Méthodes de segmentation d'images basées sur l'apprentissage profond dans le traitement des tumeurs bénignes et malignes de l'utérus

Chen Zhang

COLLEGE MATHS, TELECOMS
DOCTORAL INFORMATIQUE, SIGNAL
BRETAGNE SYSTEMES, ELECTRONIQUE

Université de Rennes

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE RENNES

ÉCOLE DOCTORALE N° 601
Mathématiques, Télécommunications, Informatique, Signal, Systèmes, Électronique
Spécialité : *Signal, Image, Vision*

Par
Chen Zhang

Deep learning-based image segmentation methods in the treatment of benign and malignant uterine tumor diseases

Thèse présentée et soutenue à Southeast University, Nanjing, China, le 26 mai 2023
Unité de recherche : LTSI, UMR INSERM U 1099 and LIST, Southeast University, Nanjing, China

Rapporteurs avant soutenance :

Julien BERT, Ingénieur de Recherche, HdR, Latim, INSERM, CHU Brest
GUI Zhiguo, Professeur, School of Information and Communication Engineering, North University of China, Taiyuan, China

Composition du Jury :

Président :	SHANGYuanyuan	Professeur, Capital Normal University, Beijing, China
Examinateurs :	Julien BERT	Ingénieur de Recherche, CHU Brest
	GUI Zhiguo	Professeur, North University of China, Taiyuan, China
	SHANG Yuanyuan	Professeur, Capital Normal University, Beijing, China
	Antoine SIMON	Maître de Conférences, HdR, Université de Rennes
Dir. de thèse :	Jean-Louis DILLENSEGER	Professeur, Université de Rennes
Co-dir. de thèse :	SHU Huazhong	Professeur, SouthEast University, Nanjing, China

ACKNOWLEDGEMENTS

This dissertation was conducted in the frame of CRIBs(Centre de Recherche en Information Biomédicale sino-français), which is an international associate French-Chinese laboratory (Université de Rennes 1 – France, INSERM – France, Southeast University – China).

Looking back, I recall everything since I came to the imaging lab of Southeast University in the summer when I was 22 years old. I am grateful to my lovely classmates and amiable faculty members, and I sincerely thank everyone who accompanied and supported me along the way.

First of all, I would like to thank my supervisor, Prof. Huazhong Shu, for his rigorous and well-coached approach, for his guidance when I encountered difficulties and anxieties in my research, for his tireless work attitude greatly influenced my attitude towards research. He is funny, generous and loving. I am grateful to Mr. Shu for being such a good teacher in my doctoral life and for his tolerance and care for me. I would also like to thank my French supervisors, Jean-Louis DILLENSEGER and Antoine SIMON. I was still anxious before leaving for France because of the covid. However, I was very touched by the care and concern of the two teachers, and I still remember the illustrated transportation guide drawn up for me before my departure in 2020, and the warm company of Miss. Bertille. After that, the two teachers always asked me if I had any difficulties, so I also felt at home in the LTSI lab. In terms of my studies, my two teachers provided me with guidance and revision of my projects and papers through online and offline meetings. In addition, I have had many opportunities to present myself, whether it was at the PhD Day when I first arrived, the annual CAMI conference, or the summer school last June, all of which gave me the challenge of academic exchange!

Secondly, I would like to express my gratitude to Mr. Guan Yu Yang for his helpful guidance in my paper writing and slide-making. His meticulous work, attention to experimental details, and kind and modest personality have significantly influenced me. I would also like to thank Mr. Youyong Kong, Mr. Jiasong Wu, and Mr. Chunfeng Yang in the lab for their generous and gentle advice on both my studies and personal life.

I am also thankful to the other faculty members in the LTSI lab, Soizic Charpentier and Patricia Bernabé, for practicing French with me and for their kindness and support. I am grateful to Professor Lotfi Senhadji for his care and encouragement. Additionally, my colleague Alban has helped me greatly during my time in France, and I hope our friendship will endure. I want to express my appreciation to my friends Meng Chen, Fuzhi Wu, and Qixiang Ma. We were lucky enough to travel together from Nanjing to Rennes and study together at both Southeast University and Rennes University. Their companionship has been invaluable to me, and I am

grateful for their unwavering support.

I would like to express my heartfelt gratitude to all of my friends in the LIST lab, including Dr. Jie Liu, Xingran Zhao, Rongjun Ge, Haichen Zhu, Yunbo Gu, Yuan Gao, Yuting He, and Weiya Sun, for their invaluable assistance. I also extend my gratitude to Dr. Yuncheng Hua, Dr. Na Yang, and Dr. Mengke Xu in the Ph.D. program, and I wish them all the best in their research endeavors. I am grateful to my dear friends Ziwei Lu, Rongqing Xia, and Fangsu Fan, who have been by my side for many years, for their unwavering support and encouragement throughout my journey.

Lastly, I want to express my deepest appreciation to my parents. Thank you for your constant love, support, and encouragement and for always believing in me. I am grateful to have grown up in such a happy and nurturing family, and I could not have come this far without your guidance and trust. You are my favorite people and my favorite friends!

Résumé étendu : Guidage par l'image pour les traitements des tumeurs malignes et bénignes de l'utéru

Contexte de l'étude

Les travaux de cette Thèse portent sur le guidage par l'image de traitements de tumeurs bénignes et malignes de l'utérus.

L'utérus est un organe appartenant à l'appareil génital féminin. Plusieurs affections peuvent atteindre l'utérus et parmi elles, les tumeurs utérines. Ces tumeurs peuvent être bénignes ou malignes (cancéreuses). Nous nous intéresserons en particulier à deux types de tumeurs : le fibrome utérin une tumeur bénigne, représentée par les fibromes, et les tumeurs malignes, représentées par le cancer du col de l'utérus.

Traitements des fibromes utérins

Les fibromes utérins sont de petites tumeurs bénignes qui se développent au niveau de l'utérus. Ils sont très fréquents (ils concernent environ 1 femme sur 3 en âge de procréer en Europe) Ils n'évoluent pas en cancer, mais peuvent entraîner des symptômes gênants (saignements importants, douleur, envies fréquentes d'uriner,...), voire des problèmes de stérilité. En l'absence de symptômes, une simple surveillance régulière suffit. Par contre, si des symptômes existent, et selon la fréquence et la gravité de ceux-ci, des traitements peuvent être proposés en fonction de la taille du fibrome, de sa localisation, de l'âge de la patiente (ménopausée ou non) et de son désir d'avoir un enfant : médicaments, intervention chirurgicale, embolisation des artères utérines, **ablation par ultrasons focalisés de haute intensité (HIFU).** Notre travail de Thèse va concerner cette dernière catégorie de traitements qui présente l'énorme avantage d'être non-invasive et compatible avec le désir d'enfant. En deux mots, le diagnostic et la localisation des fibromes sont génériquement menés sur des volumes d'Imagerie par Résonance Magnétique (IRM) acquis en mode préopératoire. Sur cette IRM, le chirurgien délimite la zone de la lésion à traiter, mais également les zones à risques c'est-à-dire les organes environnants

à préserver (tissus sains de l'utérus, moelle épinière, ...). La thérapie se déroule ensuite de la façon suivante (Figure 1) : la patiente est allongée à plat ventre sur la machine de traitement. Cette machine contient deux dispositifs à base d'ultrasons : un transducteur HIFU qui focalise l'énergie ultrasonore vers la cible ce qui entraîne une hausse rapide de la température au point focal qui nécrose le tissu en ce point par coagulation et un dispositif d'imagerie échographique qui permet de guider le point focal de la sonde HIFU sur la zone à traiter. Comme le fibrome n'est pas visible sur l'image échographique, le médecin s'aide également de l'IRM préopératoire pour le localiser (écran à droite de la Figure 1).

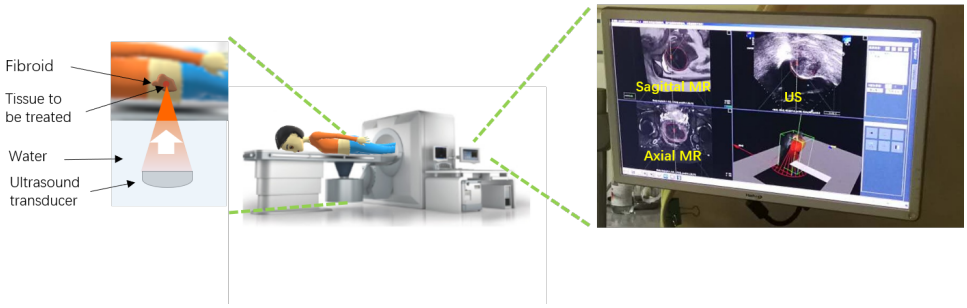

Figure 1 – Thérapie des fibromes utérins par HIFU.

Un des points clés de la thérapie concerne son planning et plus particulièrement la description de l'anatomie spécifique à la patiente à partir de l'IRM préopératoire de diagnostic, c'est-à-dire à la délinéation de l'utérus et des fibromes mais également d'autres organes à risque telle la colonne vertébrale (moelle épinière),... Actuellement cette délinéation est faite manuellement par un médecin. C'est une opération laborieuse, coûteuse en temps, dépendante de l'expertise de la personne et sujette à une grande variabilité de résultats entre experts, voire pour deux occurrences d'un même expert. Une délinéation (une segmentation) automatique est par contre actuellement difficile à réaliser du fait de a) des grandes variabilités de forme et de taille de l'utérus et les fibromes d'une patiente à l'autre; b) un faible contraste entre les organes et tissus adjacents. Le contraste entre l'utérus et les fibromes utérins est par exemple, assez faible, de sorte que les limites entre les organes sont difficiles à distinguer ; c) du nombre de fibromes et leurs formes qui sont inconnus. Une méthode de segmentation automatique et précise capable d'extraire toutes ces structures serait d'une grande importance pour le planning de la thérapie et le traitement lui-même. Récemment, l'apprentissage profond (Deep Learning) a réalisé d'énormes progrès dans la segmentation des images médicales. Ces méthodes basées sur l'apprentissage entièrement supervisé (FSL) peuvent traiter diverses tâches de segmentation d'images médicales. Cependant, la précision et la robustesse des méthodes d'apprentissage profond dépendent d'un grand nombre de données d'apprentissage annotées par des experts. L'acquisition de bonnes

annotations précises nécessite un travail laborieux, et les résultats de la délimitation inter-experts varient. Nous nous sommes intéressés à la segmentation du volume pelvien à l'aide d'apprentissage profond selon deux approches :

1) **L'adaptation d'une méthode classique basée sur de l'apprentissage supervisé pour la segmentation multi-classes des régions utérines à partir d'images IRM.** Cette méthode sera décrite dans le chapitre 3. 2) Par contre, comme nous sommes dans le domaine des données cliniques, il est extrêmement difficile d'obtenir la très grande quantité de données annotées nécessaires aux méthodes d'apprentissage profond entièrement supervisées. Nous allons donc nous concentrer sur l'utilisation d'un nombre limité de données annotées pour l'apprentissage tout en obtenant des performances de segmentation similaires à celles des méthodes entièrement supervisées. Nous avons donc développé **une nouvelle méthode semi-supervisée pour cette segmentation multi-classes des régions utérines à partir d'images IRM.** Cette méthode utilise un petit nombre de données étiquetées pour établir un premier modèle. Ensuite, des données non étiquetées sont introduites dans le modèle et co-entraînées avec les données étiquetées pour rendre le modèle plus efficace. Cette méthode sera décrite dans le chapitre 4.

Traitements des cancers du col de l'utérus

Le cancer du col de l'utérus est la quatrième tumeur maligne féminine la plus fréquente dans le monde, avec plus de 500 000 femmes diagnostiquées chaque année et la maladie causant plus de 300 000 décès dans le monde. Suite à un diagnostic, différents traitements (seuls ou associés) peuvent être proposés en fonction du stade du cancer, de l'âge et de l'état général de la patiente, de son désir d'avoir un enfant : chimiothérapie , intervention chirurgicale et/ou radiothérapie. Le cadre médical de notre travail concerne la **radiothérapie adaptative des cancers du col de l'utérus.** En deux mots, la radiothérapie adaptative vise à calculer la dose optimale à délivrer, en direct, séance après séance, en fonction de l'imagerie et des modifications de positionnement, de forme ou de volume de la tumeur et des organes adjacents. La radiothérapie adaptative permet une irradiation à haute dose et de haute précision de la zone cible de la tumeur tout en réduisant l'irradiation des tissus normaux environnants afin de minimiser la toxicité. Avec le développement de la technologie, la radiothérapie a été mise en œuvre dans la pratique clinique sur de nombreuses cibles thérapeutiques, notamment la tête et le cou, le poumon, la prostate, la vessie et, dans notre cas, le col de l'utérus. Nous nous placerons dans le cas de la thérapie adaptative offline, c'est-à-dire qui adapte le plan de thérapie quand nécessaire. C'est généralement en début de chaque séance. En effet, les changements de remplissage de la vessie et du rectum affectent grandement la position spatiale de l'utérus et donc souvent entraînent des erreurs dans la délivrance de la dose. L'émergence de la radiothérapie guidée par l'image (CBCT ou plus récemment l'IRM) a rendu possible la visualisation de la morphologie

des tissus mous pendant la radiothérapie. Grâce à cette méthode, il est possible de surveiller les changements dans la vessie, l'intestin et le rectum, garantissant ainsi une dose de rayonnement élevée dans la zone cible. La figure 2 illustre le déroulement de l'administration en ligne d'une radiothérapie guidée par CBCT avec replanification quotidienne. Il se compose de deux parties : la planification et le traitement. Dans la phase de planification, le plan de radiothérapie initial est généré sur la base des contours des scanners X (CTs) de planification . Dans la phase de traitement, le plan de traitement est constamment optimisé en fonction des informations sur la taille et l'emplacement de la tumeur et des organes à risque (obtenues grâce à l'acquisition continue d'images anatomiques CBCT avant chaque traitement). Dans notre cas particulier nous nous sommes basés sur la stratégie de replanification proposée au LTSI [1]. Cette stratégie est basée sur : 1) la génération d'une bibliothèque de plans de traitements, comprenant plusieurs plans de traitements optimisés sur la base de plusieurs CT de planification acquis avec différents remplissages de la vessie (Figure 2 – gauche) ; 2) à chaque fraction de traitement, l'acquisition de l'image CBCT du jour (Figure 2 – haut) ; 3) la sélection du plan de traitement du jour le plus approprié pour maximiser la couverture de la cible. Cette sélection se fait par la mise en correspondance entre l'image CBCT et la bibliothèque de plans de traitements (Figure 2 – bas).

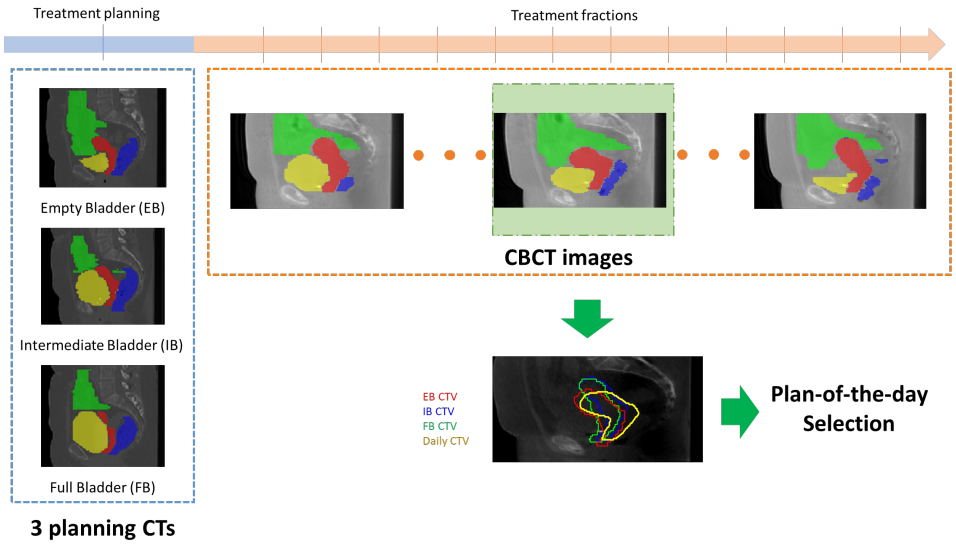

Figure 2 – Workflow de la radiothérapie pour le cancer du col de l'utérus. Le processus se compose de trois étapes (1) Planification : acquisition de plusieurs tomodensitométries de planification avec des volumes de vessie à différents stades de remplissage. (2) Acquisition de l'image CBCT du jour. (3) Sélection du plan de traitement le plus approprié pour maximiser la couverture de la cible. L'utérus, la cavité abdominale, la vessie et le rectum sont représentés par des contours de rouge, vert, jaune et bleu.

Nous nous sommes intéressées à la mise en place de la dernière étape de cette stratégie c'est-à-dire le sélection du plan de traitement du jour le plus approprié à partir du CBCT

du jour et la bibliothèque de plans de traitements. Une des préalables en est la segmentation automatisée des organes dans le CBCT. Actuellement, en routine clinique, comme la délinéation manuelle des images CBCT pour chaque patient est en fait peu pratique, le praticien effectue une sélection manuelle du plan du jour pour chaque fraction par seule comparaison visuelle. Ce processus fastidieux limite les apports d'une radiothérapie adaptative. Or l'image CBCT joue un rôle important, car elle fournit les dernières informations anatomiques sur le patient ou le repositionnement du patient. Cependant, la qualité des images CBCT est relativement faible en raison du bruit, des artefacts et du faible contraste des tissus mous. Ces problèmes rendent l'annotation manuelle difficile et chronophage. Par conséquent, la segmentation automatique des images CBCT pour la sélection du plan du jour est essentielle pour la radiothérapie. Nous proposons un processus automatique permettant de sélectionner le plan de traitement optimal. Il s'appuie sur une segmentation des images CBCT basée sur l'apprentissage profond. Le but est d'ensuite de sélectionner le plan de traitement optimal qui maximise la couverture de l'utérus par le traitement. Cette méthode sera décrite dans le chapitre 5.

Segmentation multi-classes des régions utérines par apprentissage profond supervise : HIFUNet

L'extraction automatique et précise des structures nécessaires au planning et au guidage de la thérapie des fibromes par HIFU à partir des images IRM est particulièrement complexe car : l'utérus et certains des organes périphériques (vessie, rectum) sont extrêmement déformables ; il y a de très grandes variations de forme et de taille d'une patiente à l'autre ; le contraste en IRM entre l'utérus et les fibromes utérins est assez faible, de sorte que les limites entre les organes sont difficiles à distinguer ; et, le nombre de fibromes utérins et leur forme sont inconnus. Ainsi, développer une méthode par apprentissage profond est un véritable défi. Parmi les différentes méthodes proposées dans la littérature, les modèles basés sur des architectures en encodeur-decodeur avec saut de connexion du type U-Net sont très performants. En effet, cette architecture avec sauts de connexions perte permet de fusionner les caractéristiques à différentes échelles et d'améliorer la précision des résultats du modèle. Par contre la très grande variabilité des formes et de leurs tailles et le nombre inconnu de fibromes demandent de très grands champs réceptifs pour capturer les caractéristiques de l'image. Ceci nous a amené à modifier le schéma général d'un U-net sous la forme d'un nouveau réseau appelé HIFUNet pour segmenter automatiquement l'utérus, les fibromes utérins et la colonne vertébrale. Les principales contributions de la méthode peuvent être résumées comme suit : 1) Pour remédier aux erreurs de segmentation (par exemple la mùauvaise classification du col utérin comme fibrome utérin en raison d'un champ réceptif insuffisant), nous introduisons un module de réseau convolutif global (global convolutional network - GCN) capable d'élargir le champ réceptif de manière efficace.

2) Nous intégrons le réseau convolutif global et les convolutions profondes de type ‘’atrous multiples” (deep multiple atrous convolutions – DMAC pour extraire davantage d’informations sémantiques basées sur le contexte et générer des caractéristiques plus abstraites pour les fibromes utérins de grande taille.

De manière plus précise, comme annoncé précédemment, HIFUNet est basée sur une structure de réseau convolutionnel global de type encodeur-décodeur mais avec les particularités suivantes (Figure 3). En fait, HIFUNet se compose de trois parties principales : 1) un module d’encodage des caractéristiques (basé sur un backbone ResNet101 pré-entraînée), 2) une partie d’extraction des caractéristiques (avec le réseau de convolution global et les convolutions profondes atrous multiples) et 3) un module de décodage des caractéristiques.

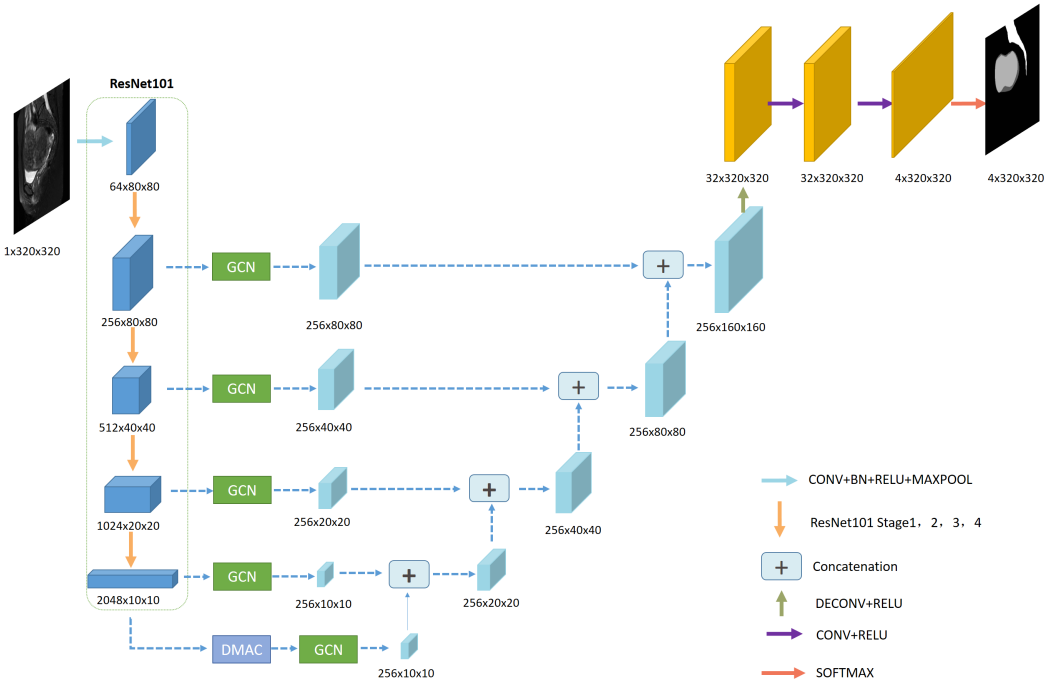

Figure 3 – Architecture de HIFUNet : le réseau se compose d’un backbone Resnet101 en tant que module d’encodage, d’un module GCN et d’un module DMAC en tant que partie extracteur de caractéristiques, et de couches de suréchantillonnage, de couches de concaténation et d’une couche de sortie en tant que partie du module décodeur de caractéristiques. Les paramètres et les tailles des caractéristiques de sortie dans les différentes couches sont présentés dans des couleurs différentes.

Le module encodeur des ResNet-101 pré-entraînés. Dans [2], les auteurs ont démontré que l’utilisation de connexions résiduelles favorise la propagation de l’information à la fois vers l’avant et vers l’arrière, ce qui permet d’améliorer considérablement la vitesse d’apprentissage et les performances. Dans notre cas nous avons simplement transformé les 3 canaux (RGB) de ResNetr en un seul canal (niveau de gris).

Dans la partie extraction de caractéristiques, deux principes sont utilisés pour augmenter les champs réceptifs : 1) La tendance actuelle en matière de conception d'architecture est à l'empilement de petits noyaux de convolution, car cette option est plus efficace que l'utilisation d'un gros noyau de convolution pour la même quantité de calcul. Cependant, compte tenu du fait que les tâches de segmentation sémantique nécessitent une prédiction de la segmentation pixel par pixel, Peng *et al.* [3] ont proposé un réseau convolutif global (GCN) pour améliorer la précision de la classification et de la localisation simultanément. Dans le GCN, une couche entièrement convolutive est adoptée pour remplacer la couche de mise en commun globale afin de conserver les informations de localisation. En outre, de grands noyaux sont introduits pour augmenter le champ réceptif valide. 2) Les convolutions atrous résolvent le problème de la résolution réduite causée par les réseaux neuronaux convolutifs profonds (DMAC) tout en ajustant le champ réceptif du filtre. L'idée principale de la convolution atrous est d'insérer des "trous" (zéros) entre les pixels dans les noyaux de convolution afin d'augmenter la résolution de l'image, permettant ainsi une extraction dense des caractéristiques dans les DMAC. Dans notre cas, nous avons mis en œuvre cinq couches convolutives avec des noyaux 3×3 avec différents taux d'échantillonnage pour extraire les différentes caractéristiques. Enfin, nous fusionnons toutes les caractéristiques avec l'image d'entrée pour générer le résultat final. L'idée derrière cette architecture est d'extraire des caractéristiques multiples et fournir des champs réceptifs de tailles multiples (au détriment du temps de calcul).

Le module de décodage utilise principalement l'opération de concaténation pour fusionner les caractéristiques multi-échelles dans notre cas les sorties de GCN avec les cartes correspondantes de caractéristiques issues de suréchantillonnage. La sortie est assez classique acec une opération de déconvolution pour agrandir l'échelle de l'image jusqu'à la taille initiale et pour restaurer les caractéristiques avec des informations plus détaillées. Enfin, le masque de sortie est obtenu après l'application de deux opérations de convolution et de softmax.

L'apprentissage et la validation du modèle ont été effectués sur une base de données clinique de volumes IRM préopératoires pondérées en T2 avec suppression de graisse de 297 patients et recueillies au First Affiliated Hospital of Chongqing Medical University. Chaque volume d'IRM est constitué de 25 coupes. La vérité de terrain a été générée par un processus d'annotation approprié. Pour garantir une référence clinique objective et cohérente, deux radiologues (un senior et un junior) ont été sollicités et ont définis les annotations après accord consensuel. Les IRMs de 260 patients ont été utilisées pour l'apprentissage et les images des 37 autres patients ont été utilisées pour les tests. La fonction de coût de l'apprentissage était la minimisation de l'entropie croisée entre les résultats de la segmentation et la vérité terrain.

L'évaluation a consisté d'une part à estimer l'apport des différents modules à la segmentation finale. L'utilisation de ResNet101, du GCN et du DMAC a permis à chaque fois d'améliorer les performances du modèle, ceci de manière statiquement démontrable.

Dans un second temps, nous avons comparé les performances en segmentation par rapport à des méthodes classiques (notre modèle a des performances plus élevées statistiquement démontrées) et d'autres méthodes basées sur l'apprentissage profond. Dans ce dernier cas, notre modèle obtenait des résultats légèrement supérieurs et surtout moins sensibles à l'organe segmenté au meilleur des modèles existants (HRNet), mais statistiquement l'apport n'était pas significatif.

En conclusion, HIFUNet donne des résultats similaires aux experts cliniques et il est plus performant que de nombreux réseaux de segmentation sémantique existants.

Notre travail sur la segmentation de l'utérus et des fibromes utérins est, à notre connaissance, la première tentative d'utilisation de réseaux neuronaux convolutifs dans la thérapie HIFU. L'inclusion de la segmentation de la colonne vertébrale, un organe critique dans la thérapie HIFU, est une autre caractéristique majeure de notre approche.

HIFUNet a fait l'objet d'une publication dans IEEE transactions on Medical Imaging [4].

Segmentation semi-supervisée des régions autours de l'utérus à partir d'IRM pour le traitement HIFU

L'étude précédente a utilisé une approche de segmentation de la zone utérine par un modèle d'apprentissage profond complètement supervisé. Ce type de modèle demande un grand nombre de données annotées pour l'apprentissage. L'accès à des données cliniques et plus encore à des données annotées de manière précise par des experts médicaux est souvent extrêmement difficile. En effet l'annotation est un processus répétitif, chronophage et peu valorisant pour les experts médicaux avec pour résultats des données en relativement petit nombre, assez imprécises avec une grande variabilité intra et inter-experts. Ce manque de données en grand nombre et relativement imprécises pose alors des problèmes de généralisation des modèles par apprentissage complètement supervisé avec des phases de réapprentissage an cas de changements d'appareils ou de paramètres d'acquisitions.

Une solution peut être apportée par des méthodes d'apprentissage semi-supervisé, qui essayent d'utiliser des données non annotées pour améliorer la précision des modèles appris sur un nombre insuffisant de données annotées. Une des stratégies pour cet apprentissage semi-supervisé consiste à produire des pseudo-labels à partir des données non annotées et de les injecter dans la phase d'apprentissage. Classiquement, le modèle est dans un premier temps entraîné à partir d'un petit nombre de données annotées. Puis les données non-annotées sont pseudo-labélisées ensuite injectées dans le modèle en utilisant le premier entraînement pour un apprentissage itératif co-joint avec des données annotées et pseudo-labélisées [5]. Un tel mécanisme est assez similaire à la régularisation entropique.

Selon nous, outre ce principe général, plusieurs apports devraient améliorer ce modèle. D'une part, une augmentation de données devrait permettre de gagner en généralisation. Nous pensons

que la méthode du mixup est une façon assez simple et efficace pour rendre un modèle plus robuste. D'autre part, nous savons que la qualité des pseudo-labels affecte l'apprentissage. Une stratégie est généralement d'associer une carte de confiance à ces données et d'écarter les pseudo-labels par seuillage de la carte de confiance. Un tel seuillage est généralement défini de manière globale et empirique. Dans notre cas nous pensons qu'un seuillage adaptatif évoluant au cours de l'apprentissage pourrait grandement améliorer la phase d'apprentissage en précision et en convergence.

Nous avons donc proposé un nouveau modèle appelé "Pseudo-label Refinement Network" (PLRNet) qui combine la stratégie d'apprentissage par pseudo-labels, le seul adaptatif des cartes de confiance de ces pseudo-labels et l'augmentation de données par mixup.

De manière plus précise, la structure de notre modèle est décrite dans la figure 4. Il est divisé en deux phases : un pipeline d'apprentissage et un pipeline d'inférence.

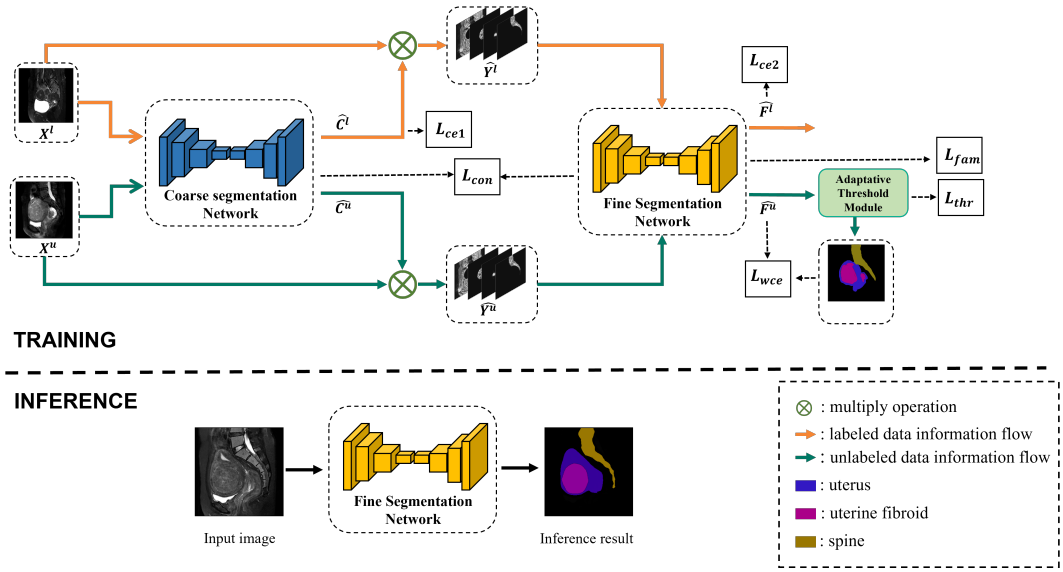

Figure 4 – Structure de PLRNet.

Pipeline d'apprentissage. Afin de gagner en précision nous avons choisi d'utiliser deux réseaux de segmentation, un réseau de segmentation plus grossier ("Coarse Segmentation Network-CSNet") suivi d'un réseau de segmentation plus fine ("Fine Segmentation Network-FSNet"). La sortie de chacun des deux réseaux est une carte de probabilités à 4 canaux (un canal par classe (fond, utérus, fibrome et colonne vertébrale). Les données d'entrée de FSNet sont issus du produit des données en niveau de gris avec les 4 canaux donnés de CSNet. Ceci permet de donner une région d'intérêt par classe à FSNet.

Pour l'entraînement, nous disposons de deux jeux de données : un jeu de données labellisés, X^l, et un jeu de données sans labels X^u. L'entraînement se fait en plusieurs

phases : 1) Un premier entraînement supervisé des deux réseaux en utilisant une partie de X^l (fonction de perte par corrélation croisée entre labels estimés et vérités terrain) ; 2) un second entraînement avec à la fois des données de X^l et de X^u. X^u donne en sortie des réseaux des pseudo-labels qui sont utilisés également en apprentissage par minimisation d'une fonction de perte spécifique ; 3) Une fonction de confiance est attachée aux pseudo-labels. Les pseudos-labels dont la confiance est en dessous d'un certain seuil T ne sont pas utilisés. Dans notre cas, T est déterminé de manière adaptative à l'aide d'une fonction de perte durant l'évolution de l'entraînement, ceci afin d'améliorer la qualité des pseudo-labels durant l'apprentissage.

L'architecture des deux réseaux est un point clés de notre méthode. Comme pour le HIFUNet de la solution totalement supervisée, nous avons utilisé le réseau convolutif global (GCN) dans la structure en U afin d'extraire efficacement les données complexes d'une scène en augmentant le champ valide de réception. Nous avons aussi remplacé les méthodes de sous-échantillonnage/sur échantillonnage habituellement utilisées dans les résaux (par ex. : max- ou mean-pooling) par des opérateurs basés sur la transformée en ondelettes et de la transformée en ondelettes inverse.

Nous avons également introduit lors de l'apprentissage une méthode d'augmentation de données basée sur le mélange de données caractéristiques alignées (Feature-aligned Mixup).

Pipeline d'inférence. Le réseau de segmentation fin FSNet après la phase d'entraînement est utilisé pour comme réseau d'inférence.

L'apprentissage et la validation du modèle ont été effectués sur la même base de données clinique de volumes IRM de 297 patients que l'étude précédente. La encore, les IRMs de 260 patients ont été utilisées pour l'apprentissage et les images des 37 autres patients ont été utilisées pour les tests. Par contre, dans les données d'entraînements, nous n'avons utilisé qu'un pourcentage des annotations pour former le jeu de données labellisés X^l, le reste composant les données non labellisées X^u. Plusieurs pourcentages de données labellisées ont été évalués allant de 100 % (apprentissage totalement supervisé) à 40%, 25% et 10%.

L'évaluation a consisté d'une part à estimer l'apport des différentes innovations à la segmentation finale. L'utilisation du GCN et des transformées en ondelettes dans les réseaux ainsi que le seuillage adaptatif durant l'apprentissage du seuil de confiance aux pseudo-labels et que l'augmentation des données par Mixup ont amélioré chacun les performances de la segmentation, ceci pour un taux de 25% de données labellisées.

Nous avons ensuite comparé les performances de notre méthode par rapport à des méthodes totalement supervisées (U-Net et HIFUNet), ainsi qu'à 4 autres méthodes d'apprentissage semi-supervisée U-Net, ASDNet [6], Latent Mixup [7], et Cross-Consistency Training (CCT) [8], ceci pour les 3 différents pourcentages de données étiquetées/non étiquetées. D'une part, notre

méthode, avec des pourcentages de 40% et 25% de données annotées, avait des performances supérieures à celles de U-Net totalement supervisée, et du même ordre de grandeur que U-Net totalement supervisée avec un taux seulement de 10% de données annotées. Notre méthode avait également des performances significativement meilleures que 4 autres méthodes d'apprentissage semi-supervisée, et ceci quel que soit le taux de données annotées. L'analyse visuelle qualitative des résultats de segmentation des différentes méthodes a confirmé cette tendance.

En conclusion, nous avons proposé un nouveau pipeline de segmentation par apprentissage semi-supervisé appelé PLRNet. Ce pipeline inclus différentes contributions (architecture d'apprentissage à deux réseaux grossier et fin, réseaux incorporant un réseau convolutif global et du sous- et sur- échantillonnage basés sur des transformées en ondelette, seuil adaptatif de confiance aux pseudo-lebels, augmentation de données par Mixup) qui permettent d'incorporer des données non annotées lors de la phase d'apprentissage afin d'améliorer la performance de segmentation du réseau. Nous avons validé notre méthode sur des données utilisées pour la planification du traitement des fibromes par HIFU. Cette évaluation a démontré que notre réseau de segmentation était plus performant que les méthodes d'apprentissage semi-supervisé de l'état de l'art. Et avait des performances proches voir supérieure à U-Net avec nettement moins de données d'apprentissage.

Le travail futur le plus important consiste à améliorer la qualité des pseudo-labels en concevant des seuils par classe plutôt qu'un seuil global, pour générer des pseudo-étiquettes non biaisées. En outre, nous prévoyons d'étendre notre approche à d'autres ensembles de données provenant de différents sites afin d'étudier la façon de sélectionner et annoter des données représentatives et comment extraire une segmentation plus riche à partir d'une annotation de données limitée.

PLRNet a fait l'objet d'une publication dans IRBM [9].

