

HABILITATION À DIRIGER DES RECHERCHES

présentée devant

*l'Institut National des Sciences Appliquées de Lyon
et l'Université Claude Bernard LYON I*

Titre :

Approches non-supervisées et supervisées de filtrage et segmentation pour l'analyse longitudinale en imagerie médicale

Spécialité :

Traitement de l'Image

par :

Thomas GRENIER

Soutenue le 17/03/2023 devant la commission d'examen

Examineur	Pr. Hugues BENOIT-CATTIN	INSA Lyon (CREATIS)
Président	Pr. Christophe DUCOTTET	Université Jean Monnet (LaHC)
Examineur	Pr. Charles GUTTMANN	Harvard Medical School
Rapporteur	Pr. Ludovic MACAIRE	Université de Lille (CRISAL)
Examineur	Pr. Emmanuel PERRIN	Université de Lyon (CREATIS)
Rapporteuse	Pr. Caroline PETITJEAN	Université de Rouen (LITIS)
Rapporteuse	Pr. Nicole VINCENT	Université Paris Cité (LIPADE)

Laboratoire CREATIS

CREATIS

INSA LYON

Résumé

Habilitation à Diriger des Recherches

Approches non-supervisées et supervisées de filtrage et segmentation pour l'analyse longitudinale en imagerie médicale

par Thomas GRENIER

Dans ce document, je résume mes 16 ans d'activités en tant que maître de conférences au sein de l'INSA Lyon au laboratoire CREATIS et au département Génie Electrique.

Mon enseignement de plus de 5000h est concentré sur l'informatique et le traitement du signal et des images pour lesquels j'ai proposé de nombreux contenus, projets et assumé plusieurs responsabilités d'option, de modules et d'un parcours international. J'ai aussi pu m'impliquer dans plusieurs instances du département allant de la communication et la formation aux outils numériques à la réhabilitation des locaux.

Ma recherche est focalisée sur le filtrage et la segmentation d'images médicales en non-supervisé puis en supervisé (19 publications).

Mes contributions ont d'abord porté sur les méthodes de *mean-shift* et de croissance de régions afin d'intégrer dans leurs formalismes des spécificités d'images médicales, des contraintes d'adaptations locales et de suivi temporel. Le premier enjeu a été de proposer des stratégies pour déterminer les paramètres d'échelles de ces approches, d'abord séparément (filtrage *mean-shift* et croissance de régions) puis lorsqu'elles sont utilisées simultanément. Appliquées au contexte de la segmentation de la zone de pénombre en IRM de l'accident vasculaire cérébral, zone qui fournit une information sur la propagation de la région nécrosée et donc sur l'évolution des symptômes, ces méthodes se sont révélées efficaces sur des critères de qualité de segmentation. Ensuite, nous avons ajouté la notion de temps dans les formalismes de *mean-shift* et de croissance de régions. Ceci nous a permis d'obtenir des segmentations et classification de lésions de sclérose en plaques ainsi que des segmentations des zones du myocarde lésées pertinentes.

Je me suis ensuite tourné vers les approches de segmentation supervisées et plus particulièrement vers les méthodes d'apprentissage profond (ou *deep learning*). Ainsi, l'architecture U-Net - la référence en *deep learning* pour la segmentation d'images médicales - a été mise en œuvre dans plusieurs contextes de segmentation dont, dans un premier temps, en échocardiographie où les performances obtenues ont dépassé l'état de l'art. Mes contributions à ces méthodes se sont orientées sur la réduction du nombre d'images nécessaires à leur entraînement en utilisant des techniques d'augmentation de données dont une est sélective et s'appuie sur une mesure originale de morphologie. Pour deux approches de segmentation supervisée (par recalage multi-atlas et par réseau de neurones U-Net), notre mesure de morphologie permet de définir des stratégies de sélection de données spécifiques à chacune de ces méthodes. Pour la segmentation des muscles des jambes et des épaules en IRM, ces stratégies nous ont permis d'améliorer le comportement des méthodes en terme de temps de calcul et de nombre d'images utilisé pour l'apprentissage, à qualité de segmentation équivalente.

Cette problématique de segmentation par apprentissage profond fait l'objet de mes perspectives de recherche qui visent à contribuer à l'aide au diagnostic et à terme au pronostic.

Remerciements

Merci aux membres du jury d'avoir accepté cette charge supplémentaire de travail dans les conditions complexes actuelles.

J'adresse notamment mes sincères remerciements à Mme Vincent de l'université Paris Cité, pour le travail précis de relecture de ce manuscrit.

Je témoigne aussi de ma grande reconnaissance à Mme Petitjean de l'université de Rouen, déjà présente pour le jury de thèse d'Hoai-Thu Nguyen, qui a aussi accepté de rapporter sur ce mémoire.

Mes remerciements particuliers vont à Mr Macaire de l'université de Lille, pour le temps consacré à la relecture de ce mémoire et pour ceux des thèses de Ting Li et Simon Mure. Je remercie aussi Mr Perrin de l'université de Lyon, pour les nombreux échanges autour, notamment, de la pédagogie et qui a accepté de participer au jury de cette HDR.

Pour ce jury, je suis aussi très heureux de pouvoir compter sur la présence Mr Ducottet de l'université Jean Monnet. Une présence qui m'est chère : l'enseignant d'informatique lorsque j'étais étudiant, rapporteur pour ma thèse puis de plusieurs doctorants co-encadrés, collaborateur sur différents projets... Merci une nouvelle fois pour le temps et les conseils. Je suis aussi honoré d'avoir Mr Guttman d'Harvard Medical School, dans ce jury. Son expertise -entre autre- sur la sclérose en plaques, son anticipation et la cohérence sur les études à conduire ont été de véritables sources d'inspiration pour bâtir des projets, notamment une thèse dirigée ensemble.

Enfin, merci à Hugues Benoit-Cattin pour tout le temps passé, les discussions, les propositions, les relectures, la patience lors des différents travaux dont cette HDR ! Mais il y a aussi tant d'autres choses hors du cadre recherche : les cours et la pédagogie, l'INSA, tes qualités de communications et humaines ... C'est un véritable plaisir de pouvoir échanger et réfléchir avec toi.

Je tiens aussi à remercier les collègues qui m'ont permis de m'épanouir dans ce métier. Je remercie en premier lieu Mr Christophe Odet qui m'a transmis un bon nombre de démarches pour la pédagogie, l'évaluation et les soutenances des étudiants, en plus de tous les échanges en traitements du signal et des images, puis de l'informatique domaines dans lesquels sa maîtrise et ses expériences ont été (sont encore !) de véritables références. Je tiens à remercier Mr Tanneguy Redarce et Mr Claude Richard pour la confiance et l'autonomie qu'ils m'ont laissées au sein du département Génie Electrique afin de conduire les projets que j'avais à cœur. J'adresse aussi tout mes remerciement à Chantal Muller qui m'a encadré pendant ma thèse puis encore un peu après ;)

Puis il y a les -autres- collègues du 401 : Jean-Loïc Rose, Boubakeur Belaroussi, Jérôme Velut, Sorina Pop, Delphine Charpigny... merci pour les échanges, les rigolades et toutes les interactions qu'on a depuis !

Je tiens aussi à remercier les doctorants avec qui j'ai travaillé (ou travaille encore) et sans qui pas grand chose ne serait possible : Ting, Simon, Sarah, Clément, Thu, Léo, Valentine, Enyi et Emile. Je n'oubli pas non plus les post-doctorants et stagiaires. La liste, sera donnée plus loin.

Enfin, j'adresse mes remerciements aux collègues et personnels de CREATIS qui, depuis 2001, apportent un cadre scientifique et technique de haut niveau, stimulant et épanouissant.

*À ma famille et mes proches,
À Myriam,
À Jade et Éthan, mes deux petits remèdes contre le travail.*

Table des matières

1	Introduction Générale	21
----------	------------------------------------	-----------

I	Synthèse des activités	
----------	-------------------------------	--

2	Introduction et présentation	27
----------	---	-----------

3	Curriculum Vitae	29
----------	-------------------------------	-----------

4	Activités d'enseignement	33
----------	---------------------------------------	-----------

4.1	Enseignements	33
------------	----------------------	-----------

4.1.1	Informatique	34
-------	--------------------	----

4.1.2	Traitement du signal et des images	35
-------	--	----

4.1.3	Tutorats au département Génie Electrique	36
-------	--	----

4.2	Responsabilités	37
------------	------------------------	-----------

4.2.1	Responsable de la formation IMESI	38
-------	---	----

4.2.2	Chargé de la communication au département Génie Électrique	38
-------	--	----

4.2.3	Application stage	39
-------	-------------------------	----

4.2.4	Jouvence informatique	39
-------	-----------------------------	----

4.2.5	Responsable de l'option Traitement des Signaux et des Images (TdSI)	40
-------	---	----

4.2.6	Membre élu au conseil de département Génie Electrique	40
-------	---	----

4.2.7	Responsable du module "Traitement des Images - Approfondissement"	40
-------	---	----

4.3	Bilan	41
------------	--------------	-----------

5	Activités de recherche	43
5.1	Résumé des travaux	44
5.2	Participation à l'encadrement de post-doctorants	45
5.3	Participation à l'encadrement et à la codirection de doctorants	45
5.4	Encadrements d'étudiants en master et en projet de fin d'études	48
5.5	Projets	50
5.6	Contrats industriels	51
5.7	Responsabilités	52
6	Production scientifique	55

II Contributions au filtrage et à la segmentation non supervisés

7	Introduction	63
8	Filtrage <i>mean-shift</i> et segmentation	65
8.1	Introduction	65
8.2	Filtrage d'images et <i>mean-shift</i>	65
8.2.1	Brève histoire du filtrage <i>mean-shift</i>	65
8.2.2	Notations et principes <i>mean-shift</i>	66
8.2.3	Processus <i>mean-shift</i>	67
8.2.4	Filtrage d'images par <i>mean-shift</i>	68
8.3	Optimisations des paramètres d'échelles	70
8.4	Du filtrage <i>mean-shift</i> à la segmentation d'images	71
8.4.1	Croissance de région pour l'espace joint	72
8.4.2	Approche MS+RG	73
8.4.3	Échelles pour MS+RG et carte de probabilités	75
8.5	Application à l'accident vasculaire cérébral	77
8.5.1	Résultats sur le rat	78
8.5.2	Résultats sur l'homme	79
8.6	Conclusion	80
9	Filtrage et segmentation spatio-temporels	83
9.1	Introduction	83
9.2	Diffusion Anisotrope temporelle	83
9.2.1	Introduction	83
9.2.2	Méthodes AD+t et RAD+t	84
9.2.3	Evaluation et Application	86
9.2.4	Conclusion	88
9.3	<i>Mean-shift</i> Spatio-Temporel : STMS	90
9.3.1	Introduction	90
9.3.2	Méthode STMS	91

9.3.3	Evaluation sur des données de synthèse	93
9.3.4	Application à l'étude de lésions SEP	94
9.3.5	Conclusion	96
9.4	Croissance de région spatio-temporelle : STRG	97
9.4.1	Introduction	97
9.4.2	Méthode STRG	98
9.4.3	Application à la segmentation de lésions ischémiques du myocarde	98
9.4.4	Conclusion	101
9.5	Conclusion du chapitre	101
10	Conclusion de la seconde partie	103

III

Contributions en segmentation supervisée

11	Introduction	115
12	Segmentation par recalage d'atlas	117
12.1	Introduction	117
12.2	Principe du recalage d'atlas	117
12.3	Approche multi atlas <i>JLF + CL</i>	119
12.3.1	Fusion <i>jointe</i> des étiquettes - JLF	119
12.3.2	Apprentissage des corrections - CL	122
12.4	Conclusion	123
13	Réseau U-Net	125
13.1	Introduction	125
13.2	Segmentation par apprentissage automatique et profond	125
13.3	Architecture U-Net	126
13.4	Les coulisses d'U-Net	126
13.4.1	Jeu d'images	127
13.4.2	Fonction de coût	128
13.4.3	Les différentes couches du réseau	128
13.4.4	Fonction d'activation	132
13.5	Quelques variantes d'U-Net	133
13.6	Conclusion	136
14	Applications à l'imagerie médicale	137
14.1	Introduction	137
14.2	Segmentation des muscles en IRM avec JLF+CL	137
14.2.1	Problématique de segmentation de la base MUST	137
14.2.2	Optimisation des paramètres JLF+CL	138
14.2.3	Comparaison des méthodes basées atlas	140
14.2.4	Généralisation pour le longitudinal	141

14.2.5	Généralisation à la base Epaule	142
14.2.6	Discussion	144
14.3	Segmentations avec U-Net	145
14.3.1	Segmentation d'images écho-cardiographique avec U-Net	145
14.3.2	Segmentation des muscles de l'épaule	148
14.4	Contribution à la correction automatique	150
14.5	Conclusion	153
15	Augmentation de données	155
15.1	Introduction	155
15.2	Augmentations d'images par recalages	156
15.3	Sélection d'atlas basée sur la morphologie	157
15.3.1	Sélection d'atlas pour la segmentation multi-atlas	159
15.3.2	Augmentation de données basée morphologie pour U-Net	161
15.4	Conclusion	164
16	Conclusion de la troisième partie	165

IV

Conclusion Générale

17	Conclusion et projet de recherche	177
17.1	Conclusion	177
17.2	Le bilan des contributions	177
17.3	Perspectives de recherche	179
17.3.1	Couplage fort entre méthodes de deep learning et d'autres méthodes	179
17.3.2	Apprentissage continu, apprentissage actif	180
17.3.3	Complexité du problème de segmentation et dimensionnement	180
17.3.4	Apprentissage faiblement supervisé de données hétérogènes pour le pronostic	180
17.4	Bilan de rédaction de mon HDR	181

Annexes

A	Métriques pour l'évaluation quantitative	185
A.1	Evaluations quantitatives du filtrage d'image	185
A.2	Evaluations quantitatives de la segmentation d'image	186
A.2.1	Indice de Sorensen-DICE	186
A.2.2	Coefficient de Jaccard ou Intersection de l'union	187
A.2.3	Distance de Hausdorff, d_H	188
A.2.4	Distance absolue moyenne, d_m	188
A.2.5	Similarité de volume	189

B	Descriptions des bases de données	191
B.1	Base de données MUST	191
B.1.1	Imagerie IRM	191
B.1.2	Base de données	192
B.1.3	Pré-traitements	193
B.2	Base de données Epaule	194
B.2.1	Imagerie IRM	195
B.2.2	Base de données	195
B.2.3	Pré-traitements	197
C	Description de l'algorithme AdaBoost	199

Table des figures

4.1	Répartition du service d'enseignement	34
4.2	Exemples de Projets de Réalisation Technologique	37
4.3	Exemple de jouvence salle informatique du département Génie Electrique de l'INSA Lyon	40
5.1	Synthèse des activités de recherche	43
5.2	Illustration de la segmentation par recalage d'atlas (en haut) puis détails des labels des os dans deux coupes (en bas), Projet PIPER.	51
5.3	Bandeau du site web de l'école d'été co-organisée.	53
5.4	Bandeau du site web de la seconde édition de l'école d'été.	53
5.5	Bandeau du site web de la troisième édition de l'école d'été.	53
8.1	Fonction de pondération du filtrage anisotrop de Barash.	66
8.2	Description intuitive de l'évolution itérative <i>mean-shift</i> .	67
8.3	Processus <i>mean-shift blurring</i> et <i>non-blurring</i>	69
8.4	Illustration du filtrage Anisotrope et <i>mean-shift</i>	69
8.5	Filtrage <i>mean-shift</i> de l'image de synthèse IRM avec différents paramètres d'échelles.	70
8.6	Evolution du <i>PSNR</i> en fonction de h_s et h_r	71
8.7	Processus de croissance de région.	72
8.8	Illustration du voisinage de y dans un ensemble de points.	73
8.9	Pipeline de segmentation pour l'approche MS+RG.	74
8.10	Segmentation par l'approche MS+RG de la matière blanche sur l'image IRM simulée et corrompue avec du flou et du bruit gaussiens.	74
8.11	Illustration des trois stratégies d'échantillonnage des plages de paramètres d'échelles.	75
8.12	Cartes de probabilités et leurs seuillage pour les 3 différents stratégies d'échantillonnage des paramètres d'échelle.	76
8.13	Ischémie cérébrale (à gauche) et son évolution (à droite).	77
8.14	Segmentations optimales obtenues avec les paramètres ADC et peak sur 7 coupes du rat 'pMCAO2'.	79

8.15	Cartes de probabilités de segmentation de la pénombre du rat 'pMCAO2' pour trois combinaisons de cartographies IRM.	80
8.16	Illustrations des séquences IRM pour un AVC chez l'homme (11 ^e coupe du patient 171). 81	
8.17	Cartes probabilités sur les 20 coupes du patient 171 calculées avec les cartographies IRM : TTP, MTT, CBF, ADC et DWI. La référence à 30 jours est donnée pour chacune des coupes.	81
9.1	Modèle synthétique pour les tests avant floutage et bruitage.	87
9.2	Résultats de filtrage pour le 20 ^e instant de la séquence CS_2	88
9.3	Comparaison des filtrages BM3D et RAD+t.	89
9.4	Principe de STMS, la sélection spatio-temporelle et son évolution itérative	93
9.5	Données synthétiques pour l'évaluation de l'approche STMS	94
9.6	Données synthétiques et résultats des approches MS et STMS	95
9.7	Coupe IRM pondérée T2 et image de différence avec l'instant initial d'un patient atteint de SEP	95
9.8	Filtrage <i>mean-shift</i> spatio-temporel d'une lésion sur 22 instants.	96
9.9	Filtrage et clustering STMS puis STMS et DTW pour 4 lésions de SEP. L'algorithme de Déformation Temporelle Dynamique (DTW) permet de regrouper les régions dont l'évolution est similaire malgré une déformation du temps.	97
9.10	Etapes du processus de segmentation des lésions ischémiques en IRM de perfusion.	99
9.11	Segmentations de 6 patients avec l'approche proposée dans (DAVILLER et al., 2019a). 100	
9.12	Etudes de la variabilité des valeurs de MBF	101
12.1	Principe de la segmentation par recalage d'atlas.	118
12.2	Principe de la segmentation par recalage d'atlas JLF+CL de (WANG et al., 2013b). 119	
13.1	Architecture du réseau U-Net.	126
13.2	Schéma de principe de l'apprentissage supervisé.	127
13.3	Schéma de principe des couches de convolution.	130
13.4	Schéma de principe d'une couche de sous échantillonnage.	130
13.5	Schéma de principe de la convolution par image dilatée pour le sur-échantillonnage. 131	
13.6	Principales fonctions d'activation pour les réseaux de neurones.	132
13.7	Schéma de principe du calcul de la fonction softmax.	134
14.1	Définition des 4 muscles à segmenter pour l'étude des quadriceps.	138
14.2	Évolution des scores de DSC et HD en fonction du nombre d'atlas pour l'approche JLF+CL sur les 4 muscles étudiés (VL, RF, VM et VI).	139
14.3	Résultats de segmentation des quadriceps obtenus avec les approches de (GILLES et al., 2016), JLF (WANG et al., 2013b) et JLF+CL pour un coureur.	140
14.4	Evolution des scores de DSC et du temps de calcul en fonction du nombre d'atlas pour les approches JLF (en haut) et JLF+CL (en bas).	143
14.5	Exemple de segmentations des muscles de l'épaule obtenues avec les approches JLF et JLF+CL en utilisant 5 atlas tirés au hasard.	143
14.6	Influence du nombre d'atlas sur la qualité de segmentation et le temps de calcul pour les données MUST.	145
14.7	Echocardiographie (à gauche) et sa segmentation experte (à droite) de l'épicaarde du ventricule gauche (en rouge), l'endocarde du ventricule gauche (en vert) et de l'oreillette gauche (en bleue).	146
14.8	Exemples de résultats de segmentation du ventricule gauche obtenus avec U-Net sur la base de données électrocardiographie CAMUS.	146
14.9	Exemple de segmentations par U-Net obtenues par l'apprentissage suivant chacun des plans de référence anatomique pour la base de données Epaule.	149

14.10	Résultats de segmentation des meilleures méthodes basées U-Net sur la base de données Epaule.	149
14.11	Segmentations manuelles de 3 coupes de la base de données Epaule et les résultats de segmentations automatiques par JLF+CL avec les 5 meilleurs atlas, et par ResU-Net101 et supervision profonde.	151
14.12	Segmentations de trois coureurs MUST avec les approches JLF+CL, U-Net et U-Net+CL.	152
14.13	Utilisation de réseaux U-Net pour remplacer les étapes JLF de l'approche de (WANG et al., 2013b).	153
15.1	Comparaison de segmentations d'un examen de la base MUST avec les approches JLF et U-Net sans augmentation de données.	155
15.2	Stratégies d'augmentation de données pour la base de données MUST.	157
15.3	Illustration sur un sujet de MUST de la caractérisation morphologique.	158
15.4	Projection bi-dimensionnelle par ACP des vecteurs morphologiques des jambes droites des coureurs de MUST.	159
15.5	Résultats de segmentation des approches JLF et JLF+CL (WANG et al., 2013b) avec et sans sélection d'atlas basée sur la morphologie.	160
15.6	Schéma de construction du jeu d'apprentissage pour la stratégie <i>target-driven</i> de U-Net.	163
A1	Illustration pour le calcul de l'indice de DICE.	186
A2	Illustration du coefficient de Jaccard ou IoU.	187
A3	Illustration des distances de Hausdorff.	188
A4	Illustration de la distance absolue moyenne	189
B1	Acquisition IRM (à gauche), segmentation manuelle sur une coupe (au centre) et rendu 3D des segmentations des quadriceps et des os (à droite).	192
B2	Illustration des 5 types d'images utilisés pour l'étude de l'inflammation des muscles de la cuisse.	192
B3	Illustration de l'évolution de l'inflammation des muscles de la cuisse sur les cartographies T2*.	193
B4	Définition des 4 muscles à segmenter pour l'étude des quadriceps.	193
B5	Correction du biais avec l'algorithme N4.	194
B6	Pipeline proposé pour étudier l'impact du recalage sur l'extraction de caractéristiques radiomiques.	195
B7	Comparaison des approches de recalage pour la mise en correspondance des images pondérées T1 et des cartographies T2 et T2*.	195
B8	Exemple d'acquisitions IRM de la base de données Epaule.	196
B9	Exemple de segmentation manuelle des 5 muscles de la base de données Epaule.	196
B10	Illustration des différents pré-traitements effectués sur la base de données Epaule afin d'obtenir des images homogènes et correspondant à des épaules droites.	197

Liste des tableaux

8.1	Comparaison des valeurs optimales de <i>PSNR</i> avec les méthodes de <i>plug in</i> et de recherche exhaustive	72
8.2	Plages des paramètres d'échelles pour la recherche exhaustive.	74
8.3	Intervalles des paramètres d'échelle pour les deux stratégies de tirage aléatoire.	76
8.4	Scores de <i>DICE</i> suite au seuillage optimal des cartes de probabilités sur l'IRM corrompue.	77
8.5	Influence des combinaisons de paramètres IRM sur les segmentations produites par la croissance de région (RG) et l'approche MS+RG	79
9.1	Paramètres utilisées pour les expériences.	88
9.2	Résultats optimaux obtenus avec RAD+ <i>t</i> et BM3D pour la MSE (à minimiser) et le SSIM (à maximiser). Pour les deux mesures, les optimisations ont été faites sur la MSE.	89
14.1	Evaluation quantitative des segmentations des quadriceps basées sur le recalage d'atlas.	141
14.2	Détails des scores de DSC obtenus par les approches JLF et JLF+CL pour la base MUST.	141
14.3	Comparaison quantitative des approches JLF, JLF+CL et U-Net en longitudinal sur les données MUST.	142
14.4	Résultats de segmentation des approches JLF et JLF+CL pour les données Epaule, avec tous les patients. Moyennes et écarts types calculés pour 10 patients.	144
14.5	Résultats de segmentation des approches JLF et JLF+CL pour les données Epaule, sans les 3 patients faisant échouer les recalages. Moyennes et écarts types calculés pour 10 patients.	144
14.6	Comparaison des performances de huit méthodes de segmentation du ventricule gauche en échocardiographie.	147
14.7	Evaluations quantitatives des segmentations U-Net obtenues par l'apprentissage suivant chacun des plans de référence anatomique pour la base de données Epaule avec l'image "e8" comme entrée.	148
14.8	Evaluations quantitatives des segmentations obtenues par les différentes variantes d'U-Net sur la base de données Epaule.	150

14.9	Evaluation quantitative des segmentations JLF, JLF+CL, U-Net et U-NET+CL sur la base de données MUST.	153
15.1	Evaluation quantitative des stratégies d'augmentation d'images pour la segmentation de MUST.	157
15.2	Evaluations quantitatives des segmentations JLF, JLF+CL et JLF+CL6 avec un choix aléatoire ou une sélection basée sur la morphologie des atlas.	161
15.3	Evaluation quantitative de l'augmentation de données aléatoires et par sélection morphologique pour U-Net.	162
15.4	Evaluation quantitative des stratégies morphologies spécifiques d'entraînement d'U-Net pour chaque atlas de MUST.	163
15.5	Rappel des grandeurs d'inférence pour les méthodes basées sur la morphologie. . .	164
B1	Evaluation de la segmentation inter-experts de la base MUST.	194

CreatisCREATISINSA
LYONINSA
INSTITUT NATIONAL
DES SCIENCES
APPLIQUÉES
LYONcnrs
dépasser les frontièresInserm

1. Introduction Générale

Maître de conférences depuis 2006 à l'INSA Lyon, il me semble naturel de présenter mon habilitation à diriger des recherches (HDR) afin de continuer d'évoluer dans mon engagement institutionnel. Ce document dresse un bilan des expériences acquises en pédagogie, en recherche et dans les différentes responsabilités exercées lors de ces 16 premières années d'enseignant chercheur qui, je l'espère, convaincront pour obtenir le diplôme d'HDR.

Mes motivations à présenter l'HDR sont multiples.

- Avoir une reconnaissance des compétences acquises pendant ces années,
- Avoir plus d'autonomie pour l'encadrement de doctorants, tout en préservant un encadrement partagé qui me semble particulièrement efficace et synergique,
- Postuler dans les années à venir aux postes et aux responsabilités de professeur des universités,
- Accroître ma légitimité au travers de ce diplôme, notamment pour le pilotage d'équipes, projets ou work-packages,
- Prendre des responsabilités permettant de contribuer aux évolutions dans mes domaines d'enseignements et de recherche.
- Plus "institutionnelle" : l'enjeu pour les laboratoires du nombre d'HDR a toujours été et -a priori restera- une quantité critique pour la reconnaissance auprès des instances d'évaluation et des écoles doctorales, ne serait-ce que pour l'obtention de bourses de thèse.

Je vais présenter dans la suite de ce document un résumé de mes activités d'enseignements et de recherche ainsi que de mes implications dans les différents instances ou projets.

Depuis 2006, mon enseignement est réalisé au sein du département Génie Electrique de l'INSA Lyon. C'est dans ce département que j'ai pu conduire de nombreux projets pédagogiques, de communications, de gestion et organisationnels. J'y ai aussi développé ma pratique pédagogique au travers d'une diversité d'enseignements et de création de contenus.

Pour la partie recherche, dans ce document, j'ai choisi de résumer mes activités depuis 2011. Elles couvrent un travail sur le filtrage et la segmentation d'images médicales. On considère, à mon sens très justement, que la segmentation d'images est l'étape la plus commune pour

permettre une analyse d'images, notamment médicales. La segmentation d'images est donc une limite aux analyses. Cependant, le filtrage est un préliminaire critique pour obtenir des segmentations de qualité et de manière fiable sur un grand nombre d'images. Ces deux thèmes, couplés ensemble, m'apparaissent depuis la fin de ma thèse comme centraux et constituent mon projet de recherche.

Appliquer ces méthodes à l'imagerie médicale est particulièrement stimulant par rapport à la finalité applicative, aux spécificités de l'imagerie médicale, à la diversité des problèmes puis à la rigueur nécessaire.

Depuis ma thèse, je teste et améliore des méthodes sur des images médicales. J'ai pu me confronter à plusieurs modalités 2D ou 3D (principalement TEP, IRM, CT, US) et à différentes pathologies (cancers, AVC, sclérose en plaques, inflammations musculaires) en se focalisant sur des régions anatomiques diverses (cerveau, jambes, cœur, corps entier) pour différents buts (mesures diagnostiques, compréhension de la pathologie, prédictions, alimentation de simulateurs). Malgré la diversité apparente de ces applications, la démarche est en fait assez similaire.

1. Bien comprendre l'objectif de l'étude, apprendre le contexte médical, étudier l'état de l'art, déterminer les paramètres pertinents qui seront analysés, proposer une méthode d'évaluation. Comprendre et arrêter le protocole d'acquisition des données.
2. Appréhender les données et savoir les lire, les représenter ; de même avec les annotations des experts... quand il y en a (donc souvent : mobiliser et accompagner les experts pour créer ou valider les annotations).
3. Proposer une chaîne de traitements d'images originale permettant d'arriver à l'extraction des paramètres plus efficacement que dans l'état de l'art. Cette chaîne de traitements doit prendre en compte les spécificités de l'étude (nombre d'images, bruit, artéfacts, complexité anatomiques, reproductibilité, biais ...) ce qui conduit à une part d'innovation et de recherche méthodologique plus ou moins fastidieuse en filtrage et en segmentation.
4. Implémenter, tester, vérifier, re-tester, quantifier, comprendre ses erreurs d'analyse, faire évoluer ses savoirs et intuitions sur les méthodes de segmentation et de filtrage.
5. Valoriser les travaux.

Dans ce manuscrit, seules les contributions en filtrage et segmentation sont présentées et seulement pour quelques applications.

Ce manuscrit est organisé en trois parties. La première partie dresse un bilan de mes activités pédagogiques (chapitre 4) et de recherche (chapitre 5) en incluant mes implications collectives et les projets conduits. La liste des mes publications est donnée en fin de cette première partie (chapitre 6).

La seconde partie est dédiée à mes contributions à l'imagerie médicale en filtrage et segmentation non-supervisés basées sur le formalisme *mean-shift*. Ce formalisme et son optimisation dans le cadre du filtrage et de la segmentation sont donnés au chapitre 8. Puis, les extensions au domaine spatio-temporel en filtrage (STMS) et en segmentation par croissance de région (STRG), sont détaillées au chapitre 9.


La troisième partie adresse mes recherches plus récentes en segmentation supervisées. Cette partie commence par une étude d'une méthode de segmentation basée sur le recalage d'atlas (chapitre 12). Puis, au chapitre 13, on présente le fonctionnement d'une méthode de segmentation par apprentissage profond (U-Net). Nos principales applications de ces méthodes à différents contextes de segmentation d'images médicales sont données au chapitre 14. Puis, le chapitre 15 présente nos travaux sur la sélection et l'augmentation des données particulièrement critique pour les deux approches de segmentation supervisée étudiées dans cette troisième partie.

Enfin, la conclusion de ce document (chapitre 17) donne des pistes de recherche que j'aimerais conduire autour de ces méthodes. Ces pistes concerneraient notamment les possibles couplages entre les approches "classiques" (*mean-shift*, croissance de région,...) et d'apprentissage automatique, puis l'interaction des algorithmes d'apprentissage avec les experts. Et enfin, comment mieux comprendre le comportement des méthodes d'apprentissage profond afin de pouvoir répondre aux questions sur les performances atteignables avec un réseau et une base d'images donnée.



Synthèse des activités

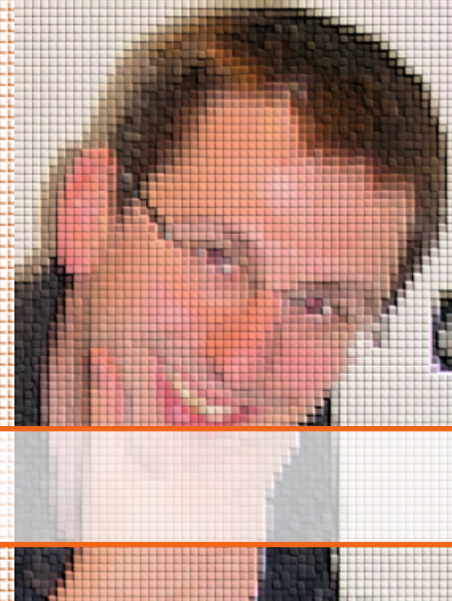
2	Introduction et présentation	27
3	Curriculum Vitae	29
4	Activités d'enseignement	33
4.1	Enseignements	
4.2	Responsabilités	
4.3	Bilan	
5	Activités de recherche	43
5.1	Résumé des travaux	
5.2	Participation à l'encadrement de post-doctorants	
5.3	Participation à l'encadrement et à la codirection de doct-rants	
5.4	Encadrements d'étudiants en master et en projet de fin d'études	
5.5	Projets	
5.6	Contrats industriels	
5.7	Responsabilités	
6	Production scientifique	55



2. Introduction et présentation

Cette partie est consacrée au bilan de mes activités de maître de conférences des universités depuis ma nomination en 2006. Après avoir exposé mon *curriculum vitae* au chapitre 3, les chapitres suivants donnent les détails de mes activités d'enseignement (chapitre 4) et de recherche (chapitre 5). Cette partie se termine par une liste de ma production scientifique (chapitre 6).

3. Curriculum Vitae



Thomas GRENIER

NUMEN : 10S0600268NYL
Né le 2 mai 1978 à Vénissieux (69)
Nationalité : Française
Concubinage, 2 enfants

Coordonnées professionnelles

e-mail : thomas.grenier@creatis.insa-lyon.fr
Téléphone : 04 72 43 64 70
Web : www.creatis.insa-lyon.fr/~grenier
CREATIS Unité CNRS UMR 5220 – INSERM U1206
Bâtiment Léonard de Vinci
21 avenue Jean Capelle
69621 Villeurbanne cedex FRANCE

Status actuel

Grade : Maître de conférences, Hors Classe
Section CNU : 61
Unité de recherche : CREATIS, Unité CNRS UMR 5220 – INSERM U1206
Enseignement : Département Génie Électrique, INSA Lyon
Distinction honorifique : **Chevalier** de l'ordre des palmes académiques, 2021

Diplômes

- ▶ **2005 - Thèse de Doctorat** INSA Lyon, Spécialité Image, obtenue le 12/12/2005. Président : P. Réfrégier. Rapporteurs : M. Jourlin, J. M. Nicolas, D. Comaniciu. Directeur : G. Gimenez. Co-directeur : C. Müller. « *Apport de l'espace des caractéristiques et des paramètres d'échelle adaptatifs pour le filtrage et la segmentation d'images* »,
- ▶ **2001 - Ingénieur ISTASE** option imagerie et instrumentation médicale, mention Bien, Université Jean Monnet - Saint Etienne,
- ▶ **2001 - DEA** spécialité Image, mention assez bien, Université Jean Monnet - Saint Etienne, sujet « Transformée de Hough et généralisations »,
- ▶ **1998 - DUT** Génie Électrique et Informatique Industrielle, Université Jean Monnet - Saint-Étienne.

Domaines de recherche

- ▶ **Filtrage d'images** : approches par diffusion anisotrope et *mean-shift*, multiparamétriques et temporelles,
- ▶ **Segmentation d'images** : approches par croissance de région et clustering, approche supervisée (apprentissage profond),
- ▶ **Applications médicales** : analyse longitudinale d'images IRM, accident vasculaire cérébral, sclérose en plaques, inflammation musculaire.

Publications

- ▶ 19 publications dans des revues internationales,
- ▶ 1 publication dans une revue nationale,
- ▶ 38 communications dans des conférences (dont 9 communications nationales),
- ▶ 2 brevets (FR2108162, déposé à l'INPI; PCT/EP2022/070697, déposé à l'OMPI)

Encadrements et co-encadrements de jeunes chercheurs

- ▶ 3 thèses co-dirigées (50%) et soutenues 2012, 2016 et 2021,
- ▶ 1 thèse co-encadrée (30%) soutenue en 2019,
- ▶ 2 thèses co-dirigées (40% et 50%) en cours (fin en octobre 2024 et 2025),
- ▶ 1 thèse co-encadrée (30%) en cours (fin en juin 2023),
- ▶ 22 encadrements d'étudiants en stage de master recherche et projets de fin d'études,
- ▶ 2 post-doctorants.

Collaborations nationales et internationale

- ▶ Laboratoire LaHC (Saint-Etienne), 2019-2021, *Segmentation et suivi de nanoparticules, imagerie ETEM, analyse de processus chimiques*,
- ▶ Hospice Civils de Lyon, 2019- , *Segmentation de muscles de l'épaule, en IRM, prédiction intérêt opératoire*,
- ▶ Hospice Civils de Lyon, 2019- , *Segmentation de muscles du tronc, en CT, évaluation de la sarcopénie et des risques en réanimation*,
- ▶ Laboratoire LYOS, 2019- , *Segmentation des os et des métastases pour la prédiction du risque de fracture par simulation numérique*,
- ▶ Laboratoire INL (Lyon), 2019-2020, *Classification de séquences vidéos de déformation de globules rouges, microscopie rapide, identification de pathologies (Paludisme)*,
- ▶ Laboratoire Mateis et IRCELYON (Lyon), 2015-2020, *Reconstruction, filtrage et suivi de région en imagerie ETEM, acquisition ultra-rapide de processus de catalyse*,
- ▶ Laboratoire IFSTTAR (Bron), 2013-2015, *Segmentation et étiquetage des os, imagerie CT, création de modèle pour la simulation de crash test*,
- ▶ Harvard Medical School, USA, 2013-2016, *Analyse longitudinale de lésions de sclérose en plaques, IRM, compréhension de la pathologie*.

Projets et actions contractuelles

- ▶ 2019-2020, Contrat de collaboration avec les Hospice Civils de Lyon, projet CovidIA visant à analyser les muscles du tronc afin d'évaluer les risques en réanimation, données + 1k€ de dotation,
- ▶ 2015- , 12 contrats d'expertise industrielle, montant 65k€,
- ▶ 2014-2015, PEPS CNRS INS2I *Classification Spatio-Temporelle d'ensemble de données multiparamétriques. Application à la caractérisation de la phase chronique des lésions SEP en suivi longitudinal IRM*, 16k€,
- ▶ 2011, Bonus Qualité Formation INSA, porteur du projet *calculs sur systèmes embarqués*, dotation de 9k€, 2011.

Animations et responsabilités

- ▶ 2019- , Membre du Comité de Pilotage pédagogique de l'EUR MANUTECH SLEIGHT et référent pour le département Génie Electrique,
- ▶ 2019- , Co-responsable du projet transversal sur la Sclérose en plaque MUSIC au laboratoire Creatis,
- ▶ 2018- , Co-responsable du Work Package Formation du Labex PRIMES,
- ▶ 2016-2020, Responsable de l'option de 5^e année du département Génie Electrique en traitement des signaux et des images TdSI, 20 étudiants, 12 ECTS,
- ▶ 2016 et 2017, Coprésident des commissions d'élaboration des sujets de baccalauréat Technologique d'enseignements technologiques transversaux STI2D, (DEC 7),
- ▶ 2013-, Responsable des outils numériques pour le département Génie Electrique,
- ▶ 2009-2013, Chargé de la communication au département Génie Electrique (et membre de l'équipe de direction),
- ▶ 2007-2014, Responsable de la formation IMESI de L'INSA Lyon.

Enseignements

- ▶ 2006 - 2022, Maître de conférences, département Génie Electrique, 5000h de CM/TD/TP/Projet en informatique, traitement du signal et traitement des images,
- ▶ 2005 - 2006, ATER 100% IUT B Lyon, département Génie Mécanique et Productique, 192h TP Electrotechnique,
- ▶ 2004 - 2005, ATER 50% IUT B Lyon, département Génie Mécanique et Productique, 96h TD informatique et bureautique,
- ▶ 2001 - 2004, Moniteur, IUT B Lyon, département Génie Industriel et Maintenance, 96h projet électronique.



4. Activités d'enseignement

Dans cette partie, je liste mes activités liées à la pédagogie en commençant par les différents enseignements, puis les tutorats effectués et enfin les différentes responsabilités occupées.

4.1 Enseignements

J'enseigne depuis septembre 2006 au département Génie Électrique de l'INSA Lyon ce qui correspond à un total cumulé de plus de 5500h résumées dans les trois sous-parties ci-après. Mes premières expériences pédagogiques ont été réalisées à l'IUT B de Lyon pendant mes années de monitorat et d'ATER (soit environ 480h) au sein des départements Génie Industriel et Maintenance puis Génie Mécanique et Productique. Elles consistaient principalement à proposer et superviser des projets de robotique et d'électronique (192h sur 3 ans), des TP d'électrotechnique (192h) et des cours/TD de bureautique et informatique (96h).

Depuis ma nomination en 2006 au département Génie Electrique de l'INSA Lyon, mes enseignements sont regroupés dans deux thématiques : l'informatique et le traitement du signal et de l'image. Génie Electrique est un département de formation au métier de l'ingénieur sur 3 ans et dont l'effectif par promotion est de 120 étudiants en filière classique et de 25 étudiants en filière par apprentissage (depuis 2012). Mes enseignements sont principalement réalisés en quatrième année (dans les filières classique et en apprentissage) et dans deux options de cinquième année : TdSI (Traitement du Signal et des Images) et SEC (Systèmes Embarqués Communicants). L'essentiel de mon service est effectué dans la formation d'ingénieur du département Génie Électrique.

Pendant les années 2008 à 2013, une partie de mon service s'effectuait dans la formation IMESI¹ de l'INSA Lyon. Il s'agissait d'un parcours international en anglais, orienté recherche sur les domaines des systèmes embarqués et du traitement de l'image médicale, et dont j'étais le responsable. Les différentes responsabilités exercées seront données dans la partie 4.2.

1. International Master of Embedded Systems and Medical Image Engineering

Enfin, j'effectue quelques heures dans le parcours *Systèmes et Images* du Master 2 recherche MISS² de l'université de Lyon 1 (anciennement Master GE/GP puis EEAP).

La figure 4.1 détaille la répartition (CM, TD, TP, projet, encadrement de stagiaires/PFE, primes) de mes services d'enseignement depuis ma nomination à l'INSA Lyon.

Je détaille ensuite les différents enseignements réalisés et mes contributions.

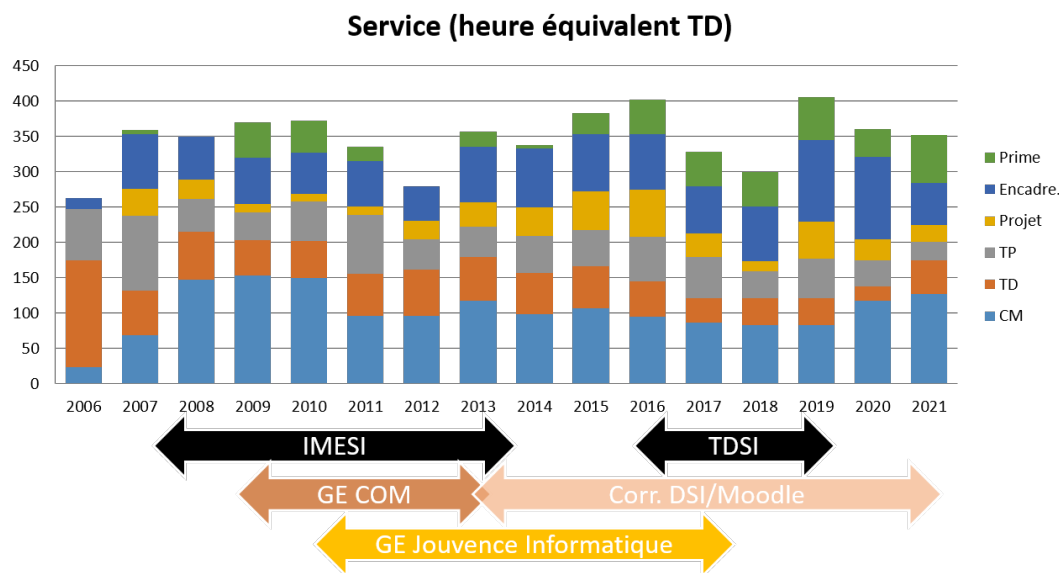


FIGURE 4.1 – Répartition du service d'enseignement depuis la nomination. Le volume horaire de chaque activité est ramené à la rémunération en heure équivalent TD.

4.1.1 Informatique

- ▶ **Programmation en langage C**, 3^e année département Génie Électrique, 120 étudiants par promotion, $8 \times 2\text{hTD}$
Ces séances de TD permettent de pratiquer les fondamentaux de la programmation en langage C (types, fonctions et passage de paramètres, tableaux statiques et dynamiques) puis la conception et réalisation de projet (organisation du code, utilisation de bibliothèques). J'ai réalisé une très grande parties des différentes versions des sujets de TD de ce module.
- ▶ **Projet de Conception Logiciel en Analyse NUMérique (CLANU)**, 3^e année département Génie Électrique, Projet promotion entière 120 étudiants
Il s'agit d'un projet dans la suite de la formation au langage C. Ce projet est co-animé par les enseignants de mathématiques et d'informatique du département. Les étudiants ont à concevoir et réaliser un programme informatique autour de l'analyse numérique : Transformée de Fourier, équation de conduction de la chaleur, *in-painting*, transformée de Radon, régression et classifications par réseau de neurones... Mon implication se situe dans la définition du projet et plus spécifiquement sur la partie informatique (réalisation du code de départ), puis dans le suivi et l'évaluation de la partie informatique.
- ▶ **Micro-contrôleur**, 3^e année département Génie Électrique et filière DUT+3 (fin en 2012), 10 étudiants, $15\text{hCM} + 8 \times 1\text{hTD} + 2 \times 3\text{hTP}$
Il s'agit de transmettre aux étudiants le fonctionnement et l'utilisation des micro-contrôleurs. Nous nous appuyons sur une plateforme Microchip PIC16F. J'ai réalisé

le cours pour les DUT+3 ainsi que les TD pour les deux filières qui, pour la filière classique, ont évolué ces dernières années. Depuis 2017, je suis chargé d'introduire, en 5h de cours, l'utilisation du langage C sur des plateformes à base de micro-contrôleurs (MSP430, STM32, RaspberryPi, Arietta G25, ...) et les interruptions (sans système d'exploitation). J'ai aussi piloté l'étude de l'évolution des plateformes à base de micro-contrôleurs pour les 3 ans de formation au département.

- ▶ **Modélisation Orientée Objets et Programmation en C++**, 4^e année département Génie Électrique, filière classique et apprentissage, 120 étudiants par promotion, $4 \times 2\text{hCM} + 6 \times 2\text{hTD}$

Ces 20h de face à face doivent permettre aux étudiants de connaître les quatre principaux paradigmes de l'orienté objet et de les maîtriser pour modéliser des systèmes en UML et les programmer en C++. J'ai réalisé le cours et les différentes versions des séances de TD. Depuis 2014, j'utilise, pendant les cours, une plateforme de questionnaire³ disponible sur appareils mobiles ou PC. En plus des effets stimulants, cela me permet par exemple de voir rapidement les concepts mal acquis lors du questionnaire de début de cours et ainsi faire les rappels nécessaires pour aborder plus sereinement les éléments du cours à venir.

- ▶ **Digital Signal Processor FPGA/DSP/GPU**, 5^e année département Génie Électrique, options Traitement des signaux et des Images et Systèmes Embarqués Communicants, 40 étudiants, $5 \times 2\text{hCM} + 4 \times 4\text{hTP}$

L'objectif est de présenter les optimisations d'architectures des processeurs pour réaliser efficacement les opérations élémentaires du traitement du signal et de l'information. Quatre TP de 4h permettent aux étudiants de manipuler un DSP et un FPGA dans le cadre du traitement du signal (filtrages IIR et FIR, MACD, accès mémoires et pipeline) et un GPU dans le cadre du calcul en concurrence (collaboration avec l'ENS Lyon et Emmanuel Quemener). J'ai repris le cours de DSP et inclus une partie sur les nouvelles infrastructures de calculs (GPU, cluster et *cloud computing...*) ainsi que les TP DSP.

- ▶ **Plateforme OMAP**, 5^e année département Génie Électrique, 100 étudiants par promotion, 4hTP

Grâce à un Bonus Qualité Formation INSA de 9k€, j'ai monté en 2011 un TP de 4h sur l'architecture OMAP de Texas Instruments permettant ainsi d'étudier la programmation d'un ARM et d'un DSP (initiation à l'architecture des DSP et SOC⁴, comparaison des performances des architectures, utilisation de codec vidéo, développement sans disque) via un linux embarqué.

- ▶ **Multi-tache et calcul en concurrence**, 5^e année département Génie Électrique, 120 étudiants par promotion, $2 \times 2\text{hTD}$

Les séances de TD permettent aux étudiants de se confronter aux mécanismes intellectuels de programmation concurrente (du *fork* à OpenMP en passant par des "*pipe*") présentés en cours. J'ai rédigé ces TD et proposé une séance de 2h d'introduction à l'utilisation d'OpenMP.

4.1.2 Traitement du signal et des images

Traitement du signal

- ▶ **Analyse Spectrale et filtrage numérique**, 4^e année département Génie Électrique, 120 étudiants par promotion, $2 \times 4\text{hTP}$

Ces deux TP ont pour objectif de faire maîtriser les matériels et leurs utilisations

3. Socrative

4. System On Chip

afin que les étudiants puissent comprendre et s'approprier les concepts théoriques vus en cours et TD : repliement de spectre, échantillonnage temporel et périodisation spectrale, modulations, filtrages numériques. En plus des activités d'animations, remise en route, petites modifications et rédaction/correction des évaluations, j'ai formé régulièrement des vacataires et ATER sur ces TP. Depuis 2021, j'ai repris le cours (6h) et les TD (4h) de filtrages analogiques et numériques.

- ▶ **Théorie de l'information**, 5^e année département Génie Électrique, option TdSI (Traitement des Signaux et des Images) et SEC (Systèmes Embarqués Communicants), 40 étudiants, $4 \times 2\text{hCM} + 4 \times 2\text{hTD}$

L'objectif est de faire découvrir la théorie de l'information, autre grande contribution de Claude Shannon et ses applications. Les étudiants apprennent les fondamentaux des codages de source et de canal (entropie, capacité) puis à manipuler les méthodes classiques de codage (Huffmann, algorithmes LZW, codage de Hamming, codage-décodage convolutif, ...). Enfin, ce cours est aussi le prétexte d'une introduction sur la sécurité (cryptage, connexion sécurisée, faille, ...). J'ai repris cette formation en ajoutant le turbo codage et les enjeux de sécurité pour les systèmes embarqués.

Traitement et des images

- ▶ **Traitement et analyse d'images**, 5^e année Génie Électrique, option TdSI (15 à 20 étudiants) et IMESI (5 étudiants), $8 \times 2\text{hTD}$

Il s'agit d'une introduction aux traitements des images : de la représentation numérique des images (dont les espaces couleurs) et leur acquisition, aux filtrages (linéaires ou non, morphologiques) et les transformations (Fourier, Hough, et géométriques) puis l'extraction de caractéristiques macroscopiques (textures). Une seconde partie se consacre à la segmentation des images (*level sets*, croissance de régions, bassin versant, k-means, ...).

- ▶ **Traitements de l'image - approfondissements**, parcours SI Master 2, 15 étudiants, $2 \times 10\text{hCM}$

L'objectif de ce cours de Master est de donner les bases des méthodes récentes en classification supervisée et non-supervisée, les descripteurs locaux, l'imagerie 3D et l'évaluation quantitative. Je me charge de la partie "clustering" avec les espaces de caractéristiques et les échelles, avec comme support l'approche *mean-shift*. L'idée principale de ce module est de fournir une formation allant de la théorie à la pratique (utilisation et codage des approches). Depuis 2019, je suis chargé de la formation théorique en filtrage, segmentation et en *apprentissage profond* que je complète par 2 TP (classification de données médicales par réseau profond ; segmentation par croissance de régions et par réseau U-Net).

4.1.3 Tutorats au département Génie Electrique

- ▶ **PFE/Stagiaires**. Par an, j'encadre 4 à 5 étudiants en projet de fin d'étude (5^e année, durée de 5 à 6 mois) et de 2 à 4 étudiants en stage (4^e année, durée 6 mois). Ces tutorats sont très enrichissants pour la culture industrielle et observer l'évolution des étudiants. Ils permettent aussi d'établir des liens avec les industriels qui débouchent parfois sur des collaborations ou des partenariats (voir section 5.6).
- ▶ **Apprentis**. Depuis 2012, j'ai suivi 7 apprentis, dont un étudiant en formation continue. Le suivi sur les trois ans de formation avec des visites et rencontres régulières avec l'étudiant et le tuteur industriel permet de construire une relation tripartite intéressante mais qui nécessite cependant une vigilance objective sur l'évolution et la progression des attendus au cours des trois années de formation.

- **Projets d'option TdSI et SEC**, 5^e année en Traitement des signaux et des Images et Systèmes Embarqués Communicants. Chaque année, je propose 3 à 4 projets se voulant originaux dans les contextes des communications, des systèmes embarqués et du traitement d'images à destination d'étudiants en 5^e année. Notamment : localisation par qualité de signal de modules Zigbee, codage LDPC, étude des noyaux linux temps réel, utilisation de HackRF One, détection de visages, correction d'artéfacts en IRM, création de version en réalité virtuelle de TP de GE, comptage de personnes dans différents contextes urbains, segmentation de muscles en IRM... Certains de ces projets ont été supportés par des partenariats industriels (Gambro, Bioclinica, EFi automotive, Alstom, Biomae), d'autres ont contribué à l'avancement de projets de recherche et d'autres ont permis l'élaboration de TP au département.
- **Projet de Réalisation Technologique**, 4^e année. Ces projets sont effectués pendant un semestre par deux étudiants de 4^e année. J'ai proposé de nombreux projets notamment autour de la musique, d'un quadri-coptère et de l'apprentissage profond notamment pour la reconnaissance automatique de 10 mots du domaine « Génie Electrique » (*i.e.* bobine, condensateur, ...) en langage des signes (figure 4.2) ou dans la parole.
- **Projet HuTech**, 3^e année, 1 projet par an. Ces projets de "humanité et technique" sont réalisés par un groupe d'une dizaine d'étudiants sur toute l'année. Il s'agit de simuler la création d'une entreprise avec un produit innovant. Depuis deux ans, ces projets doivent répondre, en plus des contraintes standards, aux enjeux de Développement Durable et de Responsabilité Sociétale. Pour ma part, il s'agit d'accompagner les étudiants sur la démarche scientifique et technique, les parties "entreprise" (analyse fonctionnelle, ...) et "DDRS" (analyse du cycle de vie, bilan carbone, ...) étant assurées par des intervenants du centre des Humanités et des professionnels.

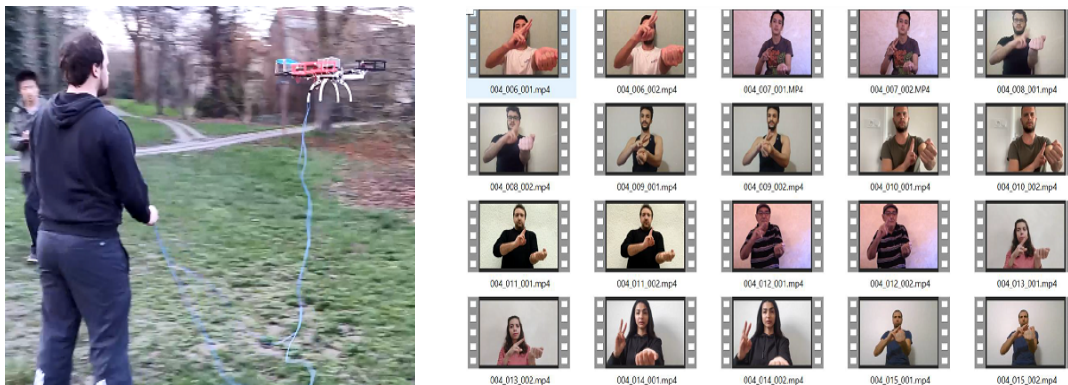


FIGURE 4.2 – Exemples de deux Projets de Réalisation Technologique : quadri-coptère à gauche (développement sur base ATMEL) et sur la reconnaissance du langage des signes avec un extrait de la base de données des signes GE.

4.2 Responsabilités

Cette partie précise les principales responsabilités pédagogiques que j'ai assurées au sein de l'INSA Lyon. Pour beaucoup d'entre elles, je tiens à remercier la direction du département Génie Electrique (Claude Richard, Tanneguy Redarce et plus récemment Lionel Petit) pour la confiance qu'ils m'ont accordée pour conduire ces différentes actions.

4.2.1 Responsable de la formation IMESI

J'ai été le responsable de la formation (*International Master of Embedded Systems and Medical Image Engineering*) de 2007 (création) à 2014 (sortie de la dernière promotion).

Il s'agissait d'un parcours de formation intégralement en anglais, équivalent M1 et permettant à des étudiants étrangers, de niveau Bachelor et ayant un projet professionnel orienté recherche, de se préparer à un Master 2 recherche en Systèmes et Images ou Génie des Systèmes Automatisés ou Systèmes Electroniques Intégrés.

Cette responsabilité comprenait la création du parcours et le recrutement des candidats (avec l'appui de la direction Internationale de l'INSA Lyon puis de CPE), le rapprochement de partenaires (formation : INSA Lyon Génie Électrique et CPE ; recherche : INL, Ampère, DISP, CREATIS), l'animation pédagogique (60 ECTS, équipes pédagogiques, jurys, projets, stages, ...) et l'accompagnement des étudiants (4 par promotion, cause principale de l'arrêt de cette formation).

Pour assurer cette mission, je pouvais compter sur Emil Dumitrescu, co-responsable IMESI, puis Renaud Daviot co-responsable coté CPE, et Christelle Coche secrétaire (1/4 ETP) ainsi que sur Eric NIEL, responsable des formations « *International Master* » de l'INSA Lyon. DUPUY et al., 2011.

La construction et direction de cette formation internationale, en anglais et avec des intervenants de différents laboratoires et écoles (INSA Lyon et CPE Lyon) a été particulièrement enrichissante. Pour ma part, le point le plus difficile et complexe a été le recrutement à l'international, et ce malgré les différents réseaux sollicités (privés et publics) et malgré les appuis des services de communication des écoles et de leurs collaborations internationales.

4.2.2 Chargé de la communication au département Génie Électrique

J'ai assuré cette mission de 2009 à 2013.

Ayant identifié des points faibles au département, j'ai proposé à la direction de définir une nouvelle mission de chargé de communication (interne et externe) pour laquelle je me suis porté volontaire.

Les objectifs : améliorer la communication en interne (personnels, étudiants) et accroître la visibilité du département à l'extérieur (candidats, entreprises) ainsi qu'augmenter la diffusion et l'utilisation d'outils numériques pour la pédagogie, l'administration, le secrétariat.

Tous les supports de communications (site Web, plaquettes de formations, posters, présentations du département lors de forums) ont été refaits en respectant une charte commune. Plusieurs rapprochements entre équipes pédagogiques ou direction et les étudiants ont été faits pour faciliter les échanges et des réflexions de changements profonds sur la formation (implication d'anciens diplômés, équivalence de crédits ECTS entre modules et activités associatives reconnues et suivies, organisation de la scolarité en 4 et 5^e années, illustrations du site web, identification et discussion de points forts et faibles du département, ...).

J'ai mis en place un système de gestion de projet et de gestion de version des documents de communication (Redmine avec GIT, ainsi qu'une charte pour les noms de fichiers), la création d'une lettre d'information impliquant la direction, les étudiants et les services administratifs (GEco', 5 numéros).

J'ai aussi contribué à la diffusion et à l'utilisation des outils pédagogiques (moodle, environnements informatiques en lien avec la DSI) au sein du département (missions toujours en cours).

Enfin, cette mission incluait la participation au conseil de direction du département (4 personnes autour du directeur) dont les réunions hebdomadaires consistaient à partager les informations et établir des positions et stratégies sur les éléments de la vie du département

(budget, taxe apprentissage, travaux, évolution de la formation et de la pédagogie, création filière apprentissage, cours en anglais, campagne emploi, ...).

4.2.3 Application stage

De 2015 à 2019, un projet important fut celui de créer une base de données des stages, entreprises, RH et ingénieurs dans les buts d'améliorer les relations entreprises, établir des statistiques précises demandées par les différents organismes et directions, mieux cibler notre campagne de recherche de taxe d'apprentissage.

Une demande importante du responsable des stages et PFE pour cette application était qu'elle automatise la génération des conventions de stages : les étudiants devenaient les acteurs principaux des démarches, la secrétaire assurant un rôle de supervision et de validation des données avec l'aide des tuteurs académiques. J'ai piloté la réalisation de ce projet, de la formalisation du cahier des charges à son développement informatique et à sa mise en production (et la correction de bugs). Deux versions ont été réalisées par de trois stagiaires d'IUT en informatique, en étroite collaboration avec la direction du département, le responsable des stages (notamment Philippe Delachartre), le secrétariat du département et la DSI de l'INSA Lyon. Ces versions ont été utilisées 4 semestres (2017 et 2018) et ont permis qu'une nouvelle version soit développée par un prestataire externe. Cette dernière version est aujourd'hui utilisée pour l'ensemble des stages et PFE du département (plus de 200 conventions par an).

4.2.4 Jouvence informatique

De 2010 à 2017, j'ai assuré pour le département Génie Electrique la maîtrise d'ouvrage de la réhabilitation de 4 salles de formation (salles informatiques et espaces de formation pour le Labex PRIMES) et de l'infrastructure informatique du département en appui aux services DSI et DirPAT (direction du patrimoine) de l'INSA Lyon. Cette responsabilité a fait suite à différents problèmes rencontrés lors de travaux effectués au département : coûts exagérés, délais non tenus et non respect des demandes formulées. Mon implication a porté sur la recherche de financements (plus de 600kEuros), l'étude de solutions afin de garantir une adéquation avec les besoins du département (accès distants et BYOD⁵) et les politiques et stratégies de l'INSA, les rencontres avec les différents corps de métiers et les choix et arbitrages, l'estimation des coûts et collecte/négociation des devis, la planification des travaux et leurs suivis ainsi qu'une partie de la réception. Les projets couvraient les lots de réhabilitation des locaux (dont gros oeuvre), l'infrastructure (courants forts et faibles), le mobilier et les machines. Les travaux ce sont principalement déroulés pendant les étés 2013, 2014 et 2015. Ils ont été menés à bien en temps et en heure aussi grâce à l'investissement de Lucas Hiltgun, technicien au département Génie Electrique notamment pour le suivi en courant fort et la réalisation des plans des salles qui était un enjeu majeur pour respecter les objectifs d'effectifs et les contraintes de normes et d'accès. Ceci a permis de définir un type de salle TP et un type de salle TD avec leur mobilier ergonomique sur mesure, respectant les contraintes d'occupation et les normes (dont PMR). Ces deux modèles équiperont aujourd'hui 9 salles du département.

Coté logiciel et matériel (PC), encore aujourd'hui, je suis référent auprès de la DSI pour le département. Avec l'appuie de la DSI de l'INSA, cette infrastructure et politique d'accès distants ont rendus possible de généraliser en quelques jours l'utilisation à distance des logiciels métiers du département pour les étudiants et enseignants, permettant ainsi d'assurer une disponibilité des outils dès le début du premier confinement.

5. *Bring Your Own Device*



FIGURE 4.3 – Exemple de jouvence : salle informatique E104 du département Génie Electrique de l'INSA Lyon en septembre 2013 avril (à gauche) puis en septembre 2014 (à droite).

4.2.5 Responsable de l'option Traitement des Signaux et des Images (TdSI)

TdSI est une option en 5^e année de formation en traitement du signal et des images au département GE de 12 ECTS. Elle est suivie par une vingtaine d'étudiants.

Pour cette responsabilité de quatre ans (de 2016 à 2020, en biseau avec mon successeur la dernière année), j'ai réorganisé le programme (200h de face à face), de l'animation et organisation de l'équipe pédagogique (5 enseignants chercheurs) et des intervenants (entre 8 et 12 personnes académiques et industriels) de manière à faire face aux technologies récentes dans cette thématique (deep learning, GPU, réalité virtuelle et augmentée, RGB-D) et au renouvellement de l'équipe pédagogique.

Les principaux changements de contenu des modules ont été la création d'un projet supplémentaire axé sur l'ingénierie des technologies, l'ajout de formations en imagerie X et ultrason, l'ajout de la composante apprentissage automatique et « réalités virtuelle et augmentée » (organisation d'une journée à Châlon-sur-Saône au Le2i puis visite d'une entreprise spécialisée et du musée Nicéphore Niépce)

4.2.6 Membre élu au conseil de département Génie Electrique

Elu depuis mai 2015, je participe aux conseils de département mensuels.

Je me suis impliqué dans plusieurs réflexions notamment le suivi du projet professionnel des étudiants (définir un échéancier de rencontres en phase avec les choix académiques importants des étudiants), le règlement des études du département (qui est maintenant une annexe au règlement commun des études à l'INSA Lyon) et l'utilisation de la plateforme pédagogique moodle pour centraliser les contenus et diversifier les pratiques pédagogiques puis faciliter la gestion de la scolarité (choix d'options de 5^e année par exemple, rendus de travaux, évaluations).

4.2.7 Responsable du module "Traitement des Images - Approfondissement"

Ce module de 20h de face à face est une formation du Master *Medical Imaging Signal and System* (MISS), parcours Systèmes et Images (depuis 2007). Définition des contenus, recherche des intervenants (académiques et industriels), animation de l'équipe pédagogique et intervenants ont été mes principales activités pour ce module dans lequel j'assurai la formation en filtrage. Depuis 2019, nous avons fortement fait évoluer le contenu de manière à intégrer la *machine learning* et pour ma part le *deep learning* et à renforcer les fondamentaux en filtrage, segmentation et recalage.

4.3 Bilan

Dans ce chapitre, j'ai présenté mes activités d'enseignements et les responsabilités pédagogiques que j'ai exercées depuis 2006. Elles occupent une moitié de mon temps. Même si certaines ont été autant difficiles et stressantes qu'éloignées des missions initiales d'un enseignant-chercheur, dans la réalité, pour avancer au sens large de la formation, il fallait les assurer et les conduire à termes. Elles sont surtout très enrichissantes et stimulantes sur les aspects de renouvellement intellectuel, de gestion des relations humaines, de capacités de communication et de pédagogie puis organisationnel. Le chapitre suivant présente l'autre moitié des mes activités d'enseignant chercheur : mes activités en recherche.



5. Activités de recherche

Mon activité de recherche est centrée sur le traitement et l'analyse d'images médicales pour l'étude de pathologies. Plus précisément, je contribue dans ce domaine aux méthodes de filtrage, segmentation et clustering de données multi paramétriques spatio-temporelles. La figure 5.1 illustre les principaux travaux de recherche conduits depuis 2010.

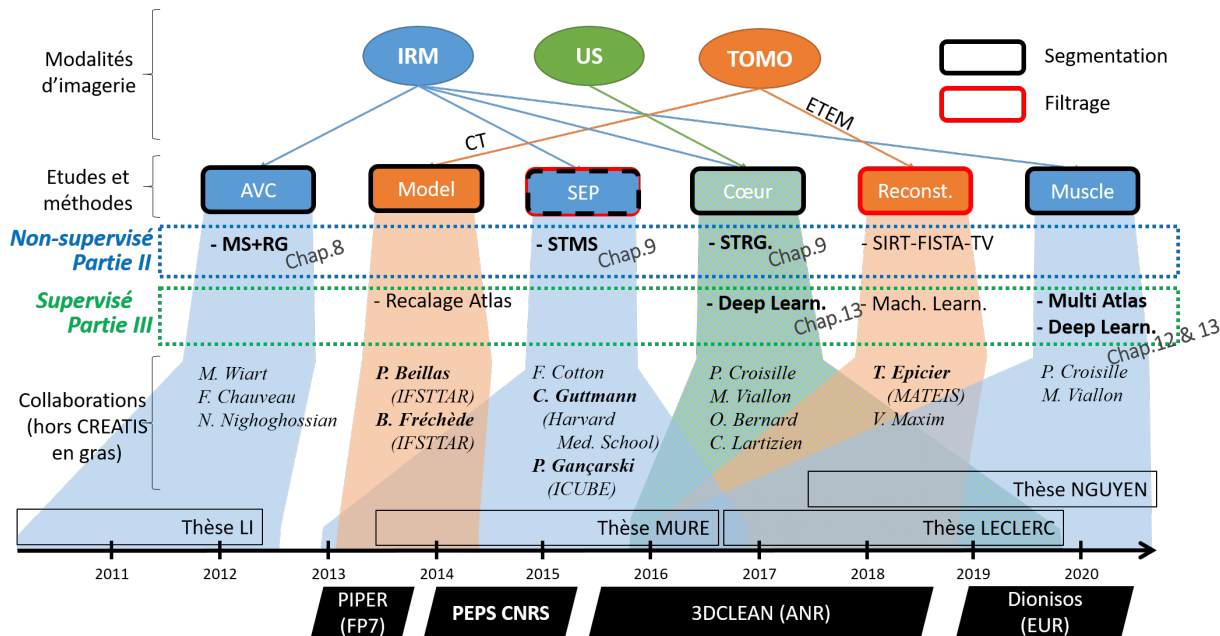


FIGURE 5.1 – Synthèse des activités de recherche de 2010 à 2020 en filtrage et segmentation. Abréviations : **MS+RG** : mean-shift avec croissance de région ; **STMS** : mean-shift spatio-temporel ; **STRG** : croissance de région spatio-temporelle ; **Deep Learn.** : apprentissage profond

Mes activités de recherche sont résumées ci-après, suivies des encadrements effectués, des principales collaborations et enfin des différentes responsabilités exercées.

5.1 Résumé des travaux

La description de mes travaux post doctoraux est organisée en quatre périodes chronologiques pendant lesquelles j'ai codirigé quatre doctorants et co-encadré deux doctorants, été impliqué sur les travaux de thèse de deux doctorants, et encadré environ 25 élèves ingénieurs ou étudiants de master.

Période 2001-2008

Pendant ma thèse, dans le cadre de l'étude de tumeurs osseuses, je me suis intéressé à la segmentation automatique par croissance de régions en imagerie TEP (Tomographie à Emission de positons) au FDG (18F-FDG : fluorodesoxyglucose marqué au fluor 18) et NaF (18F-NaF : fluorure de sodium marqué au fluor 18), traceurs qui permettent de visualiser respectivement l'activité métabolique du sucre et l'activité fonctionnelle du squelette. En imagerie TEP, les volumes 3D sont relativement bruités, les contours ne sont pas nets et les structures à segmenter peu homogènes. Afin de rendre la segmentation par croissance de régions plus automatique et afin d'améliorer les performances, je me suis tourné vers le filtrage *mean-shift* et l'analyse de ses paramètres d'échelle ainsi que de l'espace des caractéristiques spatial-amplitude utilisés pour représenter et traiter les données (GRENIER et al., 2005a). J'ai proposé dans ce cadre un formalisme de l'approche de croissance de région permettant de généraliser cette méthode de segmentation aux différentes spécificités observées sur les images TEP et aux données multiparamétriques, notamment en imagerie par ultrasons. Ainsi, il est possible de faire des liens entre clustering et segmentation et on pouvait espérer compléter le formalisme pour exprimer le critère optimisé par les approches de croissance de régions. Ce travail a été conduit dans le cadre de la thèse de Jean-Loïc Rose (soutenue en 2008) et a aussi conduit à l'ajout d'a priori géométrique (ROSE et al., 2010).

Période 2008-2012

Un autre élément important est l'optimisation des paramètres d'échelles utilisés pour régler le fonctionnement des approches de filtrage *mean-shift* (LI et al., 2011) et de segmentation par croissance de région (REVOL-MULLER et al., 2012a). La compréhension fine des espaces utilisés et la normalisation effectuée sur les données ont été au cœur de la thèse de Ting Li (soutenue en 2012) avec comme application médicale, la prédiction de zones nécrosées suite à un Accident Vasculaire Cérébral (AVC) à partir de données multiparamétriques IRM. D'après les modèles biologiques, cette zone, dite de pénombre, peut se prédire à partir des images IRM acquises aux premiers instants suivants la survenue de l'AVC. Malgré de très bonnes performances sur des rats avec des AVC provoqués, l'approche a donné des résultats mitigés sur l'homme. L'impact de l'instant de l'imagerie par rapport à la survenue de l'AVC, du lieu de l'AVC et de la variabilité des patients expliquent ces résultats. D'un point de vue méthodologique, l'apport de l'information temporelle dans ce type d'approche serait d'un enjeu capital et original que nous avons étudié par la suite.

Période 2012-2016

Comprendre l'évolution des caractéristiques dans l'espace joint spatial-amplitude, puis la formulation d'une méthode de clustering permettant de prendre en considération ces évolutions temporelles, ont guidé mes travaux pendant cette période. Nous avons ainsi étendu le formalisme *mean-shift* aux séries temporelles et introduit la notion de voisins temporels (MURE et al., 2015b). Cette approche de filtrage (STMS) a été appliquée avec succès à l'analyse de l'évolution des lésions de Sclérose En Plaques (SEP) au cours d'études longitudinales en IRM. Elle permet notamment d'obtenir une segmentation fiable des tissus évoluant de même manière et a été appliquée avec succès à d'autres problématiques

d'évolutions (thèse Simon Mure soutenue en 2016). Cependant, cette approche ajoute un nouveau paramètre d'échelle pour l'aspect temporel et de nouvelles contraintes sur les données pour garantir un fonctionnement robuste sur une base de données.

Pendant cette période, une problématique de segmentation et d'étiquetage automatique de tous les os sur des scanners X corps entier, pour réaliser des modèles anatomiques en vue de simulations, m'a conduit à utiliser un recalage d'atlas qui, contrairement aux méthodes précédentes, est une méthode supervisée. En effet, ce problème était purement insoluble avec les approches non-supervisées développées dont les *mean-shift* et la croissance de région.

Période 2016-2021

Pendant cette période, les problèmes de segmentation rencontrés ont conduit à l'utilisation quasi-systématique de méthodes supervisées, soit via des recalages d'atlas soit par apprentissage machine –en plein essor sur cette période– puis à des contributions sur ces approches (LECLERC et al., 2019c ; NGUYEN et al., 2019b). L'enjeu principal est d'obtenir des segmentations suffisamment précises et robustes pour permettre des études longitudinales (comme l'inflammation musculaire en IRM) ou temporelles (séquences échocardiographiques ou d'IRM cardiaque). Nos approches d'augmentation de données veillant à la diversité morphologique (thèse de Hoai-Thu Nguyen, soutenue le 15 octobre 2021) et différentes contributions sur les réseaux de neurones basés sur l'architecture U-Net (thèse Sarah Leclerc, soutenue en 2019) permettent d'obtenir des performances souvent suffisantes pour les études médicales.

