

Rapport de stage

Étude des modèles de fondation Segment Anything (SAM) pour l'analyse d'images médicales 3D.

Structure d'accueil : Centre de Recherche en Acquisition et Traitement de l'Image pour la Santé (CREATIS).

Adresse de la structure d'accueil : Bâtiment Léonard de Vinci, 21 Avenue Jean Capelle, 69621 Villeurbanne Cedex

Tuteur de stage : Razmig KÉCHICHIAN

Formation : Institut National des Sciences Appliquées de Lyon - Télécommunications, services et usages

Adresse de la formation : Bâtiment Hedy Lamarr - INSA Lyon 6 avenue des arts, 69621 Villeurbanne cedex

Enseignant référent : Tristan ROUSSILLON

Nom : Marie FRIOT

Intitulé du stage : Stage en laboratoire de recherche.

Période du stage : du 22/04/2025 au 29/08/2025

Table des matières

1 Présentation de ma structure d'accueil	3
1 Contexte	3
2 Le laboratoire CREATIS	3
2 Étude des modèles de fondation Segment Anything (SAM) pour l'analyse d'images médicales 3D.	6
1 Introduction	7
2 État de l'Art	8
2.1 Les modèles de type Segment Anything Model (SAM)	8
2.2 Génération automatique d'indications	9
3 Méthode	10
3.1 MedSAM2	10
3.2 FROG : Méthode de recalage d'ensemble d'images 3D	11
3.3 Segmentation multi-organes automatique avec MedSAM2	12
3.3.1 Recalage et génération automatique d'indications	12
3.3.2 MedSAM2 : Première segmentation	13
3.3.3 Segmentation en cascade : affinage de la première segmentation	13
4 Résultats	14
4.1 Base de données d'évaluation : AMOS22	14
4.2 Performance de référence de MedSAM2 : Segmentation en générant les boîtes englobantes avec les masques de vérité.	14
4.3 Segmentation en générant les boîtes englobantes de façon automatique.	15
5 Discussion	16
5.1 Segmentation en générant les boîtes englobantes avec les masques de vérité : performance de référence de MedSAM2.	16
5.2 Segmentation en générant les boîtes englobantes de façon automatique.	17
5.2.1 Stratégie A et B : Une seule segmentation avec une boîte englobante ou trois boîtes englobantes	17
5.2.2 Stratégie C : Deux segmentations en cascade	18
5.2.3 Stratégie D : Deux segmentations en cascade en augmentant l'aire des boîtes englobantes et la hauteur de propagation pour la deuxième segmentation	20
5.2.4 Comparaison des performances de notre méthode de segmentation à celles de modèles spécialisés.	21
5.3 Traitement multi-organes	22
5.4 Impact énergétique, environnemental et temps d'exécution	22
6 Conclusion et perspectives	23
7 Annexe	25
7.1 Détermination des seuils	25
7.2 Architecture matérielle utilisée.	25
3 Retour d'expérience	28
1 Compétences acquises et développées	28
2 Ma perception du milieu de la recherche	29
3 Conclusion	29

Table des figures

1.1	Bibliométrie de l'équipe MYRIAD	5
2.1	SAM : Exemples d'images naturelles avec leurs masques de segmentations utilisées pour l'entraînement.	9
2.2	SAM : Différents types d'indications.	9
2.3	MedSAM2 : Architecture et mécanisme de propagation.	11
2.4	Schéma illustrant la méthode FROG (Fast Registration Of image Groups).	11
2.5	Schéma de notre méthode de segmentation automatique multi-organes avec le modèle de fondation MedSAM2.	12
2.6	MedSAM2 : Stratégie de génération des boîtes englobantes et de propagation des masques de segmentation.	14
2.7	Segmentations avec MedSAM2 : comparaison entre l'ajout d'une boîte englobante et de trois boîtes englobantes définies avec les masques de vérité.	17
2.8	Segmentations avec MedSAM2 : comparaison entre l'ajout d'une boîte englobante et de trois boîtes englobantes définies avec les cartes de score.	18
2.9	Segmentations générées après deux inférences : la première pour localiser l'organe et la deuxième pour affiner la première.	19
2.10	Segmentations en cascade : limitations.	19
2.11	Segmentations en cascade : cas du duodénum et du pancréas.	20
2.12	Segmentations générées après deux inférences en augmentant la taille des boîtes englobantes et la hauteur de propagation pour la deuxième segmentation.	21
2.13	Segmentations générées après deux inférences en augmentant la hauteur de propagation pour la deuxième segmentation : limitations pour l'aorte, la veine cave inférieur, et les glandes surrénales.	21
2.14	Avantage du traitement multi-organes.	22

Chapitre 1

Présentation de ma structure d'accueil

1 Contexte

Lors de mon parcours en TC, j'ai choisi de suivre le parcours recherche pour me familiariser avec le domaine de la recherche. C'est dans la continuité de ce parcours, que j'ai choisi de faire mon stage de 4TC au laboratoire de recherche CREATIS. Cela m'a permis dans un premier temps de pouvoir avancer sur mon sujet de parcours recherche en ayant de plus grandes ressources en temps mais aussi en termes de matériel (ressources de calcul). Dans un second temps, ce stage en laboratoire m'a permis de découvrir de plus près le milieu de la recherche au travers des échanges que j'ai pu avoir avec les membres permanents du laboratoire, mais aussi avec les doctorants ou les post-doctorants. De plus, j'ai eu la chance de pouvoir participer à la vie du laboratoire via des séminaires et des réunions d'équipes.

2 Le laboratoire CREATIS

CREATIS (Centre de Recherche en Acquisition et Traitement du Signal pour la Santé) est un laboratoire de recherche spécialisé en imagerie médicale. Ses compétences couvrent toute la chaîne du processus d'imagerie allant de l'acquisition des images jusqu'à l'analyse des images pouvant faciliter le diagnostic médical. Le laboratoire comporte des experts en imagerie par résonance magnétique, en ultrasons, en rayons X et en optique. CREATIS mène des recherches en partenariat avec de nombreux acteurs académiques et industriels, que ce soit au niveau local, national ou international. En moyenne, le laboratoire compte 20 partenaires industriels, leur permettant de réaliser de la recherche translationnelle.

La structure est une [unité mixte de recherche](#) du CNRS (UMR 5220), de l'Inserm (U1294), de l'INSA Lyon, de l'université Claude Bernard Lyon 1 et de l'Université Jean-Monnet de Saint-Étienne. Il compte actuellement 92 membres permanents. Parmi les membres permanents, le laboratoire compte 16 chercheurs qui se dédient seulement à la recherche et 51 maîtres de conférences ou professeurs des universités qui partagent leur activité entre la recherche et l'enseignement.

Les membres du laboratoire chargés de recherche sont répartis en quatre équipes avec des objectifs différents :

- MYRIAD (mon équipe) : L'objectif de l'équipe MYRIAD est de développer des méthodologies en traitement d'images, modélisation biomécanique et simulation d'images dans le domaine de l'imagerie médicale, qu'elle soit cardio-vasculaire, dentaire, abdominale, pulmonaire ou neurologique.
- MAGICS : Cette équipe se concentre sur le développement de nouvelles méthodes de mesure des paramètres du corps humain qui ne sont pas directement visibles grâce à l'IRM ou à d'autres technologies optiques. De plus, ils travaillent sur la recherche de nouveaux [biomarqueurs](#) robustes, reproductibles et fiables et qui prennent en compte les mouvements physiologiques. Les biomarqueurs ciblés proviennent des propriétés mécaniques, biochimiques ou structurelles des tissus étudiés. Cela les mène à développer de nouvelles stratégies d'acquisition d'image ou de nouvelles instrumentations.

- **ULTIM** : L'équipe **ULTIM** s'intéresse à l'imagerie médicale par ultrasons, de la conception des capteurs à l'analyse en temps réel des tissus ou du flux sanguin par exemple.
- **TOMORADIO** : Cette équipe a été créée pour rassembler des recherches autour du traitement et reconstruction d'images par **tomodensitométrie** (CT) et des thérapies de radiation.

Ces quatre équipes de recherche contribuent à la recherche sur la médecine personnalisée et prédictive grâce au développement de nouvelles modalités d'imagerie et de différentes méthodologies d'analyse d'images. Celles-ci interviennent dans le diagnostic et dans le suivi thérapeutique des patients, tout comme dans la prédiction de l'évolution des pathologies et de leur réponse au traitement.

Le laboratoire possède en plus de ces quatre équipes de recherche, trois plateformes. La première est une plateforme d'imagerie appelée **PILoT**. Elle regroupe les modalités IRM, optique et échographie. Cette plateforme est ouverte à tous les personnels académiques ou industriels qui ont besoin du matériel mis à disposition pour réaliser leurs expériences et donner des réponses à leurs problèmes.

En plus de la plateforme **PILoT**, le laboratoire se charge de la plateforme **VIP** (Virtual Imaging Platform). C'est une plateforme web permettant d'offrir un accès libre à des applications scientifiques pour des chercheurs du monde entier. En 2025, la plateforme compte plus de 1550 utilisateurs et 25 applications. Elle permet le partage et l'accès aux ressources informatiques tout en répondant aux enjeux d'interopérabilité et de reproductibilité.

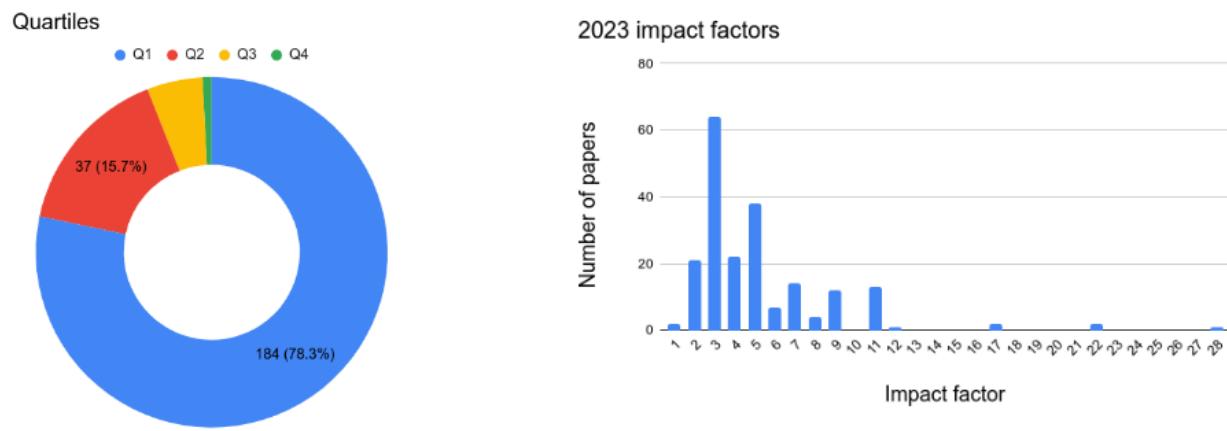
Pour finir, le laboratoire dispose de la plateforme **PRISM** située sur le site de l'hôpital Nord à Saint-Étienne. **PRISM** est une plateforme de recherche en imagerie et **spectroscopie** multi-noyaux clinique et pré-clinique. Cette plateforme à moitié dédiée à un usage clinique et à moitié dédiée à la recherche permet de développer de nouvelles recherches avec un fort accent sur la prévention (nutrition, activité physique,...) et le diagnostic.

Concernant l'évaluation des performances d'un laboratoire public plusieurs indicateurs sont mis en perspective. Le premier est la qualité des papiers de recherche publiés face à leur quantité. La qualité d'un papier de recherche est évaluée à travers le nombre de citations qu'il possède mais aussi en fonction de la revue dans laquelle il a été publié. En effet, pour chaque revue on peut calculer le facteur d'impact. Il mesure la fréquence moyenne à laquelle les articles publiés dans une revue sont cités sur une période de 2 ans. Par exemple, si une revue a un facteur d'impact de 8, cela veut dire que, en moyenne, chaque article publié dans les deux années précédentes a été cité 8 fois en un an. Le facteur d'impact est quand même un indicateur à prendre avec des pincettes car il dépend du domaine. Les revues spécialisées ou de niche vont avoir un facteur d'impact faible, mais cela ne signifie pas qu'elles sont mauvaises car elles touchent un public plus restreint. Les revues sont également classées en quartiles avec un indicateur appelé **SJR** (SCImago Journal Rank). Il mesure l'influence d'une revue, en prenant en compte le nombre de citations reçues mais aussi l'importance des revues dans lesquelles sont publiés les articles qui font ces citations. La période d'analyse est elle de trois ans. Cet indicateur permet ensuite de classer les revues triées en catégories par quartile. Une des limitations de cet indicateur est qu'il provient d'un algorithme dont les poids peuvent être remis en question. De plus, cet indicateur ne permet pas de comparer deux revues de domaines différents car il dépend de la concurrence au sein du domaine. Les performances bibliométriques de l'équipe **MYRIAD** sont présentées sur la Figure 1.1. L'équipe publie majoritairement dans des revues de premier quartile selon l'indicateur **SJR** et le facteur d'impact moyen des revues dans lesquelles sont publiés les papiers est de 5.8, ce qui est satisfaisant pour le domaine.

Un deuxième élément permettant d'évaluer un laboratoire est sa capacité à obtenir des financements. Le montant et la diversité des financements compétitifs (ANR, financements européens, financements industriels) sont des indicateurs de l'attractivité et de la qualité de la recherche d'un laboratoire.

