , Rattray, R. , Hutton, A. , Cintineo, J., et Chaudhary, V. (2017). Intervertebral Disc Detection in X-ray Images using Faster R-CNN. In *39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 564–567. DOI:10.1109/EMBC.2017.8036887.

[Sadrach (2020)] Sadrach, P. (2020). How GANs Can Improve Healthcare Analytics. *Towards Data Science*. https://towardsdatascience.com/how-gans-can-improve-healthcare-analytics-7d2379eff19e.

[Shin *et al.* (2016)] Shin, H.-C. , Roth, H. R. , Gao, M. , Lu, L. , Xu, Z. , Nogues, I. , Yao, J. , Mollura, D., et Summers, R. M. (2016). Deep Convolutional Neural Networks for Computer-aided Detection : CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, 35(5) :1285–1298. DOI:10.1109/TMI.2016.2528162.

[Shorten and Khoshgoftaar (2019)] Shorten, C. et Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1) :1–48. DOI:10.1186/s40537-019-0197-0.

[Shrivastava *et al.* (2017)] Shrivastava, A. , Pfister, T. , Tuzel, O. , Susskind, J. , Wang, W., et Webb, R. (2017). Learning from Simulated and Unsupervised Images Through Adversarial Training. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2107–2116. DOI:10.1109/CVPR.2017.241.

[Simonyan and Zisserman (2014)] Simonyan, K. et Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *Computer Vision and Pattern Recognition*. arXiv:1409.1556.

[Sindhu *et al.* (2018)] Sindhu, R. S. , Jose, G. , Shibon, S., et Varun, V. V. (2018). Using YOLO based Deep Learning Network for Real Time Detection and Localization of Lung Nodules from Low Dose CT Scans. In *Medical Imaging 2018 : Computer-Aided Diagnosis*, volume 10575, page 105751I. International Society for Optics and Photonics. doi:10.1117/12.2293699.

[Su *et al.* (2015)] Su, H. , Qi, C. R. , Li, Y., et Guibas, L. J. (2015). Render for CNN : Viewpoint Estimation in Images using CNNs Trained with Rendered 3D Model Views. In *IEEE International Conference on Computer Vision*, pages 2686–2694. DOI:10.1109/ICCV.2015.308.

[Tajbakhsh *et al.* (2016)] Tajbakhsh, N. , Shin, J. , Gurudu, S. , Hurst, R. , Kendall, C. , Gotway, M., et Liang, J. (2016). Convolutional Neural Networks for Medical Image Analysis : Full Training or Fine Tuning ? *IEEE Transactions on Medical Imaging*, 35(5) :1299–1312. doi:10.1109/TMI.2016.2535302.

[Thomas (2019)] Thomas, L. (2019). Où les reins et le foie sont localisés ? *skalapendra /shutterstock.com*. Consulté le 20 janvier 2020, https://www.news-medical.net/health/Where-are-the-Kidneys-and-Liver-Located-(French).aspx.

[Tofighi *et al.* (2018)] Tofighi, M. , Guo, T. , Vanamala, J. K., et Monga, V. (2018). Deep Networks with Shape Priors for Nucleus Detection. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 719–723. DOI:10.1109/ICIP.2018.8451797.

[Uijlings *et al.* (2013)] Uijlings, J. R. , Van De Sande, K. E. , Gevers, T., et Smeulders, A. W. (2013). Selective Search for Object Recognition. *International Journal of Computer Vision*, 104(2) :154–171. https://doi.org/10.1007/s11263-013-0620-5.

[Viola and Jones (2001)a] Viola, P. et Jones, M. (2001a). Fast and Robust Classification using Asymmetric AdaBoost and a Detector Cascade. *Advances in Neural Information Processing System*, 14 :1311–1318. https://proceedings.neurips.cc/paper/2001/file/0b1ec366924b26fc98fa7b71a9c249cf-Paper.pdf.

[Viola and Jones (2001)b] Viola, P. et Jones, M. (2001b). Robust Real-Time Object Detection. *International Journal of Computer Vision*, 4 :34–47. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.110.4868.

[Wang *et al.* (2004)] Wang, Z. , Bovik, A. C. , Sheikh, H. R., et Simoncelli, E. P. (2004). Image Quality Assessment : From error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4) :600–612. DOI:10.1109/TIP.2003.819861.