5.2 Participation à l'encadrement de post-doctorants

J'ai participé à l'encadrement de 2 post-doctorants.

Hussein Banjak, 24 mois (2017 - 2019)

Actuellement en CDI chez Intom GMBH.

Dans le cadre du projet ANR *3DCLEAN*, le travail portait sur l'amélioration des méthodes de reconstruction de tomographie en microscopie électronique environnementale ultra rapide. (BANJAK et al., 2018 ; EPICIER et al., 2018 ; EPICIER et al., 2019 ; KONETI et al., 2019). Co-encadré avec Voichita Maxim (Creatis).

Khuram Faraz, 18 mois (2019-2021)

Dans le cadre du projet EUR SLEIGHT-MANUTECH *Dionisos*, le projet porte sur le suivi de nanoparticules en microscopie électronique environnementale. Khuram Faraz¹ est co-encadré avec Christophe Ducottet (LaHC) et Thierry Epicier (Mateis, IRCeLyon). (FARAZ et al., 2021 ; FARAZ et al., 2022)

5.3 Participation à l'encadrement et à la codirection de doctorants

Jean-Loïc Rose

« *Croissance de région variationnelle et contraintes géométriques tridimensionnelles pour la segmentation d'images* »

Thèse soutenue le 5 décembre 2008. Implication sur les travaux de thèse (direction Chantal Revol-Muller et Christophe Odet). Actuellement responsable recherche chez ORTEN Lyon. Une partie de ce travail était dans la continuité de mes travaux de thèse. Nous avons entre autre pu formuler le processus de croissance de région avec les notations ensemblistes sous

1. cet étudiant est un ancien étudiant IMESI

forme variationnelle en s'appuyant sur le formalisme des *level-sets*. (REVOL-MULLER et al., 2012b; REVOL-MULLER et al., 2013; ROSE et al., 2010).

Delphine Charpigny

« *Quantification des nanoparticules à base d'oxyde de fer pour l'IRM Approche basée sur la déconvolution du défaut de champ magnétique* »

Thèse soutenue le 11 janvier 2011. Implication sur les travaux de thèse (direction Hugues Benoit-Cattin). Actuellement experte chez Ayming Lyon.

Pour le problème de la dé-convolution dans ce contexte, j'ai identifié deux approches de la littérature permettant de résoudre le problème directement dans l'espace de Fourier et qui ont permis de résoudre ce point bloquant (dé-convolution régularisée par moyenne des erreurs quadratique (Wiener) et contraint par moindres carrées (CLS)). (CHARPIGNY et al., 2009; CHARPIGNY et al., 2010; CHARPIGNY et al., 2008).

Ting Li

« *Contributions to Mean Shift filtering and segmentation, Applications to MRI ischemic data* »

Thèse soutenue le 4 avril 2012. Financement China Scholarship Council (CSC). Codirection (50%) avec Hugues Benoit-Cattin. Actuellement chez ShangHai United-imaging Healthcare Co., Ltd, China (développement de robots pour le médical).

Ces travaux étudient l'optimisation des paramètres d'échelle pour le filtrage *mean-shift* et la segmentation basée sur la croissance de région pour des données multi-paramétriques (principalement sur des images naturelles couleurs et sur des données issues de différentes séquences d'IRM). Ces paramètres sont critiques pour la qualité du résultat produit par l'une ou l'autre de ces approches ou un couplage des deux. Ces travaux établissent des liens sur l'organisation des données dans l'espace des caractéristiques lors de l'ajout de flou et de bruit afin de comprendre les limitations des approches et le choix des paramètres d'échelle. Une optimisation exhaustive des paramètres d'échelle pour chacune des approches et leur couplage a été conduite et les rapprochements avec l'estimation optimale proposée dans le contexte statistique et la normalisation des données ont été discutés. Ces stratégies ont été appliquées pour segmenter les différentes régions d'un AVC afin de prédire l'évolution de la région nécrosée, principal facteur de l'évolution du handicap et du choix du traitement pharmaceutique. (LI et al., 2010a; LI et al., 2011; REVOL-MULLER et al., 2012a).

Simon Mure

« *Classification non supervisée de données spatio-temporelles multidimensionnelles. Applications à l'imagerie* »

Thèse soutenue le 2 décembre 2016, financement MENRT école doctorale EEA. Codirection (50%) avec Hugues Benoit-Cattin. Actuellement ingénieur chez Dessintey Saint-Etienne.

Il s'agissait d'apporter une formulation à l'approche *mean-shift* pour prendre en compte l'information disponible dans des séquences d'images pour du filtrage et du clustering. L'application qui a guidé cette étude est l'analyse de l'évolution de lésions de sclérose en plaques observées par imagerie IRM. Cette méthodologie s'est révélée efficace pour de nombreux autres problèmes d'analyse d'évolution (imagerie satellitaire, imagerie spectrale optique). (AMELI et al., 2016; MURE et al., 2016a; MURE et al., 2015a; MURE et al., 2016b; MURE et al., 2016c; MURE et al., 2016d; MURE et al., 2015b; PORTEJOIE et al., 2015), (DOLET et al., 2016; DOLET et al., 2018).

Sarah Leclerc

« *Automatisation de la segmentation sémantique de structures cardiaques en imagerie ultrasonore par apprentissage supervisé* »

Thèse soutenue le 11 décembre 2019, financement Labex PRIMES (début octobre 2016). Co-encadrement 30% (codirecteurs Olivier Bernard et Carole Lartizien). Actuellement MCU 63^e à Dijon.

Ces travaux se focalisent sur la segmentation supervisée de 2 instants de la séquence cardiaque en imagerie par ultrasons. Ceci a été rendu possible par la construction d'une base de données (CAMUS) d'images et d'expertises sur les différentes régions du coeur. Les approches étudiées ont été les forêts aléatoires et surtout les réseaux de neurones au travers de l'architecture UNET qui a révolutionné la qualité des segmentations obtenues sur ces images. L'optimisation de l'architecture UNET ainsi que des contributions originales sur cette architecture constituent les principaux développements de Sarah Leclerc. (LECLERC et al., 2019a; LECLERC et al., 2018; LECLERC et al., 2019b; LECLERC et al., 2019c).

Hoi-Thu Nguyen

« *Méthodes semi-supervisées pour l'étude de la variation fonctionnelle à partir de données d'imagerie médicale multiparamétriques et longitudinales* »

Thèse soutenue le 15 octobre 2021, financement MENRT école doctorale SIS. Codirection (50%) avec Pierre Croisille.

Les contributions consistent à élaborer une approche de segmentation 3D capable de segmenter les muscles (jambes, épaules) en IRM, et ce même quand un faible nombre d'annotations est disponible. L'approche de segmentation par multi-atlas a été étudiée et plusieurs contributions lui ont été apportées notamment le remplacement de l'étape de *fusion jointe des étiquettes* par des réseaux U-Net et des stratégies de sélection d'atlas basée sur une mesure morphologique originale. (JOUVENCEL et al., 2022; NGUYEN et al., 2018a; NGUYEN et al., 2019a; NGUYEN et al., 2019b; NGUYEN et al., 2021a; NGUYEN et al., 2019c; NGUYEN et al., 2019d; NGUYEN et al., 2021b; NGUYEN et al., 2018b).

Valentine Wargnier-Dauchelle

« *Approches d'apprentissage profond pour la détection en IRM de lésions de sclérose en plaques actives au gadolinium, sans injecter de gadolinium* »

En quatrième année de thèse, financement MENRT école doctorale EDISS. Participation à l'encadrement de thèse (codirecteurs Michael Sdika et François Cotton).

Cette thèse vise à développer des méthodes basées sur l'apprentissage profond permettant de détecter les lésions SEP actives au Gadolinium en évitant d'injecter ce produit controversé au patient. Une collaboration avec OFSEP², nous permet de disposer d'environ 1000 patients respectant un protocole d'imagerie parfaitement adapté à cette étude. Cependant, cette base de données n'est pas annotée et les travaux sont orientés vers les méthodes de segmentation faiblement supervisée. Pour plus de robustesse, une étude sur l'explicabilité des réseaux est nécessaire afin de s'assurer de la cohérence des résultats avec le fonctionnement interne du réseau. (WARGNIER-DAUCHELLE et al., 2021a; WARGNIER-DAUCHELLE et al., 2021b)

2. Observatoire Français de la Sclérose en Plaques : www.ofsep.org

Enyi Chen

« Développement des techniques d'analyse de la connectivité morphologique pour l'étude des phénomènes neuro-dégénératifs dans le vieillissement normal et dans la sclérose en plaques »

Thèse commencée le 25 octobre 2021, financement Labex PRIMES *medical board*. Codirection (40%) avec Dominique Sappey-Marinier.

L'objectif principal de ce projet est de caractériser les phénomènes de neuro-dégénérescence dans le vieillissement normal et dans les différentes formes cliniques de SEP par la mesure de la connectivité morphologique par IRM conventionnelle, et l'analyse des graphes en corrélation avec les paramètres cliniques des sujets sains (âge, QI, personnalité) et des patients SEP (forme clinique, charge lésionnelle, EDSS, MSFC, ...). Une nouvelle collaboration avec OFSEP nous permet de disposer du suivi de plus de 1000 patients pour lesquels les informations sur la forme de la pathologie et le score d'EDSS sont présents.

Emile Saillard

« Approches d'apprentissage profond pour l'estimation du risque de fracture par simulation numérique de vertèbres métastasées à partir d'images CT *in vivo* »

Thèse commencée le 1 octobre 2022, financement école doctorale MEGA. Codirection (50%) avec Hélène Follet (Lyos/LBMC). Collaboration avec les Hospices Civils de Lyon.

Ces travaux de thèse visent à développer une approche automatique de segmentation permettant la prédiction, par simulation du numérique, du risque de fracture des fémurs et des vertèbres. Les contributions en segmentations, basées deep learning, seront dans l'introduction de contraintes propres aux images, à l'application ainsi qu'aux maillages utilisés pour la simulation numérique. Il s'agira aussi d'être capable de détecter et de segmenter les métastases osseuses afin de parfaire les simulations numériques. Une contribution en augmentation de données est aussi attendue afin d'étudier la sensibilité des approches et accroître la variabilité des images d'entraînement. (SAILLARD et al., 2021)

5.4 Encadrements d'étudiants en master et en projet de fin d'études

De 2003 à 2022, j'ai encadré 28 étudiants en stage de master, stage ingénieur ou projet de fin d'études :

- ▶ 2003 - Ilias AHRAZEM, stage master recherche (DEA). « Etude des stratégies de recalage rigide et affine de données TEP 3D avec la bibliothèque ITK ».
- ▶ 2004 - Aïssata MAÏGA ABDOULAYE, stage master recherche (DEA). « Segmentation par croissance de régions, étude de différentes approches. Implémentation et optimisation de l'approche généralisée de croissance de régions pour les données TEP ».
- ▶ 2004 - Tomas LAFUENTE-HERNANDEZ, stage master recherche (DEA). « Recalage affine et local en imagerie TEP optimisé par la méthode DIRECT ».
- ▶ 2006 - Inma SORLI et Franceso VARRESE, stage ingénieur. « Plateforme logicielle générique dédiée au développement de méthodes de segmentation basées sur la croissance de régions ».
- ▶ 2009 - Pierre JEGOUZO, projet de fin d'études. « Recalage et clustering de données IRM pour le diagnostic de la sclérose en plaques ».
- ▶ 2009 - Matthieu LAUZIER, stage master recherche. « Quantification of ischemic brain damage with the *mean-shift* clustering method ».

- ▶ 2009 - Leow JIAMIN, Stage d'échange (Nanyang Technological University, Singapour). « Quantification of Arterial Input Function using Phase Imaging in MRI ».
- ▶ 2010 - Obioma Chigozie UDOBATA, stage master recherche. « Approche multi-paramétrique pour la segmentation de la pénombre en ischémie cérébrale ».
- ▶ 2011 - Ionel AVRAM, stage d'échange (université Oradea, Roumanie). « Methodological Contribution to Image Restoration Framework for USPIO Quantification ».
- ▶ 2013 - Pierre PORTEJOIE, stage ingénieur. « Mosaicing et segmentation des os du corps humain à partir de volumes CT ». Projet FP7 PIPER. (MOREAU et al., 2016)
- ▶ 2014 - Pierre PORTEJOIE, projet de fin d'études. « Anisotropic Diffusion for Spatio-Temporal Data Filtering ». (PORTEJOIE et al., 2015)
- ▶ 2015 - Shaojie WANG, stage master recherche. « Cardiac Structure Detection on Ultrasound 3-D Image Using Structured Random Forest ».
- ▶ 2016 - Khanh TRAN, projet de fin d'études. « Automatisation de l'acquisition de séquences de projections en tomographie électronique ». Projet ANR 3DCLEAN. (FENG et al., 2016).
- ▶ 2016 - Yuemeng FENG, projet de fin d'études et stage master recherche. « Correction de flou pour la tomographie en microscopie électronique environnementale ». Projet ANR 3DCLEAN. (FENG et al., 2016).
- ▶ 2016 - Hoai Thu NGUYEN, stage master recherche. « Study and characterize functional changes of skeletal muscles from longitudinal multiparametric MRI datasets ».
- ▶ 2017 - Hoai Thu NGUYEN, stage master recherche. « Segmentation robuste des muscles par recalage multi-atlas couplé avec apprentissage automatique ».
- ▶ 2018 - Daniel IANNI, stage master recherche. « Réalignement de projections acquises (très) rapidement en microscopie électronique transmission électronique (ETEM) ».
- ▶ 2019 - Paul DESJARDIN, projet de fin d'études. « Apprentissage profond pour la segmentation de lésions de sclérose en plaque à partir d'examen IRM ».
- ▶ 2019 - Guillaume MARTINOD, projet de fin d'études et stage master recherche. « Apprentissage profond pour la segmentation du liquide céphalorachidien en tomographie X ».
- ▶ 2020 - Malick KANDJI, projet de fin d'études. « Segmentation automatique par méthodes supervisées des muscles de l'épaule à partir d'images IRM ». Collaboration avec les HCL.
- ▶ 2020 - Enyi CHEN, projet de fin d'études et stage master recherche. « Analyse de la déformation des globules rouges pour l'identification de pathologies ». Collaboration avec l'INL.
- ▶ 2021 - David OSOWIECHI, projet de fin d'études et stage master recherche. « Réseau d'attention pour la segmentation automatique des muscles de l'épaule en IRM et de muscles du tronc pour l'étude de la sarcopénie en CT. Portage des outils sous 3D Slicer. », projet avec les HCL.
- ▶ 2021 - Emile SAILLARD, projet de fin d'études et stage master recherche. « Segmentation automatique du fémur et de métastases osseuses pour la prédiction par simulation du risque de fracture. », en collaboration avec le LBMC/LYOS et les HCL. (SAILLARD et al., 2021)
- ▶ 2021 - Duc Toan NGUYEN, projet de fin d'études. « Système d'acquisition et de détection automatique et embarqué de pestes sur des plantes en vue de la planification de leur destruction ». Collaboration avec Ampère, ANR Greenshield. (LACOTTE et al., 2022)

- ▶ 2022 - Claire LEMOINE, projet de fin d'études. « Segmentation et parcellisation automatique du corps calleux de rats pour l'étude longitudinale de la re-myélinisation ». Collaboration avec l'équipe MAGICS de CREATIS et le projet transversal MUSIC de CREATIS.
- ▶ 2022 - Maylis JOUVENCEL, projet de fin d'études. « Segmentation automatique par apprentissage profond des muscles des jambes en IRM : apport de l'utilisation de plusieurs séquences et généralisation à d'autres bases de données ». Collaboration avec l'équipe MAGICS de CREATIS et le projet transversal IDM4 de CREATIS (JOUVENCEL et al., 2022).
- ▶ 2022 - Léon PEYRAT, stage optionnel 4ème année. « Optimisation de la segmentation par module d'attention dans les réseaux de neurones pour l'étude de la sarcopénie en scanner 3D ». Collaboration avec les HCL.
- ▶ 2022 - Zhencheng ZHANG, stage d'été (1,5 mois). « Intégration dans 3DSlicer des outils de segmentation automatique basés réseaux de neurones et recalage multi-atlas avec les pré- et post- traitements usuels ». Collaboration avec les HCL et le projet transverse IDM4 de CREATIS.

5.5 Projets

PEPS INS2I CNRS - AAP 2013-2014

Durée de 24 mois. Montant de 16kEuros. « Classification Spatio-Temporelle d'ensemble de données multiparamétriques. Application à la caractérisation de la phase chronique des lésions SEP en suivi longitudinal IRM ». J'ai été porteur de ce projet qui a permis d'accompagner les travaux de thèse de Simon Mure (achat d'ordinateur, échange avec Harvard Medical School, différents frais de missions).

Projet Européen (FP7) PIPER

<http://www.piper-project.eu/>

Le budget total est de 3,8 millions d'Euros dont 2,9 financés par la commission Européenne. Projet coordonné par Philippe Beillas (Université de Lyon 1 - IFSTTAR) et qui regroupe 10 partenaires provenant de 5 pays différents pour une période de 3 ans et demi. Ma contribution à ce projet a porté sur les traitements d'images médicales (scanner X corps entier de cadavre) dans le but de produire des segmentations et d'identifier automatiquement les os afin d'alimenter des simulateurs physiques de crash-test avec des morphologies différentes (MOREAU et al., 2016).

ANR 3DClean

http://www.clym.fr/3DCLEAN_web/3DCLEAN-ANR.html.

Le budget total est d'environ 1 950k€ dont une aide de l'ANR de 618k€. Projet coordonné par Thierry Epicier (MATEIS) de 2015 à 2019 regroupant 5 unités partenaires dont l'objectif était le développement expérimental de la microscopie électronique environnementale appliqué à la catalyse.

J'ai été impliqué pour les problématiques liées à la tomographie : débruitage et déconvolution des acquisitions (FENG et al., 2016), optimisation de l'acquisition par analyse d'images (EPICIER et al., 2018; EPICIER et al., 2019; KONETI et al., 2019; KONETI et al., 2016) et algorithme de reconstruction 3D (BANJAK et al., 2018).

EUR SLEIGHT-MANUTECH Dionisos

2019 à 2021

Budget de 87,5k€. Projet coordonné par Christophe Ducottet (laboratoire Hubert Curien St Etienne) et Thierry Epicier (MATEIS - IRCELyon). Projet sur le « Suivi de la diffusion

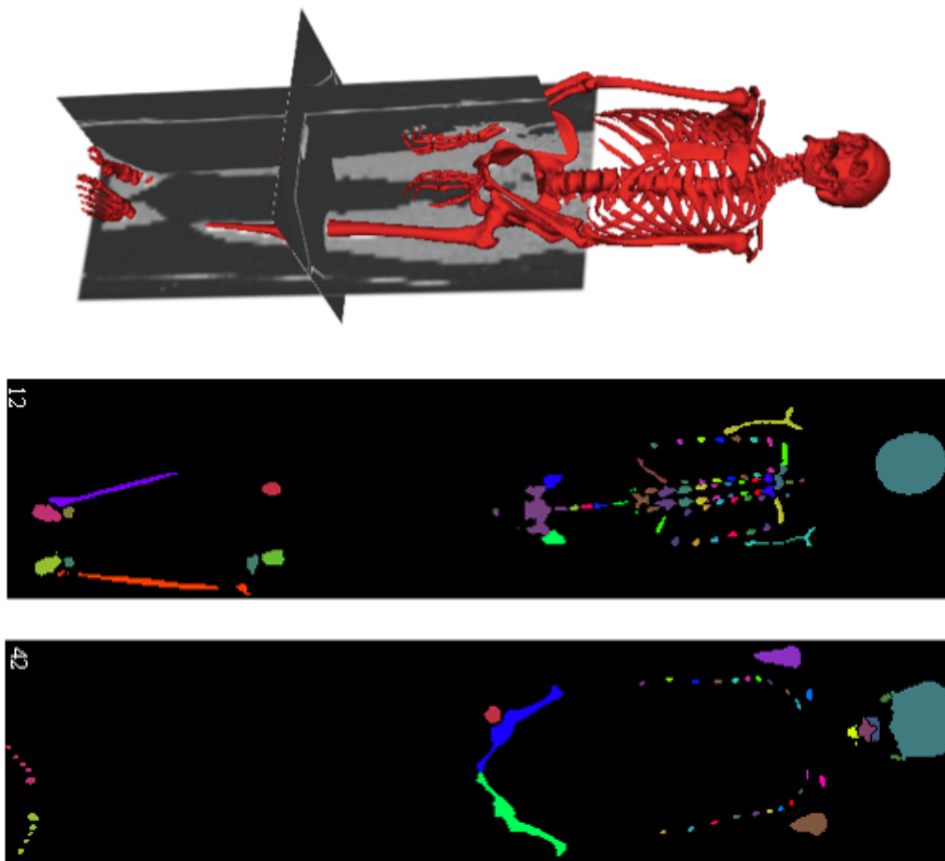


FIGURE 5.2 – Illustration de la segmentation par recalage d’atlas (en haut) puis détails des labels des os dans deux coupes (en bas), Projet PIPER.

de nanoparticules sur des surfaces pour l’étude de la nanocatalyse imagée par microscopie en transmission électronique ». Je collabore pour la segmentation et détection des nanoparticules (apprentissage profond) ainsi que les prétraitements (recalages) (FARAZ et al., 2021 ; FARAZ et al., 2022).

5.6 Contrats industriels

J’ai conduit une douzaine de contrats d’expertises avec 8 sociétés de 2015 à 2021 pour un montant total d’environ 65k Euros via la filiale de valorisation de l’INSA Lyon, INSAValor. Si le traitement d’images est le point central de ces expertises, l’intérêt pour mon profil est dans la maîtrise des environnements matériel, électronique et informatique. Voici la liste des principaux contrats.

- ▶ Bioclinica (2015). Etude d’algorithmes de correction de biais dans les images IRM.
- ▶ EFI automotive (2016-2019). Analyse d’images pour la simulation de la diffusion de la chaleur sur des cartes électroniques. Intérêt des bibliothèques GPU d’apprentissage profond pour la résolution numérique d’équations différentielles.
- ▶ Ayming (2018). Formation et potentiel de l’apprentissage profond pour l’analyse d’images médicales.
- ▶ SNCF (2019). Appuis scientifiques et techniques pour la réalisation d’un détecteur de défauts sur des cartes électroniques.

- ▶ Alstom (2018-2019). Reconnaissance temps réel d'ouvrages d'art en milieu urbain, estimation du nombre de passagers à quai et dans les voitures (tram et métro) à partir de vidéos.
- ▶ Biomae (2019-2021). Détection et classification de gammars. Appuis à la conception d'un système de triage automatiques de gammars (caméras, systèmes de pilotages, interfaces des systèmes).
- ▶ STEF (2021-2022). Apport de l'analyse d'images pour la supervision de chaînes de co-packing majoritairement manuelles. Conception et dimensionnement de solutions.

5.7 Responsabilités

- ▶ Depuis le démarrage du labex PRIMES, je suis impliqué dans le Work Package Formation et j'en suis depuis 2018 le co-responsable. J'ai géré : le choix des 8 plateformes et leurs achats puis à leurs mises en oeuvre, la rénovation de deux salles au sein du département GE pour accueillir ces plateformes, à la diffusion de l'offre de formation et l'organisation son utilisation. Le budget matériel mis en place fut d'environ 200k€. Cette salle assure 200h/an de TP spécifiques pour une dizaine de formations universitaires différentes. Je suis à l'origine de la création de l'école d'été en *deep learning* pour l'imagerie médicale. J'ai organisé celles qui se sont tenues à Lyon (département Génie Electrique) en 2019 (figure 5.3) et en virtuel en avril 2021 (figure 5.4). Une troisième édition de cette école a eu lieu début juillet 2022 à Montréal avec nos collaborateurs Canadiens et pour laquelle je fais parti des comités d'organisation et scientifique 5.5. Je co-organise aussi la quatrième édition à Lyon en avril 2023. Ces écoles se déroulent sur une semaine, incluent 4 TP et des événements scientifiques (session posters avec buffet, table ronde, ...) et sociaux (sorties, soirée gala, ...). Ces écoles ont chacune accueillis entre 80 et 150 participants inscrits de 20 nationalités différentes.
- ▶ Depuis 2019, je suis le co-responsable du projet transversal MUSIC (MUltiple Sclerosis and neuro Inflammation) du laboratoire CREATIS. Mon rôle est d'animer l'équipe de recherche et d'organiser l'articulation des axes de recherches et des demandes de financements. La dotation annuelle par le laboratoire est de 3,5k€. Ce projet regroupe 13 permanents de CREATIS (ingénieurs, chercheurs en clinique, chercheurs en IRM, chercheurs en traitement d'images) réunis tous les deux mois pendant 2h. En 2020, 13 réponses à appel à projets ont été déposés et 5 ont obtenus un financements pour 2021 (une bourse de thèse, 80k€, temps ingénieurs et post-doc, temps machine IRM et animaux)
- ▶ Membre du CU de CREATIS de 2016 à 2021. Principalement impliqué pour les problématiques de travaux (plan campus, réhabilitation des nouveaux locaux du laboratoire) et de calculs GPU (recherche de fonds (projet PPAIR 40kEuros), choix des machines GPU (Creatis et IN2P3), recommandations d'usages, mise à disposition de bibliothèques).
- ▶ Relecteur pour IEEE (5 relectures par ans depuis 2013 pour ICIP, 7 revues pour TIP) et Elsevier (environ 20 revues pour Pattern Recognition).



FIGURE 5.3 – Bandeau du site web de l'école d'été co-organisée *deepimaging2019*. *sciencesconf.org* ayant accueilli 150 participants, 70 inscrits à l'ensemble des activités.



FIGURE 5.4 – Bandeau du site web de la seconde édition de l'école d'été *deepimaging2021*. *sciencesconf.org* en mode virtuel suivie par 80 participants inscrits à toutes les activités.



FIGURE 5.5 – Bandeau du site web de la troisième édition de l'école d'été *deepimaging2022*. *sciencesconf.org* à Montréal ayant avec 65 participants sur place et 70 en asynchrone.



6. Production scientifique

Ma production scientifique est de :

- ▶ 19 articles dans des revues internationales
- ▶ 1 revue nationale
- ▶ 1 chapitre de livre
- ▶ 38 conférences dont 9 nationales
- ▶ 4 présentations pour des workshops et des journées thématiques

Articles

- AMELI, R., S. MURE, C. R. GUTTMANN, T. **Grenier**, H. BENOIT-CATTIN et F. COTTON (2016). « Analyse dynamique hebdomadaire du développement péri-veinulaire des lésions actives de SEP par imagerie de susceptibilité magnétique. » In : *Journal of Neuroradiology* 43.2, p. 91-93.
- BANJAK, H., T. **Grenier**, T. EPICIER, S. KONETI, L. ROIBAN, A.-S. GAY, I. MAGNIN, F. PEYRIN et V. MAXIM (2018). « Evaluation of noise and blur effects with SIRT-FISTA-TV reconstruction algorithm : Application to fast environmental transmission electron tomography ». In : *Ultramicroscopy* 189, p. 109-123.
- DAVILLER, C., T. **Grenier**, H. RATINEY, M. SDIKA, P. CROISILLE et M. VIALON (2019a). « Automatic myocardial ischemic lesion detection on magnetic resonance perfusion weighted imaging prior perfusion quantification : A pre-modeling strategy ». In : *Computers in Biology and Medicine*, p. 108-119.
- DOLET, A., F. VARRAY, S. MURE, T. **Grenier**, Y. LIU, Z. YUAN, P. TORTOLI et D. VRAY (2018). « Spatial and spectral regularization to discriminate tissues using multispectral photoacoustic imaging ». In : *EURASIP Journal on Advances in Signal Processing* 2018.1, p. 39.
- DUMORTIER, L., F. GUÉPIN, M. L. DELIGNETTE-MULLER, C. B. BOULOCHER et T. **Grenier** (2022). « Deep learning in veterinary medicine, an approach based on CNN to detect pulmonary abnormalities from lateral thoracic radiographs in cats ». In : *Scientific Reports* 12.

- EPICIER, T., H. BANJAK, A.-S. GAY, T. **Grenier**, S. KONETI, V. MAXIM et L. ROIBAN (2018). « Very fast tomography in the (E)TEM to probe dynamics in materials during operando and in situ experiments ». In : *Microscopy and Microanalysis* 24.S1, p. 1814-1815.
- EVAIN, E., K. FARAZ, T. **Grenier**, D. GARCIA, M. DE CRAENE et O. BERNARD (2020). « A pilot study on convolutional neural networks for motion estimation from ultrasound images ». In : *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 67.12, p. 2565-2573.
- FARAZ, K., T. **Grenier**, C. DUCOTTET et T. EPICIER (2021). « A machine learning pipeline to track the dynamics of a population of nanoparticles during in situ Environmental Transmission Electron Microscopy in gases ». In : *Microscopy and Microanalysis* 27, p. 2236-2237.
- (2022). « Deep learning detection of nanoparticles and multiple object tracking of their dynamic evolution during in situ ETEM studies ». In : *Scientific Reports* 12, p. 2484.
- FLAUS, A., T. DEDDAH, A. REILHAC, N. D. LEIRIS, M. JANIER, I. MERIDA, T. **Grenier**, C. J. MCGINNITY, A. HAMMERS, C. LARTIZIEN et N. COSTES (2022). « PET image enhancement using artificial intelligence for better characterization of epilepsy lesions ». In : *Frontiers in Medicine* 9.
- GRENIER, T., C. REVOL-MULLER, N. COSTES, M. JANIER et G. GIMENEZ (2005a). « Automated seeds location for whole body NaF PET segmentation ». In : *IEEE Transactions on Nuclear Science* 52.5, p. 1401-1405.
- KONETI, S., L. ROIBAN, F. DALMAS, C. LANGLOIS, A.-S. GAY, A. CABIAC, T. **Grenier**, H. BANJAK, V. MAXIM et T. EPICIER (2019). « Fast electron tomography : Applications to beam sensitive samples and in situ TEM or operando environmental TEM studies ». In : *Materials Characterization* 151, p. 480-495.
- LACOTTE, V., T. NGUYEN, J. D. SEMPERE, V. NOVALES, V. DUFOUR, R. MOREAU, M. T. PHAM, K. RABENOROSOA, S. PEIGNIER, F. G. FEUGIER, R. GAETANI, T. **Grenier**, B. MASENELLI, P. da SILVA, A. HEDDI et A. LELEVÉ (2022). « Pesticide-Free Robotic Control of Aphids as Crop Pests ». In : *AgriEngineering* 4.4, p. 903-921.
- LECLERC, S., E. SMISTAD, A. OSTVIK, F. CERVENANSKY, F. ESPINOSA, T. ESPELAND, E. A. R. BERG, M. BELHAMISSI, S. ISRAILOV, T. **Grenier**, C. LARTIZIEN, P. M. JODOIN, L. LØVSTAKKEN et O. BERNARD (2020). « LU-Net : A multistage attention network to improve the robustness of segmentation of left ventricular structures in 2-D echocardiography ». In : *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 67.12, p. 2519-2530.
- LECLERC, S., E. SMISTAD, J. PEDROSA, A. OSTVIK, F. CERVENANSKY, F. ESPINOSA, T. ESPELAND, E. J. BERG, P.-M. JODOIN, T. **Grenier**, C. LARTIZIEN, J. DRHOOGHE, L. LØVSTAKKEN et O. BERNARD (2019c). « Deep learning for segmentation using an open large-scale dataset in 2D echocardiography ». In : *IEEE Transactions on Medical Imaging* 38.9, p. 2198-2210.
- MANET, R., L. GERGELÉ, T. **Grenier**, Z. H. CZOSNYKA et M. CZOSNYKA (2020). « Development of normal pressure hydrocephalus following post-traumatic external hydrocephalus in an adult patient ». In : *British Journal of Neurosurgery* 0.0, p. 1-4.
- MURE, S., T. **Grenier**, D. S. MEIER, C. R. GUTTMANN et H. BENOIT-CATTIN (2015b). « Unsupervised spatio-temporal filtering of image sequences. A mean-shift specification ». In : *Pattern Recognition Letters* 68, Part 1, p. 48-55.
- NGUYEN, H.-T., S. GRANGE, B. LEPORQ, M. VIALON, P. CROISILLE et T. **Grenier** (2021a). « Impact of distortion on local radiomic analysis of quadriceps based on

quantitative magnetic resonance imaging data ». In : *International Journal of Pharma Medicine and Biological Sciences* 10.2, p. 49-54.

NGUYEN, H.-T., T. **Grenier**, B. LEPORQ, C. L. GOFF, B. GILLES, S. GRANGE, R. GRANGE, G. P. MILLET, O. BEUF, P. CROISILLE et M. VIALON (2021b). « Quantitative magnetic resonance imaging assessment of the quadriceps changes during an extreme mountain ultramarathon ». In : *Medicine & Science in Sports & Exercise* 53.4, p. 869-881.

Chapitre de livre

REVOL-MULLER, C., T. **Grenier**, J. ROSE, A. PACUREANU, F. PEYRIN et C. ODET (2013). « Region Growing : When Simplicity meets theory. Region growing revisited in feature space and variational framework ». In : *Communications in Computer and Information Science Computer Vision, Imaging and Computer Graphics. Theory and Application*. Sous la dir. de G. CSURKA, M. KRAUS, R. S. LARAMEE, P. RICHARD et J. BRAZ. T. 359. Springer. Chap. 426, p. 426-444.

Conférences

CHARPIGNY, D., T. **Grenier**, C. ODET et H. BENOIT-CATTIN (2009). « Towards iron oxide nanoparticles quantization in molecular MR images by default field deconvolution ». In : Boston, MA, United states, p. 137-140.

— (2010). « Restoration-based iron oxide particles quantification in MR images ». English. In : Dallas, TX, United states, p. 485-488.

CHARPIGNY, D., T. **Grenier**, C. ODET, J. PAULY et H. BENOIT-CATTIN (2008). « Deconvolution approach for susceptibility map building ». In : *European Society for Magnetic Resonance in Medicine and Biology, 25th Annual Scientific Meeting ESMRMB*. Valencia, Spain, in-press.

DAVIGNON, F., T. **Grenier**, C. REVOL-MULLER, G. GIMENEZ et O. BASSET (2005). « Lissage et segmentation d'images multi-paramétriques ultrasonores par une approche 'Mean shift' ». In : *GRETSI'05*. Louvain-La-Neuve, Belgium, p. 21-24.

DOLET, A., F. VARRAY, S. MURE, T. **Grenier**, Y. LIU, Z. YUAN, P. TORTOLI et D. VRAY (2016). « Spatial and spectral regularization for multispectral photoacoustic image clustering ». In : *IEEE International Ultrasonics Symposium*. Proceedings of the 2016 IEEE International Ultrasonics Symposium (IUS). Tours, France.

DUPUY, J., T. DARNANVILLE, F. VALOIS, N. ARNESEN, M.-P. FAVRE, A. FAVE, J.-Y. CAVAILLÉ, T. **Grenier**, E. DUMITRESCU, E. NIEL et C. ODET (2011). « Le modèle International de l'INSA de Lyon : au-delà de l'échange académique - Dédié aux personnels et étudiants de l'Université du Tohoku à Sendai ». In : t. 10, p. 1009.

EPICIER, T., T. **Grenier**, H. BANJAK, V. MAXIM, S. KONETI et L. ROIBAN (2019). « Very fast acquisition of tilt series in environmental TEM tomography : tips and tricks ». In : *XVIème colloque de la Société Française des Microscopies*. Poitiers, France.

FENG, Y.-M., K. TRAN, S. KONETI, L. ROIBAN, A.-S. GAY, C. LANGLOIS, T. EPICIER, T. **Grenier** et V. MAXIM (2016). « Image deconvolution for fast tomography in environmental transmission electron microscopy ». In : *European Microscopy Congress 2016 : Proceedings*. Wiley-VCH Verlag GmbH & Co. KGaA.

GRENIER, T., C. REVOL-MULLER, N. COSTES, M. JANIER et G. GIMENEZ (2003). « Automated seeds location for whole body NaF PET segmentation ». English. In : t. 3. Portland, OR, United states, p. 2210-2214.

- GRENIER, T., C. REVOL-MULLER, N. COSTES, M. JANIER et G. GIMENEZ (2006a). « 3D robust adaptive region growing for segmenting [18F] fluoride ion PET images ». In : *2006 IEEE Nuclear Science Symposium Conference Record*. T. 5, p. 2644-2648.
- GRENIER, T., C. REVOL-MULLER, F. DAVIGNON, O. BASSET et G. GIMENEZ (2005b). « 'Mean shift' adaptatif pour le lissage d'images ultrasonores ». In : *GRETSI'05*. Louvain-La-Neuve, Belgium, p. 53-56.
- (2005c). « Multiparametric smoothing based on mean shift procedure for ultrasound data segmentation ». In : *EUSIPCO'05*. Antalya, Turkey, Article ID cr1461, 4 pages.
- (2005d). « Variable bandwidth mean shift for Smoothing ultrasonic images ». In : *EUSIPCO'05*. Antalya, Turkey, Article ID cr1454, 4 pages.
- GRENIER, T., C. REVOL-MULLER et G. GIMENEZ (2006c). « Hybrid approach for multiparametric mean shift filtering ». In : *2006 International Conference on Image Processing*, p. 1541-1544.
- JOUVENCEL, M., H.-T. NGUYEN, M. VIALON, P. CROISILLE et T. Grenier (2022). « Impact of MR sequences choice on deep learning segmentation of muscles ». In : *2022 16th IEEE International Conference on Signal Processing (ICSP)*. T. 1, p. 420-425.
- KONETI, S., L. ROIBAN, V. MAXIM, T. Grenier, P. AVENIER, A. CABIAC, A.-S. GAY, F. DALMAS et T. EPICIER (2016). « Environmental transmission electron tomography : fast 3D analysis of nano-materials ». In : *European Microscopy Congress 2016*. Lyon, France : Wiley-VCH Verlag GmbH & Co. KGaA, p. 29-30.
- LECLERC, S., T. Grenier, F. ESPINOZA et O. BERNARD (2017). « A fully Automatic and multi-structural segmentation of the left ventricle and the myocardium on highly heterogeneous 2D echocardiographic data ». In : *2017 IEEE International Ultrasonic Symposium (IUS)*. Washington, DC, United States.
- LECLERC, S., E. SMISTAD, T. Grenier, C. LARTIZIEN, A. OSTVIK, F. CERVENANSKY, F. ESPINOSA, T. ESPELAND, E. A. RYE BERG, P.-M. JODOIN, L. LØVSTAKKEN et O. BERNARD (2019a). « RU-Net : A refining segmentation network for 2D echocardiography ». In : *2019 IEEE International Ultrasonics Symposium (IUS)*. Glasgow, France : IEEE, p. 1160-1163.
- LECLERC, S., E. SMISTAD, T. Grenier, C. LARTIZIEN, A. OSTVIK, F. ESPINOSA, P.-M. JODOIN, L. LOVSTAKKEN et O. BERNARD (2018). « Deep Learning applied to multi-structure segmentation in 2D echocardiography : A preliminary investigation of the required database size ». In : *2018 IEEE International Ultrasonics Symposium (IUS)*. Kobe, France : IEEE, p. 1-4.
- LECLERC, S., E. SMISTAD, A. OSTVIK, F. CERVENANSKY, F. ESPINOSA, T. ESPELAND, E. ANDREAS, R. BERG, P.-M. JODOIN, T. Grenier, C. LARTIZIEN, L. LOVSTAKKEN et O. BERNARD (2019b). « Deep learning segmentation in 2D echocardiography using the CAMUS dataset : Automatic assessment of the anatomical shape validity ». In : *International conference on Medical Imaging with Deep Learning (MIDL 2019)*. London, United Kingdom.
- LI, T., S. CAMARASU-POP, T. GLATARD, T. Grenier et H. BENOIT-CATTIN (2010a). « Optimization of mean-shift scale parameters on the EGEE grid ». In : *Studies in health technology and informatics, Proceedings of Healthgrid 2010*. T. 159, p. 203-214.
- LI, T., T. Grenier et H. BENOIT-CATTIN (2011). « Color space influence on mean shift filtering ». In : *Proceedings - International Conference on Image Processing, ICIP*. Brussels, Belgium, p. 1469-1472.
- MOREAU, B., A.-H. DICKO, P. MAILLIEZ, P. PORTEJOIE, C. LECOMTE, M. BAH, T. Grenier, E. JOLIVET, P. PETIT, B. FRÉCHÈDE, F. FAURE, B. GILLES et P. BEILLAS (2016). « A segmentation pipeline for the creation of statistical shape models in the

- PIPER project ». In : *22nd Congress of the European Society of Biomechanics (ESB 2016)*.
- MURE, S., T. **Grenier** et H. BENOIT-CATTIN (2016a). « Unsupervised spatiotemporal video clustering a versatile mean-shift formulation robust to total object occlusions ». In : *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 1536-1540.
- MURE, S., T. **Grenier**, P. GAŃCARSKI et H. BENOIT-CATTIN (2015a). « Spécification du filtrage mean-shift pour la classification non supervisée de séries temporelles multidimensionnelles. » In : *Colloque GRETSI*, p. 1-4.
- MURE, S., T. **Grenier**, C. R. G. GUTTMANN et H. BENOIT-CATTIN (2016b). « Unsupervised time-series clustering of distorted and asynchronous temporal patterns ». In : *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 1263-1267.
- MURE, S., T. **Grenier**, C. R. G. GUTTMANN, F. COTTON et H. BENOIT-CATTIN (2016c). « Classification of multiple sclerosis lesion evolution patterns a study based on unsupervised clustering of asynchronous time-series ». In : *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, p. 1315-1319.
- MURE, S., C. R. G. GUTTMANN, T. **Grenier**, H. BENOIT-CATTIN et F. COTTON (2016d). « New insight in perivenular lesion formation in multiple sclerosis on weekly susceptibility weighted images ». In : *2016 International Society for Magnetic Resonance in Medicine*.
- NGUYEN, H.-T., P. CROISILLE, M. VIALON, C. de BOURGUIGNON, R. GRANGE, S. GRANGE et T. **Grenier** (2018a). « Robust multi-atlas MRI segmentation with corrective learning for quantification of local quadriceps muscles inflammation changes during a longitudinal study in athletes ». In : *Proc. Intl. Soc. Mag. Reson. Med. 26*. Paris, France.
- (2019a). « Variation en IRM quantitative de la moelle femorale suite à un ultra-trail ». In : *SFRMBM : Société Française de Résonance Magnétique en Biologie and Médecine, 4e Congès*. Strasbourg, France.
- NGUYEN, H.-T., P. CROISILLE, M. VIALON, S. LECLERC, S. GRANGE, R. GRANGE, O. BERNARD et T. **Grenier** (2019b). « Robustly segmenting quadriceps muscles of ultra-endurance athletes with weakly supervised U-Net ». In : *International Conference on Medical Imaging with Deep Learning – Extended Abstract Track*. London, United Kingdom.
- NGUYEN, H.-T., T. **Grenier**, B. LEPORQ, L. BEY, M. VIALON et P. CROISILLE (2019c). « Evaluation of local changes in femoral bone marrow during a mountain ultra-marathon with quantitative MRI Results & Discussion ». In : *Proc. Intl. Soc. Mag. Reson. Med. 27*. Montréal, Canada.
- (2019d). « Segmentation multi-atlas avec apprentissage automatique pour l'étude de la variation fonctionnelle des quadriceps au long d'un ultra-marathon ». In : *SFRMBM : Société Française de Résonance Magnétique en Biologie and Médecine, 4e Congès*. Strasbourg, France.
- NGUYEN, H.-T., T. **Grenier**, B. LEPORQ, C. LE GOFF, C. de BOURGUIGNON, G. GIARDINI, G. MILLET et O. BEUF (2018b). « Longitudinal study of quadriceps muscle head inflammation of athletes enrolled in extreme mountain ultra-marathon using radiomic features extracted from automatic segmentation based on atlas registration and machine learning of MR Images ». In : *ELMSK : Exercise, locomotion and musculoskeletal system*. Lyon, France.
- PORTEJOIE, P., S. MURE, H. BENOIT-CATTIN et T. **Grenier** (2015). « Locally controlled regularized spatiotemporal anisotropic diffusion ». In : *Image Processing (ICIP), 2015 IEEE International Conference on*. Québec, Canada, p. 4823-4827.

- REVOL-MULLER, C., T. **Grenier**, T. LI et H. BENOIT-CATTIN (2012a). « Feature space region growing ». In : *2012 19th IEEE International Conference on Image Processing*. Orlando, United States, p. 2585-2588.
- REVOL-MULLER, C., T. **Grenier**, J. ROSE, A. PACUREANU, F. PEYRIN et C. ODET (2012b). « Region Growing : Adolescence and adulthood ; Two visions of region growing : in feature space and variational framework ». In : *VISAPP 2012. International Conference on Computer Vision Theory and Applications*. Rome, Italy, p. 286-297.
- ROIBAN, L., S. KONETI, K. TRAN, Y.-M. FENG, T. **Grenier**, V. MAXIM et T. EPICIER (2016). « Rapid tomography in environmental TEM : How fast can we go to follow the 3D evolution of nanomaterials in situ ? ». In : t. 22. S5. Cambridge University Press, 8–9.
- ROSE, J. L., T. **Grenier**, C. REVOL-MULLER et C. ODET (2010). « Unifying variational approach and region growing segmentation ». In : *Signal Processing Conference, 2010 18th European*, p. 1781-1785.
- SAILLARD, E., C. CONFAVREUX, M. GARDEGARONT, D. JICQUEL, S. CADOT, D. MITTON, F. BERNMOND, H. FOLLET, J.-B. PIALAT et T. **Grenier** (2021). « UNet based automatic femur segmentation with few annotated data for bone fracture prediction : from pre-processing to segmentation assessments ». In : *Medical Image Analysis and Artificial Intelligence, 2nd Sino French Workshop*. Online, France, 2p.
- SUN, Y., K. FARAZ, T. **Grenier**, P. CLARYSSE, D. GARCIA et O. BERNARD (2020). « A duplex pipeline for the generation of realistic echocardiographic sequences with doppler imaging ». In : p. 1-4.
- WARGNIER-DAUCHELLE, V., T. **Grenier**, F. DURAND-DUBIEF, F. COTTON et M. SDIKA (2021a). « A more interpretable classifier for multiple sclerosis ». In : *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. Nice, France, p. 1062-1066.
- WARGNIER-DAUCHELLE, V., T. **Grenier**, F. DURAND-DUBIEF, F. COTTON et M. SDIKA (2021b). « Un classifieur plus Interprétable pour la SEP ». In : *Congrès Société Française de Résonance Magnétique en Biologie et Médecine (SFRMBM)*. Lyon, France.

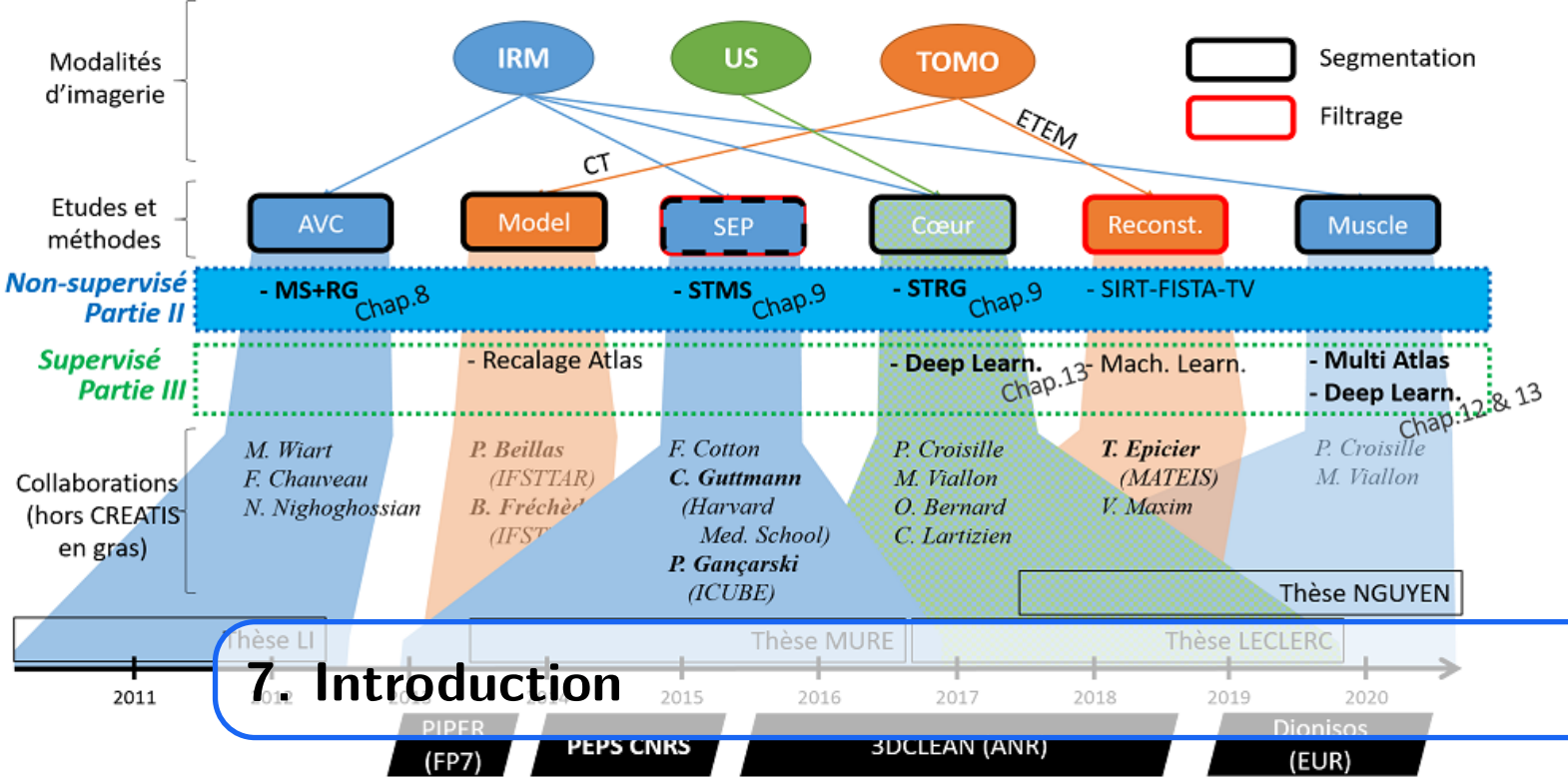
Autres

- GRENIER, T. (2020). *Machine learning en imagerie Médical – Journée Thématique Machine Learning à Lyon de la Structure de recherche IXXI, le 23-01-2020*.
- GRENIER, T. et C. LARTIZIEN (2010). *Architecture de fusion modulaire pour l'assistance au diagnostic médical, Réunion des GDR STIC-SANTE Thème : Signaux et Images en santé et GDR ISIS Thème B : Image et Vision*.
- GRENIER, T., C. REVOL-MULLER, N. COSTES, M. JANIER et G. GIMENEZ (2006b). *3D robust adaptive region growing for segmenting PET images, "1st Singaporean-French Biomedical Imaging Workshop, 2006"*.



Contributions au filtrage et à la segmentation non supervisés

7	Introduction	63
8	Filtrage <i>mean-shift</i> et segmentation ...	65
8.1	Introduction	
8.2	Filtrage d'images et <i>mean-shift</i>	
8.3	Optimisations des paramètres d'échelles	
8.4	Du filtrage <i>mean-shift</i> à la segmentation d'images	
8.5	Application à l'accident vasculaire cérébral	
8.6	Conclusion	
9	Filtrage et segmentation spatio-temporels	83
9.1	Introduction	
9.2	Diffusion Anisotrope temporelle	
9.3	<i>Mean-shift</i> Spatio-Temporel : STMS	
9.4	Croissance de région spatio-temporelle : STRG	
9.5	Conclusion du chapitre	
10	Conclusion de la seconde partie	103



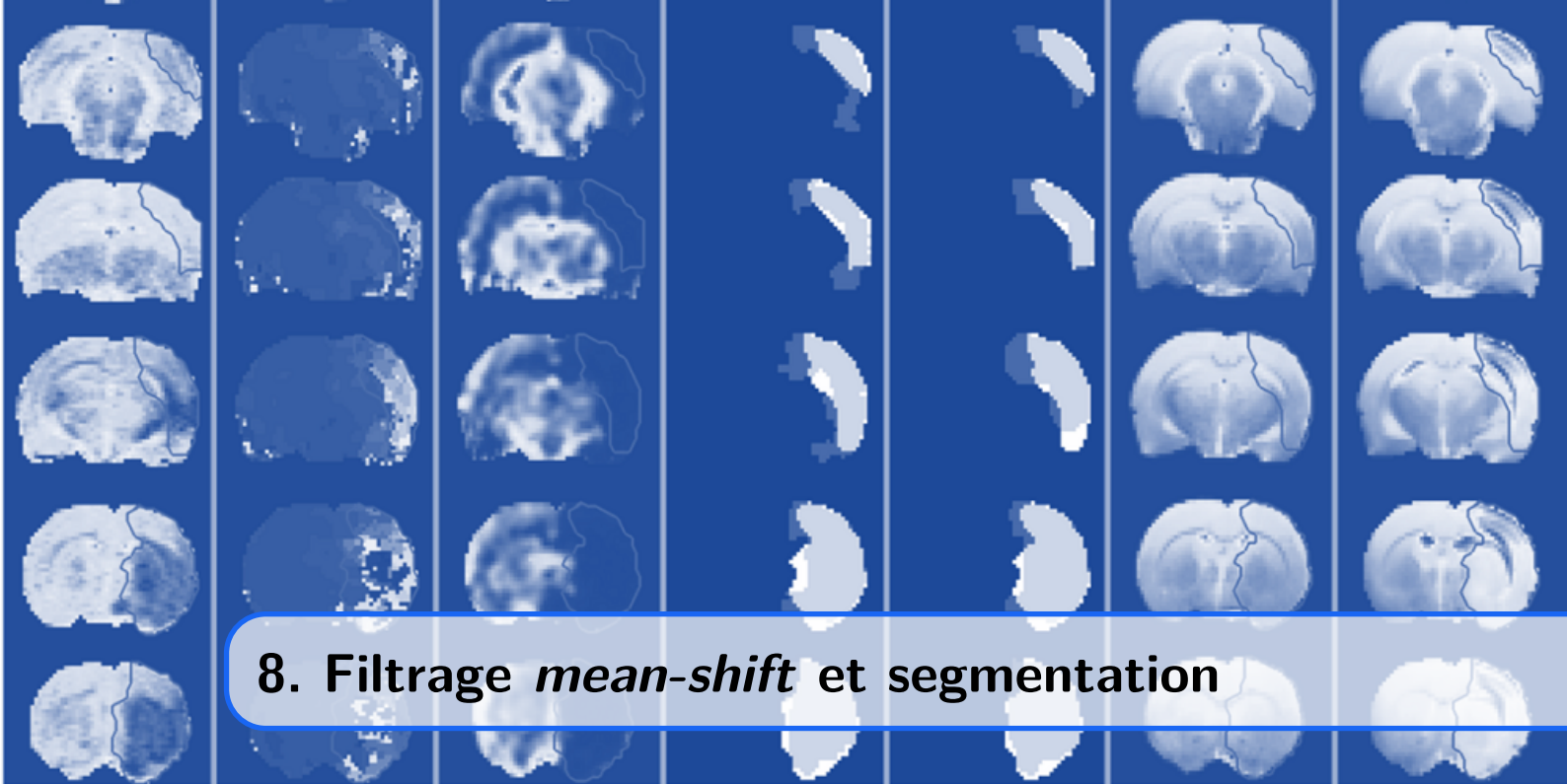
7. Introduction

Dans cette partie, mes travaux sur les *mean-shift* et leurs adaptations dans différents contextes méthodologiques sont synthétisés. Ils couvrent principalement les travaux de thèse de Ting Li (2012) et de Simon Mure (2016) que j'ai co-dirigés et qui prolongent mes travaux de thèse (2005).

Dans un premier chapitre (chapitre 8), nous présentons le *mean-shift* dans le contexte du filtrage d'images. Puis nous nous intéressons aux optimisations proposées pour cette approche et notamment comment lui associer une méthode de segmentation. Ces contributions sont appliquées à la segmentation de la zone de pénombre suite à un accident vasculaire cérébral en IRM.

Dans un second chapitre (chapitre 9), nous abordons la problématique du spatio-temporel : comment filtrer et segmenter des données issues d'un suivi médical ou d'une séquence d'images ? Les approches proposées sont appliquées au filtrage et à la segmentation de lésions de sclérose en plaques.

Nous concluons cette partie par une analyse de ces travaux et les perspectives qui en découlent (chapitre 10).



8. Filtrage *mean-shift* et segmentation

8.1 Introduction

Ce chapitre dresse le contexte du filtrage d'image *mean-shift* et introduit les deux concepts importants : l'espace des caractéristiques et les paramètres d'échelles.

Nous abordons ensuite la problématique de l'optimisation de ces paramètres d'échelles puis la segmentation.

Enfin, un concept original de carte de probabilité, développé au cours des travaux de thèse de Ting Li, est détaillé. Nous étudions ensuite l'adéquation de ces approches pour la segmentation de la zone de pénombre observée lors d'un Accident Vasculaire Cérébral (AVC) en utilisant plusieurs séquences IRM.

8.2 Filtrage d'images et *mean-shift*

Le filtrage d'images intervient en général dans les premières étapes de traitements des données. Il permet d'améliorer l'image corrompue par les différentes perturbations du signal idéal afin d'en faciliter l'analyse future.

Ce terme générique de *filtrage* englobe parfois des techniques différentes de traitement du signal ou des images : il va tantôt s'agir de dé-bruitage (notamment pour la reconstruction d'images (BANJAK et al., 2018)), de lissage, de déconvolution (CHARPIGNY et al., 2009 ; CHARPIGNY et al., 2010 ; CHARPIGNY et al., 2008), de filtrage de fréquences, *etc.* Ici, nous nous focalisons sur des méthodes ne nécessitant pas de modélisation des perturbations appliquées aux images. L'hypothèse qui sera faite est que chaque objet homogène devrait être constant dans l'image. Cette hypothèse est tout à fait adaptée à l'imagerie médicale où l'on peut considérer de nombreux organes comme homogènes à une échelle et à des caractéristiques images données.

8.2.1 Brève histoire du filtrage *mean-shift*

Les *mean-shift* ont été introduits par (FUKUNAGA et al., 1975) pour des données non-images. Il s'agissait d'obtenir les modes de la densité de probabilité d'un échantillon en utilisant

l'estimation non-paramétrique par noyau de Parzen (PARZEN, 1962).

Les travaux de (CHENG, 1995) puis ceux de (COMANICIU et al., 1997) et (COMANICIU et al., 2002) ont étendu cette approche au filtrage et à la segmentation d'images.

Pour comprendre les bases du filtrage *mean-shift*, commençons par le filtrage anisotrope. L'équation de l'évolution de l'intensité I^t au pixel \mathbf{x} pour le filtrage anisotrope (BARASH, 2004) est donnée ci-après :

$$I^{t+1}(\mathbf{x}) = \frac{\sum_{\mathbf{x}_i \in \mathcal{N}(\mathbf{x})} I^t(\mathbf{x}_i) \cdot \exp\left(-\frac{\|\nabla I^t(\mathbf{x}_i)\|^2}{2\sigma_D^2}\right)}{\sum_{\mathbf{x}_i \in \mathcal{N}(\mathbf{x})} \exp\left(-\frac{\|\nabla I^t(\mathbf{x}_i)\|^2}{2\sigma_D^2}\right)} \quad (8.1)$$

où l'on a \mathbf{x} un pixel et \mathbf{x}_i un pixel de son voisinage (avec $\mathbf{x}_i \in \mathcal{N}(\mathbf{x})$), puis la fonction de pondération $\exp\left(-\frac{\|\nabla I^t(\mathbf{x}_i)\|^2}{2\sigma_D^2}\right)$ qui va dépendre du gradient de l'intensité dans l'image à l'itération t .

Cette fonction est illustrée dans la figure 8.1 avec différentes valeurs de σ_D qui règle la largeur de la gaussienne et qui sera un paramètre à régler. Plus cette valeur augmente, plus les forts gradients vont être filtrés au risque de lisser les contours des objets (le filtre va tendre vers un filtrage moyennneur).

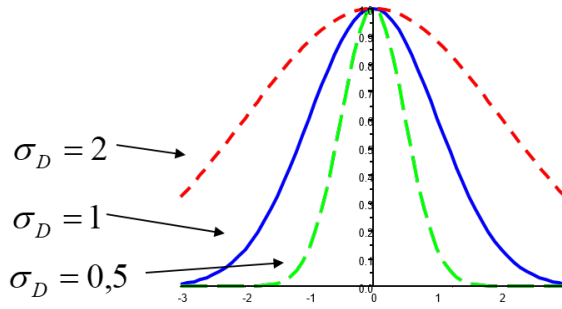


FIGURE 8.1 – Fonction de pondération du filtrage anisotrop de Barash.

L'approche *mean-shift* se distingue de cette approche à plus d'un titre. D'une part, elle converge et il ne sera donc pas nécessaire de fixer un nombre d'itérations maximum. D'autre part, elle ne modifie pas uniquement les valeurs de l'intensité de l'image.

8.2.2 Notations et principes *mean-shift*

La spécificité des *mean-shift* par rapport aux autres approches est la modification des intensités **et** des coordonnées spatiales des pixels. Ainsi l'échantillon \mathbf{x} est la concaténation de la position spatiale \mathbf{x}_s et des amplitudes \mathbf{x}_r et forme ainsi l'espace fréquemment nommé *espace joint spatial et amplitude* où s représente le nombre de dimensions spatiales et r le nombre de caractéristiques en chaque point (l'intensité, la couleur, ...). Pour une image 2D RGB, \mathbf{x} a 5 dimensions.

$$\mathbf{x} = [\mathbf{x}_s, \mathbf{x}_r] \quad \mathbf{x} \in \mathbb{R}^{s+r} \quad (8.2)$$

Pour le filtrage *mean-shift*, le calcul des nouvelles coordonnées de l'échantillon $\mathbf{x}^{[t+1]}$ est le suivant :

$$\mathbf{x}^{[t+1]} = \frac{\sum_{i=1}^n \mathbf{x}_i \cdot g\left(d^2(\mathbf{x}_i^{[t]}, \mathbf{x}^{[t]}, \mathbf{H})\right)}{\sum_{i=1}^n g\left(d^2(\mathbf{x}_i^{[t]}, \mathbf{x}^{[t]}, \mathbf{H})\right)} \quad (8.3)$$

avec :

- ▶ n le nombre d'échantillons \mathbf{x}_i permettant de faire l'estimation (n ne sera pas nécessairement le nombre de pixels de l'image).
- ▶ d est une distance comme la distance Euclidienne généralisée $d^2(\mathbf{x}, \mathbf{y}, \mathbf{H}) = (\mathbf{x} - \mathbf{y})^T \mathbf{H}^{-1} (\mathbf{x} - \mathbf{y})$. Cette distance est aussi appelée distance de Mahalanobis par la communauté *mean-shift*.
- ▶ \mathbf{H} est une matrice d'échelle ou de largeur de bande. Elle est carrée, symétrique et définie positive. Elle permet de régler les échelles entre les dimensions et il s'agit des paramètres de réglages du filtrage *mean-shift*. Pour réduire le temps de calcul, il est recommandé de pré-normaliser les données $\mathbf{x}' = \mathbf{H}^{-1/2} \mathbf{x}$. La distance calculée pour les évolutions est alors la distance Euclidienne. \mathbf{H} se mettra généralement sous la forme diagonale par bloc pour le filtrage d'images afin de spécifier indépendamment les échelles spatiales \mathbf{H}_s et d'amplitudes \mathbf{H}_r :
$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_s & 0 \\ 0 & \mathbf{H}_r \end{bmatrix}$$
- ▶ g est une fonction de pondération de \mathbb{R} dans \mathbb{R} qui se déduit du noyau utilisé pour l'estimation. Les deux noyaux les plus utilisés sont le noyau Gaussien et le noyau d'Epanechnikov qui est optimal au sens de l'erreur quadratique moyenne (g est alors la fonction porte $g(u) = 1$ si $u \leq 1$, 0 sinon). Pour le filtrage *mean-shift*, la principale contrainte pour g est d'être décroissante. Ainsi, plus deux points \mathbf{x} et \mathbf{y} sont proches, plus $g(d(\mathbf{x}, \mathbf{y}))$ est grand. Pour le filtrage d'images, g est séparé en deux composantes : une pour le spatial g_s et une pour les amplitudes g_r avec $g(\mathbf{x}) = g_s(\mathbf{x}_s) \cdot g_r(\mathbf{x}_r)$.

L'équation (8.3) correspond au calcul d'une moyenne pondérée. Puisque la pondération dépend de $\mathbf{x}^{[t]}$ et d'une mesure de similarité entre $\mathbf{x}^{[t]}$ et les autres points, on comprend que $\mathbf{x}^{[t]}$ va se déplacer en direction des échantillons qui lui sont le plus similaires. La figure 8.2 illustre cette intuition en utilisant pour g la fonction porte.

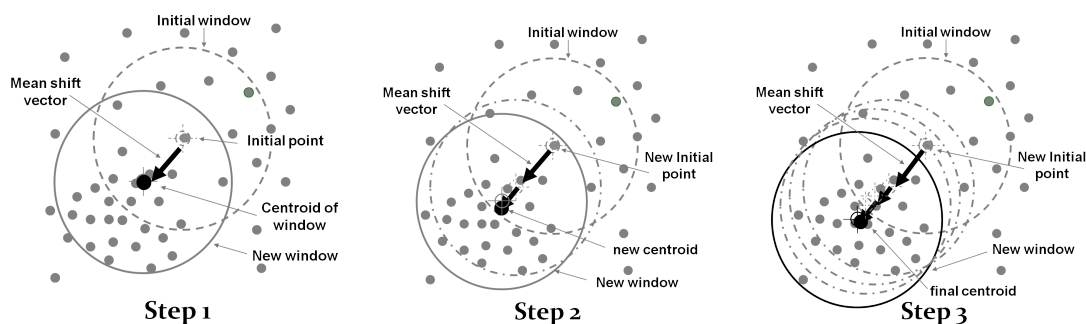


FIGURE 8.2 – Description intuitive de l'évolution itérative *mean-shift*.

8.2.3 Processus *mean-shift*

Pour filtrer des données par *mean-shift*, chaque échantillon va évoluer puis on passe à l'itération suivante et ce, tant qu'on observe un déplacement significatif dans l'espace joint spatial-amplitude. Cette description est simple mais très ambiguë : traite-t-on chaque point individuellement jusqu'à convergence en utilisant toujours les mêmes échantillons pour le calcul ou bien chaque point fait un pas et à l'itération suivante on considère les nouvelles valeurs des échantillons ?

Cette ambiguïté a été mise en évidence par (CHENG, 1995) et il a nommé respectivement ces deux processus *non-blurring* et *blurring*. Ces deux processus sont détaillés dans les algorithmes 1 et 2.

Algorithme 1 *mean-shift non-blurring***Entrée:** $\{\mathbf{x}_i\}$ avec $i = 1..n$ **Sortie:** $\{\mathbf{y}_j\}$ avec $j = 1..m$

```

for  $j = 1 \dots m$  do
   $\mathbf{y}^{[0]} \leftarrow \mathbf{y}_j$ 
   $t \leftarrow 0$ 
  repeat
     $\mathbf{y}^{[t+1]} = \frac{\sum_{i=1}^n \mathbf{x}_i \cdot g(d^2(\mathbf{y}^{[t]}, \mathbf{x}_i, \mathbf{H}))}{\sum_{i=1}^n g(d^2(\mathbf{y}^{[t]}, \mathbf{x}_i, \mathbf{H}))}$ 
     $D \leftarrow d^2(\mathbf{y}^{[t]}, \mathbf{y}^{[t+1]}, \mathbf{H})$ 
     $t \leftarrow t + 1$ 
  until  $D < \epsilon$ 
end for

```

Algorithme 2 *mean-shift blurring***Entrée:** $\{\mathbf{x}_i\}$ avec $i = 1..n$ **Sortie:** $\{\mathbf{x}_i\}$ avec $i = 1..n$

```

 $\{\mathbf{x}_i^{[0]}\} \leftarrow \{\mathbf{x}_i\}$ 
 $t \leftarrow 0$ 
repeat
   $D \leftarrow 0$ 
  for  $j = 1 \dots n$  do
     $\mathbf{x}_j^{[t+1]} = \frac{\sum_{i=1}^n \mathbf{x}_i^{[t]} \cdot g(d^2(\mathbf{x}_j^{[t]}, \mathbf{x}_i^{[t]}, \mathbf{H}))}{\sum_{i=1}^n g(d^2(\mathbf{x}_j^{[t]}, \mathbf{x}_i^{[t]}, \mathbf{H}))}$ 
     $D \leftarrow D + d^2(\mathbf{x}_j^{[t]}, \mathbf{x}_j^{[t+1]}, \mathbf{H})$ 
  end for
   $t \leftarrow t + 1$ 
until  $D/n < \epsilon$ 

```

Avec les mêmes données et les mêmes paramètres, ces deux processus ne produisent pas les mêmes résultats comme le montre la figure 8.3 sur laquelle sont représentés les déplacements itératifs de chaque échantillon filtré par *mean-shift* avec un noyau Gaussien. Dans cette illustration, chaque échantillon (les croix bleues) est dans \mathbb{R}^3 (2 coordonnées spatiales et une valeur d'intensité issue de l'image).

Il faut environ 100 itérations au processus *non-blurring* pour converger, alors que seulement 10 sont nécessaires au processus *blurring*. Le processus *blurring* converge plus rapidement et les clusters à la convergence sont plus compacts. Cependant, le processus *non-blurring*, plus facile à programmer et à paralléliser, lui a été préféré (COMANICIU et al., 2002).

La convergence pour ces processus a été étudiée dans plusieurs travaux (CARREIRA-PERPINAN, 2007; CHENG, 1995; FASHING et al., 2005; RAO et al., 2009). En 2009, (RAO et al., 2009) a montré que le processus *non-blurring* avec un noyau Gaussien minimise l'entropie croisée de Renyi dont les solutions locales sont les modes de la densité de probabilité. Quant au processus *blurring* avec un noyau Gaussien, il minimise l'entropie quadratique de Renyi et est donc par définition instable. En fait, pour le processus *blurring*, il faut absolument utiliser un noyau à support borné sinon tous les points finiront par fusionner ensemble (g ne sera jamais nul, même si la distance d est grande ce qui conduira inexorablement au regroupement de tous les points).

8.2.4 Filtrage d'images par *mean-shift*

Pour filtrer une image, quel que soit le processus *mean-shift* utilisé, les pixels filtrés verront leur position spatiale évoluer. Ainsi, il faut bien comprendre qu'après filtrage *mean-shift*, les pixels ne formeront plus une grille régulière. Afin de produire une image, il faudra donc utiliser les positions d'origine des pixels et les valeurs des intensités obtenues par filtrage. Néanmoins, il est pertinent d'utiliser la totalité de l'évolution dans l'espace joint par les traitements qui vont suivre le filtrage (souvent une segmentation).

La figure 8.4 illustre le comportement du filtrage d'images *mean-shift*. L'image (c) a été produite en affectant les coordonnées spatiales d'origine aux échantillons obtenus après filtrage *mean-shift*. On notera, par rapport à la diffusion anisotrope (b), la plus nette séparation des régions de gradients faibles (par exemple le fond à gauche et l'épaule de Lena). Sur l'image (d), on observe les déplacements des pixels et plus particulièrement la fuite des pixels du bords des régions et à l'inverse, quasi aucun déplacement dans les régions homogènes (exemple : le fond au dessus du chapeau).

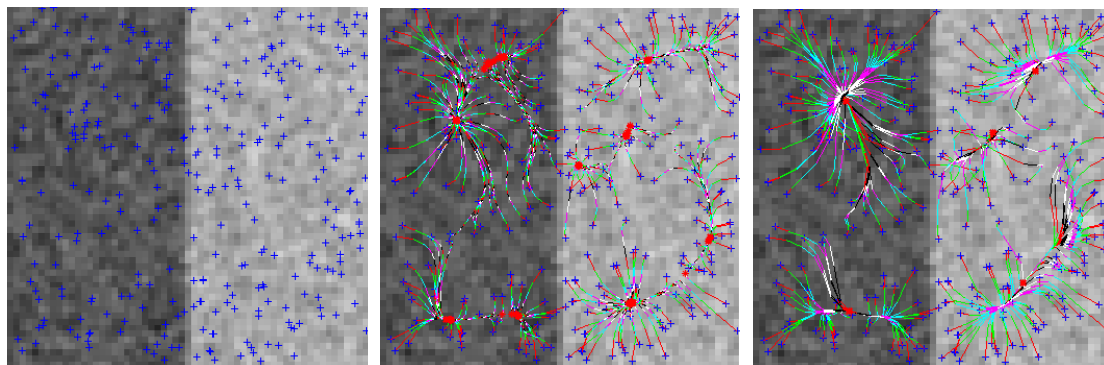


FIGURE 8.3 – Processus *mean-shift* non-blurring après 100 itérations (au centre) et blurring après 15 itérations (à droite) pour les mêmes paramètres et données (à gauche). Les croix bleues correspondent aux échantillons utilisés et filtrés, les marques rouges aux points de convergence, les segments de couleurs aux différents déplacements de chaque échantillon. Cette figure ne permet pas de voir l'évolution de l'intensité des points.

