

Titre: Méthodes d'apprentissage profond 3D en radiothérapie pour la
segmentation d'organes et la prédiction de distributions de dose

Auteur: Samuel Cros
Author:

Date: 2021

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Cros, S. (2021). Méthodes d'apprentissage profond 3D en radiothérapie pour la
segmentation d'organes et la prédiction de distributions de dose [Mémoire de
maîtrise, Polytechnique Montréal]. PolyPublie.
Citation: <https://publications.polymtl.ca/9987/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/9987/>
PolyPublie URL:

**Directeurs de
recherche:** Samuel Kadoury
Advisors:

Programme: Génie informatique
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Méthodes d'apprentissage profond 3D en radiothérapie pour la segmentation
d'organes et la prédiction de distributions de dose**

SAMUEL CROS

Département de génie informatique et génie logiciel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Génie informatique

Décembre 2021

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Méthodes d'apprentissage profond 3D en radiothérapie pour la segmentation
d'organes et la prédiction de distributions de dose**

présenté par **Samuel CROS**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Julien COHEN-ADAD, président

Samuel KADOURY, membre et directeur de recherche

Martin LACHAINE, membre externe

DÉDICACE

À ma famille.

REMERCIEMENTS

Je souhaite d’abord remercier mon directeur de recherche Samuel Kadoury pour m’avoir accueilli dans ses équipes et donné l’opportunité de travailler sur ce projet de maîtrise. Je le remercie pour son exigence et pour m’avoir encouragé à développer mon autonomie en tant que chercheur.

Je veux remercier également Eugene Vorontsov pour ses conseils avisés tout au long de mon parcours et pour son expertise scientifique.

Je remercie les membres du laboratoire MedICAL pour les discussions animées auxquelles nous avons participé et qui ont façonné ma réflexion.

Merci au Dr Nguyen et à Robert Doucet d’avoir accepté de partager leur expertise médicale avec moi.

J’aimerais remercier également le Natural Sciences and Engineering Research Council of Canada (NSERC) pour le soutien financier m’ayant permis de me focaliser pleinement sur mon travail de recherche.

Merci à mes proches, ma famille qui m’ont soutenu malgré la distance.

RÉSUMÉ

Le cancer de la tête et du cou se développe dans une région foisonnante d'organes, de tissus et de structures vitales. La radiothérapie externe est une des modalités privilégiées pour le traitement de ce cancer car les récentes innovations qu'elle a connues permettent d'administrer des doses à des tumeurs difficiles d'accès. Cette modalité nécessite cependant une longue période de planification, notamment pour les étapes de segmentation des structures et d'élaboration des plans de dose. Bien que le développement de méthodes automatiques permette de réduire ces deux étapes de plusieurs dizaines de minutes, la production de traitement pour cette région complexe se compte encore en heures. La récente évolution des modèles d'apprentissage profond et plus particulièrement les réussites des réseaux de neurones convolutifs ont accéléré la recherche de nouvelles méthodes dans ce domaine. Des tâches de segmentation et de régression sont abordées avec des réseaux profonds qui, à l'issue d'un entraînement minutieux, produisent des résultats en quelques secondes seulement.

Dans ce mémoire, nous présentons deux méthodes basées sur les encodeur-décodeurs convolutifs 3D pour aider à la planification des plans de doses. Une première méthode de segmentation automatique d'organes est développée pour une base de données interne de patients atteints du cancer de la tête et du cou. La solution proposée aborde les méthodes de transfert de poids et démontre les bienfaits de l'initialisation des réseaux avec une moyenne de poids. Le score de Dice moyen obtenu sur l'ensemble des organes considérés est de 0.75. Cela semble indiquer que les modèles entraînés pour la segmentation d'organes distincts disposent d'information transférable entre modèles. Une deuxième méthode répondant au besoin d'outils efficaces de génération de distribution de dose est ensuite proposée. La solution présentée est à la fois testée sur une base de données interne comportant des variations réalistes et sur une base de données publique standardisée. Dans les deux cas, l'architecture enrichie d'un objectif mixte incluant des aspects cliniques démontre des performances accrues face à deux architectures de l'état de l'art avec 0.95 (0.98) pour la couverture D95, 1.06 (1.07) pour la dose maximale, 0.10 (0.08) pour l'homogénéité et 0.53 (0.79) pour l'index de conformité sur la base de données interne (publique). Le modèle proposé inclut des convolutions denses et des modules d'attention portés par des couches choisies et offre des capacités de prédiction en temps réel. Les deux méthodes pourraient ainsi être intégrées en clinique pour la radio-oncologie, fournissant respectivement un outil de segmentation automatique et un outil de visualisation de plans de dose.

ABSTRACT

Head and neck cancers develop in a region of abundance and variability of structures. External radiotherapy is one of the most common type of treatment for this kind of cancer and has benefited from recent innovations that both reduces the time of the process and facilitates delivering radiations to intricate areas of tumor growth. However, that treatment modality endures long planning times, notably through the need of accurate structure segmentations and dose plans. Automatic contouring methods have shortened these steps, yet treatment planning remains very time-consuming for such a challenging site. This is in part due to the very high number of iterations needed to produce both clinically desirable and physically deliverable dose plans. The recent deep learning breakthroughs, notably illustrated by the achievements of convolutional neural networks, has fostered research for new methods in the field. Deep neural networks thoroughly trained for segmentation and regression problems even show real-time capabilities.

In this thesis, we present two 3D convolutional encoder-decoder-based methods for radiation therapy planning. A first method for automatic organ segmentation was developed and tested on an in-house dataset of head and neck cancer patients. The proposed solution relies on transfer learning methods and illustrates the superiority of a weight averaged initial state against simple transfer learning. We obtained a Dice score of 0.75 on average on all organs. This seems to indicate that models trained for different organs share several common features and could thus benefit from other predictor’s weights. A second method that addresses dose plan prediction was tested on two distinct datasets: an in-house dataset that reflects realistic variability and a more standardized dataset publicly available. The architecture was trained using a mixed objective that includes clinical knowledge and demonstrates the best performance on both datasets compared against two state of the art models reaching 0.95 (0.98) on D95 coverage, 1.06 (1.07) on max dose value, 0.10 (0.08) on homogeneity and 0.53 (0.79) on conformity index for the in-house (public) dataset. This emphasizes the relevance of domain knowledge to predict accurate dose distributions as it helps reducing the tradeoff between tumor coverage and sparing of organs. The proposed model integrates cheap dense convolutions and attention modules and offers real-time capabilities. The two methods could be easily integrated within the clinical workflow as an automatic segmentation tool and a dose visualization tool respectively.

TABLE DES MATIÈRES

DÉDICACE	iii
REMERCIEMENTS	iv
RÉSUMÉ	v
ABSTRACT	vi
TABLE DES MATIÈRES	vii
LISTE DES TABLEAUX	x
LISTE DES FIGURES	xi
LISTE DES SIGLES ET ABRÉVIATIONS	xiv
CHAPITRE 1 INTRODUCTION	1
1.1 Plan du mémoire	3
CHAPITRE 2 REVUE DE LITTÉRATURE	4
2.1 Radiothérapie des cancers de la tête et du cou	4
2.1.1 Cancer de la tête et du cou	4
2.1.2 Radiothérapie externe	5
2.1.3 Planification du traitement	6
2.2 Élaboration des plans de dose	7
2.2.1 Acquisition des contours	7
2.2.2 Génération de distributions de dose	8
2.2.3 Évaluation des plans	12
2.3 Notions d'apprentissage profond	13
2.3.1 Perceptron multicouche	13
2.3.2 Réseaux de neurones convolutifs	14
2.3.3 Réseaux de type encodeur-décodeur	16
2.3.4 Réseaux de neurones profonds	16
2.3.5 Mécanismes d'attention	19
2.3.6 Méthodes de transfert de poids	20
2.4 Applications de l'apprentissage profond	23

2.4.1	Applications à la segmentation d'organes	23
2.4.2	Applications à la génération de distributions de dose	24
2.5	Mot de synthèse	28
CHAPITRE 3 MÉTHODOLOGIE DU TRAVAIL DE RECHERCHE		29
3.1	Vue d'ensemble de la démarche	29
3.2	Étape de segmentation d'organes 3D	30
3.2.1	Explication de la démarche	30
3.2.2	Validation de la démarche	30
3.3	Étape de génération de distributions de dose 3D	31
3.3.1	Explication de la démarche	31
3.3.2	Validation de la démarche	31
CHAPITRE 4 ARTICLE 1 : MANAGING CLASS IMBALANCE IN MULTI-ORGAN CT SEGMENTATION IN HEAD AND NECK CANCER PATIENTS		33
4.1	Abstract	34
4.2	Introduction	34
4.2.1	Related work	35
4.2.2	Contribution	35
4.3	Materials and Methods	36
4.3.1	Dataset	36
4.3.2	Single Organ Segmentation	37
4.3.3	Learning from Other OAR	38
4.3.4	Network Training and Implementation	38
4.4	Results and discussion	39
4.5	Conclusion	41
4.6	Compliance with Ethical Standards	42
4.7	Acknowledgments	42
CHAPITRE 5 ARTICLE 2 : COMBINING DENSE ELEMENTS WITH ATTENTION MECHANISMS FOR 3D RADIOTHERAPY DOSE PREDICTION ON HEAD AND NECK CANCERS		43
5.1	Abstract	44
5.2	Introduction	45
5.3	Methods and materials	47
5.3.1	Data collection and formatting	47
5.3.2	HDA U-Net architecture	48

5.3.3	Training and validation	50
5.3.4	Dosimetric measurements	52
5.4	Results	52
5.4.1	Clinical in-house dataset	53
5.4.2	OpenKBP challenge public dataset	57
5.5	Discussion	60
5.6	Conclusion	62
5.7	Acknowledgments and Conflict of Interest Statement	62
CHAPITRE 6 DISCUSSION GÉNÉRALE		64
6.1	Autour de la segmentation des structures critiques	64
6.2	Autour de la génération de plans de dose	65
6.2.1	Rencontres avec des professionnels du CHUM	65
6.2.2	Développement d'un pipeline d'apprentissage profond	66
6.2.3	Intégration dans le fonctionnement clinique	68
CHAPITRE 7 CONCLUSION		69
RÉFÉRENCES		70

LISTE DES TABLEAUX

4.1	Dice similarity coefficients from the 12 organs at risk on the test set using the multi-organ segmentation (Multi), the single-organ segmentation (Single), the transfer learning method (TL) and the weight averaging method (WA). Results are shown as the mean \pm standard deviation across three runs.	39
5.1	Coverage and max dose values for each model with ground truth values. Note : * shows that values are significantly different than the proposed HDA U-Net (DVH) results on the test data ($p < 0.05$). The best values among the architectures are shown in bold in each column.	53
5.2	Homogeneity and conformity values for each model with ground truth values. Note : * shows that values are significantly different than the proposed HDA U-Net (DVH) results on the test data (p -values < 0.05). The best values among the architectures are shown in bold in each column.	53
5.3	Coverage and max dose values for each model with ground truth values. Note : * shows that values are significantly different than the proposed HDA U-Net (DVH) results on the test data (p -values < 0.05). The best values among the architectures are shown in bold in each column.	57
5.4	Homogeneity and conformity values for each model with ground truth values. Note : * shows that values are significantly different than the proposed HDA U-Net (DVH) results on the test data (p -values < 0.05). The best values among the architectures are shown in bold in each column.	57
5.5	Parameters, training and prediction time for each model.	63

LISTE DES FIGURES

2.1	Voies aérodigestives supérieures, extrait de [3]	4
2.2	Illustration d'un LINAC, extrait de [5]	5
2.3	Représentation schématique des volumes tumoraux, extrait de [9] . .	7
2.4	Illustration des trois grandes étapes de calcul du processus de Monte Carlo dans la planification du traitement en radiothérapie. S1 et S2 déterminent les trajectoires propre à la source (collimateur LINAC) et définissent le champ de radiation, S3 calcule la distribution de dose adaptée à la géométrie du patient. Adapté de [21]	9
2.5	Représentation schématique d'un Clinac de la compagnie Varian. Image extraite de [29]	11
2.6	Exemple d'histogramme de dose cumulatif, extrait de [35]	13
2.7	Illustration d'un perceptron multicouche. Image tirée de [37]	14
2.8	Illustrations des opérations de convolutions et de 'maxpooling'. Image tirée de [41]	15
2.9	Illustration d'un CNN. Image tirée de [42]	15
2.10	Illustration des fonctions de coût ReLU, Sigmoid et Tanh et leurs dérivées.	17
2.11	Illustration des bocks utilisant des 'short skip connections'.	18
2.12	Architecture du U-Net conçu par Ronneberger et al. dans [52]	19
2.13	Illustration du mécanisme d'attention dans un encodeur-décodeur séquentiel, adaptée de [59]	20
2.14	Illustration des surfaces de fonctions de perte en fonction des poids du réseau. Chaque image correspond à un plan différent. À gauche : plan montrant trois minima locaux séparés par des régions où la fonction de perte est élevée (ligne pointillée). Milieu et droite : plans présentant l'existence de chemins reliant les minima (ligne pleine), la fonction de perte reste faible le long de ces chemins qui sont accessibles par des variations simples dans les poids. Image extraite de [65].	22
2.15	Comparaison des états obtenus pour l'algorithme SGD et la moyenne de poids sur l'erreur de test et lors de l'entraînement. À gauche : W_1 , W_2 , et W_3 représentent les états de trois entraînements distincts, W_{SWA} leur moyenne. Milieu et droite : comparaison des états obtenus par l'algorithme SGD (W_{SGD}) et la moyenne de poids (W_{SWA}) sur l'ensemble de test et lors de l'entraînement. Image extraite de [66]. . .	22

2.16	Illustration d'un autoencodeur convolutif. Image tirée de [67]	23
2.17	Illustration de la phase d'entraînement et de test de l'architecture GAN développée. Image extraite de [90]	25
2.18	Architectures du U-Net standard, du HD U-Net et du DenseNet comparées dans l'étude de [91].	26
3.1	Vue d'ensemble des deux grandes étapes du traitement.	29
4.1	Visualization of reference 3D segmentations from the 12 organs at risk (OAR) in the H&N region, projected in 2D across the axial dimension.	36
4.2	The model architecture is based on the U-Net with an encoder (down-sampling) path that skips features to the decoder (upsampling) path. The <i>shallow</i> model is created using the top three blocks in each path, omitting the blocks in the dashed red; the <i>deep</i> model uses all blocks. Here, "conv" is 3D convolution, "relu" is the ReLU activation function, "BN" is batch normalization, "max pool" is max pooling, "dropout" is a dropout layer, "concat" is the concatenation of skip connections from the downsampling path to the upsampling path, and "Tconv" is transposed convolution for upscaling features. The number of convolution filters is listed in each block.	37
4.3	Sample segmentation results with CT slices of the medullary canal, oral cavity and mandible with corresponding reference segmentations, and predictions for single-organ, weight averaging (WA), and transfer learning (TL).	40
5.1	Schematic of the proposed HDA U-Net architecture that combines dense elements with attention mechanisms.	48
5.2	Schematic of the implemented attention gate using skipped features and upsampled features to produce attention coefficients.	50
5.3	Dose examples overlapped with the input CT for two sample test patients from ground truth, HD U-Net, Attention U-Net, HDA U-Net and proposed HDA DVH U-Net.	54
5.4	Dose mean percent error per structure for each model on the in-house dataset.	55
5.5	DVH curves for a sample test patient with doses from ground truth (solid line), HDA U-Net (dashed line) and proposed HDA U-Net using DVH loss (dotted line).	56

5.6	Dose examples overlapped with the input CT for two sample test patients from ground truth, HD U-Net, Attention U-Net, HDA U-Net and proposed HDA DVH U-Net.	58
5.7	Dose mean percent error per structure for each model on the public dataset.	59
5.8	DVH curves for a sample patient with doses from ground truth (solid line), HDA U-Net (dashed line) and proposed HDA U-Net using DVH loss (dotted line).	59

LISTE DES SIGLES ET ABRÉVIATIONS

AAA	Anisotropic Analytical Algorithm / Algorithme d'analyse anisotropique
ADV	Adversarial / Antagoniste
AG	Attention Gates
CH	Condensed History
CHUM	Centre Hospitalier de l'Université de Montréal
CI	Conformity Index / Index de conformité
CNN	Convolutional Neural Networks / Réseaux de neurones convolutifs
CT	Computed Tomography / Tomodensitométrie
CTV	Clinical Target Volume
Dmax	Dose maximale reçue
Dmean	Dose moyenne reçue
DVH	Dose Volume Histogram / Histogramme de Volume de Dose
D95	PTV Coverage at 95% / Couverture tumorale à 95%
FCN	Fully connected networks
GAN	Generative Adversarial Networks / Réseaux antagonistes génératifs
GPU	Graphics Processing Unit
GTV	Gross Tumor Volume
HD U-Net	Hierarchically Densely connected U-Net
HDA U-Net	Hierarchically Dense Attention U-Net
H&N	Head and Neck
IMRT	Intensity-Modulated Radiation Therapy
IRM	Image par Résonance Magnétiques
KBP	Knowledge-based Planning / Planification basée sur la connaissance
LBTE	Linear Boltzmann Transport Equation
MLP	Multilayer Perceptron / Perceptron multicouche
MSE	Mean Squared Error
NSERC	Natural Sciences and Engineering Research Council of Canada
OARs	Organs at Risk / Organes à risques
PCA	Principal component analysis / Analyse par Composantes Principales
PET	Positron Emission Tomography / Tomographie par émission de positons
PTV	Planning Target Volume
RAM	Random Access Memory
ROI	Region of interest

R-CNN	Region based convolutional neural network
SGD	Stochastic Gradient Descent / Descente de gradient stochastique
SWA	Stochastic Weight Averaging
TVs	Tumor Volumes / Volumes tumoraux
VMAT	Volumetric Modulated Arc Therapy
VMC	Voxel Monte Carlo

CHAPITRE 1 INTRODUCTION

Les cancers de la tête et du cou se développent dans une région présentant une abondance de structures critiques. Les tumeurs, aux localisations variées, doivent être traitées tout en assurant la préservation des tissus sains environnants. Pour cela, une combinaison de plusieurs types de traitement est généralement utilisée, à savoir la chirurgie, la chimiothérapie et la radiothérapie. La radiothérapie externe est la forme de traitement la plus commune et consiste à irradier la tumeur de rayons ionisants localisés à travers la peau. L'administration du traitement est rendue possible par un accélérateur linéaire qui connaît des améliorations successives avec le développement de nouvelles méthodes thérapeutiques. Si celles-ci permettent de moduler et d'administrer plus précisément les rayons au patient, elles nécessitent une longue étape de planification.

La planification du traitement de radiothérapie commence par une étape cruciale de diagnostic. Après obtention des scans, le radio-oncologue effectue l'acquisition des segmentations des structures du patient. Il s'agit de délimiter les organes à risque et les volumes tumoraux de façon précise. Ce processus est exigeant en temps et en expertise puisqu'il mobilise l'attention du professionnel de santé pendant plusieurs dizaines de minutes. À noter également qu'à chaque opération humaine est associé un certain biais : la segmentation manuelle introduit intrinsèquement de la variabilité dans les contours rendant la tâche d'optimisation à venir plus difficile encore. En réponse à cela, des méthodes de contours assistées apparaissent avec la promesse d'un gain en temps et en objectivité. Une grande majorité s'appuie sur des atlas de données regroupant plusieurs centaines de cas traités qui servent d'exemples pour générer de nouvelles segmentations. Ces méthodes réduisent effectivement le temps d'acquisition mais nécessitent une large collection de données.

Une autre étape importante de la planification est l'élaboration des plans de dose. Cette étape introduit la définition de contraintes en fonction de la pathologie et de l'anatomie du patient mais aussi de l'objectif du traitement. Les différents organes possèdent des seuils de résistance variables à ne pas dépasser et les volumes tumoraux doivent être irradiés suffisamment. Il existe donc un important compromis entre l'irradiation optimale des zones cancéreuses et la préservation des organes vitaux. En considérant ces contraintes, le dosimétriste s'appuie sur les méthodes d'optimisation Monte Carlo pour simuler les trajectoires et interactions des rayons qui seront administrés au patient. L'élaboration des plans de dose est ainsi le fruit du travail conjoint d'un médecin oncologue et d'un expert dosimétriste qui doivent itérer jusqu'à l'obtention d'un plan acceptable cliniquement et réalisable physiquement. Ce

processus itératif nécessite une expertise et un temps précieux pour produire des plans de qualité.

L'utilisation de l'apprentissage machine en radiothérapie évolue rapidement avec l'émergence de l'apprentissage profond et de ses nombreuses applications dans les domaines de vision par ordinateur, notamment grâce à l'introduction des réseaux de neurones convolutifs. Ces modèles mettent à profit de multiples opérations de filtrage réparties dans plusieurs couches et pondérées par des poids appris lors de nombreuses itérations. Les architectures développées s'illustrent dans les problèmes de classification et de régression ce qui font d'elles de bons candidats dans le domaine médical pour les tâches de segmentation d'organes et de prédiction de plans de dose. Les méthodes d'apprentissage profond présentent une grande flexibilité dans leur implémentation. Elles bénéficient par exemple de mécanismes d'attention qui regroupent des processus permettant d'améliorer la capacité de régénération des réseaux en donnant plus d'importance à certaines parties de l'image d'entrée. Aussi, ces méthodes parviennent à obtenir de meilleures performances en utilisant un pré-entraînement sur des bases de données riches avant d'être adaptées pour une tâche spécifique. Elles peuvent ainsi être employées sur des bases de données plus réduites.

Cependant, le recours à de tels ajouts ne permet pas toujours d'obtenir un modèle viable pour ces tâches complexes. Toute modification apportée aux réseaux de neurones convolutifs impacte la complexité numérique et temporelle, surtout en présence de larges données 3D en imagerie médicale. Dans le cadre du traitement du cancer par radiothérapie, ces données présentent des caractéristiques très spécifiques selon la pathologie des patients et les architectures ne bénéficient donc pas forcément d'un pré-entraînement sur des données génériques. Ainsi, le but de ce projet de maîtrise est de trouver des méthodes efficaces pour accélérer et uniformiser les étapes les plus exigeantes de la planification du traitement du cancer de la tête et du cou. Il s'agira pour cela de développer des modèles 3D d'apprentissage profond pour 1) segmenter des structures critiques et 2) produire des plans de dose pour des patients atteints du cancer de la tête et du cou. Cela passe par le développement d'un pipeline d'apprentissage profond capable d'exploiter efficacement de la ré-utilisation de données et des modules avancés.

1.1 Plan du mémoire

Le chapitre 2 présente la revue de littérature. Les sujets sont répartis entre radiothérapie des cancers de la tête et cou, élaboration des plans de dose, notions d'apprentissage profond et applications de l'apprentissage profond. Les notions médicales abordées comprennent le cancer de la tête et du cou, le traitement par radiothérapie externe, la planification de ce traitement. L'élaboration des plans de dose comprend l'acquisition des contours, la génération de distributions de dose et l'évaluation des plans. Les notions d'apprentissage profond abordées comprennent le perceptron multicouche, les réseaux de neurones convolutifs, les réseaux de type encodeur-décodeur, les réseaux de neurones profonds, les mécanismes d'attention, les méthodes de transfert de poids. Les applications de l'apprentissage profond étudiées comprennent la segmentation d'organes et la génération de distributions de dose.

Le chapitre 3 présente la méthodologie du travail de recherche et explique les différentes étapes nécessaire au développement et à la validation des solutions proposées pour les deux tâches abordées. Ce chapitre liste aussi les publications intégrées aux chapitres 4 et 5.

Le chapitre 4 présente le premier article portant sur l'élaboration d'une méthode de segmentation d'organes par apprentissage profond. Des réseaux de type encodeur-décodeur convolutifs profonds y sont développés et des méthodes de transfert de poids sont explorées.

Le chapitre 5 présente le deuxième article portant sur l'élaboration d'une méthode de prédiction de distributions de dose par apprentissage profond. Une architecture de type U-Net intégrant des convolutions denses et des modules d'attention est développée et l'ajout d'un objectif clinique mixte est exploré.

Le chapitre 6 porte sur une discussion des méthodes développées. Il présentent les défis rencontrés, les solutions apportées, les limitations restantes et les améliorations futures envisagées.

Le mémoire se conclut au chapitre 7.

CHAPITRE 2 REVUE DE LITTÉRATURE

2.1 Radiothérapie des cancers de la tête et du cou

2.1.1 Cancer de la tête et du cou

Les cancers de la tête et du cou (H&N) réfèrent au développement de tumeurs au niveau de la bouche, du nez, de la gorge, des sinus, du larynx, de la cavité orale et des glandes salivaires. Ils montrent ainsi une grande diversité de localisations et sont entourés de nerfs, de tissus et d'organes assurant des fonctions essentielles. Dans 75% des cas, ils résultent d'une consommation d'alcool ou de tabac et touchent principalement les personnes de 55 à 65 ans, avec une incidence deux fois supérieure chez les hommes [1, 2].

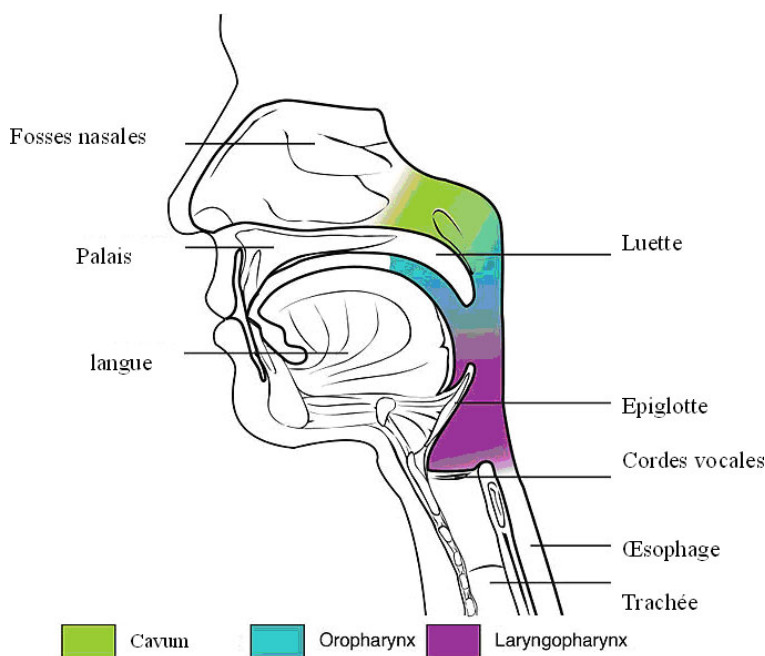


FIGURE 2.1 Voies aérodigestives supérieures, extrait de [3]

Ceux-ci peuvent être soignés s'ils sont détectés suffisamment tôt mais présentent néanmoins un véritable enjeu de préservation des organes vitaux proches. Un des objectifs majeurs est ainsi de protéger au mieux la qualité de vie du patient après traitement en limitant l'incidence de celui-ci sur ses sensations, son apparence, son humeur ou encore sa respiration [4]. Pour cela, il est courant d'avoir recours à une combinaison des types de traitement suivants : chirurgie, chimiothérapie et radiothérapie. Dans le cadre de cette étude, nous nous concentrons sur le traitement par radiothérapie externe.

2.1.2 Radiothérapie externe

La radiothérapie externe est la forme la plus communément utilisée dans le traitement du cancer. Elle consiste à envoyer à travers la peau des radiations localisées sous forme de rayons ionisants à haute intensité. L'accélérateur linéaire (LINAC) représenté en [5] est une des machines permettant l'administration du traitement. Il utilise des technologies micro-ondes pour accélérer des électrons dans un guide d'onde. Ces électrons sont ensuite amenés à entrer en collision avec une plaque de métal pour produire les rayons X à haute intensité destinés à irradier la tumeur. En sortant de la machine, ces rayons passent par plusieurs collimateurs visant à modifier leur forme, pour qu'ils suivent la géométrie ciblée, avant d'être dirigés vers la tumeur. Le patient est couché sur une table de traitement mobile et sa position est guidée par l'utilisation de lasers. La liberté de mouvement de la table de traitement couplée au grand angle de rotation du LINAC permettent un ciblage précis de la tumeur [6, 7].