Les distinctions obtenues par les chercheurs (prix, invitations à des conférences prestigieuses, postes éditoriaux dans des revues), la participation à des réseaux scientifiques internationaux et la réputation auprès des pairs sont aussi des indicateurs traduisant la qualité de la recherche d'un laboratoire.

Pour finir, l'impact socio-économique de la recherche d'un laboratoire est aussi un critère d'évaluation. Il se traduit par le nombre de brevets déposés, les licences et logiciels élaborés, la collaboration avec le secteur privé ou les institutions publiques ou la contribution à des politiques publiques. Concernant le laboratoire **CREATIS**, 7 brevets ont été déposés ces 5 dernières années et 25 thèses CIFRE, en partenariat avec une entreprise, se sont déroulées.



Chapitre 2

Étude des modèles de fondation Segment Anything (SAM) pour l'analyse d'images médicales 3D.

Génération automatique d'indications pour la segmentation multi-organe avec le modèle MedSAM2.

Résumé. La segmentation automatique d'images médicales est actuellement une tâche cruciale pour les cliniciens. Toutefois, son coût énergétique est assez élevé. Cela est dû à l'augmentation continue du nombre de paramètres des modèles par apprentissage profond, qui sont maintenant devenus l'état de l'art dans ce domaine. Cette croissance de la complexité des modèles mène à des entraînements et des inférences plus longues et plus gourmandes en ressources. Ce phénomène est accentué par la démultiplication des modèles spécialisés. À travers ce manuscrit est présentée une méthode de segmentation automatique utilisant un modèle de fondation déjà entraîné sur des images médicales, MedSAM2 [Ma et al., 2025]. Ce modèle a montré de bonnes performances lorsqu'il est guidé avec des indications (prompts en anglais) précises. Dans ce contexte, les indications sont des contraintes pour la réalisation de la segmentation et font partie du fonctionnement de l'algorithme. Le but de notre démarche a été de générer automatiquement ces indications en utilisant un recalage d'images. Cette méthode de segmentation se veut frugale car il n'y a aucun entraînement ou réentraînement du modèle MedSAM2. Elle a été testée sur des images tomodensitométriques (images CT) 3D de la base de données AMOS22 [Ji et al., 2022] et a été comparée à des modèles spécialisés avec une architecture de type U-Net. Elle a montré des performances similaires ou du même ordre de grandeur que les modèles spécialisés sur les organes abdominaux ayant une forme et une localisation peu variables mais a des limitations pour les organes à la forme et la localisation plus instables.

1 Introduction

La segmentation des organes abdominaux fournit des informations cruciales telles que les interrelations entre organes, leurs tailles, leurs positions et formes individuelles, essentielles à la prise de décision clinique, que ce soit pour l'analyse de structures, le diagnostic de pathologies, la planification d'opérations ou le suivi longitudinal des patients. Par exemple, la segmentation multi-organes abdominale joue un rôle essentiel dans le traitement de certains cancers nécessitant une radiothérapie. Cette méthode consiste à cibler les cellules cancéreuses et à leur envoyer des radiations afin de les détruire. En revanche, il est nécessaire que les radiations n'atteignent pas les organes sains situés autour au risque de les endommager. C'est là où la segmentation multi-organes intervient, permettant de définir les positions et formes des organes abdominaux afin de planifier le trajet des faisceaux d'irradiation.

Dans la plupart des applications cliniques, les structures saines à risque sont délimitées dans les images à la main par des spécialistes. Cette tâche est cependant fastidieuse et coûteuse en temps. C'est pour cela que de nombreux chercheurs ont mené des études permettant de développer des stratégies automatiques permettant de segmenter les organes abdominaux.

Les premières études, [Lee et al., 2014], utilisaient des méthodes de détection automatiques de caractéristiques comme les lignes ou les coins ainsi que des modèles mathématiques pour suivre le gradient de l'image le long des limites des objets. Des méthodes basées sur des atlas ont également été des approches très utilisées pour la segmentation automatique. Ces méthodes consistent à transposer les structures pré-définies sur l'image cible à l'aide d'un recalage [Park et al., 2003]. Enfin, des méthodes basées sur des modèles statistiques

de forme pour la segmentation automatisée ont également été proposées. En revanche, ces méthodes ont été effacées avec l'essor de l'apprentissage profond. De nombreux algorithmes dérivant de ce concept ont alors été développés [Wang et al., 2022]. L'une des méthodes de segmentation automatique qui est désormais une méthode de référence pour la segmentation des images tomodensitométriques est U-Net [Ronneberger et al., 2015].

Cependant, un changement de paradigme est observé récemment dans la segmentation d'images médicales. Jusqu'à présent, pour résoudre un problème de segmentation, des modèles spécialisés comme U-Net, entraînés pour une tâche particulière, étaient donc utilisés. Désormais, de plus en plus de modèles de fondation comme les modèles de type Segment Anything Model (SAM) [Kirillov et al., 2023] sont développés. Ces modèles, entraînés sur un très grand nombre d'images, sont capables de générer de bonnes segmentations, à condition d'être guidés par des indications fournies par l'utilisateur (clics positifs/négatifs, boîte englobante ou masque binaire de l'objet cible). Cela est dû à la bonne capacité de généralisation de ces modèles, qui les rend capables de segmenter correctement des objets ou structures dans des images qui diffèrent de celles utilisées pendant leur entraînement. Lors de cette étude, nous nous sommes plus particulièrement intéressés au modèle de fondation MedSAM2 [Ma et al., 2025], entraîné sur des images médicales et capable de traiter des images 3D. Nous avons voulu comparer les performances de ce modèle sans entraînement à celles d'un modèle spécialisé d'architecture U-Net pour la segmentation multi-organes abdominale sur des images tomodensitométriques (images CTs) 3D de la base de données AMOS22 [Ji et al., 2022]. De plus, nous avons voulu rendre la segmentation abdominale à l'aide du modèle MedSAM2 automatique en générant les indications à l'aide d'un recalage. Ce recalage permet d'aligner des masques de segmentation de référence sur l'image cible à segmenter. Ces masques recalés servent à générer les boîtes englobantes des organes à segmenter.

L'un des principaux atouts d'un modèle de fondation réside dans sa capacité de généralisation, qui lui permet d'aborder des tâches de segmentation spécifiques sans recourir à un réentraînement. Il peut être utilisé comme un modèle généraliste, capable de résoudre de nombreuses tâches de traitement d'images médicales, sans être entraîné à chaque fois. C'est intéressant du point de vue consommation énergétique et émission de gaz à effet de serre [Barbierato and Gatti, 2024].

2 État de l'Art

2.1 Les modèles de type Segment Anything Model (SAM)

Le modèle Segment Anything Model (SAM) [Kirillov et al., 2023] est un modèle de Vision basé sur une architecture de type Transformer [Dosovitskiy et al., 2020]. Il a été entraîné sur 11 millions d'images naturelles et sur 1 milliard de masques de segmentation. La Figure 2.1 illustre des masques de segmentation. Le modèle SAM permet de générer de bonnes segmentations sur des images naturelles, tant qu'il est guidé par des indications fournies par l'utilisateur (clics positifs/négatifs, boîte englobante ou masque binaire de l'objet cible, voir Figure 2.2). Comme il a été entraîné sur des images naturelles, certaines études ont évalué ses capacités à généraliser aux problèmes de segmentation sur des images médicales, sans nouvel entraînement [Roy et al., 2023], [Mazurowski et al., 2023], [Huang et al., 2024]. Celles-ci ont montré que le modèle SAM, en raison de l'écart de domaine entre les images médicales et les images naturelles, présentait des performances plus faibles sur les images médicales. Roy et al. [2023] ont montré également que sur les images tomodensitométriques 3D du jeu de données AMOS22, le modèle SAM avait des performances significativement plus faibles que des modèles spécialisés basés sur des architectures de type U-Net pour la segmentation des organes abdominaux. Pour cette raison, plusieurs travaux ont mis en place des méthodes permettant d'adapter le modèle SAM aux problèmes de segmentation sur des images médicales. Par exemple, Wu et al. [2025] ont intégré des blocs d'adaptation à l'intérieur desquels les paramètres ont été entraînés lors d'une phase de réentraînement alors que les paramètres initiaux du modèle ont été gelés. Wei et al. [2024] ont également ajouté des blocs d'adaptation, mais ont en plus remplacé la dernière partie du modèle, le décodeur, par un réseau de neurones capable d'apprendre à prédire les frontières de segmentation sous forme de fonction continue. D'autres études ont utilisé des techniques de "fine-tuning", c'est-à-dire de réentraînement complet du modèle SAM, lui permettant d'avoir des performances dépassant celles des modèles spécialisés pour des tâches de segmentation sur des images médicales [Ma et al., 2024].

Ces méthodes permettent de résoudre le problème d'écart de domaine entre les images naturelles et les images médicales. Cependant, leur application est limitée par le fait que le modèle sous-jacent, SAM, ne peut



FIGURE 2.1 – SAM : Exemples d’images naturelles avec leurs masques de segmentations utilisées pour l’entraînement. [Kirillov et al., 2023].



FIGURE 2.2 – SAM : différents types d’indication. De gauche à droite : boîte englobante, cliques positifs et négatifs, masque binaire. [Ravi et al., 2024].

traiter que des images 2D, alors que de nombreuses images médicales, dont les images tomodensitométriques abdominales, sont des images 3D. Ainsi, pour appliquer les méthodes dérivant du modèle SAM pour la segmentation d’organes dans des images 3D, il est nécessaire que l’utilisateur donne une indication sur chacune des coupes de l’image, ce qui rend l’application clinique de ces méthodes plus longue et plus coûteuse.

Par la suite, pour permettre le traitement de vidéos, un nouveau modèle de fondation, ayant une architecture principale similaire à celle de SAM, a été développé : SAM2 [Ravi et al., 2024]. Ce modèle possède une banque de mémoire qui lui permet de segmenter un objet présent sur plusieurs trames d’une vidéo en permettant à l’utilisateur de donner une indication sur une seule trame. Ce modèle peut donc permettre de segmenter des images 3D avec une seule indication par objet en considérant les différentes coupes de l’image comme une suite d’image 2D composant une vidéo. Pour appliquer SAM2 sur des images médicales, il a été proposé de réentraîner le modèle SAM2 sur une large base de données d’images médicales de différentes modalités [Ma et al., 2025]. Plus de détails concernant le modèle MedSAM2 seront donnés dans la Section 3.1 car il est une brique fondamentale de notre méthode. Ce modèle, MedSAM2, a montré de meilleures performances sur les images médicales que le modèle de fondation SAM2, non adapté au domaine médical. En revanche, ce modèle n’a pas été évalué face aux performances d’un modèle spécialisé pour la segmentation des organes abdominaux. De plus, ce modèle nécessite l’intervention humaine pour la génération d’indications permettant de définir quel(s) organe(s) doivent être segmenté(s) dans l’image.

2.2 Génération automatique d’indications

Les modèles reposant sur l’architecture SAM, dont MedSAM2 fait partie, nécessitent des indications de la part de l’utilisateur, telles que des clics positifs/négatifs, une boîte englobante ou un masque binaire. Ils requièrent donc l’intervention d’un opérateur. Certaines études ont développé des méthodes permettant de générer automatiquement les indications nécessaires aux modèles de type SAM, afin de réduire le temps de traitement des images par les cliniciens et de standardiser la génération des indications. Ces méthodes peuvent se scinder en deux catégories.

La première catégorie de méthode regroupe les méthodes utilisant un classifieur entraîné sur les images cibles permettant de générer une boîte englobante ou une segmentation grossière afin de localiser l'organe à segmenter. Par exemple Pandey et al. [2023] et Yin et al. [2025] ont utilisé un détecteur (YOLO) pour générer les boîtes englobantes autour des organes cibles. Na et al. [2024] et Colbert et al. [2024] ont quant à eux utilisé un réseau neuronal convolutif avec une architecture de type U-Net pour produire une segmentation grossière, utilisée ensuite pour générer les indications nécessaires à un modèle de type SAM permettant d'affiner la segmentation. Borgli et al. [2025] ont utilisé un classifieur (DenseNet) entraîné sur des images similaires aux images cibles. Le classifieur a été ensuite utilisé pour générer pour chaque image et chaque organe une carte d'activation de classe permettant de déterminer les boîtes englobantes des organes à segmenter.

La deuxième catégorie de méthode concerne celles utilisant le recalage d'images. Xu et al. [2024] proposent de trouver l'image la plus similaire à l'image cible dans une banque d'images annotées et de la recaler avec son masque de vérité sur l'image à segmenter. Le masque de vérité recalé sur l'image cible est alors utilisé pour générer les indications nécessaires aux modèles de type SAM. Cette méthode a l'avantage d'être rapide et ne nécessite pas d'entraîner un modèle spécifique. C'est pour cela que nous avons décidé d'utiliser également un recalage d'images pour générer automatiquement les indications.

3 Méthode

La méthode de segmentation multi-organe que nous avons mise en place utilise le modèle MedSAM2 comme segmentateur. Ce modèle nécessite des indications que nous avons générées à l'aide d'un recalage d'image, obtenu avec la méthode FROG. Dans cette section, le modèle MedSAM2 ainsi que la méthode FROG sont d'abord détaillés en amont de notre approche expliquée dans la Section 3.3.