Segmentation automatique pour la sélection du plan du jour dans la radiothérapie adaptative du cancer du col de l'utérus guidée par CBCT

Ce chapitre porte sur le traitement du cancer du col de l'utérus par radiothérapie adaptative externe dont l'objectif est d'irradier la tumeur lors de différentes fractions de traitement tout en essayant de limiter au maximum la toxicité sur les tissus normaux environnants. Le traitement est assez complexe du fait des fortes variations anatomiques intrapelviennes survenant entre les fractions de traitement. La position et la forme du volume cible clinique (clinical target volume – CTV) comprenant le col de l'utérus, l'utérus et le haut du vagin dépendent fortement du remplissage de la vessie et du rectum, et de la régression tumorale le long du traitement. Le traitement doit donc s'adapter à la morphologie lors de chaque fraction. Une des stratégies

permettant de prendre en compte la morphologie lors de la fraction est de faire une acquisition CBCT avant la séance et à adapter le planning de dose (replanifier) en fonction des modifications de positionnement, de forme ou de volume de la tumeur et des organes adjacents. Une des stratégies est alors celle du plan du jour (plan of the day - PoD) qui est basée sur la génération au préalable d'une bibliothèque de plans de traitement comprenant plusieurs plans de traitement optimisés en fonction de multiples scanners X de planification (pCT) acquises avec différents remplissages de la vessie. Ensuite, lors de la séance de traitement, le plan de traitement est sélectionné parmi ceux de la bibliothèque ("plan du jour") sur la base de l'image CBCT. Bien que cette stratégie semble adéquate pour compenser les mouvements utérins, elle reste complexe dans un flux de travail clinique en raison du faible contraste des images CBCT et des grandes déformations anatomiques. Le choix du plan du jour est généralement fait par l'expert médical, et, comme le montre l'étude de Gobeli *et al.*, il existe une forte variabilité inter-expert (dans cette étude, le plan du jour optimal n'a été choisi en moyenne que par 60% des experts). Dans ce contexte notre objectif était de proposer une stratégie pour sélectionner automatiquement le plan de traitement optimal. Cette stratégie repose sur une segmentation des images CBCT basée sur l'apprentissage profond suivi d'une procédure de sélection du plan de traitement optimal maximisant la couverture du volume cible clinique sur la base d'un critère géométrique.

De manière plus précise, la structure de notre stratégie est décrite dans la figure 5.

Noua avons à notre disposition 3 scanners X de planification avec 3 remplissages de vessie (vessie vide -EB-, vessie intermédiaire -IB- et vessie pleine -FB-). Ces scanners ont été segmentés manuellement (col de l'utérus, utérus et haut du vagin, ainsi que rectum, vessie et sac intestinal) et un planning de doses a été établi sur chacun de ces volumes. Les volumes segmentés sont nommés ($\mathbf{CTV}_{EB}$, $\mathbf{CTV}_{IB}$ et $\mathbf{CTV}_{FB}$).

Lors d'une session, un volume CBCT est acquis. Ce volume passe alors par plusieurs étapes pour la sélection du plan du jour :

1. La segmentation du CBCT par un modèle d'apprentissage profond. Nous avons choisi un réseau existant, le nnU-Net 3d_fullres [10], qui s'est avéré être l'un des modèles les plus performants dans de nombreuses tâches de segmentation d'images. Nous avons entraîné ce réseau sur les données CBCT de 17 patients (environ 200 volumes) et testé sur 6 patients (environ 70 volumes). Une évaluation en 4-fold cross-validation nous a donné un Dice médian de l'ordre 0,8 pour les différents organes. Ce Dice est de l'ordre de grandeurs des deux autres études de segmentations en CBCT mais avec des gains en vitesse de calcul et en nombre d'organes segmentés ;
2. Un recalage rigide du CBCT sur chacun des 3 scanners X de planification. Le recalage a été effectué sur les structures osseuses à l'aide de la bibliothèque Elastix. Le recalage rigide se justifie car nous ne devions pas déformer les organes abdominaux pour passer à l'étape suivante ;

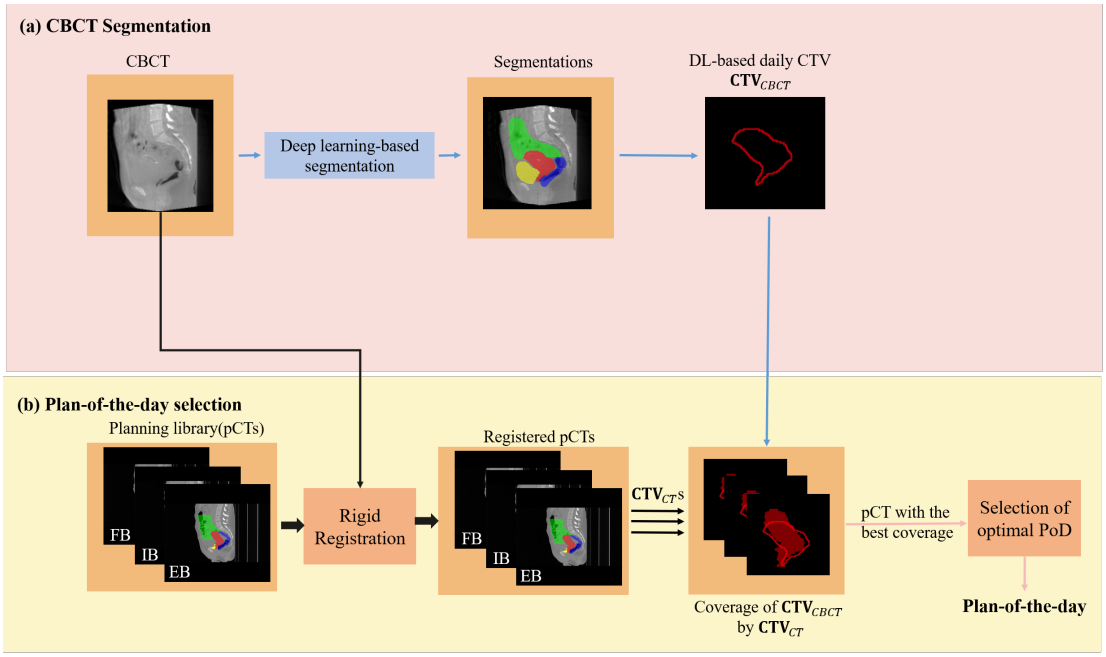

Figure 5 – Organigramme de la méthode de choix du plan du jour optimal : (1) segmentation CBCT à l'aide de l'apprentissage profond et (2) sélection du plan du jour (PoD) à l'aide des contours du volume cible clinique (CTV). La sélection du PoD s'appuie sur : (a) le recalage rigide basé sur l'os du CBCT per-opératoire du jour sur les 3 CTs de planification (pCTs) de la librairie ; (b) le calcul de la couverture entre le CTV du jour ($\mathbf{CTV}_{CBCT}$) et les 3 CTVs de la librairie ($\mathbf{CTV}_{EB}$,$\mathbf{CTV}_{IB}$,$\mathbf{CTV}_{FB}$) ; (c) la sélection du meilleur plan de traitement sur la base de la couverture cible : le pCT correspondant à la couverture la plus élevée est sélectionné. (EB : vessie vide ; IB : vessie intermédiaire ; FB : vessie pleine ; cov : valeur de couverture).

3. Sélection du plan du jour. Le CBCT est reprojeté sur chacun des 3 scanners X de planification. Pour chacun des 3 scanners le taux de recouvrement entre les organes segmentés automatiquement du CBCT et ceux manuellement du scanner X de planification est calculé. Le scanner X de planification correspondant à la valeur de taux de recouvrement le plus élevé sera sélectionné comme plan du jour de la fraction de traitement.

Pour l'évaluation nous avons comparé le résultat de notre choix automatique avec celui de l'expert dans une étude a posteriori sur les 272 CBCTs à notre disposition. Notre méthode était en concordance stricte avec l'expert dans 91.5% (seuls 23 cas sur 272 avaient un plan du jour sélectionné sous-optimal par rapport à la référence). Par contre, nous avons constaté que pour certains patients, les scanners X de planification pouvaient être assez similaires avec moins de 5% de différences des volumes des organes entre deux remplissages de vessie. Si l'on considère cette tolérance de 5%, la concordance est alors de 99.6%, avec un seul cas qui présentait un plan du jour sélectionné sous-optimal par rapport à la référence. C'était le cas d'une patiente avec une poute petite vessie que notre algorithme de segmentation n'arrivait pas à estimer. Nous

pensons que dans un contexte clinique, ce type de cas serait facilement détecté visuellement et ajusté manuellement

L'algorithme de sélection du plan du jour a fait l'objet d'une publication dans Physics in Medicine and Biology [11].

Conclusion

Dans cette thèse, nous nous sommes concentré sur la recherche et développement de nouveaux algorithmes pour la segmentation d'images de la région utérine dans le cadre des traitements des fibromes utérins par HIFU et tumeurs du col de l'utérus par radiothérapie adaptative. Plus particulièrement nous avons proposé 3 contributions dans ces domaines :

1. La solution HIFUNet pour la segmentation automatique des images IRM de l'utérus avant le traitement HIFU. Une segmentation entièrement automatique et précise de l'utérus, des fibromes utérins et de la colonne vertébrale dans la région utérine été proposée. À notre connaissance, il s'agit de la première tentative de méthode d'apprentissage profond pour la segmentation multi-classes dans la région utérine. L'évaluation a montré que HIFUNet est plus robuste et plus précis que les méthodes traditionnelles ou basées Deep Learning sur cet organe. L'avantage de HIFUNet est qu'il utilise des grands noyaux convolutifs afin d'étendre le champ réceptif, ce qui permet au réseau d'extraire les caractéristiques de la cible de segmentation dans l'arrière-plan complexe de l'image médicale de la scène. Les résultats expérimentaux indiquent que HIFUNet a bonne précision de segmentation des fibromes utérins quels que soient leurs nombres ou leurs tailles ;
2. La solution PLRNet qui permet de gérer la rareté des données annotées pour l'apprentissage. C'est un réseau basé sur de l'apprentissage semi-supervisé. Apres un premier apprentissage sur un nombre restreint de données annotées, l'apprentissage se poursuit en y incluant également des données non annotées. Lors de l'apprentissage, le réseau propose des pseudo-labels à partir des données non-annotées, pseudo-labels qui sont réinjectés puis raffinés dans le processus d'apprentissage. Pour cela, nous utilisons deux étages de réseaux à noyau convolutif de grande taille avec des opérateurs de sous- et sur-échantillonnage basés sur des transformées en ondelettes. Nous avons également proposé un mécanisme de détermination de seuil de confiance aux pseudo-labels qui s'adapte durant l'apprentissage. Nous avons également intégré un mécanisme d'augmentation des données par mélanges de caractéristiques dans les couches cachées, ceci permettant d'éviter le surapprentissage. Ces différentes contributions permettent à PLRNet de résoudre le problème de la rareté des données étiquetées dans la segmentation des images médicales et ainsi permettre des réapprentissages rapides à partir de peu de données annotées en cas de changement de paramètres ou de machine d'acquisition ;

3. La sélection du plan du jour lors d'une fraction de radiothérapie adaptative du cancer du col de l'utérus. Pour cela nous avons proposé une méthode basée sur 1) la segmentation des images CBCT par un réseau de type nnU-Net ; 2) le recalage rigide entre CBCT et les scanner X de planificattion d'une librairie ; et 3) le choix du plan du jour en choisissant le scanner X de planification dont les organes ont le plus grand recouvrement avec ceux du CBCT. Cette méthode constitue la première tentative de sélection automatique du plan du jour en radiothérapie adaptative.

Les méthodes de segmentation présentées dans cette thèse ont obtenu de bons résultats dans la segmentation d'images soit sur l'IRM préopératoire pour le traitement HIFU, soit sur le CBCT pour la radiothérapie adaptative, mais il reste encore quelques problèmes de traitement d'images dans ces deux domaines d'applications qui méritent d'être explorés et résolus. Pour le traitement des fibromes par HIFU, le guidage de cette thérapie se fait sous échographie 3D. Il nous faudra donc développer une nouvelle technique de segmentation de la zone anatomique mais sur l'image échographique. Dans un deuxième temps, la fusion des deux modalités (IRM/ultrasons) par recalage élastique devra être envisagée si possible avec de l'apprentissage profond pour des questions de vitesse de traitement. Pour la radiothérapie adaptative, là encore un recalage entre CBCT et le CT de planification pourrait être envisagé pour associer directement la morphologie du patient au moment du traitement à celle utilisée pour établir le planning de dose.

BIBLIOGRAPHY

[1] B. Rigaud, A. Simon, M. Gobeli, C. Lafond, J. Leseur, A. Barateau, N. Jaksic, J. Castelli, D. Williaume, P. Haigron et al., "CBCT-guided evolutive library for cervical adaptive IMRT," Medical physics, vol. 45, no. 4, pp. 1379–1390, 2018.

[2] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in Thirty-first AAAI conference on artificial intelligence, 2017.

[3] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters improve semantic segmentation by global convolutional network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4353–4361.

[4] C. Zhang, H. Shu, G. Yang, F. Li, Y. Wen, Q. Zhang, J.-L. Dillenseger, and J.-L. Coatrieux, "HIFUNet: multi-class segmentation of uterine regions from MR images using global convolutional networks for HIFU surgery planning," IEEE transactions on medical imaging, vol. 39, no. 11, pp. 3309–3320, 2020, publisher: IEEE.

[5] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in ICML 2013 Workshop : Challenges in Representation Learning (WREPL), vol. 3, 2013, p. 896, issue: 2.

[6] D. Nie, Y. Gao, L. Wang, and D. Shen, "Asdnet: Attention based semi-supervised deep networks for medical image segmentation," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, ser. Lecture Notes in Computer Science, vol. 11073, 2018, pp. 370–378.

[7] P. K. Gyawali, S. Ghimire, P. Bajracharya, Z. Li, and L. Wang, "Semi-supervised medical image classification with global latent mixing," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, ser. Lecture Notes in Computer Science, vol. 12261, 2020, pp. 604–613.

[8] Y. Ouali, C. Hudelot, and M. Tami, "Semi-supervised semantic segmentation with cross-consistency training," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), jun 2020, pp. 12 674–12 684.

[9] C. Zhang, G. Yang, F. Li, Y. Wen, Y. Yao, H. Shu, A. Simon, J.-L. Dillenseger, and J.-L. Coatrieux, "CTANet: confidence-based threshold adaption network for semi-supervised segmentation of uterine regions from MR images for HIFU treatment," IRBM, vol. 44, no. 3, p. 100747, jun 2023.

[10] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," Nature methods, vol. 18, no. 2, pp. 203–211, 2021.

[11] C. Zhang, C. Lafond, A. Barateau, J. Leseur, B. Rigaud, D. B. Chan Sock Line, G. Yang, H. Shu, J.-L. Dillenseger, R. de Crevoisier, and A. Simon, "Automatic segmentation for plan-of-the-day selection in CBCT-guided adaptive radiation therapy of cervical cancer," Physics in Medicine and Biology, vol. 67, no. 24, p. 245020, dec 2022.

TABLE OF CONTENTS

List of Figures

List of Tables

INTRODUCTION

The uterus is a pear-shaped hollow organ in the female pelvis, located between the bladder and the rectum. Uterine tumors may occur in the body, isthmus, and cervix and may be benign or malignant (cancerous). Among these, benign tumors, represented by fibroids, and malignant tumors, represented by cervical cancer, are currently becoming a serious health risk for women. With the development of medical technology, there have been more advanced surgical treatments for benign and malignant tumors of the uterus. Computer-aided methods to improve surgery's accuracy, efficiency, and safety are essential to women's health.

In the treatment of benign and malignant tumors of the uterus, accurate annotation of the lesions in the uterine region and the surrounding crisis organs is an essential part of the diagnosis and treatment planning: 1) In treating uterine fibroids, lesion annotation helps the surgeon determine the fibroid's size, shape, location, and, thus, the type of fibroid. 2)When treating adaptive radiotherapy (ART) procedures for cervical cancer, the doctor can develop the radiotherapy process and the prescribed dose based on the results of the delineation. 3) In high-intensity focused ultrasound surgery (HIFU) for uterine fibroids, the target area in the preoperative images is mapped to the intraoperative images to guide the surgery. However, annotating target areas and organs at risk in medical images are time-consuming and labor-intensive, and factors such as image noise make accurate annotation difficult. Therefore, exploring automatic and accurate annotating methods for uterine images has significant clinical value for treating benign and malignant uterine diseases.

To this end, this paper investigates the multimodal image segmentation algorithms and automated surgical techniques involved in HIFU and ART to treat uterine fibroids and cervical cancer, respectively. This paper aims to improve the automation and precision of these two treatments in clinical practice. The main work and contributions are as follows:

1. Multi-Class Segmentation of Uterine Regions From MR Images Using Global Convolutional Networks for HIFU Surgery Planning

To address the problem that existing state-of-the-art (SOTA) deep learning segmentation methods are not effective enough for complex multi-level feature extraction, we propose a novel convolutional neural network called HIFUNet to segment the uterus, uterine fibroids, and spine. The network is an end-to-end encoder-decoder architecture designed with a global convolutional network (GCN) module to expand the valid receptive field and extract multi-scale contextual information. In addition, combining GCN with our proposed deep multiple atrous convolution (DMAC) module can further extract contextual semantic information and denser feature maps.

Our approach is compared to both conventional and other deep learning methods and the experimental results conducted on a large dataset show its effectiveness.

2. Semi-supervised uterine MR image segmentation method based on pseudo-label Refinement for HIFU procedure planning The fully supervised semantic segmentation approach is effective but requires a high amount of annotation on the training data, so we propose a semi-supervised deep learning approach to segment MRI images. This method aims to refine the pseudo-label generated in the semi-supervised method named: pseudo-label refinement network (PLRNet). Inspired by the fully supervised uterine segmentation method HIFUNet, feature extraction can be improved by expanding the valid reception field in segmenting uterine fibroids with different sizes and shapes. Therefore, in semi-supervised feature extraction, we consider a network with large convolutional kernels to extract contextual features for each target class in MR images. In addition, the feature dilution problem caused by the typical pooling operation in deep learning is improved, and wavelet pooling is utilized to suppress the image noise. Our semi-supervised segmentation network has two components based on two cascaded large convolutional kernel networks containing wavelet pooling, a coarse segmentation network and a fine segmentation network. The coarse segmentation network is pretrained to fix the model parameters on the labeled images, and after the initial rough segmentation of the unlabeled data, the inaccurate prediction results are fed into the second network for further segmentation, and the obtained prediction results are named as the "pseudo-label" of the unlabeled data. Unlike the traditional semi-supervised approach of setting fixed thresholds, we use an adaptive method based on confidence thresholds for the first time in semi-supervised segmentation to improve the quality of the pseudo-label. As the network is trained, the threshold value decreases, automatically shifting the network's attention from the less difficult spine region to the more difficult fibroid region, with changes in the threshold value corresponding to changes in the segmentation target features. In addition, inspired by "Mixup" method, we extend Mixup operations to each hidden layer of the fine segmentation network, which helps data augmentation and avoid the overfitting phenomenon that tends to occur in semi-supervised learning, improving the generalization and robustness of the model.

3. Automatic segmentation for plan-of-the-day selection in CBCT-guided adaptive radiation therapy of cervical cancer Plan-of-the-day (PoD)-based ART is based on a library of treatment plans. At each treatment fraction, the PoD is selected based on daily images. However, this strategy is limited by the optimal PoD selection due to visual uncertainties. This work proposes a workflow to automatically and quantitatively determine the PoD of ART for cervical cancer based on daily CBCT images. The quantification is performed by segmenting the main structures of interest (Clinical target volume (CTV), rectum, bladder, and bowel bag) in CBCT images using a deep learning model. Then, the PoD is selected from the treatment plan library according to the geometrical coverage of the CTV. The resulting PoD is compared

to the one obtained considering reference CBCT delineations for the evaluation.

Chapter 1

Background-uterine tumors and therapies

1.1 Anatomy of uterus

The uterus is one of the female internal reproductive organs and is located behind the bladder and in front of the rectum. The normal adult uterus measures 6 to 9 cm in length [1]. As is shown in the Figure 1.1, the uterus is pear-shaped and has three sections: the cervix, the body and the fundus. The uterine body narrows to form a waist (the isthmus), which extends into the cervix. The uterine canal passes through the internal os and emerges as the external os at the vaginal vault [2]. The uterus holds the growing fetus during pregnancy. The cervix connects the lower part of the uterus to the vagina and together with the vagina, forms the birth canal.

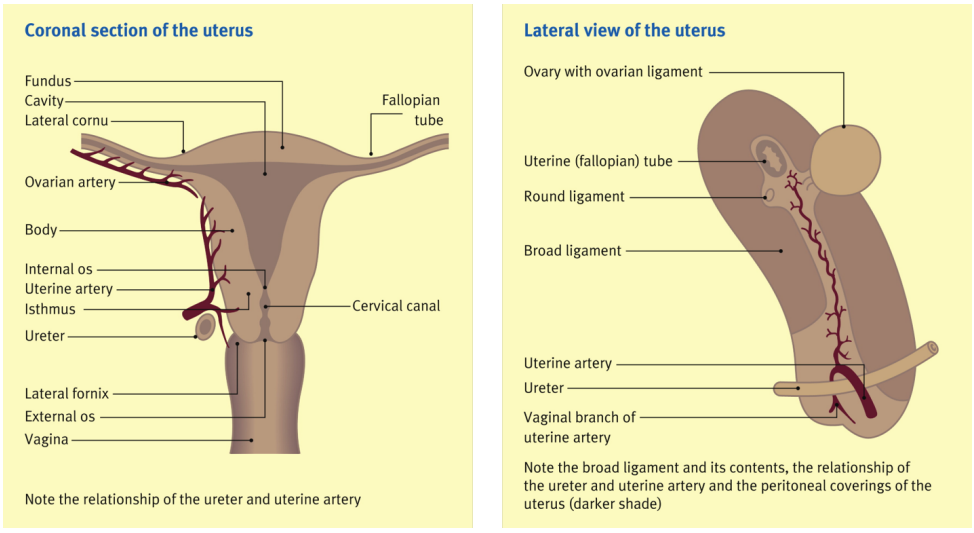

Figure 1.1 – Coronal and lateral views of the anatomy of the uterus. Reprinted from "Anatomy of the uterus." by Ellis, Harold. 2011, Anaesthesia & Intensive Care Medicine,12(3), 99-101. Copyright (2022) by Elsevier.License number: 5305830077364.

The examination by medical imaging of the uterus includes the evaluation of the endometrial stripe, the junction zone, the myometrium, and the cervix. The cervix, endometrium, and

myometrium are well visualized by both ultrasound and Magnetic Resonance Imaging(MRI). However, the junctional zone is better assessed on MRI. The contour and anomalies of the uterus can also be assessed by ultrasound and MRI [3].

Although the interpretation of imaging methods (e.g. MRI) is consistent with the anatomical description, the precise anatomy of the uterus, must often be analyzed in combination with multi-modal images. For example, in ultrasound images, imaging of the internal details of the uterus is often not achieved because of the low resolution of ultrasound imaging. In particular, in real-time ultrasound video images, the uterine borders are sometimes blurred due to factors such as the patient's respiratory. Therefore, in such cases, the doctor will combine the patient's MRI images with the ultrasound to make a diagnosis.

1.2 Uterine benign and malignant tumors

The uterus is one of the most important organs in relation to women's reproductive health. Uterine tumors, caused by changes in the growth of cells in the uterus and thus uncontrolled growth of the uterus, are a threat to the women's health. Tumors are classified as benign or malignant.

Benign tumors, also called noncancerous tumors, are tumors that grow but do not spread to other parts of the body. There are four types of noncancerous growths of uterus: uterine fibroids, benign polyps, endometriosis and endometrial hyperplasia.

Malignant tumors, also known as cancerous tumors, can spread to other parts of the body and can be life-threatening. The two main types of uterine cancer are adenocarcinoma and sarcoma [4].

Although the upper end of the cervix is attached to the body of the uterus and is only a few centimeters away from it, cervical cancer is not classified as an uterine cancer mainly because (1) the causes of development are different: cervical cancer is due to human papillomaviruses (HPV) infection; uterine cancer is due to genetic factors, overweight, etc.; (2) The treatment modalities are different: for early stages of cervical cancer, surgery or radiation combined with chemotherapy can be used and but for late stage, radiation combined with chemotherapy is usually the main treatment; uterine cancer is usually treated by surgical removal of the uterus, fallopian tubes and ovaries.

In this article, we focus on one benign and one malignant disease of the uterus: uterine fibroids and cervical cancer. These two diseases and their treatments are described in detail below.

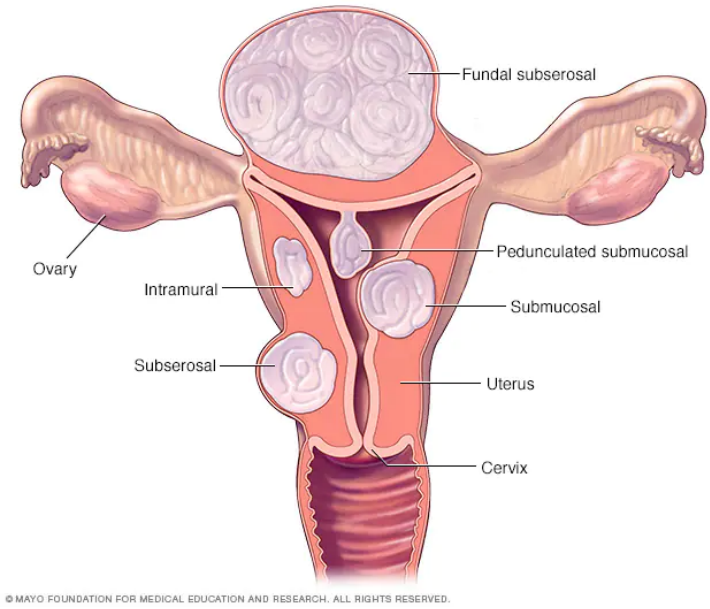

Figure 1.2 – Fibroid locations (MAYO 2019). "Uterine fiborids" by Mayo Clinic staff, accessed 16 May 2022, <https://www.mayoclinic.org/diseases-conditions/uterine-fibroids/symptoms-causes/syc-20354288#dialogId66869006>.

1.2.1 Uterine fiborids and high-intensity focused ultrasound (HIFU) therapy

Uterine fibroids

Uterine fibroids (UF) are also called uterine leiomyomas. UF are benign smooth muscle tumors of the uterus. They affect women of childbearing age. Uterine fibroids are not associated with an increased risk of uterine cancer and almost never develop into cancer. The incidence of uterine fibroids has continued to increase in recent years. In 2001, they are clinically apparent in up to 25% of women[5]. Twenty years later, the prevalence has increased to more than 75% [6]. At age 50, nearly 70% of white women and more than 80% of black women have at least one uterine fibroid [7].

As can be seen from Figure 1.2 (MAYO 2019), there are three main types of uterine fibroids. Intramural fibroids that develop within the muscular uterine wall. Submucosal fibroids that protrude into the uterine cavity. Subserosal fibroids that project to the outside of the uterus [8].

Uterine fibroids diagnosis

In most cases, the diagnosis is not timely because patients with fibroids are asymptomatic or their symptoms develop slowly. Most findings of fibroids are due to routine pelvic examinations or incidental imaging [9]. Ultrasound is then the standard confirmatory imaging modality because

it can easily and inexpensively distinguish fibroids from the gravid uterus or adnexal masses. The need for additional imaging depends on the clinical findings of the patient [10]. Transvaginal ultrasonography is as efficient as MRI in detecting myoma presence, but its ability to accurately map myomas is inferior to MRI, especially in the case of large multiple-myoma [11]. Diagnostic imaging is used to confirm clinically suspected uterine fibroids. More details will be discussed in section 1.3.

Uterine fibroids treatment

When the symptoms of a fibroid become bothersome, a treatment option may be considered. However, the choice of an option is quite complicated because it depends on the size of the fibroma, its location, the age of the patient (menopausal or not) and her desire to have a child. This choice is further complicated by the fact that only few randomized trials have compared various therapies for fibroids and there is a lack of data to provide information on different intervension strategies. Different treatment strategies should be used depending on the size and symptoms of the fibroids. Indeed, many fibroids are relatively small and asymptomatic. Several factors should be considered when proposing a management plan for benign uterine fibroids, such as the woman's preference, severity of symptoms, fertility desires, and the patient's age. By assessing the fibroid symptom and the woman's preferences, recommended decision trees for the management of symptomatic UFs are provided in professional guidelines (See Figure 1.3) [10]. Hysterectomy, laparoscopic myomectomy and hysteroscopic myomectomy are the most commonly used surgical interventions for myomas. Alternatives to surgical intervention include uterine artery embolization (UAE), magnetic resonance-guided high intensity focused ultrasound surgery (MRgFUS) and vaginal uterine arteries occlusion [12]. Specifically, compared with hysterectomy, focused ultrasound procedures result in rapid recovery and low risk of complications and may provide effective treatment [9]. In this article, we will focus on the application of focused ultrasound surgery in uterine fibroids.

High-intensity focused ultrasound (HIFU) therapy

High-intensity focused ultrasound (HIFU) is a high-precision medical procedure for local heating and ablation of diseased tissue. It has been widely used to treat uterine fibroids. Compared to other surgical therapies, HIFU has the advantage of being non-invasive and having a low number of complications. HIFU can be considered as a promising treatment option for women who wish to conceive a child [13]. HIFU can be either guided by MRI (Magnetic Resonance-guided HIFU - MRgHIFU) or ultrasound (Ultrasound-guided HIFU - USgHIFU). However, the clinical use of MRgHIFU is limited due to the requirements of a dedicated MR device to guide the treatment and the length of the procedure [14].

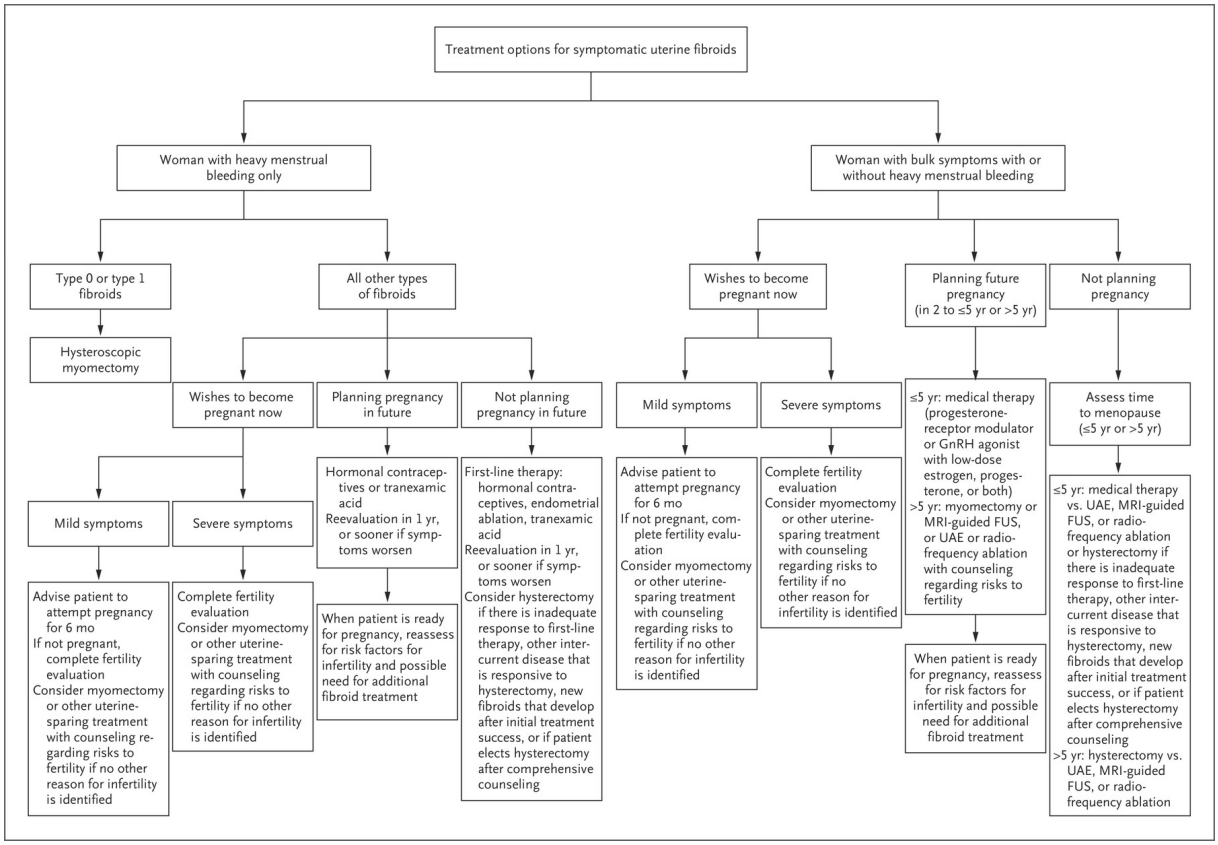

Figure 1.3 – Algorithm for the Management of Symptomatic Uterine Fibroids. FUS denotes focused ultrasound surgery, GnRH gonadotropin-releasing hormone, MRI magnetic resonance imaging, and UAE uterine-artery embolization. Reproduced with permission from "Clinical practice. Uterine fibroids." by Stewart, Elizabeth A. 2015, The New England journal of medicine, 372(17), 1646-1655. Copyright Massachusetts Medical Society.

Figure 1.4 shows the treatment process of the HIFU procedure on the patient. During the treatment, the patient is lying prone on the treatment machine. An extracorporeal focused piezoelectric transducer converges the ultrasound energy to the target. The conversion of ultrasound into thermal energy induces irreversible cell death by coagulation necrosis if the temperature exceeds 56°C and is maintained for more than 2 seconds. In fact, during ultrasonic ablation (UA), the temperature at the focal volume can rapidly exceed 80°C [15, 16].

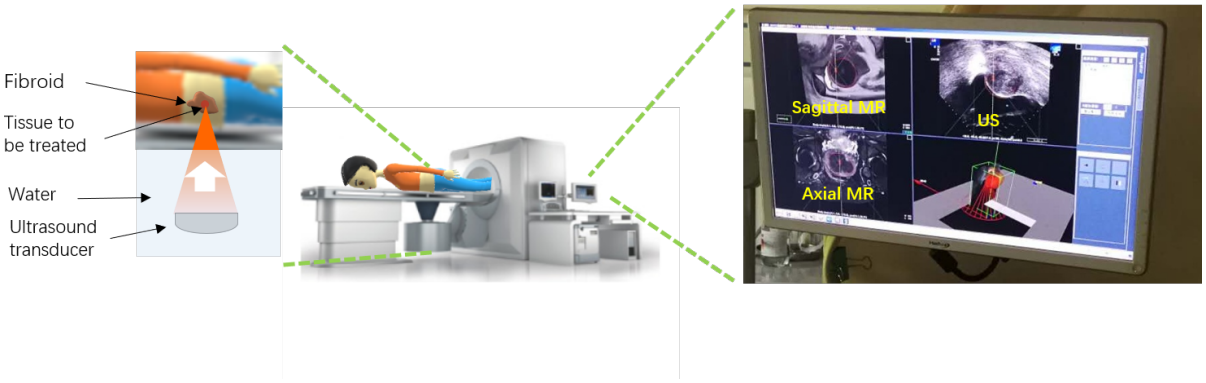

Figure 1.4 – High-intensity focused ultrasound (HIFU) therapy for fibroid ablation.

The flowchart for treating uterine fibroids with HIFU is shown in Figure 1.5 (with a focus on USgHIFU as an example). During the diagnostic and planning phase, the patient first undergoes an MRI scan. On this MRI, the surgeon delineates the area of the lesion (typically the uterus and the fibroid area). During the operation, the HIFU spot (the area to be treated) is guided by means of a real time ultrasound. In order to transfer the area of the fibroids delineated in the MRI to the ultrasound, the physician performs a manual registration of the preoperative MRI image with the intraoperative ultrasound real time acquisition. This registration then allows the guidance of the fibroid ablation in the HIFU procedure.

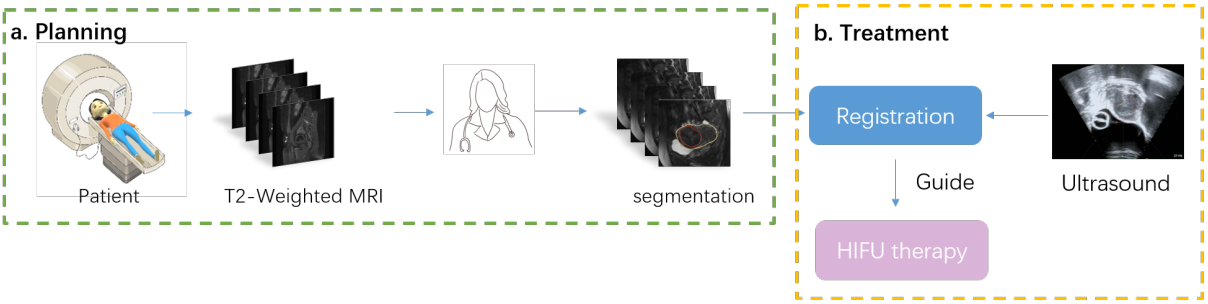

Figure 1.5 – The flowchart of USgHIFU for treating uterine fibroids. The figure was partly modified from Servier Medical Art, licensed under a Creative Common Attribution 3.0 Generic License. http://smart.servier.com/.

1.2.2 Cervical cancer and adaptive radiotherapy (ART)

Cervical cancer diagnose

Cervical cancer is the fourth most common female malignancy worldwide [17], with more than 500,000 women diagnosed with cervical cancer each year and the disease causing more than 300,000 deaths worldwide [18]. Most cases occur in the less developed countries where no effective screening systems is available. Risk factors include exposure to human papillomavirus, smoking, and immune-system dysfunction [19]. The usual tests for diagnosing cervical cancer are: colposcopy with biopsy and large loop excision of the transformation zone (LLETZ) or cone biopsy.

Adaptive radiotherapy (ART)

Adaptive radiation therapy (ART) is a closed-loop radiation treatment process where the treatment plan can be modified through systematic feedback of measurements. It was first introduced and discussed conceptually by Yan *et al.* in 1997 [20]. ART is mainly designed to solve the problem of the impact of target area location and morphological changes between radiotherapy fractions on the actual dose distribution. ART allows for high-dose, high-precision irradiation of the tumor target area while reducing irradiation of the surrounding normal tissue to minimize toxicity.

With the development of technology, ART has been implemented in clinical practice on many therapeutic targets, including head-and-neck, lung, prostate, bladder and cervix. Volumetric imaging and automated segmentation allow the calculation of daily doses so that adaptation decisions can be made based on dosimetric information rather than geometric information alone [21]. ART can be classified into three categories: adapt when necessary (offline ART), adapt before or during the treatment of that day (online ART), or adapt in real time to changes and movements (real-time ART).