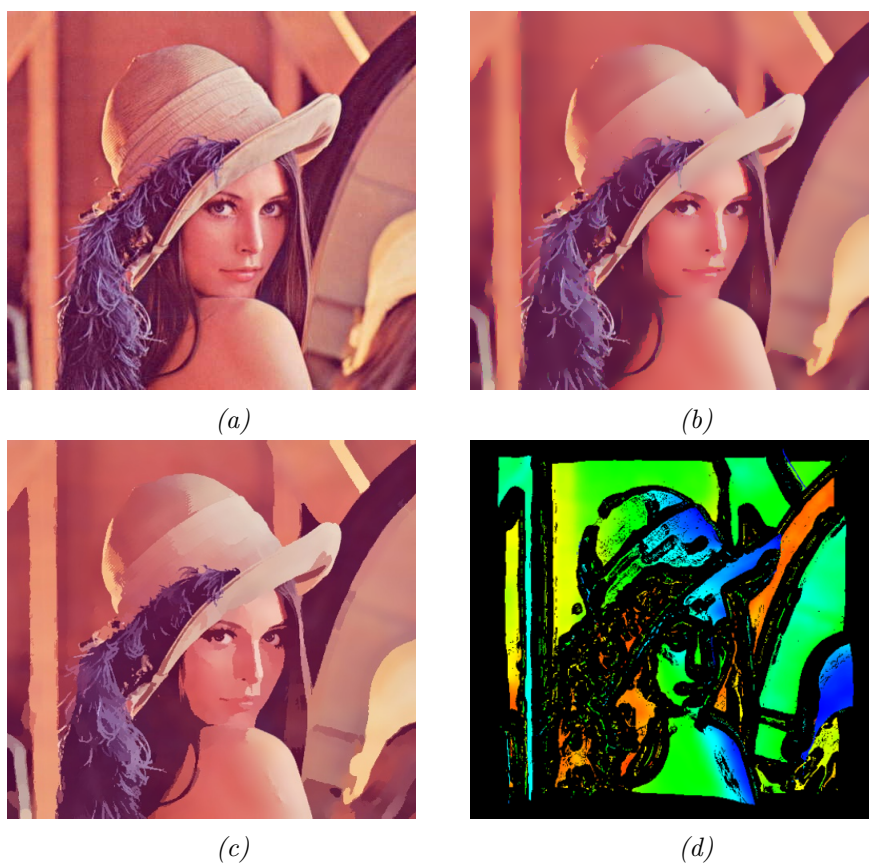


FIGURE 8.4 – Illustration du filtrage Anisotrope (b) et *mean-shift* (c) sur une image couleur Lena (a). L'image (d) permet d'observer la position des pixels après filtrage *mean-shift* (chaque pixel est un point en couleur qui représente la valeur de la composante rouge, le noir est une absence de point).

Afin de filtrer des images et obtenir les résultats escomptés, il est nécessaire de régler les paramètres d'échelles qui sont les seuls paramètres influençant fortement les résultats. Leur optimisation fait l'objet de la prochaine partie.

8.3 Optimisations des paramètres d'échelles

Le contexte applicatif des travaux de thèse de Ting Li était le filtrage et la segmentation d'images issues de plusieurs séquences IRM afin de déterminer précocement le volume de la pénombre suite à un AVC.

Le filtrage *mean-shift* était tout indiqué pour son efficacité et sa capacité à utiliser des espaces de caractéristiques quelconques sans *a priori* ou supervision. Cependant, afin de garantir la robustesse de l'approche, nous avons étudié plusieurs optimisations du filtrage *mean-shift* dont : le choix de l'espace de représentation des caractéristiques (espaces couleurs : (LI et al., 2011)), le sous-échantillonnage aléatoire de l'image et la détermination des paramètres échelles en vue de l'optimisation de la segmentation. Cette dernière étude est résumée ici où seulement l'optimisation du *PSNR* est considérée. Les métriques d'évaluation sont données en annexes A.

Il s'agit plus précisément d'étudier la forme et le choix des valeurs de la matrice d'échelle \mathbf{H} . Pour nous, cette matrice et le noyau sont exprimés de manière séparable et indépendante en spatial et en amplitude. On peut ainsi fixer l'un ou l'autre des paramètres d'échelles et étudier l'impact de l'autre paramètre sur les résultats du filtrage comme le montre la figure 8.5 où une image de synthèse IRM (représentant trois grandeurs IRM (T1, T2 et densité de protons) des différents tissus du cerveau) de taille 256 x 256 pixels est corrompue par un flou et du bruit puis filtrée par approche *mean-shift* avec différents paramètres d'échelles.

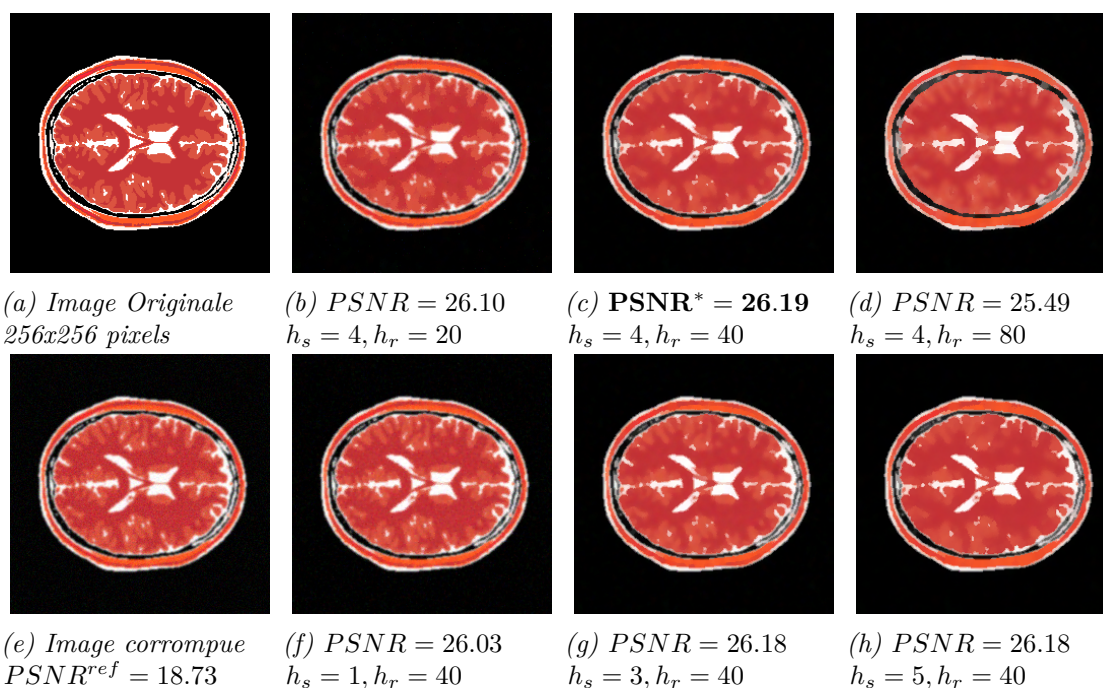


FIGURE 8.5 – Filtrage *mean-shift* de l'image de synthèse IRM avec différents paramètres d'échelles. (A) image originale de 256 x 256 pixels; (E) image corrompue avec un flou gaussien ($\sigma = 1$ rayon du filtre de 63 pixels) et bruit gaussien ($\sigma = 10$); (B), (C), (D), (F), (G) et (H) montrent différents résultats obtenus en faisant varier h_s et h_r . Pour chaque image, le *PSNR* avec l'image d'origine est donné. Le cas (C) correspond au filtrage optimal par rapport au *PSNR*.

On notera que le *PSNR* évolue peu dans la plage des paramètres d'échelles utilisée dans la figure 8.5. Cette évolution est illustrée dans la figure 8.6. Ce type d'évolution a été constatée dans toutes les images étudiées.

Une première analyse de ces résultats conduit à privilégier les petites valeurs de h_s (de 1 à 7) et à chercher à optimiser les valeurs de \mathbf{H}_r .

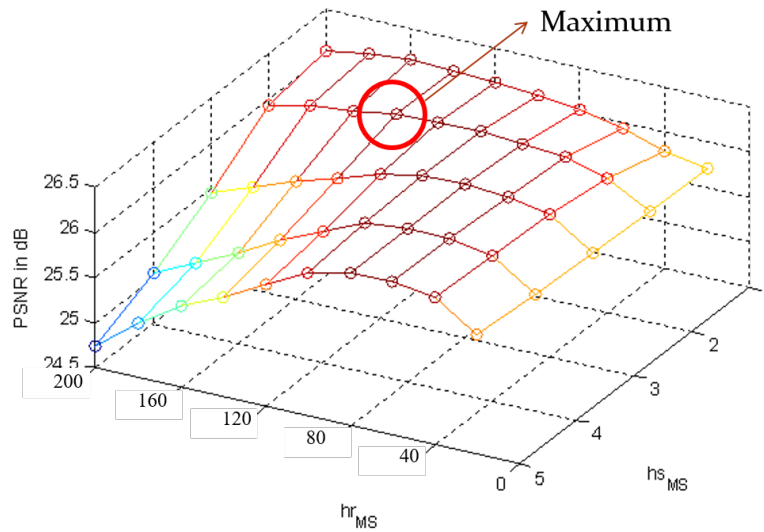


FIGURE 8.6 – Evolution du PSNR en fonction de h_s et h_r

En s'appuyant sur les travaux de (DUONG, 2003 ; DUONG, 2007 ; WAND et al., 1995), il est possible de déterminer, au sens de l'erreur quadratique moyenne intégrée, les valeurs optimales de la matrice d'échelle dans le cadre de l'estimation non paramétrique (et non du filtrage *mean-shift*). Cette méthode s'appelle la méthode du *plug in*.

Nous avons confronté les résultats obtenus avec la méthode du *plug in* avec ceux issus d'une optimisation par recherche exhaustive des paramètres dans le cadre du filtrage *mean-shift* d'images couleurs naturelles (Lena, Mandrill, Pepper, Lake) et de deux images de synthèse (Tito et IRM) corrompues par du flou et du bruit.

Pour cela, nous avons considéré des valeurs fixées h_s de [1; 5] et 3 formes pour le calcul de \mathbf{H}_r par l'approche du *plug in* :

- PI_{1D} lorsque le calcul du *plug in* est effectué pour chacune des trois composantes indépendamment (\mathbf{H}_r est diagonale),
- PI_{Diag} lorsque \mathbf{H}_r est optimisé par le *plug in* en utilisant simultanément les trois composantes d'amplitude (\mathbf{H}_r est diagonale).
- PI_{Full} lorsque \mathbf{H}_r est optimisé par le *plug in* en utilisant les neuf paramètres.

Pour la recherche exhaustive, les paramètres h_s et h_r (scalaires) sont optimisés simultanément.

Dans les travaux de thèse de Ting Li, l'optimisation a été faite avec deux mesures : le PSNR (présentée dans la table 8.1) et le SSIM de (WANG et al., 2004b). Ici, seuls les résultats avec le PSNR sont présentés.

On observe qu'une bonne alternative à la recherche exhaustive (extrêmement coûteuse en ressource (LI et al., 2010a)) est le PI_{Diag} . Sans pouvoir se généraliser à toutes les applications de filtrage d'images, le *plug in* semble pouvoir déterminer efficacement des paramètres pertinents. Cette étude a aussi permis de valider la séparation h_s et h_r pour l'optimisation, ainsi que l'utilisation de la forme diagonale de \mathbf{H}_r dans le cas d'images avec trois composantes d'amplitude (images couleurs (LI et al., 2011) ou 3 séquences IRM).

L'objectif est maintenant d'obtenir une segmentation à partir du filtrage *mean-shift*.

8.4 Du filtrage *mean-shift* à la segmentation d'images

Le filtrage *mean-shift* n'inclut pas de segmentation ni d'étiquetage des points. Il rapproche itérativement les points qui se ressemblent. A la convergence, les points très similaires

<i>Images</i>	PSNR (en dB)				
	$PSNR^{ref}$	PI_{1D}	PI_{Diag}	PI_{Full}	$PSNR_{RE}^*$
<i>Lena</i>	27.62	27.65	37.03	36.65	37.05
<i>Mandrill</i>	28.13	28.16	31.80	31.70	31.95
<i>Pepper</i>	28.13	28.05	36.41	36.06	36.48
<i>Lake</i>	28.06	28.08	34.89	34.47	34.95
<i>Tito</i>	27.58	27.58	33.21	33.19	33.88
<i>MRI</i>	25.70	25.70	26.18	26.04	26.20

TABLE 8.1 – Comparaison des valeurs optimales de PSNR (en dB) obtenues à l'aides du plug in et de la Recherche Exhaustive des paramètres d'échelle. $PSNR^{ref}$ correspond à la valeur du PSNR de l'image corrompue.

partagent des positions si proches que la tâche de regroupement (ou *clustering*) est grandement facilitée... c'était l'objectif de cette étape de filtrage. En utilisant les positions spatiales d'origine des échantillons, ce clustering permettra d'obtenir une segmentation de l'image.

Plusieurs approches permettent de fusionner les points ayant convergés et ainsi effectuer le clustering. Parce qu'elles nécessitent de fixer le nombre de cluster k , les approches de type k-moyennes (et ses dérivées) sont rarement utilisées (CABRIA et al., 2012). Les deux approches les plus pertinentes sont :

- la fusion des échantillons proches lors des itérations *mean-shift* (cette approche sera détaillée plus loin),
- une approche de type croissance de région après le filtrage *mean-shift* dans l'espace joint spatial-amplitude.

8.4.1 Croissance de région pour l'espace joint

La croissance de région a été introduite par (ZUCKER, 1976). Il s'agit d'un processus itératif, rapide et intuitif initialement dédié à la segmentation d'images. La segmentation d'images a pour objectif d'associer à chaque pixel \mathbf{x}_i une région R identifiée par une étiquette l .

Le principe de la croissance de région est simple. Autours d'une région R à l'itération t , on cherche les points qui vérifient un critère de similarité en intensité. Ces points sont ajoutés à la région pour l'itération suivante (figure 8.7). Ce processus est répété tant que la région évolue.

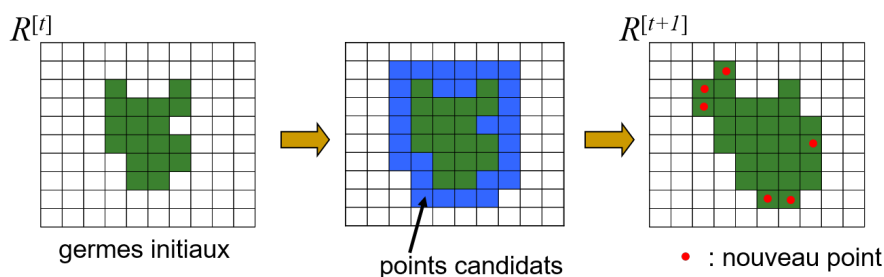


FIGURE 8.7 – Processus de croissance de région. Dans cet exemple, les points en rouge valident le critère de similarité en intensité et sont ajoutés à la région pour l'itération suivante.

Au démarrage, la région initiale est constituée de quelques points souvent nommés *germes*.

Il est possible d'utiliser de nombreux critères de similarité et de modéliser ce processus sous une forme variationnelle (ROSE et al., 2010).

Il est aussi possible de fusionner la notion de similarité d'intensité et de proximité spatiale en utilisant, comme pour les *mean-shift*, l'espace joint spatial amplitude.

Ainsi, le voisinage d'un point \mathbf{y} parmi l'ensemble des points dans X s'exprime : $\mathcal{N}(\mathbf{y}; \epsilon) = \{\mathbf{x} \in X | d(\mathbf{y}, \mathbf{x}) \leq \epsilon\}$, avec d une distance et ϵ un seuil sur cette distance permettant de définir une hypersphère contenant l'ensemble des points voisins (figure 8.8) dans l'espace.

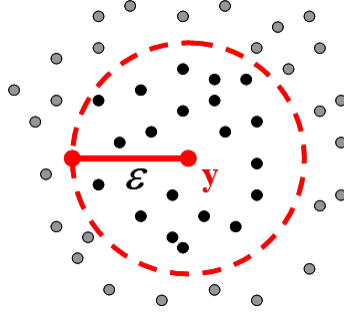


FIGURE 8.8 – Illustration du voisinage de \mathbf{y} dans un ensemble de points. Les points noirs correspondent aux voisins de \mathbf{y} à une distance ϵ .

En utilisant la distance Euclidienne généralisée, plus adaptée aux mélanges des grandeurs et dynamiques, cette expression peut se mettre sous la forme :

$$\mathcal{N}(\mathbf{y}; \epsilon) = \{\mathbf{x} \in X | (\mathbf{y} - \mathbf{x})^T \epsilon^{-1/2} \mathbf{I} (\mathbf{y} - \mathbf{x}) \leq 1\} \quad (8.4)$$

puis, plus généralement :

$$\mathcal{N}(\mathbf{y}; \epsilon) = \{\mathbf{x} \in X | d(\mathbf{y}, \mathbf{x}; \mathbf{H}_{RG}) \leq 1\} \quad (8.5)$$

Ceci permet de *clusteriser* des données filtrées par *mean-shift* en utilisant des paramètres d'échelles de croissance de région \mathbf{H}_{RG} dans le même espace que ceux des *mean-shift* \mathbf{H}_{MS} . Ainsi, pour la croissance de région, on notera \mathbf{H}_{sRG} le paramètre spatial et \mathbf{H}_{rRG} le paramètre pour les amplitudes.

8.4.2 Approche MS+RG

Le filtrage par *mean-shift* suivi de la segmentation par croissance de région (MS+RG) dans l'espace joint se place dans la chaîne de traitements illustrée sur la figure 8.9.

Une étape de pré-traitements effectue les recalages des différentes composantes d'amplitude ainsi que leur normalisation. Cette étape est utile en IRM où les séquences ne sont pas parfaitement alignées et où la dynamique des intensités peut grandement changer d'un patient à l'autre.

Les post-traitements consistent en une étape d'érosion et dilatation (ouverture morphologique avec un élément structurant de rayon 2) sur la segmentation afin de supprimer les points orphelins et lisser les contours des régions.

La figure 8.10 montre deux résultats de segmentation de la matière blanche sur l'image IRM de synthèse corrompue par un flou et un bruit gaussien obtenus avec des paramètres d'échelles différents. Dans cet exemple, les amplitudes normalisées lors du pré-traitement permettent d'utiliser des matrices d'échelles en amplitude \mathbf{H}_{rMS} et \mathbf{H}_{rRG} de la forme : $\mathbf{H}_{rMS} = hr_{MS} \cdot \mathbf{I}$ et $\mathbf{H}_{rRG} = hr_{RG} \cdot \mathbf{I}$. Il en sera de même pour le spatial où les pixels sont isotropes.

La table 8.2 précise les intervalles et les pas de recherche pour les paramètres d'échelles.

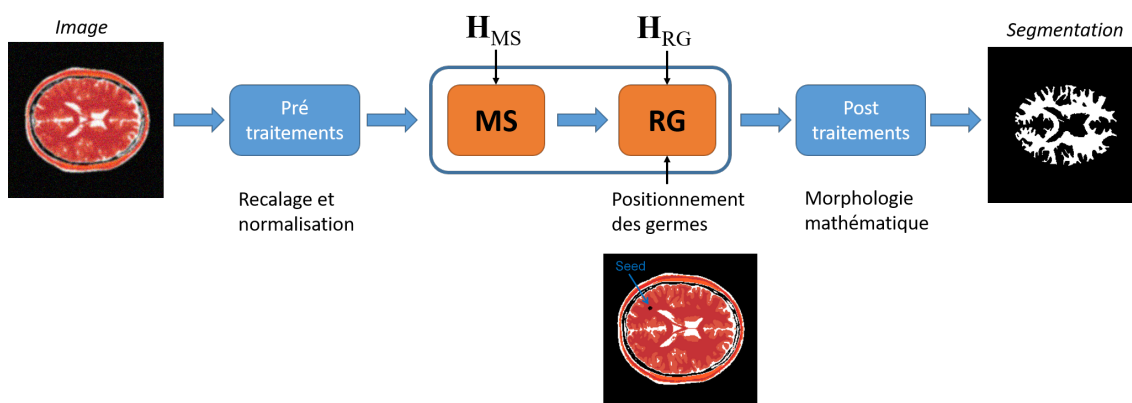


FIGURE 8.9 – Pipeline de segmentation pour l'approche MS+RG.

Paramètres	Intervalles	Pas
hs_{MS}	1 à 5	0.5
hr_{MS}	20 à 100	20
hs_{RG}	1 à 3	1
hr_{RG}	2 à 10	2

TABLE 8.2 – Plages des paramètres d'échelles pour la recherche exhaustive.

La méthode de croissance de région est initialisée avec 9 germes dans la matière blanche. Les couleurs sur les images (b) et (c) permettent d'évaluer visuellement la sur- et sous-segmentation de la matière blanche, respectivement en rouge et en blanc. L'évaluation quantitative est fait avec le score de DICE (DICE, 1945).

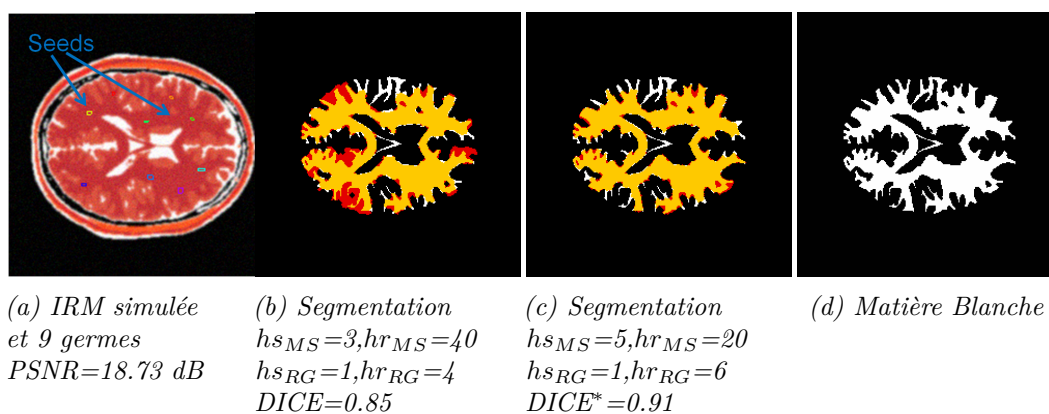


FIGURE 8.10 – Segmentation par l'approche MS+RG de la matière blanche sur l'image IRM simulée et corrompue avec du flou et du bruit gaussiens.

Les qualités de segmentation dépendent grandement du choix des 4 paramètres d'échelles. Ici, la segmentation optimale avec un DICE de 0.91 a été obtenue avec une recherche exhaustive des 4 paramètres. Cependant, les valeurs de ces paramètres sont propres à cette image et à la gamme de perturbations apportées, ainsi qu'à la segmentation de la matière blanche.

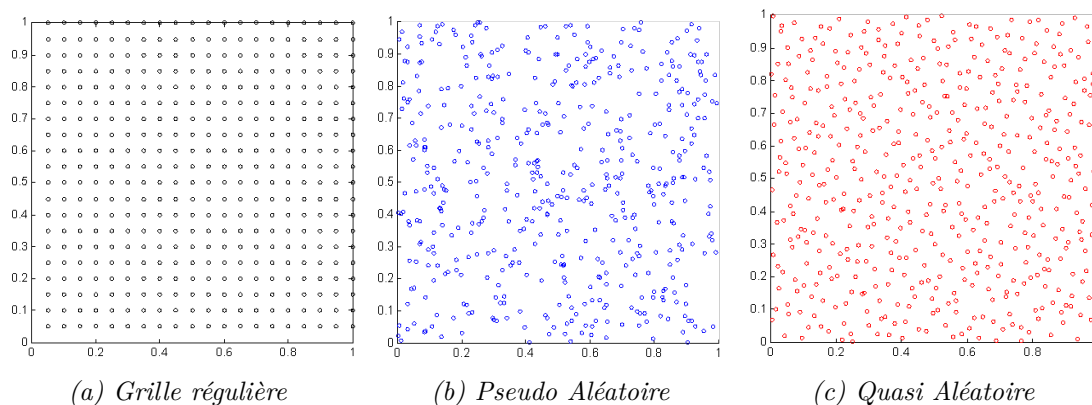


FIGURE 8.11 – Illustration des trois stratégies d'échantillonnage des plages de paramètres d'échelles.

8.4.3 Échelles pour MS+RG et carte de probabilités

La recherche exhaustive précédente est basée sur un échantillonnage régulier de l'espace des 4 échelles à optimiser et a nécessité 750 filtrages et segmentations pour déterminer les valeurs optimales des échelles. Ceci est long mais surtout nécessite une référence pour trouver l'optimalité et malheureusement aucune règle ni lien avec le *plug in* n'ont été observés.

Cependant, à partir des remarques faites précédemment sur la stabilité des résultats pour des paramètres d'échelles dans une plage donnée (voir 8.6), l'idée a été de moyenniser les résultats de segmentation obtenus avec les différents quadrinômes de paramètres d'échelles et de construire ainsi une carte de probabilité. Il faudra ensuite seuiller cette carte (un seul paramètre utilisateur, traitement très rapide) afin obtenir la segmentation finale.

Cependant, 750 filtrages et segmentations sont toujours nécessaires avec la stratégie d'échantillonnage par grille régulière.

Afin de diminuer le temps de calcul en réduisant ce nombre de filtrages et segmentations sans diminuer d'avantage les plages ou les pas d'exploration des paramètres, nous avons étudié l'influence de deux tirages aléatoires (pseudo aléatoire uniforme et quasi aléatoire (KENG et al., 1981 ; KUIPERS, 1974 ; NIEDERREITER, 1992)) comme stratégies d'échantillonnage des paramètres d'échelles.

Ces tirages aléatoires apportent aussi une solution au problème de distribution massive des calculs. En effet, le temps de calcul nécessaire aux 750 filtrages et segmentations est important mais peut être très fortement réduit s'ils sont exécutés en parallèle. Nous nous sommes appuyés sur la plateforme VIP¹ pour distribuer leur exécution sur de nombreuses grilles de calcul. Avec un très grand nombre de processeurs disponibles, et une distribution complète des 750 travaux, le temps d'obtention des résultats pourrait être juste supérieur à celui du travail le plus long. Malheureusement, dans la pratique, le temps d'exécution d'un calcul sur les grilles n'est pas déterministe et certains peuvent être extrêmement retardés voir annulés bloquant l'obtention du résultat final. Mais en tirant aléatoirement un très grand nombre de quadrinômes de paramètres d'échelles, plus que nécessaire, il sera possible d'obtenir le résultat final quand un nombre suffisants de calculs seront finis et non la totalité (LI et al., 2010a).

Les trois stratégies d'échantillonnage étudiées sont illustrées sur la figure 8.11.

1. *Virtual Imaging Platform* du laboratoire Creatis <https://www.creatis.insa-lyon.fr/vip/>

Pour les deux stratégies aléatoires, les paramètres d'échelles ont été choisis dans les intervalles donnés dans la table 8.3. Pour la stratégie d'échantillonnage régulier, les plages des paramètres sont les mêmes que précédemment (table 8.2).

Paramètres	Intervalles
hs_{MS}	1 à 5
hr_{MS}	5 à 100
hs_{RG}	1, $\sqrt{2}$, 2, $\sqrt{3}$, 3
hr_{RG}	1 à 10

TABLE 8.3 – Intervalles des paramètres d'échelle pour les deux stratégies de tirage aléatoire.

Avec ces plages de paramètres, les cartes de probabilités obtenues pour 750 calculs, ainsi que le seuillage optimal des cartes de probabilités pour chacune des approches, sont données sur la figure 8.12.

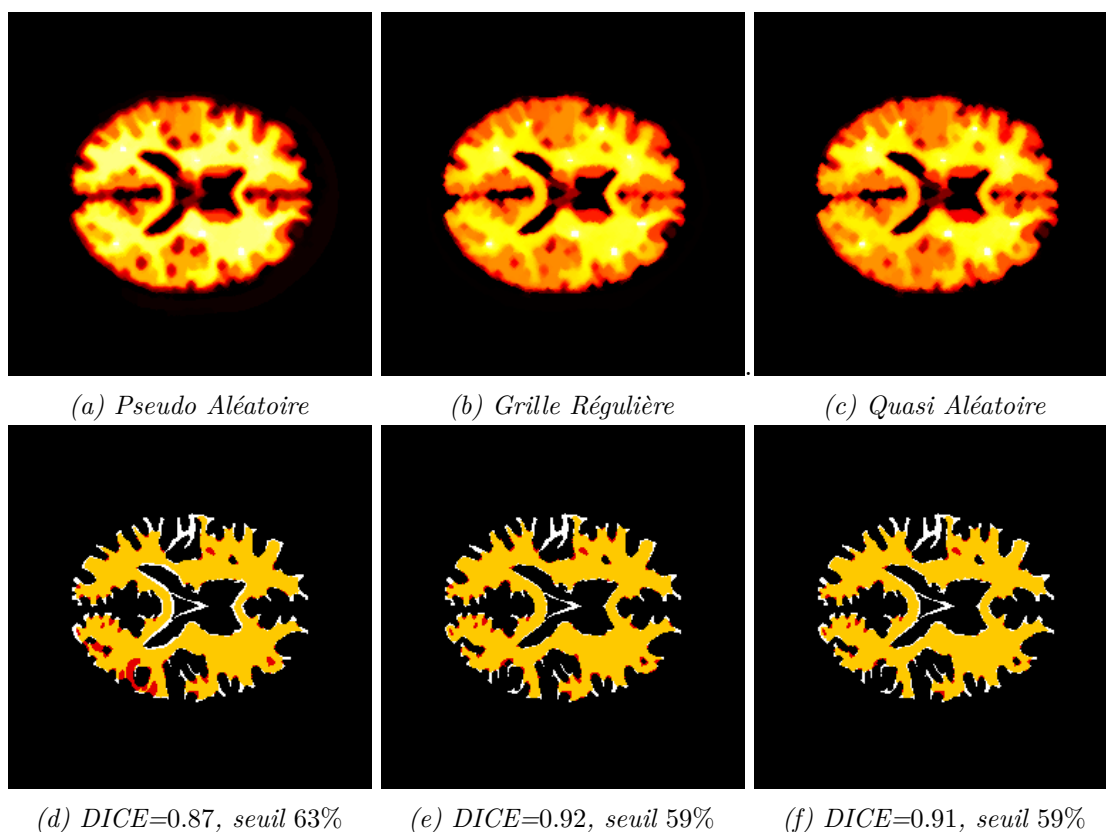


FIGURE 8.12 – Cartes de probabilités et leurs seuillage pour les 3 différents stratégies d'échantillonnage des paramètres d'échelle.

Les résultats obtenus avec 100, 500 et 750 filtrages et segmentations pour les tirages aléatoires et 750 calculs avec la grille régulière sont présentés sur la table 8.4. Pour les tirages aléatoires, les expériences ont été répétées 30 fois pour calculer les moyennes et les écarts types du *DICE*. Le temps moyen d'un calcul de filtrage et segmentation est aussi indiqué (tous les calculs ont été faits sur la même machine).

Dans cet exemple, la stratégie de grille régulière est la plus performante mais aussi la plus longue. Pour la suite de ce manuscrit, et notamment l'application à l'AVC, nous utiliserons

la stratégie d'échantillonnage quasi aléatoire plus rapide et dont les résultats sont très proches de ceux de la grille régulière.

	nb_{h_b}	\bar{t}_{MSRG} (en minutes)	$DICE^*$
Grille Régulière	750	10	0.92
	100	4.38	0.8646 ± 0.0094
Pseudo Aléatoire	500	4.12	0.8707 ± 0.0027
	750	3.6	0.8721 ± 0.0008
Quasi Aléatoire	100	4.29	0.9097 ± 0.0083
	500	3.79	0.9135 ± 0.0020
	750	3.52	0.9137 ± 0.0003

TABLE 8.4 – Scores de $DICE$ suite au seuillage optimal des cartes de probabilités sur l'IRM corrompue pour les trois stratégies différentes d'échantillonnage et pour trois nombres d'échantillons. \bar{t}_{MSRG} est le temps moyen en minute pour un filtrage et segmentation.

8.5 Application à l'accident vasculaire cérébral

L'accident vasculaire cérébral de type ischémique est une maladie grave qui nécessite une prise en charge très rapide pour limiter ses conséquences (GONZALEZ, 2011) (figure 8.13 gauche). Un des traitements consiste, par des moyens chimiques ou mécaniques, à désobstruer le vaisseau responsable de l'accident pour assurer la re-perfusion des tissus cérébraux. Dans ce contexte, une identification et une prédiction des tissus à risque d'infarctus serait une aide précieuse à la décision clinique.

Ces tissus à risque sont situés dans la zone de pénombre (figure 8.13 droite) qu'il est possible d'observer peu de temps après la survenue de l'AVC en IRM via notamment les séquences de diffusion (DWI) et de perfusion (PWI) (ASHOK et al., 2006 ; OSTERGAARD, 2005 ; ZAFAR, 2008).

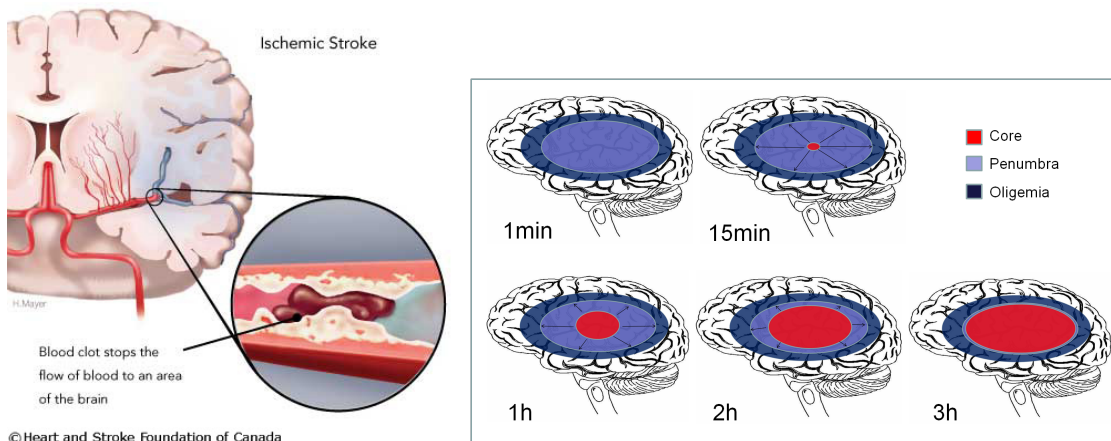


FIGURE 8.13 – Ischémie cérébrale (à gauche) et son évolution (à droite).

Nous avons utilisé la DWI et l'ADC (*Apparent Diffusion Coefficient*) avec des paramètres quantitatifs et semi-quantitatifs issus de l'imagerie PWI en IRM. On parlera aussi de cartographies quand toute une image est créée avec les valeurs de ces paramètres. On se focalise ici sur les cinq paramètres suivants :

- CBV (*Cerebral Blood Volume*) : le volume sanguin des capillaires et des veinules cérébraux par volume de tissu cérébral,
- CBF (*Cerebral Blood Flow*) : le volume sanguin délivré à une unité de masse de tissu définie par unité de temps,
- TTP (*Time To Peak*) : délai au maximum de l'effet créé par le traceur,
- peak : valeur du maximum de l'effet créé par le traceur,
- MTT (*Mean Transit Time*) : le temps moyen qu'il faut au traceur pour parcourir une région donnée du cerveau. Ce temps de transit dépend de la distance parcourue entre l'entrée artérielle et la sortie veineuse et est lié au CBV et au CBF selon le principe du volume central (OSTERGAARD, 2005), qui stipule que : $MTT = CBV/CBF$.

Différentes combinaisons de ces paramètres ont été testées en s'appuyant sur le formalisme de l'espace des caractéristiques. Les vecteurs \mathbf{x}_i étant construits ainsi pour le filtrage et la croissance de région :

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_r \end{bmatrix} = \begin{bmatrix} x_x \\ x_y \\ x_z \\ x_{ADC} \\ \vdots \\ x_{MTT} \\ x_{CBV} \end{bmatrix} \quad (8.6)$$

8.5.1 Résultats sur le rat

Grâce à Marlène Wiart et Fabien Chauveau, nous avons pu disposer de quatre rats pour lesquels un AVC a été provoqué par un procédé chirurgical et qui ont ensuite été imagés par IRM au jour 0, puis au jour 1 et jour 12. Deux de ces rats ont été reperfusés afin de se rapprocher du traitement chez l'homme. Toutes les données ont été recalées sur le jour 0. Les images IRM pondérées T2 réalisées au jour 1 et 12 ont permis de faire la segmentation experte de la zone de nécrose finale (1 seul rat à survécu jusqu'au jour 12). C'est cette zone qu'il faut être capable de prédire dès le jour 0. Les acquisitions IRM en DWI et PWI ne sont disponibles qu'au jour 0.

Les germes pour la segmentation ont été placés manuellement dans le coeur de l'ischémie à l'aide de la cartographie TTP. La figure 8.14 illustre les cartographies d'ADC, TTP et *peak* pour un rat, ainsi que les segmentations optimales obtenues par croissance de région (notée *onlyRG* sur la figure) et par filtrage *mean-shift* suivi de la croissance de région (MS+RG) en utilisant les cartographies ADC et *peak*. L'optimisation a été faite sur les paramètres d'échelles (échantillonnage par grille régulière) en cherchant à maximiser le score de DICE. La table 8.5 donne les scores de DICE optimaux obtenus sur les quatre rats.

On observe que les résultats sont meilleurs avec l'utilisation des cartographies d'ADC et de *peak* et que le filtrage *mean-shift* avant la croissance de région permet d'améliorer le score DICE. Les valeurs de DICE optimales sont assez encourageantes pour l'application. On calcule maintenant les cartes de probabilités en utilisant un échantillonnage pseudo aléatoire ce qui évite de régler les quatre paramètres d'échelles. Pour le même rat, les cartes de probabilités sont données sur la figure 8.15.

En seillant ces cartographies, seul réglage qui reste à faire à l'opérateur, le meilleur DICE que l'on peut obtenir pour le rat 'pMCAO2' est de 0,80 avec les cartographies ADC et *peak*. Ce résultat est proche de ceux obtenus avec un réglage optimal des paramètres d'échelles (table 8.5 : 0.81).

Nous avons ensuite appliqué cette approche à l'homme.

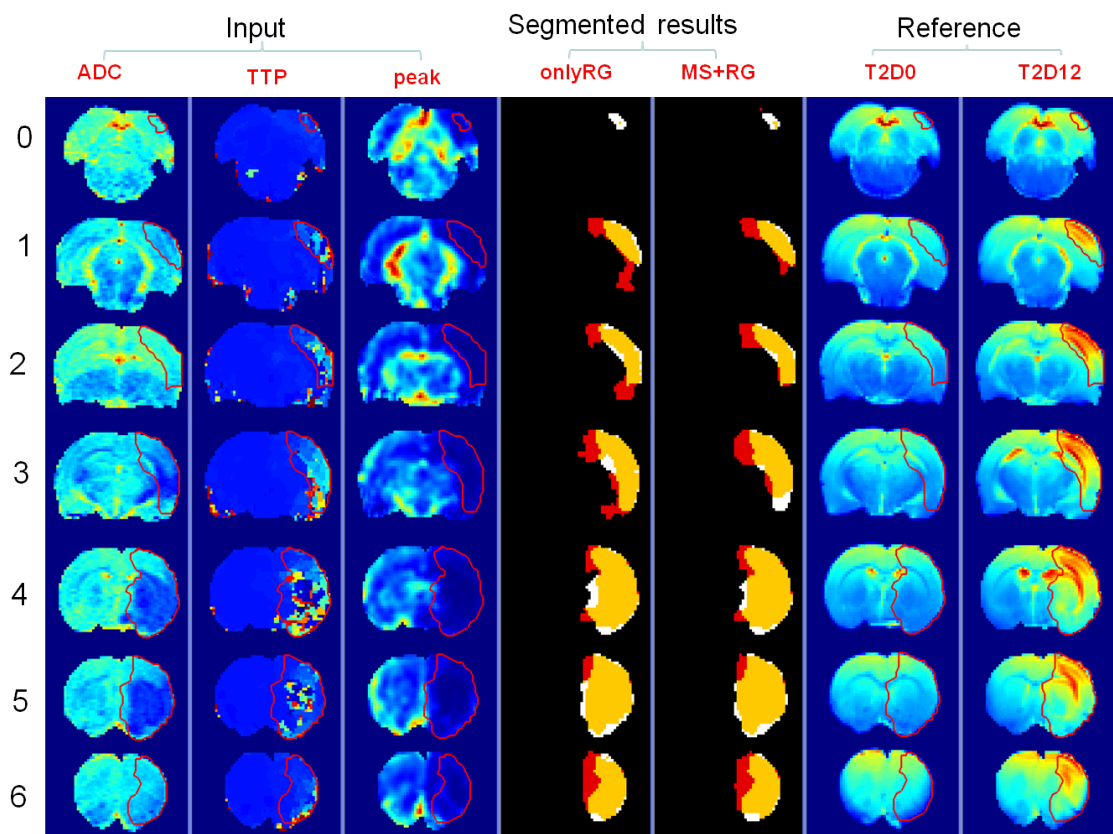


FIGURE 8.14 – Segmentations optimales obtenues avec les paramètres ADC et peak sur 7 coupes du rat 'pMCAO2'.

Rat	Méthode	DICE*		
		ADC, peak	ADC, TTP	ADC, peak, TTP
<i>pMCAO2</i>	RG seule	0.67	0.29	0.59
	MS+RG	0.81	0.66	0.65
<i>pMCAO4</i>	MS+RG	0.86	0.72	0.78
<i>dha81</i> (r)	MS+RG	0.79	0.49	0.74
<i>dha82</i> (r)	MS+RG	0.81	0.53	0.76

TABLE 8.5 – Influence des combinaisons de paramètres IRM sur les segmentations produites par la croissance de région (RG) et l'approche MS+RG. Les deux derniers rats marqués avec un (r) sont les deux rats avec reperfusion.

8.5.2 Résultats sur l'homme

Nous avons pu utiliser la base de données I-KNOW qui rassemblait, au moment de notre étude, 80 patients atteints de différentes formes d'ischémie cérébrale. Nous nous sommes intéressés qu'aux patients atteints d'ischémie non-hémorragique. Contrairement aux rats, tous les patients ont été traités et généralement, ils reperfusionnent. Cependant, nous avons identifié deux patients pour lesquels il n'y a pas eu de reperfusion : le patient 171 et le patient 255.

Concernant les données IRM disponibles, au jour 0, les séquences T2 FLAIR, DWI et PWI étaient faites, puis une IRM T2 FLAIR au 30^e jour (figure 8.16). Ainsi nous disposons des

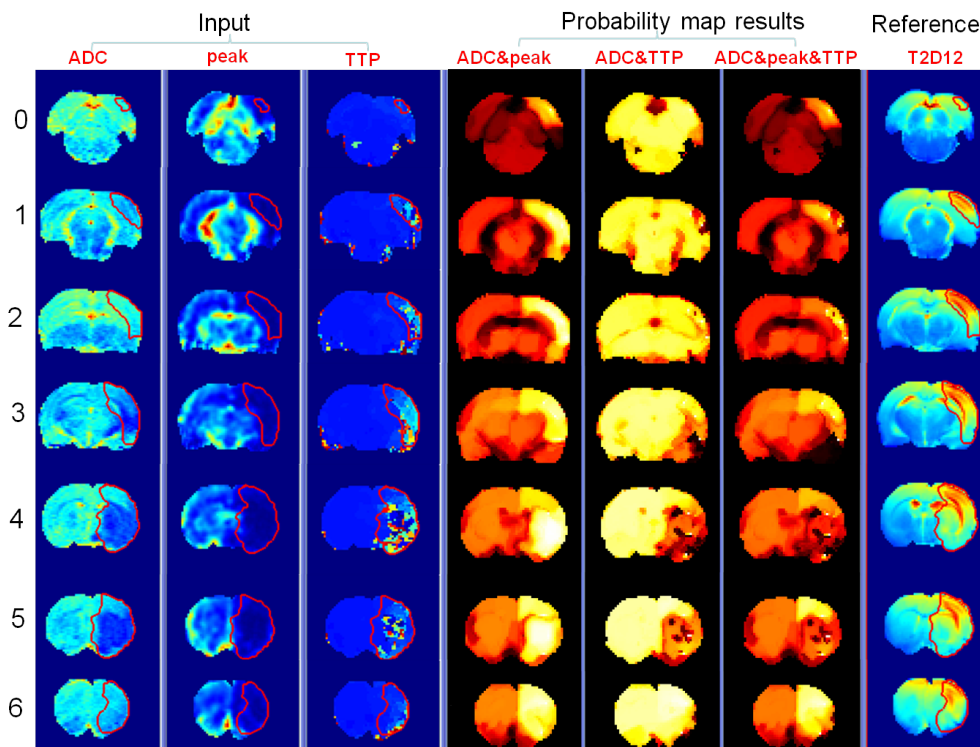


FIGURE 8.15 – Cartes de probabilités de segmentation de la pénombre du rat 'pMCAO2' pour trois combinaisons de cartographies IRM.

mêmes données que pour l'étude des rats.

La segmentation manuelle du coeur de l'ischémie au jour 0 sur l'IRM T2 FLAIR a servi de germes aux méthodes de croissance de région. De même, la zone nécrosée finale qui servira de référence, a été délimitée sur l'IRM T2 FLAIR acquise au jour 30.

Sur la figure 8.17 on observe les cartes de probabilités obtenues en utilisant les cinq séquences IRM (TTP, MTT, CBF, ADC et DWI). Les meilleurs *DICE* obtenus sont entre 0.50 et 0.55 ce qui n'est pas acceptable.

Malgré l'optimisation du choix des séquences IRM et des paramètres, nous n'avons pas pu obtenir des résultats convenables sur les deux patients. De plus, dans le cas de ces deux patients sans reperfusion, nous sommes dans un cas idéal pour l'estimation de la zone de la pénombre par l'approche des cartes de probabilités.

8.6 Conclusion

Dans ce chapitre nous avons introduit les principales notions sur le filtrage *mean-shift* : l'espace des caractéristiques (domaine joint spatial et amplitude), les paramètres d'échelles et le formalisme *mean-shift* (LI et al., 2011).

Nous avons aussi cherché à optimiser les paramètres d'échelles soit au sens des statistiques (avec la méthode du *plug in*), soit avec une recherche exhaustive (LI et al., 2010a ; REVOL-MULLER et al., 2012a). Puis, afin de s'affranchir de la référence nécessaire à cette optimisation, le concept de cartes de probabilités a été introduit et appliqué à des données de synthèses et à des données réelles en IRM pour l'AVC.

Les résultats chez le rat ont montré l'intérêt de ces méthodes en produisant des segmentations de bonnes qualité de la zone de pénombre.

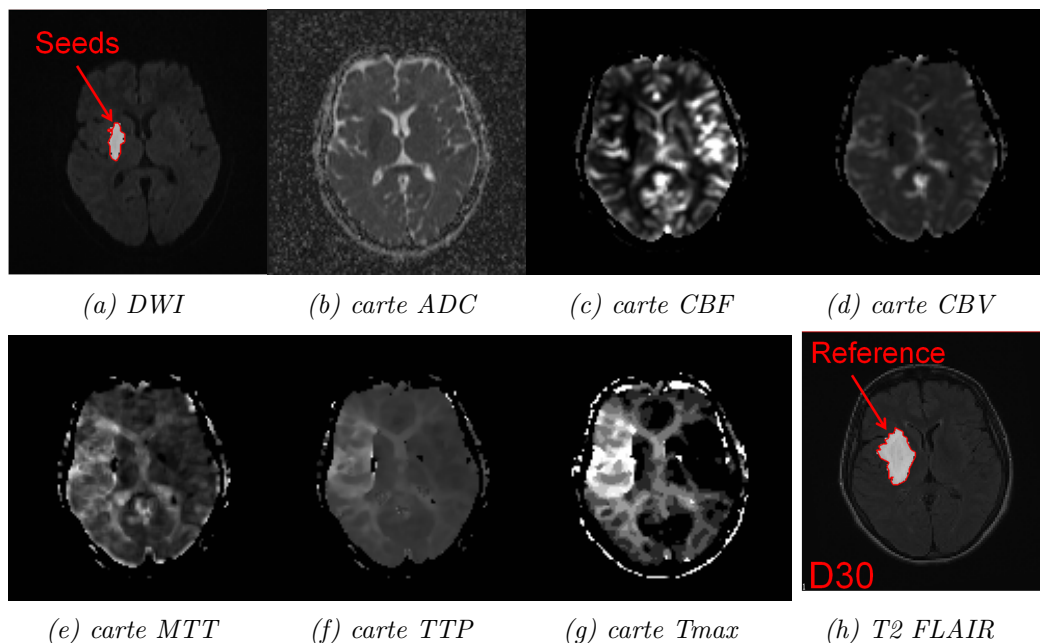


FIGURE 8.16 – Illustrations des séquences IRM pour un AVC chez l'homme (11^e coupe du patient 171).

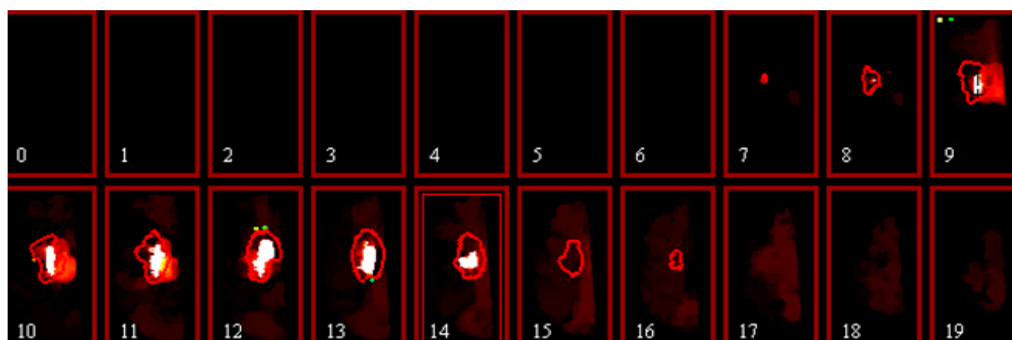


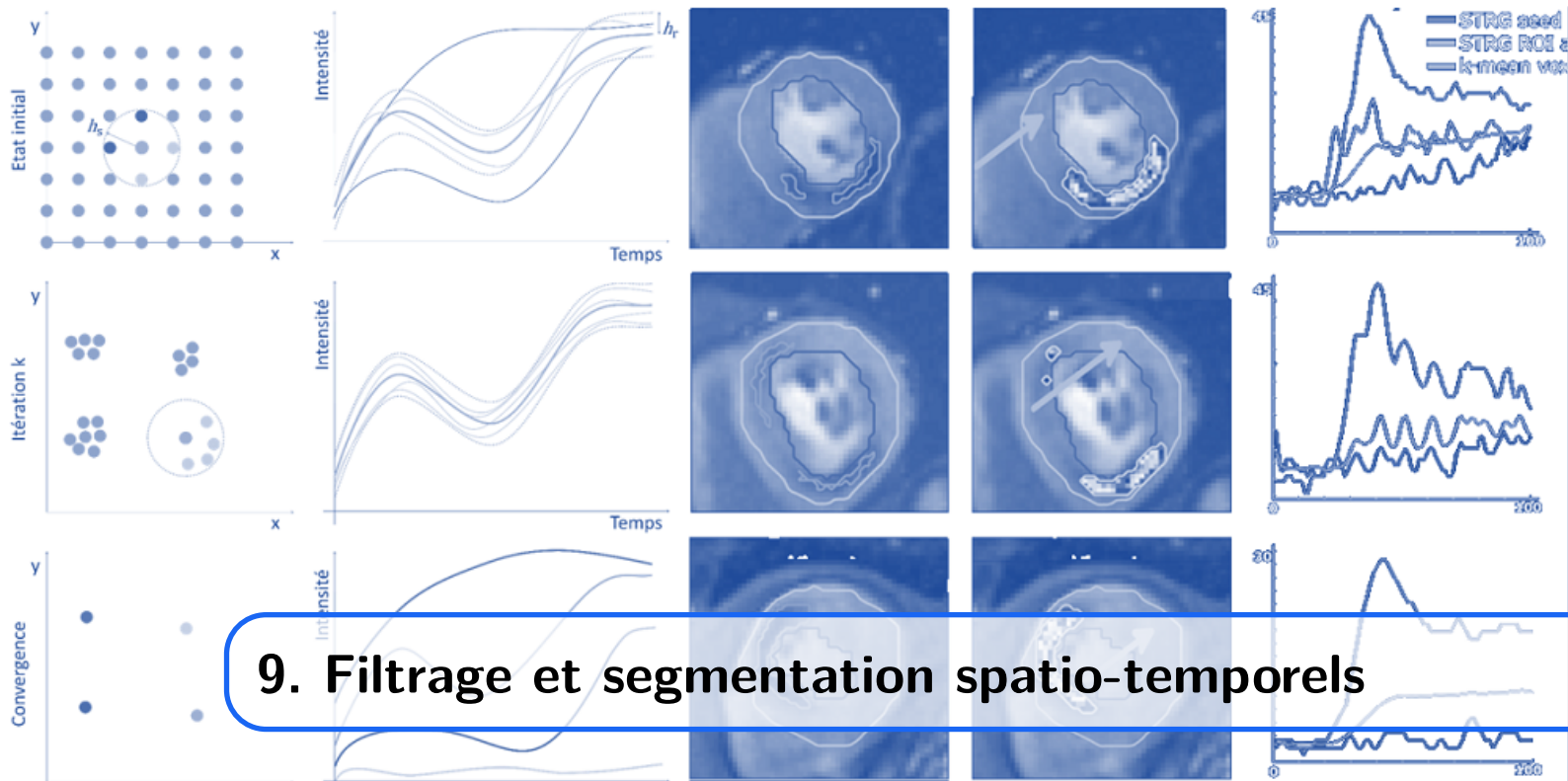
FIGURE 8.17 – Cartes probabilités sur les 20 coupes du patient 171 calculées avec les cartographies IRM : TTP, MTT, CBF, ADC et DWI. La référence à 30 jours est donnée pour chacune des coupes.

Cependant, transférées à l'homme et malgré de nombreuses expériences, les cartes de probabilités n'ont pas réussi à donner avec un bon taux de succès une segmentation acceptable en terme de *DICE*.

Ceci est principalement dû au points suivants

- la diversité des AVC : lieux de l'ischémie, taille des vaisseaux impactés, antécédents du patient, ...
- l'instant de l'AVC par rapport à l'imagerie faite qui n'est ni contrôlable, ni connu.

Le premier point semble complexe à considérer dans son ensemble et nécessiterait un complément d'information non disponible dans les bases de données existantes. Le second, lui, pourrait être abordé en observant sur plusieurs instants du temps l'évolution des images IRM. C'est cette stratégie que nous avons suivi et qui est présentée dans le chapitre suivant.



9. Filtrage et segmentation spatio-temporels

9.1 Introduction

Le chapitre précédent a montré l'intérêt du filtrage pour l'amélioration de la segmentation mais aussi le besoin de l'intégration de la dimension temporelle pour certaines applications. Dans ce chapitre nous résumons trois contributions sur cette problématique.

Les deux premières sont basées sur l'intégration de la composante temporelle dans les filtrages anisotrope et *mean-shift*. Une troisième est consacrée à la croissance de région.

Ces contributions ne seront pas appliquées à l'AVC mais aux lésions de sclérose en plaques (SEP). En effet, ces approches où l'on va considérer une évolution observée sur plusieurs acquisitions IRM échantillonnées dans le temps, sont incompatibles avec les acquisitions préalablement utilisées pour l'AVC.

9.2 Diffusion Anisotrope temporelle

Ces travaux ont été présentés dans (PORTEJOIE et al., 2015).

9.2.1 Introduction

Le débruitage des images est étudié depuis plusieurs décennies dans le cadre du traitement des images. Il s'agit d'une tâche fondamentale visant à récupérer l'image la plus propre possible à partir d'une entrée corrompue. Un grand nombre de techniques différentes ont été conçues pour résoudre ce problème (BUADES et al., 2005b; PRASATH et al., 2014). Si la plupart d'entre elles traitent des images en deux dimensions, d'autres techniques étendent leur formulation à des données en trois dimensions en considérant le temps comme la troisième dimension, comme dans les séquences d'images ou le traitement vidéo.

L'un des moyens les plus élémentaires de filtrer les images corrompues est évidemment de les faire passer à travers un filtre gaussien, bien que cela conduise à des données floues. La diffusion anisotrope est un moyen de résoudre ce problème : le processus de filtrage est lié à la norme locale de gradient. Plus la norme de gradient est basse, plus l'impact du filtre passe-bas est important. Il s'agit d'un processus itératif sans critère d'arrêt basé

sur l'image. La première description de la diffusion anisotrope pour le filtrage d'images a été introduite par Perona et Malik (PERONA et al., 1990) mais Barash en a proposé une plus robuste dans (BARASH, 2002), facilement extensible au filtrage d'images en couleur, comme le filtrage bilatéral introduit par Tomasi dans (TOMASI et al., 1998). Le débruitage des images peut également être réalisé par des approches basées sur des patches comme les *NL-means*, proposées par Buades et al. (BUADES et al., 2005b).

La plupart de ces méthodes ont été étendues pour traiter des données longitudinales. Montagnat (MONTAGNAT et al., 2003) a été le premier à proposer une application de diffusion anisotrope au filtrage de séquences d'images échocardiographiques. Plus récemment, les méthodes BM3D et BM4D ont été appliquées aux séquences d'images et aux vidéos dans (BOULANGER et al., 2010; BUADES et al., 2008; MALM et al., 2007), et également utilisées pour l'amélioration des images (JIANG et al., 2013) et la restauration (BOULANGER et al., 2007; LI et al., 2009). L'idée que l'estimation du mouvement n'était pas nécessaire dans ces approches a été introduite dans (BUADES et al., 2005a).

De nombreuses autres approches ont été proposées pour l'élimination du bruit ou le filtrage des images, telles que des méthodes plus avancées basées sur les champs de Markov (CHEN et al., 2007; MAGGIONI et al., 2012), le filtrage collaboratif (DABOV et al., 2007), des transformées dans des espaces peu denses (BHAGAVATHY et al., 2007) et les variations totales (CHAMBOLLE, 2004).

En ce qui concerne l'imagerie médicale, on note les travaux en IRM de filtrage par diffusion anisotrope dans (HE et al., 2009b) et (ARDIZZONE et al., 2003), puis pour l'aide à la reconstruction d'images TEP dans (BRANKOV et al., 2000).

Ici, nous nous concentrons sur les images médicales associées à des études longitudinales et les considérons comme des séquences d'images d'objets statiques : on pourra exprimer l'évolution des intensités des pixels comme des séries temporelles. Ainsi, nous étendons la méthode de diffusion anisotrope en incluant la dimension temporelle dans le processus ce qui permettra d'obtenir un comportement joint aux dimensions spatiales et temporelles (méthodes $AD+t$ et $RAD+t$). Notre solution est basée sur la forme de Barash (BARASH, 2002), très intuitive et très stable, et est décrite à l'aide de matrices d'échelles pour les dimensions spatiales, d'intensité et temporelles.

Classiquement, un des inconvénients du filtrage par diffusion anisotrope était sa faible performance pour éliminer les points très bruités dans les régions homogènes. Pour cela, nous ajoutons un terme de régularisation à la procédure itérative pour supprimer les valeurs aberrantes restantes.

9.2.2 Méthodes $AD+t$ et $RAD+t$

L'approche proposée est basée sur la diffusion anisotrope de Perona and Malik (PERONA et al., 1990), initialement proposée pour des images 2D.

Pour une image données $I(\mathbf{x})$, la diffusion anisotrope de la chaleur au pixel \mathbf{x} est :

$$\frac{\partial I(\mathbf{x})}{\partial t} = c(\mathbf{x}) \Delta I(\mathbf{x}) + \nabla c(\mathbf{x}) \cdot \nabla I(\mathbf{x}). \quad (9.1)$$

où la fonction $c(\mathbf{x})$ choisie est celle recommandée par Perona and Malik :

$$c(\mathbf{x}) = \exp\left(-\frac{\|\nabla I(\mathbf{x})\|^2}{K^2}\right) \quad (9.2)$$

avec K une constante qui permet de régler la largeur de la gaussienne.

En discrétisant l'équation (9.2) avec les relations de Barash (BARASH, 2002), on a l'équation d'évolution de la diffusion anisotrope (AD) :

$$I^{k+1}(\mathbf{x}) = \frac{\sum_{i \in \mathcal{N}} I^k(\mathbf{x}_i) \cdot c(\mathbf{x}_i)}{\sum_{i \in \mathcal{N}} c(\mathbf{x}_i)} \quad (9.3)$$

avec \mathcal{N} l'ensemble des voisins de \mathbf{x} et avec initialement $I^0(\mathbf{x}) = I(\mathbf{x})$.

La méthode proposée prend comme entrée une séquence d'images, où la valeur d'intensité d'un pixel \mathbf{x} à l'instant t est notée $I(\mathbf{x}, t)$. Dans les expressions suivantes, k désignera le numéro d'itération du processus de filtrage et \mathcal{N} l'ensemble contenant les voisins spatiotemporels de (\mathbf{x}, t) . Ainsi, nous étendons la forme de Barash à notre approche temporelle AD+t :

$$I^{k+1}(\mathbf{x}, t) = \frac{\sum_{i \in \mathcal{N}} I^k(\mathbf{x}_i, t_i) \cdot C(\mathbf{x}_i, t_i, I^k(\mathbf{x}_i, t_i))}{\sum_{i \in \mathcal{N}} C(\mathbf{x}_i, t_i, I^k(\mathbf{x}_i, t_i))} \quad (9.4)$$

avec $C(\mathbf{x}_i, t_i, I^k(\mathbf{x}_i, t_i))$ la fonction de pondération suivante :

$$C(\mathbf{x}_i, t_i, I^k_{\mathbf{x}_i, t_i}) = G_s(\mathbf{x}, \mathbf{x}_i) \cdot G_t(t, t_i) \cdot G_r(I^k_{\mathbf{x}, t}, I^k_{\mathbf{x}_i, t_i}) \quad (9.5)$$

qui combine trois fonctions de pondération : une pour le domaine spatial G_s , une pour le domaine temporel G_t et une pour les intensités G_r . Ces trois fonctions sont définies ainsi :

$$G_s(\mathbf{x}, \mathbf{x}_i) = g_s(d_s(\mathbf{x} - \mathbf{x}_i)) \quad (9.6)$$

$$G_t(t, t_i) = g_t(d_t(t - t_i)) \quad (9.7)$$

$$G_r(I^k_{\mathbf{x}, t}, I^k_{\mathbf{x}_i, t_i}) = g_r(d_r(I^k_{\mathbf{x}, t} - I^k_{\mathbf{x}_i, t_i})) \quad (9.8)$$

où g (également appelée fonctions de noyau ou de profil) est une fonction positive, qui somme à un et décroissante, et où les fonctions d sont des distances, typiquement euclidiennes généralisées (ou de Mahalanobis). Dans le cas de la distance de Mahalanobis, des matrices d'échelles pour les caractéristiques spatiales, temporelles et d'intensité doivent être définies. On peut noter que notre fonction $C(\cdot)$ permet toujours une pré-normalisation des données d'entrée par les matrices d'échelles.

Comme le filtrage de diffusion anisotrope n'est pas capable d'éliminer les pixels aberrants dans les régions homogènes, dont les écarts d'intensité par rapport aux intensités moyennes des régions dépassent la valeur de la bande passante, nous introduisons un terme de régularisation pour réduire le bruit restant. Cela conduit à l'équation d'évolution de notre approche RAD+t :

$$I^{k+1}(\mathbf{x}) = (1 - \kappa_{\mathbf{x}}^k) \frac{\sum_{i \in \mathcal{N}} I^k(\mathbf{x}_i, t_i) \cdot C(\mathbf{x}_i, t_i, I^k(\mathbf{x}_i, t_i))}{\sum_{i \in \mathcal{N}} C(\mathbf{x}_i, t_i, I^k(\mathbf{x}_i, t_i))} + \kappa_{\mathbf{x}}^k \frac{\sum_{i \in \mathcal{N}} I^k(\mathbf{x}_i, t_i)}{|\mathcal{N}|} \quad (9.9)$$

où $|\mathcal{N}|$ représente la cardinalité de l'ensemble \mathcal{N} .

Le terme de régularisation est la moyenne des intensités sur le voisinage de (\mathbf{x}, t) noté $\mathcal{N}(\mathbf{x}, t)$. Il est important que cette régularisation ne soit pas appliquée systématiquement afin de conserver la propriété de la diffusion anisotrope de préservation des contours des objets.

Ainsi, ce terme doit être pondéré par un scalaire $\kappa_{\mathbf{x}}^k \in [0; 1]$ qui va dépendre des valeurs d'intensités de $\mathcal{N}(\mathbf{x}, t)$. Afin d'obtenir les propriétés de filtrage désirées, $\kappa_{\mathbf{x}}^k$ doit être proche de 1 quand l'intensité de $I^k(\mathbf{x}, t)$ est considérée comme aberrante, et proche de 0 sinon.

Ici, on utilise $\kappa_{\mathbf{x}}^k \in \{0; 1\}$ et la « norme » zéro $\|\cdot\|_0$, qui compte le nombre de dimensions non nulles dans un vecteur, pour mesurer le comportement dans le voisinage.

Le vecteur $\mathbf{I}^k(\mathbf{x}, t) \in \mathbb{R}^{|\mathcal{N}|}$ est construit à partir de toutes les intensités de $\mathcal{N}(\mathbf{x}, t)$ ainsi : $\mathbf{I}^k(\mathbf{x}, t) = [I^k(\mathbf{x}_1, t_1), I^k(\mathbf{x}_2, t_2), \dots, I^k(\mathbf{x}_{|\mathcal{N}|}, t_{|\mathcal{N}|})]$. Le vecteur $\Delta\mathbf{I}^k(\mathbf{x}, t)$ est défini comme la soustraction entre $\mathbf{I}^k(\mathbf{x}, t)$ et $I^k(\mathbf{x}, t)$. En utilisant ces notations, $\kappa_{\mathbf{x}}^k$ se calcule de la manière suivante :

$$\kappa_{\mathbf{x}}^k = \begin{cases} 1 & \text{if } \|\Delta\mathbf{I}^k(\mathbf{x}, t)\|_0 + 1 > \tau \\ 0 & \text{otherwise} \end{cases} \quad (9.10)$$

avec τ un seuil fixant le nombre maximum de voisins aux intensités dissimilaires à $I^k(\mathbf{x}, t)$ pour remplacer le calcul de l'équation de diffusion anisotrope (9.4) par la moyenne des intensités des voisins de (\mathbf{x}, t) .

Pour calculer $\|\cdot\|_0$, il est recommandé de multiplier par un facteur supérieur à un puis de tronquer à la valeur entière inférieure les valeurs de $\Delta\mathbf{I}^k(\mathbf{x}, t)$ pour accroître la robustesse aux petites variations dans les régions homogènes.

Dans le cas multidimensionnel, quand $I^k(\mathbf{x}, t)$ est un vecteur de valeurs comme dans les images couleurs, la j^{th} composante de $\Delta\mathbf{I}^k(\mathbf{x}, t)$ est obtenue en tronquant d , la distance euclidienne entre $I^k(\mathbf{x}_j, t_j)$ et $I^k(\mathbf{x}, t)$, divisée par h_τ , le facteur d'échelles optimal ($h_\tau \geq 1$) défini ainsi :

$$d(I_{\mathbf{x}_j, t_j}^k, I_{\mathbf{x}, t}^k; h_\tau) = \frac{1}{h_\tau} \|I^k(\mathbf{x}_j, t_j) - I^k(\mathbf{x}, t)\|. \quad (9.11)$$

En pratique, plusieurs réglages doivent être faits :

- les fonctions décrites dans les équations (9.6), (9.7) et (9.8) doivent être spécifiées (avec leurs propres paramètres),
- la forme et la taille du voisinage \mathcal{N} doivent être choisies,
- les valeurs de τ et de h_τ doivent être données,
- le nombre d'itérations *#it* du filtre doit être fixé.

Par rapport aux approches de diffusion anisotrope spatiale, deux paramètres sont ajoutés : τ et h_τ pour le terme de régularisation. Il pourra cependant être pertinent d'ajouter un paramètre pour le contrôle de la contribution de la dimension temporelle sur le filtrage.

9.2.3 Evaluation et Application

La méthode proposée a été évaluée à la fois quantitativement sur une séquence d'images synthétiques et qualitativement sur des acquisitions cérébrales IRM longitudinales 2D de patients atteints de lésions de sclérose en plaques (SEP).

Le modèle synthétique est de taille $256 \times 256 \times 40$ (figure 9.1) et est composé de plusieurs régions spatio-temporelles constantes puis d'une région évoluant en intensité. Tous les instants temporels sont ensuite indépendamment corrompus par un flou gaussien d'écart-type σ_β puis par un bruit gaussien additif d'écart-type σ_η . Ainsi, trois séquences d'images corrompues sont créées : CS_1 avec $[\sigma_\beta=2, \sigma_\eta=10]$, CS_2 avec $[\sigma_\beta=4, \sigma_\eta=20]$, CS_3 avec $[\sigma_\beta=6, \sigma_\eta=30]$.

Les acquisitions IRM longitudinales réelles font partie d'une étude de patients atteints de sclérose en plaques et ont été acquises sur 25 instants temporels (GUTTMANN et al., 1995). Le pré-traitement réalisé sur ces données est basé sur celui proposé par (MEIER et al., 2003) : les données sont recalées et les intensités sont normalisées sur la première acquisition. Ces données sont 3D et ici, une seule coupe 2D est extraite des IRM.

Pour évaluer quantitativement la qualité de nos résultats de filtrage, nous avons choisi de comparer les images obtenues avec l'image originale non altérée en calculant la MSE et le

SSIM (WANG et al., 2004a). Pour les images de données réelles, l'évaluation est uniquement qualitative car il n'existe aucune vérité terrain.

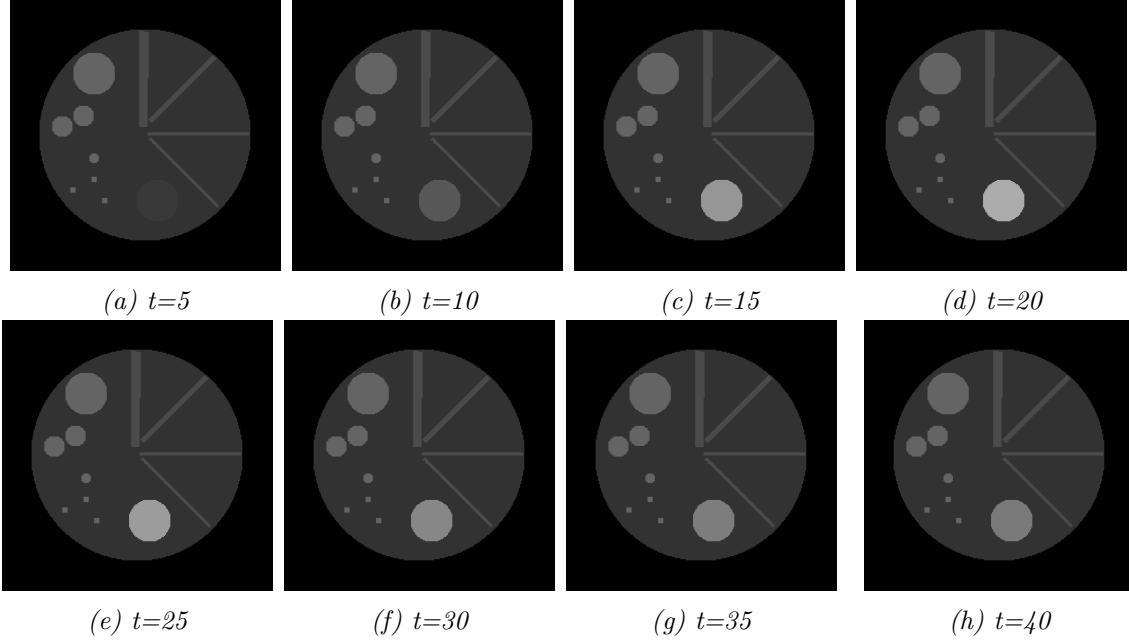


FIGURE 9.1 – Modèle synthétique pour les tests avant floutage et bruitage.

Expérimentations

Les approches de diffusion anisotrope proposées sont comparées à la méthode BM3D (DABOV et al., 2007) qui prend en compte le temps sur les données synthétiques. Ensuite, les résultats de ces approches sur les données IRM réelles sont présentés.

Nous commençons par décrire les choix communs à toutes nos expériences (principalement les fonctions décrites dans les équations (9.6), (9.7) et (9.8)), puis nous précisons les expériences.

Nous utilisons la distance euclidienne généralisée au carré d_M^2 pour toutes les mesures de distance d_s , d_t et d_r avec $\mathbf{H}_s = h_s^2 \cdot \mathbb{I}_s$, h_t^2 et $\mathbf{H}_r = h_r^2 \cdot \mathbb{I}_r$ les facteurs d'échelles respectifs avec \mathbb{I}_s et \mathbb{I}_r les matrices identités de dimensions correspondantes aux dimensions des espaces spatial et d'intensité. Les fonctions $g_s(\cdot)$, $g_t(\cdot)$ et $g_r(\cdot)$ sont toutes la même fonction de profil $g(u) = \exp(-u)$ avec le scalaire u égal à d_M^2 . Ces choix nous permettent d'utiliser le même $c(\mathbf{x})$ que (9.2) généralisé à notre approche spatiotemporelle.

Pour les séquences synthétiques corrompues, les paramètres ont été optimisés par une recherche exhaustive en minimisant l'erreur quadratique moyenne (MSE, voir l'annexe A). Grâce aux valeurs obtenues sur les séquences corrompues, nous avons fixé manuellement les paramètres pour les données IRM réelles. Tous les paramètres sont décrits dans le tableau 9.1. Nous précisons que \mathcal{N} a été fixé en fonction des paramètres h_s et h_t de sorte que $|\mathcal{N}| = (2h_s + 1) \times (2h_s + 1) \times (2h_t + 1)$.