3.1 MedSAM2

MedSAM2 [Ma et al., 2025] est un transformeur de vision (Vision Transformer) [Dosovitskiy et al., 2020]. Il a été développé en réentraînant SAM2 [Ravi et al., 2024] sur des images médicales. La base de données de réentraînement était constituée d'images médicales 3D et de vidéos imageant une grande variété de structures anatomiques saines et pathologiques (organes abdominaux, cœur, cerveau,...) issues d'une multitude de modalités (images CT, IRM, ultrasons,...). La modalité la plus représentée était la tomodensitométrie 3D (images CTs) avec 363 161 images. Concernant l'architecture du modèle, MedSAM2 a donc la même structure que SAM2 : un encodeur d'images ("image encoder"), un module d'attention avec mémoire ("memory attention"), un encodeur d'indications ("prompt encoder") et un décodeur de masque ("mask decoder"). L'architecture du modèle ainsi que son mécanisme de propagation est présenté sur la Figure 2.3. L'encodeur d'image permet d'extraire des caractéristiques multi-échelles de chaque coupe 2D des images 3D en utilisant un Vision Transformer hiérarchique (Hier) [Ryali et al., 2023]. Il a l'avantage d'être plus rapide et plus performant que le Vision Transformer initial utilisé dans le modèle SAM [Dosovitskiy et al., 2020]. Le module d'attention avec mémoire performe un mécanisme d'auto-attention mais aussi d'attention croisée entre la coupe à segmenter et les coupes précédentes déjà segmentées gardées dans la mémoire dynamique. Cela permet de conditionner la prédiction en cours en fonction des prédictions précédentes et d'éviter à l'utilisateur de donner une indication sur l'organe à segmenter à chaque coupe. L'encodeur d'indications permet de convertir les indications de l'utilisateur (ex : points, boîtes englobantes et masques) en représentation vectorielle. Le modèle MedSAM2 a été réentraîné avec seulement des boîtes englobantes car c'est la méthode qui est la moins ambiguë pour spécifier la cible de segmentation tout en étant beaucoup moins contraignante que le masque binaire. Lors du réentraînement du modèle, pour les images 3D, les boîtes englobantes ont été données au milieu de l'organe cible et le masque de segmentation a été alors propagé de façon bidirectionnelle jusqu'au bout du volume. Enfin, le décodeur de masque incorpore les caractéristiques conditionnées par la mémoire de la coupe 2D ainsi que la représentation vectorielle de l'indication (prompt) pour permettre de générer un masque de segmentation.

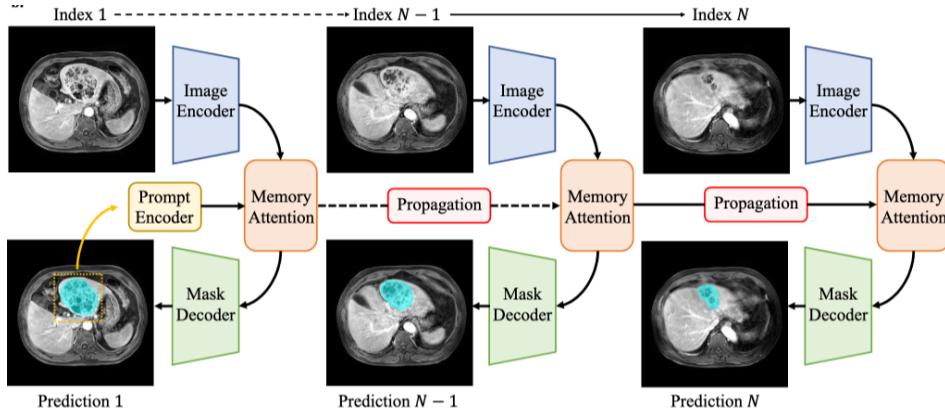


FIGURE 2.3 – MedSAM2 : Architecture et mécanisme de propagation : [Ma et al., 2025].

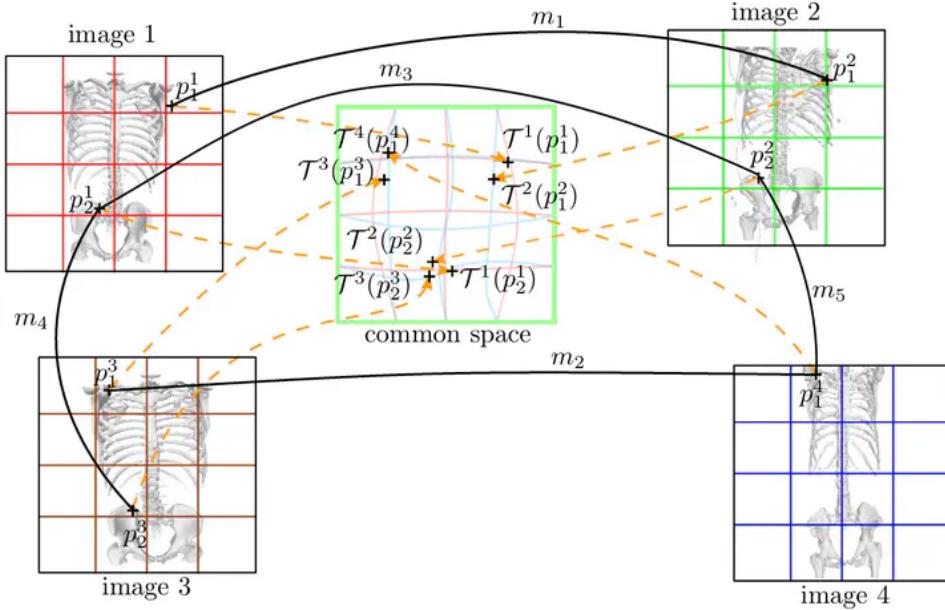


FIGURE 2.4 – Schéma illustrant la méthode FROG (Fast Registration Of image Groups) : [Agier et al., 2020] .

3.2 FROG : Méthode de recalage d'ensemble d'images 3D

FROG (Fast Registration Of image Groups) [Agier et al., 2020], est un algorithme de recalage d'image permettant de recaler plusieurs images simultanément, dans un espace commun. Cette méthode commence par extraire des points d'intérêts dans les zones homogènes des images à l'aide de l'algorithme SURF [Bay et al., 2006]. Ensuite, le voisinage de chaque point d'intérêt est représenté par un vecteur de caractéristiques appelé descripteur. Les descripteurs des points d'intérêt sont alors mis en correspondance deux par deux puis le recalage optimise les transformations pour chaque image afin de minimiser les distances entre les paires de points ayant des descripteurs similaires dans un espace commun. Un schéma représentant ce fonctionnement est présenté dans la Figure 2.4. Cette méthode de recalage a pour avantage d'être rapide et de faible complexité car le recalage se fait sur un petit nombre de points d'intérêt et non en résolution pleine.

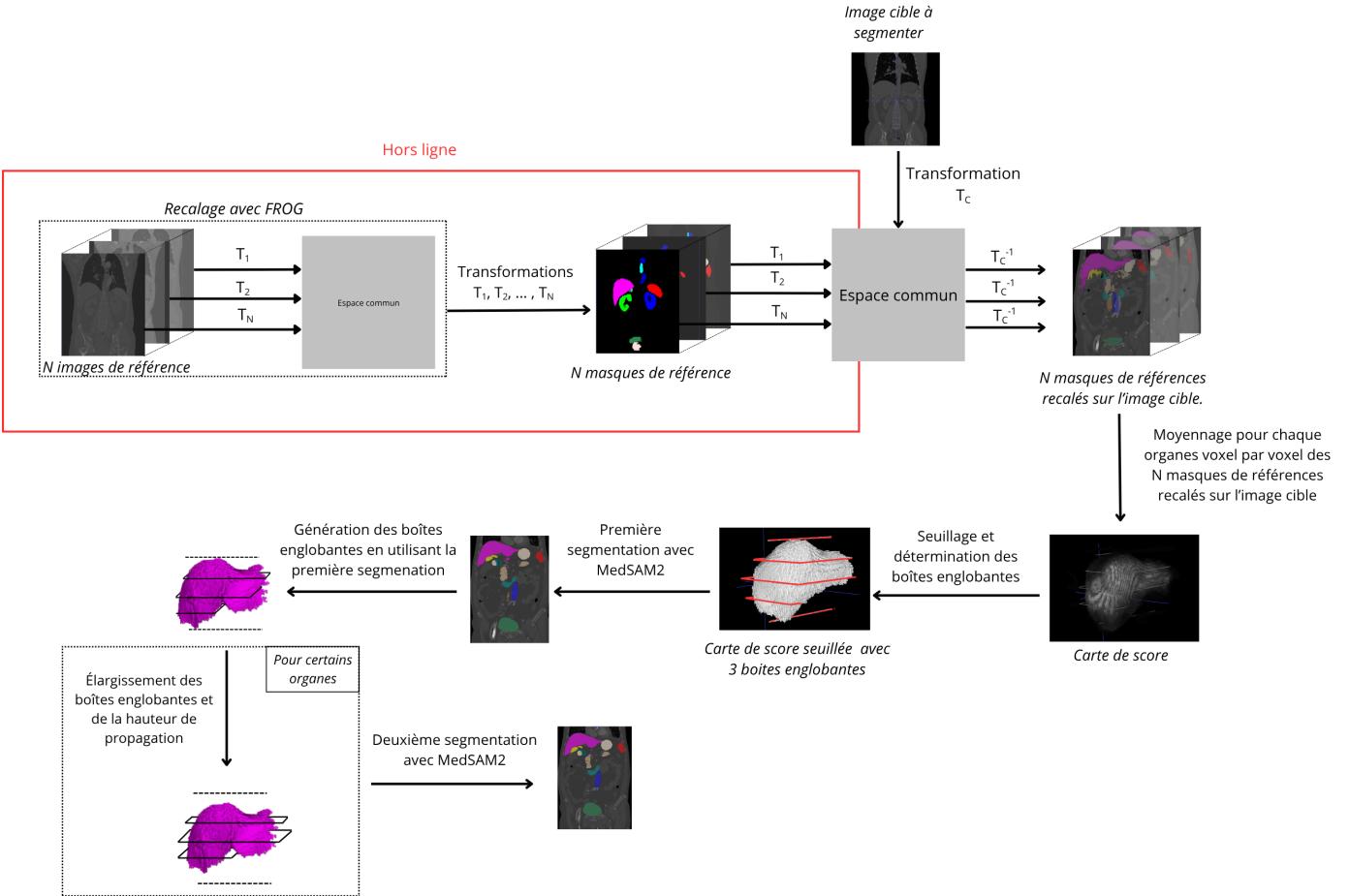


FIGURE 2.5 – Schéma de notre méthode de segmentation automatique multi-organes avec le modèle de fondation MedSAM2.

3.3 Segmentation multi-organes automatique avec MedSAM2

Dans cette section est détaillée notre démarche de segmentation automatique multi-organes d’images avec MedSAM2. Notre méthode utilise le recalage de masques de référence sur l’image à segmenter pour la génération d’indications nécessaires au modèle MedSAM2.

3.3.1 Recalage et génération automatique d’indications

Pour mettre en œuvre cette méthode, la première étape consiste à isoler un groupe de 15 images de référence. Ces images présentent un champ de vue complet et proviennent de patients ne présentant pas de pathologies susceptibles d’influencer la forme, la position ou l’apparence des organes abdominaux à segmenter. Ces 15 images ont ensuite été recalées toutes ensemble dans un espace commun à l’aide de FROG. Il en résulte 15 transformations notées $T_1, T_2, T_3, \dots, T_{15}$, comme montré dans le cadre rouge de la Figure 2.5. De la même façon, l’image cible à segmenter est également recalée dans l’espace commun, on obtient donc une transformation T_c . Les masques multi-organes du groupe des images de référence, appelés masques de référence et notés $M_1, M_2, M_3, \dots, M_{15}$ sont ensuite tous recalés dans l’espace commun à l’aide des transformations $T_1, T_2, T_3, \dots, T_{15}$, puis alignés sur l’image cible à l’aide de la transformation inverse de T_c , T_c^{-1} .

$$M'_i = T_c^{-1}(T_i(M_i)).$$

Les organes à segmenter sont alors isolés un par un des masques de références recalés sur l’image cible $M'_1, M'_2, M'_3, \dots, M'_{15}$. Les masques de références sont ensuite moyennés voxel par voxel. Il en résulte autant de cartes de scores que d’organes à segmenter. La Figure 2.5 illustre cette démarche ainsi qu’une carte de

score. Ces cartes de scores sont des volumes dans lesquels chaque voxel contient une valeur comprise entre 0 et 1. Par exemple, une valeur de 1 pour un voxel signifie que tous les masques recalés attribuent ce voxel au même organe, tandis qu'une valeur de 0,5 indique que la moitié des masques recalés l'attribuent à cet organe. Ainsi, un score élevé pour un organe dans un voxel signifie une forte probabilité de présence de l'organe en question dans le voxel.

Ces cartes de scores sont ensuite seuillées différemment en fonction des organes. Les différents seuils ont été déterminés expérimentalement et traduisent de la performance du recalage sur les différents organes ainsi que de leur proximité avec d'autres organes ou d'autres structures. Plus de détails sur les seuils sont donnés dans la Section 7. Une fois les cartes de scores seuillées (les voxels avec un score inférieur ou égal au seuil prennent comme valeur 0 et les autres prennent la valeur 1), il reste pour chaque organe une image 3D binaire. De cette image binaire on peut alors déterminer l'altitude maximale Z_{max} ainsi que l'altitude minimale Z_{min} sur le plan axial (plan perpendiculaire à l'axe crâno-caudal) de l'organe à segmenter. Ensuite, les boîtes englobantes 2D peuvent être générées sur certaines coupes en prenant le plus petit rectangle défini avec 2 coordonnées, ayant les bords parallèles aux bordures de l'image, contenant la carte de score binarisée. Concernant le nombre de boîtes englobantes par organe, deux stratégies ont été testées. La première consiste à générer une seule boîte englobante par organe à l'altitude $\frac{Z_{max}-Z_{min}}{2}$ (désignée ensuite par la **Stratégie A**) et la seconde revient à générer 3 boîtes englobantes par organe : à l'altitude $\frac{Z_{max}-Z_{min}}{4}$, $\frac{Z_{max}-Z_{min}}{2}$ et $\frac{3(Z_{max}-Z_{min})}{4}$. Cette seconde stratégie, désignée **Stratégie B** ensuite, est la stratégie représentée sur la Figure 2.6.