In this thesis, we will focus on the treatment of cervical cancer. Hereafter, ART will refer to offline ART, except unless otherwise noted. During ART for cervical cancer, changes in bladder and rectal filling can affect the spatial position of the uterus and thus often lead to errors in dose delivery. The emergence of image-guided radiotherapy (IGRT) has made it possible to visualize the morphology of the soft tissues during ART. With this method it is possible to monitor changes in the bladder, bowel and rectum, ensuring a high radiation dose within the target area [22]. MR- [23] or cone-beam computed tomography(CBCT)-guided [24] ART is widely used in the clinic. The workflow of CBCT-guided online ART delivery with daily replanning is shown in Figure 1.6. It consists of two parts: planning and treatment. In the planning phase, the initial radiotherapy plan is generated based on the contours of the planning CTs. In the treatment phase,the treatment plan is constantly optimized based on information about the size

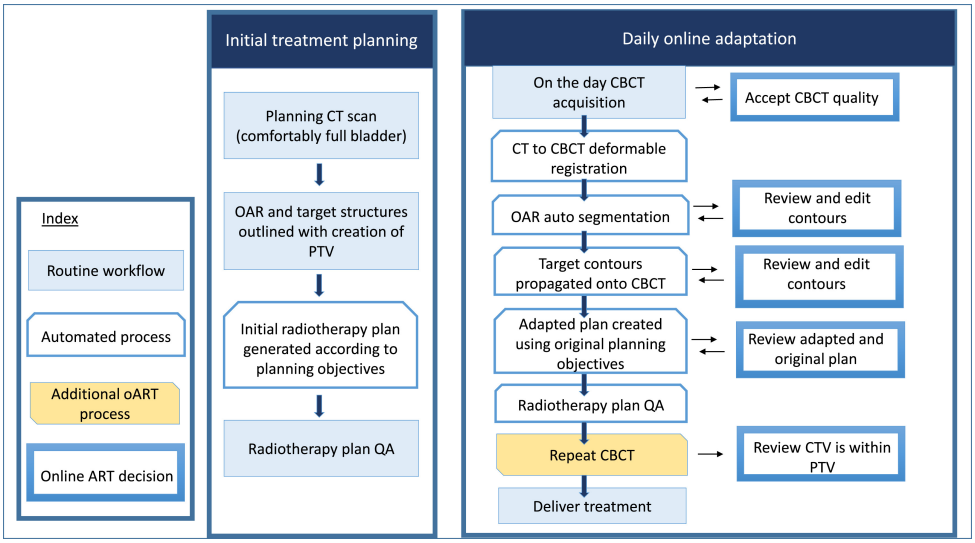

Figure 1.6 – Workflow for cone beam computed tomography (CBCT)-guided online adaptive radiotherapy with daily replanning for cervical cancer. Reproduced with permission from "Adaptive Radiotherapy in the Management of Cervical Cancer: Review of Strategies and Clinical Implementation." by Shelley, C. E.,(2021). Clinical Oncology, 33(9), 579-590. Copyright Clearance Center's RightsLink service. License number: 5318700298103.

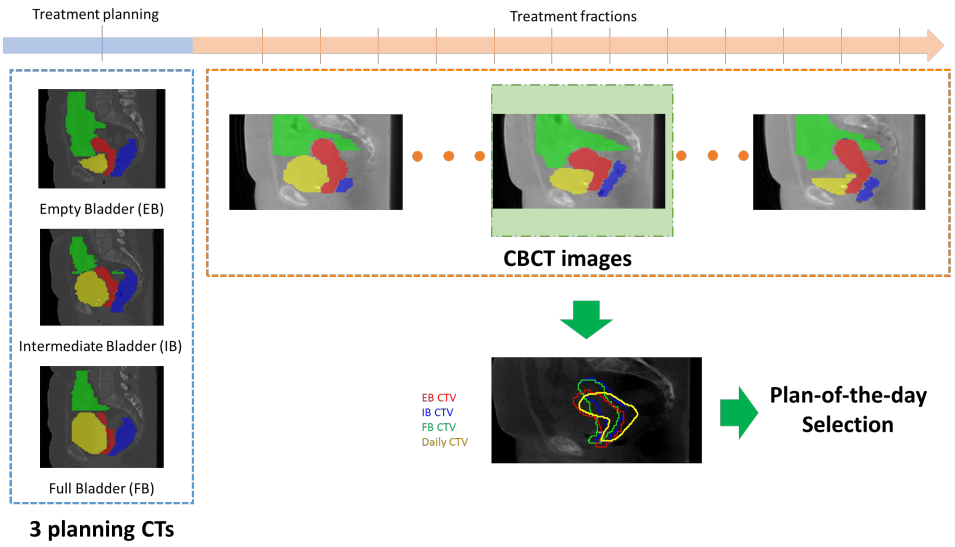

Figure 1.7 – Flowchart of plan-of-the-day ART for cervical cancer. The process consists of three steps (1) Planning: acquisition of multiple planning CT scans with variable bladder volumes. (2) Acquisition of the CBCT image of the day. (3) Selection of the most appropriate treatment plan to maximize the target coverage. The CTV, bowel sac, bladder and rectum are represented as red, green, yellow and blue filled contours. For plan-of-the-day selection, empty bladder (EB), intermediate bladder (IB), full bladder (FB) and daily CTV are represented on the daily CBCT as red, blue, green and yellow contours, respectively.

and location of the tumor and organs at risks (OARs) obtained from the continuous acquisition of online anatomical images.

In the context of offline CBCT-guided ART, plan-of-the-day (PoD) strategies have been proposed, based on the generation of a treatment plan library, including several treatment plans optimized based on multiple planning CTs (pCT) acquired with various bladder fillings. At each treatment fraction, the treatment plan is then selected among those of the library ("plan-of-the-day") based on an in-room image (*e.g.*, CBCT image, see Figure 1.7).

1.3 Uterine imaging

Medical images are currently essential for diagnosis and in image-guided surgery. In the treatment of uterine fibroids and cervical cancer, diagnostic images are used for the planning of both HIFU therapy and ART. Also, both of these treatments are image-guided therapies. Therefore, in this section we present the different types and specificities of uterine imaging.

1.3.1 Magnetic Resonance (MR) Imaging

Figure 1.8 – MRI of a case with uterine fibroids. (a) fat-suppressed T2-weighted in sagittal direction, (b) fat-saturated T2-weighted in sagittal direction, (c) fat-saturated T2-weighted in axial direction.

MRI is the reference modality because of the high tissue spatial resolution and absence of ionizing radiation. The different layers of the uterus are visualized distinctly. The advantage of MRI is that it can also evaluate the adjacent anatomical structures such as the ovaries, fallopian tubes, bladder and rectum. Compared to transvaginal ultrasound, MRI provides better soft-tissue contrast resolution and is preferred for preoperative mapping to determine location, size, and number of fibroids [25]. Multiple sequences in different planes are obtained. The standard pelvic examination begins with a sagittal T2 weighted image, followed by coronal and axial views.

It is followed by a sagittal and/or axial fat-saturated T2 weighted sequence. Other sequences may be helpful in some cases, such as angled axis to evaluate the cervix and Mullerian duct anomalies. An oblique plane perpendicular to the long axis of the cervix can also be useful in staging cervical carcinoma [26].

As shown in Figure 1.8, different MR sequences have different imaging information about the organs.

1.3.2 Ultrasound (US) imaging

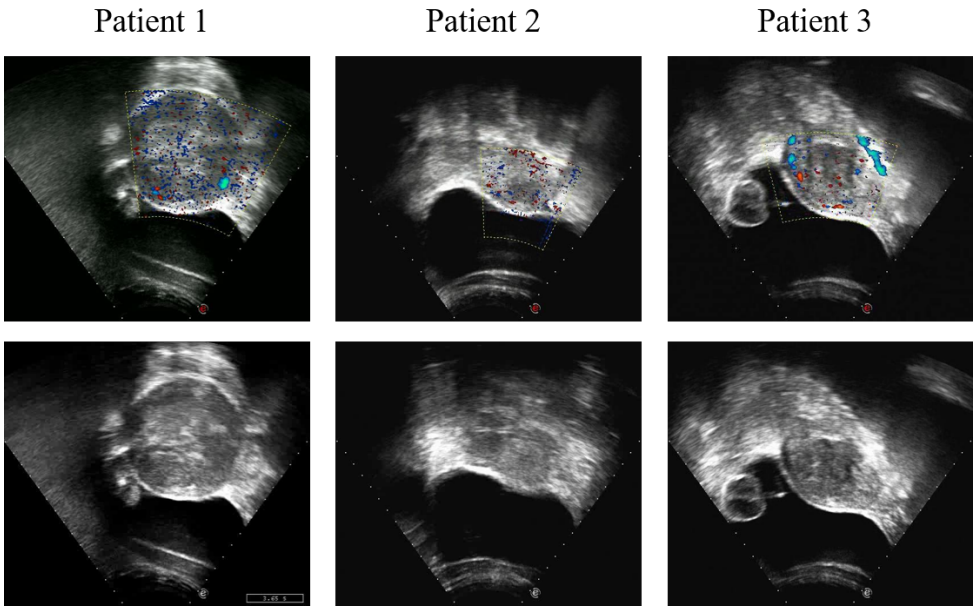

Figure 1.9 – Examples of ultraound images in USgHIFU. The doctor uses Doppler to evaluate the vascularization inside the fibroids (above). The shape of the fibroids is difficult to distinguish and only the unclear contour of the uterus are shown (below).

Ultrasound (US) is the most common and usually the primary modality for evaluating the uterus because to its availability and low cost [27]. US provides a good assessement of anatomy and contour. Transvaginal images offer a better view of the endometrium and the entire myometrium and are better to detect fibroids near the cervix. Transabdominal US can locate large leiomyomas. Color Doppler can be used to evaluate the vascularity within the lesions, while whereas leiomyomas will show an absence of flow [28].

As mentioned in Section 1.2.1 and shown in Figure 1.4 and Figure 1.5, transabdominal US imaging is used in the USgHIFU treatment procedure. During treatment, the doctors monitor the changes and movements of the uterus and observe the ablation of the uterine fibroids.

Figure 1.9 shows examples of US images in USgHIFU. These images are frames of the real-

time US video. Before HIFU ablation, the doctor uses Doppler to assess the vascularity of the fibroids (above). The shape of the fibroids is difficult to distinguish in the US image and usually only unclear contours of the uterus can be seen (below). This is why the preoperative MRI is important to overlay the segmented contours of the fibroids to the US images during the treatment.

In addition, transvaginal ultrasound (TVUS) could be used for cervical cancer staging. Several dedicated imaging centers report that the accuracy of TVUS is comparable to that of MRI for cervical cancer staging and assessment of parametrial involvement [29, 30]. In CBCT guided cervical cancer ART, a first finding is that the combination of CBCT and US further improves the accuracy of the detection of the uterus [31].

1.3.3 Computerized tomography (CT) and cone-beam computed tomography(CBCT)

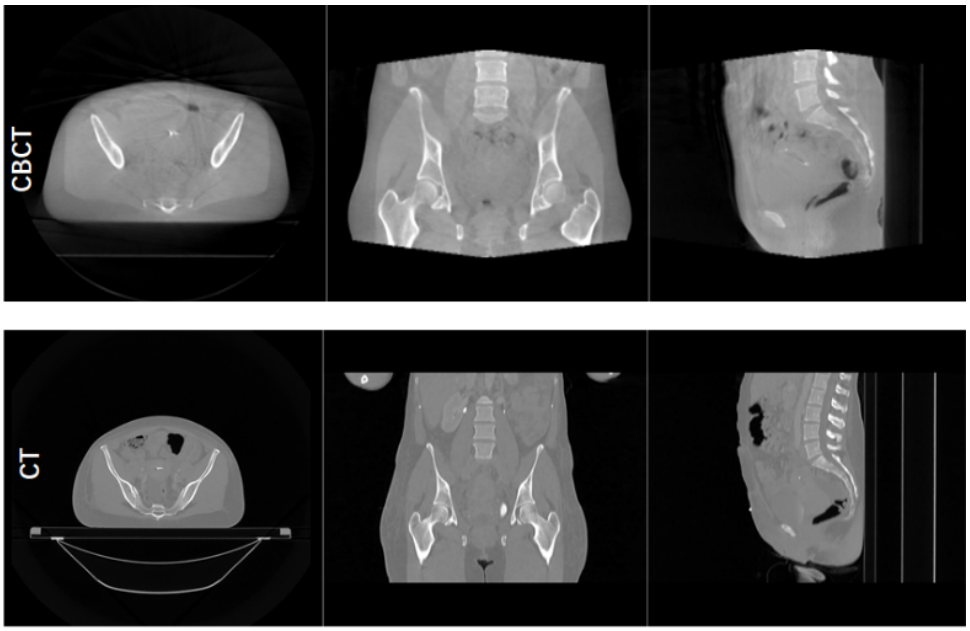

Figure 1.10 – CBCT and CT images of cervical cancer in 3 different planes. Comparing the quality of CBCT (above) and CT (below) images, CBCT images are of lower quality and have lower contrast in the soft tissues.

Computed tomography (CT) is a widely used imaging modality. CT is more cost-effective than MRI and is more common worldwide. In cervical cancer staging, CT is used primarily to evaluate the size of the cervix and to assess the recurrence assessment of patients. In cervical cancer ART, planning CTs are used to allow Radiophysicist to obtain 3D images of the area being treated in order to create individualized radiotherapy plans including dose distributions planning, patient alignment and radiation beam optimization [32].

Cone-beam CT (CBCT) is an effective image-guided radiotherapy (IGRT) tool for verifying the patient position. It also allows the real-time re-evaluation of the treatment plans for adaptive radiation therapy (ART). However, the relatively poor image quality of CBCT and the particularly large variability in Hounsfield unit values (HU) pose problems for its use as a valid tool for ART because these fluctuations in HU values can affect the accuracy of dose calculations [33]. As shown in Figure 1.10, CBCT images are of quality and have lower soft tissues contrast.

1.4 Challenges

In this Section, we focus on a review and discussion of the different image segmentation and alignment algorithms for the treatment of uterine fibroids and cervical cancer using HIFU and ART respectively.

Here we present the main scientific challenges, which we address in this thesis.

1.4.1 Multi-class segmentation of uterine Regions from MR images

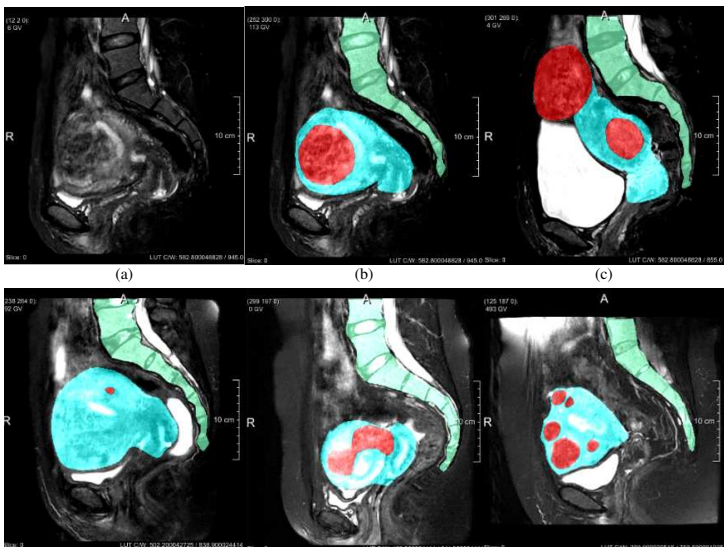

Figure 1.11 – MR images of the uterine regions in different patients. Red indicates the fibroids, blue the uterus, and green the spine. (a) Raw MR image of Patient 71, slice14. The labeled images of: (b) Patient 71, slice14; (c) Patient 84 slice14; (d) Patient 93, slice12; (e) Patient 26 slice12; and (f) Patient 8, slice13. We can observe: 1) large variations in shape and size between individuals; 2) low contrast between adjacent organs and tissues; 3) uterine fibroids that vary greatly in numbers and shapes.

The segmentation of uterus and uterine fibroids is a prerequisite step for the planning of a HIFU treatment. However, the segmentation of the spine is also important in order to avoid

any injury to the spinal cord. Manual delineation of the uterus, fibroids, and spine is a tedious, time-consuming task and is subject to intra- and inter-expert variability during both pre- and post-treatment. Thus, an automatic and accurate segmentation method capable to extract all these structures is of great importance.

Such an objective is challenging because of **1) the large shape and size variations among individuals.** As shown in Figure 1.11, uterine and fibroids are highly variable between patients; **2) a poor contrast between adjacent organs and tissues.** The contrast between uterus and uterine fibroids is quite low, so the boundaries between the organs are difficult to distinguish; **3) the number of uterine fibroids and their shapes are unknown.** For the above mentioned reasons, the existing methods dealing with uterine fibroid segmentation are often applied after treatment, while the pre-treatment is always performed manually by an operator to mark the uterus, fibroids and surrounding organs.

Recently, deep learning (DL) has achieved tremendous progress in medical image segmentation. These fully-supervised learning (FSL)-based methods can handle various medical images segmentation tasks. However, the accuracy and robustness of the DL methods depend on a large number of learning data annotated by experts. Acquiring good and accurate annotations requires laborious work, and the results of inter-expert delineation vary.

1.4.2 Semi-supervised learning-based multi-class image segmentation

When solving segmentation problems with deep learning methods, it is often necessary to annotate a large amount of data to satisfy the training of the neural network. However, in clinical practice, it is difficult to obtain data due to the privacy-protective nature of medical data and the reliance on specialist doctors for accurate annotation. In recent years, semi-supervised learning has also been used in medical image segmentation. Semi-supervised methods require only a small amount of data to be annotated. A small amount of annotated data is fed into the network with a large amount of unannotated data during training. Existing semi-supervised methods that use pseudo-labels obtained from training with unlabelled data to expand the trainable dataset are widely used. To further promote the use of HIFU surgery in clinical practice, this paper considers using semi-supervised methods to segment the uterine region.

However, the quality of the pseudo-label affects the accuracy of the segmentation method, and the segmentation quality of each target on the pseudo-label varies due to the different difficulties of the segmentation targets in the uterine region. Therefore, optimizing the quality of pseudo-labels for multiple classes is the main challenge in this paper in segmenting uterine images using semi-supervised segmentation.

1.4.3 Automated segmentation for plan-of-the-day (PoD) selection in CBCT-guided cervical cancer ART

The complex ART flowchart (as shown in Figure 1.7), including daily image observation and analysis, can increase the workload and hinder the deployment of advanced ART strategies in clinical routine. In the standard flowchart, the manual delineation of CBCT images for each patient is actually impractical. And the manual selection of the PoD for each fraction relys on the visual comparison. These time-consuming process limits rapid progress of ART. CBCT image plays an important role in ART, it can provide the latest anatomical information about the patient or the patient repositioning. However, the quality of CBCT images is relatively low due to noise, artifacts and low soft tissue contrast. These challenges make manual annotation difficult and time-consuming.

Therefore, the automatic segmentation of CBCT images and PoD selection are essential in ART.

1.5 Thesis aims

The main aim of this thesis is to develop new methods for image-guided surgeries to treat uterine fibroids and cervical cancer. Moreover, the proposed methods may contribute to improve the accuracy, efficiency and robustness of these clinical procedures.

Our objective is to address the above challenges, so we list the following aims:

1) HIFU therapy: To address the segmentation problem in HIFU therapy. To facilitate the development of HIFU therapy with automatic and accurate segmentation of the uterine region in preoperative MR images in multiple categories. This work on segmentation in MRI will be described in chapter 3.

2) HIFU: To solve the problem of semi-supervised segmentation in HIFU therapy. It is difficult to obtain the large amount of annotated data required for fully supervised deep learning methods in the field of clinical data. Therefore, investigate how semi-supervised learning can be performed using limited annotated data. This work on segmentation in MRI will be described in chapter 4.

2) ART: We propose an automatic workflow to select the optimal treatment plan. It relies on a deep learning-based segmentation of the CBCT images, enabling to select the optimal treatment plan (PoD selection) regarding the CTV coverage based on a geometrical criterion. This contribution on segmentation in CBCT will be presented in chapter 5.

BIBLIOGRAPHY

[1] J. Hodler, R. A. Kubik-Huch, and G. K. von Schulthess, "Diseases of the abdomen and pelvis 2018-2021: Diagnostic imaging-idkd book," 2018.

[2] H. Ellis, "Anatomy of the uterus," Anaesthesia & Intensive Care Medicine, vol. 12, no. 3, pp. 99–101, 2011.

[3] R. Balasubramanya and C. Valle, "Uterine imaging. treasure island (fl): Statpearls publishing," 2019.

[4] C. E. Board, "Uterine Cancer: Introduction," https://www.cancer.net/cancer-types/uterine-cancer/introduction, 2021, accessed: 2022-05-13.

[5] E. A. Stewart, "Uterine fibroids," The Lancet, vol. 357, no. 9252, pp. 293–298, 2001.

[6] E. A. Stewart and R. A. Nowak, "Uterine fibroids: hiding in plain sight," Physiology, vol. 37, no. 1, pp. 16–27, 2022.

[7] D. D. Baird, D. B. Dunson, M. C. Hill, D. Cousins, and J. M. Schectman, "High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence," American journal of obstetrics and gynecology, vol. 188, no. 1, pp. 100–107, 2003.

[8] M. Clinic, "Uterine fibroids," https://www.mayoclinic.org/diseases-conditions/uterine-fibroids/symptoms-causes/syc-20354288, 2019, accessed: 2022-05-16.

[9] S. S. Singh and L. Belland, "Contemporary management of uterine fibroids: focus on emerging medical treatments," Current medical research and opinion, vol. 31, no. 1, pp. 1–12, 2015.

[10] E. A. Stewart, "Clinical practice. uterine fibroids." The New England journal of medicine, vol. 372, no. 17, pp. 1646–1655, 2015.

[11] M. Dueholm, E. Lundorf, E. S. Hansen, S. Ledertoug, and F. Olesen, "Accuracy of magnetic resonance imaging and transvaginal ultrasonography in the diagnosis, mapping, and measurement of uterine myomas," American journal of obstetrics and gynecology, vol. 186, no. 3, pp. 409–415, 2002.

[12] J. Donnez and M.-M. Dolmans, "Uterine fibroid management: from the present to the future," Human reproduction update, vol. 22, no. 6, pp. 665–686, 2016.

[13] K. J. Anneveldt, H. J. van't Oever, I. M. Nijholt, J. R. Dijkstra, W. J. Hehenkamp, S. Veersema, J. A. Huirne, J. M. Schutte, and M. F. Boomsma, "Systematic review of reproductive outcomes after high intensity focused ultrasound treatment of uterine fibroids," European Journal of Radiology, vol. 141, p. 109801, 2021.

[14] Y. Wang, Z.-B. Wang, and Y.-H. Xu, "Efficacy, efficiency, and safety of magnetic resonance-guided high-intensity focused ultrasound for ablation of uterine fibroids: comparison with ultrasound-guided method," Korean Journal of Radiology, vol. 19, no. 4, pp. 724–732, 2018.

[15] H. E. Cline, J. Schenek, K. Hynynen, and R. D. Watkins, "MR-guided focused ultrasound surgery," Journal of computer assisted tomography, vol. 16, pp. 956–956, 1992.

[16] R. Clarke and G. Ter Haar, "Temperature rise recorded during lesion formation by high-intensity focused ultrasound," Ultrasound in medicine & biology, vol. 23, no. 2, pp. 299–306, 1997.

[17] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: a cancer journal for clinicians, vol. 68, no. 6, pp. 394–424, 2018.

[18] P. A. Cohen, A. Jhingran, A. Oaknin, and L. Denny, "Cervical cancer," The Lancet, vol. 393, no. 10167, pp. 169–182, 2019.

[19] S. E. Waggoner, "Cervical cancer," The Lancet, vol. 361, no. 9376, pp. 2217–2225, 2003.

[20] D. Yan, F. Vicini, J. Wong, and A. Martinez, "Adaptive radiation therapy," Physics in Medicine & Biology, vol. 42, no. 1, p. 123, 1997.

[21] K. K. Brock, "Adaptive radiotherapy: moving into the future," in Seminars in radiation oncology, vol. 29, no. 3. NIH Public Access, 2019, p. 181.

[22] A. Webster, A. Appelt, and G. Eminowicz, "Image-guided radiotherapy for pelvic cancers: a review of current evidence and clinical utilisation," Clinical Oncology, vol. 32, no. 12, pp. 805–816, 2020.

[23] A. Hunt, V. Hansen, U. Oelfke, S. Nill, and S. Hafeez, "Adaptive radiotherapy enabled by MRI guidance," Clinical Oncology, vol. 30, no. 11, pp. 711–719, 2018.

[24] P. Sibolt, L. M. Andersson, L. Calmels, D. Sjöström, U. Bjelkengren, P. Geertsen, and C. F. Behrens, "Clinical implementation of artificial intelligence-driven cone-beam computed tomography-guided online adaptive radiotherapy in the pelvic region," Physics and imaging in radiation oncology, vol. 17, pp. 1–7, 2021.

[25] Y. Griffin, V. Sudigali, and A. Jacques, "Radiology of benign disorders of menstruation," in Seminars in Ultrasound, CT and MRI, vol. 31, no. 5. Elsevier, 2010, pp. 414–432.

[26] H. Hricak, C. Alpers, L. E. Crooks, and P. E. Sheldon, "Magnetic resonance imaging of the female pelvis: initial experience," American Journal of Roentgenology, vol. 141, no. 6, pp. 1119–1128, 1983.

[27] K. E. Rice, J. R. Secrist, E. L. Woodrow, L. M. Hallock, and J. L. Neal, "Etiology, diagnosis, and management of uterine leiomyomas," Journal of Midwifery & Women's Health, vol. 57, no. 3, pp. 241–247, 2012.

[28] S. Wilde and S. Scott-Barrett, "Radiological appearances of uterine fibroids," Indian Journal of Radiology and Imaging, vol. 19, no. 03, pp. 222–231, 2009.

[29] D. Fischerova, D. Cibula, H. Stenhova, H. Vondrichova, P. Calda, M. Zikan, P. Freitag, J. Slama, P. Dundr, and J. Belacek, "Transrectal ultrasound and magnetic resonance imaging in staging of early cervical cancer," International Journal of Gynecologic Cancer, vol. 18, no. 4, 2008.

[30] F. Moloney, D. Ryan, M. Twomey, M. Hewitt, and J. Barry, "Comparison of mri and high-resolution transvaginal sonography for the local staging of cervical cancer," Journal of Clinical Ultrasound, vol. 44, no. 2, pp. 78–84, 2016.

[31] S. A. Mason, I. M. White, T. O'Shea, H. A. McNair, S. Alexander, E. Kalaitzaki, J. C. Bamber, E. J. Harris, and S. Lalondrelle, "Combined ultrasound and cone beam CT improves target segmentation for image guided radiation therapy in uterine cervix cancer," International Journal of Radiation Oncology* Biology* Physics, vol. 104, no. 3, pp. 685–693, 2019.

[32] D. Paquin, D. Levy, and L. Xing, "Multiscale registration of planning CT and daily cone beam CT images for adaptive radiation therapy," Medical physics, vol. 36, no. 1, pp. 4–11, 2009.

[33] K. Srinivasan, M. Mohammadi, and J. Shepherd, "Applications of linac-mounted kilovoltage Cone-beam Computed Tomography in modern radiation therapy: A review," Polish journal of radiology, vol. 79, p. 181, 2014.

Chapter 2

Deep learning-based Medical Image segmentation

2.1 Overview

2.1.1 Image segmentation

In computer vision, image segmentation is the process of subdividing a digital image into multiple sub-regions of the image. The purpose of image segmentation is to simplify or change the image representation to make the image easier to understand and analyze. An image is a collection of different pixel points, so image segmentation can also be seen as a grouping or classification of these pixel points. In general, image segmentation can be divided into semantic segmentation [1, 2] and instance segmentation [3]. The semantic segmentation assigns a class to each pixel of the image, but objects of the same category are not distinguished. Instance segmentation, on the other hand, only classifies specific objects. It is similar to target detection, except that target detection outputs the bounding box and category of the target, while instance segmentation outputs the mask and category of the target.

Symbolically, let the set R represent pixels point of the whole image. The segmentation of R can be seen as partitioning R into N nonempty subsets [1] R_1, R_2, R_3,..., R_N according to some uniformity predicate P (some uniformity or similarity rules related to the values of the pixel) such as [4, 5, 6]:

$$\bigcup_{i=1}^{N} R_i = R \tag{2.1}$$

$$R_i, \quad i = 1, 2, \ldots, N \text{ is connected} \tag{2.2}$$

$$P(P_i) = \text{ TRUE for } i = 1.2 \ldots N \tag{2.3}$$

1. A subset R_i consists of contiguous pixel points.

$$P\left(R_i \cup R_j\right) = \text{FALSE for } i \neq j \tag{2.4}$$

where R_i and R_j are adjacent subsets and P is a uniform predicate which is true for each subset. The first condition implies that every picture point must be in a region. This means that the segmentation algorithm should not terminate until every point is processed. The second condition implies that regions must be connected, i.e. composed of contiguous lattice points. The third condition determines what kind of properties the segmented regions should have, for example, uniform gray levels. The fourth condition expresses the maximality of each region in the segmentation.

Numerous image segmentation methods have been proposed in the literature. In addition to the deep learning (DL)-based methods described in detail in 2.2.1, the other techniques are as follows:

1. **Thresholding** [7, 8, 9]: Thresholding is one of the most widely used and simplest image segmentation methods. Based on the pixel intensity of the original image, we set a threshold value to select the region of interest in one image. In thresholding setting process, we can choose a appropriate threshold by considering the feature histogram of the image to be segmented. In particular, feature histograms can include: grayscale histograms, gradient histograms, texture histograms, etc.
2. **Region-growing**[10, 11, 12, 13, 14]: The main idea of region-growing is to merge adjacent pixel points with similar properties. For each region, a seed point is assigned as the starting point for growth, then the pixel points in the field around the seed point are compared with the seed point, and the points with similar properties are merged together and continue to grow outward until no pixels satisfying the conditions are included.
3. **Clustering** [15, 16]: Clustering is the process of finding different groups in the feature space, where the pixels in each group are more closely related to each other than the pixels that are assigned to different groups. The general steps of the clustering method are: 1) initialize a coarse cluster; 2) cluster pixel points with similar features to the same superpixel in an iterative manner until convergence to obtain the final segmentation. The clustering-based segmentation methods are: K-means segmentation, Fuzzy C-means clustering, mountain clustering method and substractive clustering method [15].
4. **Watershed methods** [17, 18, 19]: The concept of a watershed comes from a topographical analogy. Think of an image as a representation of three-dimensional(3D) topography: a two-dimensional (2D) land base (image space) and the third dimension of height (image grayscale). We can represent areas of high-intensity as peaks and areas of low-intensity as valleys. To separate the objects in the image, we will fill each valley with water of different colors. Slowly, the water will rise to a point where the water from the different valleys

begins to merge. At this point, we build barriers on the tops of the mountains to prevent them from being flooded with water. These barriers are the segmentation boundaries.

5. **Active contours**[20, 21]: The main principle of the active contour model is to build curves that fit the edges of the objects with a minimum of energy. This minimization allows to find a compromise between the attachment to the data (attraction towards the edges) and the complexity of the curve (torsion, ...). The contour curve gradually approaches the edge of the object to be detected and finally segments the target.
6. **Graph cuts**[22, 23]: The graph cut is an optimization method for energy functions. Representing an image as an undirected graph, the pixel points in the graph are the vertices of the graph, and the connection of every two four-neighborhood vertices is an edge (called n-links). Two terminal vertices (foreground target and background) are connected to the vertices representing each pixel and form edges (called t-links). Each edge has a cost, and a cost function is defined such that the sum of the costs of a cut (a subset of the set of edges) is minimized, which is the result of the graph cut. Common algorithms are: Normalized cuts segmentation [24] and MRFs (Markov Random Fields) graph cuts segmentation [25].

2.2 Deep Learning-based medical image segmentation

Medical images share the basic image properties with natural images, so some of the segmentation methods mentioned in section 2.1 can be applied to medical images. For example: image segmentation using thresholding [26], region-growing [27], k-means [28], watershed [29]. Pham *et al.* summarized the conventional methods of medical images segmentation.

However, some differences between natural and medical images make some of the methods that work well on the former, fail when applied to the latter. These differences are in the following aspects:

1. Medical images have non-uniform noise distributions and artifacts due to the single light source and the thickness of the body. On the other hand, the noise distribution in natural images approximates Gaussian noise because the light field distribution can be considered as uniform.
2. Medical images have many forms of information, such as 2D grayscale, 2D with 4 channels, 3D volumes and even 4D. They also have spatial resolution, scan parameters, field of view (FOV), and other information. Natural images are generally 2D RGB images.
3. Medical images have multi-modal and multi-view information. These are not available in natural images.
4. Medical images are more difficult to acquire (from patients) and more challenging to delineate due to the need of medical expertise.

5. Medical images analysis requires greater details. For example, for lung nodules detection, the target lesion area is tiny compared to the background, so it is challenging to accurately detect the location of lung nodules.

In the past decade, the development of DL has led to advances and great success in medical image processing [30]. Since DL-based image segmentation methods can automatically extract a huge number of meaningful features from the characteristics of the data itself, DL methods are simpler and more adaptable to process medical images of different modalities than traditional methods designed to deal with morphological features or intensity information.

2.2.1 Fully-, semi- and self-supervised learning

With the development of DL, which has played a considerable role in advancing the field of image processing, heavy reliance on large amounts of labeled data has frustrated some image tasks. Therefore, gradually reducing the reliance on annotated datasets, including the need for large data volumes and fine-grained annotations, has become a hot concern in the industry. DL has gradually evolved from traditional fully supervised learning to semi-(weakly-)supervised, self-supervised learning.

In fully supervised learning, a large amount of labeled data is needed to train the model, and the model's prediction and the data's annotation generate losses followed by backpropagation (calculating gradients, updating parameters). The above process is repeated until the model obtains the expected learning capability.

Semi-supervised learning [31, 32] attempts to learn from both unlabeled and labeled samples, usually assuming sampling from the same or similar distributions. Weakly supervised learning [33] includes incomplete supervision (only some of the labels are given), inexact supervision (the labels of the training data are coarse-grained), and inaccurate supervision (the labels given are not always correct). Semi-supervised learning belongs to incomplete supervision. In this paper, we focus on semi-supervised learning.

Self-supervised learning [34] constructs semantically meaningful image representations by using a pretext task that does not require semantic annotation. The pretext task is typically performed by transforming the input image and requiring the learner to predict the properties of the transformation from the transformed image.

2.2.2 Fully-supervised learning (FSL) for medical image segmentation and limitations

Convolutional neural networks (CNNs) are the most widely used architectures for processing medical images. They are composed of convolutional layers, pooling layers, normalization layers and fully connected layers. Hesamian *et al.* [35] summarized the popular DL techniques for

medical image segmentation and categorized the approaches in the following aspects:

1. **CNN**: 2D CNN refer to CNN networks that accept input images in the form of 2D images. Zhang *et al.* [36] used a deep 2D CNN to segment infant brain tissue from input T1, T2 and fractional anisotropy (FA) images. Furthermore, in order to extract more spatial information, some methods [37, 38] developed 2.5D CNNs that feed the network with orthogonal 2D patches of XY,YZ,XZ planes. Then, the convolutional kernel was extended from 2D to 3D thanks to the improvement of the hardware and allowed to get better performance than 2.5D CNN. The 3D CNN can now support 3D patches of the data with the original spatial information. Urban *et al.* [39] proposed a 3D CNN model to segment the brain tumor.
2. **Fully convolutional network (FCN)**: FCN replaces the classical last fully connected layer with a fully convolutional layer to get a dense pixel-wise prediction. Nie *et al.* [40] proposed multi-FCNs to segment isointense-phase brain images from three modality images (T1, T2 and FA).
3. **U-Net**: U-Net was first proposed by Ronneberger *et al.* [41] in 2015 to segment 2D biomedical image and then it has been widely used in medical image segmentation with success. Its shape is like U-shape with a symmetrical encoder-decoder architecture. Based on the U-Net, other modified methods are developed to solve more medical image tasks, such as: V-Net [42], 3D U-Net [43], UNet++[44].
4. **Convolutional Residual Networks (CRNs)**: Based on the residual block in ResNet [45], 2D CRNS [46] and 3D CRNs [47] are proposed to successfully improve the segmentation accuracy in medical images.
5. **Recurrent Neural Networks (RNNs)**: The recurrent neural block is designed to extract contextual information from sequential data. But it can also be applied to volumetric medical images to memorize inter-slice spatial information across adjacent slices. The most popular type is the long short-term memory (LSTM)-based CNN [48].

Fully supervised segmentation methods achieve significant performance in medical images. However, fully supervised segmentation methods also face challenges and limitations. These methods rely on a huge number of annotated data. While in the medical field, the clinical data should be collected from the patients, and the annotation process is tedious and time-consuming with the involvement of medical experts.

2.2.3 Semi-supervised and self-supervised learning (S^4L) for medical image segmentation and limitations

In traditional FSL methods, the training of the model relies on a large amount of highly accurate labeled data. The emergence of S^4L gets rid of the reliance on large amounts of labeled

data. Therefore, the S^4L approach is closer to the real-world application.

In semi-supervised learning methods, the main idea is to improve the model performance by training the labeled data and then using the unlabeled data as constraints. The algorithms can be divided into the 4 following categories:

1. **Graph-based methods**: The graph has convexity, scalability and effectiveness in modeling relationships between different entities [49]. Kipf *et al.* [50] presented an approach using the graph topology and the nodes side information for semi-supervised classification.
2. **Generative Adversarial Nets (GAN)-based methods**: GAN-based semi-supervised learning methods can generate a perfect discriminator by learning imperfect generators on labeled and unlabeled data. Hung *et al.* [51] applied adversarial learning for semi-supervised semantic segmentation by combining two semi-supervised loss terms to leverage the unlabeled data.
3. **Self-training or co-training**: Self-training or co-training is a proxy label method that produces proxy labels on unlabeled data without supervision. Self-training methods use labeled data to pre-train a model and then predict the pseudo-labels on unlabeled data. Yalniz *et al.* [52] trained a teacher-student model to exploit the large-scale unlabeled data and achieved a 4.8% accuracy improvement compared to ResNet-50. Co-training was originally proposed to describe a model in which unlabeled data is used to augment labeled data based on two views of an example [53]. Inspired by this model, Qiao *et al.* [54] extended co-training to deep co-training for semi-supervised image recognition. Especially, adversarial examples are used in different views to prevent a model from collapsing.
4. **Consistency training**: Consistency regularization allows to obtain similar output results for the same input with different data enhancements or networks. Π-Model and temporal ensembling [55] are typical implementations of consistency regularization. However, temporal ensembling reaches its limits for large datasets because each target is updated once per epoch. Tarvainen *et al.* [56] overcame the problem by averaging model weights instead of predictions. The method is called Mean Teacher, which includes two networks: the teacher network and the student network. The two networks have the same structure but are updated in different ways. The student network updates parameters by back-propagating gradient descent and the teacher network updating parameters by exponential moving average (EMA) of the student network parameters. Compared with the temporal ensembling, Mean Teacher can update the moving average of the network parameters once per backpropagation, which is more efficient. Ouali *et al.* [57] proposed a cross-consistency training (CCT) network, in which predictions invariance is enforced over different perturbations applied to the encoder outputs. Besides, the adoption of ad-

versarial learning can enforce the segmentation distributions of unannotated images to be similar to those of the annotated images.