Résultats

Pour faciliter la lecture, les données et les résultats (animés) sont disponibles en ligne¹. La figure 9.2 présente les résultats obtenus par le filtrage de diffusion anisotrope de base (AD), puis par notre approche de filtrage anisotrope spatiotemporel (AD+t) et par notre

1. <http://www.creatis.insa-lyon.fr/%7egrenier/research/PortejoieICIP2015/>

Séquences	Paramètres					
	h_s	h_t	h_r	$\#it$	$\tau/ \mathcal{N} $	h_τ
CS_1	1	3	5	10	0.88	2
CS_2	1	3	10	15	0.88	2
CS_3	1	3	25	20	0.88	2
Real Data	1	3	5	10	0.88	2

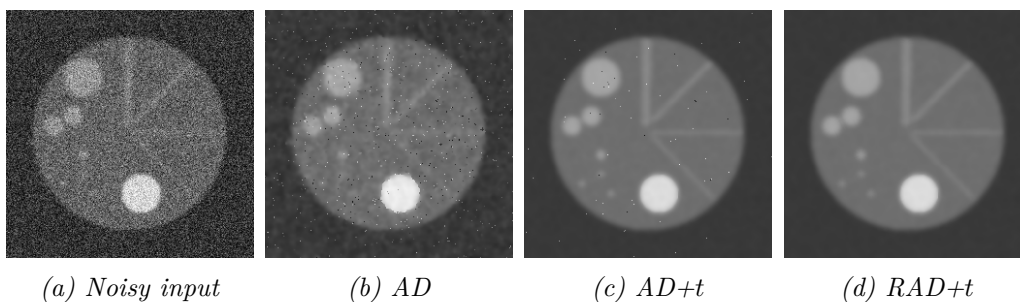
TABLE 9.1 – Paramètres utilisées pour les expériences.

méthode régularisée (RAD+ t). La figure 9.3 présente les résultats obtenus pour les données de synthèses et les données réelles.

Sur la base de l'évaluation visuelle et de la MSE, nous pouvons déduire que nos filtres AD+ t et RAD+ t sont plus performants que AD pour la suppression du bruit et la préservation des contours. Mais, pour obtenir de bons résultats de filtrage, *six* paramètres doivent être ajustés. Cependant, d'après la recherche exhaustive sur ces données, il semble que seuls h_r et $\#it$ doivent être réglés avec soin, car les autres restent les mêmes pour les différentes images. Nous vérifions également que la suppression du bruit sur la première et la dernière image des séquences est affectée par les effets de bords et que ce nombre d'images est lié à h_t .

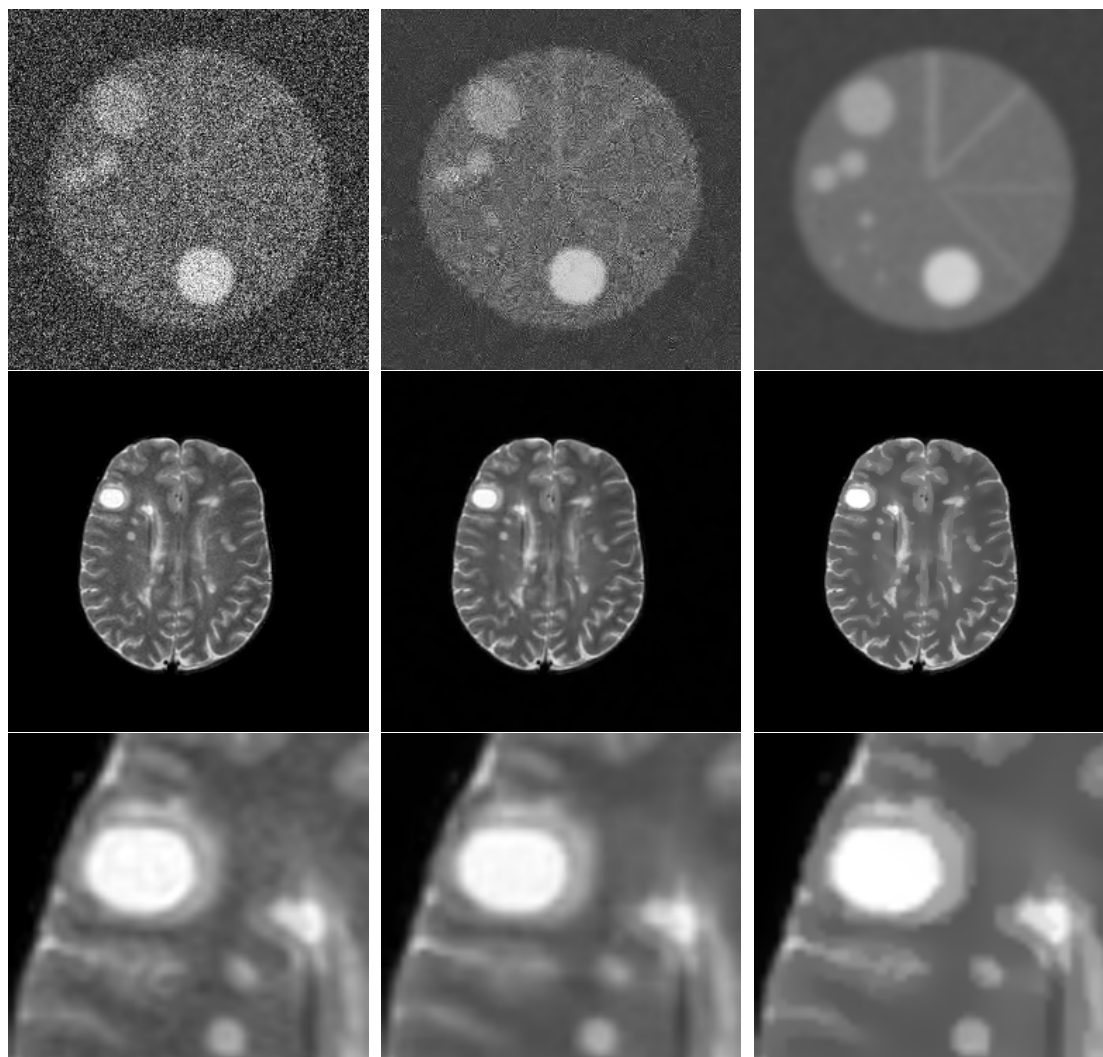
Nous avons comparé quantitativement notre approche RAD+ t à BM3D (DABOV et al., 2007) qui est l'une des approches de filtrage les plus efficaces. Nous avons appliqué cet algorithme à nos données sans optimiser les paramètres par défaut car cette approche ne nécessite pas de réglage. Les valeurs MSE et SSIM obtenues sont présentées dans le tableau 9.2. Nous observons que notre approche est plus adaptée que BM3D pour ce type particulier de séquences d'images.

Qualitativement, nous observons sur les données IRM réelles que notre approche est efficace pour éliminer les artefacts de bruit et améliorer la qualité de l'image, nous pouvons également noter que les détails sont plus nets.

FIGURE 9.2 – Résultats de filtrage pour le 20^e instant de la séquence CS_2 .

9.2.4 Conclusion

Nous avons proposé une nouvelle méthode pour filtrer les séquences d'images, en tenant compte des dimensions spatiales et temporelles dans le processus de filtrage par diffusion anisotrope. Nous avons étendu la forme de Barash (BARASH, 2002) en ajoutant des données temporelles et un terme régularisation contrôlé localement dans le processus pour supprimer les bruits aberrants dans les régions homogènes.



(a) Image

(b) BM3D

(c) RAD+t

FIGURE 9.3 – Comparaison des filtrages BM3D et RAD+t. Image d'entrée (colonne de gauche) et les résultats de filtrage pour BM3D (colonne du centre) et RAD+t (colonne de droite) pour le 20^e instant de CS_3 (première ligne) et le 4^e instant pour les données IRM réelles (ligne au centre) et un détail sur une région d'intérêt (ligne du bas).

	CS_1	CS_2	CS_3
RAD+t MSE	25	59	181
BM3D MSE	32	64	222
RAD+t SSIM	0.93	0.88	0.87
BM3D SSIM	0.91	0.85	0.85

TABLE 9.2 – Résultats optimaux obtenus avec RAD+t et BM3D pour la MSE (à minimiser) et le SSIM (à maximiser). Pour les deux mesures, les optimisations ont été faites sur la MSE.

On a montré que ces méthodes améliorent la suppression du bruit et préserve les contours des régions dans les séquences d'images. Aussi, nous avons également montré l'efficacité de l'utilisation du terme de régularisation pour améliorer le débruitage.

D'après nos expériences sur nos données synthétiques, nous obtenons des résultats supérieurs à ceux obtenus avec l'une des meilleures approches (BM3D).

L'utilisation de trois paramètres d'échelles pour contrôler la diffusion anisotrope permet un réglage intuitif. Cependant, notre approche nécessite un ajustement fin de 2 à 6 paramètres (trois paramètres d'échelles, τ et le nombre d'itérations) qui n'est pas décrit ici.

Une partie de ces paramètres pourraient être supprimée par l'utilisation du formalisme *mean-shift* à condition de l'étendre aux données spatio-temporelles. Puisque le filtrage *mean-shift* a aussi montré sa supériorité par rapport au filtrage anisotrope, on peut aussi espérer une amélioration des performances.

9.3 Mean-shift Spatio-Temporel : STMS

Cette partie présente la contribution d'extension des *mean-shift* aux données spatio-temporelle (STMS) que l'on va considérer comme des séries temporelles. Ce travail a fait l'objet de la thèse de Simon Mure (MURE et al., 2015b).

9.3.1 Introduction

Ces dernières années les acquisitions longitudinales ont augmenté de façon spectaculaire. Les séquences vidéo, le suivi par système de positionnement mondial (GPS) ou le suivi médical, ont conduit au développement de nombreuses applications d'exploration de données de séries temporelles. Ainsi, l'analyse non supervisée de séries temporelles est devenue très pertinente dans le but de détecter et d'identifier automatiquement des modèles ou comportements.

Plusieurs méthodes de regroupement de séries temporelles ont été proposées pour la prédiction, basées sur l'étude des corrélations de signaux (PAPADIMITRIOU et al., 2007), des attributs de forme (HAUTAMAKI et al., 2008 ; MEESRIKAMOLKUL et al., 2012) ou des modèles d'évolution (KALPAKIS et al., 2001). Les travaux présentés dans (MATSUBARA et al., 2014) introduisent une méthode non supervisée et sans paramètre pour exploiter les régimes (ou patrons ou modèles) et les transitions (discontinuités) dans de grandes séries temporelles co-évolutives, mais ne regroupent pas les évolutions similaires.

Bien que d'après (AGGARWAL et al., 2013), les seules méthodes connues qui peuvent être généralisées au regroupement de séries temporelles multivariées sont celles proposées dans le domaine des trajectoires spatiales (BENKERT et al., 2008 ; JEUNG et al., 2008 ; LI et al., 2010b ; ZHENG et al., 2013), on note que des travaux exploitant la combinaison de l'information spatiale et couleur ont déjà été publiés dans le cadre de l'analyse de vidéos. Notamment, de nombreuses études portant sur le filtrage ou la restauration de séquences d'images ont été menées depuis le début des années 90 (voir (BHAGAVATHY et al., 2007)). Certaines approches s'appuient sur un estimateur de mouvement. (VARGHESE et al., 2010) et (MAGGIONI et al., 2014) ont proposé de telles approches exploitant à la fois les redondances spatiales et temporelles des données et les ont comparées aux méthodes de filtrage vidéo spatio-temporel les plus efficaces connues à l'époque. Néanmoins, la performance de ces méthodes dépend principalement du choix de l'estimateur de mouvement.

D'autres approches s'appuient sur le formalisme *mean-shift* pour la segmentation de vidéos en couleurs (DEMENTHON et al., 2005 ; KE et al., 2005 ; WANG et al., 2004a). Dans ces travaux, les pixels de chaque image ont été considérés comme des échantillons indépendants, c'est-à-dire que le *mean-shift* n'a pas été utilisé pour filtrer l'évolution temporelle des caractéristiques associées à un pixel mais a plutôt été utilisé pour filtrer des volumes vidéo (2D+t) multi-canaux (par exemple RGB) ou des caractéristiques précédemment calculées pour chaque pixel.

Les travaux (ANJUM et al., 2008; FENG et al., 2003) ont décrit comment étendre le *mean-shift* au domaine espace-temps afin de filtrer des séquences vidéo et des données spatio-temporelles. Puis pour les applications médicales, (AI et al., 2014; CHENG et al., 2009; LEUNG et al., 2006) traitent de l'analyse longitudinal d'IRM avec le *mean-shift*. Cependant dans ces travaux, l'information spatio-temporelle n'est pas formulée explicitement dans le processus.

Ici, nous allons étendre le *mean-shift* aux données spatio-temporelles en ajoutant une contrainte sur l'évolution des échantillons dans le temps. Seuls les échantillons dans le voisinage du point à filtrer partageant une évolution similaire de leurs caractéristiques contribueront au filtrage de ce point.

Ce principe est détaillé dans la section suivante. Puis il sera testé sur des données de synthèse et des données réelles d'IRM pour de l'analyse longitudinale de lésions SEP.

9.3.2 Méthode STMS

L'approche STMS, qui permet le filtrage spatio-temporel de séries temporelles et basée sur le concept *mean-shift*, est détaillée dans cette partie.

On considère un ensemble de n échantillons de coordonnées spatiales $\{\mathbf{x}_{s,i}\}_{i=1\dots n}$ et dont les caractéristiques évoluent au cours du temps $\{\mathbf{x}_{t,i}\}_{i=1\dots n}$. Le nombre de dimensions spatiales et de points temporels sont respectivement S et T . L'ensemble des échantillons $\mathbf{X} = \{\mathbf{x}_i\}_{i=1\dots n}$ est défini ainsi :

$$\mathbf{x}_i = \begin{bmatrix} \mathbf{x}'_{s,i} & \mathbf{x}'_{t,i} \end{bmatrix} \in \mathbf{X} \quad \text{with} \quad \begin{array}{l} \mathbf{x}_{s,i} \in \mathbb{R}^S : \text{domaine spatial} \\ \mathbf{x}_{t,i} \in \mathbb{R}^T : \text{domaine temporel} \\ i = 1, \dots, n : \text{index de l'échantillon} \end{array} \quad (9.12)$$

En utilisant ces notations, on propose l'équation suivante pour calculer l'évolution *mean-shift* spatio-temporel de chaque échantillon :

$$\mathbf{x}_i^{[k+1]} = \frac{\sum_{j=1}^n Sp_{i,j}(\mathbf{x}_{s,i}^{[k]}, \mathbf{x}_{s,j}^{[k]}) \cdot Ra_{i,j}(\mathbf{x}_{t,i}^{[k]}, \mathbf{x}_{t,j}^{[k]}) \cdot \mathbf{x}_j^{[k]}}{\sum_{j=1}^n Sp_{i,j}(\mathbf{x}_{s,i}^{[k]}, \mathbf{x}_{s,j}^{[k]}) \cdot Ra_{i,j}(\mathbf{x}_{t,i}^{[k]}, \mathbf{x}_{t,j}^{[k]})} \quad (9.13)$$

où $Sp_{i,j}(\cdot)$ et $Ra_{i,j}(\cdot)$ sont respectivement les fonctions de pondération basées sur les distances spatiale et temporelle entre l'échantillon d'intérêt \mathbf{x}_i et un autre échantillon \mathbf{x}_j (\mathbf{x}_i et $\mathbf{x}_j \in \mathbb{R}^{S+T}$) :

$$Sp_{i,j}(\mathbf{x}_{s,i}^{[k]}, \mathbf{x}_{s,j}^{[k]}) = g_s \left(d_s^2(\mathbf{x}_{s,i}^{[k]}, \mathbf{x}_{s,j}^{[k]}, \mathbf{H}_s) \right) \quad (9.14)$$

$$Ra_{i,j}(\mathbf{x}_{t,i}^{[k]}, \mathbf{x}_{t,j}^{[k]}) = g_r \left(d_r^2(\mathbf{x}_{t,i}^{[k]}, \mathbf{x}_{t,j}^{[k]}, \mathbf{H}_r) \right) \quad (9.15)$$

Contrairement à l'approche *mean-shift* classique, la même distance n'est pas utilisée pour toutes les caractéristiques dans le cadre *mean-shift* spatio-temporel. Nous allons faire une distinction entre les dimensions spatiales et de temps.

Pour la caractéristique spatiale, la distance euclidienne généralisée $d_s(\mathbf{u}_s, \mathbf{v}_s, \mathbf{H}_s)$ est calculée pour deux échantillons \mathbf{u}_s et \mathbf{v}_s avec \mathbf{H}_s la matrice d'échelle spatiale de taille $S \times S$.

En revanche, pour la caractéristique temporelle nous utilisons la norme infini :

$$d_r(\mathbf{u}_t, \mathbf{v}_t, \mathbf{H}_r) = \|\mathbf{H}_r^{-\frac{1}{2}}(\mathbf{u}_t - \mathbf{v}_t)\|_\infty \quad (9.16)$$

avec \mathbf{H}_r la matrice d'échelle pour la caractéristique temporelle qui est carrée et de taille $T \times T$. De cette manière, cette norme permet d'obtenir le plus grand écart (réduit par la matrice \mathbf{H}_r) qu'il existe au cours du temps entre deux échantillons. On va ainsi pouvoir conserver pour le calcul de la moyenne (9.13) uniquement les échantillons qui ne s'éloignent jamais trop de l'échantillon à filtrer, ni en spatial, ni en temps.

Dans ce travail, on propose de conserver la même fonction de profil g pour pondérer les deux distances :

$$g_s(d_s^2(\cdot)) = g_r(d_r^2(\cdot)) = \begin{cases} 1 & \text{if } d_s^2(\cdot), d_r^2(\cdot) \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad (9.17)$$

L'équation (9.15) fait en sorte que $Ra_{i,j}^{[k]}$ devienne nul si la distance entre deux trajectoires, à un moment donné, est supérieure à 1. Par conséquent, les trajectoires éloignées de celle du point à filtrer \mathbf{x}_i seront exclues pour le filtrage de \mathbf{x}_i .

Une illustration de la sélection d'échantillons décrite ci-dessus est illustrée dans la figure 9.4. Bien que les échantillons rouges soient inclus dans le voisinage spatial de l'échantillon à filtrer (en bleu), ils ne seront pas utilisés pour le filtrage car leurs évolutions s'éloignent hors de la limite définies par l'évolution de l'échantillon à filtrer. *A contrario*, les échantillons verts sont à la fois suffisamment proches de l'échantillon de référence en spatial et dans le temps pour participer à sa mise à jour.

Pour le *mean-shift* spatio-temporel (9.13), nous avons retenu le processus *mean-shift blurring*. Ainsi, l'approche STMS est illustrée sur la figure 9.4 et son algorithme est décrit dans 3.

Algorithme 3 Algorithme de filtrage *mean-shift* spatio-temporel *blurring*

Entrée: h_s et h_r réglé par l'utilisateur

Entrée: $X = \{\mathbf{x}_i\}_{i=1\dots n}$

- 1: $k \leftarrow 0$
 - 2: $X^{[0]} \leftarrow X$
 - 3: **repeat**
 - 4: **for all** $\mathbf{x}_i \in X^{[k]}$ **do**
 - 5: Calculer $\mathbf{x}_i^{[k+1]}$ avec (9.13)
 - 6: **end for**
 - 7: $k \leftarrow k + 1$
 - 8: **until** $|X^{[k]} - X^{[k-1]}| < \epsilon$
 - 9: $\hat{X} \leftarrow X^{[k]}$
 - 10: **return** \hat{X}
-

Le processus *blurring* permet aussi une optimisation par fusion des échantillons similaires. Cette optimisation va permettre d'économiser du temps de calcul et aussi produire, en plus du filtrage, un clustering de tous les échantillons.

Quand deux échantillons sont suffisamment proches, c'est à dire $Sp_{i,j} \cdot Ra_{i,j} \neq 0$ calculés avec des paramètres d'échelles 10 fois plus petits que ceux utilisés pour le filtrage, on peut supposer que ces deux échantillons vont converger au même endroit et qu'il n'est donc pas pertinent de s'occuper des deux échantillons. Une solution est de les fusionner en créant un nouvel échantillon qui sera la moyenne des deux échantillons et qui aura deux fois plus de poids dans le calcul de l'évolution *mean-shift* ou *mean-shift* spatio-temporel.

En pratique, il est possible de fusionner tous les échantillons suffisamment proches (au sens du critère précédent) d'un autre échantillon et ceci à chaque itération. Une telle approche permet en moyenne de diviser le nombre d'échantillons par deux à chaque itération et

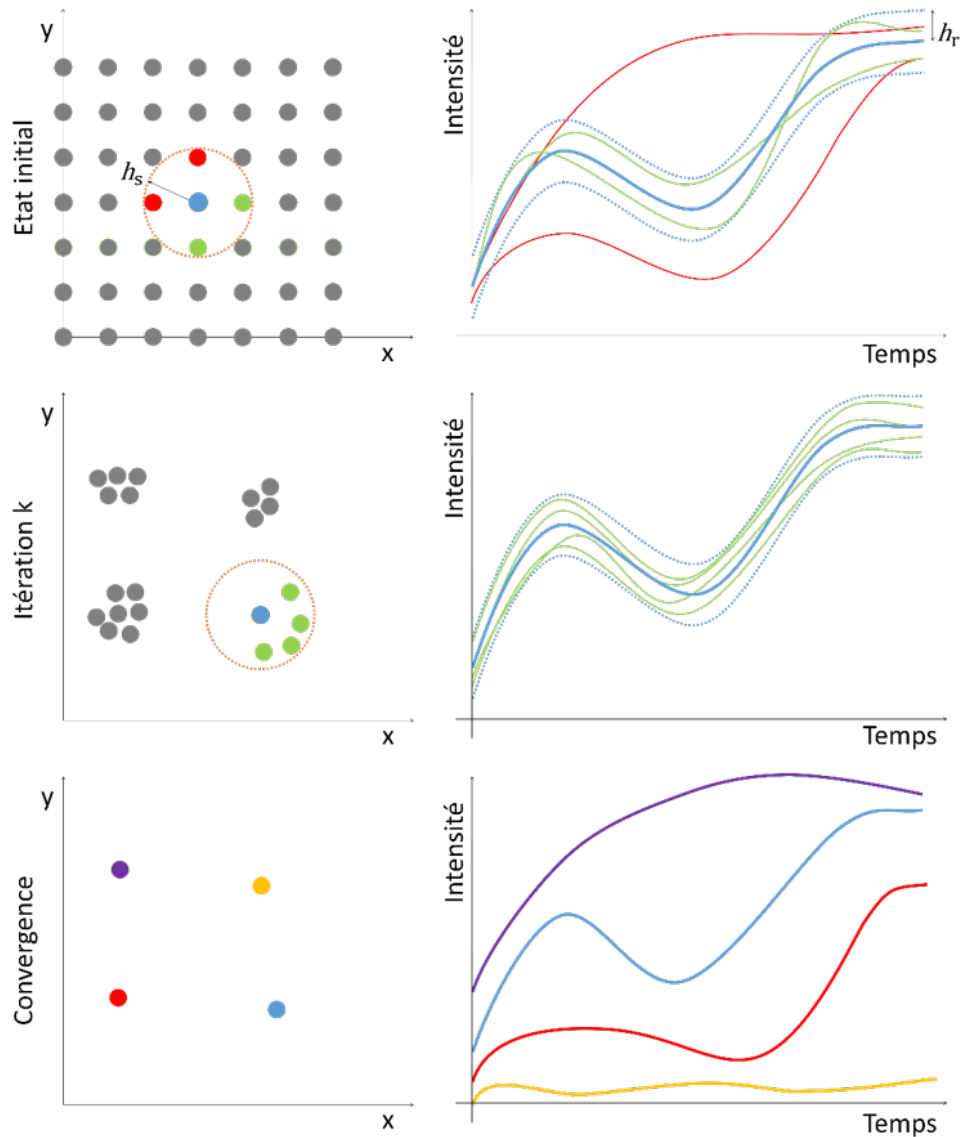


FIGURE 9.4 – Principe de STMS, la sélection spatio-temporelle et son évolution itérative. La première ligne illustre le voisinage spatial (à gauche) et en temps (à droite) d'un point à filtrer (en bleu). La ligne du milieu illustre l'évolution des positions spatiales et en temps des échantillons. La ligne du bas illustre la convergence de ce processus.

ainsi de réduire considérablement le temps de calcul. Cependant, elle modifie de manière substantielle les résultats obtenus car cette fusion peut impacter l'évolution des échantillons à proximité des échantillons fusionnés.

Ce regroupement (ou clustering) peut aussi se faire à la convergence du processus *mean-shift* spatio-temporel. Son intérêt est de regrouper les pixels ayant eu la même évolution. On peut ainsi obtenir une segmentation des régions partageant la même évolution au cours de la séquence.

9.3.3 Evaluation sur des données de synthèse

Des données spatio-temporelles synthétiques ont été créées pour évaluer l'apport du *mean-shift* spatio-temporel par rapport au *mean-shift*. Elles sont construites à partir du fantôme Shepp-Logan (SHEPP et al., 1974) où chacune des cinq régions voit sa valeur d'intensité

évoluer à 8 instants du temps. Le fantôme et les évolutions des régions sont détaillés dans la figure 9.5. On obtient ainsi une séquence de $256 \times 256 \times 8$ qui va ensuite être corrompue d'abord par une convolution avec un filtre gaussien de taille 10×10 et d'écart-type 0,4 puis par l'ajout d'un bruit normal d'écart-type 0,2 (figure 9.6). Sans l'information temporelle, il n'est pas possible de discriminer les cinq régions.

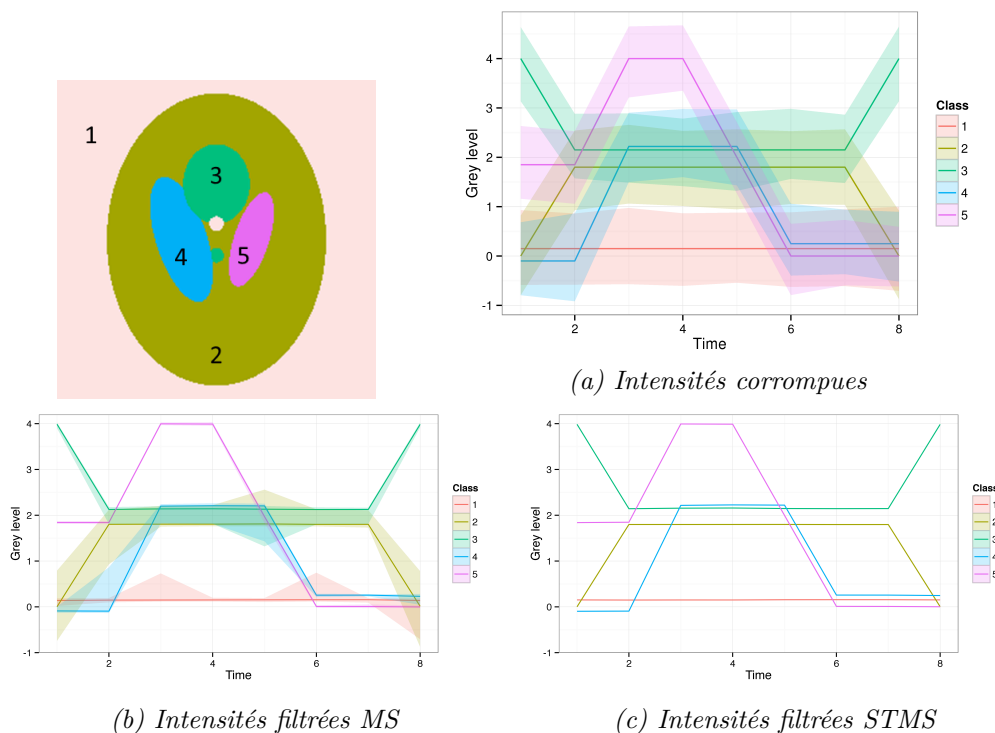


FIGURE 9.5 – Données synthétiques pour l'évaluation de l'approche *mean-shift* spatio-temporel. Ligne du haut : à gauche la représentation des 5 régions et l'évolution sur les 8 instants du temps de leurs intensités (à droite) avec les perturbations de flou et de bruit. Ligne du bas : intensités filtrées par approche *mean-shift* (à gauche) et *mean-shift* spatio-temporel (à droite).

On observe que l'apport de la prise en compte du temps dans le formalisme *mean-shift* permet d'obtenir les améliorations souhaitées par rapport à l'approche *mean-shift*.

Sur nos images de synthèses, on observe une amélioration de la séparation des objets (flèches bleues et rouges de la figure 9.6), une amélioration de la suppression des points aberrants (flèches vertes de la même figure), une discrimination et une homogénéisation plus précises des niveaux de gris de chaque région. Les évolutions des intensités données sur la figure 9.5 permettent d'apprécier le filtrage quasi parfait réalisé par l'approche *mean-shift* spatio-temporel.

Pour les deux méthodes, ces résultats sont les résultats de filtrage optimum obtenus par une recherche exhaustive des paramètres d'échelles maximisant la moyenne du *PSNR* sur la séquence. L'approche *mean-shift* spatio-temporel permet d'obtenir un *PSNR* moyen optimal de 62dB, 45dB pour le filtrage *mean-shift* et les données corrompues étaient à 29dB.

9.3.4 Application à l'étude de lésions SEP

Données réelles

L'algorithme a été testé sur quatre régions d'intérêt (ROI) 3D (figure 9.9), chacune contenant une lésion de sclérose en plaques évoluant dans le temps sur des images IRM pondérées en

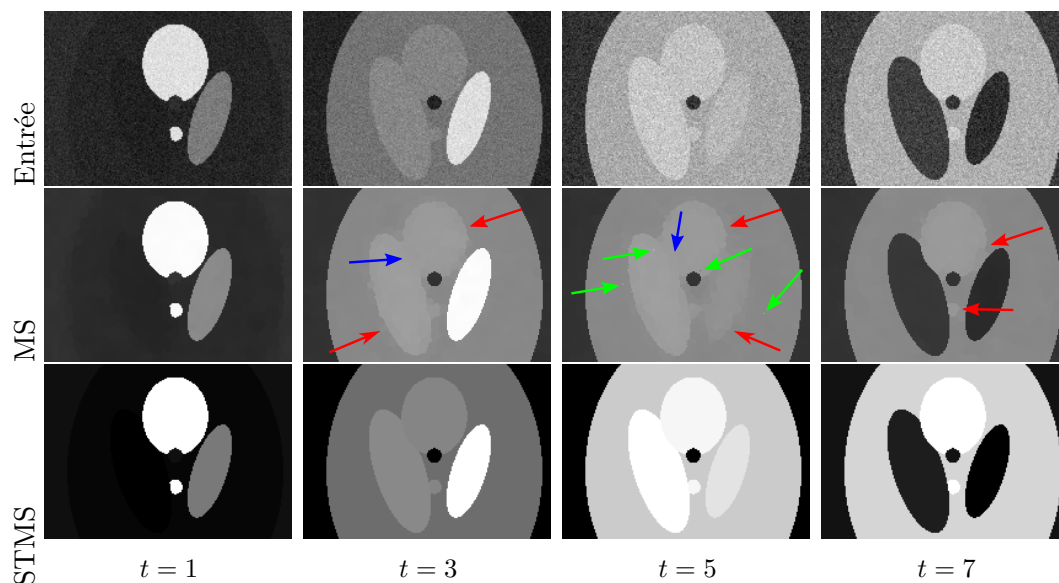


FIGURE 9.6 – Données synthétiques et résultats des approches MS et STMS. Sur la première du haut : les données corrompues par ajout de flou gaussien et de bruit gaussien pour 4 instants. Lignes du milieu et du bas : résultats du filtrage mean-shift puis mean-shift spatio-temporel.

T2 (figure 9.7) issues de l'étude de (GUTTMANN et al., 1995).

Dans cette étude, quarante patients atteints de sclérose en plaques ont été imagés 24 fois en IRM au cours d'une année. Les huit premières acquisitions ont été réalisées chaque semaine, puis les huit suivantes ont été réalisées toutes les deux semaines et les dernières acquisitions ont été réalisées chaque mois.

Les pré-traitements effectués avant filtrage *mean-shift* spatio-temporel sont inspirés de (MEIER et al., 2003). Leurs réglages ainsi que les choix des méthodes de recalage, corrections des artefacts, extraction de la boîte crânienne et normalisation ont été optimisés afin d'être robustes sur l'ensemble de la cohorte de patients. On notera, qu'après ces pré-traitements, le premier examen IRM d'un patient est soustrait à chacun des examens IRM suivants.

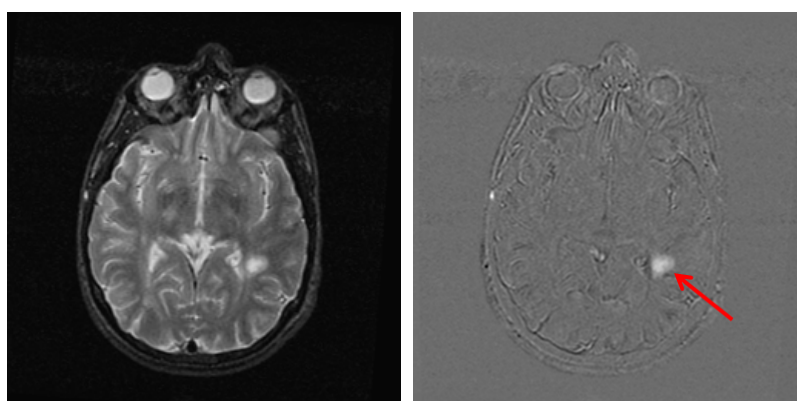


FIGURE 9.7 – Coupe d'une IRM pondérée T2 d'un patient atteint de lésions de sclérose en plaques (à gauche) et l'image de différence avec la première acquisition du suivi de ce patient (à droite). La flèche rouge indique la lésion. La région d'intérêt autour de cette lésion est étudiée en détail sur la figure 9.8.

La figure 9.8 permet d'observer l'efficacité du filtrage STMS sur les évolutions des intensités : il ne reste que quatre évolutions différentes après filtrage *mean-shift* spatio-temporel. La

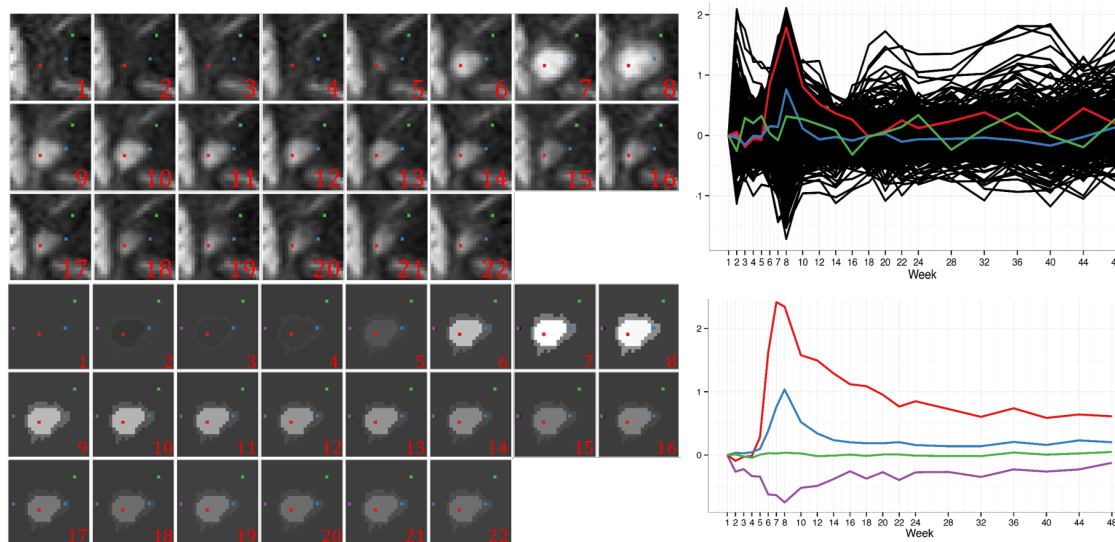


FIGURE 9.8 – Filtrage STMS d'une lésion sur 22 instants. En haut, représentation en séquence et du profil des intensités d'une lésions et sa périphérie sur les 22 instants. En bas, la même zone et les évolutions des intensités après filtrage mean-shift spatio-temporel.

figure 9.9 illustre le filtrage et le clustering sur quatre lésions. Ce clustering permet d'obtenir une segmentation des zones de l'images se comportant de manière similaire.

9.3.5 Conclusion

Nous avons introduit l'approche *mean-shift* spatio-temporel et nous avons observé son efficacité sur le filtrage de données spatio-temporelles synthétiques et réelles. Cette approche permet aussi d'obtenir une segmentation des séquences d'images permettant d'identifier les régions partageant un comportement similaire au long de la séquence.

Dans (AMELI et al., 2016 ; MURE et al., 2016d) ceci nous a permis d'identifier un processus de sténose des veinules dans les lésions SEP qui impactent seulement quelques pixels. L'approche *mean-shift* spatio-temporel a aussi été utilisée pour des données multi-spectrale en photoacoustique pour la discrimination de tissus (DOLET et al., 2016 ; DOLET et al., 2018).

Nous avons aussi adapté STMS pour répondre à des problématiques spécifiques. Afin de pouvoir comparer les évolutions de lésions entre elles, nous avons ajouté une étape de calcul de déformation temporelle dynamique (ou DTW pour *dynamic time warping*) qui va permettre de fusionner les régions dont le comportement est similaire à une déformation du temps prêt (MURE et al., 2016c). Sur les quatre lésions étudiées précédemment (figure 9.8), l'algorithme de DTW permet de regrouper sous la même classe les zones périphériques des lésions, ce qui signifie qu'elles partagent un comportement similaire, à contrario du coeur des lésions dont un d'entre eux est spécifique (retour à zéro de l'intensité). Cet algorithme a aussi permis de regrouper efficacement des régions en fonction des activités humaines dans le cadre de la surveillance du territoire par imagerie satellitaire (MURE et al., 2016b).

Ce formalisme a aussi été adapté aux vidéos en ajoutant une fenêtre temporelle (proche de ce qui a été fait pour la diffusion anisotrope) *STMS++* pas nécessairement symétrique par rapport à l'image à filtrer (MURE et al., 2016a) et la segmentation obtenue permet de ré-identifier des objets ayant été occultés.

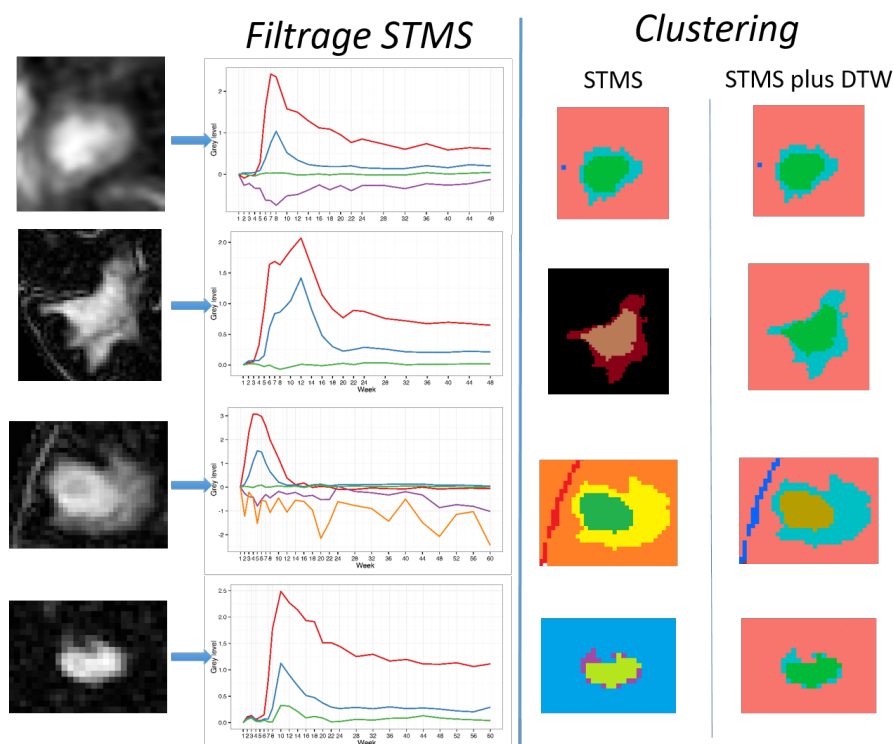


FIGURE 9.9 – Filtrage et clustering STMS puis STMS et DTW pour 4 lésions de SEP. L’algorithme de Déformation Temporelle Dynamique (DTW) permet de regrouper les régions dont l’évolution est similaire malgré une déformation du temps.

Le code STMS (écrit en C++ et basé sur la bibliothèque ITK) est disponible sur git². En 2020, une version GPU a été écrite par Claire Mouton de l’équipe *info-dev* de CREATIS. Ces approches de filtrage et de clustering permettent de segmenter toute l’image. Cependant, ceci n’est pas toujours nécessaire : on peut souhaiter se focaliser sur une seule région. Le premier avantage est une économie en temps de calcul puisque seulement une petite partie des pixels est analysée. Un second avantage est de faciliter les réglages de paramètres d’échelles : quand on s’intéresse à toute l’image, il faut que les paramètres d’échelles soient compatibles avec tout ce qui est observé dans l’image et pas uniquement l’objet d’intérêt. Nous avons utilisé cette stratégie dans les cas des lésions de sclérose en plaques en restreignant l’analyse spatio-temporelle à des régions d’intérêt autour des lésions.

La partie suivante présente une extension de la croissance de région aux données spatio-temporelles.

9.4 Croissance de région spatio-temporelle : STRG

Ces développements ont été faits dans le cadre des travaux de thèse de Clément Daviller (dirigé par Pierre Croisille et Magalie Viallon) et ont été valorisés dans (DAVILLER et al., 2019a).

9.4.1 Introduction

Nous considérons, encore une fois, des séquences d’images d’objets statiques dont les intensités évoluent. Ainsi, à chaque pixel on peut associer une évolution temporelle de ses

2. [git://git.creatis.insa-lyon.fr/STMS.git](https://git.creatis.insa-lyon.fr/STMS.git)

intensités. Il s'agit alors d'une série temporelle.

La segmentation de séries temporelles est englobée dans la problématique de *clustering* spatio-temporel qui est largement couverte dans la littérature (voir les revues (ANSARI et al., 2019 ; SHI et al., 2019)). Ces méthodes sont automatiques (seul le réglage de leurs quelques paramètres est à faire) et il n'y a pas d'interaction avec l'utilisateur pour choisir, à priori, une région.

Dans le cadre des images et pour des approches interactives, (GAMBOTTO, 1992) a été le premier à proposer une segmentation de vidéos basée sur la croissance de région. Il existe maintenant de nombreuses méthodes pour segmenter les vidéos mais, comme pour le filtrage, cette problématique est directement liée à l'estimation du mouvement dans les images.

Dans notre cas, il n'y a pas de mouvement : soit l'objet étudié est naturellement statique, soit il est rendu statique par recalage rigide ou non-rigide (pour le mouvement cardiaque (PETIBON et al., 2013 ; XUE et al., 2008)).

Ainsi, il est possible de définir une approche de croissance de région spatio-temporelle à partir des développements précédents en croissance de région et en filtrage spatio-temporel.

9.4.2 Méthode STRG

Pour formaliser l'approche de croissance de région spatio-temporelle (STRG), on utilise les notations introduites pour STMS, $\mathbf{x}_i = [\mathbf{x}'_{s,i} \ \mathbf{x}'_{t,i}]' \in \mathbf{X}$, où \mathbf{X} représente l'ensemble des pixels.

Soit une région spatiale $R^{[k]}$, on définit l'ensemble des voisins par $\mathbf{x}_{s,i} \in \mathbb{N}(R^{[k]})$. En s'inspirant du filtrage STMS et de la norme infini, le critère permettant d'ajouter des points à la région $R^{[k]}$ est :

$$\|\mathbf{x}_{r,i} - \mathbf{x}_r^{ref}\|_{\infty} \leq h_r \quad (9.18)$$

où le paramètre de l'échelle des amplitudes h_r est directement utilisé comme seuil sur les évolutions temporelles et \mathbf{x}_r^{ref} représente une évolution de référence pour $\mathbf{x}_{r,i}$.

Pour cette référence, il pourra s'agir, par exemple, de la moyenne temps à temps des $\mathbf{x}_{r,i}$ appartenant à $R^{[k]}$. Ce processus de moyenne temps à temps est légitime si on fait l'hypothèse que la région $R^{[k]}$ est constituée de pixels qui partagent un comportement similaire et que l'on souhaite ajouter à cette région uniquement des pixels d'évolution proche. Par récurrence, cela nécessite de partir de germes définissant le comportement souhaité. L'utilisation de la norme infini, très stricte, permet de ne pas s'éloigner de ce comportement.

9.4.3 Application à la segmentation de lésions ischémiques du myocarde

STRG a été utilisé dans le cadre de la segmentation d'image IRM de lésions ischémiques du myocarde pour améliorer les résultats obtenus par *k - means*.

L'ensembles des traitements proposés par (DAVILLER et al., 2019a) sont présentés dans la figure 9.10. Les prétraitements sont détaillés dans (DAVILLER et al., 2019a). Cette méthode nécessite une segmentation préalable du myocarde (ici faite manuellement). Ensuite, les étapes d'extraction des caractéristiques, de sélection des germes puis de segmentation de la lésion sont automatiques. Nous nous focalisons ici sur la problématique de segmentation. Pour cette étude, quatre caractéristiques permettant de décrire globalement l'évolution des intensités dans chaque pixel du myocarde sont calculées à partir des séquences d'images de perfusion IRM :

- P_m , la valeur maximum de l'intensité,
- Δt_{max} : l'instant du maximum d'intensité,

- δ_m : la plus grande augmentation d'intensité,
- AUC : aire sous la courbe de perfusion,

Ces caractéristiques forment un vecteur noté \mathcal{F} .

Sur ces images de caractéristiques, une première segmentation de la lésion est réalisée par un k -means à trois classes. Ensuite, l'approche STRG est appliquée sur la région correspondant à la lésion afin d'affiner cette segmentation. Cependant, STRG utilise, pour les valeurs de $\mathbf{x}_{r,i}$, les intensités des pixels dans les séquences d'images de perfusion IRM et non les quatre caractéristiques globales.

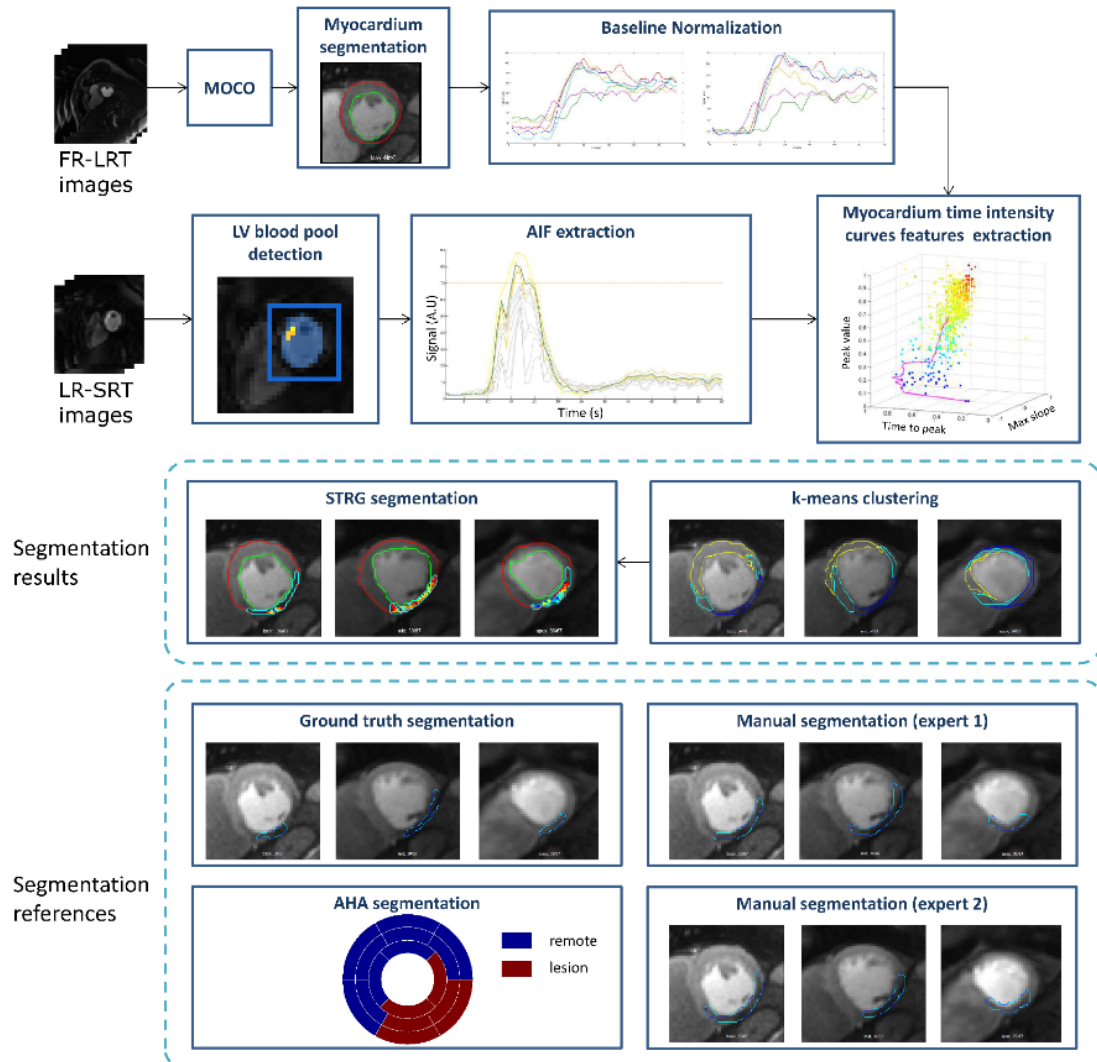


FIGURE 9.10 – Etapes du processus de segmentation des lésions ischémiques en IRM de perfusion.

L'approche a été rendue moins dépendante au paramètre d'échelle h_r en utilisant une hypothèse de ressemblance aux quatre caractéristiques globales de la région d'origine. Ainsi, on va chercher la plus grosse région R qui satisfasse ce critère de ressemblance. Cette recherche est faite en augmentant progressivement le seuil h_r comme détaillé dans l'algorithme 4.

Cette approche a été testé sur 30 patients présentant diverses formes d'ischémie du myocarde et sur trois coupes du myocarde. Six segmentations sont présentées sur la figure 9.11.

Algorithme 4 Algorithme d'amélioration de segmentation basé sur STRG**Entrée:** R région initiale (k -means)**Entrée:** λ seuil de similarité sur les caractéristiques**Sortie:** h_r^* seuil optimal**Sortie:** R_λ^*

- 1: $\overline{\mathbf{x}}_{r, \text{seed}} \leftarrow \frac{1}{|R|} \sum_{i|\mathbf{x}_{s,i} \in R} \mathbf{x}_{r,i}$
- 2: $\mathcal{F}_{\text{seed}} \leftarrow$ calcul des caractéristiques $[P_m, \Delta t_{max}, \delta_m, AUC]$ à partir de $\overline{\mathbf{x}}_{r, \text{seed}}$
- 3: $h_r \leftarrow 0$
- 4: **repeat**
- 5: $h_r \leftarrow h_r + 1$
- 6: $R \leftarrow \text{STRG}(R, h_r)$
- 7: $\overline{\mathbf{x}}_R \leftarrow \frac{1}{|R|} \sum_{i|\mathbf{x}_{s,i} \in R} \mathbf{x}_{r,i}$
- 8: $\mathcal{F}_R \leftarrow$ calcul des caractéristiques $[P_m, \Delta t_{max}, \delta_m, AUC]$ à partir de $\overline{\mathbf{x}}_R$
- 9: **until** $\frac{\|\mathcal{F}_R - \mathcal{F}_{\text{seed}}\|}{\|\mathcal{F}_{\text{seed}}\|} < \lambda$
- 10: $h_r^* \leftarrow h_r$
- 11: $R_\lambda^* \leftarrow R$
- 12: **return** R_λ^*, h_r^*

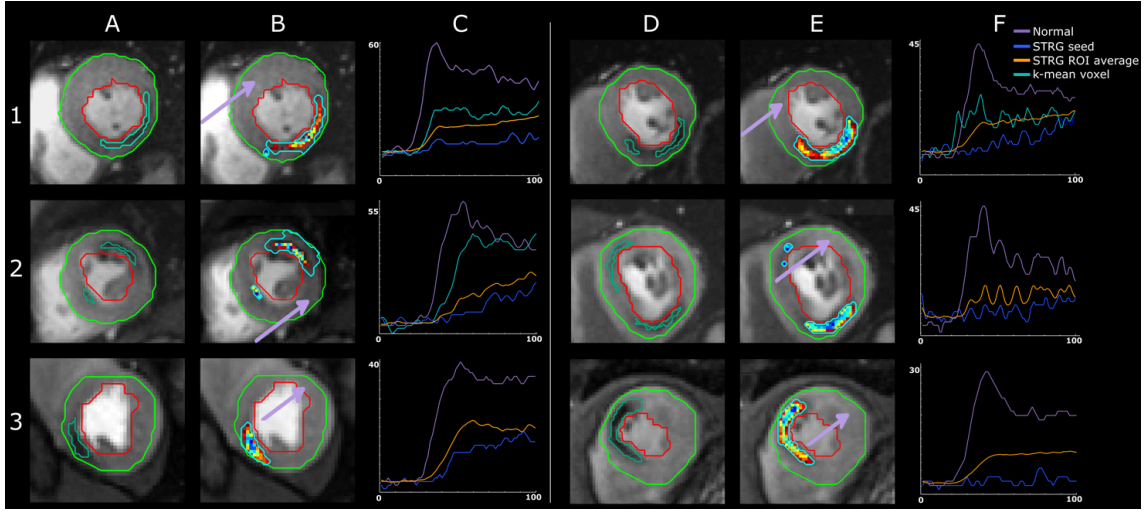


FIGURE 9.11 – Segmentations de 6 patients avec l'approche proposée dans (DAVILLER et al., 2019a), (A) et (D) segmentations manuelles du myocarde (vert et rouge) et de la lésion (vert foncé), (B) et (E) résultats de segmentation par k -means (cyan) et STRG (pixels en couleurs qui représente la valeur de h_r pour laquelle le pixel a été inclus dans la région), (C) et (F) donnent les évolutions des intensités dans les différentes régions, la flèche violette désigne la région normale.

Les meilleurs scores de DICE ont été obtenus pour $\lambda = 0.5, 0.6$ et étaient de l'ordre de 0.64 quelle que soit la forme pathologique et la coupe étudiée, STRG conduisant à une sur-segmentation des lésions. Ce faible DICE s'explique par la petite taille de la région recherchée et par la difficulté à définir les bords de la lésion malgré le fait que la vérité terrain soit un consensus de deux experts.

Pour évaluer l'apport de STRG pour cette application et critiquer ce faible DICE, la mesure du flux sanguin myocardique (MBF) a été calculée pour six propositions de lésions obtenues par : le consensus, les 2 experts individuellement, les segmentations STRG et k -means et via le schéma standard de l'American Heart Association (AHA). La variabilité observée pour l'approche STRG est la plus similaire à celle obtenue par le consensus d'experts. Ce

qui permet d'affirmer que, en terme d'activité fonctionnelle des pixels, les sur-segmentations proposées par STRG sont très homogènes et compatibles avec les celles définies par le consensus, et ainsi qu'il est possible que les sur-segmentations proposées par STRG soient cohérentes.

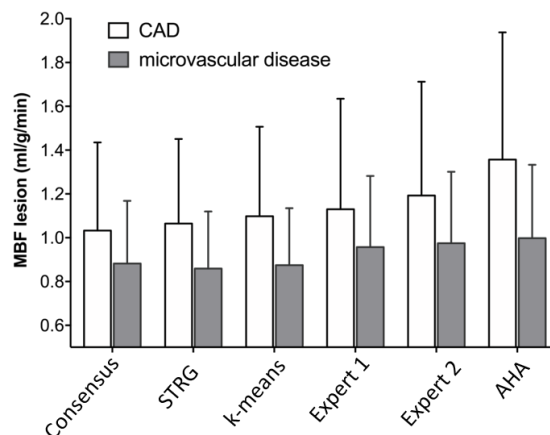


FIGURE 9.12 – Etudes de la variabilité des valeurs de MBF, pour deux types d'atteintes (micro-vasculaire et des artères coronaires (CAD)) calculées dans les régions des lésions segmentées par : le consensus, STRG, k-means, les deux experts médicaux individuellement et les segments du modèle AHA.

9.4.4 Conclusion

Nous avons proposé ici une contribution à la croissance de région lui permettant de traiter des séries temporelles. Cette contribution s'appuie sur le même formalisme que celui utilisé pour *mean-shift* spatio-temporel.

Utilisé pour segmenter des lésions ischémiques sur des IRM de perfusion, cette approche a permis, sur 30 patients, d'obtenir des résultats très cohérents au sens des évolutions des intensités.

On note que pour cette méthode, une étape manuelle de segmentation du myocarde est nécessaire. Ensuite, la méthode est complètement automatique.

9.5 Conclusion du chapitre

Dans ce chapitre, nous avons présenté nos contributions au filtrage d'images spatio-temporel ($AD+t$, $RAD+t$ et $STMS$) ainsi qu'à la segmentation par croissance de région (STRG). Nous avons pour cela utilisé le formalisme de l'espace des caractéristiques et des paramètres d'échelles introduits au chapitre précédent. Ainsi, nous avons pu nous appuyer sur le formalisme *mean-shift* dont les performances en filtrage et en clustering sont parmi les meilleures.

L'utilisation de paramètres d'échelles rend la compréhension des paramètres des méthodes assez intuitive. Cependant, une certaine expertise est nécessaire pour appréhender la contribution de la composante temporelle dans ces approches.

Ces méthodes non-supervisées sont automatiques, une fois le réglage des paramètres effectué, voir interactives dans le cas de la croissance de région quand le germe ne peut être positionné automatiquement. Mais elles ne permettent pas facilement d'intégrer des a priori diverses tels que la forme ou la localisation des objets.

Ces approches originales ont montré leurs intérêts sur différentes données médicales (suivi IRM, IRM de perfusion, multi-spectral) et pour différentes applications (SEP, ischémie cardiovasculaire, caractérisation de tissus). L'approche *mean-shift* spatio-temporel a aussi été utilisée avec succès sur des images satellitaires et des vidéos.



10. Conclusion de la seconde partie

Les méthodes vues dans cette partie tirent toutes leur formalisme de l'espace des caractéristiques et des paramètres d'échelles qui ont été étendus pour prendre en compte la dimension temporelle. Avec seulement le réglage de quelques paramètres, ces méthodes ont montré leur efficacité sur des données médicales en filtrage et segmentation d'images, pouvant inclure plusieurs composantes d'intensité ainsi que l'évolution temporelle.

Ces approches permettent d'obtenir rapidement des segmentations avec très peu de connaissances. De part leur caractère non-supervisées, elles permettent aussi de chercher dans les images des relations inconnues. Enfin, ces approches, avec l'utilisation des paramètres d'échelles, sont des approches *bottom-up* qui effectuent des regroupements de pixels et qui pourront être plus tard fusionnées ou séparées par une approche bénéficiant de plus d'information ou connaissance.

Cependant, malgré le fait que les paramètres soient assez intuitifs, leur réglage peut nécessiter une certaine expertise, surtout pour la dimension temporelle. Il faut aussi souligner que pour une étude rigoureuse d'une cohorte, ces paramètres doivent être les mêmes pour tous les patients. Or, pour garantir cette stabilité, il faut apporter une très grande attention aux pré-traitements et particulièrement au recalage et à la normalisation des intensités.

Une autre approche est de déterminer les valeurs des paramètres d'échelles automatiquement comme nous l'avons étudié avec le *plug in* dans 8.3 pour le filtrage et les cartes de probabilité dans 8.4.3.

Pour s'assurer de bonnes performances en filtrage et segmentation pour des paramètres d'échelles donnés, il est recommandé de restreindre les traitements à des régions d'intérêt. En effet, pour avoir un fonctionnement homogène sur tout un volume, le réglage des paramètres d'échelles peut s'avérer extrêmement complexe. Dans les études précédentes, on avait par exemple la définition d'un volume d'intérêt autour des lésions SEP ou la localisation du coeur et, ensuite, la segmentation du myocarde pour le travail sur l'ischémie. Or, ces tâches demandent soit un travail manuel, soit une expertise ou des connaissances a priori pour guider ces méthodes. Intégrer efficacement ces a priori dans les méthodes

non-supervisées n'est pas évident et est presque un non-sens.

Ainsi, la suite de mes recherches se tourne vers les méthodes supervisées et semi-supervisées qui font l'objet de la troisième partie de ce manuscrit.

Contributions personnelles pour cette partie

Articles

- AMELI, R., S. MURE, C. R. GUTTMANN, T. **Grenier**, H. BENOIT-CATTIN et F. COTTON (2016). « Analyse dynamique hebdomadaire du développement péri-veinulaire des lésions actives de SEP par imagerie de susceptibilité magnétique. » In : *Journal of Neuroradiology* 43.2, p. 91-93.
- BANJAK, H., T. **Grenier**, T. EPICIER, S. KONETI, L. ROIBAN, A.-S. GAY, I. MAGNIN, F. PEYRIN et V. MAXIM (2018). « Evaluation of noise and blur effects with SIRT-FISTA-TV reconstruction algorithm : Application to fast environmental transmission electron tomography ». In : *Ultramicroscopy* 189, p. 109-123.
- DAVILLER, C., T. **Grenier**, H. RATINEY, M. SDIKA, P. CROISILLE et M. VIALON (2019a). « Automatic myocardial ischemic lesion detection on magnetic resonance perfusion weighted imaging prior perfusion quantification : A pre-modeling strategy ». In : *Computers in Biology and Medicine*, p. 108-119.
- DOLET, A., F. VARRAY, S. MURE, T. **Grenier**, Y. LIU, Z. YUAN, P. TORTOLI et D. VRAY (2018). « Spatial and spectral regularization to discriminate tissues using multispectral photoacoustic imaging ». In : *EURASIP Journal on Advances in Signal Processing* 2018.1, p. 39.
- MURE, S., T. **Grenier**, D. S. MEIER, C. R. GUTTMANN et H. BENOIT-CATTIN (2015b). « Unsupervised spatio-temporal filtering of image sequences. A mean-shift specification ». In : *Pattern Recognition Letters* 68, Part 1, p. 48-55.

Conférences

- CHARPIGNY, D., T. **Grenier**, C. ODET et H. BENOIT-CATTIN (2009). « Towards iron oxide nanoparticles quantization in molecular MR images by default field deconvolution ». In : Boston, MA, United states, p. 137-140.
- (2010). « Restoration-based iron oxide particles quantification in MR images ». English. In : Dallas, TX, United states, p. 485-488.
- CHARPIGNY, D., T. **Grenier**, C. ODET, J. PAULY et H. BENOIT-CATTIN (2008). « Deconvolution approach for susceptibility map building ». In : *European Society for Magnetic Resonance in Medicine and Biology, 25th Annual Scientific Meeting ESMRMB*. Valencia, Spain, in-press.
- DOLET, A., F. VARRAY, S. MURE, T. **Grenier**, Y. LIU, Z. YUAN, P. TORTOLI et D. VRAY (2016). « Spatial and spectral regularization for multispectral photoacoustic image clustering ». In : *IEEE International Ultrasonics Symposium*. Proceedings of the 2016 IEEE International Ultrasonics Symposium (IUS). Tours, France.
- LI, T., S. CAMARASU-POP, T. GLATARD, T. **Grenier** et H. BENOIT-CATTIN (2010a). « Optimization of mean-shift scale parameters on the EGEE grid ». In : *Studies in health technology and informatics, Proceedings of Healthgrid 2010*. T. 159, p. 203-214.
- LI, T., T. **Grenier** et H. BENOIT-CATTIN (2011). « Color space influence on mean shift filtering ». In : *Proceedings - International Conference on Image Processing, ICIP*. Brussels, Belgium, p. 1469-1472.
- MURE, S., T. **Grenier** et H. BENOIT-CATTIN (2016a). « Unsupervised spatiotemporal video clustering a versatile mean-shift formulation robust to total object occlusions ».

-
- In : *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 1536-1540.
- MURE, S., T. **Grenier**, C. R. G. GUTTMANN et H. BENOIT-CATTIN (2016b). « Unsupervised time-series clustering of distorted and asynchronous temporal patterns ». In : *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 1263-1267.
- MURE, S., T. **Grenier**, C. R. G. GUTTMANN, F. COTTON et H. BENOIT-CATTIN (2016c). « Classification of multiple sclerosis lesion evolution patterns a study based on unsupervised clustering of asynchronous time-series ». In : *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, p. 1315-1319.
- MURE, S., C. R. G. GUTTMANN, T. **Grenier**, H. BENOIT-CATTIN et F. COTTON (2016d). « New insight in perivenular lesion formation in multiple sclerosis on weekly susceptibility weighted images ». In : *2016 International Society for Magnetic Resonance in Medicine*.
- PORTEJOIE, P., S. MURE, H. BENOIT-CATTIN et T. **Grenier** (2015). « Locally controlled regularized spatiotemporal anisotropic diffusion ». In : *Image Processing (ICIP), 2015 IEEE International Conference on*. Québec, Canada, p. 4823-4827.
- REVOL-MULLER, C., T. **Grenier**, T. LI et H. BENOIT-CATTIN (2012a). « Feature space region growing ». In : *2012 19th IEEE International Conference on Image Processing*. Orlando, United States, p. 2585-2588.
- ROSE, J. L., T. **Grenier**, C. REVOL-MULLER et C. ODET (2010). « Unifying variational approach and region growing segmentation ». In : *Signal Processing Conference, 2010 18th European*, p. 1781-1785.



Bibliographie de la seconde partie

- AGGARWAL, C. et C. REDDY (2013). *Data clustering : Algorithms and applications*. CRC Press.
- AI, L., X. GAO et J. XIONG (2014). « Application of mean-shift clustering to Blood oxygen level dependent functional MRI activation detection ». In : *BMC Medical Imaging* 14.1, p. 6.
- ANJUM, N. et A. CAVALLARO (2008). « Multifeature object trajectory clustering for video analysis ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 18.11, p. 1555-1564.
- ANSARI, M. Y., A. AHMAD, S. KHAN, G. BHUSHAN et MAINUDDIN (2019). « Spatiotemporal clustering : a review ». In : *Artificial Intelligence Review* 53, p. 2381-2423.
- ARDIZZONE, E., R. PIRRONE et O. GAMBINO (2003). « Automatic segmentation of MR images based on adaptive anisotropic filtering ». In : *Image Analysis and Processing, 2003.Proceedings. 12th International Conference on*, p. 283-288.
- ASHOK, S., G. MAYANK, A. FA et L. CHEEMUN (2006). « State-of-the-art imaging of acute stroke ». In : *Stroke* 26(suppl 1), S75-S95.
- BARASH, D. (2002). « Fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.6, p. 844-847.
- BARASH D. Comaniciu, D. (2004). « A common framework for nonlinear diffusion, adaptive smoothing, bilateral filtering and mean shift ». In : *Image and Vision Computing* 22.1, p. 73-81.
- BENKERT, M., J. GUDMUNDSSON, F. HÜBNER et T. WOLLE (2008). « Reporting flock patterns ». In : *Computational Geometry* 41.3, p. 111-125.
- BHAGAVATHY, S. et J. LLACH (2007). « Adaptive spatio-temporal video noise filtering for high quality applications ». In : *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*. T. 1, p. I-761-I-764.

- BOULANGER, J., C. KERVRANN et P. BOUTHEMY (2007). « Space-time adaptation for patch-based image sequence restoration ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29.6, p. 1096-1102.
- BOULANGER, J., C. KERVRANN, P. BOUTHEMY, P. ELBAU, J.-B. SIBARITA et J. SALAMERO (2010). « Patch-based nonlocal functional for denoising fluorescence microscopy image sequences ». In : *Medical Imaging, IEEE Transactions on* 29.2, p. 442-454.
- BRANKOV, J., M. WERNICK, Y. YANG et M. NARAYANAN (2000). « Spatially-adaptive temporal smoothing for reconstruction of dynamic and gated image sequences ». In : *2000 IEEE Nuclear Science Symposium. Conference Record (Cat. No.00CH37149)*. T. 2. IEEE, p. 15/146-15/150.
- BUADES, A., B. COLL et J.-M MOREL (2005a). « Denoising image sequences does not require motion estimation ». In : *Advanced Video and Signal Based Surveillance, 2005. AVSS 2005. IEEE Conference on*, p. 70-74.
- BUADES, A., B. COLL et J.-M. MOREL (2008). « Nonlocal image and movie denoising ». English. In : *International Journal of Computer Vision* 76.2, p. 123-139.
- BUADES A. Coll, B. et J. MOREL (2005b). « A review of image denoising algorithm, with a new one ». In : *Multiscale Modeling & Simulation* 4(2), p. 490-530.
- CABRIA, I. et I. GONDRA (2012). « A mean shift-based initialization method for k-means ». In : *2012 IEEE 12th International Conference on Computer and Information Technology*, p. 579-586.
- CARREIRA-PERPINAN, M. A. (2007). « Gaussian mean-shift is an EM algorithm ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.5, p. 767-776.
- CHAMBOLLE, A. (2004). « An algorithm for total variation minimization and applications ». English. In : *Journal of Mathematical Imaging and Vision* 20.1-2, p. 89-97.
- CHEN, J. et C.-K. TANG (2007). « Spatio-temporal Markov random field for video denoising ». In : *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, p. 1-8.
- CHENG, J., F. SHI, K. WANG, M. SONG, J. JIANG, L. XU et T. JIANG (2009). « Nonparametric mean shift functional detection in the functional space for task and resting-state fMRI ». In : *Workshop on fMRI data analysis : statistical modeling and detection issues in intra- and inter-subject functional MRI data analysis, in conjunction with the MICCAI 2009*. London, United Kingdom.
- CHENG, Y. (1995). « Mean shift, mode seeking, and clustering ». In : *Pattern Analysis and Machine Intelligence, IEEE Transaction on* 17.8, p. 790-799.
- COMANICIU, D. et P. MEER (1997). « Robust analysis of feature spaces : Color image segmentation ». In : *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Juan, PR, USA, p. 750-755.
- (2002). « Mean shift : A robust approach toward feature space analysis ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.5, p. 603-619.
- DABOV, K., A. FOI, V. KATKOVNIK et K. EGIAZARIAN (2007). « Image denoising by sparse 3-D transform-domain collaborative filtering ». In : *Image Processing, IEEE Transactions on* 16.8, p. 2080-2095.
- DEMENTHON, D. et D. DOERMANN (2005). « Video retrieval of near-duplicates using k-nearest neighbor retrieval of spatiotemporal descriptors ». In : *Multimedia Tools Applicat.*
- DICE, L. R. (1945). « Measures of the amount of ecologic association between species ». In : *Ecology* 26.3, p. 297-302.
- DUONG Tarn ; Hazelton, M. (2003). « Plug-in bandwidth matrices for bivariate kernel density estimation ». In : *Journal of Nonparametric Statistics Volum* 15(1), p. 17-30.

- DUONG, T (2007). « ks. Kernel density estimation and kernel discriminant analysis for multivariate data in R ». In : *J Stat Soft* 21.7, p. 1-16.
- FASHING, M. et C. TOMASI (2005). « Mean shift is a bound optimization ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27.3, p. 471-474.
- FENG, W. et R.-C. ZHAO (2003). « Non-rigid objects detection and segmentation in video sequence using 3D mean shift analysis ». In : *International Conference on Machine Learning and Cybernetics*. T. 5, p. 3134-3139.
- FUKUNAGA, K. et L. D. HOSTETLER (1975). « Estimation of the gradient of a density function with applications in pattern recognition. » In : *IEEE Transaction on Information Theory* 21.1, p. 32-40.
- GAMBOTTO, J. (1992). « A region-based spatio-temporal segmentation algorithm ». In : *11th IAPR International Conference on Pattern Recognition. Vol. III. Conference C : Image, Speech and Signal Analysis*, t. 1. Los Alamitos, CA, USA : IEEE Computer Society, p. 189-192.
- GONZALEZ, R. (2011). *Acute ischemic stroke : imaging and intervention*. Sous la dir. de L. SCHWAMM. Springer.
- GUTTMANN, C., S. S. AHN, L. HSU, R. KIKINIS et F. A. JOLESZ (1995). « The evolution of multiple sclerosis lesions on serial MR. » In : *American journal of neuroradiology* 16.7, p. 1481-1491.
- HAUTAMAKI, V., P. NYKANEN et P. FRANTI (2008). « Time-series clustering by approximate prototypes ». In : *19th International Conference on Pattern Recognition*, p. 1-4.
- HE, L. et R. GREENSHIELDS IAN (2009b). « A nonlocal maximum likelihood estimation method for Rician noise reduction in MR images ». In : *Medical Imaging, IEEE Transactions on* 28.2, p. 165-172.
- JEUNG, H., M. L. YIU, X. ZHOU, C. S. JENSEN et H. T. SHEN (2008). « Discovery of convoys in trajectory databases ». In : *Proceedings of the Very Large Data Bases Endowment* 1.1, p. 1068-1080.
- JIANG, X., H. YAO, S. ZHANG, X. LU et W. ZENG (2013). « Night video enhancement using improved dark channel prior ». In : *Image Processing (ICIP), 2013 20th IEEE International Conference on*, p. 553-557.
- KALPAKIS, K., D. GADA et V. PUTTAGUNTA (2001). « Distance measures for effective clustering of ARIMA time-series ». In : *Data Mining, IEEE Proceedings of the International Conference on*, p. 273-280.
- KE, Y., R. SUKTHANKAR et M. HEBERT (2005). « Efficient temporal mean shift for activity recognition in video ». In : *19th Annual Conference on Neural Information Processing Systems Workshop on Activity Recognition and Discovery, Whistler, Canada*. Citeseer.
- KENG, H. L. et W. YUAN (1981). *Applications of number theory to numerical analysis*. Springer Berlin Heidelberg.
- KUIPERS L. Niederreiter, H (1974). *Uniform distribution of sequences*. A WILEY-INTERSCIENCE PUBLICATION.
- LEUNG, K., N. SAEED, K. CHANGANI, S. CAMPBELL et D. HILL (2006). « Spatio-temporal segmentation of rheumatoid arthritis lesions in serial MR images of joints ». In : *Computer Vision and Pattern Recognition Workshop, 2006. IEEE Conference on*, p. 91-91.
- LI, X et Y. ZHENG (2009). « Patch-based video processing : A variational Bayesian approach ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 19.1, p. 27-40.