3.3.2 MedSAM2 : Première segmentation

Après que les boîtes englobantes soient générées, elles sont injectées comme indications (ou prompts) dans le modèle MedSAM2 afin de segmenter l'image cible. Quel que soit le nombre de boîtes englobantes définies, les masques de segmentation sont toujours propagés en partant de la coupe du milieu jusqu'à la coupe supérieure d'altitude Z_{max} (flèche supérieure sur la Figure 2.6) puis de la coupe du milieu à la coupe inférieure d'altitude Z_{min} (flèche inférieure sur la Figure 2.6). Même si plusieurs boîtes englobantes sont données entre le milieu de l'organe et son extrémité, la banque de mémoire n'est pas réinitialisée pour assurer la continuité de la segmentation en permettant au modèle d'utiliser les segmentations précédentes pour prédire les nouvelles segmentations.

Dans le cadre d'un problème de segmentation multi-organes avec N organes, une stratégie d'inférence en deux temps est adoptée. D'abord l'image 3D est encodée une seule fois dans l'encodeur. Ensuite de façon indépendante pour chaque organe, les boîtes englobantes sont propagées dans l'encodeur d'indications et le masque de segmentation 3D est généré grâce au module d'attention avec mémoire et au décodeur de masque. A chaque fois, le modèle donne une carte de prédiction ayant les mêmes dimensions que l'image à segmenter, contenant les valeurs de prédictions brutes pour chaque voxel. Plus la valeur de prédiction pour un voxel est élevée, plus le voxel a de chances de contenir l'organe indiqué avec la boîte englobante. Une fois les N cartes de prédictions générées (une par organe), une fonction Softmax est appliquée voxel par voxel à travers les N cartes, de façon à normaliser les valeurs de prédiction en probabilité d'appartenance à chacune des N classes. Finalement, pour chaque voxel, la classe assignée correspond à celle ayant la plus grande probabilité, produisant la segmentation finale multi-organe. Les plus grandes composantes connexes de chaque organe sont ensuite retenues. Cela permet d'éliminer les petites régions isolées, qui sont souvent des bruits ou des faux positifs.

3.3.3 Segmentation en cascade : affinage de la première segmentation

Une fois une première segmentation déterminée, une deuxième segmentation est générée toujours avec MedSAM2 mais en utilisant cette fois-ci les masques de vérités de la première segmentation pour déterminer les altitudes Z_{min} et Z_{max} de chaque organe ainsi que les boîtes englobantes (**Stratégie C**). Le mécanisme de génération de prompt est le même qu'avec la carte de score et la méthode de segmentation reste identique. Cette segmentation en cascade permet d'affiner la première segmentation car les boîtes englobantes sont mieux placées autour des organes. De plus, pour régler certaines sous-segmentations lors de la première inférence, qui seraient répétées lors de la deuxième inférence à cause du positionnement des boîtes englobantes et des

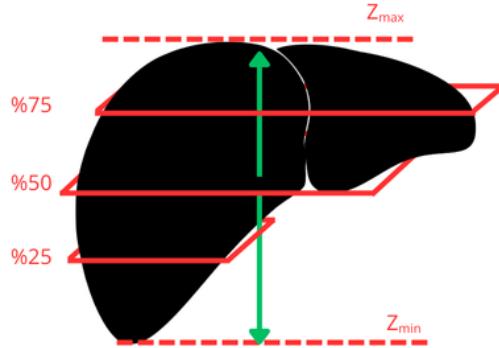


FIGURE 2.6 – MedSAM2 : Stratégie de génération des boîtes englobantes et de propagation des masques de segmentation.

altitudes maximales et minimales avec la première inférence, l'aire des boîtes englobantes est agrandie de 10% et la hauteur de propagation $Z_{\max} - Z_{\min}$ est augmentée de 20%. Ceci est appliqué à tous les organes sauf l'aorte, la veine cave inférieure et les glandes surrénales et est désigné ensuite par la **Stratégie D**. Cette stratégie est d'autant plus pertinente sachant que MedSAM2 a été entraîné en prenant des boîtes soit de la taille de l'organe soit plus larges, mais jamais plus petites. De plus, si les bornes de propagation dépassent l'organe et que les segmentations sur les coupes précédentes se font bien dans l'organe cible, le modèle MedSAM2 est capable de détecter la non-présence de l'organe en ne segmentant rien, car il a été entraîné de la sorte. Plus de détails concernant cette stratégie sont donnés dans la section 5.2.3.

4 Résultats

4.1 Base de données d'évaluation : AMOS22

Notre démarche et ses différentes stratégies ont été évaluées sur 30 images de la base de données AMOS22 [Ji et al., 2022]. Cette base de données contient des images tomodensitométriques 3D (images CT) et des images IRMs (non étudiées ici) abdominales. Les volumes de la base de données sont annotés avec 15 structures citées dans le Tableau 2.1 avec leurs abréviations utilisées par la suite. Dans la coupe axiale (plan perpendiculaire à l'axe crâno-caudal) les images ont une résolution de 512x512 voxels ou 768x768 voxels, avec un espacement (dimensions des voxels) d'environ 0.75mm. Selon l'axe crâno-caudal (axe vertical), les volumes ont un espacement de 2 ou 5mm.

4.2 Performance de référence de MedSAM2 : Segmentation en générant les boîtes englobantes avec les masques de vérité.

Afin d'évaluer les performances maximales de MedSAM2 lorsque les indications sont parfaitement générées, nous avons en premier segmenté 30 images de la base de données AMOS22 [Ji et al., 2022] en déterminant les altitudes maximales et minimales des organes ainsi que leur(s) boîte(s) englobante(s) à l'aide des masques de vérité. Nous avons utilisé deux stratégies pour donner les indications au modèle définies dans la Section 3.3.1. La première était d'utiliser une seule boîte englobante par organe (**Stratégie A**) et la deuxième était d'en utiliser trois (**Stratégie B**). La qualité de la segmentation a été mesurée à l'aide de l'indice de Dice reporté dans la deuxième partie du Tableau 2.2. Cet indice permet d'évaluer le recouvrement de deux masques A et B par la formule suivante :

$$\text{Dice}(A, B) = \frac{2 \times |A \cap B|}{|A| + |B|}$$

Une valeur de 1 correspond à un recouvrement parfait et une valeur de 0 correspond à aucun recouvrement.

Abréviation	Organe
SPL	Rate (Spleen)
RKI	Rein droit (Right Kidney)
LKI	Rein gauche (Left Kidney)
GBL	Vésicule biliaire (Gallbladder)
ESO	(Esophage (Esophagus)
LIV	Foie (Liver)
STO	Estomac (Stomach)
AOR	Aorte (Aorta)
IVC	Veine cave inférieure (Inferior Vena Cava)
PAN	Pancréas (Pancreas)
RAG	Angle colique droit (Right Colon Angle)
LAG	Angle colique gauche (Left Colon Angle)
DUO	Duodénum (Duodenum)
BLA	Vessie (Bladder)
PRO/UTE	Prostate / Utérus (Prostate / Uterus)

TABLE 2.1 – Organes annotés dans AMOS22 et leur abbréviations.

4.3 Segmentation en générant les boîtes englobantes de façon automatique.

Nous avons exploré plusieurs stratégies de segmentation dérivant de la méthode présentée précédemment sur les mêmes 30 images de la base de données AMOS22 :

- **Stratégie A : Une segmentation avec une seule boîte englobante** par organe générée en utilisant la carte de score seuillée obtenue après le recalage. Les boîtes englobantes sont générées au milieu de chaque organe sur la coupe axiale.
- **Stratégie B : Une segmentation avec trois boîtes englobantes** par organe générées en utilisant la carte de score seuillée obtenue après le recalage. Les boîtes englobantes sont générées au quart, au milieu et aux trois quarts de la carte de scores seuillée de chaque organe sur la coupe axiale comme schématisé sur la Figure 2.6.
- **Stratégie C : Segmentations en cascade** : Deux segmentations en générant à chaque fois trois boîtes englobantes par organes. Pour la seconde segmentation les boîtes englobantes sont générées avec les masques de la première segmentation. La seconde segmentation permet d'affiner la première.
- **Stratégie D : Segmentations en cascade** : deux segmentations en générant à chaque fois trois boîtes englobantes par organes et en agrandissant de 10% l'aire des boîtes englobantes et de 20% la hauteur de la propagation pour la deuxième segmentation sauf pour les glandes surrenales, l'aorte et la veine cave inférieure.

Les résultats de ces 4 stratégies sont présentés dans la dernière partie du Tableau 2.2.

Organes	U-Net		Performances de références : boîte(s) englobante(s) définie(s) avec les masques de vérité.		Stratégies automatiques de détermination des boîtes englobantes.			
	U-Net 2D	U-Net 3D	1 boîte	3 boîtes	Stratégie A	Stratégie B	Stratégie C	Stratégie D
SPL	0.91	0.92	0.96 ± 0.02	0.96 ± 0.02	0.86 ± 0.14	0.85 ± 0.14	0.85 ± 0.14	0.89 ± 0.14
RKI	0.90	0.91	0.92 ± 0.02	0.95 ± 0.02	0.91 ± 0.18	0.90 ± 0.18	0.91 ± 0.18	0.92 ± 0.18
LKI	0.93	0.94	0.96 ± 0.02	0.95 ± 0.02	0.95 ± 0.03	0.93 ± 0.05	0.94 ± 0.03	0.95 ± 0.03
GBL	0.76	0.77	0.84 ± 0.04	0.91 ± 0.03	0.35 ± 0.28	0.37 ± 0.27	0.37 ± 0.27	0.43 ± 0.27
ESO	0.77	0.77	0.81 ± 0.07	0.82 ± 0.07	0.66 ± 0.24	0.61 ± 0.25	0.62 ± 0.25	0.63 ± 0.25
LIV	0.95	0.95	0.94 ± 0.04	0.93 ± 0.04	0.91 ± 0.17	0.85 ± 0.16	0.87 ± 0.15	0.87 ± 0.15
STO	0.77	0.81	0.84 ± 0.08	0.88 ± 0.07	0.60 ± 0.20	0.60 ± 0.20	0.63 ± 0.22	0.65 ± 0.22
AOR	0.91	0.92	0.89 ± 0.09	0.93 ± 0.05	0.87 ± 0.11	0.88 ± 0.12	0.88 ± 0.12	0.88 ± 0.1
IVC	0.82	0.87	0.81 ± 0.17	0.85 ± 0.11	0.78 ± 0.16	0.80 ± 0.12	0.81 ± 0.12	0.80 ± 0.12
PAN	0.76	0.81	0.74 ± 0.16	0.80 ± 0.12	0.56 ± 0.22	0.63 ± 0.21	0.60 ± 0.21	0.63 ± 0.21
RAG	0.66	0.70	0.63 ± 0.16	0.69 ± 0.11	0.45 ± 0.24	0.46 ± 0.22	0.50 ± 0.23	0.50 ± 0.23
LAG	0.67	0.71	0.70 ± 0.10	0.75 ± 0.08	0.55 ± 0.23	0.54 ± 0.22	0.55 ± 0.22	0.54 ± 0.22
DUO	0.69	0.74	0.61 ± 0.15	0.67 ± 0.09	0.47 ± 0.21	0.52 ± 0.19	0.49 ± 0.22	0.50 ± 0.22
BLA	0.82	0.82	0.87 ± 0.12	0.88 ± 0.15	0.61 ± 0.30	0.57 ± 0.29	0.60 ± 0.31	0.61 ± 0.31
PRO/UTE	0.74	0.73	0.85 ± 0.05	0.87 ± 0.04	0.44 ± 0.25	0.49 ± 0.24	0.51 ± 0.28	0.53 ± 0.28
Moyenne	0.80 ± 0.09	0.82 ± 0.09	0.82 ± 0.11	0.86 ± 0.09	0.66 ± 0.19	0.67 ± 0.18	0.68 ± 0.18	0.69 ± 0.17

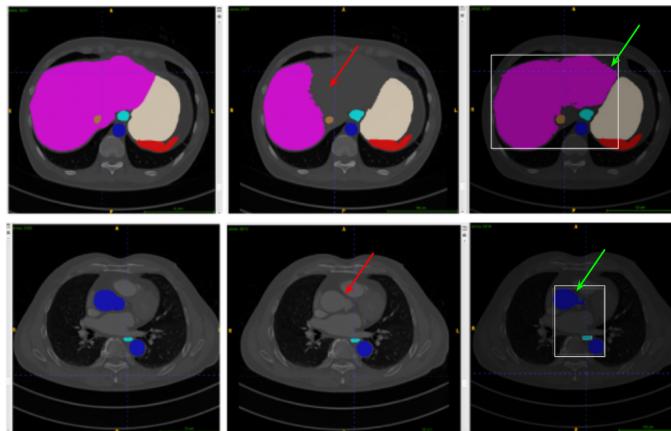
TABLE 2.2 – Colonnes 1 et 2 : Indices de Dice par organes obtenus à l'aide de modèles spécialisés d'architecture type U-Net 2D et U-Net 3D. En bleu sont indiquées la meilleure performance par organe pour ces deux architectures. Colonnes 3 et 4 : Indices de Dice obtenus avec MedSAM2 en générant les indications à l'aide des masques de vérité. En rouge sont indiquées la meilleure performance par organe en fonction du nombre de boîte englobante. Colonnes 5 à 9 : Indices de Dice obtenus avec MedSAM2 selon les stratégies A,B,C et D. Les scores en vert représentent la meilleure performance par organe pour ces 4 stratégies.

