Sorbonne Université

École Doctorale Informatique, Télécommunications et Electronique
(ED130)

Laboratoire LIP6

Analyse d'images par des méthodes d'intelligence artificielle en endoscopie digestive, cas de la CPRE

Par Garance MARTIN

Thèse de doctorat en Informatique

Soutenue le 13 décembre 2024

Devant le jury composé de :

Nadjia	Kachenoura	Directrice de Recherche	Examinatrice, Présidente
Aymeric	Histace	Professeur	Rapporteur
Hugues	Talbot	Professeur	Rapporteur
Odyssée	Merveille	Maîtresse de Conférences	Examinatrice
Andrea	Pinna	Professeur	Directeur
Isabelle	Bloch	Professeure	Co-Directrice
Jérôme	Szewczyk	Professeur	Co-Directeur

Except where otherwise noted, this work is licensed under
http://creativecommons.org/licenses/by-nc-nd/3.0/

Remerciements

Mes premiers remerciements s'adressent aux membres du jury. Un grand merci à Aymeric Histace et Hugues Talbot qui ont accepté la lourde tâche de rapporteurs de ce travail de thèse. Je remercie Aymeric Histace pour son rapport consciencieux et constructif mais également pour l'enthousiasme qu'il a manifesté pour le sujet. Merci à Hugues Talbot d'avoir non seulement rapporté cette thèse mais également d'être la personne qui m'a donnée l'envie saugrenue de faire une thèse et qui m'a initiée aux joies de la morphologie mathématique et du traitement d'images lorsque j'étais en école d'ingénieurs.

Je remercie également les examinatrices, Nadjia Kachenoura et Odyssée Merveille d'avoir accepté de faire partie de mon jury. Je remercie particulièrement Nadjia Kachenoura d'avoir présidé le jury. Je remercie spécialement Odyssée Merveille qui a suivi mon travail de thèse en tant que membre du comité de suivi. J'ai énormément appris grâce à ses remarques toujours pertinentes.

Une thèse ne peut se dérouler sans encadrement, et je suis très reconnaissante d'avoir été choisie par ce trio pour mener à bien ce travail passionnant. J'ai pu bénéficier d'un encadrement complet à la fois scientifiquement et humainement. Merci Andrea d'avoir cru en moi pour porter cette thèse et de m'avoir poussée quand je n'avais pas confiance. Merci pour ta bienveillance et d'avoir su trouver les bons mots notamment la veille de la soutenance (merci pour les graphes de dernière minute aussi). Merci Jérôme d'avoir su créer et coordonner un tel projet interdisciplinaire, mais aussi pour ta gentillesse et ton implication. Merci Isabelle d'être tellement investie dans le parcours de tes doctorants et doctorantes, merci d'avoir cru en moi et de m'avoir tant appris. Merci d'avoir été et de toujours être un modèle. Ma gratitude dépasse l'étendue de mon vocabulaire. Merci également de m'avoir permis de tant voyager pendant ces trois années.

Ce projet ne serait rien sans l'équipe médicale qui le porte avec entrain et fierté. Merci à Aymeric Becq, Marine Camus, Grégoire Salin et Dominique Anelone-Ake pour leur disponibilité malgré leurs emplois du temps chargés, leur pédagogie incroyable, leur passion autour du sujet et leur bienveillance.

Merci à l'équipe de l'UTC, Isabelle Claude, Jean-Matthieu Prot et Abdelhadi Essamlali avec qui c'est toujours un plaisir d'échanger et de constater l'avancée du projet MAAGIE.

Merci aux ingénieurs de l'ISIR également, Philippe Gauthier et Clément Alberge, qui ont toujours été très disponibles pour expliquer et avancer sur l'algorithme de localisation du fil guide dans les voies biliaires.

Je remercie également Yuxuan qui a mené à bien son stage de master et m'a permis

d'avancer sur la segmentation du fil guide et dont une partie des résultats est utilisée dans ce travail.

Je remercie grandement Emmanuel Zarpas qui a été mon mentor depuis la deuxième année de thèse, malgré un début assez bouleversant, tu es resté à mes côtés pour me prodiguer des conseils et me rassurer. Je ne sais pas si je l'ai assez exprimé mais j'ai beaucoup appris de toi et je te remercie pour le temps que tu m'as accordé. Merci à Sorbonne Université et au Collège Doctoral d'avoir mis en place ce programme plus que nécessaire pour les doctorants et doctorantes et de m'avoir sélectionnée pour la première édition.

J'aimerais ensuite remercier les deux équipes dont je fais partie. Merci à l'équipe SYEL d'avoir toujours été accueillante et de m'avoir fait découvrir tout l'aspect embarqué de l'informatique. Merci à l'équipe LFI de créer cette atmosphère si bienveillante dans laquelle il fait bon de travailler et de s'épanouir en tant que doctorante. Un merci spécial à Christophe qui a instauré le goûter post-réunion d'équipe. Et un énorme merci à Marie-Jeanne qui est toujours disponible lorsque j'ai une question et qui m'a beaucoup soutenue et aidée durant la dernière phase de rédaction et quand je peinais à faire mes diapositives de soutenance.

Je tiens également à remercier les membres qui font tourner le laboratoire, qui ont toujours été d'une grande sympathie à mon égard et avec qui j'adore échanger, merci à Fabrice, Aline, Hélène, Laurence.

Un énorme merci à Nora de m'avoir accompagnée tout au long de cette thèse, merci pour ta disponibilité, ton organisation impeccable et ta gentillesse.

Je tiens ensuite à remercier mes trois piliers de la thèse, celle et ceux qui ont rendu l'expérience de thèse inoubliable.

Mélina, les mots me manquent pour décrire à quel point je te suis reconnaissante. Merci d'être toi tout simplement et ma meilleure alliée dans la thèse mais aussi en dehors. Merci d'avoir créé la semaine des memes et les cafés philo. Merci pour tout ce qu'on a partagé, la nourriture (élément de survie indispensable au doctorat), les fous rires, les discussions philosophiques et la colère contre le système. Je te remercie pour tous ces bons moments qui vont énormément me manquer mais auxquels je repenserais toujours avec nostalgie. Merci de m'avoir inspirée par tes qualités, je suis sûre que tu vas devenir une grande chercheuse.

Yann, mon presque-jumeau de thèse, merci pour ta patience et ta bonne humeur quotidienne. J'ai compté tu as fait 151 blagues pas drôles et tu as dit 42 fake news durant la thèse, ce qui reste grandement raisonnable sur 3 ans. Merci de m'avoir écouté râler nuit et jour contre le monde entier, merci d'avoir expliqué les règles des jeux 70 fois d'affilée sans te plaindre parce que je les oubliais. Merci pour les discussions sans fin sur la nourriture et le Japon (et souvent la nourriture au Japon). Merci pour ta générosité, tu as fait honneur aux gens du Sud en tant que digne représentant au LIP6. Je te nomme officiellement meilleur co-bureau, ça sera difficile de faire mieux pour les suivants ou suivantes.

Leandro, merci pour ta sensibilité, merci d'être une personne à qui on peut parler ouvertement de tout. Merci pour ta convivialité, ton partage de musique, de bonne humeur, d'anecdotes linguistiques et tes questions toujours plus aléatoires. Merci aussi pour ta gentillesse et ton réconfort. Et merci d'être mon canard de débogage ! Je ne sais pas comment je vais faire plus tard. Je tiens à souligner ta patience et ta disponibilité pour m'aider et te moquer de mon code (ce qui était justifiable). Merci de m'avoir fait me sentir moins seule en imagerie médicale.

Je tiens également à remercier dans le désordre, mes collègues doctorants et docto-

rantes. Merci à Adulam de m'avoir accueillie dans le bureau 516, le bureau de la joie et des en-cas. Merci de m'avoir montré la voie et de m'avoir inspirée. Merci à Guimauve d'avoir partagé le bureau de la joie également depuis son stage. Merci pour les fous rires complètement absurdes, merci d'apporter ta touche de folie et de sensibilité avec tes références obscures et de militantisme.