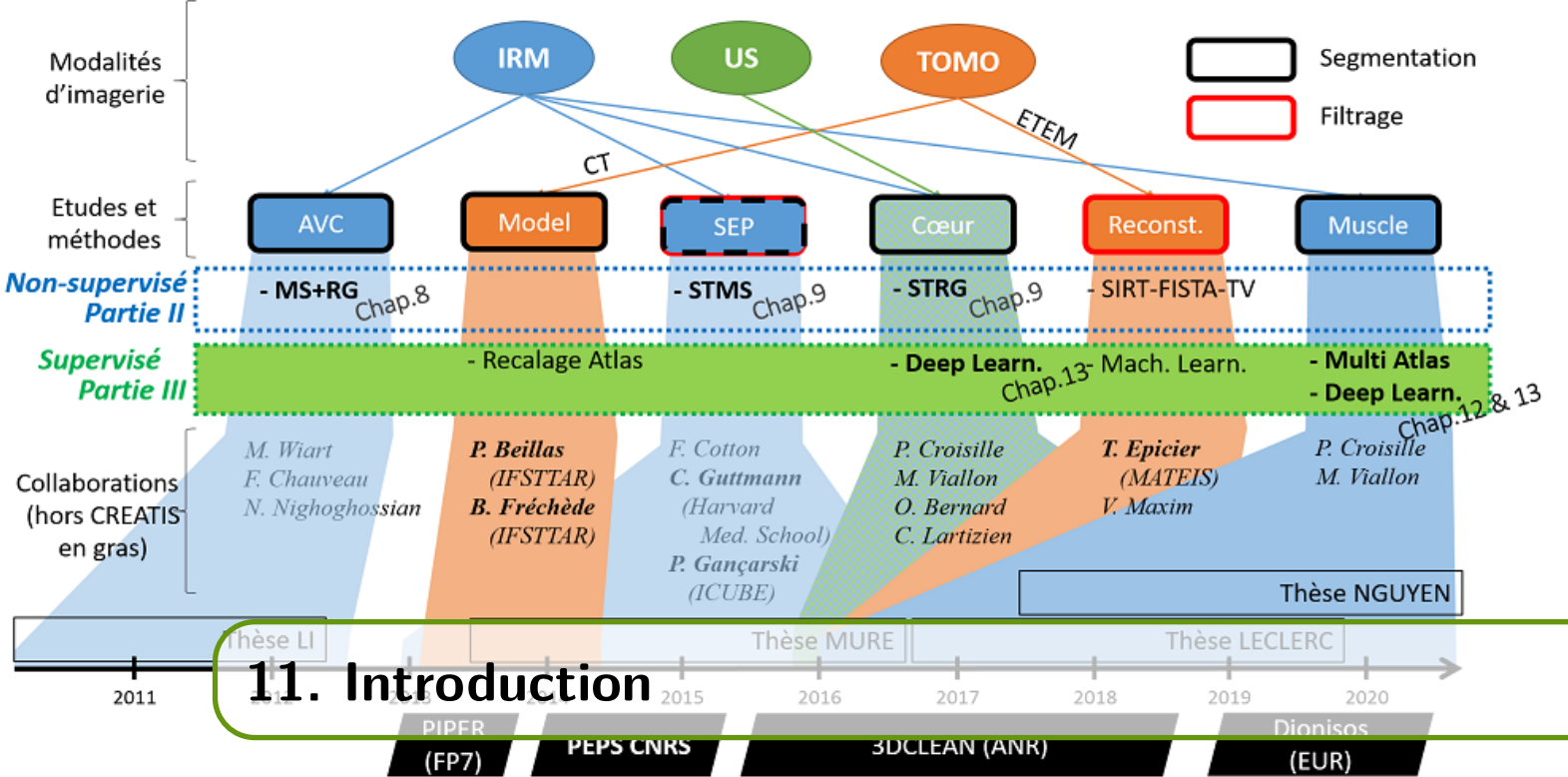
- LI, Z., B. DING, J. HAN et R. KAYS (2010b). « Swarm : Mining relaxed temporal moving object clusters ». In : *Proceedings of the Very Large Data Bases Endowment* 3.1-2, p. 723-734.
- MAGGIONI, M., G. BORACCHI, A. FOI et K. EGIAZARIAN (2012). « Video denoising, deblocking, and enhancement through separable 4-D nonlocal spatiotemporal transforms ». In : *Image Processing, IEEE Transactions on* 21.9, p. 3952-3966.
- MAGGIONI, M., E. SANCHEZ-MONGE et A. FOI (2014). « Joint removal of random and fixed-pattern noise through spatiotemporal video filtering ». In : *Image Processing, IEEE Transactions on* 23.10, p. 4282-4296.
- MALM, H., M. OSKARSSON, E. WARRANT, P. CLARBERG, J. HASSELGREN et C. LEJDFORS (2007). « Adaptive enhancement and noise reduction in very low light-level video ». In : *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, p. 1-8.
- MATSUBARA, Y., Y. SAKURAI et C. FALOUTSOS (2014). « AutoPlait : Automatic mining of co-evolving time sequences ». In : *Proceedings of the ACM SIGMOD International Conference on Management of Data. SIGMOD '14*. Snowbird, Utah, USA : ACM, p. 193-204.
- MEESRIKAMOLKUL, W., V. NIENNATTRAKUL et C. A. RATANAMAHATANA (2012). « Shape-Based clustering for time series data ». In : *Advances in knowledge discovery and data mining*. Springer, p. 530-541.
- MEIER, D. et C. GUTTMANN (2003). « Time-series analysis of MRI intensity patterns in multiple sclerosis ». In : *NeuroImage* 20.2, p. 1193-1209.
- MONTAGNAT, J., M. SERMESANT, H. DELINGETTE, G. MALANDAIN et N. AYACHE (2003). « Anisotropic filtering for model-based segmentation of 4D cylindrical echocardiographic images ». Anglais. In : *Pattern Recognition Letters - Special Issue on Ultrasonic Image Processing and Analysis* 24.4-5, p. 815-828.
- NIEDERREITER, H. (1992). *Random number generation and quasi-Monte Carlo methods*. SOCIETY FOR INDUSTRIAL et APPLIED MATHEMATICS.
- OSTERGAARD, L. (2005). « Principles of cerebral perfusion imaging by bolus tracking ». In : *JOURNAL OF MAGNETIC RESONANCE IMAGING* 22(6), p. 710-717.
- PAPADIMITRIOU, S., J. SUN et C. FALOUTSOS (2007). « Dimensionality reduction and forecasting on streams ». In : *Data Streams*. Springer, p. 261-288.
- PARZEN, E. (1962). « On estimation of a probability density function and mode ». In : *The annals of mathematical statistics* 33.3, p. 1065-1076.
- PERONA, P. et J. MALIK (1990). « Scale-space and edge detection using anisotropic diffusion ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 12.7, p. 629-639.
- PETIBON, Y., J. OUYANG, X. ZHU, C.-C. HUANG, T. G. REESE, S. Y. CHUN, Q. LI et G. E. FAKHRI (2013). « Cardiac motion compensation and resolution modeling in simultaneous PET-MR : a cardiac lesion detection study. » In : *Physics in medicine and biology* 58 7, p. 2085-102.
- PRASATH, V. S. et D. VOROTNIKOV (2014). « Weighted and well-balanced anisotropic diffusion scheme for image denoising and restoration ». In : *Nonlinear Analysis : Real World Applications* 17.0, p. 33 -46.
- RAO, S., A. de MEDEIROS MARTINS et J. C. PRINCIPE (2009). « Mean shift : An information theoretic perspective ». In : *Pattern Recognition Letters* 30.3, p. 222-230.
- SHEPP, L et B. F. LOGAN (1974). « The fourier reconstruction of a head section ». In : *Nuclear Science, IEEE Transactions on* 21, p. 21-43.
- SHI, Z. et L. S. C. PUN-CHENG (2019). « Spatiotemporal data clustering : A survey of methods ». In : *ISPRS Int. J. Geo-Information* 8, p. 112.

-
- TOMASI, C. et R. MANDUCHI (1998). « Bilateral filtering for gray and color images ». In : *Computer Vision, 1998. Sixth International Conference on*, p. 839-846.
- VARGHESE, G. et Z. WANG (2010). « Video denoising based on a spatiotemporal gaussian scale mixture model ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 20.7, p. 1032-1040.
- WAND, M. et M. JONES (1995). *Kernel smoothing*. Monographs on statistics and applied probability. Chapman & Hall.
- WANG, J., B. THIESSON, Y. XU et M. COHEN (2004a). « Image and video segmentation by anisotropic kernel mean shift ». In : *Computer Vision-ECCV 2004*. Springer, p. 238-249.
- WANG, Z., A. BOVIK, H. SHEIKH et E. SIMONCELLI (2004b). « Image quality assessment : from error visibility to structural similarity ». In : *IEEE Transactions on Image Processing* 13.4, p. 600-612.
- XUE, H., J. GUEHRING, L. SRINIVASAN, S. ZUEHLSORFF, K. A. SADDI, C. CHEFD'HOTEL, J. V. HAJNAL et D. RUECKERT (2008). « Evaluation of rigid and non-rigid motion compensation of cardiac perfusion MRI ». In : *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention* 11 Pt 2, p. 35-43.
- ZAFAR, S. (2008). « Perfusion imaging in ischaemic stroke ». In : *Journal of Pakistan Medical Association* 58(7), p. 391-394.
- ZHENG, K., Y. ZHENG, N. J. YUAN et S. SHANG (2013). « On discovery of gathering patterns from trajectories ». In : *Data Engineering, IEEE Proceedings of the 29th International Conference on*, p. 242-253.
- ZUCKER, S. (1976). « Region growing : Childhood and adolescence ». In : *Computer Graphics and Image Processing* 5(3), p. 382-399.



Contributions en segmentation supervisée

11	Introduction	115
12	Segmentation par recalage d'atlas	117
12.1	Introduction	
12.2	Principe du recalage d'atlas	
12.3	Approche multi atlas $JLF + CL$	
12.4	Conclusion	
13	Réseau U-Net	125
13.1	Introduction	
13.2	Segmentation par apprentissage automatique et profond	
13.3	Architecture U-Net	
13.4	Les coulisses d'U-Net	
13.5	Quelques variantes d'U-Net	
13.6	Conclusion	
14	Applications à l'imagerie médicale	137
14.1	Introduction	
14.2	Segmentation des muscles en IRM avec JLF+CL	
14.3	Segmentations avec U-Net	
14.4	Contribution à la correction automatique	
14.5	Conclusion	
15	Augmentation de données	155
15.1	Introduction	
15.2	Augmentations d'images par recalages	
15.3	Sélection d'atlas basée sur la morphologie	
15.4	Conclusion	
16	Conclusion de la troisième partie	165



11. Introduction

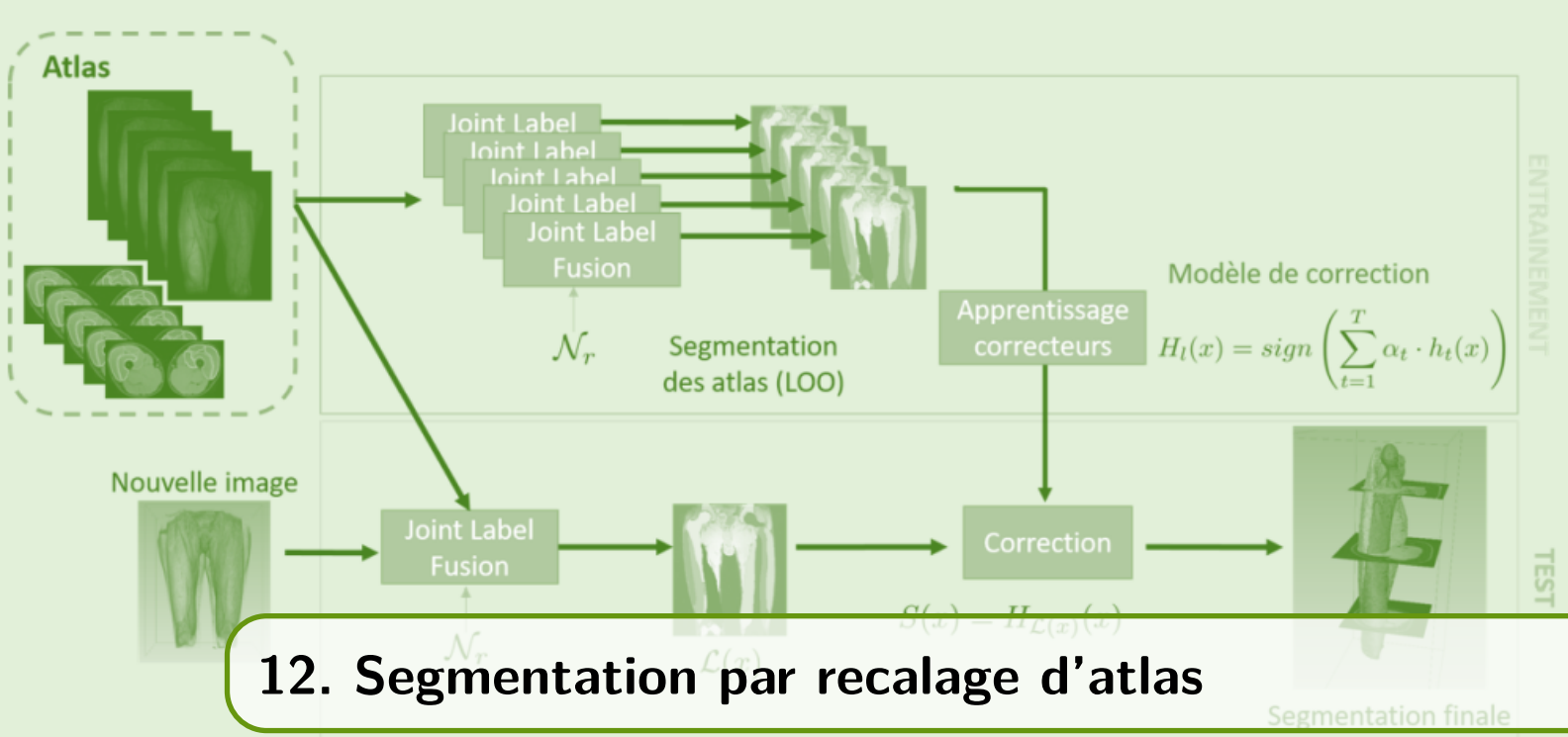
La partie précédente a montré l'efficacité des approches non-supervisées pour la segmentation et le filtrage d'images médicales. Elle a aussi montré leurs limites : les réglages des paramètres et le besoin d'intégrer des a priori dans les formalismes.

Plusieurs projets nous ont conduit à considérer des approches plus automatiques et de nouvelles pistes pour atteindre les objectifs attendus en clinique, notamment pour la segmentation d'images écho-cardiographiques et la segmentation de muscles en IRM.

Dans cette partie, nous présentons nos travaux en segmentation supervisée. Deux approches ont été étudiées : la segmentation par recalage d'atlas et par réseaux de neurones artificiels. Ces deux approches extraient des images et de segmentations de référence, des connaissances sur la géométrie, l'organisation spatiale et les textures des objets à segmenter. Ces images de références sont appelées atlas ou images annotées mais correspondent à la même chose. Il s'agit toujours d'un couple : une image d'intensité $A_F(\mathbf{x})$, aussi appelée atlas d'intensité ou image de référence, et une image étiquetée $A_S(\mathbf{x})$, ou image annotée ou encore segmentée. Pour les deux approches étudiées, ces atlas sont critiques mais pour des raisons différentes. Pour la première, leur nombre doit resté petit sous peine de demander un temps de calcul énorme comme nous le verrons par la suite. A l'inverse, les réseaux de neurones doivent disposer de beaucoup d'images annotées afin d'apprendre à reproduire la tâche de segmentation. Dans les deux cas, il y a un enjeu à bien sélectionner les images utiliser comme atlas et pour d'entraînement.

Dans le chapitre 12, les approches de segmentations basées sur le multi-atlas sont présentées. Le chapitre 13 expose la famille de réseaux de neurones utilisée. Les études médicales conduites avec ces approches sont décrites dans le chapitre 14. Enfin, le chapitre 15 traite des problématiques d'augmentation de données et des contributions proposées.

Cette partie concerne des travaux réalisés au cours des thèses de Sarah Leclerc et Hoai-Thu Nguyen. Certains sont en cours de valorisation.



12.1 Introduction

Dans ce chapitre, nous nous intéressons à la segmentation par recalage d'atlas et multi-atlas. Le recalage d'images est une technique qui a montré sa robustesse quant aux changements dans les images (résolution, modalité, zones imagées, ...). Les approches de segmentation par recalage d'atlas sont particulièrement efficaces pour proposer des segmentations de qualité lorsque peu de données annotées sont disponibles (on parle de 1 à 10 atlas).

Nous commençons par rappeler les principes du recalage et l'intérêt d'utiliser plusieurs atlas (section 12.2). Cependant, cette multiplicité des atlas peut conduire à différents biais et difficultés pour lesquels (WANG et al., 2013b) propose une solution élégante et efficace que nous détaillons dans la section 12.3.

Nos applications seront décrites dans le chapitre 14.

12.2 Principe du recalage d'atlas

La segmentation par recalage d'atlas est largement couverte dans les chapitres des livres (BACH CUADRA et al., 2015; ROHLFING et al., 2005). La figure 12.1 en résume le principe. Un atlas A sera le regroupement de $A_F(\mathbf{x})$ l'image d'intensité, et de $A_S(\mathbf{x})$ la segmentation manuelle : $A = (A_F, A_S)$.

Tout d'abord, rappelons que le recalage d'image consiste à mettre en correspondance deux images : F l'image fixe ou de référence, et M , l'image à déformer. Cela correspond à déterminer la transformation T^* des coordonnées spatiales qui maximise une mesure de similarité \mathcal{S} entre F et M déformée par une transformation T de \mathcal{T} .

$$T^* = \arg \max_{T \in \mathcal{T}} \mathcal{S}(F, M; T) \quad (12.1)$$

Il existe de nombreuses mesures de similarité \mathcal{S} et de types de transformations spatiales \mathcal{T} . Pour les mesures de similarités \mathcal{S} , dans le cadre du recalage basé sur les intensités de l'image, les plus courantes sont basées sur l'erreur quadratique pour des images de même modalité, et sur des ratios de corrélation ou des mesures issues de la théorie de l'information

(principalement : l'information mutuelle) pour des images provenant de modalités différentes (MAES et al., 2015). Dans la suite, nous utilisons la mesure d'information mutuelle, même si les modalités des images sont les mêmes, car nous avons observé qu'elle permettait d'obtenir des recalages satisfaisants pour nos différents projets.

Pour les transformations spatiales \mathcal{T} , nous retenons les trois suivantes :

- la transformation rigide qui permet d'effectuer un recalage en translation et en rotation (soit 6 paramètres en 3D),
- la transformation affine qui ajoute à la transformation rigide, des échelles et des cisaillements suivant chacune des dimensions (12 paramètres en 3D),
- la transformation déformable basée sur des B-splines qui va déformer localement l'image (RUECKERT et al., 1999).

La principale difficulté du recalage d'atlas est de mettre en correspondance le plus précisément possible l'image à segmenter avec celle de l'atlas. Ce besoin de précision implique d'effectuer un recalage déformable. Pour que ce recalage déformable soit robuste, il est recommandé de recalcr préalablement les images avec une transformation rigide puis avec une transformation affine.

Ces différentes transformations sont ensuite appliquées à la segmentation manuelle en utilisant, pour son interpolation, une méthode ne créant pas de nouvelles étiquettes comme les k plus proches voisins. Avec un seul atlas, des segmentations multi-organes pertinentes peuvent être obtenues, comme dans le cas des os en scanner X (MOREAU et al., 2016).

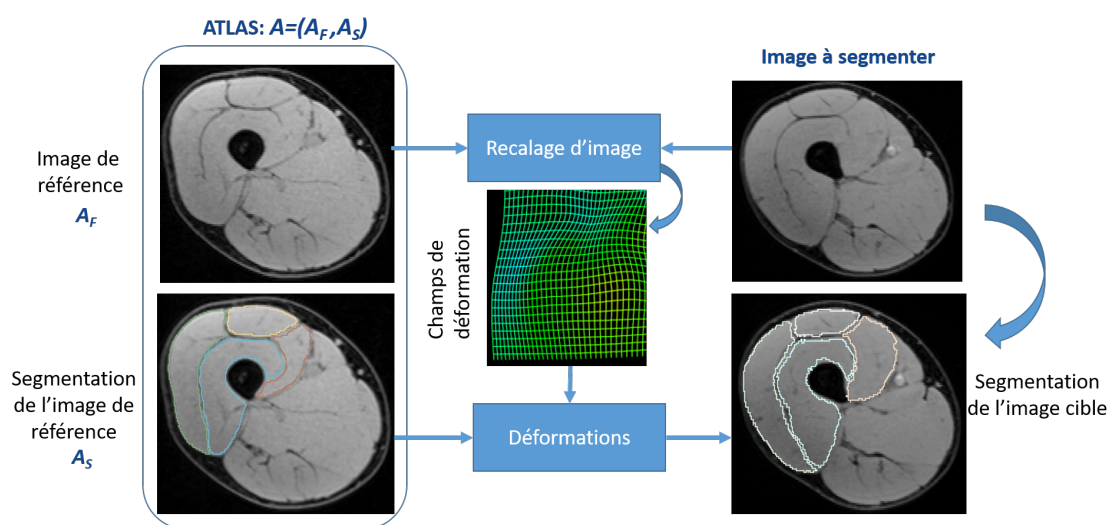


FIGURE 12.1 – Principe de la segmentation par recalage d'atlas. L'image atlas (à gauche) est recalée sur l'image à segmenter (à droite) et le champ de déformation obtenu est appliqué à la segmentation de l'atlas afin d'obtenir une segmentation de l'image cible.

Ainsi, la segmentation par recalage d'atlas est très dépendante du recalage : les erreurs de segmentation viennent principalement d'erreurs liées au fait que le recalage a échoué en partie ou en totalité. Afin d'améliorer la capacité de cette méthode, une approche consiste à utiliser plusieurs atlas (IGLESIAS et al., 2015) pour représenter plus de diversité anatomique et de variabilité des intensités.

Les deux principales stratégies pour prendre en considération les différents atlas sont : soit de déterminer et de conserver uniquement l'atlas le plus représentatif de l'image à segmenter, soit de fusionner les segmentations obtenues avec les différents atlas. Nous avons retenu l'approche de segmentation multi-atlas de (WANG et al., 2013b) qui propose une fusion des segmentations suivie par un algorithme de correction basé sur un apprentissage

automatique des erreurs. Cette approche a été proposée pour segmenter des régions du cerveau et des muscles de jambes de chien dans des images IRM. Cette méthode est détaillée dans la section suivante.

12.3 Approche multi atlas *JLF + CL*

Cette approche comporte deux étapes consécutives : une étape de fusion jointe des étiquettes (JLF) (WANG et al., 2013a) et une étape -optionnelle- de correction par apprentissage automatique (CL) (WANG et al., 2012). Ces deux étapes sont décrites ci-après et illustrées dans la figure 12.2. Le JLF correspond à l'étape de recalage multi-atlas. La correction par apprentissage vise à supprimer une partie du biais de segmentation.

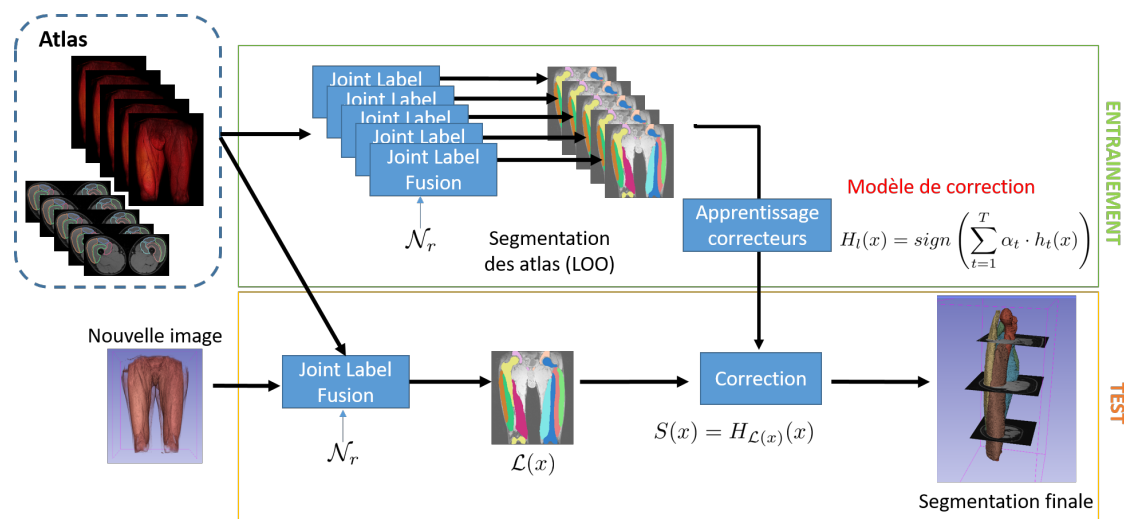


FIGURE 12.2 – Principe de la segmentation par recalage d'atlas *JLF+CL* de (WANG et al., 2013b). Cette approche comporte deux parties : la fusion jointe des étiquettes (JLF) puis une partie de correction par apprentissage (CL). Pour cette dernière partie, une étape d'entraînement est nécessaire.

12.3.1 Fusion jointe des étiquettes - JLF

Dans le cas du recalage multi-atlas, plusieurs segmentations sont proposées et un pixel peut se voir attribuer à différentes régions. Le problème est de fusionner ces résultats. Classiquement, cela est fait avec un vote à la majorité qui, malgré sa simplicité, a deux inconvénients. D'une part, ce vote ne garanti pas de cohérence spatiale : on peut ainsi obtenir des résultats extrêmement bruités. D'autre part, il crée un biais de segmentation dû au fait que les atlas et leur recalage peuvent produire des erreurs corrélées (WANG et al., 2012).

Pour remédier à ces problèmes, Wang *et al.* proposent un vote pondéré où le problème est l'optimisation des poids de pondération, de manière à minimiser l'erreur faite sur la segmentation de l'image cible. Cependant, cette erreur étant inconnue, Wang propose de l'estimer en utilisant la similarité des intensités dans le voisinage des pixels. Cette solution permet d'adresser efficacement le problème de résultats de segmentations bruités.

Cette méthode, qui fusionne les étiquettes en utilisant conjointement le domaine spatial et la similarité des niveaux de gris, est nommée *Joint Label Fusion* et abrégée JLF. Cette approche est détaillée ci-après.

L'algorithme JLF produit une carte de probabilités pour chaque étiquette et à chaque pixel ce qui permettra d'effectuer un vote par consensus. Pour obtenir une segmentation,

l'algorithme détermine pour chaque pixel l'étiquette qui a recueilli le plus grand nombre de voix. La principale différence entre les méthodes de fusion d'étiquettes et celle proposée par Wang est qu'elle prend en compte les corrélations entre les atlas lors du calcul des cartes de poids. Ainsi, dans le cas extrême où un atlas serait inclus deux fois (ou avec deux atlas extrêmement similaires), cette approche conduira au même résultat que s'il n'avait été inclus qu'une seule fois.

Dans la suite, on exprime ces concepts sous forme d'expressions mathématiques. Cela permet d'identifier les paramètres critiques de l'approche et, pour de futures contributions, d'avoir une compréhension précise des éléments intermédiaires, notamment les cartes de probabilités.

Soit T_F une image test à segmenter et $A^1 = (A_F^1, A_S^1), \dots, A^n = (A_F^n, A_S^n)$ les n atlas dont A_F^i est la i^e image atlas recalée sur T_F et A_S^i la segmentation atlas recalée correspondante. Pour chaque pixel \mathbf{x} , il est possible de modéliser l'erreur de segmentation pour l'étiquette l (avec $l \in 1, \dots, L$ et L est le nombre d'étiquettes) par :

$$\delta_l^i(\mathbf{x}) = \mathcal{I}[T_S(\mathbf{x}) = l] - \mathcal{I}[A_S^i(\mathbf{x}) = l] \quad (12.2)$$

où $\mathcal{I}[\cdot]$ représente la fonction indicatrice (qui vaut 1 si la condition est vraie, 0 sinon) et ainsi $\delta^i(x)$ ne peut prendre que trois valeurs $\delta^i(x) \in \{-1; 0; 1\}$ et $T_S(\mathbf{x})$ représente la segmentation inconnue à obtenir. La distribution de cette erreur pour les n atlas peut s'exprimer à partir des images :

$$q_l^i(\mathbf{x}) = p(|\delta_l^i(\mathbf{x})| = 1 \mid T_F, A_F^1, \dots, A_F^n) \quad (12.3)$$

Pour produire la segmentation consensus \bar{S} , la stratégie de vote pondéré est utilisée. Dans le cas binaire, cette pondération s'exprime ainsi :

$$\bar{S}(\mathbf{x}) = \sum_{i=1}^n w^i(\mathbf{x}) A_S^i(\mathbf{x}) \text{ avec } \sum_{i=1}^n w^i(\mathbf{x}) = 1. \quad (12.4)$$

Le but est de déterminer les poids w^i qui minimisent l'erreur moyenne entre la segmentation proposée \bar{S} et la segmentation de référence T_S :

$$\begin{aligned} E_{\delta^1(\mathbf{x}), \dots, \delta^n(\mathbf{x})} & \left[\left(T_S(\mathbf{x}) - \bar{S}(\mathbf{x}) \right)^2 \mid T_F, A_F^1, \dots, A_F^n \right] \\ &= E_{\delta^1(\mathbf{x}), \dots, \delta^n(\mathbf{x})} \left[\left(\sum_{i=1}^n w^i(\mathbf{x}) \delta^i(\mathbf{x}) \right)^2 \mid T_F, A_F^1, \dots, A_F^n \right] \\ &= \sum_{i=1}^n w^i(\mathbf{x}) \sum_{j=1}^n w^j(\mathbf{x}) E_{\delta^i(\mathbf{x}), \delta^j(\mathbf{x})} \left[\delta^i(\mathbf{x}) \delta^j(\mathbf{x}) \mid T_F, A_F^1, \dots, A_F^n \right] \\ &= \mathbf{w}_x^T \mathbf{M}_x \mathbf{w}_x \end{aligned} \quad (12.5)$$

où \mathbf{w}_x représente le vecteur $[w^1(\mathbf{x}); \dots; w^n(\mathbf{x})]$ et \mathbf{M}_x est la matrice des appariements d'atlas i et j .

A partir de \mathbf{M}_x , les poids optimum \mathbf{w}_x^* sont déterminés par la minimisation de :

$$\begin{aligned} \mathbf{w}_x^* &= \underset{\mathbf{w}_x}{\operatorname{argmin}} \mathbf{w}_x^T \mathbf{M}_x \mathbf{w}_x + \alpha \|\mathbf{w}_x\|_2 \quad (12.6) \\ &\text{avec } \sum_{i=1}^n w^i(\mathbf{x}) = 1 \end{aligned}$$

α est le terme de régularisation dont la valeur est généralement fixée à 0,1 (WANG et al., 2013b).

Wang propose de calculer les éléments de $\mathbf{M}_{\mathbf{x}}$ en utilisant la similarité locale dans les images des atlas i et j avec l'image à segmenter T_F . C'est à dire que l'expression :

$$\begin{aligned} \mathbf{M}_{\mathbf{x}}(i, j) &= E_{\delta^i(\mathbf{x}), \delta^j(\mathbf{x})} \left[\delta^i(\mathbf{x}) \delta^j(\mathbf{x}) \mid T_F, A_F^1, \dots, A_F^n \right] \\ &= p \left(\delta^i(\mathbf{x}) \delta^j(\mathbf{x}) = 1 \mid T_F, A_F^1, \dots, A_F^n \right) \end{aligned} \quad (12.7)$$

va pouvoir s'exprimer avec les images A_F^i et A_F^j . En assumant l'indépendance aux autres atlas, on peut écrire :

$$\mathbf{M}_{\mathbf{x}}(i, j) = p \left(\delta^i(\mathbf{x}) \delta^j(\mathbf{x}) = 1 \mid T_F, A_F^i, A_F^j \right) \quad (12.8)$$

En supposant maintenant que les pixels éloignés de \mathbf{x} n'ont pas d'influence sur cette probabilité, alors l'élément $\mathbf{M}_{\mathbf{x}}(i, j)$ peut être exprimé uniquement sur le voisinage $\mathcal{N}(\mathbf{x})$. On a alors :

$$\mathbf{M}_{\mathbf{x}}(i, j) = p \left(\delta^i(\mathbf{x}) \delta^j(\mathbf{x}) = 1 \mid \left\{ T_F, A_F^i, A_F^j \mid \mathbf{y} \in \mathcal{N}(\mathbf{x}) \right\} \right) \quad (12.9)$$

$$\propto \left[\sum_{\mathbf{y} \in \mathcal{N}(\mathbf{x})} |T_F(\mathbf{y}) - A_F^i(\mathbf{y})| \cdot |T_F(\mathbf{y}) - A_F^j(\mathbf{y})| \right]^\beta \quad (12.10)$$

avec β , un paramètre du modèle qui sera classiquement fixé à 2.

Dans le cas multi-classes, l'erreur moyenne se généralise ainsi :

$$E_{\delta^1(\mathbf{x}), \dots, \delta^n(\mathbf{x})} \left[\left(\mathcal{I}[T_S(\mathbf{x}) = l] - \sum_{i=1}^n w^i(\mathbf{x}) \mathcal{I}[A_S^i(\mathbf{x}) = l] \right)^2 \mid T_F, A_F^1, \dots, A_F^n \right] \quad (12.11)$$

$$= \mathbf{w}_{\mathbf{x}}^t \mathbf{M}_{\mathbf{x}} \mathbf{w}_{\mathbf{x}} \quad (12.12)$$

et le calcul de $\mathbf{M}_{\mathbf{x}}$ conduit à la la formulation :

$$\mathbf{M}_{\mathbf{x}}(i, j) \sim \left[\langle |A_F^i(\mathcal{N}(\mathbf{x})) - T_F(\mathcal{N}(\mathbf{x}))|, |A_F^j(\mathcal{N}(\mathbf{x})) - T_F(\mathcal{N}(\mathbf{x}))| \rangle \right]^\beta \quad (12.13)$$

avec $\langle \cdot, \cdot \rangle$ le produit scalaire et $|A_F^i(\mathcal{N}(\mathbf{x})) - T_F(\mathcal{N}(\mathbf{x}))|$ le vecteur des écarts absolus sur le patch \mathcal{N} centré en \mathbf{x} des intensités entre l'image à segmenter T_F et l' image d'atlas A_F^i recalée sur T_F .

Les erreurs de recalage peuvent perturber ce calcul : le contexte autour de \mathbf{x} dans l'image à segmenter pourrait ne pas correspondre parfaitement à ceux dans les atlas recalés. Afin d'améliorer l'estimation de $\mathbf{M}_{\mathbf{x}}$, plusieurs patchs $\mathcal{N}(\mathbf{x} + \epsilon)$ centrés à différentes positions ϵ autour de \mathbf{x} vont être testés. Celui qui minimise l'erreur entre $A_F^i(\mathcal{N}(\mathbf{x} + \epsilon))$ et $T_F(\mathcal{N}(\mathbf{x}))$ sera conservé pour le calcul de \mathbf{M} ainsi que $A_S^i(\mathcal{N}(\mathbf{x} + \epsilon))$ pour le vote pondéré. La dynamique d'exploration du paramètre ϵ va définir le voisinage de recherche qu'on notera \mathcal{N}_r .

Pour rendre l'algorithme plus robuste au cas d'images avec différentes dynamiques d'intensité, ce qui est commun en IRM, les intensités dans les patchs sont normalisées avant d'être comparées.

Les tailles des patchs \mathcal{N} et du voisinage de recherche \mathcal{N}_r dépendent de la taille des structures à segmenter.

La probabilité d'avoir l'étiquette l à la position \mathbf{x} pour l'image à segmenter T_F est :

$$p(l|\mathbf{x}, T_F) = \sum_{i=1}^n \mathbf{w}_x^i \mathcal{I}(A_S^i(\mathbf{x}) = l) \quad (12.14)$$

Pour réaliser la segmentation, il faut déterminer l'étiquette la plus probable pour chaque \mathbf{x} . C'est aux frontières des objets que la probabilité sera la plus critique mais aussi là où l'approche à le plus de chance de se tromper à cause de ses propres atlas. Comment être sûr que, malgré la variation anatomique et le recalage, tous les atlas se superposent bien à l'image à segmenter et sont bien cohérents ? Ou autrement : le contexte local des intensités de l'image à segmenter est-il vraiment en accord avec celui des atlas ? Dans ces approches de recalage multi-atlas, il existe nécessairement ce type de biais systématique qu'il est possible d'apprendre comme nous allons le voir dans la section suivante.

12.3.2 Apprentissage des corrections - CL

L'algorithme de *Corrective Learning* (CL) a pour but de détecter puis de corriger les erreurs de segmentation commises systématiquement par un algorithme de segmentation automatique. Contrairement aux erreurs aléatoires, provenant du bruit ou de variations anatomiques aléatoires, les erreurs systématiques sont prévisibles étant donné un ensemble de conditions (par exemple : de forme, de localisation, d'organe et d'intensité dans l'image). Elles ont plusieurs origines : des petites variations sur les définitions anatomiques, des discontinuités sur les segmentations manuelles, des biais entre les connaissances a priori incluses dans les méthodes automatiques et les données à segmenter.

Une approche de correction, basée sur de l'apprentissage automatique, est proposée dans (WANG et al., 2011). Appliquée à l'imagerie IRM, une diminution de 20% à 70 % du nombre de pixels mal segmentés est observée sur quatre algorithmes de segmentation automatique, dont le JLF. Dans la suite, cet algorithme de correction pour le JLF est expliqué en commençant par la phase d'apprentissage.

Phase d'apprentissage

Chaque atlas va être segmenté par le JLF comme s'il s'agissait d'une image à segmenter. Cependant, le JLF ne s'appuiera que sur les autres atlas disponibles pour effectuer cette segmentation. Puisque pour cet atlas testé on dispose de la bonne segmentation, il va être possible de déterminer les erreurs faites par le JLF. En appliquant ceci pour l'ensemble des atlas (approche Leave One Out), il est possible de construire une base de données permettant d'apprendre un modèle de correction des erreurs systématiques du JLF pour la base d'atlas et ceci pour chacune des étiquettes.

Pour ne considérer que les zones où il faut potentiellement corriger, une région de travail est créée pour chaque étiquette. Chaque région de travail est obtenue par dilatation, de taille r_d , de la région associée à l'étiquette.

Pour chaque point \mathbf{x} de cette région de travail, la position spatiale relative, des caractéristiques d'apparence (valeur des intensités) et de contexte de segmentation sont extraites dans un voisinage \mathcal{N}_f . La position spatiale relative est la coordonnée relative de \mathbf{x} au barycentre de la région de travail. L'apparence et les caractéristiques contextuelles sont aussi réduites par chacune des composantes spatiales afin d'augmenter la corrélation spatiale. Wang propose un voisinage de taille $5 \times 5 \times 5$ et obtient ainsi un vecteur $\mathcal{F}(\mathbf{x})$ de caractéristiques en \mathbf{x} de $3 + 125 + 125 + 3 \times 125 + 3 \times 125 = 1003$ dimensions.

Tous les pixels de cette région servent à l'apprentissage et le jeu de données d'entraînement sera construit par les paires $(\mathcal{F}(\mathbf{x}), \mathbf{x}_l)$ où \mathbf{x}_l correspond à l'étiquette du pixel \mathbf{x} . L'algorithme utilisé par Wang est l'algorithme de classification binaire AdaBoost (FREUND et al., 1996)

dont une description précise est donnée dans (ZHOU, 2012, p. 24) et expliqué dans l'annexe C de ce document. Cet algorithme va produire pour chaque étiquette et chaque pixel un modèle de correction.

Phase d'inférence

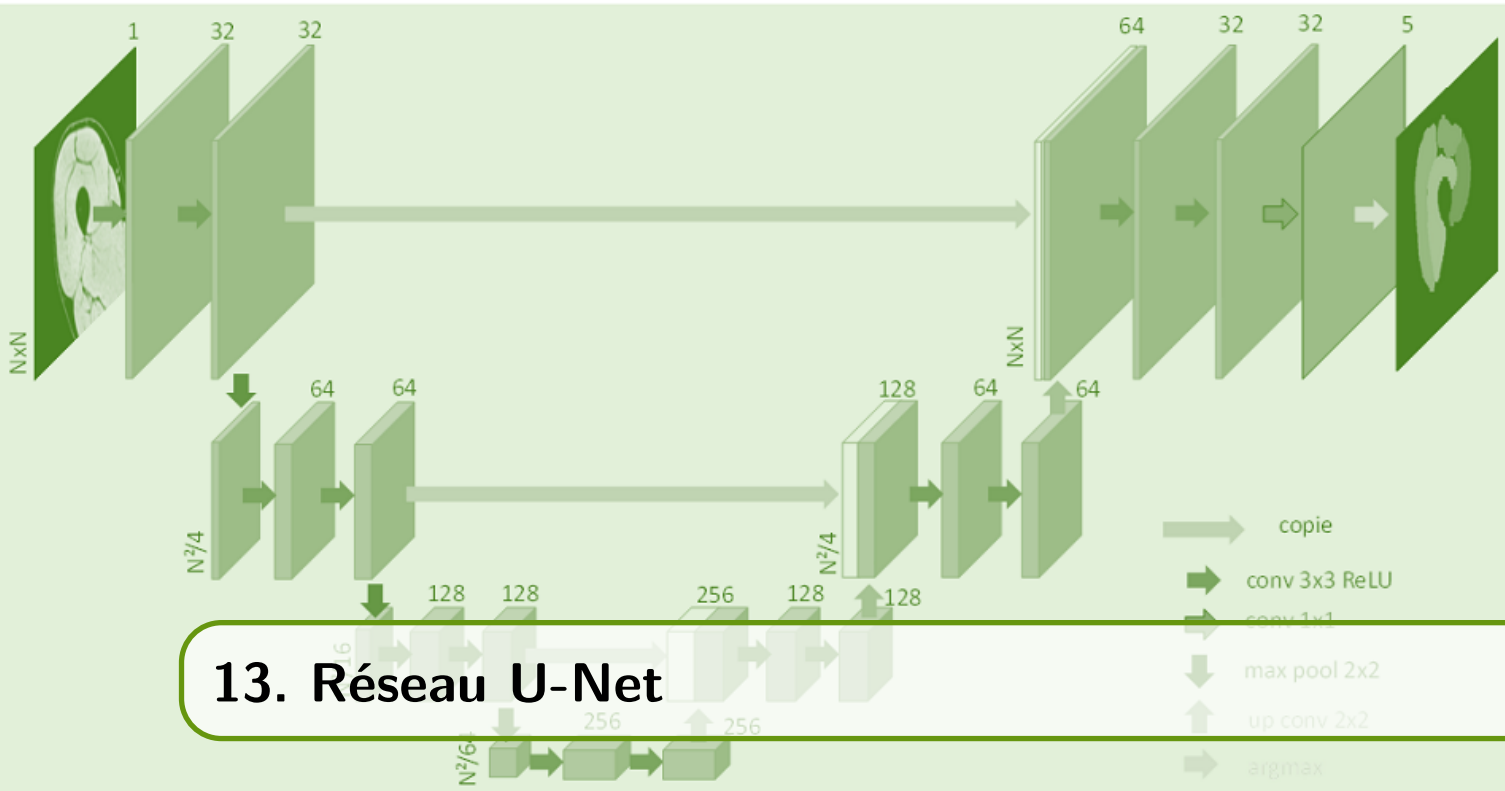
Lors de la correction d'une nouvelle image, chaque pixel préalablement étiqueté par le JLF, va être testé en utilisant la même définition de région de travail et les mêmes caractéristiques. Le pixel sera ré-affecté à l'étiquette dont le modèle de correction aura donné la plus grande confiance.

12.4 Conclusion

Dans ce chapitre nous avons étudié les approches de segmentation par recalage d'atlas et notamment l'approche multi-atlas de Wang (WANG et al., 2013b). Ces approches permettent d'obtenir des segmentations de bonnes qualités d'un ou plusieurs organes, sur différents instants du temps et ceci à partir d'un petit nombre d'atlas car elles tirent profit de toute la connaissance anatomique et de l'information image contenue dans les atlas. Ces approches sont à privilégier pour fournir des segmentations de qualité quand peu d'images avec une segmentation de référence sont disponibles.

Cependant, ces approches sont assez lourdes en terme de calcul. De plus, lorsque le nombre d'atlas utilisé devient grand, le temps de calcul augmente rapidement et on n'observe pas nécessairement d'amélioration des résultats.

Pour réduire les temps de calculs, l'utilisation d'une approche multi-échelle a été proposée (WANG et al., 2018a). Nous avons préféré proposer une alternative au JLF qui serait basée non plus sur des recalages mais sur des réseaux de neurones profonds. Le prochain chapitre présente l'approche de segmentation par réseaux de neurones.



13. Réseau U-Net

13.1 Introduction

Dans ce chapitre, on s'intéresse à la segmentation par apprentissage profond qui est devenue, en moins de 10 ans, une méthode incontournable. Nous commençons par une introduction à la segmentation d'images médicales par apprentissage profond. Puis nous détaillons l'architecture de référence en segmentation d'images médicales : le réseau U-Net. Nous expliquons ensuite le fonctionnement interne de ce réseau et son l'entraînement. Enfin, nous donnons un bref état de l'art de quelques une des variantes d'U-Net.

13.2 Segmentation par apprentissage automatique et profond

Depuis plusieurs décennies, l'apprentissage supervisé apporte des solutions alternatives à de nombreux problèmes d'analyse de données et d'images médicales. Depuis 2010, les ouvrages annuels de *Machine Learning in Medical Imaging*¹ en lien avec la conférence MICCAI² en donne une liste conséquente.

Ces dernières années, pour le traitement des images, nous avons assisté à l'explosion des approches d'apprentissage profond, fruit du rapprochement des communautés du *big data* et du *computer vision* avec la démocratisation de l'utilisation des ressources de calcul de type GPU. Ces approches ont rapidement été appliquées aux données médicales. Seulement quelques réseaux et traitements sont issus de problématiques médicales intégrant les spécificités des données, notamment les images 3D, la dynamique, la résolution, les artéfacts et la problématique de l'augmentation de données (KER et al., 2018 ; LITJENS et al., 2017) qui est critique dans ce contexte médical car il est souvent difficile de disposer de suffisamment de données pour l'entraînement.

Concernant la segmentation d'images médicales par *deep learning*, les différentes approches et les difficultés sont largement discutées dans les travaux de (RIZWAN I HAQUE et al.,

1. https://www.springer.com/gp/search?dnc=true&facet-type=type__book&query=Machine+Learning+in+Medical+Imaging&submit=Envoyer

2. Medical Image Computing and Computer Assisted Interventions

2020 ; TAJBAKSHI et al., 2020). Le réseau le plus emblématique pour la segmentation d'images médicales est le réseau U-Net introduit en 2015 par (RONNEBERGER et al., 2015). C'est ce réseau que nous avons utilisé pour différentes applications et que nous allons présenter dans les prochains paragraphes.

13.3 Architecture U-Net

A l'origine, le réseau U-Net est un auto-encodeur entièrement convolutionnel (RONNEBERGER et al., 2015 ; SHELHAMER et al., 2017), qui se distingue des auto-encodeurs classiques par l'ajout de connexions de sauts entre l'encodeur et le décodeur (figure 13.1).

Le réseau U-Net peut être appliqué dans plusieurs contextes : à la segmentation (RONNEBERGER et al., 2015) (application d'origine), au filtrage, à la restauration (LEHTINEN et al., 2018) et à la synthèse d'images (ISOLA et al., 2017). On ne se focalise ici que sur la segmentation. La littérature est très riche sur ce thème.

La suite de ce chapitre est une synthèse de travaux fondamentaux sur cette architecture :

- Le paragraphe 13.4 décrit les principaux éléments du réseau U-Net et d'apprentissage.
- Le paragraphe 13.5 présente quelques variantes du réseau U-Net pour la segmentation.

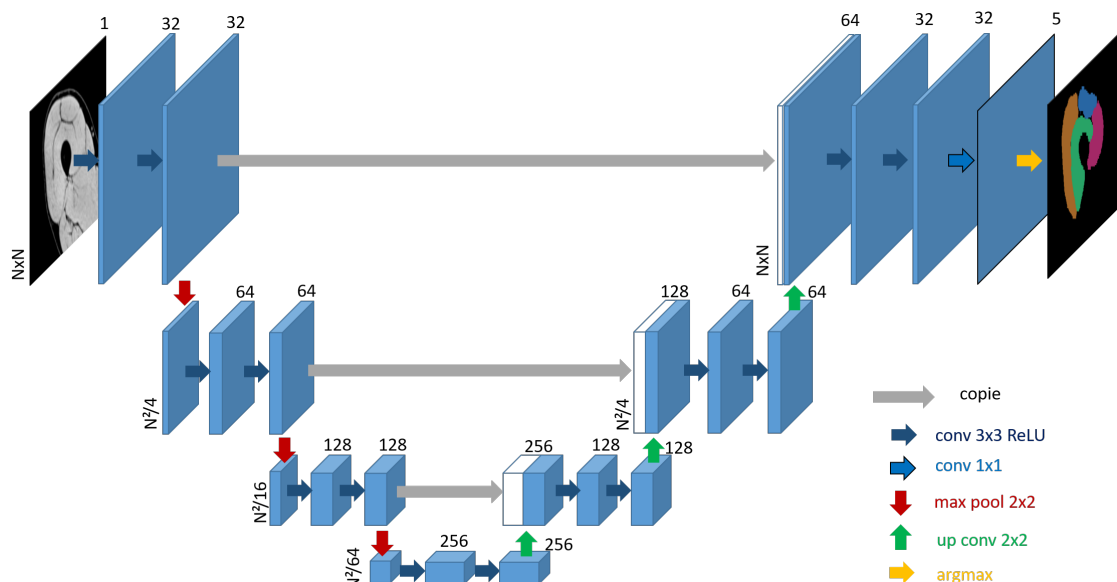


FIGURE 13.1 – Architecture du réseau U-Net avec 4 échelles et 32 filtres sur la première couche. L'encodeur est la partie de gauche où la taille de l'image est diminuée, le décodeur est la partie de droite où la taille de l'image est augmentée. Illustration du réseau utilisé pour la segmentation de 4 muscles et du fond à partir d'une image IRM.

13.4 Les coulisses d'U-Net

Un réseau de neurones est un modèle paramétrique. Les paramètres du modèle sont des valeurs (ou poids) qui doivent être appris afin d'accomplir une tâche donnée. Cet apprentissage est réalisé pendant la phase d'entraînement. Une fois appris, le réseau est capable de reproduire la tâche sur de nouvelles données. Il s'agit des phases de test ou d'inférence.

La phase d'apprentissage est itérative et implique de nombreux *hyper-paramètres*³.

3. les "hyper-paramètres" correspondent aux paramètres à régler pour l'optimisation des poids du réseau.

Dans notre contexte, le réseau est U-Net, la tâche est la segmentation d'images et l'apprentissage est supervisé.

L'apprentissage supervisé signifie, ici, que l'entraînement va reposer sur un jeu d'images pour lesquelles on dispose des "vraies" segmentations. L'apprentissage peut ainsi être supervisé par une mesure d'erreur entre les vraies segmentations et celles proposées par le réseau. Le principe de la mise à jour des poids lors de l'apprentissage d'un réseau de neurones est illustré sur la figure 13.2. Les différents éléments sont détaillés ensuite.

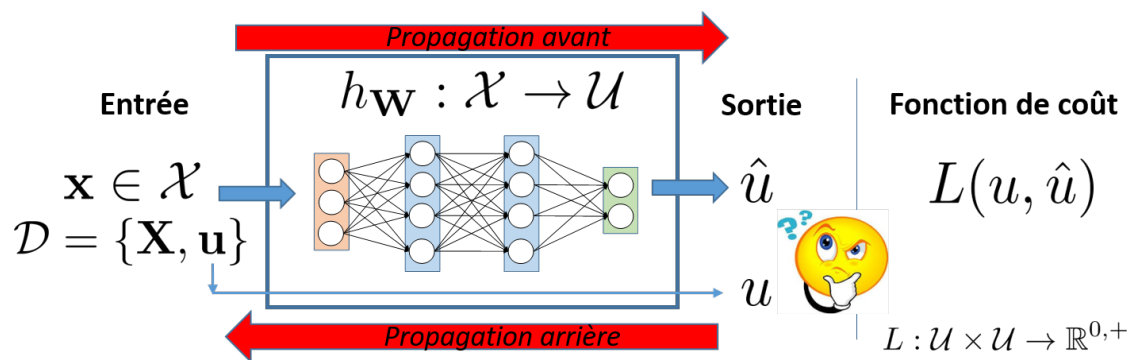


FIGURE 13.2 – Schéma de principe de l'apprentissage supervisé.

13.4.1 Jeu d'images

On note une image $\mathbf{x} \in \mathcal{X}$ (\mathcal{X} représente l'espace des images) et sa segmentation $u \in \mathcal{U}$ (où \mathcal{U} est l'espace des segmentations). Le jeu de données \mathcal{D} constitué de n couples (image, segmentation) se représente ainsi :

$$\begin{aligned} \mathcal{D} &= \{(\mathbf{x}_1, u_1), (\mathbf{x}_2, u_2), \dots, (\mathbf{x}_n, u_n)\} \\ &= \{\mathbf{X}, \mathbf{u}\} \end{aligned} \quad (13.1)$$

Afin de pouvoir apprendre et évaluer la capacité du réseau à se généraliser sur de nouvelles images, le jeu de données \mathcal{D} est séparé en un jeu d'apprentissage \mathcal{D}_A et un jeu de test \mathcal{D}_T , généralement dans les proportions 80%-20% ou 90%-10% bien que cela soit fonction de l'application et du jeu de données disponible. Le jeu d'apprentissage est souvent divisé en deux dans les mêmes proportions :

- le jeu d'entraînement \mathcal{D}_E va permettre l'évolution des poids du réseau (80% du jeu d'apprentissage),
- le jeu de validation \mathcal{D}_V va permettre d'ajuster les hyper-paramètres du réseau.

Le jeu de test est utilisé uniquement après l'entraînement pour vérifier le comportement du réseau sur des images qu'il n'a jamais 'vues'. Le réseau va estimer la segmentation \hat{u} de chaque image \mathbf{x} de \mathcal{D}_T . Puisqu'on dispose pour les images \mathbf{x} de \mathcal{D}_T de leurs segmentations u , il est possible de quantifier la performance du réseau en utilisant les métriques habituelles A.

Lors de l'entraînement, ces métriques sont utilisées pour l'optimisation itérative des poids du réseau. Par convention et uniformisation, ces métriques sont modifiées afin d'être minimisées lors de l'optimisation. Elles sont alors dénommées "fonctions de coût" : elles expriment le coût des erreurs commises par le réseau. Les poids du réseau, organisés en couche, doivent être modifiés de manière à diminuer le coût de l'erreur.

13.4.2 Fonction de coût

Une fonction de coût $L(u, \hat{u})$ permet de mesurer l'erreur entre la prédiction du réseau \hat{u} et la segmentation u à obtenir. Ces fonctions sont définies ainsi $L : \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}^{0,+}$ et on cherche à minimiser $\min_W L(u, \hat{u})$. Pour une optimisation basée sur une descente de gradient, les fonctions de coût doivent être dérivables.

Il existe de nombreuses fonctions de coût. Pour le réseau U-Net, les principales sont :

- **Le score de Dice** transformé en fonction de coût (1-DICE) et la fonction de coût de Dice généralisée (cas multi-classes) (SUDRE et al., 2017). La classe représentant le fond est ignorée quand le fond est sur-représenté : elle nuirait à l'apprentissage des autres classes,
- **L'entropie croisée** (*cross-entropy*) est définie ainsi dans le cadre multi-classes (avec C classes) pour un échantillon o :

$$L_{CE}(u_o, \hat{u}_o) = - \sum_{c=1}^C u_{o,c} \log(\hat{u}_{o,c}) \quad (13.2)$$

avec $u_{o,c}$ valant 1 si la classe de l'échantillon o est c , sinon 0, et $\hat{u}_{o,c}$ est la probabilité prédite par le réseau pour que o soit de classe c . Cette fonction est sensible à l'équilibre des représentations des classes principalement quand on effectue une moyenne sur un grand nombre d'échantillons et que les classes majoritaires sont bien traitées : les éléments rares seront d'influence négligeables. Pour réduire ce problème, le *log* est souvent pondéré par l'inverse de la probabilité de la classe c .

- **La fonction de coût focale** (*focal loss*) a été introduite pour répondre au problème de non équi-représentation des classes. Elle est définie ainsi :

$$L_{FL}(u_o, \hat{u}_o) = - \sum_{c=1}^C (1 - u_{o,c})^\gamma \log(\hat{u}_{o,c}) \quad (13.3)$$

13.4.3 Les différentes couches du réseau

On décrit ici, les principales couches d'un réseau de neurones artificiels comme U-Net.

U-Net est constitué d'un encodeur et d'un décodeur avec, entre eux, des connexions de saut (figure 13.1).

A chaque étage de l'encodeur, deux couches de convolution de même nombre de filtres c sont appliquées successivement. Le résultat est une image de sensiblement la même taille que l'image d'entrée mais disposant de c canaux. Cette sortie est d'une part transmise au décodeur puis sous-échantillonnée par un coefficient 4 en 2D. Ce sous échantillonnage est à l'origine un *max pooling*. Puis on rapplique deux convolutions successives dont le nombre de filtres est deux fois plus importants que pour l'étage précédent (soit $2c$). Ce processus est répété jusqu'à obtenir un support spatial cohérent par rapport aux objets à segmenter et à leur organisation spatiale.

Commence ensuite le décodeur. Une opération de sur-échantillonnage est d'abord réalisée pour accroître le support spatial d'un coefficient 4. Le nombre de filtres utilisé lors de cette opération permet de diviser par 2 le nombres de canaux de l'étage inférieur. Puis deux couches de convolution successives sont appliquées (comme pour l'encodeur). Cette architecture se répète jusqu'à retrouver la taille de l'image d'origine.

Enfin, une dernière couche de convolution comportant autant de filtres que de classes à segmenter (ne pas oublier la classe pour "le fond") est appliquée. On obtient ainsi une image de taille proche de l'image d'entrée mais comportant n canaux, n étant le nombre de classes incluant le fond.

Il existe de nombreuses combinaisons de ces couches permettant d'obtenir des comportements particuliers (bloc d'attention, Inception, ...) et ainsi les variantes d'U-Net. Voici les détails des plus communes.

- **Couche complètement connectée** (*dense, fully connected*). Ce type de couche complètement connectée, n'est pas utilisé dans U-Net, mais est la couche fondamentale des réseaux de neurones qui connecte les entrées \mathbf{x} aux k neurones de la couche considérée, produisant ainsi un vecteur de sortie \mathbf{z} de taille k . On note \mathbf{w}_k , les $|\mathbf{x}|$ poids d'un neurone k et sa sortie se calcule par $z_k = \mathbf{w}_k \mathbf{x} + b$ où b correspond au biais. Pour une couche complètement connectée, on a $\mathbf{z} = \mathbf{W}\mathbf{x} + \mathbf{b}$ et \mathbf{W} est une matrice $|\mathbf{x}| \times k$.
- **Couche de convolution**. Afin de réduire le nombre de poids nécessaires aux couches complètement connectées, les poids des neurones de convolution ne sont connectés qu'à une partie de la données d'entrée, on parle de champs récepteurs, mais appliqués à toute la donnée d'entrée en déplaçant le champs récepteur et en conservant les poids. Cette donnée d'entrée nécessite d'avoir un support spatial ou temporel. La sortie produite conservera ce support. Ce mécanisme est très proche de l'opération de convolution, d'où son nom. En 2D, les poids \mathbf{W} d'une couche correspondent aux k filtres de taille $f_h \times f_w \times f_c$ appliqués à l'entrée \mathbf{x} ayant f_c canaux. Ainsi, on exprime la sortie $z_{i,j,k}$ d'un filtre d'une couche de convolution 2D par :

$$z_{i,j,k} = b_k + \sum_{u=0}^{f_h-1} \sum_{v=0}^{f_w-1} \sum_{c=0}^{f_c-1} x_{i',j',c} \cdot \mathbf{W}_{u,v,c,k} \quad (13.4)$$

avec $i' = i \cdot s_h + u$ et $j' = j \cdot s_w + v$

s_h et s_w représentent les pas dans les 2 dimensions spatiales. On remarque que cette expression n'est pas celle d'un produit de convolution mais d'un produit de corrélation croisée. La terminologie n'est donc pas complètement rigoureuse. Cependant, vu que les poids \mathbf{W} sont appris, cette distinction est de faible importance et on pourra toujours observer les *vrais* filtres en inversant les indexes. Concernant la contrainte de validité du support de convolution, on utilise de plus en plus des marges de valeurs (*padding*) permettant d'obtenir en sortie un support spatial identique à celui d'entrée. Les stratégies de padding sont les stratégies classiques : complétion par une valeur (souvent 0), par miroir ou circulaire. La figure 13.3 illustre deux couches de convolution préservant la taille de l'entrée par padding.

Ainsi, la sortie d'une couche de k filtres correspond à k images, ou cartes de caractéristiques, regroupées en une seule image de k canaux et ceci quel que soit le nombre de canaux de l'image d'entrée.

- **Sous-échantillonnage** (*pooling*). Afin d'accroître le champ réceptif des couches de convolution, sans augmenter la taille des filtres et ainsi le coût de calcul, un sous échantillonnage a été proposé après un enchaînement de couches de convolution. Ce sous-échantillonnage est classiquement réalisé soit par moyenne (on ne conserve que la valeur moyenne pour représenter un ensemble de pixels) - il s'agit de l'*average pooling*, soit par conservation de la valeur maximale des intensités d'un ensemble de pixels, *max pooling*. Nous sommes loin des approches rigoureuses du traitement du signal visant à ne pas introduire de repliement de spectre lors du sous-échantillonnage. Les objectifs ne sont pas les mêmes. Pour les réseaux de neurones, le but est de réduire le support en un minimum de temps et en ne conservant que l'information la plus significative. Classiquement, le *pooling* est réalisé sur des blocs de 2×2 pixels et ceci tous les 2 pixels afin de diviser par 4 la taille de l'image. Cette opération ne s'applique qu'aux dimensions spatiales de l'image, pas à ses f_c canaux ou cartes de

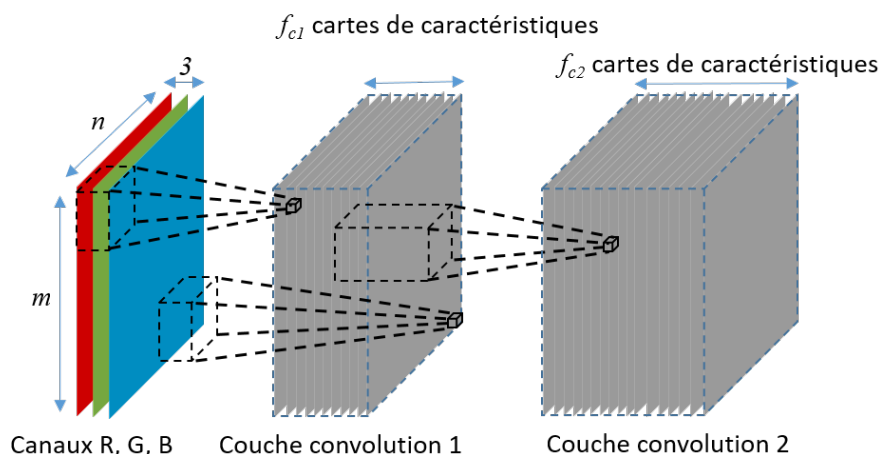


FIGURE 13.3 – Schéma de principe des couches de convolution. L'entrée est une image RGB de 3 canaux. Il y a f_{c1} filtres de taille $(f_h, f_w, 3)$ pour la première couche de convolution et f_{c2} filtres de taille (f_h, f_w, f_{c1}) pour la seconde.

caractéristiques. Leur nombre reste donc inchangé comme le montre la figure 13.4. Il n'y a aucun paramètre à apprendre pour cette couche.

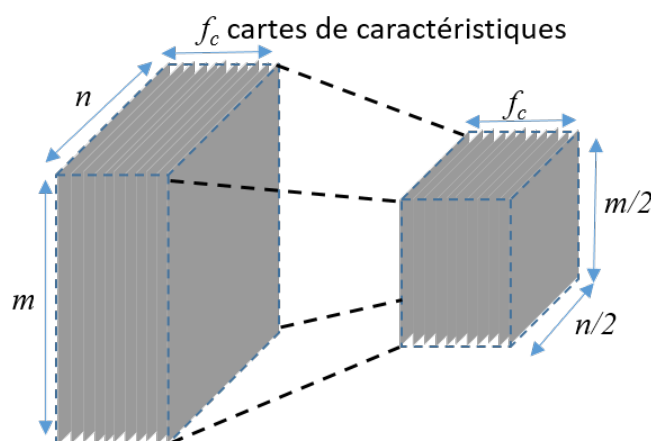


FIGURE 13.4 – Schéma de principe d'une couche de sous-échantillonnage (pooling). L'entrée de taille (n, m, f_c) est réduite à une taille de $(\lfloor n/2 \rfloor, \lfloor m/2 \rfloor, f_c)$. On note que le nombre de canaux f_c (ou de cartes de caractéristiques) est inchangé.

- **Sur-échantillonnage** (*upsampling*). Il s'agit de l'opération inverse à la précédente afin de remonter à la taille d'origine de l'image d'entrée. Il existe trois méthodes : la répétition d'une valeur sur un bloc, une interpolation classique (voisin le plus proche, linéaire, ...), une *convolution inverse* ou *déconvolution*. Ces deux derniers termes sont impropres aux notions relatives à l'opération de déconvolution et sont maintenant proscrits. Les termes de *up-convolution* ou de convolution transposée leurs sont préférés. Pour ces approches, il s'agit de réaliser plusieurs convolutions dont le support spatial de sortie sera plus grand que le support d'entrée, la totalité des canaux en entrée étant utilisée. Les deux techniques reviennent presque au même : soit on intercale des 0 entre les valeurs de l'entrée puis on convole, soit on applique une convolution (3×3) par exemple, voire une simple multiplication, à toutes les valeurs en entrée, en conservant les résultats sur le support du filtre et en se déplaçant

de 2 en 2 pixels sur la sortie. Les valeurs dans les zones de recouvrement sont sommées ou moyennées. Le principe de la convolution à trous, pour une seule convolution et pour trois canaux en entrée, est illustré sur la figure 13.5.

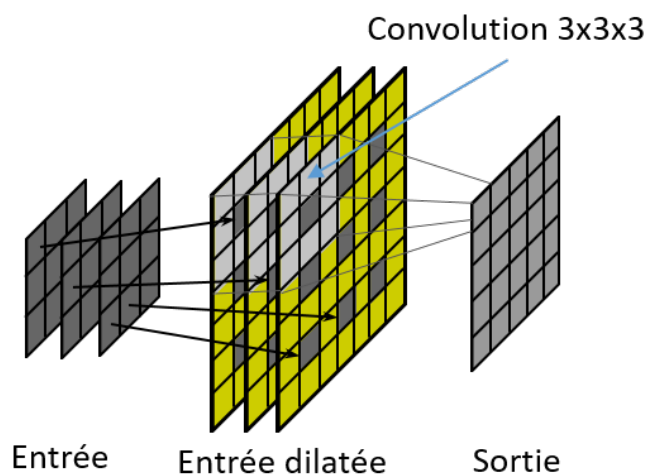


FIGURE 13.5 – Schéma de principe de la convolution par image dilatée pour le sur-échantillonnage (convolution à trous). L'entrée de taille $(3 \times 3 \times f_c)$, ici représentée avec trois canaux $f_c = 3$, est d'abord dilatée spatialement (ici d'un facteur quatre) et son support étendu. Puis une convolution de noyau $3 \times 3 \times f_c$ est appliquée sur l'image dilatée permettant d'obtenir une sortie de support spatial plus grand $(5 \times 5 \times f_n)$ où f_n est le nombre de convolution (ici une seule convolution soit : $f_n = 1$).

- **Concaténation et résidu.** Les connexions de sauts de U-Net nécessitent une fusion de deux informations de même support spatial $n \times m$. La concaténation consiste à créer un tenseur où les canaux f_c sont concaténés. Ainsi, si les tenseurs de deux entrées a et b sont de taille $n \times m \times f_{c,a}$ et $n \times m \times f_{c,b}$, alors le tenseur résultat sera de taille $n \times m \times (f_{c,a} + f_{c,b})$. Une opération d'addition terme à terme entre les deux tenseurs d'entrée est aussi possible quand $c_a = c_b$. Dans les deux cas, les informations sont fusionnées et les gradients directement propagés sur chacune des couches à l'entrée, ce qui permet de limiter la disparition du gradient sur les couches profondes et ainsi de favoriser leurs évolutions.
- **Normalisation par lots (batch normalization).** Au cours de l'entraînement, les poids des couches sont itérativement mis à jour conduisant nécessairement à une évolution des distributions des sorties qui vont donc impacter la couche suivante. IOFFE et al., 2015 a proposé la normalisation des sorties par la moyenne μ_B et l'écart type σ_B du lot traité afin de supprimer cet effet. Ainsi, la sortie \mathbf{z} normalisée par lot est obtenue à partir de l'entrée \mathbf{x} comme suit (\mathbf{x} appartient au lot considéré) :

$$\hat{\mathbf{x}} = \frac{\mathbf{x} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (13.5)$$

$$\mathbf{z} = \gamma \hat{\mathbf{x}} + \beta$$

avec γ et β deux paramètres à apprendre. La normalisation par lot permet d'utiliser des taux d'apprentissage plus importants et ainsi de converger plus vite, notamment avec des fonctions d'activation saturantes.

- **Dropout** (HINTON et al., 2012; SRIVASTAVA et al., 2014). Cette couche masque aléatoirement un pourcentage des sorties des neurones de la couche précédente. Concrètement, la sortie d'un neurone temporairement masquée sera de 0 et ainsi il

ne contribuera ni à la passe avant, ni à la passe arrière : ses poids n'auront aucune influence et n'évolueront pas. On dit aussi que le *dropout* désactive ou éteint des neurones. Cependant cette désactivation n'est pas permanente et ne dure généralement qu'une itération (passe avant et passe arrière pour un mini lot). L'objectif est de forcer toutes les parties du réseaux de neurones à être efficaces et autonomes, en suppléant au besoin une partie défaillante. Cette technique permet de gagner quelques pourcents sur les performances du réseau en utilisant des ratios souvent proches de 0,5. Aussi, le *dropout* agit comme une régularisation en prévenant du sur-apprentissage. Il est aussi vu comme une technique de méthode d'ensembles où un réseau se comporte comme une myriade de petits réseaux qui agglomèrent leurs résultats. Pour les couches de convolution, le dropout classique est mal approprié car s'appliquant sur des éléments (pixels) il ne prévient pas du sur-apprentissage et allonge simplement la durée d'apprentissage. Dans le cas de couches de convolution, il est pertinent d'appliquer le dropout sur la totalité des sorties des filtres. Ainsi, le *spatial dropout* désactive aléatoirement certains filtres pour obtenir plus d'indépendance dans les cartes de caractéristiques et prévenir du sur-apprentissage (TOMPSON et al., 2015).

13.4.4 Fonction d'activation

Une fonction d'activation non linéaire est appliquée aux sorties d'une couche de convolution ou d'une couche complètement connectée. Il existe de nombreuses fonctions d'activation ayant toutes un avantage par rapport aux autres : plage de sortie, complexité de calcul, dérivabilité, efficacité d'apprentissage, auto-normalisation, ... Voici une liste non exhaustive de fonctions représentant cette diversité et représentée sur la figure 13.6.

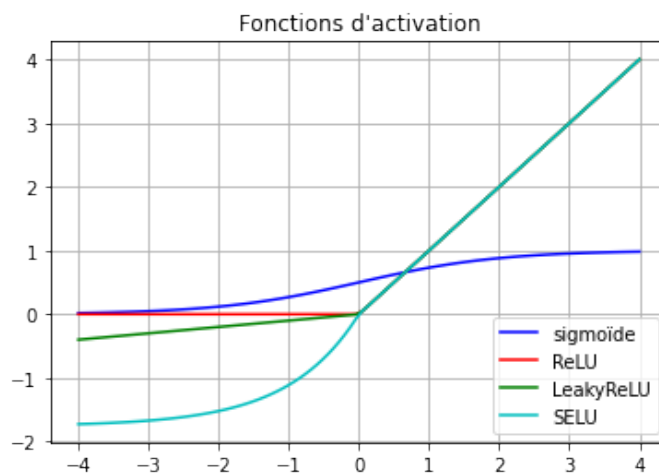


FIGURE 13.6 – Principales fonctions d'activation pour les réseaux de neurones.

- **Sigmoïde.** Fonction historique qui permet d'obtenir des valeurs entre 0 et 1 et dont la dérivée est facilement exprimable. Elle est cependant biaisée (0,5) et saturante. Elle conduit fréquemment aux problèmes de disparition ou explosion des gradients. Pour U-Net, elle aurait un intérêt que pour la couche de sortie dans le cadre d'une segmentation binaire.

$$f_{\text{sigmoïde}}(z_k) = \frac{1}{1 + \exp(-z_k)} \quad (13.6)$$

- **ReLU** ou rectifieur ou *Rectified Linear Unit*. Fonction non saturante, extrêmement utilisée, rapide, facile à dériver (sauf en 0). Excellente alternative à la sigmoïde et à

la tangente hyperbolique.

$$f_{ReLU}(z_k) = \max(0, z_k) = \begin{cases} 0 & \text{si } z_k < 0 \\ z_k & \text{si } z_k \geq 0 \end{cases} \quad (13.7)$$

A cause de cette constante 0 pour les valeurs négatives, elle peut causer la *mort de neurones* pendant l'apprentissage, notamment si les poids des neurones sont mal initialisés : les sorties des neurones resteront à 0 et n'influenceront pas les évolutions des neurones des couches précédentes.