5 Discussion

5.1 Segmentation en générant les boîtes englobantes avec les masques de vérité : performance de référence de MedSAM2.

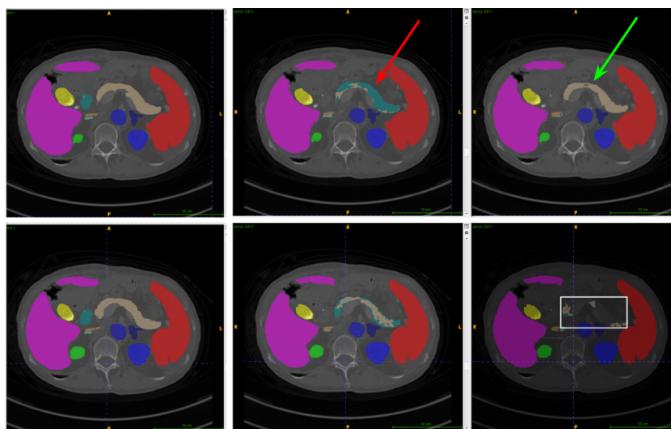
Comme le montre la deuxième partie du Tableau 2.2, concernant la segmentation à l'aide des masques de vérité, la segmentation est meilleure lorsque l'on donne 3 boîtes englobantes au modèle. La différence de performance est d'autant plus importante pour les organes qui n'ont pas une forme patatoïde et régulière comme le pancréas, le duodénum, l'estomac, la vésicule biliaire et les glandes surrénales ainsi que pour les organes ayant une forme tubulaire comme l'aorte ou la veine cave inférieure. Ainsi, pour les formes complexes ou allongées, le modèle est plus performant lorsqu'il a plus d'indications. L'ajout d'indications compense l'ambiguïté des boîtes englobantes des organes à la forme non convexe qui peuvent dépasser sur une structure voisine. On peut en effet voir sur les images de la Figure 2.7b que la boîte englobante positionnée aux trois quarts de la carte de score seuillée du pancréas, a permis au modèle de le différencier du duodénum. Cette confusion est due à la proximité du duodénum avec le pancréas et donc au chevauchement de leurs boîtes englobantes. L'ajout de boîtes englobantes au quart et trois quarts de la carte des score seuillée de l'organe permet également d'injecter un a priori de forme sur l'organe, comme on peut le constater pour le foie sur la Figure 2.7a en haut à droite. De plus, l'ajout de boîtes englobantes permet de capturer différentes parties de l'organe en cas de discontinuité comme c'est le cas pour l'aorte dans la Figure 2.7a en bas à droite.

Concernant les performances de segmentation de MedSAM2, on peut observer avec le Tableau 2.2, que le modèle, lorsqu'il est guidé avec des indications précises, dépasse les performances d'un modèle spécialisé entraîné sur des images similaires aux images cibles pour la tâche de segmentation. Le modèle de fondation permet donc, lorsqu'il a de bonnes indications, de réaliser des segmentations de qualité similaire, voire plus importante qu'un modèle spécialisé, sans entraînement.



(a) A gauche, masques de vérité. Au milieu, masques prédits avec une boîte englobante par organe. A droite : masques prédits avec trois boîtes englobantes par organe. Les boîtes englobantes placées au trois quart de la carte de score seuillée du foie et de l'aorte sont représentées respectivement sur la ligne du haut et du bas.

■ : Aorte, ■ : Foie



(b) Sur la ligne du haut : A gauche masques de vérité. Au milieu, masques prédits avec une boîte englobante par organe. A droite : masques prédits avec trois boîtes englobantes par organe. Sur la ligne du bas : les mêmes images, une coupe en dessous. Sur l'image de droite est représentée la boîte englobante placée au trois quart de la carte de score seuillée du pancréas. ■ : Pancréas, ■ : Duodénium

FIGURE 2.7 – (a) et (b) Segmentations avec MedSAM2 : comparaison entre l’ajout d’une boîte englobante et de trois boîtes englobantes par organes définies avec les masques de vérité.

5.2 Segmentation en générant les boîtes englobantes de façon automatique.

5.2.1 Stratégie A et B : Une seule segmentation avec une boîte englobante ou trois boîtes englobantes

Tout d’abord, avec la troisième partie du Tableau 2.2, le même constat que pour la génération de boîtes englobantes à l’aide des masques de vérité peut être tiré : la moyenne des scores de segmentation est meilleure avec 3 boîtes qu’avec une seule boîte, pour les mêmes raisons qu’énoncées précédemment. En revanche, pour certains cas, l’ajout de boîtes englobantes en haut et en bas de l’organe peut engendrer des erreurs de segmentation lorsque le recalage est mauvais à ces endroits. On peut par exemple voir sur la Figure 2.8a à gauche, que dans la partie supérieure du foie, la carte de score, même après seuillage, déborde sur le poumon. Cela introduit une forme de confusion dans le processus de prédiction du modèle qui va alors segmenter le poumon à la place du foie. Même constat sur la Figure 2.8b où l’on peut voir que les recalages des reins gauches se sont faits trop haut par rapport au patient cible et donc l’ajout de la boîte englobante en haut du rein gauche fausse les prédictions.

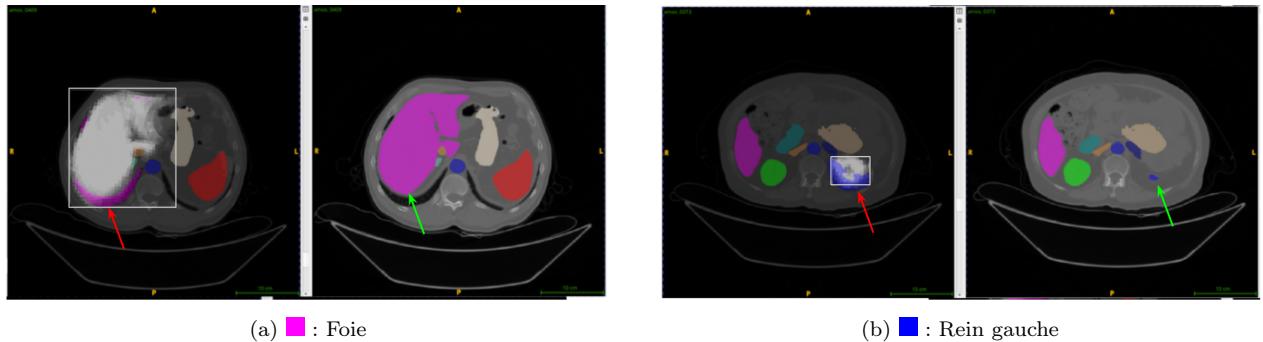
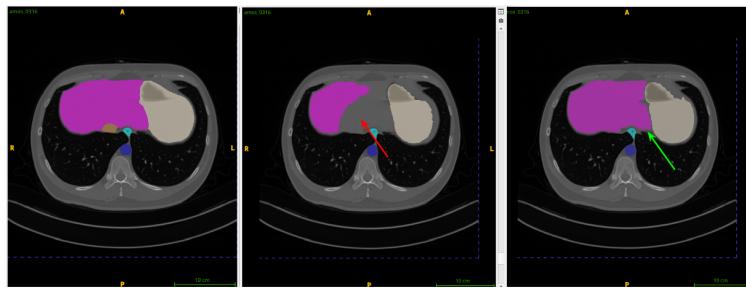


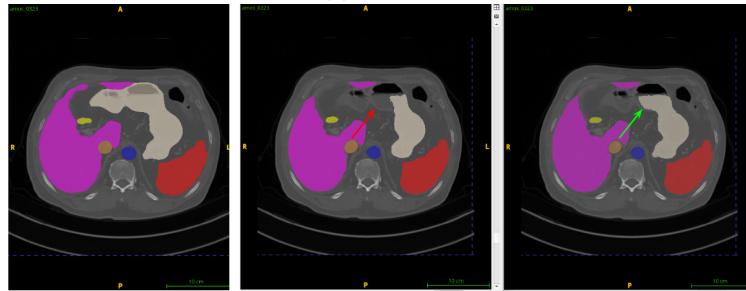
FIGURE 2.8 – (a) et (b) A gauche, segmentations générées avec trois boîtes englobantes avec la carte de score et la boîte englobante supérieure du foie (a) et de du rein gauche (b) respectivement. A droite, segmentations générées avec une boîte englobante.

5.2.2 Stratégie C : Deux segmentations en cascade

Le fait d'affiner la première segmentation à l'aide d'une deuxième segmentation en utilisant les masques de vérités générés par la première inférence permet d'améliorer les scores de segmentation et d'affiner les masques obtenus. En effet, la première segmentation permet de localiser les organes cibles de façon plus précise que la carte des scores. Ainsi, les boîtes englobantes pour la deuxième segmentation et les altitudes maximales et minimales de propagation sont placées de façon plus juste et ont plus de chances d'entourer l'organe. Cela permet d'avoir une deuxième segmentation plus fine car mieux guidée comme c'est le cas pour le foie et l'estomac dans la Figure 2.9. En revanche, pour les organes mal segmentés dont les boîtes englobantes de la première segmentation sont mal définies, la deuxième segmentation ne permet pas d'améliorer la première segmentation. Par exemple, dans la Figure 2.10 la rate et l'estomac sont toujours sous-segmentés après la deuxième inférence. Cependant, la deuxième segmentation n'est pas plus mauvaise que la première. Le pancréas et le duodénum en revanche s'écartent de ce constat. En effet, le pancréas et le duodénum sont deux structures très proches. Ce qui fait que les boîtes englobantes du pancréas lors de la première segmentation peuvent déborder sur le duodénum et inversement. Ainsi, lors de la première segmentation, les deux organes sont confondus, comme on peut le voir sur la Figure 2.11 à gauche : la segmentation du duodénum a débordé sur celle du pancréas. Lors de la seconde segmentation, cette confusion s'empire car la boîte englobante dépasse franchement sur le pancréas (Figure 2.11 au milieu) ce qui fait que sur les coupes supérieures à la boîte englobante, le duodénum est segmenté dans le pancréas (Figure 2.11 à droite).

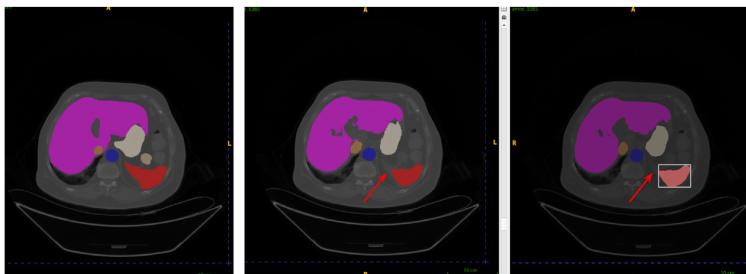


(a) ■ : Foie

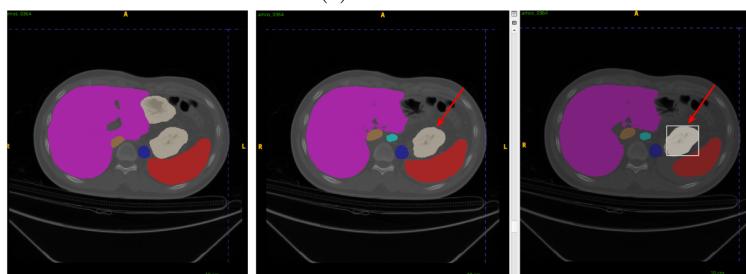


(b) □ : Estomac

FIGURE 2.9 – (a) et (b) A gauche, les masques de vérité. Au milieu, segmentations générées avec trois boîtes englobantes par organes définies avec les cartes de score (Stratégie B). A droite, segmentations générées après une deuxième inférence en utilisant les résultats de la première inférence pour générer les boîtes englobantes (Stratégie C).



(a) ■ : Rate



(b) □ : Estomac

FIGURE 2.10 – (a) et (b) A gauche, masques de vérité. Au milieu, segmentations générées avec les boîtes englobantes définies avec les cartes de score (Stratégie B). A droite, segmentations générées après une deuxième inférence en utilisant les résultats de la première inférence pour générer les boîtes englobantes (Stratégie D). Une boîte englobante de la rate (a) et de l'estomac (b) y sont respectivement présentés. A droite, les masques de vérité.