Maxime Maxime et Ali, merci pour ses longues pauses café à râler de la thèse (principalement). Je suis heureuse d'avoir pu partager ces moments de thèse avec vous ainsi que ces buffets à volonté. Merci à Mathuran pour ta gentillesse la première fois qu'on s'est rencontré qui n'a cessé de diminuer après. Merci d'être aussi attentionné malgré tout.

Je tiens à mentionner et remercier spécialement Arun, qui a passé beaucoup de temps à m'installer des logiciels sur mon ordinateur mais aussi de nous avoir partagé sa passion pour l'aquarelle. Merci à Ege de me faire me sentir utile avec tes questions administratives. Merci de faire partie de l'équipe des doctorants d'Isabelle qui font de l'imagerie et de me faire sentir moins seule.

Merci à Giammarco d'avoir été disponible et à l'écoute pour parler de son code et m'aider à le ré-utiliser. Merci aux autres membres de l'équipe de Necker Enzo et Thomas qu'il est toujours agréable de croiser en conférence.

Merci à Camilo et Inès pour le partage de leur passion pour l'opéra mais aussi leur convivialité. Merci à tous les anciens et anciennes qui ont rendu la vie au laboratoire meilleure durant leur passage : Clara (Bove), Jérémie, les Adams, Cecilia, Oriane et Aymeric. Merci à la relève Thomas, Nour, Garance de faire vivre l'équipe et le laboratoire. Merci à Clara (Charon) de m'avoir fait me sentir moins seule à ma première conférence ! Merci pour tous ces bons moments notamment avec l'équipe DECISION. Merci à Margot, Mahdi Thibault et les autres.

Je suis très reconnaissante d'avoir rencontré grâce à la thèse des femmes aussi brillantes et inspirantes que Frosso et Céline. Merci de votre joie et votre force. C'est une chance d'avoir pu partager ces moments avec vous. Merci d'être des alliées dans la lutte pour nos droits.

Merci à Anna-Marie et Aurore qui ont rendu le laboratoire plus chaleureux et la thèse plus agréable.

Merci aux doctorants et doctorantes des autres équipes que j'ai pu rencontrer durant ces années notamment grâce au conseil des doctorants et doctorantes : Guillaume (Moinard), Bastien, Aymeric, Baptiste, Etienne, Nicolas.

Un grand merci également à Matthieu, Kevin et Hugo pour tous ces bons moments.

Merci à Sophie de m'avoir aidée depuis mon stage au CREATIS et de m'avoir fait sentir intégrée malgré la distance. C'est toujours un plaisir de pouvoir te voir en vrai dans les évènements de la communauté d'IA pour l'imagerie biomédicale.

Merci à Maud qui m'accompagne depuis la seconde. On a fini toutes les deux en thèse à la Sorbonne, n'en déplaise à M. Salonne. Je suis fière de nous.

Merci à mes amies est colocs incroyables qui illuminent mes journées et mes soirées, merci d'avoir cru en moi et de m'avoir soutenue depuis toutes ces années.

Merci à mes parents de m'avoir toujours laissé faire les études que je voulais. Merci à Tatoune de m'avoir poussée à persévérer en mathématiques. Merci à Marek d'avoir relu mon manuscrit et de m'avoir fait des commentaires. Merci aussi de ne pas avoir posé de questions à ma soutenance. Merci à ma Tante de m'avoir toujours encouragée. Merci à ma petite sœur Louise, d'être la sœur la plus incroyable depuis que tu es toute petite. Tu m'as beaucoup appris par ton grand cœur.

Édouard, merci d'être là quoiqu'il arrive. Merci de t'occuper de moi quand je n'ai plus

de forces, de m'accueillir et de m'accompagner dans la vie.

Résumé

La Cholangiopancrétographie Rétrograde Endoscopique (CPRE) est une intervention mini-invasive des voies biliaires à visée diagnostique et thérapeutique. Elle permet notamment de drainer les voies biliaires obstruées par des calculs ou des tumeurs à l'aide d'instruments endoscopiques (endoscope et fil guide) insérés via le tube digestif. Ces instruments sont suivis par radioscopie 2D et opacification à l'aide de produit de contraste durant l'intervention.

La difficulté majeure de cette intervention est de placer le fil guide dans la voie d'intérêt. Les endoscopistes se heurtent ici à plusieurs difficultés : la complexité de l'anatomie des voies biliaires (présentant une grande variabilité inter-patients), la perte d'information (repères) dans la lecture d'une image bi-dimensionnelle refigurant un objet tridimensionnel, l'utilisation limitée à la fois des prise d'images radioscopiques (rayons X) et du produit de contraste.

Ainsi demeure une grande incertitude sur la localisation du fil guide dans les voies biliaires par rapport à la zone à drainer. Cela entraîne des risques pour les patients. Le but du projet MAAGIE (Méthodes Avancées pour l'Assistance au Geste Interventionnel en Entérologie) est de réunir des endoscopistes et des chercheurs pour améliorer la CPRE grâce à des outils numériques comme par exemple un algorithme de fusion des informations 2D et 3D dont nous disposons pour mieux repérer le fil guide dans l'arbre biliaire. Dans ce contexte mes travaux de thèse s'intéressent à la segmentation automatique du fil guide sur des images radioscopiques de fluoroscopie 2D. Ces informations seront ensuite fournies à l'algorithme de fusion. Plusieurs difficultés méthodologiques sont à relever : la quantité limitée de données d'imagerie, la faible qualité des images et l'absence de base de données annotées. De plus, aucune méthode dédiée à la segmentation des fils guides de CPRE n'existe actuellement.

Afin de surmonter ces défis, plusieurs contributions sont proposées dans cette thèse. Tout d'abord, les données de deux centres hospitaliers ont été collectées. Ensuite, nous avons porté une attention particulière à la qualité de ces données en vue de mettre au point un modèle de prédiction automatique. Une première contribution a été d'harmoniser le jeux des données collectées, puis d'annoter ces données afin de construire une base exploitable pour les futures expériences d'apprentissage supervisé. La structure du fil guide étant très fine, les méthodes et métriques usuelles d'évaluation de la segmentation n'apparaissent pas suffisantes ni pertinentes pour évaluer les modèles utilisés et les comparer. Nous avons alors exploré ce qui se fait dans d'autres domaines en termes de méthodes adaptées à l'évaluation de la segmentation de structures comparables, notamment en nous appuyant sur le domaine de la segmentation vasculaire. Nous discutons également des méthodes d'évaluation d'un point de vue clinique afin de complémenter les approches issues des sciences des données et d'offrir aux médecins une meilleure compréhension et évaluation des modèles proposés.

Pour la segmentation du fil guide proprement dite, nous nous inspirons des méthodes développées pour les interventions endovasculaires. Ces méthodes s'appuient principalement sur l'architecture U-Net qui est un réseau de neurones typiquement utilisé en imagerie médicale. Notre approche est de comparer différentes fonctions de coût pour la segmentation du fil guide. Nous obtenons une segmentation imparfaite du point de vue numérique mais qui peut suffire à l'algorithme de fusion pour apporter l'information additionnelle sur l'image 3D au médecin. Enfin, nous segmentons aussi l'endoscope qui porte le fil guide à l'entrée des voies biliaires, et nous évaluons si la connaissance de sa position

dans l'image peut influencer la segmentation du fil guide. En conclusion, ce travail constitue une première approche de segmentation du fil guide pour améliorer le geste de CPRE. Les résultats obtenus à l'issue de cette thèse apparaissent prometteurs avec des scores de segmentation comparables à l'état de l'art dans d'autres domaines et seront poursuivis dans le cadre du projet MAAGIE.