- **Leaky ReLU**. Cette fonction a été proposée notamment pour palier à "la mort des neurones" causée par ReLU. Une *fuite* de pente α est proposée sur la partie négative. Classiquement, on choisit $\alpha = [0.01, 0.2]$ mais il est possible de tirer au hasard cet hyper-paramètre, voire de l'optimiser pendant l'apprentissage (il devient un paramètre du réseau) (XU et al., 2015a).

$$f_{LeakyReLU}(z_k; \alpha) = \max(\alpha z_k, z_k) = \begin{cases} \alpha z_k & \text{si } z_k < 0 \\ z_k & \text{si } z_k \geq 0 \end{cases} \quad (13.8)$$

- **ELU** (*exponential Linear Unit*) et **SELU** (*scaled ELU*). ELU, proposée par (CLEVERT et al., 2016) pour résoudre l'extinction de neurones, permet aussi d'obtenir une convergence à l'entraînement plus rapide et une meilleure généralisation sur le jeu de test. Il n'y a plus d'explosion ou de disparition des gradients. Elle est cependant plus complexe à calculer que ReLU. Le paramètre α est généralement fixé à 1.

$$f_{ELU}(z_k; \alpha) = \begin{cases} \alpha(\exp(z_k) - 1) & \text{si } z_k < 0 \\ z_k & \text{si } z_k \geq 0 \end{cases} \quad (13.9)$$

Les travaux de (KLAMBAUER et al., 2017) ont montré que par un paramètre α bien choisi, il était possible de forcer le réseau à s'auto-normaliser et ainsi conserver une même variance en sortie de chaque couche lors de l'entraînement. Il s'agit de la fonction SELU.

$$f_{SELU}(z_k) = 1,0507 \cdot f_{ELU}(z_k; 1,67326) \quad (13.10)$$

Les auteurs de SELU ont aussi proposé une adaptation du *dropout* (nommée *alpha dropout*) qui permet de conserver les propriétés de SELU lors de la désactivation de neurones.

- **Softmax**. Contrairement à toutes les fonctions précédentes, cette fonction d'activation est partagée par tous les K neurones d'une couche (voir la figure 13.7). Elle permet de représenter la loi de probabilité sur les K éléments discrets possibles.

$$f_{softmax}(y_k) = \frac{\exp(z_k)}{\sum_{i=1}^K \exp(z_i)} \quad (13.11)$$

Elle est la fonction d'activation de nombreux réseaux où une décision doit être prise. Typiquement pour U-Net, elle permet de calculer la probabilité d'appartenance d'un pixel à toutes les classes et facilite ainsi l'utilisation de nombreuses fonctions de coût. Pour déterminer la classe à laquelle associer le pixel, elle sera suivie de la fonction *argmax*.

13.5 Quelques variantes d'U-Net

Il existe de nombreuses approches de segmentation par apprentissage profond (MINAEE et al., 2022). On se limite ici à l'architecture "U-Net".

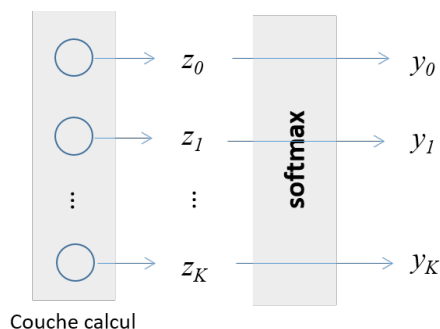


FIGURE 13.7 – Schéma de principe du calcul de la fonction softmax. Toutes les sorties de la couche de neurones partagent cette fonction.

Introduit en 2015 (RONNEBERGER et al., 2015), il existe maintenant de nombreuses variantes et il est laborieux d'en faire une liste exhaustive.

Pour les variantes liées à l'imagerie médicale, on pourra se référer à (SIDDIQUE et al., 2021) qui liste les travaux indexés chez les trois principaux groupes éditoriaux (Elsevier, IEEE et Springer) sur la période de 2017 à décembre 2020. Ces variantes d'U-Net consistent souvent à modifier l'architecture et/ou à proposer une fonction de coût adaptée à l'apprentissage pour cette nouvelle architecture.

Au niveau de l'architecture, plusieurs types d'améliorations ont été proposés. Un premier type consiste à améliorer l'encodeur notamment en s'inspirant des architectures des réseaux de classification. Par exemple, ResU-Net (ZHANG et al., 2018) s'appuie sur les blocs résiduels (introduits dans le réseau de classification ResNet (HE et al., 2016)), pour sa partie encodeur et qui utilise la moyenne des erreurs quadratiques comme fonction de coût. L'intérêt des blocs résiduels est qu'ils facilitent l'entraînement en permettant aux couches résiduelles de modéliser les résidus et non la fonction complète. La propagation du gradient est aussi moins atténuée par ces blocs, ce qui permet de créer des réseaux extrêmement profonds (plus de 100 couches) et ainsi d'accroître la capacité du réseau à acquérir des concepts de haut niveau et une grande capacité de généralisation. ResU-Net sera étendu à ResU-Net++ (JHA et al., 2019) qui introduit, entre autre, l'utilisation d'unité d'attention dans le réseau ResU-Net. Ces unités déterminent quelles sous parties (spatiales et/ou caractéristiques (HU et al., 2020)) du réseau et des données doivent avoir le plus "d'attention", c'est à dire plus d'influence sur la décision finale. Elles permettent aussi de réduire le temps de calcul. Un second type de modifications de l'architecture U-Net consiste à accroître le nombre de connexions de sauts pour augmenter la prise en compte des caractéristiques à différentes échelles. U-Net++ (ZHOU et al., 2020) introduit la fonction de coût focale et redéfinit les connexions de saut avec une connectivité dense pour permettre une meilleure optimisation et améliorer la généralisation. U-Net3+ (HUANG et al., 2020) propose d'exploiter la diversité de l'information provenant des différentes échelles et propose aussi une fonction de coût hybride qui capture la structure sous différentes formes (SSIM, focale) et échelles. U-Net++ et U-Net3+ s'appuient tous deux sur une supervision profonde (LEE et al., 2015) pour apprendre la représentation à partir de caractéristiques agrégées à toutes les échelles. De même, les réseaux comme BCDU-Net (AZAD et al., 2019) intègrent des cellules LSTM⁴ sur les connexions de saut et la concaténation des caractéristiques permettent ainsi de fusionner les informations provenant de deux échelles différentes de manière non-linéaire.

4. *Long Short Term Memory* : cellule permettant aux réseaux de neurones de considérer une cohérence temporelle

Un troisième type de modifications est la mise en cascade d'une succession de réseaux auto-encodeurs, tels que des U-Net. Ceci a été proposé initialement dans (NEWELL et al., 2016) puis spécifié à l'imagerie cardiaque dans (VIGNEAULT et al., 2018). Ce type d'architecture permet d'affiner progressivement le résultat de segmentation, que l'on peut voir comme une sorte d'attention progressive. On note que pour accroître la qualité de la segmentation finale, (VIGNEAULT et al., 2018) ajoutent avant le réseau de segmentation, un réseau de transformation spatiale permettant d'orienter et de mettre à l'échelle l'objet d'intérêt (ici le coeur) et ainsi de limiter la diversité de représentation des objets présentés au réseau de segmentation. Nous avons aussi proposé une contribution dans ce sens (LECLERC et al., 2020).

U-Net a été proposé initialement pour des images bidimensionnelles. Pour segmenter un volume 3D, il est possible de segmenter chaque coupe 2D puis de ré-empiler les résultats afin de produire la segmentation 3D. Cependant, ce n'est potentiellement pas optimal car le réseau doit être capable de se généraliser aux différents contextes anatomiques et ne peut apprendre du contexte 3D. Plusieurs architectures 3D ont été proposées telles que 3D U-Net (ÇIÇEK et al., 2016) et V-Net (MILLETARI et al., 2016) mais nécessitent une grande quantité de RAM GPU et des GPU très performants pour bénéficier d'un entraînement en temps raisonnable avec des convolutions spatiales 3D. La quantité de volumes 3D nécessaires à l'entraînement est aussi importante, conduisant à un volume d'information lourd à transférer. Aussi, il est souvent difficile de réaliser une augmentation de données "à la volée" basée sur des transformations rigides ou affines pour les volumes 3D : le temps nécessaire à l'interpolation et la construction du nouveau volume étant particulièrement important. Ainsi, ces architectures 3D sont souvent entraînées sur des petits sous volumes, réduisant les performances de ces approches à celles obtenues en 2D (voir moins (WANG et al., 2020)). Néanmoins, utiliser l'information tridimensionnelle reste un enjeu pour améliorer U-Net. Des travaux visant à minimiser le besoin de ressources ont été publiés notamment dans :

- (LI et al., 2019) qui propose une nouvelle stratégie de division en patch et utilisant des convolutions séparables (exprimables en 1D).
- (LI et al., 2018) où les représentations 2D intra-coupes et les caractéristiques inter-coupes 3D sont optimisées conjointement grâce à une couche de fusion hybride.
- (ALKADI et al., 2019) qui propose un « 2.5D U-Net » où les coupes supérieure et inférieure de la coupe centrale à segmenter sont ajoutées comme canaux d'entrée supplémentaires.
- (HAQUE et al., 2019) qui entraîne 3 réseaux pour chacun des axes, puis fusionne les résultats avec un vote à la majorité, et (PERSLEV et al., 2019) qui généralise le concept aux coupes multi directionnelles et effectue un vote pondéré où la pondération est apprise automatiquement.

Enfin, le récent *framework* nnUNet (ISENSEE et al., 2021) permet de faire conjointement des tests d'architectures UNet (2D ou 3D), une recherche des hyper-paramètres et des pré- et post- traitements les plus optimaux pour un problème donné en fusionnant les résultats obtenus par différentes architectures. nnUnet intègre notamment une architecture en cascade de 2 réseaux U-Net 3D dont le premier travaille sur des volumes sous échantillonnés (segmentation grossière) afin d'alimenter le second réseau qui utilisera cette segmentation grossière et des patch 3D à résolution initiale. Ce framework s'impose de plus en plus comme une référence permettant ainsi d'évaluer l'apport d'une nouvelle contribution. Et... il est souhaitable qu'une référence émerge !

En effet, deux causes rendent l'état de l'art en segmentation par *deep learning* difficiles à appréhender et à exploiter efficacement. Tout d'abord, l'effervescence autour du *deep*

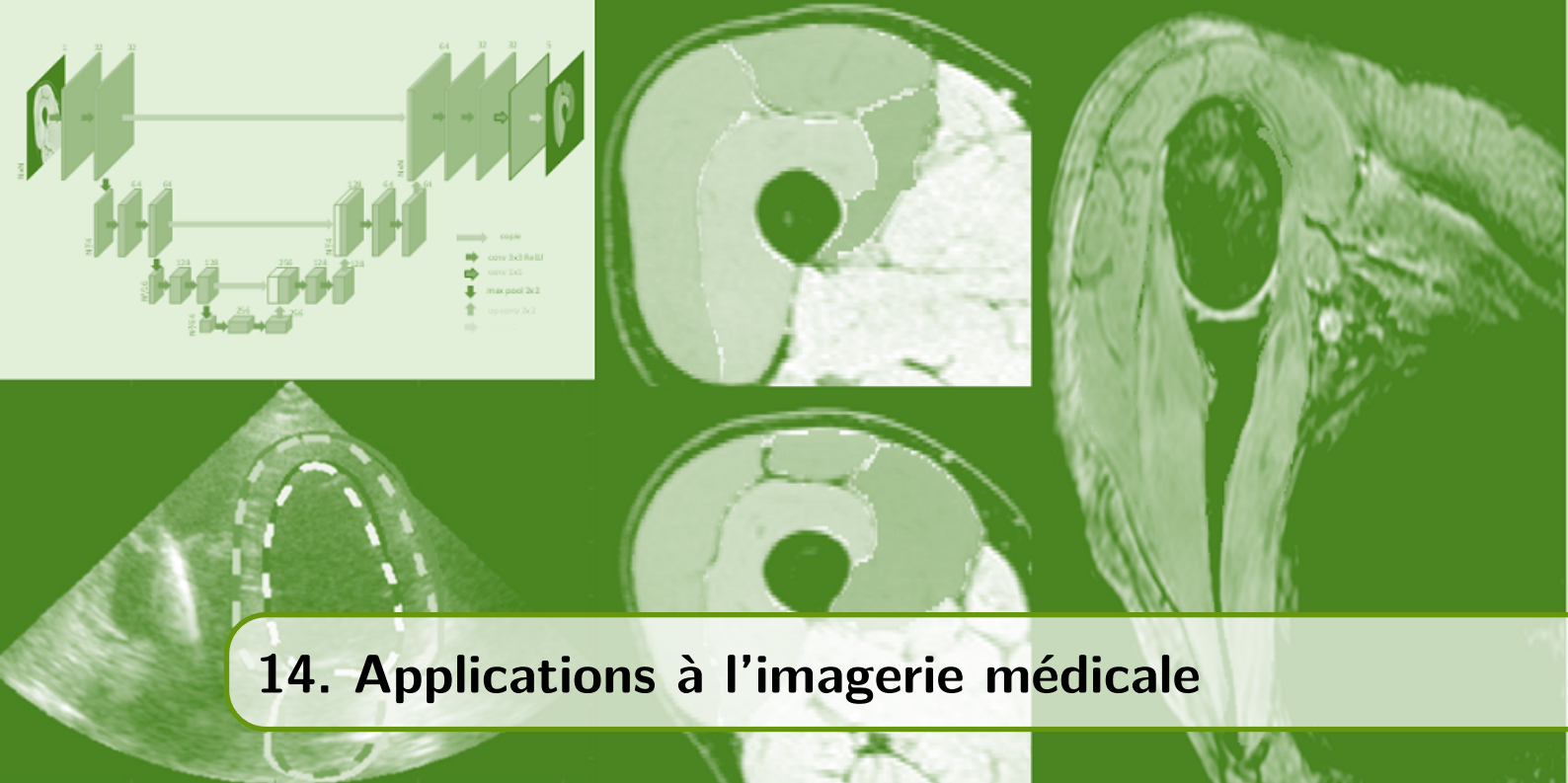
learning et la faciliter de proposer de nouvelles architectures, puis les tester, a conduit à une explosions des contributions en segmentation. Ensuite, il est quasiment impossible de savoir si une architecture fait vraiment mieux qu'une autre ou pas et encore moins sur un nouveau jeu de d'images. Enfin, il est clairement impossible de tester toutes les méthodes et encore moins d'optimiser leurs hyper-paramètres, ainsi que les différents pré et post traitements, afin de comparer quantitativement et statistiquement les approches entre elles. Avoir une méthode de référence, comme nnUNet, est alors un premier élément pertinent de comparaison.

Pour le reste du manuscrit, nous utiliserons l'architecture U-Net 2D et quelques une de ses variantes qui nous ont semblé pertinentes pour nos applications. Nos contributions utilisant U-Net sont décrites dans le chapitre 14.

13.6 Conclusion

Dans ce chapitre, nous avons présenté le réseau de référence en segmentation d'images médicales : U-Net. Nous avons aussi présenté ses variantes et les structures fondamentales de son architecture qui constituent aussi des éléments communs pour de nombreux réseaux de neurones. Actuellement, il n'existe pas de "recette" permettant de déterminer l'architecture et les hyper-paramètres les plus appropriés à un problème de segmentation donné. Il n'existe pas non plus de règles permettant de dimensionner les jeux de données.

Dans la suite, nous allons utiliser et donc spécifier -au mieux- cette approche pour différents problèmes de segmentation d'images médicales. Une partie de ces travaux va être exposée dans le prochain chapitre.



14. Applications à l'imagerie médicale

14.1 Introduction

Dans ce chapitre, nous présentons nos contributions en segmentation supervisée avec les approches JLF, JLF+CL et U-Net. Ces approches ont été appliquées à différentes modalités et contextes de segmentation :

- la segmentation des quadriceps et des muscles de l'épaule en IRM.
- la segmentation du coeur en échocardiographie.

Nous présentons aussi une contribution mélangeant les approches JLF+CL et U-Net pour la correction automatique de la segmentation des quadriceps.

Ces recherches ont été conduites dans le cadre des travaux de thèse de Sarah Leclerc et Hoai-Thu Nguyen.

14.2 Segmentation des muscles en IRM avec JLF+CL

Le premier travail de segmentation avec les approches JLF et JLF+CL a porté sur les images IRM des jambes d'athlètes ayant fait un ultra-marathon de l'étude MUST (voir annexe B). L'objectif est de proposer une approche de segmentation capable d'étudier la variation fonctionnelle dans les quadriceps. Donc de produire une segmentation suffisamment précise et reproductible pour les différents athlètes et différents instants du temps.

14.2.1 Problématique de segmentation de la base MUST

La segmentation des quadriceps en IRM reste un problème difficile d'une part à cause des in-homogénéités d'intensité, du défaut de champs, des artefacts d'acquisition et d'autre part car les frontières entre les muscles sont souvent difficiles à observer (PRESCOTT et al., 2011) notamment chez les personnes ayant un taux de graisse très faible et des muscles bien développés, comme par exemple les athlètes d'ultra endurance qui font justement l'objet de l'étude longitudinale MUST.

Une segmentation automatique robuste est attendue pour cette étude dont le but est de permettre l'analyse longitudinale des changements fonctionnels des quadriceps, car la

segmentation manuelle est très longue et mentalement épuisante, alors que les variations observées dans les images sont souvent très subtiles.

Certaines études récentes ont abordé la segmentation automatique des quadriceps (GILLES et al., 2016; LE TROTTER et al., 2016; PRESCOTT et al., 2011), mais aucune n'a décrit une segmentation précise des frontières des muscles, ce qui est très important pour la quantification des changements de volume dans les études longitudinales.

Comme l'état de l'art et les difficultés énumérées ci-avant concluaient que les méthodes non supervisées étaient peu adaptées, nous avons opté pour une méthode basée sur le recalage d'atlas : JLF+CL.

L'objectif étant de proposer un cadre de segmentation précis qui nécessiterait un nombre d'atlas aussi réduit que possible.

L'approche JLF+CL a été appliquée sur la base de données MUST de quadriceps (voir annexe B) où 7 atlas 3D sont disponibles. Pour cette application, il s'agit de segmenter 4 muscles sur des images IRM pondérée T1 comme le montre la figure 14.1 et d'être capable de réaliser une segmentation fiable et reproductible sur les trois instants du temps malgré un faible nombre d'atlas et une certaine variabilité anatomique entre les coureurs. On note que le score de DSC entre experts est de 0,910.

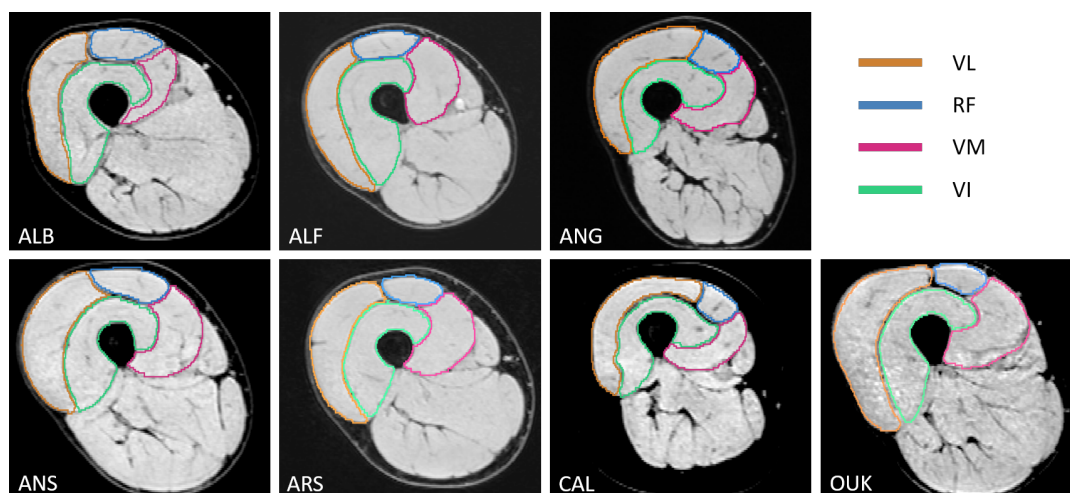


FIGURE 14.1 – Définition des 4 muscles à segmenter pour l'étude des quadriceps. Illustration de segmentations manuelles de la coupe centrale sur les 7 atlas. Les abréviations sont : VM – vectus medialis, VL – vectus lateralis, VI – vectus intermedius, RF – rectus femoris.

14.2.2 Optimisation des paramètres JLF+CL

Un première étude a porté sur l'impact des paramètres de la méthode sur les résultats de segmentation. Les paramètres ayant une influence significative sont :

- la taille des patches \mathcal{N} ,
- la taille du voisinage de recherche \mathcal{N}_r ,
- le rayon de la dilatation r_d de la zone de travail,
- la taille du patch pour l'extraction des caractéristiques \mathcal{N}_f .

Après avoir étudié leur impact sur le score de Dice (DSC), tout en considérant les temps de calcul (thèse H-T. Nguyen), les paramètres que nous avons retenus pour la suite sont : $\mathcal{N} = 5 \times 5 \times 5$, $\mathcal{N}_f = \mathcal{N}_r = 8 \times 8 \times 8$ pixels et $r_d = 15$.

Ensuite, nous avons étudié l'impact du nombre d'atlas sur les résultats de segmentation. Les évolutions des scores de DSC et de la distance de Hausdorff (HD) sont donnés sur la figure 14.2.

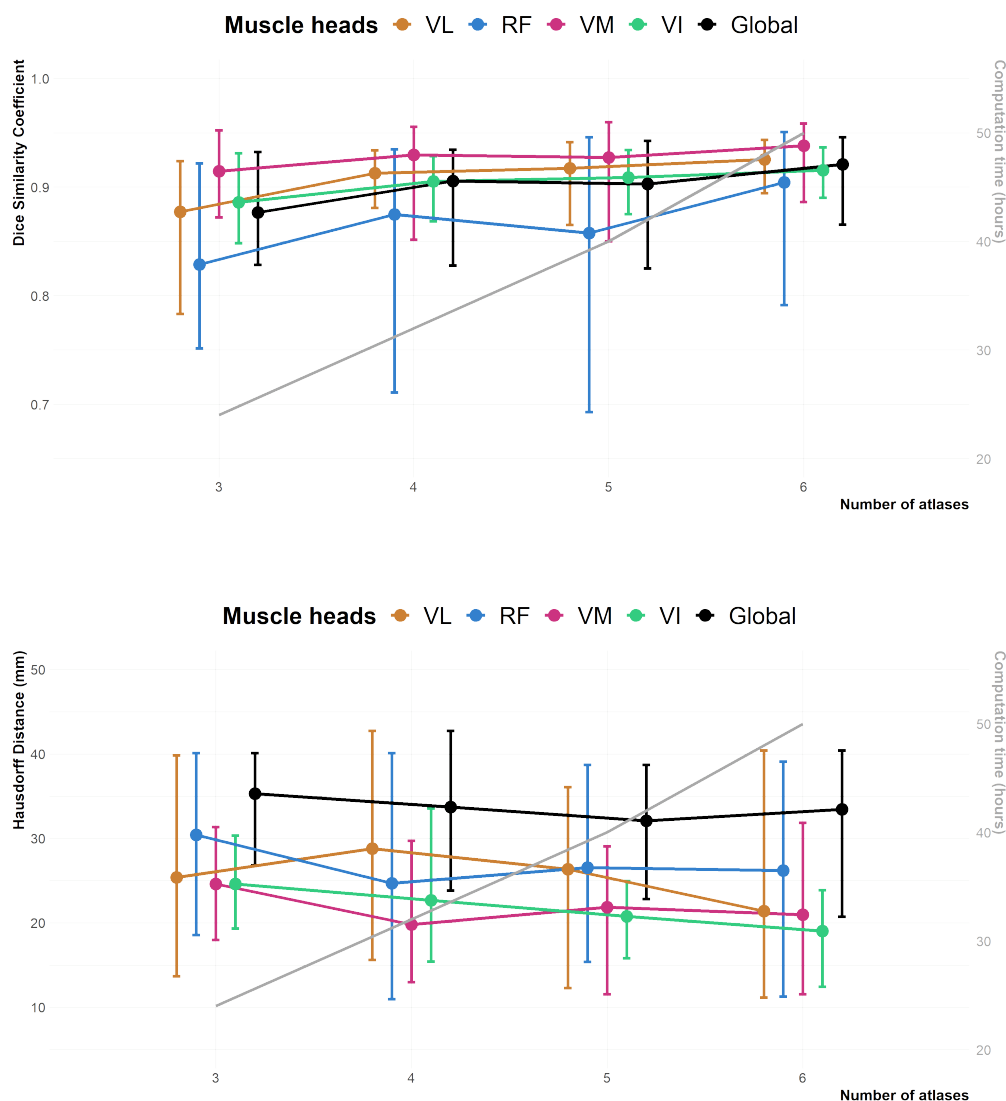


FIGURE 14.2 – Évolution des scores de DSC et HD en fonction du nombre d'atlas pour l'approche JLF+CL sur les 4 muscles étudiés (VL, RF, VM et VI).

On observe, avec l'augmentation du nombre d'atlas, une convergence assez rapide des performances de segmentation en moyenne comme en dispersion : le DSC augmente et sa variance diminue, la distance de Hausdorff tend à diminuer pour chacun des muscles. Ainsi, plus on utilise d'atlas, plus le résultat est bon et fiable quel que soit le choix des atlas. Cela signifie aussi que le choix des 3 atlas est critique pour obtenir de bons résultats. On observe aussi une augmentation quasi linéaire du temps de calcul avec le nombre d'atlas : il est de 23h pour 3 atlas et d'environ 50h pour 6 atlas.

Dans cette étude où nous disposons de 7 atlas, les meilleurs résultats sont logiquement obtenus avec l'utilisation de 6 atlas (1 atlas étant utilisé pour le test). Cependant l'amélioration apportée par le sixième atlas est faible et peut être discutée relativement au temps de calcul supplémentaire nécessaire (10h). Pour cette application, on peut extrapoler que les performances convergeraient très rapidement si on utilisait plus de 6 atlas.

L'analyse plus spécifique des résultats (figure 14.2) montre qu'un muscle, le *rectus femoris*

(RF), est moins bien segmenté que les autres et cela avec une grande disparité des résultats. Ceci peut se comprendre en observant les variabilités de forme et de position de ce muscle dans la base de données comme illustré sur la figure 14.1 : le recalage déformable n'arrive pas à le positionner convenablement et, ces variabilités étant peu représentées dans les atlas, la *corrective learning* cherche à le supprimer lorsque sa position n'est pas conforme aux données d'entraînement.

14.2.3 Comparaison des méthodes basées atlas

Nous avons comparé les DSC obtenus par plusieurs approches basées sur le recalage d'atlas : le recalage d'atlas avec un seul atlas, les approches JLF et JLF+CL avec 6 atlas, et l'approche de (GILLES et al., 2016; GILLES et al., 2010; GILLES et al., 2008) basée sur du recalage déformable où des points de contrôle sont positionnés manuellement. La figure 14.3 illustre un exemple de résultats. Les scores sont donnés dans la table 14.1. Il s'agit de valeurs moyennes obtenues sur les 7 segmentations des atlas.

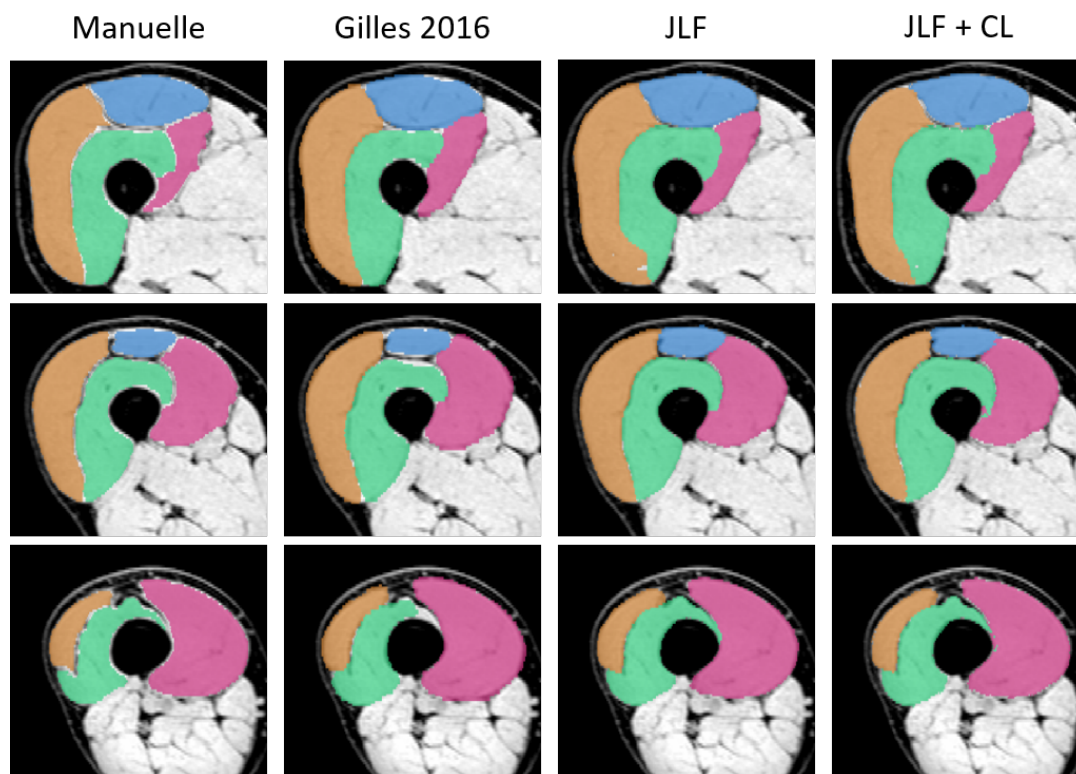


FIGURE 14.3 – Résultats de segmentation des quadriceps obtenus avec les approches de (GILLES et al., 2016), JLF (WANG et al., 2013b) et JLF+CL pour un coureur.

Parmi ces approches, JLF+CL permet d'obtenir les meilleurs résultats. On note l'apport de l'utilisation de plusieurs atlas : les approches JLF et JLF+CL permettent d'obtenir de meilleurs résultats qu'avec un seul atlas ou avec l'aide d'un utilisateur.

Sur la figure 14.3 et la table 14.2, on observe que l'apport de la correction CL est modéré. L'amélioration sur le muscle *rectus femoris* est la plus prononcée (+0.017 DICE).

Cette approche de correction détecte bien l'erreur liée au modèle et corrige en partie ce biais, mais elle peut aussi introduire de petites erreurs. C'est notamment le cas pour le muscle VI qui est aussi le muscle où la concordance entre les experts est la plus petite (voir table B1 dans l'annexe).

Méthode	DSC	DSC _w	HD (mm)	MAD (mm)	VS
<i>Recalage déformable</i>	.821	.829	39.13	3.69	.139
(GILLES <i>et al.</i> , 2016)	.864	.870	30.43	2.53	.114
JLF (WANG <i>et al.</i> , 2013b)	.914	.917	34.77	1.65	.080
JLF+CL (WANG <i>et al.</i> , 2013b)	.921	.923	33.44	1.46	.056

TABLE 14.1 – Evaluation quantitative des segmentations des quadriceps basées sur le recalage d’atlas pour les jambes droite des 7 sujets. L’approche de recalage déformable utilise un seul atlas. Les approches JLF et JLF+CL utilisent 6 atlas. La méthode de (GILLES *et al.*, 2016) est semi-automatique.

	ALB-2725	ALF-4529	ANG-2014	ANS-3229	ARS-4026	CAL-4223	OUK-2927	Moyenne
<i>VL</i>								
JLF	.931	.914	.876	.951	.944	.883	.938	.920
JLF+CL	.931	.932	.896	.942	.943	.894	.939	.925
<i>RF</i>								
JLF	.928	.926	.786	.946	.937	.737	.950	.887
JLF+CL	.936	.941	.824	.939	.948	.791	.951	.904
<i>VM</i>								
JLF	.924	.953	.923	.954	.953	.880	.950	.934
JLF+CL	.937	.959	.930	.952	.956	.886	.945	.938
<i>VI</i>								
JLF	.903	.917	.906	.932	.934	.890	.918	.914
JLF+CL	.899	.935	.901	.927	.937	.890	.920	.916
<i>Global</i>								
JLF	.921	.927	.873	.945	.942	.848	.939	.914
JLF+CL	.926	.942	.888	.940	.946	.865	.939	.921

TABLE 14.2 – Détails des scores de DSC obtenus par les approches JLF et JLF+CL pour la base MUST. Les valeurs en gras indiquent les améliorations obtenues avec l’approche CL. Abréviations des muscles : VL – Vastus Lateralis, RF – Rectus Femoris, VM – Vastus Medialis, VI – Vastus Intermedius.

14.2.4 Généralisation pour le longitudinal

Afin d’évaluer la capacité des approches JLF et JLF+CL à produire de bonnes segmentations sur les 3 instants du temps, et ainsi permettre une étude longitudinale, nous avons calculé les scores de DICE et de similarité de volume (VS) entre les segmentations manuelles de 17 coupes de 4 coureurs aux trois instants disponibles (avant la course (Pré), juste après leur arrivée (Post), puis 3 jours après (Post+3)) avec les segmentations produites par JLF et JLF+CL ainsi que par U-Net.

Il faut rappeler que tous les atlas utilisés pour réaliser les segmentations proviennent d’images acquises avant la course et que ce jeu d’images ne change pas pour l’entraînement. Pour compléter cette étude, nous avons aussi recalé l’atlas de l’instant Pré sur les autres instants du temps d’un même coureur. Nous avons fait de même avec les segmentations obtenues à l’instant Pré par les approches JLF, JLF+CL et U-Net .

Les résultats des métriques sont donnés dans le tableau 14.3.

Ces DSC sont calculés en 2D sur 17 coupes et ne peuvent être comparés aux résultats précédents calculés en 3D sur tout le volume. Néanmoins, ils permettent d’observer la bonne capacité de ces deux approches à se généraliser aux différents instants malgré l’utilisation d’atlas provenant uniquement de données acquises avant la course (Pré). Précisons que

sur les IRM pondérées T1 utilisées, le changement principal observable suite à un ultra marathon est une évolution du volume musculaire (FLECKENSTEIN, 1993).

L'apport du CL est discutable et cela s'explique car cette étape n'a pu apprendre les corrections à faire pour prendre en considération l'évolution du volume suite à l'effort physique. L'apprentissage des corrections ne pourra vraisemblablement pas se faire par cette approche car l'évolution est très patient spécifique.

Un autre projet nous a permis de disposer de plus de données et ainsi vérifier la généralisation de l'approche et de nos observations. Il s'agit du projet de segmentation de muscles de l'Epaule qui est expliqué dans la prochaine section.

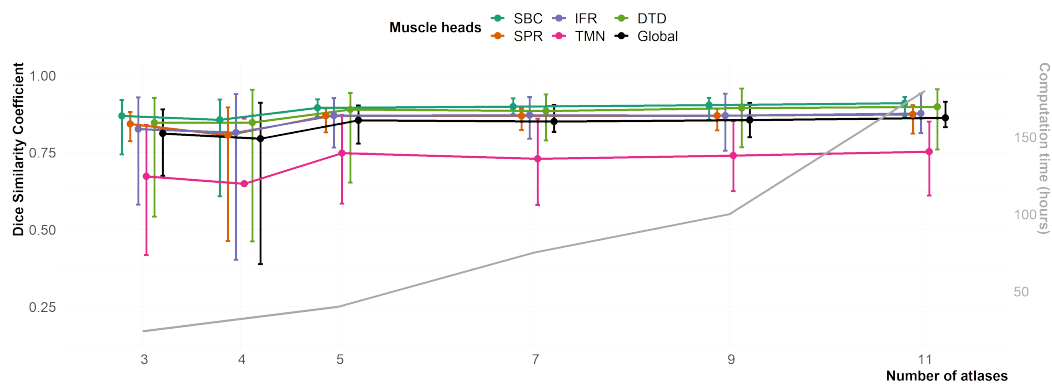
Instant/Méthode	DSC				VS			
	ALB	ALF	ARS	OUK	ALB	ALF	ARS	OUK
<i>Pré</i>								
JLF	.914	.933	0.947	0.945	.108	.084	.044	.044
JLF+CL	.922	.950	.953	.946	.079	.028	.027	.047
U-Net	.917	.929	.949	.946	.120	.079	.065	.030
<i>Post</i>								
Manuelle Pré, recalé	.914	.929	.934	.921	.111	.070	.066	.039
JLF+CL Pré, recalé	.925	.923	.935	.923	.040	.047	.059	.043
UNet Pré, recalé	.936	.914	.943	.926	.032	.047	.018	.025
JLF	.935	.929	.945	0.929	.024	0.019	0.034	0.044
JLF+CL	.923	.923	.942	.920	.060	.049	.056	.062
U-Net	.934	.914	.945	.934	.018	.072	.029	.021
<i>Post+3</i>								
Manuelle Pré, recalé	.919	.924	.934	.937	.085	.060	.058	.038
JLF+CL Pré, recalé	.924	.918	.935	.941	.046	.068	.061	.038
UNet Pré, recalé	.934	.909	.948	.942	.036	.052	.015	.020
JLF	.924	.923	.945	.943	.029	.053	.032	.042
JLF+CL	.911	.912	.944	.940	.052	.068	.062	.025
U-Net	.933	.931	.952	.946	.034	.032	.024	.027

TABLE 14.3 – Comparaison quantitative des approches JLF, JLF+CL et U-Net en longitudinal pour 4 coureurs de MUST (ALB, ALF, ARS and OUK). Le DSC et le VS ont été calculés sur les même 17 coupes pour les instant *Pre*, *Post*, and *Post+3*. Les cellules en gris indiquent les meilleurs résultats pour chaque instant.

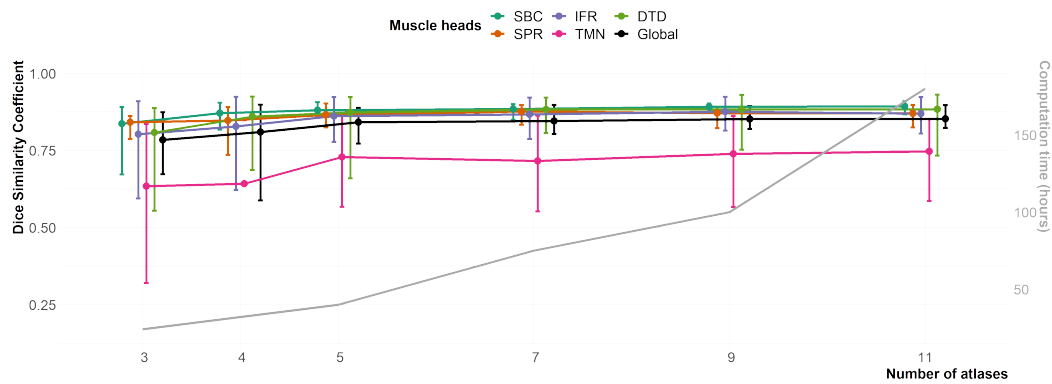
14.2.5 Généralisation à la base Epaule

Sur la base de données Epaule (voir annexe B) où 27 atlas 3D sont disponibles, les résultats obtenus ont confirmé les observations précédentes. Les approches JLF et JLF+CL ont été appliquées aux volumes 3D de la séquence *Dixon echo 8* après pré-traitements (décrits dans l'annexe).

Sur cette base de données, nous avons étudié l'évolution de la qualité des segmentations en fonction de l'augmentation du nombre d'atlas (figure 14.4). Les valeurs moyennes et d'écart types ont été calculées pour 10 patients. On observe qu'au delà de 5 atlas, il n'y a plus d'améliorations nettes des valeurs moyennes ou de la dispersion des valeurs. La figure 14.5 illustre un exemple de segmentations obtenues par ces approches avec 5 atlas tirés au hasard parmi les 17 disponibles.



(a) DSC JLF



(b) DSC JLF+CL

FIGURE 14.4 – Evolution des scores de DSC et du temps de calcul en fonction du nombre d'atlases pour les approches JLF (en haut) et JLF+CL (en bas).

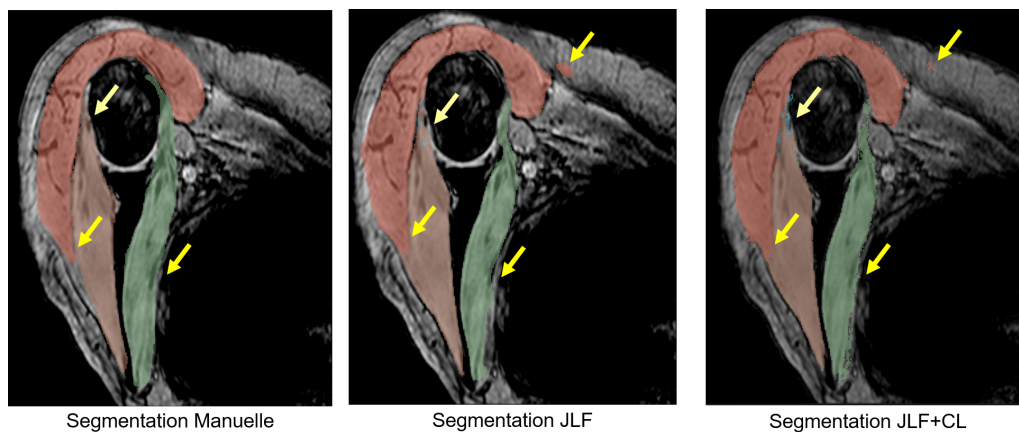


FIGURE 14.5 – Exemple de segmentations des muscles de l'épaule obtenues avec les approches JLF et JLF+CL en utilisant 5 atlas tirés au hasard. Les temps de calcul pour les 2 approches sont de 24h pour JLF et de 120h+30h pour le JLF+CL. Les flèches pointent des différences de segmentation.

Les expériences sur cette base ont aussi permis d'observer la sensibilité de la méthode au recalage. En effet, l'approche JLF suppose un très bon recalage de chaque examen vers tous

les autres. Or ceci n'est pas toujours vrai. Typiquement dans cette étude, pour 3 patients il n'est pas évident de garantir le bon fonctionnement du recalage. La table 14.4 donne les métriques des segmentations obtenues en considérant la totalité de la base de données pour l'apprentissage des approches JLF et JLF+CL, tandis que la table 14.5 donne ces mêmes métriques lorsque les 3 patients problématiques sont retirés de la base pour l'entraînement. L'amélioration des résultats est quasi systématique.

		DSC	HD (mm)	MAD (mm)
3 atlas	JLF	0.84±0.09	41.9±20.0	3.19±1.69
	JLF + CL	0.81±0.09	33.1±11.9	2.87±1.50
4 atlas	JLF	0.83±0.14	42.7±15.4	3.96±4.92
	JLF + CL	0.85±0.06	29.7±8.6	3.20±3.72
5 atlas	JLF	0.88±0.05	37.1±16.8	2.01±0.91
	JLF + CL	0.87±0.05	27.3±6.3	1.91±0.80

TABLE 14.4 – Résultats de segmentation des approches JLF et JLF+CL pour les données Epaule, avec tous les patients. Moyennes et écarts types calculés pour 10 patients.

		DSC	HD (mm)	MAD (mm)
3 atlas	JLF	0.88±0.03	36.9±10.1	2.46±1.21
	JLF + CL	0.80±0.11	32.0±9.5	2.70±1.73
4 atlas	JLF	0.89±0.03	36.6±8.7	2.07±0.69
	JLF + CL	0.86±0.04	28.95±6.5	2.08±0.76
5 atlas	JLF	0.90±0.02	32.8±10.3	1.85±0.66
	JLF + CL	0.87±0.05	28.2±6.5	1.98±0.89

TABLE 14.5 – Résultats de segmentation des approches JLF et JLF+CL pour les données Epaule, sans les 3 patients faisant échouer les recalages. Moyennes et écarts types calculés pour 10 patients.

14.2.6 Discussion

Les approches JLF et JLF+CL sont capables de fournir de très bonnes segmentations avec un nombre très raisonnable d'atlas. Cependant, lorsqu'on utilise peu d'atlas, le choix des atlas utilisés influence énormément les résultats de segmentation. Or, les approches JLF et JLF+CL sont extrêmement longues et le temps dépend directement du nombre d'atlas utilisés, comme le montre la figure 14.6.

Pour les paramètres retenus ici, en moyenne, JLF nécessite 50h sur un CPU 16 cœurs pour effectuer une segmentation avec 6 atlas. Pour l'entraînement du JLF+CL avec 6 atlas, cela correspond à 50×6 h pour les JLF et 7h par étiquette pour l'étape de CL, soit pour les quadriceps (4 muscles plus le fond) un total de $300 + 35 = 335$ h¹. L'inférence dure ensuite 48h pour le JLF et 50h pour le CL.

Dans le but de diminuer ces temps de calcul, Wang a proposé une approche multi-échelle (WANG et al., 2018a) qui réalise les JLF sur des images de plus faible résolution. En supposant que les erreurs de segmentations ajoutées par l'utilisation du sous-échantillonnage sont des erreurs systématiques, il est alors possible d'apprendre à corriger ces erreurs avec

1. Ces temps n'incluent pas les recalages affines et déformables réalisés au préalable avec la stratégie un vers tous. Un recalage affine puis déformable d'un atlas (image et segmentation) durent en moyenne 10 minutes. Soit 6h pour 6 atlas.

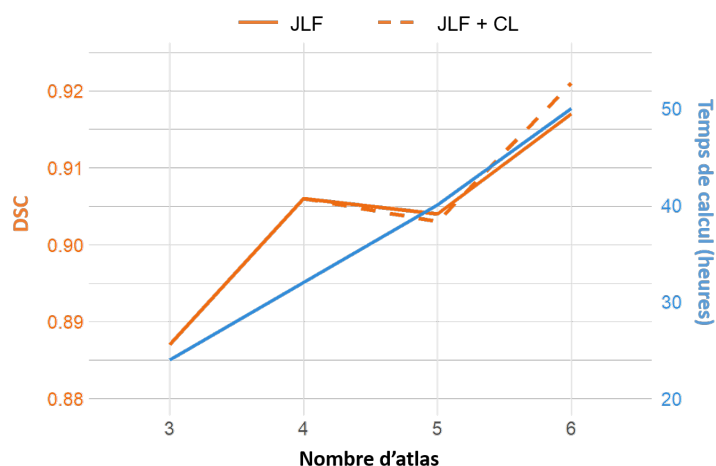


FIGURE 14.6 – Influence du nombre d’atlas sur la qualité de segmentation (DSC) et le temps de calcul pour les données MUST. Le temps de calcul est donné uniquement pour l’étape JLF étant donné que le nombre d’atlas n’influence pas le temps de calcul de l’étape de CL.

l’algorithme CL qui lui utilisera les données de haute résolution. Nos études sur les différentes stratégies multi-échelles n’ont pas permis d’obtenir des résultats de segmentations et des gains de temps convaincants.

Une autre voie pour diminuer très significativement le temps de calcul est de réduire le nombre d’atlas utilisés : passer de 6 à 3 atlas divise par deux le temps de calcul. Une solution d’une telle sélection est présentée dans le chapitre 15. Cependant, elle reste longue car basée sur 3 atlas (25h).

Nous avons donc testé les approches basées U-Net qui seront développées dans la prochaine section.

14.3 Segmentations avec U-Net

Nous présentons ici les travaux réalisés en segmentation avec U-Net sur les images échocardiographiques et IRM de la base Epaule.

14.3.1 Segmentation d’images écho-cardiographique avec U-Net

Dans le cadre de l’imagerie écho-cardiographique, nous avons proposé et optimisé l’utilisation d’U-Net pour la segmentation du myocarde (LECLERC et al., 2019a ; LECLERC et al., 2019c). Cette tâche est rendue particulièrement complexe par la qualité des images ultrason caractérisée par la variabilité des textures dans les différents tissus, l’absence de certains contours et la présence d’artéfacts (figure 14.7).

Comme le montre la table 14.6, cette approche d’apprentissage profond surpasse considérablement les résultats de segmentation obtenus par les meilleures approches de la littérature, comme celle des BEASM basée sur des modèles B-spline de contours actifs appliquées de manière automatique ou semi-automatique. Deux résultats de segmentation du ventricule gauche sont donnés sur la figure 14.8.

Les deux modèles U-Net1 et U-Net2 que nous avons proposés, utilisent des architectures et régularisations différentes. Le modèle U-Net1 comporte 2 millions de paramètres avec un schéma de décodeur simpliste (répétition), est entraîné sans normalisation des mini-lots et avec une fonction de coût basée sur le DICE. Le modèle U-Net2 comporte 18 millions de paramètres, utilise des *déconvolutions* (voir 13.4.3) pour le décodeur, une normalisation

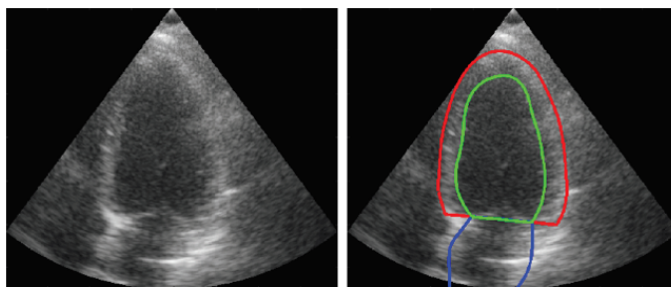


FIGURE 14.7 – Echocardiographie (à gauche) et sa segmentation experte (à droite) de l'épicaide du ventricule gauche (en rouge), l'endocarde du ventricule gauche (en vert) et de l'oreillette gauche (en bleue).

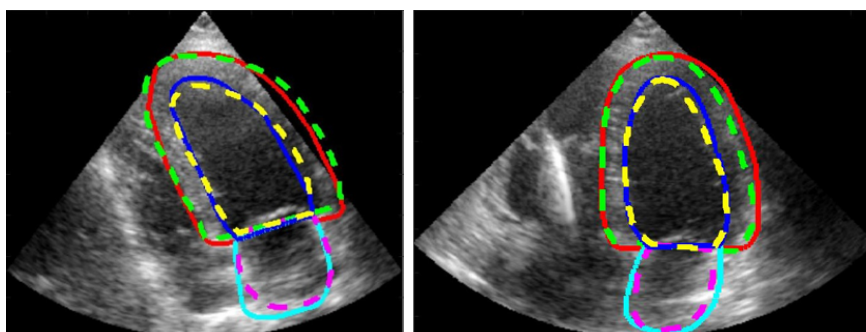


FIGURE 14.8 – Exemples de résultats de segmentation du ventricule gauche obtenus avec U-Net sur la base de données électrocardiographie CAMUS. La segmentation experte est en pointillée, la segmentation U-Net (réseau U-Net2) est en trait plein.

des mini-lots et est entraîné avec une fonction d'entropie croisée où le taux d'apprentissage décroît progressivement. Cependant, malgré ces différences, on peut constater sur la table 14.6 que les performances de ces 2 réseaux sont très proches.

Actuellement, les résultats de segmentation obtenus avec U-Net sont de qualité comparable (voir meilleure) à ceux obtenus entre deux experts et la variabilité des segmentations automatiques est de l'ordre de la variabilité intra-expert.

Pour arriver à de tels résultats, l'entraînement des réseaux a été fait sur une base de données de 500 patients où deux instants caractéristiques du cycle cardiaque ont été manuellement segmentés (LECLERC et al., 2019c). La construction de cette base de données CAMUS² a demandé un investissement important et il n'est pas toujours possible d'avoir une telle quantité de données avec une variabilité de qualité d'acquisition et de patients ainsi qu'une annotation manuelle fiable et riche.

Nos études ont montré qu'à partir de 300 patients utilisés pour l'apprentissage, les segmentations ne sont plus beaucoup améliorées et ainsi augmenter la taille de cette base de données n'améliorerait que marginalement les résultats de segmentation. La stagnation des scores vient principalement de quelques rares images qu'il n'est pas possible d'apprendre avec les architectures U-Net testées. Ces images aberrantes sont liées à une anatomie, une prise de vue, une qualité d'image ou un artefact très particulier.

Pour quantifier ces segmentations aberrantes, nous avons proposé des métriques d'évaluation de la qualité de segmentation complémentaires aux mesures classiques et permettant de prendre en compte ces erreurs flagrantes de géométrie et d'anatomie du ventricule (LECLERC et al., 2019b).

2. <https://www.creatis.insa-lyon.fr/Challenge/camus/>

TABLE 14.6 – Comparaison des performances de huit méthodes de segmentation du ventricule gauche (LV_{endo} et LV_{epi}) sur 406 patients dont la qualité de l'image échocardiographique a été jugée bonne ou moyenne. Les valeurs en gras correspondent aux meilleures performances obtenues pour chaque mesure. Les p-values sont basées sur le test signé de Wilcoxon entre les méthodes U-Net 1 et U-Net 2 pour chaque métrique. Les mesures inter et intra experts sont réalisées sur 40 patients dont la qualité des images échocardiographiques a été jugée bonne ou moyenne.

Méthodes *	Fin Diastole						Fin Systole					
	LV_{endo}			LV_{epi}			LV_{endo}			LV_{epi}		
	DSC	MAD	HD	DSC	MAD	HD	DSC	MAD	HD	DSC	MAD	HD
val.	mm	mm	val.	mm	mm	val.	mm	mm	val.	mm	mm	
O_{1a} vs O_2 (inter-exp)	0.919 ± 0.033	2.2 ± 0.9	6.0 ± 2.0	0.913 ± 0.037	3.5 ± 1.7	8.0 ± 2.9	0.873 ± 0.060	2.7 ± 1.2	6.6 ± 2.4	0.890 ± 0.047	3.9 ± 1.8	8.6 ± 3.3
O_{1a} vs O_3 (inter-exp)	0.886 ± 0.050	3.3 ± 1.5	8.2 ± 2.5	0.943 ± 0.018	2.3 ± 0.8	6.5 ± 2.6	0.823 ± 0.091	4.0 ± 2.0	8.8 ± 3.5	0.931 ± 0.025	2.4 ± 1.0	6.4 ± 2.4
O_2 vs O_3 (inter-exp)	0.921 ± 0.037	2.3 ± 1.2	6.3 ± 2.5	0.922 ± 0.036	3.0 ± 1.5	7.4 ± 3.0	0.888 ± 0.058	2.6 ± 1.3	6.9 ± 2.9	0.885 ± 0.054	3.9 ± 1.9	8.4 ± 2.8
O_{1a} vs O_{1b} (intra-exp)	0.945 ± 0.019	1.4 ± 0.5	4.6 ± 1.8	0.957 ± 0.019	1.7 ± 0.9	5.0 ± 2.3	0.930 ± 0.031	1.3 ± 0.5	4.5 ± 1.8	0.951 ± 0.021	1.7 ± 0.8	5.0 ± 2.1
BEASM-auto	± 0.074 0.879 ± 0.065	± 3.6 3.3 ± 1.8	± 10.2 9.2 ± 4.9	± 0.057 0.895 ± 0.051	± 2.0 3.9 ± 2.1	± 9.1 10.6 ± 5.1	± 0.137 0.826 ± 0.092	± 7.8 3.8 ± 2.1	± 13.6 9.9 ± 5.1	± 0.078 0.880 ± 0.054	± 4.7 4.2 ± 2.0	± 11.1 11.2 ± 5.1
BEASM-semi	0.920 ± 0.039	2.2 ± 1.2	6.0 ± 2.4	0.917 ± 0.038	3.2 ± 1.6	8.2 ± 3.0	0.861 ± 0.070	3.1 ± 1.6	7.7 ± 3.2	0.900 ± 0.042	3.5 ± 1.7	9.2 ± 3.4
U-Net 1	0.934 ± 0.042	1.7 ± 1.0	5.5 ± 2.9	0.951 ± 0.024	1.9 ± 0.9	5.9 ± 3.4	0.905 ± 0.063	1.8 ± 1.3	5.7 ± 3.7	0.943 ± 0.035	2.0 ± 1.2	6.1 ± 4.1
U-Net 2	0.939 ± 0.043	1.6 ± 1.3	5.3 ± 3.6	0.954 ± 0.023	1.7 ± 0.9	6.0 ± 3.4	0.916 ± 0.061	1.6 ± 1.6	5.5 ± 3.8	0.945 ± 0.039	1.9 ± 1.2	6.1 ± 4.6
ACNN	0.932 ± 0.034	1.7 ± 0.9	5.8 ± 3.1	0.950 ± 0.026	1.9 ± 1.1	6.4 ± 4.1	0.903 ± 0.059	1.9 ± 1.1	6.0 ± 3.9	0.942 ± 0.034	2.0 ± 1.2	6.3 ± 4.2
SHG	0.934 ± 0.034	1.7 ± 0.9	5.6 ± 2.8	0.951 ± 0.023	1.9 ± 1.0	5.7 ± 3.3	0.906 ± 0.057	1.8 ± 1.1	5.8 ± 3.8	0.944 ± 0.034	2.0 ± 1.2	6.0 ± 4.3
U-Net ++	0.927 ± 0.046	1.8 ± 1.1	6.5 ± 3.9	0.945 ± 0.026	2.1 ± 1.0	7.2 ± 4.5	0.904 ± 0.060	1.8 ± 1.0	6.3 ± 4.2	0.939 ± 0.034	2.1 ± 1.1	7.1 ± 5.1
p-values	$\ll 0.05$	$\ll 0.05$	$\ll 0.05$	$\ll 0.05$	$\ll 0.05$	≈ 0.83	$\ll 0.05$	$\ll 0.05$	$\ll 0.05$	$\ll 0.05$	$\ll 0.05$	≈ 0.48

* LV_{endo} : Contour de l'endocarde du ventricule gauche ; LV_{epi} : Contour de l'épicarde sur ventricule gauche
 FD : Fin diastole ; FS : Fin systole ; DSC : Indice de Dice
 MAD : Distance absolue moyenne ; HD : Distance de Hausdorff

Puis, afin de réduire ces erreurs, nous avons proposé l'ajout d'un mécanisme d'attention avec deux réseaux U-Net consécutifs, permettant de restreindre l'image à la région du myocarde (LECLERC et al., 2020 ; LECLERC et al., 2019a). Sur 500 patients, le traitement des images avec cette dernière approche (LU-Net) permet de réduire de 71 à 20 le nombre de segmentations aberrantes.

Conclusion

Ces premières études ont montré l'intérêt de l'approche U-Net pour les problèmes de segmentation anatomique, c'est à dire où les objets à segmenter respectent une géométrie et une organisation (ici anatomique).

Elles ont aussi mis en évidence que différentes architectures fournissent des résultats assez

proches (U-Net1, U-Net2 et U-Net++) et que seulement quelques hyper-paramètres d'architecture sont critiques (U-Net1 vs. U-Net2) mais dont la variation impacte modérément les résultats. Ces paramètres sont cependant intriqués les uns avec les autres.

En revanche, la base d'images pour l'apprentissage et les pré-traitements sont les éléments importants. Rappelons que la construction de cette base de données a été difficile et longue et que de nombreuses autres études de segmentation ne pourront pas disposer d'autant d'images et encore moins annotées. Il est critique de comprendre les besoins d'U-Net et d'utiliser des méthodes d'augmentation de données (ce qui sera abordé au chapitre 15).

14.3.2 Segmentation des muscles de l'épaule

La base de données Epaule dispose de 27 segmentations 3D qui, grâce aux pré-traitements, peuvent toutes être considérées pour entraîner et évaluer un réseau de neurones quelle que soit l'épaule (gauche ou droite), la résolution et l'orientation des acquisitions. Dix épaules servent de tests, les 17 autres vont constituer la base d'entraînement de différents réseaux. Pour cette base de données, on dispose ainsi de 17 volumes de taille $384 \times 384 \times 180$. Les réseaux de neurones sont soit 2D et les segmentations obtenues sont empilées pour reformer le volume 3D, soit les réseaux sont 3D. Les métriques quantitatives sont calculées sur les volumes 3D.

Etude des plans 2D

Plusieurs réseaux ont été testés dans le cadre des stages de Malick Kandji et David Osowiechi. On étudie ici que les réseaux U-Net (RONNEBERGER et al., 2015) et quelques une de ses dérivées (ResU-Net (ZHANG et al., 2018) et U-Net3+ (HUANG et al., 2020)). Dans un premier temps un réseau U-Net classique a été utilisé avec l'image "echo 8" comme unique entrée. Nous avons comparé les performances de trois réseaux U-Net appris chacun sur un plan de référence anatomique différent (transversal, coronal et sagittal). Nous avons aussi fusionné ces trois segmentations : chaque pixel est affecté à la classe du réseau ayant la probabilité la plus importante. Ce qui correspond donc au choix le plus sûr, ou dans lequel on peut avoir le plus confiance. Les résultats sont donnés dans la table 14.7 et illustrés dans la figure 14.9.

U-Net	DSC _w	HD (mm)	MAD (mm)	VS
<i>U-Net Transversal</i>	.849 ± 0.07	75.3 ± 13.5	3.74 ± 2.46	0.170 ± 0.08
<i>U-Net Coronal</i>	.875 ± 0.04	67.9 ± 20.5	2.42 ± 0.72	0.116 ± 0.07
<i>U-Net Sagittal</i>	.857 ± 0.06	85.0 ± 17.8	3.50 ± 1.47	0.189 ± 0.13
<i>Au Plus Confiant (APC)</i>	0.892 ± 0.04	48.6 ± 14.8	2.12 ± 0.70	0.146 ± 0.08

TABLE 14.7 – Evaluations quantitatives des segmentations U-Net obtenues par l'apprentissage suivant chacun des plans de référence anatomique pour la base de données Epaule avec l'image "e8" comme entrée.

On remarque que les résultats sont différents et que le plan coronal donne de meilleurs résultats. Ceci n'est pas forcément évident car les pré-traitements ont rendu isotropiques les pixels et l'acquisition IRM a été faite dans le plan transversal. Il est possible que l'organisation musculaire ou une meilleure discrimination du contexte suivant ce plan par le réseau U-Net, explique ceci.

Il apparaît aussi que la segmentation "au plus confiant" (abrégé APC), qui fusionne ces trois segmentations, donne les meilleurs résultats. Cette approche d'ensemble fonctionnant mieux que chaque méthode indépendamment, les apprentissages sont donc complémentaires.

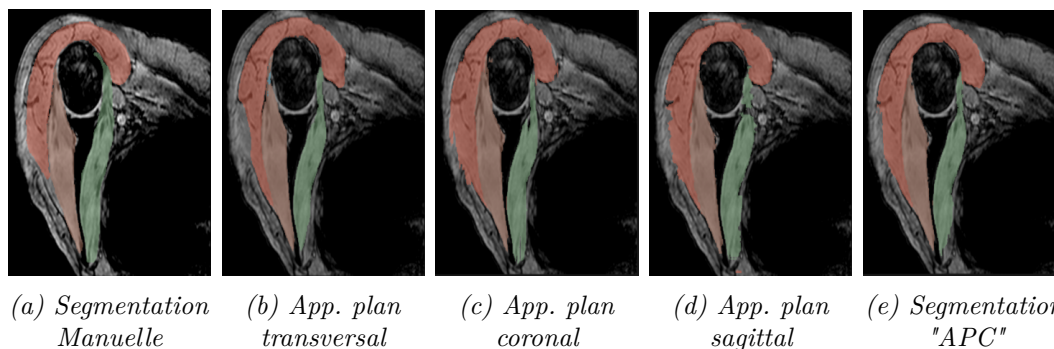


FIGURE 14.9 – Exemple de segmentations par U-Net obtenues par l'apprentissage suivant chacun des plans de référence anatomique pour la base de données Epaule. La segmentation APC ("au plus confiant") est obtenue en prenant la classe ayant la probabilité la plus grande parmi les segmentations issues de chacun des plans.

Tests d'architectures

Dans le cadre de cette étude, afin d'améliorer les résultats et de comprendre l'impact de certaines modifications, nous avons testé des changements sur le nombre de séquences IRM en entrée du réseau ainsi que des variantes d'U-Net récentes et l'apprentissage par transfert. Incrémentant les travaux de (ZHOU et al., 2020), l'approche U-Net3+ de (HUANG et al., 2020) propose plusieurs améliorations dont :

- des sauts de connexions entre toutes les échelles du réseau U-Net,
- la systématisation d'un réseau résiduel (ResNet101) pour l'encoder d'U-Net et l'utilisation de poids provenant d'un entraînement de classification d'images naturelles et couleurs de la base ImageNet,
- l'utilisation d'une fonction de coût sommant trois métriques (le SSIM (WANG et al., 2004b), l'intersection sur l'union (IoU) et la perte focale) et ceci à chaque changement de résolution du décodeur. Cette fonction de coût est nommée la supervision profonde (DS).

Pour la base de données Epaule, les résultats de ces améliorations successives sont donnés dans la table 14.8 et un exemple de segmentation de deux méthodes est illustré sur la figure 14.10

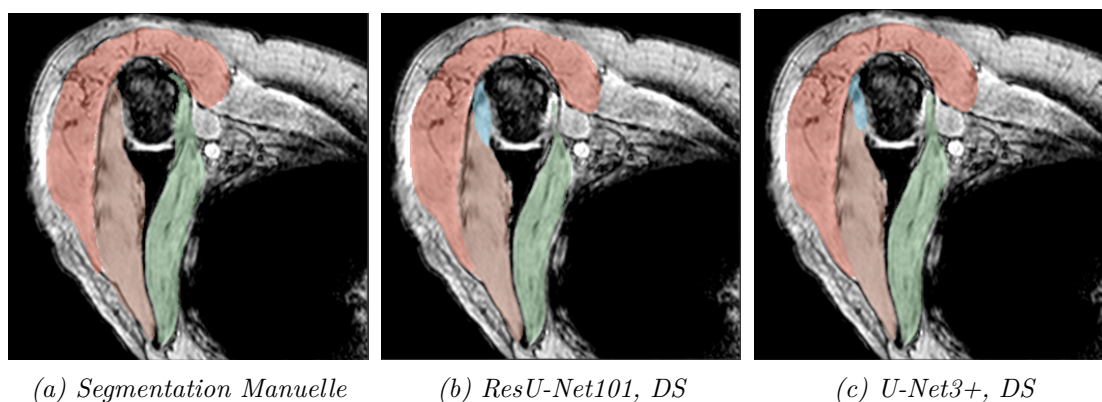


FIGURE 14.10 – Résultats de segmentation des meilleures méthodes basées U-Net sur la base de données Epaule. DS l'abréviation pour supervision profonde.

	Séquences	Poids initiaux	Supervision profonde	DSCw	HD (mm)	MAD (mm)	VS
U-Net APC	e8	aléatoire		.892 ± 0.04	48.6 ± 14.8	2.12 ± 0.70	0.146 ± 0.08
ResU-Net101	e8	ImageNET		.883 ± 0.05	72.1 ± 18.0	3.44 ± 2.75	0.145 ± 0.09
	e2, e5, e8	aléatoire		.857 ± 0.07	76.8 ± 12.9	5.56 ± 5.24	0.202 ± 0.15
	e2, e5, e8	ImageNET		.891 ± 0.05	79.4 ± 8.6	5.39 ± 5.58	0.165 ± 0.14
	e2, e5, e8	ImageNET	oui	.893 ± 0.04	63.2 ± 28.4	3.22 ± 2.80	0.148 ± 0.10
U-Net3+	e2, e5, e8	ImageNET		.893 ± 0.05	62.7 ± 23.7	4.16 ± 3.98	0.151 ± 0.12
	e2, e5, e8	ImageNET	oui	.887 ± 0.05	59.7 ± 25.6	3.35 ± 2.72	0.178 ± 0.11
AttentionU-Net	e2, e5, e8	aléatoire	oui	.910 ± 0.01	50.9 ± 10.0	2.42 ± 0.857	0.111 ± 0.02
3D Cascade*	e8	aléatoire	oui	.891	12.41	-	-
3D FullRes*	e8	aléatoire	oui	.903	9.80	-	-
TransU-Net*	e8 - 2D	aléatoire	oui	.856	15.18	-	-
	e8 - 3D	aléatoire	oui	.891	13.09	-	-

TABLE 14.8 – *Evaluations quantitatives des segmentations obtenues par les différentes variantes d'U-Net sur la base de données EpauLe. Les 3 dernières architectures marquées d'une '*' ont été testées via le framework nnU-Net avec pré- et post- traitements sur 5 blocs.*

Discussion

L'analyse de ce tableau montre que, toujours pour ces images, les améliorations notables viennent de combinaisons de changements :

- le passage à ResU-Net101 et l'utilisation de poids initiaux venant d'une base de données d'images couleurs améliorent un peu le DSCw, mais pas les autres métriques,
- pour l'utilisation de trois canaux en entrée, il faut utiliser un réseau pré-entraîné,
- la supervision profonde n'améliore pas toujours les résultats,
- les post-traitements (conservation de la plus grande région) et les approches 3D améliorent considérablement les métriques notamment la distance de Hausdorff (division par 5),

Ces résultats sont propres à cette étude portant sur la segmentation multi-classes de muscles en IRM. Des études en cours (stage de master) visent à mieux comprendre le rôle des canaux en entrée et leur utilité. On explore aussi l'approche nnU-Net et les enjeux de post- et pré-traitements, ce qui permet de faire des liens avec la partie précédente sur les filtres anisotropes et *mean-shift*.

On explore aussi la localisation des erreurs faites par les réseaux par rapport aux annotations manuelles. La base de données EpauLe a permis de mettre en évidence l'extrême rigueur que l'on doit avoir lors de la création des annotations manuelles. Sans cela, l'apprentissage est plus complexe et les résultats quantitatifs moins bons. Sur la figure 14.11, on met en évidence quelques erreurs d'annotation manuelle. Les tendons (flèche sur le petit rond) n'ont pas été systématiquement segmentés manuellement. L'artéfact (entouré) a sûrement conduit à une sous-segmentation du deltoïde par l'expert sur la coupe 72. La correction de ces annotations manuelles amélioreront substantiellement les scores des approches de segmentation automatiques. Il reste quand même une vraie difficulté pour les algorithmes à déterminer précisément la frontière entre le petit rond (en bleu) et l'infra épineux (en marron).

14.4 Contribution à la correction automatique

On revient sur les résultats obtenus sur la base MUST.

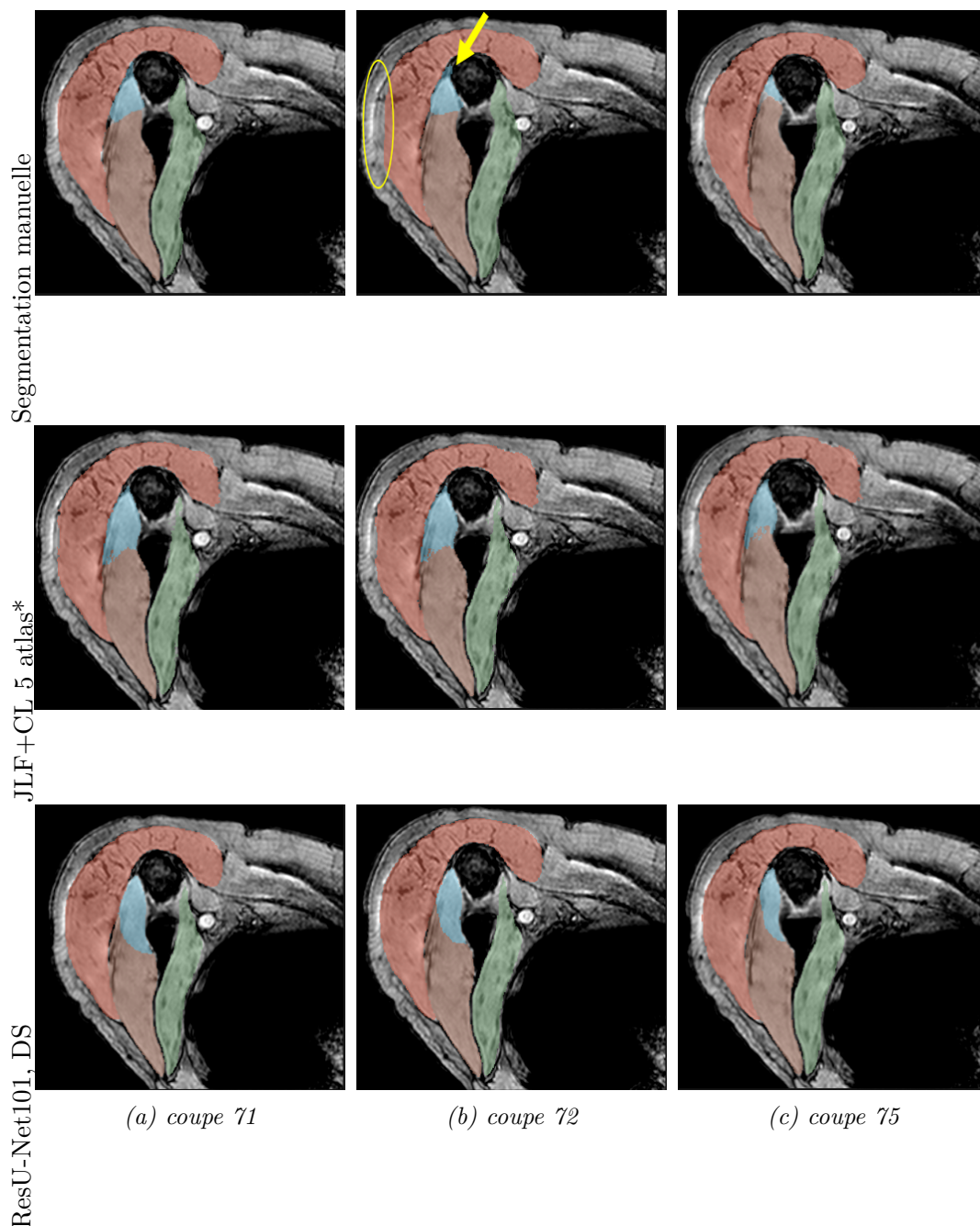


FIGURE 14.11 – Segmentations manuelles de 3 coupes de la base de données Epaule et les résultats de segmentations automatiques par JLF+CL avec les 5 meilleurs atlas, et par ResU-Net101 et supervision profonde.

Comme on peut l’observer sur la figure 14.12, les résultats d’U-Net peuvent être très bons comme pour le coureur ARS-4026 où le DSC est proche de 0,94, à mauvais comme pour le coureur CAL-4223. On note aussi la création d’erreurs anatomiques (non connexité des régions par exemples) comme on peut l’observer sur le coureur ALB-2725.