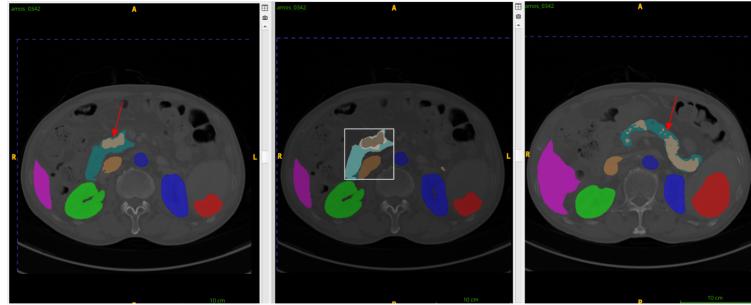


FIGURE 2.11 – A gauche : segmentations générées avec les boîtes englobantes définies avec les cartes de score (Stratégie B). Au milieu : Mêmes segmentations avec la boîte englobante du duodénum utilisée à la deuxième inférence. A droite : image trois coupes au dessus après une deuxième inférence (Stratégie D). ■ : Pancréas, ■ : Duodénum

5.2.3 Stratégie D : Deux segmentations en cascade en augmentant l'aire des boîtes englobantes et la hauteur de propagation pour la deuxième segmentation

La segmentation en cascade admet quelques limites. Dans la majorité des cas, les bornes maximales et minimales de l'organe sont mal déterminées avec la carte de score seuillée. Cela induit des sous-segmentations selon l'axe crano-caudal lors de la première inférence. Ces sous-segmentations se reproduisent lors de la deuxième segmentation car les bornes de propagation sont définies avec les masques prédits lors de la première inférence. De plus, si les bornes de propagation dépassent l'organe et que les segmentations sur les coupes précédentes se font bien dans l'organe cible, le modèle MedSAM2 est capable de détecter la non-présence de l'organe en ne segmentant rien. C'est pour cela que la deuxième segmentation est améliorée en augmentant de 20% la hauteur de propagation. Cette augmentation doit se faire proportionnellement à la hauteur de la première segmentation car plus celle-ci est longue, plus l'organe a de chances d'être étiré et donc un plus grand nombre de coupes sont à explorer au-dessus et en dessous des bornes de segmentation.

Lors de la première inférence, on peut également observer des sous-segmentations sur le plan axial, lorsque les boîtes englobantes définies par les cartes de scores se trouvent à l'intérieur de l'organe. En plus de cela, le modèle MedSAM2 a été entraîné en faisant varier la taille des boîtes englobantes, mais celles-ci ont toujours été définies systématiquement plus larges que l'organe. Ainsi, le modèle est capable de générer de meilleures segmentations lorsque les boîtes englobantes sont trop grandes et non serrées autour de l'organe que lorsque les boîtes englobantes sont trop petites et tombent à l'intérieur de l'organe. Ce sont pour ces deux raisons que lors de la deuxième segmentation, les résultats sont améliorés lorsque l'aire des boîtes englobantes est augmentée de 10%.

L'augmentation de la taille des boîtes englobantes ainsi que de la hauteur de propagation permet donc pour la plupart des organes d'améliorer les résultats de segmentation comme on peut observer sur le Tableau 2.2 et sur la Figure 2.12. En revanche, pour les structures très petites, l'augmentation de la hauteur de la segmentation mène à des moins bonnes performances. En effet, le recalage des glandes surrénales est très mauvais : ces organes sont assez petits pour qu'il n'y ait peu de chances que des points d'intérêts soient tirés à l'intérieur. Ainsi, la carte de score est très étirée selon l'axe crano-caudal (Figure 2.13a au milieu), ce qui fait que les bornes de propagation sont trop grandes et que des boîtes englobantes sont données à des endroits où l'organe n'est pas présent. Cela mène à des sur-segmentations lors de la première inférence selon l'axe crano-caudal (Figure 2.13a au milieu). Ainsi, l'augmentation de la hauteur de propagation lors de la deuxième segmentation empire les cartes de segmentation (Figure 2.13a à droite).

Le cas de la veine cave inférieure et de l'aorte est un peu particulier. En effet, les masques de vérité donnés dans la base de données AMOS22 s'arrêtent dans le cas de l'aorte à la bifurcation aortique et dans le cas de la veine cave inférieure à la confluence des deux veines iliaques. Or, comme ces deux organes se prolongent encore après ces deux éléments, lorsque l'on augmente la hauteur de propagation, MedSAM2 va segmenter l'aorte et la veine cave inférieure en dessous de la bifurcation aortique et de la confluence des veines iliaques respectivement : Figure 2.13b à droite. La segmentation est juste mais cela ne se reflète pas lorsque l'on calcule les indices de Dice.

Les résultats présentés dans le Tableau 2.2 sont ceux obtenus lorsque la hauteur de propagation n'est pas augmentée pour la deuxième segmentation pour les deux glandes surrénales, l'aorte et la veine cave inférieure.

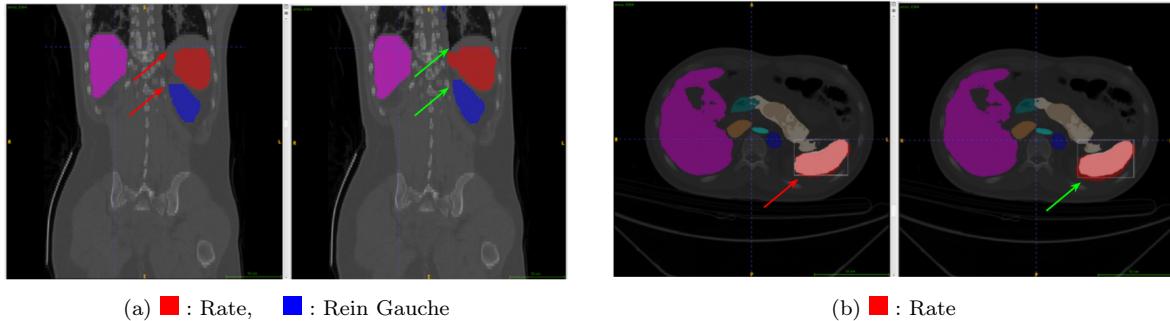


FIGURE 2.12 – (a) et (b) A gauche, segmentations après deux inférences sans augmenter la taille des boîtes englobantes ou la profondeur de propagation (Stratégie C). A droite, segmentations après deux inférences en augmentant la hauteur de propagation de 20% et l'aire des boîtes englobantes de 10% définies avec la première segmentation (Stratégie D).

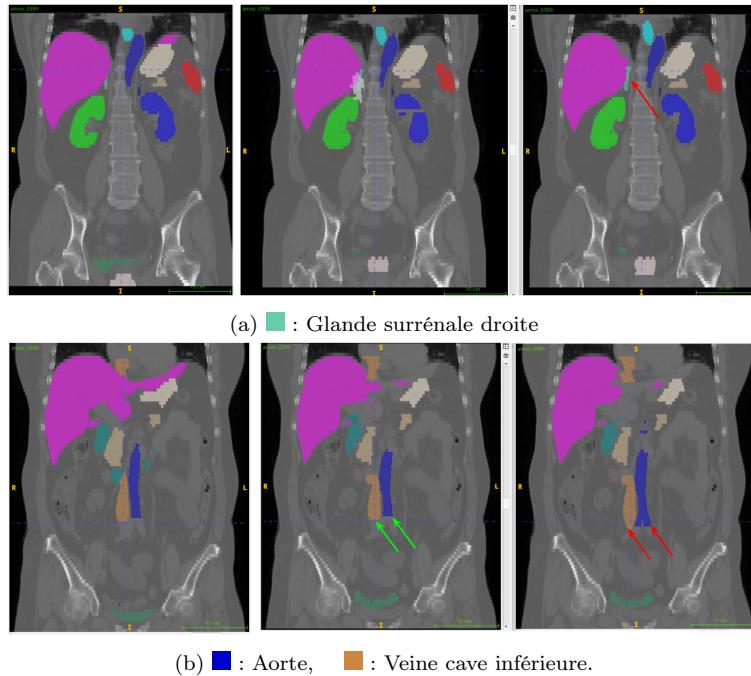


FIGURE 2.13 – (a) et (b) À gauche : Masques de vérité. Au milieu, segmentations obtenues à partir des boîtes englobantes générées via les cartes de score (stratégie B). A droite : segmentations après une seconde inférence en utilisant les résultats de la première pour générer de nouvelles boîtes englobantes et en augmentant la hauteur de propagation de 20%. (a) Sur la figure du milieu, la carte des scores de la glande surrénale droite y est représenté.

5.2.4 Comparaison des performances de notre méthode de segmentation à celles de modèles spécialisés.

Comparé aux modèles spécialisés d'architecture U-Net 2D et U-Net 3D évalués sur les mêmes 30 images, notre méthode a des performances similaires pour les reins et dans les mêmes ordres de grandeur pour la rate, le foie et la veine cave inférieure : Tableau 2.2. En revanche, pour les autres organes, les scores obtenus sont inférieurs. Cela se comprend par le fait que les organes comme l'estomac, le pancréas, le duodénum ou la vessie ont des formes et des positions très variables d'un patient à l'autre. L'estomac, par exemple, peut être vide ou rempli, ce qui fait qu'il peut avoir un volume très variable. Pour la vessie, sa position et sa forme varient en fonction du sexe des patients. Pour un homme, la vessie se situe devant le rectum, et au-dessus de

la prostate alors que pour une femme, la vessie se trouve devant le vagin, et sous l'utérus. Pour les organes cités précédemment, la création d'un atlas de référence est alors très compliquée au vu de leur variabilité anatomique. Concernant la vésicule biliaire ainsi que les glandes surrénales, ce sont des organes très petits. Les glandes surrénales, par exemple, ont une épaisseur d'environ 3 cm, ce qui correspond à environ 6 voxels dans les images ayant un espacement de 5 mm et 15 voxels pour celles ayant un espacement de 2 mm selon l'axe crano-caudal. Le recalage de ces organes est donc assez mauvais car peu de points d'intérêts tombent à l'intérieur de ces organes. Pour la prostate et l'utérus, ces deux organes sont labellisés identiquement dans la base de données AMOS22, alors qu'ils ont une forme et une position différentes. Il n'est donc pas possible de générer avec un recalage des boîtes englobantes assez précises pour ces organes pour permettre à MedSAM2 de réaliser des segmentations aussi performantes que celles de l'état de l'art.

5.3 Traitement multi-organe

MedSAM2 ne permet de traiter qu'un seul objet à la fois. Une inférence distincte est donc effectuée pour chaque organe. Lors de chaque inférence, le modèle génère une carte de prédiction, dans laquelle les valeurs de prédiction représentent le degré de confiance quant à la présence de l'organe ciblé. Ces cartes de prédiction sont ensuite normalisées à l'aide de la fonction Softmax, produisant une carte de probabilités. La classe correspondante est finalement déterminée en identifiant, pour chaque voxel, la probabilité maximale. Ce choix se base sur l'a priori qu'un voxel ne peut appartenir qu'à une seule classe et permet de rendre la segmentation plus robuste. En effet, lorsque la boîte englobante d'un organe déborde sur un autre organe, le modèle a tendance à donner des valeurs de prédiction positives sur cet autre organe. Par exemple, lorsque la boîte englobante de la rate déborde sur le rein gauche, comme dans la Figure 2.14a, le modèle prédit des valeurs de prédiction pas très grandes mais positives au niveau du rein gauche, Figure 2.14b. Ainsi, si la rate est traitée toute seule, la segmentation de la rate engloberait le rein gauche et la rate. En revanche, lorsque le rein gauche est traité simultanément, le modèle génère des scores plus élevés avec la boîte englobante du rein gauche au niveau du rein gauche qu'avec celle de la rate. On peut observer ce phénomène sur la Figure 2.14c. En prenant les valeurs de prédiction maximales, le rein gauche et la rate seront segmentés correctement.

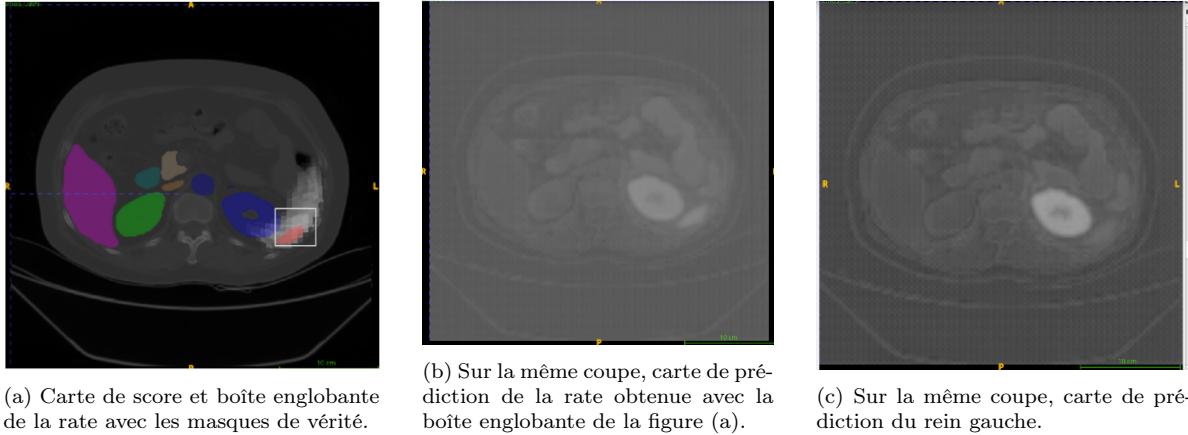


FIGURE 2.14 – Avantage du traitement multi-organe. ■ : Rate, ■ : Rein Gauche

5.4 Impact énergétique, environnemental et temps d'exécution

L'avantage de notre méthode du point de vue énergétique et en termes de ressources de calcul est qu'elle ne nécessite pas d'entraînement contrairement à un modèle spécialisé. Par exemple, pour obtenir les résultats présentés dans le Tableau 2.2 le modèle d'architecture U-Net 2D a été entraîné 13h30 et le modèle U-Net 3D 13h00 sur 300 images de la base de données d'entraînement d'AMOS22 rééchantillonnée à une taille de 256x256 sur la coupe axiale. On peut estimer que ces entraînements ont émis respectivement 0.351 kg eq CO₂ et 0.336 kg eq CO₂. Les émissions de gaz à effet de serre liées à l'entraînement sont notées E_{entraînement}. Ces estimations sont réalisées par le centre de calcul dans lequel les modèles ont été entraînés (Jean Zay) et se

basent sur la méthodologie développée par [Benaben et al. \[2024\]](#) du collectif Labos 1point5. Ils prennent en compte les infrastructures de calcul et les autres équipements du centre, le type d'alimentation, le système de refroidissement, le bâtiment et le personnel. En revanche, cette estimation de coût d'entraînement doit être mise en perspective avec le coût d'inférence des modèles.

