



HAL
open science

Quantification de l'aération pulmonaire sur des images CT de patients atteints du syndrome de détresse respiratoire aiguë

Ludmilla Penarrubia

► **To cite this version:**

Ludmilla Penarrubia. Quantification de l'aération pulmonaire sur des images CT de patients atteints du syndrome de détresse respiratoire aiguë. Médecine humaine et pathologie. Université Claude Bernard - Lyon I, 2022. Français. NNT : 2022LYO10164 . tel-04496018

HAL Id: tel-04496018

<https://theses.hal.science/tel-04496018>

Submitted on 8 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**THÈSE de DOCTORAT DE
L'UNIVERSITÉ CLAUDE BERNARD LYON 1**

**École Doctorale N°205
École Doctorale Interdisciplinaire Science Santé (EDISS)**

Discipline : Traitement du signal et des images

Soutenue publiquement le 05/12/2022, par :

Ludmilla Penarrubia

**Quantification de l'aération pulmonaire sur des images CT
de patients atteints du syndrome de détresse
respiratoire aiguë**

Devant le jury composé de :

Alain Lalande	Maître de conférences - Praticien Hospitalier, Université de Bourgogne Franche-Comté	Rapporteur
Diana Mateus	Professeure des universités, Centrale Nantes	Rapporteuse
Loïc Bousset	Professeur des Universités - Praticien Hospitalier, Université Claude Bernard Lyon 1	Examineur
Caroline Petitjean	Professeur des universités, Université de Rouen Normandie	Présidente
Maria A. Zuluaga	Maître de conférences, EURECOM	Examinatrice
Jean-Christophe Richard	Professeur des Universités - Praticien hospitalier, Université Claude Bernard Lyon 1	Directeur de thèse
Maciej Orkisz	Professeur des universités, Université Claude Bernard Lyon 1	Co-directeur de thèse
Emmanuel Roux	Maitre de conférences, Université Claude Bernard Lyon 1	Co-directeur de thèse

La médecine, la loi, le commerce et l'industrie sont de nobles occupations, et nécessaires pour la survie de l'humanité.

Mais la poésie, la beauté et le dépassement de soi, l'amour : c'est tout ce pour quoi nous vivons. Écoutez ce que dit Whitman : « Ô moi ! Ô vie !... Ces questions qui me hantent, ces cortèges sans fin d'incrédules, ces villes peuplées de fous. Quoi de bon parmi tout cela ? Ô moi ! Ô vie ! ». Réponse : que tu es ici, que la vie existe, et l'identité. Que le spectacle continue et que tu peux y apporter ta rime. Que le spectacle continue et que tu peux y apporter ta rime...

Quelle sera votre rime ?

Professeur Keating - Le cercle des poètes disparus

Remerciements

Depuis aussi longtemps que je me souviens, apprendre a toujours fait partie de mes occupations préférées. Il y a quelque chose d'exaltant dans le fait d'acquérir des connaissances, je trouve. D'un côté, l'excitation de découvrir des choses par soi-même, sur le monde, sur soi et de l'autre, la reconnaissance envers ceux qui nous partagent leurs savoirs. À l'échelle de la vie, la thèse est un condensé des deux.

C'est ainsi chargée de leçons que j'aimerais remercier ceux qui m'ont tant appris ces trois dernières années : Jean-Christophe, Maciej et Emmanuel. Je n'aurais pu espérer meilleur encadrement pour m'essayer au travail de recherche. Merci Jean-Christophe de t'être toujours rendu disponible, malgré des contraintes importantes, notamment avec cette pandémie que tu as affronté aux premières loges. T'entendre parler des patients m'a toujours permis de garder en tête l'importance du projet que j'ai rejoint. Merci Maciej d'avoir été toujours à l'écoute, le précieux guide dans la longue aventure qu'est la thèse. Et enfin, merci Emmanuel d'avoir rejoint l'encadrement de mon travail. Ton expertise et ton enthousiasme m'ont été d'une aide considérable. La bienveillance avec laquelle vous m'avez, chacun, accompagnée dans ce travail de recherche m'a profondément marquée, un grand merci.

La recherche est faite d'essais, de réussites et d'échecs, mais surtout d'analyses, car c'est l'esprit critique qui permet d'apporter du poids aux réponses que l'on propose. Je tiens alors à remercier les membres de mon jury pour avoir accepté d'évaluer mon travail. Merci à Loïc Bousset, Caroline Petitjean et Maria A. Zuluaga pour leurs retours et leur grand intérêt en tant qu'examineurs. Et un merci particulier à Alain Lalande et Diana Mateus pour leur relecture attentive de mon manuscrit de thèse.

Cette aventure de trois ans n'aurait pas eu la même saveur si elle n'avait pas eu lieu à CREATIS. Je remercie ainsi tous celles et ceux que j'ai pu croiser au cours de ces années pour leur accueil et leur contribution à faire de ce laboratoire un havre pour la recherche.

Il faut évidemment que je mentionne celles et ceux que je n'ai pas juste croisés, mais qui ont rendu mon quotidien si agréable ces trois années durant.

Merci à mes premières rencontres, mes co-bureaux de BP, Antonio et Florian pour leur accueil et l'introduction à la passion des mots fléchés. J'en profite alors pour remercier tous les adhérents de la team mots fléchés pour les belles pauses (mention spéciale au bureau 13 qui en est devenu le siège).

Année 2021, déménagement vers LdV : j'ai rejoint de nouveaux co-bureaux avec qui cela a été un plaisir de partager la distanciation sociale comme le retour à la normale. Merci Charles, Raoul, Zexian, Nicolas (non ! Loiseau) et Julia (et Théo ?) pour ces belles discussions et la bonne humeur ! Celles et ceux avec qui je n'ai pas partagé de bureau, mais beaucoup de mon temps en salle de pause ou ailleurs, par exemple à la K-fêt ou au Ninkasi ou au Shrubbery ou encore au Toi-Toi : Anne-Lise, Audrey, Maxime, Benoît, Frank, PJ, Matthis, Sophie, Juliette, Mehdi, Samaneh, Cyril, Suzanne, Antoine, Kannara. Durant ces trois années entrecoupées de distanciation sociale, échanger, sortir et profiter de la vie à vos côtés a été un réel plaisir.

Un merci tout particulier aux deux personnes qui ont marqué ma dernière année de thèse : Pierre et Gassoupaalou. Merci pour cette année mémorable à vos côtés : les soirées, les

balades, les rires, les films et les pizzas, les photos, les défis rap et j'en passe. Un grand merci Pierre pour ton soutien durant la belle période de rédaction!

Et pour finir, un merci tout particulier à ma coloc, Louise, sans qui je n'aurais pas survécu pareil aux nombreux confinements (courage, la thèse est bientôt finie)!

Aussi, il y a celles et ceux que je n'ai pas vus durant ma thèse, mais que j'ai toujours gardé à l'esprit. Notamment tous les professeurs qui ont marqué ma scolarité par leur passion, enseignements et bienveillance. Du primaire à l'école d'ingénieur, en passant par le collège et les classes prépa, je remercie ces enseignants qui ont su me partager leurs savoirs et le goût des études.

Celles et ceux que j'ai peu vus : merci Julie et Thomas d'être toujours dans ma vie malgré la distance. Merci Mamie d'avoir été là pour moi dans les moments difficiles. Merci à tous ces amis que j'ai revus comme au bon vieux temps, qui m'ont presque fait oublier les années qui passent : Sara, Jules, Raphou et Lolo, Camille et Coco, Maud, Val, Clément, Jessy, Nico, Robin, Darchi, Arthur.

J'aimerais également adresser mes remerciements à Aleski et Senena, qui même loin de moi me portent dans leur cœur : c'est réciproque!

Enfin, je remercie du fond du cœur mes parents pour leur soutien, durant cette thèse, mais pas seulement. Si j'en suis arrivée jusque-là, c'est grâce à vous, merci d'avoir toujours cru en moi.

Résumé

Le syndrome de détresse respiratoire aiguë (SDRA) provoque un dérèglement de l'équilibre vital entre oxygène apporté et dioxyde de carbone évacué par le système respiratoire. En effet, les alvéoles qui composent les poumons sont collabées ou remplies de liquide en cas de SDRA, ce qui empêche le processus normal de ventilation. Pour maintenir en vie les patients diagnostiqués du SDRA, le recours à la ventilation mécanique est systématique. Cette dernière permet de rétablir l'équilibre oxygène-dioxyde de carbone, mais peut provoquer des lésions pulmonaires supplémentaires. Dans le but de réduire ces lésions et adapter la prise en charge de chaque patient, il est nécessaire de quantifier l'aération pulmonaire. L'image tomодensitométrique (ou CT pour *computed tomography*), déjà utilisée en clinique dans le processus diagnostique, permet cette quantification, car elle contient des informations de densité, mais exige une segmentation préliminaire des poumons. Cette thèse propose d'automatiser la tâche de segmentation des poumons sur les images CT. L'apprentissage profond supervisé est utilisé pour répondre au défi que constitue la segmentation de poumons peu contrastés, car présentant des lésions denses. L'accent est mis sur l'importance des données et de la manière dont celles-ci sont présentées au modèle lors de l'entraînement. Divers aspects de la gestion des données, tels que le transfert d'apprentissage, les détails du contexte, la diversité ou encore la pertinence de l'information traitée, sont explorés en utilisant des architectures de U-net 2D ou 3D. Enfin, dans le contexte pré-clinique, un modèle de production est proposé pour la segmentation des poumons de patients avec SDRA afin d'améliorer leur prise en charge par le réglage personnalisé de la ventilation mécanique.

Abstract

Acute respiratory distress syndrome (ARDS) induces a disturbance in the vital balance between oxygen intake and carbon dioxide output from the respiratory system. Indeed, alveoli composing the lungs are collapsed or filled with fluid in case of ARDS, impeding the normal ventilation process. To maintain patients diagnosed with ARDS alive, mechanical ventilation is routinely used. The latter restores the oxygen-carbon dioxide balance, but can cause ventilation induced lung injuries (VILI). In order to reduce VILI and to provide patient-specific management, aeration quantification is necessary. Computed tomography (CT) image, already used in the clinical diagnostic process, allows this quantification because it contains density information. Yet, preliminary lung segmentation is needed. This thesis proposes to automate lung segmentation on CT images. Supervised deep learning is used to address the challenge of segmenting poorly contrasted lungs, caused by the presence of high density lesions. Focus is made on data prevalence and on the way of presenting them to the model during training. Hence, various aspects of data management, as transfer learning, amount of details in the context, diversity or relevance, are explored with 2D or 3D U-net architectures. Eventually, in the pre-clinical context, a 3D production model is provided to segment lungs of patients with ARDS and therefore improve their care through personalized mechanical ventilation setting.

Table des matières

Remerciements	iii
Résumé	v
Abstract	v
Table des matières	vii
Introduction	2
1 Contexte médical et enjeux	4
1.1 Le système respiratoire humain	4
Fonctions	4
Anatomie	5
La ventilation pulmonaire	6
1.2 Le syndrome de détresse respiratoire aiguë (SDRA)	8
Description du contexte clinique	8
Prise en charge	10
SDRA résultant de la maladie COVID-19	14
1.3 Conclusion	14
2 Etat de l'art	16
2.1 La segmentation en traitement d'images	16
L'image numérique	17
La segmentation	19
Évaluer la segmentation	20
2.2 Les méthodes traditionnelles	24
2.3 Les méthodes par apprentissage profond	26
Le principe	26
Les différents modèles	28
Les méthodes spécifiques au SDRA	30
Les bases de données publiques de poumons segmentés	33
2.4 Conclusion	33
3 Approche 2D : gestion de données	35
3.1 Introduction	35
3.2 Données	37
3.3 Description d'expérience	40
Modèle	40

Stratégie d'entraînement	40
Évaluation	42
3.4 Résultats	43
Entraînements de référence	43
Entraînements combinés	44
Entraînements avec transfert d'apprentissage	45
Résultats globaux	46
3.5 Discussion	47
3.6 Conclusion	52
4 Approche 3D : exploiter l'information dans les données	54
4.1 Introduction	54
4.2 Données	55
4.3 Description d'expérience	59
Plan d'expérience	59
Modèle, entraînement et inférence	63
Évaluation	64
4.4 Résultats	66
Expériences sur l'initialisation	66
Expériences sur la quantité d'information	69
Expérience sur l'apport d'information physiologique	73
4.5 Discussion	77
4.6 Conclusion	82
5 Modèles de production	84
5.1 Introduction	84
5.2 Données	85
5.3 Méthode	87
Modèles	88
Intégration au logiciel de segmentation semi-automatique : creaSDRA	88
Évaluation	89
5.4 Performances	89
5.5 Discussion	93
5.6 Conclusion	94
Conclusion	96
Liste des communications	99
Bibliographie	100

Liste des figures

1.1	Plans de coupes et directions anatomiques de référence.	5
1.2	Zone de conduction du système respiratoire.	5
1.3	Situation des alvéoles pulmonaires dans le poumon (zone respiratoire).	6
1.4	La cage thoracique et la plèvre.	6
1.5	Vue sagittale de la cage thoracique à l’inspiration et l’expiration.	7
1.6	Volumes pulmonaires et capacités volumiques respiratoires.	7
1.7	Radiographies diagnostiques (vue coronale).	9
1.8	TDM diagnostiques (vue axiale).	9
1.9	Schéma des différentes lésions alvéolaires possibles entre l’inspiration et l’expiration. (A) la surdistension (B) contraintes entre alvéoles aérées et non aérées (C) recrutement cyclique.	11
1.10	Principe du scanner CT.	12
1.11	Coupe axiale annotée d’une image CT des poumons (patient avec SDRA). Encerclé, une calcification dans le cœur.	13
1.12	Coupe coronale d’une image CT des poumons (patient avec SDRA).	13
1.13	Coupe sagittale d’une image CT des poumons (patient avec SDRA).	13
1.14	Illustration de lésions et mécanismes associés sur une coupe axiale d’image CT, pour un patient atteint du SDRA.	13
1.15	Aération quantifiée sur coupe axiale.	14
2.1	Représentation d’une image numérique 2D de taille $M \times N$	17
2.2	Formation d’une image par réflexion de la source (ici la lumière).	17
2.3	Formation d’une image CT par transmission des rayons X par l’objet 3D.	17
2.4	Exemple d’une image numérique 2D de taille 7×8 dans laquelle les pixels sont séparés par des lignes blanches pour cette représentation.	18
2.5	Différences entre les tâches de classification, localisation, détection et la segmentation sémantique.	19
2.6	Processus classique du traitement d’images	20
2.7	Illustration de la segmentation de référence sur une image CT de poumons.	20
2.8	Illustration d’une bonne segmentation S_b (à gauche) et d’une mauvaise segmentation S_o (à droite), labellisée selon les quatre catégories de la matrice de confusion.	21
2.9	Intervalle de valeur des métriques selon leurs catégories – En vert les valeurs d’une bonne segmentation, en rouge celles d’une mauvaise segmentation	21
2.10	Voisinage (orange) du pixel (bleu) – (à gauche) 4-voisins (à droite) 8-voisins.	24
2.11	Relation entre les domaines de l’intelligence artificielle, l’apprentissage automatique et les réseaux de neurones.	26
2.12	Représentation d’un réseau neuronal à trois couches, dont une couche cachée.	27
2.13	Représentation d’un neurone.	27
2.14	Effet d’un filtre de convolution 2D sur le pixel d’une image en entrée	27
2.15	Effet d’un filtre de convolution 3D sur le voxel d’une image 3D en entrée	28
2.16	Découpage de données en trois sous-groupes : celui d’entraînement, de validation (V) et de test.	28
2.17	Détail d’une couche de convolution dans un CNN.	29
2.18	Architecture d’un réseau CNN.	29
2.19	Architecture d’un réseau fCNN.	29

2.20	Architecture d'un encodeur-décodeur.	30
2.21	Représentation simplifiée de l'architecture U-net (seules deux couches de convolutions représentées sur les quatre).	30
2.22	Compte des articles internationaux contenant les mots <i>deep learning</i> , <i>lung segmentation</i> et <i>CT</i> entre 2016 et 2022 (Source : PubMed®, 28/08/22).	30
2.23	Compte des articles internationaux contenant les mots <i>deep learning</i> , <i>lung segmentation</i> , <i>CT</i> et <i>ARDS</i> entre 2016 et 2022 (Source : PubMed®, 28/08/22).	30
2.24	Compte des articles internationaux contenant (<i>deep learning</i> , <i>lung segmentation</i> , <i>CT</i>) et (<i>ARDS</i> ou <i>COVID</i>) entre 2016 et 2022 (Source : PubMed®, 28/08/22).	31
2.25	Architecture ConvNet proposée par Gerard et al. [Ger18].	31
2.26	Illustration de différentes techniques d'échantillonnage d'une image permettant de réduire la mémoire nécessaire pour l'entraînement d'un modèle d'apprentissage profond. – Résolution : nombre de points par unité de longueur physique; Etendue : champ de vue; Mémoire : espace mémoire en machine. – Image reproduite depuis [Ger18].	32
3.1	Vue d'un même cochon avec modèle de SDRA à 4 niveaux d'atteintes différents (PK HIDRA).	37
3.2	Vue d'un même cochon avec modèle de SDRA à 4 niveaux d'atteintes différents (Hamilton).	38
3.3	Vue de 2 patients avec SDRA en fin d'expiration (en haut) et en fin d'inspiration (en bas) de la base de données CT4-ARDS.	39
3.4	Illustration de la segmentation semi-automatique avec le masque semi-automatique en rouge (à gauche) et les corrections manuelles en bleu (à droite).	39
3.5	Schéma du modèle multi-coups utilisé avec, pour chaque axe de coupe, un U-net 2D modifié comme décrit dans la section Modèle.	41
3.6	Résultats des modèles de référence testés sur 3 sous-ensembles de la base de données CT4-ARDS.	43
3.7	Courbes d'entraînement du modèle $A.3^{C4A}$ sur les 5 sous-ensembles de sa base de données d'entraînement.	44
3.8	Résultats des modèles entraînés sur des combinaisons de bases de données, testés sur la base CT4-ARDS en validation croisée sur 3 sous-ensembles.	44
3.9	Courbes d'entraînement du modèle $B.4^{\Omega}$ sur sa base de données d'entraînement.	45
3.10	Résultats des modèles entraînés avec transfert d'apprentissage testés sur la base CT4-ARDS.	45
3.11	Courbes d'entraînement du modèle $C.3^{PK}$ sur la base de données CT4-ARDS (en validation croisée de 3 sous-ensembles).	46
3.12	Résultats de tous les modèles sur les 3 sous-ensembles de la base de données CT4-ARDS.	47
3.13	Exemples des erreurs observées dans les pires et les meilleures segmentations des modèles de référence.	48
3.14	Exemple des pires (en rouge) et des meilleures (en violet) segmentations des modèles $B.n$ sur la base de données CT4-ARDS, superposées avec la segmentation manuelle de référence en bleu.	50
3.15	Exemple des pires (en rouge) et des meilleures (en violet) segmentations des modèles $C.n$ sur la base de données CT4-ARDS, superposées avec la segmentation manuelle de référence en bleu.	51
3.16	Décomposition de la prédiction du vote majoritaire superposé avec la segmentation manuelle (en bleu).	52

4.1 Exemple de deux patients de la base CT4-ARDS (vue axiale, coronale et sagittale) avec leur segmentation manuelle en bleu – Sur la première ligne, un patient avec SDRA modéré et un épanchement pleural indiqué par la flèche sur la coupe axiale; sur la deuxième ligne, un patient avec SDRA sévère. . . .	56
4.2 Exemple de patients avec SDRA modéré, sévère ou sévère sous ECMO extrait de la base de données CT4-ARDS2 (Lyon-2).	57
4.3 Exemple de deux patients de la base CT4-ARDS2 (Grenoble-2) avec SDRA modéré ou sévère.	58
4.4 Arborecence des fichiers <i>hdf5</i> de description des bases de données.	59
4.5 Coupe axiale à trois résolutions différentes, illustrant la perte de détails lors du sous-échantillonnage. – Pour des voxels isotropes de taille 1 mm (gauche), 2 mm (centre) et 3 mm (droite).	61
4.6 Coefficients γ dans les trois mélanges de fonction de perte évalués.	63
4.7 Schéma de l'architecture du U-net 3D utilisé. – Chaque pavé représenté correspond à la sortie d'une couche de convolution composée de deux convolutions et d'une étape de réduction (ou augmentation) de dimension pour l'encodeur (ou le décodeur). Le nombre de filtres de chaque sortie de couche de convolution est indiqué sous chaque pavé.	64
4.8 Performances des modèles D_x évaluant le transfert d'apprentissage en validation croisée sur la base de données Lyon-1.	66
4.9 Dice par compartiment d'aération segmentés et leur proportion dans le volume total pour les modèles D_x sur la base de données Lyon-1 en validation croisée.	67
4.10 Tracé Bland-Altman du biais de recrutement en validation croisée sur la base de données Lyon-1.	68
4.11 Performances des modèles E_r^β en validation croisée sur la base de données Lyon-1. – Les résolutions d'entraînement 1 mm, 2 mm et 3 mm sont respectivement symbolisées par les couleurs bleu, vert et violet.	69
4.12 Performances détaillées sur les compartiments d'aération des modèles E_r^β sur la base de données Lyon-1 en validation croisée.	71
4.13 Tracés de Bland-Altman illustrant le biais entre les modèles E_r^β et la référence pour le calcul du recrutement, en validation croisée sur la base de données Lyon-1. – σ l'écart-type des différences.	72
4.14 Performances globales des modèles F évalués en validation croisée sur la base de données Lyon-1 avec les métriques géométriques.	73
4.15 Performances détaillées sur les compartiments d'aération des modèles F en validation croisée sur la base de données Lyon-1.	74
4.16 Tracé Bland-Altman du biais de recrutement en validation croisée sur la base de données Lyon-1.	75
4.17 Courbes d'entraînement du modèle F_α et tracé des performances du Dice global et du Dice NA sur les 5 sous-ensembles de la base de données Lyon-1. – La fonction de perte utilisée utilise les deux termes de Dice comme défini dans l'équation Equation 4.2	76
4.18 Courbes d'entraînement du modèle F_{masse1} sur les 5 sous-ensembles de la base de données Lyon-1.	76
4.19 Courbes d'entraînement du modèle F_{masse2} sur les 5 sous-ensembles de la base de données Lyon-1.	76
4.20 Courbes d'entraînement du modèle F_{masse3} sur les 5 sous-ensembles de la base de données Lyon-1.	76
4.21 Image CT quantifiée. – HI pour le compartiment hyperinflaté.	78

4.22 Exemples des pires, moyennes et meilleures segmentations produites par les modèles D_x en validation croisée sur la base de données Lyon-1. – En bleu, les masques de segmentation de référence, en rouge les échecs de segmentations, en orange les segmentations moyennes et en vert les meilleures segmentations.	79
4.23 Exemples de la pire, la meilleure et une segmentation moyenne prédite par le modèle D_t sur les bases de données de test Lyon-2 et Grenoble-2.	80
4.24 Exemple de segmentation prédite par le modèle E_3^2 sur une image de test de la base de données Lyon-1. – En bleu le masque de segmentation de référence, en blanc la segmentation prédite.	80
4.25 Exemple de segmentation prédite par le modèle E_2^8 sur une image de test de la base de données Lyon-1. – En bleu le masque de segmentation de référence, en blanc la segmentation prédite.	80
4.26 Exemple de la pire segmentation du modèle F_{masse1} sur une image de test de la base de données Lyon-1. – En bleu la segmentation de référence, en rouge la prédiction du modèle F_{masse1}	81
5.1 Exemple de trois cochons pris dans les bases de données Hamilton et PK HIDRA.	86
5.2 Exemples de patients présents dans la base de données d’entraînement Lyon-12E avec leurs segmentations de référence. – Sur la première ligne un patient avec SDRA modéré et un pneumothorax, sur la deuxième un patient avec SDRA sévère sans ECMO et sur la dernière un patient avec SDRA sévère sous ECMO.	87
5.3 Organigramme des tâches réalisées à l’interface (entrée et sortie) du logiciel de segmentation semi-automatique. En orange les tâches réalisées dans le logiciel, en bleu celles réalisées par le code fourni dans un environnement virtuel.	89
5.4 Performances géométriques des modèles $U-net_{SDRA}^{3D}$ et $U-net_{EA}^{2D}$ sur la base de données Lyon-2T.	91
5.5 Performances avec la métrique ASSD détaillées par niveau de sévérité du SDRA sur la base de données Lyon-2T.	91
5.6 Détail des performances par compartiment d’aération sur la base de données Lyon-2T.	91
5.7 Tracé de Bland-Altman sur la mesure du recrutement sur la base de données Lyon-2T, par les deux méthodes, en comparaison avec celle de référence.	91
5.8 Performances géométriques des modèles $U-net_{SDRA}^{3D}$ et $U-net_{EA}^{2D}$ sur la base de données Grenoble-2.	92
5.9 Performances avec la métrique ASSD détaillées par niveau de sévérité du SDRA sur la base de données Grenoble-2.	92
5.10 Détail des performances par compartiment d’aération sur la base de données Grenoble-2.	92
5.11 Tracé de Bland-Altman sur la mesure du recrutement sur la base de données Grenoble-2, par les deux méthodes en comparaison avec celle de référence.	92
5.12 Exemples de pires (en rouge) et meilleures (en vert) segmentations fournies par les deux modèles $U-net_{EA}^{2D}$ et $U-net_{SDRA}^{3D}$ sur les bases de données Lyon-2T et Grenoble-2. – La croix jaune sur l’image CT à la 2e ligne indique le début du diaphragme.	94

Liste des tableaux

1.1 Pressions durant la respiration.	7
1.2 Définition des niveaux de sévérité du SDRA.	8
2.1 Matrice de confusion comparant la segmentation à la vérité terrain	21
2.2 Bases de données publiques avec masques de segmentation des poumons.	33
3.1 Caractéristiques des bases de données disponibles	40
3.2 Description des expériences menées avec les combinaisons de bases de données utilisées à l'entraînement.	41
3.3 Répartition des sous-ensembles pour la validation croisée en k sous-ensembles.	42
3.4 Résultats des modèles de référence animaux sur les deux bases de données animales.	43
3.5 Résultats des modèles utilisant une combinaison de bases de données sur les deux bases de données animales.	45
3.6 Résultats des modèles $C.1^{Ha}$ et $C.3^{PK}$ utilisant le transfert d'apprentissage uniquement sur la base de données animales évalués en validation croisée sur ces deux bases (PK et Ha).	46
3.7 Comparaison des modèles en fonction de leur base de données d'entraînement et du type d'image utilisée.	47
4.1 Détails des bases de données CT4-ARDS et CT4-ARDS2 utilisées respectivement pour l'entraînement ou le test.	59
4.2 Modèles utilisés pour l'évaluation de l'initialisation.	60
4.3 Modèles E_r^β entraînés avec des images de résolution spatiale r et par lot de taille β	61
4.4 Seuils des compartiments d'aération selon leur densité en unité Hounsfield.	61
4.5 Performances globales des modèles D_x sur les bases de données de test Lyon-2 et Grenoble-2.	67
4.6 Performances sur les compartiments peu aéré et non aéré des modèles D_x sur les bases de données de test Lyon-2 et Grenoble-2.	68
4.7 Biais de recrutement (R) et de cohérence de masse (C) des modèles D_x sur les bases de données de test Lyon-2 et Grenoble-2.	68
4.8 Performances globales des modèles E_r^β sur les bases de données de test Lyon-2 et Grenoble-2.	71
4.9 Durées d'entraînement moyennes par modèle en heure. – La durée correspond à la moyenne des 5 modèles entraînés parallèlement pour la validation croisée.	73
4.10 Performances globales des modèles F_x sur les bases de données de test Lyon-2 et Grenoble-2.	74
4.11 Biais sur le recrutement (R) et la cohérence de masse (C) des modèles F_x sur les bases de données de test Lyon-2 et Grenoble-2.	75
5.1 Détails des bases de données CT4-ARDS et CT4-ARDS2 utilisées respectivement pour l'entraînement ou le test.	86

Liste des acronymes

AECC	<i>American-European Consensus Conference</i>
ASSD	<i>Average Symmetric Surface Distance</i>
CHU	Centre Hospitalo-Universitaire
CNN	<i>Convolutional Neural Network</i>
COVID-19	<i>Coronavirus disease 2019</i>
CPU	<i>Central Processing Unit</i>
CREATIS	Centre de REcherche en Acquisition et Traitement de l'Image pour la Santé
CT	<i>Computed Tomography</i>
DH	Distance de Hausdorff
DH-95	95 ^e centile de la distance de Hausdorff
ECMO	<i>Extracorporeal membrane oxygenation</i>
FN	Faux Négatif
FP	Faux Positif
GPU	<i>Graphical Processing Unit</i>
hdf	<i>hierarchical data format</i>
IRM	Imagerie par Résonance Magnétique
LIVM	Lésions pulmonaires Induites par la Ventilation Mécanique
PEP	Pression Expiratoire Positive
SDRA	Syndrome de Détresse Respiratoire Aiguë
TDM	Tomodensitométrie
UH	Unité Hounsfield
VC	Volume Courant
VN	Vrai Négatif
VP	Vrai Positif
2D	Deux dimensions
3D	Trois dimensions

Introduction

L'expression « les poumons de la Terre » est souvent utilisée pour désigner ces écosystèmes qui sont le siège d'échanges gazeux entre dioxyde de carbone et oxygène. Cette métaphore fait écho au rôle essentiel que jouent les poumons dans le corps humain. Au centre du système respiratoire, les poumons sont le lieu des échanges gazeux entre l'oxygène fraîchement apporté de l'air extérieur aux cellules du corps et le dioxyde de carbone à expulser provenant du sang. La quantité d'oxygène qui parvient aux cellules, grâce à l'air inspiré, et celle de dioxyde de carbone qui est rejetée via l'air expiré, permet de maintenir un équilibre vital¹. Lorsque cet équilibre est rompu et que le système respiratoire² ne parvient pas à le rétablir, le pronostic vital est engagé. C'est ce qui survient dans le cadre du syndrome de détresse respiratoire aiguë (SDRA).

En effet, ce syndrome désigne un dysfonctionnement pulmonaire sévère caractérisé par une hypoxémie profonde et un processus inflammatoire, entre autres. Les patients atteints du SDRA sont ainsi pris en charge dans les services de réanimation respiratoire. Un traitement de support leur est apporté pour réduire l'hypoxémie en les plaçant sous ventilation mécanique, pendant que la cause du SDRA est cherchée puis traitée. L'imagerie tomodensitométrique (ou CT pour *Computed Tomography*) est souvent utilisée pour la confirmation diagnostique et permet d'observer finement les lésions présentes dans les poumons. Les lésions pulmonaires observées sont le plus souvent des atélectasies³ ou des œdèmes pulmonaires⁴ qui sont alors des zones de fortes densités (relativement aux tissus sains) au sein du poumon. Le ventilateur mécanique, réglé par le réanimateur, injecte de l'air dans les poumons avec comme objectif d'aérer les zones lésées, qui sont comme expliqué non aérées. Cependant, en fonction des lésions du patient, la ventilation mécanique peut causer des lésions supplémentaires. Ceci survient par exemple lorsque des zones déjà aérées sont hyperinflatées.

Afin d'éviter ces lésions qui causent une sur-mortalité, un réglage personnalisé de la ventilation mécanique est une perspective prometteuse. Pour cela, la quantification de l'aération de chaque patient est nécessaire. Cette dernière est disponible par analyse des images CT, mais demande au préalable de délimiter précisément dans l'image CT la zone des poumons.

C'est à ce niveau qu'intervient le travail de cette thèse, dont l'objectif est de fournir la segmentation des poumons de patients atteints du SDRA sur des images CT.

Le Chapitre 1 permet de présenter en détail le contexte médical dans lequel s'intègre ce travail de thèse. Il définit dans un premier temps le système respiratoire pour en comprendre le fonctionnement. Puis dans un deuxième temps, le syndrome de détresse respiratoire aiguë, la compréhension de ses mécanismes et sa prise en charge sont expliqués. Ainsi, ce premier chapitre contextualise et précise les attentes de cette thèse en termes de segmentation, notamment les contraintes liées à la pratique clinique qui ont orienté la méthodologie adoptée.

1: Appelé homéostasie.

2: De paire avec le système cardiaque.

3: Partie du poumon qui se vide d'air, qui est collapsée.

4: Accumulation de liquide dans les alvéoles pulmonaires.

Le Chapitre 2 apporte les définitions nécessaires pour la compréhension des techniques de traitement d'images utilisées pour répondre aux objectifs énoncés. Les difficultés de la segmentation sur cette application sont présentées via une synthèse de l'état de l'art dans le domaine. En effet, les méthodes traditionnelles ne parviennent pas à dépasser la limite de perte de contraste des poumons dans les cas les plus sévères du SDRA. L'apprentissage profond est identifié selon la littérature comme capable de résoudre des tâches difficiles. C'est pourquoi utiliser l'apprentissage profond a été choisi pour répondre au problème technique de cette thèse.

Les bases de données structurées⁵ constituent un élément essentiel dans l'entraînement et l'efficacité des modèles d'apprentissage profond. Cette thèse s'attache particulièrement à évaluer quelles stratégies permettent d'exploiter de manière optimale les données à disposition.

Le Chapitre 3 décrit la première expérience menée en ce sens sur une architecture multi-2D. Différentes bases de données⁶ sont utilisées pour identifier s'il existe une gestion optimale de présentation de ces dernières aux modèles durant l'apprentissage. Le transfert d'apprentissage est notamment évalué entre apprentissage sur anatomie porcine vers l'anatomie humaine.

Le Chapitre 4 poursuit la démarche en évaluant d'autres caractéristiques des données avec une architecture 3D. Le contexte apporté par l'utilisation de l'architecture tridimensionnelle permet d'étudier l'impact d'éléments tels que la diversité des contextes volumiques considérés ou encore l'ajout d'information quantitative physiologique⁷.

Enfin, le Chapitre 5 présente les modèles produits pour la segmentation des poumons de cochons, d'une part, et de patients atteints du SDRA, d'autre part, et intégrés dans un logiciel de segmentation semi-automatique. Leur utilisation dans un contexte pré-clinique est présentée.

La volonté de rendre cette thèse didactique est illustrée par l'utilisation d'encadrés colorés permettant de mettre en évidence les éléments importants à la compréhension du lecteur. En particulier, les encadrés bleus servent à résumer les informations importantes décrites dans le texte qui les précède. Les encadrés orange servent quant à eux à donner une idée numérique des valeurs des concepts cités dans le texte principal.

5: Les images et leurs annotations.

6: Animales et humaines.

7: Telle que la masse ou la réaération des zones non aérées.

Résumé des informations à retenir

Ce type d'encadré est utilisé pour mettre en évidence des concepts importants pour la compréhension de cette thèse comme :

- des mots clés,
- des définitions,
- des objectifs,
- l'état d'avancement en fin d'expérience.

Exemple numérique

Ce type d'encadré sert à fournir des informations numériques, soit sur des concepts présentés ou bien en tant que rappel pour appuyer des interprétations.

Ce chapitre présente le contexte médical dans lequel se place cette thèse afin d'en faire ressortir les enjeux et motivations. Dans un premier temps, une description du système respiratoire est présentée qui permet de comprendre le fonctionnement et l'importance de ce système pour le corps humain. Dans un deuxième temps, les atteintes de la pathologie ciblée : le syndrome de détresse respiratoire aiguë (SDRA), sont détaillées au regard de la première partie. Enfin, l'exposé des diverses techniques de prise en charge actuelles met en lumière les limitations rencontrées en clinique pour le traitement de ce syndrome. Les objectifs cliniques de la thèse sont alors décrits.

Table des matières

1.1	Le système respiratoire humain	4
	Fonctions	4
	Anatomie	5
	La ventilation pulmonaire	6
1.2	Le syndrome de détresse respiratoire aiguë (SDRA)	8
	Description du contexte clinique	8
	Prise en charge	10
	SDRA résultant de la maladie COVID-19	14
1.3	Conclusion	14

Les descriptions anatomiques, fonctionnelles, mécaniques et physiologiques de cette section s'appuient sur divers livres et cours du domaine médical [Tor+00; Net11; Gor+13; Sar13; Dup15]

1.1 Le système respiratoire humain

Le corps humain peut être divisé en trois grands systèmes¹ qui remplissent les fonctions suivantes : (1) la nutrition, (2) la relation et (3) la reproduction. Le système respiratoire fait partie du groupe servant la fonction de nutrition². Le rôle du système respiratoire sera défini dans un premier temps. Les organes le composant seront présentés dans un second temps. Enfin, le fonctionnement de ce système sera expliqué.

1: Un système biologique est composé des cellules, tissus et organes qui remplissent une fonction commune.

2: Entre autres avec les systèmes digestifs, circulatoire, lymphatique et excréteur.

Fonctions

Le système respiratoire assure un rôle essentiel dans le corps humain puisqu'il apporte l'oxygène (O_2) présent dans l'air au réseau vasculaire et évacue le dioxyde de carbone (CO_2) présent dans le sang via l'air expiré³. Ceci est réalisé en deux étapes : la première est anatomique, c'est la ventilation et la deuxième est physiologique, c'est la diffusion. La ventilation a pour but d'acheminer l'air dans les bonnes conditions jusqu'à la zone d'échange gazeux où pourra avoir lieu la diffusion. Cette dernière est un phénomène vital puisque l'oxygène (en réalité le dioxygène), qui est véhiculée dans le corps à travers le sang, est une molécule nécessaire au fonctionnement cellulaire. De plus, la privation d'oxygène⁴ peut mener

3: Ce processus est appelé l'hématose.

4: Ou plus simplement l'hypoxémie, le faible taux d'oxygène dans le sang.

à des lésions irréversibles et être létale pour les cellules. La réaction chimique dans laquelle est impliqué l'oxygène au sein des cellules génère du dioxyde de carbone, qui lui est toxique pour l'organisme. Alors le système respiratoire, de pair avec le système cardio-vasculaire, doit garantir l'équilibre adéquat entre la quantité d'oxygène apportée et celle de dioxyde de carbone évacuée du corps. Cela se traduit par l'ajustement, d'un côté, du débit cardiaque, et de l'autre côté, de la fréquence respiratoire ou du volume de gaz expiré. Le premier permet d'adapter le flux circulatoire du sang et le deuxième permet d'adapter les échanges gazeux et ainsi correspondre aux besoins du corps à tout moment. Cet état d'équilibre vital est appelé l'homéostasie respiratoire et se traduit par l'harmonie de plusieurs paramètres sanguins, notamment la quantité de gaz (i.e. O_2 , CO_2) dans le sang ou encore le pH sanguin. Le bon déroulement des échanges gazeux est également garanti par les muqueuses qui tapissent les parois du système respiratoire. Ces dernières assurent la filtration de l'air conduit jusque dans les poumons en même temps que son humidification et son réchauffement.

En plus du maintien de l'homéostasie, le système respiratoire sert d'autres fonctions non vitales comme l'odorat ou la production de son (la parole).

Anatomie

Avant de décrire l'anatomie du système respiratoire, définir les différents termes du repérage anatomique dans l'espace est nécessaire. Les positions et axes anatomiques de référence sont présentés dans la Figure 1.1 (le corps de gauche représente la position standard). Les termes cranial et caudal sont aussi respectivement appelés supérieur et inférieur. Les plans de coupes de référence dans lesquels sont observés ou représentés les organes y sont également illustrés. Il s'agit des plans axial, coronal et sagittal qui sont définis par rapport à la position standard et sont orthogonaux entre eux.

Le système respiratoire s'étend du nez, au niveau de la tête, jusqu'au diaphragme, au centre de l'abdomen. On peut le séparer en deux zones : la zone de circulation et la zone respiratoire. La zone de circulation sert uniquement à transporter l'air extérieur jusqu'à la zone d'échange gazeux (la zone respiratoire) dans les poumons. C'est un réseau de voies aériennes principalement droites (voir Figure 1.2) qui permettent une libre circulation de l'air. Dans sa partie supérieure, il commence par les narines et les cavités nasales, communique avec la bouche par le pharynx jusqu'au larynx. La partie inférieure de la zone de circulation commence à la trachée qui se subdivise ensuite successivement en un arbre bronchique, composé de bronches primaires, secondaires et tertiaires, jusqu'aux bronchioles terminales⁵. La zone de circulation se termine aux bronchioles terminales et comme elle ne participe pas aux échanges gazeux, elle est appelée l'espace mort anatomique⁶.

À partir de la fin des deux bronches principales, l'arbre bronchique est situé à l'intérieur des poumons. Siège de l'hématose, les poumons constituent l'organe principal de l'appareil respiratoire. Au niveau des bronchioles terminales commence la zone de transition via les bronchioles respiratoires où vont commencer les échanges gazeux (environ le 16^e niveau de ramification dans la Figure 1.3). Cette zone de transition comprend environ trois

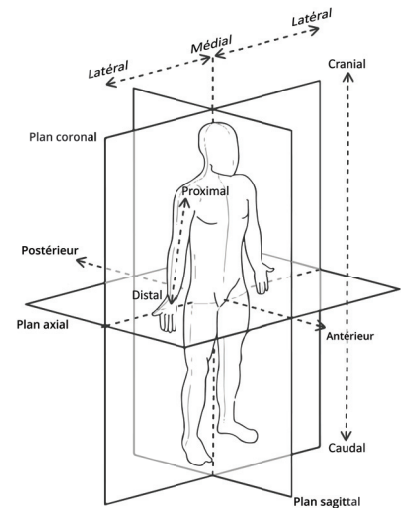


FIGURE 1.1 – Plans de coupes et directions anatomiques de référence. – Image traduite de Edoardo / CC-BY-SA-3.0.

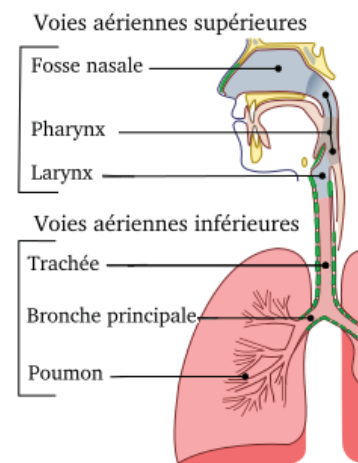


FIGURE 1.2 – Zone de conduction du système respiratoire. – Traduite depuis Jmarchn / CC-BY-SA-3.0.

5: Il y a 15 niveaux de division des bronches principales aux bronchioles terminales.

6: Le volume qui ne participe pas aux échanges gazeux représente environ 150mL (2mL/kg).

niveaux de divisions jusqu'aux bronchioles respiratoires terminales. Attachés aux bronchioles respiratoires terminales par les canaux alvéolaires, se trouvent les sacs alvéolaires composés d'alvéoles (voir cercle dans la Figure 1.3). C'est cette partie des poumons qu'on appelle le parenchyme respiratoire, car il s'agit du tissu fonctionnel des poumons. Les parois alvéolaires sont recouvertes d'un film d'eau et sont élastiques. Elles s'étirent lorsque l'air pénètre la zone respiratoire, afin d'augmenter la surface des échanges gazeux. Ceci est également facilité par la présence d'un liquide recouvrant l'intérieur des alvéoles qui empêche qu'elles ne s'affaissent : le surfactant⁷. Les alvéoles communiquent entre elles par des pores qui garantissent une pression égale dans les alvéoles et le poumon. Comme l'illustre la Figure 1.3, de petits vaisseaux sanguins (les capillaires) entourent les alvéoles, ce qui permet un échange gazeux entre le sang et l'air présent dans les poumons. Les capillaires sont la jonction de deux systèmes de circulation sanguine au sein des poumons. Le premier, composé des artères et veines pulmonaires, est fonctionnel et participe à l'oxygénation du sang. L'artère pulmonaire achemine le sang chargé en CO_2 jusqu'aux poumons, c'est la fonction de perfusion, tandis que les veines pulmonaires font circuler le sang fraîchement oxygéné. Le second système, composé des artères et des veines bronchiques, est nutritif pour les organes. Le sang oxygéné est amené aux poumons par l'artère bronchique puis repart au cœur par les veines bronchiques.

Tous ces vaisseaux communiquent avec le cœur qui est logé entre les deux poumons. Autour du cœur, on trouve aussi l'œsophage, la trachée et les deux bronches principales dont une partie rentre dans les poumons. Cet ensemble d'organes est appelé le médiastin. Il est touché par la membrane qui entoure les poumons. Il s'agit de la plèvre pariétale qui adhère à la face interne de la cage thoracique (les côtes). La face externe des poumons est quant à elle enveloppée par une autre membrane : la plèvre viscérale. Ces deux membranes forment la cavité pleurale, à l'intérieur de laquelle se trouve le fluide pleural (voir Figure 1.4). Ce dernier facilite le glissement entre les deux plèvres lors des mouvements de respiration. Les plèvres gauches et droites (associées respectivement au poumon gauche et droit) sont séparées par le médiastin.

Tous les organes du système respiratoire, à partir de la trachée, sont entourées par la cage thoracique et les muscles respiratoires (muscles intercostaux et diaphragme). Le diaphragme est le dernier organe qui marque la limite inférieure de ce système⁸. C'est un muscle plat qui participe majoritairement à l'inspiration, l'une des étapes de la ventilation.

La ventilation pulmonaire

Le processus de ventilation pulmonaire spontané (au repos), qui permet à l'air de rentrer dans les poumons à l'inspiration et d'en ressortir à l'expiration, est en partie régi par des différences de pression. Il est divisé en deux étapes : (1) l'inspiration et (2) l'expiration. Durant la première étape, l'air rentre dans les poumons sous l'action des muscles inspiratoires, c'est un processus actif. La deuxième étape est alors passive puisqu'elle correspond au relâchement de ces mêmes muscles. La pression atmosphérique (P_{atm}) sert de référence dans la respiration, ainsi une pression respiratoire qui lui est égale vaut zéro. Tandis qu'une pression plus faible que la pression atmosphérique sera négative et une pression plus élevée sera positive. Afin

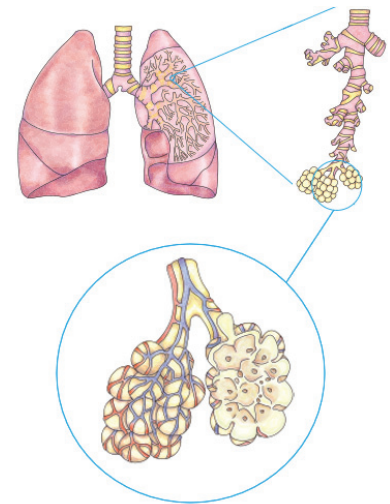


FIGURE 1.3 – Situation des alvéoles pulmonaires dans le poumon (zone respiratoire). – Image modifiée à partir de DataBase Center for Life Science (DBCLS) / CC-BY-4.0

7: Il diminue la tension superficielle dans les alvéoles, ce qui permet lorsque leur taille décroît (à l'expiration) qu'elles ne s'affaissent pas.

Caractéristiques moyennes des poumons

La masse moyenne des deux poumons est de 1.2kg. Leur capacité respiratoire est d'environ 5-6L. À tous moments ils contiennent 1/4 de la masse sanguine circulante.

8: Il marque la limite entre les cavités thoraciques (qui contiennent les poumons) et abdominales

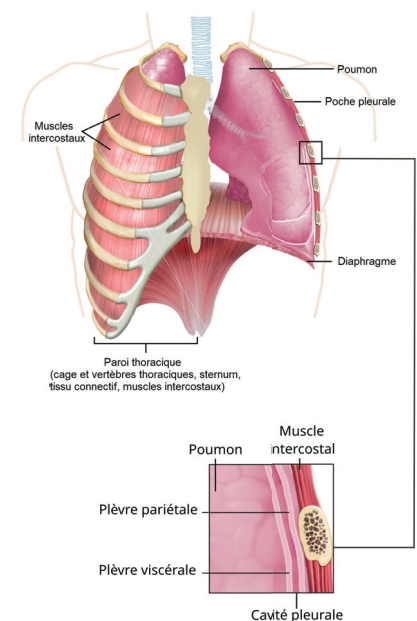


FIGURE 1.4 – La cage thoracique et la plèvre. – Image modifiée de OpenStax College / CC-BY-3.0

de créer des gradients de pression entre les différentes zones du système respiratoire, trois pressions sont à considérer : la pression atmosphérique, la pression intra-alvéolaire (P_{alv} , ou pression intra-pulmonaire) et la pression intra-pleurale⁹ (P_{pl}).

À l'inspiration, un influx nerveux en provenance des neurones inspiratoires provoque la contraction des muscles de la respiration. Le diaphragme déclenche en grande partie (75%) l'entrée de l'air dans les poumons. En effet, en se contractant¹⁰, ce muscle agrandit la cage thoracique de paire avec les muscles intercostaux externes qui augmentent le diamètre antéro-postérieur (voir Inspiration sur la Figure 1.5). Le volume étant agrandi, la pression intra-alvéolaire (égale à P_{atm} au repos) diminue¹¹ et devient négative (voir Tableau 1.1). Au même moment, la pression intra-pleurale diminue également du fait de l'étirement de la cage thoracique à laquelle est attachée la plèvre pariétale. Les poumons, enveloppés dans la plèvre viscérale, suivent alors le mouvement d'élargissement dans le même sens. Tant que cette différence de pression entre l'air extérieur et les poumons (la pression transpulmonaire), provoquée à l'inspiration, existe et est positive, l'air rentrera dans les poumons.

Lorsque les muscles respiratoires se relâchent, la phase d'expiration commence alors. En effet, l'influx nerveux qui déclenche l'inspiration est simplement arrêté, ce pourquoi l'expiration est dite passive, ce qui provoque le relâchement des muscles. Le volume de la cage thoracique diminue alors selon les axes antéro-postérieur et vertical. Les tissus pulmonaires n'étant plus soumis à une force qui les tire vers l'extérieur, leur élasticité leur fait reprendre leur volume initial. La pression intra-alvéolaire augmente alors, dépassant la pression atmosphérique. Le gradient de pression est inversé et l'air est alors expulsé vers l'extérieur du corps.

Les étapes précédemment décrites correspondent à la respiration dite calme, lorsque le corps est au repos. Selon le mode de respiration (calme, effort, forcée), l'apport des divers muscles respiratoires sera plus ou moins important. La capacité des poumons à se gonfler dépendra aussi de la compliance des parois thoraciques et de la résistance des voies respiratoires. La compliance pulmonaire, décrivant la facilité du poumon à se gonfler, dépend à la fois de l'élasticité du tissu pulmonaire et de la présence de surfactant dans le liquide alvéolaire. Cette dernière permet également d'éviter les phénomènes d'atélectasie, qui sont l'affaissement des alvéoles lorsque leur volume décroît (à l'expiration par exemple).

Pendant le processus de respiration, différents volumes d'air déplacé sont mis en jeu. Évaluer ces différents volumes permet de suivre l'état de santé d'une personne. Ceci est habituellement réalisé via un spiromètre. Durant la respiration calme, le volume d'air qui rentre dans les poumons est appelé le volume courant (VC). Les volumes maximaux qu'il est possible d'inspirer ou expirer sont respectivement notés VRI et VRE pour volume de réserve inspiratoire ou expiratoire. Enfin, même en cas d'expiration maximale, il reste toujours de l'air dans les poumons qui correspond au volume résiduel (VR). Tous ces volumes permettent de définir des capacités respiratoires qui sont illustrées dans la Figure 1.6. L'écart par rapport aux valeurs usuelles sera notamment indicateur d'un dysfonctionnement pulmonaire. À l'échelle des alvéoles, mesurer le rapport de la ventilation alvéolaire sur le débit sanguin des poumons est aussi informatif sur l'état de la fonction respiratoire.

9: Respectivement la pression dans les alvéoles et la pression dans la cavité pleurale.

10: L'amplitude de mouvement du diaphragme va de 1cm en respiration calme à près de 10cm en respiration profonde.

11: Voir la loi de Boyle.

TABLEAU 1.1 – Pressions durant la respiration.

	Pressions (mmHg)	
	Inspiration	Expiration
P_{alv}	-2	+2
P_{pl}	-6	-3

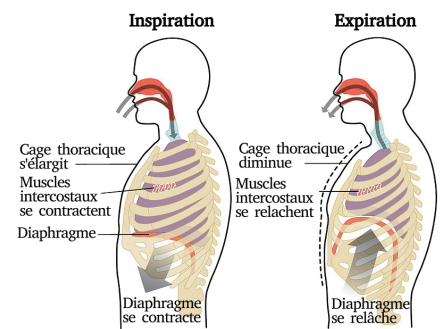


FIGURE 1.5 – Vue sagittale de la cage thoracique à l'inspiration et à l'expiration. – Image traduite de OpenStax College / CC-BY-3.0

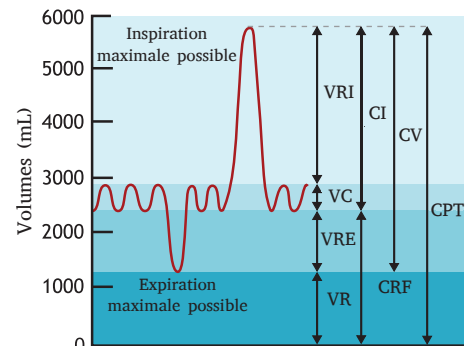


FIGURE 1.6 – Volumes pulmonaires et capacités volumiques respiratoires. – VRI : Volume de réserve inspiratoire, VC : Volume courant, VRE : Volume de réserve expiratoire, VR : Volume résiduel, CI : Capacité inspiratoire, CRF : Capacité résiduelle fonctionnelle, CV : Capacité vitale, CPT : Capacité pulmonaire totale. – Image traduite de Michal Komorniczak / CC-BY-SA-3.0

Respiration au repos en quelques chiffres

- Volume courant = 500mL
- Fréquence respiratoire : 12 cycles/min
- Espace mort anatomique : 150mL
- Air frais renouvelé (alvéoles) : 350mL
- Ventilation alvéolaire : 4200mL/min

1.2 Le syndrome de détresse respiratoire aiguë (SDRA)

Description du contexte clinique

Définition Le syndrome de détresse respiratoire aiguë (SDRA) désigne un dysfonctionnement pulmonaire sévère, manifesté par une perte de compliance pulmonaire, une hypoxémie profonde, une augmentation du shunt intra-pulmonaire¹² et de l'espace mort anatomique. C'est un processus inflammatoire qui affecte la barrière alvéolo-capillaire, diminue la quantité de surfactant et de ce fait réduit l'aération pulmonaire. L'inflammation, entre autres, cause une augmentation de la perméabilité vasculaire qui provoque une accumulation de liquide dans les alvéoles pulmonaires. Le poids et la densité du poumon augmentent alors, car celui-ci ne se gonfle plus autant qu'en temps normal.

La première définition du SDRA est datée de 1967 [Ash+67] puis révisée par la AECC (*American-European Consensus Conference*) en 1994, et finalement remaniée à nouveau en 2011 pour aboutir à la définition de Berlin [Ran+12] aujourd'hui utilisée.

Définition de Berlin du SDRA [Ran+12]

Délai d'installation	7 jours après l'agression pulmonaire ou l'aggravation des symptômes respiratoires.
Imagerie thoracique (Radiographie ou CT)	Opacités bilatérales, non expliquées totalement par l'effusion, l'atélectasie ou les nodules.
Origine de l'œdème	Défaillance respiratoire non expliquée totalement par l'insuffisance cardiaque ou le surplus de fluide.
Oxygénation	Rapport PaO_2/FiO_2 ¹³ qui permet de classer en trois niveaux de sévérité (voir Tableau 1.2) pour un patient ventilé avec une pression expiratoire positive (PEP) supérieure ou égale à $5\text{cmH}_2\text{O}$.

Au fil des années, les méthodes et outils disponibles pour le diagnostic ont évolué, cependant la modalité de référence reste la radiographie thoracique puisqu'elle est disponible au lit du malade. Le recours à l'imagerie tomodensitométrique a lieu dans des cas où les critères cités précédemment ne permettent pas d'établir un diagnostic clair. Cela s'explique par le fait que la radiographie peut parfois sous-estimer l'étendue des lésions ou bien être difficilement interprétable¹⁴.

Epidémiologie Selon la définition utilisée du SDRA, les incidences relayées par la littérature peuvent varier [Pap08; Pap+21]. L'étude multicentrique LUNG SAFE réalisée en 2016, a permis d'estimer plus précisément l'impact de ce syndrome dans la population mondiale. Parmi les 459 unités de soins intensifs¹⁵ (USI), 10.4% des patients et 23.4% des patients sous

12: Circulation anormale de sang d'un système circulatoire à un autre. Affecte les zones dont le rapport ventilation-perfusion est inférieur à 1.

13: PaO_2 est la pression partielle artérielle d'oxygène et FiO_2 la fraction d'oxygène dans l'air inspiré.

TABLEAU 1.2 – Définition des niveaux de sévérité du SDRA.

Sévérité	Intervalle PaO_2/FiO_2 (mmHg)
Léger	$200 < PaO_2/FiO_2 \leq 300$
Modéré	$100 < PaO_2/FiO_2 \leq 200$
Sévère	$PaO_2/FiO_2 \leq 100$

14: En cas d'obésité extrême ou d'anomalies anatomiques par exemple.

15: Provenant de 50 pays différents réparties sur 5 continents.

ventilation mécanique répondaient à la définition du SDRA selon la définition de Berlin [Bel+16]. Cela correspond à environ 3 millions de patients chaque année atteints de ce syndrome dans le monde. En France, ce nombre s'élevait à 24.6 pour 100 000 personnes.an¹⁶ en 2017 [Pap+21]. Cependant, il semble que ces chiffres soient tout de même sous-estimés, avec près de 40% des SDRA non diagnostiqués (20% des SDRA sévères) [Bel+16]. Cela peut alors mener à une prise en charge non adaptée (réglages ventilatoires non adaptés) en plus de la limitation matérielle [Pap08; Mey+21].

L'âge moyen des patients atteints du SDRA est de 61,5 ans¹⁷ et 62% d'entre eux sont des hommes [Bel+16].

Malgré la prise en charge, la mortalité hospitalière du SDRA reste élevée puisqu'elle se situe autour de 40% selon le niveau de sévérité¹⁸. Ajouté à cela les lésions pulmonaires induites par la ventilation mécanique (LIVM) qui causent une surmortalité [CIS15; Nie+17].

Étiologie Les causes du SDRA sont multiples et peuvent être classées selon qu'elles sont des agressions directes ou indirectes des poumons. Dans la première catégorie, les plus fréquentes sont la pneumonie infectieuse¹⁹ (bactérienne ou virale), l'inhalation de liquide gastrique, l'inhalation (gaz toxiques²⁰), la contusion pulmonaire et la noyade. Les causes extrapulmonaires sont communément le sepsis extrapulmonaire, les traumatismes, la pancréatite, les brûlures graves, l'overdose médicamenteuse et les transfusions multiples [Fan+13; Ben+18; Dia+21]. Présenter l'une de ces atteintes ne mène pas systématiquement au développement d'un SDRA. Les facteurs de risque que sont l'âge avancé, le sexe, la consommation d'alcool et le tabagisme peuvent aussi influencer l'apparition du SDRA [Dia+19]. L'identification de la cause spécifique du SDRA reste capitale dans la prise en charge thérapeutique de ce syndrome.

Physiopathologie²¹ Le SDRA provoque une atteinte de la membrane entre les alvéoles et les capillaires. Inflammation et œdème apparaissent alors dans les poumons par réponse inflammatoire et augmentation de la perméabilité capillaire, ce qui affecte le volume pulmonaire aéré. La capacité résiduelle fonctionnelle (voir CRF dans la Figure 1.6) est très fortement diminuée puisqu'elle passe de 2500mL en moyenne chez l'adulte sain à 500mL chez un patient atteint du SDRA. Ceci s'explique par les phénomènes d'atélectasie qui apparaissent de façon hétérogène dans les poumons (gradient ventro-dorsal) [CIS15]. Sont alors juxtaposées des alvéoles normalement aérées à d'autres non aérées, peu aérées ou surdistendues. La prise en charge, via ventilation mécanique, vise à homogénéiser l'aération à travers les poumons en ouvrant les zones collabées (recrutement) sans pour autant surdistendre les zones normalement aérées²².

Diagnostic Le diagnostic du SDRA s'appuie dans un premier temps sur les critères de la définition de Berlin (citée plus haut). Une série d'examen²³ et l'analyse de l'historique clinique permettent de répondre en partie à une suspicion diagnostique du clinicien. L'imagerie médicale, soit de radiographie (voir Figure 1.7), soit de tomodensitométrie (TDM, voir Figure 1.8) permet d'appuyer également le diagnostic positif. Enfin, l'identification des causes primaires et secondaires de l'agression alvéolo-capillaire est réalisée après le diagnostic différentiel²⁴.