Pour corriger ces erreurs, nous pouvons envisager des post-traitements ad-hoc, comme cela a été évoqué dans la section précédente, notamment via nnU-Net. Nous avons préféré nous appuyer sur une approche générale qui apprendrait les erreurs systématiques et les corrigerait, typiquement : l’approche de d’apprentissage des corrections CL (*corrective*

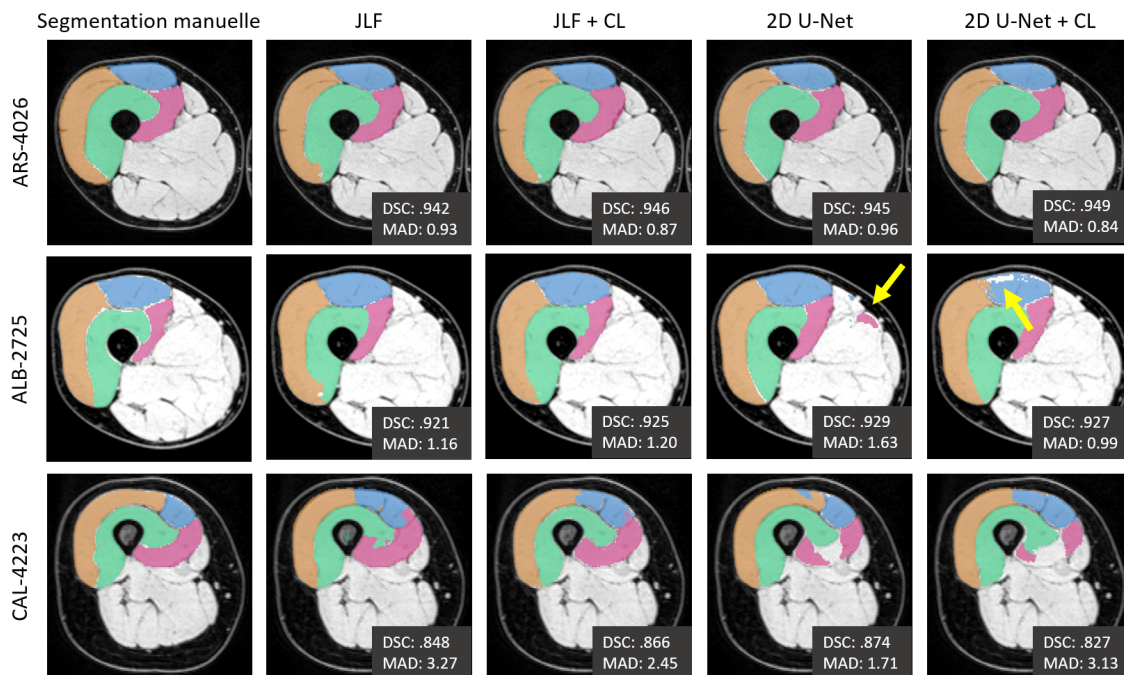


FIGURE 14.12 – Segmentations de trois coureurs MUST avec les approches JLF+CL, U-Net et U-Net+CL. On remarquera les erreurs globales réalisées par les différentes méthodes sur les coureurs ALB-2725 et CAL-4223 ainsi que la capacité du CL à réduire l'erreur grossière de U-Net sur le coureur ALB-2725 mais créant aussi des erreurs aux frontières (flèches jaunes).

learning) proposée par (WANG et al., 2011) et utilisée précédemment.

Méthode proposée

En remarquant que l'approche JLF produit une carte de probabilité de segmentation, comme U-Net, il est alors possible de substituer les longues étapes JLF par des U-Net dans l'approche de (WANG et al., 2013b). Nous avons proposé dans (NGUYEN et al., 2019b) de modifier la méthode JLF+CL en U-Net+CL comme illustrée dans la figure 14.13. Pour cette approche, un réseau U-Net doit être entraîné pour chaque atlas. Puis l'étape de *Corrective Learning* apprend les erreurs systématiques commises par les réseaux U-Net. Enfin, un dernier réseau U-Net est entraîné en utilisant tous les atlas. Ce dernier sert à la phase de test : il permet d'obtenir une segmentation d'une nouvelle données, et ce sera cette segmentation qui sera corrigée par l'approche CL.

Résultats

La dernière ligne de la table 14.9 et la dernière ligne de la figure 14.12 donnent les résultats obtenus par l'approche U-Net+CL.

Discussion

Comparés aux approches multi-atlas JLF, les résultats basés sur U-Net sont assez similaires, sauf pour la distance de Hausdorff. L'intérêt des réseaux est le temps de calcul bien plus raisonnable : environ 2 heures pour un entraînement et moins d'une minute pour une inférence, plus de 50 heures pour l'approche JLF avec 6 atlas. Les résultats de l'approche U-Net+CL montrent la capacité à réduire les erreurs grossières (diminution significative de la distance de Hausdorff par rapport à l'approche U-Net). Néanmoins, les autres métriques sont toujours satisfaisantes mais globalement plus mauvaises que celles obtenues avec le réseau U-Net seul.

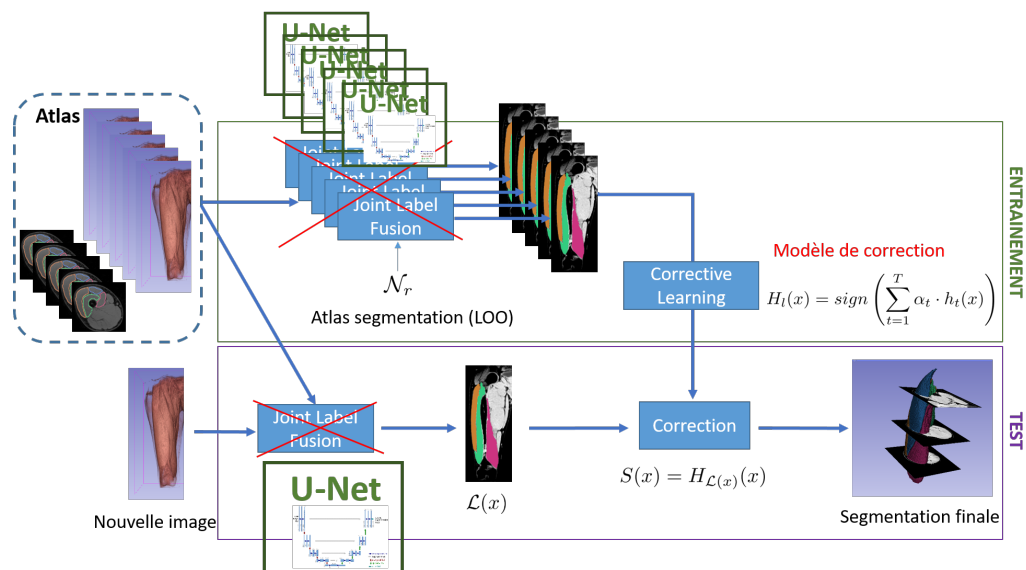


FIGURE 14.13 – Utilisation de réseaux U-Net pour remplacer les étapes JLF de l’approche de (WANG et al., 2013b).

Méthode	DSC	DSCw	HD (mm)	MAD (mm)	VS
JLF	.914 [.848, .945]	.917 [.866, .945]	34.77 [18.67, 48.59]	1.65 [0.82, 3.27]	.080 [.028, .141]
JLF + CL	.921 [.866, .946]	.923 [.879, .945]	33.44 [20.73, 40.40]	1.46 [0.88, 2.74]	.056 [.024, .104]
U-Net	.921 [.874, .945]	.923 [.878, .944]	85.32 [52.97, 132.57]	1.46 [0.95, 1.99]	.064 [.034, .095]
U-Net + CL	.915 [.827, .949]	.917 [.835, .943]	51.71 [18.04, 120.97]	1.55 [0.84, 3.13]	.062 [.025, .137]

TABLE 14.9 – Evaluation quantitative des segmentations JLF, JLF+CL, U-Net et U-Net+CL sur la base de données MUST sur les 7 atlas en utilisant une validation croisée un contre tous. Les valeurs entre crochets correspondent respectivement aux valeurs minimum et maximum obtenues sur les 7 tests.

On remarque aussi que pour les coureurs difficiles à segmenter (exemple du coureur CAL-4223 de la figure 14.12) les approches d’apprentissage profond ne permettent pas d’améliorer les résultats, et que l’étape CL ne parvient pas à les corriger. Cela conduit à l’hypothèse que la morphologie de ce type de coureur n’est pas apprise lors de la phase d’apprentissage d’U-Net ou de CL. Puisque sur ces coureurs les approches JLF et JLF+CL ne produisent pas non plus de bonnes segmentations, on peut conclure que la morphologie de ces quelques coureurs est très spécifique et n’est pas assez représentée dans la base de données.

Il serait pertinent de pouvoir caractériser ces morphologies et représenter la variabilité présente dans la base de données afin d’en tenir compte pour l’apprentissage. On pourrait ainsi optimiser cette base de données en incluant des morphologies pertinentes, ou mieux appréhender les limites de cette base pour les segmentations de certaines morphologies. Nous développerons ces idées dans le chapitre suivant (chapitre 15).

14.5 Conclusion

Dans ce chapitre, nous avons exposé nos travaux de segmentation supervisée d’images ultrasonores et IRM avec les approches multi-atlas JLF+CL et de *deep learning* à base d’architecture U-Net. Nous avons aussi présenté une contribution mélangeant les deux approches U-Net+CL. Concernant les approches multi-atlas, elles permettent d’obtenir,

avec un faible nombre d'atlas, des résultats de segmentation des muscles très intéressants sur les bases MUST et Epauls (IRM).

En segmentation echo-cardiographique 2D, U-Net permet d'obtenir des résultats bien meilleurs que ceux obtenus jusqu'à maintenant par des approches classiques de segmentation. Pour les segmentations de muscles en IRM de la base Epauls, les résultats obtenus sont de qualité comparables à ceux produits par les approches multi-atlas, mais avec des temps de calcul bien plus acceptables. Toujours sur cette base de données, nous avons constaté que les performances des réseaux et fonctions de coût ne se généralisent pas nécessairement d'un problème à un autre, conduisant pour l'instant à systématiser l'expérimentation d'architectures ainsi que leurs optimisations. Ce travail long et fastidieux fait l'objet de nombreuses recherches et une méthode nnU-Net s'impose pour s'affranchir d'une étude exhaustive. Néanmoins, les pré- et post- traitements sont des étapes permettant d'améliorer très significativement les segmentations, souvent avec moins d'effort que l'optimisation d'architectures complexes.

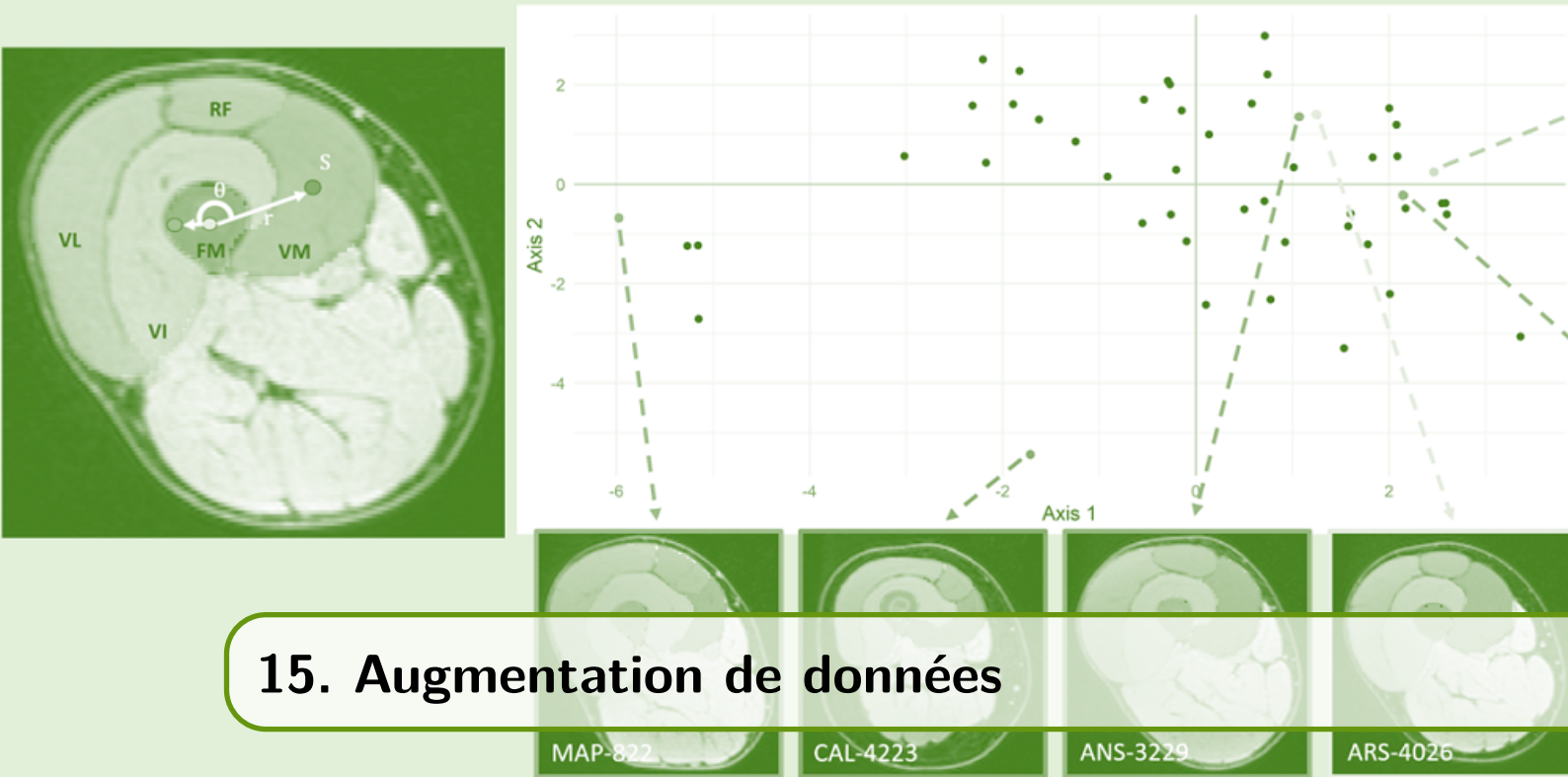
Dans ce chapitre, nous avons aussi exposé une approche basée sur le *Corrective Learning* visant à réduire les erreurs importantes de segmentation (quantifiées par la distance de Hausdorff) souvent créées par le réseau U-Net 2D, et ceci de manière supervisée et généralisable, c'est à dire sans la création de post-traitements ad-hoc. Notre contribution U-Net+CL réduit bien ces erreurs mais celles-ci restent d'une part, encore presque deux fois supérieures à celles produites par l'approche JLF+CL et d'autre part, elle réduit sensiblement la qualité de segmentation globale (diminution du score de Dice).

De manière plus globale, nos observations sur les erreurs de segmentations sont de deux types :

- soient des petites erreurs qui peuvent être rapprochées à des difficultés locales de segmentation (contours peu marqués par exemple), voir à des incohérences dans les annotations (notamment 3D) ou encore à des artefacts dans l'image,
- soient des erreurs plus globales, qui s'expliquent par une mauvaise représentativité lors de l'apprentissage de la morphologie ou spécificité anatomique et pour lesquelles le réseau n'est pas capable de bien généraliser.

En tout cas, la qualité des résultats produits par les approches d'apprentissage profond a changé notre manière d'appréhender la problématique de segmentation d'images. L'enjeu principal est d'abord la base d'images annotées.

Le chapitre suivant traite du problème de l'augmentation de données et des stratégies de sélection des atlas pour améliorer les résultats et diminuer le temps nécessaire aux approches multi-atlas.



15. Augmentation de données

15.1 Introduction

Pour les applications précédentes, nous avons toujours eu le problème de la sélection d'atlas pour l'approche JLF+CL et de l'augmentation de données pour les approches U-Net. Dans le premier cas, nous avons dû tester les différentes combinaisons d'atlas et nous avons constaté une grande variabilité des résultats quand le nombre d'atlas est faible. Pour U-Net, malgré le fait de traiter le problème de segmentation en 2D, donc disposer potentiellement de beaucoup d'images, pour la base MUST il a fallu augmenter le nombre d'images annotées car les coupes d'un même patient sont fortement corrélées entre elles ce qui conduit à un sur-apprentissage néfaste à la généralisation du réseau à d'autres images.

Pour illustrer sur la base de données MUST, sans augmentation de données, et à partir des 6 atlas disponibles, U-Net permet d'obtenir des scores de DSC de 0,7 - peu satisfaisants comme le montre la figure 15.1.

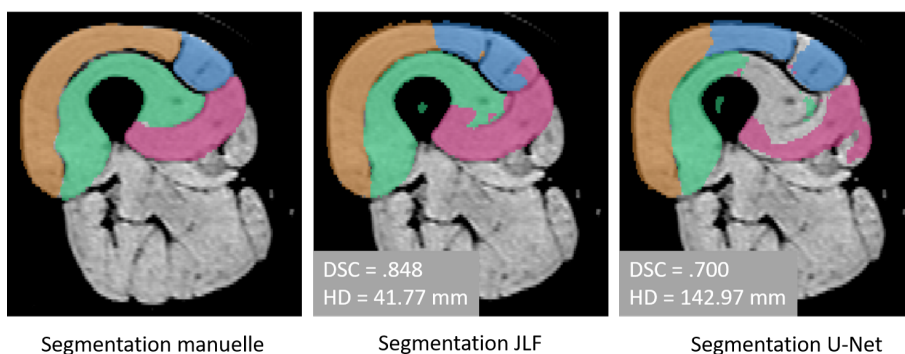


FIGURE 15.1 – Comparaison de segmentations d'un examen de la base MUST avec les approches JLF et U-Net sans augmentation de données.

Ce chapitre traite des méthodes que l'on a expérimentées et proposées pour les approches de recalage d'atlas et d'apprentissage profond pour les bases MUST et Epaule.

Cette problématique est bien couverte dans la littérature, surtout depuis l'explosion "deep learning" qui requiert pour l'entraînement un *grand* nombre de données. Les travaux de (SHORTEN et al., 2019) dresse une liste des stratégies d'augmentation de données pour plusieurs problématiques d'apprentissage profond sur des images naturelles. Le versant médical a été traité par (CHLAP et al., 2021).

Cependant, il existe aussi des approches essayant de réaliser un entraînement avec un faible nombre d'exemple, voir un seul. On parle de "*few-shot*" et de "*one-shot*". (ZHAO et al., 2019) décrit de telles approches pour la segmentation d'images IRM du cerveau qui reposent en fait sur des stratégies d'augmentation de données à partir d'un seul atlas (le plus proche de la moyenne) pour une approche basée sur le recalage et une basée sur un réseau U-Net spécifique. Dans leur travail, sans atteindre les performances où les 101 annotations sont utilisées, l'augmentation de données est cruciale et le réseau de neurones permet d'obtenir les meilleurs performances.

Ainsi, quand le nombre de patients annotés dont on dispose est petit, il est capital d'augmenter la diversité des images pour l'apprentissage. Nous présentons ici nos stratégies d'augmentation de données d'abord basées sur des recalages aléatoires, puis une contribution optimisant la sélection des atlas ou des images à augmenter en fonction de la morphologie à segmenter.

15.2 Augmentations d'images par recalages

Afin de disposer d'assez d'images pour entraîner convenablement un réseau et améliorer nos résultats de segmentation (NGUYEN et al., 2018a ; NGUYEN et al., 2018b), nous avons utilisé dans un premier temps deux stratégies d'augmentation de données :

- des déformations aléatoires des atlas,
- des recalages d'atlas sur des coureurs non-segmentés afin de bénéficier de leur morphologie (NGUYEN et al., 2019b)

Dans ce dernier cas, un recalage entre un atlas (où l'on dispose du couple IRM et segmentation manuelle) et un examen IRM d'un coureur non segmenté, est réalisé sur les images IRM. Le champs de déformation obtenu est ensuite appliqué sur l'IRM et sur la segmentation de l'atlas afin de produire une nouvelle image, et sa segmentation, qui ressemble à celle du coureur non segmenté. La figure 15.2 illustre des données obtenues par ces deux stratégies.

Les champs de déformations aléatoires doivent être modérés pour rester réalistes (en terme d'anatomie et de texture d'images) ce qui conduit à des changements subtils dans les images donc une faible variabilité des données d'entrée. Il s'agira de ne pas trop générer d'images avec cette méthode au risque de ne pas observer d'amélioration des segmentations sur une base de test (sur-apprentissage).

L'approche de recalage déformable permet une plus grande dispersion des morphologies en restant cohérent par rapport à la variabilité de la base de données car elle s'appuie sur les données de la base. Puisque l'on dispose de 7 atlas et 50 coureurs, il est possible de générer un grand nombre d'images afin d'entraîner le réseau U-Net.

La table 15.1 donne les résultats obtenus avec les approches JLF, JLF+CL et U-Net utilisant les augmentations de données proposées ici. Les tests que nous avons réalisés sont basés sur 42 volumes obtenus en combinant les 2 stratégies d'augmentation de données. Ces stratégies permettent de passer de 7 volumes annotés à 77 (5 déformations aléatoires de chaque atlas et 5 recalages sur les coureurs non segmentés). Systématiquement, les résultats ont été obtenus sur chacun des 7 atlas segmentés avec l'approche de validation croisée de un contre tous, l'examen testé étant exclu des stratégies d'augmentation de données.

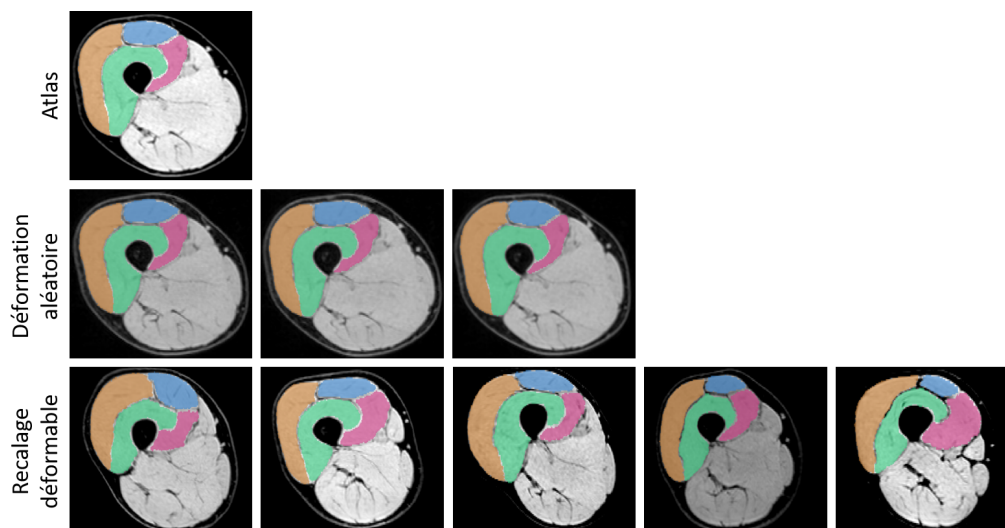


FIGURE 15.2 – Stratégies d’augmentation de données pour la base de données MUST. L’atlas est donné sur la première ligne. La ligne du milieu donne des trois exemples de déformations aléatoires. La dernière ligne illustre les recalages déformables sur cinq coureurs non segmentés.

Méthode	DA	DSC	HD (mm)	MAD (mm)	VS
<i>JLF</i>	None	.914 [.848, .945]	34.77 [18.67, 48.59]	1.65 [0.82, 3.27]	.080 [.028, .141]
<i>JLF + CL</i>	None	.921 [.866, .946]	33.44 [20.73, 40.40]	1.46 [0.88, 2.74]	.056 [.024, .104]
<i>U-Net</i>	D	.892 [.774, .946]	98.73 [79.66, 141.49]	2.22 [0.87, 3.78]	.097 [.047, .172]
<i>U-Net</i>	R	.915 [.842, .947]	79.49 [32.97, 141.62]	1.77 [0.82, 3.72]	.095 [.043, .190]
<i>U-Net</i>	D + R	.921 [.874, .945]	85.32 [52.97, 132.57]	1.46 [0.95, 1.99]	.064 [.034, .095]
<i>U-Net + CL</i>	D + R	.917 [.842, .947]	48.83 [13.43, 89.84]	1.48 [0.84, 2.70]	.061 [.022, .140]

TABLE 15.1 – Evaluation quantitative des stratégies d’augmentation d’images pour la segmentation de MUST. *U-Net* est testé pour les différentes stratégies d’augmentation de données (DA) : D - Déformations aléatoires, R - recalages, D + R - déformations aléatoires et recalages. Les résultats sont la moyenne et les valeurs minimum et maximum pour chacun des 7 atlas (en validation croisée leave-one-out).

Grâce à ces stratégies d’augmentation de données, on obtient une nette amélioration de la qualité des segmentations obtenues avec l’approche *U-Net*. Le gain de temps est à prendre en compte : avec un GPU, l’apprentissage d’*U-Net* dure 2h et le temps d’inférence d’un volume 3D est de l’ordre de la minute.

Nous avons aussi remarqué que le choix des images pour l’augmentation de données influence les performances à l’inférence. Notre contribution suivante traite de la sélection des images pour l’augmentation de données.

15.3 Sélection d’atlas basée sur la morphologie

Dans le cadre de ses travaux de thèse, Hoai-Thu Nguyen a proposé une sélection des images et atlas à utiliser pour l’augmentation de données. Cette sélection s’effectue sur la base d’un critère original de morphologie.

En effet, au chapitre 14 nous avons observé qu’une grande partie des erreurs de segmentation provenait d’anatomies très particulières, faiblement représentées dans la base d’apprentissage. Notre hypothèse est que de prendre en considération cette caractéristique

morphologique permettrait de définir des stratégies pour améliorer les approches multi-atlas et d'apprentissage profond.

Afin de tester cette hypothèse, nous caractérisons les morphologies musculaires de manière quantitative, avec une description simple basée sur une segmentation d'une coupe des muscles. La position de la coupe considérée est définie de manière anatomique afin d'être cohérente pour toute la base de données. On désigne cette coupe par "coupe centrale" pour la base MUST.

On prend ici l'exemple de la base MUST, où 5 régions (les 4 quadriceps et le fémur) servent à la caractérisation de la morphologie. Pour chacune des régions, la surface S et le barycentre sont calculés. Le barycentre du fémur est utilisé comme origine et on considère les distances des barycentres des régions à celui du fémur comme le montre la figure 15.3. La mesure est ainsi invariante en translation. Pour la rendre invariante en rotation, on considère chaque angle entre le barycentre d'une région et le vecteur formé par les barycentres du fémur et de la région du Vastus Intermedius (VI).

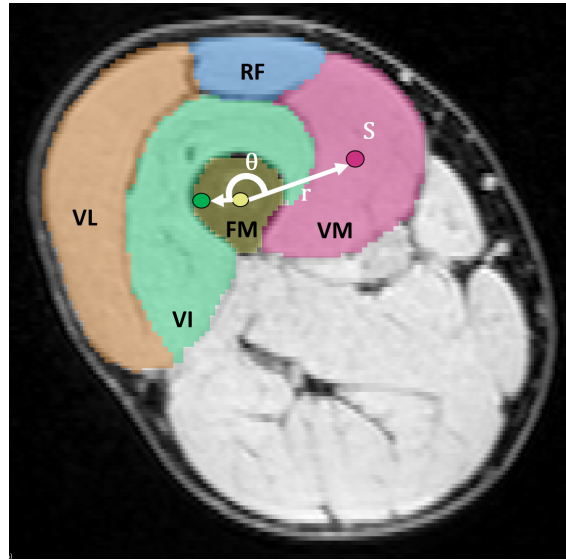


FIGURE 15.3 – Illustration sur un sujet de MUST de la caractérisation morphologique. Les surfaces et barycentres des régions sont calculés. Le barycentre du fémur sert d'origine et l'angle avec le barycentre de VI, de référence pour le calcul des angles. Sont représentés ici l'angle et le module pour la région VM.

On obtient un vecteur f de 12 dimensions :

$$f_X = [S_X^{VI}, 0, r_X^{VI}, S_X^{VM}, \theta_X^{VM}, r_X^{VM}, S_X^{VL}, \theta_X^{VL}, r_X^{VL}, S_X^{RF}, \theta_X^{RF}, r_X^{RF}] \quad (15.1)$$

La distance morphologique d_{AB} entre deux images A et B se calcule avec la distance Euclidienne : $d_{AB} = \|f_A^* - f_B^*\|_2$

où f_X^* est le vecteur des caractéristiques morphologiques de X , $f_X^* \in \mathbb{R}^{11}$ centré et réduit par les moments calculés sur l'ensemble de la base de données. On retire aussi l'angle θ_X^{VI} qui est, par définition, égal à 0.

La figure 15.4 montre la projection 2D par Analyse en Composantes Principales (ACP) des jambes droites de 48 sujets de la base MUST avec quelques exemples d'images et segmentations de la coupe centrale. Ces 2 dimensions représentent 70% de la variance du jeu de données. Cette méthode -simple- de caractérisation des morphologies reflète assez bien les observations, notamment pour les coureurs en bleu et vert qui se retrouvent assez éloignés des autres qui visuellement se ressemblaient. L'un de ces deux coureurs (CAL-4223)

est le coureur le plus difficile à segmenter par les approches testées (voir la figure 14.12 pour la comparaison JFL, JLF+CL, U-Net et U-Net+CL).

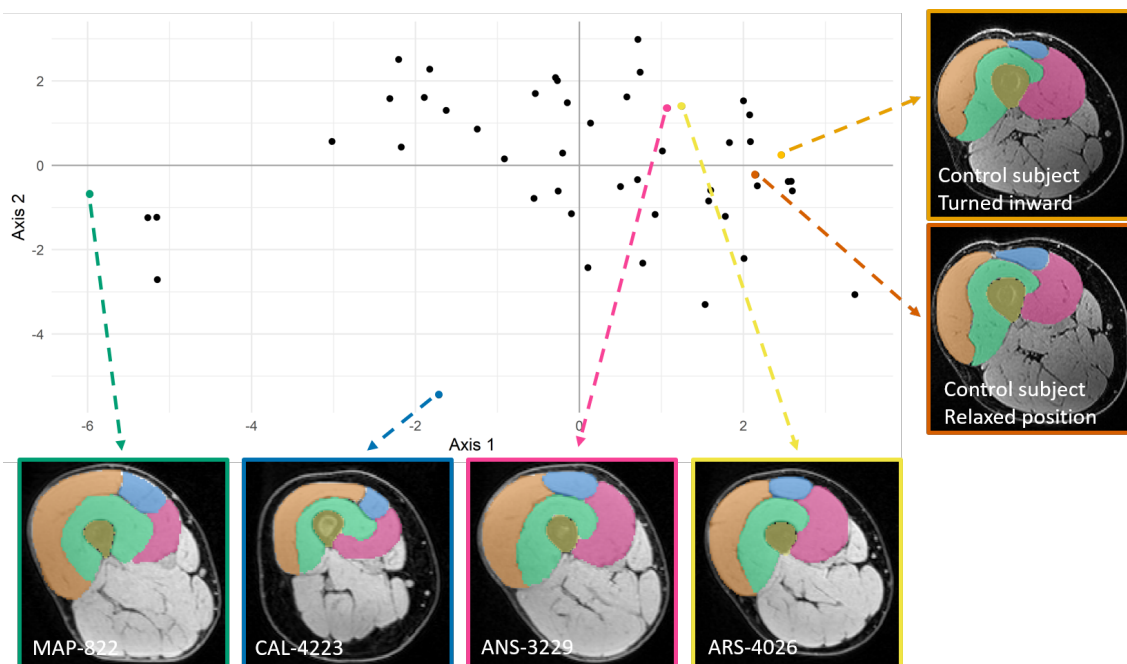


FIGURE 15.4 – Projection bi-dimensionnelle par ACP des vecteurs morphologiques des jambes droites des coureurs de MUST avec 6 exemples d'IRM pondérées T1 de la coupe centrale et sa segmentation ayant permis d'extraire les caractéristiques morphologiques. En vert, bleu, rose et jaune on peut observer les correspondances entre morphologie et proximité des vecteurs projetés en 2D. En orangé, les deux positions d'un même sujet de contrôle pour apprécier la rotation biomécanique.

Nous avons défini le vecteur de caractéristiques morphologiques invariant en translation et rotation. Nous avons évalué la sensibilité de la distance pour une rotation biomécanique qui n'est pas équivalente à une rotation de l'image. Deux images de la jambe droite d'un même sujet ont été acquises dans deux positions : normale (relâchée) et tournée vers l'intérieur (15.4, à droite). La distance (calculée sur les 11 dimensions, pas sur la projection ACP bi-dimensionnelle) entre ces deux positions est la plus faible observée pour la base de données.

Ainsi, avec ces caractéristiques et distances morphologiques, nous disposons d'une mesure permettant de quantifier la ressemblance anatomique des coureurs. Grâce à cela, nous pouvons définir des stratégies de sélection d'atlas afin d'améliorer les méthodes de segmentation étudiées.

15.3.1 Sélection d'atlas pour la segmentation multi-atlas

Les deux principaux inconvénients que nous avons observés pour les approches JLF et JLF+CL sont : le temps de calcul nécessaire à l'utilisation de 6 atlas (50h) et les performances réduites pour des sujets dont la morphologie diffère de celle de la plupart des atlas. Nous proposons d'optimiser la méthode en effectuant une sélection d'atlas basée sur la morphologie pour réduire le temps de calcul tout en améliorant la qualité de la segmentation.

Pour la segmentation d'un nouveau volume, il faut délimiter approximativement les quadriceps sur la coupe centrale afin de pouvoir déterminer les atlas les plus proches en utilisant la distance morphologique précédemment présentée.

Expérimentations

Avec 7 segmentations manuelles, nous avons adopté le schéma LOO (Leave-One-Out) pour évaluer la méthode de segmentation. Chaque coureur, parmi les 7 avec des segmentations manuelles, a servi de test alors que les 6 autres ont servi d'atlas. Sur la base des caractéristiques morphologiques, nous trions les atlas du plus proche au plus éloigné du volume de test et choisissons d'utiliser 3 à 5 atlas les plus proches du sujet test.

Résultats & Discussion

Comme nous l'avons vu 14.2, le temps de calcul de la fusion conjointe d'étiquettes (JLF) augmente rapidement avec le nombre d'atlas, d'environ 24h avec 3 atlas à environ 50h avec 6 atlas.

La figure 15.5 montre les évolutions de DSC de la méthode de (WANG et al., 2013b) avec une sélection d'atlas aléatoire ou basée sur la morphologie.

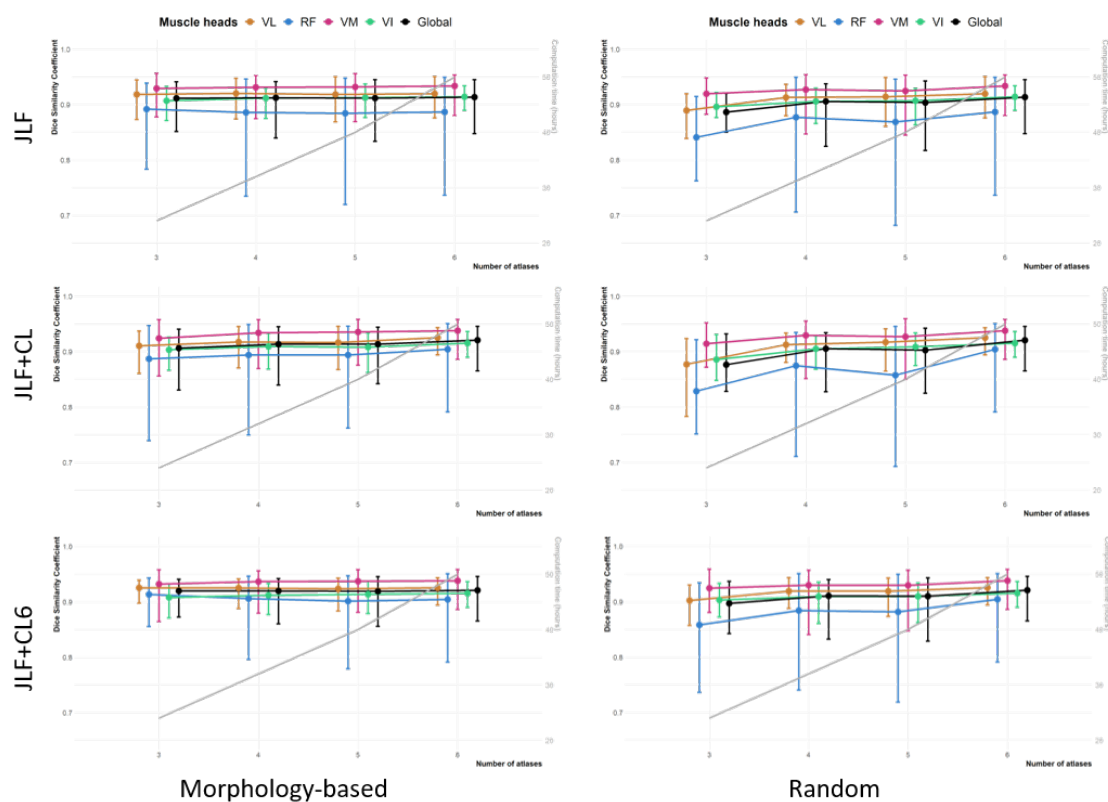


FIGURE 15.5 – Résultats de segmentation des approches JLF et JLF+CL (WANG et al., 2013b) avec et sans sélection d'atlas basée sur la morphologie. Les résultats JLF+CL6 correspondent à l'approche JLF+CL ou les 6 atlas disponibles ont servi à l'apprentissage de l'algorithme de correction (quel que soient les atlas utilisés pour le JLF). Chaque valeur par muscle et pour tous les muscles est donnée en valeur moyenne (sur 7 réalisations) et la barre verticale donne le score minimal et maximal.

L'observation principale est que la sélection basée sur la morphologie permet d'obtenir avec 3 atlas des résultats très proches de l'utilisation de 6 atlas (voir tableau 15.2). On observe bien l'apport de la sélection morphologique sur le tirage aléatoire lors de l'utilisation d'un petit nombre d'atlas : de 0,877 de DSC en moyenne avec 3 atlas choisis aléatoirement à 0,912 avec les 3 atlas les plus proches morphologiquement. Cette sélection morphologique permet aussi à l'apprentissage correctif (CL) d'être plus pertinent, ce qui a confirmé

l'observation faite dans des études antérieures (NGUYEN et al., 2019b; NGUYEN et al., 2019c) que le CL ne convient pas pour corriger les erreurs importantes causées par la variation morphologique du quadriceps.

Parallèlement, étant donné que l'augmentation du nombre d'atlas pour le CL n'augmente que le temps d'apprentissage (apprentissage du modèle de correction) mais pas le temps d'inférence, nous avons appliqué le modèle correcteur appris sur 6 atlas (CL6) sur les résultats de JLF. Nous avons obtenu un DSC moyen similaire à JLF + CL avec 6 atlas. De plus, pour la segmentation d'un nouveau volume, corriger la segmentation automatique avec un modèle pré-entraîné sur tous les atlas disponibles est plus pratique que de ré-entraîner un modèle correctif basé sur les 3 atlas les plus proches.

	Aléatoire			Basée morphologie		
	JLF	JLF + CL	JLF + CL6	JLF	JLF + CL	JLF + CL6
<i>DSC</i>						
3 atlas	.887 [.850, .921]	.877 [.828, .932]	.897 [.842, .937]	.912 [.852, .942]	.906 [.831, .941]	.920 [.873, .941]
4 atlas	.906 [.825, .938]	.906 [.828, .934]	.911 [.833, .940]	.912 [.840, .942]	.914 [.840, .945]	.920 [.860, .942]
5 atlas	.904 [.817, .943]	.903 [.825, .943]	.910 [.829, .943]	.912 [.834, .945]	.914 [.842, .944]	.919 [.856, .946]
6 atlas	.914 [.848, .945]	.921 [.866, .946]				
<i>MAD (mm)</i>						
3 atlas	2.15 [1.32, 3.46]	2.50 [1.09, 3.75]	1.94 [1.01, 3.29]	1.62 [0.97, 2.90]	1.72 [1.03, 3.13]	1.43 [0.99, 2.66]
4 atlas	1.85 [1.14, 3.87]	1.91 [1.09, 3.82]	1.79 [1.06, 3.09]	1.67 [0.95, 3.21]	1.61 [0.88, 3.00]	1.48 [0.99, 2.69]
5 atlas	1.95 [0.90, 4.36]	1.93 [0.91, 4.05]	1.75 [0.91, 3.55]	1.72 [0.88, 3.57]	1.58 [0.90, 2.97]	1.53 [0.87, 3.01]
6 atlas	1.65 [0.84, 3.27]	1.45 [0.88, 2.74]				
<i>VS</i>						
3 atlas	.104 [.069, .143]	.117 [.046, .238]	.085 [.044, .169]	.076 [.030, .138]	.079 [.035, .121]	.064 [.037, .088]
4 atlas	.080 [.045, .152]	.079 [.051, .121]	.062 [.019, .090]	.079 [.032, .151]	.076 [.036, .137]	.060 [.033, .091]
5 atlas	.087 [.025, .175]	.082 [.031, .148]	.066 [.028, .136]	.083 [.025, .164]	.077 [.029, .155]	.062 [.030, .118]
6 atlas	.080 [.028, .141]	.056 [.024, .104]				

TABLE 15.2 – *Evaluations quantitatives des segmentations JLF, JLF+CL et JLF+CL6 avec un choix aléatoire ou une sélection basée sur la morphologie des atlas. Chaque mesure est donnée en moyenne, minimum et maximum pour les 7 atlas disponibles. Les meilleures valeurs moyennes ou la plus petite dynamique est indiquée en gras pour chaque métrique.*

Cette stratégie de sélection des 3 atlas morphologiquement les plus proches réduit de moitié le temps d'exécution des approches JLF et JLF+CL avec 6 atlas tout en conservant, voire en améliorant, la qualité de la segmentation. Nous présentons ensuite notre stratégie basée sur la sélection d'atlas pour l'approche U-Net faiblement supervisée.

15.3.2 Augmentation de données basée morphologie pour U-Net

Comme évoqué dans l'introduction de ce chapitre (15.1) et illustré sur la figure 15.1, entraîner U-Net en 2D avec seulement 6 atlas, malgré les centaines de coupes que cela représente, ne permet pas de réaliser un apprentissage généralisable à de nouvelles données car les coupes sont très similaires entre elles. L'augmentation de données est nécessaire. Nous avons proposé d'utiliser les approches par déformations aléatoires ou de recalage sur des coupeurs choisis aléatoirement (voir le paragraphe 15.2). Avec les caractéristiques morphologiques, nous pouvons définir des stratégies plus spécifiques soit pour apprendre uniformément la diversité morphologique présente dans la base de données, soit pour apprendre spécifiquement une morphologie donnée afin d'obtenir de meilleurs résultats. Ces stratégies ont été proposées dans les travaux de thèse de Hoai-Thu Nguyen.

Caractéristiques morphologiques pour apprendre la diversité

A partir de distances morphologiques ordonnées dans l'ordre croissant entre chacun des 7 volumes annotés et les 41 volumes non annotés, nous créons, pour chacun des volumes

annotés, 5 groupes de 6 images, du plus similaire au moins similaire, après avoir exclu les 11 volumes les plus similaires. Ces 11 volumes sont jugés trop proches et donc conduiraient à une sur-représentation de la morphologie, c'est à dire un possible sur-apprentissage de celle-ci. Parmi chacun des 5 groupes, on sélectionne pour l'apprentissage un volume dans les groupes un, trois et cinq, puis pour la validation, un volume dans les groupes deux et quatre. Chaque volume sélectionné sert alors pour le recalage avec le volume annoté pour l'augmentation de données. Il est alors possible d'entraîner un réseau avec une bonne représentation de la diversité morphologique. Cette stratégie est nommée *diversity*.

Les résultats sur la base de données MUST sont présentés dans la table 15.3 toujours en utilisant une validation croisée sur le schéma *leave-one-out*. Chaque test inclut les 6 volumes annotés (les 7-1 atlas) avec, pour l'entraînement, 3 déformations aléatoires et 3 recalages sur les volumes sélectionnés (groupes 1, 3 et 5). La validation s'effectue sur 12 volumes (groupes 2 et 4). Cela représente respectivement environ 2700 et 770 coupes. Afin de comparer les approches d'augmentation de données, nous avons entraîné 5 fois le même réseau U-Net avec approximativement le même nombre de coupes en utilisant l'approche de déformations aléatoires (section 15.2).

Les deux approches donnent des résultats en faveur de la sélection basée sur la diversité morphologique. On note que les intervalles pour chacune des métriques sont plus petits avec la morphologique.

Cependant, si on utilise la totalité des images, comme dans la table 15.1, les performances sont alors comparables.

Strategy	DSC	HD (mm)	MAD (mm)	VS
<i>Random</i>	.918 [.822, .954]	76.19 [37.26, 139.96]	1.51 [0.84, 3.60]	.076 [.036, .168]
<i>Diversity</i>	.920 [.849, .951]	77.34 [47.62, 105.82]	1.37 [0.76, 2.68]	.084 [.042, .151]

TABLE 15.3 – Evaluation quantitative de l'augmentation de données aléatoires et par sélection morphologique pour U-Net (*Diversity*). Les résultats sont donnés sous forme moyenne [minimum, maximum] sur les 7 volumes annotés en utilisant la stratégie de *leave-one-out* pour la base de données MUST.

La plus petite valeur de DSC est encore obtenue pour le coureur CAL-4223. Il manque toujours de l'information pour obtenir des performances acceptables pour ce coureur.

Entraînement morphologie-spécifique pour UNet

Afin d'accroître l'information pour une morphologie spécifique, nous proposons deux stratégies d'entraînement du réseau.

La première, nommée *target-driven*, consiste à créer un jeu d'entraînement très proche de la morphologie ciblée. Typiquement, 10 volumes sont créés par rapport à la morphologie ciblée :

- chacun des deux plus proches atlas sont recalés sur les 2 volumes non-annotés les plus proches de la morphologie ciblée,
- le volume ciblé est déformé aléatoirement 4 fois et chacun des 2 atlas les plus proches sont recalés sur ces déformations.

Le schéma de construction du jeu d'entraînement du *target-driven* est illustré sur la figure 15.6. Le jeu de validation est constitué des 2 atlas les plus proches recalés sur le volume ciblé.

La seconde approche est une approche d'affinage d'entraînement, nommée *fine tuned U-Net*. A partir du réseau U-Net entraîné avec la stratégie basée sur la diversité (voir le paragraphe 15.3.2), on continue l'apprentissage sur les 2 atlas les plus proches du volume ciblé recalés

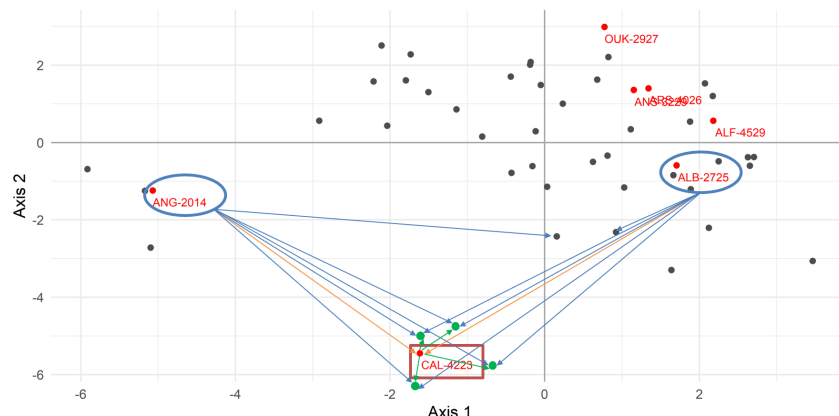


FIGURE 15.6 – Schéma de construction du jeu d'apprentissage pour la stratégie target-driven de U-Net illustré pour le coureur CAL-4223. Les deux atlas les plus proches sont entourés en bleu. Les 4 déformations aléatoires de CAL-4223 sont les points en vert. Les flèches bleues indiquent les recalages réalisés à partir des 2 atlas les plus proches pour le jeu d'entraînement. Les flèches oranges, les recalages pour le jeu de validation.

sur celui-ci. Le jeu de validation est constitué du troisième atlas le plus proche recalé sur le volume ciblé. Nous avons observé que l'entraînement est optimisé avant la cinquième époques et que ce *fine tuning* ne nécessite que 5 à 10 minutes.

La table 15.4 donne les résultats obtenus par les différentes stratégies.

	ALB-2725	ALF-4529	ANG-2014	ANS-3229	ARS-4026	CAL-4223	OUK-2927	Moyenne
<i>DSC</i>								
JLF+CL6	.917	.941	.900	.934	.941	.872	.933	.920
Diversity UNet	.927	.893	.924	.951	.949	.849	.943	.920
Target-driven UNet	.913	.920	.902	.937	.931	.880	.934	.916
Fine-tuned UNet	.929	.926	.925	.948	.947	.915	.945	.934
<i>HD (mm)</i>								
JLF+CL6	44.80	18.03	36.47	23.89	34.74	28.37	43.75	32.86
Diversity UNet	100.00	63.33	96.01	70.55	47.62	105.82	58.02	77.34
Target-driven UNet	103.96	137.11	96.61	120.50	94.55	132.27	77.77	108.40
Fine-tuned UNet	42.48	102.89	100.64	56.60	46.81	57.77	58.73	66.56
<i>MAD (mm)</i>								
JLF+CL6	1.33	0.99	2.66	1.13	1.01	1.72	1.19	1.43
Diversity UNet	1.23	1.91	1.01	0.76	0.85	2.68	1.11	1.37
Target-driven UNet	2.79	2.29	1.88	1.00	1.18	3.05	2.18	2.05
Fine-tuned UNet	1.14	1.20	1.08	0.81	0.96	0.89	1.00	1.01
<i>VS</i>								
JLF+CL6	.088	.037	.078	.049	.045	.088	.065	.064
Diversity UNet	.106	.129	.073	.042	.045	.151	.042	.084
Target-driven UNet	.115	.071	.121	.041	.055	.070	.020	.070
Fine-tuned UNet	.095	.103	.088	.044	.035	.064	.026	.065

TABLE 15.4 – Evaluation quantitative des stratégies morphologies spécifiques d'entraînement d'U-Net pour chaque atlas de MUST. Pour comparaison sont donnés : les résultats du JLF+CL6 basé sur les 3 atlas les plus proches, le U-Net entraîné sur un jeu sélectionné par les stratégies de diversité (15.3.2), target-driven et fine-tuned (paragraphe 15.3.2). Les valeurs en gras sont les meilleurs scores obtenus pour chaque coureur.

On rappelle aussi le nombre de volumes et les temps nécessaires à ces approches dans le tableau 15.5 qui montre bien le gain en temps réalisé par l'approche *fine-tuned*.

Méthodes	Nombres de volumes	Temps d'inférence
JLF+CL6	3 (6)	24 heures
Diversity U-Net	42	1 minutes
Target-driven U-Net	10	30 minutes
Fine-tuned U-Net	2 (42)	10 minutes

TABLE 15.5 – Rappel des grandeurs d'inférence pour les méthodes basées sur la morphologie. Les valeurs entre parenthèses précisent qu'une partie de la méthode est entraînée préalablement aux inférences sur un jeu d'entraînement plus large. Diversity U-Net nécessite 2 heures d'entraînement.

On observe sur la table 15.4 que la stratégie *target-driven* donne les moins bonnes performances, illustrant bien le compromis tenu entre la spécificité morphologique et le nombre d'atlas. Ainsi, l'approche combinant la diversité et son affinage sur le volume ciblé (approche *fine-tuned*) donne des résultats très intéressants, notamment sur la cas le plus problématique (CAL-4223) montrant la bonne prise en compte de cette morphologie particulière, avec des temps d'inférence de l'ordre de la dizaine de minutes.

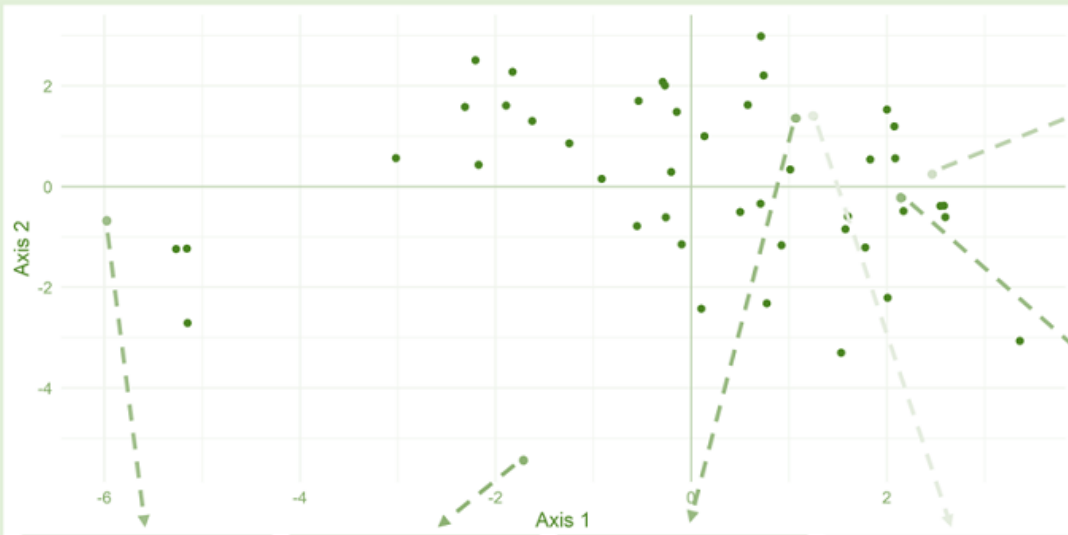
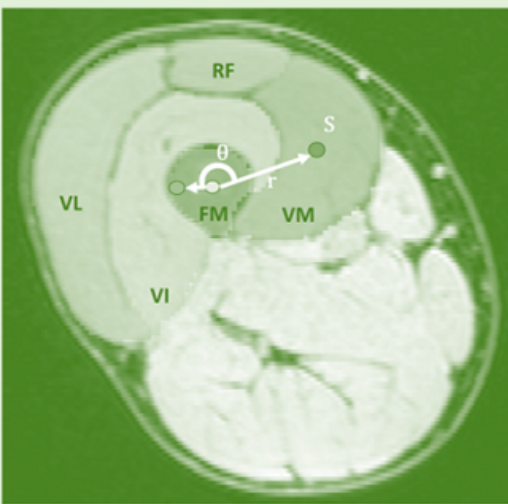
15.4 Conclusion

Dans ce chapitre, nous avons montré l'importance des images utilisées pour l'entraînement des approches de segmentation basées sur le JLF et U-Net. L'augmentation de données est critique pour les réseaux de neurones qui sont incapables de fournir des résultats pertinents sans une grande quantité d'images, capacité dont dispose la segmentation basée sur le multi-atlas.

Dans ce chapitre, nous avons aussi proposé une mesure basée sur la morphologie qui nous a permis d'établir des stratégies de sélection d'atlas et d'augmentation de données permettant d'améliorer significativement le comportement des méthodes de segmentation testées. La contrainte à cette approche est de réaliser la segmentation d'une coupe 2D de l'examen à traiter.

Néanmoins, cette mesure ouvre de nombreuses pistes pour améliorer notre compréhension du fonctionnement des réseaux de neurones. On rappelle qu'à ce jour il n'existe pas de formules permettant de dimensionner les bases de données pour un problème de segmentation, ni d'estimer les performances atteignables pour une base d'images donnée.

Nos récents travaux, non détaillés ici, ont aussi mis en évidence que les réseaux ne se généralisent pas à des changements des conditions d'acquisition IRM, ni à des morphologies et qualité musculaire différentes. Les approches comme le JFL, malgré un temps de calcul long, montrent de bien meilleurs comportements face à ces modifications.



16. Conclusion de la troisième partie

MAP-822

CAL-4223

ANS-3229

ARS-4026

Dans cette partie nous avons présenté nos contributions en segmentation d'images par approches supervisées.

Les deux approches choisies sont la segmentation par recalage multi-atlas avec apprentissage de correction et la famille des réseaux de neurones U-Net. Après les avoir détaillées, nous avons présenté nos résultats sur des applications médicales en échocardiographie et en imagerie IRM des muscles.

Le recalage multi-atlas (JLF et JLF+CL) a montré sa capacité à fournir des segmentations anatomiques de qualité surtout si on ne dispose que d'un petit nombre d'atlas. Sur nos données, les améliorations apportées par l'utilisation de plus de 5 atlas sont très modérées voir insignifiantes. Cependant, leurs temps de calcul longs, la sensibilité à l'erreur de recalage et, intrinsèquement, à la mise en correspondance globale (échec possible si perturbations locales) constituent des limitations importantes dans le cadre d'études sur un grand nombre d'images.

Nos contributions à la segmentation d'images basées sur le réseau U-Net ont montré des avancées significatives dans les problématiques de segmentation des muscles imagés par échographie et IRM. Une fois les problèmes de la quantité et qualité de données résolus, cette approche permet d'obtenir, en temps très raisonnable, des segmentations de bonne qualité.

Le problème d'augmentation de données a aussi été adressé et nous avons présenté une mesure originale qui nous a permis de dériver différentes stratégies d'augmentation d'images utilisant l'information morphologique. Ces stratégies optimisent les images servant de supervision aux approches (choix des atlas pour le recalage, les jeux d'entraînement et de validation pour U-Net) et les rendent ainsi plus efficaces en temps et en performances.

De nombreux projets ont été conduits avec succès grâce à ces approches (DUMORTIER et al., 2022 ; EVAÏN et al., 2020 ; FARAZ et al., 2022 ; SAILLARD et al., 2021), mais de nombreux autres ne sont pas achevés, voire ne sont pas résolus.

On se heurte à un problème simple à exprimer : avec combien d'images d'entraînement, on va pouvoir atteindre la précision spécifiée dans *tous les cas* ?

Avant toutes choses, dans une étude, il faut nécessairement spécifier rigoureusement le périmètre de *tous les cas* et s'y tenir. Ensuite, les problèmes résident d'une part dans le dimensionnement de la complexité de la tâche de segmentation et d'autre part dans la prédiction du comportement des méthodes.

Pour la complexité de la tâche de segmentation, il n'y a pas de méthode permettant de quantifier les variabilités des images (artéfacts, résolution, paramètres d'acquisition et de reconstruction, ...) et des segmentations (différences morphologiques, apparitions aléatoires de pathologies, tailles de régions d'intérêts, ...). L'expérience est nécessaire pour demander aux experts médicaux un nombre approximatif de données annotées, et définir les pré-traitements et être vigilant au respect du protocole d'acquisition pour minimiser la variabilité dans les images.

Pour la prédiction du comportement des méthodes, les problématiques de compréhension et de prédiction des erreurs sont particulièrement étudiées actuellement. Les travaux de thèses en cours de Valentine Wargnier-Dauchelle (WARGNIER-DAUCHELLE et al., 2021a) adressent notamment le problème de l'interprétation des décisions des réseaux de neurones au travers des cartes d'attribution. Il serait aussi intéressant de corréliser notre métrique morphologique aux performances d'un réseau et ainsi prévoir la fiabilité ou la précision de ce réseau sur une nouvelle image. Il s'agirait d'une définition orientée morphologie du domaine d'applicabilité des approches garantissant une précision.

Enfin, il faut revenir sur le nombre de données et comprendre pourquoi les réseaux de neurones ont besoin d'une si grande quantité d'images alors que le recalage atlas, pour la problématiques de segmentation anatomique, n'en nécessite qu'un nombre très raisonnable. Des couplages de ses approches ont été proposées...

Ces idées sont connexes aux thèmes d'apprentissage semi-supervisé, voir *few-shot* et *one-shot*, et d'apprentissage actif. Dans la partie de conclusion qui suit, en plus des bilans sur mes activités, je propose des projets de recherches dans ces thématiques. Notre proposition de mesure morphologique peut s'inscrire dans cette optique de sélection de données à annoter et pourrait être un élément vers une estimation de la complexité de segmentation.

Contributions personnelles

Articles

- DUMORTIER, L., F. GUÉPIN, M. L. DELIGNETTE-MULLER, C. B. BOULOCHER et T. Grenier (2022). « Deep learning in veterinary medicine, an approach based on CNN to detect pulmonary abnormalities from lateral thoracic radiographs in cats ». In : *Scientific Reports* 12.
- EVAIN, E., K. FARAZ, T. Grenier, D. GARCIA, M. DE CRAENE et O. BERNARD (2020). « A pilot study on convolutional neural networks for motion estimation from ultrasound images ». In : *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 67.12, p. 2565-2573.
- FARAZ, K., T. Grenier, C. DUCOTTET et T. EPICIER (2022). « Deep learning detection of nanoparticles and multiple object tracking of their dynamic evolution during in situ ETEM studies ». In : *Scientific Reports* 12, p. 2484.
- LECLERC, S., E. SMISTAD, A. OSTVIK, F. CERVENANSKY, F. ESPINOSA, T. ESPELAND, E. A. R. BERG, M. BELHAMISSI, S. ISRAILOV, T. Grenier, C. LARTIZIEN, P. M. JODOIN, L. LØVSTAKKEN et O. BERNARD (2020). « LU-Net : A multistage attention network to improve the robustness of segmentation of left ventricular structures in 2-D echocardiography ». In : *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 67.12, p. 2519-2530.

LECLERC, S., E. SMISTAD, J. PEDROSA, A. OSTVIK, F. CERVENANSKY, F. ESPINOSA, T. ESPELAND, E. J. BERG, P.-M. JODOIN, T. **Grenier**, C. LARTIZIEN, J. DRHOOGHE, L. LØVSTAKKEN et O. BERNARD (2019c). « Deep learning for segmentation using an open large-scale dataset in 2D echocardiography ». In : *IEEE Transactions on Medical Imaging* 38.9, p. 2198-2210.

Conférences

LECLERC, S., E. SMISTAD, T. **Grenier**, C. LARTIZIEN, A. OSTVIK, F. CERVENANSKY, F. ESPINOSA, T. ESPELAND, E. A. RYE BERG, P.-M. JODOIN, L. LØVSTAKKEN et O. BERNARD (2019a). « RU-Net : A refining segmentation network for 2D echocardiography ». In : *2019 IEEE International Ultrasonics Symposium (IUS)*. Glasgow, France : IEEE, p. 1160-1163.

LECLERC, S., E. SMISTAD, A. OSTVIK, F. CERVENANSKY, F. ESPINOSA, T. ESPELAND, E. ANDREAS, R. BERG, P.-M. JODOIN, T. **Grenier**, C. LARTIZIEN, L. LOVSTAKKEN et O. BERNARD (2019b). « Deep learning segmentation in 2D echocardiography using the CAMUS dataset : Automatic assessment of the anatomical shape validity ». In : *International conference on Medical Imaging with Deep Learning (MIDL 2019)*. London, United Kingdom.

MOREAU, B., A.-H. DICKO, P. MAILLIEZ, P. PORTEJOIE, C. LECOMTE, M. BAH, T. **Grenier**, E. JOLIVET, P. PETIT, B. FRÉCHÈDE, F. FAURE, B. GILLES et P. BEILLAS (2016). « A segmentation pipeline for the creation of statistical shape models in the PIPER project ». In : *22nd Congress of the European Society of Biomechanics (ESB 2016)*.

NGUYEN, H.-T., P. CROISILLE, M. VIALON, C. de BOURGUIGNON, R. GRANGE, S. GRANGE et T. **Grenier** (2018a). « Robust multi-atlas MRI segmentation with corrective learning for quantification of local quadriceps muscles inflammation changes during a longitudinal study in athletes ». In : *Proc. Intl. Soc. Mag. Reson. Med. 26*. Paris, France.

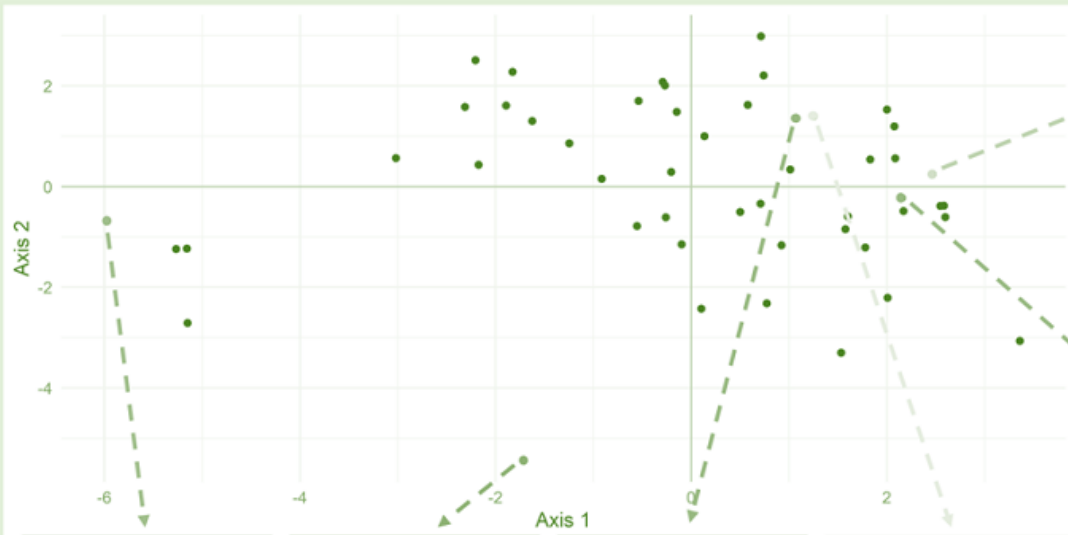
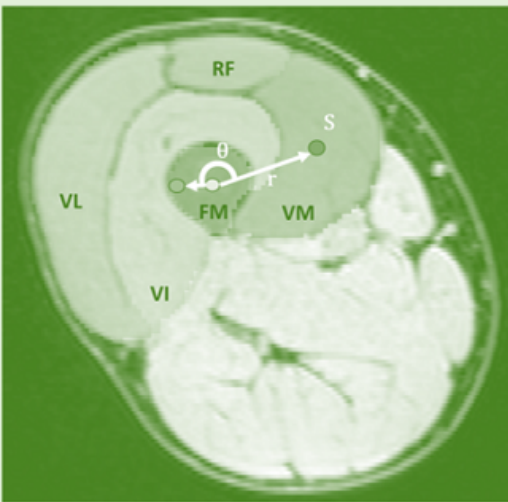
NGUYEN, H.-T., P. CROISILLE, M. VIALON, S. LECLERC, S. GRANGE, R. GRANGE, O. BERNARD et T. **Grenier** (2019b). « Robustly segmenting quadriceps muscles of ultra-endurance athletes with weakly supervised U-Net ». In : *International Conference on Medical Imaging with Deep Learning – Extended Abstract Track*. London, United Kingdom.

NGUYEN, H.-T., T. **Grenier**, B. LEPORQ, L. BEY, M. VIALON et P. CROISILLE (2019c). « Evaluation of local changes in femoral bone marrow during a mountain ultra-marathon with quantitative MRI Results & Discussion ». In : *Proc. Intl. Soc. Mag. Reson. Med. 27*. Montréal, Canada.

NGUYEN, H.-T., T. **Grenier**, B. LEPORQ, C. LE GOFF, C. de BOURGUIGNON, G. GIARDINI, G. MILLET et O. BEUF (2018b). « Longitudinal study of quadriceps muscle head inflammation of athletes enrolled in extreme mountain ultra-marathon using radiomic features extracted from automatic segmentation based on atlas registration and machine learning of MR Images ». In : *ELMSK : Exercise, locomotion and musculoskeletal system*. Lyon, France.

SAILLARD, E., C. CONFAVREUX, M. GARDEGARONT, D. JICQUEL, S. CADOT, D. MITTON, F. BERNMOND, H. FOLLET, J.-B. PIALAT et T. **Grenier** (2021). « UNet based automatic femur segmentation with few annotated data for bone fracture prediction : from pre-processing to segmentation assessments ». In : *Medical Image Analysis and Artificial Intelligence, 2nd Sino French Workshop*. Online, France, 2p.

WARGNIER-DAUCHELLE, V., T. **Grenier**, F. DURAND-DUBIEF, F. COTTON et M. SDIKA (2021a). « A more interpretable classifier for multiple sclerosis ». In : *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. Nice, France, p. 1062-1066.



Bibliographie de la troisième partie

ALKADI, R., A. EL-BAZ, F. TAHER et N. WERGI (2019). « A 2.5D deep learning-based approach for prostate cancer detection on T2-weighted magnetic resonance imaging ». In : *Computer Vision – ECCV 2018 Workshops*. Sous la dir. de L. LEAL-TAIXÉ et S. ROTH. Cham : Springer International Publishing, p. 734-739.

AZAD, R., M. ASADI-AGHBOLAGHI, M. FATHY et S. ESCALERA (2019). « Bi-directional ConvLSTM U-Net with densely connected convolutions ». In : *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*.

BACH CUADRA, M., V. DUAY et J.-P. THIRAN (2015). « Atlas-based Segmentation ». In : *Handbook of Biomedical Imaging : Methodologies and Clinical Research*. Sous la dir. de N. PARAGIOS, J. DUNCAN et N. AYACHE. Boston, MA : Springer US, p. 221-244.

CHLAP, P., H. MIN, N. VANDENBERG, J. DOWLING, L. HOLLOWAY et A. HAWORTH (2021). « A review of medical image data augmentation techniques for deep learning applications ». In : *Journal of Medical Imaging and Radiation Oncology* 65.5, p. 545-563. eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1111/1754-9485.13261>.

ÇIÇEK, Ö., A. ABDULKADIR, S. S. LIENKAMP, T. BROX et O. RONNEBERGER (2016). « 3D U-Net : Learning dense volumetric segmentation from sparse annotation ». In : *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Sous la dir. de S. OURSELIN, L. JOSKOWICZ, M. R. SABUNCU, G. UNAL et W. WELLS. Cham : Springer International Publishing, p. 424-432.

CLEVERT, D.-A., T. UNTERTHINER et S. HOCHREITER (2016). « Fast and accurate deep network learning by exponential linear units (ELUs) ». In : *arXiv : Learning*.

FLECKENSTEIN, J. L. (1993). « Muscle water shifts, volume changes, and proton T2 relaxation times after exercise ». In : *Journal of Applied Physiology* 74.4, p. 2047-2048.

FREUND, Y. et R. E. SCHAPIRE (1996). « Experiments with a new boosting algorithm ». In : *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*. ICML'96. Bari, Italy : Morgan Kaufmann Publishers Inc., 148-156.

GILLES, B., C. DE BOURGUIGNON, P. CROISILLE, G. MILLET, O. BEUF et M. VIALON (2016). « Automatic segmentation for volume quantification of quadriceps muscle head :

- a longitudinal study in athletes enrolled in extreme mountain ultra-marathon ». In : *ISMRM : International Society for Magnetic Resonance in Medicine*.
- GILLES, B. et N. MAGNENAT-THALMANN (2010). « Musculoskeletal MRI segmentation using multi-resolution simplex meshes with medial representations ». In : *Medical Image Analysis* 14.3, p. 291 -302.
- GILLES, B. et D. K. PAI (2008). « Fast musculoskeletal registration based on shape matching ». In : *Med Image Comput Comput Assist Interv* 5242.Pt 2, p. 822-829.
- HAQUE, H., M. HASHIMOTO, N. UETAKE et M. JINZAKI (2019). « Semantic segmentation of thigh muscle using 2.5D deep learning network trained with limited datasets ». In : *ArXiv* abs/1911.09249.
- HE, K., X. ZHANG, S. REN et J. SUN (2016). « Deep residual learning for image recognition ». In : *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 770-778.
- HINTON, G. E., N. SRIVASTAVA, A. KRIZHEVSKY, I. SUTSKEVER et R. R. SALAKHUTDINOV (2012). *Improving neural networks by preventing co-adaptation of feature detectors*. arXiv : 1207.0580 [cs.NE].
- HU, J., L. SHEN, S. ALBANIE, G. SUN et E. WU (2020). « Squeeze-and-excitation networks ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42.8, p. 2011-2023.
- HUANG, H., L. LIN, R. TONG, H. HU, Q. ZHANG, Y. IWAMOTO, X. HAN, Y.-W. CHEN et J. WU (2020). « UNet 3+ : A full-scale connected UNet for medical image segmentation ». In : ii, p. 1055-1059. arXiv : 2004.08790.
- IGLESIAS, J. E. et M. R. SABUNCU (2015). « Multi-atlas segmentation of biomedical images : A survey ». In : *Medical Image Analysis* 24.1, p. 205-219.
- IOFFE, S. et C. SZEGEDY (2015). « Batch normalization : Accelerating deep network training by reducing internal covariate shift ». In : *Proceedings of the 32nd International Conference on Machine Learning*. Sous la dir. de F. BACH et D. BLEI. T. 37. Proceedings of Machine Learning Research. Lille, France : PMLR, p. 448-456.
- ISENSEE, F., P. F. JAEGER, S. A. A. KOHL, J. PETERSEN et K. H. MAIER-HEIN (2021). « nnU-Net : Self-adapting framework for U-Net-based medical image segmentation ». In : *Nature Methods* 18, p. 203-211.
- ISOLA, P., J.-Y. ZHU, T. ZHOU et A. A. EFROS (2017). « Image-to-image translation with conditional adversarial networks ». In : *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 5967-5976.
- JHA, D., P. H. SMEDSRUD, M. A. RIEGLER, D. JOHANSEN, T. DE LANGE, P. HALVORSEN et H. D. JOHANSEN (2019). « ResUNet++ : An advanced architecture for medical image segmentation ». In : *Proceedings - 2019 IEEE International Symposium on Multimedia, ISM 2019*, p. 225-230. arXiv : 1911.07067.
- KER, J., L. WANG, J. RAO et T. LIM (2018). « Deep learning applications in medical image analysis ». In : *IEEE Access* 6, p. 9375-9389.
- KLAMBAUER, G., T. UNTERTHINER, A. MAYR et S. HOCHREITER (2017). « Self-normalizing neural networks ». In : *Proceedings of the 31st international conference on neural information processing systems*, p. 972-981.
- LE TROTIER, A., A. FOURÉ, M. GUYE, S. CONFORT-GOUNY, J.-P. MATTEI, J. GONDIN, E. SALORT-CAMPANA et D. BENDAHAN (2016). « Volume measurements of individual muscles in human quadriceps femoris using atlas-based segmentation approaches ». In : *Magnetic Resonance Materials in Physics, Biology and Medicine* 29.2, p. 245-257.
- LEE, C.-Y., S. XIE, P. GALLAGHER, Z. ZHANG et Z. TU (2015). « Deeply-supervised nets ». In : sous la dir. de G. LEBANON et S. V. N. VISHWANATHAN. T. 38. Proceedings of Machine Learning Research. San Diego, California, USA : PMLR, p. 562-570.