Semi-supervised learning approaches have been widely applied in medical image segmentation tasks. Nie *et al.* [58] employed an adversarial network named ASDNet to produce unannotated high-confidence data to train the segmentation network. Li *et al.* [59] used transformation consistency learning to do the semi-supervised skin lesion segmentation and got a new record on the ISIC2017 skin lesion segmentation challenge. Bai *et al.* [60] proposed an semi-supervised learning framework in which the pseudo labels of the unlabeled data were obtained by non-rigid image registration in the cardiac cycle. This method was evaluated on a short-axis cardiac MR image dataset and obtained mean Dice values of 0.92, 0.85, and 0.89 for the left ventricular (LV) cavity, LV myocardium, and right ventricular (RV) cavity, respectively. Similarly, Ito *et al.* [61] also used the registration-based semi-supervised learning method to achieve brain tissue segmentation and evaluated it on human and marmoset brain image datasets to show the effectiveness of the method.

Self-supervised learning is a recent training paradigm that does not require labeled data. Specifically, it involves extracting supervised information from unlabeled data and thus learning robust data representations. This can be considered as an effective approach to solve the problem of sparse annotated medical data. In self-supervised learning, two tasks need to be defined. One is the pretext task, which is used to perform useful feature learning from unlabeled data. The second, called the down-stream task, is used to transfer and fine-tune the concepts learned in the pretext task to achieve the final task goal. Shurrab *et al.* [62] reviewed and analysed the recent self-supervised medical image analysis methods.They classified self-supervised learning pretext tasks into three categories, predictive, generative, and contrastive.

The predictive pretext task learns latent features in the input data by treating the pretext task as a classification problem. Taleb *et al.* [63] used Jigsaw puzzle as the pretext method. Specifically, they introduced a multimodal puzzle task, which is beneficial for learning rich representations from multiple image modalities. The final down-stream tasks include brain tumor segmentation and prostate segmentation, as well as liver segmentation using unregistered CT and MRI modalities, demonstrating the effectiveness of the method.

The generative pretext task aims to learn latent features throughout the reconstruction process. Hervella *et al.* [64] proposed the use of the multimodal reconstruction between retinography and fluorescein angiography as a common self-supervised pre-training task. After fine-tuning of the pre-trained network, the self-learning model can address the localization and segmentation of the main anatomical structures of the eye fundus.

The contrastive task is one of the most popular scheme recently in self-supervised learning for its comparable to supervised learning methods. It aims to develop robust representations from the input data. The model maximizes the consistency between different transformed views of

the same image while minimizing the consistency between transformed views of different images, thereby acquiring general representations. MoCo [65]and SimCLR [66] are two widely used constrastive learning approaches and also have also been shown to achieve significant performance in self-learning medical image analysis tasks [67, 68]. MoCO is mainly used in classification tasks. SimCLR can support the transfer to different down-stream tasks due to its simple but effective design. In medical image segmentation, Chaitanya *et al.* [69] improved the SimCLR by using new contrastive strategy and contrastive loss to learn distinctive representations of local regions that are useful for per-pixel segmentation. The experimental evaluation was performed on the cardiac and prostate segmentation tasks.

S^4L methods have achieved segmentation performance comparable to FSL segmentation methods, although they are free to some extent from strong data dependence. However, they still face some challenges and limitations.

In semi-supervised learning method, semi-supervised learning relies on labeled data distribution features. For multi-center medical image data, the unlabeled data maybe misaligned with the labeled data, the semi-supervised learning segmentation method maybe not robust and generalized enough. Besides, in the process of pseudo-labels generation, the quality of the pseudo-labels will affect the optimization of the model when it is updated. In the existing methods, the generation of pseudo-labels relies on time-consuming artificial offline selection, usually based on experience or after experiments on a small validation set. Then a threshold is set to generate a credible confidence map. Therefore, an adaptive threshold strategy needs to be developed for an automatic adaptation to different semi-supervised data distributions. This should help to improve the generalization and robustness of the semi-supervised methods.

In terms of self-supervised learning, there are few studies focusing on embedding medical knowledge into pretext tasks. If prior medical knowledge can be taken into account in the design of pretext tasks, the model can be closer to the down-stream task. In addition, studying how to combine transfer learning with self-supervised learning can help to get better results when transferring data representations to down-stream tasks.

2.3 Conclusion

This chapter introduces the basic concepts of image segmentation and related algorithms. Traditional image segmentation methods are first investigated, followed by a description of recent deep learning-based medical image segmentation algorithms of fully supervised and semi-(self-) supervised, and an analysis and summary of the current drawbacks and limitations of each type of segmentation algorithm.

BIBLIOGRAPHY

[1] Y. Guo, Y. Liu, T. Georgiou, and M. S. Lew, "A review of semantic segmentation using deep neural networks," International journal of multimedia information retrieval, vol. 7, no. 2, pp. 87–93, 2018.

[2] S. Asgari Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad, and G. Hamarneh, "Deep semantic segmentation of natural and medical images: a review," Artificial Intelligence Review, vol. 54, no. 1, pp. 137–178, 2021.

[3] A. M. Hafiz and G. M. Bhat, "A survey on instance segmentation: state of the art," International journal of multimedia information retrieval, vol. 9, no. 3, pp. 171–189, 2020.

[4] S. L. Horowitz, "Picture segmentation by a directed split-and-merge procedure," in IJCPR, 1974, pp. 424–433.

[5] T. Pavlidis, "Structural pattern recognition," Springer Series in Electrophysics, 1977.

[6] K.-S. Fu and J. Mui, "A survey on image segmentation," Pattern recognition, vol. 13, no. 1, pp. 3–16, 1981.

[7] N. Otsu, "A threshold selection method from gray-level histograms," IEEE transactions on systems, man, and cybernetics, vol. 9, no. 1, pp. 62–66, 1979.

[8] A. Perez and R. C. Gonzalez, "An iterative thresholding algorithm for image segmentation," IEEE transactions on pattern analysis and machine intelligence, no. 6, pp. 742–751, 1987.

[9] H. Cai, Z. Yang, X. Cao, W. Xia, and X. Xu, "A new iterative triclass thresholding technique in image segmentation," IEEE transactions on image processing, vol. 23, no. 3, pp. 1038–1046, 2014.

[10] S. Hojjatoleslami and J. Kittler, "Region growing: a new approach," IEEE Transactions on Image processing, vol. 7, no. 7, pp. 1079–1084, 1998.

[11] R. Adams and L. Bischof, "Seeded region growing," IEEE Transactions on pattern analysis and machine intelligence, vol. 16, no. 6, pp. 641–647, 1994.

[12] Y.-L. Chang and X. Li, "Adaptive image region-growing," IEEE transactions on image processing, vol. 3, no. 6, pp. 868–872, 1994.

[13] T. Pavlidis and Y.-T. Liow, "Integrating region growing and edge detection," IEEE transactions on pattern analysis and machine intelligence, vol. 12, no. 3, pp. 225–233, 1990.

[14] S.-Y. Wan and W. E. Higgins, "Symmetric region growing," IEEE Transactions on Image processing, vol. 12, no. 9, pp. 1007–1015, 2003.

[15] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image segmentation using K-means clustering algorithm and subtractive clustering algorithm," Procedia Computer Science, vol. 54, pp. 764–771, 2015.

[16] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," science, vol. 315, no. 5814, pp. 972–976, 2007.

[17] A. P. Mangan and R. T. Whitaker, "Partitioning 3D surface meshes using watershed segmentation," IEEE Transactions on Visualization and Computer Graphics, vol. 5, no. 4, pp. 308–321, 1999.

[18] L. Shafarenko, M. Petrou, and J. Kittler, "Automatic watershed segmentation of randomly textured color images," IEEE transactions on Image Processing, vol. 6, no. 11, pp. 1530–1544, 1997.

[19] G. Hamarneh and X. Li, "Watershed segmentation using prior shape and appearance knowledge," Image and Vision Computing, vol. 27, no. 1-2, pp. 59–68, 2009.

[20] M. Kass, A. P. Witkin, and D. Terzopoulos, "Snakes: Active contour models," International Journal of Computer Vision, vol. 1, pp. 321–331, 2004.

[21] S. Osher and R. Fedkiw, "Level set methods and dynamic implicit surfaces," in Applied mathematical sciences, 2003.

[22] F. Yi and I. Moon, "Image segmentation: A survey of graph-cut methods," in 2012 International Conference on Systems and Informatics (ICSAI2012), 2012, pp. 1936–1941.

[23] S. Vicente, V. Kolmogorov, and C. Rother, "Graph cut based image segmentation with connectivity priors," in 2008 IEEE conference on computer vision and pattern recognition. IEEE, 2008, pp. 1–8.

[24] J. Shi and J. Malik, "Normalized cuts and image segmentation," IEEE Transactions on pattern analysis and machine intelligence, vol. 22, no. 8, pp. 888–905, 2000.

[25] P. Kohli and P. H. Torr, "Efficiently solving dynamic Markov random fields using graph cuts," in Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, vol. 2. IEEE, 2005, pp. 922–929.

[26] N. Senthilkumaran and S. Vaithegi, "Image segmentation by using thresholding techniques for medical images," Computer Science & Engineering: An International Journal, vol. 6, no. 1, pp. 1–13, 2016.

[27] R. Pohle and K. D. Toennies, "Segmentation of medical images using adaptive region growing," in Medical Imaging 2001: Image Processing, vol. 4322. SPIE, 2001, pp. 1337–1346.

[28] M.-N. Wu, C.-C. Lin, and C.-C. Chang, "Brain tumor detection using color-based k-means clustering segmentation," in Third international conference on intelligent information hiding and multimedia signal processing (IIH-MSP 2007), vol. 2. IEEE, 2007, pp. 245–250.

[29] Y.-L. Huang and D.-R. Chen, "Watershed segmentation for breast tumor in 2-d sonography," Ultrasound in medicine & biology, vol. 30, no. 5, pp. 625–632, 2004.

[30] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," Annual review of biomedical engineering, vol. 19, p. 221, 2017.

[31] X. J. Zhu, "Semi-supervised learning literature survey," 2005.

[32] J. E. Van Engelen and H. H. Hoos, "A survey on semi-supervised learning," Machine Learning, vol. 109, no. 2, pp. 373–440, 2020.

[33] Z.-H. Zhou, "A brief introduction to weakly supervised learning," National science review, vol. 5, no. 1, pp. 44–53, 2018.

[34] I. Misra and L. v. d. Maaten, "Self-supervised learning of pretext-invariant representations," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 6707–6717.

[35] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: achievements and challenges," Journal of digital imaging, vol. 32, no. 4, pp. 582–596, 2019.

[36] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," NeuroImage, vol. 108, pp. 214–224, 2015.

[37] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. B. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention, vol. 16 Pt 2, pp. 246–53, 2013.

[38] K. Kushibar, S. Valverde, S. González-Villà, J. Bernal, M. Cabezas, A. Oliver, and X. Lladó, "Automated sub-cortical brain structure segmentation combining spatial and deep convolutional features," Medical Image Analysis, vol. 48, p. 177–186, 2018.

[39] G. Urban, M. Bendszus, F. Hamprecht, and J. Kleesiek, "Multi-modal brain tumor segmentation using deep convolutional neural networks," MICCAI BraTS (brain tumor segmentation) challenge. Proceedings, winning contribution, pp. 31–35, 2014.

[40] D. Nie, L. Wang, Y. Gao, and D. Shen, "Fully convolutional networks for multi-modality isointense infant brain image segmentation," 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1342–1345, 2016.

[41] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," ArXiv, vol. abs/1505.04597, 2015.

[42] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571, 2016.

[43] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," ArXiv, vol. abs/1606.06650, 2016.

[44] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," IEEE Transactions on Medical Imaging, vol. 39, pp. 1856–1867, 2020.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.

[46] E. Gibson, M. R. Robu, S. A. Thompson, P. J. E. Edwards, C. Schneider, K. S. Gurusamy, B. R. Davidson, D. J. Hawkes, D. C. Barratt, and M. J. Clarkson, "Deep residual networks for automatic segmentation of laparoscopic videos of the liver," in Medical Imaging, 2017.

[47] H. Chen, Q. Dou, L. Yu, J. Qin, and P.-A. Heng, "VoxResNet: Deep voxelwise residual networks for brain segmentation from 3d mr images," NeuroImage, vol. 170, pp. 446–455, 2018.

[48] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, "Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation," ArXiv, vol. abs/1506.07452, 2015.

[49] Y. Chong, Y. Ding, Q. Yan, and S. Pan, "Graph-based semi-supervised learning: A review," Neurocomputing, vol. 408, pp. 216–230, 2020, publisher: Elsevier.

[50] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," arXiv preprint arXiv:1609.02907, 2016.

[51] W.-C. Hung, Y.-H. Tsai, Y.-T. Liou, Y.-Y. Lin, and M.-H. Yang, "Adversarial learning for semi-supervised semantic segmentation," arXiv preprint arXiv:1802.07934, 2018.

[52] I. Z. Yalniz, H. Jégou, K. Chen, M. Paluri, and D. Mahajan, "Billion-scale semi-supervised learning for image classification," arXiv preprint arXiv:1905.00546, 2019.

[53] A. Blum and T. Mitchell, "Combining Labeled and Unlabeled Data with Co-Training y," p. 10.

[54] S. Qiao, W. Shen, Z. Zhang, B. Wang, and A. Yuille, "Deep co-training for semi-supervised image recognition," in Proceedings of the european conference on computer vision (eccv), 2018, pp. 135–152.

[55] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," arXiv preprint arXiv:1610.02242, 2016.

[56] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," arXiv preprint arXiv:1703.01780, 2017.

[57] Y. Ouali, C. Hudelot, and M. Tami, "Semi-supervised semantic segmentation with cross-consistency training," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 12 674–12 684.

[58] D. Nie, Y. Gao, L. Wang, and D. Shen, "Asdnet: Attention based semi-supervised deep networks for medical image segmentation," in International conference on medical image computing and computer-assisted intervention. Springer, 2018, pp. 370–378.

[59] X. Li, L. Yu, H. Chen, C.-W. Fu, and P.-A. Heng, "Semi-supervised skin lesion segmentation via transformation consistent self-ensembling model," arXiv preprint arXiv:1808.03887, 2018.

[60] W. Bai, O. Oktay, M. Sinclair, H. Suzuki, M. Rajchl, G. Tarroni, B. Glocker, A. King, P. M. Matthews, and D. Rueckert, "Semi-supervised learning for network-based cardiac MR image segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2017, pp. 253–260.

[61] R. Ito, K. Nakae, J. Hata, H. Okano, and S. Ishii, "Semi-supervised deep learning of brain tissue segmentation," Neural Networks, vol. 116, pp. 25–34, 2019, publisher: Elsevier.

[62] S. Shurrab and R. Duwairi, "Self-supervised learning methods and applications in medical imaging analysis: A survey," PeerJ Computer Science, vol. 8, p. e1045, 2022.

[63] A. Taleb, C. Lippert, T. Klein, and M. Nabi, "Multimodal self-supervised learning for medical image analysis," in IPMI, 2021.

[64] Á. S. Hervella, J. Rouco, J. Novo, and M. Ortega, "Learning the retinal anatomy from scarce annotated data using self-supervised multimodal reconstruction," Appl. Soft Comput., vol. 91, p. 106210, 2020.

[65] K. He, H. Fan, Y. Wu, S. Xie, and R. B. Girshick, "Momentum contrast for unsupervised visual representation learning," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9726–9735, 2020.

[66] T. Chen, S. Kornblith, M. Norouzi, and G. E. Hinton, "A simple framework for contrastive learning of visual representations," ArXiv, vol. abs/2002.05709, 2020.

[67] X. Chen, L. Yao, T. Zhou, J. Dong, and Y. Zhang, "Momentum contrastive learning for few-shot COVID-19 diagnosis from chest ct images," Pattern Recognition, vol. 113, pp. 107 826 – 107 826, 2021.

[68] Y. N. T. Vu, R. Wang, N. Balachandar, C. Liu, A. Ng, and P. Rajpurkar, "MedAug: Contrastive learning leveraging patient metadata improves representations for chest x-ray interpretation," in MLHC, 2021.

[69] K. Chaitanya, E. Erdil, N. Karani, and E. Konukoglu, "Contrastive learning of global and local features for medical image segmentation with limited annotations," ArXiv, vol. abs/2006.10511, 2020.

Chapter 3

HIFUNet: Multi-class Segmentation of Uterine Regions from MR Images Using Global Convolutional Networks for HIFU Surgery Planning

3.1 Introduction

Uterine fibroids are benign tumors, common and present in up to 25% of women [1]. High intensity focused ultrasound (HIFU) is a new noninvasive surgery method for treating uterine fibroids. Magnetic Resonance (MR) image is clinically used for their diagnosis and the guidance of the HIFU procedure. The segmentation of uterus and uterine fibroids is a prerequisite step for the planning of HIFU treatment. However, the segmentation of the spine is also important in order to avoid any injury to the spinal cord. Manual delineation of the uterus, fibroids, and spine is a time-consuming, tedious task and subject to intra-expert and inter-expert variability during both pre- and post- treatment. Thus, an automatic and accurate segmentation method capable to extract all these structures is of great importance.

Such an objective is challenging because of **1) large shape and size variations among individuals.** As it is shown in Figure 3.1, uterine and fibroids are highly variable in different patients; **2) a low contrast between adjacent organs and tissues.** The contrast among uterus and uterine fibroids is quite low, so the boundaries between organs are difficult to distinguish; **3) the number of uterine fibroids and their shapes are unknown.** These issues are illustrated in Figure 3.1. Due to the above reasons, the existing methods dealing with uterine fibroid segmentation are often applied after treatment, while the pre-treatment is still performed manually by an operator to mark uterus, fibroids and surrounding organs. Therefore, in order to facilitate the development of a treatment plan, a preoperative segmentation is required.

In recent years, deep learning (DL) methods have been widely used in medical image seg-

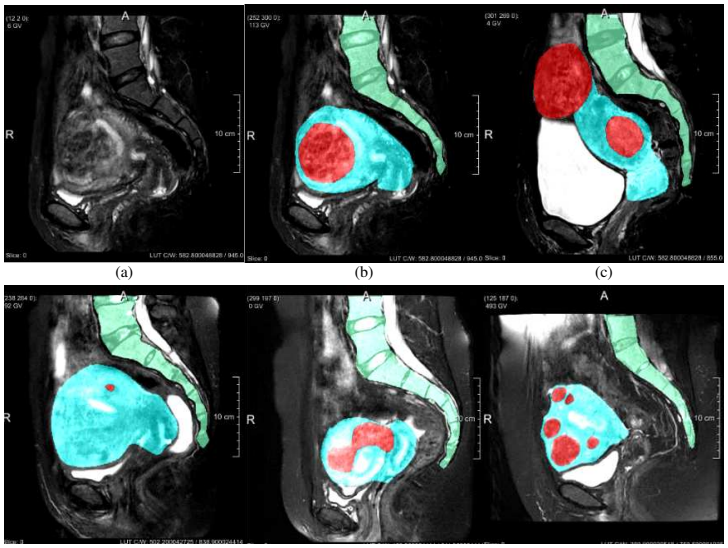

Figure 3.1 – MR images of uterus regions in different patients. Red denotes the fibroids, blue the uterus, and green the spine. (a) Patient 71 slice14 of raw MR image. (b-f) The labeled images of Patient 71 slice14, Patient 84 slice14, Patient 93 slice12, Patient 26 slice12, Patient 8 slice13. We can observe 1) large shape and size variations among individuals; 2) a low contrast between adjacent organs and tissues; 3) highly variable uterine fibroids numbers and shapes.

mentation [2, 3]. However, they have to face the overall complexity of the scenes under study. We propose here to derive comprehensive anatomical information through a global convolutional network (GCN) module based on a large valid receptive field and deep multiple atrous convolutions (DMAC) for hierarchically structuring the information. By doing so, the performance in locating and classifying the structures of interest can be improved.

Such semantic segmentation can be built upon the Encoder-Decoder architecture already widely utilized. Inspired by Fully Convolutional Network (FCN) [4] which was initially designed for image classification, U-Net was proposed for medical image segmentation by Ronneberger *et al.* [5] where the pooling operators in FCN are replaced by upsampling operators so that the output resolution can be retained at the same size as the input. The state-of-the-art results of U-Net in segmenting medical images, especially with small training dataset, show a promising ability of this Encoder-Decoder architecture. Basically, the Encoder aims to capture features and reduce the spatial dimensions while the Decoder aims to recover the object details and spatial dimension. Therefore, in order to improve the performance of image segmentation, more high-level features need to be automatically captured in the encoder and more spatial information can be saved in the decoder.

The U-Net was later extended in order to tackle different problems. Cicek *et al.* [6] modified the initial U-Net architecture by replacing all 2D operations with their 3D counterparts. Milletari *et al.* [7] presented a novel 3D segmentation approach (called V-Net) that leverages the power

of a fully convolutional neural network based on the Dice coefficient for processing volumetric medical images such as MR images. In addition, in contrast with 3D U-Net, the V-net formulates each stage by using a residual function which can accelerate the convergence rate. Many other U-Net based segmentation schemes have been further reported for retinal vessels, liver and tumors in CT scans, ischemic stroke lesion, intervertebral disc and pancreas [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18].

The U-Net shows a good segmentation performance with the usage of skip connections which can concatenate two feature maps of the same size in the corresponding parts of the encoder and decoder. The concatenated feature maps contain the information from both high and low levels, thus achieving feature fusion under different scales to improve the accuracy of model results. Even so, the complex anatomical scene involved in our HIFU therapy application remains a challenge. Large valid receptive fields play an important role in global scene observation. Global convolutional network [19] enables dense connections within a large region by using spatial decomposed convolution with a large kernel. It can capture multi-scale context cues with less computational cost than a general convolution with a large kernel. Therefore, we introduce layer-by-layer the GCN which has an efficient kernel parameter number to enlarge the receptive field in our Encoder-Decoder architecture.

In addition, getting the hierarchical structural information can help to provide more contextual information at various levels by using atrous convolutions. The key element of this method is to insert holes into the convolution kernels, which allows preserving the resolution and enlarging the receptive field. Recently, atrous convolution has been widely used in many deep learning architectures. DeepLab [20], based on FCN and atrous convolutions, maintains the receptive field unchanged. Besides, in order to get a better object segmentation at multiple scales, in DeepLabV2 [21], Chen *et al.* proposed a module called atrous spatial pyramid pooling (ASPP) which uses multiple parallel atrous convolutional layers with different sampling rates. The use of atrous convolutions preserves the spatial resolution of the final map and thus leads to higher performance when compared to most methods in Encoder-Decoder schemes. DeepLabV3+ [22] combines the advantages of Xception [23] and Encoder-Decoder, which employs DeepLabV3 [24] as the encoder.

However, the uncertainty regarding the location, the numbers and the sizes of uterine fibroids leads to an increase of complexity for segmentation and many existing deep learning segmentation models lack using features from different levels efficiently. Subsequently, in some cases, the targets can be segmented incorrectly. More effective feature extraction approaches are required for uterine fibroid segmentation.

Motivated by the above discussions and ResNet [25] structures, we propose a novel network named HIFUNet to segment uterus, uterine fibroids and spine automatically. The main contributions of the method can be summarized as follows:

1. To address the segmentation errors (*i.e., classifying uterine neck as uterine fibroid because of insufficient receptive field*), we introduce a global convolutional network module able to enlarge the receptive field effectively.
2. We integrate the global convolutional network and deep multiple atrous convolutions to further extract context-based semantic information and generate more abstract features for large scaled uterine fibroids.
3. The proposed HIFUNet behaves similarly to clinical experts and, as it will be shown through a large number of experiments, performs better than many existing semantic segmentation networks.
4. The segmentation of the uterus and uterine fibroids is, to the best of our knowledge, the first methodological attempt using convolutional neural networks in HIFU therapy. The inclusion of the spine segmentation, a critical organ in HIFU therapy, is another major feature of our approach.

3.2 Related Work

We sketch here the conventional methods proposed so far for segmenting the uterus and uterine fibroids and we review the state-of-the-art MR image segmentation methods based on CNN architectures.

3.2.1 Conventional Methods of Uterus and Uterine Fibroid Segmentation

Very few contributions have been reported for segmenting uterus and uterine fibroids from MR images. The main methods are summarized below:

Approaches based on level-set: Ben-Zadok *et al.* [26] presented an interactive level set segmentation framework that allows user feedback. It is a semi-automatic method where the users have to select seed-points. Khotanlou *et al.* [27] proposed a two-stage method combining the region-based level set [28] and the hybrid Bresson methods [29]. Yao *et al.* [30] employed a method based on a combination of fast marching level-set and Laplacian level set.

Approaches based on Fuzzy C-Means (FCM): Fallahi *et al.* [31] segmented the uterine fibroids by combining a fuzzy C-means method with some morphological operations. Later, on the basis of [31], a two-step method [32] was proposed by employing a Modified Possibilistic Fuzzy C-Means (MPFCM) [33] in a second step.

Approaches based on region-growing: Militello *et al.* [34] used a semi-automatic approach based on region-growing and reported a quantitative and qualitative evaluation of the HIFU treatment by providing the 3D model of the fibroid area. Rundo *et al.* [35] presented a two-phase method where the first phase is an automatic seed-region selection and region detection while the second one is aimed at uterine fibroid segmentation.

Other mixed methods: Antila *et al.* [36] designed an automatic segmentation pipeline without user input. They applied the active shape model (ASM) to get the deformed surface, and classified PV (perfused volume: the untreated tissue) and NPV (nonperfused volume: the treated tissue) by an expectation maximization (EM) algorithm. Militello *et al.* [37] proposed a novel fully automatic method based on the unsupervised Fuzzy C-Means clustering and iterative optimal threshold selection algorithms for uterus and fibroid segmentation.

Recently, Rundo *et al.* [38] evaluated the above mentioned two computer-assisted segmentation methods [37, 35] and provided a quantitative comparison on segmentation accuracy in terms of area-based and distance-based metrics. Their results show that both methods remarkably outperform the other ones.

However, there are still some limitations and drawbacks in the conventional methods and a fully-automatic and accurate method, able to reduce or even to remove pre-processing/ post-processing procedures as well as the interventions of the medical physicists, is still expected. For this purpose, a detailed comparison between the methods reported in [35] and [37] and our method will be shown in Section 3.4.4.

3.2.2 Deep Learning Methods of MR Image Segmentation

Only a few attempts have been reported for the uterus segmentation using CNN-based methods. Kurata *et al.* [39, 40] evaluated the clinical feasibility of fully automatic uterine segmentation on T2-weighted MR images based on an optimized U-Net. The segmentation of uterus in this research was focused on the staging of uterine endometrial cancer and on estimating the extent of tumor invasion to the uterine myometrium. To the best of our knowledge, there is no literature published on the uterine fibroid segmentation using CNN-based methods. Even so, it is important to highlight that many innovative deep learning methods have been proposed for MR image processing [41, 42]. The most common applications concern segmentation of organs, substructures, or lesions, often as a preprocessing step for feature extraction and classification. Deep learning methods for MR image segmentation can be divided into two different categories.

DL based on image patches: Features are extracted from a local patch for every voxel using convolutional layers. These features are then classified with a fully connected neural network to obtain a label for every voxel. This method is for instance widely used in brain tumor [43], white matter segmentation in multiple sclerosis patients [44], normal components of brain anatomy [45] and rectal cancer segmentation [46]. However, such methods have some disadvantages. The main problem is that their computational efficiency is very low because they have to process overlapping parts of the image. Another disadvantage is that each voxel is segmented based on a finite size context window, ignoring the broader context. In some cases, more global information may be needed to properly assign these labels to pixels or voxels.

Fully convolutional neural network (FCNN): In this case, the entire image or a large portion

is processed, the output being a segmentation result instead of a label of a single pixel or voxel. Such an approach solves the shortcomings of the former method and improves the efficiency of the algorithm. Many architectures can be considered for segmentation among which, as mentioned in Section 3.1, encoder-decoder ones such as U-Net and its modified versions [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18]. For MR images, we refer to [41] for a full survey. Zhang *et al.* [47] used CNN for segmenting the infant brain tissues by combining T1, T2, and FA images into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). Brain tumor segmentation was addressed in [48]. Avendi *et al.* [49] associated DL algorithms with deformable models for the left ventricle segmentation of the heart. Milletari *et al.* [7] proposed a 3D image segmentation based on a volumetric, fully convolutional, neural network. Their CNN was trained end-to-end on MR image volumes depicting the prostate and learned to predict segmentation for the whole volume at once. Some universal architectures were also proposed (for instance CE-Net by Gu *et al.* [50]) to address different clinical applications.

However, our target presents significant differences with these examples (*i.e.* brain, prostate, and heart). The deformation of the uterus shape is very large among the patients. The uterus position is also varying a lot. The high number of surrounding organs together with their similarity in tissue features makes more challenging the segmentation. In addition, different kinds of uterine fibroids (such as subseries fibroids, submucosal fibroids, intramural uterine fibroid tumors, pedunculated leiomyomas, and parasitic uterine fibroids) may be located in different regions of the uterus, and the gray level of these fibroids are affected by the signal intensity and other experimental factors. All these considerations have guided the design of our approach.

3.3 Method

To accurately segment the uterus, uterine fibroids and spine from the raw MR images, we propose an Encoder-Decoder global convolutional network scheme. The whole pipeline is illustrated in Figure 3.2. This network (called HIFUNet) consists of three major parts: the feature encoder module (based on a pre-trained ResNet101 backbone), the feature extractor part (with the global convolution network and deep multiple atrous convolutions) and the feature decoder module.

3.3.1 Encoder Module

The encoder part uses pre-trained ResNet-101 [25]. In [51], the authors demonstrated that the use of residual connections promotes information propagation both forward and backward, so it helps to improve significantly both the training speed and the performance. Because we have only one channel in our raw 2D input image (instead of RGB channels like in natural images), we change the original first portion which forms three input channels to one channel and we

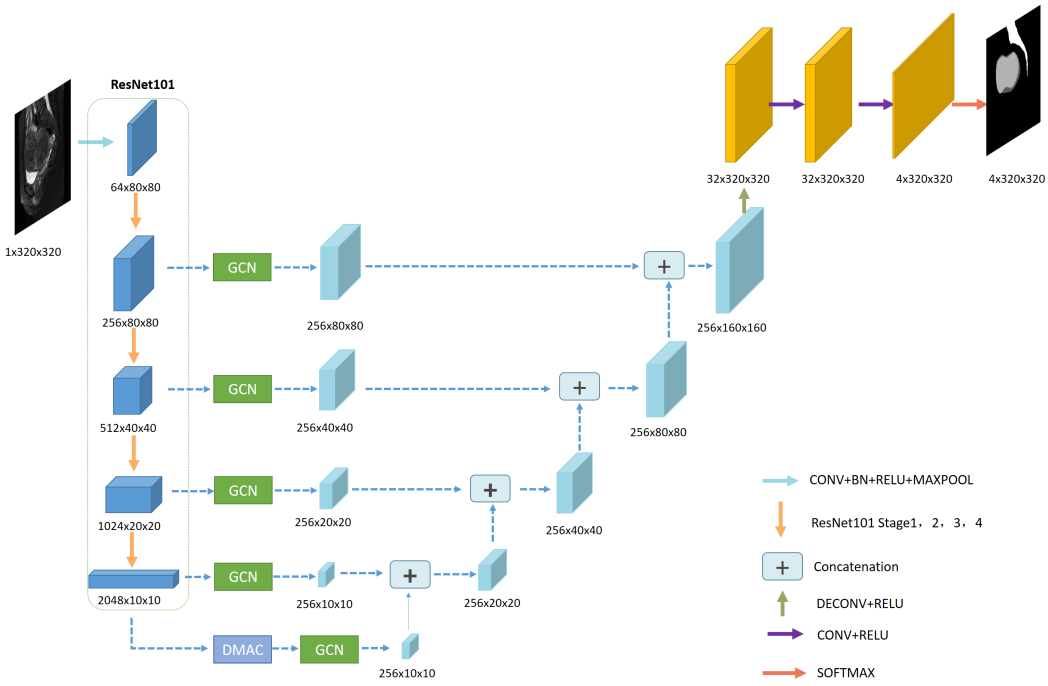

Figure 3.2 – The architecture of our proposal network (HIFUNet). The network consists ofa Resnet101 backbone as Encoder Module; GCN module and DMAC module as feature extractor par; and upsampling layers, concatenation layers, and an output layer as part of the feature decoder module. The parameters and sizes of output features in different layers are presented in different colors.

obtain 64 channels after the first 'Conv1'. Then, four feature extracting blocks are employed. The first, second, third, and fourth stages contain 3, 4, 23, and 3 bottlenecks respectively and each block has no average pooling layer or fully connected layers.

3.3.2 Global Convolution Network

The current trend in architecture design goes toward stacking small convolution kernels because this option is more efficient than using a large convolution kernel with the same amount of computation. However, considering that semantic segmentation tasks require pixel-by-pixel segmentation prediction, Peng *et al.* [19] proposed a global convolutional network to improve the accuracy of classification and localization simultaneously. In GCN, a fully-convolutional layer is adopted to replace the global pooling layer in order to keep the localization information. Besides, large kernels are introduced to increase the valid receptive field (VRF). However, using a large kernel or a global convolution directly is inefficient. To further improve the computational efficiency, GCN uses a combination of two large 1D convolutional kernels to replace a single 2D kernel for the skip-connector layer. The architecture of GCN is shown in Figure 3.3. The kernel size we use in our segmentation approach is 11×11.

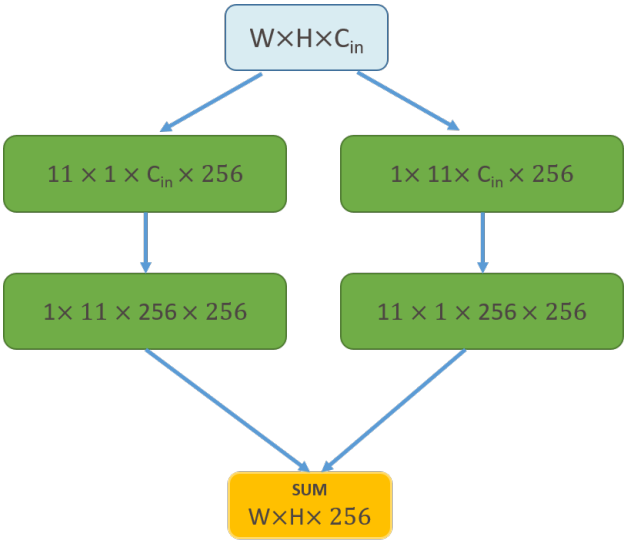

Figure 3.3 – Global Convolutional Network

3.3.3 Deep Multiple Atrous Convolutions

Atrous convolutions solve the problem of reduced resolution caused by the Deep Convolutional Neural Networks (DCNNs) while adjusting the receptive field of the filter. Figure 3.4 illustrates the atrous convolution. The main idea of atrous dilation rate convolution is to insert 'holes' (zeros) between pixels in convolutional kernels to increase the image resolution, enabling thus dense feature extraction in DCNNs. The atrous convolution was initially proposed to efficiently compute the undecimated wavelet transform [52] and the wavelet decomposition [53] in the atrous scheme. In recent years, atrous convolution has been widely used in tasks such as semantic segmentation and object detection. The Deeplab series [20, 21, 22, 24] and dense upsampling convolution (DUC) [54] made thorough studies of atrous convolution. Figure 3.5 shows our proposed deep multiple atrous convolution scheme to achieve multi-scale representations. We implement five convolutional layers with 3×3 kernels with different sampling rates to extract the different features. Finally, we fuse all features with the input image to generate the final result.

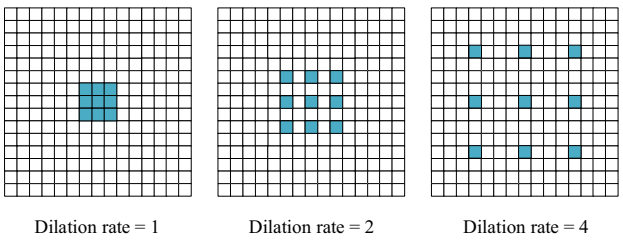

Figure 3.4 – Atrous convolutions with 3×3 kernel (blue blocks) and rates 1, 2 or 4.

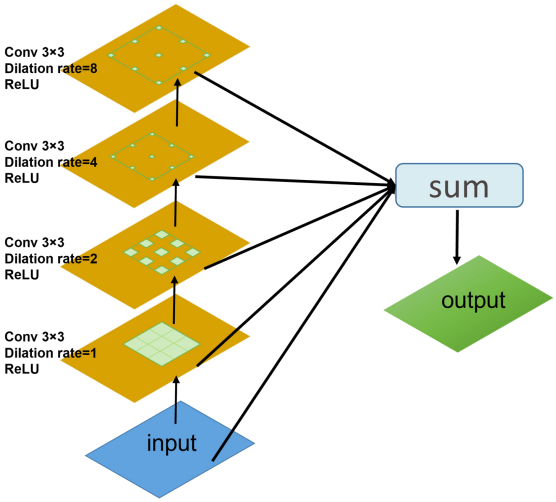

Figure 3.5 – Deep multiple atrous convolutions (DMAC) consist of five atrous convolutional layers.