16: L'unité personnes.an désigne le fait de prendre en compte une quantité de personnes pendant une durée d'une année. Par exemple, les données associées à 100 personnes suivies pendant 1 an sont de 100 personnes.an tandis que celles de 100 personnes suivies pendant 10 sont de 1000 personnes.an.

17: (62,8 ans en France avec 67,7% qui sont des hommes [Pap+21].

18: SDRA léger : 34.9%; SDRA modéré : 40.3%; SDRA sévère : 46.1% [Bel+16]

19: Pour 61.5% des patients en France [Pap+21].

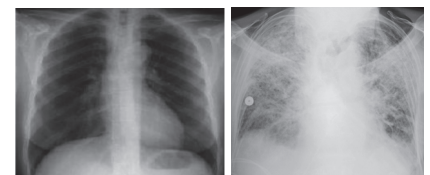
20: La cigarette électronique par exemple [Mey+21].

21: C'est l'étude des troubles fonctionnels causés par une affection particulière.

22: Ceci pour éviter les LIVM

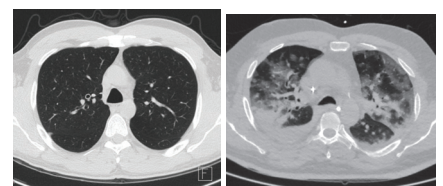
23: Pour exclure notamment les causes d'agressions non prévues par la définition de Berlin.

24: Exclusion des autres problèmes possibles pouvant expliquer les symptômes du patient.



(a) Poumons sains. (b) Poumons avec SDRA. – Image by James Heilman, MD, CC BY-SA 3.0

FIGURE 1.7 – Radiographies diagnostiques (vue coronale).



(a) Poumons sains. (b) Poumons avec SDRA.

FIGURE 1.8 – TDM diagnostiques (vue axiale).

Prise en charge

Le SDRA est pris en charge en soignant les symptômes pendant que la cause du syndrome est identifiée puis traitée. L'objectif principal consiste alors à améliorer l'oxygénation qui n'est plus assurée naturellement par le système respiratoire. Selon le niveau de sévérité d'atteinte du patient, plusieurs stratégies peuvent être adoptées. Parmi elles, la réduction de la consommation d'oxygène via la sédation et l'apport d'oxygène par ventilation non invasive²⁵ ou, au contraire, par ventilation mécanique [Sie+22]. Le réglage de la ventilation mécanique est le point critique de ce traitement de soutien, car il a été montré qu'une mauvaise ventilation mécanique causait des lésions pulmonaires et extrapulmonaires et une surmortalité [Nie+17]. Dès le niveau modéré, le recours à la ventilation mécanique est pourtant recommandé avec une stratégie de protection pulmonaire [Pap+19; Fer+21]. La suite de la stratégie ventilatoire est ensuite actualisée en fonction de l'aggravation ou de l'amélioration de l'état de santé du patient.

De nouvelles méthodes pour pallier les LIVM ou tout simplement le recours à la ventilation mécanique sont à l'étude dans la littérature. Elles portent entre autres sur d'autres stratégies de ventilation (invasives ou non), sur le ciblage par patient des méthodes à utiliser ou l'utilisation de traitements pharmacologiques [Bei+16; Fer+21]. Le recours à l'assistance extracorporelle (ECMO – *extracorporeal membrane oxygenation*) permet notamment dans les cas de SDRA très sévères de minimiser la ventilation et ainsi soustraire le patient à l'action délétère de la ventilation mécanique. Cependant, bien que de nombreux essais cliniques aient eu lieu, aucun médicament n'est à ce jour connu pour être efficace dans le traitement du SDRA [Fan+18; Mey+21].

En complément de la ventilation, la prise en charge peut également porter sur l'apport de surfactant, le traitement de la réponse inflammatoire pulmonaire et la prévention des LIVM.

Ventilation mécanique L'enjeu du traitement de soutien de ventilation porte sur le réglage d'une ventilation mécanique qui garantisse une oxygénation acceptable tout en étant protectrice²⁶.

Pour ce faire, il existe des recommandations [Pap+19; Mey+21] qui concernent les principaux réglages du ventilateur mécanique et qui s'appuient sur l'anatomie du patient ou d'autres paramètres d'aération mesurés. Les réglages recommandés concernent alors le volume courant, la pression plateau (P_{plat}), la pression expiratoire positive. Le premier correspond au volume d'air couramment inspiré entre l'inspiration et l'expiration. Le second est la pression mesurée par le ventilateur mécanique pendant la phase passive de l'inspiration (juste avant l'expiration) et fixer son seuil permet de diminuer la mortalité [Pap+19; Fer+21]. La pression expiratoire positive (PEP) correspond, quant à elle, à la pression appliquée à la fin de l'expiration. En effet, contrairement à la respiration normale (voir section La ventilation pulmonaire), la pression va ici être maintenue au-dessus de zéro afin d'éviter que les alvéoles qui se sont ouvertes à l'inspiration ne s'effondrent à la fin de l'expiration. Ce processus, appelé recrutement, est capital dans le traitement ventilatoire du SDRA, puisqu'il mesure l'aération (l'ouverture) des zones précédemment non aérées (effondrées) entre

25: Cela désigne par exemple les canules nasales ou les masques faciaux qui ne nécessitent pas d'intubation trachéale contrairement à la ventilation mécanique.

26: L'amélioration du pronostic vital en appliquant une ventilation protectrice est connu depuis déjà une vingtaine d'années [Pas+98].

Réglages recommandés en ventilation mécanique [Mey+21; Fer+21]

- Faibles VC : 4-8mL/kg du poids prédit par la taille
- PEP : élevée si SDRA modéré ou sévère (pour augmenter le recrutement)
- $P_{plat} < 30\text{cmH}_2\text{O}$
- FiO_2 (fraction d'oxygène inspiré) = 100%

l'expiration et l'inspiration. Le réglage de ces paramètres, entre autres, est soit fixé (P_{plat}), soit calculé à partir du poids prédit par la taille (VC), soit déterminé par échelle glissante (PEP). Malgré ces pratiques, tous les patients ne sont pas recrutables face aux mêmes réglages initiaux [Cop+19]. Adapter la ventilation de manière individualisée selon les caractéristiques d'aération est encouragé [Pap08; Bei+16; Nie+17; Cop+19; Mey+21] par la littérature et peut se faire de différentes façons.

Lésions pulmonaires induites par la ventilation mécanique Dans un premier temps, afin de pouvoir réduire les lésions causées par la ventilation, à savoir des blessures inflammatoires secondaires, il est capital de comprendre comment se forment ces lésions [Nie+17]. Il existe trois principaux mécanismes causant les LIVM. Ils sont causés par l'aération hétérogène et les interactions que cela cause entre les différentes zones²⁷ durant la respiration. On peut aussi différencier ces mécanismes selon qu'ils soient statiques, indépendants du cycle respiratoire ou dynamiques, causés par l'alternance entre inspiration et expiration.

Premièrement, la présence de concentrateurs de tension est mise en évidence comme un mécanisme majeur causant des lésions. Ceux-ci apparaissent lorsque la ventilation est hétérogène, car les alvéoles aérées avoisinant des zones effondrées ou remplies d'œdème génèrent des contraintes élevées sur les parois alvéolaires (voir (B) dans la Figure 1.9). Ajouté à cela, les tensions dynamiques (recrutement cyclique, (C) dans la Figure 1.9) sont jugées plus nocives que les tensions statiques (surdistension alvéolaire, (A) dans la Figure 1.9) [Nie+17]. Le recrutement dynamique est notamment lié à la PEP et au volume courant. En effet, un faible volume courant et une haute PEP causeront de faibles tensions dynamiques tandis que l'inverse causera de hautes tensions dynamiques. Cela s'explique par la diminution de surfactant, qui maintient habituellement les alvéoles ouvertes en fin d'expiration. Augmenter la PEP permet donc de garder les alvéoles susceptibles de s'effondrer (dérecrutement), ouvertes pendant l'inspiration et l'expiration [Pap08]. De même, le mécanisme de surdistension est diminué par l'application d'un faible volume courant et d'une PEP adéquate. Ce mécanisme décrit la compression des alvéoles par le gonflement excessif des tubes alvéolaires.

Diminuer l'apparition de ces lésions peut s'appuyer sur la quantification des diverses zones d'aérations [Pap08; Nie+17].

27: On peut identifier trois types d'aérations : les alvéoles normalement aérées, celles effondrées et celles hyperinflatées (voir Figure 1.9).

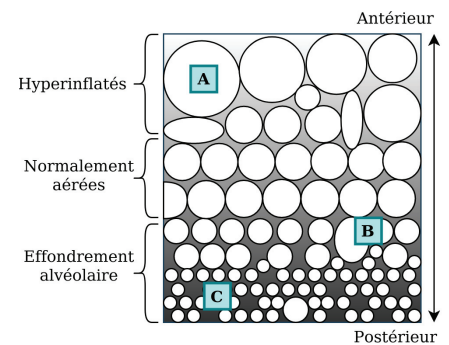


FIGURE 1.9 – Schéma des différentes lésions alvéolaires possibles entre l'inspiration et l'expiration. (A) la surdistension (B) contraintes entre alvéoles aérées et non aérées (C) recrutement cyclique.

Notions d'aération à retenir

Mécanisme d'aération	Etat des alvéoles	
	Expiration	Inspiration
Recrutement	non aérées	aérées
Hyperinflation cyclique	aérées	trop aérées
Surdistension	tendues	

Mesure de l'aération Adapter la stratégie ventilatoire est envisagé via l'étude de la réponse pulmonaire. Par exemple, identifier une importante hyperinflation cyclique conduit à la diminution du volume courant. Le réglage d'un autre paramètre important, la PEP, est possible grâce à la

mesure de l'aération et au calcul de la recrutabilité [Nie+17].

Le *gold standard* actuel, utilisé pour le réglage de la PEP, consiste à utiliser une échelle glissante basée sur les changements d'oxygénation. La quantification de l'espace mort anatomique permet aussi de régler ce paramètre du ventilateur mécanique. En calculant le ratio espace mort sur volume courant, il est possible d'identifier une ventilation dangereuse²⁸.

Une autre approche est d'utiliser les connaissances mécaniques des poumons, dont les liens avec la mortalité sont mieux connus [Nie+17]. Il est ainsi possible de régler la PEP par suivi de la compliance dynamique des poumons, puisque le point de décroissance de la compliance est associé au début de l'effondrement alvéolaire. Une autre technique de réglage s'appuie sur le tracé de la courbe pression/volume. En effet, identifier deux points d'inflexion sur la courbe permet de connaître la pression au début du recrutement alvéolaire (premier point) et le début de la surdistension (deuxième point). Cependant, la génération de cette courbe est une procédure complexe et dangereuse pour le patient. Une autre méthode moins éprouvante est d'utiliser la pression transpulmonaire, obtenue grâce à une sonde œsophagienne, et régler la PEP de telle sorte que la pression transpulmonaire soit supérieure à $0\text{ cmH}_2\text{O}$. Enfin, un autre tracé de courbe permet de régler la PEP, il s'agit de l'index de tension qui trace la courbe pression en fonction du temps. L'imagerie médicale est une autre possibilité qui fournit de nombreuses solutions [Pes+16].

Tout d'abord, la tomographie par impédance électrique permet de mesurer la ventilation pulmonaire au lit du patient entre chaque respiration. Elle permet aussi de quantifier le recrutement et dérecrutement alvéolaire pour différentes valeurs de PEP. Cependant, cette modalité qui nécessite le placement d'une ceinture d'électrodes autour du thorax, est peu résolue et ne permet d'imager qu'une trache de 5 à 10 cm à la fois [Pes+16]. L'utilisation de l'imagerie ultrasonore est aussi largement étudiée ces dernières années [Mon+19]. Le suivi se fait par identification et comptage de lignes spécifiques (*B-lines*). Cette modalité a l'avantage d'être mobile, peu coûteuse et non ionisante. Cependant, c'est une modalité de périphérie et d'artéfact, ainsi il n'est possible d'accéder qu'aux couches superficielles des poumons. Une autre modalité a aussi été utilisée pour suivre l'influence de la PEP sur la micromécanique des poumons, il s'agit de la tomographie par radiation au synchrotron [Sca+19]. Cependant, cette modalité encore émergente n'est pas disponible dans tous les hôpitaux.

Imagerie tomодensitométrie aux rayons X En 2001, l'équipe de Gattinoni [Gat+01] montre l'intérêt de l'imagerie CT dans la compréhension et le traitement du SDRA²⁹.

L'imagerie tomодensitométrie (TDM) aux rayons X ou CT (pour *computed tomography*) est le résultat de l'assemblage de plusieurs images radiographiques planaires qui permettent de reconstruire un volume. En effet, le scanner³⁰ CT est composé d'un émetteur de rayons X et d'un récepteur placé en face. Cet ensemble tourne autour du patient au fur et à mesure que les rayons X sont envoyés et réceptionnés (voir Figure 1.10). L'image CT est la reconstruction résultante de tous ces signaux. L'image volumique est reconstruite en coupes représentant une certaine épaisseur de la réalité. Comme l'imagerie aux rayons X est une imagerie de transmission³¹, l'imagerie CT est directement liée à la densité des tissus observés. Ainsi, une image CT est exprimée en unité Hounsfield (UH) et peut être vue

28: Sachant qu'un ratio élevé est caractéristique du SDRA.

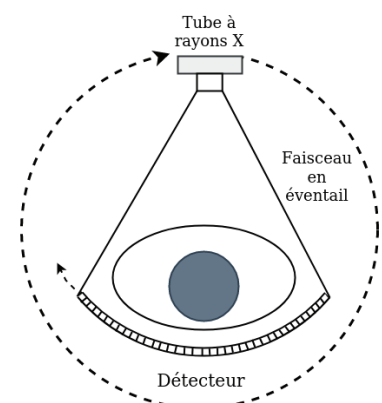


FIGURE 1.10 – Principe du scanner CT.

29: Intérêt depuis largement accepté dans la littérature [She+12; Cop+19]

30: Nom donné à l'appareil qui émet et crée l'image, ici avec des rayons X.

31: Qui crée l'image à partir de l'absorption de la source par les tissus.

comme une carte de densités, allant de -1000 UH pour l'air à +1000 UH ou plus pour les os. À l'image, les éléments qui sont très denses sont blancs (les os, les calcifications), tandis que ceux qui le sont peu sont noirs, avec les nuances de gris pour les éléments intermédiaires (voir Figures 1.11, 1.12, 1.13). La formule qui fait le lien entre la densité du tissu et son unité Hounsfield est la suivante :

$$UH = 1000 \times \frac{\mu - \mu_{eau}}{\mu_{eau}} \quad (1.1)$$

avec μ_{eau} la densité de l'eau et μ celle du tissu traversé. Selon les réglages de précision du scanner CT utilisé, l'unité Hounsfield sera la moyenne des densités des tissus contenus dans la zone imagée. La taille des détails représentés dans une image dépendent à la fois de l'appareil et des réglages, notamment de la dose de rayons X. Étant ionisante, la tendance, pour cette modalité, est aujourd'hui de réaliser des images à faible dose de rayons X, tout en maintenant une qualité de l'image suffisante pour son analyse.

L'utilisation de cette modalité, par rapport à la simple radiographie³², permet dans le cas du SDRA d'identifier plus précisément des lésions lors du diagnostic [Gat+10; She+12; Zom+14; Bar22]. Une variété de lésions sont ainsi évocatrices du SDRA (voir exemples dans la Figure 1.14). Leur présence, leur étendue et leur durée dans le temps permettent à la fois de déterminer la phase du syndrome dans laquelle se situe le patient et diriger son pronostic [She+12; Zom+14].

Son niveau de définition permet également de l'utiliser pour la prise en charge, d'autant plus qu'elle apporte de manière non invasive des informations sur l'état d'aération des alvéoles. La comparaison des caractéristiques des poumons observées chez des sujets sains et dans le cas du SDRA a notamment permis de mettre en évidence des différences significatives de la distribution des densités [Gat+01]. Une échelle a par la suite été définie séparant les unités Hounsfield pour définir quatre niveaux d'aérations du parenchyme pulmonaire (voir Figure 1.15) : non aéré (-100 UH à +100 UH), peu aéré (-500 UH à -100 UH), normalement aéré (-900 UH à -500UH) et trop aéré (-1000 UH à -900 UH). Cette échelle est aujourd'hui largement utilisée pour quantifier l'aération à partir des images CT [McK+16; Pes+16]. L'information volumique des images CT permet également de visualiser la répartition dans l'espace des œdèmes et inflammations, qui ont tendance à se concentrer du côté postérieur des poumons sous l'effet de la gravité [Gat+01].

En résumé, différentes méthodes sont disponibles afin de régler la PEP dans une démarche de ventilation protectrice. Ces méthodes s'appuient sur diverses informations biologiques et fournissent une compréhension du développement du SDRA qui reste limitée en termes de fiabilité. C'est pour quoi l'imagerie CT semble être celle qui permette de dépasser ces freins, au prix du transport du patient. Cependant, cette modalité implique une complexité de traitement pour tirer le plus d'informations anatomiques [Gat+10; She+12; Zom+14] et ainsi améliorer la compréhension et par la même occasion le traitement du SDRA.

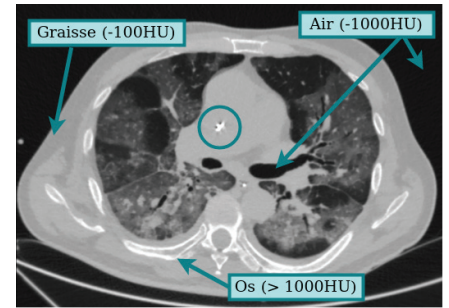


FIGURE 1.11 – Coupe axiale annotée d'une image CT des poumons (patient avec SDRA). Encerclé, une calcification dans le cœur.

32: Même si la tomodensitométrie requiert le déplacement du patient, ce qui est contraignant et peut être dangereux pour le patient.

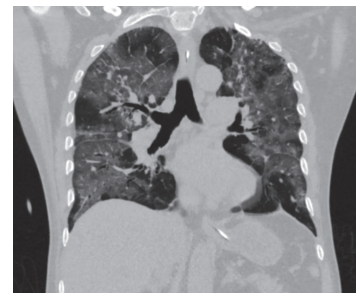


FIGURE 1.12 – Coupe coronale d'une image CT des poumons (patient avec SDRA).

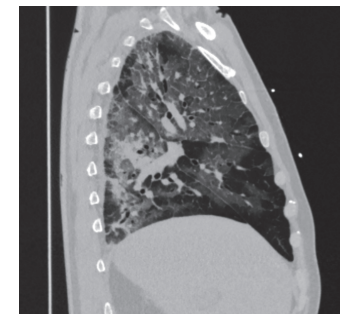


FIGURE 1.13 – Coupe sagittale d'une image CT des poumons (patient avec SDRA).

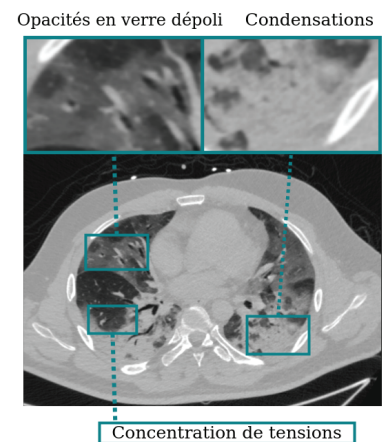


FIGURE 1.14 – Illustration de lésions et mécanismes associés sur une coupe axiale d'image CT, pour un patient atteint du SDRA.

SDRA résultant de la maladie COVID-19

La cause principale provoquant l'apparition d'un SDRA est la pneumonie infectieuse. De ce fait, les épisodes de grippes sévères sont facteurs d'augmentation du nombre de SDRA. Depuis le début du 21^e siècle, trois coronavirus³³ ont causé des épidémies ou pandémies, dont les formes graves prenaient la forme de syndromes respiratoires. En 2009, une autre pandémie causée par le virus A(H1N1) comptait 60% de cas de SDRA parmi les patients admis en réanimation [Rég09; Van14].

Plus récemment, la contagiosité du virus SARS-CoV-2 a fait émerger et se répandre à grande échelle la maladie COVID-19. La forme grave de celle-ci prend la forme du SDRA selon la définition de Berlin. Cependant, le SDRA découlant de la maladie COVID-19 (SDRA COVID+) présente des caractéristiques spécifiques qui le différencient du SDRA classique (SDRA COVID-) [Tir+20; Cha+20; Bel+21]. En effet, le SDRA COVID+ peut survenir sans perte de compliance pulmonaire, comme c'est le cas en temps normal [Gat+20; Gar+21]. Les lésions en verre dépoli sont les plus fréquentes sur les images radiographiques. Malgré les différences structurelles observées chez les patients développant un SDRA des suites d'une infection au SRAS-CoV-2, la prise en charge majoritairement adoptée a été celle du SDRA classique [Fer+21].

L'imagerie tomodensitométrique a été largement utilisée dans la prise en charge du virus à l'hôpital, notamment pour la détection précoce des patients infectés [Mah+20; Ye+20; Lod+20; Bar22]. Les lésions les plus courantes étaient semblables au SDRA classique, à savoir les opacités en verre dépoli et les condensations. Cependant, d'autres manifestations lésionnelles pouvaient apparaître au cours du développement de la maladie.

1.3 Conclusion

En somme, les besoins cliniques identifiés et les limitations actuellement présentes motivent les travaux menés dans cette thèse. À savoir, fournir une solution technique fiable qui permette au clinicien d'étudier in vivo la réponse pulmonaire à la ventilation. Ceci permettra par la suite d'évaluer la corrélation entre ces réglages ventilatoires et l'issue clinique.

L'objectif de cette thèse est de répondre à la première attente quant à la solution technique fiable de mesure de la réponse pulmonaire via l'imagerie CT. Pour cela, des objectifs de segmentation sont définis afin d'être compatibles avec une utilisation pré-clinique dans l'encadré suivant.

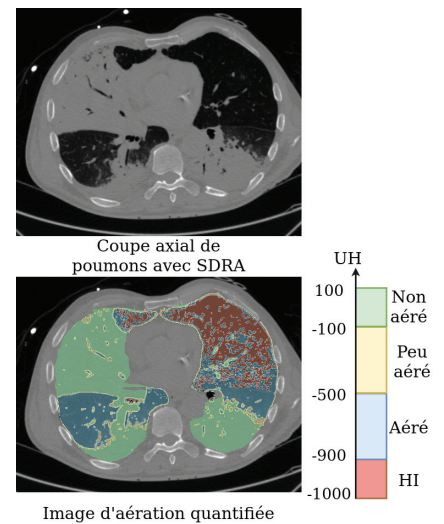


FIGURE 1.15 – Aération quantifiée sur coupe axiale.

33: Le SARS-CoV-1 en 2003, le MERS-CoV en 2012 et le SARS-Cov-2 en 2019 [Tes21].

Problématique et enjeux de la thèse

Tâche	Algorithme capable d'identifier (localiser) les poumons sur une image CT 3D.
Délai de calcul	< 1h, pour être compatible avec une utilisation clinique, au vu de l'urgence de traitement du SDRA.
Précision	Nécessite peu de retouches manuelles (si présentes).
Robustesse	Capacité de réaliser la tâche peu importe le niveau de sévérité du SDRA ou les lésions présentes dans l'image.
Cohérence	Fournit peu de différences entre deux images traitées du même patient.
Interactivité	Automatique.
Cadre d'utilisation	Mise à disposition en clinique.

Ce chapitre présente quelques points clés du domaine du traitement d'images afin d'explorer les solutions existantes dans la littérature scientifique qui répondent à la problématique présentée dans le chapitre précédent. Une présentation générale des différentes familles de méthodes est proposée. Ensuite, pour chacune d'entre elles, l'accent sur les méthodes appliquées au domaine d'intérêt (segmentation du poumon pathologique, en particulier avec SDRA) est progressivement présenté. À l'issue de ce chapitre, la direction empruntée, au regard de l'état de l'art, pour le travail de recherche est décrite.

Les descriptions du domaine du traitement d'images, de l'image numérique puis de la tâche de segmentation s'appuient dans cette section majoritairement sur ces deux livres [Gon+07; Bol+95]

Table des matières

2.1	La segmentation en traitement d'images	16
	L'image numérique	17
	La segmentation	19
	Évaluer la segmentation	20
2.2	Les méthodes traditionnelles	24
2.3	Les méthodes par apprentissage profond	26
	Le principe	26
	Les différents modèles	28
	Les méthodes spécifiques au SDRA	30
	Les bases de données publiques de poumons segmentés	33
2.4	Conclusion	33

2.1 La segmentation en traitement d'images

Le traitement d'images, sous-catégorie du domaine du traitement du signal, apparaît de la nécessité de traiter les images numériques pour l'utilisation informatique. Ce domaine réunit une variété de tâches qui concernent soit la qualité de l'image¹, soit le traitement de l'information² contenue dans l'image. C'est ainsi que l'on distingue deux types d'opérations, celles dites bas niveau qui ne sont attachées à aucune compréhension de l'image et à l'inverse, celles dites de haut niveau qui servent à interpréter et comprendre l'image. Dans la première catégorie, on peut citer entre autres les tâches d'acquisition, de transmission ou compression³, de filtrage et d'amélioration, de détection de contours ou segmentation ou encore de suivi de formes. Les tâches de haut niveau seront notamment celles de reconnaissance de formes ou de classification.

Afin de comprendre les différents concepts utilisés dans l'une de ces tâches, la segmentation, une définition⁴ de l'image numérique dans le domaine du traitement du signal est fournie dans un premier temps. Ensuite, la tâche d'intérêt, la segmentation, est définie. Enfin, les diverses manières⁵ d'évaluer la qualité d'une segmentation sont présentées.

1: À son acquisition ou a posteriori (débruitage, restauration, compression).

2: Sémantique, colorimétrique ou géométrique entre autres.

3: Utilisation de format de représentation de l'image qui permet de réduire sa taille en mémoire.

4: Celle-ci permet de mettre en évidence les mots clés qui seront utilisés tout au long de cette thèse.

5: Au regard du domaine de l'imagerie pulmonaire.

L'image numérique

Dans le domaine de la vision par ordinateur, le terme image désigne une représentation sur un plan⁶ (deux dimensions - 2D) d'une scène ou d'un objet de la réalité à trois dimensions (3D). D'un point de vue mathématique, c'est une fonction f qui associe à des coordonnées (i, j) du plan, une valeur d'intensité $f(i, j)$. Lorsque $f(i, j)$ a valeur dans un espace discret et fini, on parle d'image numérique. La Figure 2.1 illustre cette notation d'une image numérique. Les images numériques n'ont pu être gérées par les ordinateurs⁷ qu'à partir du milieu du siècle dernier tant leur manipulation et les calculs associés sont coûteux en mémoire. Comprendre la formation de l'image numérique puis sa représentation en machine sont deux concepts qui seront développés dans cette section.

Tout d'abord, l'acquisition d'une image est réalisée via un ou plusieurs capteurs sensibles à une grandeur donnée. En vision par ordinateur, la formation de l'image est le plus souvent basée sur la réflexion, le capteur mesurant la réflexion de la grandeur imagée. Dans le cadre de cette thèse, la modalité utilisée⁸ est basée sur le principe de la transmission, puisque le capteur reçoit le résultat transmis de la grandeur absorbée par l'objet imagé. Dans le premier cas, on peut citer en exemple l'appareil photo (le capteur) qui reçoit la lumière (la grandeur) réfléchi par ce qu'il capture (la scène ou l'objet imagé). Une illustration de ce principe est présentée dans la Figure 2.2. Dans le deuxième cas, présenté ici en Figure 2.3, le capteur reçoit les restes transmis d'une source absorbée par l'objet d'intérêt. Le scanner CT fonctionne selon ce principe et permet de créer une image représentant la transmission des tissus traversés par les rayons X émis. Quel que soit le principe utilisé pour former l'image, les capteurs excités vont générer un ensemble de valeurs numériques qui permettent alors de représenter les interactions grandeur-matière sur un écran après reconstruction.

L'image alors acquise est définie comme une matrice de valeurs, dont chaque case de coordonnées discrètes (i, j) est appelée un élément ou un pixel (pour une image 2D) et dont la taille définit les dimensions de l'image. On note x la dimension de la matrice qui correspond aux lignes de l'image et y la dimension de la matrice qui correspond aux colonnes de l'image. Les dimensions de l'image (2D) sont alors le nombre de lignes et le nombre de colonnes qui compose sa matrice, et sont notées $M \times N$ (voir Figure 2.1). Chaque pixel contient l'information d'intensité qui permet de représenter l'image sur un écran. Lorsque l'image est en niveaux de gris (comme l'exemple en Figure 2.4), une seule valeur numérique par pixel permet de coder sa représentation. Si l'image est en couleurs, trois valeurs par pixel, correspondant par exemple aux trois couleurs primaires en synthèse additive, permettent de reproduire sur un écran toutes les couleurs du spectre visible. La quantité de valeurs que comporte chaque pixel est enregistrée dans une nouvelle dimension de la matrice représentant l'image : le canal. La schématisation d'une image numérique en Figure 2.4 comporte donc un seul canal, car uniquement composée de niveaux de gris. D'ailleurs, seuls quatre niveaux de gris sont utilisés dans cette image : blanc, gris clair, gris foncé et noir, d'où son aspect très contrasté⁹. En effet, l'étendue des intensités que peuvent prendre chaque pixel va influencer l'apparence de l'image en termes de couleurs. Ceci est déterminé entre autres par (1) le type de valeurs, (2) leur taille en mémoire et (3) leur valeur minimale et maximale. Le premier point désigne le fait que l'intensité peut être

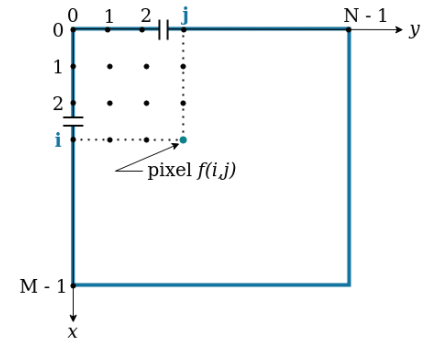


FIGURE 2.1 – Représentation d'une image numérique 2D de taille $M \times N$.

6: En imagerie médicale, certaines modalités permettent de représenter cette réalité directement en trois dimensions.

7: Stockage en mémoire, affichage et traitement.

8: L'imagerie aux rayons X, en particulier la tomodensitométrie aux rayons X. Voir explication de cette modalité dans la sous-section Prise en charge du Chapitre 1.

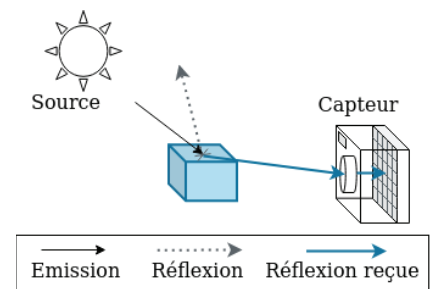


FIGURE 2.2 – Formation d'une image par réflexion de la source (ici la lumière).

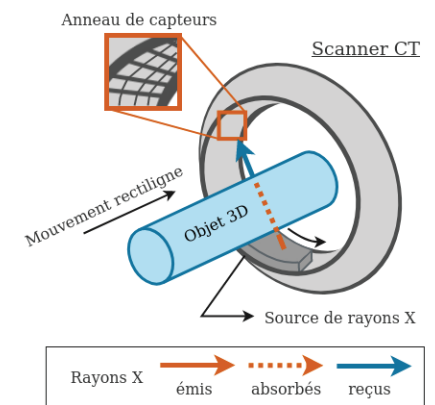


FIGURE 2.3 – Formation d'une image CT par transmission des rayons X par l'objet 3D.

9: Définition du contraste dans une image : différence de luminosité entre les parties claires et les parties sombres de l'image.

représentée aussi bien par des nombres entiers que par des nombres réels, et en fonction de leur type, leur espace en mémoire n'est pas le même. C'est ce deuxième point, l'espace en mémoire de l'image, qui est directement impacté par le nombre de bits¹⁰ attribués au canal du pixel. On parle de profondeur de bit, pour indiquer le nombre de bits utilisés pour encoder la valeur d'un pixel, et celle-ci est soit directement dépendante du capteur, soit choisie par l'utilisateur. Enfin, le troisième point, désigne la valeur minimale et maximale des pixels d'une image. Cette caractéristique, appelée plage dynamique, est limitée par la profondeur de bit des pixels, mais peut aussi être diminuée par l'utilisateur. Ces trois paramètres, entre autres, permettent ainsi d'influencer la taille en mémoire de l'image.

Enfin, un autre paramètre qui influence à la fois l'apparence et la taille en mémoire d'une image est la résolution spatiale de l'image. Il s'agit de la capacité de l'image à représenter précisément la réalité qui dépend du capteur utilisé pour acquérir l'image. Premièrement, comme indiqué plus tôt, une image numérique a valeur dans un espace discret et fini à la différence des grandeurs physiques qui sont continues. Cela suppose donc une étape de discrétisation ou échantillonnage qui permet de faire le lien entre le monde analogique et le monde numérique. Cette étape est paramétrée par la fréquence d'échantillonnage (densité spatiale) qui induit la quantité¹¹ d'éléments de la matrice qui vont être utilisés pour représenter le signal reçu. La réalité physique ainsi représentée par chaque case de la matrice est appelé résolution spatiale¹² et est mesurée en unité de longueur physique¹³. Plus le nombre de cases est grand pour représenter les mêmes dimensions physiques, plus la résolution spatiale de l'image sera petite. L'écart entre le centre de chaque pixel (l'intervalle d'échantillonnage), dans chaque direction x et y pour une image 2D, devient donc de plus en plus petit, plus une image est résolue.

Tout ceci peut être appliqué aux volumes ou images dites 3D ou volumique, qui peuvent être vues comme l'empilement de matrices d'images 2D. La matrice d'une image 3D comporte alors trois dimensions (x, y, z) avec z la nouvelle direction selon laquelle on accède aux matrices 2D. Chaque case de la matrice de coordonnées (i, j, k) , semblable au pixel, est nommée voxel et représente alors une tranche de la réalité. Cette épaisseur est retranscrite dans la résolution spatiale par une nouvelle dimension, dans la nouvelle direction z . Le scanner CT présenté dans la Figure 2.3 produit des images 3D par reconstruction tomographique des différentes projections 2D obtenues lors de la rotation de la source de rayons X. Un résumé des caractéristiques présentées dans cette section est présenté ci-dessous.

Caractéristiques principales d'une image 2D ou 3D

- dimensions (x, y) ou (x, y, z)
- résolution spatiale (mm)
- nombre de canaux
- plage dynamique (min, max)
- profondeur de bit (bit)
- taille en mémoire (Mo)

10: Un bit est la plus petite unité d'information basée sur le système binaire. Un ordinateur ne traite que des données binaires.

Caractéristiques de l'image numérique de la Figure 2.4

- taille 7×8
- canal : 1
- plage dynamique : (0, 255)
- profondeur de bit : 8-bit
- taille théorique : 0.5 Mo

11: Ainsi, le nombre de cases dans la matrice va influencer la quantité d'information à coder.

12: La résolution d'une image est néanmoins limitée physiquement par la résolution optique, c'est-à-dire la capacité du capteur à saisir les détails de la scène imagée.

13: Ou bien en points par pouce (ppp) qui est l'unité de résolution d'une image numérique.

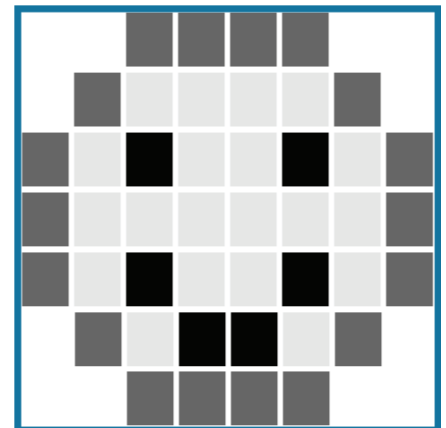


FIGURE 2.4 – Exemple d'une image numérique 2D de taille 7×8 dans laquelle les pixels sont séparés par des lignes blanches pour cette représentation.

La segmentation

La tâche d'intérêt de cette thèse est la segmentation. C'est une tâche du traitement d'images, qui est souvent utilisée à la jonction des tâches bas et haut niveau, puisqu'elle permet d'identifier un ou des éléments d'intérêt dans une image. Dans le premier cas, elle sert de pré-traitement permettant de définir une zone d'intérêt réduite pour une tâche suivante. Dans le second cas, elle est une fin en soi. Il s'agit tantôt d'identifier des régions selon des critères purement numériques (intensité des pixels, forme), tantôt de segmenter des éléments en fonction de leur nature, on parle alors de segmentation sémantique. Dans le second cas, la tâche de segmentation s'apparente à de la classification pour chaque pixel de l'image. La classification désigne généralement le fait d'attribuer un ou plusieurs labels à une image, pour indiquer si un ou des objets d'intérêt sont présents ou non dans l'image. Il n'y a pas d'information sur la position¹⁴ de l'élément. Lorsque plusieurs labels existent dans la segmentation d'une image, on parle de segmentation multi-classes. Autrement, la segmentation binaire désigne l'attribution d'un seul label en dehors de celui attribué au fond de l'image. C'est l'exemple illustré par la Figure 2.5 dans laquelle la segmentation sémantique contient le label poumon et tout le reste est considéré comme le fond de l'image. Le résultat de la segmentation présenté dans cette figure est d'ailleurs appelé masque, puisqu'il permet par superposition de colorer l'objet d'intérêt dans une image.

Le formalisme mathématique définit la segmentation comme une partition de l'image I en n régions R_i homogènes selon un critère logique, telles que :

Définition mathématique de la segmentation

- (1) $\forall i \in [1, n], R_i \neq \emptyset$
- (2) $\forall i, j$ et $i \neq j, R_i \cap R_j = \emptyset$
- (3) $I = \bigcup_i^n R_i$
- (4) $C(R_i) = \text{VRAI}$ pour $i \in [1, n]$
- (5) $C(R_i \cup R_j) = \text{FAUX}$ pour R_i et R_j adjacents

avec R_i des ensembles de pixels connectés et $C(R_i)$ le critère logique de segmentation défini sur la région R_i . Ce formalisme traduit que chaque région soit non vide et que les régions soient deux à deux disjointes¹⁵. Chaque pixel de l'image doit faire partie d'une région¹⁶ et vérifie un critère homogène dans sa région¹⁷. Enfin, deux régions adjacentes, c'est-à-dire qui partagent une frontière commune, ne peuvent pas vérifier le même critère logique¹⁸. Ainsi, une segmentation regroupe des pixels selon leur similarité ou leur différence. On peut ainsi rassembler les méthodes basées sur la similarité comme des approches régions et celles basées sur la différence comme des approches frontière. Dans le premier cas, les pixels vérifiant le critère d'une région sont directement affectés à celle-ci, tandis que dans le deuxième cas, une étape supplémentaire pour passer d'une frontière à une région est souvent nécessaire. En effet, pour respecter toutes les définitions de la segmentation¹⁹, identifier les limites entre chaque région n'est pas suffisant, bien que cela puisse être une approche plus simple selon l'image.

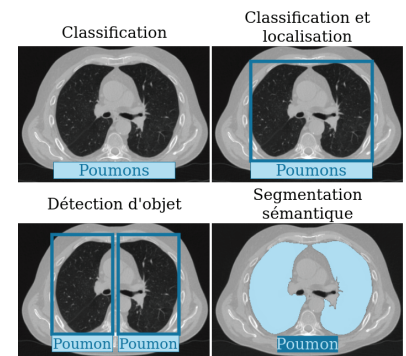


FIGURE 2.5 – Différences entre les tâches de classification, localisation, détection et la segmentation sémantique.

14: A contrario de la localisation ou de la détection d'objet, une autre tâche du traitement d'images.

15: Conditions (1) et (2).

16: Condition (3).

17: Condition (4).

18: Condition (5).

19: Voir l'encadré **Définition mathématique de la segmentation**.

Il existe ainsi une série d'étapes classiques appliquées dans la tâche de segmentation²⁰ (voir Figure 2.6) qui visent à l'optimiser en amont et en aval. La première étape est l'acquisition de l'image²¹ sur laquelle l'utilisateur peut ou non agir. Vient ensuite une étape désignée comme du pré-traitement, puisqu'elle consiste à modifier l'image pour qu'elle soit dans les conditions optimales de traitement pour la tâche d'intérêt. Cette étape consiste par exemple à ajuster les intensités de l'image soit pour augmenter les contrastes ou au contraire lisser l'image pour effacer le bruit²².

Le traitement principal peut alors être appliqué, l'image préparée pour correspondre aux spécificités de l'algorithme choisi. Selon la méthode employée, l'image pourra être considérée comme une matrice²³, un processus stochastique, un vecteur aléatoire ou une surface. Enfin, il convient aussi d'appliquer des corrections simples en fin de traitement pour pallier les éventuelles erreurs retournées à l'étape précédente ou simplement vérifier les définitions de la segmentation. Choisir un algorithme adéquat pour la tâche d'intérêt dépend d'abord de l'image et de son acquisition, mais aussi des ressources de calcul disponibles. En effet, la nature de l'image ou bien des objets présents dans cette dernière sont à prendre en considération, mais les contraintes d'exploitation que sont, entre autres, le temps de calcul nécessaire et les ressources disponibles orientent également les traitements possibles.

Évaluer la segmentation

Une fois la segmentation produite, il convient d'en évaluer la qualité. Pour cela, connaître la segmentation cible, celle qui identifie parfaitement et objectivement les éléments recherchés, est nécessaire. Cette dernière est appelée la vérité terrain. Néanmoins, dans le domaine de l'imagerie médicale, obtenir une vérité objective est impossible. Pour pallier ce frein, la segmentation de référence ou *gold standard* est utilisée. Il s'agit de la segmentation qu'aurait produite un expert²⁴ du domaine. Cette dernière est soit réalisée manuellement par l'expert du domaine, qui associe un label à chaque pixel de l'image, soit de manière semi-automatique auquel cas l'expert corrige le résultat d'un algorithme (avec initialisation). Dans le cas d'une segmentation binaire²⁵, la référence l'est aussi et comporte uniquement deux labels (ou classes) : l'objet segmenté (positif au critère de segmentation) et tout le reste (le fond, négatif au critère de segmentation). La Figure 2.7 illustre à la fois la segmentation de référence et sa superposition avec l'image CT sur laquelle est elle basée.

La qualité de la segmentation produite est alors dépendante de sa ressemblance ou dissemblance avec la vérité terrain, selon des critères propres à chaque application. Quantifier les éléments bien et mal segmentés par la méthode peut être fait en classant chaque pixel du masque selon qu'il est ou non dans la même classe que la vérité terrain. On décompte alors quatre catégories détaillées dans le Tableau 2.1 qui sont les vrais négatifs et vrais positifs (bonne segmentation) et les faux négatifs et faux positifs (mauvaise segmentation). Les notations VN, FN, FP et VP désignent le nombre de pixels (ou voxels) appartenants à la classe correspondante. Les termes négatif et positif désignent respectivement un label attribué au fond de l'image ou à l'objet segmenté. Quantifier chacune de ces catégories permet de déterminer principalement la superposition du masque produit

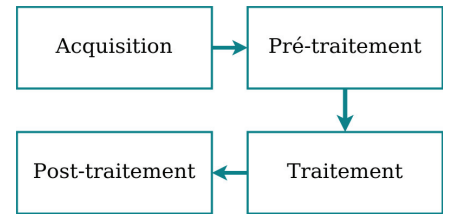


FIGURE 2.6 – Processus classique du traitement d'images

20: Aussi communes aux autres tâches du traitement d'images.

21: Lorsque ce n'est pas la tâche d'intérêt, souvent cette étape ne dépend pas de l'utilisateur.

22: Définition du bruit numérique dans une image : irrégularité ou dégradation subie par l'image entre son acquisition et son enregistrement.

23: Voir la représentation classique d'une image numérique dans la section 5.

Les exemples présentés seront en 2D même si toutes les métriques discutés sont applicables aux images 3D.

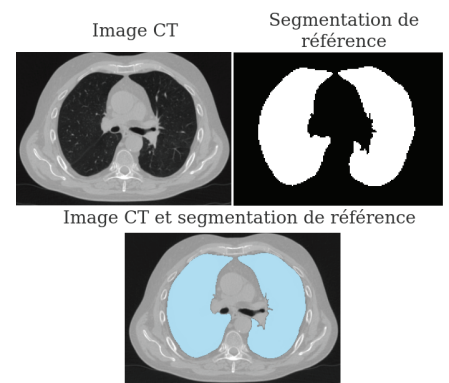


FIGURE 2.7 – Illustration de la segmentation de référence sur une image CT de poumons.

24: Un médecin ou un interne en médecine dans le cas de l'imagerie médicale.

25: Cadre choisi pour cette thèse.

TABEAU 2.1 – Matrice de confusion comparant la segmentation produite à la vérité terrain – Négatif désigne tout pixel qui ne répond pas au critère de segmentation ; Positif désigne tout pixel qui répond au critère de segmentation.

		Vérité terrain	
		Négatif	Positif
Segmentation	Négatif	Vrai négatif : VN	Faux négatif : FN
	Positif	Faux positif : FP	Vrai positif : VP

par la méthode et la vérité terrain en tant que facteur de qualité. L'intérêt peut aussi porter sur la présence de valeurs aberrantes (éléments isolés et éloignés de l'objet à segmenter), la similarité géométrique (formes proches), etc. [Tah+15]. Caractériser la superposition, la forme, les contours, l'alignement, le volume ou la tendance à fournir peu de valeurs aberrantes de la segmentation reste au choix de l'utilisateur. Il n'existe pas de métrique qui permette de mesurer toutes les caractéristiques précédemment citées, ce pourquoi la littérature s'attache à utiliser un ensemble de métriques génériques qui balayent suffisamment de propriétés pour les besoins d'une application donnée [Tah+15; Riz+20]. La Figure 2.8 illustre schématiquement d'un côté une segmentation acceptable et de l'autre une mauvaise segmentation. Il est simple de qualifier une segmentation d'échec dès lors que celle-ci ressemble peu à l'objet à segmenter. Pourtant, la qualifier de succès dépend de la précision souhaitée et de la métrique utilisée.

L'étude de [Tah+15] dénombre six catégories de métriques appliquées à la segmentation d'images médicales, parmi lesquelles trois seront détaillées ici de façon non exhaustive : celles basées sur la superposition, celles basées sur les distances spatiales et celles basées sur le volume.

Métrique de volume Cette catégorie compte une seule métrique répandue dans la littérature [Tah+15; Yeg+18], qui consiste à calculer le ratio de volume entre la segmentation et la référence. Cette métrique, appelée différence relative (absolue) de volume (DRV), se calcule en utilisant les éléments de la matrice de confusion Tableau 2.1.

$$DRV = 1 - \frac{|FN - FP|}{2VP + FP + FN} \tag{2.1}$$

Il est impossible de différencier le cas où FP et FN sont grands et égaux (mauvaise segmentation) du cas où ils sont nuls (segmentation parfaite), ce qui en fait une métrique peu fiable. Si l'on prend l'exemple de la segmentation S_o de la coupe en Figure 2.8 la DRV indiquerait une bonne segmentation.

Métriques de superposition Elles sont toutes basées sur le compte des pixels (ou voxels) classés dans les catégories VP, VN, FN et FP et de ce fait bornées entre 0 et 1, car calculant toutes des proportions. L'échec total vaut donc 0 et le succès vaut 1 (voir Figure 2.9). Parmi les métriques basées sur la superposition, la plus basique est nommée *accuracy* et elle évalue la proportion de pixels bien classés par l'algorithme de segmentation.

$$Accuracy = \frac{VP+VN}{VP+FP+FN+VN} \tag{2.2}$$

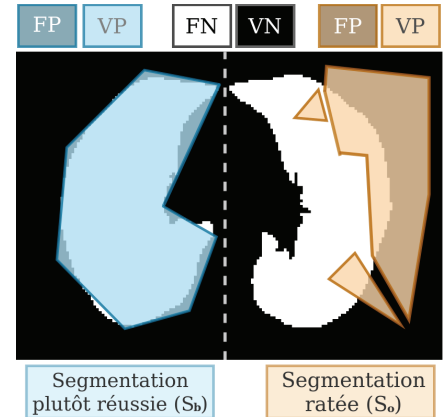


FIGURE 2.8 – Illustration d'une bonne segmentation S_b (à gauche) et d'une mauvaise segmentation S_o (à droite), labellisée selon les quatre catégories de la matrice de confusion.

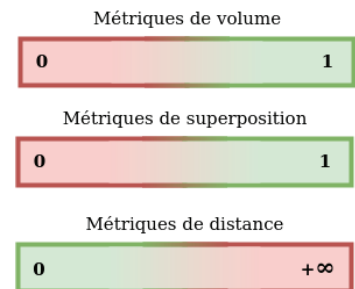


FIGURE 2.9 – Intervalle de valeur des métriques selon leurs catégories – En vert les valeurs d'une bonne segmentation, en rouge celles d'une mauvaise segmentation

Du fait de sa formulation (voir Equation 2.2), elle est très sensible au déséquilibre de classes et peut donner un avis biaisé en cas de classe dominante dans l'image. De ce fait, elle convient lorsque les classes à segmenter dans l'image sont bien équilibrées. C'est pour quoi dans l'image présentée en Figure 2.8, la valeur obtenue par la segmentation S_b est plutôt bonne (au-dessus de 0.9, voir encadré **Performances des segmentations S_b et S_o**).

La précision compte la proportion de pixels bien segmentés par rapport à la totalité des pixels segmentés.

$$\text{Précision} = \frac{VP}{VP+FP} \quad (2.3)$$

C'est une métrique qui permet de mettre en évidence²⁶ la sur-segmentation. Cependant, elle ne tient pas compte des zones non segmentées, par exemple les faux négatifs dans la segmentation S_o . Au contraire, pour évaluer la sous-segmentation, il faut compter la proportion de pixels bien segmentés sur la totalité des pixels de la référence (à segmenter). C'est une information qui est tantôt désignée par le terme rappel ou bien sensibilité et est définie comme suit :

$$\text{Rappel} = \text{Sensibilité} = \frac{VP}{VP+FN} \quad (2.4)$$

Ici, seuls les pixels²⁷ superposés à la référence sont considérés. Cette métrique ne rend donc pas compte des erreurs de segmentation à l'extérieur²⁸ de la référence. La sensibilité est donc toujours couplée avec la spécificité, qui mesure, elle, tout ce qui se situe à l'extérieur de la référence.

$$\text{Spécificité} = \frac{VN}{VN+FP} \quad (2.5)$$

Elles témoignent ainsi, à elles deux, de la capacité de la méthode de segmentation à reconnaître l'objet à segmenter (sensibilité) tout en ne segmentant pas les zones à l'extérieur (spécificité). Néanmoins, la spécificité n'est pas souvent utilisée en imagerie médicale [Tah+15], car dépendante de la taille relative de la référence à l'image. Une combinaison de la sensibilité et de la précision permet de définir une nouvelle métrique qui s'appelle score $F1_{score}$ ²⁹, coefficient de similarité de Dice ou score de Dice (Dice) selon le domaine d'utilisation.

$$\text{Dice} = \frac{2VP}{2VP+FP+FN} \quad (2.6)$$

Elle correspond à la proportion de pixels (ou voxels) correctement identifiés dans la référence sur le nombre moyen de pixels (ou voxels) composant la référence et la segmentation produite. C'est l'une des métriques les plus utilisées en traitement d'images médicales [Yeg+18] puisqu'elle permet d'évaluer à la fois les frontières de la segmentation en même temps que la superposition avec la vérité terrain. Une variation de sa définition³⁰ permet d'obtenir une autre métrique, l'indice de similarité de Jaccard (JI, voir Equation 2.7, qui est aussi appelée IoU pour *Intersection over Union*), qui va considérer l'intersection de la segmentation et de la référence sur l'union des deux.

$$\text{JI} = \frac{VP}{VP + FP + FN} \quad (2.7)$$

Cette métrique est toujours inférieure au Dice sauf aux extrémités. Malgré l'information de superposition prise en compte par ces dernières, la dis-

26: Ainsi visible dans la valeur obtenue par la segmentation S_b , dans l'encadré **Performances des segmentations S_b et S_o**

27: Segmentés ou pas.

28: Raison pour laquelle S_b obtient un score presque parfait.

29: $F1_{score} = 2 \times \frac{\text{Précision} \times \text{Rappel}}{\text{Précision} + \text{Rappel}}$

30: $\text{JI} = \frac{\text{Dice}}{2 - \text{Dice}}$

tribution spatiale des pixels segmentés reste ignorée. Ceci ne permet pas d'évaluer finement les erreurs aux contours, ni celles très éloignées comme les erreurs isolées et déconnectées du volume à segmenter.

Métriques de distances De ce fait, il est intéressant de considérer les distances entre les contours, le masque produit et le masque de référence. L'une de ces métriques mesure la distance moyenne entre les contours de la segmentation et ceux de la référence, il s'agit de la distance moyenne symétrique entre surfaces (*average symmetric surface distance* – ASSD), définie ci-dessous :

$$\text{ASSD}(V_S, V_r) = \frac{\sum_{x \in V_r} \mathbf{d}(x, V_S) + \sum_{y \in V_S} \mathbf{d}(y, V_r)}{|V_r| + |V_S|}, \quad (2.8)$$

avec V_r le volume de la vérité terrain et V_S le volume de la segmentation produite par la méthode qui est évaluée, x et y deux points de l'espace et \mathbf{d} la plus petite distance euclidienne entre un point donné et tous les points d'un volume. Cette métrique intègre la réalité physique dans l'image via le calcul de la distance et la résolution spatiale de l'image (voir Section 5). Une autre métrique permet d'accorder plus d'importance aux erreurs isolées, puisqu'elle mesure la plus grande distance (symétrique) entre la segmentation et la vérité terrain. Cette distance est définie dans l'Equation 2.9 et se nomme la distance de Hausdorff (DH).

$$\text{DH}(V_S, V_r) = \max\{\max_{x \in V_S} \mathbf{d}(x, V_r), \max_{y \in V_r} \mathbf{d}(y, V_S)\} \quad (2.9)$$

Le 95^e centile³¹ de cette métrique est considéré comme plus robuste et davantage utilisé dans la littérature, car il ne pénalise pas le cas où un faible nombre de très grandes erreurs existerait dans la segmentation produite.

$$\mathbf{d}_k = K_{x \in V_S}^{me} \left(\min_{y \in V_r} \mathbf{d}(x, y) \right), \quad (2.10)$$

avec $K_{x \in V_S}^{me}$ le k-ième centile des distances ordonnées pour $x \in V_S$. Cependant, cette distance n'est pas symétrique, de ce fait la définition complète du 95^e centile de la distance de Hausdorff (DH-95) est :

$$\text{DH-95}(V_S, V_r) = \max(\mathbf{d}_{95}(V_S, V_r), \mathbf{d}_{95}(V_r, V_S)) \quad (2.11)$$

Cette catégorie de métriques apporte des informations sur les contours et les erreurs isolées entre autres, mais ne permet pas de prendre en compte l'information volumique.

En résumé, chaque catégorie de métrique permet d'accentuer une ou plusieurs caractéristiques à évaluer. Elles sont donc à considérer de manière complémentaire dans l'évaluation des performances d'une méthode de segmentation. Notamment, l'utilisation combinée d'au moins une métrique de superposition et de distance sera privilégiée, car elles permettent d'évaluer à la fois la superposition, mais aussi la position des segmentations.

Performances des segmentations S_b et S_o

Exemple numérique correspondant à la Figure 2.8 :

Métrique	S_b	S_o
Accuracy	0.933	0.530
Précision	0.862	0.311
Sensibilité	0.992	0.242
Spécificité	0.894	0.690
Dice	0.922	0.272
Jl	0.856	0.157

Performances des segmentations S_b et S_o

Exemple numérique correspondant à la Figure 2.8 :

Métrique	S_b	S_o
ASSD (mm)	3.32	11.5
DH (mm)	10.5	35.9
DH-95 (mm)	7.70	25.5

31: Seuil tel que 95% des valeurs mesurées sont en dessous et 5% au-dessus.

2.2 Les méthodes traditionnelles

Le traitement d'images est un domaine qui est depuis le début dépendant des capacités de calcul des ordinateurs. On peut identifier, entre autres, deux grands groupes de méthodes de segmentation selon que celles-ci se basent sur la théorie du traitement du signal ou sur l'intelligence artificielle. La première catégorie sera désignée comme traditionnelle et étudiée dans cette partie. Les méthodes de segmentation 2D et 3D seront balayées du point de vue général avec chaque fois un accent mis sur l'imagerie médicale, en particulier sur l'imagerie CT et les poumons avec pathologies ou SDRA.

Méthodes basées sur le seuillage La méthode la plus basique de segmentation est basée sur le seuillage, c'est-à-dire associer chaque pixel à une région selon que sa valeur est inférieure ou supérieure à un seuil. Ce seuil est soit connu et défini par l'utilisateur ou bien déterminé par un algorithme qui s'appuie sur l'histogramme de l'image³². Cette méthode est utilisée pour l'une des premières fois en 1964 pour la reconnaissance de motifs sur des photos de satellite [Kat64]. Les méthodes basées sur le seuillage ont l'avantage d'être simples à implémenter, intuitives et rapides en temps de calcul. En imagerie médicale, elles sont bien adaptées aux structures bien contrastées, comme c'est le cas des poumons sains et uniformément aérés par exemple³³ [Hu+01]. Cependant, par nature, les méthodes basées seuillage ne sont pas robustes aux multiples variations locales d'intensités, qui surviennent en cas de lésions, de densités ou d'artéfacts [Men+12; Rik+13; Man+15]. C'est pourquoi le seuillage sera plutôt utilisé comme prétraitement pour définir une zone d'intérêt [Zha+12; Shi+16; Jav+16; Car+18; Ast+22].

Méthodes basées sur les régions Pour pallier la sensibilité aux variations d'intensités, une autre approche simple est de considérer le voisinage des pixels (voir Figure 2.10) en plus de leurs valeurs respectives. On définit le voisinage d'un pixel selon que l'on considère uniquement les pixels qui le touchent dans les directions orthogonales (voir Figure 2.10 (a)) ou toutes les directions (voir Figure 2.10 (b)). Les méthodes suivant cette caractéristique, sont dites basées sur les régions, restent rapides et sont plus robustes au bruit ou aux petites lésions [Sun+06; Man+15] que celles basées sur le seuillage. Divers attributs des régions peuvent être pris en compte pour définir les critères de segmentation.

Le critère le plus simple associe à la même région tous les pixels voisins qui vérifient la même condition sur l'intensité (par exemple l'homogénéité), c'est l'algorithme dit par croissance de régions [Bri+70]. D'autres critères nécessitent de représenter l'image de façon compacte comme un graphe, pour lui appliquer des méthodes de la théorie des graphes [Boy+98; Cha+17] et résoudre la segmentation comme une optimisation. Chaque pixel est représenté comme un nœud d'un graphe et segmenter revient à trouver le découpage qui maximise la cohérence des nœuds d'un groupe³⁴. Lorsque sont ajoutées la forme et la distance entre les pixels au critère de segmentation, on parle de méthode par connexité floue. Appliquée à la segmentation de divers organes sur des images CT ou IRM, cette méthode est assez robuste [Udu+03; Shi+16]. Pourtant, dans l'étude de Shi et al., utiliser une combinaison de méthodes simples basées sur les régions fournit des résultats de segmentation de poumons comparables aux

32: Il s'agit de la distribution des intensités de l'image.

33: En effet le seuil peut être facilement défini sur des images CT puisque l'aération est directement liée aux unités Hounsfield

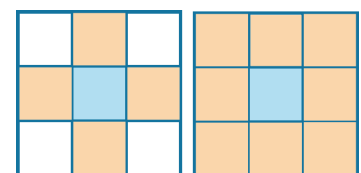


FIGURE 2.10 – Voisinage (orange) du pixel (bleu) – (à gauche) 4-voisins (à droite) 8-voisins.

34: C'est la méthode des *graph cut*.

méthodes plus complexes³⁵ et dans un temps plus court que ces dernières [Shi+16]. En 2020, Davila et al. [Dáv+20] montrent l'efficacité en utilisation clinique de segmenter les zones aérées des poumons avec SDRA par des méthodes de croissance de régions. Spécifiquement en ce qui concerne la segmentation des poumons sur des images CT, les méthodes basées sur les régions ont l'avantage de fournir des segmentations qui ne débordent pas hors du corps [Man+15]. Néanmoins, ces méthodes ne suffisent pas toujours, même sur des organes sains. Alors une augmentation de la complexité et de l'information traitée permet d'améliorer les segmentations dans les cas non réussis. En effet, considérer le contexte spatial de la zone à segmenter permet d'ajouter une information spatiale élargie au critère de segmentation.