Abstract

Endoscopic retrograde cholangiopancreatography (ERCP) is a minimally invasive diagnostic and therapeutic procedure of the bile ducts. In particular, it is used to drain bile ducts obstructed by stones or tumors using endoscopic instruments (endoscope and guidewire) inserted through the digestive tract. These instruments are monitored during the operation by 2D radioscopy and contrast opacification. The main difficulty in this procedure is the placement of the guidewire in the tract of interest. The endoscopist faces a number of difficulties : the complexity of the anatomy of the bile ducts (with considerable inter-patient variability), the loss of information (landmarks) when reading a two-dimensional image reflecting a three-dimensional object, and the limited use of both fluoroscopic images (X-rays) and contrast medium. As a result, there is still a great deal of uncertainty about where to place the guide wire in the bile ducts in relation to the area to be drained. This poses a risk to patients. The goal of the MAAGIE project is to bring together endoscopists and researchers to improve ERCP using digital tools, such as an algorithm to merge the 2D and 3D information to better locate the guide wire in the biliary tree. In this context, my dissertation focuses on the automatic segmentation of the guidewire on 2D fluoroscopic images. This information will then be provided to the fusion algorithm.

There are several methodological difficulties : the limited amount of imaging data, the low quality of the images, and the lack of an annotated database. In addition, there is currently no method dedicated to the segmentation of ERCP guidewires. To overcome these challenges, several contributions are proposed in this thesis. First, we collected data from two hospitals. Second, we paid special attention to the quality of these data in order to develop an automatic prediction model. A first contribution was to harmonize the collected data sets and then to annotate the data in order to build a usable base for future supervised learning experiments. Because the guidewire structure is very fine, the usual methods and metrics for evaluating segmentation do not seem to be sufficient or relevant for evaluating and comparing the models used. We have therefore reviewed what is being done in other fields in terms of methods suitable for evaluating the segmentation of comparable structures, with particular reference to the field of vascular segmentation. We also discuss evaluation methods from a clinical perspective to complement data science approaches and to help physicians better understand and evaluate the proposed models. For guidewire segmentation itself, we draw on methods developed for endovascular procedures. These methods are mainly based on the U-Net architecture, a neural network typically used in medical imaging. Our approach is to compare different cost functions for guidewire segmentation. We obtain a segmentation that is imperfect from a numerical point of view, but may be sufficient for the fusion algorithm to provide additional information to the physician on the 3D image. Finally, we also segmented the endoscope carrying the guidewire at the papilla and evaluated whether knowing its position in the image could influence guidewire segmentation. In conclusion, this work represents a first approach to guidewire segmentation to improve the ERCP procedure. The results obtained at the end of this thesis appear promising, with segmentation scores comparable to the state of the art in other fields, and will be pursued as part of the MAAGIE project.

Acronymes

AP-HP *Assistance Publique des Hôpitaux de Paris*

BMBI *Laboratoire de Biomécanique et de Bioingénierie*

CPRE *CholangioPancréatographie Rétrograde Endoscopique*

FGAC *Frangi Geodesic Active Contours*

FN *Faux Négatifs*

FP *Faux Positifs*

HM *Henri-Mondor*

ISIR *Institut des Systèmes Intelligents et Robotique*

MAAGIE *Méthodes Avancées pour l'Assistance au Geste Interventionnel en Entérologie*

ReLU *Unité Linéaire Rectifiée*

SA *Saint-Antoine*

SNFGE *Société Nationale Française de Gastro-Entérologie*

SU *Sorbonne Université*

UTC *Université Technologique de Compiègne*

VB *Voies biliaires*

VBIH *Voies biliaires intra-hépatiques*

VBP *Voie biliaire principale*

VN *Vrais Négatifs*

VP *Vrais Positifs*

Table des matières

1 Introduction **15**
1.1 Contexte médical . . . 15
1.1.1 Anatomie des voies bilio-pancréatiques . . . 16
1.1.2 Pathologies . . . 17
1.1.3 Déroulement de la CPRE . . . 18
1.2 Limites et difficultés de la CPRE . . . 20
1.3 Objectifs du projet MAAGIE . . . 21
1.4 Contexte méthodologique . . . 21
1.4.1 Segmentation . . . 22
1.4.2 Difficultés de la segmentation d'images de CPRE . . . 23
1.5 Questions de recherche . . . 24
1.6 Organisation du manuscrit . . . 24

2 État de l'art **25**
2.1 Évaluation de la difficulté de l'intervention . . . 25
2.2 Recalage 2D-3D . . . 26
2.3 Segmentation du fil guide . . . 28
2.3.1 Segmentation d'images biomédicales . . . 29
2.3.2 Segmentation du fil guide . . . 34
2.4 Discussion et conclusion . . . 39

3 Données **41**
3.1 Collecte des données . . . 41
3.2 Création de la base de données . . . 42
3.2.1 Nomenclature des images . . . 42
3.2.2 Pré-traitement . . . 42
3.3 Caractéristiques des images . . . 45
3.4 Annotation des images . . . 46
3.5 Conclusion . . . 48

4 Méthodes d'évaluation **49**
4.1 Cas limites . . . 49
4.2 Méthodes usuelles . . . 50

4.3 Segmentation . . . 52
4.3.1 Score de Dice . . . 52
4.3.2 Distances . . . 54
4.3.3 Courbes . . . 55
4.3.4 Limites des méthodes usuelles . . . 56
4.4 Méthodes spécifiques . . . 56
4.4.1 Mesures de chevauchement . . . 56
4.4.2 Coefficient de Corrélation de Matthews (MCC) . . . 59
4.4.3 Évaluer la topologie . . . 59
4.4.4 Mesures utilisées pour la segmentation de vaisseaux . . . 60
4.5 Évaluation clinique . . . 61
4.6 Conclusion . . . 61

5 Segmentation du fil guide 63
5.1 Fonctions de coût . . . 64
5.1.1 Notations . . . 64
5.1.2 Entropie croisée . . . 64
5.1.3 Fonction de coût de Dice . . . 65
5.1.4 Fonction de coût de Tversky . . . 66
5.1.5 Fonction de coût Focal . . . 66
5.1.6 Fonction de coût *ClDice* . . . 67
5.1.7 Du filtre à la fonction de coût pour les structures tubulaires . . . 67
5.1.8 Contours actifs et chemins minimaux . . . 70
5.1.9 Conclusion . . . 73
5.2 Données . . . 74
5.3 Méthode proposée . . . 74
5.3.1 Méthode de référence . . . 75
5.3.2 Fonction de coût combinant contour actif géodésique et filtre de Frangi . . . 75
5.3.3 Paramètres de l'expérience . . . 75
5.4 Résultats . . . 77
5.4.1 Saint-Antoine . . . 77
5.4.2 Henri-Mondor . . . 79
5.5 Reconstruction . . . 82
5.6 Conclusion . . . 84

6 Endoscope 87
6.1 Segmentation de l'endoscope seul . . . 87
6.1.1 Données . . . 87
6.1.2 Méthode . . . 87
6.1.3 Résultats . . . 88
6.1.4 Conclusion . . . 89
6.2 Segmentation multi-objets . . . 89
6.2.1 Données . . . 90
6.2.2 U-Net multi-objets . . . 90
6.2.3 IterNet . . . 91
6.2.4 Résultats . . . 92
6.3 Discussion . . . 93
6.4 Conclusion . . . 94

7 Conclusion et perspectives **95**
7.1 Conclusion 95
7.2 Perspectives 96
7.2.1 Perspectives cliniques 96
7.2.2 Perspectives méthodologiques 97

A Base de données **101**

B Application d'un filtre de Frangi sur les données **102**

C Augmentation des données **103**

CHAPITRE 1

Introduction

Ce travail de doctorat s'inscrit dans le projet *Méthodes Avancées pour l'Assistance au Geste Interventionnel en Entérologie* (MAAGIE), coordonné par le Pr. Jérôme Szewczyk. Dans ce projet une équipe pluridisciplinaire a été constituée afin d'améliorer l'intervention de **Cholangio-Pancréatographie Rétrograde Endoscopique** (CPRE) grâce à des outils informatiques. Ce projet résulte d'une collaboration entre les quatre entités de l'Alliance Sorbonne Université suivantes (figure 1.1) :

- l'Institut des Systèmes Intelligents et Robotique (ISIR), à Sorbonne Université (SU),
- les services de Gastroentérologie, des hôpitaux Saint-Antoine et Henri Mondor - réunis au sein de l'Assistance Publique Hôpitaux de Paris (AP-HP),
- le laboratoire de Biomécanique et de Bioingénierie (BMBI) - à l'Université de Technologie de Compiègne (UTC),
- et le laboratoire d'informatique LIP6, à SU.