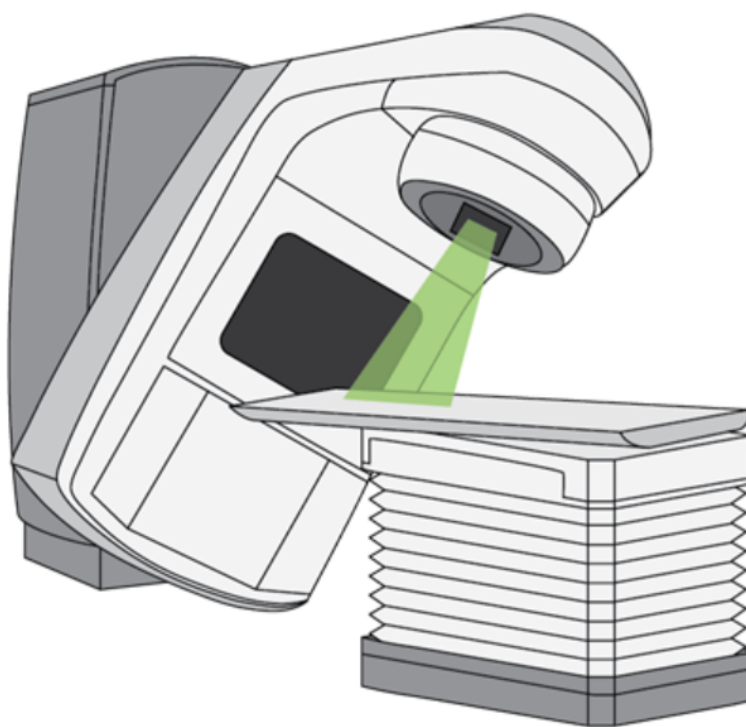


FIGURE 2.2 Illustration d'un LINAC, extrait de [5]

Il existe plusieurs types de radiothérapie externe dont deux traités dans cette étude : la radiothérapie conformationnelle avec modulation d'intensité (IMRT) et l'arthérapie volumétrique modulée (VMAT). L'IMRT est une méthode avancée utilisant les pleines capacités du LI-

NAC pour administrer le traitement de façon modulaire. L'image par résonance magnétique (IRM) ou l'image tomodensitométrique (CT) du patient est utilisée pour créer une image précise de la tumeur en 3D et sa localisation est paramétrée dans la machine. L'IMRT permet de recourir à des rayons d'intensités variables et réduit ainsi le compromis entre traitement des volumes tumoraux et préservation des organes à risque. Cette méthode peut être utilisée pour traiter des tumeurs difficiles d'accès comme dans les zones de la tête et du cou, de la prostate, des poumons ou encore du cerveau. Le traitement VMAT reprend ce fonctionnement en y ajoutant une capacité de rotation à 360 degrés et contribue à réduire le temps de traitement [5].

2.1.3 Planification du traitement

La planification du traitement en radiothérapie démarre après réception du ou des scans du patient. Il est courant d'utiliser plusieurs types de scanners à la fois parmi IRM, CT et la tomographie par émission de positons (PET). Leur superposition donne une meilleure précision au radio-oncologue car certaines structures sont plus faciles à distinguer et donc à détourner sur certains types de scanners [8].

Selon la pathologie du patient, le traitement peut avoir différents objectifs [8]. Il peut être curatif quant il s'agit d'éliminer définitivement les volumes tumoraux en employant des doses de radiation élevées (jusqu'à 70 Gy). Il est palliatif quant il s'agit de ralentir la progression du cancer sans risquer d'endommager les structures adjacentes. Des doses plus faibles sont alors utilisées (jusqu'à 20 à 30 Gy). Enfin, après une opération par chirurgie, un traitement par radiothérapie peut être envisagé pour éliminer les derniers restants de tumeur avec des doses localisées (jusqu'à 60 Gy). Ce choix dépend d'abord de la proximité entre les organes à risques et les volumes tumoraux mais il doit aussi prendre en compte d'autres paramètres comme l'âge et l'état de santé du patient [8]. Chez les patients les plus jeunes, la priorité est souvent de se débarrasser de la tumeur entièrement et de prévenir toute propagation future et toute récurrence. Chez les plus âgés, c'est habituellement le choix du traitement palliatif qui est fait, où la priorité est alors de protéger les structures critiques pour éviter toute complication.

Ces questions émergent généralement de la première discussion avec le patient. Cet échange a pour but d'informer le patient des différentes possibilités de traitement étant donné sa pathologie et permet de fixer des objectifs et des attentes réalisables.

L'élaboration du traitement nécessite plusieurs étapes importantes qui comprennent notamment la segmentation des structures du patient. Cette opération fastidieuse est réalisée à la main par des médecins expérimentés. La zone tumorale est d'abord subdivisée en différents volumes cibles [8,9]. Le volume tumoral brut (GTV) représente la partie visible, palpable de la

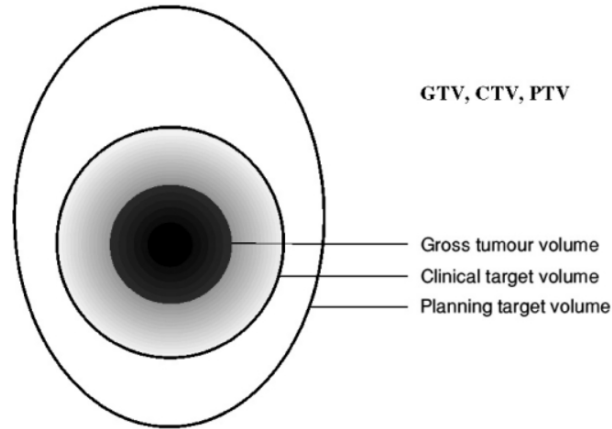


FIGURE 2.3 Représentation schématique des volumes tumoraux, extrait de [9]

tumeur. Le volume tumoral clinique (CTV) étend le volume du GTV pour inclure les parties de la tumeur qui n'apparaissent pas sur les images. Dans le cadre du cancer de la tête et du cou, ce volume tient également compte du risque de propagation du cancer [8]. Le volume de planification (PTV) intègre les incertitudes liées à la planification et à la future administration. Il assure le traitement adéquat du CTV en fonction de l'anatomie et de la position du patient. Enfin, les organes à risques (OAR) à proximité de la tumeur sont détournés. Ces organes présentent des seuils de tolérance aux radiations variables qui formeront une partie des contraintes à respecter dans l'optimisation des plans.

Les contraintes formulées par le médecin oncologue portent plus précisément sur les doses moyennes et les doses maximales reçues par chaque structure. Si de trop fortes radiations administrées aux organes à risque peuvent causer de graves séquelles, une trop faible dose sur les tumeurs peut entraîner la persistance et la propagation du cancer. L'objectif est donc d'administrer la dose prescrite aux volumes tumoraux tout en minimisant l'irradiation des organes et des tissus sains.

2.2 Élaboration des plans de dose

2.2.1 Acquisition des contours

Les deux dernières décennies amènent une prolifération des modalités d'imagerie médicale (CT, IRM, PET) et permettent l'intégration de nouvelles méthodes dans l'étape de planification [10–12]. Le développement des méthodes d'administration IMRT et VMAT en radiothérapie externe permet d'améliorer considérablement la qualité des plans administrés. En revanche, ces nouvelles modalités demandent une attention accrue portée sur l'acquisition

des segmentations des structures critiques. Les médecins sont amenés à se conformer à des seuils d'exigence toujours plus élevés ce qui impacte fortement le temps d'acquisition. La segmentation manuelle introduit aussi un certain nombre de biais par l'opérateur humain. Les méthodes de contours assistées visent ainsi à homogénéiser et accélérer la production des tracés.

Une première catégorie de méthodes utilise la radiodensité de Hounsfield qui donne des valeurs d'intensité variable selon la composition de la matière traversée par un rayon X [13]. L'image de densité obtenue est traitée par des opérations de filtrage tels que le filtrage par motifs ou la classification de composantes connexes [14]. L'image filtrée permet alors de distinguer et classifier les pixels pour former les segmentations souhaitées. Cette méthode utilise exclusivement des images CT et doit généralement s'accompagner d'autres procédés pour donner des contours satisfaisants.

Une autre catégorie de méthodes de contours allie le recalage d'images à l'utilisation d'atlas de données [15–19]. Le recalage d'images considère deux entrées distinctes et cherche à établir des correspondances entre chaque point des deux images. En imagerie médicale, cette méthode sera donc utilisée pour déterminer les déformations à appliquer aux segmentations d'un patient déjà traité, issu de l'atlas, pour obtenir une segmentation acceptable pour un nouveau patient. Un tel procédé nécessite cependant une large collection de données fiables. Un patient présentant une pathologie trop éloignée des données accessibles dans l'atlas demandera une implication plus grande du médecin et donc une perte d'objectivité sur les contours obtenus.

Dans les deux cas, ces méthodes permettent de réduire le temps d'acquisition de plusieurs dizaines de minutes montrant ainsi l'impact des méthodes de contours assistées dans le domaine médical.

2.2.2 Génération de distributions de dose

Méthode Monte Carlo

Les méthodes de Monte Carlo ont pour but de calculer une valeur numérique approchée en utilisant des procédés aléatoires. Cela se traduit généralement par la simulation d'un grand nombre d'évènements dont sont extraits des statistiques utiles comme la moyenne, la variance et l'écart type. Celles-ci permettent d'obtenir une notion de risque sur la tâche simulée au travers d'intervalles de confiance par exemple.

Ces méthodes se distinguent des approches analytiques et numériques classiques qui ont parfois des difficultés à rendre compte de problèmes complexes tels que le calcul d'intégrale

en dimensions élevées, les simulations financières ou encore la physique des particules.

Dans le cas du calcul de dose propre à notre étude, la méthode Monte Carlo permet de simuler la trajectoire de chaque particule (généralement des photons ou des électrons en radiothérapie). Cette simulation englobe de nombreuses notions physiques tel que le type d'interactions, la composition de la matière irradiée, le dépôt d'énergie et dépend autant de la machine utilisée que de la géométrie du patient [20,21]. Malgré le recours à des approximations de calcul, ce procédé requiert un grand nombre d'itérations et demeure particulièrement chronophage.

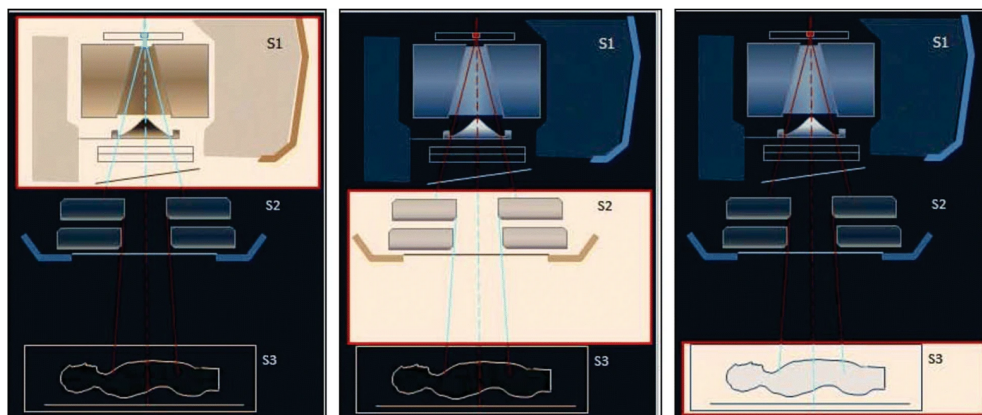


FIGURE 2.4 Illustration des trois grandes étapes de calcul du processus de Monte Carlo dans la planification du traitement en radiothérapie. S1 et S2 déterminent les trajectoires propre à la source (collimateur LINAC) et définissent le champ de radiation, S3 calcule la distribution de dose adaptée à la géométrie du patient. Adapté de [21]

Dans la pratique, l'optimisation des plans peut prendre jusqu'à quatre heures d'après nos échanges avec un dosimétriste [22] du Centre Hospitalier de l'Université de Montréal (CHUM). Leur procédé nécessite l'élaboration de deux à trois plans : un premier plan est généré étant donné la géométrie du patient (CT et zones d'intérêt) et une liste de contraintes acceptables dressée par le médecin oncologue. Ce plan comporte généralement les zones de plus forte dose. Jusqu'à deux plans supplémentaires sont produits en réduisant la dose administrée (au travers de contraintes plus sévères) dans les régions déjà suffisamment traitées par le premier plan. Le premier plan est généralement plus long à produire car les deux autres peuvent être obtenus depuis une étape intermédiaire du calcul du premier.

Méthode Voxel Monte Carlo VMC++

L'utilisation de ces méthodes pour la prédiction de distributions de dose s'appuie sur l'utilisation d'un Condensed History (CH). Ce sont des simulations Monte Carlo du transport

d'électrons dans lequel les effets cumulés de multiples collisions sont approximés pour un pas d'une distance donnée [23]. Ces approximations introduisent des erreurs statistiques et de troncature qui dépendent de la longueur des pas simulés et donc de la précision de la méthode. Les développements apportés par cette nouvelle méthode Voxel Monte Carlo (VMC++) permet des pas de simulation plus longs générant proportionnellement moins d'artefacts. Des techniques de réduction de la variance sont employées pour réduire les incertitudes. Le 'photon splitting' divise artificiellement une particule en deux quand une région d'intérêt est traversée pour donner un meilleur échantillonnage de la région [24]. À l'inverse, la 'russian roulette' associe un poids à chaque particule et combine plusieurs d'entre elles quand leur poids chute en dessous d'une valeur seuil. Cette technique ne diminue pas la variance du problème mais permet de réduire fortement le temps de calcul en faisant disparaître les particules ayant moins d'intérêt [24]. L'effet combiné de ces techniques permet d'améliorer l'efficacité du calcul de dose par un facteur oscillant entre 5 et 10 en moyenne en comparaison avec les méthodes Monte Carlo traditionnelles qui n'emploient pas de tels ajouts.

Algorithme d'analyse anisotropique

Des méthodes de simulations plus précises modélisant la dispersion latérale des faisceaux ont été développées, ils utilisent d'abord des modèles gaussiens [25, 26] avant de s'appuyer sur des fonctions exponentielles [27]. Dans l'algorithme d'analyse anisotropique (AAA) introduit par l'entreprise Varian, la particule source est divisée en trois sources : la source primaire, les photons dispersés (collimateur primaire et filtre d'aplatissement) et les électrons et photons créés par le dispositif de limitation du faisceau [28]. Chaque source est associée à un modèle de Monte Carlo. Le spectre initial des photons juste en dessous de la cible est généré par une simulation Monte Carlo disponible dans la base de données du logiciel Eclipse. Les modifications entraînées par le passage dans le filtre d'aplatissement sont modulées en atténuant le spectre initial. Des filtres de convolutions sont appliqués sur chaque subdivision du faisceau en plus petits éléments ('beamlets'). La dose finale est obtenue en superposant la contribution de chaque 'beamlet'. Les filtres sont calculés en utilisant la somme de six exponentielles pondérées qui dépendent de la profondeur pénétrée. Plusieurs facteurs de corrections portant sur la densité des photons, les convolutions et la profondeur sont appliqués pour rendre la dose plus homogène. La dispersion des photons finale dans le fantôme du patient est calculée grâce à la simulation dans l'eau par le logiciel EGSnrc.

Cette nouvelle méthode montre de meilleures performances dans une variété de cas (incluant des traitements par IMRT) [28] face aux méthodes Monte Carlo traditionnelles et aux premiers développements sous Eclipse.

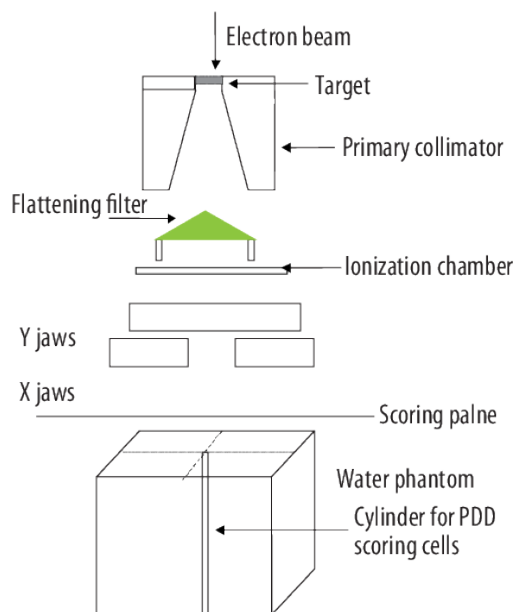


FIGURE 2.5 Représentation schématique d'un Clinac de la compagnie Varian. Image extraite de [29]

Algorithme Acuros XB

L'algorithme Acuros XB appartient à la classe des 'solvers' d'équation de Boltzmann de transport linéaire (LBTE) qui modélise le comportement statistique de phénomènes thermodynamiques. Les 'solvers' LBTE permettent la modélisation précise des interactions lors du transport de radiations dans un média. Les méthodes de modélisation présentent généralement des imprécisions et dans le cas des méthodes Monte Carlo, leur ampleur dépend du niveau d'échantillonnage choisi pour la simulation. Les 'solvers' LBTE introduisent un bruit statistique dans la dose calculée ce qui tend à faire disparaître les incertitudes [30].

L'algorithme Acuros [31] s'appuie sur l'algorithme Attila [32], une méthode de calcul de transport 3D permettant d'obtenir une solution numérique à des problèmes d'ordonnées discrètes comme l'équation de Boltzmann. Les capacités de modélisation d'un tel algorithme permettent de créer efficacement des géométries 3D complexes. L'Acuros XB est l'implémentation de la compagnie Varian sous Eclipse de l'algorithme Acuros. Elle comporte deux éléments principaux : un modèle pour la source du faisceau de photon et un modèle de transport de radiations. Le modèle de la source est commun à tous les algorithmes disponibles sous Eclipse, à savoir Acuros XB, AAA et VMC++ et s'appuie sur des méthodes d'optimisation d'un modèle à sources multiples [33]. Le modèle de transport de radiations utilise une représentation sous forme de grille du problème de l'équation de Boltzmann [31]. Acuros XB offre un outil de simulation du transport de radiations avec de nombreux hyperparamètres

permettant ainsi de modéliser au mieux les tissus humains traversés. Les résultats montrent des prédictions de haute qualité faisant de l’algorithme Acuros XB une alternative efficace aux méthodes Monte Carlo.

2.2.3 Évaluation des plans

Quelle que soit la méthode employée, l’évaluation des plans passe par l’utilisation de nombreuses métriques que l’on peut classer en trois catégories principales.

Les métriques de couverture (D95, D98 et D99) permettent de s’assurer que la vaste majorité (95, 98 et 99%) des zones tumorales sont traitées à la dose prescrite. Il est admis que certaines zones reçoivent une quantité de radiations légèrement au dessus ou au dessous de la prescription (de l’ordre de 2 à 3 Gy). C’est pour cela que ces métriques sont généralement combinées avec des valeurs moyennes (Dmean) et maximales (Dmax). Mathématiquement, D95 peut être obtenu en prenant le cinquième percentile, D98 le deuxième et D99 le premier.

Les métriques d’homogénéité ($H1 = \frac{D2-D98}{D50}$ et $H2 = \frac{D95}{D50}$) visent à encourager la production de plans lisses, montrant peu de variations sur les volumes tumoraux. En pratique, leur implémentation utilise les valeurs de couverture à différents percentiles : plus celles-ci sont proches, plus le plan est jugé comme homogène.

Les métriques de conformité (CI et van’t Riet [34]) assurent que les points de hautes radiations correspondent aux volumes tumoraux. Leur calcul compare la proportion de points du plan recevant la dose prescrite face au nombre de points de la tumeur. Idéalement cette valeur devrait s’approcher de 1 et révèle si l’on irradie trop (>1) ou trop peu (<1) de points. Cette métrique est particulièrement capricieuse, puisque qu’un point recevant 99.99% de la dose prescrite sera jugé comme trop peu irradié. Il est donc important de noter que ces métriques gagnent à être interprétées conjointement.

De plus, l’attirail du médecin oncologue ne s’arrête pas là puisque d’autres outils plus visuels viennent s’ajouter à ces métriques numériques. Les histogrammes de volume de dose (DVH) sont devenus omniprésents, ils sont au coeur de la planification basée sur la connaissance (KBP) qui prend place dans l’élaboration d’un plan de dose de nos jours.

Le DVH est une représentation graphique de la dose reçue par les différentes structures du patient étant donné un plan de radiothérapie 3D. Dans le cadre de notre étude, nous nous intéressons aux DVH cumulatifs qui permettent d’observer pour chaque structure l’évolution de la quantité de radiations reçues en fonction du volume de structure considéré, voir figure 2.6. Ils offrent un moyen rapide d’analyser pour chaque structure la plage de dose administrée, permettant ainsi d’évaluer la satisfaction des contraintes définies lors de la planification.

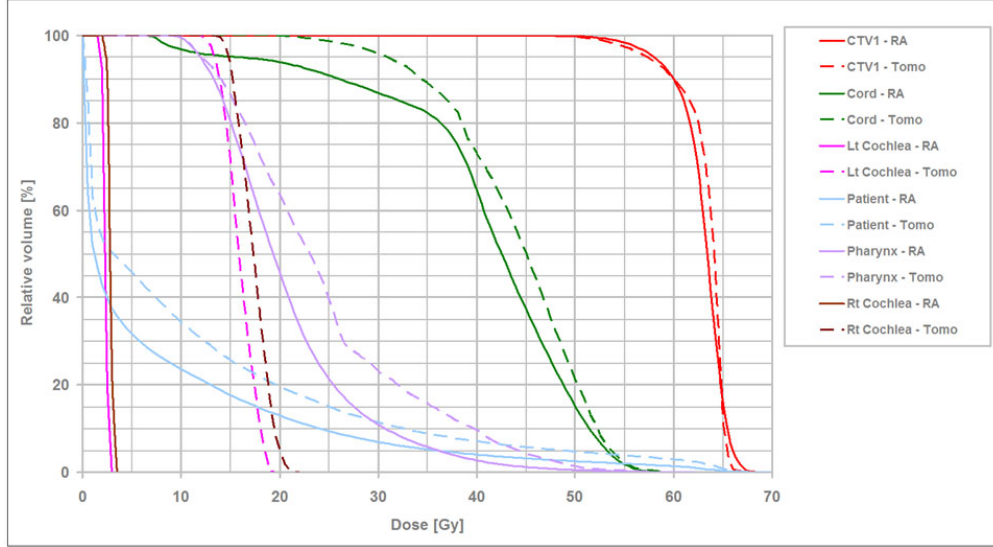


FIGURE 2.6 Exemple d'histogramme de dose cumulatif, extrait de [35]

2.3 Notions d'apprentissage profond

2.3.1 Perceptron multicouche

Le perceptron multicouche (MLP) [36] est un réseau de neurones de base permettant la propagation vers l'avant ('feedforward') de l'information des couches d'entrée vers les couches de sortie, voir figure 2.7. Chaque neurone réalise une opération matricielle entre son vecteur d'entrée x et un vecteur de poids W , il y ajoute un biais b et utilise une fonction d'activation f pour déterminer sa sortie :

$$y = f(W^T x + b) \quad (2.1)$$

Les poids d'un tel réseau peuvent être mis à jour par un processus de rétropropagation qui optimise la configuration des poids pour une tâche donnée de façon itérative. Pour cela une fonction de perte L , adaptée à la tâche effectuée, est utilisée pour calculer l'erreur commise. Le gradient de cette fonction est propagé à chaque neurone en fonction de sa contribution dans le réseau. Si on note L la fonction de perte, λ le taux d'apprentissage, w_i un poids, w'_i la mise à jour du poids, alors l'équation de mise à jour de la descente de gradient peut s'écrire :

$$w'_i = w_i - \lambda \frac{\partial L}{\partial w_i} = w_i + \Delta w_i \quad (2.2)$$

Les réseaux de neurones s'appuient sur cette équation de mise à jour pour optimiser leur configuration de poids à chaque itération durant l'entraînement.

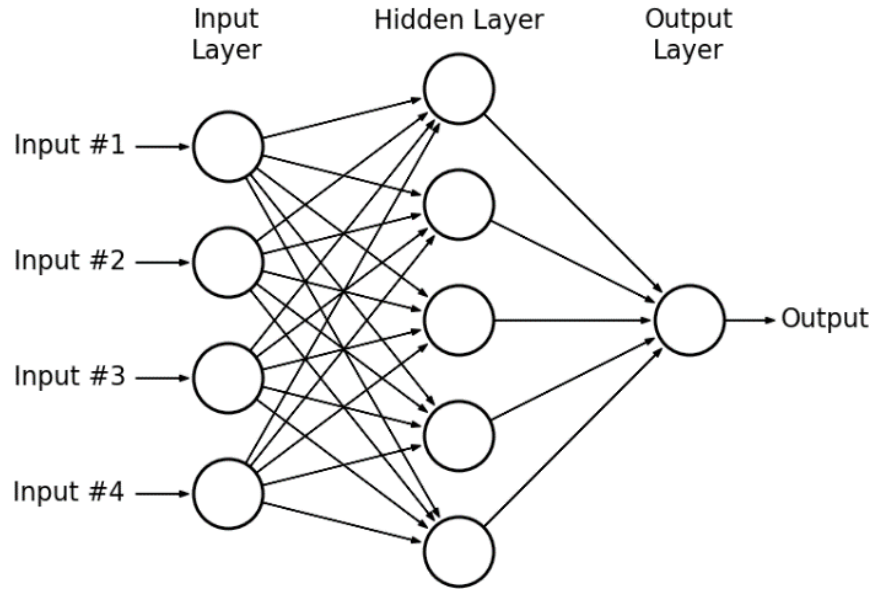


FIGURE 2.7 Illustration d'un perceptron multicouche. Image tirée de [37]

2.3.2 Réseaux de neurones convolutifs

Les réseaux de neurones convolutifs (CNN) sont un type de réseaux de neurones artificiels communément utilisés pour des tâches d'imagerie [38, 39]. Ils étendent les concepts du MLP en remplaçant l'opération matricielle par l'application d'un filtre de convolution illustré à la figure 2.8. Une opération de filtrage intuitive est par exemple la détection des contours (points de changement abrupt d'intensité) dans une image [40]. Cette opération offre une représentation utile et ordonnée d'une caractéristique précise de l'image. Les CNNs explorent cette idée en capturant une multitude d'opérations de filtrage qui sont portées par les poids des filtres de convolution. Ces poids sont répartis dans de nombreuses couches et peuvent être entraînés par un processus de descente de gradient stochastique (SGD).

À chaque convolution, les données sont parcourues localement par un filtre de taille fixe. La sortie filtrée est agrégée par valeur moyenne (AveragePooling) ou par valeur maximale

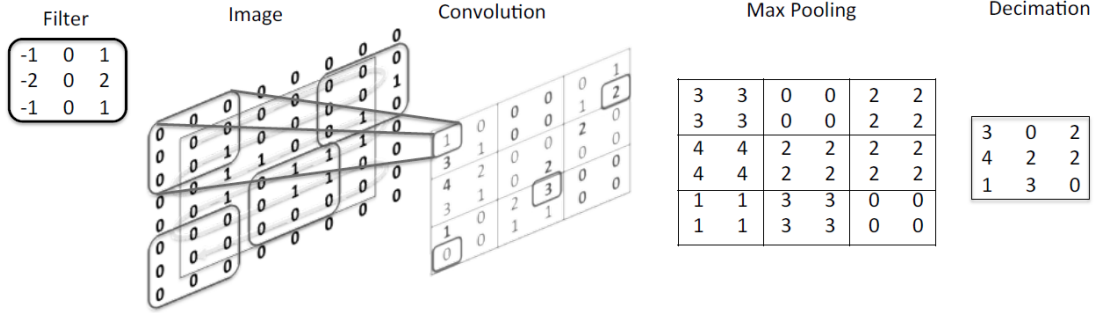


FIGURE 2.8 Illustrations des opérations de convolutions et de 'maxpooling'. Image tirée de [41]

(MaxPooling). Cette agrégation s'accompagne généralement d'un sous-échantillonnage qui réduit la résolution de la couche actuelle. Ce procédé de 'decimation', illustré à la figure 2.8, apporte un degré d'invariance supplémentaire pour de faibles déplacements des éléments de la couche [41]. Ces opérations se répètent jusqu'à l'obtention des 'feature maps' finales qui sont transmises à un MLP pour produire la prédiction.