Méthodes basées sur la forme Les distributions des niveaux de gris et les contrastes sont directement impactés par la présence de pathologies, notamment dans les poumons [Men+12]. Ce pour quoi utiliser l'information de forme semble être la solution privilégiée dans la littérature [Car+18]³⁶. En effet, la popularité de ces méthodes tient au fait qu'elles prennent en compte l'information locale et globale sur les images. Pour ce faire, deux grandes approches existent : celles utilisant un atlas ou celles utilisant un modèle de forme.

La première stratégie s'appuie sur un ensemble d'images segmentées, constituant l'atlas, par rapport auquel l'image traitée va être ajustée. Cette étape de correspondance entre l'image de référence (fabriquée à partir de l'atlas) et l'image traitée est appelée le recalage en traitement d'images³⁷. Bien que cette approche soit performante sur des poumons peu malades [Li+03], elle introduit aussi deux difficultés qui sont la constitution de l'atlas et l'étape de recalage. En particulier, la variabilité de l'anatomie des poumons et les différentes apparences des lésions pulmonaires, rendent la constitution et le recalage de l'atlas complexe [Man+15].

La deuxième stratégie est basée sur le même principe que les atlas, mais résout le problème de segmentation avec une approche statistique ou probabiliste. On y dénombre de nombreuses méthodes telles que les champs aléatoires de Markov, les modèles déformables (contours actifs, *snakes*, *level-set*). En 2017, [Kla+17] compare sur des poumons avec SDRA la segmentation manuelle et la segmentation automatique basée sur la forme³⁸. Les résultats montrent une quasi-équivalence de segmentation des deux méthodes sur les compartiments normalement aérés et hyperinflatés. Les compartiments peu aérés sont surestimés par la méthode automatique tandis que les compartiments non aérés sont quant à eux sous-estimés. Les algorithmes basés sur la forme ont l'inconvénient d'être coûteux en calcul, peu reproductibles et hautement dépendants de la qualité des caractéristiques du modèle de forme ou de l'atlas constitué.

En résumé, les nombreuses approches existantes ont été appliquées successivement à la segmentation des poumons sur des images CT. Cela s'est avéré concluant lorsque ceux-ci étaient sains et bien contrastés. Cependant, en présence de pathologies ou d'anatomie atypique, augmenter la complexité³⁹ ou les types d'informations utilisées est une solution qui n'a pas été systématiquement concluante. C'est ce que montrent les études de [Man+15] et [Men+12] sur des poumons avec pathologies, qui d'un côté comparent les erreurs produites par cinq méthodes classiques de

35: Une méthode basée sur les *graph cut* et une autre basée sur la connexité floue.

36: Avec 23 articles sur 98 étudiés, cette étude place les méthodes de segmentation 3D sur images CT basées sur la forme comme les plus répandues.

37: C'est un domaine à part entière, tout autant que la segmentation, qui comporte ses difficultés et enjeux

38: Utilisant le logiciel MALUNA [Her+02] qui intègre différents modèles de formes sur la circularité, le diamètre de disque et l'orientation d'ellipse.

39: D'une seule tâche ou en rajouter, telles que la tâche de recalage utilisée par [Slu+05] et [Mor16].

traitement d'images et de l'autre observent l'influence de la pathologie sur une méthode de segmentation simple. Parmi les lésions sources d'erreur citées par [Man+15], quatre d'entre elles sont observables chez les patients atteints du SDRA⁴⁰.

Globalement, les méthodes traditionnelles qui utilisent des informations explicites de forme, de connexité ou d'intensité, dépendent grandement de paramètres initialisés par l'utilisateur ou d'interactions manuelles.

2.3 Les méthodes par apprentissage profond

De la même manière que le traitement d'images a connu un essor avec l'augmentation des capacités de calcul des ordinateurs, le domaine de l'intelligence artificielle a grandement bénéficié du développement des cartes graphiques au début du 21^e siècle. Ce large domaine de l'informatique vise la production d'algorithmes imitant l'intelligence humaine (raisonnement, comportement, etc.).

L'une des approches, largement utilisée, pour atteindre cet objectif, est l'apprentissage automatique. Cette dernière s'appuie sur les statistiques et l'idée d'apprendre à partir des données à produire le résultat attendu plutôt que de le détailler mathématiquement. Pour ce faire, ces algorithmes apprennent à extraire ou reconnaître des caractéristiques dans les données. Malgré son nom, beaucoup des méthodes de l'apprentissage automatique restent dépendantes de l'intervention humaine⁴¹. L'une des solutions proposées à ce frein s'appuie sur l'idée d'imiter le réseau neuronal humain et donne tout naturellement naissance à la famille de méthodes que sont les réseaux de neurones. Bien que la théorie des réseaux de neurones ait précédé son utilisation massive⁴² à partir des années 2010, on peut noter la première application de cette méthode à la segmentation en 1990 [Bla+90]. Ce sous-ensemble de l'apprentissage automatique (voir Figure 2.11) permet l'extraction des caractéristiques dans les données de façon plus automatique et complexe. La quantité de couches dans un réseau de neurones permet de définir sa profondeur et la complexité (le niveau d'abstraction) des caractéristiques qu'il est capable d'extraire. Ainsi, l'apprentissage profond, méthode d'intérêt de cette partie, désigne le recours aux réseaux neuronaux avec un grand nombre de couches (plus d'une couche cachée).

Le principe du réseau de neurones est brièvement défini ci-après pour ensuite mettre l'accent sur les méthodes de l'apprentissage profond. Enfin, seront abordées les différentes avancées amenées par l'état de l'art d'une sélection de méthodes appliquées aux images médicales. En particulier, les méthodes appliquées à la segmentation des poumons seront discutées, notamment dans le cadre du SDRA.

Le principe

Le réseau de neurones (artificiels ou réseau neuronal) est le modèle de base sur laquelle est basé l'apprentissage profond. Il est inspiré du réseau neuronal humain et se compose de neurones (ou nœuds), de connexions (poids) qui les lient entre eux et de valeurs seuil. L'exemple simple d'un réseau de neurones à trois couches (voir Figure 2.12) permet de comprendre

40: Il s'agit des opacités en verre dépoli, des consolidations, des épanchements pleuraux et des nids d'abeille; avec aussi des lésions spécifiques aux patients atteints de SDRA COVID⁺.

Seul l'apprentissage profond sera traité dans cette partie. Les points essentiels à la compréhension de cette thèse seront détaillés.

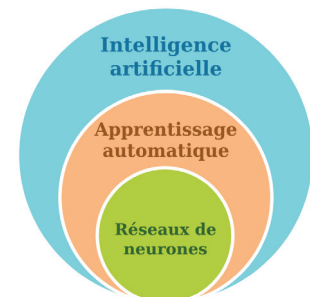


FIGURE 2.11 – Relation entre les domaines de l'intelligence artificielle, l'apprentissage automatique et les réseaux de neurones.

41: Par initialisation par exemple.

42: Dans tous les domaines utilisant une grande quantité de données. Par exemple la vision par ordinateur, le traitement du langage naturel ou encore la reconnaissance vocale.

Partie appuyée sur un ouvrage incontournable du domaine [Goo+16].

le principe de base de l'apprentissage profond. Les neurones sont organisés en couches afin d'accueillir en entrée les données (couche d'entrée), d'extraire les caractéristiques ensuite (couche(s) cachée(s)) et de fournir un résultat en sortie (couche de sortie). Chaque neurone de la couche cachée est activé ou non en fonction d'un calcul qui dépend des trois nœuds d'entrée (voir Figure 2.13). Cela signifie que le résultat de la fonction d'activation, détaillée dans la Figure 2.13, témoigne de la présence ou non de la ou les caractéristiques encodées par ce neurone. La fonction d'activation est généralement la même pour tous les neurones d'un réseau et permet d'encoder des caractéristiques complexes, grâce à la non-linéarité qu'elle introduit. Entraîner un réseau (ou modèle) actualise les poids de ses couches cachées, ce qui désigne la phase d'apprentissage. Dans le cas d'un apprentissage supervisé, ceci est fait en comparant le résultat en sortie, la prédiction, à la référence (une annotation). Lorsqu'il y a peu ou pas de référence, l'apprentissage peut être semi, faiblement, auto, ou non-supervisé en fonction de l'annotation utilisée. La comparaison est faite via une fonction de perte qui permet à un algorithme d'optimisation de modifier les poids en fonction de leur impact sur la prédiction. L'actualisation des poids qui découle du calcul de la perte calculée est appelée la rétropropagation⁴³. Une fois l'apprentissage effectué, les poids ne sont plus modifiés et prédire un nouveau résultat à partir d'une nouvelle donnée en entrée est désigné comme la phase d'inférence.

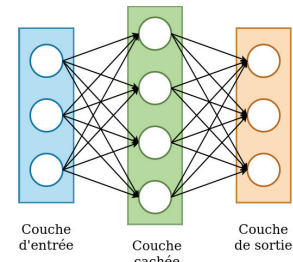


FIGURE 2.12 – Représentation d'un réseau neuronal à trois couches. – Les cercles blancs représentent les neurones liés par les flèches noires.

L'apprentissage profond désigne un réseau de neurones avec un grand nombre de couches cachées (au moins deux). Il n'est pas réservé à une tâche en particulier du traitement d'images. En effet, la nature de la prédiction dépend des annotations associées à l'image en entrée et de la définition de la couche de sortie. Appliqué à la segmentation, il se différencie des méthodes traditionnelles par le fait que le critère de segmentation⁴⁴ n'est pas utilisé de manière explicite. Le modèle d'apprentissage profond est entraîné afin de minimiser une erreur mesurée par une fonction de perte sur un jeu de données. Associer un neurone à chaque pixel ou voxel d'une image en entrée provoque cependant une quantité de calculs colossale. En effet, les poids reliant alors la couche d'entrée à la première couche cachée (pour une couche pleinement connectée) seraient très nombreux⁴⁵. Alors, pour pallier cet obstacle computationnel, l'opération de convolution est utilisée pour remplacer la fonction qui à chaque x_i en entrée associe le z correspondant au neurone de la couche cachée (voir Figure 2.13). L'image est alors considérée comme une entrée unique et les poids de chaque connexion (d'un réseau neuronal classique) deviennent les éléments des filtres utilisés dans l'opération de convolution. Ce type de réseau de neurones est appelé réseau neuronal convolutif et génère des cartes de caractéristiques Y au moyen de plusieurs filtres h de taille $M \times N$ qui balayent les pixels de l'image X et leur voisinage selon cette définition :

$$Y(i, j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} h(m, n) \cdot X(i - m, j - n), \quad (2.12)$$

avec (i, j) les coordonnées du pixel considéré et (m, n) les coordonnées⁴⁶ des poids du filtre h dans le cas d'une convolution 2D⁴⁷ (comme illustré dans la Figure 2.14). Cette opération est répétée plusieurs fois selon l'architecture, et en fonction de la taille de l'image en entrée et du nombre de filtres utilisés, peut devenir très coûteuse en mémoire pour la machine.

43: Cette étape se fait par calcul de gradients qui témoignent de l'effet de chaque poids dans le résultat obtenu avec la fonction de perte. L'actualisation des poids est donc corrélée avec leur effet (négatif ou positif) sur la perte calculée.

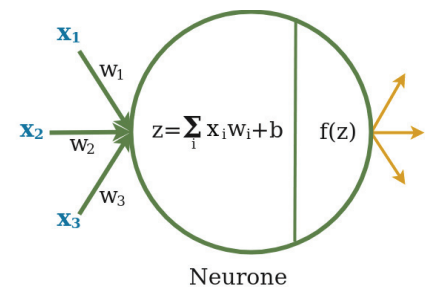


FIGURE 2.13 – Représentation d'un neurone. – Avec x_i les neurones d'entrée, w_i les poids de chaque connexion, b le biais et f la fonction d'activation.

44: Voir définition mathématique en Sous-section 13).

45: Une image de dimension 512x512 compte déjà 262 144 pixels.

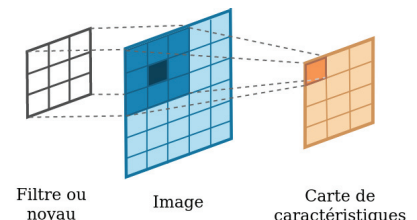


FIGURE 2.14 – Un élément (orange foncé) de la carte de caractéristiques (orange clair) obtenu par convolution 2D entre un filtre 3x3 (blanc) et le voisinage (bleu foncé) d'un pixel (noir) d'une image en entrée (bleu clair).

46: Dans le repère du filtre, la première valeur se situe aux coordonnées (0,0).

47: Sans pas de déplacement du filtre et sans remplissage des bords.

D'autant plus lorsque l'opération de convolution est 3D (comme l'illustre la Figure 2.15) il y a trois fois plus d'élément à prendre en compte pour la convolution d'un seul voxel.

$$Y(i, j, k) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \sum_{p=0}^{P-1} h(m, n, p) \cdot X(i - m, j - n, k - p) \quad (2.13)$$

Cependant, dans le but d'extraire les caractéristiques 3D dans les images (extraire de l'information du contexte volumique), ce sont souvent des modèles 3D qui sont utilisés pour la segmentation du poumon en entier [Ast+22].

Les données jouent un rôle essentiel dans l'apprentissage profond [Hof+20]. Au cours de l'entraînement (à chaque itération), elles sont présentées seules ou par lots⁴⁸. Une unité (sans dimension) est d'ailleurs définie en fonction des données, l'epoch, puisqu'elle indique chaque fois⁴⁹ que le modèle a utilisé toutes les données d'entraînement pour actualiser ses poids. À la fin d'une epoch est réalisée l'étape de validation, qui permet une fois l'entraînement terminé de choisir la meilleure optimisation des poids effectuée. La validation a lieu sur un sous-groupe des données, défini en amont de l'entraînement, qui n'est donc pas utilisé pour la rétropropagation. Ce sous-groupe, appelé sous-groupe de validation, permet aussi de vérifier en fin d'entraînement que le modèle est correctement⁵⁰ entraîné. Enfin, un dernier sous-groupe de données, jamais utilisées à l'entraînement, permet de garder des données pour tester les performances du modèle en inférence (après l'entraînement), il s'agit du sous-groupe de test. La partition usuelle d'une base de données évaluées sur elle-même est autour de 70%, 10% et 20% respectivement pour les sous-groupes d'entraînement, de validation et de test (voir Figure 2.16).

Définir ces différents paramètres, appelés hyperparamètres⁵¹, qui conditionnent l'apprentissage, est à la charge de l'utilisateur.

Mots clés de l'apprentissage profond	
Données	pré-traitement, partition en sous-groupes (entraînement, validation)
Architecture ou modèle	organisation et composition des couches de convolution
Fonction de perte	fonction dérivable qui permet le calcul des gradients
Optimisation	algorithme qui à partir du calcul des gradients produit la rétropropagation (actualisation des poids)
Stratégie d'entraînement	supervisé, faiblement supervisé, auto supervisé

Les différents modèles

Il existe une très grande variété de réseaux neuronaux dédiés à la segmentation des images [Min+21] et leur application au domaine de l'imagerie médicale a connu un essor grandissant depuis le milieu des années

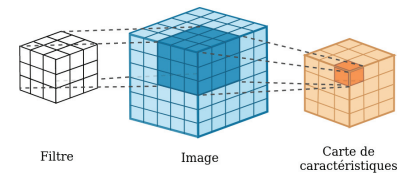


FIGURE 2.15 – Effet d'un filtre 3 × 3 × 3 de convolution 3D sur le voxel d'une image 3D en entrée, sans pas de déplacement et sans remplissage des bords.

48: Pour profiter de l'accélération des calculs matriciels sur carte graphique.

49: L'entraînement est délimité par le nombre d'epoch utilisées pour actualiser les poids du modèle.

50: Sans surapprentissage (correspondance trop forte aux données d'entraînement) et capable de généraliser (s'adapter à des données du même type que l'entraînement mais jamais vues). Le surapprentissage est observé lorsque la courbe de perte décroît pour le sous-groupe d'entraînement mais croît pour le sous-groupe de validation.

51: Par exemple la taille des lots, le nombre d'epochs, le taux d'apprentissage qui conditionne l'algorithme d'optimisation, la fonction de perte, etc.

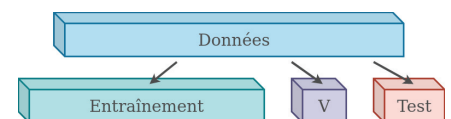


FIGURE 2.16 – Découpage de données en trois sous-groupes : celui d'entraînement, de validation (V) et de test.

2010 [Lit+17] (2018 pour la segmentation des poumons [Ast+22]). Comme rappelé ci-dessus dans les principes clés de l'apprentissage profond, nombreux sont les aspects qui peuvent amener à une amélioration. La plupart des contributions de l'état de l'art ont d'abord porté sur l'architecture, bien qu'au fil des années l'intérêt donné aux méthodes d'entraînement et d'optimisation ait grandi [Asg+21]. Ainsi, cette partie se concentre sur la présentation des catégories de modèles utilisés pour la segmentation, et plus spécifiquement celle des poumons en imagerie CT.

Les plus répandus sont les réseaux neuronaux convolutifs [Ast+22], notés CNN⁵², qui se composent successivement de couches de convolution jusqu'à fournir un vecteur de sortie.⁵³ Ce que l'on appelle couche de convolution (voir Figure 2.17) comprend un ou plusieurs filtres de convolution appliqués à l'image pour fournir une carte de caractéristiques, une fonction d'activation non linéaire qui permet d'augmenter la complexité des liens établis entre les caractéristiques et une fonction de réduction de dimension spatiale⁵⁴ pour accélérer les calculs et permettre la généralisation. La Figure 2.18 représente schématiquement l'effet de chaque élément de la couche de convolution sur une image donnée en entrée. Cet enchaînement de couche de convolution crée progressivement des cartes de caractéristiques de plus en plus abstraites à partir des caractéristiques observées dans l'image. Cependant, cette architecture (voir Figure 2.18) correspond davantage à la tâche de classification, étant donné qu'elle impose une taille d'image à l'entrée pour correspondre à la sortie unidimensionnelle. Appliquée comme telle à la tâche de segmentation, il faut considérer des patches (voir Figure 2.26 pour illustration) dans l'image pour prédire successivement la classe de chaque pixel central du patch. Un patch représente une portion⁵⁵ de l'image à traiter, qui permet de diminuer la taille de l'objet en entrée de la chaîne de traitement tout en maintenant la résolution spatiale de l'image d'origine. C'est une approche qui apparaît dans la littérature [Xu+19], mais qui a été appliquée à des poumons avec peu de lésions (cancer des poumons et maladies pulmonaires obstructives chroniques). Les performances de cette méthode sont alors de 0.99, 0.90 et 0.96 en termes de spécificité, sensibilité et score de Dice, pour un temps d'inférence de 11 s par image. Segmenter via des patches reste un défi non résolu dans certains cas. D'une part, car la consistance spatiale globale n'est pas garantie et d'autre part, car différentes zones anatomiques peuvent avoir la même apparence dans un patch.

Fully Convolutional Neural Networks (fCNN) Pour pallier la contrainte de taille d'entrée fixe des CNN, un remplacement de la dernière couche totalement connectée par une couche de convolution a été proposé [Lon+15] : c'est le fCNN (voir Figure 2.19). Cette modification permet de mieux correspondre à la tâche de segmentation puisque pour une image en entrée, le modèle peut directement prédire une image en sortie dans laquelle une classe est attribuée à chaque pixel, bien que la cohérence spatiale entre les pixels reste faible. Plusieurs modifications de l'architecture visent à pallier ce manque. On peut citer les modèles DeepLab v1 [Che+14], DeepLab v2 [Che+16], DeepLab v3 [Che+17] et v3+ [Che+18]. Utilisé pour la segmentation de poumons sains [Zho+17], c'est une architecture qui fournit de bons résultats avec $JI = 0.903 \pm 0.037$ sur les 36 images CT de 12 patients (sous-groupe de test). Appliqué à la segmentation des poumons malades sur des images CT, ce type de modèle utilise beaucoup d'ajustements

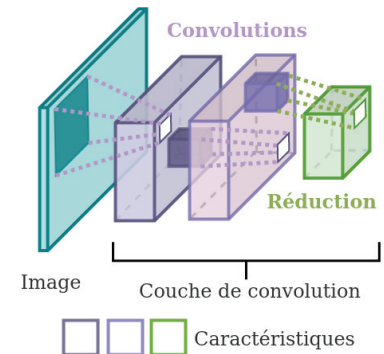


FIGURE 2.17 – Détail d'une couche de convolution dans un CNN. – Les résultats successifs de plusieurs filtres de convolution sur l'image d'entrée, puis sur la 1^{re} carte de caractéristiques (en violet foncé), puis le résultat de la fonction de réduction (en vert). Chaque pavé représente une carte de caractéristiques.

52: pour *Convolutional Neural Networks*.

53: Dit élément totalement connecté du fait que chaque élément du vecteur dépend de tous les éléments de la couche précédente.

54: Les stratégies de réduction de dimension les plus courantes sont le regroupement par moyenne ou par maximum.

55: De même dimension (2D ou 3D) que l'image dont il est issu.

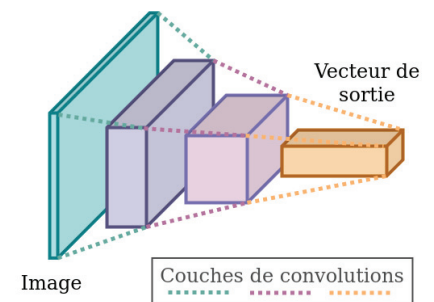


FIGURE 2.18 – Architecture d'un réseau CNN.

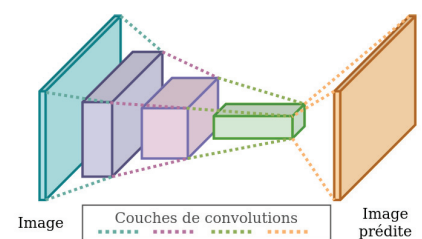


FIGURE 2.19 – Architecture d'un réseau fCNN.

[Zha+18]. Toutes les versions, améliorées au fil du temps de l'architecture DeepLab, ont l'inconvénient de demander beaucoup de ressources de calcul et d'être peu précises sur les contours des segmentations [Mal+22].

Encodeur-Décodeur Dans la même optique, un autre type de modèle intègre l'information spatiale en deux étapes pour produire une prédiction de même dimension que l'entrée. La première partie du réseau s'apparente à un *fCNN* et permet d'extraire une représentation sémantique compressée de l'image qui est appelée l'espace latent. Cette partie appelée encodeur est suivie par un ensemble de couches⁵⁶, le décodeur, qui permet de reconstruire l'image à partir de l'espace latent. Ce type de modèle est de ce fait appelé encodeur-décodeur (voir Figure 2.20).

Un modèle d'encodeur-décodeur aujourd'hui largement utilisé est le U-net [Ron+15] qui a été développé spécifiquement dans le cadre de l'imagerie médicale. Son apport à l'architecture classique des encodeur-décodeur est d'ajouter les cartes de caractéristiques de moins en moins abstraites (et de plus en plus localisées) au fur et à mesure de la reconstruction spatiale faite par le décodeur, ce qui renforce la cohérence spatiale des caractéristiques encodées. Cette architecture, initialement prévue pour des images 2D, a connu beaucoup d'ajustements donnant des modèles aujourd'hui très utilisés : U-Net 3D [Çiç+16], U-Net++ [Zho+18], Attention U-Net [Okt+18], V-Net [Mil+16] entre autres. L'une des variations du U-net a particulièrement marqué la communauté en 2020 en remportant la première place du challenge *Medical Segmentation Decathlon* [Ant+21], le modèle nommé nnUNet [Ise+21]. Ce dernier intègre les étapes usuellement fastidieuses de recherche des hyperparamètres, ainsi que le choix de l'architecture optimale entre un modèle U-net 2D, un U-net 3D et un U-net en cascade⁵⁷.

Les méthodes spécifiques au SDRA

Parmi toutes les méthodes citées précédemment, on retrouve peu de modèles qui ont été appliqués à la segmentation des poumons avec SDRA. En effet, une recherche de certains mots clés dans le moteur de recherche de données bibliographiques PubMed®[22], spécialisé dans la biologie et la médecine, permet de mettre en évidence ce manque. D'un côté, la Figure 2.22 montre le nombre d'articles scientifiques répertoriés dans la base de données MEDLINE qui contiennent les mots *lung segmentation* (segmentation du poumon), *deep learning* (apprentissage profond) et *CT* (TDM) entre 2016 et 2022 (fin août). La quantité d'articles scientifiques utilisant ces trois mots clés est croissante depuis 2016, avec une forte augmentation depuis 2020 (près de 95% des articles publiés après 2019). De l'autre côté, la quantité d'articles scientifiques qui, en plus de ces trois mots, mentionnent le SDRA (*ARDS* en anglais) est nettement plus faible (voir Figure 2.23). Si l'on ajoute à cette recherche le terme COVID (pour qu'elle contienne *ARDS* ou *COVID*), on obtient alors 142 résultats de plus (voir Figure 2.24). En effet, comme présenté dans le Chapitre 1, la pandémie du SARS-Cov-2 a introduit une maladie pulmonaire dont les formes graves sont proches du SDRA. Parmi les 146 citations ainsi répertoriées, 142 sont des articles (et 4 des résumés) et 92 sont appliquées à des patients plutôt qu'à des modèles expérimentaux. Malgré les mots clés utilisés, certains de ces articles n'ont pas pour finalité la segmentation des poumons, mais

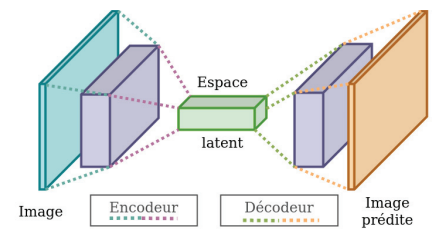


FIGURE 2.20 – Architecture d'un encodeur-décodeur.

56: Couches de déconvolution, opération inverse de la convolution, et d'augmentation de dimension. L'augmentation de dimensions est usuellement réalisée par convolution transposée.

57: Architecture utilisant deux U-net liés en série dont chacun fournit une tâche complémentaire.

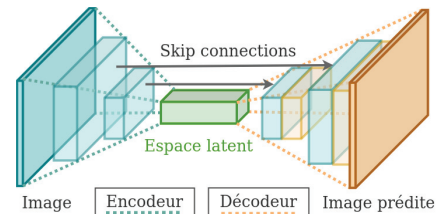


FIGURE 2.21 – Représentation simplifiée de l'architecture U-net (seules deux couches de convolutions représentées sur les quatre).

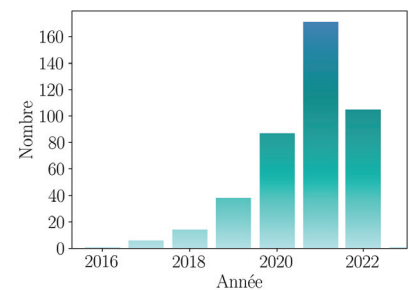


FIGURE 2.22 – Compte des articles internationaux contenant les mots *deep learning*, *lung segmentation* et *CT* entre 2016 et 2022 (Source : PubMed®, 28/08/22).

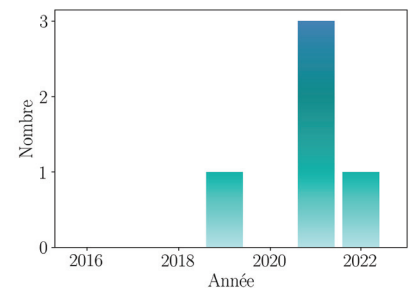


FIGURE 2.23 – Compte des articles internationaux contenant les mots *deep learning*, *lung segmentation*, *CT* et *ARDS* entre 2016 et 2022 (Source : PubMed®, 28/08/22).

plutôt celle des lésions pulmonaires (qui nécessite la segmentation des poumons en tant qu'étape préliminaire).

Ainsi, à la fin de l'année 2019 et au début du travail bibliographique de cette thèse, le seul article appliquant l'apprentissage profond à la segmentation de poumons avec SDRA est le travail de l'équipe américaine de Reinhardt [Ger18; Ger+19]. Leur objectif? Créer un algorithme robuste pour la segmentation des poumons, applicable chez différentes espèces de mammifères avec des modèles expérimentaux de SDRA. Leur stratégie se base alors sur un modèle multi-résolution, appuyé sur l'hypothèse qu'il est nécessaire d'intégrer l'information globale, à savoir les organes voisins, pour segmenter précisément les poumons avec lésions. Ils utilisent dans un premier temps la technique du transfert d'apprentissage,⁵⁸ en entraînant leur réseau de segmentation (ConvNet – voir Figure 2.25) sur une base de données privée de 3418 CT de poumons humains (avec lésions telles que BPCO⁵⁹ et FPI⁶⁰). Ces deux pathologies entraînent peu ou pas du tout de perte d'aération (contrairement au SDRA), ce qui se traduit par des poumons plus contrastés dans les images et donc plus simples à segmenter⁶¹. Un des enjeux associé à l'utilisation d'un tel réseau pour la segmentation des images médicales, est qu'il requiert une grande capacité de mémoire pour des images en haute résolution (voir Figure 2.26 a). Généralement, afin de diminuer la taille des données, changer de résolution est la solution utilisée (voir Figure Figure 2.26 b-e). Dans le cas de la basse résolution, toute l'image est utilisée, mais le nombre de voxels est réduit en combinant les intensités de plusieurs voxels voisins (sous-échantillonnage). Une autre approche consiste à rogner l'image et utiliser uniquement une région d'intérêt, un patch, cela pour garder la pleine résolution tout en diminuant le nombre de voxels analysés (voir Figure Figure 2.26 c-e). L'approche choisie par [Ger+19] est d'utiliser à la fois la pleine résolution et la basse résolution, pour garder effectivement toute l'information de l'image (contexte des poumons).

Une fois les poids du réseau pré-entraînés sur cette base de données, ils sont affinés sur une base de données plus petite (301 images CT 3D) contenant uniquement des images de mammifères avec modèle expérimental de SDRA. Le recours au transfert d'apprentissage est justifié par la faible quantité d'images CT de poumons annotés chez les mammifères, contrairement aux CT de patients utilisés dans cet article⁶². Leurs masques de référence sont constitués très finement par des experts (entre 4 et 6 heures par image CT 3D), qui corrigent sur un logiciel le résultat d'une segmentation semi-automatique fourni par PASS (*Pulmonary Analysis Software Suite*, University of Iowa Advanced Pulmonary Physiomic Imaging Laboratory). L'architecture du réseau utilisé (Seg3DNet) est un encodeur-décodeur qui prend en entrée une image CT 3D et ressort une carte de probabilité d'appartenance aux poumons. Le modèle basse résolution est d'abord entraîné avec des images sous-résolues (voir Figure 2.26 b). Puis le modèle haute résolution, utilise en entrée une combinaison de contextes fournis par la sortie du modèle basse résolution et un pavé contenant les détails à haute résolution (voir Figure 2.26 c).

En ce qui concerne l'évaluation, les méthodes haute et basse résolution sont comparées indépendamment à la méthode multi-résolution. L'index de Jaccard et l'ASSD sont utilisés pour l'évaluation des performances de la méthode. La méthode multi-résolution fournit de meilleurs résultats⁶³ sur ces trois métriques que les méthodes haute⁶⁴ et basse résolution.⁶⁵ De

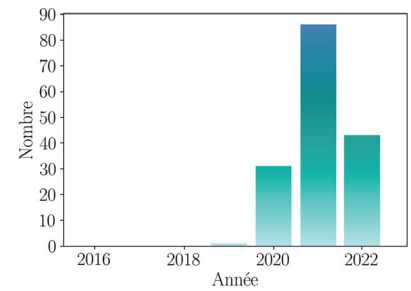


FIGURE 2.24 – Compte des articles internationaux contenant (*deep learning, lung segmentation, CT*) et (*ARDS ou COVID*) entre 2016 et 2022 (Source : PubMed®, 28/08/22).

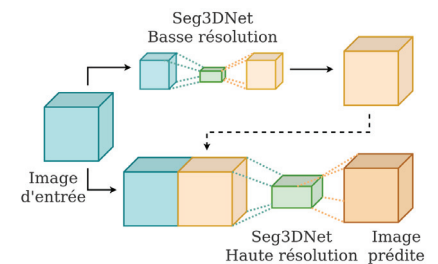


FIGURE 2.25 – Architecture ConvNet proposée par Gerard et al. [Ger18].

58: Technique qui consiste à pré-entraîner un modèle sur une jeu de données très conséquent, mais différent de la tâche d'intérêt. Le modèle est ensuite spécialisé sur un nombre inférieur de données correspondant cette fois-ci à l'application recherchée.

59: Bronchopneumopathie chronique obstructive.

60: Fibrose pulmonaire idiopathique.

61: Avec des méthodes basées sur le seuillage par exemple, voir section 2.2.

62: Leur grand nombre a notamment été facilité par le recours à la segmentation automatique sur des images de patients ne présentant que peu ou pas de lésions denses.

63: $JI=0.963 \pm 0.025$; $ASSD = 0.438 \pm 0.315mm$.

64: $JI = 0.950 \pm 0.036$; $ASSD = 0.657 \pm 0.519mm$.

65: $JI = 0.919 \pm 0.027$; $ASSD = 0.971 \pm 0.368mm$.

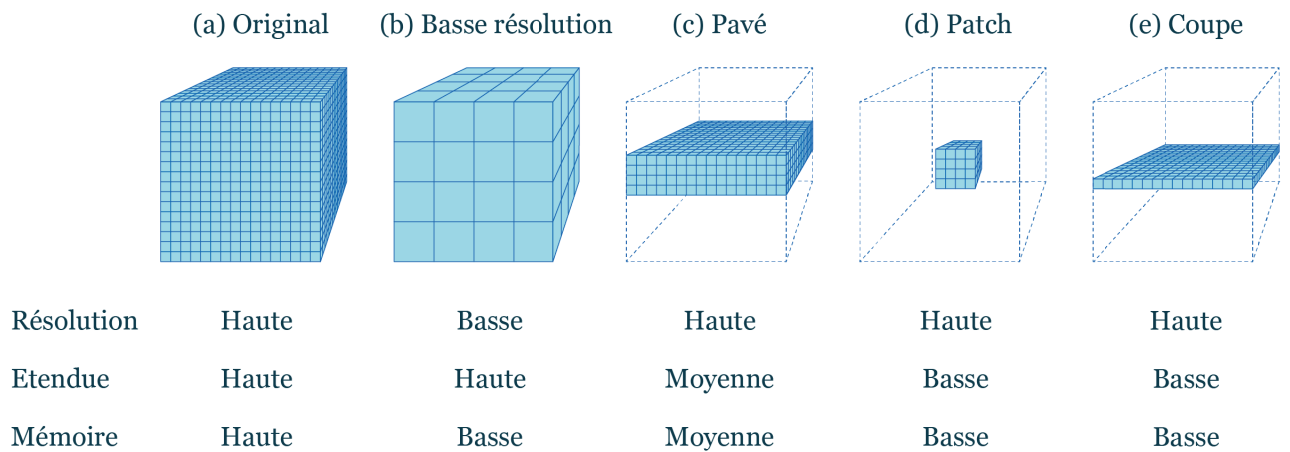


FIGURE 2.26 – Illustration de différentes techniques d'échantillonnage d'une image permettant de réduire la mémoire nécessaire pour l'entraînement d'un modèle d'apprentissage profond. – Résolution : nombre de points par unité de longueur physique ; Etendue : champ de vue ; Mémoire : espace mémoire en machine. – Image reproduite depuis [Ger18].

manière qualitative, le modèle haute résolution fournit des segmentations plus fines que les deux autres modèles. Le modèle basse résolution, quant à lui, fournit des résultats plus robustes (pour le même niveau de précision) peu importe la sévérité des lésions présentes. Cette robustesse est due au fait que le modèle basse résolution repose sur le contexte global et les caractéristiques anatomiques (comme la position relative des organes), informations qui restent cohérentes peu importe le niveau d'aération. C'est d'ailleurs ce qui permet à cette approche multi-résolution d'être aussi robuste au changement d'espèce.

Ces résultats sont encourageants quant à l'application de l'apprentissage profond pour la segmentation des poumons dans le cadre du SDRA. Pourtant, l'évaluation ne montre pas l'intérêt spécifique de l'étape de pré-entraînement via le transfert d'apprentissage. Pour cela, une comparaison avec le modèle entraîné uniquement sur les données de patients aurait été nécessaire. De plus, l'évaluation est réalisée sur des modèles expérimentaux du SDRA qui sont très différents du SDRA observé chez l'humain.

De nombreux articles sur la segmentation de poumons avec pathologies ont été publiés à partir de 2020, notamment dans le cadre de l'effort international de recherche sur la pandémie du COVID-19. Ce contexte particulier a plutôt concentré les efforts sur la segmentation et la classification des lésions du COVID-19 [Moh+20; Val+21; Zha21; Sur+21]. De plus, les images utilisées étaient souvent celles de contrôle, réalisées à l'admission à l'hôpital dans le but de détecter la maladie.

Plus récemment, des équipes internationales ont appliqué des méthodes d'apprentissage profond spécifiquement à la segmentation des poumons chez des patients avec SDRA. Cependant, au vu de leur date de publication, ces dernières n'ont pas influencé le travail de cette thèse et seront donc abordées dans la discussion.

Les bases de données publiques de poumons segmentés

Les bases de données constituent l'un des leviers des méthodes de segmentation basées sur l'apprentissage profond. La qualité des images, leur quantité ainsi que la diversité des sujets et lésions représentés sont autant de caractéristiques recherchées pour maximiser l'apprentissage des modèles [Hof+20]. Réunir toutes ces caractéristiques est pourtant difficile dans le cas des données médicales. D'une part à cause du temps et de l'expertise que demandent les annotations manuelles et d'autre part pour des raisons de protection des données. Les quelques bases de données qui sont publiques ne sont généralement pas constituées spécifiquement pour la segmentation des poumons, mais plutôt pour celle des lésions⁶⁶. De ce fait, les masques de référence des poumons sont souvent absents ou bien servent uniquement d'initialisation⁶⁷ pour une autre tâche. Lorsqu'ils sont présents, ils sont soit réalisés sur un nombre limité de coupes ou bien l'image est sous-résolue pour réduire le nombre total de coupes. Le résumé des bases de données publiques contenant des images CT de poumons est présenté dans le Tableau 2.2.

66: Nodules, lésions pulmonaires caractéristiques du COVID-19, etc.

67: Comme c'est le cas du challenge VESSEL12.

TABLEAU 2.2 – Bases de données publiques avec masques de segmentation des poumons.

Base de données	Nombre d'images	Commentaires
The LObe and Lung Analysis 2011 (LOLA11) [Gin21]	55	
VESSel SEGmentation in the Lung 2012 Challenge (VESSEL12)	20	Masques fournis comme initialisation pour la segmentation des vaisseaux (peu précis).
VISCERAL Anatomy3	20	
Lung CT Segmentation Challenge 2017 (LCTSC)	60	Format des segmentations peu commun
TCIA CT Images in COVID-19	120	Masques de référence des lésions
COVID-19 CT segmentation dataset (medicalsegmentation)	64	Tâche de segmentation des lésions

2.4 Conclusion

Après avoir présenté les concepts utiles à la compréhension des enjeux techniques de cette thèse, une vue d'ensemble de l'état de l'art a permis de mettre en évidence l'aspect prometteur de l'apprentissage profond. Cette direction s'appuie avant tout sur la difficulté présentée par la littérature à identifier une méthode de traitement d'images traditionnelle qui soit robuste aux différentes lésions et anatomies rencontrées dans l'imagerie des poumons de patients avec SDRA.

Choisir l'apprentissage profond comme solution technique implique néanmoins des enjeux supplémentaires que sont : recueillir une grande quantité de données annotées, de même que disposer de ressources de calcul importantes en fonction des architectures utilisées. En effet, l'unique article paru au début de cette thèse, appliquant l'apprentissage profond à la segmentation de poumons avec SDRA utilise des milliers d'images CT⁶⁸ avec poumons annotées. Leur architecture multi-résolution [Ger18], vise alors

68: De patients et aussi de mammifères avec modèle expérimental de SDRA.

à exploiter autant les détails présents dans les pavés en haute résolution que le contexte tiré des images basses résolutions. Ce compromis entre résolution des données utilisées à l'entraînement et robustesse souhaitée est nécessaire du fait de la charge computationnelle que représente l'entraînement d'un tel modèle.

Inspiré des stratégies de cette équipe et au regard des commentaires relevés par la littérature, l'accent sera mis dans cette thèse sur l'utilisation optimale d'une architecture légère, mais robuste. En effet, avec la quantité de données annotées disponibles, il s'agira d'étudier si une gestion optimisée de ces données permet de dépasser les limites matérielles⁶⁹ et atteindre une robustesse et une capacité de généralisation utilisable en clinique. Dans un premier temps, l'approche sera 2D⁷⁰ et évaluera l'intérêt d'utiliser des images provenant de plusieurs anatomies pour augmenter la robustesse. Puis, dans un second temps, une architecture 3D sera utilisée pour permettre d'exploiter d'autres caractéristiques des données, et garantir davantage de cohérence spatiale.

69: Ressources de calcul et quantité de données annotées.

70: Notamment pour la rapidité de traitement que cela permet.

Approche 2D : gestion de données

La recherche d'une stratégie optimale d'utilisation des données a été choisie comme direction pour cette thèse. Ce chapitre s'attache à évaluer s'il existe une manière de présenter des images d'anatomies différentes qui renforce la robustesse du modèle entraîné. Cette question s'inspire de travaux appliquant le transfert d'apprentissage pour la segmentation des poumons de mammifères avec modèles expérimentaux de SDRA [Ger+19]. Pour y répondre, deux types de combinaisons des données sont évalués : le transfert d'apprentissage et le mélange des données pendant l'entraînement (avec des données de modèles porcins expérimentaux de SDRA et de patients avec SDRA). Le modèle d'apprentissage profond proposé utilise des coupes d'images 2D dans les trois directions orthogonales (axiale, coronale, sagittale) afin d'être léger, mais d'utiliser une part de contexte volumique [Pen+22]. Les résultats ne montrent pas de gain significatif des différentes stratégies de combinaison des données évaluées. Cependant, cette expérience répond déjà en partie (temps de calcul, robustesse) aux objectifs de cette thèse présentés dans les chapitres précédents.

Table des matières

3.1	Introduction	35
3.2	Données	37
3.3	Description d'expérience	40
	Modèle	40
	Stratégie d'entraînement	40
	Évaluation	42
3.4	Résultats	43
	Entraînements de référence	43
	Entraînements combinés	44
	Entraînements avec transfert d'apprentissage	45
	Résultats globaux	46
3.5	Discussion	47
3.6	Conclusion	52

3.1 Introduction

Depuis près d'une décennie, des travaux sur la segmentation des poumons avec SDRA ont été menés au sein du laboratoire CREATIS [Mor16; Mor+17; Ork+19]. Les images utilisées provenaient de diverses études expérimentales sur modèles porcins¹ de SDRA. Une grande quantité de données par sujet était disponible du fait des protocoles expérimentaux, sur laquelle s'est appuyée la méthode de segmentation² ainsi développée [Mor+17]. Cette méthode n'étant pas répliquable dans le cas des images de patients, et au vu de la littérature, l'utilisation de l'apprentissage profond a été choisie pour poursuivre la recherche sur la problématique de la

Dans ce chapitre la stratégie d'apprentissage étant 2D, la précision entre image 2D ou 3D sera faite lorsque cela peut porter à confusion.

1: Un modèle animal décrit l'étude d'un processus pathologique, spontané ou induit, similaire chez un humain, sur un animal non humain.

2: Méthode utilisant le recalage d'une segmentation faite sur l'image la plus contrastée sur les images moins contrastées.

segmentation des poumons dans le cadre du SDRA.

Le travail d'une équipe américaine (Iowa) paru fin 2019 [Ger+19] a alors particulièrement inspiré les axes de recherche de cette thèse. Premier article de la littérature à appliquer l'apprentissage profond à la segmentation de poumons dans le cadre du SDRA, l'utilisation de données animales et patients y est proposée pour atteindre une meilleure robustesse. La méthode utilisée pour parvenir à apprendre à partir de diverses anatomies³ est le transfert d'apprentissage. Cela leur a permis d'utiliser dans un premier temps⁴ une grande quantité d'images de patients avec peu de lésions denses, puis, dans un deuxième temps⁵, d'apprendre sur une quantité réduite de données de mammifères avec modèle expérimental de SDRA. Malgré des résultats encourageants, leur étude n'évalue pas l'apport spécifique d'utiliser le transfert d'apprentissage⁶. L'anatomie ciblée dans cette étude diffère également par rapport à celle de cette thèse, les modèles expérimentaux de SDRA étant très différents dans la forme qu'ils prennent par rapport au SDRA observé chez l'humain.

L'idée développée dans ce chapitre est alors d'évaluer l'apport spécifique du transfert d'apprentissage dans le cadre du SDRA. L'utilisation de modèles expérimentaux de SDRA pour le pré-entraînement peut-elle renforcer la robustesse d'un modèle ensuite spécialisé pour la segmentation de poumons de patients avec SDRA ?

Segmenter les poumons est une tâche tridimensionnelle. Pourtant, les ressources de calcul nécessaires pour appliquer l'apprentissage profond aux images 3D sont conséquentes et cela limite⁷ la taille des données utilisables [Çiç+16] durant l'entraînement et l'inférence. Plusieurs stratégies existent dans la littérature, pour parvenir à un compromis et s'intéressent tantôt aux données, tantôt aux modèles. Lorsque le choix est porté sur les données, considérer une partie des volumes ou diminuer leur taille permet de diminuer les ressources de calcul nécessaires. C'est pourquoi une branche de l'apprentissage profond s'est axée sur l'utilisation de modèles dits 2.5D : ces modèles utilisent plusieurs coupes d'un volume pour réaliser la segmentation 2D. L'idée est de dépasser le frein des ressources de calcul lourdes⁸ sans pour autant perdre l'information tridimensionnelle contextuelle⁹.

Ce type de modèle, considéré comme une sous-catégorie des méthodes multi-flux [Lit+17], a été choisi pour répondre à la problématique de cette première expérience. Il existe différentes façons de sélectionner les coupes à traiter à l'apprentissage. On peut séparer les modèles d'apprentissage 2.5D en trois catégories selon qu'ils utilisent les trois plans orthogonaux classiques (axial, coronal, sagittal) [Pra+13; Lyk+15; Moe+16], des plans parallèles¹⁰ [War+21] ou de multiples plans d'orientations aléatoires [Rot+14; Per+19]. Quelle que soit l'orientation ou la position relative choisies des plans, le champ de vue peut lui aussi varier selon les études. Certains considèrent la totalité de la coupe choisie [Per+19; War+21], tandis que d'autres sélectionnent une sous-partie de chaque coupe choisie [Pra+13; Lyk+15; Rot+14; Moe+16]. Ces deux approches visent à réaliser le compromis entre la quantité d'information traitée et la capacité de calcul nécessaire pour le faire. Dans le premier cas, prendre en compte toute la coupe permet d'utiliser un large contexte durant l'apprentissage. La deuxième approche, bien qu'elle réduise le contexte utilisé, permet de diminuer par la même occasion la charge computationnelle. Des méthodes d'augmentation de

3: Modèles canins, ovins et porcins de SDRA et anatomie humaine avec autres pathologies pulmonaires.

4: Pour le pré-entraînement.

5: Pour la spécialisation du modèle.

6: En comparant des résultats du modèle avec ou sans pré-entraînement.

7: Un compromis est alors réalisé entre la dimension des données et la précision des modèles entraînés.

8: En utilisant des modèles 2D, beaucoup plus légers en termes de paramètres et calculs.

9: Qui est alors compensée par l'utilisation des diverses coupes.

10: Qui peuvent former des tranches si les coupes sont adjacentes.

données sont alors utilisées pour compenser cette perte d'information et améliorer la robustesse de l'apprentissage [Rot+14; Per+19].

Combiner les informations des diverses coupes peut se faire à différents stades de la chaîne de traitement. Une stratégie consiste à grouper l'information pendant l'apprentissage, en considérant chaque coupe comme un canal différent de la même image d'entrée du modèle [Rot+14; Per+19; War+21]. Une autre approche est d'apprendre séparément¹¹ sur chaque type de coupe avant d'utiliser la dernière couche du réseau pour fusionner l'information [Pra+13; Moe+16]. Enfin, une dernière stratégie, semblable à la précédente, réalise la fusion en considérant les différentes prédictions comme un ensemble dont la solution est choisie de manière déterministe [Lyk+15].

11: Via des modèles indépendants ou des branches indépendantes d'un même modèle.

3.2 Données

Les bases de données utilisées pour mener à bien cette expérience proviennent de deux études expérimentales et une étude clinique.

La première étude expérimentale, nommée PK HIDRA, compte sept cochons, ayant tous suivi le même protocole d'agression pour reproduire les effets d'un SDRA modéré à sévère, à savoir l'inhalation d'acide chlorhydrique. Les images ont été acquises sur un scanner BioGraph TruePoint ®PET/CT (Siemens, Munich, Allemagne), avec comme espacement moyen de $[0.6, 0.6, 1]$ en millimètres, respectivement dans les directions $[x, y, z]$. En fonction du cochon, le nombre de coupes axiales varie de 298 à 380 pour une taille fixe de coupe axiale de 512×512 . Pour chaque cochon, plusieurs images ont été acquises à différents niveaux d'atteinte (voir Figure 3.1) dans le but de mesurer les effets de différents réglages ventilatoires.

Toutes les bases de données utilisées dans les expériences suivantes ont été approuvées par un comité d'éthique.

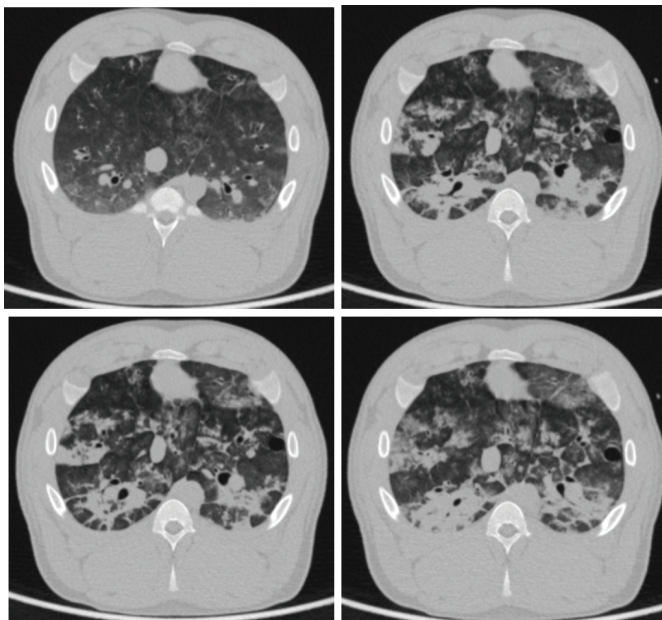


FIGURE 3.1 – Vue d'un même cochon avec modèle de SDRA à 4 niveaux d'atteintes différents (PK HIDRA).

La deuxième étude expérimentale est nommée Hamilton et compte également sept cochons. Le lavage salin y est utilisé pour reproduire un

SDRA sévère, puis différentes PEP sont appliquées de manière décroissante (16, 12, 8 et 4 cmH_2O). Pour chaque condition ventilatoire (référence, 15 min après les critères du SDRA, 10 min après stabilisation pour chaque PEP), une paire d'images CT (fin d'inspiration et fin d'expiration) est acquise sur un scanner Biograph mCT (Siemens, Munich, Allemagne), avec comme espacement moyen [0.8, 0.8, 1] en millimètres, respectivement dans les directions [x, y, z]. En fonction du cochon, le nombre de coupes axiales varie de 269 à 379 pour une taille fixe de coupe axiale de 512×512 . Les atteintes présentes dans cette base sont plus sévères que dans la base PK HIDRA (voir Figure 3.2).

Que ce soit dans la première ou la deuxième étude expérimentale, le nombre d'images acquises par sujet a varié, limité par la survie des sujets aux agressions utilisées.

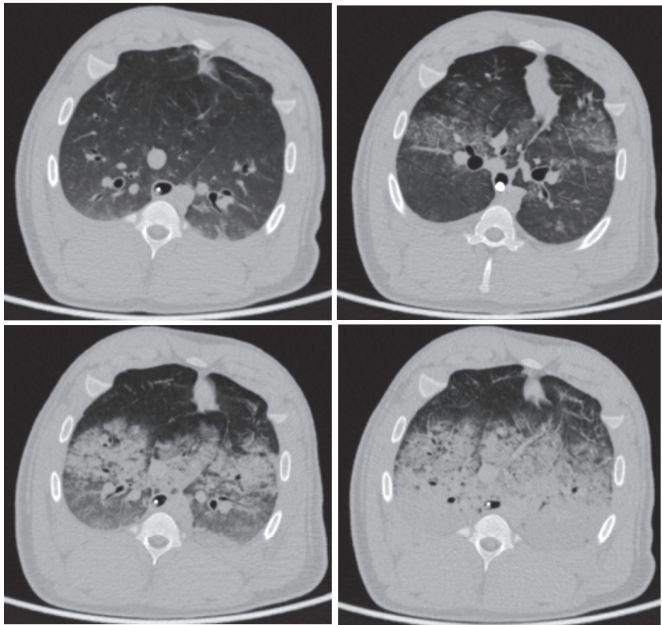


FIGURE 3.2 – Vue d'un même cochon avec modèle de SDRA à 4 niveaux d'atteintes différents (Hamilton).

L'étude clinique CT4-ARDS (enregistrée sous l'identifiant NCT03870009) est, quant à elle, composée de 30 patients admis au service de réanimation de l'hôpital de la Croix-Rousse (Hospices Civils de Lyon) [Cha+20]. Pour chaque patient, deux images ont été acquises, l'une en fin d'expiration et l'autre en fin d'inspiration, dans le but de mesurer l'hyperinflation cyclique. Réalisée entre mai 2019 et mai 2020, les patients composant cette base de données sont pour 50% atteints d'un SDRA sévère et pour 50% atteints d'un SDRA modéré. Parmi les 30 patients étudiés, 60% étaient positifs à la maladie COVID-19. Les images ont été acquises sur un scanner iCT 256 ou bien Ingenuity CT (Philips Healthcare, Eindhoven, Pays-Bas) avec comme espacement moyen [0.7, 0.7, 1] en millimètres, respectivement dans les directions [x, y, z]. En fonction du patient, le nombre de coupes axiales varie de 230 à 323 pour une taille fixe de coupe axiale de 512×512 .

Toutes les segmentations de référence ont été réalisées par des cliniciens du service de réanimation du CHU de la Croix-Rousse. Un logiciel développé au laboratoire CREATIS, creaSDRA¹², [Dáv+20] a permis aux experts de segmenter semi-automatiquement les masques des poumons.

12: Qui est utilisé pour réaliser la segmentation semi-automatique mais aussi à partir de cette dernière calculer les mesures d'aération.

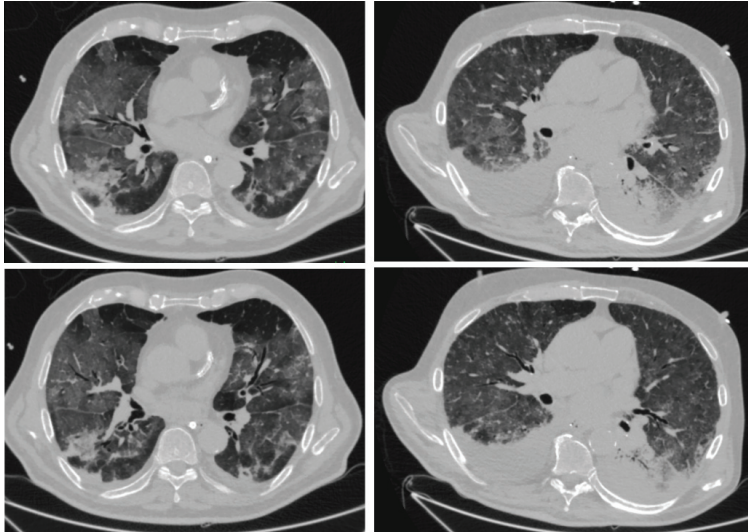


FIGURE 3.3 – Vue de 2 patients avec SDRA en fin d’expiration (en haut) et en fin d’inspiration (en bas) de la base de données CT4-ARDS.

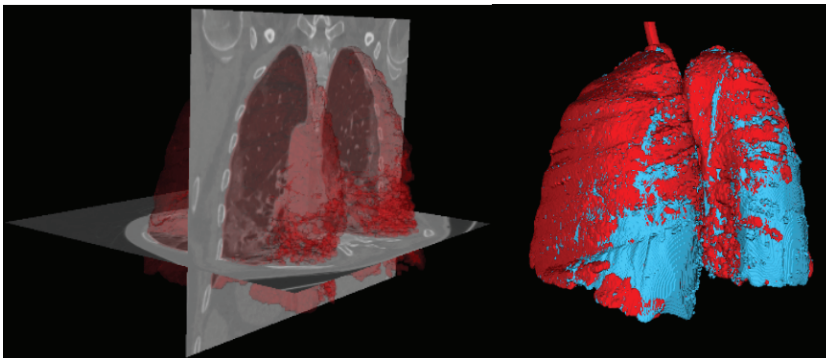


FIGURE 3.4 – Illustration de la segmentation semi-automatique avec le masque semi-automatique en rouge (à gauche) et les corrections manuelles en bleu (à droite).

Une première segmentation basée sur les niveaux de gris, la croissance des régions et la morphologie mathématique est fournie par le logiciel après initialisation manuelle. Cette étape correspond à la segmentation des zones normalement aérées avec comme unité Hounsfield des valeurs entre -900 et -500. Ensuite, l’expert corrige avec des outils d’édition spécifiquement développés pour remplir ou effacer¹³ des voxels de la segmentation, afin d’inclure dans la segmentation les zones peu aérées et non aérées (voir Figure 3.4). Sont exclues de la segmentation les effusions pleurales et les structures hilaires et médiastinales. La référence alors créée est enregistrée sous forme d’image binaire, formant ainsi un masque de segmentation pour chaque image¹⁴, de même dimension que l’image CT qui a servi pour la segmentation. Le recours à ce logiciel pour la segmentation des références permet aux cliniciens de segmenter un volume en une heure environ.

Les images et les références de segmentation ont toutes été pré-traitées avant d’être exploitées pour l’apprentissage. Dans un premier temps, elles ont subi un ré-échantillonnage¹⁵ dans le but d’imposer une résolution spatiale isotrope de 1 mm. Enfin, selon les dimensions d’entrée de chaque base, les images ont subi un recadrage afin de resserrer le champ de vue¹⁶. Le Tableau 3.1 présente les dimensions finales des images utilisées dans les expériences décrites ci-après.

13: Par connectivité ou avec un critère local, appliqué en 2D ou 3D

14: Pour toutes les conditions respiratoires d’acquisition.

15: Précédé par une étape de filtrage, par filtre gaussien, afin de garantir le critère de Shannon pour les images CT.

16: Et ainsi diminuer la quantité de voxels du fond de l’image qui ne sont pas utiles pour la segmentation et augmenteraient la mémoire de calcul nécessaire.

TABLEAU 3.1 – Caractéristiques des bases de données disponibles. – Le nombre de coupes est indiqué dans la direction axiale (A), coronale (C) et sagittale (S) donnée.

Base de données	Sujets	Volumes	Coupes	Dimensions
PK HIDRA	7	73	25 696 (A, C, S)	352 × 352 × 352
Hamilton	7	96	24 576 (A, C, S)	256 × 256 × 256
CT4-ARDS	30	60	19 200 (A) - 30 720 (C, S)	512 × 512 × 320

3.3 Description d’expérience

Modèle

Le modèle proposé dans cette expérience vise à intégrer la cohérence 3D tout en restant raisonnable en termes de temps d’entraînement et de capacité de calcul nécessaire. Pour cela, trois modèles basés sur l’architecture U-net 2D [Ron+15] sont entraînés indépendamment, respectivement avec des coupes axiales, coronales et sagittales. Chaque modèle fournit un volume reconstruit par empilement dans la direction d’apprentissage des prédictions de chaque coupe. Les trois volumes ainsi prédits sont fusionnés à l’inférence par un vote majoritaire [Pen+22]. L’illustration de la méthode proposée est visible dans la Figure 3.5. Par rapport à l’architecture originale du U-net 2D, quelques modifications sont à noter, afin d’atteindre l’objectif de réduction des ressources de calcul nécessaires. Pour cette raison, le nombre de filtres de convolution utilisé dans chaque couche est réduit d’un facteur 4 ce qui réduit le nombre de paramètres entraînaibles : par rapport à l’article original du U-net, le nombre de paramètres entraînaibles passe de 17 266 306 à 1 081 762¹⁷. Une étape de normalisation de *batch* est ajoutée après chaque opération de convolution pour améliorer la rapidité et la stabilité de l’apprentissage [Iof+15]. Les trois modèles sont identiques à la seule différence des images qu’ils utilisent pour l’entraînement ; ils sont notés θ_a , θ_c et θ_s respectivement pour les coupes axiales (a), coronales (c) et sagittales (s). Le modèle global comprenant tout le processus de prédiction par coupe dans chaque direction (modèles désignés par $\theta_{direction}$), la concaténation en volume et le vote majoritaire, est noté θ_{Vm} .