1.1 Contexte médical

La CholangioPancréatographie Rétrograde Endoscopique (CPRE) est une intervention mini-invasive à visée diagnostique et thérapeutique des pathologies des voies biliaires et pancréatiques. Elle permet notamment le drainage des voies biliaires lors d'une obstruction d'ordre lithiasique ou néoplasique. Il s'agit d'un geste d'endoscopie dite interventionnelle.

Ce geste, réalisé pour la première fois en 1968 par McCune et al. [1968], est effectué environ **78000 fois par an en France**[1]. Étant complexe, il requiert de l'expérience de la part de l'endoscopiste. C'est pourquoi la plupart des patients concernés sont dirigés vers des centres experts.

1. Livre blanc de la CNPHGE https://www.cnp-hge.fr/wp-content/uploads/2020/03/LIVRE-BLANC_HGE.pdf

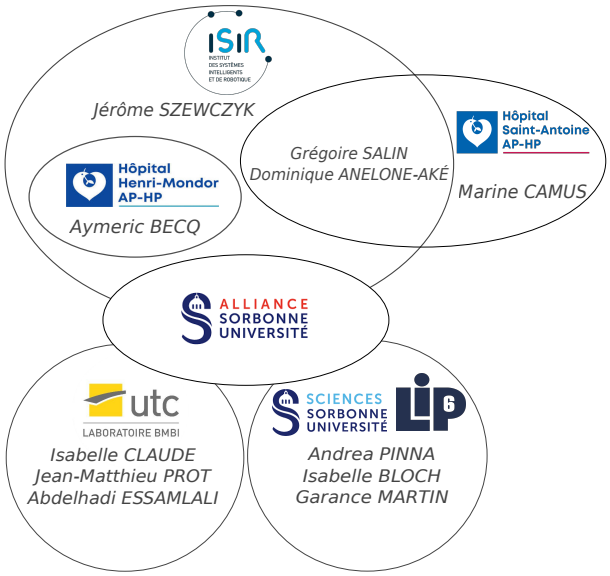


Figure 1.1 – Entités collaborant au sein du projet MAAGIE.

Dans ce chapitre nous décrivons l'anatomie des voies biliaires et les pathologies liées. Puis nous détaillons les étapes et difficultés de la CPRE. Nous exposons ensuite le contexte méthodologique, c'est-à-dire l'analyse des images de CPRE.

1.1.1 Anatomie des voies bilio-pancréatiques

Il convient d'abord de décrire l'anatomie de la zone concernée par la CPRE. Les voies bilio-pancréatiques se situent dans la partie supérieure droite de l'abdomen, près de plusieurs organes impliqués dans la digestion tels que le foie, la vésicule biliaire et le pancréas (figure 1.2). Ce sont des canaux qui acheminent la bile et les enzymes pancréatiques dans l'intestin grêle au niveau du duodénum via la papille (ou ampoule de Vater).

Commençons par le foie. Il s'agit de l'organe le plus volumineux du corps humain. Celui-ci assure plusieurs fonctions dont la production de la bile. Ce liquide jaunâtre, du fait de sa composition en bilirubine, contient également des acides biliaires utiles à la dissolution des graisses lors de la digestion. Cette bile, produite par les lobules, est directement collectée par les voies biliaires intra-hépatiques (VBIH). Le foie s'organise en deux lobes asymétriques, eux-mêmes découpés en huit segments. Les VBIH sont sectorisées selon ces segments et comprennent trois branches principales : gauche, antérieure droite et postérieure droite. Elles se rejoignent au niveau du hile hépatique pour former le canal hépatique commun.

La vésicule biliaire qui se situe en dessous du foie a pour fonction de stocker la bile. Elle est liée au canal hépatique commun par le canal cystique. La jonction de ces deux canaux donne naissance au canal cholédoque. On appelle voie biliaire principale (VBP) l'ensemble formé par les canaux hépatique commun et cholédoque.

Enfin le pancréas est une glande constituée d'un canal pancréatique principal, dit de Wirsung et de canaux secondaires. Les enzymes sécrétées par le pancréas sont ainsi drainées vers la papille et déversées dans le duodénum lors de la digestion.

Cette anatomie concerne environ 58% de la population, ce qui laisse une marge de variabilité anatomique conséquente qui touche aussi bien les voies biliaires intra-hépatiques que les voies biliaires extra-hépatiques (figure 1.3).

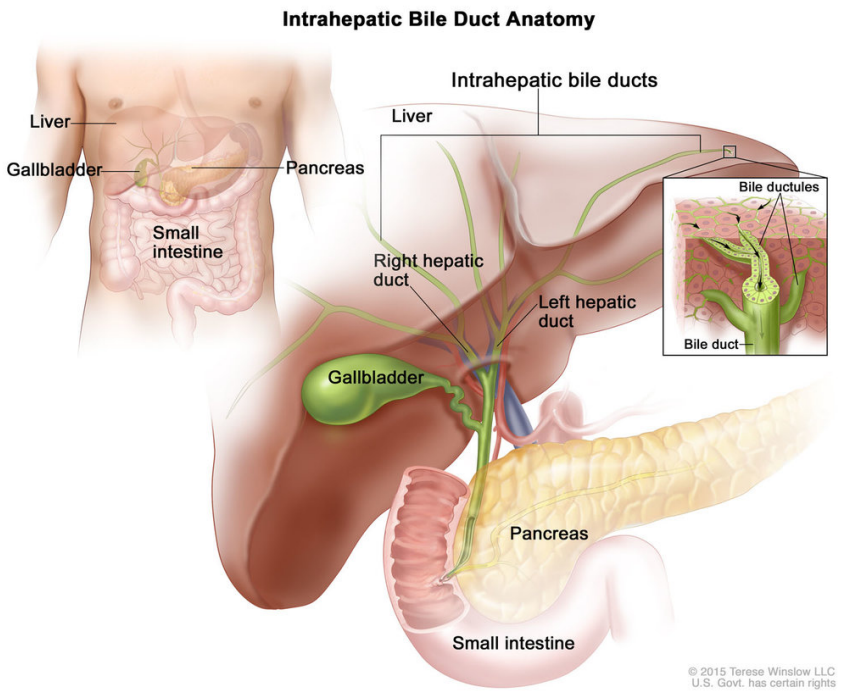


FIGURE 1.2 – Schéma de voies biliaires saines. Les voies intrahépatiques sont séparées en deux segments (gauche et droit). Elles se rejoignent au niveau du hile hépatique et forment le canal hépatique commun. Le canal cystique rejoint ce dernier pour former la voie biliaire principale. Finalement le canal pancréatique principal rejoint la voie biliaire principale et débouchent dans le duodénum via le sphincter d'Oddi. © 2015 Terese Winslow LLC.