[Welander *et al.* (2018)] Welander, P. , Karlsson, S., et Eklund, A. (2018). Generative Adversarial Networks for Image-to-Image Translation on Multi-Contrast MR Images-A Comparison of CycleGAN and UNIT. *CoRR*, abs/1806.07777. http://arxiv.org/abs/1806.07777.

[Wold *et al.* (1987)] Wold, S. , Esbensen, K., et Geladi, P. (1987). Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1-3) :37–52. https://doi.org/10.1016/0169-7439(87)80084-9.

[Wolf (2018)] Wolf, S. (2018). CycleGAN : Learning to Translate Images (Without Paired Training Data). https://towardsdatascience.com/cyclegan-learning-to-translate-images-without-paired-training-data-5b4e93862c8d.

[Wolterink *et al.* (2017)] Wolterink, J. , Dinkla, A. , Savenije, M. , Seevinck, P. , van den Berg, C., et Išgum, I. (2017). Deep MR to CT Synthesis using Unpaired Data. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 14–23. Springer. https://doi.org/10.1007/978-3-319-68127-6_2.

[Xie *et al.* (2021)] Xie, X. , Niu, J. , Liu, X. , Chen, Z. , Tang, S., et Yu, S. (2021). A Survey on Incorporating Domain Knowledge into Deep Learning for Medical Image Analysis. *Medical Image Analysis*, page 101985. https://doi.org/10.1016/j.media.2021.101985.

[Xu *et al.* (2019)] Xu, X. , Zhou, F. , Liu, B. , Fu, D., et Bai, X. (2019). Efficient Multiple Organ Localization in CT Image using 3D Region Proposal Network. *IEEE Transactions on Medical Imaging*, 38(8) :1885–1898. DOI:10.1109/TMI.2019.2894854.

[Yang *et al.* (2020)] Yang, M. , Xiao, X. , Liu, Z. , Sun, L. , Guo, W. , Cui, L. , Sun, D. , Zhang, P., et Yang, G. (2020). Deep RetinaNet for Dynamic Left Ventricle Detection in Multiview Echocardiography Classification. *Scientific Programming*, 2020. https://doi.org/10.1155/2020/7025403.

[Yap *et al.* (2018)] Yap, M. , Pons, G. , Marti, J. , Ganau, S. , Sentis, M. , Zwiggelaar, R. , Davison, A., et Marti, R. (2018). Automated Breast Ultrasound Lesions Detection using Convolutional Neural Networks. *IEEE J. Biomed. Health Inform.*, 22(4) :1218–1226. doi:10.1109/JBHI.2017.2731873.

[Yi *et al.* (2019)] Yi, X. , Walia, E., et Babyn, P. (2019). Generative Adversarial Network in Medical Imaging : A review. *Medical Image Analysis*, page 101552. http://arxiv.org/abs/1809.07294.

[Yue *et al.* (2019)] Yue, Q. , Luo, X. , Ye, Q. , Xu, L., et Zhuang, X. (2019). Cardiac Segmentation from LGE MRI using Deep Neural Network Incorporating Shape and Spatial Priors. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 559–567. Springer. https://doi.org/10.1007/978-3-030-32245-8_62.

[Zaidi *et al.* (2021)] Zaidi, S. S. A. , Ansari, M. S. , Aslam, A. , Kanwal, N. , Asghar, M., et Lee, B. (2021). A Survey of Modern Deep Learning based Object Detection Models. *arXiv preprint.* arXiv:2104.11892.

[Zhang *et al.* (2016)] Zhang, R. , Zheng, Y. , Mak, T. , Yu, R. , Wong, S. , Lau, J., et Poon, C. (2016). Automatic Detection and Classification of Colorectal Polyps by Transferring Low-level CNN Features from non Medical Domain. *IEEE J. Biomed. Health Inform.*, 21(1) :41–47. DOI:10.1109/JBHI.2016.2635662.

[Zhang *et al.* (2018)] Zhang, J. , Cain, E. , Saha, A. , Zhu, Z., et Mazurowski, M. (2018). Breast Mass Detection in Mammography and Tomosynthesis via Fully Convolutional Network-based Heatmap Regression. *Medical Imaging 2018 : Computer-Aided Diagnosis, International Society for Optics and Photonics*, 10575 :1057525. https://doi.org/10.1117/12.2295443.

[Zhao *et al.* (2020)] Zhao, J. , Li, D. , Kassam, Z. , Howey, J. , Chong, J. , Chen, B., et Li, S. (2020). Tripartite-GAN : Synthesizing Liver Contrast-Enhanced MRI to Improve Tumor Detection. *Med. Image Anal.*, 101667. DOI:10.1016/j.media.2020.101667.