Les hyperparamètres¹⁸ utilisés pour l’entraînement sont choisis après une recherche non exhaustive et fixés pour tous les modèles. L’entraînement est réalisé par lot de 32 images prises aléatoirement dans un même plan de coupe et optimisé via l’algorithme Adam avec un taux d’apprentissage fixé à 1.10^{-3} . La fonction de coût minimisée lors de l’optimisation est basée sur le score de Dice¹⁹. Les modèles sont entraînés pendant 100 epochs²⁰. Le modèle retenu est celui pour lequel la fonction de coût est la plus faible sur la sous-partie de validation.

Stratégie d’entraînement

Pour évaluer spécifiquement l’impact que peut apporter l’utilisation d’images de plusieurs anatomies pendant l’apprentissage, deux types d’association des bases de données PK HIDRA (PK), Hamilton (Ha) et CT4-ARDS (C4A) sont étudiées. L’une par transfert d’apprentissage, l’autre par apprentissage combiné²¹. L’ensemble des bases de données utilisées est noté Ω . Chaque modèle désigné dans cette section représente le modèle θ_{Vm} décrit dans la section Modèle. Ainsi, tous les entraînements listés

Informations pratiques

Tous les codes de développement sont écrits en Python 3.9 et utilisent comme bibliothèques majeures PyTorch (1.8), Numpy (1.20) et SimpleITK (2.0.2).

17: Cela représente une réduction de 94% du nombre de paramètres entraînaibles.

18: Désigne les paramètres qui contrôlent le processus d’apprentissage automatique, par exemple au niveau des couches du modèle, ou encore pendant l’apprentissage.

19: Elle est définie comme suit :

$$\mathcal{L} = 1 - Dice \tag{3.1}$$

20: Une epoch désigne l’itération pour laquelle le modèle a vu toutes les données du jeu d’entraînement et mis à jour ses poids en conséquence.

21: En utilisant les différentes bases de données de façon aléatoire pendant le même apprentissage.

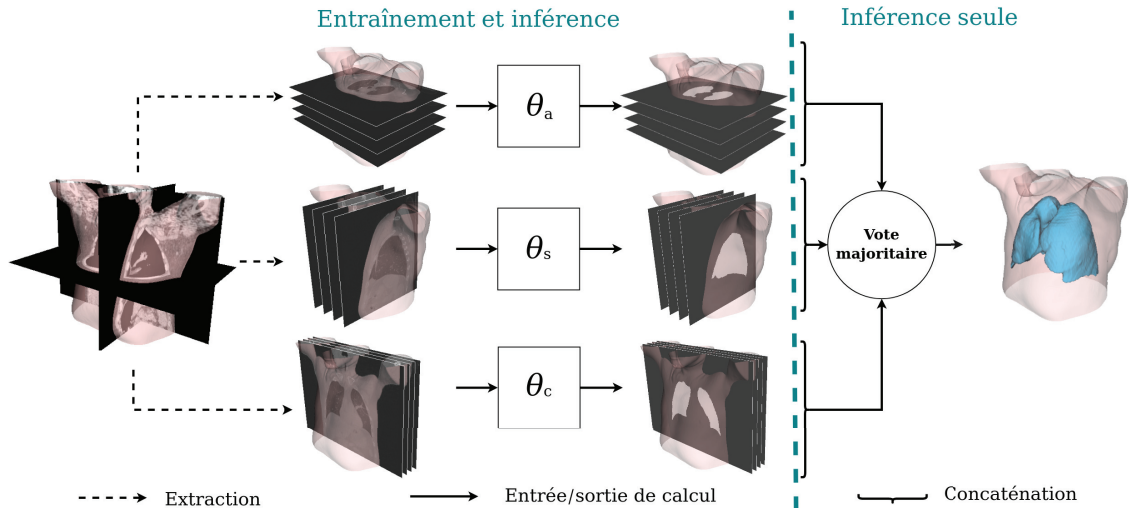


FIGURE 3.5 – Schéma du modèle multi-coupes utilisé avec, pour chaque axe de coupe, un U-net 2D modifié comme décrit dans la section Modèle.

comprennent l’apprentissage dans les trois directions orthogonales (axiale, coronale, sagittale).

Dans un premier temps, chaque base est utilisée pour entraîner séparément trois modèles de référence. Ils sont notés $A.n$, avec $n \in \{1, 2, 3\}$, dans le Tableau 3.2 récapitulant les expériences. Ces modèles serviront pour évaluer l’entraînement standard, avec initialisation aléatoire, qui n’utilise aucune gestion particulière de données. L’intégralité de chaque base, PK, Ha et C4A est utilisée pour l’entraînement de chacun des trois modèles $A.1$, $A.2$ et $A.3$ respectivement.

Dans un second temps, c’est l’augmentation de la diversité des données en mélangeant les anatomies qui est évaluée, via les modèles notés $B.n$ ($n \in \{1, 2, 3, 4\}$) reportés dans le Tableau 3.2. Pour ces modèles, toutes les combinaisons (représentées par les croix dans le Tableau 3.2) des trois bases de données disponibles ont été utilisées avec une initialisation aléatoire des poids. À chaque itération, le lot d’images utilisé est sélectionné aléatoirement dans une des bases utilisées (il n’y a pas de mélange des bases au sein d’un même lot d’images).

La troisième série d’expériences, vise à évaluer spécifiquement l’apport du transfert d’apprentissage. Le transfert d’apprentissage est réalisé en utilisant les poids d’un modèle précédemment entraîné comme initialisation des poids d’un nouveau modèle. Tous les poids sont alors actualisés lors de la nouvelle phase d’apprentissage. Les modèles de cette série d’expériences sont notés $C.n$ avec $n \in \{1, 2, 3, 4, 5, 6, 7\}$. Sont reportés dans le tableau Tableau 3.2 les identifiants des poids des modèles utilisés pour initialiser l’entraînement sur d’autres données. Les quatre premiers entraînements utilisent des modèles entraînés précédemment sur une seule base de données. Le modèle $C.5$ utilise comme poids d’initialisation l’entraînement conjoint des deux bases de données expérimentales (PK-Ha). Enfin, les deux derniers modèles $C.6$ et $C.7$ utilisent des poids d’initialisation provenant d’entraînements successifs sur les deux bases expérimentales (respectivement $PK \rightarrow Ha \rightarrow C4A$ et $Ha \rightarrow PK \rightarrow C4A$).

Cette combinaison des différentes bases de données permet d’obtenir trois types de modèles : ceux spécialisés à segmenter les poumons chez le porc ($A.1$, $A.2$, $B.3$, $C.1$, $C.3$), ceux spécialisés à segmenter les poumons chez l’humain ($A.3$, $C.2$, $C.4$, $C.5$, $C.6$, $C.7$) et ceux non spécialisés ($B.1$, $B.2$,

TABLEAU 3.2 – Description des expériences menées avec les combinaisons de bases de données utilisées à l’entraînement. – Pour chaque modèle est indiqué la ou les bases de données utilisées pour l’entraîner ainsi que s’il s’agit d’une initialisation aléatoire des poids du modèle (x) ou d’un transfert d’apprentissage (identifiant d’un autre modèle). Lorsque plusieurs bases de données sont cochées pour le même modèle, cela signifie qu’elles ont été utilisées aléatoirement et entièrement pendant l’entraînement.

Identifiant modèle	Base de données		
	PK	Ha	C4A
$A.1$	x		
$A.2$		x	
$A.3$			x
$B.1$	x		x
$B.2$		x	x
$B.3$	x	x	
$B.4$	x	x	x
$C.1$		$A.1$	
$C.2$			$A.1$
$C.3$	$A.2$		
$C.4$			$A.2$
$C.5$			$B.3$
$C.6$			$C.1$
$C.7$			$C.3$

B.4). Les combinaisons de transfert d'apprentissage sont toujours réalisées avec comme anatomie cible l'humain, sauf pour les modèles intermédiaires C.1 et C.3, qui serviront eux-mêmes d'initialisation.

La notation $X.n$ est introduite pour la concision, mais par la suite l'exposant $\beta \in \{PK, Ha, C4A, \Omega\}$ sera ajouté pour rappeler sur quelle base de données a été fait l'entraînement et faciliter ainsi la lecture.

Évaluation

Du fait de la faible quantité de sujets par base de données, une stratégie de validation croisée en k sous-ensembles a été mise en place. Les trois bases de données ont été utilisées pour l'évaluation. D'une part pour comparer les performances respectives des modèles sur leur base d'entraînement²², sans transfert d'apprentissage (modèles $A.n$ et $B.n$). D'autre part, pour quantifier l'impact du transfert d'apprentissage sur chaque combinaison de base de données effectuée.

La séparation des données est réalisée au niveau des sujets (patients ou cochons) pour qu'une image d'un même sujet ne puisse apparaître à la fois dans un sous-groupe d'entraînement et un sous-groupe de test. Ce niveau de séparation impose de considérer toutes les images d'un même sujet au sein du même groupe²³.

La base de données CT4-ARDS a été séparée en cinq sous-ensembles²⁴, pour permettre d'atteindre des proportions usuelles entre les trois sous-groupes d'entraînement, de validation et de test. À savoir 21 patients pour l'entraînement, 3 patients pour la validation et 6 patients réservés au test. Les deux bases de données sur les études expérimentales ont été quant à elles découpées en 3 sous-ensembles non équivalents entre eux²⁵. En effet, du fait de leur constitution, chaque sujet de ces deux bases ne compte pas le même nombre d'images, en plus du nombre premier de sujets qui empêche une division équilibrée. Pour deux sous-ensembles, le nombre de sujets utilisés est alors de 4, 2, et 1 respectivement pour les sous-groupes d'entraînement, de validation et de test. Le dernier sous-ensemble compte 3 cochons dans le sous-groupe d'entraînement, 1 dans celui de validation et 3 dans celui de test.

TABLEAU 3.3 – Répartition des sous-ensembles pour la validation croisée en k sous-ensembles, avec le nombre de sujets dans chacun des trois sous-groupes et le nombre de volumes correspondants entre parenthèses.

Base de données	k	Proportions		
		Entraînement	Validation	Test
PK HIDRA	3	3-4 (39-46)	1 (6-8)	2-3 (20-28)
Hamilton	3	3-4 (48-60)	1 (4-12)	2-3 (24-44)
CT4-ARDS	5	21 (42)	3 (6)	6 (12)

Les métriques utilisées²⁶ pour évaluer les performances des divers modèles sont le score de Dice, le quatre-vingt quinzième centile de la distance de Hausdorff²⁷ (DH-95) et la distance moyenne de surface symétrique (ASSD). Elles sont toutes calculées sur les masques volumiques prédits soit en sortie de chaque modèle $\theta_{direction}$ ou bien du modèle global θ_{Vm} . Tous les modèles listés dans le Tableau 3.2 sont évalués sur les trois bases de données utilisées. Les modèles qui utilisent les bases de données animales et patients²⁸ sont tous entraînés sur trois sous-ensembles choisis

22: Entraînements avec une ou plusieurs bases de données à la fois.

23: Même lorsque cela signifie déséquilibrer le ratio des sous-groupes entraînement, validation et test.

24: Avec chacun les mêmes proportions des sous-groupes d'entraînement, de validation et de test.

25: En ce qui concerne les sous-groupe d'entraînement, de validation et de test.

26: Métriques définies dans la sous-section Évaluer la segmentation.

27: Cela signifie que 95% des valeurs mesurées sont en dessous de la valeur affichée.

28: Les modèles $B.n$ avec $n \in \{1, 2, 3, 4\}$ et $C.n$ avec $n \in \{1, 3, 5, 6, 7\}$.

aléatoirement, afin de permettre la validation croisée sur les bases animales. L'accent sera cependant mis sur les résultats de test de la base CT4-ARDS, l'objectif étant la segmentation des poumons humains.

3.4 Résultats

Entraînements de référence

Les entraînements de référence sont identifiés par le code $A.n^\beta$ avec $n \in \{1, 2, 3\}$ et $\beta \in \{PK, Ha, C4A\}$. La Figure 3.6 regroupe les performances de ces trois modèles sur la base d'intérêt CT4-ARDS par validation en k sous-ensembles ($k = 3$). Le modèle entraîné sur la base d'intérêt ($A.3^{C4A}$) fournit en valeur de référence 0.970 ± 0.018 (Dice), $0.99 \pm 0.88mm$ (ASSD) et $4.00 \pm 4.81mm$ (DH-95). Lorsque $k = 5$, les performances de ce modèle sont de 0.968 ± 0.021 (Dice), $1.04 \pm 1.30mm$ (ASSD) et $4.06 \pm 4.23mm$ (DH-95). En comparaison, les deux modèles spécialisés dans la segmentation des poumons de cochons obtiennent tous deux des performances inférieures avec davantage de valeurs aberrantes. C'est le modèle $A.2^{Ha}$ qui obtient les meilleurs résultats en termes de médiane bien que ses performances soient plus étalées avec des valeurs aberrantes plus importantes. En effet, comparé au modèle $A.3^{PK}$, le modèle $A.2^{Ha}$ obtient sur le score de Dice 0.931 ± 0.060 (contre 0.920 ± 0.046), sur la métrique ASSD $2.56 \pm 3.69mm$ (contre $3.23 \pm 1.62mm$) et sur la distance de Hausdorff (DH-95) $13.6 \pm 29.8mm$ (contre $15.3 \pm 6.85mm$).

Les performances des modèles entraînés sur les bases de données animales et testés sur ces mêmes données sont présentées dans le Tableau 3.4. Sur leur base respective d'entraînement, les performances des modèles $A.1^{PK}$ et $A.2^{Ha}$ sont similaires en moyenne, bien que la répartition des résultats diffère (écart-type et valeurs aberrantes). En effet, les modèles $A.1^{PK}$ et $A.2^{Ha}$ ont des valeurs médianes d'ASSD proches respectivement de $0.45mm$ et $0.42mm$ mais leur écart-type (et leur valeur maximale) sont respectivement de $0.23mm$ ($1.61mm$) et $1.05mm$ ($4.06mm$). Testé sur la base de données PK, non utilisée pendant l'entraînement, le modèle $A.2^{Ha}$ est celui dont les performances sont les plus dégradées (-1% de Dice, $+281\%$ d'ASSD et $+375\%$ de DH-95) contre seulement (-1% de Dice, $+44\%$ d'ASSD et $+45\%$) pour le modèle $A.1^{PK}$ testé sur la base de données Ha.

TABLEAU 3.4 – Résultats des modèles de référence animaux sur les deux bases de données animales.

Base de données de test	Modèle	Dice	ASSD (mm)	DH-95 (mm)
PK HIDRA	$A.1^{PK}$	0.99 ± 9.10^{-3}	0.45 ± 0.23	2.0 ± 1.1
	$A.2^{Ha}$	0.98 ± 0.026	1.6 ± 2.6	9.5 ± 29
Hamilton	$A.1^{PK}$	0.98 ± 0.048	0.65 ± 1.1	2.9 ± 5.7
	$A.2^{Ha}$	0.99 ± 0.028	0.42 ± 1.1	2.0 ± 9.9

L'entraînement des modèles de référence a duré en moyenne respectivement 13h, 7h et 13h pour $A.1^{PK}$, $A.2^{Ha}$ et $A.3^{C4A}$. Ces temps d'entraînement sont moyennés sur les temps respectifs d'entraînement des trois modèles $\theta_{direction}$ qui composent chaque modèle global θ_{Vm} . En effet, les entraînements étant indépendants, ils ont pu être réalisés en parallèle sur des

Toutes les figures présentant les métriques seront calculées sur les sous-groupes de test de la base de données CT4-ARDS.

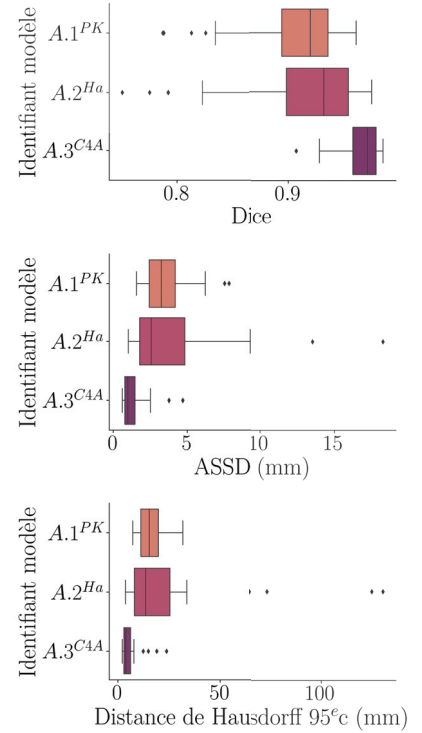


FIGURE 3.6 – Résultats des modèles de référence testés sur 3 sous-ensembles de la base de données CT4-ARDS.

machines différentes. Les courbes d’entraînement du modèle $A.3^{C4A}$, détaillées par coupe d’entraînement, sont visibles dans la figure 3.7. On observe une rapide décroissance dans les premières epochs puis une stabilisation en palier, peu importe le sous-ensemble d’entraînement. Le comportement des courbes d’entraînement des trois modèles de référence est similaire.

Entraînements combinés

Les entraînements avec bases de données combinées sont notés $B.n^\beta$ avec $n \in \{1, 2, 3, 4\}$ et β les différentes associations de bases de données possibles. La Figure 3.8 illustre les performances de ces modèles sur la base de données CT4-ARDS, en validation croisée sur 3 sous-ensembles des 5 de cette base de données. Tous les modèles qui contiennent la base CT4-ARDS à l’entraînement ($B.1^{PK-C4A}$, $B.2^{Ha-C4A}$, $B.4^\Omega$) fournissent des résultats similaires, et ce, pour toutes les métriques considérées. Certaines caractéristiques observées sur les trois métriques permettent de les discriminer, positivement : la meilleure médiane du modèle $B.1^{PK-C4A}$ (0.99 mm d’ASSD, contre respectivement 1.03mm et 1.07 mm pour les modèles $B.4^\Omega$ et $B.2^{Ha-C4A}$ par exemple) ou le plus faible écart-type du modèle $B.4^\Omega$ (± 0.83 mm d’ASSD contre respectivement ± 0.88 mm et ± 2.14 mm pour les modèles $B.2^{Ha-C4A}$ et $B.1^{PK-C4A}$); ou bien négativement : la présence de très grandes valeurs aberrantes du modèle $B.1^{PK-C4A}$ (comme l’illustre la Figure 3.8). Le modèle entraîné uniquement sur les bases de données animales ($B.3^{PK-Ha}$) rassemble les moins bonnes performances sur ces critères de comparaison avec une répartition des résultats plus large et des valeurs aberrantes plus grandes (voir Figure 3.8). Ses performances sont de 0.926 ± 0.074 pour le score de Dice, $3.13 \pm 2.84mm$ pour l’ASSD et $3.87 \pm 24.5mm$ pour la distance de Hausdorff (DH-95).

Les performances des modèles $B.n^\beta$ entraînés avec les stratégies de mélange des données (décrites dans la section Stratégie d’entraînement) et testés sur les données animales sont présentées dans le Tableau 3.5. Le modèle $B.3^{PK-Ha}$ fournit les meilleurs résultats quelle que soit la base de données utilisée pour le test. Il fournit ainsi un score de Dice de 0.99, une distance ASSD inférieure à $0.44mm$ et une distance de Hausdorff autour de $1.7mm$. Les performances des autres modèles sur le score de Dice restent semblables entre elles, avec des valeurs supérieures à 0.98. Les variations de performances sont plus importantes au niveau des métriques de distance. En effet, les écart-types des résultats d’ASSD sont au minimum de $0.22mm$ et au maximum de $2.1mm$. La même observation est faite sur la distance de Hausdorff avec au minimum un écart-type de $0.98mm$ (pour le modèle $B.1^{PK-C4A}$ testé sur la base de données PK HIDRA) et au maximum $20mm$ (pour le modèle $B.2^{Ha-C4A}$ testé sur la base de données PK HIDRA).

Les courbes d’entraînement du modèle $B.4^\Omega$ sont présentées (voir Figure 3.9) pour chaque sous-partie $\theta_{direction}$ accompagnées des valeurs de validation. Les allures des autres entraînements des modèles $B.n$ sont similaires à celles illustrées en Figure 3.9, c’est-à-dire une rapide et forte décroissance, partant de 0.8 pour devenir inférieure à 0.2 dès la première epoch. Ensuite, la perte sur les données d’entraînement et de validation se stabilise respectivement autour de 0.007 et 0.04.

Les entraînements de chaque sous-partie $\theta_{direction}$ des modèles θ_{Vm} ont pu être réalisés en parallèle, et la moyenne des trois modèles $\theta_{direction}$ a duré 29 h.

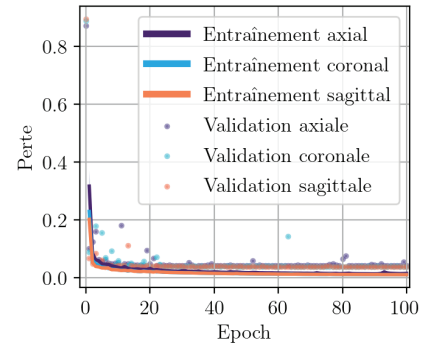


FIGURE 3.7 – Courbes d’entraînement du modèle $A.3^{C4A}$ sur les 5 sous-ensembles de sa base de données d’entraînement.

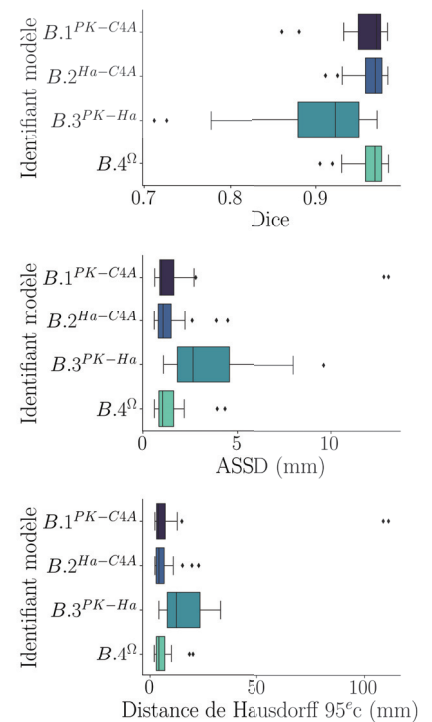


FIGURE 3.8 – Résultats des modèles entraînés sur des combinaisons de bases de données, testés sur la base CT4-ARDS en validation croisée sur 3 sous-ensembles.

TABLEAU 3.5 – Résultats des modèles utilisant une combinaison de bases de données sur les deux bases de données animales.

Base de données de test	Modèle	Dice	ASSD (mm)	DH-95 (mm)
PK HIDRA	$B.1^{PK-C4A}$	0.99 ± 9.10^{-3}	0.47 ± 0.22	2.0 ± 0.98
	$B.2^{Ha-C4A}$	0.98 ± 0.012	0.54 ± 2.1	2.2 ± 20
	$B.3^{PK-Ha}$	0.99 ± 0.013	0.44 ± 0.35	1.7 ± 1.8
	$B.4^{\Omega}$	0.98 ± 9.10^{-3}	0.50 ± 0.54	2.0 ± 0.89
Hamilton	$B.1^{PK-C4A}$	0.98 ± 0.039	0.74 ± 0.95	3.2 ± 5.6
	$B.2^{Ha-C4A}$	0.98 ± 0.034	0.53 ± 1.4	2.2 ± 11
	$B.3^{PK-Ha}$	0.99 ± 0.040	0.43 ± 0.63	1.6 ± 2.8
	$B.4^{\Omega}$	0.98 ± 0.029	0.49 ± 1.3	2.0 ± 11

Entraînements avec transfert d'apprentissage

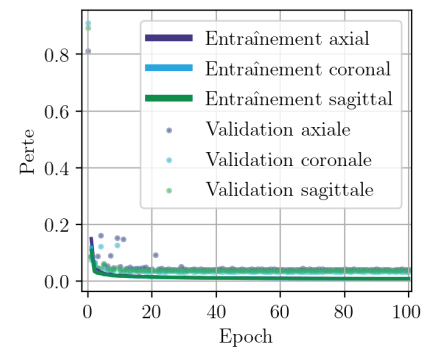
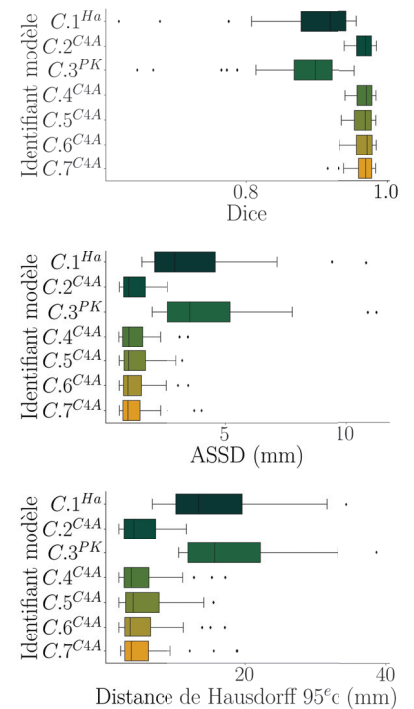
Les entraînements qui utilisent le transfert d'apprentissage sont notés $C.n^{\beta}$ avec n allant de 1 à 7 et $\beta \in \{PK, Ha, C4A\}$. Les résultats présentés dans la Figure 3.10 sont calculés sur les sous-groupes de test de trois sous-ensembles de la base de donnée CT4-ARDS. Parmi ces sept modèles, cinq sont entraînés en affinement sur la base CT4-ARDS ($C.n^{C4A}$ avec $n \in \{2, 4, 5, 6, 7\}$) tandis que les deux restants sont affinés sur les bases de données animales, en tant que modèles intermédiaires. Les modèles affinés sur la base de données patient ont des performances similaires quelle que soit la métrique observée. Ainsi, les modèles $C.n^{C4A}$ ont tous une ASSD médiane sub-millimétrique, un score de Dice médian supérieur à 0.97 et une distance de Hausdorff médiane inférieure ou égale à $4mm$. Ils produisent peu de valeurs aberrantes et celles-ci sont peu éloignées du reste des valeurs (quelle que soit la métrique). Dans cette expérience, le modèle dont la médiane et l'écart-type sont légèrement meilleurs pour toutes les métriques est le modèle $C.6^{C4A}$. Ce modèle est initialement entraîné sur la base de données PK HIDRA, puis affiné sur la base de données Hamilton et enfin affiné sur la base de données CT4-ARDS. Sur les trois sous-ensembles de test de la base CT4-ARDS, il obtient un score de Dice de 0.970 ± 0.015 , une distance ASSD de $0.954 \pm 0.704mm$ et une distance de Hausdorff au 95^e centile de 3.61 ± 3.76 .

Les modèles $C.1^{Ha}$ et $C.3^{PK}$, entraînés sur les bases de données animales, sont utilisés comme les initialisations respectives des modèles $C.6^{C4A}$ et $C.7^{C4A}$ avant leur entraînement en affinement sur la base C4A. Par rapport aux autres modèles $C.n^{C4A}$, on peut d'ailleurs observer une diminution de l'écart-type sur ces deux derniers modèles, avec $C.7^{C4A}$ qui a la plus petite variabilité.

Les performances des modèles intermédiaires sur les deux bases de données animales sont résumées dans le Tableau 3.6 en validation croisée sur les trois sous-ensembles de chaque base de données. Les deux modèles obtiennent des performances similaires sur leurs bases d'entraînement respectives. Le modèle $C.3^{PK}$ est celui dont les résultats se dégradent le moins, lorsqu'il est évalué sur la base de données utilisées pour le pré-entraînement.

L'entraînement des modèles intermédiaires $C.1^{Ha}$ et $C.3^{PK}$ a duré en moyenne $5h$ pour chaque sous-partie $\theta_{direction}$. Concernant les modèles affinés sur la base CT4-ARDS ($C.n^{C4A}$), le temps d'entraînement moyen était de $13h$, pour chaque sous-partie $\theta_{direction}$.

La stabilité de tous les entraînements utilisant le transfert d'apprentissage


FIGURE 3.9 – Courbes d'entraînement du modèle $B.4^{\Omega}$ sur sa base de données d'entraînement.

FIGURE 3.10 – Résultats des modèles entraînés avec transfert d'apprentissage testés sur la base CT4-ARDS.

TABEAU 3.6 – Résultats des modèles $C.1^{Ha}$ et $C.3^{PK}$ utilisant le transfert d'apprentissage uniquement sur la base de données animales évalués en validation croisée sur ces deux bases (PK et Ha).

Base de données de test	Modèle	Dice	ASSD (mm)	DH-95 (mm)
PK HIDRA	$C.1^{Ha}$	0.97 ± 0.017	0.60 ± 0.53	2.2 ± 3.4
	$C.3^{PK}$	$0.98 \pm 9.2 \cdot 10^{-3}$	0.40 ± 0.24	1.7 ± 1.1
Hamilton	$C.1^{Ha}$	0.98 ± 0.017	0.41 ± 0.36	1.7 ± 1.2
	$C.3^{PK}$	0.97 ± 0.032	0.49 ± 0.60	2.1 ± 2.9

est semblable à celle des modèles de référence et des modèles utilisant les bases de données combinées. Cela s'observe sur la Figure 3.11. La valeur de la perte est au départ de l'entraînement autour de 0.125 puis décroît de façon stable (quelle que soit la direction d'entraînement) jusqu'à atteindre des valeurs autour de 0.01. En ce qui concerne la perte calculée sur les données de validation, elle reste stable autour de 0.03 durant tout l'entraînement.

Résultats globaux

Le résumé de toutes les expériences est présenté sous différents aspects. La Figure 3.12 regroupe les performances de tous les modèles entraînés, évalués sur les 3 sous-ensembles de la base CT4-ARDS. Pour chaque métrique, on observe des performances nettement inférieures des modèles entraînés uniquement sur les bases de données animales. Ceux-ci ont à la fois des médianes dégradées, une répartition des résultats élargie ainsi que les erreurs aberrantes les plus éloignées. L'ensemble des modèles qui utilisent la base CT4-ARDS fournit des résultats semblables, que ce soit en termes de médiane, d'écart-type et d'erreurs aberrantes. La répartition des résultats au 25^e et 75^e centiles permet de discriminer un modèle qui obtient le meilleur score sur les trois métriques. Il s'agit du modèle $C.4^{C4A}$ qui est le résultat du transfert d'apprentissage depuis la base de données Hamilton. En effet, il obtient pour le score de Dice $0.970[0.957; 0.978]$, pour l'ASSD $0.995[0.741; 1.56]mm$ et le 95^e centile de la distance de Hausdorff $3.74[2.73; 6.22]mm$. En comparaison, le modèle de référence $A.3^{C4A}$ évalué en validation croisée sur l'ensemble de sa base de données d'entraînement ($k = 5$) atteint des performances légèrement inférieures à celles du modèle $C.4^{C4A}$.

L'apport de la méthode multi-coupes (θ_{Vm}) par rapport à chaque modèle individuel $\theta_{direction}$ est évalué dans le Tableau 3.7. Les modèles sont regroupés en deux groupes, selon que leur base de données d'entraînement est uniquement animale ou bien contient des patients. Le premier groupe est noté $X.n^{animal}$ et contient cinq modèles : $A.1^{PK}$, $A.2^{Ha}$, $B.3^{PK-Ha}$, $C.1^{Ha}$ et $C.3^{PK}$. Le second groupe noté $X.n^{patient}$ contient tous les autres modèles pour un total de neuf modèles. Quel que soit le groupe de modèle évalué ($X.n^{animal}$ ou $X.n^{patient}$), la méthode multi-coupes θ_{Vm} fournit les meilleures performances. La médiane est systématiquement améliorée et l'écart-type diminué sur les expériences du groupe $X.n^{animal}$, et même si l'amélioration est moindre sur le groupe $X.n^{patient}$, l'écart-type est systématiquement réduit par rapport aux modèles axial et sagittal. On observe davantage l'impact de la méthode multi-coupes (θ_{Vm}) sur les métriques

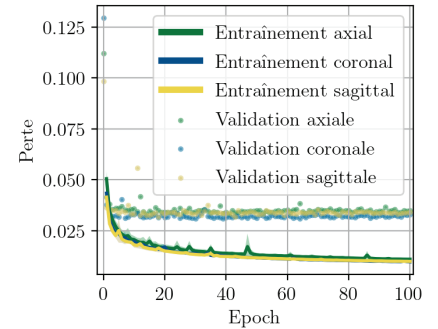


FIGURE 3.11 – Courbes d'entraînement du modèle $C.3^{PK}$ sur la base de données CT4-ARDS (en validation croisée de 3 sous-ensembles).

de distance (ASSD et DH-95). En effet, le gain est marqué à la fois sur la médiane et l'écart-type pour ces métriques, qui s'améliorent toutes deux. Aussi, l'impact de la méthode multi-coupes s'observe davantage sur le groupe $X.n^{animal}$ qui gagne en robustesse sur la base de données de test CT4-ARDS.

TABLEAU 3.7 – Comparaison des modèles en fonction de leur base de données d'entraînement et du type d'image utilisée.

Groupe	Type de modèle	Dice	ASSD (mm)	DH-95 (mm)
$X.n^{animal}$	θ_{axial}	0.87 ± 0.13	4.8 ± 8.0	22 ± 37
	$\theta_{coronal}$	0.88 ± 0.09	5.9 ± 8.0	31 ± 33
	$\theta_{sagittal}$	0.90 ± 0.08	4.1 ± 3.1	19 ± 23
	θ_{Vm}	0.92 ± 0.07	2.9 ± 2.5	14 ± 15
$X.n^{patient}$	θ_{axial}	0.96 ± 0.04	1.3 ± 3.1	5.7 ± 19
	$\theta_{coronal}$	0.97 ± 0.02	1.1 ± 1.2	4.2 ± 6.9
	$\theta_{sagittal}$	0.97 ± 0.05	1.1 ± 2.3	4.2 ± 14
	θ_{Vm}	0.97 ± 0.02	0.99 ± 1.2	4.0 ± 9.1

À l'inférence, tous les modèles prennent²⁹ en moyenne 2 minutes (calcul sur CPU) ou 15 secondes (calcul avec GPU) pour prédire un volume de taille $512 \times 512 \times 295$. Selon l'infrastructure matérielle utilisée (CPU ou GPU) la mémoire utilisée est tantôt de 4700 Mo en calcul CPU, tantôt distribuée en 3800 Mo sur la mémoire CPU et 870 Mo sur la carte graphique (GPU).

3.5 Discussion

L'objectif de cette expérience était d'évaluer l'apport potentiel de l'utilisation de données annotées additionnelles³⁰ dans l'apprentissage de la segmentation des poumons de patients avec SDRA.

Performances de référence Les modèles $A.n$ fournissent les performances de référence atteintes lors d'un entraînement classique³¹. Le modèle $A.3^{C4A}$ permet d'établir, pour la base de données CT4-ARDS, les performances auxquelles se référer dans l'évaluation des autres entraînements utilisant une gestion des bases de données différente. De même, les modèles entraînés au départ sur les deux bases de données animales fournissent la référence pour évaluer spécifiquement l'apport du transfert d'apprentissage. Leurs performances sur la base de données CT4-ARDS (ou la base de données animales non utilisée à l'entraînement) témoignent de la capacité initiale de ces modèles à généraliser sur une anatomie différente.

Ainsi, les performances de référence pour la base de données CT4-ARDS sont de 0.968 ± 0.021 (Dice), $1.04 \pm 1.30mm$ (ASSD) et $4.06 \pm 4.23mm$ (DH-95), en validation croisée sur les cinq sous-ensembles au total de la base de données. Évaluée uniquement sur les trois sous-ensembles, pour correspondre aux nombres de modèles utilisant les bases de données animales, les performances sont très semblables avec 0.970 ± 0.018 pour le score de Dice, $0.986 \pm 0.879mm$ pour la métrique ASSD et $4.00 \pm 4.81mm$ pour la métrique DH-95. Alors, il est possible de considérer l'utilisation de ces trois sous-ensembles pour la suite des entraînements comme représentative de

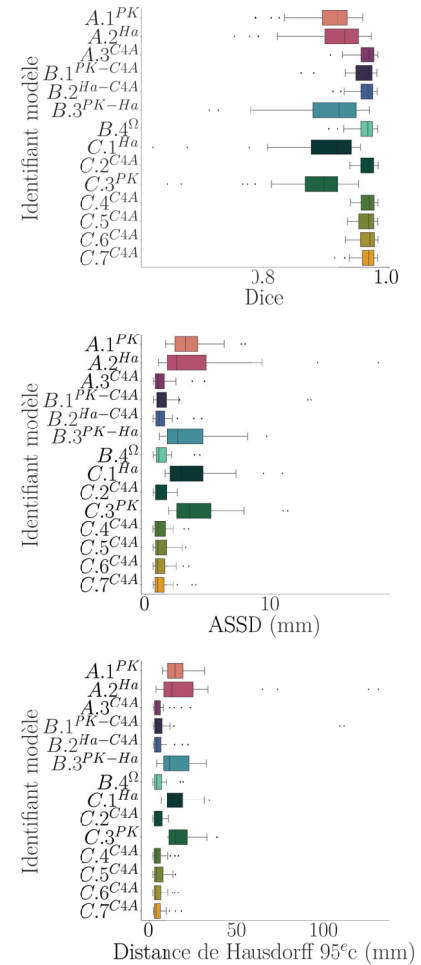


FIGURE 3.12 – Résultats de tous les modèles sur les 3 sous-ensembles de la base de données CT4-ARDS.

29: Calcul effectué en CPU avec 32Go de RAM sur un processeur Intel®Core™i5-8500 CPU @ 3.00GHz × 6 et en GPU avec une carte graphique Quadro P6000.

30: Ici des images CT de modèles expérimentaux porcins de SDRA.

31: Séparation d'une base de données contenant l'anatomie cible en trois sous-groupes (entraînement, validation et test).

la base de données CT4-ARDS (voir encadre ci-contre). Les modèles entraînés sur les bases de données animales ($A.1^{PK}$ et $A.2^{Ha}$) sont à la fois moins performants et génèrent davantage d'erreurs aberrantes sur la base de données CT4-ARDS. Leur capacité à généraliser est assez faible, car les performances passent d'au moins 0.98 pour le score de Dice sur les bases de données animales à 0.93 sur la base de données patient. Les résultats mesurés sur la distance de Hausdorff mettent davantage en exergue cette faiblesse avec une dégradation de la médiane de $2mm$ à plus de $13mm$.

L'évaluation qualitative des masques de segmentation permet de mettre en évidence le type d'erreurs observables pour chaque modèle. En effet, les métriques géométriques utilisées sont globales et ne rendent pas compte des erreurs au niveau local³². La Figure 3.13 illustre pour chaque modèle de référence un exemple d'échec et de réussite de segmentation. Les erreurs observées (encadrées en jaune) ne sont pas spécifiques à un des modèles de référence et surviennent autant dans les zones du poumon faiblement contrastées que dans celles bien contrastées. On peut lister parmi les erreurs les plus observées les faux négatifs, les trous et les contours peu réguliers en escalier. Ces derniers sont un résultat direct de la méthode multi-coupes, qui du fait de son fonctionnement 2D ne garantit pas de cohérence de segmentation entre les coupes. Concernant les faux négatifs et les trous, ceux-ci surviennent à la fois dans les régions peu aérées que dans les régions bien aérées, ce qui indique une faiblesse dans l'entraînement. Les trous pourraient être facilement comblés par une étape additionnelle de post-traitement, ce qui n'est pas le cas des faux négatifs qui demandent plus d'attention concernant l'ajustement de leur frontière.

Résultats de référence du modèle $A.3^{C4A}$
 Calculés sur les trois sous-ensembles fixés pour comparaison avec les modèles évalués dans les autres expériences B et C :
 Dice = 0.970 ± 0.018
 SSD = 0.99 ± 0.88 mm
 DH-95 = 4.00 ± 4.81 mm

32: Leur forme, leur position dans le contexte volumique dans lequel elles apparaissent.

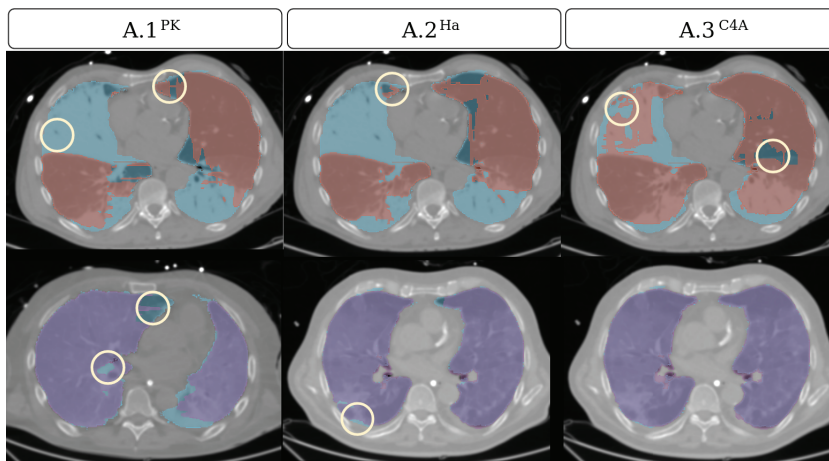


FIGURE 3.13 – Exemple des pires (en rouge en haut) et des meilleures (en violet en bas) segmentations des modèles de référence $A.1^{PK}$, $A.2^{Ha}$ et $A.3^{C4A}$ sur la base de données CT4-ARDS, superposées avec la segmentation manuelle de référence en bleu.

Performances avec augmentation du nombre de données et différentes anatomies Les modèles $B.n$ utilisent davantage de données à l'entraînement que les modèles de référence ($A.n$). Chaque sous-ensemble utilisé pendant l'entraînement est donc une composition des différentes bases de données présentées dans le Tableau 3.3. Au minimum 81 volumes et au maximum 141 volumes sont utilisés pour l'entraînement, par rapport aux 60 volumes maximum utilisés par les modèles de référence ($A.n$).

Les modèles qui utilisent la base de données patient à l'entraînement fournissent des performances similaires entre eux, quelle que soit la métrique évaluée. Celui, entraîné sur le mélange des bases de données CT4-ARDS et PK HIDRA, $B.1^{PK-C4A}$, obtient les résultats les plus proches du modèle de référence $A.3^{C4A}$ en termes de médiane. Cependant, son écart-type est systématiquement dégradé sur chaque métrique, particulièrement sur la distance de Hausdorff (DH-95) pour laquelle il vaut $24.6mm$, soit huit fois plus que la médiane obtenue sur cette même métrique.

Les modèles entraînés sur plusieurs bases de données simultanément ($B.n$) gagnent en généralisation, par rapport aux modèles de référence $A.1^{PK}$ et $A.2^{Ha}$, et ce peu importe la base de données animale testée. En effet, les écarts-types de performance entre les différents modèles $B.n$ sont diminués puisque ici le meilleur et le pire modèle obtiennent respectivement une métrique ASSD de 0.43 ± 0.63 et 0.74 ± 0.95 alors que pour les modèles de référence $A.1^{PK}$ et $A.2^{Ha}$ le meilleur obtenait 0.42 ± 1.1 et le pire 1.6 ± 2.6 . Le modèle $B.3^{PK-Ha}$ entraîné exclusivement sur les bases de données animales améliore notamment les performances sur les deux bases de données concernées. En effet, l'amélioration s'observe soit sur la médiane³³, soit sur l'écart-type³⁴.

Concernant l'évaluation qualitative des masques de segmentation, elle est observable sur la Figure 3.14 avec d'un côté une sélection des pires performances des quatre modèles et de l'autre côté les meilleures performances. On remarque que la répartition des erreurs est quelque peu modifiée par rapport aux modèles de référence $A.n$. En effet, les zones non aérées et donc non contrastées présentent quelques pixels segmentés, bien que ceux-ci ne recouvrent pas entièrement la région des poumons (voir zone encerclée du modèle $B.1^{PK-C4A}$ dans la Figure 3.14). Les faux négatifs présents dans les zones aérées restent aussi présents pour ce type de modèles. Concernant les meilleures segmentations, quelques erreurs de sous-segmentation subsistent comme l'illustrent les deux zones encerclées sur la Figure 3.14. La première erreur illustrée concerne le modèle $B.3^{PK-Ha}$ pour lequel la segmentation (en violet) n'est pas suffisamment collée aux côtes au niveau postérieur, comparé à la segmentation manuelle (en bleu). La seconde erreur est la présence d'un trou dans la segmentation du modèle $B.2^{Ha-C4A}$.

En résumé, augmenter le nombre d'images tout en intégrant une anatomie différente durant l'entraînement ne permet pas de gagner en robustesse. La majeure erreur observée est la sous-segmentation des zones non aérées, qui est difficilement corrigeable par post-traitement. Le gain amené par cette stratégie de mélange des bases de données, porte avant tout sur la capacité de généralisation des modèles qui utilisent à la fois une base de données animale et une base de données patient. En effet, si un seul modèle devait être utilisé pour segmenter à la fois les patients comme les animaux, ces modèles ont l'avantage d'être polyvalents dans le type d'anatomie qu'ils peuvent segmenter et fournissent des résultats légèrement en dessous des modèles spécialisés de référence.

33: Visible lorsque la base de données PK HIDRA est utilisée pour le test, en passant de $0.45mm$ à $0.44mm$ (ASSD) et de $2.0mm$ à $1.7mm$ (DH-95).

34: Visible lorsque la base de données Hamilton est utilisée pour le test, en passant de $1.1mm$ à $0.63mm$ et de $9.9mm$ à $2.8mm$ (DH-95).

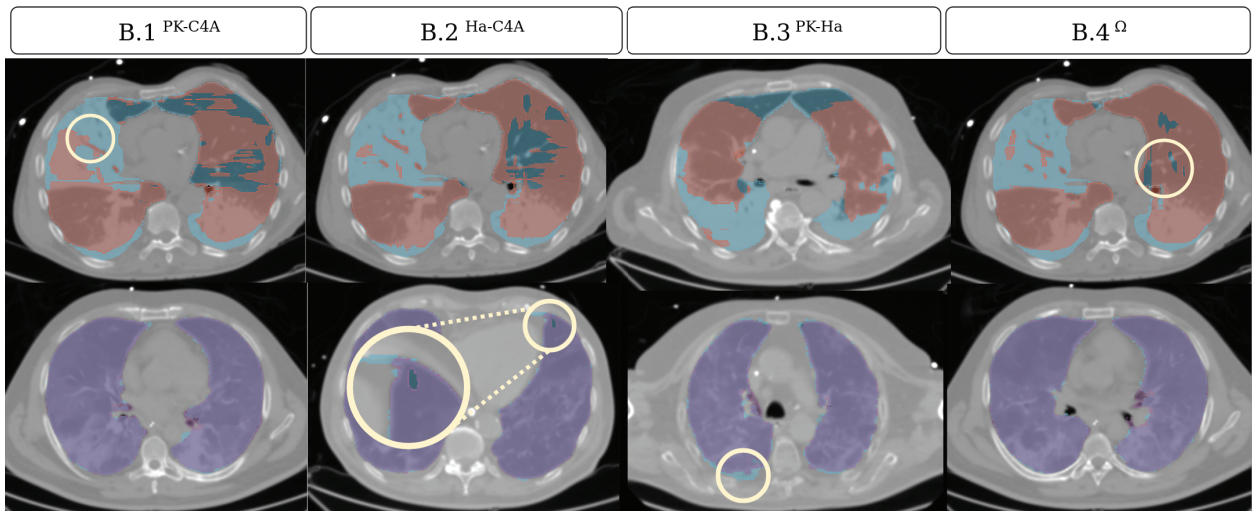


FIGURE 3.14 – Exemple des pires (en rouge) et des meilleures (en violet) segmentations des modèles $B.n$ sur la base de données CT4-ARDS, superposées avec la segmentation manuelle de référence en bleu.

Évaluation du transfert d'apprentissage Les modèles $C.n$ sont censés bénéficier des apprentissages précédents faits sur les autres bases de données sur lesquelles ils ont été entraînés. Différentes combinaisons ont été effectuées pour évaluer si l'ordre de présentation des bases de données influence l'apprentissage. Tous les précédents modèles qui n'ont pas utilisé la base de données CT4-ARDS servent alors d'initialisation pour un entraînement de spécialisation sur la base de données patient. Cela conduit donc à sept modèles utilisant le transfert d'apprentissage, dont cinq sont spécialisés sur la base de données patient. Ces résultats sont présentés dans la Figure 3.10.

Les deux modèles qui subissent le transfert d'apprentissage sur la base de données animale qu'ils n'ont pas encore utilisée, $C.1^{Ha}$ et $C.3^{PK}$, sont évalués sur ces deux mêmes bases de données dans le Tableau 3.6. Par rapport aux modèles animaux de référence, ils perdent légèrement en performances sur le score de Dice mais gagnent en robustesse sur les métriques de distance. Ce qui indique des erreurs de superposition, mais des contours plus proches des masques de référence.

Les performances des modèles spécialisés en transfert d'apprentissage sur la base de données CT4-ARDS, $C.n^{patient}$, sont semblables entre eux. En effet, les scores de Dice médians de ces modèles se trouvent tous entre 0.969 et 0.971, les distances ASSD médianes entre 0.954 et 0.995 mm et les distances DH95 entre 3.61 et 4.00 mm. En comparaison du modèle de référence $A.3^{C4A}$, tous ces modèles obtiennent une distance DH-95 médiane inférieure ou égale, de même pour son écart-type. Cela indique des prédictions plus proches des contours du masque de segmentation manuelle. Lorsque toutes les métriques sont prises en compte, les deux modèles qui obtiennent de meilleurs résultats (ou bien équivalents) au modèle $A.3^{C4A}$ sont les modèles³⁵ $C.6^{C4A}$ et $C.7^{C4A}$. Le premier obtient ainsi un score de Dice de 0.971 ± 0.015 , une distance ASSD de $0.954 \pm 0.704mm$ et une distance DH-95 de $3.61 \pm 3.76mm$. Le second obtient des résultats inférieurs, qui restent cependant soit meilleurs, soit légèrement inférieurs aux résultats du modèle de référence. Ces performances sont de 0.969 ± 0.017 (Dice), $0.954 \pm 0.817mm$ (ASSD) et $3.74 \pm 4.40mm$ (DH-95).

En ce qui concerne l'évaluation qualitative des segmentations, une sélection

Résultats de référence du modèle $A.3^{C4A}$

Calculés sur les trois sous-ensembles fixés pour comparaison avec les modèles évalués dans les autres expériences B et C :

Dice = 0.970 ± 0.018
 ASSD = 0.99 ± 0.88 mm
 DH-95 = 4.00 ± 4.81 mm

Résultats du modèle $C.6^{C4A}$

Calculés sur les trois sous-ensembles fixés :

Dice = 0.971 ± 0.015
 ASSD = 0.95 ± 0.70 mm
 DH-95 = 3.61 ± 3.76 mm

35: Pour rappel, ces deux modèles ont bénéficié successivement de deux transferts d'apprentissage.

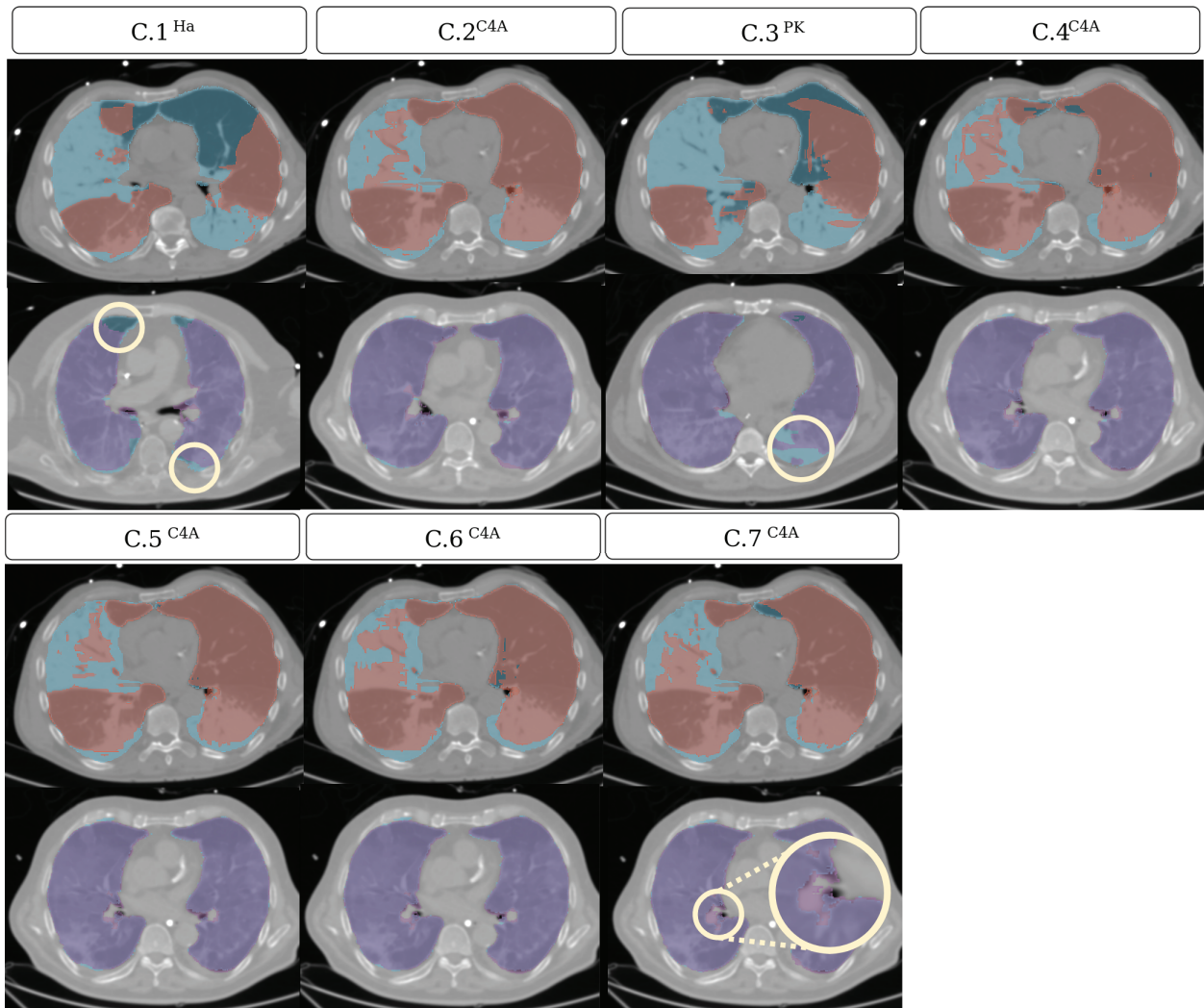


FIGURE 3.15 – Exemple des pires (en rouge) et des meilleures (en violet) segmentations des modèles $C.n$ sur la base de données CT4-ARDS, superposées avec la segmentation manuelle de référence en bleu.

d'exemples est présentée dans la Figure 3.15. Concernant les erreurs observées, elles surviennent toujours principalement dans les zones non aérées. On remarque que les modèles $C.1^{Ha}$ et $C.3^{PK}$, non entraînés sur la base CT4-ARDS, sont ceux qui échouent le plus dans ces zones qui sont largement sous-segmentées. Les autres modèles, $C.n^{patient}$, parviennent un peu mieux à segmenter la condensation postérieure du poumon gauche, bien qu'aucun d'eux ne se superpose parfaitement avec la segmentation manuelle. La grande opacité qui occupe le poumon droit demeure un obstacle pour tous les modèles. Du côté des meilleures segmentations obtenues avec ces modèles, on observe que les erreurs surviennent surtout pour les modèles non spécialisés aux patients $C.1^{Ha}$ et $C.3^{PK}$ qui sous-segmentent les condensations postérieures. Pour les autres modèles, la tendance est plutôt de sur-segmenter dans les zones médiastinales, en intégrant les bronches et les vaisseaux présents dans ces zones.

En somme, les modèles qui ont utilisé le transfert d'apprentissage ne montrent pas de gain significatif en robustesse par rapport au modèle de référence A.3. Les erreurs qui persistent sont celles de sous-segmentation dans les zones non aérées et de sur-segmentation au niveau du médiastin.

Limites de l'expérience L'objectif était d'évaluer la contribution du transfert d'apprentissage entre deux anatomies différentes des poumons. Le modèle de l'article qui a inspiré ce travail [Ger18] n'a pas été répliqué au vu de la contrainte temporelle qu'impose l'entraînement d'un tel modèle. De plus, la quantité de données à disposition et donc utilisée dans les expériences présentées ne garantit pas les proportions présentées dans le travail de Gerard et al. [Ger18]. En effet, le nombre de données utilisées en pré-entraînement n'est également pas du même ordre de grandeur, puisque ce sont 3 418 images CT qui ont été utilisées dans l'article, contre environ une centaine de volumes au maximum³⁶. La grande quantité d'expériences a été rendue possible par la légèreté et la rapidité d'entraînement des modèles multi-coupes θ_{vm} . Cependant, la faible complexité du modèle multi-coupes utilisé pourrait expliquer la difficulté à encoder efficacement les poumons des différentes anatomies et donc réduire l'apport des différentes combinaisons de données évaluées. La stratégie de fusion par vote majoritaire des résultats des trois modèles directionnels peut aussi être discutée. Bien qu'améliorant la robustesse (voir Tableau 3.7), celle-ci pourrait aussi introduire des erreurs³⁷ supplémentaires comme l'illustre la Figure 3.16.

36: Selon le sous-ensemble, voir Tableau 3.3.

37: Le cas où seul l'un des trois modèles $\theta_{direction}$ a correctement prédit la segmentation.

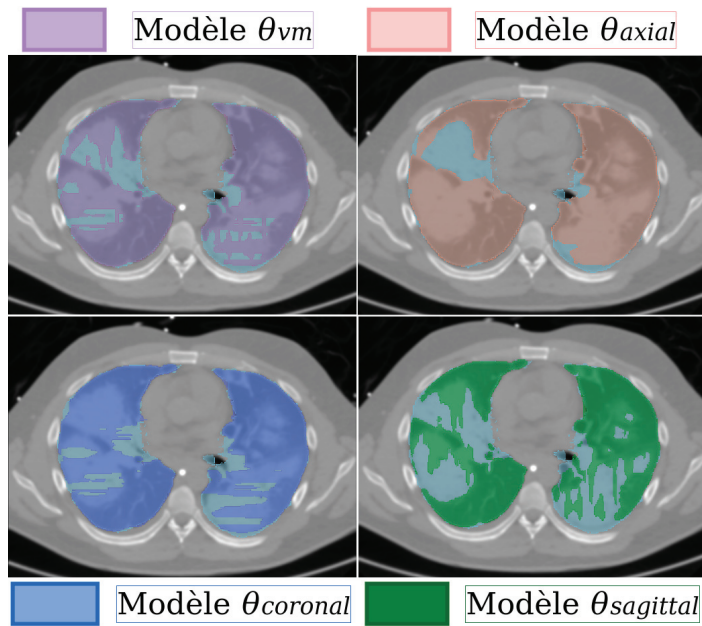


FIGURE 3.16 – Décomposition de la prédiction du vote majoritaire superposée avec la segmentation manuelle (en bleu).

3.6 Conclusion

Les différentes expériences menées durant cette étude, ne mettent pas en évidence un apport significatif des différentes stratégies de gestion des données. Seul un modèle parvient à dépasser les performances du modèle de référence, au prix de deux transferts d'apprentissage sur les deux bases de données animales.

En dehors des métriques géométriques, la surface des masques de segmentation générés par la méthode multi-coupes a tendance à être hachée. Cela est dû au fait que les masques sont générés coupes par coupes, et ce, dans

les trois directions orthogonales. L'opération de vote majoritaire, en plus d'effacer les erreurs ponctuelles que produirait un des trois sous-modèles $\theta_{direction}$, permet de lisser quelque peu cet aspect.

L'état d'avancement des objectifs de la thèse à la suite de cette première approche est présenté dans l'encadré suivant.

Objectifs initiaux et état d'avancement		
Cadre d'utilisation	Mise en application en clinique	<i>Recherche (CREATIS)</i>
Tâche	Segmentation des poumons sur image CT 3D	Méthode 2D multi-coupes
Cohérence	Fournit peu de différences entre deux images traitées du même patient	<i>Non évaluée</i>
Interactivité	Automatique	Oui
Délai de calcul	< 1h	Environ 15s (GPU) ou 2min (CPU)
Robustesse	Indépendant de la sévérité de SDRA et lésions présentes	<i>Dépendante</i>
Précision	Nécessite peu de retouches manuelles (si présentes)	<i>Erreurs de sous-segmentation à corriger, et irrégularités de surface</i>

En gras les objectifs atteints dans ce chapitre, en police normale ce qui a été atteint dans les chapitres précédents et en italique ce qu'il reste à faire.