When compared to the conventional network structure, our deep multiple atrous convolutions can extract multiple features and provide receptive fields of multiple sizes. It can be noticed that the architecture of our atrous convolution scheme adopts a serial frame instead of a parallel structure such as Inception and Atrous Spatial Pyramid Pooling (ASPP). We employ the DMAC block in the final layer of the encoder and this way more abstract information can be exploited. Within the DMAC block, as the layer is deeper, the dilation rate is getting larger. Because of the kernel discontinuity, not all pixels are used for calculation, so more atrous rate convolutions can compensate for the uncalculated information in the serial structure, which can increase the receptive field effectively. Besides, different sizes of atrous rates can help to extract different sized targets (from small fibroids to large organs like uterus or spine). The serial structure can get global distribution information from various scales of atrous convolution. The final step sums up as the output the abstract information extracted from the multiple layers. This output is then sent to the decoder phase in order to recover the object details and spatial dimensions. Therefore, in order to improve the performance of image segmentation, more low and high-level features are automatically captured in the encoder.

3.3.4 Decoder Module

The decoder module mainly uses the concatenation operation to fuse the multi-scale features. U-Net concatenates the downsampling feature maps with the corresponding upsampling feature maps. Here, this concatenation is performed between two neighboring feature maps after the GCN modules and this from the bottom to the top. After four concatenation operations, the image scale increases from 1/32 to 1/2 of the input image size. Then, we use a deconvolution operation to enlarge the image scale to the initial size and to restore features with more detailed

information. Finally, the output mask is obtained after applying two convolution operations and softmax. As illustrated in Figure 3.2, the decoder module mainly includes four concatenation operations (a 1×1 convolution, a 4×4 transposed convolution, and two 3×3 convolutions consecutively). Then, the feature decoder module outputs a mask with the same size as the original input.

3.3.5 Loss Function

The HIFUNet can be trained by minimizing the cross-entropy error between its prediction result and the ground-truth. The loss function is defined as:

$$L = \sum_{i \in \Omega} y_{c_i} \log(p_{c_i}) + (1 - y_{c_i}) \log(1 - (p_{c_i})) \tag{3.1}$$

where p_{c_i} denotes the predicted probability of c-th class for pixel i in the predicted result p, $y_{c_i} \in \{0,\ 1\}$ is the corresponding ground-truth value. If $y_{c_i} = 1$, it means that pixel i belongs to the c-th class. If $y_{c_i} = 0$, it means that pixel i does not belong to the c-th class. $c = 0$ denotes the background, $c = 1$ denotes the uterus, $c = 2$ denotes the uterine fibroids while $c = 3$ denotes the spine. Ω denotes the space of the predicted result of p and the ground-truth y. By minimizing the loss function on a training database, the parameters of HIFUNet can be optimized. Then the trained HIFUNet can be applied for automated uterus, uterine fibroids and spine segmentation on different datasets.

3.3.6 Discussion about the choice of our HIFUNET model

The main difference between our HIFUNet and other state-of-the-art deep learning networks including GCN[19], HRNet[55], U-Net[5], CE-Net [50], AttentionUNet [18], and LEDNet[56] is summarized as follows:

— GCN uses large kernels to enlarge the effective receptive field which can help classify different objects.
  Different from GCN, in order to exploit more abstract information, HIFUNet adds an original DMAC block which improves the accuracy of segmentation of key parts such as the cervix and minor fibroids.

— HRNet relies on a parallel structure enabling the model to connect multi-resolution subnetworks in a novel and effective way. It starts from a high-resolution subnetwork as the first stage and gradually adds high-to-low resolution subnetworks one by one to form more stages, the multiresolution subnetworks being connected in parallel.
  The main difference is that HIFUNet and HRNet use different ways for computing high-resolution representation. Our HIFUNet employs the way of recovering high-resolution representations from low-resolution representations outputted by a network (*e.g.*, ResNet).

While in HRNet, the authors propose another way that maintaining high-resolution representations through high-resolution convolutions and strengthening the representations with parallel low-resolution convolutions

— U-Net uses a simple downsampling way to extract features while HIFUNet uses ResNet101 as the backbone to extract more features. We add large kernels in the skip-connections to increase the valid receptive field (VRF).

— CE-Net uses the Dense Atrous Convolution (DAC) module with multi-scale convolution and the Residual Multi-kernel Pooling (RMP) with multi-scale pooling at the bottom to extract and decode multi-scale features in parallel, as well directly integrate them. It ignores the global scene content at each level which further enhance the localization effect of the skip connection, as well as the progressivity and the correlativity among the multi-scale structure.

Especially different from the CE-Net, the proposed HIFUNet adopts GCN in each skip connection between the encoder and the decoder. So that it is able to embed global scene information in the decoder, avoiding the global scene information loss in the dimension reduction during encoding. Besides, the HIFUNet also employs DMAC with the series structure and hierarchical fusion at the bottom of the encoder to progressively and correlatively extract multi-scale structure for the semantic.

— AttentionUNet proposes a novel Attention Gate (AG) model for medical imaging that automatically learns to focus on target structures of varying shapes and sizes, which brings a risk of transmitting multiplicative error along with the network.

CE-Net and AttentionUNet are both based on the U-Net and keep the way of extracting features in the encoder of U-Net. Differently, we choose to use a ResNet-101 pre-trained on Imagenet as our backbone because it can be easier to train Resnet than training simple deep convolutional neural networks and resolve the problem of accuracy degradation.

— LEDNet aims at real-time semantic image segmentation. It employs an asymmetric encoder-decoder architecture. The encoder adopts a ResNet as the backbone network, where two new operations, channel split and shuffle, are utilized in each residual block to greatly reduce the computational cost while maintaining a higher segmentation accuracy. On the other hand, an Attention Pyramid Network (APN) is employed in the decoder to further decrease the entire network complexity.

In our task, we pay more attention to the segmentation accuracy than to the efficiency of training. In the decoder part, LEDNet focuses on the last feature map from the encoder network, while some low-level features can be let out, which is not conducive to recovering detailed information. Therefore, we choose to recover the high-resolution information by concatenating low- and high-level features, which can help to identify the objects of all sizes and the details in complex medical images.

3.4 Experiment and Discussions

3.4.1 Datasets

To train and validate our work, we used preoperative T2-weighted MR images with fat suppression of 297 patients.These images were collected from the First Affiliated Hospital of Chongqing Medical University. Sagittal T2-weighted fast spin-echo images were acquired using a 3.0T MR unit (Signa HD Excite, GE Healthcare, Marlborough, MA) with an eight-channel phased-array coil. The scan parameters and characteristics of MR images are shown in Table 3.1.

Table 3.1 – The scan parameters and characteristics of the MR Dataset

Variable	**Value**
Repetition time (TR)	3040 ms
echo time (TE)	107.5 ms
field of view (FOV)	28×22.4 cm
slice thickness	6 mm
slice gap	1 mm
matrix	304×304
age (years)	$40.8 \pm 6.6^{*}$

* Age is Mean value ± S.D.

Each MR volume consists of 25 slices of 304×304 pixels. The ground truth has been generated through a proper annotation process. To ensure an objective and consistent clinical reference, two radiologists were solicited for consensus agreement. This procedure included three steps:

1. Annotations through discussions: the discussion between our two radiologists, A (7-year experience) and B (15-year experience), was held in a face-to-face mode to set the annotation rules and identify special and complicated cases. It appeared, in this application, that the variability of the annotations mainly existed on the contour of the cervix and some minor fibroids.
2. The radiologist A took 2 months in annotating (no more than 5 volumes per day). After annotating 10 volumes, a second face-to-face discussion was held to analyze the first-round annotation, and improve the annotation rule further.
3. Then the radiologist A processed all cases (297 patients). Radiologist B checked all results and marked the cases which have some divergent views. Then, they held a face-to-face discussion and solved these situations.

After the above three steps, a full agreement between the two radiologists was obtained.

The research associated with the treatment of uterine fibroids was approved by the ethics committee and has no implication on patient treatment.

3.4.2 Experimental Setup

Training and testing phase

MR images from 260 patients were used for training and images from the rest 37 patients were used for testing. The number of images in the testing set was 925. But as we know that the use of a small amount of training data can result in overfitting. To prevent this overfitting due to the limited number of images, the training data was augmented by image manipulation [57]. We applied the random shifting and scaling strategies (zoom range of 0.1, shift of 0.5mm).

Parameter settings and platform

For the optimization of our network, we use the Adam optimizer and set the initial learning rate to 2e-4. After each epoch, if we observe that the validation loss does not decrease for three consecutive times, the learning rate is reduced to 1/5 of its current value until it stops at 5e-7. Therefore, the number of training epochs is determined by the decreasing learning rate. The batch size is set to 8. All the comparative experiments adopt the same strategy for updating the hyperparameters. Besides, in the ablation study, the hyperparameters are fixed when removing parts of the network.

Our proposed network is based on the pretrained ResNet101 model on ImageNet. Notice that we adapt the first convolution operation because, as mentioned in section 3.3.1, we have a single channel input image instead of RGB channels like in natural images. The implementation is carried out on the PyTorch platform. The training and testing bed are ubuntu 16.04 system with NVIDIA Titan XP GPU (12 GB memory) and CUDA 9.0.

3.4.3 Evaluation Metrics

Different quantitative measures are used to comprehensively evaluate and compare the segmentation performance with the other methods.

Area-based indexes, which compare the predicted segmentation results (S_p) with the reference delineation (S_r) manually labeled by radiologists in terms of the mask. The following metrics are introduced : In general:

True positive (TP) = correctly identified

False positive (FP) = incorrectly identified

True negative (TN) = correctly rejected

False negative (FN) = incorrectly rejected

1. Dice coefficient (DSC) [58], also called the overlap index, is the most used metric for validating medical volume segmentation.

$$DSC = 2\frac{|s_r \cap s_p|}{|s_r| + |s_p|} = \frac{2TP}{2TP + FP + FN} \tag{3.2}$$

2. Precision (PR) [59] is able to describe the purity of our positive detections relative to the reference

$$PR = \frac{TP}{TP + FP} \tag{3.3}$$

3. Sensitivity (SE) [60]: also called true positive rate (TPR) or recall, it measures the extent to which actual positives are not overlooked.

$$SE = TPR = RR = \frac{TP}{TP + FN} \tag{3.4}$$

4. Specificity (SP) [60]: also called the true negative rate (TNR) is the extent to which actual negatives are classified.

$$SP = TNR = \frac{TN}{TN + FP} \tag{3.5}$$

5. Jaccard index (JI) [61]: also referred to as the Intersection over Union (IoU) metric, is essentially a method to quantify the percent overlap between the ground-truth and our prediction segmentation output.

$$JI = \frac{|S_r \cap S_p|}{|S_r \cup S_p|} = \frac{TP}{FP + FN + TP} \tag{3.6}$$

6. False Positive Ratio (FPR), False Negative Ratio (FNR) and False Region Ratio (FRR) [38]:

$$FPR = \frac{FP}{FP + TN} = 1 - TNR \tag{3.7}$$

$$FNR = \frac{FN}{FN + TP} = 1 - TPR \tag{3.8}$$

$$FRR = \frac{FP + FN}{FN + TP} \tag{3.9}$$

Distance-based indexes, which evaluate the segmentation in terms of both the location and shape accuracies of the extracted region boundaries. We defined two point sets A and B from S_p and S_g. N is the number of points in A.

1. Mean Absolute Distance (MAD) [38]: measures the average error of one boundary pixel the closest boundary pixels in the other segmentation.

$$MAD = \frac{1}{N} \sum_{a \in A} \min_{b \in B} \|a - b\| \tag{3.10}$$

2. Maximum Distance (MAXD) [38]: measures the maximum difference from two boundaries.

$$MAXD = \max_{a \in A} \min_{b \in B} \|a - b\| \tag{3.11}$$

3. Hausdorff Distance (HD) [62] : measures the similarity between two boundaries and can be expressed as:

$$HD = \max(h(A, B), h(B, A)) \tag{3.12}$$

where $h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|$.

Some literature report HD95, *i.e.*, the 95^{th} percentile of HD, to limit the influence of small outliers.

3.4.4 Comparison with Conventional Methods and Discussion

As mentioned in Section 3.2.1, Rundo *et al.* [35] and Militello *et al.* [37] proposed to segment uterine fibroids after treatment and evaluated them in [38]. We compare their methods with our method on the same dataset (fat-suppressed T2-weighted MR images composed of 375 slices issued from 15 patients).

It can be noticed that the above two methods are based on the fact that ablated fibroids appear as homogeneous hypo-intense regions with respect to the rest of the uterus (after contrast medium injection). Before the treatment, all kinds of fibroids appear as different states, which makes the segmentation task harder.

For all patients, area-based and distance-based indexes were computed based on a slice-by-slice comparison and were performed on each slice having a fibroid area. The results are displayed in Table 3.2. They show the superiority of the proposed method over the other two approaches and demonstrate its ability for uterine fibroid segmentation.

Table 3.2 – Values of Area-Based and Distance-Based for segmenting uterine fibroids using different methods on T2-weighted MR Images

Method	area-based								distance-based		
	DSC(%)	Precision(%)	SE(%)	SP(%)	JI(%)	FPR	FNR	FRR	MAD	MAXD	HD
IOTS [37]	80.50	76.83	89.03	98.22	69.34	0.018	0.110	0.540	2.432	7.893	8.893
SM&RG [35]	81.15	77.74	89.47	98.33	72.13	0.017	0.105	0.429	3.422	11.536	12.935
Proposed	86.58	88.17	88.45	99.53	78.45	0.005	0.116	0.709	2.955	9.365	16.372

Some visual results are depicted in Figure 3.6. It can be seen that, for the Patient 4, the gray level values around the area outlined by the circles have little difference from adjacent tissue. While in the post-treatment MR images the ablated tissue does not absorb the contrast medium and is hypo-intense with respect to the uterus, the use of simple adaptive global thresholding and region growing methods remains possible. However, the quality of the MR images is affected by noise which may lead to gray values in the regions of uterine fibroids similar to those of the surrounding tissues. As it is shown for Patient 7, there are two fibroids that appear with different signal strengths because of the different moisture contents: one is dark and the other one is bright. Thus, it is difficult for IOTS to distinguish the two different grayscale distributions of fibroids. SM&RG fails to identify the contour of fibroids and assimilates the uterus to fibroids. The segmentation provided by our DL method is close to the ground-truth segmented by the clinical experts.

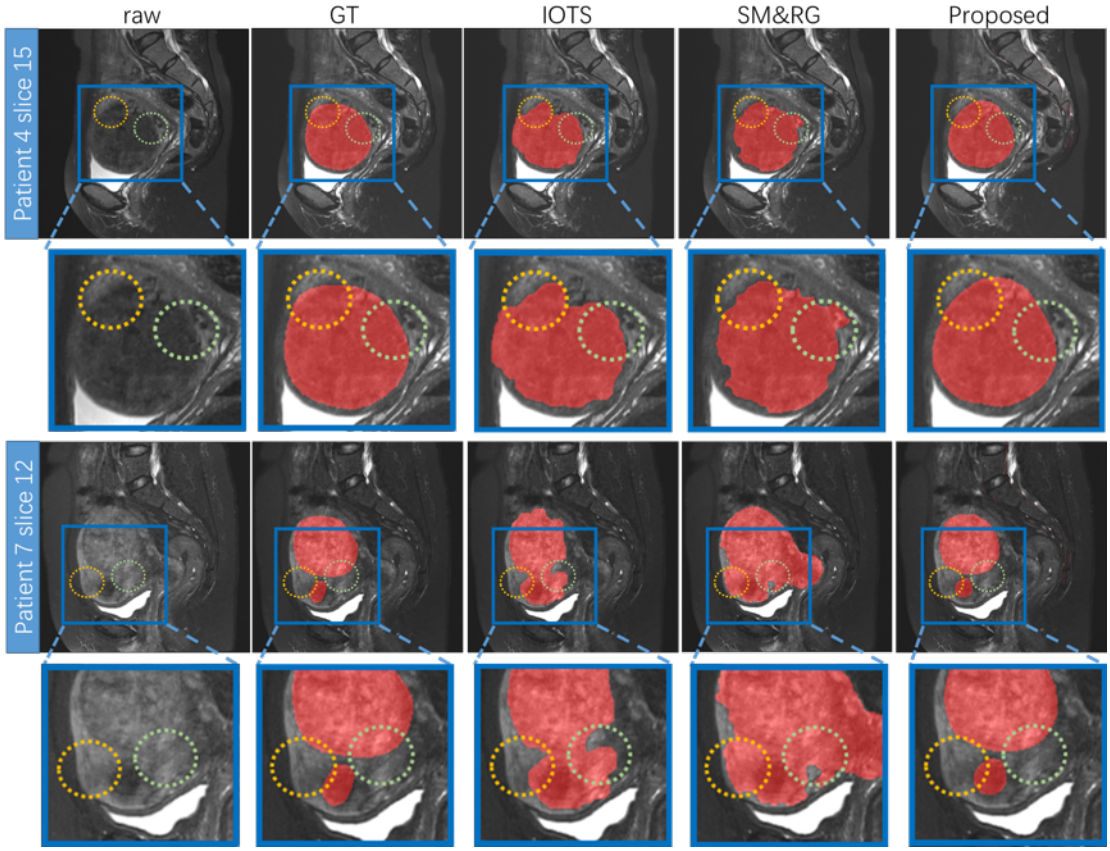

Figure 3.6 – Visualization of the uterine fibroids segmentation results on two patients using the proposed method and two conventional methods. Red denotes the fibroids, and the yellow and green circles point out incorrect segmentation of uterine fibroids due to the little gray value difference with the surrounding tissues.

Additional comments on the two methods used here for comparison are worth making. The

uterus ROI segmentation is a preliminary step for a robust fibroid detection in [38]. This task can be accomplished manually by the user to remove parts outside the uterus which are present in sagittal sections [35] or can rely on the Fuzzy C-Means (FCM) [37], which is an automatic method but where the number of clusters is set according to a visual inspection (*i.e.* anatomical properties of the analyzed pelvic images by considering image features) and experimental evidence (by means of segmentation trials). It means that the intervention of the experts is indispensable and that a complex and time-consuming preprocessing is needed before applying the intensity-based clustering technique. In conclusion, although these conventional methods have some merits in terms of performance, they show some practical limits in the clinical setting.

3.4.5 Comparison with Other Deep Learning Methods

We compare our method with six state-of-the-art (SOTA) algorithms, including U-Net [5], AttentionUNet [18], GCN [19], CE-Net [50], HRNet [55], LEDNet [56]. Their original implementations were kept and the same experimental conditions were used.

We select four of these SOTA methods (U-Net, GCN, HRNet and CE-Net) to visually compare our method in Figure 3.7 where the segmentation results are overlaid on the raw images. Different colors denote different classes (red denotes the fibroids, blue the uterus and green the spine). The images show that our method provides more accurate results. The performance of the six selected methods is presented in Table 3.3 for quantitative comparison. Among them, HRNet is the best method for segmenting uterus and fibroids. Besides, for the spine, which has a high contrast with adjacent tissues, the introduction of the attention mechanism (*i.e.* AttentionUNet) gives quite good results. However, overall, our method provides the best results.

Table 3.3 – Quantitative comparison of three evaluation indexes of different segmentation methods on the testing dataset. The best results are indicated in bold.)

Method	Uterus			Fibroids			Spine			Memory	Test time
	DSC	Precision	Recall	DSC	Precision	Recall	DSC	Precision	Recall		
GCN[19]	79.44%	79.27%	80.37%	80.43%	82.88%	80.04%	80.50%	85.14%	77.74%	464.96M	108.25ms
HRNet [55]	80.43%	78.29%	83.45%	80.88%	**85.39%**	80.76%	**85.45%**	83.77%	86.50%	561.88M	165.55ms
U-Net [5]	75.34%	76.97%	74.81%	77.58%	78.39%	79.23%	78.15%	**89.10%**	71.46%	317.97M	**14.56ms**
CE-Net [50]	74.69%	75.42%	74.99%	76.38%	75.05%	80.66%	82.48%	86.99%	79.15%	123.22M	105.77ms
AttentionUNet [18]	74.79%	76.08%	74.56%	76.24%	74.97%	81.18%	83.28%	88.54%	79.25%	927.34M	159.12ms
LEDNet [56]	77.87%	77.10%	79.46%	78.92%	83.71%	76.12%	79.02%	87.19%	74.19%	**121.37M**	73.84ms
Proposed	**82.37%**	**79.45%**	**86.00%**	**83.51%**	84.48%	**83.70%**	85.01%	82.51%	**88.69%**	503.71M	109.83ms

Regarding the computation cost, we estimated them by displaying the GPU memory requirements and the test time for segmenting each slice. Because of using ResNet as our backbone, our HIFUNet has a larger number of parameters. However, in clinical applications, the accuracy of the segmentation is much more important than the computation cost. From Table 3.3, we can see that the performance of HIFUNet is significantly better in comparison to the other methods. We found it acceptable that the increases in computational costs are negligible for the

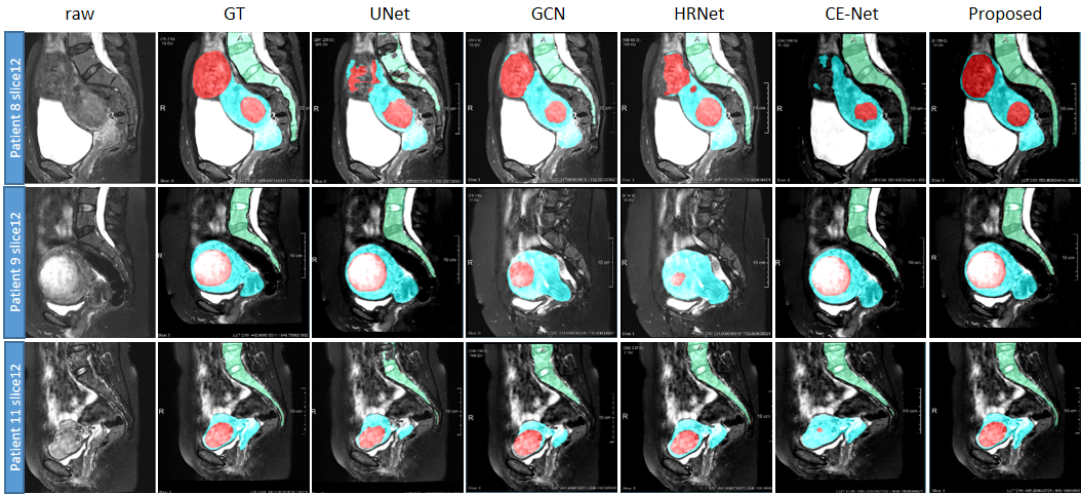

Figure 3.7 – Visualization of the segmentation results of uterus, fibroids and spine by using the proposed method and other four SOTA methods. From top to bottom are three different patients. Red denotes the fibroids, blue denotes uterus, and green denotes spine.

improvement in accuracy. The computational cost of our method at test time can be borne by a standard GPU.

Looking now organs by organs. As can be seen from Figure 3.7, the fibroids are more difficult to segment than the uterus, due to their unclear boundaries and undefined shapes. For patient 9, GCN and HRNet fail to segment the spine. For patient 8, U-Net, HRNet and CE-Net lead to incomplete segmentations. We can also observe the crucial role of the large receptive field used in our approach. Figure 3.8 show the DSC of uterus and fibroid segmentation results in the form of box plots. Our method provides the best and steadiest performance in segmenting both uterus and fibroids while the performance of HRNet is slightly weaker.

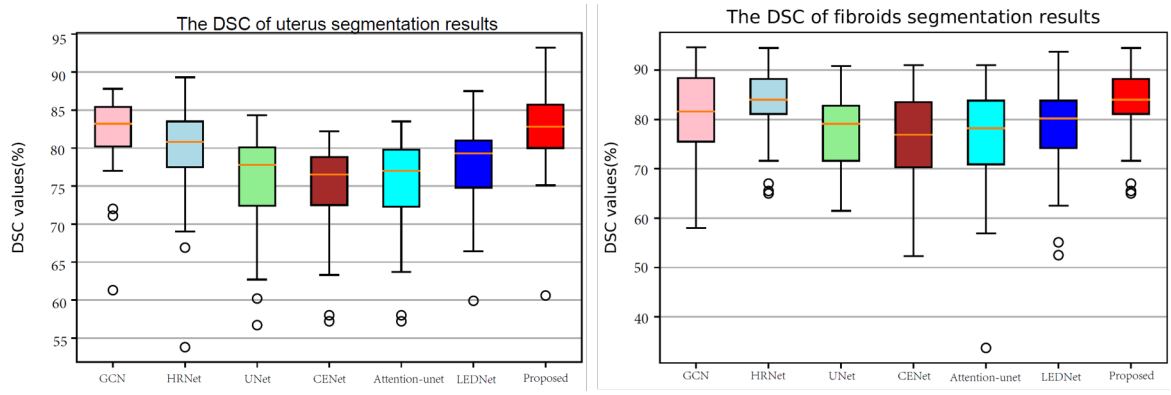

Figure 3.8 – Box plots of the qualitative performance to segment the uterus (left) and the fibroid (right). The y axis indicates the DSC values, while the x axis corresponds to the different methods (unfilled circles denote the suspected outliers).

3.4.6 Ablation Study

DMAC block We first conducted ablation studies and validated the effectiveness of our DMAC block using the same training strategy and datasets. The original GCN (GCN-no DMAC [19]) was compared with the modified GCN (GCN-DMAC) with a DMAC block added in the last layer. In the proposed HIFUNet (Proposed-DMAC), the DMAC block was put in the last layer and before the operation of global convolution. Comparisons were performed between the Proposed-DMAC, removal of DMAC block (Proposed-no DMAC) and insertion of the DMAC after the global convolutional operation (Proposed-DMAC behind). Table 3.4 shows the results of this study together with the time needed for each training epoch. They point out that the segmentation results are not significantly improved for GCN-DMAC. Concerning the DMAC position in our method, the computation time is strongly reduced when it is behind but the performance is worse than DMAC in-front (*i.e.* Proposed-DMAC). Our method is time intensive in training due to the large number of feature channels in the last layer (1024), but it also retains more features as a result. Also, HIFUNet outperforms CGN -DMAC with a p-value of $0.0031 > 0.001$.

Table 3.4 – The mean DSC and computation time of different segmentation methods using DMAC block. The best results are indicated in bold.)

Methods	DSC			Time(s)
	Uterus	Fibroids	Spine	
GCN-no DMAC	79.44%	80.43%	80.50%	164
GCN-DMAC	80.15%	81.08%	80.01%	**161**
Proposed-no DMAC	76.87%	78.84%	84.28%	479
Proposed-DMAC behind	77.72%	77.47%	80.89%	441
Proposed-DMAC	**82.37%**	**83.51%**	**85.01%**	1094

Some images are shown in Figure 3.9 for visual inspection. GCN leads to a relatively good segmentation of the uterus and the spine but the boundary of the fibroids is clearly inaccurate, and most parts of the fibroids fail to be labeled out. Adding the DMAC on GCN helps to refine the inaccurate boundary of the uterus and correct to some extent the wrong segmentation of fibroids. When replacing GCN by our proposed main structure, two fibroids are labeled out successfully with accurate boundaries (see Patient 20 slice 13) which shows the advantage of our main structure. In the same slice, by comparing GCN and Proposed-no DMAC, the boundary of the spine is corrected, which confirms the previous observation. A slightly better result can be achieved with DMAC. In all cases, our method labels both the uterus and the inside fibroids accurately which shows the effectiveness of the proposed DMAC. In particular, by comparing the last two columns, we can conclude that DMAC can extract the features of a large receptive field in a multi-scale context from multi-level feature maps.

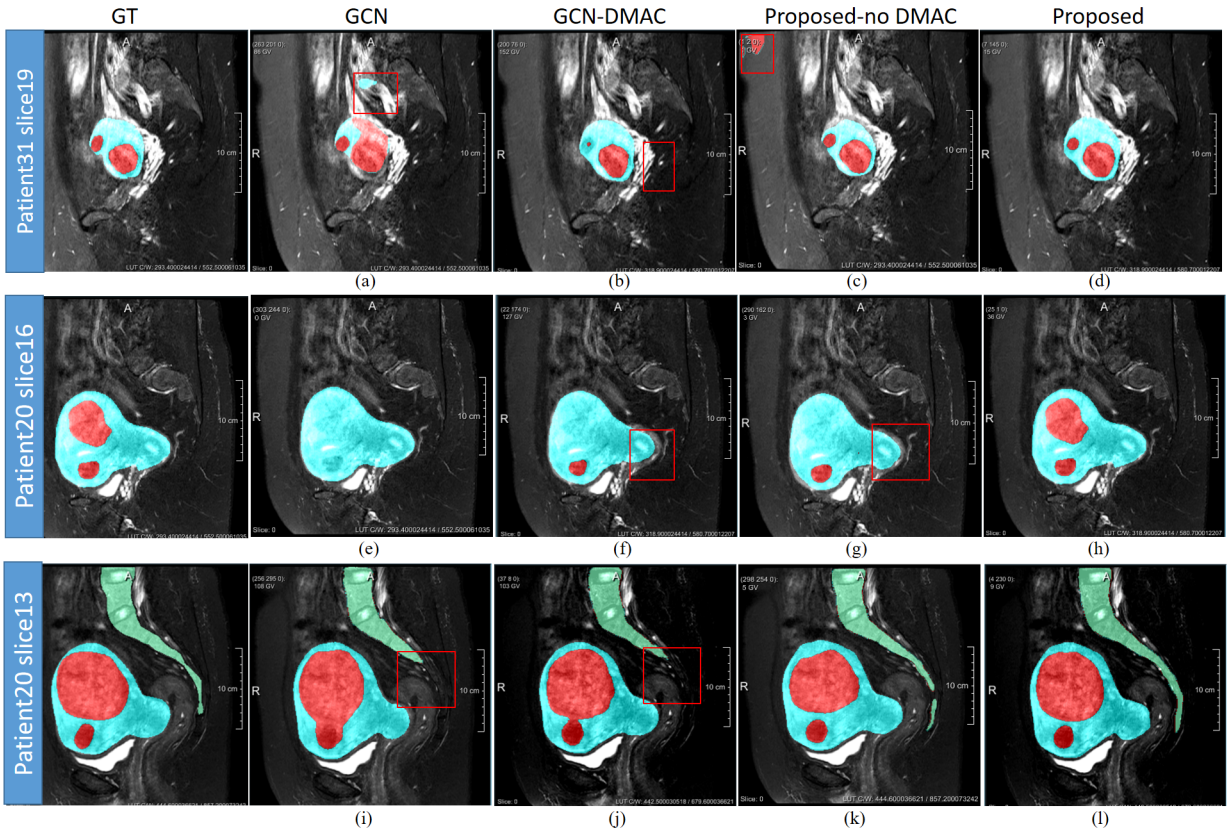

Figure 3.9 – Visualization of the segmentation results of uterus, fibroids and spine from two patients by using different methods which are mentioned in Table 3.3. From left to right: ground-truth, GCN [19], GCN with DMAC, our proposed method without/with DMAC. Red denotes the fibroids, blue denotes uterus, and green denotes spine. The places showing differences between the methods are surrounded by a red frame.

Decoder method In our approach, we replace the summation operation in GCN by a concatenation operation in U-Net. Besides, in the procedure of upsampling, the deconvolution operation is employed to recover the original image size and to get the output mask. Recent contributions focus on the use of an upsampling module to upsample a low-resolution feature map given high-resolution feature maps as guidance. For instance, Joint Pyramid Upsampling (JPU) [63] aims at generating a high-resolution target image by transferring details and structures from the guidance image. DUpsampling (DUP) [64] was also proposed to replace the standard bilinear upsampling to recover the final pixel-wise prediction. The DUP takes advantage of the redundancy in the label space of semantic segmentation and is able to recover the pixel-wise prediction from low-resolution outputs of CNNs.

We report here the experiments made in order to compare different ways of decoding. Inspired by Octave Convolution [65], in which Chen *et al.* proposed to store and process low-frequency

and high-frequency characteristics respectively, we plan to deal with low and high channels separately. Also motivated by the Inception module [66] which employed a split-transform-merge strategy, we design a Channel-Split (CS) module that splits channels of each feature map after the GCN module into high and low channels and then we use concatenation and summation operations to integrate features of different layers in a continuous way. Different from Octave Convolution in [65] which is an operation as a direct replacement of vanilla convolutions, CS is a decoder strategy to change the way of merging different channels from different layers. Another decoding method is shown in Figure 3.2. It removes the operations of summation in each layer and mainly uses deconvolution and concatenation. We name it Concatenation-Decoding (CD).

We train the three networks, *i.e.* with JPU or with CS or with CD as decoder respectively. The backbone here is the encoder of ResNet101 with GCN block and DMAC block. DUP is not trained because there is no formal code implementation of it. Experiments for method comparison were conducted on the same training parameter settings over the same training and validation dataset. The quantitative assessment was performed on the same testing dataset. The implementation of the JPU refers to the official PyTorch version on https://github.com/wuhuikai/FastFCN.

As shown in Table 3.5, the CD method is more accurate than the JPU and CS methods, with a benefit in DSC ranging from 6% and 16% for the uterus. It can be concluded that concatenation helps to recover the features especially in complex contexts and multiple targets. The summation is applied in the shortcuts (skip connections) in ResNet. It can help the network to speed up the training process and improve the gradient flow since the shortcuts are taken from previous convolution operations. Therefore, it is effective for the backpropagation to transfer error corrections to earlier layers, which can address the problem of vanishing gradient. However, due to the summation of the different channels or feature maps in CS, it may be difficult for the networks to distinguish different targets or recover the object details in the decoder. In contrast, the concatenation in CD operates on the feature maps generated by different filter sizes and keeps the information of different resolution feature maps since the information of features is not lost by summing up. JPU mainly uses the last three layers in the encoder. Therefore, the features of multiple objects in our complicated context may not be fully exploited by employing JPU.

Table 3.5 – The performance on the testing dataset by using different decoder methods: Joint Pyramid Upsampling (JPU), Channel-Split (CS) and Concatenation-Decoding (CD). The best results are indicated in bold.

Method	Uterus			Fibroids			Spine		
	DSC	Precision	Recall	DSC	Precision	Recall	DSC	Precision	Recall
Backbone+JPU	66.26%	70.44%	63.37%	67.36%	70.20%	66.74%	66.07%	76.90%	58.79%
Backbone+CS	76.37%	**80.77%**	73.27%	80.06%	79.55%	83.31%	83.88%	**87.45%**	81.37%
Backbone+CD (Proposed)	**82.37%**	79.45%	**86.00%**	**83.51%**	**84.48%**	**83.70%**	**85.01%**	82.51%	**88.69%**

3.5 Conclusions

In this study, we have proposed a global convolutional network with deep multiple atrous convolutions to segment uterus, uterine fibroids and spine automatically. The employment of the DMAC block allows capturing effectively more low and high-level features.

Experimental results on the same datasets and platform demonstrated (i) the accuracy and robustness of the proposed method, (ii) a significant improvement when compared to state-of-the-art segmentation methods and (iii) the performance could be close to radiologist level.

Although the proposed method shows promising results, some boundary inaccuracies may still be present in patients depicting multiple fibroids (see the left fibroid in the first row of Figure 3.9). We plan to improve our approach by working directly in 3D (*i.e.* 3D convolutional filters) instead of dealing with 2D slices. This will make the training issues (improving efficiency and reducing training time) more critical. Other ideas should also be explored such as the use of prior anatomical and pathological knowledge on the uterus and spine. Coupling our approach with other techniques (active contour models, for instance) to refine the boundaries of the uterus and spine may also offer a sound way to correct the remaining errors mentioned above.

BIBLIOGRAPHY

[1] E. A. Stewart, "Uterine fibroids," The Lancet, vol. 357, no. 9252, pp. 293–298, 2001.

[2] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," Medical image analysis, vol. 42, pp. 60–88, 2017.

[3] D. Shen, G. Wu, and H.-I. Suk, "Deep learning in medical image analysis," Annual review of biomedical engineering, vol. 19, p. 221, 2017.

[4] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in International Conference on Medical image computing and computer-assisted intervention. Springer, 2015, pp. 234–241.

[6] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in International conference on medical image computing and computer-assisted intervention. Springer, 2016, pp. 424–432.

[7] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in 2016 fourth international conference on 3D vision (3DV). IEEE, 2016, pp. 565–571.

[8] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes," IEEE transactions on medical imaging, vol. 37, no. 12, pp. 2663–2674, 2018.

[9] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer, 2018, pp. 3–11.

[10] J. Zhang, Y. Jin, J. Xu, X. Xu, and Y. Zhang, "Mdu-net: Multi-scale densely connected u-net for biomedical image segmentation," arXiv preprint arXiv:1812.00352, 2018.

[11] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A deformable network for retinal vessel segmentation," Knowledge-Based Systems, vol. 178, pp. 149–162, 2019.

[12] Q. Jin, Z. Meng, C. Sun, H. Cui, and R. Su, "RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans," Frontiers in Bioengineering and Biotechnology, p. 1471, 2020.

[13] J. Dolz, I. Ben Ayed, and C. Desrosiers, "Dense multi-path U-Net for ischemic stroke lesion segmentation in multiple image modalities," in International MICCAI Brainlesion Workshop. Springer, 2018, pp. 271–282.

[14] J. Guo, J. Deng, N. Xue, and S. Zafeiriou, "Stacked dense u-nets with dual transformers for robust face alignment," arXiv preprint arXiv:1812.01936, 2018.

[15] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert et al., "NNU-net: Self-adapting framework for u-net-based medical image segmentation," arXiv preprint arXiv:1809.10486, 2018.

[16] J. Dolz, C. Desrosiers, and I. Ben Ayed, "Ivd-net: Intervertebral disc localization and segmentation in MRI with a multi-modal UNet," in International workshop and challenge on computational methods and clinical applications for spine imaging. Springer, 2018, pp. 130–143.

[17] J. Zhuang, "LadderNet: Multi-path networks based on U-Net for medical image segmentation," arXiv preprint arXiv:1810.07810, 2018.

[18] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz et al., "Attention u-net: Learning where to look for the pancreas," arXiv preprint arXiv:1804.03999, 2018.

[19] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters improve semantic segmentation by global convolutional network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4353–4361.

[20] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," arXiv preprint arXiv:1412.7062, 2014.

[21] ——, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 4, pp. 834–848, 2017.

[22] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 801–818.

[23] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1251–1258.

[24] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," arXiv preprint arXiv:1706.05587, 2017.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[26] N. Ben-Zadok, T. Riklin-Raviv, and N. Kiryati, "Interactive level set segmentation for image-guided therapy," in 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE, 2009, pp. 1079–1082.

[27] H. Khotanlou, A. Fallahi, M. A. Oghabian, and M. Pooyan, "Segmentation of uterine fibroid on MR images based on Chan–Vese level set method and shape prior model," Biomedical Engineering: Applications, Basis and Communications, vol. 26, no. 02, p. 1450030, 2014.