Pour estimer les émissions de gaz à effet de serre lors de l'inférence ($E_{\text{inférence}}$) des modèles U-Net 2D et U-Net 3D ainsi que celui de MedSAM2 de façon équitable, nous avons utilisé l'outil CodeCarbon [\[Courty et al., 2024\]](#). Cet outil prend en compte la consommation GPU, la consommation CPU, la consommation de la RAM et le mix énergétique du pays dans lequel a été faite l'expérience. Nous avons évalué les différentes méthodes sur une même image de taille 512x512x235 et sur la configuration matérielle décrite dans le Tableau [2.4](#).

Dans le cadre de notre méthode et de la segmentation en cascade avec le modèle MedSAM2, une émission de gaz à effet de serre de 0.36 g eq CO₂ a été mesurée. En comparaison, la segmentation de la même image redimensionnée à une échelle réduite de résolution 256×256×235 voxels par le modèle d'architecture U-Net 2D ou U-Net 3D entraîne une émission de 0.02 g eq CO₂.

L'émission totale ($E_{\text{U-Net}}$) de gaz à effet de serre des modèles spécialisés, en prenant en compte le nombre d'inférences ($N_{\text{inférence}}$), peut être calculée de la façon suivante :

$$E_{\text{U-Net}} = E_{\text{entraînement}} + N_{\text{inférence}} \times E_{\text{inférence}}$$

Pour notre méthode, l'expression de l'émission totale de gaz à effet de serre (E_{tot}) s'affranchit du terme représentant les émissions liées à l'entraînement ($E_{\text{entraînement}}$), et s'exprime par la formule :

$$E_{\text{tot}} = N_{\text{inférence}} \times E_{\text{inférence}}$$

Sur cette base, on peut estimer qu'après environ 1000 inférences, l'émission de gaz à effet de serre des modèles spécialisés devient similaire à l'émission de gaz à effet de serre de notre méthode. L'empreinte carbone liée à l'entraînement des modèles spécialisés U-Net 2D et U-Net 3D est alors compensée. Ce nombre serait plus important si les modèles spécialisés avaient été entraînés sur les images en pleine résolution. Cette estimation n'a pas été faite dans le cadre de notre étude. De plus, cet entraînement en pleine résolution ne se justifie pas en termes de performance, car les performances obtenues avec les modèles spécialisés sont déjà similaires ou supérieures à celles de notre méthode. Ainsi, dans ce cas, l'utilisation de modèles spécialisés réentraînés présente un impact environnemental cumulé inférieur à celui d'un modèle de fondation appliquée directement sans réentraînement. Cependant, il est quand même important de considérer qu'avant d'entraîner complètement un modèle spécialisé, prêt pour réaliser des inférences, une première phase de prototypage du modèle a lieu. Cette phase peut être aussi très consommatrice en énergie et peut entraîner d'importantes émissions de gaz à effet de serre à cause du large espace de recherche des hyperparamètres dans lequel est cherchée la meilleure architecture pour la tâche de segmentation. L'utilisation d'un modèle de fondation sans réentraînement évite la réalisation de cette phase à chaque nouvelle tâche de segmentation.

Concernant le temps d'exécution de notre méthode, la phase de recalage des images de références dans l'espace commun et de transposition des masques de références dans l'espace commun (encadré rouge sur la Figure [2.5](#)) n'est pas prise en compte car cette étape se fait une seule fois hors-ligne, sans avoir besoin de l'image cible à segmenter. En revanche, la génération de boîtes englobantes impliquant le recalage de l'image cible dans l'espace commun puis le recalage des 15 masques de références sur l'image cible prend environ 7 minutes sur l'architecture décrite dans le Tableau [2.4](#) en annexe. Le temps de recalage des 15 masques de références se fait actuellement en série mais pourrait être divisé par quinze si les transformations sont parallélisées. Concernant l'inférence de MedSAM2, la segmentation des 15 organes sur la même architecture d'une image de notre base de données prend environ 30 secondes. Pour les modèles spécialisés, l'inférence d'une image avec son changement d'échelle (sous-échantillonnage de l'image puis sur-échantillonnage de la segmentation) avec les modèles U-Net 2D et U-Net 3D prend moins de 20 secondes sur la même architecture. Notre méthode nécessite donc un temps de traitement bien plus long que les modèles spécialisés.

6 Conclusion et perspectives

L'utilisation d'un modèle de fondation comme MedSAM2 permet d'obtenir des segmentations aussi bonnes voire meilleures qu'un modèle spécialisé lorsque les indications de type boîte englobantes sont données de façon

précise. De plus, l'utilisation d'un modèle de fondation permet de s'affranchir d'une phase d'entraînement, ce qui rend la segmentation plus frugale lorsqu'elle est réalisée sur un petit nombre d'images.

En revanche, pour rendre cette segmentation automatique, il est nécessaire de greffer une méthode de génération de boîte englobante. Notre méthode par recalage a montré des performances similaires ou du même ordre de grandeur sur les organes abdominaux ayant une forme et une localisation stables comparées aux performances de référence de MedSAM2 en générant les boîtes englobantes avec les masques de vérité. Cependant, notre démarche a des limitations pour les organes petits ou ayant une forme et une localisation variables car le recalage est moins fiable sur ces organes.

Pour améliorer ces résultats, de nombreuses perspectives concernant les travaux futurs peuvent être tirées. Une des premières pistes serait d'améliorer la génération des boîtes englobantes, surtout pour les organes petits ou ayant une forme et une localisation variables. Celle-ci pourrait être améliorée avec peut-être d'autres méthodes de recalage qu'il serait possible d'explorer ou alors en utilisant un classifieur ou un détecteur léger permettant de localiser grossièrement les organes cibles.

Un autre moyen d'améliorer la génération des boîtes englobantes serait d'utiliser des a priori anatomiques pour relocaliser ou redéfinir les boîtes englobantes ayant une forme et une localisation instables à partir des cartes de score déterminées pour les organes stables. Une application serait de déplacer le centre de la boîte englobante d'un organe lorsqu'elle déborde trop sur la carte des scores d'un organe voisin plus stable. Par exemple, sur la Figure 2.14a, la boîte englobante de la rate déborde sur le rein gauche et donc probablement sur la carte de score du rein gauche. Ainsi, l'a priori anatomique pouvant être utilisé est que la rate se situe derrière le rein gauche. Cela pourrait permettre de décaler le centre de la boîte englobante ou la redéfinir vers le derrière de la coupe car la rate se situe derrière le rein gauche.

Le déplacement ou la nouvelle définition des boîtes englobantes des organes moins bien recalés et plus sensibles au prompt peut se faire avec la carte des scores des organes mieux recalés ou alors après leur segmentation. Ceci pourrait permettre d'imaginer une segmentation hiérarchique, où les organes plus faciles à segmenter sont traités en premier. Ensuite, leur position et leur forme définies par leur masque de segmentation peuvent permettre, avec des a priori anatomiques, de remplacer ou redéfinir les boîtes englobantes des autres organes.

L'intégration d'a priori anatomique peut également s'imaginer au niveau de la génération des boîtes englobantes, mais aussi au niveau de la segmentation, dans le modèle MedSAM2. Cela pourrait rendre le modèle plus performant lorsque les indications données ne sont pas très précises et plus robuste à la qualité des indications. La problématique qui se pose est la suivante : comment intégrer efficacement les contraintes anatomiques dans le processus de segmentation ? Plusieurs approches peuvent être envisagées, notamment l'incorporation de ces contraintes dans la fonction de perte, l'ajout d'un module complémentaire suivi d'un réentraînement, ou encore la mise en place de mécanismes contraignant l'attention du modèle vers des régions anatomiquement pertinentes.

7 Annexe

7.1 Détermination des seuils

La détermination des seuils permettant les meilleures performances a été déterminée expérimentalement en testant différentes valeurs pour la génération des boîtes englobantes lors de la première segmentation. Les seuils retenus par organe sont définis dans le Tableau 2.3

Organe	SPL	RKI	LKI	GBL	ESO	LIV	STO	AOR	IVC	PAN	RAG	LAG	DUO	BLA	PRO/UTE
Seuil	0.6	0.6	0.4	0.3	0.2	0.2	0.6	0.3	0.3	0.4	0.1	0.1	0.3	0.7	0.1

TABLE 2.3 – Seuils appliqués pour chaque organe.

Pour les organes qui sont bien recalés (les différents masques de références recalés expriment un haut consensus), le seuil choisi est plus élevé. C'est le cas de la rate (SPL) et des reins (RKI et LKI). A noter que pour le rein droit, le seuil doit être plus restrictif pour permettre aux boîtes englobantes de ne pas dépasser sur le foie. Pour les organes mal recalés (les différents masques de références recalées n'expriment pas de consensus), la carte de score est très étalée et contient des valeurs assez faibles. Le seuil doit donc être également faible pour éviter de perdre complètement la localisation de l'organe. C'est le cas par exemple pour les glandes surrénales (RAG et LAG) ou la prostate et l'utérus (PRO/UTE).

Pour les organes comme le duodénum (DUO), le pancréas (PAN) ou l'estomac (STO), le seuil ne doit pas être trop restrictif pour permettre de ne pas générer une boîte englobante trop à l'intérieur de l'organe cible mais il ne doit pas être trop grand pour éviter que les boîtes englobantes dépassent sur les structures voisines, ces trois organes étant très rapprochés.

7.2 Architecture matérielle utilisée.

Composant	Détails
Processeur (CPU)	Intel Core i5-10500 6 cœurs / 12 threads Fréquence : 3,10 GHz
Mémoire (RAM)	Capacité : 16 Go
Système d'exploitation	Linux Fedora 41
GPU	NVIDIA Quadro RTX 4000 VRAM : 8 Go
Contexte d'exécution	Machine dédiée, Turbo Boost activé Charge minimale au moment des tests

TABLE 2.4 – Architecture matérielle utilisée.

Bibliographie

- R. Agier, S. Valette, R. Kéchichian, L. Fanton, and R. Prost. Hubless keypoint-based 3d deformable groupwise registration. *Medical image analysis*, 59 :101564, 2020.
- E. Barbierato and A. Gatti. Toward green ai : A methodological survey of the scientific literature. *IEEE Access*, 12 :23989–24013, 2024.
- H. Bay, T. Tuytelaars, and L. Van Gool. Surf : Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006.
- D. Benaben, F. Berthoud, G. Guennebaud, A.-L. Ligozat, and S. Valcke. Estimation de l’empreinte carbone d’une heure de calcul sur un cœur cpu ou sur un gpu. Technical report, Groupe de travail Infrastructures de recherche – calcul, équipe Empreinte, GDR Labos 1point5, janvier 2024. URL https://labos1point5.org/static/rapports/Estimation_empreinte_carbone_heure_de_calcul.pdf. Consulté le 12 août 2025.
- H. Borgli, H. K. Stensland, and P. Halvorsen. Automatic prompt generation using class activation maps for foundational models : A polyp segmentation case study. *Machine Learning and Knowledge Extraction*, 7 (1) :22, 2025.
- Z. M. Colbert, D. Arrington, M. Foote, J. Gårding, D. Fay, M. Huo, M. Pinkham, and P. Ramachandran. Repurposing traditional u-net predictions for sparse sam prompting in medical image segmentation. *Bio-medical Physics & Engineering Express*, 10(2) :025004, 2024.
- B. Courty, V. Schmidt, S. Luccioni, Goyal-Kamal, MarionCoutarel, B. Feld, J. Lecourt, LiamConnell, A. Saboni, Inimaz, supatomic, M. Léval, L. Blanche, A. Cruveiller, ouminasara, F. Zhao, A. Joshi, A. Bogroff, H. de Lavoreille, N. Laskaris, E. Abati, D. Blank, Z. Wang, A. Catovic, M. Alencon, Michał Stęchły, C. Bauer, L. O. N. de Araújo, JPW, and MinervaBooks. mlco2/codecarbon : v2.4.1, May 2024. URL <https://doi.org/10.5281/zenodo.11171501>.
- A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. An image is worth 16x16 words : Transformers for image recognition at scale. *arXiv preprint arXiv :2010.11929*, 2020.
- Y. Huang, X. Yang, L. Liu, H. Zhou, A. Chang, X. Zhou, R. Chen, J. Yu, J. Chen, C. Chen, et al. Segment anything model for medical images ? *Medical Image Analysis*, 92 :103061, 2024.
- Y. Ji, H. Bai, J. Yang, C. Ge, Y. Zhu, R. Zhang, Z. Li, L. Zhang, W. Ma, X. Wan, et al. Amos : A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. *arXiv preprint arXiv :2206.08023*, 2022.
- A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023.
- L. K. Lee, S. C. Liew, and W. J. Thong. A review of image segmentation methodologies in medical image. In *Advanced Computer and Communication Engineering Technology : Proceedings of the 1st International Conference on Communication and Computer Engineering*, pages 1069–1080. Springer, 2014.