Approche 3D : exploiter l'information dans les données

4

Les expériences du chapitre précédent avaient pour but d'identifier une stratégie de présentation des données qui puisse augmenter la robustesse à partir d'images provenant de différentes anatomies. Les stratégies évaluées n'ont pas mis en évidence de gain en robustesse de l'utilisation des images de modèles porcins expérimentaux, ce qui peut être expliqué par la faible capacité du modèle multi-coups basé sur l'architecture du U-net 2D. Afin de pallier cette limite, l'utilisation d'un modèle 3D¹, pour répondre aux questions de gestion optimale des données, est évaluée dans ce chapitre. Diverses expériences sont menées avec comme objectif de mesurer l'impact sur la robustesse et sur la généralisation de l'ajout d'informations provenant des images durant l'entraînement. Pour cela, sont évaluées l'initialisation (avec le transfert d'apprentissage), la diversité², la résolution des images³ et l'utilisation de valeurs physiologiques dans la fonction de perte. L'un de ces éléments peut-il contrebalancer la limite computationnelle qu'impose l'utilisation d'une architecture 3D?

1: Ceci apporte alors des limitations computationnelles dues, entre autre, à l'augmentation de la complexité des calculs des convolutions (2D vers 3D).

2: Dans les lots d'images utilisés pour calculer la perte et actualiser les poids lors de l'entraînement.

3: En évaluant alors la finesse des détails observables dans les images.

Table des matières

4.1	Introduction	54
4.2	Données	55
4.3	Description d'expérience	59
	Plan d'expérience	59
	Modèle, entraînement et inférence	63
	Évaluation	64
4.4	Résultats	66
	Expériences sur l'initialisation	66
	Expériences sur la quantité d'information	69
	Expérience sur l'apport d'information physiologique	73
4.5	Discussion	77
4.6	Conclusion	82

4.1 Introduction

À la suite des expériences menées dans le chapitre précédent, la question reste ouverte quant à la meilleure exploitation possible des images pour gagner en robustesse et en capacité de généralisation. En effet, la capacité du modèle 2D pouvant être un frein aux approches par transfert du fait de l'amnésie[De +22]. Afin de dépasser les possibles limites⁴ rencontrées dans l'expérience 2D, le recours à une architecture 3D est choisi. Il est attendu de cette augmentation de capacité et du champ de vue 3D apporté que les prédictions soient davantage cohérentes spatialement, avec un espace latent qui tienne mieux compte du contexte des zones à segmenter. Cependant, le changement des convolutions 2D aux convolutions 3D implique également une augmentation conséquente de la mémoire de calcul⁵ nécessaire. Diverses équipes [Çiç+16; Mil+16] ont

4: Peu de robustesse gagnée des transferts d'apprentissage ou du mélange des bases de données utilisées à l'entraînement.

5: A la fois à l'entraînement et à l'inférence.

montré qu'il était possible de trouver un compromis entre taille d'image d'entrée et capacité de calcul. Cependant, ce compromis se fait souvent aux dépens de la résolution spatiale⁶ ou du champ de vue⁷ de l'image d'entrée, qui perd alors en détails ou en contenu.

Pourtant, un courant d'idée du domaine de l'apprentissage profond [Hof+20; Ise+21] indique que la robustesse d'un modèle réside davantage dans la richesse des données qui ont été utilisées à l'entraînement plutôt que dans la complexité⁸ du modèle entraîné. Ceci s'exprime entre autres dans la diversité des morphologies, des lésions ou encore des anomalies représentées dans les données utilisées pour l'entraînement, mais aussi dans la quantité de données annotées. Cependant, celles disponibles pour l'application de la segmentation des poumons chez les patients atteints du SDRA est très limitée. De plus, le SDRA prenant une variété de motifs différents dans les images CT, conserver les détails contextuels (et donc une résolution élevée) semble être une priorité au vu de la difficulté de la tâche.

L'idée est alors de trouver un ou des moyens d'exploiter de manière optimale les données pour augmenter la robustesse de la segmentation des poumons⁹ tout en maintenant une architecture relativement légère. En effet, pour que le modèle soit utilisable en clinique sur du matériel accessible¹⁰ des services de réanimation des hôpitaux, il faut qu'il ne soit pas trop gourmand en mémoire de calcul. Plusieurs axes sont évalués pour tenter de répondre à cet objectif.

Le premier fait directement suite aux expériences menées dans le chapitre précédent, puisqu'il évalue le transfert d'apprentissage avec une architecture 3D. En spécialisant un modèle¹¹ à la tâche de segmentation de poumons sur des patients atteints du SDRA, l'objectif est de dépasser une initialisation aléatoire. La capacité de généralisation obtenue des diverses anatomies vues par le réseau est aussi évalué.

Le deuxième axe se concentre sur la quantité d'information traitée à chaque itération. En effet, la limite computationnelle de l'architecture 3D impose un compromis entre le nombre d'images prises par lot et la taille de ces images. Cet aspect est évalué selon deux points de vue : (1) le nombre d'images utilisées pour actualiser les poids du modèle et (2) la quantité de détails¹² présents dans une même image. L'objectif est d'évaluer lequel de ces deux aspects influence le plus la robustesse et la capacité de généralisation.

Enfin, le dernier axe concerne non pas la quantité d'information utilisée à l'entraînement, mais sa pertinence¹³. En effet, ce dernier point a pour objectif d'intégrer dans l'apprentissage des informations physiologiques¹⁴ afin de dépasser les compromis qu'impose l'utilisation d'une architecture 3D.

L'évaluation de tous ces aspects est examinée à la fois d'un point de vue géométrique, mais aussi du point de vue de l'utilisation en clinique. En effet, l'objectif final est l'intégration d'un modèle de production dans un logiciel d'aide à la décision.

4.2 Données

Davantage de données ont été utilisées dans les expériences de ce chapitre, en comparaison de celles utilisées dans le chapitre précédent. Cela s'explique par la constitution d'une nouvelle base de données durant le déroulement de cette thèse.

6: Qui est diminuée en sous-échantillonnant l'image pour réduire sa taille.

7: Avec le recours aux patches.

8: Un compromis est à trouver, une capacité minimum du modèle étant tout de même nécessaire.

9: Chez les patients atteints du SDRA.

10: Un équipement informatique qui ne nécessite pas de financement spécifique.

11: Pré-entraîné à segmenter des poumons sur des modèles expérimentaux de SDRA procin.

12: Ici faisant référence à la résolution spatiale de l'image.

13: Qui désigne ici la nature de l'information apportée et son impact dans la robustesse du modèle.

14: Usuellement utilisées comme critère de qualité ou de diagnostic par les cliniciens

Les bases de données utilisées pour les expériences suivantes sont privées et ont été acquises durant le déroulement de cette thèse.

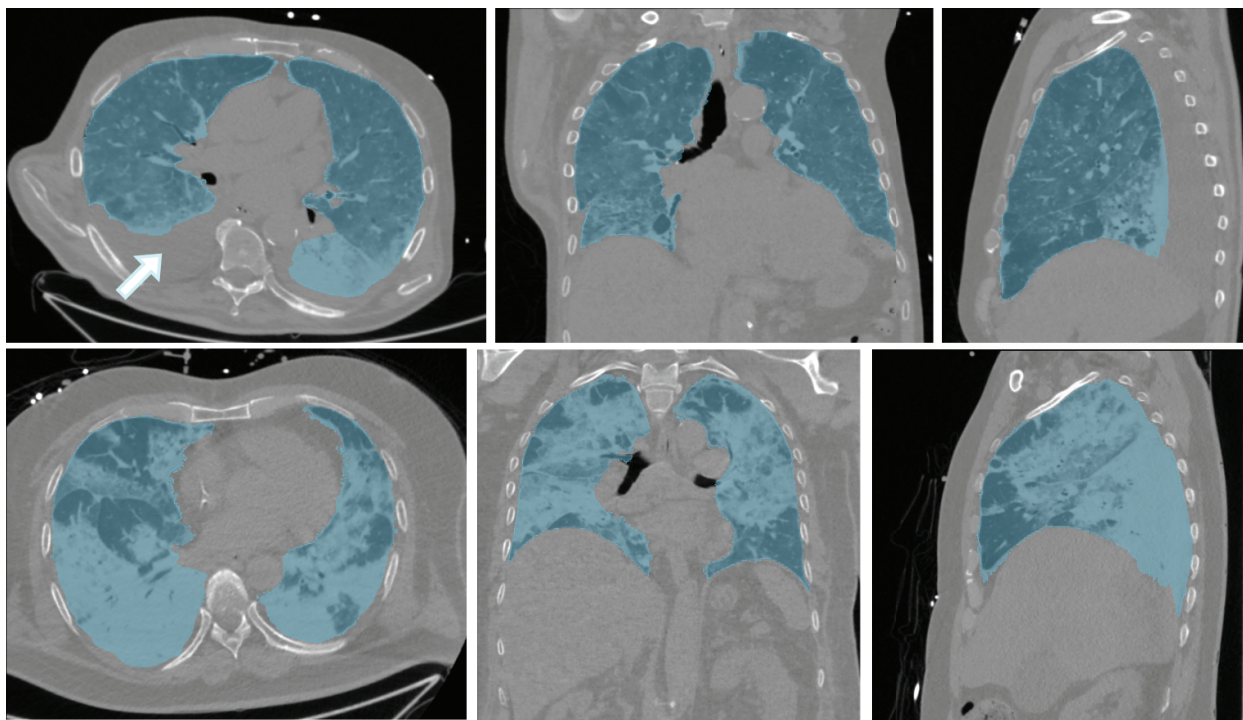


FIGURE 4.1 – Exemple de deux patients de la base CT4-ARDS (vue axiale, coronale et sagittale) avec leur segmentation manuelle en bleu – Sur la première ligne, un patient avec SDRA modéré et un épanchement pleural indiqué par la flèche sur la coupe axiale; sur la deuxième ligne, un patient avec SDRA sévère.

La première base de données, déjà présentée dans le chapitre précédent, est nommée CT4-ARDS. Réalisée dans le cadre d'une étude clinique,¹⁵ elle compte 30 patients admis au service de médecine intensive en réanimation de l'hôpital de la Croix-Rousse, entre mai 2019 et mai 2020. Pour chaque patient, deux¹⁶ images CT ont été acquises sur un scanner iCT 256 ou bien Ingeniuty CT (Philips Healthcare, Eindhoven, Pays-Bas). Les paramètres d'acquisition fournissent des volumes avec une épaisseur de coupe de 1mm et une taille de pixel d'environ $0.7\text{mm} \times 0.7\text{mm}$. Le nombre de coupes varie entre 230 et 323 pour une dimension axiale de 512×512 . Toutes les images ont été segmentées de manière semi-automatique¹⁷ par des cliniciens du service de médecine intensive réanimation. Le temps de segmentation moyen par volume est d'environ une heure. Cette base de données comporte un nombre égal de patients avec SDRA modéré ou SDRA sévère¹⁸ (voir exemple de patient pour chaque niveau de sévérité en Figure 4.1). D'autres lésions pulmonaires sont aussi présentes chez certains patients, comme c'est le cas du pneumothorax¹⁹, du pneumomédiastin²⁰ et de l'épanchement pleural²¹ (voir la zone indiquée par une flèche dans la Figure 4.1). Le résumé des informations des patients est décrit dans le Tableau 4.1.

La seconde base de données utilisée est nommée CT4-ARDS2, car elle fait suite à la première étude clinique (CT4-ARDS) avec cette fois-ci un objectif d'évaluation de la performance diagnostique de la mesure semi-automatisée de l'hyperinflation cyclique. Pour cette seconde étude, une collaboration a eu lieu avec un autre hôpital de la région Rhône-Alpes. L'étude bi-centrique compte alors deux centres d'acquisition : le service de médecine intensive réanimation de l'hôpital de la Croix-Rousse (Lyon) et le

15: Visant à évaluer une mesure semi-automatisée de l'hyperinflation cyclique pulmonaire au scanner CT. Pour rappel, cela désigne un processus qui survient lorsque des alvéoles déjà aérées en fin d'expiration sont hyperinflées en fin d'inspiration, et qui est alors cyclique au cours de la respiration.

16: L'une en fin d'inspiration et l'autre en fin d'expiration.

17: Voir section 3.2 pour la description succincte du logiciel utilisé [Dáv+20]

18: Les formes les plus graves et les plus léthales du SDRA

19: Présence d'air dans la cavité pleurale.

20: Présence d'air dans le médiastin.

21: Accumulation anormale de liquide dans la cavité pleurale.

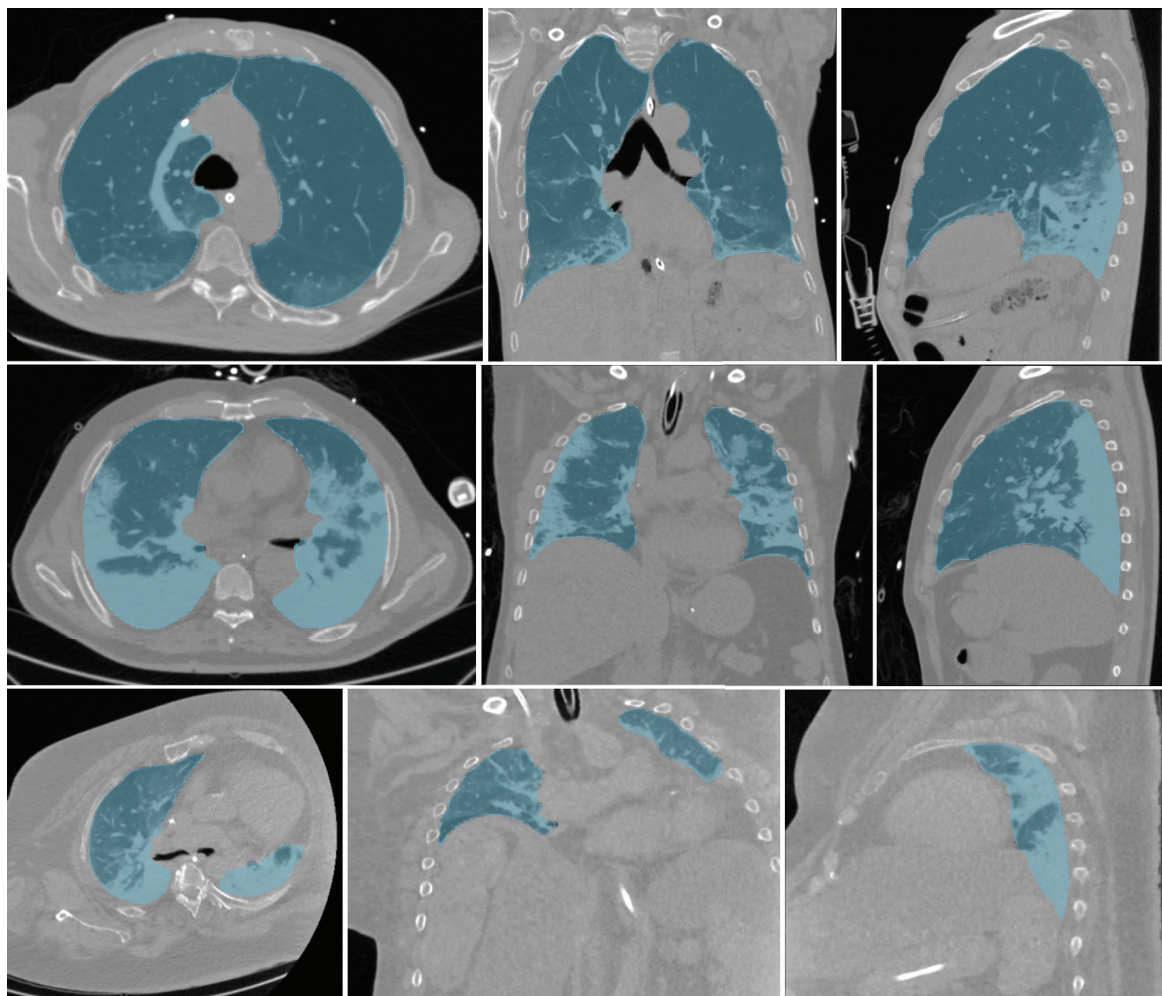


FIGURE 4.2 – Exemple de trois patients de la base CT4-ARDS2 (Lyon-2), en vue axiale, coronale et sagittale, avec leur segmentation manuelle en bleu – Sur la première ligne, un patient avec SDRA modéré; sur la deuxième ligne, un patient avec SDRA sévère; sur la troisième ligne, un patient avec SDRA sévère sous assistance extracorporelle (ECMO).

service de médecine intensive réanimation du CHU Grenoble (Grenoble). Ce sont ainsi 101 patients qui sont inclus dans l'étude représentant un total de 366 volumes. Le protocole d'acquisition conduit à quatre images CT acquises dans les conditions suivantes :

- une image en fin d'inspiration à PEP²² fixée par le clinicien,
- une image en fin d'expiration à PEP fixée par le clinicien,
- une image en fin d'expiration à PEP 5cmH₂O,
- une image en fin d'expiration à PEP 15cmH₂O.

22: Pour rappel, pression expiratoire positive qui maintient les alvéoles aérées entre la fin de l'inspiration et l'expiration.

Les images ont été acquises entre octobre 2020 et décembre 2021 sur un scanner iCT 256 ou bien Ingeniuty CT (Philips Healthcare, Eindhoven, Pays-Bas) pour le centre lyonnais et sur un scanner GE Optima CT scan (GE Medical Systems, Milwaukee, USA) pour le centre grenoblois. Les paramètres d'acquisition respectifs de chaque centre fournissent des volumes avec une épaisseur de coupe de 1mm (Lyon) et 1.25mm (Grenoble) et une taille de pixel d'environ 0.7mm × 0.7mm. Le nombre de coupes varie entre 231 et 346 pour une dimension axiale de 512 × 512. Les segmentations de référence sont également réalisées de manière semi-automatique à l'aide du logiciel développé par CREATIS [Dáv+20] par les cliniciens des différents centres. La durée de segmentation manuelle par volume

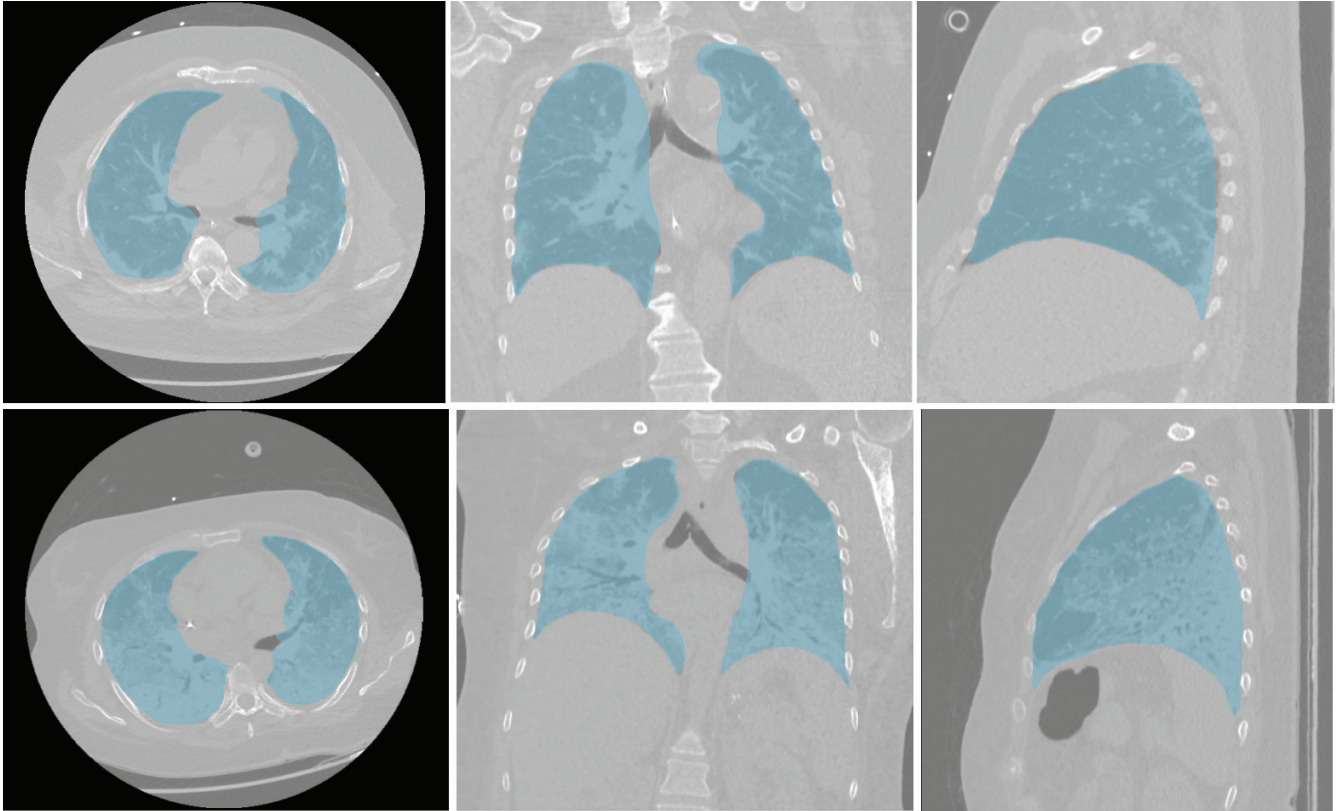


FIGURE 4.3 – Exemple de deux patients de la base CT4-ARDS2 (Grenoble-2), en vue axiale, coronale et sagittale, avec leur segmentation manuelle en bleu – Sur la première ligne, un patient avec SDRA modéré; sur la deuxième ligne, un patient avec SDRA sévère.

varie selon la sévérité du SDRA ; la difficulté de la segmentation étant liée à l'intensité de la perte d'aération pulmonaire. Elle est d'environ une heure, soit quatre heures au total par patient. Les masques de référence sont tous révisés et corrigés si nécessaire par le chef du service de réanimation de Lyon. Les patients composant cette base de données sont majoritairement atteints de SDRA sévère (84%) et positifs à la maladie COVID-19 (99%). Parmi les patients avec SDRA sévère, un tiers est sous assistance extracorporelle (ECMO) du fait de la gravité de son atteinte. Trois patients illustrant chaque niveau de sévérité rencontré dans la base de données CT4-ARDS2 (données du centre lyonnais) sont présentés dans la Figure 4.2. L'exemple du patient avec SDRA sévère et sous ECMO (en troisième ligne de la Figure 4.2) témoigne de l'effondrement alvéolaire très présent chez ce type de patients et par conséquent de la difficulté de la segmentation manuelle,²³ qui a pris jusqu'à deux fois plus de temps pour ces patients. En raison de plusieurs critères, certaines images sont exclues de la base de données : cela survient lorsqu'une image est artéfactée ou que des erreurs sont survenues durant l'acquisition. Ainsi, le nombre de volumes disponibles est inférieur²⁴ à celui initialement prévu par le protocole.

Les deux bases de données sont séparées en vue d'être utilisées soit pour l'entraînement, soit pour le test des modèles (en inférence). La base de données CT4-ARDS (Lyon-1) est utilisée pour l'entraînement de toutes les expériences de ce chapitre. Les images de cette base de données subissent ainsi un ré-échantillonnage²⁵ pour obtenir des voxels isotropes de taille 1 mm, 2 mm ou 3 mm (selon l'expérience). Afin d'avoir une taille d'image

23: Aucune base de données de la littérature ne fait état de poumons de patients sous ECMO segmentés.

24: Voir Tableau 4.1 où le nombre de volumes dans la base de données CT4-ARDS2 est inférieur à $4 \times n_{\text{patients}}$, avec n_{patients} le nombre de patients.

25: Précédé d'un filtre gaussien.

TABLEAU 4.1 – Détails des bases de données CT4-ARDS et CT4-ARDS2 utilisées respectivement pour l'entraînement ou le test.

Base de données	Entraînement		Test	
	CT4-ARDS	CT-4ARDS2	CT4-ARDS2	CT4-ARDS2
Identifiant	Lyon-1	Lyon-2	Grenoble-2	
Patients	30	13	21	
Volumes	60	50	68	
Sévérité				
Modéré	15 (50%)	2 (15%)	9 (43%)	
Sévère	15 (50%)	9 (69%)	12 (57%)	
Sévère sous ECMO	0	2 (15%)	0	
Autre lésion pulmonaire*	4 (13%)	3 (23%)	1 (5%)	

fixe par base de données, elles sont ensuite redimensionnées²⁶ à la taille 448 × 448 × 320 (voxel de 1 mm), 192 × 192 × 192 (voxel de 2 mm) et 128 × 128 × 128 (voxel de 3 mm). La base de données CT4-ARDS2 est quant à elle utilisée exclusivement pour le test. Un sous-ensemble des données acquises à Lyon est utilisé pour évaluer la généralisation, notamment sur des images de patients aux atteintes très sévères. Ce sous-ensemble, composé de 13 patients (50 volumes) est noté Lyon-2. Les images acquises au centre grenoblois (CT4-ARDS2) sont désignées par l'identifiant Grenoble-2. Sur ces dernières, un pré-traitement est appliqué afin de supprimer le cercle noir autour du corps du patient (voir coupes axiales de la Figure 4.3). Ainsi, tous les voxels qui font partie de ce fond noir sont modifiés pour correspondre à l'unité Hounsfield de l'air.

Afin de garantir la reproductibilité des expériences et surtout l'utilisation de toutes les données d'un patient dans une même tâche (l'entraînement, la validation ou le test), les pointeurs vers les données sont organisés dans des fichiers *hdf5*²⁷ (*hierarchical data format*). Cette organisation impose le recours à des scripts automatisés pour l'édition des données utilisées. Une arborescence est ainsi créée (voir Figure 4.4), qui sépare pour chaque jeu de données²⁸ les images correspondantes aux différentes conditions ventilatoires de chaque patient. Ces fichiers sont créés pour chaque base de données en fonction du nombre de sous-ensembles utilisés dans l'évaluation.

4.3 Description d'expérience

Plan d'expérience

L'objectif des expériences suivantes est d'identifier l'élément qui est déterminant sur la robustesse durant l'apprentissage. Ces éléments ont été limités à trois, et sont décrits ci-dessous :

- l'initialisation,
- la quantité d'information par itération,
- la cohérence et la pertinence de l'information traitée.

26: Coupées ou agrandies par remplissage.

* : Pneumothorax, pneumomédiastin ou épanchement pleural.

27: Format de fichier qui permet l'enregistrement de grandes quantités de données.

28: Entraînement, validation, test.

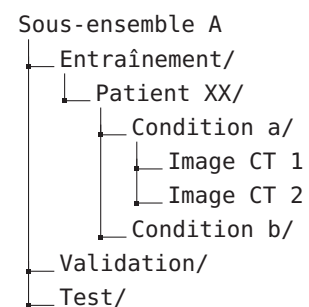


FIGURE 4.4 – Arborescence des fichiers *hdf5* de description des bases de données.

Informations pratiques

Tous les codes de développement sont écrits en Python 3.9 et utilisent comme bibliothèques majeures PyTorch (1.8 ou 1.12), Numpy (1.20) et SimpleITK (2.0.2).

Dans le but de répondre à chacun de ces axes, diverses expériences ont été menées et seront détaillées ci-dessous. Le même modèle est utilisé pour toutes les expériences afin d'évaluer l'impact des paramètres tels que la résolution de l'image d'entrée, la diversité ou la cohérence entre elles des images présentées par lot et les informations physiologiques utilisées dans la fonction de perte.

Expérience D : l'initialisation Le premier axe évalué fait suite aux expériences menées dans le chapitre précédent sur le transfert d'apprentissage. En effet, les diverses combinaisons de transfert d'apprentissage évaluées avec un modèle $3 \times 2D$ ont montré une faible capacité d'amélioration de la robustesse. L'idée est d'évaluer si l'utilisation d'un modèle à la capacité augmentée (3D) permet d'encoder plus efficacement les caractéristiques volumiques des poumons (peu importe l'anatomie) avec SDRA. Cependant, pour respecter le compromis entre temps d'entraînement et ressources disponibles, le nombre d'expériences évaluées est diminué par rapport aux expériences menées en 2D. La base de données d'entraînement sur laquelle sont affinés les poids reste la base CT4-ARDS (Lyon-1), pour sa composition équilibrée des niveaux de sévérité (voir le Tableau 4.1). Les jeux de données Lyon-2 et Grenoble-2 sont utilisées pour évaluer la capacité de généralisation de ces modèles.

Trois modèles sont ainsi comparés pour évaluer l'apport du transfert d'apprentissage, en termes de robustesse et de capacité de généralisation. Le premier, celui de référence, est l'entraînement d'un U-net 3D (pendant 100 epochs), initialisé aléatoirement et noté D_a . Le deuxième, D_t , utilise comme initialisation les poids d'un modèle entraîné (pendant 100 epochs) sur les deux bases de données animales²⁹ afin d'être spécialisé (pendant 100 epochs) sur la base de données Lyon-1. Le troisième modèle (D_a^{200}) suit le même entraînement que le premier, à l'exception qu'il est entraîné pendant deux fois plus d'epochs, pour correspondre au nombre d'itérations d'entraînement du modèle qui a subi le transfert d'apprentissage (D_t). Tous les entraînements de cette expérience utilisent la base de données Lyon-1 pré-traitée pour contenir des images de résolution spatiale isotrope de 1 mm. Le résumé des modèles est présenté dans le tableau 4.2.

Expérience E : la quantité d'information Cette deuxième expérience évalue l'impact des caractéristiques présentes dans les images utilisées à l'entraînement. La quantité d'information présente dans les images est évaluée selon deux hyper-paramètres : la taille de lot et la résolution spatiale de l'image d'entrée.

Durant l'entraînement, à chaque itération, une ou plusieurs images sont passées à travers le réseau, ce qui mène à l'actualisation des poids de ce dernier par rétropropagation du gradient de la fonction de perte. Ce processus est usuellement réalisé en regroupant les images par lot. Ceci permet d'accélérer le traitement d'une part, mais aussi d'augmenter la quantité d'informations utilisée pour actualiser les poids du réseau. En faisant varier la composition des lots fournis au réseau, l'objectif est ainsi d'évaluer comment cela influence les performances du modèle. Cependant, en fonction de la taille de l'image en entrée, la taille du lot est limitée par l'espace mémoire nécessaire aux calculs pendant l'entraînement. Ce paramètre est ainsi généralement réglé pour respecter cette limite matérielle. De la même manière, la taille des images en entrée du modèle est également

TABLEAU 4.2 – Modèles utilisés pour l'évaluation de l'initialisation.

Modèle	Initialisation	Epoch
D_a	aléatoire	100
D_a^{200}	aléatoire	200
D_t	transfert	100

29: Décrites en détail dans la section Données du Chapitre 3.

influencée³⁰ par la limite des ressources de calcul disponibles. Réduire la taille des images d'entrée est une solution qui se fait au détriment des détails³¹ qui sont alors perdus lors du ré-échantillonnage de l'image. Parmi les autres solutions observées dans la littérature, utiliser des patchs plutôt que l'image entière permet de maintenir une grande résolution spatiale tout en diminuant la taille de l'image en entrée. Cependant, cette option ne convient pas dans toutes les applications. En effet, en diminuant le champ de vue représenté dans un patch, le risque est de produire des patchs au même aspect visuel, mais représentant différentes zones. Le contexte ne peut particulièrement pas être ignoré dans le cadre de la segmentation des poumons de patients avec SDRA, la perte de contraste à l'interface des zones non aérées et des tissus étant source d'erreur. Utiliser le contexte, même dégradé en résolution spatiale, permettrait donc de gagner en robustesse [Ger18].

Cette expérience vise donc à évaluer lequel des deux éléments discutés apporte le plus de robustesse : la quantité (et donc diversité des images) vus à chaque itération ou bien la quantité de détails présents dans peu d'images. Le premier élément est représenté par la taille de lot des images, notée β . Le deuxième élément est représenté par la résolution spatiale des images en entrée, notée r . Chaque modèle est ainsi noté E_r^β avec r la résolution d'entraînement et β la taille des lots d'images utilisés à l'entraînement. Cinq tailles de lots sont évaluées, avec une particularité pour la taille $\beta = 2$, pour laquelle deux cas sont évalués : (1) deux images aléatoires de la base de données et (2) deux images du même patient³². Toutes les autres tailles de lots impliquent une composition aléatoire des images. Trois résolutions spatiales sont évaluées pour cette expérience : 1 mm, 2 mm et 3 mm pour la taille isotrope d'un voxel. L'effet de ces trois tailles d'échantillonnage est illustré dans la Figure 4.5. En raison des limites matérielles, pour la résolution spatiale de 1 mm la taille maximale de lot utilisée est de 2. Ainsi, 15 modèles résumés dans le tableau 4.3 sont entraînés sur la base de données Lyon-1.

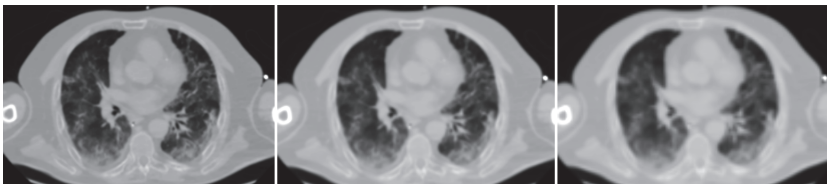


FIGURE 4.5 – Coupe axiale à trois résolutions différentes, illustrant la perte de détails lors du sous-échantillonnage. – Pour des voxels isotropes de taille 1 mm (gauche), 2 mm (centre) et 3 mm (droite).

Expérience F : l'information physiologique Cette dernière expérience a pour objectif d'intégrer davantage d'information physiologique durant l'entraînement. L'idée est d'utiliser la fonction de perte pour intégrer cette information à chaque itération de l'apprentissage. Pour cela, deux concepts sont étudiés : l'aération et la masse pulmonaire.

D'un côté, l'information d'aération est séparée en quatre compartiments, définis par des seuils (voir le tableau 4.4) sur l'unité Hounsfield de chaque voxel de l'image CT [Gat+01]. L'utilisation des différents compartiments d'aération devrait permettre de pénaliser ou avantager la segmentation de certaines zones des poumons pendant l'apprentissage. De l'autre côté, la prédiction du masque 3D à chaque itération permet de calculer la masse des

30: D'autant plus lorsqu'il s'agit d'une architecture de CNN 3D.

31: Plus petit élément qu'il est possible de distinguer. Cette information est liée à la résolution spatiale de l'image.

32: Ce choix est motivé par la cohérence apportée par deux images d'un même patient, contre la diversité de deux images provenant de deux patients.

TABLEAU 4.3 – Modèles E_r^β entraînés avec des images de résolution spatiale r et par lot de taille β .

Lot d'images	Résolution (mm)		
	1	2	3
1	E_1^1	E_2^1	E_3^1
2	E_1^2	E_2^2	E_3^2
2 (1 patient)	E_1^{2p}	E_2^{2p}	E_3^{2p}
4	-	E_2^4	E_3^4
8	-	E_2^8	E_3^8
16	-	E_2^{16}	E_3^{16}

TABLEAU 4.4 – Seuils des compartiments d'aération selon leur densité en unité Hounsfield.

Compartiment d'aération	Seuils (HU)
Non aéré (NA)	[-100; +100]
Peu aéré (PA)	[-500; -101]
Aéré (A)	[-900; -501]
Hyperinflaté (HI)	[-1000; -901]

poumons prédits en même temps. À partir de cette information, l'objectif est d'atteindre une cohérence de segmentation entre les masques de deux images du même patient³³. Ce critère a d'ailleurs déjà été utilisé dans la littérature [Yin+09; Gor+12] pour garantir la cohérence entre deux images lors de l'opération de recalage³⁴.

Afin d'exploiter ces informations physiologiques pendant l'apprentissage, elles sont individuellement intégrées à chaque itération via la fonction de perte. Une fonction de perte basée sur le score de Dice³⁵ est utilisée et définie comme suit :

$$\mathcal{L}_{Dice}(\alpha) = 1 - \text{Dice}(\alpha) \quad (4.1)$$

avec α le compartiment d'aération³⁶ sur lequel est calculé le score de Dice. Lorsque α n'est pas indiqué, le score de Dice est calculé sur toute l'image et la fonction notée \mathcal{L}_{Dice} . Un ou plusieurs termes sont ajoutés à cette fonction de base pour former la fonction de perte \mathcal{L} du modèle. Les modifications apportées à la fonction de perte sont faites en garantissant que la fonction demeure dérivable, condition indispensable à l'actualisation des poids par l'algorithme d'optimisation. En fonction de la grandeur physiologique utilisée, différentes fonctions sont définies ci-après.

Dans un premier temps, la pondération de la fonction de perte par zone d'aération est évaluée. Le compartiment non aéré est le plus compliqué à segmenter, car composé de densités similaires aux tissus environnant les poumons. L'objectif est donc de pénaliser davantage les erreurs dans cette zone, et au contraire récompenser les prédictions qui se rapprochent le plus possible de la référence. Pour ce faire, le calcul du Dice est utilisé sur le compartiment non aéré. Compris entre 0 et 1, il sert de pondération pour la fonction de Dice calculée sur toute l'image. La fonction de perte qui en découle est de la forme :

$$\mathcal{L}_{NA} = 1 - (\text{Dice} \times \text{Dice}(NA)), \quad (4.2)$$

avec NA le compartiment non aéré. Le modèle entraîné à l'aide de cette fonction de perte, basée sur le compartiment d'aération, est noté F_α .

Dans un deuxième temps, l'intégration de l'information massique des poumons dans l'apprentissage est évaluée. Le calcul de la masse se déduit directement de la définition³⁷ des unités Hounsfield (UH), qui expriment la densité des tissus. Chaque voxel de l'image représente l'atténuation moyenne des tissus considérés dans le volume. Obtenir la masse se fait via la formule suivante :

$$m_{voxel} = \begin{cases} 0 & \text{si } I_{voxel} < -1000, \\ \frac{I_{voxel} + 1000}{1000} \cdot V_{voxel} \cdot \mu_{tissu} & \text{si } -1000 \leq I_{voxel} \leq 0, \\ 1 \cdot V_{voxel} \cdot \mu_{tissu} & \text{sinon.} \end{cases} \quad (4.4)$$

avec m_{voxel} la masse d'un voxel en grammes, I_{voxel} l'intensité en unité Hounsfield du voxel, V_{voxel} le volume d'un voxel en millilitres et μ_{tissu} la densité des tissus³⁸ en grammes par millilitre.

Une première version naïve de cette fonction de perte pourrait être de comparer uniquement les masses prédites à l'inspiration et l'expiration pour un même patient. La contrainte porterait alors sur la consistance entre les deux masques prédits, sans assurance d'un quelconque lien avec la

33: Puisque la masse des poumons varie peu entre l'inspiration et l'expiration.

34: Tâche du traitement d'images qui consiste à superposer deux images de la même scène prises à deux instants différents ou avec deux modalités différentes.

35: Qui apporte l'information de superposition de la prédiction avec la référence.

36: La sous-partie du masque de référence qui correspond aux voxels du compartiment d'aération α dans l'image CT associée.

37: Intensité I d'un voxel en unité Hounsfield :

$$I_{voxel} = 1000 \times \frac{(\mu_{tissu} - \mu_{eau})}{\mu_{eau}}, \quad (4.3)$$

avec μ_{tissu} et μ_{eau} respectivement la densité du tissu présent dans le voxel et celle de l'eau.

38: Qui est supposée égale à 1 g/mL.

réalité anatomique. La fonction de perte associée serait de la forme :

$$\mathcal{L}_{masse_{naive}} = \frac{|m_{exp} - m_{ins}|}{\hat{M}} \quad (4.5)$$

avec m_{exp} la masse prédite à l'expiration, m_{ins} la masse prédite à l'inspiration et \hat{M} la masse moyenne réelle.

Une deuxième proposition, celle adoptée, est d'intégrer la réalité anatomique en comparant les différences de masse à l'inspiration et à l'expiration entre masques prédits et masques réels. La fonction de perte proposée est alors de la forme :

$$\mathcal{L}_{masse} = \frac{|M_{exp} - m_{exp}| + |M_{ins} - m_{ins}| + |m_{exp} - m_{ins}|}{\hat{M}} \quad (4.6)$$

avec m_{exp} la masse prédite à l'expiration, m_{ins} la masse prédite à l'inspiration, M_{exp} la masse réelle à l'expiration, M_{ins} la masse réelle à l'inspiration et \hat{M} la masse moyenne réelle³⁹.

L'association de la fonction de perte basée sur la masse à celle globale est proposée sous cette forme :

$$\mathcal{L} = \mathcal{L}_{Dice} + \gamma \mathcal{L}_{masse}, \quad (4.7)$$

avec γ le coefficient de pondération du terme de masse. La question se pose alors du choix du coefficient, de même que son évolution dans le temps. En effet, l'information de masse d'un masque prédit ne semble pas pertinente pour la segmentation si elle n'est pas associée avec une prédiction déjà localisée sur la zone des poumons. Pour ce faire, trois stratégies de mélange des deux termes de la fonction de perte sont comparées. La première évalue l'apprentissage conjoint à l'aide de ces deux termes (voir γ_1 dans la Figure 4.6). Les deux suivantes évaluent l'ajout du terme de masse à partir d'un seuil d'apprentissage, fixé arbitrairement lorsque la perte vaut moins de 0.1. Il y a un ajout abrupt à partir de ce seuil pour γ_2 et un ajout progressif pour γ_3 du terme \mathcal{L}_{masse} . Les trois mélanges sont notés ϕ_X avec $X \in \{1, 2, 3\}$ et sont présentés dans la Figure 4.6. Les modèles utilisant cette fonction sont notés F_{masseX} pour $X \in \{1, 2, 3\}$ indiquant le mélange des fonctions de coût utilisé.

Modèle, entraînement et inférence

À la suite des expériences menées dans le précédent chapitre, augmenter la capacité d'encodage du modèle a motivé le choix d'un modèle 3D. La différence fondamentale par rapport aux réseaux convolutifs 2D⁴⁰ réside dans l'utilisation des convolutions 3D. Au regard des performances du U-net 2D, le choix s'est tout naturellement porté sur la version 3D de celui-ci [Çiç+16].

Le modèle utilisé est quelque peu modifié par rapport à la publication originale du U-net 3D. Le nombre de couches de convolution est augmenté (de 3 à 5), mais le nombre de filtres de convolution par couche est réduit⁴¹. La sortie de la couche initiale de convolution contient ainsi 4 filtres, contre 32 dans l'article du U-net 3D. Par rapport à l'architecture originale, le nombre

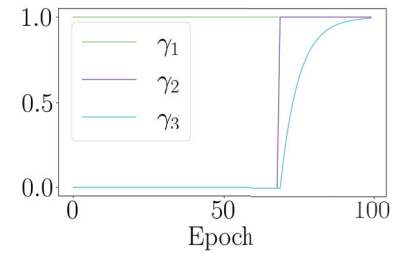


FIGURE 4.6 – Coefficients γ dans les trois mélanges de fonction de perte évalués.

39: Calculée à partir des segmentations de référence sur les images acquises en fin d'inspiration et en fin d'expiration.

40: Tels que le U-net 2D [Ron+15] utilisé dans le Chapitre 3

41: Cela dans le but de réduire des ressources de calcul nécessaires.

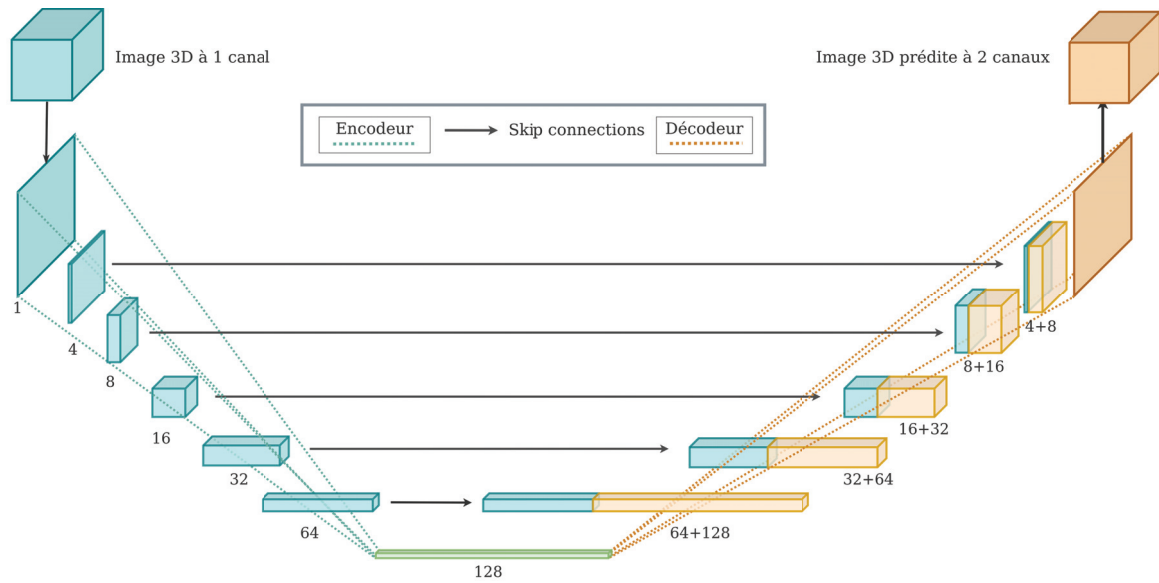


FIGURE 4.7 – Schéma de l'architecture du U-net 3D utilisé. – Chaque pavé représenté correspond à la sortie d'une couche de convolution composée de deux convolutions et d'une étape de réduction (ou augmentation) de dimension pour l'encodeur (ou le décodeur). Le nombre de filtres de chaque sortie de couche de convolution est indiqué sous chaque pavé.

de paramètres entraînaibles passe de 19 069 955 à 2 066 130. L'architecture utilisée est détaillée en Figure 4.7.

En dehors des hyper-paramètres étudiés, voici les valeurs communes à toutes les expériences : le taux d'apprentissage utilisé dans l'algorithme d'optimisation est fixé à $3 \cdot 10^{-3}$, 100 epochs sont utilisées pour l'entraînement standard. La taille de lots utilisée sur les entraînements standards⁴² est de 2 images aléatoires par lot. Le modèle retenu pour l'inférence est celui pour lequel la fonction de perte est la plus faible sur la sous-partie de validation. Tous les modèles sont entraînés sur les ressources de calcul du laboratoire CREATIS, notamment pour les modèles entraînés avec des images de résolution 1 mm, en calcul parallèle sur deux GPU de 32 Go minimum.

À l'inférence, les prédictions subissent un post-traitement dans le but d'éliminer les petits éléments isolés. Pour cela, les éléments connexes sont comptés, et en fonction de leur taille relative, un seuil est établi qui permet d'éliminer les éléments dont la taille est éloignée des deux plus gros éléments.

Évaluation

Afin d'évaluer l'ensemble des axes explorés dans les expériences présentées dans ce chapitre, plusieurs types d'évaluation sont mis en place. D'une part, l'évaluation géométrique (métriques aussi utilisées dans l'évaluation des expériences du Chapitre 3) et d'autre part une évaluation physiologique. L'objectif est de rendre compte à la fois de la qualité des masques de segmentation du point de vue du traitement d'images, mais aussi du point de vue médical en vue de leur possible utilisation en clinique. Les diverses bases de données à disposition offrent également une variété de possibilités pour l'évaluation. En effet, la base de données Lyon-1 permet d'évaluer les performances des modèles sur le jeu de données équilibré de l'entraînement⁴³, tandis que la base de données CT4-ARDS2 permet

42: Qui n'évaluent pas cet hyper-paramètres.

Ressources de calcul

Les entraînements sont lancés sur le cluster du laboratoire CREATIS. Il s'agit d'un groupe de ressources informatiques distribuées dédié au calcul informatique. Il est composé de 23 machines, 532 CPU avec 14GB en moyenne par et 15 GPU de taille 8 GB à 48 GB.

43: En termes de sévérité de SDRA représenté.

d'évaluer à la fois la robustesse⁴⁴ et la généralisation⁴⁵. La validation croisée en k sous-ensemble (avec $k=5$) est utilisée pour pouvoir évaluer les performances des modèles sur la base d'entraînement. Lorsque les modèles sont évalués sur les bases de données de test Lyon-2 et Grenoble-2, les résultats présentés sont la moyenne des performances de chaque modèle entraîné sur chacun des 5 sous-ensembles.

Géométrique Du point de vue géométrique, les métriques présentées dans l'état de l'art servent de référence. La combinaison d'une métrique de superposition et de métriques de distances permet de quantifier équitablement la qualité de la segmentation produite par rapport à la référence manuelle. De ce fait, le score de Dice, la distance ASSD et la distance de Hausdorff (DH-95) sont à nouveau utilisées.

Quels que soient le modèle évalué et la résolution utilisée à l'entraînement, les résultats sont calculés sur les données à leur résolution d'origine. Cela permet de comparer les performances de tous les modèles tels qu'ils seraient implémentés en situation réelle (clinique). Pour cela, l'inférence est réalisée à la résolution d'entraînement, seul le calcul des métriques se fait à la résolution d'origine après post-traitement.

Enfin, une autre information qui mélange à la fois l'information géométrique et physiologique est de calculer la superposition des masques de segmentation selon le niveau d'aération associé dans l'image CT. Pour cela, les seuils d'aération en unité Hounsfield⁴⁶ sont utilisés dans le calcul du score de Dice. Ce qui mène à l'utilisation de fonction $Dice(\alpha)$, avec α le niveau d'aération pris parmi les quatre compartiments d'aération. Quatre nouvelles valeurs de Dice sont alors à considérer qui indiquent plus précisément les performances des modèles selon le niveau de difficulté⁴⁷ de la zone à segmenter. Le calcul des volumes relatifs de chaque compartiment d'aération segmenté est enfin utilisé afin de mettre en perspective les performances géométriques présentées précédemment.

Physiologique La cohérence spatiale apportée par l'utilisation d'un modèle 3D permet d'évaluer certains paramètres physiologiques des poumons. En effet, l'un d'eux, la masse des poumons, est utilisée par les cliniciens pour évaluer la qualité des segmentations manuelles sur plusieurs images d'un même patient. En tenant compte des circulations sanguines qui ont lieu entre l'inspiration et l'expiration, le seuil de 5% de différence est généralement fixé entre deux images du même patient à ces deux phases respiratoires. Ce seuil est souvent exprimé en fonction de la masse totale des poumons comme suit :

$$\text{Cohérence de masse} = \frac{m^{exp} - m^{ins}}{\hat{M}}, \quad (4.8)$$

avec m^{exp} et m^{ins} respectivement les masses des poumons en fin d'expiration (exp) et en fin d'inspiration (ins), et \hat{M} la masse moyenne des poumons. Une autre information à laquelle les cliniciens tentent d'accéder par la segmentation des poumons entre la fin d'inspiration et la fin d'expiration est le recrutement⁴⁸. Cette métrique permet d'évaluer la recrutabilité⁴⁹ d'un patient et ainsi ajuster ses paramètres ventilatoires. Le recrutement

44: Car composée d'images plus compliquées à segmenter (i.e. patients sous ECMO).

45: Car comportant des images d'un autre centre.

46: Définis dans le tableau 4.4

47: Les régions peu et non aérées sont les plus difficiles à segmenter étant les zones les moins contrastées avec les tissus environnants.

48: Défini dans le Chapitre 1, sous-section Mesure de l'aération.

49: Voir sous-section Mesure de l'aération dans le chapitre Contexte médical et enjeu

courant est calculé [Ric+22] comme suit :

$$R = \frac{m_{NA}^{exp} - m_{NA}^{ins}}{\hat{M}} \quad (4.9)$$

avec m_{NA}^{exp} et m_{NA}^{ins} respectivement les masses des zones non aérées (NA) en fin d'expiration (exp) et en fin d'inspiration (ins), et \hat{M} la masse moyenne des poumons. Le biais obtenu entre la prédiction et la mesure clinique, pour les deux métriques présentées, est alors calculé. Celui-ci sert de témoin de la qualité de la prédiction, et plus il se rapproche de zéro plus la cohérence entre deux masques d'un même patient est élevée. En particulier, le biais du recrutement est calculé comme ceci :

$$\text{Biais}(R) = R_{prediction} - R_{reference} \quad (4.10)$$

et est représenté sous forme de tracé de Bland-Altman [Alt+83] pour l'analyse des résultats.

4.4 Résultats

Expériences sur l'initialisation

Les trois modèles comparés dans cette expérience visent à évaluer l'apport du transfert d'apprentissage en initialisation, sur une architecture 3D. Leurs résultats sont comparés d'un point de vue géométrique et physiologique tantôt sur la base de données d'entraînement (Lyon-1) et tantôt sur la base de données de test (Lyon-2).

Métriques géométriques La Figure 4.8 présente les performances évaluées en validation croisée ($k = 5$) avec les trois métriques Dice, ASSD et DH-95. Les trois modèles ont des performances très semblables, à la petite différence de leurs valeurs aberrantes. Le modèle D_a sert de référence dans cette expérience, puisqu'il est entraîné de façon standard en validation croisée sur la base de données Lyon-1. Il obtient un Dice médian de 0.960 ± 0.025 , une distance ASSD de $1.33 \pm 0.87mm$ et une distance DH-95 de $5.22 \pm 4.85mm$. Les résultats des modèles D_t et D_a^{200} sont très proches entre eux (et meilleurs que ceux du modèle D_a) avec respectivement : Dice= 0.963 ± 0.024 vs. $= 0.964 \pm 0.029$; ASSD= $1.12 \pm 0.81mm$ vs. ASSD= $1.15 \pm 0.84mm$; DH-95= $4.64 \pm 5.99mm$ vs. DH-95= $4.64 \pm 4.45mm$. Le modèle D_a^{200} génère toutefois le plus de valeurs aberrantes parmi les trois (lorsque les trois métriques sont prises en compte). Les résultats des modèles D_a^{200} et D_t représentent une sensible amélioration des performances du modèle de référence sur les métriques géométriques globales. Observer les performances géométriques plus en détails dans la Figure 4.9 : selon les caractéristiques (1) locales (Dice(α)) et (2) de quantité (volumes relatifs segmentés), permet de noter plus finement les différences entre les trois modèles. Le score de Dice sur chaque compartiment d'aération dans la Figure 4.9 (graphique de gauche) met en évidence les disparités de performances selon la zone à segmenter. Tout d'abord, le compartiment non aéré est unanimement celui pour lequel les performances sont les

Toutes les figures présentées dans cette partie illustrent des résultats obtenus en validation croisée sur la base de données d'entraînement Lyon-1. Les résultats calculés sur les bases de données Lyon-2 et Grenoble-2 sont présentés dans les tableaux.

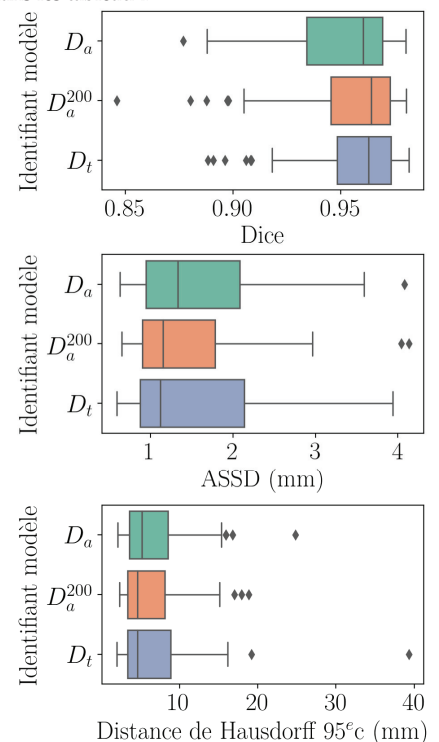


FIGURE 4.8 – Performances des modèles D_x évaluant le transfert d'apprentissage en validation croisée sur la base de données Lyon-1.

moins bonnes. C'est aussi celui pour lequel les prédictions de volume relatives varient le plus selon les modèles (graphique de droite de la Figure 4.9). Le compartiment peu aéré montre également une légère différence de performances entre les trois modèles.

Évaluées sur les jeux de données de test (Lyon-2 et Grenoble-2), les performances globales, présentées dans le Tableau 4.5, permettent de quantifier la capacité de généralisation de ces modèles. Les résultats avec les métriques géométriques sont légèrement inférieurs à ceux obtenus sur la base de données d'entraînement. Néanmoins, la tendance respective des trois modèles reste la même, leurs performances sont très proches entre elles. Le modèle D_t obtient les meilleures performances en termes de médiane, mais la distribution de ses résultats est souvent la plus large⁵⁰ aussi. Le modèle D_a , quant à lui, fournit les résultats avec la plus faible dispersion des valeurs.

50: Voir grand écart-type reporté dans le Tableau 4.5

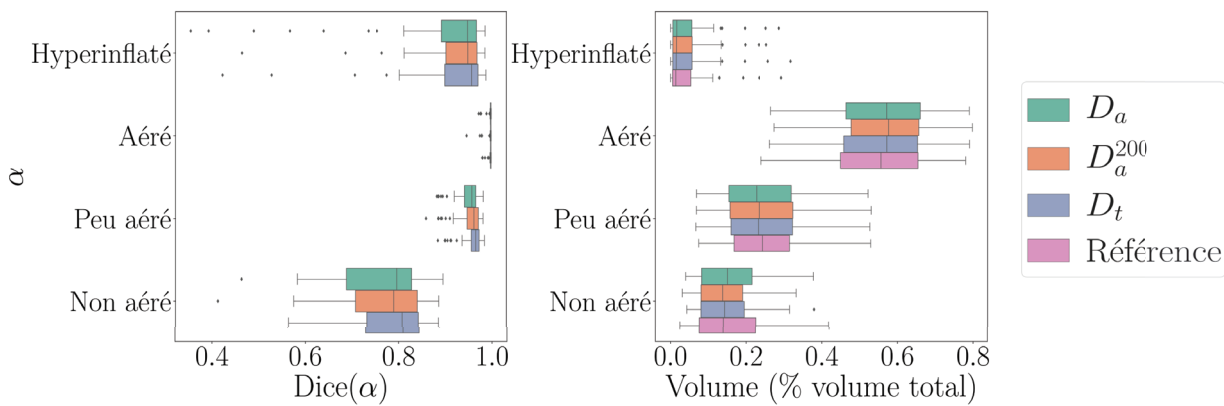


FIGURE 4.9 – Dice par compartiment d'aération segmentés et leur proportion dans le volume total pour les modèles $D.x$ sur la base de données Lyon-1 en validation croisée.

TABLEAU 4.5 – Performances globales des modèles D_x sur les bases de données de test Lyon-2 et Grenoble-2.

Base de données de test	Modèle	Dice	ASSD (mm)	DH-95 (mm)
Lyon-2	D_a	0.92 ± 0.09	2.2 ± 2.3	9.2 ± 13.6
	D_a^{200}	0.92 ± 0.10	2.2 ± 3.5	9.0 ± 19.1
	D_t	0.93 ± 0.11	1.8 ± 3.6	7.5 ± 20.4
Grenoble-2	D_a	0.93 ± 0.04	2.5 ± 1.5	9.9 ± 8.8
	D_a^{200}	0.93 ± 0.05	2.5 ± 1.8	10.7 ± 10.9
	D_t	0.94 ± 0.04	2.4 ± 1.6	9.7 ± 10.4

Lorsque sont observés les résultats sur les compartiments les plus compliqués à segmenter⁵¹ (dans le Tableau 4.6), la tendance reste la même que sur la base de données d'entraînement. Le modèle D_t est celui qui obtient les meilleurs résultats, bien que ceux-ci soient très proches des résultats des deux autres modèles.

51: Compartiments PA et NA.

Métriques physiologiques Les valeurs physiologiques calculées sur les masques prédits sont comparées à celles calculées sur les masques de référence. Le critère évalué grâce à elles est la cohérence de prédictions sur différentes images⁵² d'un même patient. Le biais calculé sur le recrutement est présenté pour les trois modèles

52: Dans le cadre des données utilisées : acquises à différentes conditions de ventilation.

TABLEAU 4.6 – Performances sur les compartiments peu aéré et non aéré des modèles D_x sur les bases de données de test Lyon-2 et Grenoble-2.