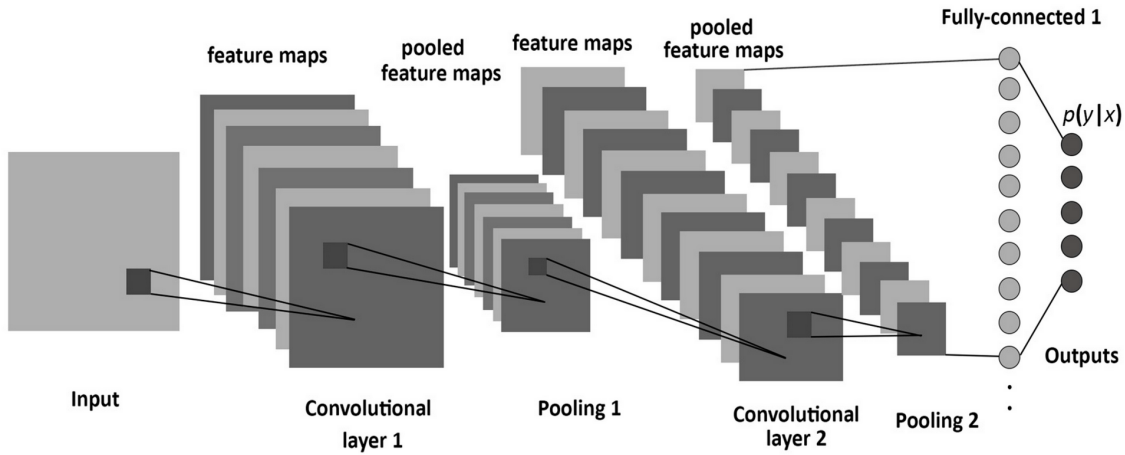


FIGURE 2.9 Illustration d'un CNN. Image tirée de [42]

Ce type de réseau possède une invariance aux translations inhérente au calcul par filtre de convolution qui est souvent renforcée par l'utilisation d'augmentation de données. Enfin, le calcul des opérations de convolution et donc l'entraînement de ces réseaux peuvent être fortement accélérés sous processeur graphique (GPU), ce qui permet d'avoir recours à des modèles considérablement plus profonds [39,43]. Dans la suite, nous noterons FCN les réseaux de neurones pleinement convolutifs [44] introduits par Long et al. pour la segmentation sémantique. Ces derniers ont servi d'architecture de base pour de nombreux problèmes dont

la segmentation d'organes étudiée lors de ce projet.

2.3.3 Réseaux de type encodeur-décodeur

Les réseaux de type encodeur-décodeur se composent d'une partie qui convertit l'information d'entrée x vers un format de plus faible dimension $f(x)$ et d'une partie qui retranscrit l'information compressée vers un format intelligible souhaité y . Un exemple d'application des réseaux de ce type est la traduction d'une langue à une autre [45]. L'objectif dans de tels modèles est de réduire la différence entre la sortie du décodeur et la sortie souhaitée. Cela passe par l'entraînement conjoint de l'encodeur et du décodeur. Le problème d'optimisation peut être formalisé comme suit :

$$\arg \min_{f,g} ||y - g \circ f(x)||^2 \quad (2.3)$$

où f représente l'opération d'encodage, g l'opération de décodage, x l'entrée et y la sortie espérée. À l'issue de l'entraînement, l'encodeur est capable de générer une représentation abstraite $f(x)$ suffisamment riche qui permet au décodeur de produire une sortie $g \circ f(x)$ proche de la référence.

L'information compressée, également appelée état interne ou 'hidden state', est similaire au résultat d'une analyse par composante principale (PCA). Cependant, la capacité de représentation d'un encodeur-décodeur est à la fois plus puissante et plus flexible [41]. En effet, pour une dimension égale de l'état interne, l'erreur de reconstruction est bien souvent plus faible pour ce type de réseau que pour une PCA. Des études ont ainsi prouvé qu'un meilleur arrangement encourageant une plus nette séparation des données pouvait être obtenu par les réseaux encodeur-décodeur [46]. Ces derniers profitent également d'une très bonne intégration dans l'architecture d'un MLP. Ils produisent ainsi des états intermédiaires successifs de plus en plus compressés jusqu'à atteindre un 'bottleneck' qui force la réduction de la dimension vers un format souhaité.

2.3.4 Réseaux de neurones profonds

Problème de disparition du gradient

Un problème récurrent dans l'entraînement des réseaux de neurones profonds est la disparition du gradient. Il survient lorsque la mise à jour Δw_i dans l'équation de descente des gradients devient trop faible. Alors, plus aucune modification significative ne s'opère dans le réseau

et l'entraînement est compromis. Ce problème est intimement lié au choix de la fonction de perte.

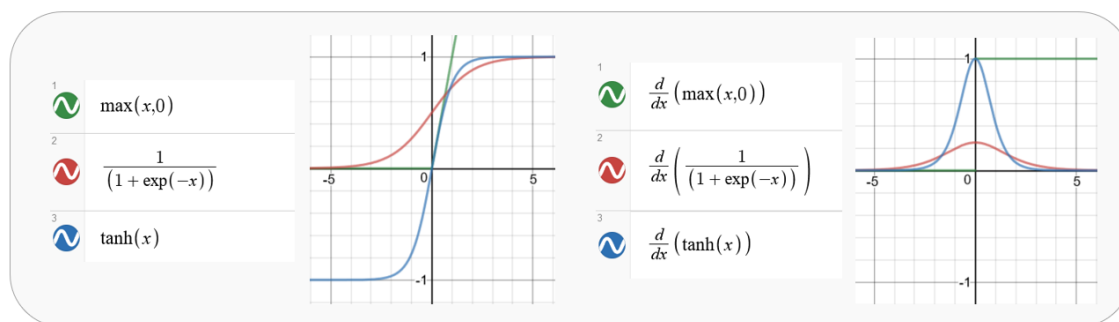


FIGURE 2.10 Illustration des fonctions de coût ReLU, Sigmoid et Tanh et leurs dérivées.

Les fonctions Sigmoid et Tanh, encore souvent utilisées comme fonctions de perte, sont respectivement plafonnées au niveau des valeurs 0, 1 et -1, 1, voir figure 2.10. Cela a pour conséquence de produire des dérivées très proches de 0 et peut faire progressivement disparaître les gradients au cours de l'entraînement. À l'inverse, la dérivée de la fonction ReLU [47] offre deux alternatives : soit elle garde le gradient intact ($\frac{d}{dx} ReLU(x) = 1, x > 0$) soit elle lui affecte une valeur nulle ($\frac{d}{dx} ReLU(x) = 0, x \leq 0$). La fonction ReLU et sa version modifiée leaky ReLU ont ainsi remplacé les fonctions Sigmoid et Tanh dans de nombreux cas, en particulier comme fonction d'activation interne aux réseaux de neurones profonds.

Normalisation par batch

Une autre façon d'endiguer le problème de disparition du gradient est le recours à la normalisation par 'batch' [48]. Il s'agit d'une technique de normalisation des données reposant sur le centrage des 'features' par le calcul de la moyenne et de l'écart type. Dans les CNNs, une couche de normalisation par batch est généralement présente après chaque opération de convolution et permet de garder les données dans une plage de valeurs qui ne saturent pas l'activation [49].

Short skip connections

Les réseaux de type ResNet [50] et DenseNet [51] introduisent le principe de 'short skip connections'. Il s'agit d'une autre solution au problème de disparition du gradient qui permet d'avoir recours à des réseaux encore plus profonds. Les 'short skip connections' sautent une partie des couches du réseau et connectent deux couches en donnant la sortie du premier comme entrée au second, voir figure 2.11. Dans le cas des architectures de type ResNet, un

terme identité est ajouté, sa dérivée étant égale à 1, il permet de maintenir la valeur du gradient dans les couches précédentes. Dans le cas des architectures de type DenseNet, toutes les couches sont connectées entre-elles vers l'avant. Cela assure une réutilisation complète des 'features' des couches antérieures qui possèdent des informations moins abstraites que les couches profondes. En général dans les réseaux convolutifs, on parle de 'short skip connections' pour des liens entre des couches consécutives qui ne modifient pas la dimension de l'entrée.

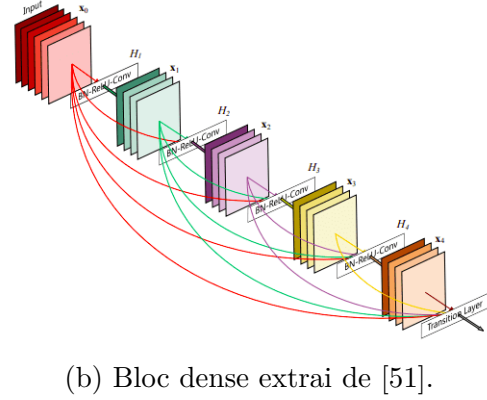
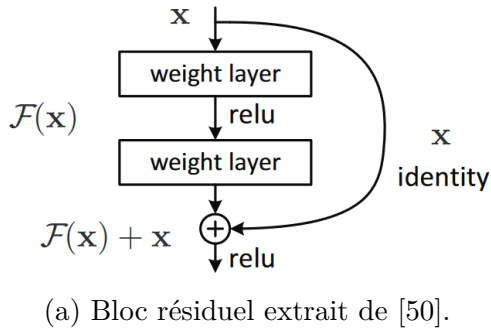


FIGURE 2.11 Illustration des bocks utilisant des 'short skip connections'.

Long skip connections

Les réseaux de type U-Net introduits par Ronneberger et al. [52] comportent un encodeur-décodeur disposant de 'long skip connections'. Celles-ci correspondent à la concaténation de 'features' de l'encodeur dans le décodeur. Elles permettent au réseau d'avoir un rappel sur l'état des données avant compression et améliorent ainsi les capacités de régénération du décodeur. Ces opérations de concaténation apparaissent de façon régulière dans l'architecture à chaque niveau de résolution et connectent les 'features' de l'encodeur à celles du décodeur après 'upsampling', voir figure 2.12. En général dans les réseaux de type encodeur-décodeur, on parle de 'long skip connections' pour des liens qui relient une couche de l'encodeur à une couche du décodeur.

Le U-Net est d'abord destiné à des tâches de segmentation en imagerie médicale 2D mais il est rapidement mis à jour pour des tâches 3D [53, 54] avant d'intégrer des blocs résiduels [55] et des blocs denses [56].

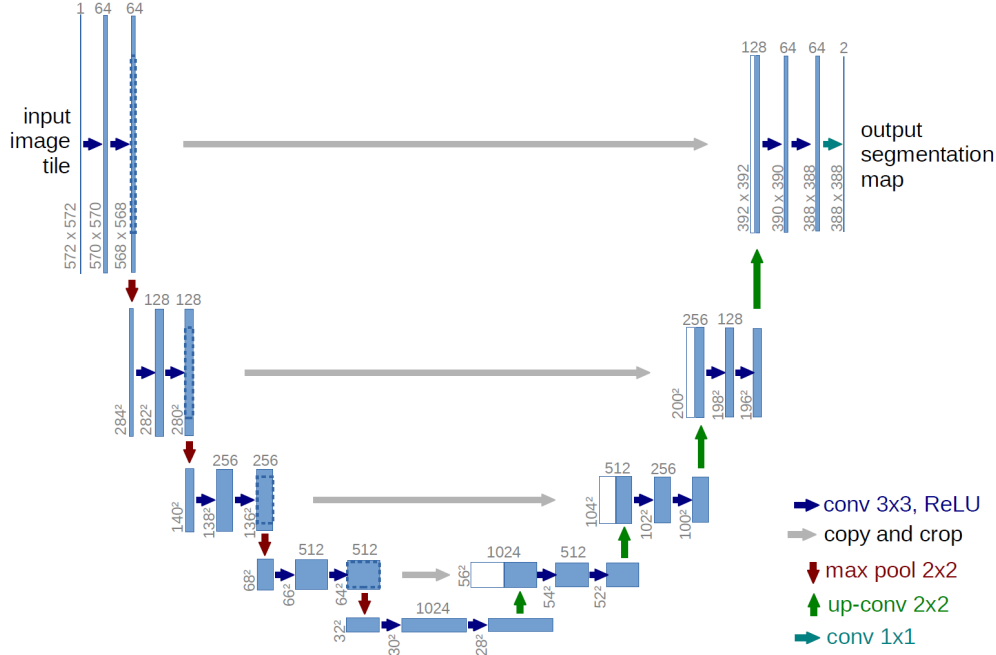


FIGURE 2.12 Architecture du U-Net conçu par Ronneberger et al. dans [52]

2.3.5 Mécanismes d'attention

Les mécanismes d'attention regroupent un ensemble de techniques visant à améliorer l'information contextuelle dans un réseau en donnant plus d'importance à certaines valeurs. Ils seront d'abord développés pour des tâches de traduction de séquence à séquence (Seq2Seq) [57] puis adaptés pour la vision par ordinateur [58], deux tâches dans lesquelles le contexte est crucial.

Introduction dans les modèles de traduction Seq2Seq

Dans une architecture Seq2Seq classique reposant sur de multiples blocs d'encodeurs-décodeurs, voir figure 2.13, un score est attribué à chaque pas de l'encodage et permet de pondérer son importance dans le traitement du décodage. Ce score est obtenu par une fonction ou un modèle d'alignement qui mesure la correspondance entre les couches de l'encodeur et du décodeur. Les deux types de fonctions d'alignement les plus courantes sont les fonctions additives (généralement la sortie d'un réseau 'feed-forward' à une couche) et le produit scalaire. Les scores obtenus sont ensuite rassemblés via Softmax pour former la distribution d'attention qui est multipliée aux 'features' de l'encodeur. La somme des 'features' de chaque encodeur modifiées par le mécanisme d'attention donne le vecteur de contexte qui est généralement concaténé au décodage actuel. Ainsi, à chaque étape, au lieu de disposer seulement du décodage précédent,

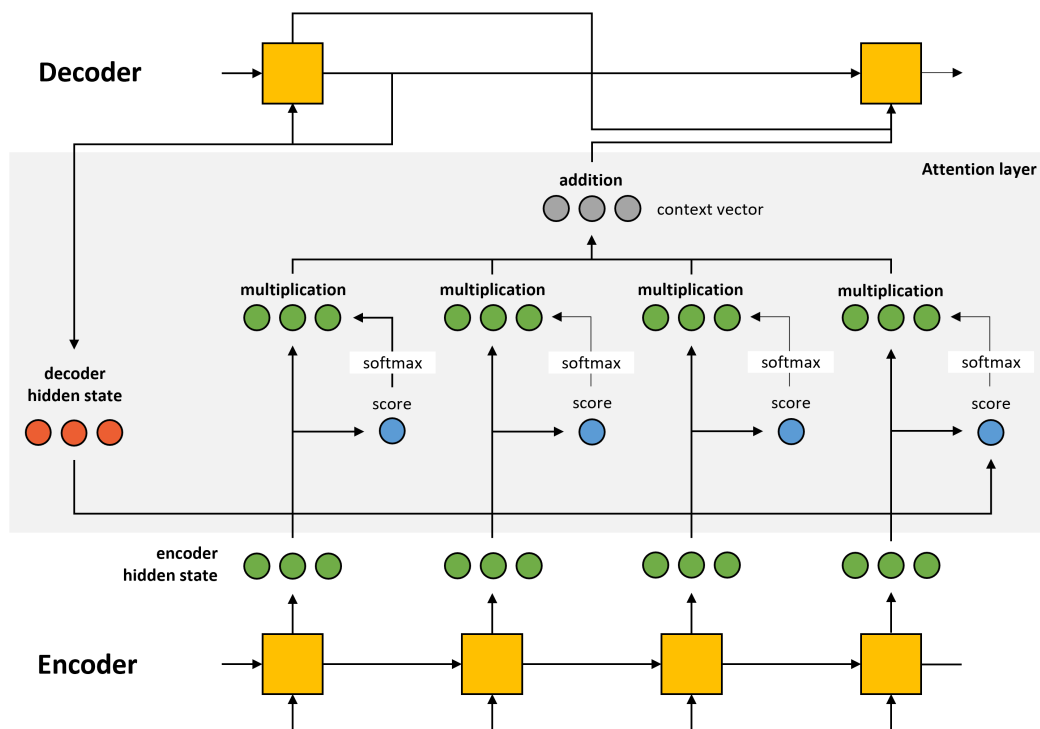


FIGURE 2.13 Illustration du mécanisme d'attention dans un encodeur-décodeur séquentiel, adaptée de [59]

le décodeur obtient le décodage actuel et un vecteur riche d'information contextuelle.

Évolution dans les réseaux convolutifs

Dans le cas du traitement d'images, le processus n'est plus séquentiel mais reprend la même idée en s'intéressant au contexte de l'image. La 'soft' attention utilise des poids d'alignement appris et répartis sur tous les patches de l'image d'entrée. Le modèle reste ainsi différentiable mais les calculs deviennent considérablement plus coûteux pour des images de grandes tailles. À l'inverse, la 'hard' attention considère un patch de l'image à la fois, ce qui réduit la complexité de calcul du modèle à la prédiction mais le rend non-différentiable [60].

2.3.6 Méthodes de transfert de poids

Apprentissage par transfert

Zhuang et al. définissent l'apprentissage par transfert comme l'idée d'améliorer la performance d'un réseau sur une tâche cible en ré-utilisant les connaissances apprises sur des domaines

différents mais apparentés [61]. Dans la pratique, cela permet au réseau d’entamer son entraînement avec une initialisation plus ciblée qu’un démarrage en suivant une loi aléatoire. Il en résulte une progression rapide du modèle vers un état avancé d’optimisation sur une seconde tâche. Cette technique est très populaire pour les problèmes de segmentation et de classification d’images puisqu’elle permet d’entraîner rapidement un modèle sur un nouvel ensemble de données (utilisation massive d’AlexNet [39] pré-entraîné sur ImageNet [62]). Cependant, dans des domaines plus restreints, comme certaines applications en imagerie médicale, le recours à des initialisations génériques n’aide pas toujours. Les modèles pré-entraînés peuvent proposer un état paramétrique trop avancé et trop éloigné de l’optimum à atteindre pour la tâche cible, on parle alors de surconfiguration [63].

Apprentissage ensembliste et moyenne de poids

L’apprentissage ensembliste considère plusieurs réseaux entraînés sur une même tâche et cherche à combiner leurs prédictions respectives. Cette méthode repose sur l’idée que différents réseaux produisent des erreurs sur différents échantillons de données. Selon le type d’application, la combinaison des différentes prédictions peut être faite sous forme de vote, de moyenne pondérée voire même en utilisant un modèle destiné à prédire la sortie finale en fonction des sorties des autres réseaux. Les réseaux rassemblés présentent généralement des architectures variées pour accentuer cet effet. Cependant, en 2017, Huang et al. démontrent les bienfaits de l’apprentissage ensembliste en utilisant une même architecture [64]. Pour cela, ils enregistrent l’état d’un modèle à différents instants pendant un entraînement. Ces états disposent d’une même architecture mais de poids différents et leur combinaison permet d’améliorer les performances sans coût supplémentaire. Dans cette version, le pas de sauvegarde choisi reste grand (20 à 40 epochs) afin de produire des modèles suffisamment distincts dans l’espoir de créer une combinaison finale robuste. Ce choix est remis en question par Garipov et al. en 2018 qui proposent une méthode similaire mais utilisant un pas de sauvegarde très court (2 à 4 epochs) [65]. Leur découverte s’appuie sur l’identification de trajectoires d’optimisation de la fonction de perte reliant les minima locaux dans certains plans, voir figure 2.14. Le long de ces trajectoires, des modèles suffisamment distincts et conservant une fonction de perte faible peuvent être extraits. Leurs sorties sont une fois de plus combinées et obtiennent de meilleures performances. Avec ces améliorations successives, l’entraînement de nombreux modèles n’est plus forcé mais l’enregistrement d’une très grande quantité d’états reste nécessaire.

En 2018, Izmailov et al. introduisent un algorithme reposant sur un ensemble de poids et non plus un ensemble de modèles [66]. Une fois entraîné, un modèle peut être vu comme un point

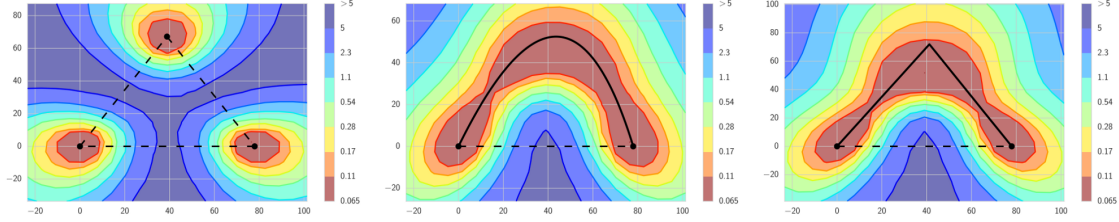


FIGURE 2.14 Illustration des surfaces de fonctions de perte en fonction des poids du réseau. Chaque image correspond à un plan différent. **À gauche** : plan montrant trois minima locaux séparés par des régions où la fonction de perte est élevée (ligne pointillée). **Milieu et droite** : plans présentant l'existence de chemins reliant les minima (ligne pleine), la fonction de perte reste faible le long de ces chemins qui sont accessibles par des variations simples dans les poids. Image extraite de [65].

dans un espace de poids multidimensionnel. Izmailov et al. observent empiriquement que les minima locaux dans cet espace tendent à s'accumuler en bordure des zones où la fonction de perte est faible, voir figure 2.15. Ils démontrent alors que la moyenne de ces poids mène à l'obtention d'une solution plus générale. Dans la pratique, les poids du réseau sont mis à jour de façon itérative en prenant la moyenne pondérée entre un nouvel état et la moyenne de poids calculée au temps précédent. Cette implémentation permet notamment de n'avoir qu'un modèle à entraîner et deux jeux de poids à sauvegarder pendant l'entraînement.

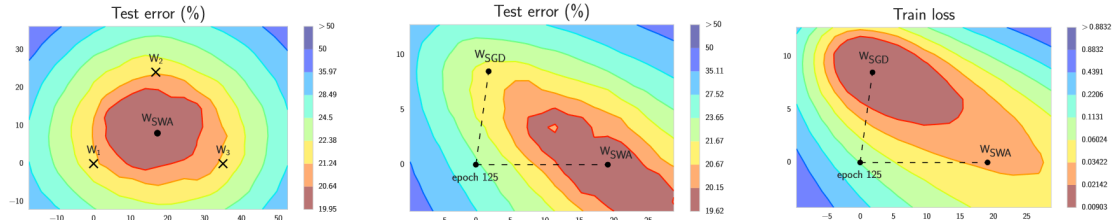


FIGURE 2.15 Comparaison des états obtenus pour l'algorithme SGD et la moyenne de poids sur l'erreur de test et lors de l'entraînement. **À gauche** : W_1 , W_2 , et W_3 représentent les états de trois entraînements distincts, W_{SWA} leur moyenne. **Milieu et droite** : comparaison des états obtenus par l'algorithme SGD (W_{SGD}) et la moyenne de poids (W_{SWA}) sur l'ensemble de test et lors de l'entraînement. Image extraite de [66].

2.4 Applications de l'apprentissage profond

2.4.1 Applications à la segmentation d'organes

Dans cette section, les principaux modèles d'apprentissage profond dédiés à la segmentation d'organes sont présentés.

Réseaux auto-encodeurs

Les auto-encodeurs sont une classe d'encodeur-décodeur dans lesquels les domaines d'entrée et de sortie sont les mêmes. Une fois entraînés, ils permettent de régénérer l'entrée en passant par une représentation de plus faible dimension, prouvant qu'ils sont capables d'en synthétiser les éléments les plus importants. Ils peuvent ainsi être pré-entraînés puis intégrés dans un réseau encodeur-décodeur avant d'être ajustés vers une tâche cible.

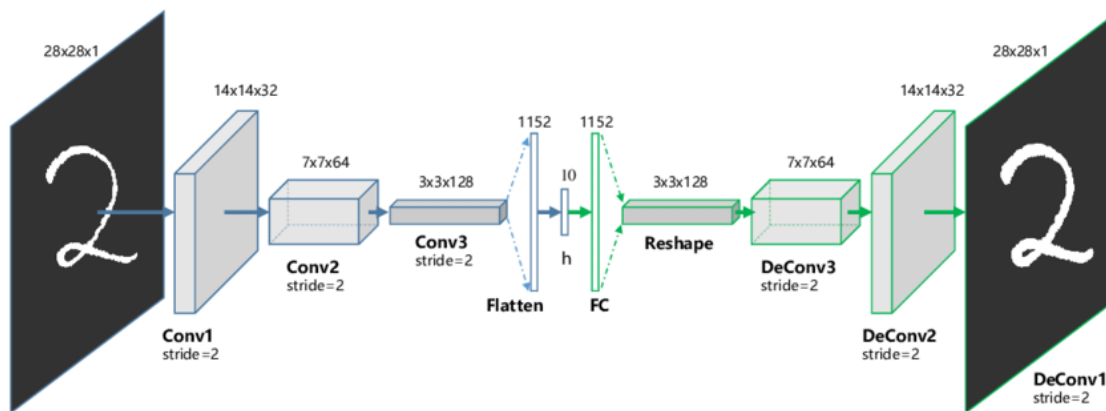


FIGURE 2.16 Illustration d'un autoencodeur convolutif. Image tirée de [67]

Ces réseaux disposent de nombreuses extensions comme les autoencodeurs de débruitage (DAE) [68]. Ce sont des modèles capables de produire une représentation robuste obtenue en obstruant volontairement des parties de l'image d'entrée. La recherche autour de l'ajout de termes de régularisation a mené aux autoencodeurs contractifs (CAE) [69] qui produisent une meilleure initialisation pour pré-entraîner des réseaux sur des tâches de classification. Les autoencodeurs variationnels (VAE) [70] proposent une étude de l'état interne et illustrent les capacités génératives des autoencodeurs. L'intégration des opérations de convolution dans ces réseaux est envisagée [71, 72] et permet encore d'améliorer les performances sur les bases de données de traitement d'images telles que MNIST, CIFAR-10 ou ImageNet [62]. Dans le domaine médical, les modalités d'imagerie privilégiées par ces méthodes sont les images CT et IRM et permettent d'obtenir des segmentations pour des régions diverses telles que

le foie [73], les vertèbres [74], le thorax [75] et les H&N [76], ces derniers restent cependant assez peu représentés. Plusieurs de ces modèles s'appuient sur des méthodes de transfert de données [73, 75] utilisant un pré-entraînement sur des bases de données génériques mais n'explorent pas les dernières innovations des techniques de transfert de poids.

Réseaux de neurones basés sur des régions

La segmentation de multiples organes dans les images médicales est une tâche difficile qui nécessite la détection et la segmentation précise de structures de tailles variées. Une approche en traitement de l'image consiste à séparer cette opération en deux étapes : une première étape de détection générant des régions d'intérêts (ROI) et une deuxième étape de classification des pixels par ROI. L'architecture region-CNN (R-CNN) [77] et ses développements successifs Fast R-CNN [78] et Mask R-CNN [79] permettent l'extraction de régions candidates qui sont ensuite passées à un CNN produisant les 'features' nécessaires à la classification. Dans le domaine médical, les évolutions 3D de la méthode démontrent de meilleures capacités de localisation et des temps de prédiction réduits, elles s'illustrent notamment sur les tâches de segmentation de nodules au poumon et du thorax [80, 81].

Réseaux antagonistes génératifs

Les réseaux antagonistes génératifs introduits par Goodfellow et al. [82] cherchent à parfaire les capacités de génération d'un réseau en l'entraînant conjointement avec un réseau qui discrimine les prédictions synthétiques des sorties de référence. Ces architectures comportent ainsi un réseau générateur et un réseau discriminateur. Leurs applications dans le domaine médical sont multiples, de la génération d'archives manquantes [83] aux tâches de détections cérébrales [84] et d'augmentation de données pour la classification du foie [85]. Dans notre domaine d'étude, la combinaison d'un réseau U-Net comme générateur et d'un FCN comme discriminateur est particulièrement répandue [86, 87]. Aussi, certaines modifications autour de l'objectif d'entraînement des GAN explorent l'intégration de considérations spatiales comme les distances inter-organes [88]. Ces modèles présentent cependant de nombreux hyperparamètres à déterminer ce qui les rend particulièrement difficiles à entraîner.