- J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang. Segment anything in medical images. *Nature Communications*, 15(1) :654, 2024.
- J. Ma, Z. Yang, S. Kim, B. Chen, M. Baharoon, A. Fallahpour, R. Asakereh, H. Lyu, and B. Wang. Medsam2 : Segment anything in 3d medical images and videos. *arXiv preprint arXiv :2504.03600*, 2025.
- M. A. Mazurowski, H. Dong, H. Gu, J. Yang, N. Konz, and Y. Zhang. Segment anything model for medical image analysis : an experimental study. *Medical Image Analysis*, 89 :102918, 2023.
- S. Na, Y. Guo, F. Jiang, H. Ma, and J. Huang. Segment any cell : A sam-based auto-prompting fine-tuning framework for nuclei segmentation. *arXiv preprint arXiv :2401.13220*, 2024.
- S. Pandey, K.-F. Chen, and E. B. Dam. Comprehensive multimodal segmentation in medical imaging : Combining yolov8 with sam and hq-sam models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 2592–2598, October 2023.
- H. Park, P. H. Bland, and C. R. Meyer. Construction of an abdominal probabilistic atlas and its application in segmentation. *IEEE Transactions on medical imaging*, 22(4) :483–492, 2003.
- N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, et al. Sam 2 : Segment anything in images and videos. *arXiv preprint arXiv :2408.00714*, 2024.
- O. Ronneberger, P. Fischer, and T. Brox. U-net : Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- S. Roy, T. Wald, G. Koehler, M. R. Rokuss, N. Disch, J. Holzschuh, D. Zimmerer, and K. H. Maier-Hein. Sam. md : Zero-shot medical image segmentation capabilities of the segment anything model. *arXiv preprint arXiv :2304.05396*, 2023.
- C. Ryali, Y.-T. Hu, D. Bolya, C. Wei, H. Fan, P.-Y. Huang, V. Aggarwal, A. Chowdhury, O. Poursaeed, J. Hoffman, et al. Hiera : A hierarchical vision transformer without the bells-and-whistles. In *International conference on machine learning*, pages 29441–29454. PMLR, 2023.
- R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi. Medical image segmentation using deep learning : A survey. *IET image processing*, 16(5) :1243–1267, 2022.
- X. Wei, J. Cao, Y. Jin, M. Lu, G. Wang, and S. Zhang. I-medsam : Implicit medical image segmentation with segment anything. In *European Conference on Computer Vision*, pages 90–107. Springer, 2024.
- J. Wu, Z. Wang, M. Hong, W. Ji, H. Fu, Y. Xu, M. Xu, and Y. Jin. Medical sam adapter : Adapting segment anything model for medical image segmentation. *Medical image analysis*, 102 :103547, 2025.
- J. Xu, X. Li, C. Yue, Y. Wang, and Y. Guo. Sam-mpa : Applying sam to few-shot medical image segmentation using mask propagation and auto-prompting. *arXiv preprint arXiv :2411.17363*, 2024.
- D. Yin, Q. Zheng, L. Chen, Y. Hu, and Q. Wang. Apg-sam : Automatic prompt generation for sam-based breast lesion segmentation with boundary-aware optimization. *Expert Systems with Applications*, 276 :127048, 2025.

Chapitre 3

Retour d'expérience

1 Compétences acquises et développées

Concernant mon retour d'expérience, ce stage m'a permis d'acquérir et de mettre en œuvre de nombreuses compétences professionnelles et personnelles.

Sur le plan scientifique, j'ai acquis de nombreuses connaissances liées aux modèles de fondations avec une architecture de type ViT (Vision Transformer). Ces connaissances vont de l'architecture des modèles, aux méthodes d'adaptation pour appliquer un modèle de fondation sur un autre domaine que celui sur lequel il a été entraîné. Cela a complété les connaissances que j'ai acquises lors du module IAT (Intelligence Artificielle pour les Télécommunications) ainsi que dans un cours que j'ai suivi en Pologne lors de mon échange ERASMUS portant sur l'apprentissage profond pour l'analyse d'images médicales. J'ai pu également me familiariser avec les compétences requises pour un chercheur que j'avais déjà pu expérimenter lors de ma formation en TC avec le module ASDS (Analyse et Synthèse de Documents Scientifiques) et PIR (Projet d'Initiation à la Recherche). Ces compétences regroupent la recherche et l'analyse de papiers de recherche, la planification et la conduite d'un projet de recherche et la synthèse de travaux de recherche sous forme scientifique.

D'un point de vue technique - applicatif, je suis montée en compétence sur plusieurs librairies Python comme PyTorch, utilisée dans le développement de modèles d'apprentissage machine et d'apprentissage profond, Nibabel et SimpleITK, utilisées dans le traitement d'images. Je me suis également familiarisée avec les infrastructures informatiques du laboratoire CREATIS et notamment avec l'utilisation d'un cluster de calcul pour réaliser certaines expériences nécessitant de plus grandes ressources de calcul. Les infrastructures informatiques du laboratoire possèdent la distribution Linux Fedora. Le module PIT (Passeport Informatique Télécoms) ainsi que les projets réalisés en utilisant une distribution Linux Ubuntu m'ont permis d'être à l'aise facilement.

Sur le plan organisationnel, j'ai travaillé de façon autonome en dehors des réunions hebdomadaires avec mon tuteur. Je n'ai pas eu de problème avec ce fonctionnement car lors de ma formation au département Télécommunications, beaucoup de projets sont réalisés en autonomie. En revanche, la durée des phases de projets en TC est assez courte et on ne travaille jamais seul (travaux de groupe). Au cours de mon stage, j'ai eu l'occasion de travailler sur un projet à plus long terme. Ainsi, il est nécessaire (en tout cas de mon point de vue) de varier les tâches exécutées (jongler entre bibliographie, analyse de résultats, développement de code, débogage de code ancien qui ne tourne plus avec les nouvelles versions de librairies). De plus, on ne sait pas toujours prédire en avance jusqu'à où va pouvoir être mené le projet ou alors la direction qu'il va prendre et les résultats que l'on va pouvoir obtenir. Cette flexibilité sur l'orientation et l'angle d'attaque du projet est selon moi l'un des grands avantages du milieu de la recherche. En revanche, cela nécessite d'être rigoureux pour ne pas s'éparpiller et avancer toujours dans la direction qui semble être la meilleure.

Ainsi, cette expérience de stage m'a permis de mettre en œuvre des compétences acquises avec la formation TC et de les renforcer.

2 Ma perception du milieu de la recherche

Lors de ce stage, j'ai pu également découvrir le monde de la recherche publique de plus près au fur et à mesure des échanges avec les doctorants et les membres permanents. J'ai beaucoup apprécié la liberté que peut offrir la recherche sur le choix des sujets que l'on souhaite approfondir ainsi que du point de vue à partir desquels on peut les aborder. J'apprécie également que, dans ce domaine, les recherches soient principalement orientées vers le bien commun, c'est-à-dire autour de sujets susceptibles d'améliorer la qualité de vie des individus, directement ou indirectement. Je suis convaincue que la recherche publique, lorsqu'elle est indépendante de tout intérêt financier lié à une entité privée, a un impact plus positif sur la société qu'une organisation privée dont l'un des objectifs principaux reste la rentabilité. Un des exemples qui me vient à l'esprit, lié au milieu de la santé, est celui de l'affaire Mediator. Le Mediator était un médicament développé par le laboratoire Sévrier et prescrit pour le traitement du diabète. Cependant, en 2010, le médicament s'est retrouvé au cœur d'un scandale sanitaire suite à la publication du livre *Mediator 150 mg. Combien de morts ?* signé par la pneumologue Irène Frachon. En effet, il s'est trouvé que le groupe pharmaceutique a commercialisé pendant des années ce médicament, qui provoquait de graves lésions des valves cardiaques et de l'hypertension artérielle pulmonaire, alors que les premières alertes sur la toxicité du médicament avaient été données dans les années 1990 [Le Monde, 2021]. Le groupe Sévrier, pendant tout ce temps, a essayé de dissimuler ou minimiser les preuves liées à la toxicité de son traitement, sans le changer ou le retirer du marché. Une entreprise ayant un rôle majeur dans la santé et dans le traitement de pathologies est donc capable de faire passer son intérêt financier au-dessus du bien commun. Bien sûr et heureusement, ce cas fait partie des extrêmes, mais je pense quand même qu'une entreprise privée doit bien souvent, soit par vénalité, soit pour permettre sa survie, considérer en avant-poste l'intérêt financier, reléguant bien souvent le bien commun.

En revanche, dans le milieu de l'intelligence artificielle, actuellement, la recherche est en pleine explosion. Cependant, j'ai l'impression que les papiers de recherche publiés sont principalement axés sur la performance des algorithmes, qui parfois se jouent à peu. Cela mène à l'augmentation des tailles de modèles, des bases de données d'entraînement et rend les modèles de plus en plus lourds. Les tailles des mémoires requises sont donc de plus en plus importantes et les modèles nécessitent des GPUs de plus en plus puissants. Cela freine déjà leur applicabilité dans des infrastructures comme les hôpitaux, car ceux-ci éprouvent des difficultés à s'équiper de matériel de calcul. De plus, cela les rend de plus en plus consommateurs en énergie, alors que les modèles se démultiplient. Je trouve ça donc dommage que pour le moment, dans ce milieu, la course aux meilleures performances occulte le travail de certains chercheurs qui développent des algorithmes plus frugaux, aux performances légèrement inférieures à l'état de l'art, mais qui pour certaines applications peuvent être déjà suffisantes. Il me semble également que, par effet de mode, la tendance est d'intégrer de l'apprentissage profond dans certaines méthodes, sans que cela soit vraiment nécessaire en termes d'amélioration significative de performances, alors que cela nuit à l'explicabilité des résultats et augmente la consommation énergétique, au détriment des ressources existantes.

Ce phénomène est renforcé par la transformation du modèle historique de la recherche publique où les études étaient financées principalement sur dotation en un modèle managérial. Désormais, un laboratoire est dirigé sur un modèle plus similaire à celui des entreprises avec le remplacement du financement structurel récurrent par des appels à projets mettant en concurrence les membres du laboratoire et les laboratoires entre eux. Cela est couplé à des systèmes de primes (de recherche ou d'encadrement) poussant les chercheurs à la performance, performance évaluée avec des métriques pouvant être remises en question. Le nombre d'articles publiés, le h-index, le facteur d'impact de la revue où est publié un article sont-ils réellement des indications d'une science de qualité ? Ces méthodes d'évaluation constituent, selon moi, un dispositif incitant le chercheur à produire dans le cadre des attentes de sa communauté scientifique, au détriment des travaux marginaux qui cherchent à penser autrement. Dans l'état du domaine actuel de l'intelligence artificielle, cela revient à privilégier la course à la performance, en occultant les approches alternatives.

3 Conclusion

Pour conclure, ce stage m'a permis de mettre un vrai pied dans le milieu de la recherche, permettant de m'y familiariser. J'ai appris beaucoup de choses d'un point de vue technique et organisationnel, mais aussi et

surtout d'un point de vue personnel. Cela a permis d'alimenter mes réflexions sur mon parcours professionnel. Je vois désormais le milieu de la recherche à travers un prisme plus réel et je pense avoir saisi un peu mieux son fonctionnement et ses enjeux.

Je suis très satisfaite et reconnaissante de ce qu'a pu m'apporter mon stage, principalement grâce aux échanges que j'ai pu avoir avec les doctorants et permanents du laboratoire, ainsi qu'avec mon tuteur. Je pars en cette fin de stage convaincue de l'utilité d'une telle démarche.

Glossaire

biomarqueurs : Caractéristiques définies et mesurées comme des indicateurs des processus biologiques normaux, des processus pathogènes ou des réactions à une exposition ou une intervention, y compris les interventions thérapeutiques . [3](#)

spectroscopie : ici, spectroscopie par résonance magnétique nucléaire (spectroscopie RMN); technique d'analyse utilisée pour déterminer la structure des molécules basée sur le comportement de certains noyaux dans un champ magnétique externe. [4](#)

tomodensitométrie (ou CT de l'anglais Computed Tomography) : Technique d'imagerie médicale qui consiste à mesurer l'absorption des rayons X par les tissus puis, par traitement informatique, à numériser et enfin reconstruire des images 2D ou 3D des structures anatomiques. Aussi appelé scanner dans le langage courant . [4](#)

unité mixte de recherche (UMR) : structure de recherche en France qui regroupe des chercheurs de différents établissements (souvent un organisme national de recherche comme le CNRS, INRAE, INSERM, etc., et une ou plusieurs universités ou grandes écoles). [3](#)

ViT (Vision Transformer) : Architecture de réseau neuronal profond qui applique le principe d'attention croisée sur une image découpée en patch. Il effectue des tâches de vision par ordinateur telles que la classification, la segmentation ou la détection . [28](#)

Bibliographie Rapport

Le Monde. Scandale du médiator : les laboratoires servier condamnés à 2,7 millions d'euros d'amende, Mar. 2021. URL https://www.lemonde.fr/societe/article/2021/03/29/scandale-du-mediator-les-laboratoires-servier-condamnes-a-2-7-millions-d-euros-d-amende_6074840_3224.html. Consulté le 12 août 2025.