Base de données de test	Modèle	Dice (Peu aéré)	Dice (Non aéré)
Lyon-2	D_a	0.94 ± 0.03	0.78 ± 0.11
	D_a^{200}	0.95 ± 0.03	0.77 ± 0.15
	D_t	0.95 ± 0.03	0.80 ± 0.16
Grenoble-2	D_a	0.92 ± 0.04	0.75 ± 0.08
	D_a^{200}	0.92 ± 0.04	0.73 ± 0.9
	D_t	0.92 ± 0.04	0.75 ± 0.08

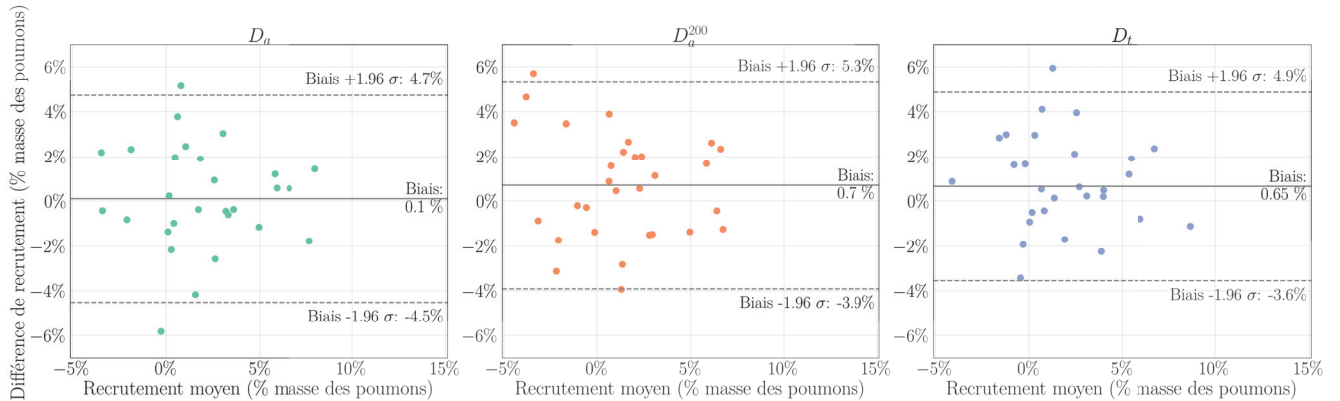


FIGURE 4.10 – Tracé Bland-Altman du biais de recrutement en validation croisée sur la base de données Lyon-1.

évalués sur la base de données Lyon-1 dans la Figure 4.10. Tout comme pour les métriques précédentes, la différence de performances entre les trois modèles est légère. Les prédictions de recrutement les plus précises et en accord avec la référence proviennent du modèle D_a , avec un biais (différence moyenne) de zéro et des limites d'agrément entre -4.5% et 4.7%. Le modèle D_t , bien que fournissant une différence moyenne inférieure, garde des limites d'agrément similaires entre -3.6% et 4.9%.

En ce qui concerne le biais et les limites d'agrément calculés sur la cohérence de masse prédite entre deux segmentations du même patient⁵³, elle est respectivement de $-0.004[-0.049; 0.041]$, $0.000[-0.050; 0.050]$ et $-0.003[-0.058; 0.53]$ pour les modèles D_a , D_a^{200} et D_t . Les résultats de ces

53: Voir équation (4.8).

TABLEAU 4.7 – Biais de recrutement (R) et de cohérence de masse (C) des modèles D_x sur les bases de données de test Lyon-2 et Grenoble-2.

Données de test	Modèle	Biais (R) en %	Biais (C) en %
Lyon-2	D_a	$-0.4[-10.7; 9.9]$	$0.1[-7.7; 7.9]$
	D_a^{200}	$-1.4[-14.8; 12.0]$	$1.0[-9.4; 11.5]$
	D_t	$2.2[-13.1; 8.6]$	$0.1[-8.3; 10.3]$
Grenoble-2	D_a	$0.7[-3.9; 5.2]$	$-0.4[-3.2; 2.2]$
	D_a^{200}	$-0.1[-3.7; 3.6]$	$-0.1[-2.8; 2.7]$
	D_t	$0.1[-3.1; 3.3]$	$-0.3[-2.8; 2.2]$

deux métriques calculés sur les données de test sont présentés dans le tableau Tableau 4.7. Ici les résultats diffèrent selon la base de données de test utilisée. En effet, le modèle le plus cohérent sur la base de données Lyon-2 est le modèle D_a tandis que sur la base de données Grenoble-2, il

s'agit du modèle D_a^{200} .

Le modèle D_a a été le plus rapide à entraîner, avec une durée moyenne d'entraînement de 43h. Les entraînements des modèles D_a^{200} et D_t ont respectivement duré environ 86h et 133h. Le temps d'inférence pour tous ces modèles est en moyenne de 11 s sur CPU⁵⁴ et 2 s sur GPU⁵⁵ pour une image en entrée de taille $512 \times 512 \times 315$. Le temps total de traitement, comprenant le pré-traitement et le post-traitement, est quant à lui, d'environ de 27 s et occupe en espace mémoire 8.5 Go. Les courbes d'entraînements pour tous ces modèles sont similaires à celles obtenues dans le chapitre précédent, décroissent de manière stable et ne montrent pas de sur-apprentissage.

Expériences sur la quantité d'information

Les modèles évalués dans cette expérience ont pour objectif d'identifier l'impact de la quantité d'information utilisée à chaque itération de l'entraînement. Notés E_r^β avec r la résolution spatiale des images utilisées à l'entraînement et β la taille des lots d'images, ils sont au nombre de quinze. Les modèles sont tous évalués en validation croisée ($k = 5$) sur leur base de données d'entraînement (Lyon-1). Les résultats des différentes performances sont détaillés ci-dessous.

Métriques géométriques Les performances géométriques globales des modèles E_r^β évalués en validation croisée sur la base de données Lyon-1 sont présentées dans la Figure 4.11. Trois tendances se dégagent des graphiques, quelle que soit la métrique observée. La première concerne les modèles allant de E_1^1 à E_2^4 (avec r et β croissant dans E_r^β) pour lesquels il est difficile d'établir une différence notable de performances. En effet, ces modèles obtiennent des résultats autour de 0.95 pour le score de Dice, 1.5 mm pour la distance ASSD et 5.5 mm pour la distance DH-95. La deuxième tendance concerne les deux derniers modèles utilisant la résolution $r = 2$ (E_2^8 et E_2^{16}) pour lesquels les performances sont quelque peu dégradées. Leurs résultats moyens sont de d'environ 0.90 pour le score de Dice, 3.3 mm pour la métrique ASSD et 12.5 mm pour la métrique DH-95. Enfin, la dernière tendance concerne tous les modèles utilisant la résolution spatiale de 3 mm à l'entraînement pour lesquels les performances changent d'ordre de grandeur (fort dégradation). La précision de ces modèles est également basse comme en témoigne l'étendue des boîtes englobantes qui représentent leur résultat (en violet dans la Figure 4.11).

Le détail des performances par compartiment d'aération en validation croisée sur la base de données Lyon-1 est présenté dans la Figure 4.12. À nouveau, plusieurs tendances apparaissent de ces résultats. Une légère différence de performances apparait entre les modèles E_1^β (bleu dans la figure) et E_2^β ($\beta \in \{1, 2, 2p, 4\}$) en vert dans la figure) sur le compartiment peu aéré en faveur de l'utilisation d'une résolution conservant les détails. Les différences de volumes relatifs par compartiment d'aération illustrent les erreurs de segmentation à la fois sur la quantité de voxels segmentés, mais aussi sur la position de ceux-ci. C'est notamment le cas des modèles E_3^β (violet) qui sous-estiment les parties aérées et non aérées, mais sur-estiment les parties hyperinflatées. Les modèles entraînés avec $r = 3$

54: Avec un processeur Intel®Core™ i5-8500 CPU @ 3.00GHz x6 et 32Go de mémoire vive.
55: Avec une carte graphique NVIDIA Tesla M60 de 8 GB.

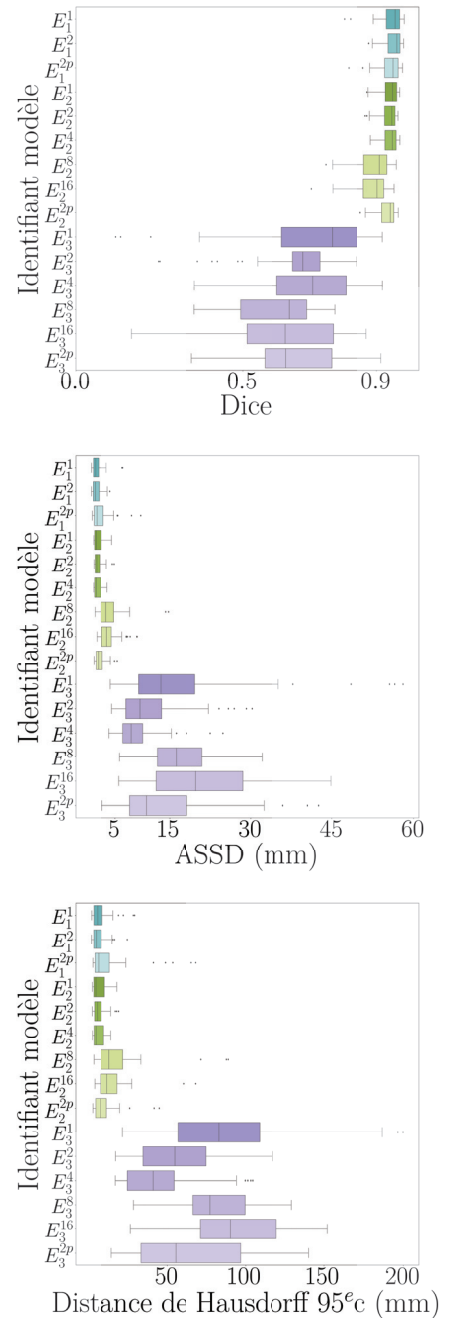


FIGURE 4.11 – Performances des modèles E_r^β en validation croisée sur la base de données Lyon-1. – Les résolutions d'entraînement 1 mm, 2 mm et 3 mm sont respectivement symbolisées par les couleurs bleu, vert et violet.

sont systématiquement très inférieurs en performances quel que soit le compartiment évalué.

La capacité de généralisation des modèles E_r^β est évaluée dans le Tableau 4.8 sur les métriques géométriques globales. Quelle que soit la base de données utilisée et le modèle évalué, les résultats sont inférieurs à ceux obtenus sur la base de données d'entraînement. Néanmoins, tout comme sur la base de données d'entraînement, les modèles E_1^β à E_2^4 obtiennent des performances très similaires entre eux et les meilleures parmi tous. Tous atteignent des scores médians d'au moins 0.9 pour le score de Dice, 3.0 mm pour la distance ASSD et 12 mm pour la distance DH-95. La différence de performance entre les deux bases de données (Lyon-2 et Grenoble-2) utilisées est peu visible sur les modèles qui obtiennent de bons scores, mais s'accroît au contraire pour les modèles qui échouent. C'est le cas des modèles E_2^β (avec $\beta \in \{8, 16\}$) et de tous les modèles E_3^β . Bien que leurs performances soient très faibles, elles sont systématiquement meilleures sur la base de données Grenoble-2.

Métriques physiologiques La cohérence des modèles pour la segmentation de deux images d'un même patient est évaluée au regard de la mesure du recrutement. Ces résultats sont présentés dans la Figure 4.13 pour la validation croisée sur la base de données Lyon-1. Globalement, il existe peu ou pas de biais entre le recrutement calculé sur les prédictions et celui calculé à partir des segmentations des cliniciens. En effet, tous les tests obtiennent une différence moyenne entre 0 et 0.04⁵⁶. L'intervalle de confiance varie davantage selon les modèles. Il est inférieur à 5% pour les modèles E_1^1, E_1^2, E_2^1 et E_2^4 , au-delà de 10% pour les modèles E_2^8, E_2^{16}, E_3^8 et E_3^{16} et entre 5% et 10% pour les autres modèles. La cohérence globale, calculée sur l'ensemble du masque de segmentation, permet davantage d'identifier les modèles les plus fiables. Le seuil⁵⁷ de 5% est respecté par les modèles E_1^2 et E_2^1 et dépassé pour tous les autres en validation croisée sur la base de données Lyon-1.

La durée d'entraînement des modèles était variable selon la taille de lot ou la résolution des images utilisées. Les modèles les plus chronophages ont été ceux qui utilisent la taille de données la plus grande, à savoir celles avec une résolution isotrope de 1 mm. Ces chiffres sont donnés à titre indicatif, étant donné que les entraînements n'ont pas tous été réalisés sur les mêmes ressources⁵⁸. De plus, les paquets (nouvelle version de cuda⁵⁹ notamment) servant à la distribution des calculs sur les cartes graphiques ont été mis à jour entre l'entraînement des modèles à la résolution 1 mm et les autres.

Les entraînements de tous les modèles ont montré une décroissance stable sans signe de sur-apprentissage d'après les courbes d'entraînement. Les temps d'inférence pour les modèles sont indépendants de la taille de lot d'entraînement, mais dépendent de la résolution utilisée à l'entraînement. Pour l'évaluer, un pré-traitement est appliqué pour redimensionner l'image de test⁶⁰ à la résolution d'entraînement. Il est ainsi de 11 s, 2 s et 1 s respectivement pour une résolution de 1 mm, 2 mm et 3 mm. Le temps total de traitement⁶¹ est, quant à lui, respectivement de 27 s, 13 s et 11 s.

56: Pour rappel le recrutement est calculé comme étant la différence des masses des compartiments non aérés entre la fin d'expiration et la fin d'inspiration. Il est normalisé par la masse des poumons (en fin d'expiration) résultante de la méthode utilisée. Les différences entre les recrutements de deux méthodes s'expriment alors aussi comme un ratio.

57: Voir définition dans la section Évaluation de ce chapitre.

58: Voir sous-partie Modèle, entraînement et inférence.

59: Librairie qui distribue les calculs sur carte graphique.

60: Mesures réalisées sur une image de taille originale $512 \times 512 \times 315$ et sur une machine équipée d'un processeur Intel®Core™ i5-8500 CPU @ 3.00GHz x6 et 32Go de mémoire vive.

61: De la lecture de l'image à l'enregistrement du masque de segmentation.

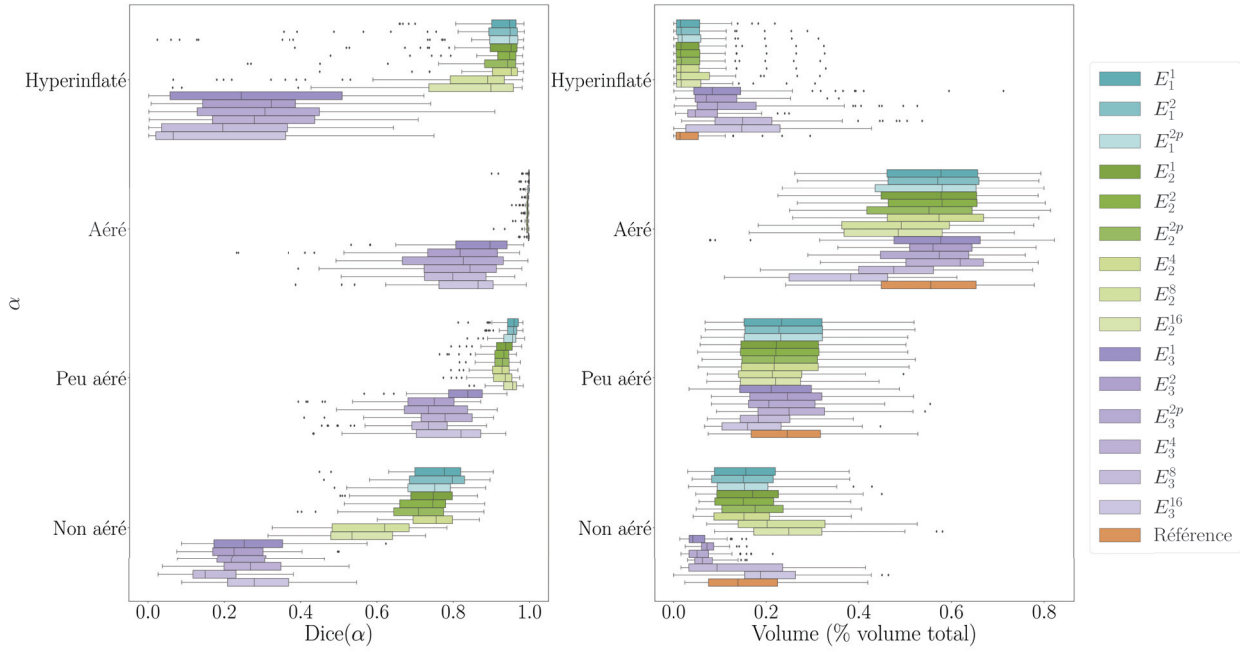


FIGURE 4.12 – Performances détaillées sur les compartiments d'aération des modèles E_r^β sur la base de données Lyon-1 en validation croisée.

TABLEAU 4.8 – Performances globales des modèles E_r^β sur les bases de données de test Lyon-2 et Grenoble-2.

Modèle	Dice		ASSD (mm)		DH-95 (mm)	
	Lyon-2	Grenoble-2	Lyon-2	Grenoble-2	Lyon-2	Grenoble-2
E_1^1	0.92 ± 0.10	0.93 ± 0.05	2.3 ± 3.2	2.5 ± 1.6	9.2 ± 17	10 ± 9.4
E_1^2	0.92 ± 0.09	0.93 ± 0.04	2.2 ± 2.3	2.5 ± 1.5	9.2 ± 14	9.9 ± 8.8
E_1^{2p}	0.91 ± 0.11	0.92 ± 0.05	2.7 ± 4.4	2.9 ± 2.4	11 ± 26	12 ± 18
E_2^1	0.92 ± 0.09	0.92 ± 0.04	2.3 ± 3.9	2.7 ± 1.6	9.1 ± 24	12 ± 10
E_2^2	0.91 ± 0.10	0.92 ± 0.04	2.4 ± 3.2	2.8 ± 1.4	10 ± 18	11 ± 7.7
E_2^{2p}	0.90 ± 0.09	0.91 ± 0.06	2.8 ± 3.2	3.0 ± 3.6	11 ± 21	12 ± 18
E_2^4	0.91 ± 0.09	0.92 ± 0.04	2.5 ± 2.8	2.7 ± 1.5	10 ± 17	11 ± 8.6
E_2^8	0.64 ± 0.24	0.89 ± 0.08	19 ± 14	3.9 ± 2.85	90 ± 10	15 ± 20
E_2^{16}	0.86 ± 0.10	0.88 ± 0.06	4.2 ± 3.0	3.9 ± 2.6	14 ± 20	13 ± 14
E_3^1	0.61 ± 0.22	0.70 ± 0.16	20 ± 17	13 ± 10	100 ± 48	71 ± 35
E_3^2	0.61 ± 0.15	0.68 ± 0.12	14 ± 10	11 ± 4.6	77 ± 33	60 ± 21
E_3^{2p}	0.63 ± 0.15	0.70 ± 0.10	15 ± 13	12 ± 5.3	77 ± 41	60 ± 25
E_3^4	0.62 ± 0.15	0.69 ± 0.11	11 ± 9.5	8.3 ± 3.9	57 ± 35	33 ± 21
E_3^8	0.48 ± 0.12	0.64 ± 0.10	17 ± 13	15 ± 5.3	81 ± 29	75 ± 17
E_3^{16}	0.52 ± 0.20	0.65 ± 0.12	24 ± 11	18 ± 3.7	106 ± 34	81 ± 23

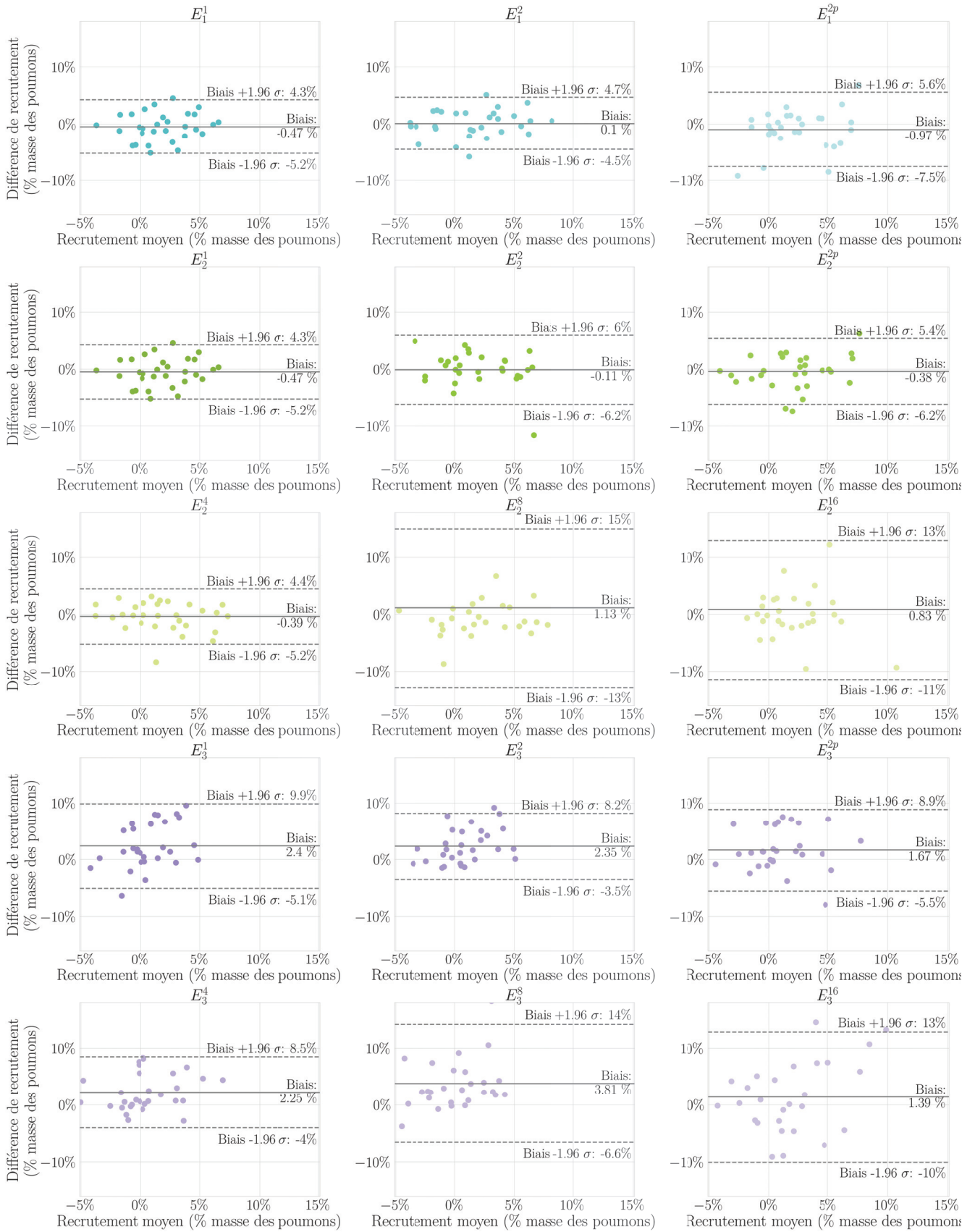


FIGURE 4.13 – Tracés de Bland-Altman illustrant le biais entre les modèles E_i^p et la référence pour le calcul du recrutement, en validation croisée sur la base de données Lyon-1. σ l'écart-type des différences.

TABLEAU 4.9 – Durées d'entraînement moyennes par modèle en heure. – La durée correspond à la moyenne des 5 modèles entraînés parallèlement pour la validation croisée.

Résolution (mm)	Taille de lot					
	1	2	2 (1 patient)	4	8	16
1	85h	43h	43h			
2	0.5h	1.3h	1h	1.25h	0.5h	0.5h
3	0.5h	0.15h	0.25h	0.3h	0.25h	0.25h

Expérience sur l'apport d'information physiologique

Les modèles évalués dans cette nouvelle expérience avaient pour objectifs d'évaluer de nouvelles formes d'intégration de l'information durant l'apprentissage. En effet, 2 types de nouvelles fonctions de perte ont été utilisées à l'entraînement, l'une basée sur les compartiments d'aération et l'autre sur le calcul de la masse pulmonaire. Le modèle qui utilise la fonction de perte de la première catégorie est noté F_α . Les modèles utilisant la masse pulmonaire associé à la perte sur le Dice sont notés F_{masseX} avec $X \in \{1, 2, 3\}$ selon le mélange utilisé des termes au cours de l'entraînement.

Métriques géométriques Les performances globales des modèles F en validation croisée sur la base de données Lyon-1 sont présentées dans la Figure 4.14. Les trois modèles F_α , F_{masse2} et F_{masse3} ont des performances du même ordre de grandeur sur les trois métriques, ce qui n'est pas le cas du modèle F_{masse1} . En effet, ce dernier obtient des performances largement inférieures aux autres modèles, avec un Dice moyen de 0.24, et des distances de 45 mm (ASSD) et 161 mm (DH-95). Il s'agit, pour rappel, du modèle qui utilise le terme de masse et de superposition à poids égal dans la fonction de perte.

En comparaison, les autres modèles fournissent des résultats bien supérieurs. Le modèle F_α obtient les meilleurs résultats parmi eux avec un score de Dice médian de 0.957 ± 0.033 , une distance ASSD médiane de 1.38 ± 3.88 mm et une distance DH-95 médiane de 5.99 ± 27.7 mm. La Figure 4.14 montre que les deux autres modèles (F_{masse2} en orange et F_{masse3} en jaune) sont légèrement moins performants, avec des boîtes englobantes plus larges.

Les résultats détaillés par compartiment d'aération permettent de mettre en évidence les zones qui impactent positivement ou négativement les valeurs du Dice. Tout d'abord, il est important de noter que les volumes relatifs estimés par le modèle F_{masse1} sont très en dessous de ceux estimés par les autres modèles ou de ceux de référence. Ceci explique les faibles performances de ce modèle sur les métriques globales. En effet, n'ayant pas prédit de volumes dont au moins la moitié contient des voxels d'intensité cible (les intensités Hounsfield des poumons définis par compartiment dans le tableau 4.4), les performances mesurées sont très faibles.

Les autres modèles, au contraire, obtiennent des volumes relatifs très proches entre eux et de ceux de référence. Des disparités s'observent sur le Dice par compartiment, notamment sur le compartiment non aéré. Ici encore, le modèle F_α fournit les meilleurs résultats, formant un écart plus important avec les modèles qui le suivent en termes de performances.

Les résultats des modèles de cette expérience sur les deux bases de données de test sont présentés dans le Tableau 4.10. Les performances de

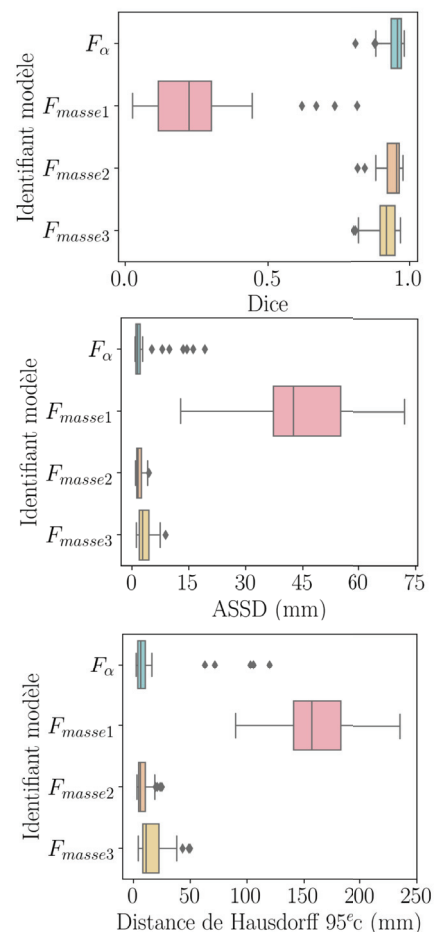


FIGURE 4.14 – Performances globales des modèles F évalués en validation croisée sur la base de données Lyon-1 avec les métriques géométriques.

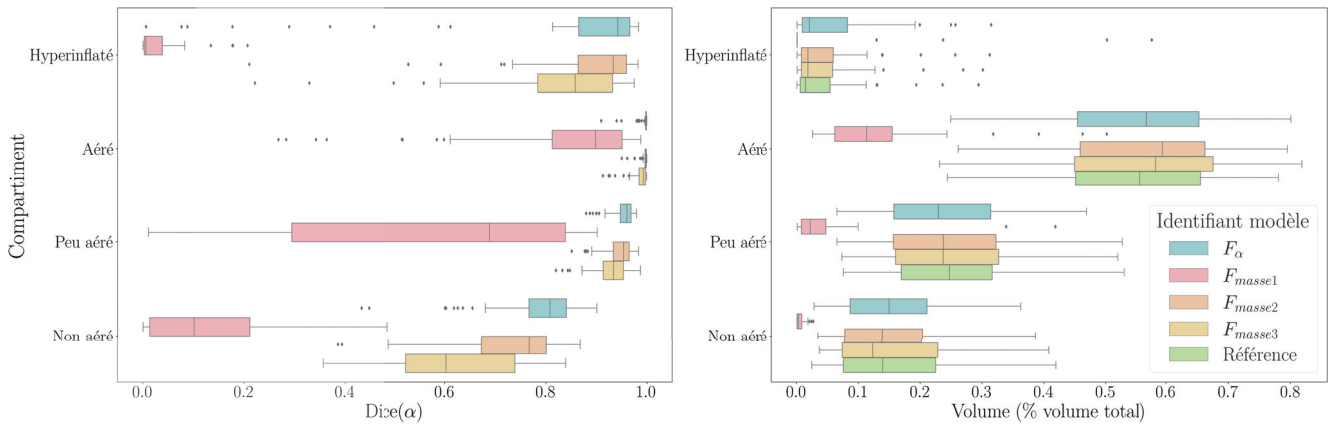


FIGURE 4.15 – Performances détaillées sur les compartiments d’aération des modèles F en validation croisée sur la base de données Lyon-1.

TABLEAU 4.10 – Performances globales des modèles F_x sur les bases de données de test Lyon-2 et Grenoble-2.

Base de données de test	Modèle	Dice	ASSD (mm)	DH-95 (mm)
Lyon-2	F_α	0.92 ± 0.10	2.7 ± 5.8	9.9 ± 32.6
	F_{masse1}	0.11 ± 0.09	60 ± 15	194.8 ± 38.4
	F_{masse2}	0.91 ± 0.11	3.0 ± 3.5	12.7 ± 19.3
	F_{masse3}	0.87 ± 0.12	4.4 ± 4.9	18.8 ± 26.9
Grenoble-2	F_α	0.93 ± 0.05	2.6 ± 2.6	10.9 ± 16.7
	F_{masse1}	0.17 ± 0.09	46 ± 7.1	154.3 ± 15.8
	F_{masse2}	0.93 ± 0.05	2.7 ± 2.2	11.2 ± 13.6
	F_{masse3}	0.90 ± 0.06	3.4 ± 2.4	14.8 ± 13.2

tous les modèles sont inférieures à celles observées en validation croisée sur les données d’entraînement. Le modèle qui généralise le mieux est le modèle F_α , bien que l’écart-type de ses résultats soit parfois très grand. Avec des résultats sensiblement inférieurs en médiane, mais davantage de précision, le modèle F_{masse2} obtient des performances presque aussi bonnes que le modèle F_α .

Peu de disparités sont observées entre les performances calculées sur la base de données Lyon-2 et Grenoble-2.

Métriques physiologiques La question qui se pose est d’évaluer si l’apport d’information physiologique durant l’apprentissage a permis de gagner en robustesse sur des métriques qui y sont liées. Les graphiques de Bland-Altman en Figure 4.16 illustrent pour chaque méthode le biais entre le recrutement prédit et celui calculé avec les segmentations de référence. Les trois modèles F_α , F_{masse2} et F_{masse3} ont tous des différences moyennes de recrutement très faibles, mais se différencient entre eux par leur intervalle de confiance. En effet, seuls les modèles F_α et F_{masse2} obtiennent un intervalle de confiance inférieur à 5%. Le modèle F_{masse3} dépasse légèrement ce seuil avec un intervalle de confiance de $[-7.3\%; 6.9\%]$.

Les observations sont identiques sur la mesure du biais pour la cohérence de masse (entre l’expiration et l’inspiration). Les modèles F_α et F_{masse2} obtiennent ainsi respectivement un biais de -0.4 $[-4.6\%; 3.9\%]$ et -0.5 $[-5.1\%; 4.2\%]$.

Enfin, la capacité de généralisation est évaluée (voir Tableau 4.11), en

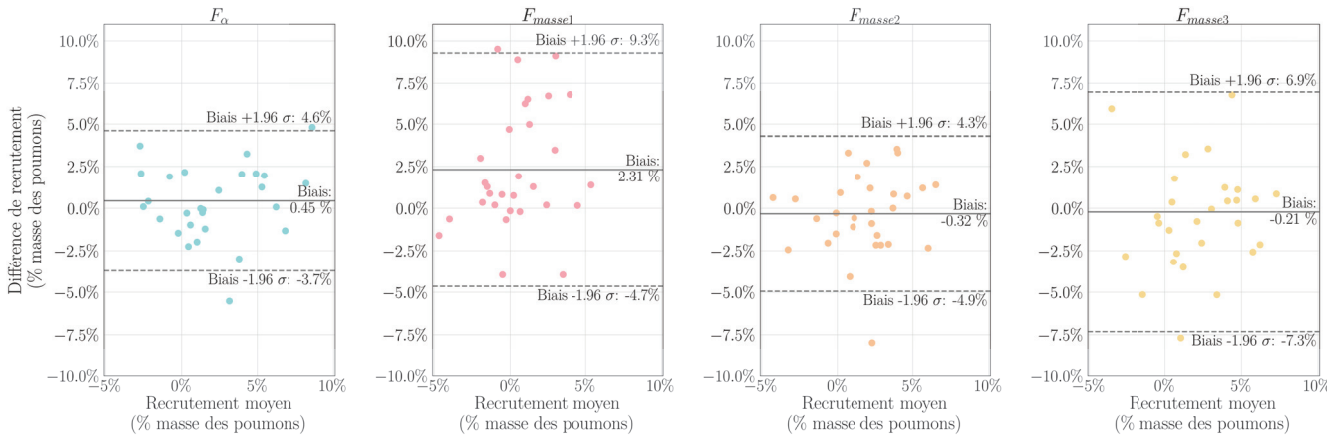


FIGURE 4.16 – Tracé Bland-Altman du biais de recrutement en validation croisée sur la base de données Lyon-1.

calculant le biais de recrutement et de cohérence de masse sur les bases de données de test. La première observation possible est que les performances diffèrent selon la base de données utilisée. En effet, sur la base de données Lyon-2, c'est le modèle F_{masse2} qui semble le plus cohérent avec la référence au vu de la faible moyenne des différences. Malgré cela, tous les modèles obtiennent des intervalles de confiance au-dessus des 5% acceptables. Lorsque la base de données Grenoble-2 est utilisée, les conclusions sont toutes autres. Trois des quatre modèles ont un intervalle de confiance inférieur à 5% et le modèle avec la plus faible moyenne des différences est le modèle F_α .

TABLEAU 4.11 – Biais sur le recrutement (R) et la cohérence de masse (C) des modèles F_x sur les bases de données de test Lyon-2 et Grenoble-2.

Données de test	Modèle	Biais (R) en %	Biais (C) en %
Lyon-2	F_α	-1.7[-11.3; 7.8]	0.7[-7.3; 8.7]
	F_{masse1}	-2.7[-19.7; 14.3]	5.4[-9.0; 19.9]
	F_{masse2}	-0.6[-9.9; 8.8]	0.1[-8.4; 8.7]
	F_{masse3}	-0.7[-10.0; 8.7]	-0.0[-7.8; 7.7]
Grenoble-2	F_α	0.3[-3.4; 4.0]	-0.2[-2.5; 2.0]
	F_{masse1}	-4.0[-9.3; 1.4]	11.1[-10.6; 32.8]
	F_{masse2}	0.8[-2.1; 3.8]	0.0[-1.6; 1.7]
	F_{masse3}	0.3[-2.2; 2.8]	0.7[-1.3; 2.7]

L'entraînement des modèles F a duré en moyenne 44 h, pour chaque sous-ensemble utilisé. Les courbes d'entraînement de tous les modèles montrent une décroissance globale (voir Figure 4.17 à Figure 4.20). Dans l'entraînement utilisant le Dice sur le compartiment non aéré, ce dernier commence l'entraînement avec de meilleures valeurs que le Dice global. Ensuite, les deux termes de Dice croissent, le terme global dépassant celui sur le compartiment non aéré, jusqu'à atteindre un palier autour de 0.96 pour le premier et 0.90 pour le second. Un léger saut dans les valeurs de la perte est observé pour les modèles où le terme de masse \mathcal{L}_{masse} est ajouté en cours d'entraînement (voir par exemple la Figure 4.19). C'est d'ailleurs le terme prédominant au début de l'entraînement du modèle F_{masse1} et qui décroît ensuite très rapidement (voir Figure 4.18). La perte calculée sur le sous-groupe de validation est un peu moins stable sur l'entraînement du modèle F_{masse3} que sur les autres modèles F_{masse} (voir Figure 4.20).

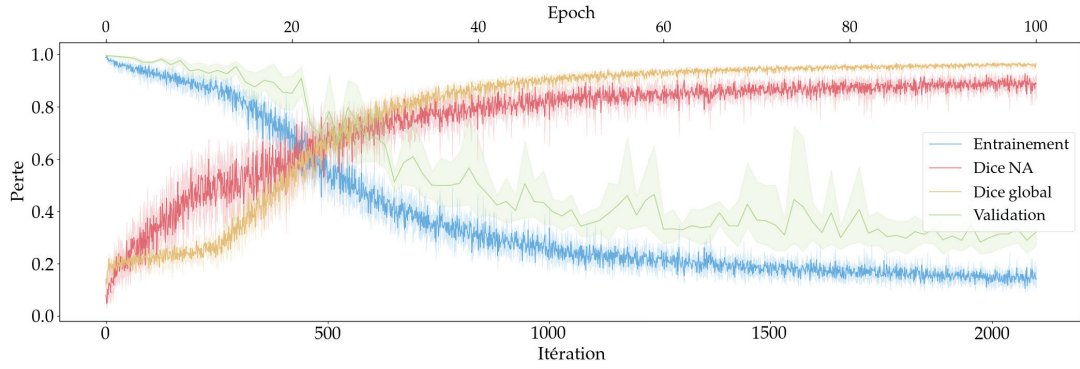


FIGURE 4.17 – Courbes d’entraînement du modèle F_α et tracé des performances du Dice global et du Dice NA sur les 5 sous-ensembles de la base de données Lyon-1. – La fonction de perte utilisée utilise les deux termes de Dice comme défini dans l’équation Equation 4.2

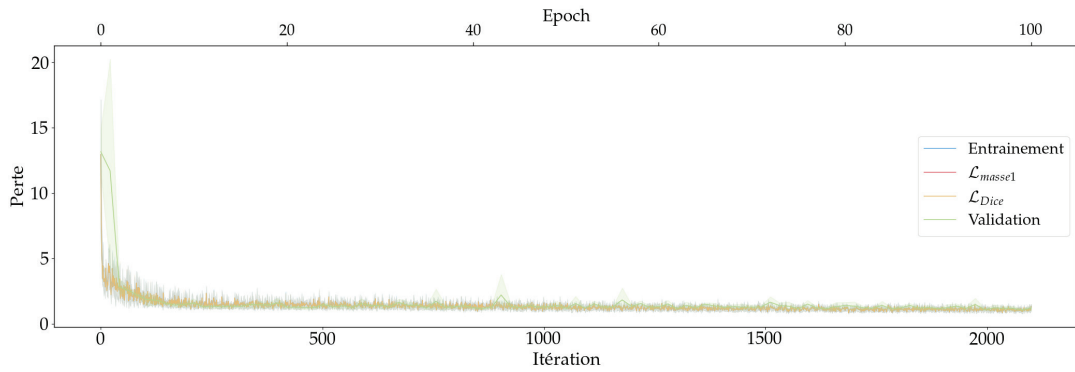


FIGURE 4.18 – Courbes d’entraînement du modèle F_{masse1} sur les 5 sous-ensembles de la base de données Lyon-1.

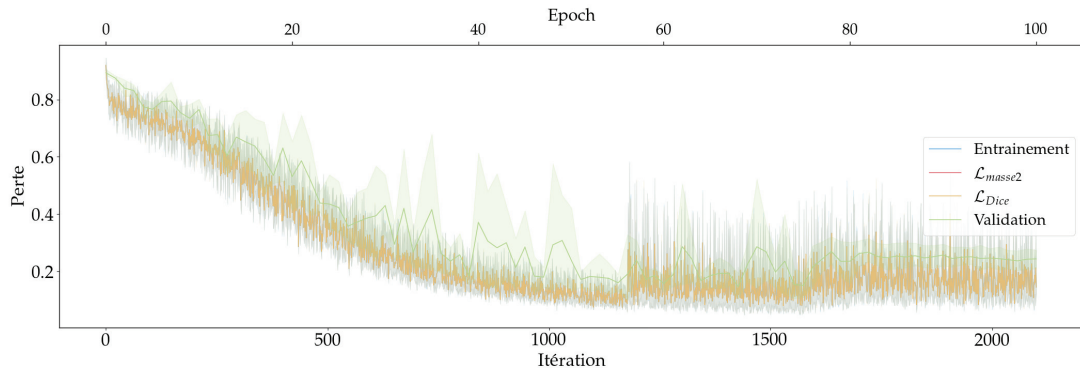


FIGURE 4.19 – Courbes d’entraînement du modèle F_{masse2} sur les 5 sous-ensembles de la base de données Lyon-1.

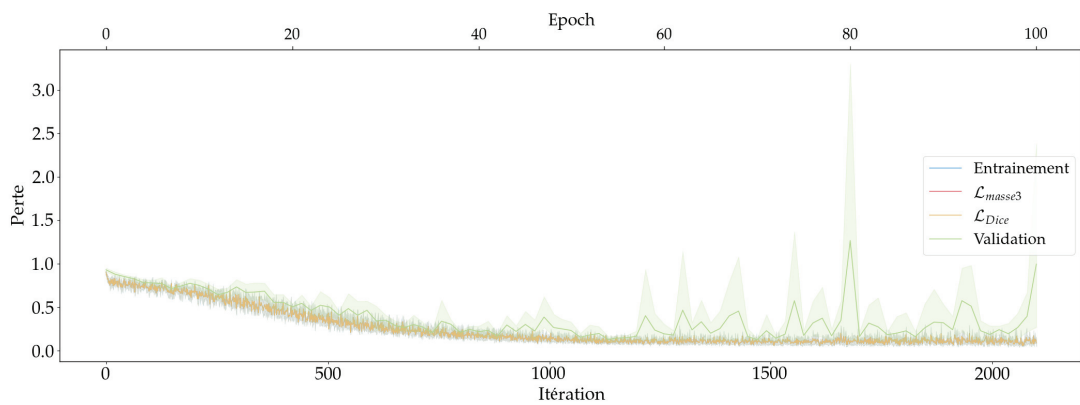


FIGURE 4.20 – Courbes d’entraînement du modèle F_{masse3} sur les 5 sous-ensembles de la base de données Lyon-1.

4.5 Discussion

Au cours des différentes expériences menées dans ce chapitre, l'objectif a été d'identifier quelle stratégie de manipulation des données apporte le plus de robustesse et de capacité de généralisation aux modèles. Trois approches ont été testées afin de répondre à cette question :

- commencer l'entraînement avec des poids entraînés sur une application proche⁶², en étant donc plus proche des poids optimaux qu'avec une initialisation aléatoire ;
- utiliser des images dégradées en résolution pour augmenter la quantité d'exemples contribuant à l'actualisation des poids, évaluant ainsi le rapport de force entre détails et diversité ;
- et enfin, intégrer des informations additionnelles à la simple superposition dans le but d'assister l'apprentissage dans les zones les plus compliquées à segmenter.

Les résultats de ces trois approches sont discutés dans les parties suivantes.

Expérience sur l'initialisation Les modèles entraînés dans cette expérience avaient pour but d'évaluer l'impact du transfert d'apprentissage dans la robustesse et la capacité de généralisation que peut atteindre un modèle d'apprentissage 3D.

Le modèle de référence D_a , entraîné en validation croisée sur la base de données Lyon-1 fournit les performances accessibles avec un entraînement standard. L'utilisation des deux bases de données animales avec modèle expérimental porcin de SDRa pour le pré-entraînement du modèle D_t , affiné sur la base de données Lyon-1 apporte une très sensible amélioration des performances⁶³. Celle-ci a cependant été obtenue en doublant le nombre d'epochs⁶⁴. De plus, le modèle D_a^{200} qui est entraîné avec initialisation aléatoire, sans transfert d'apprentissage et pendant 200 epochs obtient une amélioration comparable à celle de D_t . L'observation de ces résultats par compartiment d'aération⁶⁵ confirme unanimement la difficulté à segmenter les parties non aérées. Cependant, cette mise en forme des résultats ne permet pas de discriminer significativement un modèle par rapport à un autre.

Les métriques physiologiques⁶⁶ fournissent également des résultats très proches pour les trois modèles, sur la base de données Lyon-1. Cependant, le seuil d'acceptabilité étant fixé à 5%, les modèles D_a^{200} et D_t sont cette fois-ci pénalisés respectivement sur le biais de recrutement et le biais de cohérence de masses. Ainsi, le modèle qui garantit⁶⁷ les segmentations les plus cohérentes entre elles et mène à des mesures physiologiques proches de celles des cliniciens est celui de référence, le modèle de référence D_a .

La capacité de généralisation, ensuite évaluée, montre une dégradation des performances⁶⁸ sur les métriques géométriques globales. Les performances sont très sensiblement meilleures sur la base de données Grenoble-2. Pour rappel, cette base de données est acquise avec un imageur et des réglages différents de ceux de la base de données d'entraînement. De plus, elle contient des patients aux niveaux de sévérité de SDRa comparable à la base d'entraînement (Lyon-1), contrairement à la base Lyon-2 qui comporte des patients d'une gravité jamais vue à l'entraînement. Ces différences de performance selon que la base testée soit Grenoble-2 ou Lyon-2 indique d'un point de vue géométrique une robustesse plus sensible à la sévérité des patients qu'au changement d'imageur. Les données physiologiques

62: Tâche identique de segmentation, mais application et anatomie utilisées différentes.

63: Voir Figure 4.8.

64: 100 pour le pré-entraînement et 100 pour l'affinement dans le cas de D_t , contre 100 epochs au total pour D_a .

65: Voir Figure 4.9.

66: Voir Figure 4.10.

67: En validation croisée sur la base de données d'entraînement.

68: Qui doublent au moins, sur les métriques de distance.

calculées sur les bases de données de test sont cette fois-ci discriminantes, mais ne fournissent pas les mêmes conclusions selon la base de données testée. Aucun des modèles ne respecte le seuil de 5% pour l'intervalle de confiance du biais de recrutement ou de la cohérence des masses évalué sur la base de données Lyon-2. Pourtant, la conclusion est presque inverse sur la base de données Grenoble-2 sur laquelle seul le modèle D_a ne respecte pas le seuil sur l'intervalle de confiance du biais de recrutement.

Enfin, d'un point de vue plus pratique, l'évaluation qualitative des masques de segmentation permet d'identifier des points précis sur lesquels concentrer l'effort d'amélioration. La Figure 4.22 illustre pour chaque modèle sur la base de données Lyon-1, un exemple de la pire, de la meilleure et d'une segmentation moyenne. La colonne des pires exemples de segmentations prédits met en évidence deux types d'erreurs, celle de sous-segmentation des zones non aérées et celle de sous-segmentation des zones hyperinflatées⁶⁹. La première est observable dans les métriques géométriques présentées précédemment, mais la deuxième n'y apparaît pas clairement⁷⁰. En effet, ces zones qui sont très contrastées semblent simples à segmenter, car très différentes de leur contexte. Pourtant, les très faibles intensités de ces voxels sont souvent peu présentes⁷¹ dans les images. Intégrer une étape supplémentaire au post-traitement pourrait permettre de pallier ce type d'erreurs, en segmentant les voxels connexes au masque de segmentation dont l'intensité Hounsfield correspond au compartiment hyperinflaté. Mais cette solution introduirait potentiellement d'autres erreurs en segmentant les bronches ou même l'œsophage.

Les segmentations moyennes ou les meilleures obtenues avec ces modèles comportent aussi des erreurs, mais qui sont beaucoup plus minimes, et plus facilement corrigeables par intervention manuelle.

Les types d'erreurs observés sur les bases de données Lyon-2 et Grenoble-2 sont illustrées dans la Figure 4.23 avec le modèle D_t . Les pires segmentations surviennent sur des patients dont le SDRA est très sévère, pour lesquels les zones non aérées sont sous-segmentées. Sur les segmentations moyenne, les contours sont sous-segmentés comme le montre l'exemple du patient de la base de données Grenoble-2. Enfin, la meilleure segmentation observée sur un patient de la base de données Lyon-2 présente ici des erreurs de sur-segmentation. Ces dernières se produisent au niveau du médiastin, sur des voxels aux intensités Hounsfield qui sont attribuées au compartiment non aéré dans les poumons, mais correspondent ici à d'autres tissus. Ces erreurs témoignent plutôt d'une tendance à la sous-segmentation, avec une sur-segmentation qui est contenue dans la cage thoracique.

Pour conclure, les modèles entraînés dans cette expérience n'ont pas démontré⁷² la valeur ajoutée d'utiliser un pré-entraînement sur la segmentation de modèles expérimentaux de SDRA. Ce résultat est contraire aux conclusions amenées par la littérature, notamment appliqué à la segmentation des poumons dans le cadre du SDRA [Her+21; Mai+22]. Ceci pourrait s'expliquer par la différence de données utilisées pour la spécialisation du modèle. Les performances atteintes avec les modèles 3D permettent néanmoins d'obtenir des masques de segmentation au contour plus lisses que dans l'expérience 2D et dans un temps plus court⁷³.

Expérience sur la quantité d'information L'objectif des entraînements menés dans cette expérience étaient d'identifier laquelle des deux caracté-

69: Voir l'image quantifiée en Figure 4.21.

70: Ces zones étant le plus souvent très petites en proportions.

71: Voir le faible volume relatif que représente ce compartiment d'aération dans la Figure 4.9.

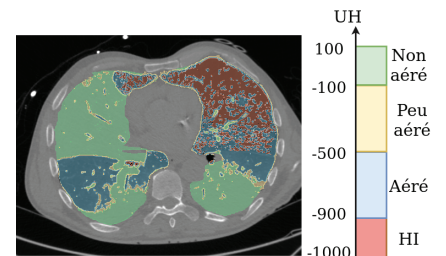


FIGURE 4.21 – Image CT quantifiée. – HI pour le compartiment hyperinflaté.

72: Tout comme les modèles entraînés en 2D, voir chapitre Approche 2D : gestion de données

73: A l'inférence. Cependant, les temps d'entraînement des modèles 2D restent un de leur plus gros atout.

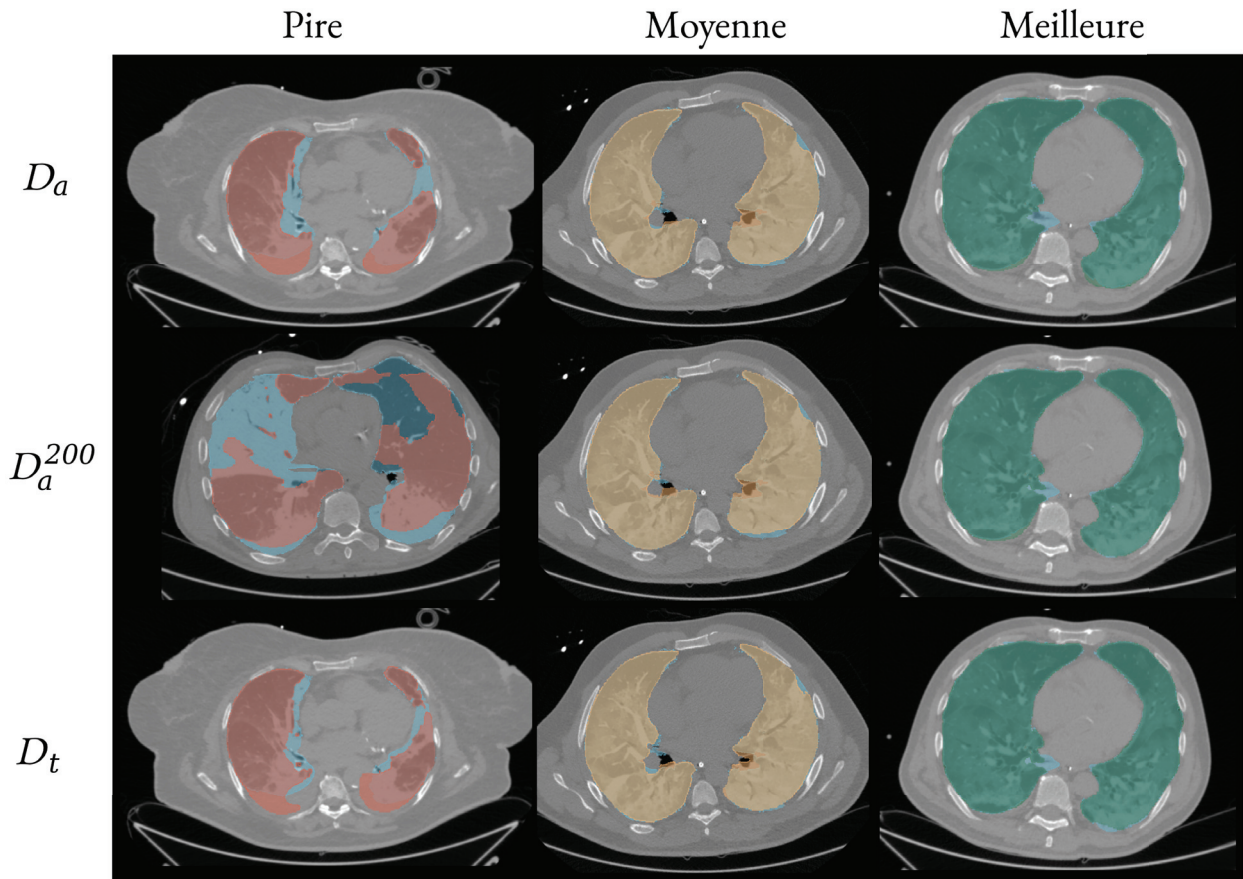


FIGURE 4.22 – Exemples des pires, moyennes et meilleures segmentations produites par les modèles D_x en validation croisée sur la base de données Lyon-1. – En bleu, les masques de segmentation de référence, en rouge les échecs de segmentations, en orange les segmentations moyennes et en vert les meilleures segmentations.

ristiques suivantes : (1) la finesse des détails, ou (2) la diversité des exemples considérés, avait le plus fort impact sur la qualité des segmentations. Les 15 modèles E_r^β évalués apportent des éléments de réponse à cette question. Leurs résultats sont discutés dans cette partie.

Tout d'abord, il est clair que les modèles utilisant la résolution spatiale de 3 mm (E_3^β) ne parviennent pas à apprendre suffisamment du contexte pour être robustes sur des images en résolution originale. Les performances de tous ces modèles sont largement en dessous des autres modèles E_1^β et E_2^β , quelle que soit la métrique utilisée (géométrique ou physiologique). Et augmenter la quantité d'images par lots ne permet pas de dépasser cette perte de détails. La réduction de la résolution ne permet pas non plus de diminuer la complexité de la tâche de segmentation puisque les zones non aérées sont tout de même mal segmentées. Ce sont les seuls modèles pour lesquels, les volumes relatifs prédits sont très éloignés de la référence. En effet, les zones non aérées et aérées sont sous-estimées tandis que les zones hyperinflatées sont sur-estimées. Cela indique alors une tendance à sous-segmenter, étant donné que les compartiments aérés et non aérés représentent⁷⁴ environ trois quarts du volume total des poumons. Cette tendance s'explique en partie par la production d'un masque de segmentation contenant des voxels de taille 3 mm, qui est ensuite redimensionné à la résolution d'origine, provoquant des contours grossiers comme l'illustre la Figure 4.24. Testés sur les bases de données Lyon-2 et Grenoble-2, les

74: Dans les données évaluées.

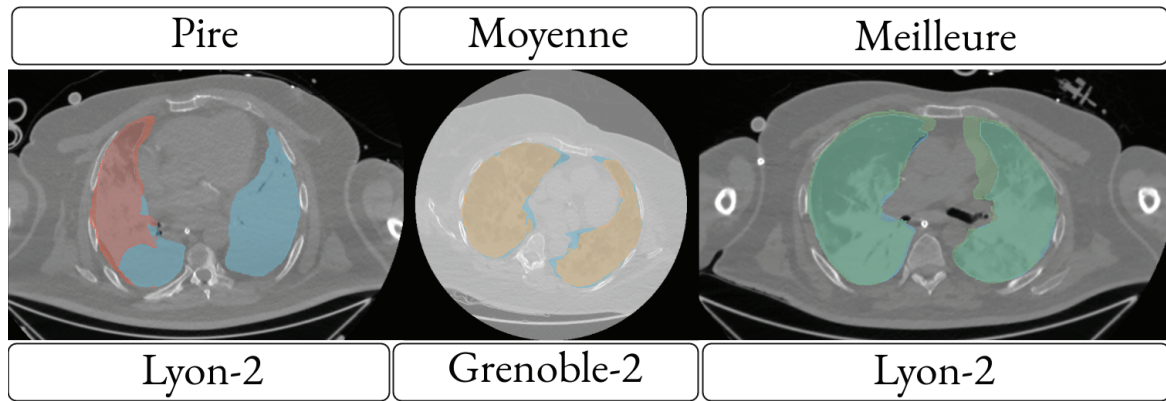


FIGURE 4.23 – Exemples de la pire, la meilleure et une segmentation moyenne prédite par le modèle D_1 sur les bases de données de test Lyon-2 et Grenoble-2.

conclusions sont les mêmes avec une dégradation encore plus importante. Deux autres modèles fournissent des résultats systématiquement inférieurs aux résultats moyens, il s'agit des modèles E_2^8 et E_2^{16} . Ces deux modèles, entraînés avec une résolution d'image de 2 mm et des lots d'images relativement élevés⁷⁵, ne parviennent pas à être robustes. Ils fournissent des segmentations aux contours moins grossiers que les modèles E_3^β mais ont tendance à sur-segmenter les zones non aérées (voir exemple en Figure 4.25 et résultats dans la Figure 4.12). Comme c'est attendu avec de telles performances sur leur base d'entraînement, ces modèles montrent peu de capacité à généraliser.

Les performances géométriques globales des autres modèles E_1^β et E_2^β (pour $\beta < 8$, avec $r=2$) sont toutes très proches entre elles. L'observation des résultats du score de Dice selon les compartiments d'aération permet de mettre en évidence une légère supériorité des modèles E_1^β sur le compartiment peu aéré⁷⁶. La prise en compte des métriques physiologiques permet de différencier un peu plus les performances de ces modèles entre eux. En effet, seuls quatre modèles obtiennent un intervalle de confiance inférieur à 5% pour le biais de recrutement. Et ce nombre se réduit à deux, lorsque le biais sur la cohérence des masses est aussi pris en compte. Ces deux modèles sont celui entraîné avec deux images par lot et une résolution spatiale de 1 mm (E_1^2) et le modèle qui utilise une image par lot et une résolution spatiale de 2 mm (E_2^1). Ces observations confirment davantage la priorité des détails dans l'image à la diversité accessible dans l'utilisation de lots composés de beaucoup d'images. Cette observation pourrait avantager l'apprentissage qui utilise les deux images d'un même patient par lot. Pourtant, cette cohérence est ici pénalisée puisque les modèles E_r^{2p} sont les moins performants⁷⁷ parmi les modèles aux performances proches. La capacité de généralisation de ces modèles reste modérée. Le modèle E_1^2 est le plus robuste, peu importe la base de données évaluée.

En résumé, les modèles entraînés pour cette expérience ont mis en évidence l'importance d'apprendre sur des données en haute résolution dans le cadre de la segmentation des poumons avec SDRA. Même lorsque la taille en mémoire le permet⁷⁸, le recours à davantage d'images par lot n'améliore pas les résultats.

75: $\beta = 8$ ou $\beta = 16$.

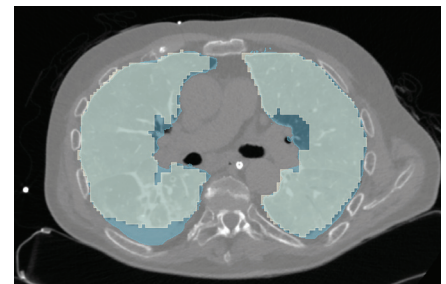


FIGURE 4.24 – Exemple de segmentation prédite par le modèle E_3^2 sur une image de test de la base de données Lyon-1. – En bleu le masque de segmentation de référence, en blanc la segmentation prédite.