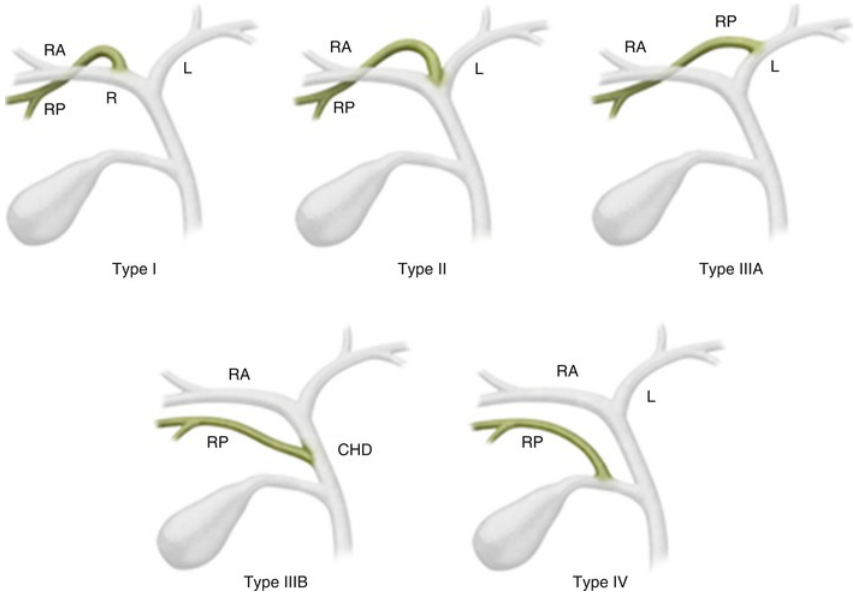


FIGURE 1.3 – Illustration de la variabilité de l'anatomie des voies biliaires. Source : https://radiologykey.com/the-biliary-tree

1.1.2 Pathologies

Plusieurs pathologies peuvent toucher les voies biliaires. On peut en distinguer deux catégories, les calculs biliaires et les tumeurs biliaires.

Les calculs biliaires peuvent être de type cholestérolique ou pigmentaire et peuvent

apparaître à différents niveaux des voies biliaires. Plus un calcul a une forte taille et plus il peut obstruer les voies biliaires en amont, ce qui crée une accumulation de bile. La bile ne pouvant plus s'écouler normalement, passe dans le sang et se diffuse dans le reste du corps. La bilirubine colore en jaune la peau et les yeux du patient, on parle alors d'ictère (jaunisse).

Un calcul engendre en général une dilatation des voies biliaires en amont de la sténose. Les calculs sont éliminés par CPRE grâce à un ballon ou par lithotripsie.

Les tumeurs biliaires peuvent être bénignes ou malignes. Les tumeurs bénignes peuvent avoir un potentiel dysplasique. Dans ce cas, une écho-endoscopie et une ponction ou un brossage permettent de déterminer le type tissulaire.

Les tumeurs malignes des voies biliaires ou cholangiocarcinomes ont une incidence faible. Ils représentent 2000 à 3000 nouveaux cas par an en France. Cependant, ils ont un pronostic sombre dû à l'apparition tardive de symptômes et donc de diagnostic. En effet le taux de survie à 5 ans ne dépasse pas 7% [Malka et al., 2022].

Il existe différents types de cholangiocarcinomes en fonction de leurs localisations. Ainsi on distingue le cholangiocarcinome intra-hépatique, péri-hilaire et extra-hépatique. La prolifération anarchique de cellules induit un rétrécissement des voies biliaires au niveau de la ou des tumeurs. L'écoulement de la bile est ainsi obstrué, ce qui se manifeste par une jaunisse. La structure des voies biliaires est grandement modifiée par le cholangiocarcinome, ce qui les rend particulièrement complexes à interpréter en utilisant l'imagerie pré-interventionnelle. Le geste de CPRE est effectué pour drainer les voies biliaires à l'aide d'une endoprothèse (*stent*). Un brossage ou autre prélèvement peut avoir lieu pour confirmer le diagnostic et initier une chimiothérapie. Dans le cas de sténose indéterminée, la technique SPYGLASS © est parfois utilisée. Il s'agit d'un fil guide muni d'une caméra, ce qui facilite l'intervention pour le brossage. Cependant cette technique est à usage unique et son coût avoisine les 3000 euros. Par conséquent il n'est pas possible de l'utiliser en routine pour toutes les CPRE.

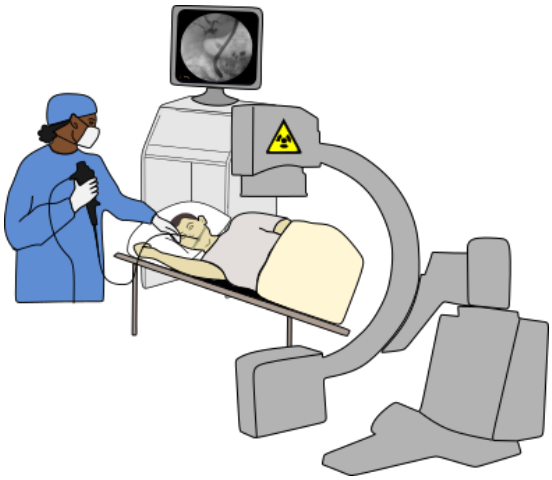
1.1.3 Déroulement de la CPRE

L'accès aux voies biliaires lors de la CPRE se fait grâce à des instruments endoscopiques, insérés par le tube digestif (figure 1.4a). Pour cette raison elle est dite mini-invasive, à l'inverse de la chirurgie. Le fait de passer par des voies pré-existantes supprime le temps de cicatrisation et réduit le risque d'infection. Cependant l'endoscopie suppose une visualisation indirecte des instruments, ce qui rend la navigation plus complexe pour l'endoscopiste [Becq, 2024].

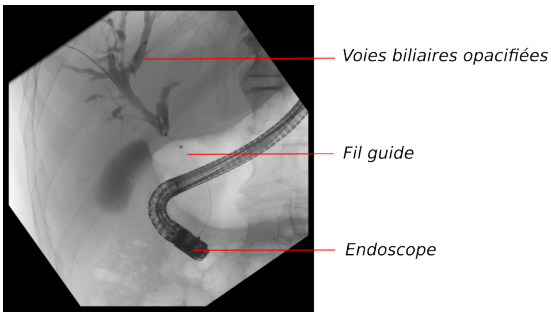
Le geste se déroule en plusieurs étapes. La première consiste à insérer le duodénoscope [2] via l'œsophage et l'estomac jusqu'au duodénum (figure 1.5a). Le duodénoscope est un tube long et flexible avec une caméra latérale. Cette caméra permet la visualisation de la papille qui indique la sortie des voies biliaires. La deuxième étape consiste à entrer dans les voies biliaires en forçant l'accès par la papille à l'aide d'un sphincterotome. Il s'agit de l'étape de canulation. Cette étape est délicate et a une durée variable en fonction du patient.

Un fil guide est ensuite introduit dans la VBP par le sphincterotome dans le sens contraire à l'écoulement normal de la bile (figure 1.5b). C'est la signification du terme

2. Le duodénoscope est un endoscope spécialisé, dans la suite nous y ferons référence également en employant le terme endoscope.

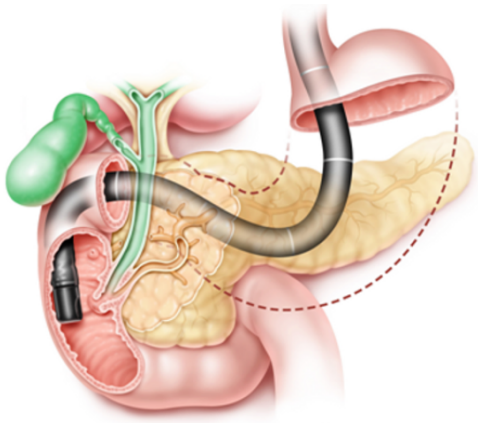


(a) *Illustration simplifiée du geste de CPRE.*

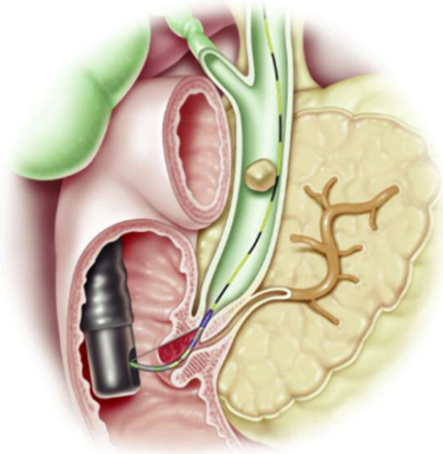


(b) *Exemple d'image de fluoroscopie contenant l'endoscope, le fil guide et les voies biliaires partiellement opacifiées.*

FIGURE 1.4 – *L'endoscopiste insère un duodénoscope jusqu'à la papille biliaire. Un bras radiographique entoure le patient et émet des rayons X lorsque l'endoscopiste appuie sur une pédale. L'image acquise est affichée sur l'écran face à l'endoscopiste. Un autre écran affiche la vue de la caméra du duodénoscope. Les instruments insérés dans les voies biliaires ainsi que le produit de contraste passent par les canaux de l'endoscope.*



(a) *Étape d'introduction du duodénoscope face à la papille.*



(b) *Introduction du fil guide dans la voie biliaire principale après la canulation.*

FIGURE 1.5 – *Principe d'insertion de l'endoscope puis du fil guide. Source : © 2013 Boston Scientific Corporation*

rétrograde dans CPRE. À cet instant, du produit de contraste est injecté pour confirmer le cathétérisme et avoir plus d'informations sur la pathologie.