2.4.2 Applications à la génération de distributions de dose

Dans cette section, les principaux modèles d'apprentissage profond dédiés à la génération de distributions de dose sont présentés. De nombreux réseaux développés initialement pour des tâches de segmentation comme le U-Net et les GAN ont été modifiés pour des tâches de

génération de plans de dose. Ils nécessitent généralement un plus grand nombre de couches et des modules spécifiques pour performer sur cette tâche complexe de régression.

Réseaux antagonistes génératifs

Babier et Mahmood et al. présentent une application des modèles GAN pour la génération de distributions de dose pour des patients atteints de cancers oropharyngiens [89, 90]. Leur générateur, basé sur un réseau U-Net, reçoit une image CT entrée qu'il abstrait dans ses couches profondes pour finalement produire une dose synthétique proche de la sortie espérée, voir figure 2.17. Le discriminateur consiste en de nombreuses couches de convolutions permettant de produire un score binaire étant donné une dose d'entrée. À l'issu de l'entraînement conjoint des deux réseaux, le générateur est séparé du discriminateur et permet la production de doses synthétiques fiables.

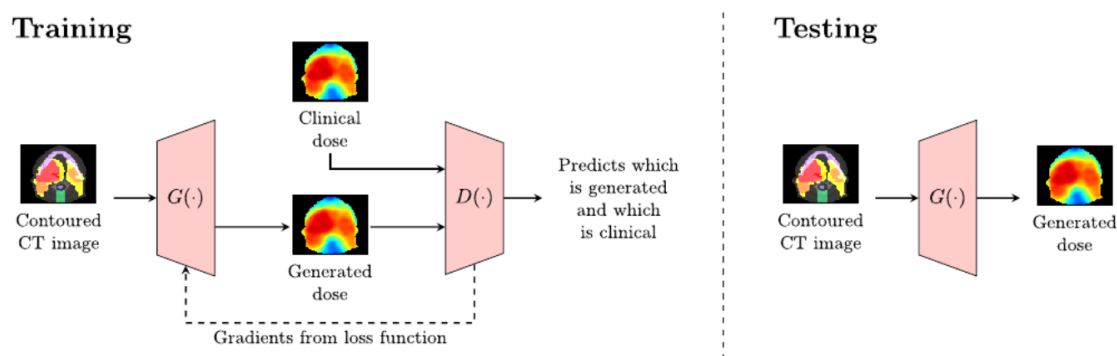


FIGURE 2.17 Illustration de la phase d'entraînement et de test de l'architecture GAN développée. Image extraite de [90]

La version 3D de la solution proposée montre les meilleures performances en satisfaisant plus de 80% des critères cliniques sur les organes à risque et sur les volumes tumoraux de la base de données. La méthode s'appuie cependant sur une optimisation des plans disjointe du module de prédiction. Théoriquement, un modèle combinant ces deux étapes devrait donner de meilleures performances en un temps réduit [90].

Réseaux de type U-Net avancés

Le réseau encodeur-décodeur U-Net introduit sur des tâches de segmentation [52] a servi de base à de multiples évolutions dans les domaines de l'imagerie [55, 56]. Dans le cadre de notre étude de la prédiction de dose pour des patients atteints du cancer de la tête et du

cou, Nguyen et al. proposent une architecture intégrant des modules denses dans un réseau U-Net [91] qu'ils comparent au U-Net [52] et au DenseNet [51], voir figure 2.18.

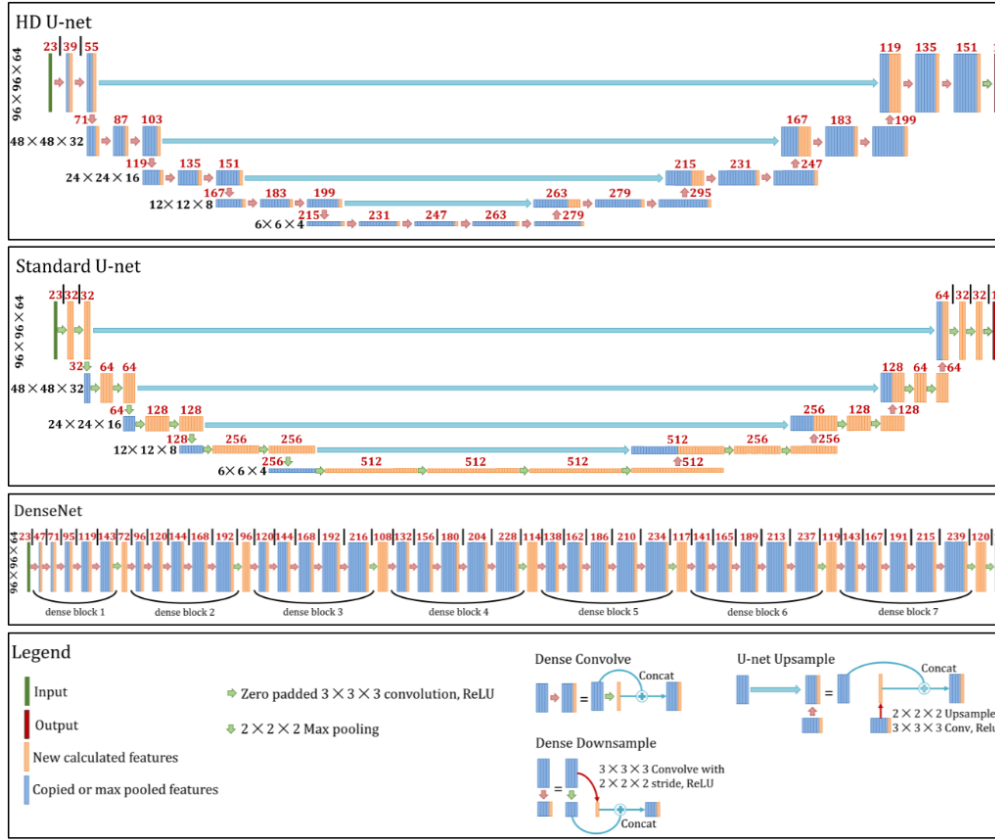


FIGURE 2.18 Architectures du U-Net standard, du HD U-Net et du DenseNet comparées dans l'étude de [91].

Dans leur 'hierarchically densely connected U-net' (HD U-Net), les convolutions denses introduites agissent de façon localisée puisqu'elles ne présentent des liens qu'avec les opérations d'une même résolution. Ainsi, l'architecture obtenue bénéficie d'un transport d'information enrichi grâce aux multiples 'short skip connections' sans pour autant surcharger le réseau. Le HD U-Net proposé montre de plus une grande stabilité avec des courbes de validations faibles et montrant peu de variations. Cela pourrait s'expliquer de par la différence dans le nombre de paramètres des réseaux. Dans une architecture encodeur-décodeur classique orientée pour le traitement d'images, la résolution est successivement divisée par deux. Parallèlement, le nombre de 'features' augmente fortement à chaque changement de résolution. Cette évolution parallèle permet au réseau d'abstraire l'information visuelle dans des 'features' de plus en plus riches. En revanche, le HD U-Net utilise une évolution constante du nombre de 'features' à chaque convolution. Ainsi, ce dernier dispose d'environ 3M de paramètres seulement là où

l’architecture du U-Net, avec le même nombre de résolutions, dispose de 40M de paramètres, soit 13 fois plus.

La méthode proposée offre un outil de visualisation de dose rapide et fiable que les auteurs pensent améliorer en y intégrant les données CT et des contraintes sur les valeurs de dose. La méthode comporte certaines limitations, à savoir l’utilisation accrue de mémoire vive (RAM), la faible taille de l’ensemble de données et l’unique mode d’administration VMAT étudié.

Objectif clinique mixte

L’erreur quadratique moyenne (MSE) est la fonction de perte la plus communément utilisée pour traiter des problèmes de régression dont la génération de distributions de dose fait partie [92]. La MSE dispose d’un bon comportement qui privilégie un entraînement stable et rapide dans la grande majorité des cas. Cependant, cet objectif reste purement mathématique puisqu’il considère uniquement la qualité des prédictions en terme de distance. Les GANs, introduits plus tôt, peuvent déjà être vu comme utilisant un objectif mixte. Le générateur optimisé par une MSE est évalué par un discriminateur, ce qui ajoute un objectif antagoniste. Toutefois, un tel objectif reste difficile à exploiter. Comme souvent en apprentissage profond, une limitation est qu’il est difficile d’en connaître l’influence exacte [90].

Nguyen et al. proposent un objectif clinique mixte [93] constitué à la fois d’une MSE, d’un objectif antagoniste (ADV) et d’un objectif clinique défini par un DVH différentiable. L’idée est d’incorporer dans la fonction de perte utilisée pour l’entraînement des considérations spécifiques au domaine. L’implémentation proposée du DVH différentiable est disponible dans l’article source [93] et est rappelée dans le chapitre 5. L’architecture proposée utilise un U-Net simple comme générateur et un FCN comme discriminateur.

À l’issue d’une ‘ablation study’ sur les différentes combinaisons d’objectifs, la méthode proposée MSE-DVH-ADV obtient les plus faibles erreur de couverture avec 1.65% (D95), 2.14% (D98), 2.43% (D99) ; d’homogénéité avec 0.026 (H2) et de conformité avec 0.038 (CI). Elle comporte néanmoins de fortes limitations. La solution proposée a été testée dans un cadre très spécifique d’administration de dose par IMRT avec 7 faisceaux sur un ensemble de test de 10 patients seulement. Aussi, le temps d’entraînement est fortement allongé avec l’utilisation de réseaux antagonistes puisque le modèle final nécessite 4 jours d’entraînement (malgré le très faible nombre de cas considérés), soit 2.5 fois plus que pour un modèle classique.

2.5 Mot de synthèse

La revue de la littérature a permis de détailler tous les éléments clés de la problématique du projet. Le traitement du cancer de la tête et du cou est une tâche complexe de par la forte variabilité de localisations que la tumeur peut occuper dans cette région dense en structures vitales. Des modalités de traitement avancées sont donc nécessaires mais demandent une étape de planification exigeante. Les méthodes d'apprentissage profond montre une vraie alternative aux méthodes de segmentation d'organes et de prédiction de distributions de dose automatiques. Elles disposent d'entraînements modulaires et de temps de prédiction réduits ce qui les rend viables pour des applications en temps réel. Toutefois, de nombreuses méthodes n'ont pas encore été rigoureusement testées pour la région complexe des H&N et d'autres sont encore limitées par leurs complexités temporelle et numérique. Une avenue de recherche est donc ouverte dans ce domaine et pourrait s'appuyer sur le développement de réseaux légers intégrant des modules avancées à moindre coût.

CHAPITRE 3 MÉTHODOLOGIE DU TRAVAIL DE RECHERCHE

Le chapitre précédent a permis de faire émerger le problème traité et les solutions à explorer. L'objectif de ce projet est alors de trouver des méthodes efficaces pour répondre au besoin d'outils de segmentation automatique et de génération de distributions de dose. La solution de segmentation proposée doit permettre l'acquisition d'un grand nombre de contours variés. La solution de génération de plans doit permettre la prédiction de doses adaptées au traitement difficile des cancers de la tête et du cou. Nous avons pour cela exploré les méthodes d'apprentissage profond et nous nous sommes tournés vers l'architecture U-Net. Cette architecture encodeur-décodeur bénéficie d'un entraînement simple et s'est illustrée aussi bien sur des tâches de classification que de régression. Le U-Net sert également de base à de nombreuses évolutions architecturales, nous laissant la possibilité d'intégrer des considérations spécifiques à nos deux tâches dans le modèle.

Voici la liste complète des publications réalisées dans le cadre de ce projet de maîtrise :

- S. Cros, E. Vorontsov and S. Kadoury, "Managing Class Imbalance in Multi-Organ CT Segmentation in Head and Neck Cancer Patients", 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), publié, Avril 2021
- S. Cros, H. Bouttier, P. F, Nguyen-Tan, E. Vorontsov, and S. Kadoury, "Combining Dense Elements with Attention Mechanisms for 3D Radiotherapy Dose Prediction on Head and Neck Cancers", Medical Physcis, soumis, Novembre 2021

3.1 Vue d'ensemble de la démarche

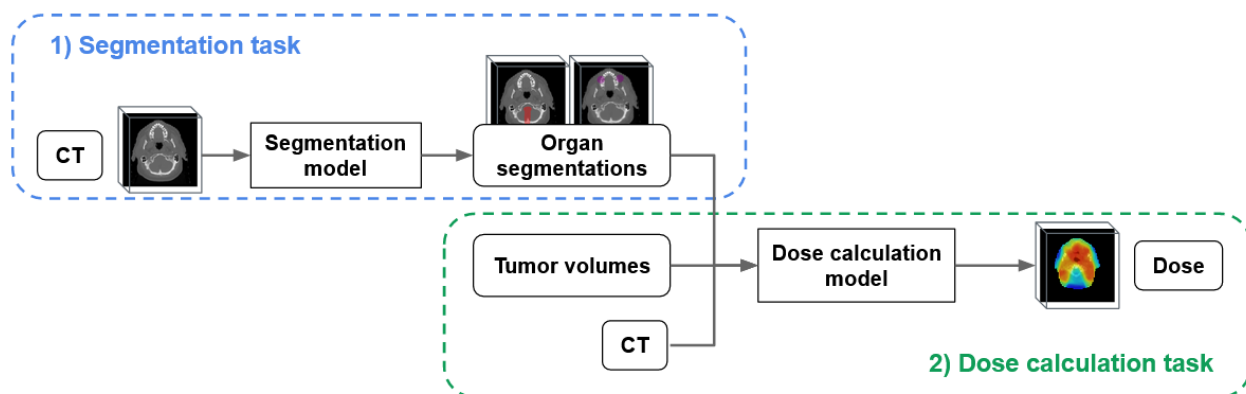


FIGURE 3.1 Vue d'ensemble des deux grandes étapes du traitement.

La planification du traitement du cancer passe par deux grandes étapes. D’abord les images CT doivent être analysées pour détourner les organes à risques et les volumes cibles. Cette étape doit tenir compte de la géométrie variable des structures d’un patient à un autre. Ensuite, les plans doivent être générés et optimisés en se basant sur les images CT annotées. Les plans doivent respecter un certain nombre de critères portant notamment sur les seuils de tolérance des organes et sur la dose prescrite aux volumes cibles.

3.2 Étape de segmentation d’organes 3D

3.2.1 Explication de la démarche

La première étape de ce projet est de concevoir un modèle de segmentation 3D. Une première approche consistait à demander à nos modèles de générer toutes les segmentations à la fois depuis une entrée CT. Cependant, dans cette configuration, seules les structures les plus volumineuses étaient correctement segmentées. En effet, malgré l’utilisation d’une fonction de perte Dice pondérée, afin de donner plus d’importance aux erreurs commises sur les petites structures, nos réseaux ne parvenaient à produire des segmentations que pour le canal médullaire, la cavité orale, la mandibule et la trachée. La solution proposée emploie donc un réseau pour chaque organe à segmenter ce qui offre de bonnes performances pour une majorité d’organes considérés mais pas pour le tronc, la trachée et l’oesophage. Dans le but d’améliorer les performances sur ces structures spécifiques, nous avons exploré des techniques de transfert de poids. Nous avons d’abord opéré des transferts entre deux réseaux entraînés sur des organes proches (en localisation, en taille, en forme) avant de générer une moyenne de poids qui semble bénéficier à une majorité de réseaux.

3.2.2 Validation de la démarche

Une base de données de 200 patients atteints du cancer de la tête et du cou a été utilisée pour valider nos résultats. Les informations disponibles étaient le CT et les segmentations des 12 organes comprenant : le canal médullaire, le canal médullaire extérieur, la cavité orale, la mandibule, l’oesophage, la trachée, le tronc, le tronc extérieur, les parotides, l’oreille interne, les yeux et les glandes sous-maxillaires. Les données ont été acquises avec une résolution de $512 \times 512 \times [100-400]$ et donnés aux réseaux sous forme de patchs d’une résolution de $256 \times 256 \times 64$, valeur maximale possible sur notre GPU. La base de données a été séparée en un ensemble d’entraînement (70%), un ensemble de validation (15%) et un ensemble de test (15%). Tous les résultats présentés ont été obtenus sur l’ensemble de test à l’issu d’entraînements répétés sur trois ‘seeds’ aléatoires différentes. De plus amples précisions sont

disponibles au chapitre 4.

La méthode proposée a ainsi été testée sur un large ensemble de données et obtient un score de Dice bonifié sur la majorité des organes considérés grâce à la moyenne de poids.

3.3 Étape de génération de distributions de dose 3D

3.3.1 Explication de la démarche

La deuxième étape de ce projet est de développer un modèle de génération de distributions de dose 3D. Cette étape a nécessité un gros travail de pré-traitement des données. Tout d’abord, la base de données a dû être régénérée de nombreuses fois dû à des imprécisions lors du transfert numérique des données. Une des conséquences constatées étaient par exemple le mauvais alignement des CT et des plans de dose. Ensuite, plusieurs rencontres avec un radio-oncologue et un expert dosimétriste ont été nécessaires pour comprendre et formater la base de données de façon cohérente. Par exemple, les traitements du CHUM sont généralement constitués de deux à trois plans de dose dont l’administration totale forme le traitement complet. Ainsi, nous avons choisi de sommer ces plans et de ne prédire que la distribution totale de dose. Une fois la base de données filtrée pour ne contenir que des cas classiques de cancers de la tête et du cou (voir chapitre 5 pour plus de précision sur les critères d’inclusion et d’exclusion), nous avons travaillé à intégrer des modules supplémentaires dans l’architecture U-Net pour parfaire ses résultats. Des modules denses ont été intégrés pour améliorer la propagation des données et ont permis l’utilisation de réseaux plus profonds. Des modules d’attention ont été ajoutés pour améliorer les capacités de localisation du modèle pour ce site riche en structures et ont permis une meilleure régénération de l’entrée de la part du modèle. Enfin, pour prendre en compte les connaissances du domaine inhérentes à l’élaboration de telles distributions, plusieurs modifications autour de l’objectif d’entraînement ont été testées (voir chapitre 5 pour plus d’information). La fonction de perte proposée utilise un objectif mixte composé d’une MSE et d’un DVH différentiable et a permis d’améliorer tous les modèles présentés pour un large panel de métriques sur deux datasets distincts.

3.3.2 Validation de la démarche

Une première base de données interne de 150 patients atteints du cancer de la tête et du cou a permis d’élaborer la majeure partie de la solution. Les informations disponibles étaient le CT, les plans de dose et les segmentations comprenant : le canal médullaire, le canal médullaire extérieur, la cavité orale, la mandibule, l’oesophage, la trachée, le tronc, le tronc extérieur, les parotides, l’oreille interne, les yeux, les glandes sous-maxillaires, le nerf optique gauche et

les volumes tumoraux. Les données ont été réduites à une résolution de $128 \times 128 \times [32-96]$ pour pouvoir être chargées sur notre GPU. La base de données a été séparée en un ensemble d'entraînement (70%), un ensemble de validation (15%) et un ensemble de test (15%). Tous les résultats présentés ont été obtenus sur l'ensemble de test à l'issu d'entraînements répétés sur trois 'seeds' aléatoires différentes. Une deuxième base de données publique de 340 patients atteints du cancer de la tête et du cou a permis de valider plus amplement la solution. Les informations disponibles étaient le CT, les plans de dose et les segmentations comprenant : le canal médullaire, le canal médullaire extérieur, la cavité orale, la mandibule, l'oesophage, la trachée, le tronc, le tronc extérieur, les parotides, l'oreille interne, les yeux, les glandes sous-maxillaires, le nerf optique gauche et les volumes tumoraux. Les données ont été acquises à une résolution de $128 \times 128 \times 128$. Les ensembles d'entraînement (200), de validation (40) et de test (100) ont été utilisés tels que définis par le challenge dont les données sont issues. De plus amples précisions sont disponibles au chapitre 5.

La méthode développée a ainsi été testée sur un première ensemble de données avec peu de standardisation, validant ainsi sa pertinence dans un contexte réaliste, puis sur une base de données publique. La méthode a été comparée à deux architectures de l'état de l'art et obtient les meilleures performances sur les deux datasets.

CHAPITRE 4 ARTICLE 1 : MANAGING CLASS IMBALANCE IN MULTI-ORGAN CT SEGMENTATION IN HEAD AND NECK CANCER PATIENTS

Cet article accepté à la conférence ISBI 2021 présente des méthodes de segmentation d'organes à partir de scan CT pour des patients atteints du cancer de la tête et du cou. L'article décrit les difficultés rencontrées, à savoir la grande variété dans les organes à segmenter, et donne les différentes étapes menant à la solution proposée qui utilise des méthodes de transfert de poids.

Auteurs

Samuel Cros¹, Eugene Vorontsov¹, Samuel Kadoury^{1,2}

Affiliations

¹ MedICAL Laboratory, Polytechnique Montréal, Montréal, Canada

² Centre de recherche du CHUM (CrCHUM), Montréal, Canada

4.1 Abstract

Radiotherapy planning of head and neck cancer patients requires an accurate delineation of several organs at risk (OAR) from planning CT images in order to determine a dose plan which reduces toxicity and salvages normal tissue. However training a single deep neural network for multiple organs is highly sensitive to class imbalance and variability in size between several structures within the head and neck region. In this paper, we propose a single-class segmentation model for each OAR in order to handle class imbalance issues during training across output classes (one class per structure), where there exists a severe disparity between 12 OAR. Based on a U-net architecture, we present a transfer learning approach between similar OAR to leverage common learned features, as well as a simple weight averaging strategy to initialize a model as the average of multiple models, each trained on a separate organ. Experiments performed on an internal dataset of 200 H&N cancer patients treated with external beam radiotherapy, show the proposed model presents a significant improvement compared to the baseline multi-organ segmentation model, which attempts to simultaneously train several OAR. The proposed model yields an overall Dice score of 0.75 ± 0.12 , by using both transfer learning across OAR and a weight averaging strategy, indicating that a reasonable segmentation performance can be achieved by leveraging additional data from surrounding structures, limiting the uncertainty in ground-truth annotations.

Keywords

Head and neck cancer, Convolutional neural networks, Organs-at-risk segmentation, Weight averaging, Transfer learning.

4.2 Introduction

Head and neck (H&N) organ segmentation remains a challenging task due to the high variation in the sizes and shapes of the many organs at risk (OAR). However, segmenting these OAR is a fundamental step to properly plan radiotherapy treatments and determine dosimetry plans. In practice, CT scans are delineated by hand, which is a slow process and requires expert knowledge; automating this process could relieve significant burden on experts' time and improve segmentation quality. Recently, fully convolutional neural networks such as the seminal U-Net [52] have dominated medical image segmentation and the semantic segmentation of natural images [94]. Methods based on the U-Net have proven to perform well for multi-organ segmentation [95], but remains sensitive to the variance in shape size between OAR.

In this work, we focus on multi-organ 3D segmentation for OAR in H&N cancer patients

from CT scans. We begin with a model based on the U-Net and propose training techniques for alleviating the effect of class imbalance on segmentation performance. Trained for multi-organ segmentation in H&N data, this model fails to segment fine organs but performs well on large organs due to the prevalence of voxels belonging to the latter classes. We propose single-organ segmentation to overcome this class imbalance and transfer learning or a simple weight-averaging strategy to use more of the available data when training the model.

4.2.1 Related work

While in recent years, 3D image segmentation approaches have shown improved performance over 2D segmentation [96, 97], they typically require working with image patches (sub-volumes) rather than full image volumes due to memory limitations during training. Extending the U-Net from 2D to 3D improves the modelling of spatial dependencies and is more appropriate for volumetric MRI or CT data. However, this may not be sufficient to achieve satisfactory performance on multi-organ segmentation. As shown by [98], some organs may be segmented worse than others when the model learns to be biased toward segmenting the larger OAR. This motivates segmentation of individual organs, instead, focused on a region of interest about the target organ. In [99], Furst *et al.* applied multiple single object segmentation models and focused on merging the outputs back together by selecting the prediction according to size and probability. More recently, Wang *et al.* in [100] proposed a two-stage method where the actual segmentation depends on a bounding box extraction and obtained satisfactory results with H&N OAR. However, a model trained to segment a single organ does not benefit from the additional highly correlated data available for the other organs—something leveraged by multi-organ segmentation models.

4.2.2 Contribution

We propose a U-Net-based multi-organ segmentation model for each of the 12 OAR in the head and neck region, trained with large 3D patches on a subset of 200 H&N cancer patients. In order to improve segmentation results by allowing segmentation models to leverage more of the available data, we propose transfer learning and weight averaging approaches. We implement a mechanism that allows a simple and efficient averaging of the weights of models trained to segment different organs, by attaining a model initialization that, after further training, results in improved performance on several OAR.

4.3 Materials and Methods

Training multi-organ segmentation models is a notoriously challenging problem due to severe class imbalance on this data. While the Dice loss was found previously to be tolerant to such an imbalance, the multi-organ segmentation in H&N regions suffers due to important variations in size between organs, obtaining reasonable performance on some of the larger structures (medullary canal, outer medullary canal, oral cavity, mandible, trachea), but failing with a near-zero Dice score on the finer structures. We trained segmentation models on 3D patches extracted from CT volumes of head and neck cancer patients. Each of twelve organs in the dataset was segmented by one model. In order for each model to leverage more of the available data, we explored transfer learning and weight averaging techniques between models.

4.3.1 Dataset

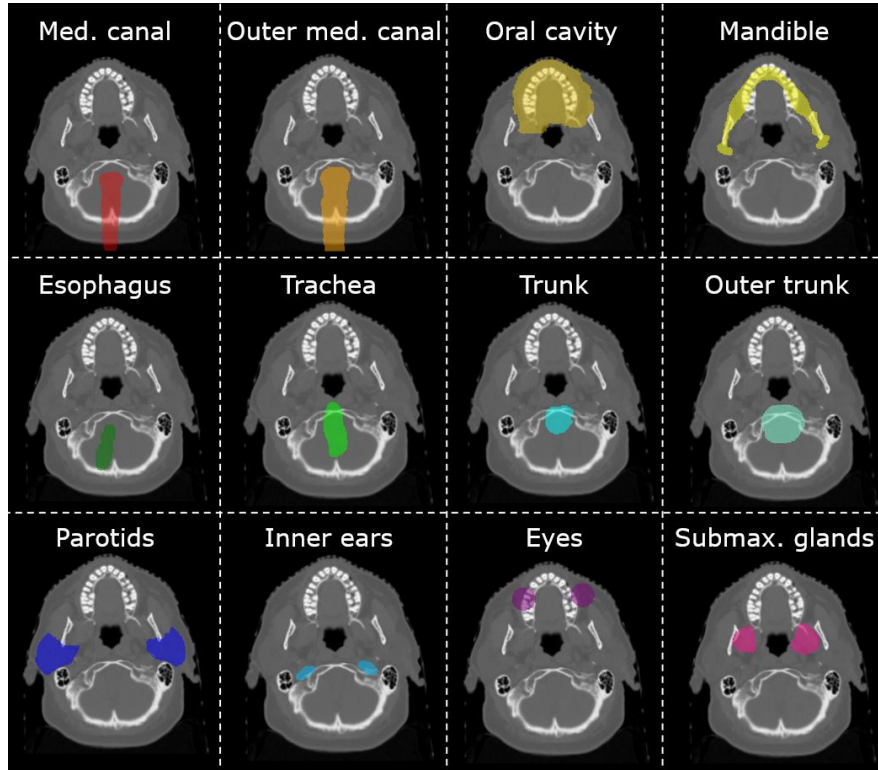


FIGURE 4.1 Visualization of reference 3D segmentations from the 12 organs at risk (OAR) in the H&N region, projected in 2D across the axial dimension.