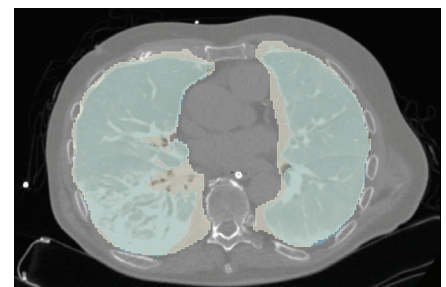


FIGURE 4.25 – Exemple de segmentation prédite par le modèle E_2^8 sur une image de test de la base de données Lyon-1. – En bleu le masque de segmentation de référence, en blanc la segmentation prédite.

76: Pour des prédictions de volumes relatifs du compartiment PA pratiquement égales des deux types de modèles.

77: Même sur les métriques qui évaluent la cohérence telles que le biais sur le recrutement ou celui sur la cohérence de masse.

78: Évalué uniquement en résolution 2 mm.

Expérience sur l'apport d'information physiologique Après avoir étudié l'influence de données additionnelles⁷⁹ et celle de la quantité d'information traitée à chaque itération, les modèles F évalués dans cette expérience ont pour objectif d'intégrer des critères physiologiques dans leur stratégie d'apprentissage. L'accent sur le compartiment d'aération NA ou l'utilisation de la masse prédite entre masques de segmentation à l'expiration et l'inspiration sont ainsi évalués.

Les performances géométriques globales des modèles F sont assez proches, à l'exception du modèle F_{masse1} . Il s'agit du modèle pour lequel le terme de masse est utilisé dès le début et tout le long de l'apprentissage. Et malgré des courbes d'entraînement qui décroissent de façon stable, l'apprentissage ne semble pas avoir été dirigé dans la bonne direction. En effet, au vu des performances très inférieures de ce modèle, cette stratégie de mélange ne permet pas d'extraire les caractéristiques descriptives des poumons. Ceci peut être expliqué par la prédominance du terme de masse dès le début de l'apprentissage, qui pour diminuer doit uniquement correspondre numériquement aux masses des poumons à l'expiration et l'inspiration. La contrainte spatiale apportée par le terme de perte sur le Dice ne parvient pas à contrebalancer le déséquilibre initial, et donc la fonction sur la masse diminue, avec des masses segmentées décorréliées des poumons. Le modèle obtient un score de Dice médian inférieur à 0.5 sur sa base d'entraînement, ce qui indique des segmentations dont plus de la moitié n'est pas superposée à la référence (voir exemple en Figure 4.26). Les volumes relatifs prédits par ce modèle confirment davantage son incapacité à identifier les poumons. Enfin, ses performances de généralisation et cohérence de segmentation restent tout aussi faibles.

Les autres modèles, en revanche, obtiennent des résultats globaux acceptables, avec un Dice moyen de 0.95. Leur performance par compartiment d'aération les différencie entre eux. En effet, le modèle F_α obtient systématiquement les meilleurs résultats (en termes de médiane et d'écart-type), en particulier sur les compartiments peu et non aéré. Les volumes relatifs prédits par ce modèle sont très proches de ceux de la référence, ce qui appuie davantage la cohérence des segmentations autant taille qu'en position. Toutefois, la répartition plus large des volumes relatifs sur le compartiment hyperinflaté indique une sur-segmentation de ces zones.

Les performances de ces modèles évalués sur les bases de données Lyon-2 et Grenoble-2 montrent une légère perte de robustesse générale sur les métriques géométriques. Sur les deux bases de données, le modèle F_α obtient les meilleures médianes pour les trois métriques géométriques, mais pour toutes celles-ci un plus grand écart-type que le modèle F_{masse2} . Ainsi, le modèle F_{masse2} semble être plus précis et robuste peu importe la base de données testée.

Lorsque les métriques de cohérence physiologique sont utilisées, les modèles F_α et F_{masse2} fournissent encore des résultats peu différenciant. Le premier est plus cohérent sur les masses globales tandis que le second l'est davantage sur le recrutement (évalués sur leur base de données d'entraînement). Sur les bases de données de test, les deux modèles sont au-dessus du seuil de 5% sur les données de Lyon-2, mais c'est le modèle F_{masse2} qui obtient la meilleure cohérence⁸⁰ sur les données de Grenoble-2.

En résumé, les modèles F_α et F_{masse2} présentent des performances proches, les unes meilleures que les autres, selon la base de données de test évaluée ou la métrique observée.

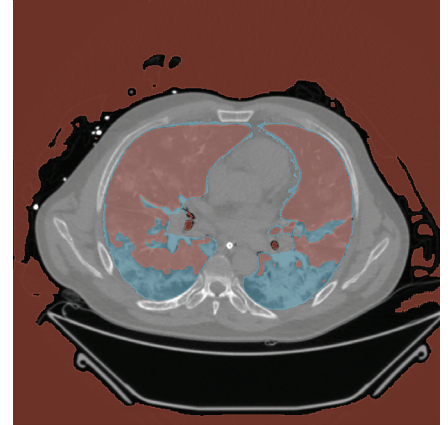


FIGURE 4.26 – Exemple de la pire segmentation du modèle F_{masse1} sur une image de test de la base de données Lyon-1. – En bleu la segmentation de référence, en rouge la prédiction du modèle F_{masse1}

79: Via le transfert d'apprentissage.

80: Sur le recrutement ou la masse des poumons.

4.6 Conclusion

Les expériences menées dans ce chapitre avaient pour but d'évaluer si la gestion de l'information provenant de l'image dans l'apprentissage 3D pouvait faire gagner en robustesse et capacité de généralisation. Les éléments suivants ont été testés :

- l'initialisation, avec l'apport d'information via le pré-entraînement ;
- la quantité d'information vue à chaque itération, à travers :
 - la finesse des détails présents dans l'image ou
 - la diversité des exemples vus ;
- la cohérence et la pertinence des informations traitées, avec l'utilisation d'information d'aération ou de masse.

La première expérience ne démontre pas d'amélioration de l'utilisation du transfert d'apprentissage. Par rapport au modèle de référence, le gain de robustesse des deux autres modèles (entraînés chacun sur 200 epochs, l'un avec transfert d'apprentissage, l'autre avec initialisation aléatoire) évalués n'est pas significatif. Ce résultat est à considérer au regard de la tâche évaluée, la segmentation des poumons de patients avec SDRA. Celle-ci diffère par deux aspects de l'application du pré-entraînement utilisé : l'anatomie (porcine dans le pré-entraînement) et la ressemblance du SDRA représenté avec celui observable chez l'humain. Ainsi, ces différences ne semblent pas ajouter en robustesse par rapport aux données déjà présentes dans la base de données d'entraînement (Lyon-1).

La seconde expérience indique une prévalence de la finesse des détails contre la diversité des exemples utilisés pour la rétropropagation durant l'apprentissage. Ce constat contredit l'idée que le contexte suffit pour segmenter des gros objets tels que les poumons dans la cage thoracique⁸¹. Les résolutions 1 mm et 2 mm fournissent des résultats très proches (la première obtenant systématiquement les meilleurs résultats), avec l'avantage pour la seconde d'être plus rapide à entraîner. Entraîner avec la résolution la plus élevée reste alors la stratégie à privilégier.

Enfin, l'utilisation d'information quantitative telle que l'aération ou la masse dans la fonction de perte dans la troisième expérience permet de gagner sensiblement en robustesse sur les compartiments les plus compliqués (peu aéré et non aéré). Cependant, les stratégies d'ajout de ces termes à la fonction de perte ne permettent pas d'affecter significativement les performances globales et ni de dépasser les performances du modèle de référence. L'utilisation de ces informations reste tout de même encourageante et à explorer, au vu du gain qui est observé dans les parties qui sont les plus sources d'erreurs.

Les objectifs complétés dans ce chapitre sont résumés dans l'encadré suivant.

81: Champ de vue habituel des images CT utilisées.

Objectifs initiaux et état d'avancement		
Cadre d'utilisation	Mise en application en clinique	<i>Recherche (CREATIS)</i>
Tâche	Segmentation des poumons sur image CT 3D	Méthode 3D
Cohérence	Fournit peu de différences entre deux images traitées du même patient	Cohérence de recrutement ou de masse
Interactivité	Automatique	Oui
Délai de calcul	< 1h	Environ 2 s (GPU) ou 11 s (CPU)
Précision	Nécessite peu de retouches manuelles (si présentes)	<i>Erreurs de sous-segmentation à corriger, aspect lisse des segmentations</i>
Robustesse	Indépendant de la sévérité de SDRA et lésions présentes	<i>Dépendante</i>

En gras les objectifs atteints dans ce chapitre, en police normale ce qui a été atteint dans les chapitres précédents et en italique ce qu'il reste à faire.

L'objectif de cette thèse est de proposer une solution robuste et généralisable de segmentation des poumons dans le cadre du SDRA. Le choix d'utiliser l'apprentissage profond, au vu de la littérature scientifique du domaine, a mené à tester divers aspects de l'entraînement de modèle d'apprentissage profond. Plusieurs approches ont été évaluées, en 2D et 3D, afin d'identifier quelle stratégie d'utilisation des données disponibles permettait d'atteindre l'objectif de cette thèse. À partir des résultats ainsi obtenus, des modèles de production ont été proposés aux cliniciens accompagnant le travail de cette thèse. Les descriptions de la mise en place de ces modèles, de leurs performances en comparaison d'un modèle de l'état de l'art et de leurs améliorations sont présentées dans ce chapitre.

Table des matières

5.1	Introduction	84
5.2	Données	85
5.3	Méthode	87
	Modèles	88
	Intégration au logiciel de segmentation semi-automatique : creaSDRA	88
	Évaluation	89
5.4	Performances	89
5.5	Discussion	93
5.6	Conclusion	94

5.1 Introduction

Les précédentes expériences, menées dans les Chapitre 3 et Chapitre 4, ont permis de tirer des conclusions qui sont utilisées pour la mise en place de deux types de modèle de production. La première concerne l'architecture du modèle, puisque le modèle 3D évalué dans le Chapitre 4 a fourni des masques aux contours lisses¹ et montré une cohérence de masse qui seront appréciables en clinique. La deuxième leçon tirée des expériences précédentes concerne la stratégie de gestion des données. En effet, le transfert d'apprentissage (ou l'apprentissage conjoint) depuis les données de modèles porcins expérimentaux n'a pas montré de gain significatif sur la robustesse ou la capacité de généralisation. Pour cette raison, deux modèles de production seront réalisés, chacun spécialisé sur une anatomie. Les performances des modèles du Chapitre 4 testés en généralisation ont également apporté permis de choisir de manière avisée les bases de données à utiliser pour l'entraînement d'un modèle avec bonne capacité de généralisation. En effet, tous les modèles testés dans le Chapitre 4 sur la base Lyon-2, montraient systématiquement de moins bonnes performances en généralisation que lorsqu'ils étaient testés sur la base Grenoble-2. La première était composée de patients avec des SDRA

1: En comparaison du modèle multi-coupes utilisé dans le Chapitre 3.

sévère, notamment des patients sous ECMO qui n'étaient pas présents dans la base de données d'entraînement. La deuxième, composée tout comme la base d'entraînement d'un nombre équilibré de patients avec SDRA modéré ou SDRA sévère, mais avec des images acquises sur un imageur différent, provoquait pourtant de meilleurs résultats de généralisation. Utiliser la base de données Lyon-2 pour l'entraînement semble donc capital pour améliorer la capacité de généralisation d'un modèle de production, d'autant plus que les patients sous ECMO n'ont jamais été segmentés automatiquement dans la littérature. L'utilisation des informations d'aération ou de masse qui semblait être également prometteuse, n'a pas abouti avec les expériences évaluées à des résultats qui permettent de la choisir pour un modèle de production. Enfin, le recours à une résolution élevée² des images utilisées pour l'entraînement a permis d'avoir systématiquement les meilleurs résultats sur les données en résolution originale.

De ces choix, deux modèles sont proposés. Le premier est appliqué à la segmentation des poumons des modèles expérimentaux porcins de SDRA. Le second est appliqué à la segmentation des poumons chez les patients atteints du SDRA. Il s'agit pour ces modèles d'être intégrés à un logiciel de segmentation semi-automatique utilisé en pré-clinique au service de réanimation de l'hôpital de la Croix-Rousse. La segmentation ainsi fournie permet de calculer directement des informations d'aération. À partir de ces informations additionnelles obtenues pratiquement en temps réel, le clinicien peut alors adapter le réglage de la ventilation mécanique [Her+21]. Faut de données de test annotées, le modèle de segmentation porcine n'est pas évalué. Sa mise en place est simplement décrite dans ce chapitre.

Le modèle de segmentation humaine est quant à lui évalué sur des données annotées laissées volontairement hors de l'entraînement. Ce choix est justifié par la volonté de valoriser ce travail dans une publication. Celle-ci évaluera notamment les variabilités de segmentation intra- et inter-observateurs en comparaison de celles fournies par la méthode automatique.

5.2 Données

Les modèles de production proposés ont pour objectif d'avoir une grande capacité de généralisation et une forte robustesse. Pour cela, un choix éclairé a été fait parmi les bases de données à disposition.

Le premier type de modèle de production a pour but une utilisation dans le cadre d'études expérimentales autour du SDRA sur modèle porcine. Les deux bases de données animales présentées dans le Chapitre 3 (Hamilton et PK HIDRA) ont ainsi été utilisées pour entraîner ce modèle et sont désignées ensemble par la notation H-PK. Cela représente 169 volumes annotés, acquis chez 14 cochons différents dans des conditions ventilatoires différentes. Des exemples des images présentes dans ces 169 volumes sont présentés dans la Figure 5.1.

Concernant le type de modèle de production destiné à l'utilisation en clinique, les bases de données animales³ n'ont pas été utilisées, car n'apportant pas significativement de capacité de généralisation, comme cela a été montré dans les chapitres 3 (en 2D) et 4 (en 3D). Les bases de données CT4-ARDS et CT4-ARDS2 ont quant à elles été utilisées. Les 67 patients non utilisés dans le chapitre précédent ont cette fois-ci été utilisés

2: En comparaison avec des résolutions inférieures.

3: Modèles porcins expérimentaux de SDRA.

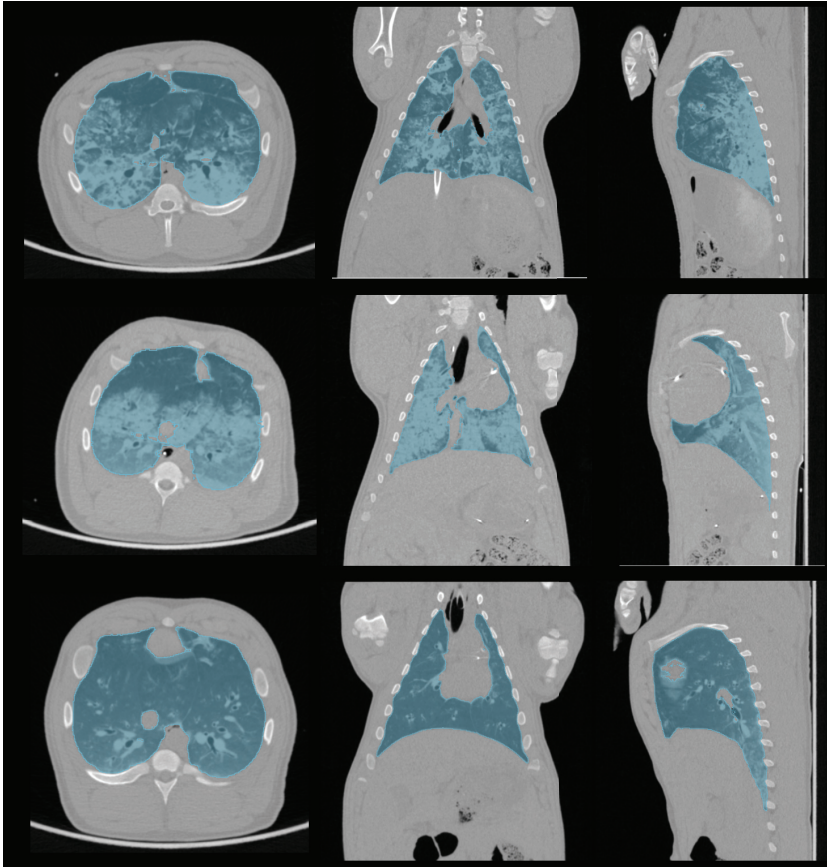


FIGURE 5.1 – Exemple de trois cochons pris dans les bases de données Hamilton et PK HIDRA.

pour l'entraînement du modèle de production. Le sous-ensemble Lyon-2 utilisé dans le Chapitre 4 n'a pas été utilisé pour l'entraînement, mais laissé pour le test⁴. Pour différencier les données de CT4-ARDS2 utilisées pour l'entraînement ou le test, un E (pour entraînement) et un T (pour test) sont ajoutés aux identifiants des bases de données. Les données utilisées pour l'entraînement sont ainsi renommées Lyon-12E. Le sous-ensemble Grenoble-2 a aussi été conservé uniquement pour le test, bien qu'il aurait pu apporter en capacité de généralisation à l'entraînement. Ce choix s'est justifié par la faible qualité des segmentations de référence, réalisées par un opérateur moins expérimenté. Ainsi, le détail des données utilisées pour l'entraînement du modèle de production est résumé dans le Tableau 5.1.

4: Avec comme objectif de publier les résultats obtenus sur ce sous-ensemble

TABLEAU 5.1 – Détails des bases de données CT4-ARDS et CT4-ARDS2 utilisées respectivement pour l'entraînement ou le test.

Base de données	Entraînement		Test	
	CT4-ARDS	CT-4ARDS2	CT4-ARDS2	CT4-ARDS2
Identifiant	Lyon-1	Lyon-2E	Lyon-2T	Grenoble-2
Patients	30	67	13	21
Volumes	60	256	50	68
Sévérité				
Modéré	15 (50%)	6 (9%)	2 (15%)	9 (43%)
Sévère	15 (50%)	40 (60%)	9 (69%)	12 (57%)
Sévère sous ECMO	0	21 (31%)	2 (15%)	0
Autre lésion pulmonaire*	4 (13%)	15 (22%)	3 (23%)	1 (5%)

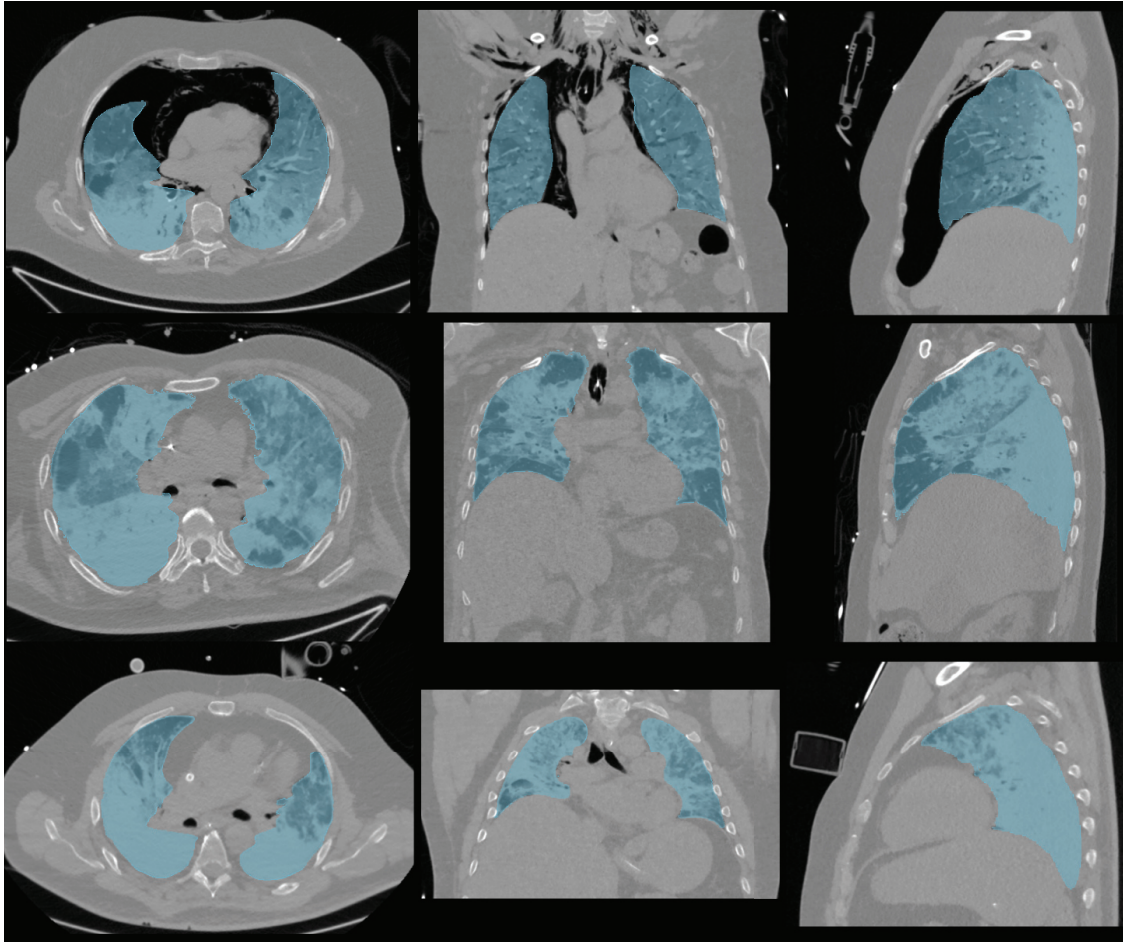


FIGURE 5.2 – Exemples de patients présents dans la base de données d’entraînement Lyon-12E avec leurs segmentations de référence. – Sur la première ligne un patient avec SDRA modéré et un pneumothorax, sur la deuxième un patient avec SDRA sévère sans ECMO et sur la dernière un patient avec SDRA sévère sous ECMO.

Les données utilisées pour l’entraînement ont toutes subies un pré-traitement visant à obtenir des images aux voxels isotropes de taille 1 mm. Le redimensionnement des images était précédé de l’application d’un filtre gaussien pour respecter le critère de Shannon. Enfin, les images sont toutes rognées ou remplies pour avoir une taille fixe de $448 \times 448 \times 320$ pour la base Lyon-1 et $448 \times 448 \times 352$ pour la base Lyon-2E.

5.3 Méthode

Les modèles entraînés sont intégrés dans un logiciel de segmentation semi-automatique utilisé par les cliniciens du centre de réanimation de l’hôpital de la Croix-Rousse. Le premier est dédié à la segmentation des poumons de cochons avec modèle expérimental de SDRA et noté $U\text{-net}_{animal}$. Le second est conçu pour la segmentation des poumons chez l’humain atteint du SDRA et noté $U\text{-net}_{SDRA}^{3D}$. Un autre modèle de l’état de l’art est également utilisé⁵ pour la comparaison des performances avec les modèles de production fournis. La description de leur mise en place et évaluation est décrite ci-dessous.

5: Uniquement en inférence.

Modèles

Deux types d'architectures ont été utilisées dans cette thèse. Dans un premier temps, l'architecture du U-net 2D [Ron+15] dans le modèle multicoups (2.5D) du Chapitre 3 et dans un deuxième temps l'architecture du U-net 3D [Çiç+16] dans le Chapitre 4. La première présente l'avantage d'être rapide à entraîner et d'être assez robuste sur sa base de données d'entraînement. Pourtant, elle fournit des segmentations aux contours peu lisses et ne garantit aucune cohérence tridimensionnelle du fait de sa construction. C'est pourquoi le modèle choisi pour le modèle de production a été l'architecture du U-net 3D. Ce modèle a l'avantage d'assurer une cohérence 3D, de produire des segmentations aux contours lisses⁶ et d'être robuste et relativement généralisable avec peu de données d'entraînement.

Comme dans le chapitre précédent, l'architecture utilisée est légèrement modifiée par rapport à la publication d'origine [Çiç+16]. Cinq couches sont implémentées au lieu de trois et seuls quatre filtres sont utilisés en entrée du modèle au lieu de trente-deux. De ces modifications résulte une architecture profonde, mais légère en termes de paramètres entraîna- bles, puisqu'elle est presque dix fois plus légère que l'architecture originale.

Les deux modèles sont entraînés pendant 100 epochs : le modèle U-net^{3D}_{SDRA} sur les données Lyon-12E⁷ et le modèle U-net_{animal} sur les données H-PK. Pour chaque modèle individuellement, un lot de 2 images entières, prises aléatoirement dans sa base de données d'entraînement, est utilisé à chaque itération. La fonction de perte basée sur le score de Dice⁸ permet à l'algorithme d'optimisation Adam d'actualiser les poids, avec un taux d'apprentissage de 3.10^{-3} .

Un modèle entraîné de l'état de l'art, disponible en ligne, est ensuite utilisé pour comparer les performances des modèles de production. Il s'agit d'un U-net⁹ 2D, entraîné par coupes axiales [Hof+20]. Le code et les poids de ce modèle sont récupérés et utilisés sans aucune modification. Il sera noté ici U-net^{2D}_{EA} avec l'indice EA pour état de l'art et l'exposant 2D pour le différencier du U-net 3D proposé. Il prévoit un redimensionnement des images traitées à l'entraînement¹⁰ à la taille 256×256 , après un rognage entourant uniquement le corps dans l'image. L'apprentissage du modèle a été réalisé sur 231 volumes représentant 108 248 coupes. Les données utilisées sont issues de 22 combinaisons de paramètres d'acquisition différentes¹¹.

Intégration au logiciel de segmentation semi-automatique : creaSDRA

Afin de permettre l'utilisation des modèles de production sur les différents systèmes d'exploitation existants, un travail de mise en forme a été nécessaire en collaboration avec un ingénieur du service de développement informatique du laboratoire. Un code accompagné de son environnement virtuel¹², permettant toutes les étapes de traitement, de la lecture à l'enregistrement du masque de segmentation¹³, a ainsi été fourni. Le détail de ces étapes est présenté dans l'organigramme en Figure 5.3.

Le logiciel fonctionne avec les deux images d'un même patient¹⁴, celle en fin d'expiration et celle en fin d'inspiration. Une fois les deux segmentations enregistrées, l'analyse des informations d'aération est alors directement

6: Plus facilement corrigeables.

7: La combinaison des données d'entraînement Lyon-1 et Lyon-2E.

8: $\mathcal{L}_{Dice} = 1 - Dice$

9: Augmenté de l'étape de normalisation des lots à chaque couche.

10: Comme à l'inférence.

11: Scanner, épaisseur de coupe, filtre de convolution.

12: Un environnement virtuel est un répertoire contenant une installation de Python autonome.

13: Aux mêmes dimensions que l'image originale.

14: Ou d'un même cochon.

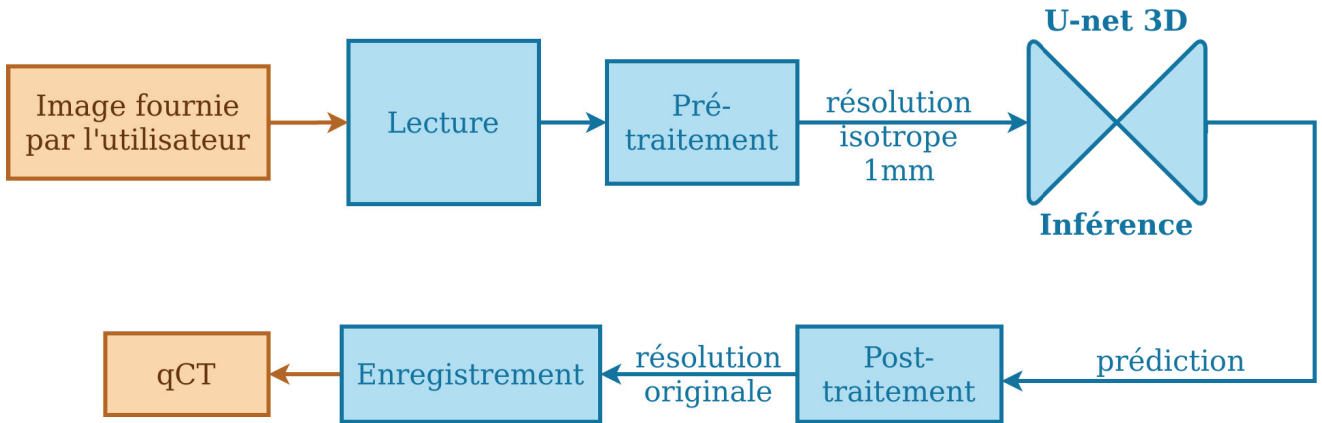


FIGURE 5.3 – Organigramme des tâches réalisées à l'interface (entrée et sortie) du logiciel de segmentation semi-automatique. En orange les tâches réalisées dans le logiciel, en bleu celles réalisées par le code fourni dans un environnement virtuel.

réalisée dans l'interface [Dáv+20]. Des outils de correction simples sont aussi prévus dans cette interface, lorsque la qualité du masque de segmentation n'est pas jugée suffisante par l'utilisateur. Des améliorations de ces outils sont d'ailleurs en cours de développement.

Évaluation

Seul le modèle destiné à la segmentation des poumons chez les patients atteints du SDRA est évalué, $U\text{-net}_{SDRA}^{3D}$. Ses performances sont comparées à celles du modèle de l'état de l'art $U\text{-net}_{EA}^{2D}$. Les deux méthodes sont évaluées en robustesse et capacité de généralisation sur les bases de données Lyon-2T et Grenoble-2. Le masque de segmentation prédit par chaque méthode est ainsi soumis aux métriques géométriques¹⁵ : score de Dice, distance ASSD et distance de Hausdorff (au 95^e centile). Le détail du score de Dice par compartiment d'aération est aussi exploité de paire avec les volumes relatifs correspondant à chacun des compartiments. Enfin, les deux méthodes de segmentation sont chacune comparées à la segmentation de référence par tracé de Bland-Altman [Alt+83].

15: Définies dans le Chapitre 2

5.4 Performances

Les performances du modèle de production $U\text{-net}_{SDRA}^{3D}$ sont comparées au modèle de l'état de l'art $U\text{-net}_{EA}^{2D}$ sur les bases de données Lyon-2T et Grenoble-2.

Les mesures géométriques des deux méthodes sont présentées dans la Figure 5.4. Quelle que soit la métrique utilisée, le modèle produit dans cette thèse, $U\text{-net}_{SDRA}^{3D}$ fournit les meilleurs résultats. La médiane, l'écart-type et la quantité de valeurs aberrantes sont des caractéristiques qui sont toutes meilleures pour le modèle $U\text{-net}_{SDRA}^{3D}$. Les performances atteintes sont ainsi de 0.957 ± 0.028 (Dice), 1.19 ± 0.87 mm (ASSD) et 4.63 ± 5.98 mm (DH-95). Les métriques de distance sont celles qui montrent le plus d'écart entre les deux méthodes.

Rappel : définition des compartiments d'aération

Compartiment d'aération	Seuils (HU)
Non aéré (NA)	[-100; +100]
Peu aéré (PA)	[-500; -101]
Aéré (A)	[-900; -501]
Hyperinflaté (HI)	[-1000; -901]

Le détail des performances sur la métrique ASSD par niveau de sévérité du SDRA des patients considérés, est présenté dans la Figure 5.5. L'écart de performances se creuse davantage entre les deux méthodes avec ce point de vue. En effet, sur les cas¹⁶ de SDRA modéré (avec le moins de lésions denses¹⁷), les deux modèles obtiennent des médianes très proches d'environ 1.2 mm. Cependant, c'est le modèle $U\text{-net}_{SDRA}^{3D}$ qui fournit les résultats les plus précis avec un écart interquartile de 0.77 mm contre 4.74 mm pour le modèle de l'état de l'art. Ces résultats se dégradent de paire avec le niveau de sévérité qui augmente. Le modèle $U\text{-net}_{SDRA}^{3D}$ fournit sur les patients avec SDRA sévère (sans ECMO) les résultats les plus précis et des valeurs aberrantes sous les 4.5 mm. Au contraire, le modèle $U\text{-net}_{EA}^{2D}$ montre une boîte englobante plus large avec davantage de valeurs aberrantes et la plus grande qui atteint 7 mm. Enfin, pour les patients atteints du SDRA le plus grave¹⁸ (nécessitant alors le recours à l'ECMO) le modèle $U\text{-net}_{SDRA}^{3D}$ dépasse encore une fois les performances du modèle $U\text{-net}_{EA}^{2D}$. Il obtient sa plus grande médiane sur cette métrique qui est de 1.38 mm contre 3.02 mm pour le modèle de l'état de l'art. Ce dernier fournit des résultats très dispersés puisque son troisième quartile se situe à 7.12 mm et son maximum à 12.97 mm.

Lorsque les résultats sont observés du point de vue des compartiments d'aération (voir Figure 5.6) on observe une différence de performance entre les deux modèles sur les compartiments les plus compliqués à segmenter. En effet, les compartiments peu aérés et non aérés sont mieux segmentés par le modèle proposé dans ce chapitre. Les volumes relatifs prédits par ce dernier sont aussi les plus proches de ceux de référence. Le modèle de l'état de l'art a plutôt tendance à sous-estimer le compartiment non aéré, ce qui se voit dans les volumes relatifs produits.

Enfin, les résultats de cohérence sur le recrutement prédit (voir Figure 5.7) par les deux méthodes montrent une plus forte cohérence du modèle $U\text{-net}_{SDRA}^{3D}$. Ce dernier obtient un très faible biais moyen de 1% pour des différences comprises pour la plupart entre -7% et 5.2%. Le biais moyen du modèle $U\text{-net}_{EA}^{2D}$ est, lui aussi, faible (1%) mais ses différences se répartissent pour la plupart entre -8.2% et 10%.

Les performances de généralisation du modèle $U\text{-net}_{SDRA}^{3D}$ sont ensuite observées sur la base de données Grenoble-2. Les résultats sur les métriques géométriques globales sont présentés dans la Figure 5.8. Le constat est le même que sur la base de données Lyon-2T, le modèle $U\text{-net}_{SDRA}^{3D}$ dépasse le modèle de l'état de l'art sur toutes les métriques globales. Il obtient ainsi 0.940 ± 0.034 sur le score de Dice, 2.18 ± 1.23 mm sur la métrique ASSD et 9.78 ± 7.90 mm sur la métrique DH-95. L'écart interquartile est sur cette base pratiquement égal pour les deux méthodes, avec des résultats un peu plus dispersés que sur la base de données Lyon-2T¹⁹ une légère perte de robustesse est observée.

La base de données Grenoble-2 ne compte que 2 niveaux de sévérité (modéré et sévère) du SDRA représentés par ses 21 patients, sur lesquels le modèle $U\text{-net}_{SDRA}^{3D}$ est le plus robuste (voir Figure 5.9). Les erreurs avec la métrique ASSD sont inférieurs de 1.82 mm pour les patients avec SDRA modéré, et inférieurs à 2.74 mm pour les patients avec SDRA sévère (sans ECMO).

Le détail des performances par compartiment d'aération est présenté dans la Figure 5.10. Ici aussi, les observations sont identiques à celles faites sur la

16: Représentant 2 patients pour 6 volumes.

17: Parmi les niveaux de sévérité représentés.

18: Ils sont au nombre de 2 représentant 8 volumes.

19: Une sous-partie de la base de données sur laquelle est entraîné le modèle $U\text{-net}_{SDRA}^{3D}$.

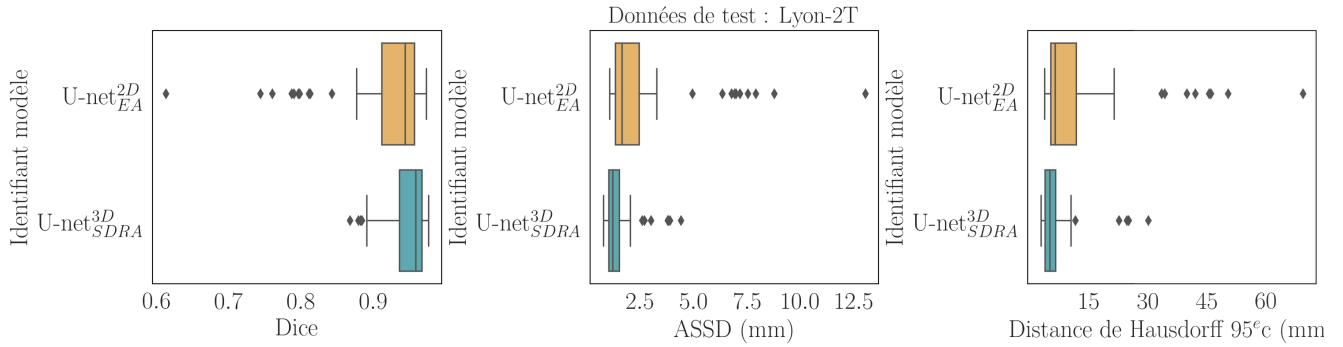


FIGURE 5.4 – Performances géométriques des modèles $U\text{-net}_{SDRA}^{3D}$ et $U\text{-net}_{EA}^{2D}$ sur la base de données Lyon-2T.

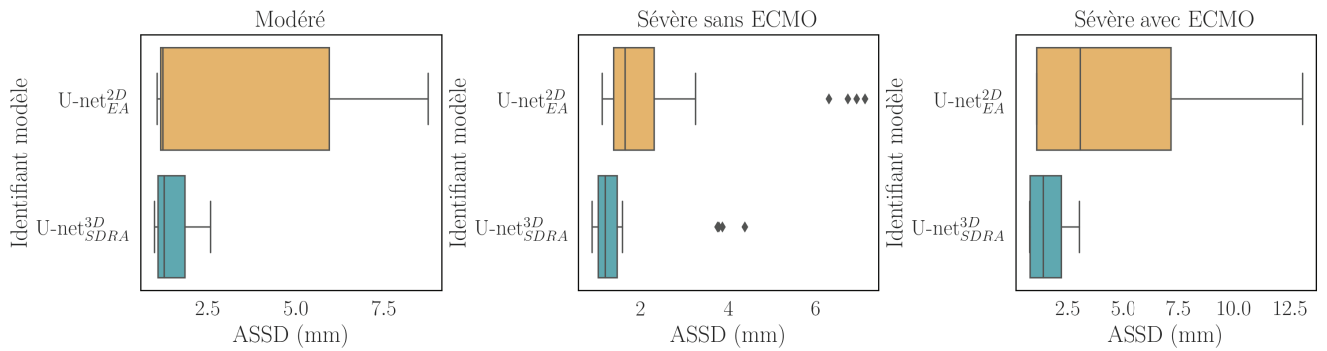


FIGURE 5.5 – Performances avec la métrique ASSD détaillées par niveau de sévérité du SDRa sur la base de données Lyon-2T.

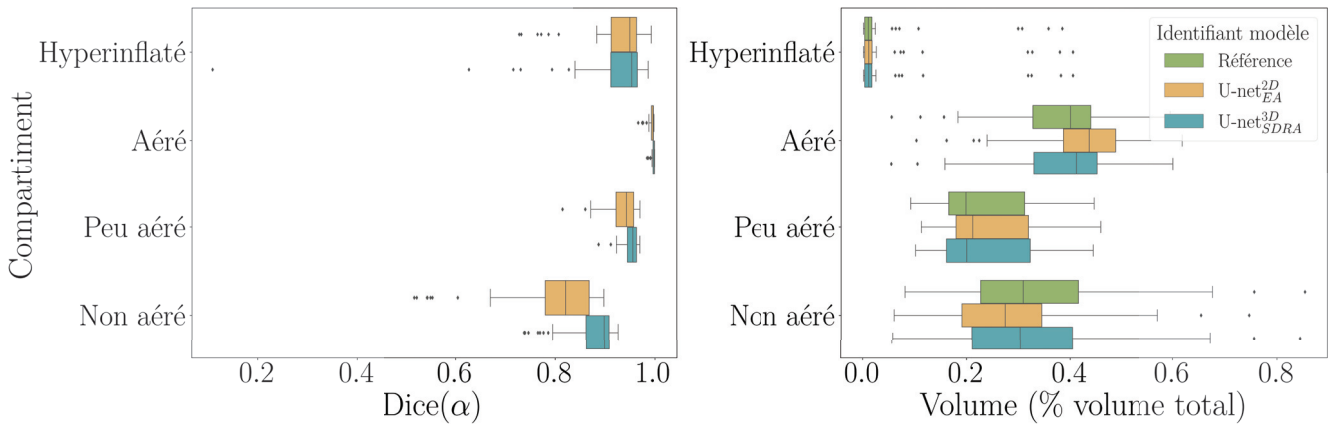


FIGURE 5.6 – Détail des performances par compartiment d'aération sur la base de données Lyon-2T.

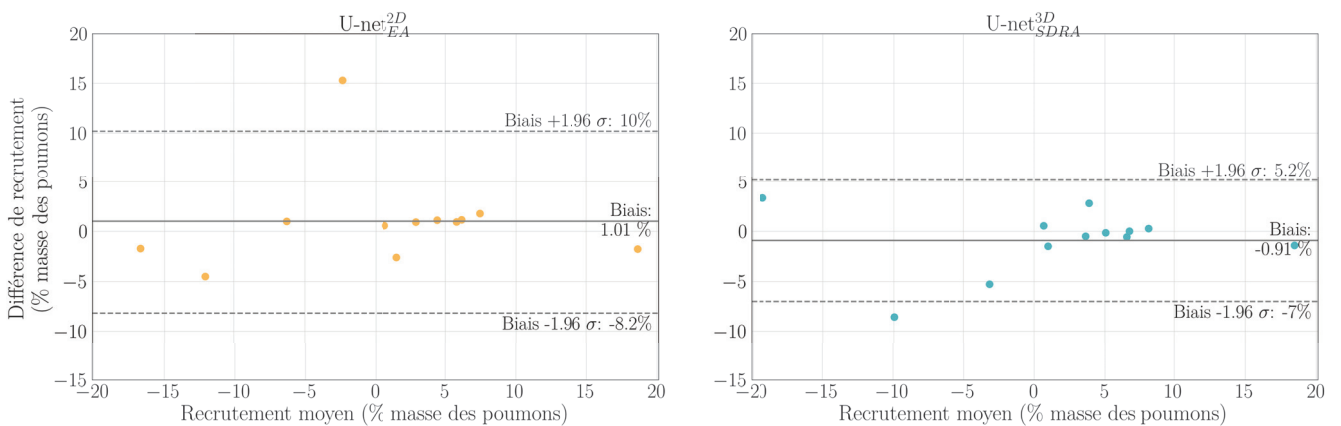


FIGURE 5.7 – Tracé de Bland-Altman sur la mesure du recrutement sur la base de données Lyon-2T, par les deux méthodes, en comparaison avec celle de référence.

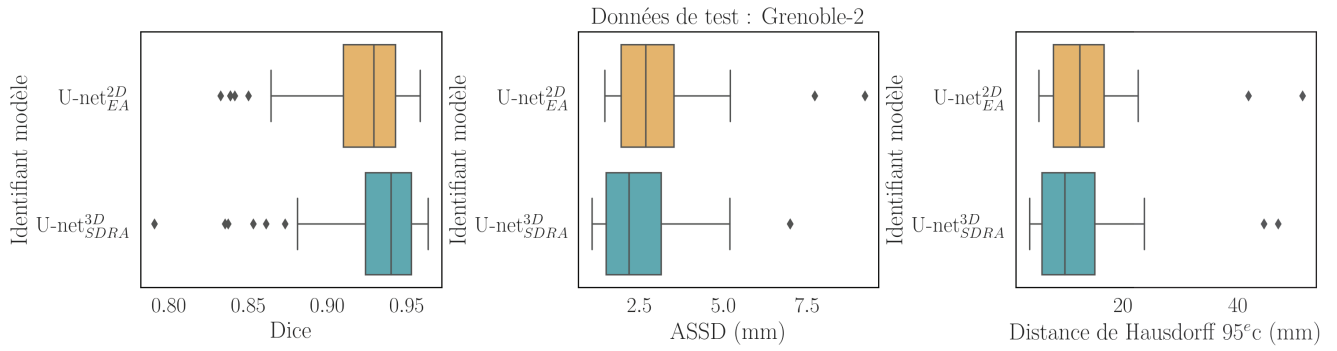


FIGURE 5.8 – Performances géométriques des modèles U-net^{3D}_{SDRA} et U-net^{2D}_{EA} sur la base de données Grenoble-2.

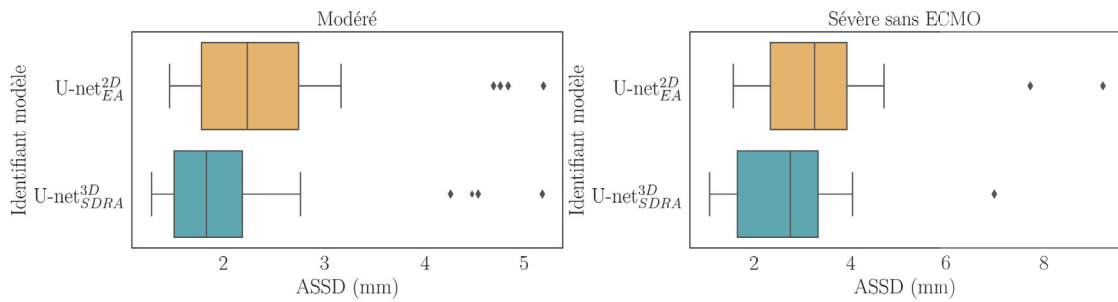


FIGURE 5.9 – Performances avec la métrique ASSD détaillées par niveau de sévérité du SDRA sur la base de données Grenoble-2.

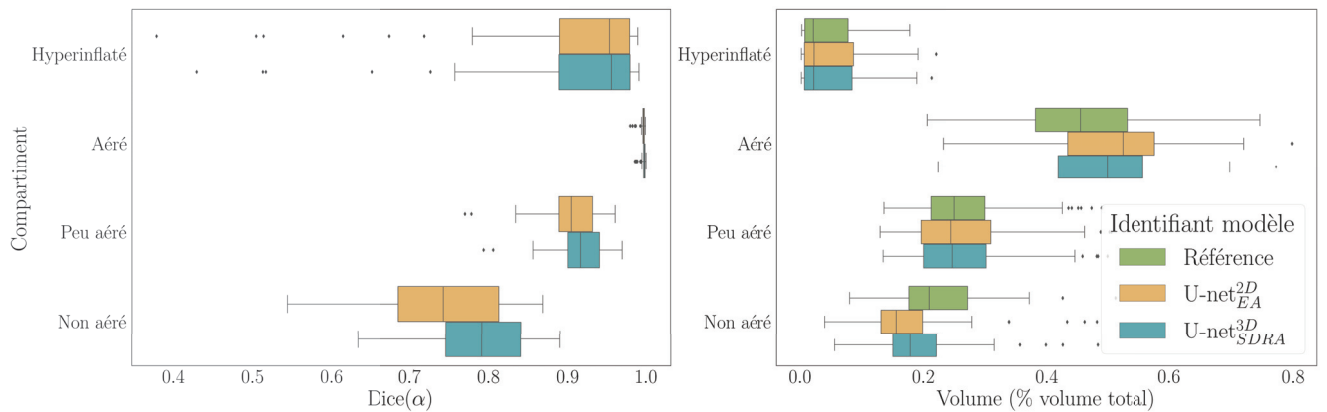


FIGURE 5.10 – Détail des performances par compartiment d'aération sur la base de données Grenoble-2.

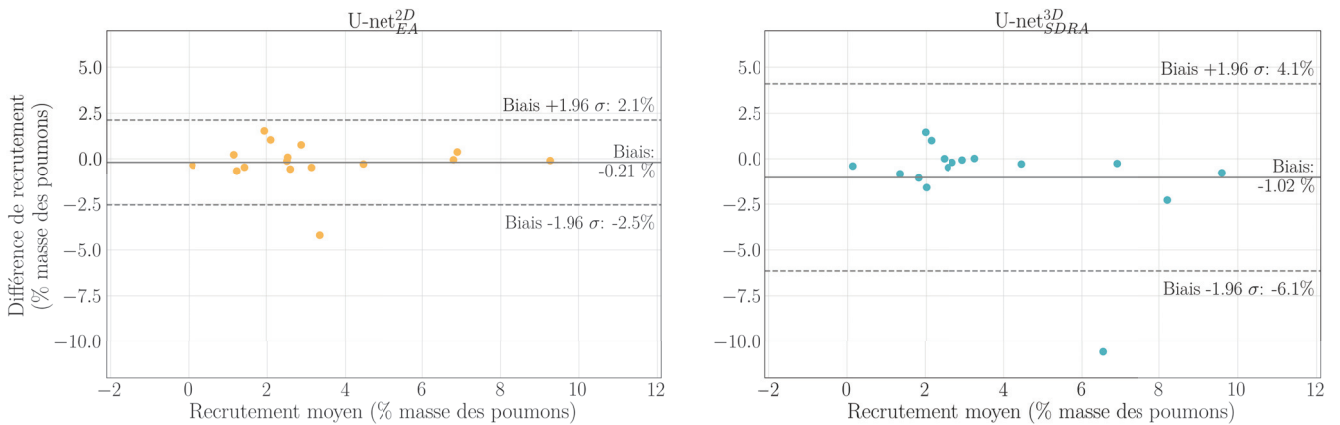


FIGURE 5.11 – Tracé de Bland-Altman sur la mesure du recrutement sur la base de données Grenoble-2, par les deux méthodes en comparaison avec celle de référence.

base de données Lyon-2T. Le modèle $U\text{-net}_{SDRA}^{3D}$ est ainsi plus robuste que le modèle de l'état de l'art sur les compartiments peu aéré et non aéré. Les volumes relatifs prédits par le modèle proposé sont aussi systématiquement les plus proches de ceux de référence.

Enfin, les résultats de cohérence entre les méthodes évaluées et la référence sont présentés sous forme de tracé de Bland-Altman dans la Figure 5.11. Cette fois-ci, c'est le modèle de l'état de l'art qui montre le plus petit intervalle de confiance de -2.5% à 2.1% autour d'un biais de recrutement nul. Le modèle proposé obtient, quant à lui, un biais de -1% et un intervalle de confiance de -6.1% à 4.1%.

Le modèle $U\text{-net}_{SDRA}^{3D}$ a été entraîné sur deux GPU de 32 Go pendant 2 semaines. L'inférence du modèle sur un volume de taille $512 \times 512 \times 315$ prend 10 secondes sur CPU et 8 Go en mémoire. Le traitement total de l'image, de sa lecture à son post-traitement, prend 27 secondes. En comparaison, le temps d'inférence du modèle de l'état de l'art est d'environ 2 minutes et 15 secondes pour le même volume avec 2 Go utilisé comme espace mémoire.

L'évaluation qualitative des masques permet de mettre en évidence des erreurs typiques commises par les deux modèles. Des exemples de ces erreurs sont affichés dans la Figure 5.12. Sur la base de données Lyon-2T, la pire segmentation illustre comment le modèle de l'état de l'art a raté la segmentation du poumon droit, alors que le modèle $U\text{-net}_{SDRA}^{3D}$ le segmente en grande partie. Et même sur la meilleure segmentation, une erreur de segmentation du diaphragme apparaît pour le modèle $U\text{-net}_{EA}^{2D}$. Sur la base de données Grenoble-2, ce sont plutôt des erreurs de sous-segmentation des zones non aérées qui sont mises en évidence pour les deux modèles. Les erreurs qui subsistent sur les meilleures segmentations se situent généralement au niveau du médiastin qui est aussi sous-segmenté.

5.5 Discussion

Le modèle proposé $U\text{-net}_{SDRA}^{3D}$, spécialisé pour la segmentation des poumons dans le cadre du SDRA, a ainsi été comparé à un modèle de l'état de l'art destiné à la segmentation des poumons en présence de pathologies. Sur les deux bases de données évaluées (Lyon-2T et Grenoble-2) le modèle proposé est toujours le plus robuste, surtout dans le cas de patients atteints d'un SDRA très sévère. La base de données Lyon-2T est issue du même centre d'acquisition que la majorité des données utilisées pour l'entraînement, la capacité de généralisation évaluée sur celle-ci se concentre alors uniquement sur les nouveaux patients testés. Les résultats ainsi obtenus sur ce sous-ensemble de test sont satisfaisants au vu de la représentation de sévérité des patients évalués²⁰. Ajouté à cela le fait que le modèle $U\text{-net}_{SDRA}^{3D}$ est entraîné sur les images de 97 patients, contre 231 pour le modèle de l'état de l'art, la robustesse et la capacité de généralisation observées sont satisfaisantes. La capacité de généralisation évaluée sur la base Grenoble-2, qui concerne alors des images acquises avec des réglages différents, est légèrement inférieure à celle observée sur la base Lyon-2²¹. La diminution des performances est toutefois assez faible puisque par exemple l'écart entre le Dice calculé sur Lyon-2 et calculé sur Grenoble-2 est

20: 36 volumes de patients avec SDRA sévère, 8 avec SDRA sévère sous ECMO et seulement 6 avec SDRA modéré.

21: Résultat opposé à celui observé dans le Chapitre 4.

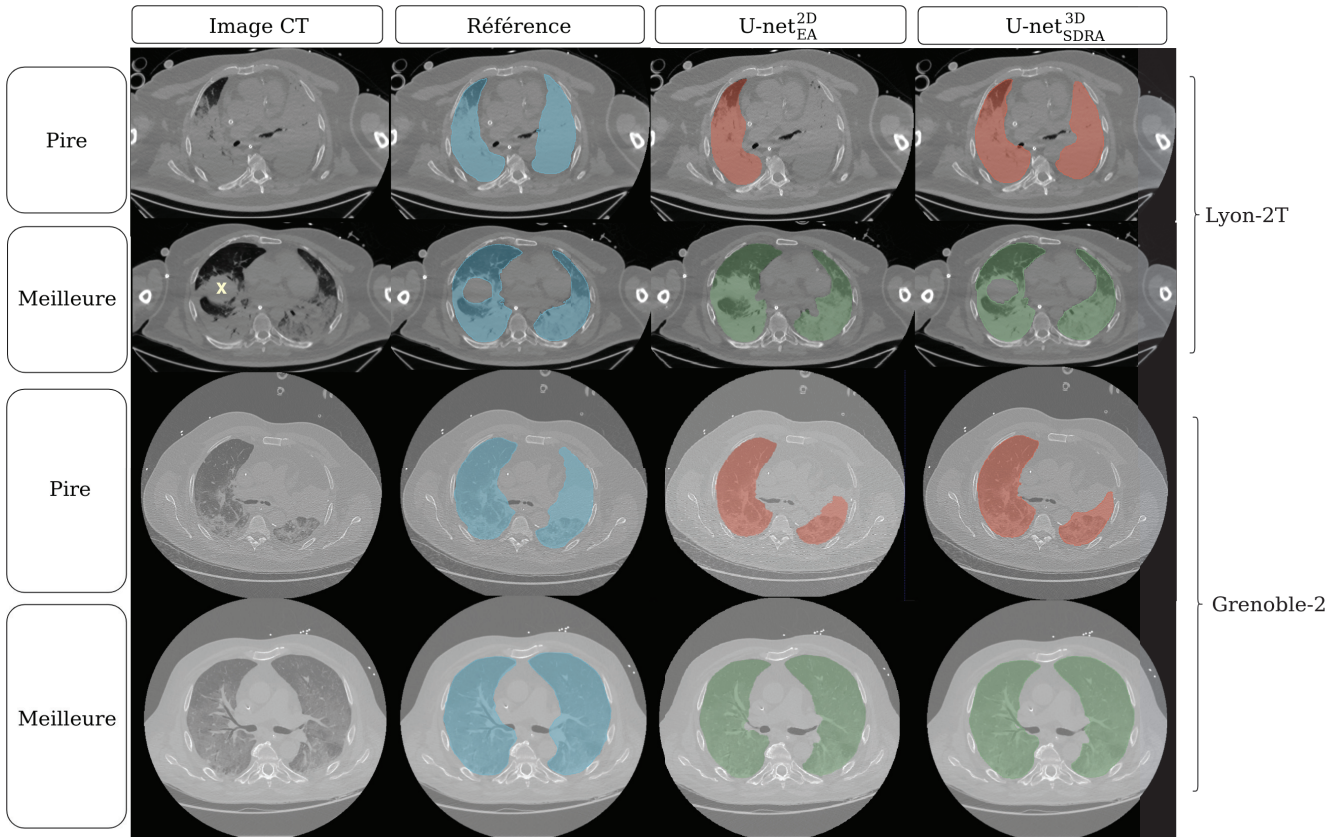


FIGURE 5.12 – Exemples de pires (en rouge) et meilleures (en vert) segmentations fournies par les deux modèles $U\text{-net}_{EA}^{2D}$ et $U\text{-net}_{SDRA}^{3D}$ sur les bases de données Lyon-2T et Grenoble-2. – La croix jaune sur l’image CT à la 2e ligne indique le début du diaphragme.

de 2%. La cohérence et la robustesse ainsi montrées par les performances permettent une reproductibilité appréciable pour une utilisation clinique. Les erreurs qui subsistent sur les segmentations surviennent principalement dans les zones non aérées qui sont sous-segmentées. Des outils de correction sont alors disponibles dans l’interface du logiciel creaSDRA, qui permettent d’interagir avec la surface du masque de segmentation en 3D pour soit enfoncer ou étirer la surface. Ces outils sont encore en cours de développement, pour permettre d’améliorer la manière²² de corriger l’erreur pour l’utilisateur.

Le modèle $U\text{-net}_{animal}$ qui n’a pas pu être évalué dans ce chapitre, faute de nouvelles données annotées, est tout de même déjà utilisé dans le cadre d’autres études expérimentales. Il a ainsi permis de segmenter une centaine d’images de manière automatique sur la durée d’une après-midi.

5.6 Conclusion

Les différents modèles de segmentation produits à l’issue de cette thèse sont transférables pour une utilisation pré-clinique. En effet, le modèle spécialisé à la segmentation des modèles expérimentaux porcins de SDRA a déjà pu être utilisé dans le cadre d’une expérience animale. Le modèle spécialisé pour la segmentation des poumons humains chez des patients atteints du SDRA a aussi été utilisé pour ajuster la ventilation mécanique. En fournissant des informations d’aération auxquels les cliniciens n’ont

22: Qui ne soit pas trop coûteuse en temps au vu du temps gagné par la segmentation automatique.

habituellement pas accès, il a permis au clinicien de personnaliser la ventilation chez des patients admis au service de réanimation.

Malgré les imperfections observées, les performances des modèles sont suffisantes pour qu'il soient déjà utilisés, tant dans le cadre d'études expérimentales sur le modèle porcin que sur des données patients issues d'un protocole de recherche pré-clinique. Notamment dans les zones hyperinflatées ou bien dans les zones non aérées. La présence d'artefact génère également des erreurs.

Objectifs initiaux et état d'avancement		
Cadre d'utilisation	Mise en application en clinique	Atteinte
Tâche	Segmentation des poumons sur image CT 3D	Méthode 3D
Cohérence	Fournit peu de différences entre deux images traitées du même patient	Cohérence de recrutement ou de masse
Interactivité	Automatique	Oui
Délai de calcul	< 1h	Environ 2 s (GPU) ou 10 s (CPU)
Robustesse	Indépendant de la sévérité de SDRA et lésions présentes	<i>Dépendante</i>
Précision	Nécessite peu de retouches manuelles (si présentes)	<i>Erreurs de sous-segmentation à corriger : outils de correction disponibles dans le logiciel creaSDRA</i>

En gras les objectifs atteints dans ce chapitre, en police normale ce qui a été atteint dans les chapitres précédents et en italique ce qu'il reste à faire.

Conclusion

Le travail mené dans cette thèse avait pour but de répondre à un besoin technique dans la compréhension et le traitement du syndrome de détresse respiratoire aiguë. Pour cela, des méthodes d'apprentissage profond ont été mises en place pour la segmentation des poumons dans des images CT.

Ainsi, dans un premier temps, le contexte médical de cette thèse a été détaillé. La description du système respiratoire a permis de comprendre les effets du SDRA sur le corps et ainsi mettre en évidence les points pouvant être améliorés dans sa prise en charge. Les objectifs pratiques de la thèse ont été définis en fonction des attentes médicales et d'un besoin de transfert de la solution proposée vers le milieu clinique. La méthode de segmentation se devait ainsi d'être automatique, rapide²³, robuste et reproductible.

Dans un deuxième temps, un certain nombre de définitions et descriptions du traitement d'images, essentiels à la compréhension²⁴ de cette thèse, ont été donnés. La synthèse des travaux sélectionnés dans la littérature scientifique appliqués à la segmentation des poumons a ensuite été présentée. D'abord du point de vue des méthodes classiques, n'utilisant pas d'apprentissage. Puis l'accent a été mis sur les méthodes d'apprentissage profond supervisé, utilisant notamment des réseaux neuronaux convolutifs, qui fournissent de très bons résultats en segmentation²⁵.