[28] T. F. Chan and L. A. Vese, "Active contours without edges," IEEE Transactions on image processing, vol. 10, no. 2, pp. 266–277, 2001.

[29] X. Bresson, P. Vandergheynst, and J.-P. Thiran, "A variational model for object segmentation using boundary information and shape prior driven by the Mumford-Shah functional," International Journal of Computer Vision, vol. 68, no. 2, pp. 145–162, 2006.

[30] J. Yao, D. Chen, W. Lu, and A. Premkumar, "Uterine fibroid segmentation and volume measurement on MRI," in Medical Imaging 2006: Physiology, Function, and Structure from Medical Images, vol. 6143. SPIE, 2006, pp. 640–649.

[31] A. Fallahi, M. Pooyan, H. Ghanaati, M. A. Oghabian, H. Khotanlou, M. Shakiba, A. H. Jalali, and K. Firouznia, "Uterine segmentation and volume measurement in uterine fibroid patients' MRI using fuzzy C-mean algorithm and morphological operations," Iranian Journal of Radiology, vol. 8, no. 3, p. 150, 2011.

[32] A. Fallahi, M. Pooyan, H. Khotanlou, H. Hashemi, K. Firouznia, and M. A. Oghabian, "Uterine fibroid segmentation on multiplan MRI using FCM, MPFCM and morphological operations," in 2010 2nd International Conference on Computer Engineering and Technology, vol. 7. IEEE, 2010, pp. V7–1.

[33] L. Ma and R. C. Staunton, "A modified fuzzy C-means image segmentation algorithm for use with uneven illumination patterns," Pattern recognition, vol. 40, no. 11, pp. 3005–3011, 2007.

[34] C. Militello, S. Vitabile, G. Russo, G. Candiano, C. Gagliardo, M. Midiri, and M. C. Gilardi, "A semi-automatic multi-seed region-growing approach for uterine fibroids segmentation in MRgFUS treatment," in 2013 Seventh International Conference on Complex, Intelligent, and Software Intensive Systems. IEEE, 2013, pp. 176–182.

[35] L. Rundo, C. Militello, S. Vitabile, C. Casarino, G. Russo, M. Midiri, and M. C. Gilardi, "Combining split-and-merge and multi-seed region growing algorithms for uterine fibroid segmentation in MRgFUS treatments," Medical & biological engineering & computing, vol. 54, no. 7, pp. 1071–1084, 2016.

[36] K. Antila, H. J. Nieminen, R. B. Sequeiros, and G. Ehnholm, "Automatic segmentation for detecting uterine fibroid regions treated with MR-guided high intensity focused ultrasound (MR-HIFU)," Medical Physics, vol. 41, no. 7, p. 073502, 2014.

[37] C. Militello, S. Vitabile, L. Rundo, G. Russo, M. Midiri, and M. C. Gilardi, "A fully automatic 2d segmentation method for uterine fibroid in MRgFUS treatment evaluation," Computers in Biology and Medicine, vol. 62, pp. 277–292, 2015.

[38] L. Rundo, C. Militello, A. Tangherloni, G. Russo, R. Lagalla, G. Mauri, M. C. Gilardi, and S. Vitabile, "Computer-assisted approaches for uterine fibroid segmentation in MRgFUS treatments: quantitative evaluation and clinical feasibility analysis," in Italian Workshop on Neural Nets. Springer, 2017, pp. 229–241.

[39] Y. Kurata, M. Nishio, K. Fujimoto, M. Yakami, A. Kido, H. Isoda, and K. Togashi, "Automatic segmentation of uterus with malignant tumor on mri using U-net," in Proceedings of the Computer Assisted Radiology and Surgery (CARS) 2018 congress (accepted), 2018.

[40] Y. Kurata, M. Nishio, A. Kido, K. Fujimoto, M. Yakami, H. Isoda, and K. Togashi, "Automatic segmentation of the uterus on MRI using a convolutional neural network," Computers in biology and medicine, vol. 114, p. 103438, 2019.

[41] J. Liu, Y. Pan, M. Li, Z. Chen, L. Tang, C. Lu, and J. Wang, "Applications of deep learning to MRI images: A survey," Big Data Mining and Analytics, vol. 1, no. 1, pp. 1–18, 2018.

[42] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical image analysis using convolutional neural networks: a review," Journal of medical systems, vol. 42, no. 11, pp. 1–13, 2018.

[43] F. Milletari, S.-A. Ahmadi, C. Kroll, A. Plate, V. Rozanski, J. Maiostre, J. Levin, O. Dietrich, B. Ertl-Wagner, K. Bötzel et al., "Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound," Computer Vision and Image Understanding, vol. 164, pp. 92–102, 2017.

[44] S. Valverde, M. Cabezas, E. Roura, S. González-Villà, D. Pareto, J. C. Vilanova, L. Ramió-Torrentà, À. Rovira, A. Oliver, and X. Lladó, "Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach," NeuroImage, vol. 155, pp. 159–168, 2017.

[45] C. Wachinger, M. Reuter, and T. Klein, "Deepnat: Deep convolutional neural network for segmenting neuroanatomy," NeuroImage, vol. 170, pp. 434–445, 2018.

[46] S. Trebeschi, J. J. van Griethuysen, D. M. Lambregts, M. J. Lahaye, C. Parmar, F. C. Bakers, N. H. Peters, R. G. Beets-Tan, and H. J. Aerts, "Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR," Scientific reports, vol. 7, no. 1, pp. 1–9, 2017.

[47] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," NeuroImage, vol. 108, pp. 214–224, 2015.

[48] J. Bernal, K. Kushibar, D. S. Asfaw, S. Valverde, A. Oliver, R. Martí, and X. Lladó, "Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review," Artificial intelligence in medicine, vol. 95, pp. 64–81, 2019.

[49] M. R. Avendi, A. Kheradvar, and H. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI," Medical image analysis, vol. 30, pp. 108–119, 2016.

[50] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "Ce-net: Context encoder network for 2D medical image segmentation," IEEE transactions on medical imaging, vol. 38, no. 10, pp. 2281–2292, 2019.

[51] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in Thirty-first AAAI conference on artificial intelligence, 2017.

[52] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, "A real-time algorithm for signal analysis with the help of the wavelet transform," in Wavelets. Springer, 1990, pp. 286–297.

[53] J.-M. Combes, A. Grossmann, and P. Tchamitchian, Wavelets: Time-Frequency Methods and Phase Space Proceedings of the International Conference, Marseille, France, December 14–18, 1987. Springer Science & Business Media, 2012.

[54] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou, and G. Cottrell, "Understanding convolution for semantic segmentation," in 2018 IEEE winter conference on applications of computer vision (WACV). Ieee, 2018, pp. 1451–1460.

[55] K. Sun, Y. Zhao, B. Jiang, T. Cheng, B. Xiao, D. Liu, Y. Mu, X. Wang, W. Liu, and J. Wang, "High-resolution representations for labeling pixels and regions," arXiv preprint arXiv:1904.04514, 2019.

[56] Y. Wang, Q. Zhou, J. Liu, J. Xiong, G. Gao, X. Wu, and L. J. Latecki, "Lednet: A lightweight encoder-decoder network for real-time semantic segmentation," in 2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019, pp. 1860–1864.

[57] Z. Hussain, F. Gimenez, D. Yi, and D. Rubin, "Differential data augmentation techniques for medical imaging classification tasks," in AMIA annual symposium proceedings, vol. 2017. American Medical Informatics Association, 2017, p. 979.

[58] L. R. Dice, "Measures of the amount of ecologic association between species," Ecology, vol. 26, no. 3, pp. 297–302, 1945.

[59] F. C. Monteiro and A. C. Campilho, "Performance evaluation of image segmentation," in International Conference Image Analysis and Recognition. Springer, 2006, pp. 248–259.

[60] R. Trevethan, "Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice," Frontiers in public health, vol. 5, p. 307, 2017.

[61] P. Jaccard, "The distribution of the flora in the alpine zone. 1," New phytologist, vol. 11, no. 2, pp. 37–50, 1912.

[62] J. Henrikson, "Completeness and total boundedness of the Hausdorff metric," MIT Undergraduate Journal of Mathematics, vol. 1, no. 69-80, p. 10, 1999.

[63] H. Wu, J. Zhang, K. Huang, K. Liang, and Y. Yu, "Fastfcn: Rethinking dilated convolution in the backbone for semantic segmentation," arXiv preprint arXiv:1903.11816, 2019.

[64] Z. Tian, T. He, C. Shen, and Y. Yan, "Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3126–3135.

[65] Y. Chen, H. Fan, B. Xu, Z. Yan, Y. Kalantidis, M. Rohrbach, S. Yan, and J. Feng, "Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 3435–3444.

[66] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2818–2826.

Chapter 4

Semi-supervised segmentation of uterine regions from MR images for HIFU treatment

4.1 Introduction

Chapter 3 addressed the problem of segmentation of the uterine region, which is the prerequisite for defining the HIFU treatment planning. However, segmentation methods based on fully supervised learning (FSL) require a large amount of accurate annotated data to support the network during learning. But, clinically accurate annotated data is often difficult to obtain because it is a time-consuming process for physicians, often repetitive and not very rewarding for them. This results in imprecise annotation with high intra- or inter-individual variability between the experts and often a huge difficulty to access these data. One of the consequences is the lack of generalization of these networks based on fully supervised learning which requires a new round of learning in case of change or improvement of the image acquisition devices or even more simply of changes of the acquisition parameters.

One solution would be to use more limited sets of labeled data and develop models that could obtain segmentation performances close to those of fully supervised learning models. In our opinion, the development of such models is a critical problem to solve in the field of medical image segmentation.

Semi-supervised learning (SSL) could be one of the answers to this problem. Unlike the FSL methods, SSL methods can take advantage of large numbers of unlabeled data to improve network performance when labeled data is insufficient. One popular SSL approach is to adopt consistency learning, which regularizes the network to be consistent with the predictions of perturbation [1]. Another common way for SSL is pseudo-labeling by producing an artificial label for unlabeled images. Specifically, a pre-trained model is first used on a small number of labeled data. Then, the unlabeled data is fed to the model, and the class with the maximum predicted probability is selected and called pseudo-labels. After that, the labeled data is co-trained with the pseudo-labeled data. The above procedures are repeated to make the model more efficient. The

effectiveness of this mechanism is similar to that of entropy regularization. The semi-supervised model can achieve state-of-the-art performance when combined with Denoising Auto-Encoder and Dropout [2].

Because of the lack of annotated data, some data augmentation methods have to be considered in SSL. One simple and efficient method is Mixup [3] which can improve the generalization and robustness of the model by mixing the data pairs. This strategy can also be interpreted as an empirical risk minimization on modified data with random perturbations [4]. Based on the Mixup, CutMix [5] and Mixmatch [6] were developed to further improve the performance of the SSL.

However, the quality of the pseudo-labels will affect the optimization of the model when it is updated. In existing approaches, the generation of pseudo-labels relies on time-consuming manual offline selection, usually based on experience or after experimentation on a small validation set, followed by setting a threshold to generate a confidence map. This threshold is usually task-specific and is not universal. We believe that an adaptive thresholding strategy needs to be developed to adapt automatically to different semi-supervised data distributions. This should improve the generality and robustness of semi-supervised methods.

Besides, the utilization of limited annotated data also affects the quality of the pseudo-label generation. We plan to improve the quality of pseudo-label generation by using a powerful feature extraction network to extract features from segmented targets in limited data and noise reduction of pseudo-labels. In addition, we plan to extend the regularisation in Mixup to adapt the framework to more complex semi-supervised medical image segmentation tasks.

In order to combine all of these proposals, we have developed a new method called the Pseudo-label Refinement Network (PLRNet). The contributions of our approach are as follow:

— PLRNet utilizes an efficient segmentation noise reduction network to enhance the quality of pseudo-label generation while performing efficient feature extraction on the labeled data.
— PLRNet introduces a new online threshold adaptation strategy to generate high-confidence graphs for pseudo-labels, which improves the performance of the model even for different ratios of the amount of annotated and unannotated data.
— PLRNet uses a data augumentation method inspired by Mixup. This method called ”Feature-aligned Mixup”, will improve generalization across different patient data distributions.
— Experiments on the uterine dataset show that PLRNet outperforms other state-of-the-art semi-supervised methods and has the potential to apply to other segmentation tasks.

4.2 Methods

In this study, we aim to exploit unlabelled data by: improving feature extraction from labelled data, optimizing the generation of pseudo-labels and improving the generalisation capability of the model. In this section, we will first describe the main structure of PLRNet, then the employed pseudo-label optimization strategy, and finally the four basic components of the framework: the segmentation module, the confidence-based threshold adaptation module, the feature-aligned mixup module, and the consistency regularization operation.

4.2.1 Overview of PLRNet

PLRNet has two quite different parts as shown in Figure 4.1. On the one hand, a training pipeline that integrates annotated data and unannotated data in several phases while proposing a data augmentation. On the other hand, an inference pipeline that takes one of the networks trained in the learning pipeline to segment new MR images.

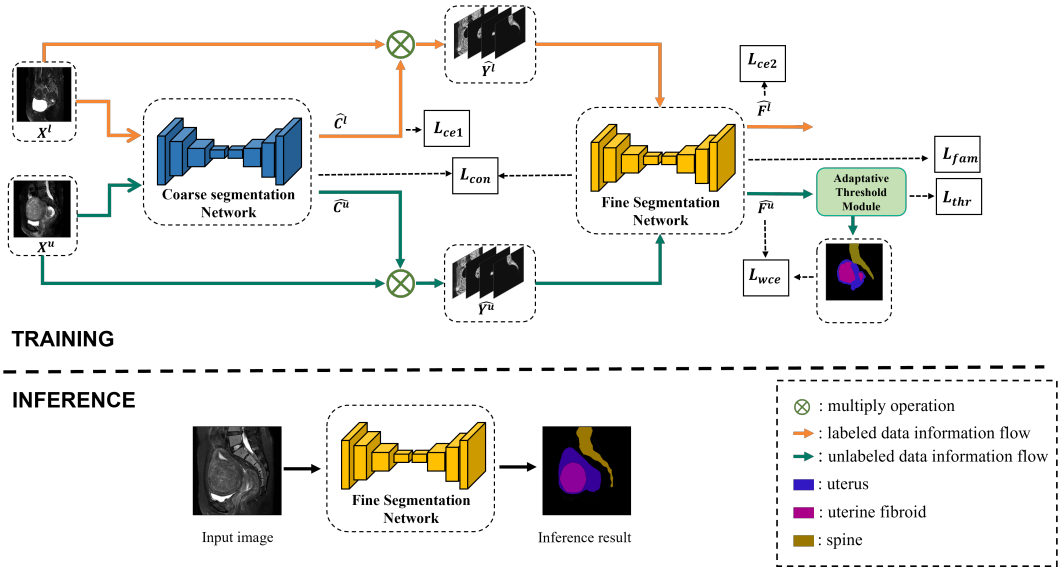

Figure 4.1 – The framework of PLRNet.

The training pipeline. The structure of the training pipeline is shown on the top of Figure 4.1). In order to gain in accuracy, we have chosen a coarse to fine approach. So our training pipeline is backed by two networks: a Coarse Segmentation Network (CSNet) and a Fine Segmentation Network (FSNet). For both, we used a CNN with large convolutional kernels as backbone network. Both underlying networks are trained from scratch. The output of both these 2 networks is a 4 channels probability map (one channel for each class: background, uterus, fibroids and spine). Note that for the cascade between these two networks, each of the probability

maps given by CSNet is pixel-multiplied with the original image of the input so that each channel has only the region of interest for the current category. We now delve into the details.

1. We first train the CSNet on labeled data X^l supervised by the cross-entropy loss L_{ce1} between $\widehat{C}^l$, the predicted outputs of CSNet, and GT^l, the ground-truth annotated by experts. In the second stage, FSNet is also trained on the labeled data X^l supervised by the cross-entropy loss L_{ce2} between $\widehat{F}^l$, the predicted outputs of FSNet, and GT^l.
2. Similarly, for the unlabeled image X^u, we use $\widehat{C}^u$ and $\widehat{F}^u$ to define the output which are both the pseudo-labels of unlabeled data. $\widehat{C}^u$ is refined to obtain $\widehat{F}^u$. Specifically, $\widehat{C}^u$ and $\widehat{C}^l$, which are four-channel probability maps, are respectively associated by a dot product with X^l, X^u to generate $\widehat{Y}^l$ and $\widehat{Y}^u$. In the second stage, $\widehat{F}^u$ is produced by feeding $\widehat{Y}^u$ into FSNet.
3. However, pseudo-labels with low confidence must be discarded from the optimization process. For this, usually a confidence threshold T is set. The network is trained only if prediction confidence is over T. In most of the paper, this threshold is fixed at the start. In our case, the threshold T is designed on the loss function in L_{thr}, which continuously optimizes the estimated global threshold as training proceeds, thus continually improving the quality of the pseudo-label.

The two networks carry out cooperative training and constantly update under four losses: the feature-aligned mixup loss L_{fam}, the weighted cross-entropy loss L_{wce}, the threshold loss L_{thr}, and consistency regularization loss L_{con}. Specifically, L_{fam} is used to improve the generalization performance of the model, L_{wce} and L_{thr} are related to the loss of unlabeled data to achieve the automatic online generation of high-confidence pseudo-labels and L_{con} helps the two networks to produce similar reference results. These losses will be detailed later on.

The inference pipeline. The bottom part of Fig.(4.1) shows the inference process. Simply, a new input image to be segmented is fed into the trained FSNet to obtain the inference result.

We will now detail some of the key points of PLRNet: the segmentation network, the confidence-based threshold adaptation and the feature-aligned Mixup.

4.2.2 Segmentation network

In semi-supervised learning, it is important to extract as many abstract feature representations as possible for annotated data. In current semi-supervised learning, U-Net has been widely used as a feature extraction network. However, in some multi-class medical image segmentation with complex contexts, using U-Net networks to extract features tends to introduce noise and neglect feature extraction for some fine targets (*e.g.* small-sized fibroids). In addition, in the cascaded CSNet-FSNet network used in this paper, CSNet acts as a pre-trained network to

generate coarse pseudo-labels for unlabelled data, and the noise in the pseudo-labels affects the optimisation of the pseudo-labels in FSNet.

To solve these challenges, we consider the use of Global Convolutional Network (GCN), an important component in HIFUNet, which can effectively extract the complex data features of a scene by increasing the valid reception field. In addition, wavelet sampling, consisting of wavelet transform and inverse wavelet transform, is used instead of traditional upsampling to suppress noise in the image to extract representative and effective features.

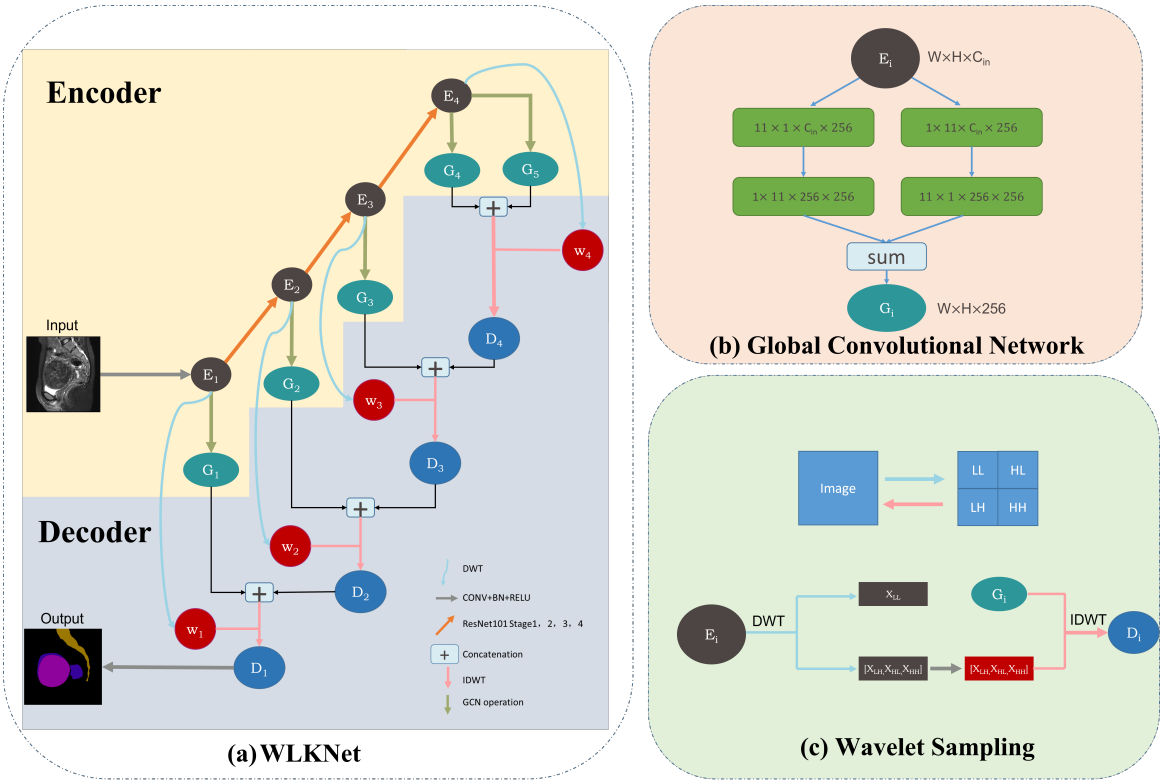

Figure 4.2 – (a) The architecture of WLKNet, (b) Global convolutional network using large convolutional kernels, (c) Wavelet Sampling

The Wavelet-Based Large Kernel Network (WLKNet) can be seen in Figure 4.2 (a). It shows the main structure of WLKNet, which is an end-to-end network with an encoder-decoder structure. E_1 - E_4 in the figure represent features encoded using Resnet101. G_1 - G_5 are features extracted using GCN and W_1 - W_4 represent wavelet coefficients extracted using the Harr Discrete Wavelet Transform (DWT). Fig. 4.2 c) is a demonstration of the process of wavelet sampling in WLKNet. Specifically, Resnet101 was performed for an input image to obtain four stages of encoding features, after which the encoding features were processed with large convolution kernels using GCN to obtain G_1 - G_5 respectively. Meanwhile, the wavelet decomposition was performed for E_1 - E_4 to obtain wavelet coefficients for W_1 - W_4 respectively. In the decoding

stage, G_4 and G_5 are concatenated, and the features obtained are then subjected to the Inverse Discrete Wavelet Transform (IDWT) together with W_4 to obtain the decoding feature D_4, and the decoded feature D_i is then merged with the previous layer's large convolution kernel feature G_{i-1} at each step. Finally, convolution and other operations are performed to recover the size of the original input image. The application of GCN and wavelet sampling in WLKNet is described in details below.

Large kernel network

The successful application of networks with large convolutional kernels in HIFUNet demonstrated that enhancing the valid receptive field size can accurately segment multi-class targets in the uterine region. Therefore, we use an 11×11 convolutional kernel as in Figure 4.2 (b) for feature extraction of the image. The input encoded features E_i are summed pixel by pixel after two separate large convolutional kernel operations. The result is a feature G_i with the same size as the input E_i and 256 channels.

Wavelet sampling

In existing deep learning networks, downsampling operations usually use either max pooling or average pooling, which have some limitations. For example, max pooling can cause the loss of primary features when their magnitude values are lower than the values of unimportant features. The use of average pooling allows for a balance between important and unimportant features, thus diluting the crucial features. In conclusion,both of these traditional pooling operations result in lower feature extraction, making segmentation less efficient. In addition, operations such as max pooling, average pooling lead to aliasing between data components in different frequency intervals. The noise in the data is mainly in the high-frequency components, while the low-frequency components contain the main information, such as the underlying object structure. As a result, aliasing introduces residual noise in the downsampled data and corrupts the underlying structure, thus reducing the accuracy and noise immunity of the CNN [7].

On the other hand, in computer vision tasks requiring high-resolution image recovery, including semantic segmentation and super-resolution recovery in encoder-decoder network structures, upsampling operations such as inverse pooling, linear interpolation, and deconvolution have traditionally been used. Unpooling fills the gaps in low-resolution feature maps containing semantic information with "0", linear interpolation fills the low-resolution feature maps with adjacent approximations, and deconvolution convolves the low-resolution features maps also with "0". These methods recover only a limited amount of detail, making it challenging to recover edge texture information from an image.

To solve the above problem, Williams *et al.* [8] recently proposed a sampling operation from the wavelet domain. Specifically, in the encoder part, wavelet pooling would replace traditional

max pooling. Wavelet pooling decomposes features into a low-frequency component and a high-frequency component after a two-dimensional wavelet decomposition. The low-frequency component stores the primary information such as the underlying feature structure, including the image contour, which is transferred to the subsequent layers to extract robust high-dimensional features. On its side, the high-frequency component stores the detailed part of the image for rounding during downsampling. Wavelet pooling discards only the detailed part, which effectively avoids removing important features as max pooling does, and solves the problem of dilution of all features as in average pooling. Wavelet pooling operation allows to maintain spatial information as much as possible while suppressing the correlation among frequency bands, and it can also effectively suppress high-frequency noise. The combination of wavelet transform and pooling operation can achieve a more powerful feature extraction capability than spatial domain downsampling. In the decoding module, wavelet upsampling will replace the traditional max unpooling. Specifically, the low-resolution high semantic features in the decoder are treated as wavelet low-frequency components, and the high-resolution texture and edge detail information in the encoder module at the lower layers are treated as wavelet high-frequency components. The inverse wavelet transform recovers the high-resolution features containing more detailed information. In short, the high-frequency components in the encoder network are stored and transmitted to the decoder for resolution recovery, which can achieve more efficient detail recovery than traditional interpolation and deconvolution.

Based on this idea, we will carry out in this project a small wavelet transform (upper diagram in Figure 4.2 (c)). A first order wavelet decomposes an image in two dimensions to obtain four sub-bands, LL, LH, HL, HH. The 3 sub-bands LH, HL, HH represent the image details including most of the noise, while LL is the low-resolution version of the image in which the primary object structure is represented.

In WLKNet, the four sub-bands are obtained by first performing DWT on E_i. These four sub-bands are half the size of the input E_i. This operation is therefore similar to a downsampling operation using max pooling. After separating the high and low-frequency sub-bands, the image details including noise such as LH, HL, and HH. are retained and then convolved to make the number of channels to 256. This result is called W_i. Afterward, W_i and G_i, which have been abstracted by the GCN features, are subjected to IDWT to recover the image features D_i.

4.2.3 Confidence-based Threshold Adaptation

As told previously, in the pseudo-label generation these with low confidence must be discarded from the optimization process. For this, usually a confidence threshold T is set. The determination of this threshold is one of the key points of the method. Usually, in previous publications, this threshold is fixed and is determined by experimentation or by a manual grid search which are time-consuming and have limitations in the multi-class medical image segmentation

tasks.

Inspired by [9], which was the first attempt of an online weighted pseudo-label in unsupervised domain adaptation (UDA), we introduce an adaptive method as an alternative to the manual grid search method to obtain the threshold T. Specifically, for each FSNet output pixel, if it is lower than T, we set the weight of this pixel as 0. On the contrary, if the pixel output of FSNet is higher than T, the pseudo-label weight ω of this pixel will be calculated by:

$$\omega = \begin{cases} \frac{\max\left(\widehat{p^u}\right) - T}{1 - T} & \text{if } \max\left(\widehat{p^u}\right) \geq T \\ 0 & \text{otherwise} \end{cases} \tag{4.1}$$

where $\max(\widehat{p^u})$ refers to the maximum confidence value for each pixel $\widehat{p^u}$ in $\widehat{F^u}$. Notice that $\widehat{F^u}$ is actually a feature map containing multiple channels, where each channel represents a segmentation category.

In this way, the pixels with higher confidence can be used to calculate the loss, while a lower ones will be discarded. By using pixel-by-pixel weighting, the network can pay more attention to pixels with correct predictions in pseudo-labels, and reduce the negative impact of pixels with inaccurate predictions. The loss function of unlabeled data, namely L_u, is defined by:

$$L_u = L_{wce} + L_{thr} \tag{4.2}$$

where L_{wce} is the weighted cross-entropy loss function, and L_{thr} is the adaptive threshold loss function. They are given by:

$$L_{wce} = -\frac{1}{N} \sum_{i \in I} \omega p_i^u \log\left(\widehat{p_i^u}\right) \tag{4.3}$$

$$L_{thr} = \log^2(1 - T) \tag{4.4}$$

Here N is the number of pixels in one image and p_i^u represents each pixel of the channel i in one pseudo-label P^u generated by FSNet. I is the number of channels. For the initialization of T, in the early training process, we first chose a threshold value that is high enough to quickly get a good result for easy-to-segment targets. Considering the training efficiency, we set the initial value of T to 0.8, which provides acceptable results. Then, as the training process proceeds, the threshold value is gradually reduced so that high weights are learned for the hard-to-segment

pixels.

4.2.4 Feature-aligned Mixup

Mixup [3] aims to improve the network generalization by a linear combination of paired input data and their labels. Recently, [10] extended this regularization strategy to both the input space and the latent space to regularize different parts of the network. Considering the semi-supervised multi-category image segmentation task, we should use limited labels to generate more data to avoid over-fitting and achieve generalization over different patients.

Feature-aligned Mixup in FSNet achieves this goal by regularizing the output of each layer in the decoder part. The Feature-aligned Mixup loss (Fig.4.3) is computed in the hidden layers of FSNet. For the encoder input in FSNet, we first generate a multiple-channel attention map $\widehat{Y^l}$ by multiplying the labeled image X^l with its predicted result of CSNet, then the unlabeled image gets its corresponding attention map $\widehat{Y^u}$ in the same way . After that, we linearly mix the two attention maps as follows:

$$\lambda \sim \text{Beta}(\alpha, \alpha) \tag{4.5}$$

$$\lambda' = \max(\lambda, 1-\lambda) \tag{4.6}$$

$$\widehat{Y^{mix}} = \lambda' \widehat{Y^l} + \left(1-\lambda'\right) \widehat{Y^u} \tag{4.7}$$

where Beta is a Beta distribution with α its positive shape parameter. α is considered as a hyperparameter in this work.

In order to make feature-aligned Mixup, we also realized the mixup with attention maps from the ground truth and pseudo-labels :

$$\widehat{R^{mix}} = \lambda' R^l + \left(1-\lambda'\right) R^u \tag{4.8}$$

where R_l is X^l multiplied by GT^l and R^u is X^u multiplied by P^u.

$\widehat{Y^{mix}}$ and $\widehat{R^{mix}}$ are respectively sent into the FSNet, and their outputs at each layer k of the decoder are marked as two sets $\hat{F} = \left\{\hat{f}_k\right\}$ and $F = \{f_k\}$ where $1 \leq k \leq K$. K is the depth of FSNet. Here we set K to 3. We can the use a Cross entropy loss function for computing Feature-aligned Mixup

$$L_{fam} = -\sum_{k=1}^{K} f_k \log\left(\widehat{f_k}\right) \tag{4.9}$$

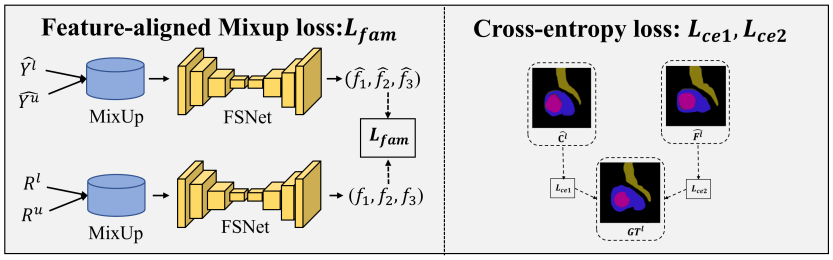

Figure 4.3 – The illustration of the Feature-aligned Mixup loss (L_{fam}) and cross-entropy loss in the PLRNet.

4.2.5 Consistency Regularization and Dropout

Due to the existence of pseudo-labels, some noise will inevitably be introduced. Therefore, regularization plays an important role in our task. We use Kullback-Leibler Divergence (D_{KL}) as the consistency regularization here. The purpose of consistency regularization is to ensure that the sample and the extended version of its network prediction have the same conceptual meaning as possible in the method.

The dropout layer is another technique to prevent our model from over-fitting. It randomly drops neurons from the network during training. The consistency regularization loss is defined as:

$$L_{\text{con}} = 0.5^* D_{KL}\left(\widehat{C} \| \widehat{F}\right) + 0.5^* D_{KL}\left(\widehat{F} \| \widehat{C}\right) \tag{4.10}$$

where $\widehat{C}$ and $\widehat{F}$ represent the prediction outputs of CSNet and FSNet for both labeled and unlabeled data, respectively.

4.2.6 Loss function

The training of PLRNet is divided into three steps: 1) the CSNet is trained with a limited proportion of labeled data; 2) then the parameters are shared to the FSNet; 3) the whole network is trained with all training data, including labeled and unlabeled data. The loss function is as follows

$$L = L_l + L_u + L_{fam} + L_{con} \tag{4.11}$$

where the L_u, L_{fam}, L_{con} were introduced in sections 4.2.3, 4.2.4 and 4.2.5 respectively. L_l is the loss function of the labeled data and is composed of two standard cross-entropy loss functions (Fig.(4.2)):

$$L_l = 0.5 * (L_{ce1} + L_{ce2}) \tag{4.12}$$

Here L_{ce1} and L_{ce2} are the cross-entropy of the output for labeled data of respectively CSNet and FSNet with the Ground-truth. Both CSNet and FSNet can improve the predicted segmentation region under the supervision of the loss function of labeled data.

4.3 Experimental configurations

4.3.1 Data Description

The dataset is the same as the experiments in the previous study (see section 3.4).

As a reminder, the HIFU dataset was collected at the State Key Laboratory of Ultrasound in Medicine and Engineering (Chongqing Medical University, Chongqing, China). Sagittal T2-weighted images were performed using a 3.0-T MRI system (Signa HD, GE Healthcare, Milwaukee, WI, USA). The standardized parameters of the T2WI sequence were as follows: Repetition time (TR) 3040ms/Echo time(TE) 107.5ms, slice thickness 6mm, slice gap 1mm. The median age of the patients was 40.8 years.

The MR dataset is a uterine fibroids dataset containing 297 labeled 3D fat-suppressed T2-weighted MRI scans with the uterus, uterine fibroids, and spine. Each MR volume consists of 25 slices of 304 × 304 pixels. We split them into 260 scans for training and the remaining 37 scans for testing. The ground-truth was manually annotated and confirmed by two radiologists through a proper annotation process to ensure the consensus agreement of the annotation.

The ethics committee approved the study at Chongqing Medical University. The patients signed an informed consent form before each procedure.

4.3.2 Experimental Setup and details

The framework is implemented using Pytorch and trained by the Adam optimizer. In the training data, the segmentation models of CSNet and FSNet are called M1, M2, respectively. Firstly, M1 is pre-trained with labeled data for 50 epochs, and the model parameters are saved.

We then import the pre-trained model parameters and train M1 and M2 with both labeled and unlabeled data for 50 epochs. The training set is further divided into training and validation sets in an 8:2 ratio. For example, if 10% of the training data is used as labeled data, we use data from 26 patients for training (21 patients as the training set and 5 patients as the validation set). The data division is randomized and different slices of the same patient do not appear in both the training or validation sets. All the models are trained with an initial learning rate of 0.001, which decays by 1/2 after every 10 epochs. The best model is saved based on the validation accuracy, and then the best weights are saved to infer the test data. The Mixing parameter α for the datasets was set to 1.0. The batch size was 4. The computation was performed on an RTX 2080Ti GPU.

The data and hyperparameters are fixed during the comparison experiments and ablation experiments.

4.3.3 Evaluation Criteria

To evaluate the performance of the segmentation, we employed some of the most commonly used metrics such as the DSC similarity coefficient (DSC), precision (PR), and recall rate (RR) (see Section 3.4.3).

4.3.4 Comparison with Other Deep Learning Methods

We compared our PLRNet with four SOTA semi-supervised learning approaches, including ASDNet [11], Latent Mixup [10], and Cross-Consistency Training (CCT) [12], this for 3 different labeled/unlabeled data ratios. Besides, we added two fully-supervised methods: the classical Vanilla U-Net [13] and HIFUNet [14] with the whole set of labeled data as the performance reference. All the experiments were conducted in a fair way with the same training, test data, and network hyperparameters.

For quantitative comparison, Table 4.1 shows the DSC, PR, and RR indices obtained on the HIFU dataset by the different methods. In order to test the impact of the ratio of labeled/unlabeled data on the results of the methods, we used respectively 10%, 25%, and 40% of the training data as labeled data (26, 65, and 104 patients) and the remainder as unlabeled data. As shown in this table, our method is better than other semi-supervised learning methods, and this is for all ratios of labeled/unlabeled data. This trend is also more pronounced for low ratios of labeled/unlabeled data.

As expected, the segmentation performance is improved when the number of labeled data increases. However, it should be noted that our method still shows segmentation performance close to that of a fully supervised U-Net (100% of the labeled data) even when the number of labeled datasets is only 65 scans (25% of the labeled data).