The input data consists of 200 CT images of head and neck cancer patients provided by our radiation oncology department, with an in-plane resolution of 512×512 , a voxel size

of $1 \times 1 \times 1\text{mm}$ and a number of slices varying between 100 and 400. The available reference segmentations created by experienced radiation oncologists, include the following twelve structures : medullary canal, outer medullary canal, oral cavity, mandible, esophagus, trachea, trunk, outer trunk, parotids, inner ears, eyes and sub-maxillary gland, as shown in Fig. 4.1. There is no prior published segmentation work on this recent dataset.

4.3.2 Single Organ Segmentation

Our network architecture is a 3D extension of the U-Net [52], defined in Fig. 4.2, with the inclusion of transposed convolution and Relu blocks in the upsampling path. For each organ, we selected one of two models according to the performance on the validation set : a *shallow* network working with four feature map resolutions (389k parameters) and a *deep* network working with seven feature map resolutions (40M parameters). As Perone *et al.* presented in a previous work [101], features tend to get noisy in the bottom layers and removing them sometimes leads to better performance.

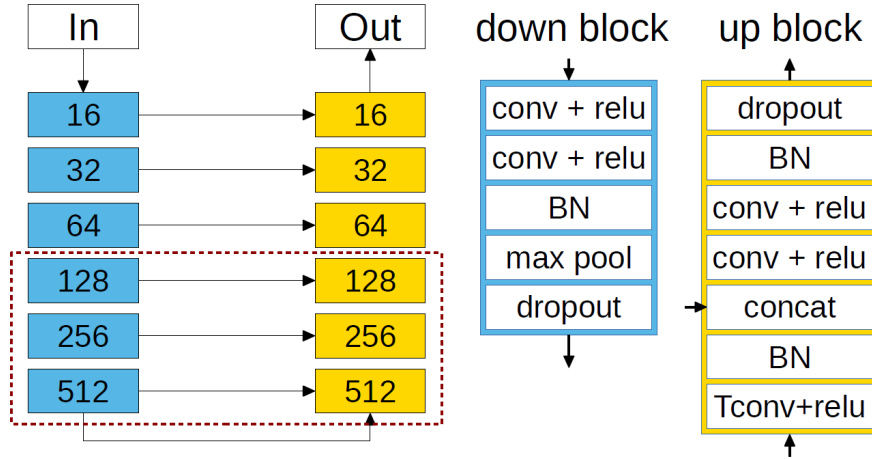


FIGURE 4.2 The model architecture is based on the U-Net with an encoder (downsampling) path that skips features to the decoder (upsampling) path. The *shallow* model is created using the top three blocks in each path, omitting the blocks in the dashed red ; the *deep* model uses all blocks. Here, "conv" is 3D convolution, "relu" is the ReLU activation function, "BN" is batch normalization, "max pool" is max pooling, "dropout" is a dropout layer, "concat" is the concatenation of skip connections from the downsampling path to the upsampling path, and "Tconv" is transposed convolution for upscaling features. The number of convolution filters is listed in each block.

To deal with the imbalance in the number of target organ pixels with respect to the number of background pixels, we used the Dice loss function which is resistant to class imbalance [102].

This function is defined as :

$$L_{dice} = -\frac{2}{|K|} \sum_{k \in K} \frac{\sum_i u_{i,k} v_{i,k}}{\sum_i u_{i,k} + \sum_i v_{i,k}}, \quad (1)$$

where u and v are respectively the predicted segmentation patch and the reference segmentation patch with shape I by K where $i \in I$ are the voxels in the 3D patch and $k \in K$ the different classes.

4.3.3 Learning from Other OAR

Models trained separately for the segmentation of individual organs each benefit only from the data associated with its one organ. In order to allow models to leverage the additional data for improved performance, we explored transfer learning from surrounding organs and model averaging between models trained on different OAR. For this strategy, we selected target organs for segmentation that had poor segmentation performance and performed transfer learning from similar organs for which segmentation results were above 0.65. We achieved transfer learning by first training a model for segmenting one organ and then fine-tuning the entire resulting network on another. This proved very useful when dealing with a limited number of data and resulted in faster training convergence as also demonstrated by Chen *et al.* in [103].

Additionally, we combined models trained to segment different organs using a simple average of the models' weights. Typically, weight averaging refers to averaging the weights across multiple networks throughout training, where it has shown to significantly improve the ability of networks to generalize both on regular image segmentation [66] and on multi-organ segmentation [104]. In our model, we take the simple average of weights across models trained on different OAR once each model is already fully trained. Then, we retrain the averaged model on the organ of interest, seeking to improve performance and generalization on that organ.

4.3.4 Network Training and Implementation

To train the single-class segmentation model, we split the 200 patient cases in 140 training cases, 30 validation cases and 30 test cases. Due to GPU memory constraints, we trained on $256 \times 256 \times 64$ patches of the images. For each structure, we extracted patches within the largest bounding box that contains that structure across all cases. Data was pre-processed only by min/max intensity normalization, mapping intensity values to span $[0, 1]$. We tried

data augmentation with random spline warps, small rotations and shear but it did not improve segmentation results. We used batch normalization throughout the network and set the dropout rate to 0.3 in the *deep* network. We used the Adam optimizer [105] for training with a learning rate of 5×10^{-4} and a batch size of 1. For both models, model states were recorded for each epoch and the state corresponding to the best validation score was kept. The *shallow* model was trained for 200 epochs while the *deep* model was trained for 50 as it was more prone to overfitting. Training hyper-parameters and transfer learning results were selected on the validation subset. For transfer learning and weight averaging, fine-tuning on the target organ was done with the RMSprop optimizer [106] with a learning rate of 10^{-4} and a batch size of 1.

4.4 Results and discussion

	Multi	Single	TL	WA
Med. canal	0.85 ± 0.01	0.84 ± 0.01	0.85 ± 0.05	0.87 ± 0.05
Outer med. canal	0.87 ± 0.02	0.86 ± 0.02	0.87 ± 0.02	0.86 ± 0.00
Oral cavity	0.70 ± 0.03	0.66 ± 0.01	0.51 ± 0.11	0.66 ± 0.02
Mandible	0.89 ± 0.01	0.86 ± 0.01	0.90 ± 0.01	0.90 ± 0.02
Esophagus	≈ 0	0.54 ± 0.01	0.54 ± 0.07	0.60 ± 0.03
Trachea	0.67 ± 0.03	0.61 ± 0.07	0.62 ± 0.06	0.67 ± 0.02
Trunk	≈ 0	0.63 ± 0.09	0.68 ± 0.03	0.67 ± 0.08
Outer trunk	≈ 0	0.67 ± 0.09	0.67 ± 0.06	0.70 ± 0.08
Parotids	≈ 0	0.74 ± 0.04	0.74 ± 0.04	0.76 ± 0.01
Inner ears	≈ 0	0.67 ± 0.03	0.68 ± 0.04	0.71 ± 0.04
Eyes	≈ 0	0.74 ± 0.09	0.75 ± 0.08	0.77 ± 0.07
Submax glands	≈ 0	0.70 ± 0.01	0.73 ± 0.06	0.74 ± 0.06
Overall	0.33 ± 0.04	0.71 ± 0.08	0.72 ± 0.09	0.75 ± 0.12

TABLEAU 4.1 Dice similarity coefficients from the 12 organs at risk on the test set using the multi-organ segmentation (Multi), the single-organ segmentation (Single), the transfer learning method (TL) and the weight averaging method (WA). Results are shown as the mean \pm standard deviation across three runs.

The first two columns of Table 4.1 compare single-organ segmentation with baseline performance using multi-organ segmentation, with the latter attempting to learn separate outputs from the segmentation model for each organ, along with a separate and independent Dice objective function for each OAR; all losses were simply averaged together. The higher complexity of the multi-organ segmentation task yielded low performance for several OAR, namely esophagus, parotids, trunk, outer trunk, ears, eyes and sub-maxillary glands. We tried a class-weighted objective (Dice and cross-entropy) with weights that were inversely

proportional to the frequency of the class in the dataset to help mitigate that problem but it resulted in lower Dice scores.

Models trained on a single organ were then improved by using the annotated data from surrounding OAR using the proposed approach, as these are likely to share several image features. To achieve this, we explored transfer learning where we initialized the model with the weights trained on a similar organ. For example, we selected inner and outer parts of an organ, or more arbitrarily, organs with a comparable shape, volume or position within the CT. This essentially provides the model with additional training data. Transfer learning results are shown in Table 4.1. The models trained to segment the medullary canal, the trunk, and the esophagus were respectively initialized with the weights of models trained on the outer medullary canal, the outer trunk and the trachea. We obtained a noticeable improvement in segmentation Dice score for trunk (0.63 to 0.68) but not medullary canal (0.84 to 0.85) or esophagus (0.54 to 0.54). This shows that transfer learning may be a reasonable approach for improving single organ segmentation performance but the improvement is inconsistent.

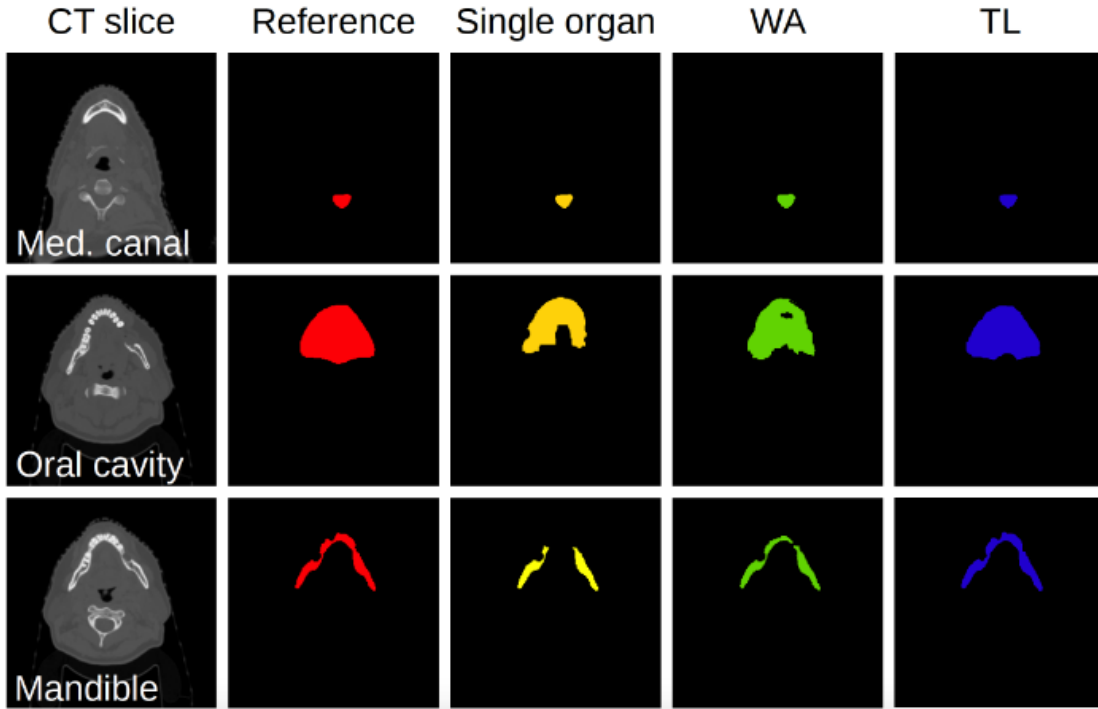


FIGURE 4.3 Sample segmentation results with CT slices of the medullary canal, oral cavity and mandible with corresponding reference segmentations, and predictions for single-organ, weight averaging (WA), and transfer learning (TL).

The following experiment therefore explored a weight averaging strategy. Specifically, we initialized the model by averaging the weights of multiple predictors trained on other organs.

This informs the model about the remainder of the dataset and helps it to find a more general solution. We first created a weight averaged model from segmentation predictors for the outer medullary canal, oral cavity, mandible, esophagus, trachea, outer trunk and parotids, then trained it on each of the six target organs as shown in Table 4.1. We found a noticeable improvement in segmentation Dice score for the esophagus (0.54 to 0.60), mandible (0.86 to 0.90), and trachea (0.61 to 0.67) but not the outer medullary canal (0.86 to 0.86). Results show that weight averaging may outperform transfer learning. Averaging weights across previously trained models is simple but not ideal as it could discard much of the information learned by those models; however, since this method yields improved segmentation performance, this suggests that the weights learned by models trained on different organs may be both similar and arranged in a similar order. Example segmentation outputs from single-organ segmentation are shown in Figure 4.3.

One should note that ground-truth segmentations obtained from expert annotators, suffer as well from a high variability across cases since the data was aggregated from several expert annotators. This inevitably affects the convergence and evaluation of segmentation model performance.

4.5 Conclusion

We address the problem of class imbalance and multi-organ segmentation of head and neck structures in CT scans used for radiotherapy planning by simultaneously training several single-organ segmentation models with weight sharing as opposed to a single multi-organ segmentation model. To overcome GPU memory constraints, we trained the models on large 3D patches and explored strategies for improving segmentation performance by allowing single-organ segmentation models to leverage data from surrounding organs. Results show that transfer learning from similar organs (based on shape and features) yields an improvement in segmentation performance, using a simple averaging of the weights across models trained from different organs. This leads to a reliable and possibly improved initialization for a multi-organ segmentation model. Results demonstrate that these approaches may be a promising direction to pursue for medical image segmentation when dealing with limited data or class imbalance. Further work on transfer learning will involve a learned weighted average of model outputs, including the model trained on the target organ and all the models transferred to this organ from others. We expect this should boost performance and this will remove the arbitrary selection of source organs for transfer learning. Similarly, further work on weight averaging will explore a learned combination of weights in each layer. This is important because features in a layer may be ordered differently in different models or may

not be comparable; it is then necessary to mix and select useful features. We also plan on expanding the experiments to more data, possibly doubling the number of patient cases. In addition, we plan to explore an improved patch sampling strategy and full-volume prediction with post-processing by selecting the largest connected component in the prediction.

4.6 Compliance with Ethical Standards

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the local Institutional Review Board.

4.7 Acknowledgments

This research has been funded in part by the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors have no relevant financial or non-financial interests to disclose.

CHAPITRE 5 ARTICLE 2 : COMBINING DENSE ELEMENTS WITH ATTENTION MECHANISMS FOR 3D RADIOTHERAPY DOSE PREDICTION ON HEAD AND NECK CANCERS

Cet article soumis au journal Medical Physics présente une méthode de génération de distributions de dose pour des patients atteints du cancer de la tête et du cou. L'article décrit les modules d'apprentissage profond intégrés dans l'architecture finale et l'objectif mixte implémenté.

Auteurs

Samuel Cros¹, Hugo Bouttier¹, Phuc-Felix Nguyen Tan², Eugene Vorontsov¹, Samuel Kadoury^{1,3}

Affiliations

¹ MedICAL Laboratory, Polytechnique Montréal, Montréal, Canada

² Centre Hospitalier de l'Université de Montréal (CHUM), Montréal, Canada

³ Centre de recherche du CHUM (CrCHUM), Montréal, Canada

5.1 Abstract

Purpose : External radiation therapy planning is a highly complex and tedious process as it involves treating large target volumes, prescribing several levels of doses, as well as avoiding irradiating critical structures such as organs at risk close to the tumor target. This requires highly trained dosimetrists and physicists to generate a personalized plan, and adapting it as treatment evolves, thus affecting the overall tumor control and patient outcomes. Our aim is to achieve accurate dose predictions for head and neck (H&N) cancer patients on a challenging in-house dataset that reflects realistic variability and to further compare and validate the method on a public dataset.

Methods : We propose a 3D deep neural network that combines a hierarchically dense architecture with an attention U-Net (HDA U-Net). We investigate a domain knowledge objective, incorporating a weighted mean squared error (MSE) with a dose volume histogram (DVH) loss function. The proposed HDA U-Net using the MSE-DVH loss function is compared with two state of the art U-Net variants on two radiotherapy datasets of H&N cases. These include reference dose plans, computed tomography (CT) information, organs at risk (OARs) and planning target volume (PTV) delineations. All models were evaluated using coverage, homogeneity and conformity metrics as well as mean dose error and DVH curves.

Results : Overall, the proposed architecture outperformed the comparative state of the art methods, reaching 0.95 (0.98) on D95 coverage, 1.06 (1.07) on max dose value, 0.10 (0.08) on homogeneity, 0.53 (0.79) on conformity index and attaining the lowest mean dose error on PTVs of 1.7% (1.4%) for the in-house (public) dataset. The improvements are statistically significant ($p < 0.05$) for the homogeneity and max dose value compared with the closest baseline. All models offer real-time prediction, measured between 0.43s and 0.88s per volume.

Conclusion : The proposed method achieved top performance on both realistic in-house data and public data compared to prior state of the art methods. The use of the DVH objective for training showed consistent improvements to the baselines on most metrics, supporting its added benefit in H&N cancer cases. The quick prediction time of the proposed method allows for real-time applications, providing physicians a method to generate an objective end goal for the dosimetrist to use as reference for planning. This could considerably reduce the number of iterations between the two professionals thus reducing the overall treatment planning time.

Keywords : Convolutional neural networks, dense architectures, attention mechanisms, radiation dose prediction methods, knowledge-based planning.

5.2 Introduction

Head and neck cancers develop in regions with several critical structures demonstrating a high degree of appearance variability. The head and neck region is a very challenging site due to the proximity of organs at risk to cancers, the presence of intersecting structures, and the need for variable dose levels. Successive technology leaps in external radiation therapy have improved the manner in which this sensitive region is treated. First through Intensity-Modulated Radiation Therapy [107–110] (IMRT) that uses varying beam intensities to better reach tumors in intricate areas. Subsequently with Volumetric Modulated Arc Therapy [111–115] (VMAT) that dynamically administers the dose in a spiral pattern around the patient, allowing for a shorter treatment time. Both methods rely on imaging modalities (CT, MRI) for a personalized planning that largely improves dose delivery and tumor targeting. However, the rising complexity in delivery method leads to a significant increase in time to generate dosimetric plans. The standard process of knowledge-based planning (KBP) involves a time-consuming process between an oncologist and a dosimetrist that requires iterating between clinically desirable and physically deliverable dose plans. This iterative work requires tremendous expertise and can take several hours to complete.

In recent years, the use of machine learning in radiation oncology has grown exponentially with the rise of deep learning and its numerous applications in computer vision through convolutional neural networks (CNN) [89–91, 93, 116]. In 2015, Ronneberger et al. proposed the U-Net [52] architecture for medical image segmentation. This neural network model is an encoder-decoder architecture with long skip connections from the downsampling encoder to the upsampling decoder that enable a recovery of spatial detail during the upsampling process. First designed for 2D slices, 3D variants were developed in the following years [53, 54], motivated by the increased availability of volumetric medical imaging data. These 3D implementations proved highly relevant in dose prediction [91, 116] as they allow correct dose prediction across 2D slices avoiding notable errors near the edge of a planning target volume. Several alternatives have allowed to improve the U-Net training capabilities with the development of residual networks [50, 55] and dense CNNs [51, 56], both aiming for efficient deep networks that would achieve higher performance. However, working with a 3D variant of a U-Net requires more memory and computational time.

In 2018, Oktay et al. and Schempler et al. proposed a modular implementation of attention gates (AG) in a standard U-Net architecture (Attention U-Net) developed for medical image

segmentation tasks [117,118]. They relied on additive gates that only filter information flowing through the long skip connections, resulting in a memory efficient attention module for the U-Net. This led to improved results and demonstrated the benefit of AGs to identify and localize specific structures. In 2019, Nguyen et al. developed a 3D neural network architecture [91] that incorporates dense elements within the quintessential U-Net architecture (HD U-Net) while retaining a reasonable memory usage. They demonstrated the superiority of their network compared with their counterparts, namely the standard U-Net and the DenseNet. However, this approach is patch-based which is inherently prone to reconstruction artifacts.

The tendency of the mean squared error loss to converge rapidly regardless of the target domain made it a prime candidate for most deep learning regression problems. However, its inability to capture domain-specific knowledge remains a strong limitation for medical imaging tasks, especially when dealing with dose prediction which still relies heavily on knowledge-based and physics-driven approaches [92]. In 2018, Mahmood and Babier et al. developed a knowledge-based dose planning pipeline using adversarial learning [89,90]. Even though such an architecture effectively incorporates domain knowledge, it tends to rely on a high number of hyperparameters, making it difficult to train. In 2019, Nguyen et al. proposed a dosimetric objective based on the dose volume histogram [93], a commonly used metric and a defining tool in KBP for cancer treatment. They adapted the DVH metric into a differentiable objective and included an adversarial objective to capture the remaining domain-knowledge. They observed substantial improvements on their prostate cancer task, mainly in the form of a reduced tradeoff between PTV coverage and sparing of OARs. However, the training time was 2.5 times slower than with a typical MSE loss.

We propose a predictive model that produces high quality dose plans on annotated volumetric data by introducing a Hierarchically Dense Attention U-Net (HDA U-Net) that can process annotated CT scans with OAR and PTV segmentations. We assess the compatibility between dense and attention-based elements within a fully-connected encoder-decoder-like architecture to generate dose distributions. We also investigate the integration of a weighted MSE-DVH loss and validate the method on two separate datasets of H&N cancer patients achieving top performance on each. We compare our network against two state of the art models, namely Attention U-Net and HD U-Net, and we assess dose coverage, homogeneity and conformity performance.

5.3 Methods and materials

5.3.1 Data collection and formatting

Two separate and independent datasets were used to train and validate the proposed architecture : an in-house clinical dataset and a public dataset used for an open challenge.

First, the in-house dataset consists of 150 H&N patients treated with IMRT or VMAT. The selection criteria for the in-house dataset included two main categories of plans : curative plans for pharynx and neck cancers and curative plans for advanced oropharynx cancers ; it also excluded three categories of doses, namely : unilateral doses, highly focused doses and supplementary doses. These exceptional case scenarios require very specific tuning such as the addition of phantom delineations by the dosimetrist and fell outside the scope of our study. The original axial-plane resolution was reduced from 512x512 to 128x128 (1mm³ voxel size) due to memory constraints during training. This was done via downscaling by a factor of 4 using spline interpolation. Dose plans in full resolution could only be processed in a patch-based approach which lead to reconstruction issues at test time. When presented with a patch without tumor, models tend to predict plans with low levels of gradients and homogeneous dose values, without considering the actual proximity with cancerous tissue just beyond the patch. This is especially relevant in H&N cancers where preventive radiation is commonly sent to regions of future cancer growth. The same problem can occur when using 2D data and has been tackled with 2.5D methods that include additional slices as context for the model [119].

In our case, downscaling allowed us to use full 3D data and ensured that the model gets broader in-plane context information as well as an additional cross-plane axis to build stronger correlations. The final processed images had a resolution of 128 x 128 x [32-96]. The manually segmented structures include : medullary canal, outer medullary canal, esophagus, oral cavity, mandible, trachea, trunk, outer trunk, left and right parotids, left and right inner ear, left and right eye, left and right submaxillary glands, left optic nerve as well as the planning target volume (PTV), clinical target volume (CTV) and gross tumor volume (GTV). Segmentations were performed by experienced radiation oncologists. Patients were prescribed a dose varying from 66 to 70 Gy which was administered in multiple fractions.

The second dataset is a public dataset available through the OpenKBP challenge [120]. We utilized the available 340 H&N patients as 128 x 128 x 128 volumes treated with IMRT only. The manually segmented structures include : brain stem, spinal cord, right and left parotids, esophagus, larynx, mandible as well as the PTV 70, PTV 63 and PTV 56. Segmentations were also performed by experienced radiation oncologists. The latter refers to the high-dose

target volume receiving 70 Gy, the mid-dose volume receiving 63 Gy and the low-dose volume receiving 56 Gy. Each patient was prescribed a dose of 70 Gy administered in 35 fractions.

For both datasets, we included the computed tomography as the first input channel and kept the structure masks in a binary format, where the ones indicate the voxels assigned to the structure and where each structure was represented by a separate channel. In total, the input data was represented by 21 channels for the in-house dataset (one CT and 20 structure channels) and by 11 channels for the public dataset (one CT and 10 structure channels).

5.3.2 HDA U-Net architecture

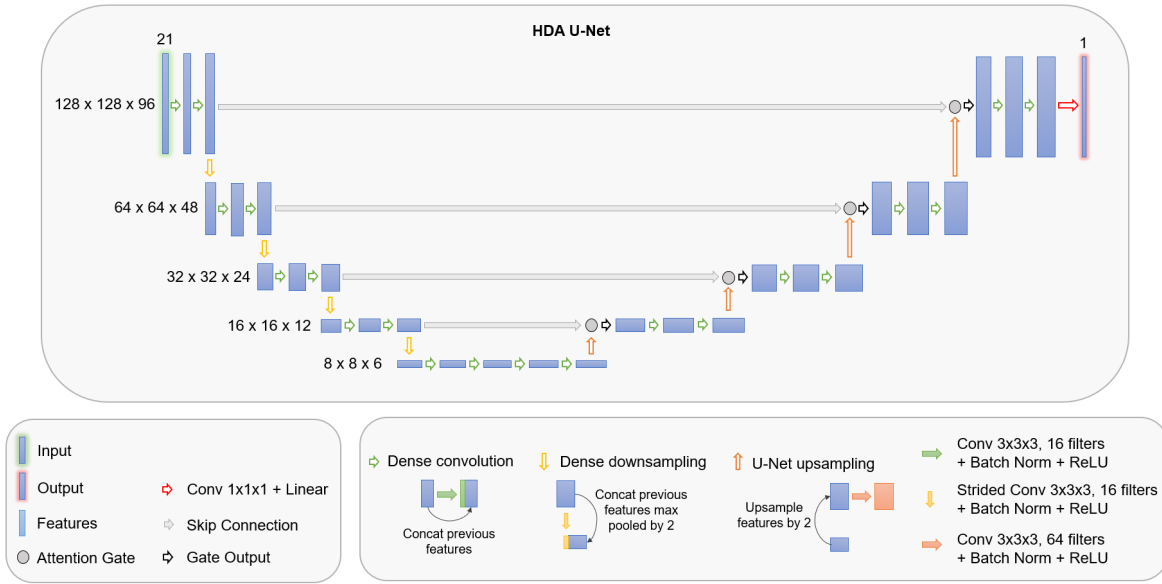


FIGURE 5.1 Schematic of the proposed HDA U-Net architecture that combines dense elements with attention mechanisms.

Standard U-Net aspects

Standard U-Net architectures [52] consist of an encoder that skips features to a decoder at each resolution in the network. The encoder acts as a downsampling path where the original input resolution is successively halved while the number of features grows. That way the model progressively integrates more spatial information as its receptive field grows. At each stage in the decoder, the features go through an upsampling operation followed by a convolution and are then combined with features drawn from the encoder. These long skip connections are key in helping the network predict high resolution detail.

Dense elements

The proposed model is based on hierarchically dense connected U-Nets which are used to capture the different sections of the model within a hierarchy [91], where the input’s resolution remains constant—namely between two max pooling and two upsampling operations. This type of model introduces two key operations : the dense convolution and the dense downsampling operations. The dense convolution consists of the the application of a 3x3x3 convolution followed by a ReLU, the result of which is concatenated to the previous features similarly to short skip connections in a DenseNet architecture [51]. This sequence happens twice at each hierarchy except in the bottom-most stage where it is repeated four times. It is a widespread practice in encoder-decoder architectures that improves the representation capabilities of the network at a low computational cost given the low input resolution in these layers (8x8x6 in our case). The dense downsampling operation concatenates the result of a 3x3x3 convolution with stride 2 followed by a ReLU with the previous features downsampled by a factor of 2 by max pooling. Together, this ensures that the number of features grows linearly with each operation while staying relatively low compared with a standard 3D U-Net architecture [53]. The global architecture is presented in Fig 5.1.

Attention mechanisms

Soft attention gates specifically designed for U-Nets were incorporated in the proposed architecture [117, 118]. Fig 5.2 depicts the implementation of the attention gate used in the proposed model. Instead of directly concatenating the skipped information with the upsampled features, the attention gate aims at tuning coefficients to suppress less relevant activations in the skipped features. More precisely, skipped features and upsampled features are convolved and summed before a ReLU, then the ReLU output is reweighted into attention coefficients in $[0, 1]$; the exact implementation details can be found in the figure. The expanding path of the U-Net then resumes, the skipped features are multiplied by these coefficients and concatenated with the upsampled features.