Le fil guide est un câble fin de 0,89 mm de diamètre inséré via l'endoscope. Il est composé d'une extrémité hydrophile d'environ 5 cm, très souple. Le reste du fil est hydrophobe et légèrement plus rigide. Le fil guide est inséré dans les voies à drainer pour guider les autres instruments comme les ballons ou les prothèses. La difficulté réside dans le fait qu'il soit passif, car l'extrémité n'est pas contrôlable, ainsi l'endoscopiste ne peut pas contrôler précisément sa trajectoire.

La visualisation des instruments et des voies biliaires se fait par fluoroscopie. Le bras radiographique placé autour du patient est actionnable via une pédale. Il produit une image appelée cholangiogramme (figure C.1d). Tandis que les instruments sont radio-opaques les voies biliaires ne sont pas directement visibles par radiographie. L'intervention nécessite donc l'injection de produit de contraste dans les voies biliaires. Cependant les voies biliaires sont rarement entièrement opacifiées par ce procédé, notamment lors d'une obstruction.

Il faut également noter que les doses de produit injecté et de radiations sont limitées. Tout d'abord, l'absence de circulation active (contrairement aux vaisseaux) dans les voies hépatiques conduit le produit de contraste à stagner et fait courir un risque important d'infection au patient. La dose de rayons X doit être limitée pour la sécurité à la fois du patient mais aussi des praticiens. En effet les rayons X sont ionisants. Pour se protéger de leurs effets les opérateurs portent un tablier de plomb ainsi qu'un protège thyroïde.

Une bili-IRM peut être réalisée chez le patient avant l'intervention. Il s'agit d'une imagerie par résonance magnétique avec une hyperpondération en T2 qui permet d'annuler le signal des tissus et d'obtenir seulement le signal de la bile. Cette image permet aux endoscopistes de se représenter les voies biliaires en trois dimensions pour mieux appréhender l'anatomie du patient et préparer l'intervention (figure 1.6). Elle permet par exemple de connaître l'anatomie du patient et s'il y a des variations. Cependant, le signal peut être atténué ou nul en aval d'une sténose ce qui empêche la visualisation de certaines parties de l'arbre biliaire.

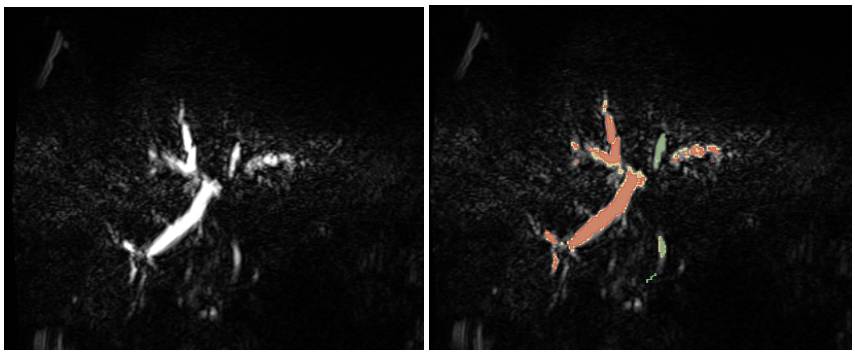


Figure 1.6 – Exemple de bili-IRM à gauche et de la segmentation manuelle effectuée par des endoscopistes et radiologues experts à droite.

1.2 Limites et difficultés de la CPRE

La réalisation d'une CPRE requiert de la part des endoscopistes un haut niveau de dextérité et d'analyse, du fait de la complexité du geste et de la pauvreté du retour visuel, en particulier dans les cas difficiles de sténoses hilaires. Plusieurs aspects sont limitants et entraînent une augmentation du taux d'échec et par conséquent des risques pour les patients.

Premièrement, le suivi radiographique se fait sur des images 2D, ce qui induit une perte d'information sur la structure 3D de l'arbre biliaire. Deuxièmement, la vision radiographique est de mauvaise qualité du fait de la limite de dose. Enfin l'impossibilité d'opacifier certaines voies obstruées empêche carrément leur visualisation.

Ces difficultés engendrent des conséquences non négligeables comme l'allongement du temps d'intervention ou le drainage d'une autre voie que celle d'intérêt. Ainsi, il a été

montré qu'une incertitude planait autour des voies biliaires drainées lors d'une CPRE et certaines imageries post-interventionnelles ont montré une erreur de localisation dans la pose de prothèse [Becq et al., 2023].

1.3 Objectifs du projet MAAGIE

La section précédente a mis en évidence les difficultés majeures de la procédure de CPRE qui a un taux de succès compris entre 80% et 90% pour des endoscopistes expérimentés sur des CPRE avec peu de difficultés [Baron et al., 2006]. Le projet MAAGIE a pour but d'améliorer la réalisation de cette intervention grâce à la collaboration d'experts pluridisciplinaires.

Les endoscopistes ont en particulier besoin de savoir avec précision et certitude dans quelles voies se situent les instruments lors d'une CPRE, notamment pour parvenir à drainer le bon secteur. Ainsi, le projet MAAGIE vise-t-il à proposer un outil d'aide à la navigation qui s'appuie sur l'imagerie IRM pré-opératoire et les images de fluoroscopie per-opératoires.

Le premier objectif est d'être capable de fournir automatiquement un modèle 3D des voies biliaires à partir de l'image de bili-IRM pré-opératoire du patient. Le fait de disposer de ce type de modèle améliorerait la pose de prothèse dans plus de deux cas sur trois [Becq et al., 2023]. Cependant, segmenter l'image de bili-IRM à la main est un travail fastidieux qui prend plus de 25h par patient, sans compter la vérification avec des radiologues experts. Une reconstruction manuelle n'est par conséquent pas envisageable. La thèse d'Abdelhadi Essamlali à l'UTC porte sur la résolution de ce premier objectif [Essamlali et al., 2023].

Le deuxième objectif est d'identifier avec certitude dans quelles voies biliaires se situe le fil guide. La solution envisagée est de créer un algorithme de fusion des informations 2D et 3D pour augmenter la CPRE et calculer la probabilité de présence du fil dans chaque voie.

Le dernier objectif est d'explorer l'utilisation de fils guides actifs contrairement aux fils guides actuels. Ce type d'instruments a pu être testé par la modélisation et l'impression 3D des VB durant la thèse d'Aymeric Becq (voir [Becq, 2024]).

Dans le cadre de cette thèse, nous nous sommes particulièrement intéressés à la segmentation automatique du fil guide sur les images de fluoroscopie 2D afin de les fournir en entrée de l'algorithme de fusion mentionné dans le deuxième objectif. Un tel algorithme suppose trois informations en entrée : premièrement, la reconstruction 3D provenant de la bili-IRM ; deuxièmement, l'image de fluoroscopie 2D ; enfin, les pixels appartenant au fil guide sur cette dernière. L'idéal serait d'obtenir cette dernière information automatiquement à partir de l'image de fluoroscopie 2D. Cette tâche est rendue particulièrement difficile par la finesse de la structure à segmenter et par la faible qualité et la faible quantité d'images. C'est l'objectif principal de cette thèse.