Table 4.1 – Quantitative comparison of three evaluation metrics of different segmentation methods on the testing dataset (best results are indicated in bold)

Method	#Scans used		DSC(%)			PR(%)			RE(%)		
	Labeled	Unlabeled	Uterus	Fibroid	Spine	Uterus	Fibroid	Spine	Uterus	Fibroid	Spine
U-Net	260 (100%)	0	75.34	77.58	78.15	76.97	78.39	89.10	74.81	79.23	71.46
HIFUNet	260	0	82.37	83.51	85.01	79.45	84.48	82.51	86.00	83.70	88.69
U-Net	104 (40%)	156 (60%)	73.32	70.73	83.76	68.43	**80.35**	76.44	80.49	65.43	**93.39**
CCT	104	156	58.28	42.34	73.97	48.05	65.11	76.60	76.87	33.02	75.16
ASDNet	104	156	71.05	69.76	85.60	71.95	66.09	86.53	71.41	81.10	85.44
Latent Mixup	104	156	73.72	74.17	84.38	69.92	69.87	81.77	79.00	83.95	87.99
PLRNet	104	156	**76.40**	**77.32**	**85.43**	**76.38**	77.99	**86.62**	**80.53**	**85.27**	88.03
U-Net	65 (25%)	195 (75%)	67.39	65.72	84.56	68.1	63.24	82.22	68.27	74.56	88.03
CCT	65	195	51.50	37.75	72.94	59.69	67.29	**82.51**	48.33	28.17	68.33
ASDNet	65	195	67.52	69.67	84.72	67.44	63.47	79.14	69.3	82.28	91.97
Latent Mixup	65	195	71.41	73.36	82.47	71.02	68.91	73.96	73.48	**82.83**	**94.07**
PLRNet	65	195	**76.01**	**75.97**	**85.98**	**75.22**	**76.88**	80.00	**78.90**	81.42	**93.76**
U-Net	26 (10 %)	234 (90%)	58.94	55.48	80.18	54.88	58.35	**79.56**	66.24	60.84	82.16
CCT	26	234	52.68	36.55	70.29	43.30	62.25	65.49	69.67	27.24	78.46
ASDNet	26	234	64.76	52.48	81.25	59.14	66.85	86.37	**74.07**	52.40	77.81
Latent Mixup	26	234	65.02	64.66	84.42	66.24	61.23	83.10	65.43	**76.31**	86.52
PLRNet	26	234	**72.95**	**71.71**	**85.96**	**72.83**	**74.97**	79.53	**74.39**	**80.04**	**87.10**

If we now look at the segmentation performance on an organ by organ basis, we can see that the spine has the highest segmentation accuracy due to the large contrast difference between the spine and the surrounding organs, which makes segmentation less difficult. As we suspected, the performance is lower for the 2 organs with a smaller contract and/or greater variability in shape (uterus and fibroid).

Fig. 4.4 compares the segmentation results on two different slices (one with one fibroid -bottom- and one with multiple fibroids -top-) by using the different SOTA methods at different labeled/unlabeled ratios to the corresponding ground-truth. On these images, we can make several qualitative observations on the behavior of the different methods:

1. The spine is segmented with higher accuracy than the uterus and the fibroid. This is due to the more fixed size and shape of the spine, and the more obvious difference in contrast with the surrounding tissues on the MRI image.
2. The case of multiple fibroids is much more complex than that of single fibroids. One explanation could be that in cases of multiple fibroids, the contrast and size of the fibroids are sometimes not the same. As shown in the case of multiple-fibroids at the top of Fig. 4.4, the two fibroids have different intensities, and it is easy to confuse them with the surrounding tissues whose contrast is similar to those of the fibroids.
3. As the ratio of annotated data increases, it does not always improve the segmentation results. For example, on the case of multiple-fibroids (the top of Fig. 4.4), we can see that from 10% to 25%, the segmentation result of each method is improved significantly. However, from 25% to 40%, the performance of all the methods decreases except U-Net (b), CCT (c) and our PLRNet (f). However, in the single-fibroid case (bottom of Fig. 4.4), from 10% to 40%, the segmentation results of all methods are significantly improved.
4. Some of the methods show relatively poor results. For example, the CCT (c) method shows jagged boundaries. On the other hand, our method shows a segmentation behavior relatively consistent with the ground truth.

4.3.5 Ablation Studies

We also conducted a series of ablation studies to justify the effectiveness of the proposed approach.

First, the effectiveness of the large convolutional kernel structure used in CSNet and FSNet is validated. Using 25% of the data as labeled data, we compare the traditional U-Net, the large convolutional kernel network LKNet, and the large convolutional kernel WLKNet with wavelet sampling used in this paper to extract features from the data. Figure 4.5 shows the results of the pseudo-label generated for the unlabeled input data when these three baseline models are used as CSNet. It can be seen that using U-Net as the feature extraction network, the feature extraction

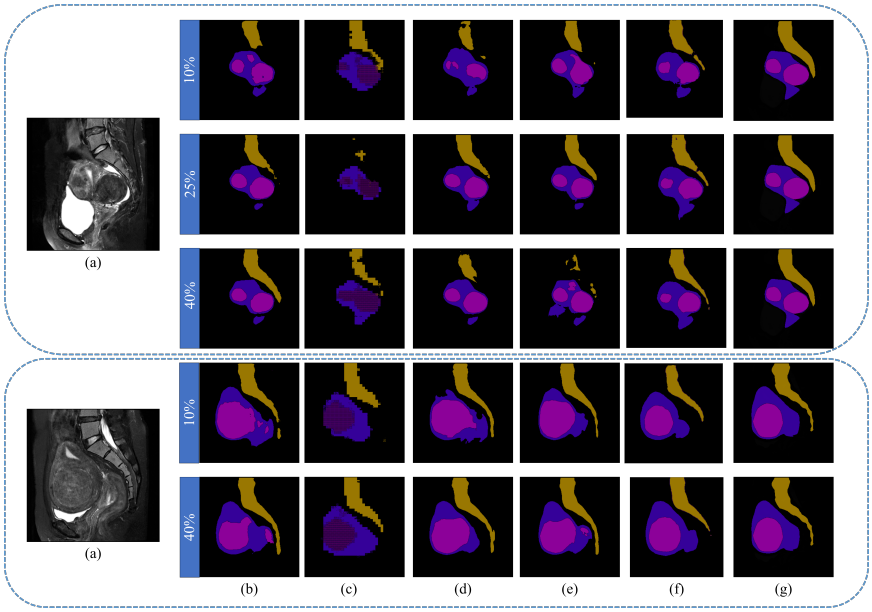

Figure 4.4 – Segmentation results of 2 slices obtained by different SOTA methods with 3 different percentages of labeled data (10%, 25%, and 40%) and the corresponding ground-truth. From left to right are the (a) raw image, (b) results of U-Net, (c) CCT, (d) ASDNet, (e) Latent Mixup, (f) our PLRNet and, (g) the ground-truth. Blue represents the uterus, pink the fibroids and yellow the spine.

capability is not sufficient and the quality of the generated pseudo-labels is low, whereas with the large convolutional kernel, the abstract feature extraction capability is improved due to the expansion of the valid receptive field. In addition, the quality of the pseudo-label is substantially improved after using the wavelet sampling operation. Our network removes the possible signal interference in the original feature extraction operation, thus maximizing the preservation of spatial information. Table 4.2 quantitatively evaluates the quality of the pseudo-labels.

Table 4.2 – Results of pseudo-labels generated by different segmentation networks of CSNet using 25% of the labeled data (best results are in bold)

Baseline	DSC(%)			PR(%)			RR(%)		
	Uterus	Fibroid	Spine	Uterus	Fibroid	Spine	Uterus	Fibroid	Spine
U-Net	53.25	58.45	46.74	63.88	69.53	97.84	47.90	56.48	31.76
LKNet	64.40	70.21	69.83	66.24	75.46	95.83	64.55	68.18	55.18
WLKNet	**69.53**	**76.31**	**77.66**	**72.53**	**76.03**	**95.98**	**68.06**	**78.04**	**66.06**

Then, we analyzed the confidence-based threshold adaptation under 25% training data of the HIFU dataset. We compared our automatic adaptive threshold strategy with different offline fixed threshold settings, ranging from 0.1 to 0.8. The results in Table 4.3 show that some of the fixed thresholds can give a good segmentation performance for one specific organ. For

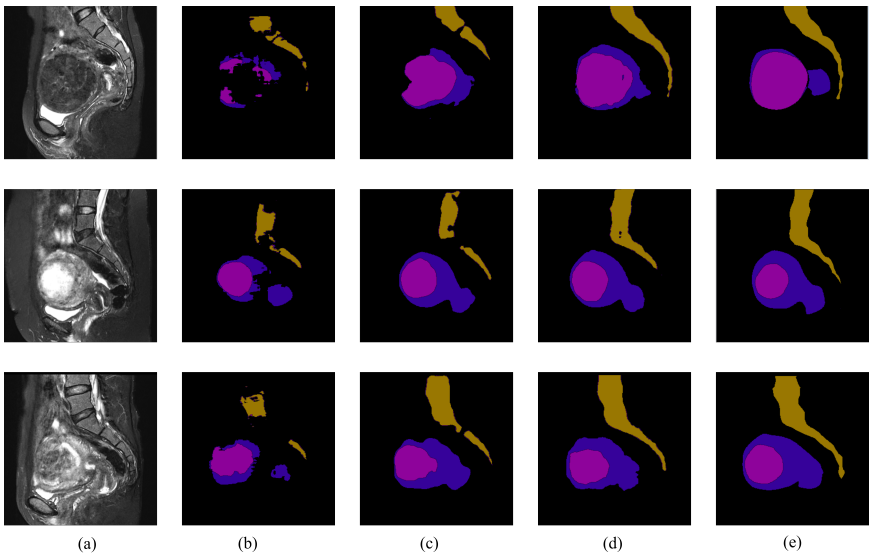

Figure 4.5 – Segmentation results of pseudo-label generated by different segmentation models with 25% labeled data. From left to right are the (a) raw image, (b) results of U-Net, (c) LKNet, (d) WLKNet, and, (e) the ground-truth. Blue represents the uterus, pink the fibroids and yellow the spine.

example, 0.4 is the best threshold for fibroid segmentation, while a threshold of 0.3 shows better performance for uterus and spine segmentation. However, our method achieves the best average performance and the best performance for almost all the target organs.

Table 4.3 – DSC(%) of the proposed Confidence-Based Threshold Adaptation module on a 25% labeled dataset (best results are indicated in bold)

Threshold	Uterus	Fibroid	Spine	Average
0.1	69.87	69.43	81.34	73.55
0.2	69.66	70.05	82.33	74.01
0.3	72.36	74.40	84.00	76.92
0.4	72.18	75.03	83.47	76.89
0.5	71.07	70.32	83.23	74.87
0.6	70.43	71.98	82.65	75.02
0.7	69.38	70.35	80.29	73.34
0.8	70.29	69.21	79.56	73.02
Adaptive	**76.01**	**75.97**	**85.98**	**79.32**

Fig. 4.6 shows the adaptation of the threshold value during the training process. It can be seen that the threshold gradually converges from 0.80 to about 0.25, and there is a sharp to slow decrease during the training process. This finding indicates that more training rounds are needed when the network learns regions that are difficult to segment.

We also wanted to estimate the impact of our several improvements on the segmentation re-

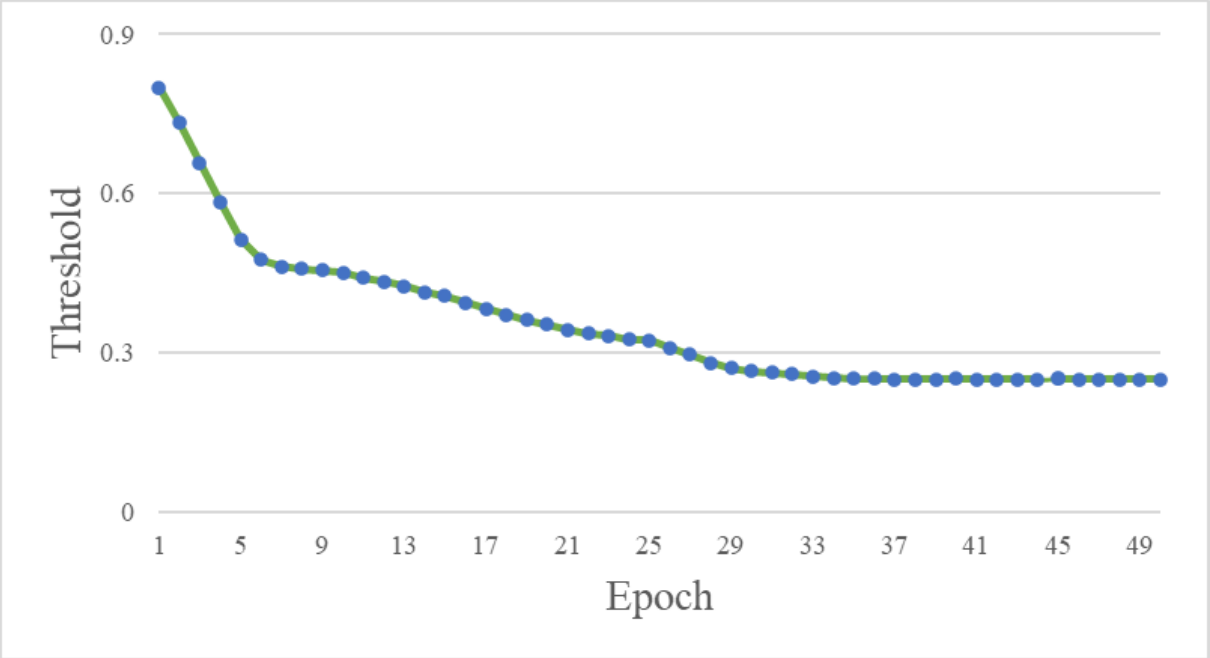

Figure 4.6 – PLRNet threshold adaptation during the training process (from 0.8 to 0.25).

sults. Table 4.4 presents the ablation study of our PLRNet with the several components or variants introduced in Section 4.2.1. All these experiments were performed using the same dataset with a 25%/75% labeled/unlabeled ratio. First, we compared the classic Vanilla U-Net (Net1) with our CSNet-FSNet architecture based on 2 U-Nets but without any other improvements (Net2). The 2 U-Nets slightly improved the results. Next, we added the CTA (Confidence-based Threshold Adaptation) to the architecture (Net3). CTA brought steady improvements in the segmentation of the uterus and fibroids. Based on this, we then compared the original Mixup (Net4) with our FAM (Feature-aligned MixUp, Net5) and found that the addition of our FAM improved the average segmentation accuracy by more than 1%. Next, the full solution (Net6) with the addition of RAD (consistent regularization and dropout) gives better segmentation results. Finally, the use of an encoder-decoder structure containing wavelet sampling with a large convolution kernel as a feature extraction scheme for CSNet and FSNet (Net7) can further improve the segmentation results. The Table shows that each component plays an important role in our semi-supervised scheme. Moreover, each of these components is independent and can be applied to other semi-supervised learning networks.

Fig. 4.7 shows the segmentation results for 3 different images obtained after adding several components or variants, *i.e.*, with the networks Net1 to Net7 (respectively Fig. 4.7. (b) to (h)). Visually comparing Net 1 (b) and Net2 (c), we can see that the segmentation accuracy is improved for all the three segmentation targets when we add the semi-supervised mechanism.

Table 4.4 – Effectiveness of the proposed techniques on the HIFU dataset using 25% of the labeled data (best results are indicated in bold)

Method	Baseline	Components					DSC(%)			
		Semi	CTA	Mixup	FAM	RAD	Uterus	Fibroid	Spine	Average
Net1	U-Net						67.39	65.72	84.56	72.56
Net2		✓					69.46	69.71	85.01	74.73
Net3		✓	✓				70.21	71.04	84.24	75.16
Net4		✓	✓	✓			71.34	72.83	81.00	75.05
Net5		✓	✓		✓		73.01	72.86	82.76	76.21
Net6		✓	✓		✓	✓	74.72	73.51	84.61	77.62
Net7	WLKNet	✓	✓		✓	✓	**76.01**	**75.97**	**85.98**	**79.32**

Looking at Net3 (d), we see that although the adaptive threshold may have a negative impact on the segmentation of the spine. It allows solving the "hole" problem that occurs in the segmentation. When comparing Net4 (e) and Net5 (f), we can see that FAM can further refine the segmentation of the uterus and improve the boundary segmentation problem compared to the traditional Mixup. One explanation could be the data augmentation at each scale of the feature map, especially for the continuous segmentation of the spine, or, more obviously, the refinement of the cervical part of the uterus. Finally, using the strategy of incorporating regularization Net6 (g), the generalization and robustness of the model are further improved. This allows us to better estimate the details of the segmentation targets, to reduce the false segmentation caused by the noise in the pseudo-labels, and to make the morphology of the segmentation targets (smoothness of the fibroid, integrity of the uterine shape, and continuity of the spine) more consistent with the real situation. Finally, replacing the segmentation network with WLKNet (Net7 (h)) yields results comparable to the reference result (i) with better details of the fibroid as well as the spine.

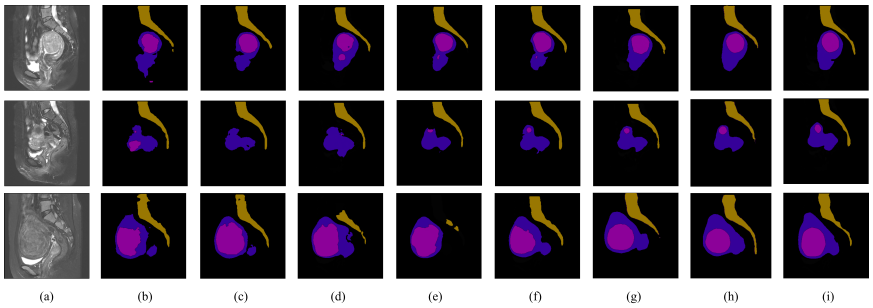

Figure 4.7 – Visualization segmentation results of ablation study on 25% of labeled data. From left to right corresponds to the network with the different components in Table 4.4: (a) the raw image, (b)-(h) segmentation results using Net1 to Net7 and (i) the ground-truth. Blue represents uterus, pink fibroids and yellow spine.

4.4 Discussion and Conclusion

We started from the semi-supervised learning method principle which consists in producing pseudo-labels from unannotated data, pseudo-labels which are then re-injected into the model during the learning process to make it more efficient. We have made several improvements to this scheme.

In order to perform feature extraction on finite labeled data as large as possible, a large convolution kernel operation is used instead of the traditional small convolution kernel operation to extend the valid receptive field. In addition, wavelet transform is used to preprocess the original signal and then laterally subsume multiple consecutive low-frequency components to reduce the dimensionality and information distortion of the image. DWT and IDWT are used to replace the down-sampling operation (pooling) and up-sampling (unpooling) operations in conventional CNN networks to effectively suppress high-frequency noise and achieve more powerful feature extraction capability than signal processing in the spatial domain. Also, our framework has a better denoising effect on pseudo-label of unlabeled data.

In order to give more importance to the reliable pixels, we defined a weighting at pixel level (see (4.1)). Thus more higher-confidence pixels correspond to higher weights. In this way, the network focuses more attention on regions with high weights, thus improving the quality of the training data. Second, the thresholds are determined using an adaptive approach, which can alleviate the problem of time-consuming manual grid search. The threshold value gradually converges from the initial set value of 0.8 to around 0.25. In the early epochs, the network focuses on pixels with higher confidence, corresponding to the easy-to-learn areas in an image. During the learning progression, the threshold gradually decreases, and the network starts to focus on some areas with lower confidence, which correspond to the hard-to-learn areas. The network gradually learns so the features of the images by the threshold adaptive method from easy to difficult.

The introduction of the Feature-aligned Mixup improves the generalizability of the model and effectively avoids over-fitting the data. The core of the strategy is to add complexity control to the space that is not covered by the training data. Our strategy performs linear interpolation using the data points generated in the encoding-decoding structure. This data augmentation model allows the neural network to learn a simple linear interpolation function in the ”blank region”, thus reducing the complexity of the uncovered space. The input of FSNet is the product of the original image and the four-channel feature. This ensures that the network pays more attention to the region of interest by adding soft attention to each channel, reducing the effect of irrelevant regions on the segmentation results.

By introducing the consistency regularization loss, the intermediate representation of the same data in two networks tends to be consistent, which improves the robustness and generalization of the model.

One possible limitation would be that the model is biased toward the dominant class. The model is trained based on labeled data distribution at the early training stage. Thus, a biased distribution of categories for the pretrained labeled data may directly affect the quality of pseudo-label generation and may impact the training results. Second, we used a simple U-Net for feature extraction, ignoring the difficulties of low tissue contrast around the uterus and fibroids with varying sizes and shapes, which is insufficient for feature extraction.

In conclusion, we have proposed a novel semi-supervised framework named PLRNet to improve the quality of pseudo-labels. The main contribution of our approach is the adaptive thresholds learning to automatically generate high-quality pseudo-labels for semi-supervised learning. This allows us to abandon offline threshold tuning. We validated our method on data used for HIFU fibroid treatment planning. This evaluation demonstrated that our segmentation network outperformed the SOTA semi-supervised learning methods.

The most prominent future work is to improve the quality of pseudo-labels by designing class-wise thresholds to generate unbiased pseudo-labels. In addition, we plan to extend our approach to external datasets from different sites. We will explore how to select and annotate representative data and extract richer feature representations from limited data annotation.

BIBLIOGRAPHY

[1] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," arXiv preprint arXiv:1610.02242, 2016. [Online]. Available: https://arxiv.org/abs/1610.02242

[2] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in ICML 2013 Workshop : Challenges in Representation Learning (WREPL), vol. 3, 2013, p. 896, issue: 2.

[3] H. Zhang, M. Cissé, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," arXiv preprint arXiv:1710.09412, 2017. [Online]. Available: http://arxiv.org/abs/1710.09412

[4] L. Carratino, M. Cissé, R. Jenatton, and J.-P. Vert, "On mixup regularization," arXiv preprint arXiv:2006.06049, 2020. [Online]. Available: https://arxiv.org/abs/2006.06049

[5] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in 2019 IEEE/CVF International Conference on Computer Vision (ICCV), oct 2019, pp. 6023–6032.

[6] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. Raffel, "Mixmatch: A holistic approach to semi-supervised learning," arXiv preprint arXiv:1905.02249, 2019. [Online]. Available: https://arxiv.org/abs/1905.02249

[7] Q. Li, L. Shen, S. Guo, and Z. Lai, "Wavelet integrated CNNs for noise-robust image classification," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 7245–7254.

[8] T. Williams and R. Li, "Wavelet pooling for convolutional neural networks," in International conference on learning representations, 2018.

[9] F. Pizzati, R. d. Charette, M. Zaccaria, and P. Cerri, "Domain bridge for unpaired image-to-image translation and unsupervised domain adaptation," in 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), mar 2020, pp. 2990–2998.

[10] P. K. Gyawali, S. Ghimire, P. Bajracharya, Z. Li, and L. Wang, "Semi-supervised medical image classification with global latent mixing," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, ser. Lecture Notes in Computer Science, vol. 12261, 2020, pp. 604–613.

[11] D. Nie, Y. Gao, L. Wang, and D. Shen, "Asdnet: Attention based semi-supervised deep networks for medical image segmentation," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, ser. Lecture Notes in Computer Science, vol. 11073, 2018, pp. 370–378.

[12] Y. Ouali, C. Hudelot, and M. Tami, "Semi-supervised semantic segmentation with cross-consistency training," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), jun 2020, pp. 12 674–12 684.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2015, ser. Lecture Notes in Computer Science, vol. 9351. Springer, 2015, pp. 234–241.

[14] C. Zhang, H. Shu, G. Yang, F. Li, Y. Wen, Q. Zhang, J.-L. Dillenseger, and J.-L. Coatrieux, "HIFUNet: multi-class segmentation of uterine regions from MR images using global convolutional networks for HIFU surgery planning," IEEE Transactions on Medical Imaging, vol. 39, no. 11, pp. 3309–3320, nov 2020.

Chapter 5

Automatic segmentation for plan-of-the-day selection in CBCT-guided adaptive radiation therapy of cervical cancer

5.1 Introduction

The standard treatment for locally advanced cervical cancer (LACC) is external beam radiotherapy (EBRT) with chemotherapy, followed by brachytherapy. Although Intensity-Modulated Radiation Therapy (IMRT) is used to reduce normal tissue toxicity [1, 2], it is limited by large and complex intrapelvic anatomical variations occurring between the treatment fractions [3, 4]. The position and shape of the clinical target volume (CTV, including the cervix, uterus, upper-vagina, and parametrium) are highly dependent on the bladder and rectum filling, and on tumor regression during treatment [5, 6, 7]. In the context of adaptive radiation therapy (ART), plan-of-the-day (PoD) strategies have been proposed based on the generation of a treatment plan library, including several treatment plans optimized according to multiple planning CTs (pCT) acquired with various bladder fillings [8, 9, 10]. At each treatment fraction, the treatment plan is then selected among those of the library ("plan-of-the-day") based on an in-room image (*e.g.*, CBCT image, see Figure 5.1). Although this strategy appears to be adequate to compensate for uterine motions [9, 10, 11], it remains complex in a clinical workflow.

Different factors limit the rapid advancement of PoD ART. The PoD selection is actually a difficult process, mainly due to the poor contrast of CBCT images and large anatomical deformations. Thus, the expert needs to visualize the full 3D volume to assess the coverage of the whole target by the available treatment plans. The PoD selection is, therefore a time-consuming process which is submitted to interobserver variability, as demonstrated in [12] where 26 operators manually selected PoD on 24 CBCT images. This study showed a high inter-observer variability since the optimal PoD was chosen on average by 60% of users.

In order to automatize PoD selection, the automatic segmentation of CBCT images has

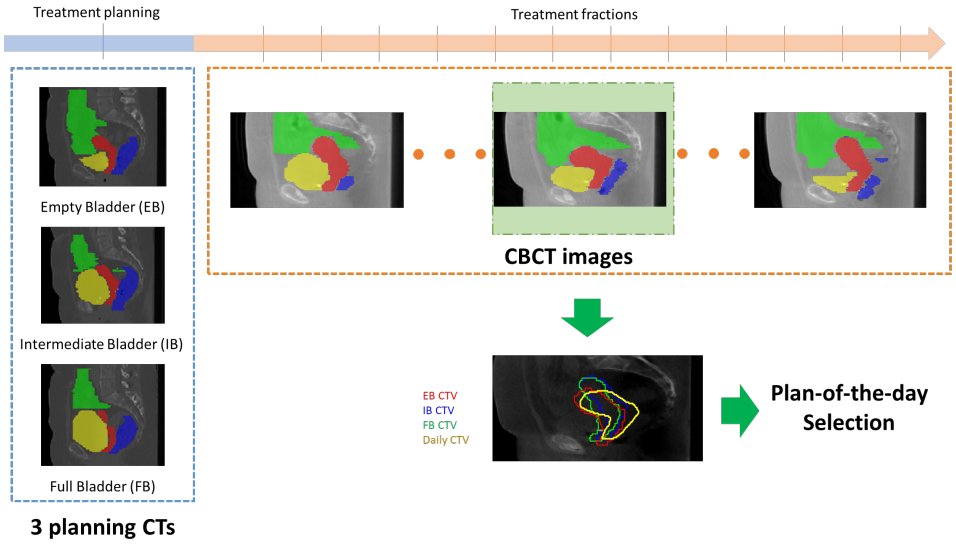

Figure 5.1 – Flowchart of plan-of-the-day ART for cervical cancer. The process consists of three steps (1) Planning: acquisition of multiple planning CT scans with variable bladder volumes. (2) Acquisition of the CBCT image of the day. (3) Selection of the most appropriate treatment plan to maximize the target coverage. The CTV, bowel sac, bladder and rectum are represented as red, green, yellow and blue filled contours. For plan-of-the-day selection, empty bladder (EB), intermediate bladder (IB), full bladder (FB) and daily CTV are represented on the daily CBCT as red, blue, green and yellow contours, respectively.

been proposed to measure, at each treatment fraction, the ability of the treatment plans to treat the target [13]. Langerak *et al.* [14] proposed to use a multi-atlas-based segmentation method. On a total of 224 CBCT, the CBCT images corresponding to low confidence levels were firstly removed, resulting to 187 images on which the Dice values were 0.85, 0.81, and 0.80 for the uterus, bladder, and rectum, respectively. However, the CBCT image quality is limited by noise, artifacts, and low soft-tissue contrast, making automatic segmentation very challenging. Recently, with the widespread use of deep learning (DL) in medical imaging, Beekman *et al* [15] compared different DL models, performing either direct segmentation or a segmentation prior deformation by diffeomorphic image registration. The deformation-based model performed the best on the CBCT test set, with a median Dice score of 0.80.

In the context of PoD-based ART for LACC, this work aimed to propose a strategy to automatically select the optimal treatment plan. It relies on a deep learning-based segmentation of the CBCT images, enabling the selection of the optimal treatment plan regarding the CTV coverage based on a geometrical criterion. This strategy was simulated and compared to a reference obtained from expert manual delineations.

5.2 Materials and methods

Figure 5.2 describes the flowchart of the study. Based on a set of three planning CT scans (corresponding to an empty, an intermediate and a full bladder) and on daily CBCT scans, it contains two parts applied to each daily CBCT image: (1) segmentation of the CBCT image using a deep learning network in order to obtain the segmented daily CTV ($\mathbf{CTV}_{CBCT}$); (2) selection of the most appropriate PoD among the 3 available pCTs. For this second step, the three planning CTs (pCTs) in the planning library were registered to the daily CBCT image to simulate the patient repositioning. Then, the coverage of the segmented daily $\mathbf{CTV}_{CBCT}$ by the CTVs of the 3 pCTs was computed. Finally, the best treatment plan was selected as the one corresponding to the highest daily CTV coverage. Our goal in this project was to implement this global scheme with the integration of specific image segmentation modules and the novel automatic PoD selection strategy based on the registration result. We will now present you in detail the several part of the framework and the evaluation of these parts.

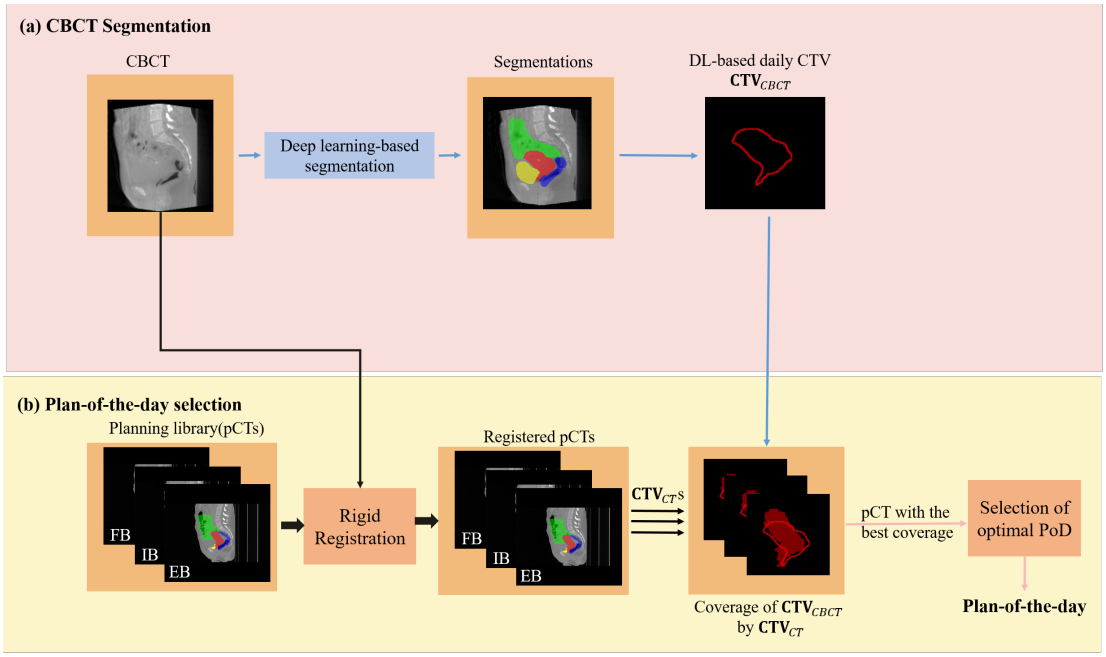

Figure 5.2 – Flowchart of the study. The steps are: (1) CBCT segmentation using deep learning and (2) plan-of-the-day (PoD) selection using clinical target volume (CTV) contours. The PoD selection relies on: (a) bone-based rigid registration of the planning CTs (pCTs) with the daily CBCT to simulate patient repositioning; (b) computation of the coverage between the daily CTV ($\mathbf{CTV}_{CBCT}$) and the CTVs of the pCTs ($\mathbf{CTV}_{EB}$, $\mathbf{CTV}_{IB}$ and $\mathbf{CTV}_{FB}$); (c) selection of the best treatment plan based on target coverage: the pCT corresponding to the highest coverage was selected. (EB: empty bladder; IB: intermediate bladder; FB: Full bladder; cov: coverage value).

5.2.1 Data acquisition and experimental settings

In this study, we collected 272 CBCT scans and 63 CT scans from 23 patients. All patients were treated with a combination of external beam radiation therapy (EBRT) and pulse-dose-rate brachytherapy (PDR-BT). EBRT delivered a total dose of 45 Gy to the pelvis (supine position), at 1.8 Gy per fraction, using IMRT along with concomitant weekly cisplatin (40 mg/ mm^2). PDR-BT was delivered following the GEC-ESTRO recommendations. All patients provided signed informed consent.

Each patient underwent two or three planning CTs with different bladder volumes: empty bladder (EB), intermediate bladder (IB), and full bladder (FB). One hour before the first IB CT, the patient had to drink 250 mL of water. Then another 500 mL of water should be consumed to obtain the FB CT after 20 minutes. For EB CT, the patient emptied her bladder. All CTs were scanned (Big Bore, Philips) with voxel spacing ranging from $0.87 \times 0.87 \times 3.00$ mm^3 to $1.21 \times 1.21 \times 3.00$ mm^3. The dimensions range of CT volumes was from $75 \times 512 \times 512$ to $168 \times 512 \times 512$. During the radiation therapy treatment, 5 to 16 CBCT images were acquired (XVI mounted on a Synergy linac, Elekta) for each patient with voxel spacing of $1.00 \times 1.00 \times 2.00$ mm^3 and dimensions of $132 \times 410 \times 410$.

All images (*i.e.*, planning CTs and CBCTs) were manually contoured slice-by-slice by one radiation oncologist. These contours were the primary CTV, including the cervix, uterus, and upper-vagina, and the rectum, bladder, and bowel bag (including the sigmoid). These delineations were considered as the reference in this study.

5.2.2 CBCT segmentation using deep-learning

The segmentation model was trained using the deep learning-based method nnU-Net, which has been demonstrated to be efficient in multiple medical image segmentation tasks [16]. NnU-Net's automatic configuration runs without human intervention when it is applied to a new dataset. The nnU-Net focuses on pre-processing, training, inference strategies, and post-processing. Although there are several methods available in nnU-Net: 2D U-Net (2d), 3D full resolution U-Net (3d_fullres), 3D low resolution U-Net (3d_lowres) and 3D cascade U-Net (3d_cascade), we didn't want to train all the models because the training process is time-consuming. We only trained nnU-Net 3d_fullres as our base model which has been shown to be one of the best performing models in many medical image segmentation tasks [17, 16].

All images were cropped to the region of nonzero values with a cropping size of $84 \times 410 \times 410$ voxels. The resampling voxel size was the median voxel spacing of the dataset, *i.e.* $1.00 \times 1.00 \times 2.00$ mm^3. A z-score normalization was applied to each image. In the post-processing procedure, a connected component analysis was performed to eliminate the detection of spurious false positives. The model was trained using Pytorch and stochastic gradient descent (SGD) optimizer.

The initial learning rate was set to 0.01 and decreased according to the "Poly" scheme [18]. The loss function was the sum of Dice similarly coefficient and cross-entropy with the same weight.

We randomly divided all 23 patients into four-fold using a cross-validation scheme. Table 5.1 presents the partition of the dataset. Multiple images (*i.e.* planning and daily images) from individual patients were not distributed among datasets. The test data was only used to evaluate the performance of the model in this fold and was not involved in training. nnU-Net further divided the training data into training and validation sets and performed a five-fold cross-validation to automatically select the best network configuration. In the end, the model (or ensemble) which got the best performance was chosen to perform the inference on the test sets of this fold. The number of epochs during training was 1000 for every fold. The evaluation of the segmented volumes is described in part 5.2.4. Table 5.2 reports the network configurations generated by nnU-Net for the considered dataset.

Table 5.1 – Partition of the dataset for the deep learning network training and testing. The population was randomly separated into four folds using a cross-validation scheme to evaluate the model performance. Multiple images (*i.e.* planning and daily images) from individual patients were not distributed among datasets.

Model	#Patient		#CBCT volume	
	Train	Test	Train	Test
1	18	5	216	56
2	17	6	197	75
3	17	6	206	66
4	17	6	199	73

Table 5.2 – Network configurations generated by nnU-Net.

Parameters	**3D full resolution U-Net**
Normalization	Z normalization
Patch size	$56 \times 192 \times 160$
Batch size	2
Downsampling strides	[[1,2,2],[2,2,2],[2,2,2],[2,2,2],[1,2,2]]
Convolution kernel sizes	[[1,3,3],[3,3,3],[3,3,3],[3,3,3],[3,3,3],[3,3,3]]

5.2.3 Plan-of-the-day selection

After obtaining the segmentation results for each CBCT image, the following three steps were performed: (1) Bone-based rigid registration between each pCT with different bladder fillings (FB, IB, EB) and the daily CBCT to simulate patient repositioning; (2) computation of the coverage between the CTV of CBCT ($\mathbf{CTV}_{CBCT}$) and CTV of the CTs ($\mathbf{CTV}_{CBCT}$) and CTV of the pCTs ($\mathbf{CTV}_{EB}$, $\mathbf{CTV}_{IB}$,$\mathbf{CTV}_{FB}$); (3) selection of the pCT corresponding to the best coverage.

Rigid registration

For each patient, a bone-based rigid registration was performed between each CBCT and each pCT, using the ROI of the CBCT which was limited to the treated region. In the clinical setup, the rigid transformation would result from moving the patient between the CT and the EBRT device with the CBCT. The library Elastix [19] was used on thresholded images to keep only the bones to estimate the rigid transformation (translation and rotation) with normalized correlation as the metric. The resulting rigid transformation was visually validated by checking the bone alignment and was applied to the pCT's corresponding delineations.

Selection of the PoD

To evaluate the ability of each treatment plan to treat the CTV in its daily position, a coverage index was computed between the $\mathbf{CTV}_{CBCT}$ segmented by the deep learning model and the different CTV of the patients's library $\mathbf{CTV}_{CT} \in \{\mathbf{CTV}_{EB}, \mathbf{CTV}_{IB}, \mathbf{CTV}_{FB}\}$:

$$\text{cov} = \frac{|CTV_{CBCT} \cap CTV_{CT}|}{|CTV_{CT}|} \tag{5.1}$$

where $|.|$ is the cardinality of the set. In this way, each CBCT had a coverage value associated with each of the three pCTs (IB, EB, FB) of the patient. All the three pCTs of the considered patient were ranked according to the corresponding coverage index. The pCT corresponding to the highest coverage value was selected as the PoD of the considered treatment fraction.

5.2.4 Evaluation protocol

The data of all the 23 patients described previously were used to evaluate the proposed PoD selection process. As said previously the rigid registration was visually validated by checking the bone alignment. However we developed the following protocol to evaluate the segmentation and the PoD selection methods.