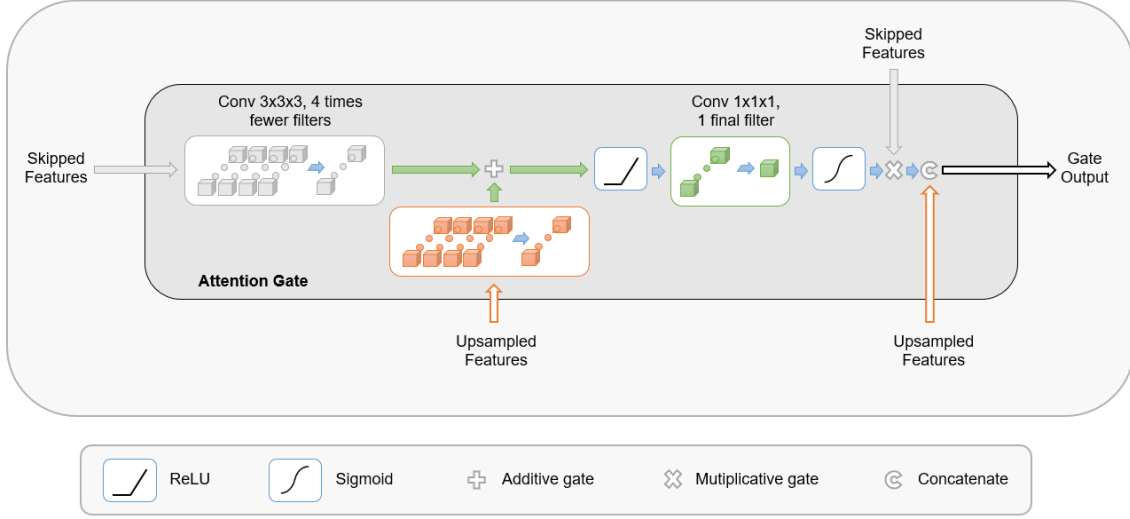


FIGURE 5.2 Schematic of the implemented attention gate using skipped features and upsampled features to produce attention coefficients.

Proposed HDA U-Net architecture details

The proposed HDA U-Net architecture implements the dense convolution and dense down-sampling operations, with the addition of a batch normalization layer after each convolution, to produce a combined output which is then used by the attention gates. The proposed model uses a growth rate of 16 to match the number of features added at each convolution within a hierarchy in the HD U-Net. The number of features entering the attention gate after upsampling and convolution was set to 64 to concord with previous studies [91].

To better compare performance, each model was designed to have the same number of down-sampling operations resulting in 3.4M parameters for the HD U-Net, 26M for the Attention U-Net and 3.5M for the HDA U-Net.

5.3.3 Training and validation

We repeated the experiments randomly initialized using three different random seeds on our in-house dataset, each time splitting the data into training (70%), validation (15%) and test (15%) subsets. For training on the public data, we used the same training, validation and test sets as defined in the challenge setup which consists in a training set of 200 patients, a validation set of 40 patients and a test set of 100 patients. In both cases, the best version of the model's weights was determined by the epoch that obtained the lowest validation loss and then evaluated on the testing data. Both datasets had their input CT images and output

radiation doses standardized via min-max normalization.

We implement the differential approximation of the DVH [93] which given a structure s , its associated binary mask M_s and a volumetric dose D can be defined as :

$$\widetilde{DVH}_s(D, M_s) = (v_{s,d_1}, v_{s,d_2}, \dots, v_{s,d_n}) \quad (5.1)$$

where the volume v_{s,d_t} is a volume v with voxel intensities at or above a threshold value d_t for a given structure s , which can be approximated as :

$$v_{s,d_t}(D, M_s) = \frac{\sum_{i,j,k} \text{Sigmoid}(\frac{m}{\beta_t}(D_{i,j,k}) - d_t) M_s(i, j, k)}{\sum_{i,j,k} M_s(i, j, k)} \quad (5.2)$$

where $\text{Sigmoid}(x) = \frac{1}{1+e^{-x}}$ is the sigmoid function, m is a steepness parameter, β_t is the bin width of the histogram and i, j, k are voxel indices for the 3D arrays. The t is an index for the dose threshold values and bin widths.

We have chosen a maximum dose threshold d_t by searching the range $[0, 80]$ Gy, with an increment step size of 1 Gy. Following the results of the original paper, we set the steepness value to $m = 1$. This small value ensures model convergence while retaining a fair approximation of the DVH. For more information on parameter choices, please refer to the source material [93].

This leads to the definition of the DVH objective as :

$$L_{DVH}(D_{true}, D_{pred}, M) = \frac{1}{n_s} \frac{1}{n_t} \sum_s \left\| \widetilde{DVH}_s(D_{true}, M_s) - \widetilde{DVH}_s(D_{pred}, M_s) \right\|_2^2 \quad (5.3)$$

where D_{true} and D_{pred} are the ground truth and predicted doses respectively, n_s is the number of structures and n_t the number of threshold values.

This objective was included in the loss function of the networks in a weighted manner as $L_{Total} = w_1 L_{MSE} + w_2 L_{DVH}$ with $w_1 = 1$ and $w_2 = 0.1$. These were chosen so that a trained model would exhibit the same magnitude for both terms. Also, we observed that when dealing with additional losses, keeping a strong driving MSE was key in achieving proper convergence. We used the Adam optimizer [105] with a learning rate of 1×10^{-3} , a batch size of 1 and default parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-7}$. All models were

trained for 200 epochs. Neural networks using the simple MSE objective were trained and evaluated using an NVIDIA V100 SXM2 GPU with 16 GB dedicated RAM. Neural networks using the weighted MSE-DVH objective were trained and evaluated using an NVIDIA Titan RTX GPU with 24 GB dedicated RAM.

5.3.4 Dosimetric measurements

Models were evaluated using the following dosimetric measurements : coverage values (D99, D98, D95), max dose value (Dmax), homogeneity ($H1 = \frac{D2-D98}{D50}$ and $H2 = \frac{D95}{D50}$), conformity index, van't Riet conformation number [34] ($\frac{(V_{PTV} \cap V_{100\%Iso})^2}{V_{PTV} \times V_{100\%Iso}}$), mean dose error (Dmean), as well as sample DVH curves. The coverage metrics are computed within the tumor, measuring the percentage of predicted dose relative to the prescribed dose, considering the 1st (D99), 2nd (D98) and 5th (D95) percentile of predicted doses. These metrics ensure that the vast majority of cancerous tissues are treated at a high percentage of the prescribed dose. Similarly, the max dose value shows the percentage of max predicted dose relative to the max prescribed dose (Dmax). The homogeneity indices are computed within the PTV as well. They depend on the difference between the worst predicted dose value and the best one, normalized over a median predicted dose value. They convey how homogeneously the tumor is treated across its entire volume. The conformity index and conformation number measure how many voxels meet the expected radiation threshold on the target volumes while the mean dose error acts as a broad measurement of quality and gives an idea of the error to expect for each structure. Finally, given a dose in Gray, the DVH shows the volume of tissue receiving at least that amount of radiation. It allows easily verifying the satisfaction of constraints on the maximum and minimum dosage for each structure. To measure the statistical significance of these metrics, we performed a two-tailed student's t-test in Microsoft Excel between the predictions' scores of the proposed method and the predictions' scores of each other presented method.

5.4 Results

The experiments were divided into two categories. The first set of experiments focused on the in-house dataset using clinical data from our institution, while the second set of experiments were performed on the public data from the OpenKBP challenge.

5.4.1 Clinical in-house dataset

TABLEAU 5.1 Coverage and max dose values for each model with ground truth values. Note : * shows that values are significantly different than the proposed HDA U-Net (DVH) results on the test data ($p < 0.05$). The best values among the architectures are shown in bold in each column.

Architectures	D99	D98	D95	Dmax
HD U-Net	$0.86^* \pm 0.03$	$0.88^* \pm 0.03$	$0.90^* \pm 0.02$	1.08 ± 0.02
Attention U-Net	$0.86^* \pm 0.03$	$0.88^* \pm 0.03$	$0.90^* \pm 0.03$	$1.09^* \pm 0.03$
HDA U-Net	$0.89^* \pm 0.03$	$0.90^* \pm 0.02$	$0.93^* \pm 0.02$	1.07 ± 0.04
HD U-Net (DVH)	$0.89^* \pm 0.02$	$0.91^* \pm 0.02$	$0.93^* \pm 0.01$	$1.08^* \pm 0.01$
Attention U-Net (DVH)	0.91 ± 0.01	0.92 ± 0.01	0.95 ± 0.01	$1.19^* \pm 0.18$
HDA U-Net (DVH)	0.92 ± 0.01	0.93 ± 0.01	0.95 ± 0.01	1.06 ± 0.02
Ground truth	0.96 ± 0.00	0.97 ± 0.00	0.98 ± 0.00	1.03 ± 0.00

Table 5.1 shows the PTV coverage and max dose values for the in-house dataset. All values are reported as a percentage of the prescribed dose. The Attention U-Net, the HD U-Net and the proposed HDA U-Net all receive a noticeable performance boost with the addition of the DVH loss on all coverage metrics. Overall the Attention U-Net and the proposed method outperform every other architecture on all coverage metrics. The proposed method does so while maintaining a significantly lower max dose value.

TABLEAU 5.2 Homogeneity and conformity values for each model with ground truth values. Note : * shows that values are significantly different than the proposed HDA U-Net (DVH) results on the test data (p -values < 0.05). The best values among the architectures are shown in bold in each column.

Architectures	H1	H2	CI	van't Riet
HD U-Net	$0.18^* \pm 0.03$	$0.87^* \pm 0.01$	$0.40^* \pm 0.06$	$0.28^* \pm 0.03$
Attention U-Net	$0.19^* \pm 0.01$	$0.86^* \pm 0.01$	0.44 ± 0.14	$0.32^* \pm 0.13$
HDA U-Net	$0.15^* \pm 0.04$	$0.89^* \pm 0.03$	0.45 ± 0.18	$0.33^* \pm 0.10$
HD U-Net (DVH)	$0.14^* \pm 0.02$	$0.90^* \pm 0.02$	$0.41^* \pm 0.03$	$0.36^* \pm 0.03$
Attention U-Net (DVH)	$0.17^* \pm 0.09$	$0.90^* \pm 0.04$	0.55 ± 0.01	0.42 ± 0.06
HDA U-Net (DVH)	0.10 ± 0.02	0.93 ± 0.01	0.53 ± 0.05	0.45 ± 0.03
Ground truth	0.06 ± 0.00	0.96 ± 0.00	0.85 ± 0.03	0.70 ± 0.03

Table 5.2 shows the homogeneity and conformity values for the in-house dataset. Similarly, the addition of the DVH loss improves the performance of every architecture. The proposed

method significantly outperforms all methods on the homogeneity metrics H1 and H2 and competes with the Attention U-Net on the conformity metrics. Performance remains quite far from reference scores for both the conformity index and the van't Riet conformation number regardless of the network.

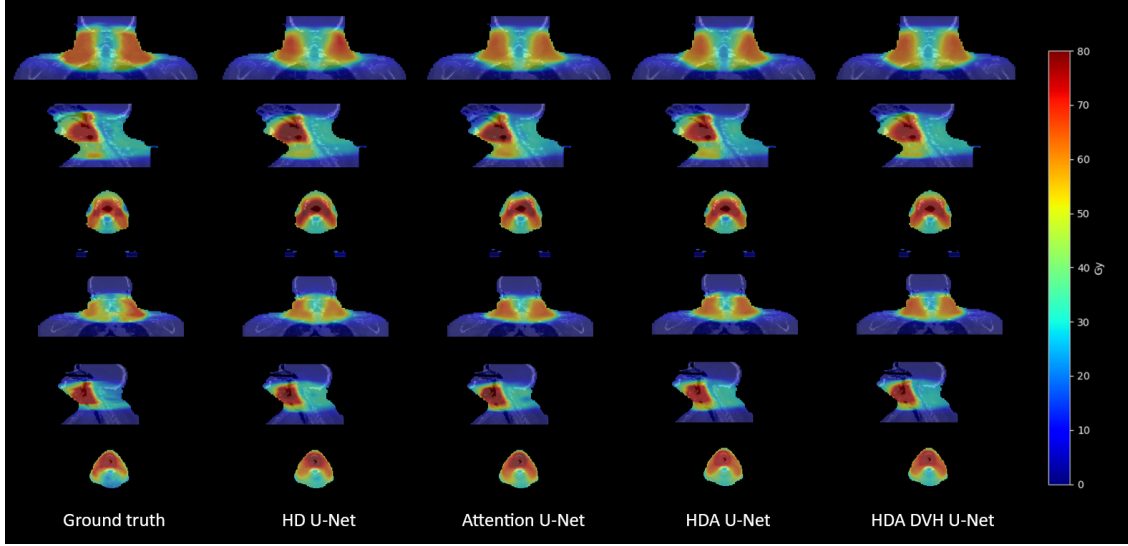


FIGURE 5.3 Dose examples overlapped with the input CT for two sample test patients from ground truth, HD U-Net, Attention U-Net, HDA U-Net and proposed HDA DVH U-Net.

Fig 5.3 shows predicted dose volumes on two sample patients from the in-house dataset. Every architecture is able to generate visually satisfying plans compared with the reference which justifies the extensive use of several metrics to properly differentiate them. However for the first example patient we can observe a shortage of dose in Gy around the mouth for the Attention U-Net and the HD U-Net, this is more visible on the sagittal and axial views. Overall, the proposed architecture yields the lowest average dose error on both examples with a plan within 3.8% and 2.4% average error of the reference distribution.

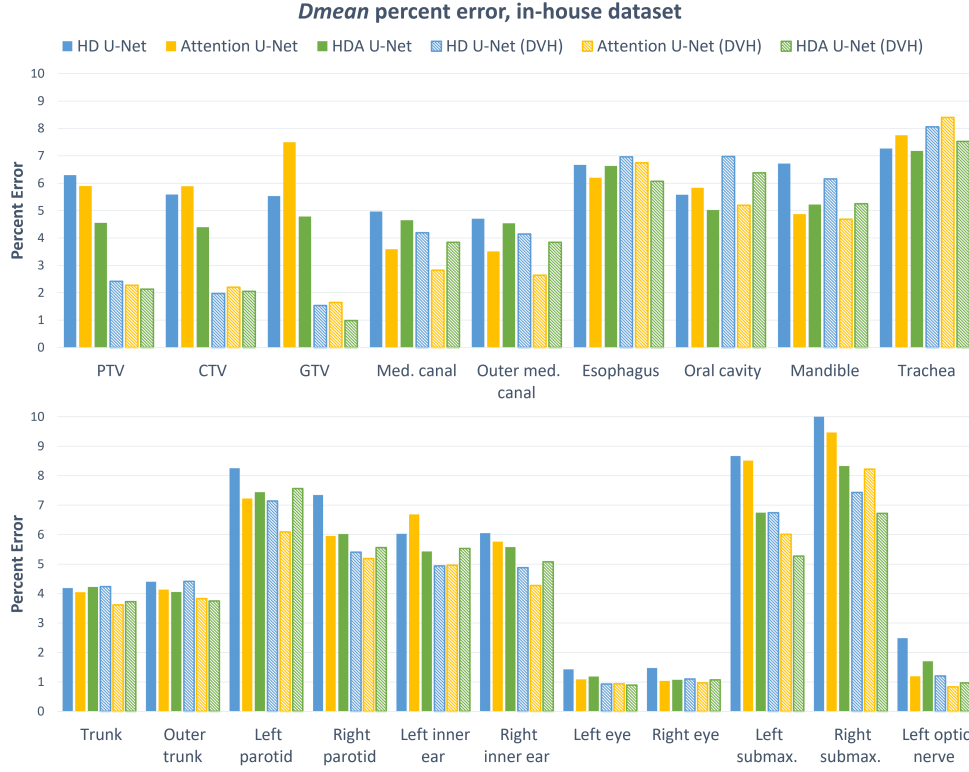


FIGURE 5.4 Dose mean percent error per structure for each model on the in-house dataset.

Fig 5.4 shows the average mean dose error for each architecture, on each structure of the in-house dataset. These are reported as a percentage of the prescription dose which gives an approximate idea of the quality to expect per structure. The addition of the DVH loss in all three architectures helped reduce the dose error in most structures. We observed up to a 2% reduction in dose error for the submaxillary glands, and up to a 4% reduction for the PTV, CTV and GTV. The Attention U-Net yields the lowest error on average on the OARs with a mean dose error of 4.4% (4.6% for proposed) while the proposed method yields the lowest error on average on TVs with a mean dose error of 1.7% (2.0% for Attention U-Net).

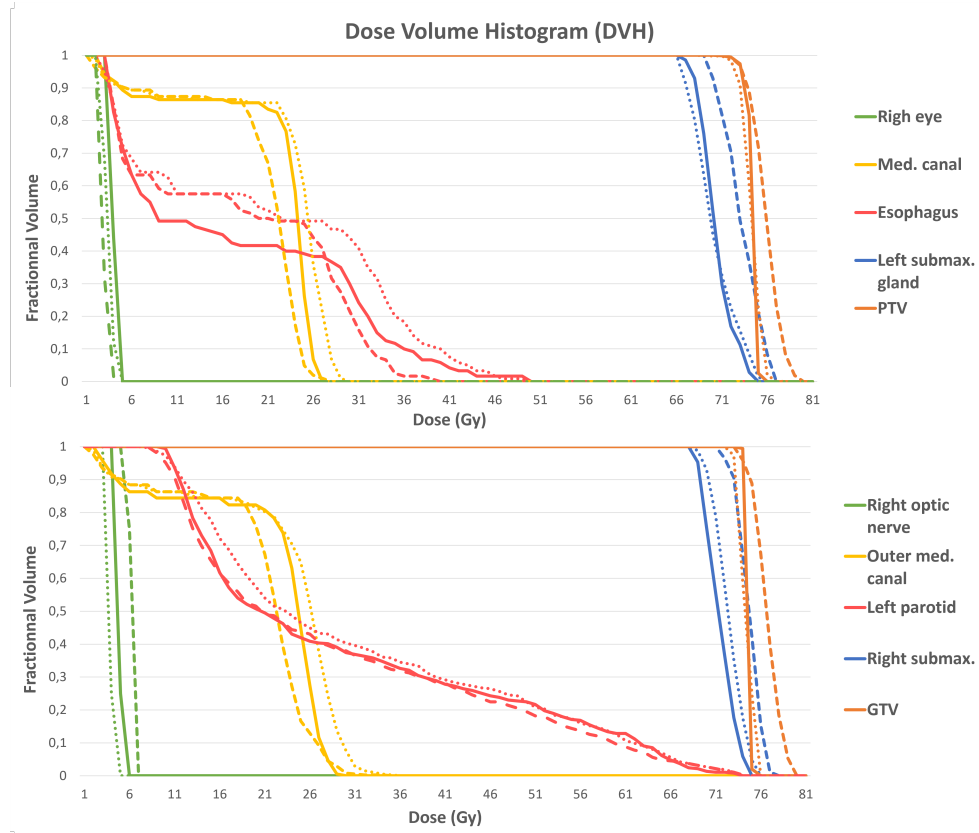


FIGURE 5.5 DVH curves for a sample test patient with doses from ground truth (solid line), HDA U-Net (dashed line) and proposed HDA U-Net using DVH loss (dotted line).

Fig 5.5 shows some sample DVH curves from a patient from the in-house dataset. To demonstrate the influence of the DVH loss on the performance, and for the sake of readability, we have chosen to include only the ground truth (in solid line), the HDA U-Net trained with MSE (in dashed line) and the proposed HDA U-Net trained with a weighted sum of MSE and DVH loss (in dotted line). For this sample patient, the curves of the proposed model are closer to the ground truth ones on a vast majority of structures (with the exception of the esophagus and left parotid).

5.4.2 OpenKBP challenge public dataset

TABLEAU 5.3 Coverage and max dose values for each model with ground truth values. Note : * shows that values are significantly different than the proposed HDA U-Net (DVH) results on the test data (p-values < 0.05). The best values among the architectures are shown in bold in each column.

Architectures	D99	D98	D95	Dmax
HD U-Net	0.93* \pm 0.02	0.94* \pm 0.02	0.96* \pm 0.01	1.07 \pm 0.03
Attention U-Net	0.93* \pm 0.03	0.94* \pm 0.02	0.96* \pm 0.02	1.05* \pm 0.02
HDA U-Net	0.94* \pm 0.03	0.95* \pm 0.02	0.97* \pm 0.02	1.06 \pm 0.02
HD U-Net (DVH)	0.92* \pm 0.03	0.94* \pm 0.03	0.96* \pm 0.02	1.07 \pm 0.02
Attention U-Net (DVH)	0.94 \pm 0.02	0.96 \pm 0.02	0.97 \pm 0.01	1.10* \pm 0.05
HDA U-Net (DVH)	0.95 \pm 0.03	0.96 \pm 0.02	0.98 \pm 0.02	1.07 \pm 0.03
Ground truth	0.96 \pm 0.05	0.97 \pm 0.04	0.99 \pm 0.03	1.06 \pm 0.03

Table 5.3 shows the PTV coverage and max dose values for the public dataset. The Attention U-Net and the proposed HDA U-Net receive a noticeable performance boost with the addition of the DVH loss on all coverage metrics. Overall, the Attention U-Net and the HDA U-Net using the DVH loss achieve significantly higher coverage scores than the HD variant. The max dose value is similar for all methods except Attention U-Net which scores lower without the DVH loss and higher with it.

TABLEAU 5.4 Homogeneity and conformity values for each model with ground truth values. Note : * shows that values are significantly different than the proposed HDA U-Net (DVH) results on the test data (p-values < 0.05). The best values among the architectures are shown in bold in each column.

Architectures	H1	H2	CI	van't Riet
HD U-Net	0.12* \pm 0.03	0.91* \pm 0.02	0.67* \pm 0.10	0.65* \pm 0.10
Attention U-Net	0.09 \pm 0.02	0.94 \pm 0.01	0.62* \pm 0.17	0.60* \pm 0.17
HDA U-Net	0.09* \pm 0.04	0.93* \pm 0.01	0.78 \pm 0.12	0.76 \pm 0.12
HD U-Net (DVH)	0.10* \pm 0.03	0.93* \pm 0.02	0.77 \pm 0.12	0.76 \pm 0.12
Attention U-Net (DVH)	0.09* \pm 0.02	0.93* \pm 0.01	0.80 \pm 0.10	0.78 \pm 0.10
HDA U-Net (DVH)	0.08 \pm 0.02	0.94 \pm 0.01	0.79 \pm 0.15	0.75 \pm 0.14
Ground truth	0.08 \pm 0.04	0.96 \pm 0.05	0.87 \pm 0.19	0.77 \pm 0.19

Table 5.4 shows the homogeneity and conformity values for the public dataset. The HD U-Net and the proposed HDA U-Net receive a noticeable performance boost with the addition

of the DVH loss on all homogeneity and conformity metrics. The proposed method scores the highest in terms of homogeneity with near ground truth level of quality while retaining reasonable conformity values.

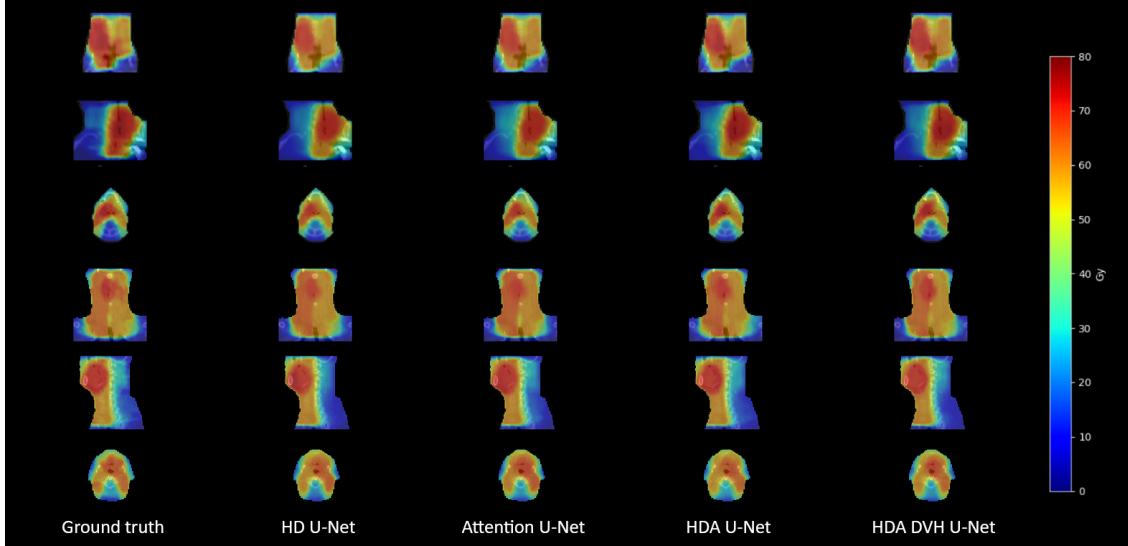


FIGURE 5.6 Dose examples overlapped with the input CT for two sample test patients from ground truth, HD U-Net, Attention U-Net, HDA U-Net and proposed HDA DVH U-Net.

Fig 5.6 shows dose prediction examples on sample patient cases from the public dataset. The dose distribution are close to ground truth across all architectures but tend to present higher levels of radiation around the throat for the first patient. On the same note, predictions are way smoother than the reference plans which demonstrate the inability of the MSE-trained models to replicate perfect beam-like trajectories. Overall, the proposed architecture yields the lowest average dose error on both examples with a plan within 4.0% and 2.2% average error of the reference distribution.

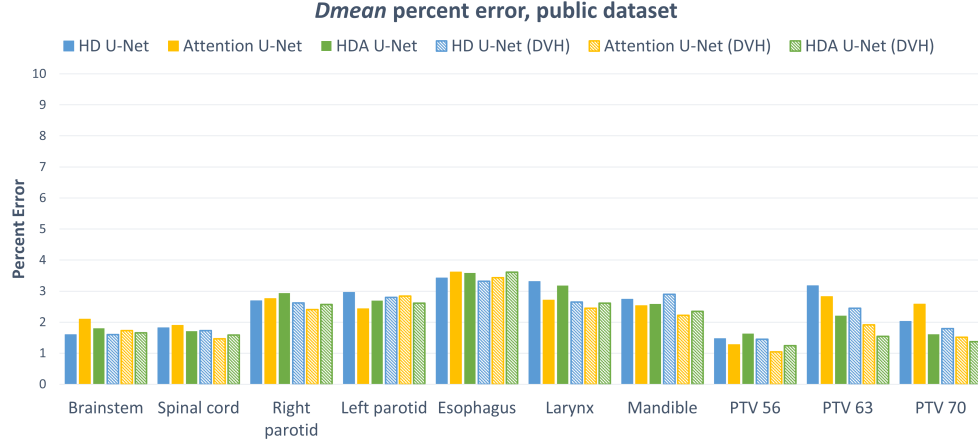


FIGURE 5.7 Dose mean percent error per structure for each model on the public dataset.

Fig 5.7 shows the average mean dose error for each architecture, on each structure of the public dataset. These are reported as a percent of the prescribed dose. The error values are considerably lower for the public dataset yet the same observations hold true : the DVH loss positively impacts the dose error across all three architectures. The Attention U-Net and the proposed method yield the lowest error on average on the OARs with a mean dose error of 2.4% while the proposed method yields the lowest error on average on TVs with a mean dose error of 1.4% (1.5% for Attention U-Net).