1.4 Contexte méthodologique

L'intelligence artificielle (IA) regroupe un ensemble de théories et de méthodes qui tendent à automatiser certaines tâches considérées complexes bien qu'exécutées de manière naturelle et très satisfaisante par les humains comme par exemple l'utilisation du langage, la conduite, la pose d'un diagnostic ou la création de musique. Ce domaine cherchant à mimer une intelligence humaine artificiellement, s'appuie sur des disciplines variées allant

des mathématiques à la psychologie et trouve des applications dans pléthore de domaines, notamment la médecine.

Plusieurs approches ont vu le jour dans l'histoire de l'IA et deux paradigmes ont émergé parallèlement : **l'IA symbolique et l'IA connexionniste**. La première approche, pour simplifier, cherche à déterminer un ensemble de règles explicites qui modélisent les phénomènes observés en s'appuyant sur les connaissances de ces derniers. Cette approche vise à représenter de manière formelle les connaissances et le raisonnement. Elle est utilisée dans les outils d'aide à la décision par exemple. Elle a pour avantage d'être en théorie compréhensible pour un humain. Cependant, il est difficile de déterminer toutes les règles qui caractérisent un phénomène. La deuxième approche s'appuie sur l'apprentissage de règles implicites en analysant un grand nombre de données. Ainsi, elle a l'avantage de ne pas nécessiter beaucoup de connaissances a priori. Cette approche s'est révélée être très performante dans les années 2010, à la suite du développement de processeurs graphiques de plus en plus efficaces, ainsi que de nouveaux algorithmes d'optimisation. L'IA connexionniste s'appuie sur les réseaux de neurones artificiels qui sont des modèles très puissants mais très lourds et gourmands en ressources et temps de calcul. Une autre limite des réseaux de neurones est leur opacité pour un utilisateur humain. Il est difficile de dire avec certitude qu'un modèle a prédit un résultat Y pour la raison X. On parlera d'apprentissage par ordinateur par la suite pour désigner ce type de méthode d'IA.

Un troisième paradigme est en train d'émerger, celui de l'IA hybride. Ce nouveau paradigme prend conscience des forces et des faiblesses des deux autres, et tente de les combiner pour obtenir des méthodes ou systèmes d'IA à la fois performants et dont les décisions sont explicables, intelligibles et compréhensibles pour un humain.

Dans ce travail de thèse, nous cherchons à utiliser les méthodes d'analyse d'image pour la reconnaissance automatique des instruments sur les images de fluoroscopie 2D. Cette tâche s'inscrit dans le domaine de la vision par ordinateur. Elle s'appuie sur un des fondements de la vision par ordinateur qui est la segmentation, décrite plus en détail dans la suite de cette section.

1.4.1 Segmentation

Alors qu'il est aisé pour un humain de reconnaître une structure sur une image et de la détourer, ce processus est plus difficile du point de vue de l'imagerie numérique. De nombreux travaux d'analyse d'images portent sur cette tâche d'extraction des pixels appartenant à un même objet, que l'on désigne par **segmentation**. Le processus de segmentation automatique ou semi-automatique est un processus complexe qui demande la perception et l'identification de l'objet sur l'image. Notons que nous ne nous intéressons pas à la segmentation manuelle qui est soumise à de forts biais d'annotateurs et qui est fastidieuse à mettre en place.

Au fil du temps, diverses approches ont été développées, nous les divisons en plusieurs grandes catégories :

- **Segmentation par seuillage**

  Seuiller une image consiste à fixer une condition d'appartenance d'un pixel donné à l'objet en fonction de son intensité. Si l'intensité est supérieure au seuil alors le pixel appartient à l'objet sinon il n'y appartient pas. Le seuil peut être déterminé manuellement mais cette méthode n'est pas très généralisable. Des approches automatiques comme la méthode d'Otsu [Otsu, 1979] trouvent le seuil optimal pour chaque image à partir de leur histogramme.

- **Segmentation par contours**

  Les contours matérialisent la limite physique et perceptuelle des objets. Ainsi, plusieurs méthodes d'analyse d'images s'appuient sur la détection de ces derniers. Certains filtres comme celui de Sobel [Kanopoulos et al., 1988] mettent en avant la différence subite d'intensité qui correspond au contour. D'autres méthodes, comme les contours actifs ou *snakes* [Kass et al., 1988] consistent à minimiser l'énergie d'une courbe paramétrique, initialisée par l'utilisateur, proche du contour, et la faire converger vers ce dernier.

- **Segmentation par régions**

  Ce type d'approche est dual par rapport à celle par contours. Il s'agit de regrouper des régions selon un critère de similarité, comme la méthode de partage et réunion de régions (*Split and merge*) [Cheevasuvit et al., 1986]. On peut également citer les méthodes par croissance de régions, qui forment des régions en partant de germes. Les pixels sont ajoutés aux régions en fonction de critères définis au préalable par l'opérateur. Citons aussi la ligne de partage des eaux [Beucher and Meyer, 1993], les filtres connexes, les champs aléatoires de Markov [Geman and Geman, 1984] ou encore les modèles déformables s'appuyant sur l'homogénéité des régions [Chan and Vese, 1999].

- **Segmentation par apprentissage statistique**

  Ces méthodes s'appuient sur peu de connaissances a priori et « apprennent » des modèles à partir des données qui leurs sont fournies. On peut citer les k-moyennes [Forgy, 1965] qui est un algorithme de regroupement non supervisé. En parallèle, Rosenblatt [1958] développait le Perceptron qui est le premier neurone artificiel. Cette approche est améliorée par Rumelhart et al. [1986] qui ajoutent des couches et le mécanisme de rétropropagation du gradient. Les réseaux de neurones artificiels se sont peu à peu complexifiés pour donner des architectures telles que celles de He et al. [2016], Huang et al. [2017], et bien d'autres. Ce type de méthode peut être supervisé ou non. L'inconvénient des réseaux de neurones est qu'ils nécessitent un nombre conséquent de données pour les entraîner.

1.4.2 Difficultés de la segmentation d'images de CPRE

Le principal frein à la segmentation automatique d'un fil guide de CPRE est sa surface réduite sur l'image car il représente moins de 1% de l'image et sa structure fine et allongée est difficile à percevoir. La structure du fil guide est telle qu'elle s'apparente plutôt à un contour qu'à une région. Les méthodes de segmentation par régions ne sont donc pas adaptées à ce cas.

Les images contiennent par ailleurs beaucoup de bruit à cause de la limitation de la dose de radiations et des mouvements pendant l'intervention. Des éléments différents du fil guide, par exemple l'endoscope, des électrodes, ou les vertèbres, empêchent les méthodes de seuillage d'être effectives car elles ne permettent pas d'isoler le fil guide.

Enfin, la quantité de données est très limitée, ce qui n'est pas optimal pour l'utilisation de réseaux de neurones artificiels qui nécessitent un nombre conséquent de données pour généraliser.

1.5 Questions de recherche

L'intérêt que nous portons à la segmentation d'images de CPRE et à ses difficultés nous amène à réfléchir aux questions suivantes :

- Comment segmenter des structures fines et allongées avec une base de données réduite ?
- Comment évaluer la segmentation de telles structures ?
- La segmentation d'autres structures comme l'endoscope peut-elle améliorer celle du fil ?

1.6 Organisation du manuscrit

La suite de ce document est organisée en sept chapitres.

Le chapitre 2 dresse un état de l'art de la CPRE assistée par ordinateur, en abordant différentes étapes de l'intervention et donc différents besoins. Ce chapitre évoque les méthodes d'évaluation de la difficulté de l'intervention, l'aide à la navigation par recalage 2D-3D et la segmentation des instruments sur les images de fluoroscopie. L'absence d'articles concernant la CPRE pour la dernière section nous conduit à nous intéresser à un domaine qui présente des similarités, la segmentation de cathéters dans le cadre d'interventions endovasculaires. Par ailleurs, les méthodes générales de segmentation d'images biomédicales sont également introduites.

Le chapitre 3 détaille la collecte et l'homogénéisation des données des deux centres hospitaliers avec lesquels nous avons collaboré. Il décrit la construction de la base de données, les données et leur annotation.