[Zhou *et al.* (2007)] Zhou, S. K. , Zhou, J., et Comaniciu, D. (2007). A Boosting Regression Approach to Medical Anatomy Detection. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. DOI:10.1109/CVPR.2007.383139.

[Zhu *et al.* (2017)] Zhu, J.-Y. , Park, T. , Isola, P., et Efros, A. (2017). Unpaired Image–to–Image Translation Using Cycle–Consistent Adversarial Networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2223–2232. http://arxiv.org/abs/1703.10593.

[Zou *et al.* (2019)] Zou, Z. , Shi, Z. , Guo, Y., et Ye, J. (2019). Object Detection in 20 years : A Survey. *Computer Vision and Pattern Recognition (cs.CV)*. arXiv:1905.05055.

FOLIO ADMINISTRATIF

THESE DE L'UNIVERSITE DE LYON OPEREE AU SEIN DE L'INSA LYON

NOM : HAMMAMI DATE de SOUTENANCE : 14/10/2021
(avec précision du nom de jeune fille, le cas échéant)
Prénoms : Maryam

TITRE : A priori anatomique et augmentation de données pour la détection multi-organe en imagerie médicale.

NATURE : Doctorat Numéro d'ordre : AAAALYSEIXXXX

Ecole doctorale : ELECTRONIQUE, ELECTROTECHNIQUE, AUTOMATIQUE (EEA)
Spécialité : Traitement du signal et de l'image

RESUME :

La détection d'objet, l'un des problèmes fondamentaux en vision par ordinateur, vise à localiser et à classer les instances d'objets. Elle peut constituer la première étape avant l'application d'autres méthodes de traitement d'images telles que la segmentation et le recalage. En imagerie médicale, elle est utile pour diverses applications, de la planification d'opérations chirurgicales à la recherche de pathologies.

Nous proposons une solution d'apprentissage profond au problème de la détection d'objets dans les images médicales. L'état de l'art nous a conduit à baser nos travaux sur le détecteur "You Only Look Once" (YOLO) qui fournit un bon compromis vitesse/précision. Malheureusement cette méthode, comme toutes les méthodes d'apprentissage profond, s'avère être sensible à la dimension réduite de l'ensemble d'apprentissage, problème fréquemment rencontré en imagerie médicale car l'étiquetage manuel à réaliser par les experts pour chaque organe est long et coûteux en temps.

Dans ce cadre, notre première contribution a consisté à développer une approche d'augmentation des données basée sur un "Cycle Generative Adversarial Network" (CycleGAN). Nous montrons à partir des résultats expérimentaux obtenus sur des données TDM et IRM que cette augmentation de données permet de régulariser l'apprentissage du détecteur YOLO en conduisant à des performances de détection significativement meilleures. Ces résultats montrent cependant également que cette performance peut encore être améliorée, dans la mesure où ils comportent un certain nombre de détections anatomiquement aberrantes.

Notre deuxième contribution nous a donc conduit à intégrer un a priori dans le processus de détection afin de pénaliser les valeurs aberrantes. Cet a priori est basé sur les relations spatiales existantes entre les structures anatomiques et est intégré sous la forme d'un terme supplémentaire dans la fonction de perte du détecteur YOLO. Les résultats expérimentaux obtenus montrent clairement que cette contrainte joue pleinement son rôle en diminuant significativement les erreurs de détection.

MOTS-CLÉS : imagerie médicale, apprentissage profond, détection, synthèse d'images, modèle d'a priori.

Laboratoire (s) de recherche : Centre de Recherche en Acquisition et Traitement de l'Image pour la Santé (CREATIS).

Directeur de thèse : Denis Friboulet

Président de jury :
Composition du jury :

| | | | |
|---|---|---|---|
| Caroline Petitjean | Professeur | Université de Rouen | Rapporteure |
| Michel Desvignes | Professeur | Grenoble INP | Rapporteur |
| Véronique Eglin | Professeur | INSA de Lyon | Examinatrice |
| Mireille Garreau | Professeur | Université de Rennes | Examinatrice |
| Denis Friboulet | Professeur | INSA de Lyon | Directeur de thèse |
| Razmig Kéchichian | Maître de conférences | INSA de Lyon | Co-encadrant de thèse |