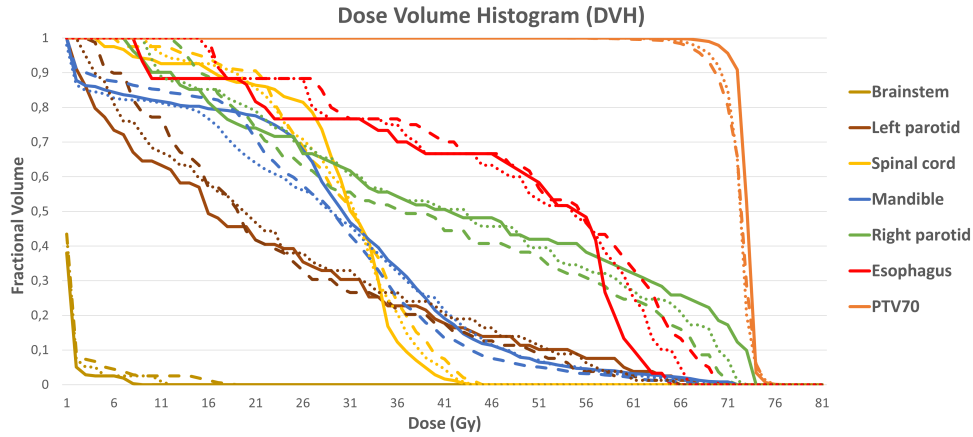


FIGURE 5.8 DVH curves for a sample patient with doses from ground truth (solid line), HDA U-Net (dashed line) and proposed HDA U-Net using DVH loss (dotted line).

Fig 5.8 shows some sample DVH curves from a patient of the public dataset. For this sample patient, the proposed model's curves are consistently closer to the ground truth ones (except for the mandible between 15 and 30 Gy). Overall, the proposed method follows much closely

the ground truth both in terms of value range and curve shape.

5.5 Discussion

The proposed HDA U-Net demonstrated the best performance across most dosimetric measurements with a significant improvement over every other architecture in both Dmax and homogeneity scores. It is important to note that this is true on both datasets as well as with and without the DVH loss function. Furthermore, in most cases, the weighted MSE-DVH improved the performance of the HD U-Net, the Attention U-Net and the proposed method. We believe this is an indication of the importance of using domain knowledge in deep learning when dealing with dosimetric predictions from medical imaging data. However, both the dense elements and the DVH objective significantly increase memory usage during training. Every tested architecture is able to produce accurate dose plans within less than a second (depending on the size of the scans), which makes them very relevant for real-time applications, such as for adaptive radiotherapy procedures. As part of the clinical workflow, the developed method allows physicians to quickly access a feasible dose distribution and gives the dosimetrist a concrete objective. This should considerably decrease the number of iterations and thus shorten the overall treatment planning phase.

Adequately choosing what to give the model as inputs plays a big part in developing a deep learning architecture. In recent dose prediction tasks, the patient CT is not always included as an input [91]. In our experiments however, its inclusion made a noticeable difference in terms of convergence. With the additional information included in the tomography, our networks were able to converge faster and more smoothly, achieving better performance. We ran a small experiment in which the CT was omitted at prediction time and the network would produce very diffuse plans, as if it lacked boundaries. We believe the CT may offer additional points of reference for the network to accurately set radiation values.

The public dataset allows a reproducible comparison to many other methods, whereas the in-house dataset is more realistically variable and contains twice as many structures. We observe on the latter a wider score gap between the predicted dose and the reference dose. For example, the highest D95 coverage is within 3% of the reference score (0.95 vs 0.98) while it is within 1% on the public data (0.98 vs 0.99). The same is true for most metrics, especially the conformity measurements, and underlines the impact of standardization on deep learning methods. One of the study goals is to evaluate the performance of the proposed model in various contexts. Systematically identifying and dealing with outliers remains an open problem in cancer treatment, meaning that outlier-resistant architectures ought to be developed.

The addition of the DVH loss did improve performance overall, however it seems to mostly benefit the PTVs, especially on the in-house dataset. The PTV channels may hold much needed information on the localization of the tumor and may drive the network to focus on them. On the other hand, nearby vital organs are also part of the dosimetric objective. That is why most knowledge-based planning methods rely on constraints on both tumors and organs at risk to generate viable plans. Different ways of conveying this concept with deep learning methods include giving the model expected values for each structure [121] or distance to target information [122, 123]. In practice, instead of having structure masks containing binary values, the channels can hold these meaningful values for the given structure. The network can learn to pay attention to these values and hopefully improve the plan quality. In our experiments, both inclusions of expected values and distance-to-target information were effective with shallower 2D predictors. They were harder to fine-tune for our deeper 3D architectures and often lead to convergence and saturation issues. Another way to integrate these constraints in our deep learning pipeline would be to weight errors differently depending on the region they occur in. Liu et al. explored this idea through a modified MSE that gives higher weights to points that belong to smaller structures [124]. As a future study, we plan to investigate this idea further by applying various weighting strategies to both the standard MSE and the DVH loss. We think that having more power over specific organs might be particularly relevant for body areas that present a large variety of structures such as the H&N region.

The in-house data is particularly challenging because it shows variable prescribed doses, different fractions, different treatment objectives, artifacts, etc. It contains a realistic variety of cases, each delineated by one of thirteen different physicians. This represents a variety of annotation styles, all of which may be equally correct, that the model must learn to generalize from when producing a dose prediction. Similarly, there may not be a single “ground truth” for dose prediction as different physicians may differ in their planning approach. For a model to produce useful outputs for a physician, it must be capable of producing outputs that match that physician’s style with specific case-by-case considerations and thus require minimal adjustment. In future work, we would like to include the ability to modify outputs to match the physician’s style. Balagopal et al. explored this idea by collecting annotations from different physicians and training a model with a separate decoder for each physician [125]. Further customization of the output can be achieved by developing a semi-automated method for the physician to submit corrections of the output to the model in an iterative fashion. One could even allow the physician to set constraints which are input into the model, as in hypernetworks that take hyperparameters as input [126]. These constraints could be per-region target dose values or even target dose values for specific pixels.

5.6 Conclusion

We developed a hierarchically dense attention U-Net architecture, HDA U-Net, to predict 3D dose distributions for H&N cancer patients. Our architecture integrates dense elements and soft attention gates in a 3D U-Net architecture. We explored the addition of a clinical objective with a weighted MSE-DVH loss function and proved it could be relevant in H&N cancers. We compared our proposed model with two variants of the U-Net : the HD U-Net and the Attention U-Net. The proposed method yielded the best results on two distinct datasets, showing improved generalization capabilities. Using the DVH objective, the HDA U-Net outperformed the HD U-Net on all metrics and outperformed the Attention U-Net on homogeneity while maintaining high conformity levels and a significantly lower max dose. The dose distributions can be generated within one second, making it very applicable for real-time purposes and allowing physicians to quickly produce a feasible goal for the dosimetrist to use as reference for planning. In future work, we plan to explore different weighting strategies to further improve the quality of the plans around specific organs and include the ability to match a physician’s style to better cope with variability within delineations.

5.7 Acknowledgments and Conflict of Interest Statement

We thank Robert Doucet from CHUM, who provided very valuable expertise on the in-house dataset and shared precious domain knowledge on the subject. This research has been funded in part by the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors have no relevant financial or non-financial interests to disclose.

Appendix

TABLEAU 5.5 Parameters, training and prediction time for each model.

	Total parameters	Average training time per epoch		Average prediction time (excluding environment loading time)	
		In-house ± 5s	Public ± 10s	In-house ± 0.05s	Public ± 0.07s
HD U-Net	3.4M	134s	402s	0.50s	0.74s
Attention U-Net	26M	120s	315s	0.43s	0.67s
HDA U-Net	3.5M	141s	402s	0.51s	0.75s
HD U-Net (DVH)	3.4M	145s	404s	0.52s	0.87s
Attention U-Net (DVH)	26M	123s	320s	0.44s	0.67s
HDA U-Net (DVH)	3.4M	148s	414s	0.53s	0.88s

CHAPITRE 6 DISCUSSION GÉNÉRALE

6.1 Autour de la segmentation des structures critiques

L'article présenté au chapitre 4 montre le développement d'une solution pour segmenter automatiquement les structures critiques sur un CT scan d'un patient. Afin de tenir compte des dépendances existantes entre les organes d'un même patient (distance, position, etc), une segmentation conjointe de tous les organes est d'abord envisagée mais ne permet pas de répondre entièrement au problème.

En effet, un fort déséquilibre dans l'apparition des classes à segmenter mène à l'obtention de résultats acceptables pour les organes les plus volumineux mais quasi-nuls pour les plus petites structures. Pour relever ce défi, les segmentations des douze structures ne sont plus prédites par un même modèle mais par autant de réseaux. Deux architectures semblables sont alors utilisées : une architecture peu profonde à trois niveaux et une architecture profonde à six niveaux. Dans le but de ré-intégrer les dépendances inter-organes dans ces modèles de prédiction et d'inclure plus de données, deux technologies de transfert de poids sont explorées : un transfert croisé entre organes et un calcul de moyenne de poids. Si le simple transfert de poids obtient des résultats inconsistants, l'état généré par la moyenne de poids semble donner une initialisation suffisamment spécifique pour accélérer et améliorer les performances des modèles.

Cette solution présente néanmoins de fortes limitations. Tout d'abord elle a été testée uniquement pour segmenter les structures critiques et non les volumes tumoraux. Ces derniers jouent pourtant un rôle prépondérant dans l'étape suivante d'élaboration des plans de doses. En effet, une expérience simple consiste à remplacer les champs réservés à ces volumes dans nos modèles de génération de plans de dose par des champs vides. Les plans générés dans ces conditions ne présentent presque plus de radiations, prouvant que ces volumes portent une très grande partie de l'information. La segmentation des volumes tumoraux est également moins documentée que la segmentation d'organes (presque trois fois moins de résultats sur les bases de données Compendex et Inspec entre 1980 et 2021) et introduit encore plus de variabilités dans les contours selon l'opérateur, l'hôpital et même le pays [8]. Cela reste ainsi une tâche importante qui repose généralement sur plusieurs modalités d'imagerie et sur plusieurs contrastes pour parfaire les segmentations des réseaux de neurones [127, 128].

Aussi, une autre limitation est apparue avec l'échantillonnage des patchs d'entraînement qui avait tendance à produire une grande proportion de patchs vides. En plus de ralentir

l'entraînement des modèles, cela a pu également favoriser les problèmes rencontrés lors de la prédiction conjointe des segmentations. Un nouveau mode d'échantillonnage a été développé et adopté pour entamer le projet suivant la génération de plans de dose. D'autres améliorations futures possibles comprennent le développement d'une moyenne pondérée des prédictions et d'une moyenne pondérée des poids, avec une combinaison apprise par un MLP par exemple. Le but est de s'assurer que les prédictions ou les poids extraits des réseaux agissent de façon constructive. Nous pensons également inclure plus de données pour vérifier que les performances restent bonnes sur un plus large panel de cas.

Enfin, au vue de la grande diversité de structures considérées, le score Dice n'est peut-être pas suffisant. Cette métrique possède en effet une grande sensibilité à la taille de l'objet considéré. Un même écart de quelques pixels peut avoir un faible impact sur les valeurs de score d'un gros volume mais un fort impact sur les valeurs d'un volume plus faible. L'utilisation d'une fonction de perte composée du score Dice et d'une distance résistante à de telles variations, comme la distance de Hausdorff par exemple, pourrait être envisagée.

6.2 Autour de la génération de plans de dose

L'article présenté au chapitre 5 détaille l'élaboration d'une solution pour générer efficacement des plans de dose. Les premiers obstacles rencontrés pour ce volet du projet ont été la compréhension et l'analyse de la base de données suivis des problématiques liées à la génération des plans. Pour éclaircir ces points, plusieurs rencontres avec des professionnels du CHUM ont été nécessaires.

6.2.1 Rencontres avec des professionnels du CHUM

Les échanges avec le radio-oncologue Phuc Felix Nguyen-Tan ont permis d'introduire des notions de base en radiothérapie portant notamment sur : les différents types d'images ; le déroulement classique de la planification, des premières phases de diagnostic à l'élaboration de contraintes précises ; l'évaluation des plans, la compréhension et le choix des métriques ; les connaissances préalables disponibles comme la dose prescrite et les seuils d'irradiation maximum par organe. Toutes ces informations ont alimenté les discussions sur le projet et auront sans conteste façonné la solution proposée.

Un certain nombre de défis ont également émergé des rencontres avec le radio-oncologue, notamment sur : les objectifs de traitement variés (curatif, palliatif, post-chirurgie) ; la multiplicité des localisations des tumeurs et la nécessité d'irradier des zones d'expansion futures (sans tumeur apparente) ; la variabilité inter-observateur ; la distinction entre cas typiques et

cas particuliers de traitements. La formulation de ces défis a mené à un centrage de l'étude sur un échantillon viable pour des expériences d'apprentissage machine et des solutions complémentaires ont été exposées en discussion du chapitre 5.

Les échanges avec l'expert dosimétriste Robert Doucet ont eux fait émerger des considérations plus techniques sur le processus d'optimisation, à savoir : l'identification des entrées (CT, segmentations, contraintes enrichies) et des sorties de l'algorithme Monte Carlo (deux à trois plans dans le cas du CHUM) ; la durée du processus (une à quatre heures) ; des considérations matérielles et physiques qui impactent la production de plan par la machine (simulation des trajectoires et interactions des particules, vitesse de rotation du bras de l'accélérateur linéaire, vitesse des lames permettant la modulations des rayons, etc). Notre étude ayant privilégié l'intégration d'aspects cliniques, ces aspects physiques n'ont pas été abordés. Néanmoins certains travaux incluent ces aspects et se basent même sur des méthodes similaires à celles explorées dans cette étude (réseaux convolutifs et fonction de perte modifiée) [129].

6.2.2 Développement d'un pipeline d'apprentissage profond

Approche en pleine résolution et 'downscaling'

Le deuxième enjeu important pour cette partie du projet a été de développer un pipeline d'apprentissage profond tout en gardant un coût raisonnable. Pour cela, une approche par patch a d'abord été développée pour utiliser les images du dataset interne en pleine résolution à coût réduit. Cependant, la reconstruction des plans de dose dans cette configuration laisse apparaître des artefacts indésirables qui rompent la continuité des plans. Rétrospectivement, après avoir constaté la grande importance accordée aux volumes tumoraux par le modèle, une solution de pré-traitement aurait pu être d'ajouter une entrée supplémentaire avec des volumes élargis. Cela aurait permis de s'assurer que tous les patches disposent d'information sur la localisation des tumeurs et aurait peut-être réduit les ruptures de continuité entre patches. Une solution de post-traitement a été succinctement testée : les patches prédits ont été superposés et des valeurs moyennes calculées aux frontières. Si cette solution a permis de donner un aspect plus lisse aux plans, elle a introduit un compromis entre temps de reconstruction et qualité des plans, ce qui a notamment éloigné la méthode globale d'une application en temps réel.

Finalement l'approche par patch a été abandonnée au profit d'une prédiction des volumes en une seule fois. Pour cela, les données CT, les segmentations et les plans ont été réduits de taille d'un facteur quatre avec une interpolation par spline. Les algorithmes ont ainsi bénéficié d'un entraînement accéléré et ne nécessitent plus de reconstruction finale. La méthode développée

reste un outil de visualisation, ainsi la perte de résolution a semblé négligeable face au gain en temps et en qualité. Aussi, avec ce changement, les images provenant du dataset interne se sont rapprochées du format des images du dataset public ce qui nous a permis de faire une comparaison plus claire des performances sur les deux bases de données. Cela reste une limitation de la méthode proposée qui pourrait par exemple être surmontée avec l'intégration de modules de super résolution [130] appliqués en radiothérapie [131] ou simplement avec l'évolution des capacités GPU.

Fonctions de perte

La solution développée utilise un objectif mixte combinant une erreur quadratique moyenne à un histogramme de volume de dose différentiable. Cependant, deux autres objectifs ont été envisagés utilisant respectivement les surfaces d'isodose et la détection des contours. Le long d'une surface d'isodose, la dose administrée reste constante. Ces surfaces sont généralement définies à intervalle régulier (tous les 10 Gy par exemple) et permettent de donner une plage de dose souhaitée par pixel. La détection des contours d'une image peut être obtenue à moindre coût par convolution avec des filtres de Sobel [40]. Cette opération révèle les zones de fortes variations d'intensité dans une image en approximant les gradients.

Dans notre étude, nous avons essayé d'intégrer ces deux notions sous forme de fonction de perte. Pour cela, les surfaces d'isodose ont été obtenues par seuillage des distributions de référence et ont donné un intervalle cible pour chaque pixel. Il suffit alors de renvoyer un coût nul quand le pixel est dans l'intervalle souhaité ; un coût proportionnel à la distance à l'intervalle quand le pixel est en dehors. En ce qui concerne les contours, les distributions de référence et prédites sont filtrées par des opérateurs de Sobel 3D puis comparées via MSE. Les deux objectifs associés à une MSE ont permis d'améliorer légèrement les performances d'un réseau U-Net de base mais pas celles de nos architectures plus avancées, montrant des résultats variables d'un entraînement à l'autre.

Finalement, l'objectif basé sur la combinaison MSE-DVH a montré de bien meilleurs gains de performance et ce sur toutes les architectures testées, il a donc été conservé. L'objectif actuel comporte néanmoins des limites puisque les réseaux développés semblent produire plus d'erreurs sur les organes à risque que sur les volumes tumoraux. Pour une amélioration future, nous envisageons de rajouter des poids par structure dans les fonctions de perte pour avoir plus de contrôle sur les structures ciblées lors de l'entraînement [124].

Pistes d'amélioration de l'architecture globale

D'un point de vue architectural, la méthode proposée combine des éléments denses à des modules d'attention. Ces derniers sont employés de façon restreinte puisque seules les convolutions d'une même résolution (ou hiérarchie) sont densément reliées et seules les 'long skip connections' de l'encodeur-décodeur présentent les modules d'attention. L'architecture U-Net a ainsi pu être enrichie tout en gardant un nombre de paramètres, des temps d'entraînement et des temps de prédiction faibles. Toutefois, le modèle pourrait servir de base de travail et connaître des améliorations dans son architecture globale. Pour ce type d'applications, les réseaux antagonistes génératifs (GAN) [82] utilisent habituellement un générateur de type encodeur-décodeur, un rôle que notre 'HDA U-Net' pourrait jouer. Malgré leur grande quantité d'hyperparamètres et leur entraînement difficile, ils donnent généralement des performances accrues [89, 90, 93]. Aussi, certains papiers de recherche incluent une distribution de dose dans l'entrée de leur réseau. Ce plan peut être une distribution grossière générée rapidement [132] ou même une distribution optimale issue d'un atlas de données [133, 134]. Cette nouvelle entrée permettrait au réseau de disposer d'une dose initiale qu'il pourrait apprendre à moduler pour un patient spécifique.

6.2.3 Intégration dans le fonctionnement clinique

Reste la question de l'intégration dans la pratique clinique en radio-oncologie. Malgré les limitations énoncées, la méthode de segmentation de structures critiques proposée est pleinement fonctionnelle et pourrait facilement intégrer les logiciels d'acquisition utilisés dans les hôpitaux. La méthode de génération de distribution de doses proposée reste une preuve de concept puisqu'elle ne permet pas de produire les paramètres nécessaires à la machine pour délivrer le plan. Néanmoins, elle offre un outil de visualisation que nous pensons utile, permettant de produire une distribution concrète et réalisable pour un patient donné. Sa capacité de prédiction en temps réel devrait permettre d'accélérer les échanges entre les professionnels concernés en les assistant dans la conception du plan.

CHAPITRE 7 CONCLUSION

La région de la tête et du cou est un foyer propice à l'apparition des tumeurs cancéreuses, accueillant des tumeurs aux localisations variables et toujours à proximité de structures critiques. Selon l'état de l'art pour région étudiée, le compromis de traitement repose toujours sur un objectif double : il est impératif d'irradier les volumes tumoraux à la dose prescrite tout en préservant les organes sains du patient. Malgré les avancées technologiques ayant eu un impact sur les étapes cruciales d'élaboration des contours et des distributions de dose, le processus de planification exige encore aujourd'hui un temps considérable et une expertise précieuse de la part des professionnels de santé. L'objectif aura donc été de proposer des méthodes efficaces pour accélérer et uniformiser ces étapes exigeantes du traitement du cancer de la tête et du cou.

Nous avons pour cela exploré des modèles d'apprentissage profond de type encodeur-décodeur convolutifs. Ils nous ont permis de synthétiser et transformer les données d'imagerie médicale pour obtenir des contours cliniques et des distributions de dose de bonne qualité. Le recours à des méthodes de transfert de poids a permis de ré-utiliser l'information acquise par certains modèles pré-entraînés sur des données spécifiques pour initialiser et améliorer l'entraînement de nos réseaux profonds. Aussi, nous avons développé une architecture légère intégrant des convolutions denses et des modules d'attention. Elle a démontré des performances accrues face à deux architectures de l'état de l'art sur des bases de données internes et publiques. Enfin, l'implémentation d'un objectif mixte incluant des aspects cliniques a permis d'améliorer les performances de tous les réseaux, démontrant une fois de plus l'intérêt des connaissances du domaine pour le traitement du cancer de la tête et du cou.

Cependant, les méthodes développées comportent plusieurs limitations. En particulier, la méthode de segmentation a été testée sur les structures critiques uniquement et pourrait nécessiter l'utilisation d'autres modalités pour segmenter aussi précisément les volumes tumoraux. Aussi, la méthode de génération de distributions de dose proposée reste un simple outil de visualisation pour assister les professionnels concernés dans l'élaboration d'un plan délivrable. Des améliorations futures de nos méthodes sont envisageables. D'abord avec l'utilisation d'une combinaison apprise des prédictions ou des poids de nos réseaux de segmentation qui permettrait d'assurer une interaction constructive entre les informations extraites de plusieurs réseaux. Pour la génération de distributions de dose, nous pensons que l'ajout d'un objectif antagoniste et d'une dose initiale pourrait bonifier l'entraînement de nos modèles.

RÉFÉRENCES

- [1] “Reports on cancer,” 2018. [En ligne]. Disponible : <https://seer.cancer.gov/statistics/reports.html>
- [2] A. Teymoortash et J. A. Werner, “Current advances in diagnosis and surgical treatment of lymph node metastasis in head and neck cancer,” *GMS Current topic in Otorhinolaryngology - Head and Neck Surgery*, vol. 11, 2012.
- [3] “Généralités sur les cancers : Les cancers des voies aérodigestives supérieures,” 2021. [En ligne]. Disponible : <https://www.chu-dijon.fr/cancerologie/cancers-tete-cou/generalites-sur-cancers>
- [4] “Head and neck cancer,” 2021. [En ligne]. Disponible : <https://www.cancer.net/cancer-types/head-and-neck-cancer/types-treatment>
- [5] “External radiation therapy,” 2021. [En ligne]. Disponible : <https://cancer.ca/en/treatments/treatment-types/radiation-therapy/external-radiation-therapy>
- [6] “Linear accelerator,” 2021. [En ligne]. Disponible : <https://www.radiologyinfo.org/en/info/linac>
- [7] E. B. Podgorsak, *Treatment machines for external beam radiotherapy*. Austria : INTERNATIONAL ATOMIC ENERGY AGENCY, 2005.
- [8] F. Nguyen, 2021, communication personnelle avec un médecin oncologue.
- [9] B. NG, T. SJ, B. KE et J. SJ, “Defining the tumour and target volumes for radiotherapy,” *Cancer Imaging*, vol. 4, n^o. 2, p. 153–161, 2004.
- [10] R. Alonzi et P. J. Hoskin, “Functional imaging in clinical oncology : magnetic resonance imaging- and computerised tomography-based techniques.” *Clinical oncology (Royal College of Radiologists (Great Britain))*, vol. 18 7, p. 555–70, 2006.
- [11] M. Pinkawa, M. J. Eble et F. M. Mottaghy, “Pet and pet/ct in radiation treatment planning for prostate cancer,” *Expert Review of Anticancer Therapy*, vol. 11, n^o. 7, p. 1035–1041, 2011. [En ligne]. Disponible : <https://doi.org/10.1586/era.11.51>
- [12] T. Gupta et S. Beriwal, “PET/CT-guided radiation therapy planning : From present to the future,” *Indian Journal of Cancer*, vol. 47, n^o. 2, p. 126–133, 2010.
- [13] S. Kamalian, M. H. Lev et R. Gupta, “Chapter 1 - computed tomography imaging and angiography – principles,” dans *Neuroimaging Part I*, ser. Handbook of Clinical Neurology, J. C. Masdeu et R. G. González, édit. Elsevier, 2016, vol. 135,

- p. 3–20. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/B9780444534859000015>
- [14] N. Caria, B. Engels, S. Bral, V. V. Hung, P. Doornaert, A. Muraglia, P. Meskell, A. R. Razavi et T. Morgas, “Varian smartsegmentation ®knowledge-based contouring clinical evaluation of an automated segmentation module,” 2013.
 - [15] D. N. Teguh, P. C. Levendag, P. W. Voet, A. Al-Mamgani, X. Han, T. K. Wolf, L. S. Hibbard, P. Nowak, H. Akhiat, M. L. Dirkx, B. J. Heijmen et M. S. Hoogeman, “Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck,” *International Journal of Radiation Oncology*Biology*Physics*, vol. 81, n°. 4, p. 950–957, 2011. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S0360301610009326>
 - [16] G. Strassmann, S. Abdellaoui, D. Richter, F. Bekkaoui, M. Haderlein, E. Fokas, N. Timmesfeld, B. M. Vogel, M. Henzel et R. Engenhart-Cabillic, “Atlas-based semiautomatic target volume definition (ctv) for head-and-neck tumors,” *International Journal of Radiation Oncology, Biology and Physics*, vol. 78, n°. 4, 11 2010. [En ligne]. Disponible : <https://www.osti.gov/biblio/21438064>
 - [17] C.-C. Teng, L. G. Shapiro et I. J. Kalet, “Head and neck lymph node region delineation with image registration,” *BioMedical Engineering OnLine*, vol. 9, 2010. [En ligne]. Disponible : <https://doi.org/10.1186/1475-925X-9-30>
 - [18] R. Speight, J. Sykes, R. Lindsay, K. Franks et D. Thwaites, “The evaluation of a deformable image registration segmentation technique for semi-automating internal target volume (itv) production from 4dct images of lung stereotactic body radiotherapy (sb rt) patients,” *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology*, vol. 98, p. 277–83, 02 2011.
 - [19] V. K. Reed, W. A. Woodward, L. Zhang, E. A. Strom, G. H. Perkins, W. Tereffe, J. L. Oh, T. K. Yu, I. Bedrosian, G. J. Whitman, T. A. Buchholz et L. Dong, “Automatic segmentation of whole breast using atlas approach and deformable image registration.” *International journal of radiation oncology, biology, physics*, vol. 73 5, p. 1493–500, 2009.
 - [20] N. Reynaert, S. van der Marck, D. Schaart, W. van der Zee, M. Tomsej, C. van Vliet-Vroegindewij, J. Jansen, M. Coghe, C. D. Wagter et B. Heijmen, “Monte carlo treatment planning : An introduction,” Netherlands Commission on Radiation Dosimetry, Netherlands, Rapport technique 16, 2006.
 - [21] P. Andreo, “Monte carlo simulations in radiotherapy dosimetry,” *Radiation Oncology*, vol. 13, 2018.