- LEHTINEN, J., J. MUNKBERG, J. HASSELGREN, S. LAINE, T. KARRAS, M. AITTALA et T. AILA (2018). « Noise2Noise : Learning image restoration without clean data ». In : sous la dir. de J. DY et A. KRAUSE. T. 80. Proceedings of Machine Learning Research. Stockholmsmässan, Stockholm Sweden : PMLR, p. 2965-2974.
- LI, P., X.-Y. ZHOU, Z. WANG et G. YANG (2019). « Z-Net : an asymmetric 3D DCNN for medical CT volume segmentation ». In : *ArXiv* abs/1909.07480.
- LI, X., H. CHEN, X. QI, Q. DOU, C. FU et P. HENG (2018). « H-DenseUNet : Hybrid densely connected UNet for liver and tumor segmentation from CT volumes ». In : *IEEE Transactions on Medical Imaging* 37.12, p. 2663-2674.
- LITJENS, G., T. KOOI, B. E. BEJNORDI, A. A. SETIO, F. CIOMPI, M. GHAFORIAN, J. V. D. LAAK, B. GINNEKEN et C. SÁNCHEZ (2017). « A survey on deep learning in medical image analysis ». In : *Medical image analysis* 42, p. 60-88.
- MAES, F., D. LOECKX, D. VANDERMEULEN et P. SUETENS (2015). « Image registration using mutual information ». In : *Handbook of Biomedical Imaging : Methodologies and Clinical Research*. Sous la dir. de N. PARAGIOS, J. DUNCAN et N. AYACHE. Boston, MA : Springer US, p. 295-308.
- MILLETARI, F., N. NAVAB et S. AHMADI (2016). « V-Net : Fully convolutional neural networks for volumetric medical image segmentation ». In : *2016 Fourth International Conference on 3D Vision (3DV)*, p. 565-571.
- MINAEI, S., Y. BOYKOV, F. PORIKLI, A. PLAZA, N. KEHTARNAVAZ et D. TERZOPOULOS (2022). « Image segmentation using deep learning : A survey ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.7, p. 3523-3542.
- NEWELL, A., K. YANG et J. DENG (2016). « Stacked hourglass networks for human pose estimation ». In : *Computer Vision – ECCV 2016*. Sous la dir. de B. LEIBE, J. MATAS, N. SEBE et M. WELLING. Springer International Publishing, p. 483-499.
- PERSLEV, M., E. B. DAM, A. PAI et C. IGEL (2019). « One network to segment them all : A general, lightweight system for accurate 3D medical image segmentation ». In : *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Sous la dir. de D. SHEN, T. LIU, T. M. PETERS, L. H. STAIB, C. ESSERT, S. ZHOU, P.-T. YAP et A. KHAN. Springer International Publishing, p. 30-38.
- PRESCOTT, J. W., T. M. BEST, M. S. SWANSON, F. HAQ, R. D. JACKSON et M. N. GURCAN (2011). « Anatomically anchored template-based level set segmentation : Application to quadriceps muscles in MR images from the osteoarthritis initiative ». In : *Journal of Digital Imaging* 24.1, p. 28-43.
- RIZWAN I HAQUE, I. et J. NEUBERT (2020). « Deep learning approaches to biomedical image segmentation ». In : *Informatics in Medicine Unlocked* 18, p. 100297.
- ROHLFING, T., R. BRANDT, R. MENZEL, D. B. RUSSAKOFF et C. R. MAURER (2005). « Quo vadis, Atlas-based segmentation ? » In : *Handbook of Biomedical Image Analysis : Volume III : Registration Models*. Sous la dir. de J. S. SURI, D. L. WILSON et S. LAXMINARAYAN. Boston, MA : Springer US, p. 435-486.
- RONNEBERGER, O., P. FISCHER et T. BROX (2015). « U-Net : Convolutional networks for biomedical image segmentation ». In : *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Sous la dir. de N. NAVAB, J. HORNEGGER, W. M. WELLS et A. F. FRANGI. Cham : Springer International Publishing, p. 234-241.
- RUECKERT, D., L. I. SONODA, C HAYES, D. L. HILL, M. O. LEACH et D. J. HAWKES (1999). « Nonrigid registration using free-form deformations : application to breast MR images. » In : *IEEE Trans. Med. Imaging* 18.8, p. 712-721.

- SHELHAMER, E., J. LONG et T. DARRELL (2017). « Fully convolutional networks for semantic segmentation ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.4, p. 640-651.
- SHORTEN, C. et T. M. KHOSHGOFTAAR (2019). « A survey on image data augmentation for deep learning ». In : *Journal of big data* 6.1, p. 1-48.
- SIDDIQUE, N., S. PAHEDING, C. P. ELKIN et V. DEVABHAKTUNI (2021). « U-Net and its variants for medical image segmentation : A review of theory and applications ». In : *IEEE Access* 9, p. 82031-82057.
- SRIVASTAVA, N., G. HINTON, A. KRIZHEVSKY, I. SUTSKEVER et R. SALAKHUTDINOV (2014). « Dropout : A simple way to prevent neural networks from overfitting ». In : *Journal of Machine Learning Research* 15.56, p. 1929-1958.
- SUDRE, C. H., W. LI, T. VERCAUTEREN, S. OURSELIN et M. JORGE CARDOSO (2017). « Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations ». In : *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Sous la dir. de M. J. CARDOSO, T. ARBEL, G. CARNEIRO, T. SYEDA-MAHMOOD, J. M. R. TAVARES, M. MORADI, A. BRADLEY, H. GREENSPAN, J. P. PAPA, A. MADABHUSHI, J. C. NASCIMENTO, J. S. CARDOSO, V. BELAGIANNIS et Z. LU. Cham : Springer International Publishing, p. 240-248.
- TAJBAKHS, N., L. JEYASEELAN, Q. LI, J. N. CHIANG, Z. WU et X. DING (2020). « Embracing imperfect datasets : A review of deep learning solutions for medical image segmentation ». In : *Medical Image Analysis* 63, p. 101693.
- TOMPSON, J., R. GOROSHIN, A. JAIN, Y. LECUN et C. BREGLER (2015). « Efficient object localization using convolutional networks ». In : *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- VIGNEAULT, D. M., W. XIE, C. Y. HO, D. A. BLUEMKE et J. A. NOBLE (2018). « Omega-Net (Omega-Net) : Fully automatic, multi-view cardiac MR detection, orientation, and segmentation with deep neural networks ». In : *Medical Image Analysis* 48, p. 95-106.
- WANG, G., X. LIU, C. LI, Z. XU, J. RUAN, H. ZHU, T. MENG, K. LI, N. HUANG et S. ZHANG (2020). « A noise-robust framework for automatic segmentation of COVID-19 pneumonia lesions from CT images ». In : *IEEE Transactions on Medical Imaging* 39.8, p. 2653-2663.
- WANG, H., J. W. SUH, S. R. DAS, J. B. PLUTA, C. CRAIGE et P. A. YUSHKEVICH (2013a). « Multi-atlas segmentation with joint label fusion ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.3, p. 611-623.
- WANG, H. et P. A. YUSHKEVICH (2012). « Spatial bias in multi-atlas based segmentation ». In : *2012 IEEE Conference on Computer Vision and Pattern Recognition*, p. 909-916.
- WANG, H., S. R. DAS, J. W. SUH, M. ALTINAY, J. PLUTA, C. CRAIGE, B. AVANTS et P. A. YUSHKEVICH (2011). « A learning-based wrapper method to correct systematic errors in automatic image segmentation : Consistently improved performance in hippocampus, cortex and brain segmentation ». In : *NeuroImage* 55.3, p. 968-985.
- WANG, H., D. KAKRANIA, H. TANG, P. PRASANNA et T. SYEDA-MAHMOOD (2018a). « Fast anatomy segmentation by combining coarse scale multi-atlas label fusion with fine scale corrective learning ». In : *Computerized Medical Imaging and Graphics* 68, p. 16 -24.
- WANG, H. et P. YUSHKEVICH (2013b). « Multi-atlas segmentation with joint label fusion and corrective learning - an open source implementation ». In : *Frontiers in Neuroinformatics* 7, p. 27.

-
- WANG, Z., A. BOVIK, H. SHEIKH et E. SIMONCELLI (2004b). « Image quality assessment : from error visibility to structural similarity ». In : *IEEE Transactions on Image Processing* 13.4, p. 600-612.
- XU, B., N. WANG, T. CHEN et M. LI (2015a). *Empirical evaluation of rectified activations in convolutional network*. arXiv : 1505.00853 [cs.LG].
- ZHANG, Z., Q. LIU et Y. WANG (2018). « Road extraction by deep residual U-Net ». In : *IEEE Geoscience and Remote Sensing Letters* 15.5, p. 749-753.
- ZHAO, A., G. BALAKRISHNAN, F. DURAND, J. V. GUTTAG et A. V. DALCA (2019). « Data augmentation using learned transformations for one-shot medical image segmentation ». In : *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- ZHOU, Z., M. M. R. SIDDIQUEE, N. TAJBAKHSH et J. LIANG (2020). « UNet++ : Redesigning skip connections to exploit multiscale features in image segmentation ». In : *IEEE Transactions on Medical Imaging* 39.6, p. 1856-1867.
- ZHOU, Z.-H. (2012). *Ensemble methods : Foundations and algorithms*. 1st. Chapman & Hall/CRC.

IV

Conclusion Générale

17	Conclusion et projet de recherche	177
17.1	Conclusion	
17.2	Le bilan des contributions	
17.3	Perspectives de recherche	
17.4	Bilan de rédaction de mon HDR	



17. Conclusion et projet de recherche

17.1 Conclusion

Dans ce manuscrit, j'ai présenté mes activités en enseignement et en recherche conduites depuis ma nomination en septembre 2006 en tant que maître de conférences. Notamment, j'ai détaillé deux thèmes de recherche. Le premier, sur le filtrage à base de *mean-shift* et la segmentation par croissance de région. Le second, sur les approches de segmentation supervisée multi-atlas et d'apprentissage profond (réseau U-Net). Ces deux thèmes de recherche ont été appliqués à différents problèmes de segmentation et filtrage d'images médicales dans le but d'améliorer le diagnostic ou la compréhension de processus pathologiques inflammatoires.

Ces deux thèmes semblent s'opposer. Il est intéressant de comparer les changements de paradigmes entre eux. Le premier, *mean-shift* et croissance de région, appartiendrait "au passé" alors que le second, l'apprentissage profond, est en pleine effervescence. Le premier s'appuie sur une écriture algorithmique (modèle) contrôlant son comportement en filtrage ou en segmentation d'images. Le second apprend son comportement en s'appuyant sur des exemples d'images.

Cette conclusion est orientée sur le bilan et les perspectives de recherche. Le bilan des contributions sur ces thématiques est dressé dans la section 17.2.

Dans la section 17.3, je détaille les poursuites que j'envisage pour ces travaux et les projets de recherche.

Enfin je dresserai un bilan de ma rédaction d'HDR.

17.2 Le bilan des contributions

Une partie de mes contributions personnelles (8 articles sur 17) alimentent les parties II et III de ce manuscrit. Elles sont listées en fin de chacune des parties et toutes données à la fin du chapitre 6.

- Pour la partie II, sur le filtrage et la segmentation non-supervisée, les contributions proposées permettent de "boucler la boucle" avec le spatio-temporel : le formalisme

de filtrage spatio temporel *mean-shift* étendu au formalisme de croissance de région. Il reste néanmoins quelques frustrations liées à la non valorisation de certains travaux sur les énergies minimisées en croissance de région (la suite des travaux de thèse de Jean-Loïc Rose) et de résultats sur l'application de ces approches à l'AVC (travaux de thèse de Ting Li). Malgré plusieurs tentatives pour relancer et finaliser ces travaux, ils deviennent de plus en plus obsolètes et difficiles à publier.

- Pour la partie III, les travaux sur les optimisations et améliorations d'U-Net et de l'approche JLF me permettent aujourd'hui encore de répondre efficacement à des collaborations appliquées ou recherche. Une valorisation des contributions méthodologiques sur la mesure morphologique est en cours.

De manière globale, ces dernières années, on observe une augmentation du nombre et une diversification du type de mes communications. Cela correspond beaucoup à l'autre partie de mes contributions (9 articles sur 17, 2 brevets déposés) qui résulte principalement des avancées de différents projets sur lesquels je me suis impliqué en apportant mon expertise en traitements d'images notamment pour les méthodes relevant du *deep learning*.

Il faut aussi reconnaître que l'effet de mode autour du *deep learning* (voire "IA") a facilité les possibilités de publications, aubaine pour les sociétés savantes qui ont diversifié et augmenté les supports de communication (nouvelles conférences, workshops, revues, ...).

Cependant, il est de plus en plus complexe de publier des avancées méthodologiques dans cette thématique, notamment pour les raisons suivantes :

- l'état de l'art est très riche et il évolue rapidement. Explorer l'important état de l'art est lourd car il devient de plus en plus difficile de discerner les travaux d'importance des autres : les informations partielles, les biais d'étude, la réelle capacité de l'approche proposée à se généraliser à un autre problème n'étant pas évidents à déterminer rapidement. La science ouverte ne résout pas tout. A mon sens, nous manquons de modélisations théoriques et notamment d'éléments de caractérisation du problème de traitement d'images et de sa complexité, notamment en segmentation.
- l'implication de grands groupes (Facebook, Google, NVidia, Microsoft, IBM...) peut rendre rapidement des travaux de recherche hautement compétitifs et difficiles à soutenir par une petite équipe qui ne disposera pas des mêmes ressources humaines, en données et en calculs que ces groupes.
- il est demandé de plus en plus de tests : en termes de quantité d'images, de bases de données étudiées, d'explorations informatiques et statistiques lourdes à réaliser. Cela implique une intégration continue de codes de la communauté (quels que peu modifiés), l'accès à des centres de calcul très puissants (exemple de l'IN2P3 ou Jean Zay du CNRS), et de privilégier l'utilisation de bases de données 'libres' afin de faciliter la comparaison aux autres travaux. Un enjeu important est de maintenir ses compétences informatiques ou de transférer cette charge à un service dédié (à l'échelle du laboratoire ou externalisé ?).
- l'effervescence de cette discipline ne permet pas d'avoir un cadrage de ses ramifications méthodologiques, théoriques et applicatives et ceci même au niveau des mots clés utilisés. Il faut donc prospecter très largement l'immense état de l'art. Un travail d'échanges scientifiques en équipe à l'échelle du laboratoire est important, mais est crucial à des échelles plus grande de manière à couvrir d'autres domaines que le médical dont le très large *computer vision*. Un enjeu serait d'aller vers une taxonomie des méthodes et être capable de la faire accepter et de la maintenir à jour.

17.3 Perspectives de recherche

Je présente dans cette partie les éléments de recherche qui me semblent pertinents de poursuivre afin de contribuer aux développements des méthodes d'analyse d'images médicales, notamment par apprentissage.

A court et moyen termes, la recherche en segmentation d'images est ce que je privilégie de poursuivre car cette étape est : i) quasi systématique sur toutes les études médicales, ii) spécifique aux différentes études et iii) est aussi la plus fastidieuse à réaliser manuellement. Mon objectif est de rendre les méthodes suffisamment fiables pour aider au diagnostic et au pronostic. Cela passe d'une part, par une meilleure compréhension des méthodes et, d'autre part, par une meilleure caractérisation des données et de la tâche à réaliser.

Pour mieux appréhender le fonctionnement des réseaux, vu la complexité du problème, une piste judicieuse serait d'établir des **couplages forts entre méthodes de deep learning et d'autres méthodes de segmentation supervisées et non supervisées**.

Ensuite, les récents travaux de thèse de Hoai-Thu Nguyen ont permis d'élaborer des approches de segmentation supervisées apprenant sur une base de données d'entraînement très petite (3 à 5 données). Les segmentations obtenues avec si peu d'atlas ne sont certes pas parfaites mais pourraient servir de base aux experts qui n'apportent alors que leurs corrections, que l'on souhaite les plus minimales possibles. Développer des **méthodes d'apprentissage en continu et d'apprentissage actif** pour faciliter l'annotation manuelle notamment en segmentation est ainsi une piste de recherche pertinente.

Enfin, pour la caractérisation des atlas, une autre proposition élaborée dans les travaux de thèse de Hoai-Thu Nguyen, est la sélection des meilleurs atlas, par mesure morphologique, à utiliser pour l'entraînement afin de minimiser les risques d'erreur de segmentation. Étendre ces travaux à la variabilité des images et des annotations permettraient d'appréhender le **dimensionnement en données nécessaires à l'apprentissage et de prévoir les limites** de fonctionnement des réseaux en termes de précision de segmentation, changement dans les images (artéfacts, paramètres IRM,...), etc. Ceci permettrait aussi de proposer des **solutions d'apprentissages faiblement supervisées** notamment pour les études exploitant des données hétérogènes de suivi et démographiques, qui obligent à considérer des effectifs homogènes très petits, mais qui sont particulièrement utiles pour obtenir un pronostic pertinent.

Ces pistes de recherches sont détaillées dans les paragraphes suivants.

17.3.1 Couplage fort entre méthodes de deep learning et d'autres méthodes

Nous avons déjà observé que les architectures de type U-Net sont capables de produire des généralisations 3D à partir d'apprentissages 2D. Aussi, nous avons fréquemment recours à l'interpolation lors de la création manuelle des références 3D. Or il existe de nombreuses méthodes semi-automatiques qui permettent de proposer des segmentations spécifiques (multi-atlas, croissance de région, level-sets) capables de prendre en compte des a priori de géométrie ou d'intensité -comme le fait l'interpolation- et qui faciliteraient le travail des experts.

Une piste intéressante et originale serait de coupler ces méthodes en s'inspirant de l'approche *teacher-student* (ou *knowledge distillation*) très efficace en apprentissage profond mais pour l'instant réservée à l'utilisation de deux réseaux de neurones.

Ceci pourrait être facilité par les nombreuses propositions de *deep unrolling* (MONGA et al., 2021), voir aussi en *deep unfolding*, qui réalisent des traitements issus de modèles algorithmiques à l'aide de réseaux de neurones. Ces réseaux, et leurs fonctions de coût, sont l'implémentation de l'expression mathématique adressée par le modèle algorithmique.

17.3.2 Apprentissage continu, apprentissage actif

L'approche précédente ajoute du déterminisme pour améliorer l'apprentissage et diminuer les erreurs de généralisation liées aux *a priori*. Ici, l'objectif serait de continuer d'apprendre dès que l'expert fait une correction ou propose une nouvelle annotation de donnée (apprentissage continu) puis de lui proposer les nouvelles données pour lesquelles le réseau estime que sa connaissance est insuffisante pour faire une prédiction fiable (apprentissage actif) (REN et al., 2021). Un tel outil serait particulièrement efficace en clinique, mais nécessite des réseaux capables de s'auto-estimer et surtout légers pour pouvoir s'apprendre facilement (exemples d'architecture : MobileNet pour la classification d'image et ENet pour la segmentation d'images).

Ce problème d'intérêt fait l'objet de développements importants notamment par NVidia (NATH et al., 2021) et MONAI Label (DIAZ-PINTO et al., 2022a ; DIAZ-PINTO et al., 2022b) et d'intégration dans les outils d'annotation comme 3DSlicer ou la plateforme web d'Open Health Imaging Foundation¹ ce qui facilite la recherche de nouvelles approches. Plusieurs scénarios incluant des interactions plus ou moins temps réel ont déjà été élaborés (BUDD et al., 2021).

17.3.3 Complexité du problème de segmentation et dimensionnement

"Thomas, de combien d'images et de segmentation as tu besoin ?"

En 2022, à l'heure des mécanismes d'attention, des transformers, des réseaux de diffusion, GAN, etc nous ne sommes pas capable de répondre à cette question.

Elle est peut être mal posée. Reformulons : "Etant donné une tâche de segmentation et une précision attendue, ainsi qu'une gamme de variabilité d'images et de segmentations, quelles sont les données à apprendre ?".

Les liens avec l'apprentissage actif sont évidents, mais peut-on, a priori, évaluer la complexité du problème de segmentation ? Cela permettrait de déterminer une méthode ou une architecture adaptée au problème et possiblement les besoins en images, en pré-traitements (dans un but d'homogénéiser les images : adapter le domaine) et en augmentation de données.

Ainsi, l'enjeu majeur est la capacité à mesurer la complexité du problème de segmentation dans son ensemble. Des travaux initiés par Hoai-Thu Nguyen, la variabilité de la morphologie est un facteur important de cette mesure. Il faudrait maintenant la coupler avec les images et étudier leur répartition pour comprendre la variabilité des données et la complexité à traiter une nouvelle image.

17.3.4 Apprentissage faiblement supervisé de données hétérogènes pour le pronostic

La concurrence sur les approches de segmentation en *deep learning* en médical va devenir de plus en plus forte, voir insoutenable par une petite équipe de recherche. Il faudra s'appuyer sur les myriades d'autres réseaux, bases de données tierces (annotées ou pas)... c'est-à-dire faire de l'apprentissage semi-supervisé (au sens large) pour être compétitif. Pour innover dans le contexte médical de pronostic, il s'agira, à mon sens, plus de tirer profit de données spécifiques (séquences IRM dédiées à des pathologies, imagerie multi spectrales) et surtout de données de suivi hétérogènes (données images + biologiques + démographiques + ...) afin d'obtenir un résultat personnalisé.

Conceptuellement, cela ne semble pas compatible avec nos besoins en fiabilité ni avec ceux du machine learning... Il faudra ainsi être capable d'innover sur la prise en compte de ces informations et aussi de proposer des méthodes de généralisation fiables.

1. OHIF <https://ohif.org/>, consultée le 14/10/2022

Un premier pas serait de proposer des réseaux fiables pour l'aide au diagnostic (sain ou malade, tumeur maligne ou bénigne, sévérité de l'infarctus, importance de l'inflammation, ...) sans nécessairement disposer de toutes les segmentations manuelles des images mais simplement du compte rendu ou observations médicales. Ces recherches sur l'explicabilité et sur l'apprentissage de segmentation faiblement supervisé (les vérités terrain sont une simple classification) rentrent dans le cadre des travaux de thèse en cours de Valentine Wagnier-Dauchelle que je co-encadre.

Le challenge suivant sera ensuite d'étendre ces approches au pronostic. Il faudra nécessairement inclure des données longitudinales et hétérogènes ce qui réduira la capacité de généralisation et d'explicabilité des approches.

Effectivement, pour la capacité de généralisation, l'augmentation d'informations discriminant les individus segmente la base de données en plus petits groupes homogènes. Ceci peut être vu comme l'augmentation du nombre de dimensions en entrée sans augmenter la quantité de données : l'apprentissage devient particulièrement délicat (espace peu dense, sur-apprentissage...), surtout si on ne peut faire d'augmentation de données. L'alternative de représentation par graphes des données est une possible piste diminuant les dimensions des données d'entrée. Cette piste est explorée dans le cadre des travaux de thèse d'Enyi Chen qui adressent l'utilisation des connexions morphologiques pour améliorer la classification des formes de lésions SEP.

Puis, pour l'explicabilité, le problème est maintenant traité globalement et devient donc plus difficile à expliquer ou justifier car sa résolution s'appuie sur l'ensemble des données d'entrée et non sur une opération interprétable sur chaque instant du temps par exemple. Pour ce dernier point, les fonctions de coût pourraient permettre de contraindre le réseau à apprendre de manière progressive ou à comprendre le rôle de chaque entrée dans la décision finale.

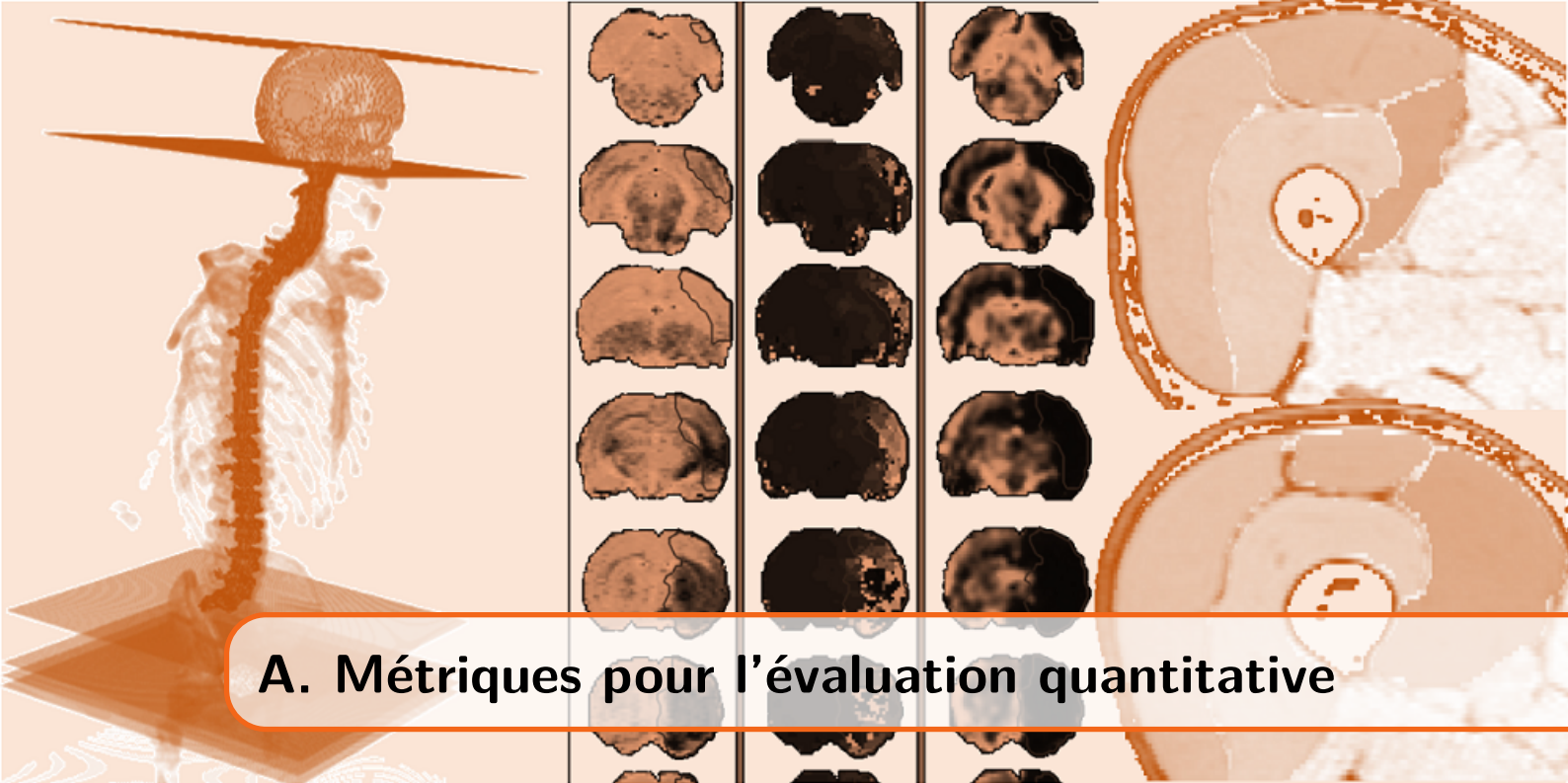
17.4 Bilan de rédaction de mon HDR

Rédigez son HDR, c'est consacrer du temps pour soi et ses projets, et revoir certains concepts parfois un peu négligés ou oubliés. C'est aussi beaucoup découvrir cet exercice et douter. J'ai pris le parti d'être assez exhaustif et de détailler certains éléments techniques à des fins de réutilisation pédagogiques et de réflexions sur des contributions en cours puis les projets à venir.

Cette rédaction a pris du temps, la tâche étant longue et alternée de périodes pleines et de périodes creuses liées notamment à des impératifs professionnels et personnels. Ces alternances n'ont pas permis d'avoir la meilleure des efficacités... Un petit exemple : il a fallu faire trois versions du projet de recherche tant les évolutions dans le domaine du *deep learning* sont rapides.

Cependant, le travail est fait et m'a été profitable en de très nombreux points, notamment sur la centralisation des informations et surtout la maturation des projets à conduire. Il reste toujours des doutes sur la pérennité de ces projets. Et de manière générale, mon principal doute est comment se positionner individuellement en recherche par rapport aux avancées tellement rapides et nombreuses de l'état de l'art et par rapport à l'efficacité de certains groupes de recherche. Fédérer une équipe de chercheurs sur une problématique critique en segmentation pour le médical sera mon principal objectif suite à cette HDR.

Ce manuscrit d'HDR n'aura donc pas servi à rien, et sait-on jamais, sur un malentendu, il pourrait servir d'exemple à suivre ou pas :)



A. Métriques pour l'évaluation quantitative

A.1 Evaluations quantitatives du filtrage d'image

Le rapport de signal à bruit crête (*Peak Signal Noise Ratio*, PSNR) est l'une des mesures de qualité d'image la plus connue et utilisée. Cette mesure s'exprime en décibel dB. Plus la valeur du *PSNR* est grande plus similaires sont les deux images X et Y comparées. Le *PSNR* est calculé ainsi :

$$PSNR(X, Y) = 10 \cdot \log_{10} \left(\frac{range^2}{MSE(X, Y)} \right) \quad (1.1)$$

avec *range* la dynamique des images. Cette mesure s'appuie sur la *Mean Square Error* (MSE) qui calcule la moyenne, sur l'ensemble des n pixels des deux images, des erreurs quadratique des amplitudes : $(\mathbf{x}_i)_r$ pour l'image X et $(\mathbf{y}_i)_r$ pour l'image Y , au même pixel i .

$$MSE(X, Y) = \frac{1}{n} \sum_{i=1}^n \|(\mathbf{x}_i)_r - (\mathbf{y}_i)_r\|^2 \quad (1.2)$$

On comparera souvent la valeur de *PSNR* obtenue entre l'image filtrée et la référence (non corrompue) et la valeur de *PSNR* dit de « référence » (noté $PSNR^{ref}$) entre l'image corrompue et l'image de référence.

Le *PSNR* et la *MSE* ne sont pas les seules mesures utilisées pour l'évaluation quantitative de filtrage. Une mesure de similarité fréquemment utilisée permettant de compléter l'interprétation des résultats basée sur le *PSNR* est le *SSIM* (WANG et al., 2009; WANG et al., 2002; WANG et al., 2004b) qui se veut plus corrélé à la sensibilité de perception humaine et apporte une alternative quand la *MSE* ne peut différencier deux images (DOSSELMANN et al., 2011). Le *SSIM* se calcule localement sur des fenêtres (taille 8×8 recommandée par les auteurs) et sur la valeur de luminance de l'image. La valeur du *SSIM* est comprise entre 0 et 1. Plus le *SSIM* est proche de 1, plus similaire sont les deux images.

On note l'arrivée de méthodes basées sur l'apprentissage profond pour réaliser des comparaisons de qualité "comme l'humain", notamment l'approche LPIPS (ZHANG et al., 2018).

A.2 Evaluations quantitatives de la segmentation d'image

Il existe de nombreuses mesures d'évaluation de la qualité de segmentation (TAHA et al., 2015 ; WANG et al., 2020). Ces mesures sont souvent décrites pour une segmentation en deux classes de l'image (ou binaire) mais elles se généralisent bien au cas multi-classes où plusieurs objets et un fond sont segmentés. Les paragraphes suivants présentent les 5 métriques utilisées dans ce manuscrit : le score de Dice, le coefficient de Jaccard, la distance de Hausdorff, la distance absolue moyenne et la similarité de volume. Comme le décrit dans (REINKE et al., 2021) pour bien appréhender un résultat ou comparer des approches de segmentation, il faut considérer simultanément plusieurs métriques (score de Dice et distance de Hausdorff par exemple).

A.2.1 Indice de Sorensen-DICE

Le coefficient, ou score ou indice, de DICE (DICE, 1945) est la mesure la plus courante permettant de comparer deux ensembles R (ou la région référence) et T (ou la région prédite).

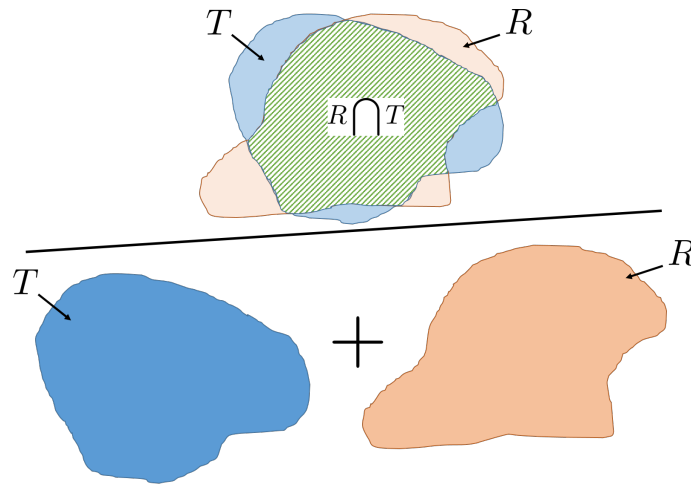


FIGURE A1 – Illustration pour le calcul de l'indice de DICE. Pour le DICE, il faut multiplier par 2.

$$DICE(R, T) = \frac{2|R \cap T|}{|R| + |T|} = \frac{2|R \cap T|}{|R \cup T| + |R \cap T|} \quad (1.3)$$

où $|\cdot|$ désigne le cardinal de l'ensemble.

LE DICE peut aussi s'exprimer avec les notations suivantes :

- TP pour *True Positive* : nombre de pixels où T et R ont la valeur *objet*,
- TN pour *True Negative* : nombre de pixels où T et R ont la valeur *fond*,
- FP pour *False Positive* : nombre de pixels où T est le *objet* et R *fond*,
- FN pour *False Negative* : nombre de pixels où T est le *fond* et R *objet*,

$$DICE = \frac{2TP}{2TP + FP + FN} \quad (1.4)$$

Plus le coefficient de DICE est grand ($0 \leq DICE(R, T) \leq 1$), plus similaires sont les 2 segmentations.

Dans le cas multi-classes, il faudra utiliser le *Dice Similarity Coefficient (DSC)* qui est la moyenne des DICE obtenus pour chacune des classes et en ignorant la classe du fond. Il est possible de pondérer cette valeur de DSC par le nombre de pixels dans chaque classe. On notera cette mesure DSCw qui sera calculée ainsi :

$$DSCw(R, T) = 2 \frac{\sum_{c=1}^N |R_c \cap T_c|}{\sum_{c=1}^N |R_c| + |T_c|} \quad (1.5)$$

avec N , le nombre de classes.

Le score de DICE est équivalent au score F_1 en utilisant les définitions précédentes. Cependant, le score F_1 est aussi utilisé pour la détection et dans ce cas, un vrai positif est défini comme tel quand la superposition entre T et R est suffisant (par exemple plus de 80%). Cette approche est formalisée avec l'intersection de l'union.

A.2.2 Coefficient de Jaccard ou Intersection de l'union

Le coefficient de Jaccard ou l'IoU (*Intersection over Union*) permettent de mesurer le rapport entre l'intersection et l'union de deux ensembles.

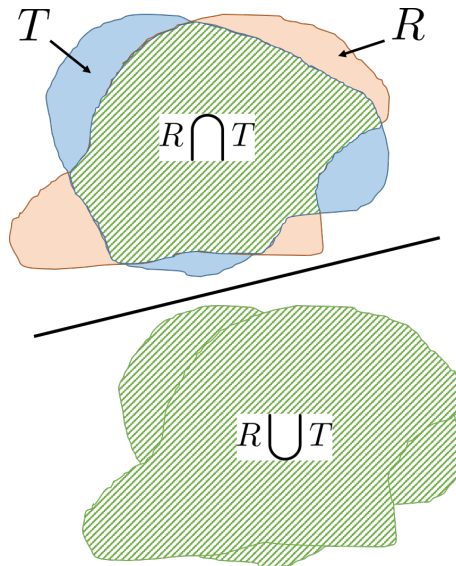


FIGURE A2 – Illustration du coefficient de Jaccard ou IoU.

Leur expression est la suivante :

$$J(R, T) = \frac{|R \cap T|}{|R \cup T|} \quad (1.6)$$

Cette expression est équivalente au coefficient de DICE mais est formulée différemment. La relation entre les coefficients de Jaccard et de DICE est :

$$J(R, T) = DSC(R, T) / (2 - DSC(R, T)) \text{ et } DSC(R, T) = 2J(R, T) / (1 + J(R, T)) \quad (1.7)$$

Plus le coefficient de Jaccard est grand, plus similaires sont les 2 segmentations ($0 \leq J(R, T) \leq 1$).

On peut noter des propositions de métriques hybrides entre Dice et IoU. Dans leurs travaux (KIRILLOV et al., 2019) sur la segmentation panoptique qui associe à chaque pixel une classe et un identifiant propre à l'instance de chaque objet d'une même classe, les auteurs

proposent une nouvelle métrique la *panoptic quality* qui est proportionnelle au produit de l'IoU par le DICE. Dans cette métrique, les auteurs utilisent l'IoU pour mesurer la qualité de segmentation et le DICE pour mesurer la qualité de reconnaissance.

A.2.3 Distance de Hausdorff, d_H

La distance de Hausdorff, notée d_H , est basée sur la distance entre les frontières des segmentations (voir illustration de la figure A3).

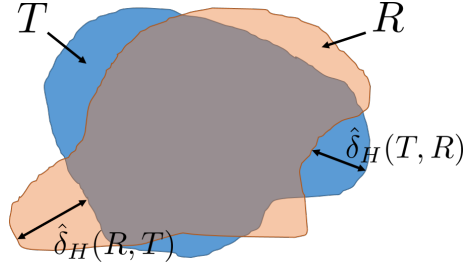


FIGURE A3 – Illustration des distances de Hausdorff.

La pseudo-distance de Hausdorff $\hat{\delta}_H$, non-symétrique, s'exprime ainsi :

$$\hat{\delta}_H(R, T) = \max_{r \in R} \min_{t \in T} \|r - t\| \quad (1.8)$$

et la distance de Hausdorff (avec la propriété de symétrie) d_H s'écrit :

$$d_H(R, T) = d_H(T, R) = \max(\hat{\delta}_H(R, T), \hat{\delta}_H(T, R)) \quad (1.9)$$

Cette distance permet de déterminer la plus grande distance entre les deux segmentations qui sera assimilée à la plus grande erreur de segmentation.

Pour obtenir une mesure globale dans le cas multi-classes, il convient de calculer la distance de Hausdorff pour chaque région et de garder le maximum de toutes ces distances.

A.2.4 Distance absolue moyenne, d_m

La distance absolue moyenne, notée MAD ou d_m , entre deux ensembles R et T s'exprime ainsi :

$$d_m(R, T) = MAD(R, T) = \frac{1}{2} (\bar{d}(R, T) + \bar{d}(T, R)) \quad (1.10)$$

avec $\bar{d}(R, T)$ la pseudo-distance directionnelle de Hausdorff moyenne :

$$\bar{d}(R, T) = \frac{1}{|R|} \sum_{r \in R} \min_{t \in T} \|r - t\| \quad (1.11)$$

L'illustration de cette métrique est représentée sur la figure A4.

Cependant, il faut noter qu'elle n'est pas équivalente à l'aire des erreurs ($|R \cup T - R \cap T|$) puisqu'en chaque point, la distance minimum à l'autre ensemble est déterminée puis accumulée.

Pour le multi-classes, comme pour la distance de Hausdorff, il est recommandé de conserver comme mesure globale, le maximum des distances calculées pour chaque classe.

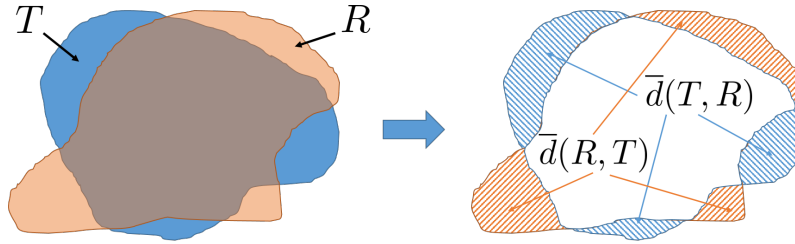


FIGURE A4 – Illustration de la distance absolue moyenne. Sur chacune des régions hachurées, la distance minimum avec l'autre région est accumulée pour le calcul des \bar{d} .

A.2.5 Similarité de volume

Une région obtenue par une méthode de segmentation peut être comparée à la vérité terrain en confrontant simplement leurs volumes. Ainsi, aucune information sur la cohérence spatiale n'est prise en compte par la similarité de volume. C'est une mesure complémentaire aux précédentes.

Cette mesure permet, lorsque les évaluations précédentes sont bonnes, d'interpréter facilement l'erreur de segmentation et décider si elle est acceptable ou non au niveau applicatif. Pour deux régions à comparer, R et T , la similarité de volume est définie :

$$VS(R, T) = 2 \frac{|T| - |R|}{|T| + |R|} \quad (1.12)$$

Si $VS(R, T) = 0$ cela signifie que les volumes sont les mêmes, mais ne garantit pas que R et T soient les mêmes ($DICE(R, T)$ pourrait valoir 0).

Dans le cas du multi-classes avec N régions, la similarité de volume pouvant être positive ou négative, la moyenne des similarités de volumes absolues par région $|VS(R_c, T_c)|$ est utilisée :

$$VS(R, T) = \frac{2}{N} \sum_{c=1}^N \frac{||T_c| - |R_c||}{|T_c| + |R_c|} \quad (1.13)$$



B. Descriptions des bases de données

Cette annexe décrit les deux bases de données MUST et Epaule sur lesquelles s'appuient les développements en segmentation supervisée développés dans ce manuscrit.

B.1 Base de données MUST

Le projet MUST est un projet international dont le but est d'étudier l'effort physiologique d'athlètes courant un ultra marathon. Il est conduit en grande partie par Pierre Croisille et Magalie Viallon (CHU Saint Etienne). Un objectif du projet MUST est de quantifier la réponse inflammatoire des muscles squelettiques dans des conditions extrêmes qui, pour certains sujets, est corrélée à celle de patients en réanimation après un poly-traumatisme ou un infarctus du myocarde. Ces études peuvent aussi être étendues à d'autres pathologies conduisant à une perte de masse des muscles squelettiques (AVC, tumeur, maladies pulmonaires obstructives chroniques, ...)

Cette étude est longitudinale et s'est appuyée à la fois sur des biomarqueurs sanguins et l'imagerie IRM quantitative de la tête, du cœur et des cuisses. Le choix de l'IRM quantitative permet une exploration non-invasive précise des altérations fonctionnelles des muscles squelettiques.

Les 51 athlètes étudiés ont été recrutés parmi les 700 coureurs du *Tor des Géants* 2014 qui est un ultra-marathon de 330km avec 24000m de dénivelé positif, se déroulant dans la vallée d'Aoste et qui doit être couru en 150 heures (6 jours et 6 heures). 27 des 51 volontaires ont fini la course, pourcentage proche de celui constaté pour tous les coureurs.

B.1.1 Imagerie IRM

On détaille ici l'imagerie IRM des cuisses utilisée dans les applications de ce manuscrit. Tous les volontaires ont eu une première IRM avant le départ de la course (notée *Pre*). Puis ceux ayant fini la course ont eu deux autres IRM : une seconde quelques minutes après leur arrivée (notée *Post*), une troisième entre 2 et 3 jours après l'arrivée (notée *Post+3*). Toutes les acquisitions ont été faites sur un scanner mobile de 1.5T (MAGNETOM Avanto, Siemens Healthcare, Erlangen, Allemagne embarqué dans un camion d'Alliance Medical,

Angleterre). Trois acquisitions ont été réalisées : écho gradient 3D isotrope (3D GRE double écho Dixon), écho gradient 3D multi-écho (8 échos) en axial (3D GRE multi-écho) et une séquence spin-écho 2D multi-echos pondérée T2 (T2 Spin Echo).

La figure B1 illustre les différentes régions des cuisses imagées par chacune de ces acquisitions. Les acquisitions 3D GRE multi-écho et T2 Spin Echo sont centrées à 15cm de la partie supérieure de la rotule.

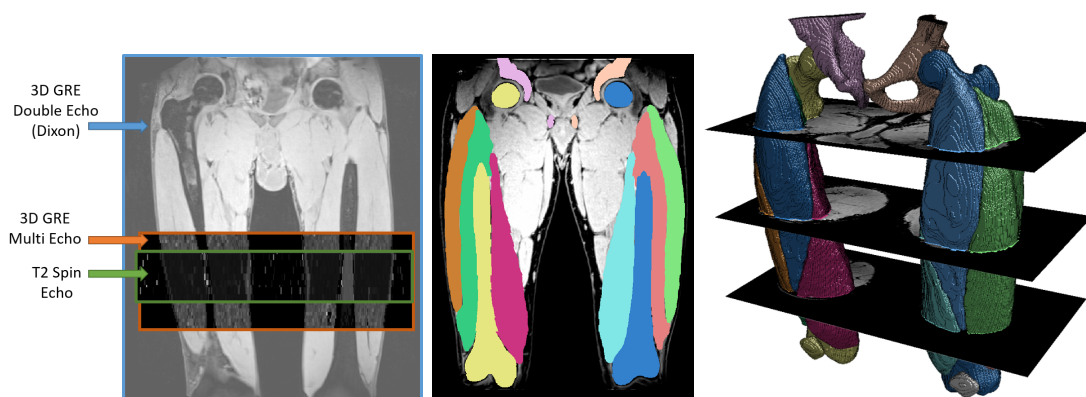


FIGURE B1 – Acquisition IRM (à gauche), segmentation manuelle sur une coupe (au centre) et rendu 3D des segmentations des quadriceps et des os (à droite).

Ces 3 acquisitions ont permis d’obtenir 8 images et cartographies IRM quantitatives dont 5 sont illustrées sur la figure B2.

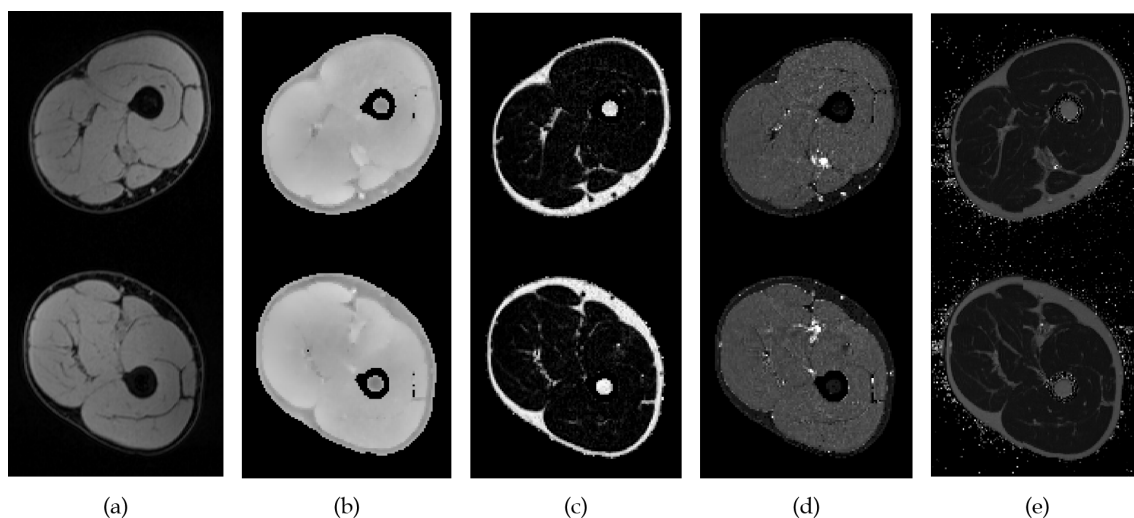


FIGURE B2 – Illustration des 5 types d’images utilisés pour l’étude de l’inflammation des muscles de la cuisse : (a) imagerie isotropique de l’eau, (b) carte de susceptibilité, (c) cartographie PDFF, (d) cartographie $T2^*$ et (e) cartographie T2.

La figure B3 illustre sur l’imagerie pondérée $T2^*$ l’évolution de l’inflammation sur les trois instants Pré, Post et Post+3.

B.1.2 Base de données

La base de données MUST comporte les images IRM de 51 coureurs dont 27 ont fini la course. Après suppression des données incomplètes ou avec artéfacts, on dispose de 42 données IRM complètes dont 20 correspondent à des coureurs ayant fini la course.

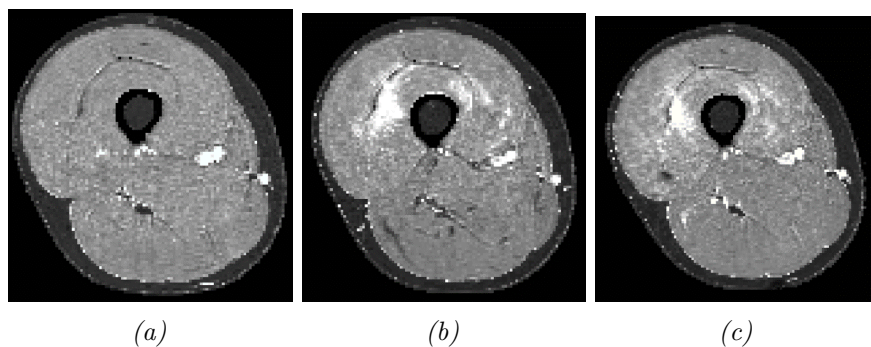


FIGURE B3 – Illustration de l'évolution de l'inflammation des muscles de la cuisse sur les cartographies $T2^*$: (a) image acquise avant la course - Pré, (b) image acquise à l'arrivée du coureur - Post, et (c) image acquise 3 jours après l'arrivée - Post+3.

Plusieurs experts ont ensuite segmenté les quadriceps (figure B4).

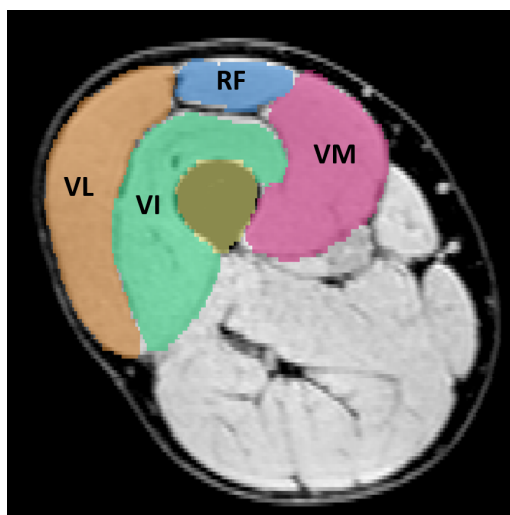


FIGURE B4 – Définition des 4 muscles à segmenter pour l'étude des quadriceps. Les abréviations sont : VM – *vectus medialis*, VL – *vectus lateralis*, VI – *vectus intermedius*, RF – *rectus femoris*.

Les jambes droites de 7 coureurs ayant fini la course, ont été segmentées en interpolant (et au besoin corrigeant) les segmentations manuelles effectuées toutes les 10 coupes. Quatre experts différents ont réalisé ses segmentations. De manière assez similaire, 5 jambes gauches ont été segmentées deux fois par deux experts. La comparaison des segmentations inter-experts donne un score de Dice de 0,910. Les comparaisons sont données en détails dans la tables B1. On note que le vastus intermedius (VI) est plus difficile à délimiter et qu'il y a une certaine in-homogénéité de concordance pour les différents coureurs.

Trois coureurs ont leurs jambes gauche et droite segmentées manuellement. Toutes ces segmentations ont été faites sur le premier instant (Pré : avant la course).

Enfin, pour réaliser une évaluation quantitative longitudinale, 17 coupes de 4 coureurs supplémentaires ont été manuellement segmentées sur les 3 instants Pré, Post et Post+3.

B.1.3 Pré-traitements

Les images IRM ont été corrigées de l'in-homogénéité de champs par l'algorithme N4 (TUSTISON et al., 2010) puis centrées et réduites avant d'être remises sur la dynamique d'un des examens.

	<i>ALB-2725</i>	<i>BRG-1924</i>	<i>CAL-4223</i>	<i>MAV-526</i>	<i>YAG-47</i>	Moyenne
DSC	.924	.877	.897	.921	.934	.910
VL	.920	.874	.906	.918	.944	.912
RF	.939	.916	.941	.940	.962	.940
VM	.945	.892	.882	.942	.934	.919
VI	.891	.825	.857	.882	.894	.870
VS	.042	.057	.054	.074	.025	.051
VL	.025	.100	.065	.049	.003	.049
RF	.025	.061	.012	.073	.021	.038
VM	.004	.023	.091	.022	.073	.043
VI	.115	.043	.050	.155	.005	.074

TABLE B1 – Evaluation par les scores de DSC et de similarité de volume (VS) des segmentations de 5 jambes gauches réalisées par 2 experts. Abréviations : VL – Vastus Lateralis, RF – Rectus Femoris, VM – Vastus Medialis, VI – Vastus Intermedius.

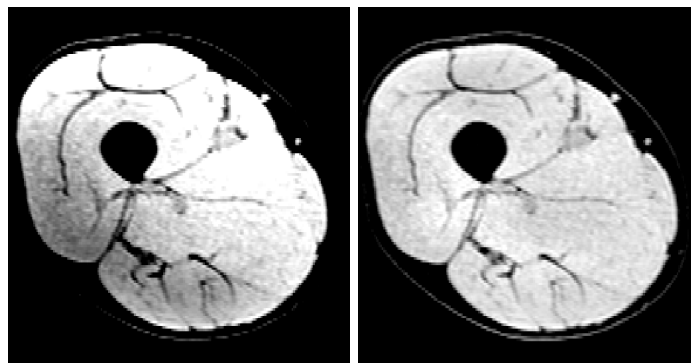


FIGURE B5 – Correction du biais avec l’algorithme N4. Image originale à gauche et corrigée à droite.

Un recalage des modalités intra-patient a aussi été fait (figure B6) pour garantir une extraction des caractéristiques radiomiques (VALLIÈRES et al., 2015) cohérentes¹.

Les résultats des différents recalages testés (rigide et déformable) sont donnés dans la figure B7. Notre étude (NGUYEN et al., 2021a) a conclu que le recalage déformable était le plus pertinent au sens de la mise en correspondance de l’anatomie, mais que les recalages ne modifiaient pas de manière statistiquement significative les valeurs radiomiques extraites à partir des régions 3D des muscles.

Après ces différents pré-traitements, nous obtenons des volumes de $280 \times 160 \times 640$ pixels.

B.2 Base de données Epaule

Cette base de données vient d’une collaboration entre Jean-Baptiste Pialat et l’hôpital de Genève. Cette base de données sur l’épaule a pour objectif d’aider à la décision de prise en charge de patients atteints de lésion du tendon de la coiffe des rotateurs. Pour cela, il est important d’estimer le volume et le pourcentage de graisse dans les différents muscles de l’épaule. L’imagerie IRM quantitative permet ce type de mesure. Pour le traitement d’images, la problématique développée ici est la segmentation précise des muscles.

1. Pour ces recalages nous nous sommes appuyés sur elastix (KLEIN et al., 2010; SHAMONIN et al., 2014)

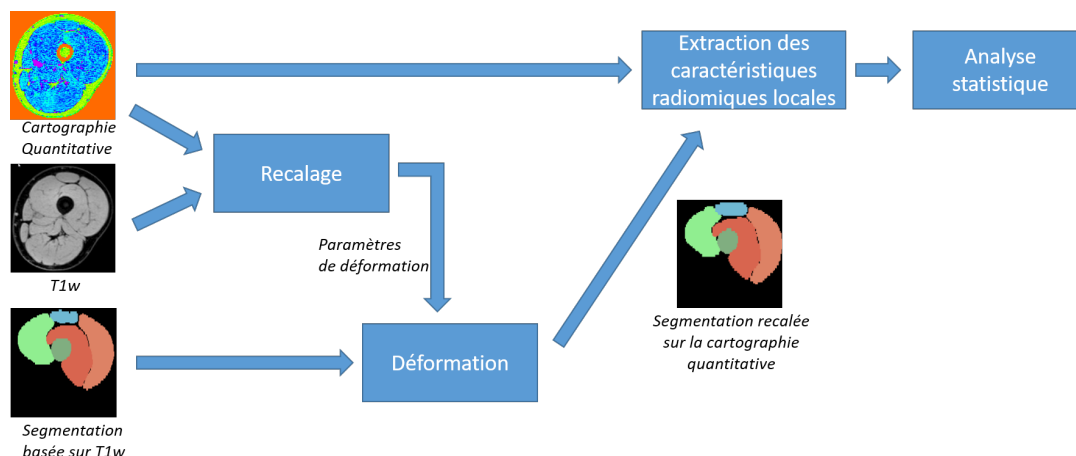


FIGURE B6 – Pipeline proposé pour étudier l'impact du recalage sur l'extraction de caractéristiques radiomiques.

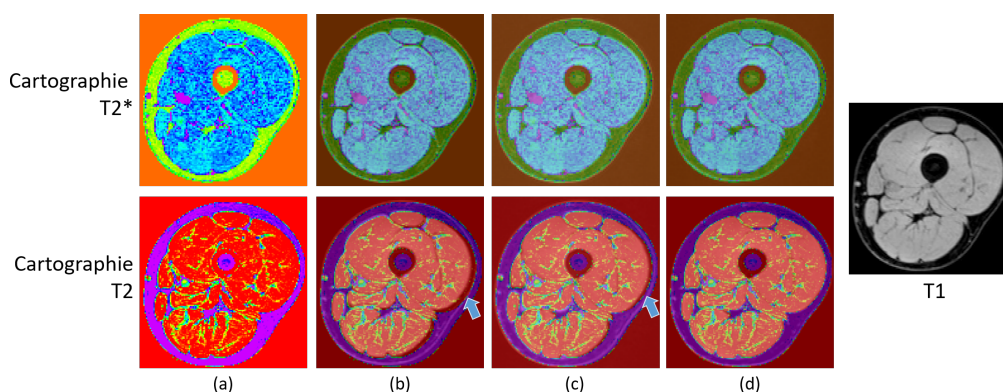


FIGURE B7 – Comparaison des approches de recalage pour la mise en correspondance des images pondérées T1 et des cartographies T2 et T2*. (à droite) image pondérée T1, (a) cartographies T2* et T2, (b) superposition T1 et des cartographies sans recalage, (c) superposition T1 et des cartographies avec recalage rigide, (d) superposition T1 et des cartographies avec recalage déformable

B.2.1 Imagerie IRM

Pour les 50 patients, l'épaule, gauche ou droite, a été imagée avec 12 séquences IRM afin d'obtenir les cartographies quantitatives notamment de graisse (PDFF) :

1. image pondérée T1,
2. image pondérée T2, suivant 3 plans de référence,
3. Dixon (permet de séparer l'eau et la graisse, basé sur les propriétés chimiques de déplacement de ces molécules) avec 8 temps d'écho différents.

Une illustration de ces acquisitions est donnée sur la figure B8.

Nous nous focaliserons sur les acquisitions Dixon. Initialement, la taille du pixel est de $0,6875mm \times 0,6875mm \times 2mm$ et les images ont une dimension de $320 \times 320 \times 60$ pixels.

B.2.2 Base de données

Sur les 50 patients, 27 ont été manuellement et entièrement segmentés en 3D. 5 muscles sont délimités : le deltoïde (DTD), le supra-épineux (SPR), le petit-rond (TMN), le sus-capulaire (SBC) et l'infra-épineux (IFR). Un exemple de segmentation manuelle est donné sur la figure B9.

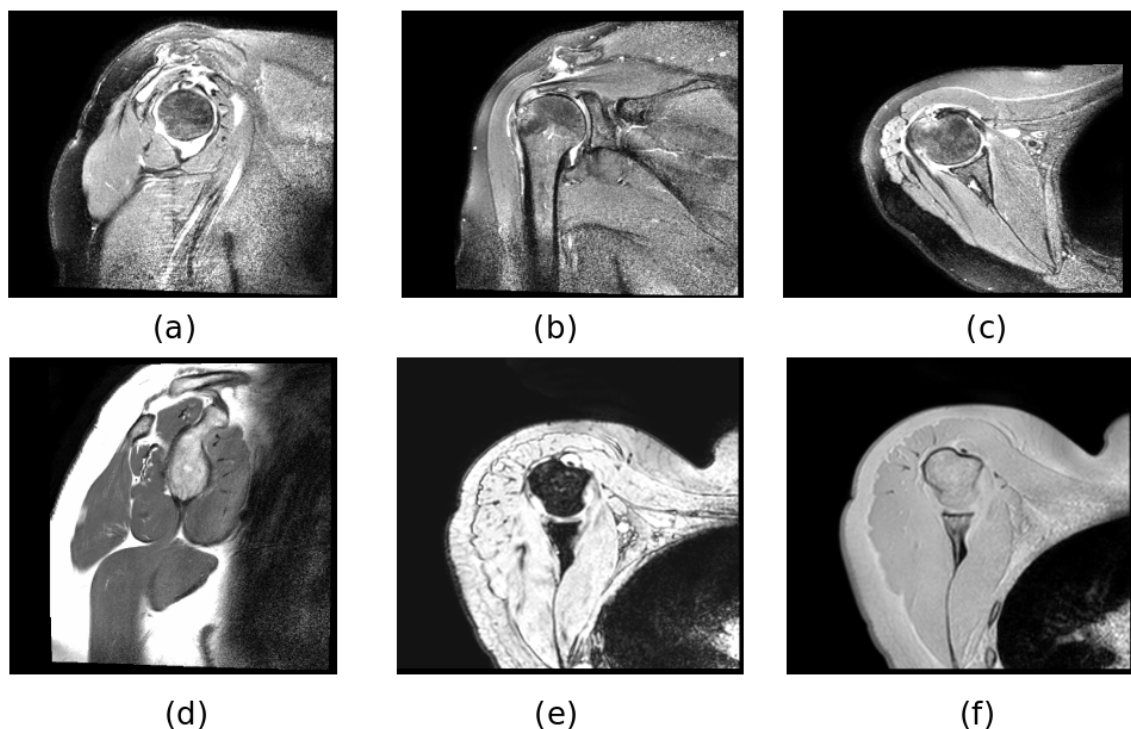


FIGURE B8 – Exemple d’acquisitions IRM de la base de données *Epaule*. (a), (b) et (c) IRM pondérée T2 coupe sagittale, frontale et transverse, (d) IRM pondérée T1 coupe sagittale, (e) IRM Dixon echo 8, (f) IRM Dixon echo 1.

L’annotation manuelle de cette base de données a nécessité les spécifications suivantes afin d’obtenir des segmentations 3D cohérentes :

- bien définir les objets anatomiques (début, fin, frontières,...) par rapport à ce qui est imagé,
- adopter une stratégie en cas d’artéfact ou d’anormalité anatomique,
- pour la 3D, essayer d’assurer une bonne continuité des annotations entre les coupes et les plans anatomiques.



FIGURE B9 – Exemple de segmentation manuelle des 5 muscles de la base de données *Epaule*. Coupe (a) transverse (b) frontale et (c) sagittale avec le deltoïde (DTD) en orange, le supra-épineux (SPR) en jaune, le petit-rond (TMN) en bleu, le sus-capulaire (SBC) en vert et l’infra-épineux (IFR) en marron.

B.2.3 Pré-traitements

Les pré-traitements de cette base de données consistent à homogénéiser les données : correction de l'in-homogénéité de champs et de l'orientation des volumes, ré-échantillonnage en pixel isotrope, modification du champs de vue, recalage sur un examen de référence, transformation de toutes les épaules en épaules droites. Ces pré-traitements sont décrits dans la figure B10.

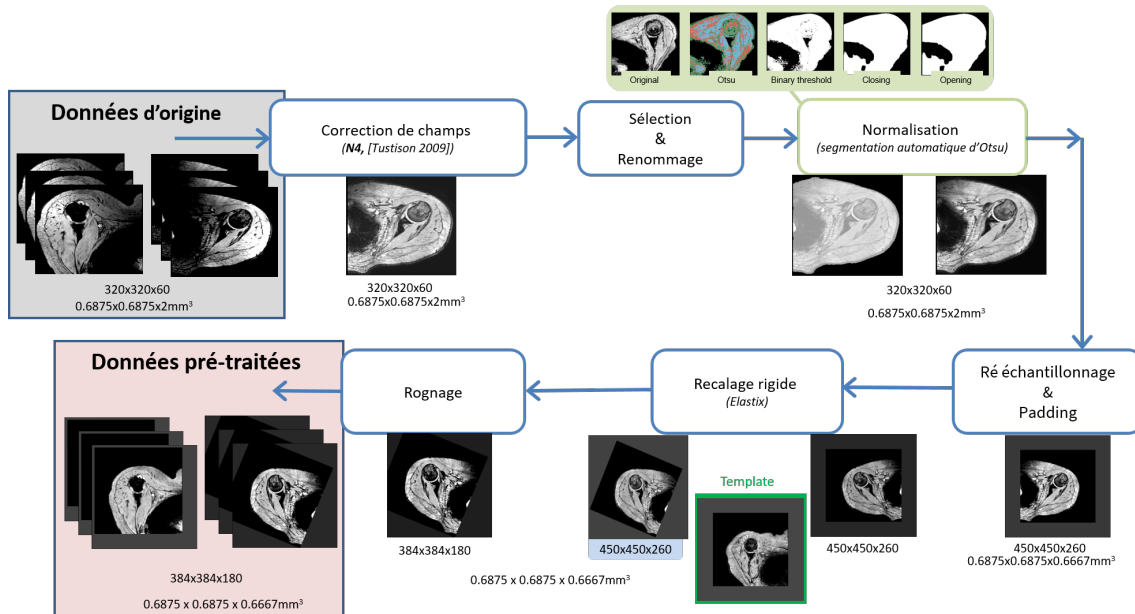


FIGURE B10 – Illustration des différents pré-traitements effectués sur la base de données Epaule afin d'obtenir des images homogènes et correspondant à des épaules droites.

Les modifications géométriques sont ensuite appliquées aux segmentations manuelles et aux autres données IRM.



C. Description de l'algorithme AdaBoost

AdaBoost (*Adaptive Boosting*) (FREUND et al., 1996) est l'un des algorithmes d'apprentissage d'ensemble de méthodes les plus célèbres. Il appartient à la famille de méthodes appelées *boosting* qui fonctionnent avec de multiples algorithmes d'apprentissage simple (ou des apprenants faibles) et tentent, en les assemblant, de faire passer leurs performances de faibles à fortes. Un apprenant faible est à minima juste meilleur qu'une estimation aléatoire (le taux d'erreur ϵ_t est légèrement inférieur à 50%) tandis qu'un apprenant fort a un taux d'erreur quasi nul.

L'algorithme d'Adaboost fonctionne ainsi. A chaque itération t , un apprenant faible h_t est choisi pour classer au mieux (taux d'erreur ϵ_t le plus petit) les m échantillons, où chaque échantillon est pondéré par \mathcal{D}_t . La distribution des poids \mathcal{D}_t des m échantillons est mise à jour en augmentant les poids des échantillons mal classés (et inversement pour les bien classés). Cet algorithme s'arrête après un nombre d'itérations T donné. L'apprenant fort est formé à partir de tous les apprenants faibles h_t obtenus lors des itérations. Ils sont pondérés par une fonction quasi inversement proportionnel au taux d'erreur ϵ_t .

L'algorithme 5 détaille AdaBoost. Les apprenants faibles sont souvent un seuillage suivant une des dimensions. Ce type d'algorithme est suffisamment simple pour éviter de trop s'adapter aux données d'entraînement (*over-fitting*).

Algorithme 5 Algorithme AdaBoost (classification binaire) (ZHOU, 2012)

Entrée: Données $D = \{(X_1, y_1), (X_2, y_2), \dots, (X_m, y_m)\}$

Entrée: Base d'algorithmes faibles de classification \mathcal{L}

Entrée: Nombre d'itérations d'apprentissage T

1: $\mathcal{D}_1(x) = 1/m$ {initialisation des pondérations des échantillons}

2: **for** $t = 1, \dots, T$ **do**

3: $h_t = \mathcal{L}(D, \mathcal{D}_t)$ {entraîner l'algo. de classification h_t avec D et \mathcal{D}_t }

4: $\epsilon_t = P_{x \sim \mathcal{D}_t}(h_t(x) \neq f(x))$ {Evaluer l'erreur de h_t }

5: **if** $\epsilon_t > 0.5$ **then**

6: **break**

7: **else**

8: $\alpha_t = \frac{1}{2} \ln\left(\frac{1 - \epsilon_t}{\epsilon_t}\right)$; {constante de pondération pour l'algorithme h_t }

9:

$$\mathcal{D}_{t+1}(x) = \frac{\mathcal{D}_t(x)}{Z_t} \times \begin{cases} \exp(-\alpha_t), & \text{if } h_t(x) = f(x) \\ \exp(\alpha_t), & \text{if } h_t(x) \neq f(x) \end{cases}$$

{Mise à jour des poids des échantillons, avec Z_t un facteur de normalisation de \mathcal{D}_{t+1} }

10: **end if**

11: **end for**

12: **return** $H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right)$



Bibliographie de la conclusion et des annexes

- BUDD, S., E. C. ROBINSON et B. KAINZ (2021). « A survey on active learning and human-in-the-loop deep learning for medical image analysis ». In : *Medical Image Analysis* 71, p. 102062.
- DIAZ-PINTO, A., S. ALLE, A. IHSANI, M. ASAD, V. NATH, F. PÉREZ-GARCÍA, P. MEHTA, W. LI, H. R. ROTH, T. VERCAUTEREN, D. XU, P. DOGRA, S. OURSELIN, A. FENG et M. J. CARDOSO (2022a). « MONAI Label : A framework for AI-assisted Interactive Labeling of 3D Medical Images ». In : *arXiv e-prints*.
- DIAZ-PINTO, A., P. MEHTA, S. ALLE, M. ASAD, R. BROWN, V. NATH, A. IHSANI, M. ANTONELLI, D. PALKOVICS, C. PINTER et al. (2022b). « DeepEdit : Deep Editable Learning for Interactive Segmentation of 3D Medical Images ». In : *MICCAI Workshop on Data Augmentation, Labelling, and Imperfections*. Springer, p. 11-21.
- DICE, L. R. (1945). « Measures of the amount of ecologic association between species ». In : *Ecology* 26.3, p. 297-302.
- DOSSELMANN, R. et X. D. YANG (2011). « A comprehensive assessment of the structural similarity index ». In : *Signal, Image and Video Processing* 5.1, p. 81-91.
- FREUND, Y. et R. E. SCHAPIRE (1996). « Experiments with a new boosting algorithm ». In : *Proceedings of the Thirteenth International Conference on International Conference on Machine Learning*. ICML'96. Bari, Italy : Morgan Kaufmann Publishers Inc., 148-156.
- KIRILLOV, A., K. HE, R. GIRSHICK, C. ROTHER et P. DOLLAR (2019). « Panoptic segmentation ». In : *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- KLEIN, S., M. STARING, K. MURPHY, M. A. VIERGEVER et J. P. PLUIM (2010). « elastix : a toolbox for intensity-based medical image registration ». In : *IEEE Transactions on Medical Imaging* 29.1, p. 196-205.
- MONGA, V., Y. LI et Y. C. ELDAR (2021). « Algorithm Unrolling : Interpretable, Efficient Deep Learning for Signal and Image Processing ». In : *IEEE Signal Processing Magazine* 38.2, p. 18-44.

- NATH, V., D. YANG, B. A. LANDMAN, D. XU et H. R. ROTH (2021). « Diminishing Uncertainty Within the Training Pool : Active Learning for Medical Image Segmentation ». In : *IEEE Transactions on Medical Imaging* 40.10, p. 2534-2547.
- REINKE, A. et al. (2021). *Common limitations of image processing metrics : A picture story*. arXiv : 2104.05642 [cs.CV].
- REN, P., Y. XIAO, X. CHANG, P.-Y. HUANG, Z. LI, B. B. GUPTA, X. CHEN et X. WANG (2021). « A Survey of Deep Active Learning ». In : *ACM Comput. Surv.* 54.9.
- SHAMONIN, D. P., E. E. BRON, B. P. LELIEVELDT, M. SMITS, S. KLEIN et M. STARING (2014). « Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease ». In : *Frontiers in Neuroinformatics* 7.50, p. 1-15.
- TAHA, A. A. et A. HANBURY (2015). « Metrics for evaluating 3D medical image segmentation : analysis, selection, and tool ». In : *BMC medical imaging* 15, p. 29.
- TUSTISON, N. J., B. B. AVANTS, P. A. COOK, Y. ZHENG, A. EGAN, P. A. YUSHKEVICH et J. C. GEE (2010). « N4ITK : Improved N3 bias correction ». In : *IEEE Transactions on Medical Imaging* 29.6, p. 1310-1320.
- VALLIÈRES, M., C. R. FREEMAN, S. R. SKAMENE et I. EL NAQA (2015). « A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities ». In : *Physics in Medicine and Biology* 60.14, p. 5471-5496.
- WANG, Z. et A. C. BOVIK (2009). « Mean squared error : Lot it or leave it ? A new look at signal fidelity measures ». In : *IEEE Signal Processing Magazine* 26.1, p. 98-117.
- WANG, Z., A. C. BOVIK et L. LU (2002). « Why is image quality assessment so difficult ? » In : *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*. T. 4. Orlando, FL, United states : Institute of Electrical et Electronics Engineers Inc., p. IV/3313-IV/3316.
- WANG, Z., A. BOVIK, H. SHEIKH et E. SIMONCELLI (2004b). « Image quality assessment : from error visibility to structural similarity ». In : *IEEE Transactions on Image Processing* 13.4, p. 600-612.
- WANG, Z., E. WANG et Y. ZHU (2020). « Image segmentation evaluation : a survey of methods ». In : *Artificial Intelligence Review* 53, p. 5637-5674.
- ZHANG, R., P. ISOLA, A. A. EFROS, E. SHECHTMAN et O. WANG (2018). « The unreasonable effectiveness of deep features as a perceptual metric ». In : *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA : IEEE Computer Society, p. 586-595.
- ZHOU, Z.-H. (2012). *Ensemble methods : Foundations and algorithms*. 1st. Chapman & Hall/CRC.