Cet état de l'art a mis en évidence la difficulté de mettre en place une méthode qui soit robuste à tous les types de lésions ou particularités anatomiques rencontrés en imagerie médicale. Notamment dans le cas de l'apprentissage profond pour la segmentation d'images médicales, la quantité de données annotées de même que le volume de chaque image à considérer constituent une limitation à cause du temps nécessaire à l'annotation. De plus, peu de méthodes ont spécifiquement été appliquées à la segmentation des poumons chez les patients atteints de SDRA sévère. Et ce même dans le contexte de la pandémie de COVID-19, qui a attiré l'attention sur ce syndrome auparavant peu traité. En effet, la littérature scientifique liée à la maladie COVID-19 a surtout été concentrée sur la détection précoce²⁶, la segmentation et la classification des lésions ou encore la prédiction de pronostic.

Pourtant, un article paru au début de cette thèse a encouragé l'utilisation de l'apprentissage profond. Les solutions, proposées par les auteurs, aux limites évoquées plus tôt ont été d'une part le recours au transfert d'apprentissage, pour augmenter la quantité de données annotées considérée. Et d'autre part, l'utilisation d'une méthode multi-résolution comme compromis computationnel qui permet de segmenter des images 3D tout en garantissant l'exploitation du contexte 3D dans l'apprentissage. L'architecture proposée, de même que la stratégie d'apprentissage, constituait ainsi l'innovation qui a motivé la direction de cette thèse. Mais plutôt que de chercher à proposer une énième nouvelle architecture au domaine, l'accent s'est porté sur l'utilisation des données. En effet, une tendance dans le domaine de l'apprentissage profond avance que la robustesse d'un modèle réside tout autant dans ses données d'entraînement que dans la complexité

23: Inférieure à une heure pour être compatible avec une utilisation clinique.

24: Adressé à des lecteurs du domaine médical comme du domaine du traitement d'images.

25: Lorsqu'une grande quantité de données annotées est utilisée.

26: Sur des images de patients peu malades.

de son architecture.

La direction choisie pour cette thèse a donc été d'identifier une gestion optimale des données à disposition pour la segmentation des poumons chez les patients atteints du SDRA. Cette idée s'est développée selon trois axes, en deux ou trois dimensions.

Le premier axe étudié dans cette thèse a été d'explorer l'impact de la gestion des données lors de l'apprentissage. Pour cela, une étude sur le transfert d'apprentissage a été réalisée en utilisant ou non des données de modèles expérimentaux porcins de SDRA et des données de patients avec SDRA. La question posée était de savoir si l'ordre de présentation des données pouvait influencer et renforcer la robustesse, en apprenant d'abord sur une anatomie donnée (porcine) puis sur une autre (humaine). Exploré sous le point de vue 2D avec une stratégie multi-coupes²⁷ ou le point de vue 3D²⁸, la conclusion est la même. Le gain en robustesse apporté par le transfert d'apprentissage n'est que peu significatif. Cette observation contredit pourtant la littérature à ce sujet. Ceci peut être expliqué par la différence de quantité de données utilisées pour la spécialisation du modèle. Mais cela renforce alors l'idée que le pré-entraînement n'est pas celui qui apporte le plus de robustesse au modèle, sinon la base de données de spécialisation. La sévérité des SDRA représentés dans les images utilisées dans cette thèse est d'ailleurs à souligner puisqu'elle contribue probablement à la difficulté de gain de robustesse du transfert d'apprentissage.

La diversité des données utilisées lors de l'apprentissage, ainsi que l'information du contexte spatial (3D) sont des éléments déterminants pour la robustesse et la capacité de généralisation des modèles entraînés. Cependant, chacun d'eux est très gourmand en ressources (mémoire et temps de calcul). Le deuxième axe exploré dans cette thèse a donc été l'évaluation du compromis entre la diversité²⁹ et les détails du contexte spatial. Le recours au sous-échantillonnage pour garantir le contexte du scanner thoracique tout en apportant la diversité d'exemples dans les lots d'images vus à l'entraînement a ainsi été évalué. Les expériences menées à différentes résolutions d'images ont ainsi montré la prévalence des détails dans le contexte sur la diversité d'exemples considérés dans l'apprentissage. Les modèles les plus robustes et avec la meilleure capacité de généralisation avaient ainsi peu d'images par lot³⁰ et une haute résolution³¹.

Le troisième et dernier axe évalué dans cette thèse a porté sur l'utilisation d'une information quantitative physiologique durant l'apprentissage. Pour être utilisable dans un contexte clinique, les segmentations produites par un modèle sur les deux images³² d'un même patient se doivent d'être cohérentes avec la référence en termes de quantité de tissus non aérés, mais aussi en termes de masse totale des poumons segmentés³³. Deux types d'information ont alors été utilisés et intégrés à la fonction de perte durant l'apprentissage. D'une part, l'information d'aération sur le compartiment non aéré et d'autre part, le calcul de la masse pulmonaire, grâce à la prédiction d'un masque de segmentation 3D. Ces données, ajoutées à l'information de superposition³⁴ durant l'apprentissage, ont permis de gagner en robustesse sur les compartiments les plus compliqués³⁵ à segmenter. La robustesse globale n'a cependant pas été impactée par ces améliorations, ce qui indique une perte de robustesse sur d'autres parties

27: Dans le Chapitre 3.

28: Dans le Chapitre 4 avec une capacité supérieure au modèle 2D.

29: Au sens du nombre d'images vues par lot.

30: Une ou deux images maximum.

31: 1 ou 2 mm.

32: A l'inspiration et à l'expiration.

33: La masse des poumons est conservée avec une variation estimée à maximum 5% entre l'inspiration et l'expiration.

34: Transcrite par la fonction de perte sur le score de Dice.

35: Peu et non aéré car très peu contrastés.

des poumons. D'autres formes de fonction de perte basée sur l'un des deux aspects étudiés pourraient être évaluées à l'avenir, de même que d'autres critères d'inclusion à la fonction de perte globale.

À partir des apprentissages acquis des expériences décrites jusqu'ici, deux modèles de production ont été proposés. L'un pour la segmentation des poumons de patients atteints du SDRA et l'autre pour la segmentation des modèles expérimentaux porcins de SDRA. Ces deux modèles ont été intégrés à un logiciel de segmentation semi-automatique d'analyse des paramètres d'aération utilisé à l'hôpital de la Croix-Rousse. Le modèle spécialisé à la segmentation des poumons humains a pu être évalué sur des données de test, et comparé à un modèle disponible de l'état de l'art. De ces résultats, il apparaît que le modèle proposé est plus robuste et celui qui généralise mieux. Cela notamment sur les patients dont le SDRA est le plus sévère, comme ceux sous ECMO. D'ailleurs, il ne semble pas y avoir eu de tentatives de segmenter ce type d'images dans la littérature, la tâche de segmentation (manuelle ou automatique) étant compliquée et très chronophage.

Le logiciel de segmentation sur lequel les modèles d'apprentissage profond de production ont été intégrés propose des outils interactifs permettant la correction des erreurs de segmentation automatique qui peuvent subsister. Cela survient notamment dans les zones hyperinflatées³⁶ ou non aérées qui peuvent être sous-segmentées.

L'utilisation du modèle proposé dans un contexte pré-clinique permettra aux cliniciens d'obtenir des informations jusqu'alors jamais exploitées en temps réel, comme le recrutement pour le réglage de la PEP. Les perspectives de ce travail sont tout d'abord la réalisation d'une étude observationnelle visant à décrire les anomalies scanographiques sur une large population de SDRA tout en adaptant les mesures scanner au contexte multicentrique. Dans un deuxième temps, il est envisagé d'étudier l'impact d'une ventilation dont les réglages seraient assistés par l'imagerie CT sur le pronostic du SDRA.

36: Qui est le compartiment le moins fréquent dans les images.

Liste des communications

Articles dans des revues internationales à comité de lecture

Ludmilla PENARRUBIA, Aude VERSTRAETE, Maciej ORKISZ, Eduardo Enrique DÁVILA SERRANO, Loic BOUSSEL, Hodane YONIS, Mehdi MEZIDI, Francois DHELFT, William DANJOU, Alwin BAZZANI, Florian SIGAUD, Sam BAYAT, Nicolas TERZI, Mehdi GIRARD, Laurent BITKER, Emmanuel ROUX et Jean-Christophe RICHARD, « Precision of CT-derived alveolar recruitment assessed by human observers and a machine learning algorithm in moderate and severe ARDS », *Intensive Care Medicine Experimental*, 2023.

Ludmilla PENARRUBIA, Nicolas PINON, Emmanuel ROUX, Eduardo Enrique DÁVILA SERRANO, Jean-Christophe RICHARD, Maciej ORKISZ et David SARRUT, « Improving motion-mask segmentation in thoracic CT with multiplanar U-nets », *Medical Physics*, 2021.

Louis CHAUVELOT, Laurent BITKER, François DHELFT, Mehdi MEZIDI, Maciej ORKISZ, Eduardo Enrique DÁVILA SERRANO, Ludmilla PENARRUBIA, Hodane YONIS, Paul CHABERT, Laure FOLLIET, Guillaume DAVID, Judith PROVOOST, Pierre LECAM, Loic BOUSSEL et Jean-Christophe RICHARD, « Quantitative-analysis of computed tomography in COVID-19 and non COVID-19 ARDS patients : A case-control study », *Journal of Critical Care*, 2020.

Présentations orales avec poster

Ludmilla PENARRUBIA, Emmanuel ROUX, Maciej ORKISZ et Jean-Christophe RICHARD, « Innovative loss function on pathological lung segmentation robustness », Journée de partenariat du LabEx PRIMES, 2022.

Présentations orales

Ludmilla PENARRUBIA, Nicolas PINON, Emmanuel ROUX, David SARRUT, Maciej ORKISZ et Jean-Christophe RICHARD, « Motion mask segmentation with deep learning approaches », eJEDISS 25^e Journée scientifique de l'EDISS, 2020.

Bibliographie

- [Alt+83] D. G. ALTMAN et J. M. BLAND. « Measurement in Medicine : The Analysis of Method Comparison Studies on JSTOR ». In : *Journal of the Royal Statistical Society. Series D (The Statistician)* 32.3 (sept. 1983), p. 307-317 (cf. p. 66, 89).
- [Ant+21] Michela ANTONELLI, Annika REINKE, Spyridon BAKAS, Keyvan FARAHANI, ANNETTEKOPP-SCHNEIDER, Bennett A. LANDMAN, Geert LITJENS, Bjoern MENZE, Olaf RONNEBERGER, Ronald M. SUMMERS, Bram van GINNEKEN, Michel BILELLO, Patrick BILIC, Patrick F. CHRIST, Richard K. G. DO, Marc J. GOLLUB, Stephan H. HECKERS, Henkjan HUISMAN, William R. JARNAGIN, Maureen K. McHUGO, Sandy NAPEL, Jennifer S. Goli PERNICKA, Kawal RHODE, Catalina TOBON-GOMEZ, Eugene VORONTSOV, Henkjan HUISMAN, James A. MEAKIN, Sebastien OURSELIN, Manuel WIESENFARTH, Pablo ARBELAEZ, Byeonguk BAE, Sihong CHEN, Laura DAZA, Jianjiang FENG, Baochun HE, Fabian ISENSEE, Yuanfeng JI, Fucang JIA, Namkug KIM, Ildoo KIM, Dorit MERHOF, Akshay PAI, Beomhee PARK, Mathias PERSLEV, Ramin REZAEIFAR, Oliver RIPPEL, Ignacio SARASUA, Wei SHEN, Jaemin SON, Christian WACHINGER, Liansheng WANG, Yan WANG, Yingda XIA, Daguang XU, Zhanwei XU, Yefeng ZHENG, Amber L. SIMPSON, Lena MAIER-HEIN et M. Jorge CARDOSO. « The Medical Segmentation Decathlon ». In : *arXiv* (juin 2021). DOI : [10.48550/arXiv.2106.05735](https://doi.org/10.48550/arXiv.2106.05735) (cf. p. 30).
- [Asg+21] Saeid ASGARI TAGHANAKI, Kumar ABHISHEK, Joseph Paul COHEN, Julien COHEN-ADAD et Ghassan HAMARNEH. « Deep semantic segmentation of natural and medical images : a review ». In : *Artificial Intelligence Review* 54.1 (2021), p. 137-178. DOI : [10.1007/s10462-020-09854-1](https://doi.org/10.1007/s10462-020-09854-1) (cf. p. 29).
- [Ash+67] DavidG. ASHBAUGH, D. BOYD BIGELOW, ThomasL. PETTY et BernardE. LEVINE. « Acute respiratory distress in adults ». In : *The Lancet* 290.7511 (1967), p. 319-323. DOI : [10.1016/S0140-6736\(67\)90168-7](https://doi.org/10.1016/S0140-6736(67)90168-7) (cf. p. 8).
- [Ast+22] Joshua R ASTLEY, Jim M WILD et Bilal A TAHIR. « Deep learning in structural and functional lung image analysis ». In : *The British Journal of Radiology* 95.1132 (avr. 2022). DOI : [10.1259/BJR.20201107](https://doi.org/10.1259/BJR.20201107) (cf. p. 24, 28, 29).
- [Bar22] Carmen Silvia Valente BARBAS. « Thoracic Computed Tomography to Assess ARDS and COVID-19 Lungs ». In : *Front. Physiol.* 0 (2022). DOI : [10.3389/fphys.2022.829534](https://doi.org/10.3389/fphys.2022.829534) (cf. p. 13, 14).
- [Bei+16] Jeremy R. BEITLER, Ewan C. GOLIGHER, Matthieu SCHMIDT, Peter M. SPIETH, Alberto ZANELLA, Ignacio MARTIN-LOECHES, Carolyn S. CALFEE et Alexandre B. CAVALCANTI. « Personalized medicine for ARDS : the 2035 research agenda ». In : *Intensive Care Medicine* 42.5 (2016), p. 756-767. DOI : [10.1007/s00134-016-4331-6](https://doi.org/10.1007/s00134-016-4331-6) (cf. p. 10, 11).
- [Bel+16] Giacomo BELLANI, John G. LAFFEY, Tàì PHAM, Eddy FAN, Laurent BROCHARD, Andres ESTEBAN, Luciano GATTINONI, Frank van HAREN, Anders LARSSON, Daniel F. MCAULEY, Marco RANIERI, Gordon RUBENFELD, B. Taylor THOMPSON, Hermann WRIGGE, Arthur S. SLUTSKY et Antonio PESENTI. « Epidemiology, Patterns of Care, and Mortality for Patients With Acute Respiratory Distress Syndrome in Intensive Care Units in 50 Countries ». In : *JAMA* 315.8 (fév. 2016), p. 788. DOI : [10.1001/jama.2016.0291](https://doi.org/10.1001/jama.2016.0291) (cf. p. 9).
- [Bel+21] François BELONCLE, Antoine STUDER, Valerie SEEGER, Jean-Christophe RICHARD, Christophe DESPREZ, Nicolas FAGE, Hamid MERDJI, Bertrand PAVLOVSKY, Julie HELMS, Sibylle CUNAT, Satar MORTAZA, Julien DEMISELLE, Laurent BROCHARD, Alain MERCAT et Ferhat MEZIANI. « Longitudinal changes in compliance, oxygenation and ventilatory ratio in COVID-19 versus non-COVID-19 pulmonary acute respiratory distress syndrome ». In : *Crit. Care* 25 (déc. 2021). DOI : [10.1186/s13054-021-03665-8](https://doi.org/10.1186/s13054-021-03665-8) (cf. p. 14).
- [Ben+18] I. BENDIB, F. SCHLEMMER, B. MAITRE, G. CARTEAUX, K. RAZAZI, A. Mekontso DESSAP et N. DE PROST. « Acute respiratory failure : The ARDS and beyond ». In : *Medecine Intensive Reanimation* 27.1 (2018), p. 25-35. DOI : [10.3166/rea-2018-0002](https://doi.org/10.3166/rea-2018-0002) (cf. p. 9).

- [Bla+90] W.E. BLANZ et S.L. GISH. « A connectionist classifier architecture applied to image segmentation ». In : *Proceedings. 10th International Conference on Pattern Recognition*. T. ii. IEEE Comput. Soc. Press, 1990, p. 272-277. DOI : [10.1109/ICPR.1990.119369](https://doi.org/10.1109/ICPR.1990.119369) (cf. p. 26).
- [Bol+95] Philippe BOLON, Jean-Marc CHASSERY, Jean-Pierre COCQUEREZ, Didier DEMIGNY, Christine GRAFFIGNE, Annick MONTANVERT, Sylvie PHILIPP, Rachid ZÉBOUJ, Josiane ZERUBIA et Henri MAÎTRE. *Analyse d'images : Filtrage et segmentation*. Sous la dir. de Jean-Pierre COCQUEREZ et Sylvie PHILIPP. Enseignement de la physique. Ouvrage publié avec l'aide du Ministère des affaires étrangères, direction de la coopération scientifique et technique. AVERTISSEMENT Le livre publié en 1995 chez MASSON (EAN13 : 9782225849237) est épuisé. Cette version pdf est une version élaborée à partie de la version préliminaire transmise à l'éditeur. La mise en page est légèrement différente de celle du livre. Malheureusement quelques figures de l'annexe C ont été perdues. MASSON, oct. 1995 (cf. p. 16).
- [Boy+98] Y. BOYKOV, O. VEKSLER et R. ZABIH. « Markov random fields with efficient approximations ». In : *Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No.98CB36231)*. 1998, p. 648-655. DOI : [10.1109/CVPR.1998.698673](https://doi.org/10.1109/CVPR.1998.698673) (cf. p. 24).
- [Bri+70] Claude R. BRICE et Claude L. FENNEMA. « Scene analysis using regions ». In : *Artificial Intelligence* 1.3-4 (1970), p. 205-226. DOI : [10.1016/0004-3702\(70\)90008-1](https://doi.org/10.1016/0004-3702(70)90008-1) (cf. p. 24).
- [Car+18] L. E. CARVALHO, A. C. SOBIERANSKI et A. VON WANGENHEIM. « 3D Segmentation Algorithms for Computerized Tomographic Imaging : a Systematic Literature Review. » In : *Journal of digital imaging* 31.6 (2018), p. 799-850. DOI : [10.1007/s10278-018-0101-z](https://doi.org/10.1007/s10278-018-0101-z) (cf. p. 24, 25).
- [Cha+17] Jung won CHA, M. Mehdi FARHANGI, Neal DUNLAP et Amir AMINI. « Volumetric analysis of respiratory gated whole lung and liver CT data with motion-constrained graph cuts segmentation ». In : *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2017, p. 3405-3408. DOI : [10.1109/EMBC.2017.8037587](https://doi.org/10.1109/EMBC.2017.8037587) (cf. p. 24).
- [Cha+20] Louis CHAUVELOT, Laurent BITKER, François DHELFT, Mehdi MEZIDI, Maciej ORKISZ, Eduardo DAVILA SERRANO, Ludmilla PENARRUBIA, Hodane YONIS, Paul CHABERT, Laure FOLLIET, Guillaume DAVID, Judith PROVOOST, Pierre LECAM, Loic BOUSSEL et Jean-Christophe RICHARD. « Quantitative-analysis of computed tomography in COVID-19 and non COVID-19 ARDS patients : A case-control study ». In : *J. Crit. Care* 60 (août 2020). DOI : [10.1016/j.jcrc.2020.08.006](https://doi.org/10.1016/j.jcrc.2020.08.006) (cf. p. 14, 38).
- [Che+14] Liang-Chieh CHEN, George PAPANDEOU, Iasonas KOKKINOS, Kevin MURPHY et Alan L. YUILLE. « Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs ». In : *arXiv* (déc. 2014). DOI : [10.48550/arXiv.1412.7062](https://doi.org/10.48550/arXiv.1412.7062) (cf. p. 29).
- [Che+16] Liang-Chieh CHEN, George PAPANDEOU, Iasonas KOKKINOS, Kevin MURPHY et Alan L. YUILLE. « DeepLab : Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs ». In : *arXiv* (juin 2016). DOI : [10.48550/arXiv.1606.00915](https://doi.org/10.48550/arXiv.1606.00915) (cf. p. 29).
- [Che+17] Liang-Chieh CHEN, George PAPANDEOU, Florian SCHROFF et Hartwig ADAM. « Rethinking Atrous Convolution for Semantic Image Segmentation ». In : *arXiv* (juin 2017). DOI : [10.48550/arXiv.1706.05587](https://doi.org/10.48550/arXiv.1706.05587) (cf. p. 29).
- [Che+18] Liang-Chieh CHEN, Yukun ZHU, George PAPANDEOU, Florian SCHROFF et Hartwig ADAM. « Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation ». In : *arXiv* (fév. 2018). DOI : [10.48550/arXiv.1802.02611](https://doi.org/10.48550/arXiv.1802.02611) (cf. p. 29).
- [Çiç+16] Özgün ÇIÇEK, Ahmed ABDULKADIR, Soeren S. LIENKAMP, Thomas BROX et Olaf RONNEBERGER. « 3D U-Net : Learning Dense Volumetric Segmentation from Sparse Annotation ». In : *CoRR* abs/1606.06650 (2016) (cf. p. 30, 36, 54, 63, 88).
- [CIS15] CISMÉF. *Le syndrome de détresse respiratoire aiguë*. 2015. URL : https://www.enseignementsup-recherche.gouv.fr/ressources-pedagogiques/notice/view/oai%7B%5C%7D253Adocumvf.crihan.fr%7B%5C%7D253Aumvf%7B%5C%7D252Fcnera%7B%5C_%7Dfr%7B%5C_%7Dmanuel%7B%5C_%7DD0C%7B%5C_%7D310362%7B%5C%7Dpresentation-ressource-pedagogique (cf. p. 9).

- [Cop+19] Silvia COPPOLA, Sara FROIO, Antonella MARINO, Matteo BRIONI, Bruno Mario CESANA, Massimo CRESSONI, Luciano GATTINONI et Davide CHIUMELLO. « Respiratory Mechanics, Lung Recruitability, and Gas Exchange in Pulmonary and Extrapulmonary Acute Respiratory Distress Syndrome ». In : *Critical Care Medicine* 47.6 (2019), p. 792-799. DOI : [10.1097/CCM.0000000000003715](https://doi.org/10.1097/CCM.0000000000003715) (cf. p. 11, 12).
- [Dáv+20] Eduardo Enrique DÁVILA SERRANO, François DHELFT, Laurent BITKER, Jean-Christophe RICHARD et Maciej ORKISZ. « Software for CT-image Analysis to Assist the Choice of Mechanical-Ventilation Settings in Acute Respiratory Distress Syndrome ». In : *Computer Vision and Graphics*. Cham, Switzerland : Springer, sept. 2020, p. 48-58. DOI : [10.1007/978-3-030-59006-2_5](https://doi.org/10.1007/978-3-030-59006-2_5) (cf. p. 25, 38, 56, 57, 89).
- [De +22] Matthias DE LANGE, Rahaf ALJUNDI, Marc MASANA, Sarah PARISOT, Xu JIA, Aleš LEONARDIS, Gregory SLABAUGH et Tinne TUYTELAARS. « A Continual Learning Survey : Defying Forgetting in Classification Tasks ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.7 (juill. 2022). Conference Name : IEEE Transactions on Pattern Analysis and Machine Intelligence, p. 3366-3385. DOI : [10.1109/TPAMI.2021.3057446](https://doi.org/10.1109/TPAMI.2021.3057446) (cf. p. 54).
- [Dia+19] M DIAMOND, HL PENISTON FELICIANO et S MAHAPATRA. « Acute Respiratory Distress Syndrome ». In : *StatPearls Publishing*. StatPearls Publishing, 2019 (cf. p. 9).
- [Dia+21] Matthew DIAMOND, Hector L. PENISTON, Devang SANGHAVI et Sidharth MAHAPATRA. *Acute Respiratory Distress Syndrome*. StatPearls Publishing, Treasure Island (FL), 2021 (cf. p. 9).
- [Dup15] Sophie DUPONT. *L'anatomie et la physiologie pour les infirmiers*. 2^e éd. Elsevier-Masson, 2015 (cf. p. 4).
- [Fan+18] Eddy FAN, Daniel BRODIE et Arthur S. SLUTSKY. « Acute respiratory distress syndrome advances in diagnosis and treatment ». In : *JAMA - Journal of the American Medical Association* 319.7 (fév. 2018), p. 698-710. DOI : [10.1001/JAMA.2017.21907](https://doi.org/10.1001/JAMA.2017.21907) (cf. p. 10).
- [Fan+13] Vito FANELLI, Aikaterini VLACHOU, Shirin GHANNADIAN, Umberto SIMONETTI, Arthur S. SLUTSKY et Haibo ZHANG. « Acute respiratory distress syndrome : New definition, current and future therapeutic options ». In : *Journal of Thoracic Disease* 5.3 (2013), p. 326-334. DOI : [10.3978/j.issn.2072-1439.2013.04.05](https://doi.org/10.3978/j.issn.2072-1439.2013.04.05) (cf. p. 9).
- [Fer+21] Shannon M FERNANDO, Bruno L FERREYRO, Martin URNER, Laveena MUNSHI et Eddy FAN. *Diagnostic et traitement du syndrome de détresse respiratoire aiguë*. 2021. DOI : [10.1503/CMAJ.202661-F](https://doi.org/10.1503/CMAJ.202661-F). URL : www.cmaj.ca/lookup/doi/10.1503/cmaj.202661 (cf. p. 10, 14).
- [Gar+21] Marc GARNIER, Christophe QUESNEL et Jean-Michel CONSTANTIN. « Atteintes pulmonaires liées à la COVID-19 ». In : *La Presse Médicale Formation* 2.1, Part 1 (fév. 2021), p. 14-24. DOI : [10.1016/j.lpmfor.2020.12.009](https://doi.org/10.1016/j.lpmfor.2020.12.009) (cf. p. 14).
- [Gat+01] Luciano GATTINONI, Pierto CAIRONI, Paole PELOSE et Lawrence R. GOODMAN. « What Has Computed Tomography Taught Us about the Acute Respiratory Distress Syndrome ? » In : *American Journal of Respiratory and Critical Care Medicine* 164.9 (2001), p. 1701-1711. DOI : [10.1164/ajrccm.164.9.2103121](https://doi.org/10.1164/ajrccm.164.9.2103121) (cf. p. 12, 13, 61).
- [Gat+20] Luciano GATTINONI, Silvia COPPOLA, Massimo CRESSONI, Mattia BUSANA, Sandra ROSSI et Davide CHIUMELLO. « COVID-19 Does Not Lead to a “Typical” Acute Respiratory Distress Syndrome ». In : *Am. J. Respir. Crit. Care Med.* (mai 2020) (cf. p. 14).
- [Gat+10] Luciano GATTINONI et Massimo CRESSONI. « Quantitative CT in ARDS : towards a clinical tool ? » In : *Intensive Care Medicine* 36.11 (nov. 2010), p. 1803-1804. DOI : [10.1007/s00134-010-2016-0](https://doi.org/10.1007/s00134-010-2016-0) (cf. p. 13).
- [Ger18] Sarah E GERARD. « Multi-scale convolutional neural networks for segmentation of pulmonary structures in computed tomography ». Thèse de doct. Iowa City, Iowa, USA : University of Iowa, déc. 2018, p. 193. DOI : [10.17077/etd.jgiv-7i2q](https://doi.org/10.17077/etd.jgiv-7i2q) (cf. p. 31-33, 52, 61).
- [Ger+19] Sarah E. GERARD, Jacob HERRMANN, David W. KACZKA, Guido MUSCH, Ana FERNANDEZ-BUSTAMANTE et Joseph M. REINHARDT. « Multi-Resolution Convolutional Neural Networks for Fully Automated Segmentation of Acutely Injured Lungs in Multiple Species ». In : *Medical Image Analysis* (2019), p. 101592. DOI : [10.1016/j.media.2019.101592](https://doi.org/10.1016/j.media.2019.101592) (cf. p. 31, 35, 36).

- [Gin21] Bram van GINNEKEN. « LOLA11 ». In : *Zenodo* (avr. 2021). DOI : [10.5281/zenodo.4708800](https://doi.org/10.5281/zenodo.4708800) (cf. p. 33).
- [Gon+07] Rafael C. GONZALEZ et Richard E. WOODS. *Digital Image Processing*. London, England, UK : Pearson, 2007 (cf. p. 16).
- [Goo+16] Ian GOODFELLOW, Yoshua BENGIO et Aaron COURVILLE. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016 (cf. p. 26).
- [Gor+12] Vladlena GORBUNOVA, Jon SPORRING, Pechin LO, Martine LOEVE, Harm A. TIDDENS, Mads NIELSEN, Asger DIRKSEN et Marleen de BRUIJNE. « Mass preserving image registration for lung CT ». In : *Medical Image Analysis* 16.4 (2012), p. 786-795. DOI : [10.1016/j.media.2011.11.001](https://doi.org/10.1016/j.media.2011.11.001) (cf. p. 62).
- [Gor+13] J. GORDON BETTS, Peter DESAIX, Eddie JONHSON, Jody E. JONHSON, Oksana KOROL, Dean KRUSE, Brandon POE, James A. WISE, Mark WOMBLE et Kelly A. YOUNG. *Neuroanatomy and physiology*. Pdf versio. OpenStax, 2013, p. 1420 (cf. p. 4).
- [Her+21] Peter HERRMANN, Mattia BUSANA, Massimo CRESSONI, Joachim LOTZ, Onnen MOERER, Leif SAAGER, Konrad MEISSNER, Michael QUINTEL et Luciano GATTINONI. « Using Artificial Intelligence for Automatic Segmentation of CT Lung Images in Acute Respiratory Distress Syndrome ». In : *Frontiers in Physiology* 12.9 (2021), p. 1-16. DOI : [10.3389/fphys.2021.676118](https://doi.org/10.3389/fphys.2021.676118) (cf. p. 78, 85).
- [Her+02] PNXP HERRMANN, XP NGUYEN, T LUECKE et M QUINTEL. « MALUNA 1.03 ein Softwaretool zur analyse computertomographischer Schnittbilder der Lunge ». In : *Virtuelle instrumente in der praxis*. Huethig Verlag, Heidelberg (2002), p. 389-395 (cf. p. 25).
- [Hof+20] Johannes HOFMANNINGER, Florian PRAYER, Jeanny PAN, Sebastian RÖHRICH, Helmut PROSCH et Georg LANGS. « Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem ». In : *European Radiology Experimental* 4.1 (2020). DOI : [10.1186/s41747-020-00173-2](https://doi.org/10.1186/s41747-020-00173-2) (cf. p. 28, 33, 55, 88).
- [Hu+01] Shiyang HU, E.A. HOFFMAN et J.M. REINHARDT. « Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images ». In : *IEEE Transactions on Medical Imaging* 20.6 (2001), p. 490-498. DOI : [10.1109/42.929615](https://doi.org/10.1109/42.929615) (cf. p. 24).
- [Iof+15] Sergey IOFFE et Christian SZEGEDY. « Batch Normalization : Accelerating Deep Network Training by Reducing Internal Covariate Shift ». In : *Proceedings of the 32nd International Conference on Machine Learning*. Sous la dir. de Francis BACH et David BLEI. T. 37. Proceedings of Machine Learning Research. Lille, France : PMLR, juill. 2015, p. 448-456 (cf. p. 40).
- [Ise+21] Fabian ISENSEE, Paul F. JAEGER, Simon A. A. KOHL, Jens PETERSEN et Klaus H. MAIER-HEIN. « nnU-Net : a self-configuring method for deep learning-based biomedical image segmentation ». In : *Nat. Methods* 18.2 (fév. 2021), p. 203-211. DOI : [10.1038/s41592-020-01008-z](https://doi.org/10.1038/s41592-020-01008-z) (cf. p. 30, 55).
- [Jav+16] Muzzamil JAVAID, Moazzam JAVAID, Muhammad Zia Ur REHMAN et Syed Irtiza Ali SHAH. « A novel approach to CAD system for the detection of lung nodules in CT images ». In : *Computer Methods and Programs in Biomedicine* 135 (2016), p. 125-139. DOI : [10.1016/j.cmpb.2016.07.031](https://doi.org/10.1016/j.cmpb.2016.07.031) (cf. p. 24).
- [Kat64] Yale H. KATZ. « Automatic pattern recognition of meteorological satellite cloud photography ». In : *RAND Corporation* (1964) (cf. p. 24).
- [Kla+17] Philipp KLAPSING, Peter HERRMANN, Michael QUINTEL et Onnen MOERER. « Automatic quantitative computed tomography segmentation and analysis of aerated lung volumes in acute respiratory distress syndrome—A comparative diagnostic study ». In : *Journal of Critical Care* 42 (2017), p. 184-191. DOI : [10.1016/j.jcrc.2016.11.001](https://doi.org/10.1016/j.jcrc.2016.11.001) (cf. p. 25).
- [Li+03] Baojun LI, Gary E. CHRISTENSEN, Eric A. HOFFMAN, Geoffrey McLENNAN et Joseph M. REINHARDT. « Establishing a Normative Atlas of the Human Lung : Intersubject Warping and Registration of Volumetric CT Images ». In : *Acad. Radiol.* 10.3 (mars 2003), p. 255-265. DOI : [10.1016/S1076-6332\(03\)80099-5](https://doi.org/10.1016/S1076-6332(03)80099-5) (cf. p. 25).
- [Lit+17] Geert LITJENS, Thijs KOOI, Babak Ehteshami BEJNORDI, Arnaud Arindra Adiyoso SETIO, Francesco CIOMPI, Mohsen GHAFORIAN, Jeroen A.W.M. van der LAAK, Bram van GINNEKEN et Clara I. SÁNCHEZ. « A survey on deep learning in medical image analysis ». In : *Medical Image Analysis* 42.1995 (2017), p. 60-88. DOI : [10.1016/j.media.2017.07.005](https://doi.org/10.1016/j.media.2017.07.005) (cf. p. 29, 36).

- [Lod+20] B. LODÉ, C. JALABER, T. ORCEL, T. MORCET-DELATTRE, N. CRESPIN, S. VOISIN, F. BILLARD, S. LUZI, T. LAPOTRE, P.-A. LENTZ, M.-P. REVEL et M. LEDERLIN. « Imagerie de la pneumonie COVID-19 ». In : *Journal D'Imagerie Diagnostique et Interventionnelle* 3.4 (sept. 2020), p. 249. DOI : [10.1016/j.jidi.2020.04.011](https://doi.org/10.1016/j.jidi.2020.04.011) (cf. p. 14).
- [Lon+15] Jonathan LONG, Evan SHELHAMER et Trevor DARRELL. *Fully convolutional networks for semantic segmentation*. IEEE Computer Society, juin 2015 (cf. p. 29).
- [Lyk+15] Mark LYKSBERG, Oula PUONTI, Mikael AGN et Rasmus LARSEN. « An Ensemble of 2D Convolutional Neural Networks for Tumor Segmentation ». In : *Proceedings of the 19th Scandinavian Conference on Image Analysis, SCIA 2015*. Berlin, Germany : Springer, 2015, p. 201-211. DOI : [10.1007/978-3-319-19665-7_17](https://doi.org/10.1007/978-3-319-19665-7_17) (cf. p. 36, 37).
- [Mah+20] Amin MAHSOULI, Melissa GRILLO, Nadia AMINI, Souad ACID, Emmanuel COCHE et Benoît GHAYE. « Imagerie thoracique du COVID-19 ». In : *Louvain Médical* (2020), p. 360-367 (cf. p. 14).
- [Mai+22] Lorenzo MAIELLO, Lorenzo BALL, Marco MICALI, Francesca IANNUZZI, Nico SCHERF, Ralf-Thorsten HOFFMANN, Marcelo Gama de ABREU, Paolo PELOSI et Robert HUHLE. « Automatic Lung Segmentation and Quantification of Aeration in Computed Tomography of the Chest Using 3D Transfer Learning ». In : *Front. Physiol.* 12 (fév. 2022), p. 725865. DOI : [10.3389/fphys.2021.725865](https://doi.org/10.3389/fphys.2021.725865) (cf. p. 78).
- [Mal+22] Priyanka MALHOTRA, Sheifali GUPTA, Deepika KOUNDAL, Atef ZAGUIA et Wegayehu ENBEYLE. « Deep Neural Networks for Medical Image Segmentation ». In : *Journal of Healthcare Engineering* (2022). DOI : [10.1155/2022/9580991](https://doi.org/10.1155/2022/9580991) (cf. p. 30).
- [Man+15] Awais MANSOOR, Ulas BAGCI, Brent FOSTER, Ziyue XU, Georgios Z. PAPADAKIS, Les R. FOLIO, Jayaram K. UDUPA et Daniel J. MOLLURA. « Segmentation and Image Analysis of Abnormal Lungs at CT : Current Approaches, Challenges, and Future Trends ». In : *RadioGraphics* 35.4 (2015), p. 1056-1076. DOI : [10.1148/rg.2015140232](https://doi.org/10.1148/rg.2015140232) (cf. p. 24-26).
- [McK+16] Andrew C. McKOWN et Lorraine B. WARE. « Quantification of lung recruitment by respiratory mechanics and CT imaging : what are the clinical implications? ». In : *Annals of Translational Medicine* 4.7 (avr. 2016), p. 20. DOI : [10.21037/atm.2016.03.20](https://doi.org/10.21037/atm.2016.03.20) (cf. p. 13).
- [Men+12] Xin MENG, Yongqian QIANG, Shaocheng ZHU, Carl FUHRMAN, Jill M. SIEGFRIED et Jiantao PU. « Illustration of the obstacles in computerized lung segmentation using examples ». In : *Medical Physics* 39.8 (2012), p. 4984-4991. DOI : [10.1118/1.4737023](https://doi.org/10.1118/1.4737023) (cf. p. 24, 25).
- [Mey+21] Nuala J. MEYER, Luciano GATTINONI et Carolyn S. CALFEE. « Acute respiratory distress syndrome ». In : *The Lancet* 398.10300 (août 2021), p. 622-637. DOI : [10.1016/S0140-6736\(21\)00439-6](https://doi.org/10.1016/S0140-6736(21)00439-6) (cf. p. 9-11).
- [Mil+16] Fausto MILLETARI, Nassir NAVAB et Seyed-Ahmad AHMADI. « V-Net : Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation ». In : *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, oct. 2016, p. 565-571. DOI : [10.1109/3DV.2016.79](https://doi.org/10.1109/3DV.2016.79) (cf. p. 30, 54).
- [Min+21] Shervin MINAEI, Yuri Y. BOYKOV, Fatih PORIKLI, Antonio J. PLAZA, Nasser KEHTARNAVAZ et Demetri TERZOPOULOS. « Image Segmentation Using Deep Learning : A Survey ». In : *IEEE Trans. Pattern Anal. Mach. Intell.* PP. (Fév. 2021). DOI : [10.1109/TPAMI.2021.3059968](https://doi.org/10.1109/TPAMI.2021.3059968) (cf. p. 28).
- [Moe+16] Pim MOESKOPS, Jelmer M. WOLTERINK, Bas H. M. van der VELDEN, Kenneth G. A. GILHUIJS, Tim LEINER, Max A. VIERGEVER et Ivana IŠGUM. « Deep Learning for Multi-task Medical Image Segmentation in Multiple Modalities ». In : *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Cham, Switzerland : Springer, oct. 2016, p. 478-486. DOI : [10.1007/978-3-319-46723-8_55](https://doi.org/10.1007/978-3-319-46723-8_55) (cf. p. 36, 37).
- [Moh+20] Youssoufa MOHAMADOU, Aminou HALIDOU et Pascaline Tiam KAPEN. « A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19 ». In : *Applied Intelligence* 50.11 (nov. 2020), p. 3913-3925. DOI : [10.1007/S10489-020-01770-9](https://doi.org/10.1007/S10489-020-01770-9) (cf. p. 32).

- [Mon+19] Silvia MONGODI, Silvia BONAITI, Andrea STELLA, Andrea COLOMBO, Erminio SANTANGELO, Rosanna VASCHETTO, Anita ORLANDO, Bélaïd BOUHEMAD et Francesco MOJOLI. « Lung Ultrasound for Daily Monitoring and Management of ARDS Patients ». In : *Clinical Pulmonary Medicine* 26.3 (2019), p. 92-97. DOI : [10.1097/CPM.0000000000000311](https://doi.org/10.1097/CPM.0000000000000311) (cf. p. 12).
- [Mor16] Alfredo MORALES PINZON. « Lung segmentation and airway tree matching : Application to aeration quantification in CT images of subjects with ARDS ». Thèse de doct. Universidad de los Andes, 2016, p. 140 (cf. p. 25, 35).
- [Mor+17] Alfredo MORALES PINZÓN, Maciej ORKISZ, Jean-Christophe RICHARD et Marcela HERNÁNDEZ HOYOS. « Lung Segmentation by Cascade Registration ». In : *Irbm* 38.5 (2017), p. 266-280. DOI : [10.1016/j.irbm.2017.07.003](https://doi.org/10.1016/j.irbm.2017.07.003) (cf. p. 35).
- [Net11] Frank Henry NETTER. *Atlas d'anatomie humaine*. 5^e éd. Elsevier Masson, juill. 2011 (cf. p. 4).
- [Nie+17] Gary F. NIEMAN, Joshua SATALIN, Penny ANDREWS, Hani AIASH, Nader M. HABASHI et Louis A. GATTO. « Personalizing mechanical ventilation according to physiologic parameters to stabilize alveoli and minimize ventilator induced lung injury (VILI) ». In : *Intensive Care Medicine Experimental* 5.1 (déc. 2017), p. 8. DOI : [10.1186/s40635-017-0121-x](https://doi.org/10.1186/s40635-017-0121-x) (cf. p. 9-12).
- [Okt+18] Ozan OKTAY, Jo SCHLEMPER, Loïc Le FOLGOC, Matthew LEE, Mattias HEINRICH, Kazunari MISAWA, Kensaku MORI, Steven McDONAGH, Nils Y HAMMERLA, Bernhard KAINZ et al. « Attention u-net : Learning where to look for the pancreas ». In : *arXiv preprint arXiv :1804.03999* (2018) (cf. p. 30).
- [Ork+19] Maciej ORKISZ, Alfredo MORALES PINZÓN, Jean-Christophe RICHARD, Claude GUÉRIN, Leslie SOLORZANO, Daniela SICARU, Camila HERNÁNDEZ, Margarita BALLÉN, B. NEYRAN, Eduardo DAVILA SERRANO et Marcela HERNANDEZ HOYOS. « Voxel-wise assessment of lung aeration changes on CT images using image registration : application to acute respiratory distress syndrome (ARDS) ». In : *International Journal of Computer Assisted Radiology and Surgery* 14 (sept. 2019). DOI : [10.1007/s11548-019-02064-3](https://doi.org/10.1007/s11548-019-02064-3) (cf. p. 35).
- [Pap08] Laurent PAPAIZIAN. *Le syndrome de détresse respiratoire aiguë*. T. 1. Paris : Springer, 2008 (cf. p. 8, 9, 11).
- [Pap+19] Laurent PAPAIZIAN, Cécile AUBRON, Laurent BROCHARD, Jean Daniel CHICHE, Alain COMBES, Didier DREYFUSS, Jean Marie FOREL, Claude GUÉRIN, Samir JABER, Armand MEKONTSO-DESSAP, Alain MERCAT, Jean Christophe RICHARD, Damien ROUX, Antoine VIEILLARD-BARON et Henri FAURE. « Formal guidelines : management of acute respiratory distress syndrome ». In : *Annals of Intensive Care* 9.1 (2019). DOI : [10.1186/s13613-019-0540-9](https://doi.org/10.1186/s13613-019-0540-9) (cf. p. 10).
- [Pap+21] Laurent PAPAIZIAN, Vanessa PAULY, Ilyes HAMOUDA, Florence DAVIET, Veronica ORLEANS, Jean Marie FOREL, Antoine ROCH, Sami HRAIECH et Laurent BOYER. « National incidence rate and related mortality for acute respiratory distress syndrome in France ». In : *Anaesthesia Critical Care and Pain Medicine* 40.1 (fév. 2021). DOI : [10.1016/j.accpm.2020.100795](https://doi.org/10.1016/j.accpm.2020.100795) (cf. p. 8, 9).
- [Pas+98] Marcelo Britto PASSOS AMATO, Carmen Silvia VALENTE BARBAS, Denise MACHADO MEDEIROS, Ricardo BORGES MAGALDI, Guilherme DE PAULA PINTO SCHETTINO, Geraldo LORENZI-FILHO, Ronaldo ADIB KAIRALLA, Daniel DEHEINZELIN, Carlos MUNOZ, Roselaine OLIVEIRA, Teresa YAE TAKAGAKI et Carlos Roberto RIBEIRO CARVALHO. « Effects of a Protective-Ventilation Strategy on Mortality in the Acute Respiratory Distress Syndrome ». In : *Survey of Anesthesiology* 42.5 (1998), p. 259. DOI : [10.1097/00132586-199810000-00009](https://doi.org/10.1097/00132586-199810000-00009) (cf. p. 10).
- [Pen+22] Ludmilla PENARRUBIA, Nicolas PINON, Emmanuel ROUX, E. E. Dávila SERRANO, Jean-Christophe RICHARD, Maciej ORKISZ et David SARRUT. « Improving motion-mask segmentation in thoracic CT with multiplanar U-nets ». In : *Med. Phys.* 49.1 (2022), p. 420-431. DOI : [10.1002/mp.15347](https://doi.org/10.1002/mp.15347) (cf. p. 35, 40).
- [Per+19] Mathias PERSLEV, Erik Bjørnager DAM, Akshay PAI et Christian IGEL. « One Network to Segment Them All : A General, Lightweight System for Accurate 3D Medical Image Segmentation ». In : *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Cham, Switzerland : Springer, oct. 2019, p. 30-38. DOI : [10.1007/978-3-030-32245-8_4](https://doi.org/10.1007/978-3-030-32245-8_4) (cf. p. 36, 37).

- [Pes+16] Antonio PESENTI, Guido MUSCH, Daniel LICHTENSTEIN, Francesco MOJOLI, Marcelo B. P. AMATO, Gilda CINNELLA, Luciano GATTINONI et Michael QUINTEL. « Imaging in acute respiratory distress syndrome ». In : *Intensive Care Medicine* 42.5 (mai 2016), p. 686-698. DOI : [10.1007/s00134-016-4328-1](https://doi.org/10.1007/s00134-016-4328-1) (cf. p. 12, 13).
- [Pra+13] Adhish PRASOON, Kersten PETERSEN, Christian IGEL, François LAUZE, Erik DAM et Mads NIELSEN. « Deep Feature Learning for Knee Cartilage Segmentation Using a Triplanar Convolutional Neural Network ». In : *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*. Berlin, Germany : Springer, 2013, p. 246-253. DOI : [10.1007/978-3-642-40763-5_31](https://doi.org/10.1007/978-3-642-40763-5_31) (cf. p. 36, 37).
- [22] *PubMed*. [Online ; accessed 28. Aug. 2022]. Août 2022. URL : <https://pubmed.ncbi.nlm.nih.gov> (cf. p. 30).
- [Ran+12] V. Marco RANIERI, Gordon D. RUBENFELD, B. Taylor THOMPSON, Niall D. FERGUSON, Ellen CALDWELL, Eddy FAN, Luigi CAMPOROTA et Arthur S. SLUTSKY. « Acute Respiratory Distress Syndrome ». In : *JAMA* 307.23 (juin 2012), p. 2526-2533. DOI : [10.1001/jama.2012.5669](https://doi.org/10.1001/jama.2012.5669) (cf. p. 8).
- [Rég09] Bernard RÉGNIER. « Le Syndrome de Détresse Respiratoire Aiguë A (H1N1), réminiscence des pandémies ». In : *BEHWeb* (déc. 2009), p. 1-5 (cf. p. 14).
- [Ric+22] Jean-Christophe RICHARD, Florian SIGAUD, Maxime GAILLET, Maciej ORKISZ, Sam BAYAT, Emmanuel ROUX, Touria AHAOUARI, Eduardo DAVILA, Loïc BOUSSEL, Gilbert FERRETTI, Hodane YONIS, Mehdi MEZIDI, William DANJOU, Alwin BAZZANI, Francois DHELFT, Laure FOLLIET, Mehdi GIRARD, Matteo POZZI, Nicolas TERZI et Laurent BITKER. « Response to PEEP in COVID-19 ARDS patients with and without extracorporeal membrane oxygenation. A multicenter case-control computed tomography study ». In : *Crit. Care* 26.1 (déc. 2022), p. 1-13. DOI : [10.1186/s13054-022-04076-z](https://doi.org/10.1186/s13054-022-04076-z) (cf. p. 66).
- [Rik+13] Eva M van RIKXOORT et Bram van GINNEKEN. « Automated segmentation of pulmonary structures in thoracic computed tomography scans : a review ». In : *Physics in Medicine and Biology* 58.17 (sept. 2013), R187-R220. DOI : [10.1088/0031-9155/58/17/R187](https://doi.org/10.1088/0031-9155/58/17/R187) (cf. p. 24).
- [Riz+20] Intisar RIZWAN I HAQUE et Jeremiah NEUBERT. « Deep learning approaches to biomedical image segmentation ». In : *Informatics in Medicine Unlocked* 18 (jan. 2020), p. 100297. DOI : [10.1016/J.IMU.2020.100297](https://doi.org/10.1016/J.IMU.2020.100297) (cf. p. 21).
- [Ron+15] Olaf RONNEBERGER, Philipp FISCHER et Thomas BROX. « U-Net : Convolutional Networks for Biomedical Image Segmentation ». In : *arXiv* (mai 2015). DOI : [10.48550/arXiv.1505.04597](https://doi.org/10.48550/arXiv.1505.04597) (cf. p. 30, 40, 63, 88).
- [Rot+14] Holger R. ROTH, Le LU, Ari SEFF, Kevin M. CHERRY, Joanne HOFFMAN, Shijun WANG, Jiamin LIU, Evrim TURKBAY et Ronald M. SUMMERS. « A New 2.5D Representation for Lymph Node Detection Using Random Sets of Deep Convolutional Neural Network Observations ». In : *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*. Cham, Switzerland : Springer, 2014, p. 520-527. DOI : [10.1007/978-3-319-10404-1_65](https://doi.org/10.1007/978-3-319-10404-1_65) (cf. p. 36, 37).
- [Sar13] Collectif SARKA-SPIP. *Le système respiratoire - Société Française des Infirmier(e)s Anesthésistes*. [Online ; accessed 7. Jun. 2022]. Mai 2013. URL : <https://sofia.medicalistes.fr/spip/spip.php?article480> (cf. p. 4).
- [Sca+19] Gaetano SCARAMUZZO, Ludovic BROCHE, Mariangela PELLEGRINI, Liisa PORRA, Savino DEROSA, Angela Principia TANNIOIA, Andrea MARZULLO, João Batista BORGES, Sam BAYAT, Alberto BRAVIN, Anders LARSSON et Gaetano PERCHIAZZI. « The Effect of Positive End-Expiratory Pressure on Lung Micromechanics Assessed by Synchrotron Radiation Computed Tomography in an Animal Model of ARDS ». In : *Journal of Clinical Medicine* 8.8 (2019), p. 1117. DOI : [10.3390/jcm8081117](https://doi.org/10.3390/jcm8081117) (cf. p. 12).
- [She+12] Sarah SHEARD, Praveen RAO et Anand DEVARAJ. *Imaging of acute respiratory distress syndrome*. Avt. 2012. DOI : [10.4187/respcare.01731](https://doi.org/10.4187/respcare.01731) (cf. p. 12, 13).

- [Shi+16] Zhenghao SHI, Jiejue MA, Minghua ZHAO, Yonghong LIU, Yaning FENG, Ming ZHANG, Lifeng HE et Kenji SUZUKI. « Many Is Better Than One : An Integration of Multiple Simple Strategies for Accurate Lung Segmentation in CT Images ». In : *BioMed Research International* 2016 (2016), p. 1-13. doi : [10.1155/2016/1480423](https://doi.org/10.1155/2016/1480423) (cf. p. 24, 25).
- [Sie+22] Mark D SIEGEL et Reed SIEMIENIUK. *Acute respiratory distress syndrome : Supportive care and oxygenation in adults - UpToDate*. [Online; accessed 15. Jun. 2022]. Mai 2022. URL : https://www.uptodate.com/contents/acute-respiratory-distress-syndrome-supportive-care-and-oxygenation-in-adults?topicRef=1637&source=related_link#H13 (cf. p. 10).
- [Slu+05] I. SLUIMER, M. PROKOP et B. van GINNEKEN. « Toward automated segmentation of the pathological lung in CT ». In : *IEEE Trans. Med. Imaging* 24.8 (août 2005), p. 1025-1038. doi : [10.1109/TMI.2005.851757](https://doi.org/10.1109/TMI.2005.851757) (cf. p. 25).
- [Sun+06] Xuejun SUN, Haibo ZHANG et Huichuan DUAN. « 3D Computerized Segmentation of Lung Volume With Computed Tomography ». In : *Academic Radiology* 13.6 (2006), p. 670-677. doi : [10.1016/j.acra.2006.02.039](https://doi.org/10.1016/j.acra.2006.02.039) (cf. p. 24).
- [Sur+21] Jasjit S. SURI, Sushant AGARWAL, Suneet K. GUPTA, Anudeep PUVVULA, Mainak BISWAS, Luca SABA, Arindam BIT, Gopal S. TANDEL, Mohit AGARWAL, Anubhav PATRICK, Gavino FAA, Inder M. SINGH, Ronald OBERLEITNER, Monika TURK, Paramjit S. CHADHA, Amer M. JOHRI, J. MIGUEL SANCHES, Narendra N. KHANNA, Klaudija VISKOVIC, Sophie MAVROGENI, John R. LAIRD, Gyan PAREEK, Martin MINER, David W. SOBEL, Antonella BALESTRIERI, Petros P. SFIKAKIS, George TSOUFAS, Athanasios PROTOGEROU, Durga Prasanna MISRA, Vikas AGARWAL, George D. KITAS, Puneet AHLUWALIA, Jagjit TEJI, Mustafa AL-MAINI, Surinder K. DHANJIL, Meyypan SOCKALINGAM, Ajit SAXENA, Andrew NICOLAIDES, Aditya SHARMA, Vijay RATHORE, Janet N. A. AJULUCHUKWU, Mostafa FATEMI, Azfa ALIZAD, Vijay VISWANATHAN, P. K. KRISHNAN et Subbaram NAIDU. « A narrative review on characterization of acute respiratory distress syndrome in COVID-19-infected lungs using artificial intelligence ». In : *Comput. Biol. Med.* 130 (mars 2021), p. 104210. doi : [10.1016/j.combiomed.2021.104210](https://doi.org/10.1016/j.combiomed.2021.104210) (cf. p. 32).
- [Tah+15] Abdel Aziz TAHA et Allan HANBURY. « Metrics for evaluating 3D medical image segmentation : Analysis, selection, and tool ». In : *BMC Medical Imaging* 15.1 (août 2015). doi : [10.1186/S12880-015-0068-X](https://doi.org/10.1186/S12880-015-0068-X) (cf. p. 21, 22).
- [Tes21] Brenda L. TESINI. « Coronavirus et syndrome respiratoire aigu sévère (COVID-19, MERS et SRAS) ». In : *Le Manuel MSD* (2021), p. 1-5 (cf. p. 14).
- [Tir+20] Jo Anna TIROLIEN et Marc GARNIER. « COVID-19, an atypical acute respiratory distress syndrome ». In : *Praticien en Anesthésie Réanimation* 24.4 (sept. 2020), p. 225-229. doi : [10.1016/J.PRATAN.2020.07.001](https://doi.org/10.1016/J.PRATAN.2020.07.001) (cf. p. 14).
- [Tor+00] Gerard J. TORTORA et Bogdan GRABOWSKI. *ÉLÉMENTS D'ANATOMIE ET D'HISTOLOGIE DU SYSTÈME RESPIRATOIRE (Part 1)*. Paris, France : De Boeck, 2000 (cf. p. 4).
- [Udu+03] Jayaram K. UDUPA et Punam K. SAHA. « Fuzzy connectedness and image segmentation ». In : *Proceedings of the IEEE* 91.10 (2003), p. 1649-1669. doi : [10.1109/JPR0C.2003.817883](https://doi.org/10.1109/JPR0C.2003.817883) (cf. p. 24).
- [Val+21] Amin VALIZADEH et Morteza SHARIATEE. « The Progress of Medical Image Semantic Segmentation Methods for Application in COVID-19 Detection ». In : *Computational Intelligence and Neuroscience* 2021 (2021). doi : [10.1155/2021/7265644](https://doi.org/10.1155/2021/7265644) (cf. p. 32).
- [Van14] David VANDROUX. « Lessons learned from pandemic influenza severe cases since 2009 ». In : *Praticien en Anesthésie Réanimation* 18.3 (2014), p. 185-191. doi : [10.1016/j.pratan.2014.03.006](https://doi.org/10.1016/j.pratan.2014.03.006) (cf. p. 14).
- [War+21] Girindra WARDHANA, Hamid NAGHIBI, Beril SIRMACEK et Momen ABAYAZID. « Toward reliable automatic liver and tumor segmentation using convolutional neural network based on 2.5D models ». In : *Int. J. CARS* 16.1 (jan. 2021), p. 41-51. doi : [10.1007/s11548-020-02292-y](https://doi.org/10.1007/s11548-020-02292-y) (cf. p. 36, 37).

- [Xu+19] Mingjie XU, Shouliang QI, Yong YUE, Yueyang TENG, Lisheng XU, Yudong YAO et Wei QIAN. « Segmentation of lung parenchyma in CT images using CNN trained with the clustering algorithm generated dataset ». In : *BioMedical Engineering OnLine* 18.2 (2019), p. 21. DOI : [10.1186/s12938-018-0619-9](https://doi.org/10.1186/s12938-018-0619-9) (cf. p. 29).
- [Ye+20] Zheng YE, Yun ZHANG, Yi WANG, Zixiang HUANG et Bin SONG. « Chest CT manifestations of new coronavirus disease 2019 (COVID-19) : a pictorial review ». In : *European Radiology* 30.8 (août 2020), p. 4381-4389. DOI : [10.1007/s00330-020-06801-0](https://doi.org/10.1007/s00330-020-06801-0) (cf. p. 14).
- [Yeg+18] Varduhi YEGHIAZARYAN et Irina VOICULESCU. « Family of boundary overlap metrics for the evaluation of medical image segmentation ». In : *Journal of Medical Imaging* 5.01 (fév. 2018), p. 1. DOI : [10.1117/1.jmi.5.1.015006](https://doi.org/10.1117/1.jmi.5.1.015006) (cf. p. 21, 22).
- [Yin+09] Youbing YIN, Eric A. HOFFMAN et Ching Long LIN. « Mass preserving nonrigid registration of CT lung images using cubic B-spline ». In : *Medical Physics* 36.9 (2009), p. 4213-4222. DOI : [10.1118/1.3193526](https://doi.org/10.1118/1.3193526) (cf. p. 62).
- [Zha21] Fengjun ZHANG. « Application of machine learning in CT images and X-rays of COVID-19 pneumonia ». In : *Medicine* 100.36 (2021), e26855. DOI : [10.1097/MD.00000000000026855](https://doi.org/10.1097/MD.00000000000026855) (cf. p. 32).
- [Zha+12] Li ZHANG, Kyungmoo LEE, Meindert NIEMEIJER, Robert F. MULLINS, Milan SONKA et Michael D. ABRÀMOFF. « Automated Segmentation of the Choroid from Clinical SD-OCT ». In : *Investigative Ophthalmology & Visual Science* 53.12 (2012), p. 7510. DOI : [10.1167/iovs.12-10311](https://doi.org/10.1167/iovs.12-10311) (cf. p. 24).
- [Zha+18] Tianyi ZHAO, Dashan GAO, Jiao WANG et Zhaozheng YIN. « Lung segmentation in CT images using a fully convolutional neural network with multi-instance and conditional adversary loss ». In : *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. T. 2018-April. Isbi. IEEE, 2018, p. 505-509. DOI : [10.1109/ISBI.2018.8363626](https://doi.org/10.1109/ISBI.2018.8363626) (cf. p. 30).
- [Zho+17] Xiangrong ZHOU, Ryosuke TAKAYAMA, Song WANG, Takeshi HARA et Hiroshi FUJITA. « Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method ». In : *Med. Phys.* 44.10 (oct. 2017), p. 5221-5233. DOI : [10.1002/mp.12480](https://doi.org/10.1002/mp.12480) (cf. p. 29).
- [Zho+18] Zongwei ZHOU, Md Mahfuzur Rahman SIDDIQUEE, Nima TAJBAKHSI et Jianming LIANG. « UNet++ : A Nested U-Net Architecture for Medical Image Segmentation ». In : *CoRR* abs/1807.10165 (2018) (cf. p. 30).
- [Zom+14] Maurizio ZOMPATORI, Federica CICCARESE et Luca FASANO. « Overview of current lung imaging in acute respiratory distress syndrome ». In : *European Respiratory Review* 23.134 (2014), p. 519-530. DOI : [10.1183/09059180.00001314](https://doi.org/10.1183/09059180.00001314) (cf. p. 13).