- [22] R. Doucet, 2021, communication personnelle avec un spécialiste en dosimétrie.
- [23] I. Kawrakow et M. Fippel, “Vmc++, a fast mc algorithm for radiation treatment planning,” 01 2010.
- [24] D. R. E. Pevey, 2020, chapter 4 Variance Reduction Techniques, lecture from the University of Tennessee Knoxville. [En ligne]. Disponible : <http://volweb.utk.edu/~rpevey/public/NE582/Chapter%204.pdf>
- [25] W. Ulmer et W. Kaissl, “The inverse problem of a gaussian convolution and its application to the finite size of the measurement chambers/detectors in photon and proton dosimetry,” *Physics in medicine and biology*, vol. 48, p. 707–27, 04 2003.
- [26] W. Ulmer, J. Pyyry et W. Kaissl, “A 3d photon superposition/convolution algorithm and its foundation on results of monte carlo calculations.” *Physics in medicine and biology*, vol. 50 8, p. 1767–90, 2005.
- [27] L. Tillikainen, H. Helminen, T. Torsti, S. Siljamäki, J. Alakuijala, J. Pyyry et W. Ulmer, “A 3d pencil-beam-based superposition algorithm for photon dose calculation in heterogeneous media,” vol. 53, n°. 14, p. 3821–3839, jun 2008. [En ligne]. Disponible : <https://doi.org/10.1088/0031-9155/53/14/008>
- [28] E. Sterpin, M. Tomsej, B. De Smedt, N. Reynaert et S. Vynckier, “Monte carlo evaluation of the aaa treatment planning algorithm in a heterogeneous multilayer phantom and imrt clinical treatments for an elekta sl25 linear accelerator,” *Medical Physics*, vol. 34, n°. 5, p. 1665–1677, 2007. [En ligne]. Disponible : <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.2727314>
- [29] A. Mesbahi, D. Thwaites et A. Reilly, “Experimental and monte carlo evaluation of eclipse treatment planning system for lung dose calculations,” *Reports of Practical Oncology & Radiotherapy*, vol. 11, p. 123–133, 05 2006.
- [30] A. Fogliata, G. Nicolini, A. Clivio, E. Vanetti et L. Cozzi, “Dosimetric evaluation of acuros xb advanced dose calculation algorithm in heterogeneous media,” *Radiation oncology (London, England)*, vol. 6, p. 82, 07 2011.
- [31] O. Vassiliev, T. Wareing, J. McGhee, G. Failla, M. Salehpour et F. Mourtada, “Validation of a new grid-based boltzmann equation solver for dose calculation in radiotherapy with photon beams,” *Physics in medicine and biology*, vol. 55, p. 581–98, 02 2010.
- [32] T. Wareing, J. McGhee, J. Morel et S. Pautz, “Discontinuous finite element sn methods on three-dimensional unstructured grids,” *Nuclear science and engineering : the journal of the American Nuclear Society*, vol. 138, p. 256–268, 07 2001.

- [33] L. Tillikainen, S. Siljamäki, H. Helminen, J. Alakuijala et J. Pyyry, “Determination of parameters for a multiple-source model of megavoltage photon beams using optimization methods,” *Physics in medicine and biology*, vol. 52, p. 1441–67, 04 2007.
- [34] A. van’t Riet, A. C. Mak, M. A. Moerland, L. H. Elders et W. van der Zee, “A conformation number to quantify the degree of conformality in brachytherapy and external beam irradiation : Application to the prostate,” *International Journal of Radiation Oncology*Biology*Physics*, vol. 37, n°. 3, p. 731–736, 1997. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S0360301696006013>
- [35] “Dose-volume histogram,” dans *Wikipédia*, 2011. [En ligne]. Disponible : https://en.wikipedia.org/wiki/Dose-volume_histogram
- [36] C. von der Malsburg, “Frank rosenblatt : Principles of neurodynamics : Perceptrons and the theory of brain mechanisms,” *Brain Theory*, p. 245–248, 01 1986.
- [37] H. Hassan, A. Negm, M. Zahran et O. Saavedra, “Assessment of artificial neural network for bathymetry estimation using high resolution satellite imagery in shallow lakes : Case study el burullus lake,” *International Water Technology Journal*, vol. 5, 12 2015.
- [38] Y. Lecun, L. Bottou, Y. Bengio et P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, n°. 11, p. 2278–2324, 1998.
- [39] A. Krizhevsky, I. Sutskever et G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” dans *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou et K. Q. Weinberger, édit., vol. 25. Curran Associates, Inc., 2012. [En ligne]. Disponible : <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>
- [40] N. Kanopoulos, N. Vasanthavada et R. L. Baker, “Design of an image edge detection filter using the sobel operator,” *IEEE Journal of solid-state circuits*, vol. 23, n°. 2, p. 358–367, 1988.
- [41] C. Pal, 2020, lecture IV : Modèles de probabilité conditionnelle d’apprentissage automatique, réseaux de neurones convolutifs et autoencodeurs.
- [42] S. Albelwi et A. Mahmood, “A framework for designing the architectures of deep convolutional neural networks,” *Entropy*, vol. 19, n°. 6, 2017. [En ligne]. Disponible : <https://www.mdpi.com/1099-4300/19/6/242>
- [43] K. Simonyan et A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” dans *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Y. Bengio et Y. LeCun, édit., 2015. [En ligne]. Disponible : <http://arxiv.org/abs/1409.1556>

- [44] J. Long, E. Shelhamer et T. Darrell, “Fully convolutional networks for semantic segmentation,” *CoRR*, vol. abs/1411.4038, 2014. [En ligne]. Disponible : <http://arxiv.org/abs/1411.4038>
- [45] I. Sutskever, O. Vinyals et Q. V. Le, “Sequence to sequence learning with neural networks,” *CoRR*, vol. abs/1409.3215, 2014. [En ligne]. Disponible : <http://arxiv.org/abs/1409.3215>
- [46] G. E. Hinton et R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” *Science*, vol. 313, n°. 5786, p. 504–507, 2006. [En ligne]. Disponible : <https://www.science.org/doi/abs/10.1126/science.1127647>
- [47] V. Nair et G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” dans *ICML*, 2010.
- [48] S. Ioffe et C. Szegedy, “Batch normalization : Accelerating deep network training by reducing internal covariate shift,” *CoRR*, vol. abs/1502.03167, 2015. [En ligne]. Disponible : <http://arxiv.org/abs/1502.03167>
- [49] L. Gupta, “Batch normalization and relu for solving vanishing gradients,” 2021. [En ligne]. Disponible : <https://medium.com/analytics-vidhya/how-batch-normalization-and-relu-solve-vanishing-gradients-3f1a8ace1c88>
- [50] K. He, X. Zhang, S. Ren et J. Sun, “Deep residual learning for image recognition,” *CoRR*, vol. abs/1512.03385, 2015. [En ligne]. Disponible : <http://arxiv.org/abs/1512.03385>
- [51] G. Huang, Z. Liu et K. Q. Weinberger, “Densely connected convolutional networks,” *CoRR*, vol. abs/1608.06993, 2016. [En ligne]. Disponible : <http://arxiv.org/abs/1608.06993>
- [52] O. Ronneberger, P. Fischer et T. Brox, “U-net : Convolutional networks for biomedical image segmentation,” *CoRR*, vol. abs/1505.04597, 2015. [En ligne]. Disponible : <http://arxiv.org/abs/1505.04597>
- [53] F. Milletari, N. Navab et S. Ahmadi, “V-net : Fully convolutional neural networks for volumetric medical image segmentation,” *CoRR*, vol. abs/1606.04797, 2016. [En ligne]. Disponible : <http://arxiv.org/abs/1606.04797>
- [54] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox et O. Ronneberger, “3d u-net : Learning dense volumetric segmentation from sparse annotation,” *CoRR*, vol. abs/1606.06650, 2016. [En ligne]. Disponible : <http://arxiv.org/abs/1606.06650>
- [55] Z. Zhang, Q. Liu et Y. Wang, “Road extraction by deep residual u-net,” *CoRR*, vol. abs/1711.10684, 2017. [En ligne]. Disponible : <http://arxiv.org/abs/1711.10684>

- [56] X. Li, H. Chen, X. Qi, Q. Dou, C. Fu et P. Heng, “H-denseunet : Hybrid densely connected unet for liver and liver tumor segmentation from CT volumes,” *CoRR*, vol. abs/1709.07330, 2017. [En ligne]. Disponible : <http://arxiv.org/abs/1709.07330>
- [57] D. Bahdanau, K. Cho et Y. Bengio, “Neural machine translation by jointly learning to align and translate,” *CoRR*, vol. abs/1409.0473, 2015.
- [58] K. Xu, J. Ba, R. Kiros, K. Cho, A. C. Courville, R. Salakhutdinov, R. S. Zemel et Y. Bengio, “Show, attend and tell : Neural image caption generation with visual attention,” *CoRR*, vol. abs/1502.03044, 2015. [En ligne]. Disponible : <http://arxiv.org/abs/1502.03044>
- [59] R. Karim, “Attn : Illustrated attention,” 2019. [En ligne]. Disponible : <https://towardsdatascience.com/attn-illustrated-attention-5ec4ad276ee3>
- [60] M. Luong, H. Pham et C. D. Manning, “Effective approaches to attention-based neural machine translation,” *CoRR*, vol. abs/1508.04025, 2015. [En ligne]. Disponible : <http://arxiv.org/abs/1508.04025>
- [61] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong et Q. He, “A comprehensive survey on transfer learning,” *CoRR*, vol. abs/1911.02685, 2019. [En ligne]. Disponible : <http://arxiv.org/abs/1911.02685>
- [62] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li et L. Fei-Fei, “Imagenet : A large-scale hierarchical image database,” dans *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, p. 248–255.
- [63] M. Raghu, C. Zhang, J. M. Kleinberg et S. Bengio, “Transfusion : Understanding transfer learning with applications to medical imaging,” *CoRR*, vol. abs/1902.07208, 2019. [En ligne]. Disponible : <http://arxiv.org/abs/1902.07208>
- [64] G. Huang, Y. Li, G. Pleiss, Z. Liu, J. E. Hopcroft et K. Q. Weinberger, “Snapshot ensembles : Train 1, get M for free,” *CoRR*, vol. abs/1704.00109, 2017. [En ligne]. Disponible : <http://arxiv.org/abs/1704.00109>
- [65] T. Garipov, P. Izmilov, D. Podoprikin, D. Vetrov et A. G. Wilson, “Loss surfaces, mode connectivity, and fast ensembling of dnns,” dans *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, ser. NIPS’18. Red Hook, NY, USA : Curran Associates Inc., 2018, p. 8803–8812.
- [66] P. Izmilov, D. Podoprikin, T. Garipov, D. P. Vetrov et A. G. Wilson, “Averaging weights leads to wider optima and better generalization,” *CoRR*, vol. abs/1803.05407, 2018. [En ligne]. Disponible : <http://arxiv.org/abs/1803.05407>

- [67] X. Guo, X. Liu, E. Zhu et J. Yin, “Deep clustering with convolutional autoencoders,” dans *Neural Information Processing*, D. Liu, S. Xie, Y. Li, D. Zhao et E.-S. M. El-Alfy, édit. Cham : Springer International Publishing, 2017, p. 373–382.
- [68] P. Vincent, H. Larochelle, Y. Bengio et P.-A. Manzagol, “Extracting and composing robust features with denoising autoencoders,” ser. ICML ’08. New York, NY, USA : Association for Computing Machinery, 2008, p. 1096–1103. [En ligne]. Disponible : <https://doi.org/10.1145/1390156.1390294>
- [69] S. Rifai, P. Vincent, X. Muller, X. Glorot et Y. Bengio, “Contractive auto-encoders : Explicit invariance during feature extraction,” dans *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ser. ICML’11. Madison, WI, USA : Omnipress, 2011, p. 833–840.
- [70] D. P. Kingma et M. Welling, “Auto-Encoding Variational Bayes,” dans *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014.
- [71] S. Tan et B. Li, “Stacked convolutional auto-encoders for steganalysis of digital images,” dans *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, 2014, p. 1–4.
- [72] A. Makhzani et B. J. Frey, “A winner-take-all method for training sparse convolutional autoencoders,” *CoRR*, vol. abs/1409.2752, 2014. [En ligne]. Disponible : <http://arxiv.org/abs/1409.2752>
- [73] M. Ahmad, J. Yang, D. Ai, S. Furqan Qadri et Y. Wang, “Deep-stacked auto encoder for liver segmentation,” 01 2018, p. 243–251.
- [74] X. Wang, S. Zhai et Y. Niu, “Automatic vertebrae localization and identification by combining deep ssae contextual features and structured regression forest,” *Journal of Digital Imaging*, vol. 32, 01 2019.
- [75] A. Elazab, F. Jia, J. Wu et Q. Hu, “Automated chest screening based on a hybrid model of transfer learning and convolutional sparse denoising autoencoder,” *BioMedical Engineering OnLine*, vol. 17, 05 2018.
- [76] E. Tappeiner, S. Pröll, M. Hönig, P. F. Raudaschl, P. Zaffino, M. F. Spadea, G. C. Sharp, R. Schubert et K. D. Fritscher, “Multi-organ segmentation of the head and neck area : an efficient hierarchical neural networks approach,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 14, p. 745–754, 2019.
- [77] R. Girshick, J. Donahue, T. Darrell et J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 11 2013.

- [78] R. B. Girshick, “Fast R-CNN,” *CoRR*, vol. abs/1504.08083, 2015. [En ligne]. Disponible : <http://arxiv.org/abs/1504.08083>
- [79] K. He, G. Gkioxari, P. Dollár et R. B. Girshick, “Mask R-CNN,” *CoRR*, vol. abs/1703.06870, 2017. [En ligne]. Disponible : <http://arxiv.org/abs/1703.06870>
- [80] E. Kopelowitz et G. Engelhard, “Lung nodules detection and segmentation using 3d mask-rcnn,” *ArXiv*, vol. abs/1907.07676, 2019.
- [81] Y. Li, L. Zhang, H. Chen et N. Yang, “Lung nodule detection with deep learning in 3d thoracic mr images,” *IEEE Access*, vol. 7, p. 37 822–37 832, 2019.
- [82] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville et Y. Bengio, “Generative adversarial nets,” dans *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, ser. NIPS’14. Cambridge, MA, USA : MIT Press, 2014, p. 2672–2680.
- [83] E. Choi, S. Biswal, B. A. Malin, J. Duke, W. F. Stewart et J. Sun, “Generating multi-label discrete electronic health records using generative adversarial networks,” *CoRR*, vol. abs/1703.06490, 2017. [En ligne]. Disponible : <http://arxiv.org/abs/1703.06490>
- [84] V. Alex, M. P., S. S. Chennamsetty et G. Krishnamurthi, “Generative adversarial networks for brain lesion detection,” 02 2017, p. 101330G.
- [85] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger et H. Greenspan, “Gan-based synthetic medical image augmentation for increased CNN performance in liver lesion classification,” *CoRR*, vol. abs/1803.01229, 2018. [En ligne]. Disponible : <http://arxiv.org/abs/1803.01229>
- [86] X. Dong, Y. Lei, T. Wang, M. Thomas, L. Tang, W. Curran, T. Liu et X. Yang, “Automatic multi-organ segmentation in thorax ct images using u-net-gan,” *Medical Physics*, vol. 46, 02 2019.
- [87] F. Mahmood, D. Borders, R. J. Chen, G. N. McKay, K. J. Salimian, A. S. Baras et N. J. Durr, “Deep adversarial training for multi-organ nuclei segmentation in histopathology images,” *CoRR*, vol. abs/1810.00236, 2018. [En ligne]. Disponible : <http://arxiv.org/abs/1810.00236>
- [88] R. Trullo, C. Petitjean, B. Dubray et S. Ruan, “Multiorgan segmentation using distance-aware adversarial networks,” *Journal of Medical Imaging*, vol. 6, p. 1, 01 2019.
- [89] R. Mahmood, A. Babier, A. McNiven, A. Diamant et T. C. Y. Chan, “Automated treatment planning in radiation therapy using generative adversarial networks,” *CoRR*, vol. abs/1807.06489, 2018. [En ligne]. Disponible : <http://arxiv.org/abs/1807.06489>

- [90] A. Babier, R. Mahmood, A. L. McNiven, A. Diamant et T. C. Chan, “Knowledge-based automated planning with three-dimensional generative adversarial networks,” *Medical Physics*, vol. 47, n^o. 2, p. 297–306, 2020. [En ligne]. Disponible : <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.13896>
- [91] D. Nguyen, X. Jia, D. Sher, M.-H. Lin, Z. Iqbal, H. Liu et S. Jiang, “3d radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected u-net deep learning architecture,” *Physics in Medicine & Biology*, vol. 64, n^o. 6, p. 065020, mar 2019. [En ligne]. Disponible : <https://doi.org/10.1088/1361-6560/ab039b>
- [92] S. Momin, Y. Fu, Y. Lei, J. Roper, J. D. Bradley, W. J. Curran, T. Liu et X. Yang, “Knowledge-based radiation treatment planning : A data-driven method survey,” *Journal of Applied Clinical Medical Physics*, vol. 22, n^o. 8, p. 16–44, 2021. [En ligne]. Disponible : <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/acm2.13337>
- [93] D. Nguyen, R. McBeth, A. Sadeghnejad Barkousaraie, G. Bohara, C. Shen, X. Jia et S. Jiang, “Incorporating human and learned domain knowledge into training deep neural networks : A differentiable dose-volume histogram and adversarial inspired framework for generating pareto optimal dose distributions in radiation therapy,” *Medical Physics*, vol. 47, n^o. 3, p. 837–849, 2020. [En ligne]. Disponible : <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.13955>
- [94] S. A. Taghanaki, K. Abhishek, J. P. Cohen, J. Cohen-Adad et G. Hamarneh, “Deep semantic segmentation of natural and medical images : A review,” *Artificial Intelligence Review*, p. 1–42, 2020.
- [95] Y. Lei, Y. Fu, T. Wang, R. L. J. Qiu, W. J. Curran, T. Liu et X. Yang, “Deep learning in multi-organ segmentation,” 2020.
- [96] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus et K. H. Maier-Hein, “Brain tumor segmentation and radiomics survival prediction : Contribution to the BRATS 2017 challenge,” *CoRR*, vol. abs/1802.10508, 2018. [En ligne]. Disponible : <http://arxiv.org/abs/1802.10508>
- [97] H. Kim, J. Jung, J. Kim, B. Cho, J. Kwak, J. Y. Jang, S.-w. Lee, J.-G. Lee et S. M. Yoon, “Abdominal multi-organ auto-segmentation using 3d-patch-based deep convolutional neural network,” *Scientific Reports*, vol. 10, n^o. 1, Apr 2020. [En ligne]. Disponible : <http://dx.doi.org/10.1038/s41598-020-63285-0>
- [98] van Rooij W, D. M, R. B. H, D. AR, S. BJ et V. WF, “Deep learning-based delineation of head and neck organs-at-risk : geometric and dosimetric evaluation,” *International Journal of Radiation Oncology • Biology • Physics*, vol. 62, p. 291–294, February 2019.

- [99] J. D. Furst, R. Susomboom et D. S. Raicu, “Single organ segmentation filters for multiple organ segmentation,” dans *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, Aug 2006. [En ligne]. Disponible : <http://dx.doi.org/10.1109/IEMBS.2006.260625>
- [100] Y. Wang, L. Zhao, Z. Song et M. Wang, “Organ at risk segmentation in head and neck ct images by using a two-stage segmentation framework based on 3d u-net,” 2019.
- [101] C. S. Perone, E. Calabrese et J. Cohen-Adad, “Spinal cord gray matter segmentation using deep dilated convolutions,” *CoRR*, vol. abs/1710.01269, 2017. [En ligne]. Disponible : <http://arxiv.org/abs/1710.01269>
- [102] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin et M. J. Cardoso, “Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations,” dans *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer, 2017, p. 240–248.
- [103] S. Chen, K. Ma et Y. Zheng, “Med3d : Transfer learning for 3d medical image analysis,” *CoRR*, vol. abs/1904.00625, 2019. [En ligne]. Disponible : <http://arxiv.org/abs/1904.00625>
- [104] R. Huang, Y. Zheng, Z. Hu, S. Zhang et H. Li, “Multi-organ segmentation via co-training weight-averaged models from few-organ datasets,” 2020.
- [105] D. P. Kingma et J. Ba, “Adam : A method for stochastic optimization,” *arXiv preprint arXiv :1412.6980*, 2014.
- [106] T. Tieleman et G. Hinton, “Lecture 6.5—RmsProp : Divide the gradient by a running average of its recent magnitude,” COURSERA : Neural Networks for Machine Learning, 2012.
- [107] A. Brahme, “Optimization of stationary and moving beam radiation therapy techniques,” *Radiotherapy and Oncology*, vol. 12, n^o. 2, p. 129–140, 1988. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/0167814088901673>
- [108] T. Bortfeld, J. Bürkelbach, R. Boesecke et W. Schlegel, “Methods of image reconstruction from projections applied to conformation radiotherapy,” *Physics in Medicine and Biology*, vol. 35, n^o. 10, p. 1423–1434, oct 1990. [En ligne]. Disponible : <https://doi.org/10.1088/0031-9155/35/10/007>
- [109] S. Webb, “Optimisation of conformal radiotherapy dose distribution by simulated annealing,” *Physics in Medicine and Biology*, vol. 34, n^o. 10, p. 1349–1370, oct 1989. [En ligne]. Disponible : <https://doi.org/10.1088/0031-9155/34/10/002>

- [110] C. X. Yu, “Intensity-modulated arc therapy with dynamic multileaf collimation : an alternative to tomotherapy,” *Physics in Medicine and Biology*, vol. 40, n°. 9, p. 1435–1449, sep 1995. [En ligne]. Disponible : <https://doi.org/10.1088/0031-9155/40/9/004>
- [111] K. Otto, “Volumetric modulated arc therapy : Imrt in a single gantry arc,” *Medical Physics*, vol. 35, n°. 1, p. 310–317, 2008. [En ligne]. Disponible : <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.2818738>
- [112] D. Palma, E. Vollans, K. James, S. Nakano, V. Moiseenko, R. Shaffer, M. McKenzie, J. Morris et K. Otto, “Volumetric modulated arc therapy for delivery of prostate radiotherapy : Comparison with intensity-modulated radiotherapy and three-dimensional conformal radiotherapy,” *International Journal of Radiation Oncology*Biophysics*, vol. 72, n°. 4, p. 996–1001, 2008. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S0360301608003428>
- [113] R. Shaffer, W. Morris, V. Moiseenko, M. Welsh, C. Crumley, S. Nakano, M. Schmuland, T. Pickles et K. Otto, “Volumetric modulated arc therapy and conventional intensity-modulated radiotherapy for simultaneous maximal intraprostatic boost : a planning comparison study,” *Clinical Oncology*, vol. 21, n°. 5, p. 401–407, 2009. [En ligne]. Disponible : <https://www.sciencedirect.com/science/article/pii/S0936655509000454>
- [114] K. Bzdusek, H. Friberger, K. Eriksson, B. Hårdemark, D. Robinson et M. Kaus, “Development and evaluation of an efficient approach to volumetric arc therapy planning,” *Medical Physics*, vol. 36, n°. 6Part1, p. 2328–2339, 2009. [En ligne]. Disponible : <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.3132234>
- [115] M. Teoh, C. H. Clark, K. Wood, S. Whitaker et A. Nisbet, “Volumetric modulated arc therapy : a review of current literature and clinical use in practice,” *The British Journal of Radiology*, vol. 84, n°. 1007, p. 967–996, 2011, pMID : 22011829. [En ligne]. Disponible : <https://doi.org/10.1259/bjr/22373346>
- [116] J. Zhou, Z. Peng, Y. Song, Y. Chang, X. Pei, L. Sheng et G. Xu, “A method of using deep learning to predict three-dimensional dose distributions for intensity-modulated radiotherapy of rectal cancer,” *Journal of Applied Clinical Medical Physics*, vol. 21, 04 2020.
- [117] O. Oktay, J. Schlemper, L. L. Folgoc, M. C. H. Lee, M. P. Heinrich, K. Misawa, K. Mori, S. G. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker et D. Rueckert, “Attention u-net : Learning where to look for the pancreas,” *CoRR*, vol. abs/1804.03999, 2018. [En ligne]. Disponible : <http://arxiv.org/abs/1804.03999>
- [118] J. Schlemper, O. Oktay, M. Schaap, M. P. Heinrich, B. Kainz, B. Glocker et D. Rueckert, “Attention gated networks : Learning to leverage salient regions

- in medical images,” *CoRR*, vol. abs/1808.08114, 2018. [En ligne]. Disponible : <http://arxiv.org/abs/1808.08114>
- [119] Z. Liu, J. Fan, M. Li, H. Yan, Z. Hu, P. Huang, Y. Tian, J. Miao et J. Dai, “A deep learning method for prediction of three-dimensional dose distribution of helical tomotherapy,” *Medical Physics*, vol. 46, n°. 5, p. 1972–1983, 2019. [En ligne]. Disponible : <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.13490>
- [120] A. Babier, B. Zhang, R. Mahmood, K. L. Moore, T. G. Purdie, A. L. McNiven et T. C. Y. Chan, “Openkbp : The open-access knowledge-based planning grand challenge,” 2021.
- [121] D. Nguyen, A. S. Barkousaraie, C. Shen, X. Jia et S. Jiang, “Generating pareto optimal dose distributions for radiation therapy treatment planning,” dans *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, P.-T. Yap et A. Khan, édit. Cham : Springer International Publishing, 2019, p. 59–67.
- [122] T. H. Maryam, B. Ru, T. Xie, M. Hadzikadic, Q. J. Wu et Y. Ge, “Dose prediction for prostate radiation treatment : Feasibility of a distance-based deep learning model,” *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, p. 2379–2386, 2019.
- [123] J. Zhang, S. Liu, H. Yan, T. Li, R. Mao et J. Liu, “Predicting voxel-level dose distributions for esophageal radiotherapy using densely connected network with dilated convolutions,” *Physics in Medicine & Biology*, vol. 65, 07 2020.
- [124] R. Liu, J. Bai, K. Zhao, K. Zhang et C. Ni, “A new deep-learning-based model for predicting 3d radiotherapy dose distribution in various scenarios,” *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, p. 748–753, 2020.
- [125] A. Balagopal, H. E. Morgan, M. Dohopolski, R. Timmerman, J. Shan, D. F. Heitjan, W. Liu, D. Nguyen, R. Hannan, A. Garant, N. Desai et S. B. Jiang, “Psa-net : Deep learning based physician style-aware segmentation network for post-operative prostate cancer clinical target volume,” *CoRR*, vol. abs/2102.07880, 2021. [En ligne]. Disponible : <https://arxiv.org/abs/2102.07880>
- [126] D. Ha, A. M. Dai et Q. V. Le, “Hypernetworks,” *CoRR*, vol. abs/1609.09106, 2016. [En ligne]. Disponible : <http://arxiv.org/abs/1609.09106>
- [127] M. A. Naser, L. V. van Dijk, R. He, K. A. Wahid et C. D. Fuller, “Tumor segmentation in patients with head and neck cancers using deep learning based-on multi-modality

- pet/ct images,” dans *Head and Neck Tumor Segmentation*, V. Andrearczyk, V. Oreiller et A. Depeursinge, édit. Cham : Springer International Publishing, 2021, p. 85–98.
- [128] L. Bielak, N. Wiedenmann, A. Berlin, N. Nicolay, D. D. Gunashekar, L. Hägele, T. Lottner, A. Grosu et M. Bock, “Convolutional neural networks for head and neck tumor segmentation on 7-channel multiparametric mri : a leave-one-out analysis,” *Radiation Oncology*, vol. 15, 07 2020.
- [129] T. Meerbothe, “A physics guided neural network approach for dose prediction in automated radiation therapy treatment planning,” mémoire de maîtrise, Delft University of Technology, 2021. [En ligne]. Disponible : <https://repository.tudelft.nl/islandora/object/uuid%3A9e6ce94e-d103-4a35-857c-648715614a9e>
- [130] C. Dong, C. C. Loy, K. He et X. Tang, “Image super-resolution using deep convolutional networks,” *CoRR*, vol. abs/1501.00092, 2015. [En ligne]. Disponible : <http://arxiv.org/abs/1501.00092>
- [131] P. Dong et L. Xing, “Deep dosenet : A deep neural network for accurate dosimetric transformation between different spatial resolutions and/or different dose calculation algorithms for precision radiation therapy,” *Physics in Medicine & Biology*, vol. 65, 12 2019.
- [132] M. Ma, N. Kovalchuk, M. K. Buyyounouski, L. Xing et Y. Yang, “Incorporating dosimetric features into the prediction of 3d VMAT dose distributions using deep convolutional neural network,” vol. 64, n°. 12, p. 125017, jun 2019. [En ligne]. Disponible : <https://doi.org/10.1088/1361-6560/ab2146>
- [133] C. McIntosh, M. Welch, A. McNiven, D. A. Jaffray et T. G. Purdie, “Fully automated treatment planning for head and neck radiotherapy using a voxel-based dose prediction and dose mimicking method,” vol. 62, n°. 15, p. 5926–5944, jul 2017. [En ligne]. Disponible : <https://doi.org/10.1088/1361-6560/aa71f8>
- [134] S. A. Yoganathan et R. Zhang, “An atlas-based method to predict three-dimensional dose distributions for cancer patients who receive radiotherapy,” vol. 64, n°. 8, p. 085016, apr 2019. [En ligne]. Disponible : <https://doi.org/10.1088/1361-6560/ab10a0>