Segmentation evaluation

The segmentation of CBCT images was evaluated using a four-fold cross-validation, considering the manual expert delineations as the reference. The following geometric metrics were computed for CTV, bowel bag, rectum, and bladder: DSC, MAD and 95HD (see section 3.4.3).

PoD selection evaluation

For each treatment fraction, the PoD resulting from the proposed automatic process was compared to the one obtained using the reference delineation of the CTV in the CBCT image.

Thus, this reference PoD corresponded to the best coverage between the manual delineation of the $\mathbf{CTV}_{CBCT}$ and $\mathbf{CTV}_{CT} \in \{\mathbf{CTV}_{EB}, \mathbf{CTV}_{IB}, \mathbf{CTV}_{FB}\}$.

To consider that multiple treatment plans may provide a satisfying coverage of the target, a 5% tolerance was applied to the maximum coverage, and the corresponding treatment plans were also selected.

The accuracy of the PoD selection was determined by calculating the number of automatically selected PoD that were identical to the reference PoD or included in the selected treatment plans.

5.3 Results

5.3.1 Performance of the segmentation

Table 5.3 reports the mean (range) values of the geometric metrics for the automatic delineations of the observed organs. Figure 5.3 summarizes also these quantitative results as boxplots. In this figure, the median DSC of primary CTV, bowel bag, rectum, and bladder were 0.79, 0.81, 0.75, and 0.84, respectively. Figure 5.4 shows some visual results of the automatic contouring for four patients in some axial and sagittal views.

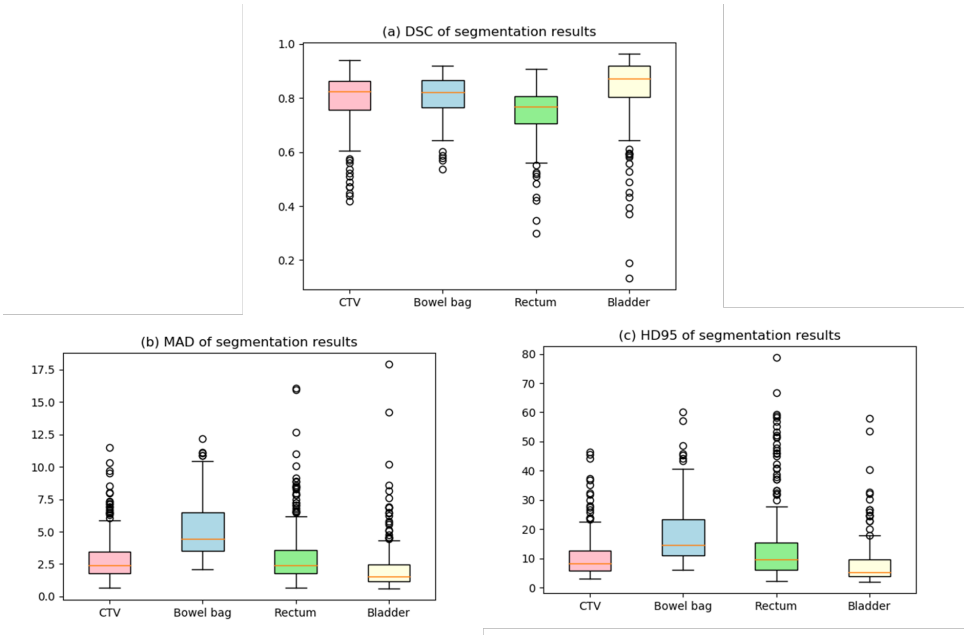

Figure 5.3 – Quantitative segmentation results of CTV (uterus), bowel bag, rectum and bladder presented as boxplots. (a) DSC = Dice similarity coefficient ; (b) MAD = Mean absolute distance ; (c) HD95 = 95^{th} percentile Hausdorff Distance

Table 5.3 – Quantitative evaluation results of CBCT segmentation.

Organ	DSC	MAD (mm)	HD95 (mm)
CTV	0.79 (0.42 - 0.94)	2.90 (0.65 - 11.48)	10.49 (3.02 - 46.43)
Bowel bag	0.81 (0.53 - 0.92)	5.22 (2.12 - 12.15)	18.40 (6.12 - 60.43)
Rectum	0.75 (0.30 - 0.91)	3.14 (0.65 - 16.04)	14.19 (2.24 - 78.79)
Bladder	0.84 (0.13 - 0.96)	2.91 (0.61 - 17.9)	9.45 (2.03 - 58.22)

DSC = Dice similarity coefficient; MAD = Mean absolute distance; HD95 = 95^{th} percentile Hausdorff Distance

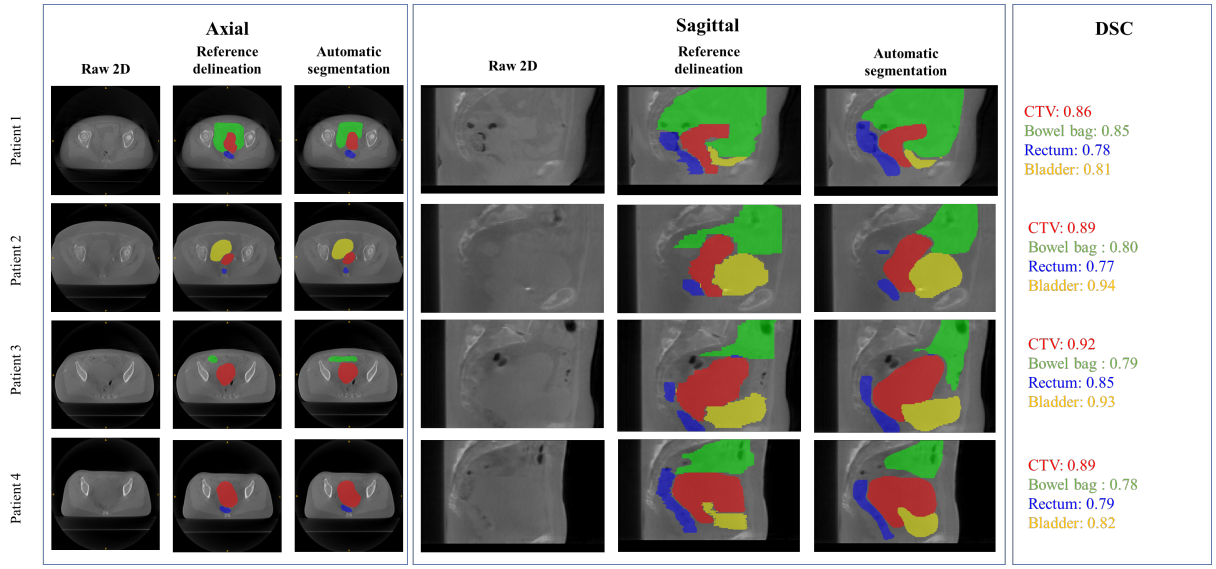

Figure 5.4 – Examples of segmentation on CBCT. The contours are represented on the axial and sagittal views in red, green, yellow and blue for the primary CTV, bowel bag, bladder and rectum, respectively. DSC: Dice similarity coefficient

5.3.2 Performance of treatment plan selections

Considering the accuracy of strict automatic PoD selection (without the 5% tolerance), an agreement between the reference PoD and the automatically identified PoD was observed for 91.5% of CBCTs. Thus, in this case, 23 out of 272 tested CBCTs having a suboptimal selected PoD compared to the reference.

Using the 5% tolerance value for PoD selection, multiple (up to three) pCTs could be selected as optimal PoD. This resulted in an increased PoD selection agreement to 99.6%, with one CBCT having suboptimal PoD selection. Figure 5.5 shows the PoD selection on four different patients corresponding to an agreement between the proposed and selected PoDs. Figure 5.6 illustrates the only case with suboptimal PoD selection. For this case, the automatic segmentation results were poor (Dice of CTV segmentation was 0.52), which resulted in the wrong proposed PoD.

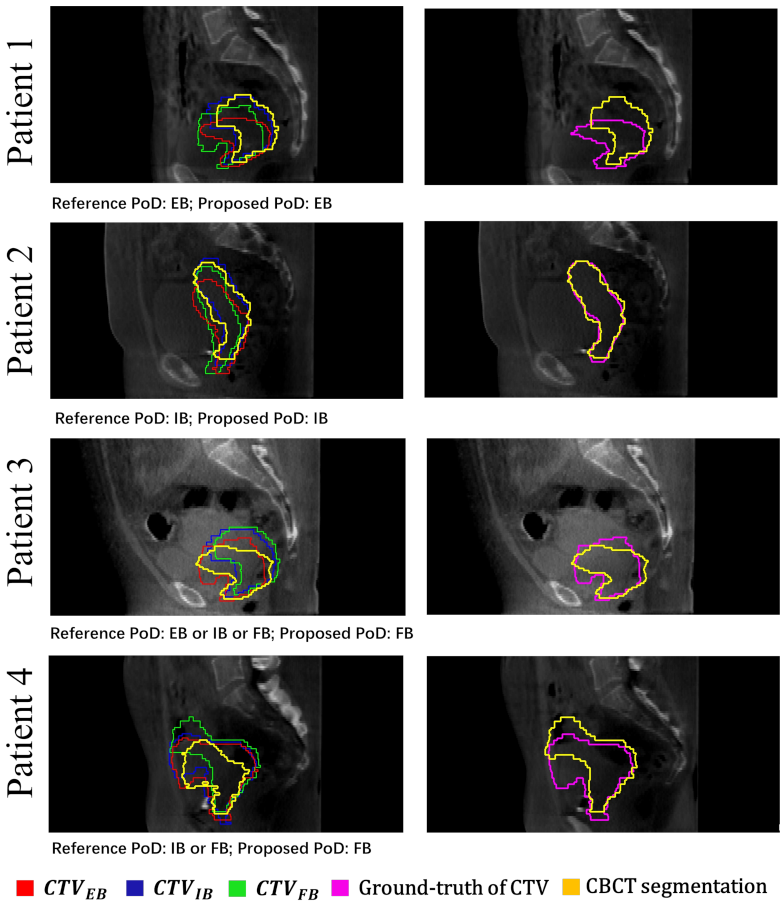

Figure 5.5 – Examples of automatic PoD selection for four patients. (Left) Result of segmentation and clinical target volumes (CTVs) corresponding to the planning library for PoD selection; (Right) Automatic and reference CTV segmentations. The cases shown here were selected based on the following criteria: single selected PoD (Patients 1 and 2); multiple PoD selections due to the tolerance value of 5% (Patient 3 and 4)

5.4 Discussion

This study aimed at improving the PoD selection for LACC ART on CBCT images. By simulating the proposed process on 272 treatment fractions, it resulted in an agreement between the reference PoD and the automatically identified PoD for all fractions except one.

The first step of the process was based on the automatic segmentation of the daily images using a deep learning network. The resulting average DSC on the 272 CBCT images were higher than 0.75 for the primary CTV and the three main organs at risk (bowel bag, rectum, and bladder).

To the best of our knowledge, Only two works in the literature have proposed automatic segmentation of the cervix in CBCT [14, 15]. Langerak *et al.* [14] used a multi-atlas-based

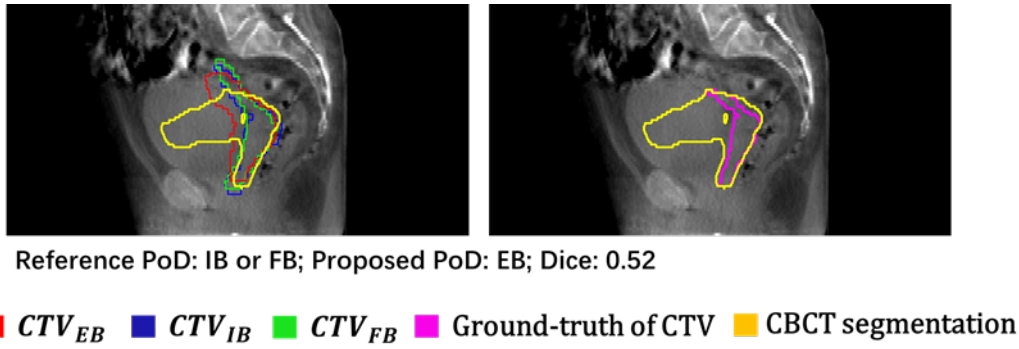

Figure 5.6 – The only case with suboptimal PoD selections. On the left, the segmentation result and clinical target volumes (CTVs) corresponding to the planning library for PoD selection; On the right the automatic CTV segmentations. The poor segmentation of CTV resulted in a suboptimal PoD selection.

segmentation method. In this study, from a total of 224 CBCT, the images corresponding to low confidence levels were firstly removed, resulting in 187 images and on these images the Dice values were 0.85, 0.81, and 0.80 for the uterus, bladder, and rectum, respectively. In our study and without discarding any CBCT image, the obtained DSC values were close to the values of the Langerak study but with lower values for the CTV and rectum, and higher values for the bladder. More recently, Beekman *et al.* [15] compared different deep learning approaches and showed that the best performances were obtained by a deformation-based registration network with a mean DSC of 0.80 on the uterus, computed on a test set of 20 CBCT images. In our study, we obtained similar DSC on 272 CBCT images. Our study and [15] are the only ones using deep learning for cervix CBCT segmentation. Although 2D and 3D U-Net (or V-Net) networks have been used on planning CT images [20, 21, 22] with a higher DSC for CTV (0.86 ± 0.08), their exploitation in CBCT remains challenging because of the lower contrast, the higher noise, but also because of a much more limited availability of reference segmentation. To our knowledge, this is the first study exploiting nnU-Net on CBCT images and on the uterine region. It confirms its good performances which have been already demonstrated on other imaging modalities [16].

In this study, the deep learning model was trained to segment not only the CTV (which is important for PoD selection), but also the main OARs. Intuitively, a binary segmentation task (segmenting only CTVs) might be easier than a multi-class segmentation task (segmenting CTVs and also OARs) and thus might yield better segmentation results. However, in terms of geometric considerations, multi-class segmentation provides a positional a priori (*e.g.* the uterus should be between the bladder and rectum). Such a positional a priori helps the convolutional neural network to constrain the position of the uterus during segmentation. Preliminary experiments have validated this assumption, with better results when OARs are included in the segmentation task.

The evaluation of the proposed process, not only in terms of segmentation accuracy (*e.g.* with Dice score) but also in terms of PoD selection, is of clinical interest since this PoD selection is the most important and difficult step in this adaptive strategy.

In the current clinical practice, the treatment plan is selected visually, resulting in potentially high inter-observer variations [12]. The identification of the optimal treatment plan appears particularly challenging in the context of complex deformations and/or limited image quality. Langerak *et al.* [14] proposed a PoD selection, after CBCT segmentation, by comparing the segmented bladder volume to the preoperative planning library (empty and full bladder). However, the shape of the CTV may be influenced by other factors than bladder filling alone, including rectum filling, tumor shrinkage, or non-moving cervix. The PoD selection should thus ideally be based primarily on the shape of the target.

In this paper, the criteria used to select the PoD was the coverage of the daily CTV by the CTVs corresponding to the different planning CTs. It enabled to consider directly the treatment of the target instead of a surrogate of the target position. Moreover, selecting the best plan as the one providing the best CTV coverage is the approach considered in the literature [12, 9]. Improving the target coverage should decrease the dose received by the OARs since the high doses would be focused on the CTV. On the other hand, since the OARs are segmented in the proposed process, it is technically possible to consider them in the PoD selection. However, this would require the definition of a decision process, or a metric, that takes into account and weights the different criteria (CTV and OARs coverage).

In our study, only one case (out of 272) resulted in a wrong PoD selection (Figure 5.6). The particularity of this case was that of a patient who had a relatively small uterus that was therefore poorly segmented. We believe that in a clinical context, this kind of result would be easily visually detected. The treatment plan could then be manually selected in the library or a backup plan could be used as proposed in [9].

Online adaptive radiotherapy has recently undergone significant improvements, especially with the development of MRI-linac and CBCT-based online optimization. Concerning the latter, although promising, only very few studies have considered this kind of systems to treat cervix cancer patients [23, 24]. Actually, the complete re-optimization workflow faces some challenges, especially related to the precise segmentation of all the considered organs on images with limited quality. If the proposed study also includes CBCT segmentation, it showed that minor segmentation uncertainties, which may be unacceptable for re-optimization, may have no impact on PoD selection. Concerning MRI-linac, very few studies have been proposed on online adaptation for cervix cancer [23], except on the precision of dose calculation [25]. Some challenges remain in the implementation of daily optimization (segmentation, pseudo-CT generation, re-optimization, and quality assurance). Long treatment time (60 min) may also be a limitation for a treatment in which hypo-fractionation is difficult when nodes have to be irradiated. Moreover,

the combination of online adaptation with PoD strategies may be of interest to optimize the clinical workflow and limit treatment times.

Our study still has some limitations. The PoD selection criterion was only based on geometrical coverage. It could be interesting to take into account dosimetric criteria, which is challenging since it would require computing the dose distribution on the CBCT images. Another limitation is related to the limited number of patients in this study. The proposed workflow will need to be evaluated on a larger cohort. In particular, the evaluation of the segmentation network will need to be further investigated, which may require improving the robustness of the model by training it with new data. Finally, the PoD ART strategy based on multiple planning CTs has shown some limitations, and some more complex strategies have been proposed to improve the libraries. For example, an "evolutive library" strategy has been proposed, enriching the library by including some CBCT anatomies into the library when the daily clinical target volume (CTV) shape differed from those in the library [26]. The inclusion of modeled anatomies resulting from population analysis was also proposed [27]. All these strategies are based on daily PoD selection, so the possibility of combining the proposed PoD selection method with them should be investigated.

5.5 Conclusion

This work proposed and evaluated an automatic workflow to select PoD for LACC ART. Based on CBCT image segmentation using a deep-learning method, it selects the optimal treatment plan based on daily CTV geometrical coverage. The evaluation on 272 treatment fractions showed a high agreement with the reference obtained by expert delineation. The proposed workflow should be further evaluated in a clinical workflow and on a larger number of patients.

BIBLIOGRAPHY

[1] A. Naik, O. Gurjar, K. Gupta, K. Singh, P. Nag, and V. Bhandari, "Comparison of dosimetric parameters and acute toxicity of intensity-modulated and three-dimensional radiotherapy in patients with cervix carcinoma: a randomized prospective study," Cancer/Radiothérapie, vol. 20, no. 5, pp. 370–376, 2016.

[2] A. K. Gandhi, D. N. Sharma, G. K. Rath, P. K. Julka, V. Subramani, S. Sharma, D. Manigandan, M. Laviraj, S. Kumar, and S. Thulkar, "Early clinical outcomes and toxicity of intensity modulated versus conventional pelvic radiation therapy for locally advanced cervix carcinoma: a prospective randomized study," International Journal of Radiation Oncology* Biology* Physics, vol. 87, no. 3, pp. 542–548, 2013.

[3] A. Buchali, S. Koswig, S. Dinges, P. Rosenthal, J. Salk, G. Lackner, D. BoÈhmer, L. Schlenger, and V. Budach, "Impact of the filling status of the bladder and rectum on their integral dose distribution and the movement of the uterus in the treatment planning of gynaecological cancer," Radiotherapy and oncology, vol. 52, no. 1, pp. 29–34, 1999.

[4] Y. Han, E. H. Shin, S. J. Huh, J. E. Lee, and W. Park, "Interfractional dose variation during intensity-modulated radiation therapy for cervical cancer assessed by weekly CT evaluation," International Journal of Radiation Oncology* Biology* Physics, vol. 65, no. 2, pp. 617–623, 2006.

[5] R. Jadon, C. Pembroke, C. Hanna, N. Palaniappan, M. Evans, A. Cleves, and J. Staffurth, "A systematic review of organ motion and image-guided strategies in external beam radiotherapy for cervical cancer," Clinical oncology, vol. 26, no. 4, pp. 185–196, 2014.

[6] P. Chan, R. Dinniwell, M. A. Haider, Y.-B. Cho, D. Jaffray, G. Lockwood, W. Levin, L. Manchul, A. Fyles, and M. Milosevic, "Inter-and intrafractional tumor and organ movement in patients with cervical cancer undergoing radiotherapy: a cinematic-MRI point-of-interest study," International Journal of Radiation Oncology* Biology* Physics, vol. 70, no. 5, pp. 1507–1515, 2008.

[7] B. M. Beadle, A. Jhingran, M. Salehpour, M. Sam, R. B. Iyer, and P. J. Eifel, "Cervix regression and motion during the course of external beam chemoradiation for cervical cancer," International Journal of Radiation Oncology* Biology* Physics, vol. 73, no. 1, pp. 235–241, 2009.

[8] M. Bondar, M. Hoogeman, J. W. Mens, S. Quint, R. Ahmad, G. Dhawtal, and B. Heijmen, "Individualized nonadaptive and online-adaptive intensity-modulated radiotherapy treatment strategies for cervical cancer patients based on pretreatment acquired variable bladder filling computed tomography scans," International Journal of Radiation Oncology* Biology* Physics, vol. 83, no. 5, pp. 1617–1623, 2012.

[9] S. T. Heijkoop, T. R. Langerak, S. Quint, L. Bondar, J. W. M. Mens, B. J. Heijmen, and M. S. Hoogeman, "Clinical implementation of an online adaptive plan-of-the-day protocol for nonrigid motion management in locally advanced cervical cancer IMRT," International Journal of Radiation Oncology* Biology* Physics, vol. 90, no. 3, pp. 673–679, 2014.

[10] A. J. van de Schoot, P. de Boer, J. Visser, L. J. Stalpers, C. R. Rasch, and A. Bel, "Dosimetric advantages of a clinical daily adaptive plan selection strategy compared with a non-adaptive strategy in cervical cancer radiation therapy," Acta oncologica, vol. 56, no. 5, pp. 667–674, 2017.

[11] M. Gobeli, A. Simon, M. Getain, J. Leseur, E. Lahlou, C. Lafond, E. Dardelet, D. Williaume, B. Rigaud, and R. de Crevoisier, "Benefit of a pretreatment planning library-based adaptive radiotherapy for cervix carcinoma?" Cancer radiotherapie: journal de la Societe francaise de radiotherapie oncologique, vol. 19, no. 6-7, pp. 471–478, 2015.

[12] M. Gobeli, B. Rigaud, C. Charra-Brunaud, S. Renard, G. De Rauglaudre, V. Beneyton, S. Racadot, K. Peignaux, J. Leseur, D. Williaume et al., "CBCT guided adaptive radiotherapy for cervix cancer: uncertainty of the choice of the plan of the day," Radiother Oncol, vol. 127, pp. S606–S607, 2018.

[13] M. L. Bondar, M. Hoogeman, W. Schillemans, and B. Heijmen, "Intra-patient semi-automated segmentation of the cervix-uterus in CT-images for adaptive radiotherapy of cervical cancer," Physics in Medicine & Biology, vol. 58, no. 15, p. 5317, 2013.

[14] T. Langerak, S. Heijkoop, S. Quint, J.-W. Mens, B. Heijmen, and M. Hoogeman, "Towards automatic plan selection for radiotherapy of cervical cancer by fast automatic segmentation of cone beam CT scans," in International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2014, pp. 528–535.

[15] C. Beekman, S. van Beek, J. Stam, J.-J. Sonke, and P. Remeijer, "Improving predictive CTV segmentation on CT and CBCT for cervical cancer by diffeomorphic registration of a prior," Medical Physics, vol. 49, no. 3, pp. 1701–1711, 2022.

[16] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," Nature methods, vol. 18, no. 2, pp. 203–211, 2021.

[17] J. Malimban, D. Lathouwers, H. Qian, F. Verhaegen, J. Wiedemann, S. Brandenburg, and M. Staring, "Deep learning-based segmentation of the thorax in mouse micro-CT scans," Scientific reports, vol. 12, no. 1, pp. 1–12, 2022.

[18] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE transactions on pattern analysis and machine intelligence, vol. 40, no. 4, pp. 834–848, 2017.

[19] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim, "Elastix: a toolbox for intensity-based medical image registration," IEEE transactions on medical imaging, vol. 29, no. 1, pp. 196–205, 2009.

[20] Z. Liu, X. Liu, B. Xiao, S. Wang, Z. Miao, Y. Sun, and F. Zhang, "Segmentation of organs-at-risk in cervical cancer CT images with a convolutional neural network," Physica Medica, vol. 69, pp. 184–191, 2020.

[21] D. J. Rhee, A. Jhingran, B. Rigaud, T. Netherton, C. E. Cardenas, L. Zhang, S. Vedam, S. Kry, K. K. Brock, W. Shaw et al., "Automatic contouring system for cervical cancer using convolutional neural networks," Medical physics, vol. 47, no. 11, pp. 5648–5658, 2020.

[22] B. Rigaud, B. M. Anderson, H. Y. Zhiqian, M. Gobeli, G. Cazoulat, J. Söderberg, E. Samuelsson, D. Lidberg, C. Ward, N. Taku et al., "Automatic segmentation using deep learning to enable online dose optimization during adaptive radiation therapy of cervical cancer," International Journal of Radiation Oncology* Biology* Physics, vol. 109, no. 4, pp. 1096–1110, 2021.

[23] C. Shelley, L. H. Barraclough, C. L. Nelder, S. Otter, and A. Stewart, "Adaptive radiotherapy in the management of cervical cancer: review of strategies and clinical implementation," Clinical Oncology, vol. 33, no. 9, pp. 579–590, 2021.

[24] A. D. Yock, M. Ahmed, D. Ayala-Peacock, A. B. Chakravarthy, and M. Price, "Initial analysis of the dosimetric benefit and clinical resource cost of CBCT-based online adaptive radiotherapy for patients with cancers of the cervix or rectum," Journal of Applied Clinical Medical Physics, vol. 22, no. 10, pp. 210–221, 2021.

[25] S. Ding, H. Liu, Y. Li, B. Wang, R. Li, B. Liu, Y. Ouyang, D. Wu, and X. Huang, "Assessment of dose accuracy for online MR-guided radiotherapy for cervical carcinoma," Journal of Radiation Research and Applied Sciences, vol. 14, no. 1, pp. 159–170, 2021.

[26] B. Rigaud, A. Simon, M. Gobeli, C. Lafond, J. Leseur, A. Barateau, N. Jaksic, J. Castelli, D. Williaume, P. Haigron et al., "CBCT-guided evolutive library for cervical adaptive IMRT," Medical physics, vol. 45, no. 4, pp. 1379–1390, 2018.

[27] B. Rigaud, A. Simon, M. Gobeli, J. Leseur, L. Duverge, D. Williaume, J. Castelli, C. Lafond, O. Acosta, P. Haigron et al., "Statistical shape model to generate a planning library for cervical adaptive radiotherapy," IEEE transactions on medical imaging, vol. 38, no. 2, pp. 406–416, 2019.

Conclusions and perspectives

Conclusion

Medical image segmentation is crucial for the diagnosis and analysis of diseases, and in the clinical setting, since detailed manual annotation of MR images and CBCT images is difficult and is medical expertise time consuming. In this thesis, we focused on the algorithmic application of image segmentation of the uterine region in the treatment of uterine fibroids and cervical cancer diseases. In particular, we developed two patient-specific deep learning-based segmentation methods of fibroids and surrounding organs from MRI data for preoperative planning of HIFU treatments and an deep-learning-based algorithm for the segmentation of the uterus and the surrounding organs from CBCT followed by an optimal choice of the plan-of-the-day in adaptive radiotherapy. For this, we made the following contributions in this Thesis:

1. HIFUNet has been proposed to address the automatic segmentation of uterine MR images before HIFU treatment. Fully automatic and accurate segmentation of the uterus, uterine fibroids, and spine in the uterine region was performed. This was the first attempt of deep learning method for multi-class segmentation in the uterine region, and according to our evaluation the proposed algorithm was more robust and accurate than the previous traditional image segmentation methods. Unlike other CNN segmentation networks that use small convolutional kernels, the large convolutional kernels with atrous in GCN and DMAC were used in HIFUNet to expand the valid receptive field, which enabled the network to extract the features of the segmentation target in the complex medical image background of the scene. Experimental results indicated that HIFUNet was really effective in extracting uterine fibroids of different sizes and numbers within the framework of preoperative planning of HIFU treatment.
2. PLRNet, a semi-supervised learning-based pseudo-label refinement network, has been proposed to solve the problem of the scarcity of labeled data for deep-learning-based medical image segmentation algorithms. Unlike the HIFUNet, which requires a large amount of labeled data to segment the uterine region, PLRNet only requires a small amount of labeled and unlabeled data for training. First, we used a large convolutional kernel network with wavelet pooling operation to efficiently extract features from MR images even when the number of labeled data is small and the pre-trained pseudo labels are coarse and contain more noise. The use of large convolutional kernels allowed so the effective extraction of abstract features. PLRNet uses DWT and IDWT instead

of downsampling and upsampling operations in CNN to preserve and recover more detailed information, which can maintain object structure and suppresses data noise during network inference. In addition, the semi-supervised method depends on pseudo-labels obtained from unlabeled data and fed back into the training. However, the integration of these pseudo-labels depends on the confidence we have in them. We have defined an adaptive process for defining a confidence threshold. This process allowed getting rid of either a manual definition or after an extensive grid search and also alleviated the threshold selection dilemma in multi-class segmentation tasks with different segmentation difficulty. Finally, based on Mixup, a data enhancement method to solve the overfitting problem caused by the difficulty of labeled data scarcity was proposed. For this, a feature-aligned mixup module in each hidden layer of the network was used to effectively enhance the robustness and generalizability of the network.

3. An automatic segmentation and plan-of-the-day framework for adaptive radiotherapy for the treatment of cervical cancer was proposed. The annotation of CBCT images in adaptive radiotherapy is difficult and time-consuming, and the CBCT segmentation results based on CBCT need to be compared with the preoperative CT images to select the daily radiotherapy plan. In this thesis, the proposed automatic framework quantified the similarity results of CBCT after automatic segmentation with the corresponding organs in the preoperative CT images. First, the segmentation of nnU-Net was used for CBCT images, which is the first attempt to use nnU-Net on CBCT images; after that and after a rigid registration between CT and CBCT, the overlap size was calculated from the CTV in the registered CT images and the CTV in CBCT. A tolerance of 5% was also set to avoid errors due to automatic segmentation. The overlap size was selected from the CTV in CBCT. The CT image with the largest overlap with the CTV of CBCT is selected as the dose reference for radiotherapy on that day. This framework was the first attempt of the automatic selection of the plan-of-the-day in adaptive radiotherapy.

The above three works on image segmentation have been used in applications in the treatment of benign or malignant uterine tumors, including MR and CBCT, fully supervised to semi-supervised segmentation and non-invasive treatment to radiotherapy. These methods have achieved competitive advantages in uterine application and have potential to be applied to other medical image segmentation tasks as well.

Prospects for future work

The segmentation methods presented in this thesis have achieved good results in image segmentation either on preoperative MR for HIFU treatment or CBCT for adaptive radiotherapy, however, there are still some image processing issues in both treatments that are worth exploring

and solving. We summarize the following future work that should be done.

1. Ultrasound segmentation problem. In HIFU treatment, intraoperative ultrasound images are used in the clinic to guide the surgical procedure. However, these images are much less rich in information than MRI images. Therefore, the intraoperative ultrasound images must be merged with the preoperative MR images. Prior to registration, the ultrasound images must be segmented. The segmentation of ultrasound images is a difficult problem to solve because of the speckle nature of transabdominal ultrasound images in HIFU surgery. The most commonly used methods are to perform some preprocessing to attenuate the noise. However, it has been shown that speckle texture can also be used as an effective feature for image segmentation, which contains information about the microstructures of the tissues. One approach is to model the spatial distribution of speckle, such as Rayleigh distribution, Rician distribution, etc., to characterize the texture of ultrasound images and then use the texture analysis results for segmentation. Since the size of speckle depends on its distance from the probe, and for circular probes, the speckle noise has a circular distribution and its orientation depends on its position in the image, the relevant texture analysis methods should use features that are independent of the size and orientation of the speckle as much as possible, which will be an effective way to improve the robustness of ultrasound image segmentation methods. The future plan is to develop scattering networks based on orthogonal moments and invariants to segment ultrasound images in HIFU surgery.
2. Multimodal image registration problem. In adaptive radiotherapy, cross-modal deformation registration of CBCT to CT images is also required in order to calculate the cumulative dose. Due to the poor image quality of CBCT, it has been found in previous work that it is difficult to perform cross-modal image registration directly using a deep-learning-based image registration network. Therefore, the technique of deep learning-based image synthesis could be considered in the future to transform the multimodal registration task into a single-modal registration task. A cycle-consistent generative adversarial neural network unsupervised used to build the style transfer framework to achieve an unsupervised end-to-end registration network. A bi-directional process would be added to the generative adversarial network that would take the source domain image input, perform a game of generators and discriminators, and finally output data under the target domain modality. The aim would be to solve the dependence of the previous method on paired datasets by the cyclic consistency in this bidirectional process. The neural network could realize the conversion generation between CBCT and CT images without changing the image structure, thus simplifying the multimodal registration problem of medical images into a simpler mono-modal registration problem.
3. Collaborative registration methods for multimodal images. Despite the use of a cycle-

consistent generative adversarial neural network to build the required style transfer framework, the network still could not guarantee the structural consistency between the input image and the synthesized image. The output image from the generator may have a certain degree of scaling and distortion, which does not destroy the cycle-consistency of the network and the registration results could appear to be fully valid, but would not maintain the original anatomical morphology of the uterine image. In order to solve the above problems and to address the large spatial resolution gap between CT and CBCT images, we add Modality Independent Neighborhood Descriptor (MIND) to the style transfer framework. MIND uses non-local small block-based self-similarity to define, which relies on the structure of local images rather than intensity values and can better measure the similarity of CBCT to CT. Therefore, these images should first be mapped into a common structural feature space by extracting modality-independent structural features, and then structural consistency should be measured in this feature space. A direct constraint would so be formed on the input and synthetic images in the generative adversarial network to ensure the structural consistency between these two images and could so improve the accuracy of the registration. In addition, spectral normalization could be used to stabilize the training process and avoid problems such as pattern collapse of the network during training.

List of publications

International Journals

1. Zhang, Chen, Huazhong Shu, Guanyu Yang, Faqi Li, Yingang Wen, Qin Zhang, Jean-Louis Dillenseger, and Jean-Louis Coatrieux. "HIFUNet: multi-class segmentation of uterine regions from MR images using global convolutional networks for HIFU surgery planning." IEEE Transactions on Medical Imaging 39, no. 11 (2020): 3309-3320, doi: 10.1109/tmi.2020.2991266.
2. Zhang, Chen, Caroline Lafond, Anaïs Barateau, Julie Leseur, Bastien Rigaud, Diane Barbara Chan Sock Line, Guanyu Yang et al. "Automatic segmentation for plan-of-the-day selection in CBCT-guided adaptive radiation therapy of cervical cancer." Physics in Medicine & Biology 67, no. 24 (2022): 245020, doi: 10.1088/1361-6560/aca5e5.
3. Zhang, Chen, Guanyu Yang, Faqi Li, Yingang Wen, Yuhan Yao, Huazhong Shu, Antoine Simon, Jean-Louis Dillenseger, and Jean-Louis Coatrieux. "CTANet: Confidence-based Threshold Adaption Network for Semi-supervised Segmentation of Uterine Regions from MR Images for HIFU Treatment." IRBM (2023): 100747, doi: 10.1016/j.irbm.2022.100747.

International Conferences

1. Zhang, Chen, Guanyu Yang, Huazhong Shu, Yinuo Liu, Yingang Wen, Qin Zhang and Jean-Louis Dillenseger. "Segmentation of uterus and uterine fibroids in MR images using convolutional neural networks for HIFU surgery planning." In 33rd International Congress and Exhibition of Computer Assisted Radiology and Surgery (CARS 2019), Jun. 18-21, 2019, Rennes.

COLLEGE DOCTORAL BRETAGNE
MATHS, TELECOMS
INFORMATIQUE, SIGNAL
SYSTEMES, ELECTRONIQUE

Université de Rennes

Titre : Méthodes de segmentation d'images basées sur l'apprentissage profond dans le traitement des tumeurs bénignes et malignes de l'utérus

Mot clés : Fibromes utérins, cancer du col de l'utérus, segmentation d'images, apprentissage profond, thérapie assistée par ordinateur.

Résumé : Cette thèse porte sur l'aide à la thérapie des fibromes utérins (tumeurs bénignes mais pouvant être douloureuses et entraîner des problèmes de fertilité) par ultrasons focalisés haute intensité (HIFU) et des cancers du col de l'utérus par radiothérapie adaptative (ART). Dans les deux cas, l'annotation précise des lésions dans la région utérine et des organes à risque environnants est une partie essentielle du diagnostic et de la planification du traitement. Dans cette thèse, nous avons proposé, d'une part deux outils de segmentations automatiques par apprentissage profond de l'utérus, des fibromes et de la colonne vertébrale en IRM préopératoire du traitement HIFU: 1) HIFUNet, un nouveau réseau neuronal convolutionnel entièrement supervisé et 2) PLRNet, une méthode basée sur de l'apprentissage semi-supervisé qui vise à obtenir des résultats de segmentation comparables aux méthodes entièrement supervisées avec seulement une petite quantité de données annotées. D'autre part, nous avons conçu une stratégie de détermination du plan du jour pour l'ART guidée par CBCT pour le cancer du col de l'utérus qui comprend un module de segmentation d'images CBCT basée sur de l'apprentissage profond suivi d'une sélection du plan du jour dans une bibliothèque de plans de traitement.

Title: Deep learning-based image segmentation methods in the treatment of benign and malignant uterine tumor diseases

Keywords: Uterine fibroids, cervical cancer, image segmentation, deep learning, computer-assisted therapy.

Abstract: This thesis deals with the therapy of uterine fibroids (benign tumors that can be painful and cause fertility problems) by high-intensity focused ultrasound (HIFU) and of cervical cancers by adaptive radiotherapy (ART). In both cases, the accurate annotation of lesions in the uterine region and surrounding organs at risk is an essential part of diagnosis and treatment planning. In this thesis, we proposed, on the one hand, two tools for automatic deep learning-based segmentations of the uterus, fibroids and spine in preoperative MRI in HIFU therapy: 1) HIFUNet, a novel fully-supervised convolutional neural network and 2) PLRNet, a method based on semi-supervised learning that aims to achieve segmentation results comparable to fully supervised methods with only a small amount of annotated data. On the other hand, for cervical cancer CBCT-guided ART, we designed an automatic plan-of-the day selection strategy that includes a deep learning-based CBCT image segmentation module followed by a day plan selection from a library of treatment plans.