Dans le chapitre 4, nous discutons des différentes méthodes d'évaluation des performances de la segmentation de structures fines. Les méthodes d'évaluation sont essentielles pour comparer les différentes approches possibles de la segmentation. Nous montrons dans ce chapitre que les méthodes classiques d'évaluation peuvent être inadaptées au cas d'étude qui nous intéresse et regroupons d'autres mesures existantes qui viennent compléter les premières. La valeur de ces méthodes d'évaluation est également discutée par rapport aux attentes et à la compréhension des médecins.

Le chapitre 5 se concentre sur la segmentation du fil guide. Nous nous inspirons des méthodes décrites pour les interventions endovasculaires et utilisons le réseau de neurones U-Net [Ronneberger et al., 2015]. Nous y comparons l'effet de différentes fonctions de coût sur le résultat de la segmentation.

Le chapitre 6 porte dans un premier temps sur la segmentation de l'endoscope dans les images des deux bases de données. La seconde partie du chapitre porte sur la segmentation conjointe de l'endoscope et du fil guide pour déterminer si les réseaux de neurones modélisent la relation spatiale entre les deux instruments et en déduisent une meilleure segmentation du fil guide.

Le chapitre 7 conclut ce travail en rappelant les contributions et en proposant des perspectives.

CHAPITRE 2

État de l'art des méthodes d'assistance à la CPRE par ordinateur

Ce chapitre dresse un état de l'art des méthodes développées pour améliorer la CPRE grâce à des outils numériques. Il se décline en trois parties distinctes qui répondent à différents enjeux de l'intervention. Premièrement, un panorama des méthodes d'évaluation de la difficulté de certaines étapes de l'intervention est dressé. Deuxièmement, sont abordées les méthodes d'aide à la navigation par recalage et fusion de modalités 2D et 3D. Finalement, une dernière partie traite des méthodes de segmentation pour le suivi automatique des instruments, qui est la tâche que nous souhaitons réaliser *in fine*.

2.1 Évaluation de la difficulté de l'intervention

La CPRE est une intervention complexe à plusieurs égards. Une des étapes clés est la canulation, moment où les endoscopistes entrent dans les voies biliaires via la papille duodénale afin d'y introduire les instruments de l'intervention. Cette étape va déterminer la réussite de la procédure et la probabilité d'effets indésirables.

Ainsi Kim et al. [2021] proposent un outil d'IA pour localiser la papille et prédire la difficulté de canulation. Ils utilisent des images provenant de l'endoscope et prédisent des boîtes englobantes et une distribution de probabilités pour localiser la papille grâce à un réseau U-Net pré-entraîné sur ImageNet qui est une grande base de données récoltée par Deng et al. [2009] et Krizhevsky et al. [2012]. Pour cette tâche, 531 images sont utilisées et ils obtiennent une Intersection sur l'Union de 64,1%, une précision de 76,2% et un rappel de 78,4%. La prédiction de la difficulté de la canulation repose sur une classification faite par réseaux de neurones convolutifs à partir des images de papille. Les images de 451 patients sont disponibles pour cette tâche. Les auteurs commencent par mener une classification binaire des « cas faciles » et « cas difficiles », les cas faciles étant ceux requérant moins de cinq minutes et les cas difficiles tous les autres. Une deuxième classification est menée cette fois-ci avec quatre classes : facile, plus de cinq minutes, plusieurs techniques de canulation et échec. Plusieurs réseaux de neurones sont comparés pour la classification : VGG 19 [Simonyan and Zisserman, 2015], Res-Net50 [He et al., 2016] et Dense-Net 161 [Huang et al., 2017]. Les meilleurs résultats sont obtenus avec le réseau ResNet pour la classification binaire avec un rappel de 71,9% pour les cas faciles et 61,9% pour les cas difficiles.

Une deuxième application par Huang et al. [2021] consiste à prédire un score de difficulté de la CPRE dans le cas des calculs de la voie biliaire principale. Le score s'appuie sur la segmentation de la voie biliaire principale, du calcul et du duodénoscope sur les images de fluoroscopie. Les auteurs déduisent de ces segmentations la taille du calcul, le diamètre et l'angle d'orientation de la voie biliaire principale. Ces paramètres permettent d'évaluer la difficulté de l'intervention consistant à retirer le calcul. La méthode utilisée consiste en deux réseaux de neurones convolutifs de bout-en-bout en cascade. D'une part un réseau D-LinkNet34 [Zhou et al., 2018] a été entraîné pour segmenter la voie biliaire principale. Puis un réseau U-Net [Ronneberger et al., 2015] en cascade pour extraire les calculs. Un deuxième réseau D-LinkNet34 a été entraîné en parallèle pour la segmentation du duodénoscope. L'algorithme de segmentation a été entraîné sur 1381 images et testé sur 573 images. La segmentation est évaluée grâce à l'intersection sur l'union moyenne (mIoU). Les auteurs obtiennent un score de 68,3%, 86,42% et 95,85% pour la segmentation du calcul, de la voie biliaire principale et du duodénoscope respectivement.

Ces applications sont des manières d'améliorer des points cruciaux de la CPRE en évaluant la difficulté à différents niveaux. Ces approches pour la CPRE demeurent néanmoins éloignées de notre objectif de rendre la navigation plus sûre pour les endoscopistes. Un autre moyen d'aider les endoscopistes durant l'intervention est de faciliter la navigation à travers les voies biliaires, en particulier grâce au recalage des modalités 2D et 3D. C'est l'objet de la section suivante.

2.2 Recalage 2D-3D pour l'aide à la navigation

Rappelons que le suivi en 2D limite les endoscopistes durant la CPRE puisqu'ils ne disposent pas de l'information de la profondeur à laquelle se situent leurs instruments dans les voies biliaires. Ce manque d'information est un frein au bon déroulé de l'intervention comme démontré dans [Becq et al., 2023]. Pour pallier cette difficulté, une image 3D des voies biliaires est quasi systématiquement réalisée en amont de l'intervention. Inspirées par le domaine endovasculaire et notamment par les travaux d'Ambrosini et al. [2015], plusieurs équipes de recherche ont cherché à faciliter la navigation des endoscopistes grâce à la fusion des données pré-opératoires en 3D avec celles en 2D. En effet la fusion bimodale a permis d'améliorer la sensibilité et le contrôle des endoscopistes en réduisant la dose de radiation ainsi que le temps d'intervention. Cette section détaille les différentes méthodes recensées.

Reuterwall et al. [2020] sont les premiers à proposer une CPRE bimodale. Une étude rétrospective est menée à partir d'images de bili-IRM (hyperpondération T2) et de fluoroscopie de 13 patients. La première étape consiste à recaler une IRM en T1 et l'image de fluoroscopie de l'abdomen grâce à des points de repère, la coupole et la colonne vertébrale. Les auteurs utilisent l'IRM en T1 car elle est de meilleure résolution spatiale que la bili-IRM. Une fois le recalage réalisé, ils projettent l'image de bili-IRM en T2 recalée avec l'image en T1 sur la fluoroscopie (figure 2.1). Le recalage dure en moyenne 11,9 minutes.

Les repères choisis sont fiables et fixes mais les voies biliaires peuvent se déformer même avec ces marqueurs du fait de la pression exercée par les instruments et de la position du patient. De plus l'équipe utilise des logiciels Siemens qui sont propriétaires. Malgré tout, ces étapes de marquage et recalage pourraient être réalisées avec d'autres logiciels. Les auteurs ne fournissent pas de mesure d'erreur de recalage. Enfin, le temps de recalage n'est pas rédhibitoire pour ce type d'intervention mais il ne s'agit certainement pas d'une méthode en temps réel. Ce point pourrait être amélioré. D'autre part, l'image utilisée est

FIGURE 2.1 – Image tirée de Reuterwall et al. [2020]. a) IRM pondérée en T1 avec le marquage du contour du foie en jaune. b) Bili-IRM recalée avec l'image (a). c) Marquage du contour du foie sur une image fluoroscopique standard. Recalage entre (a) et (c). d) Image résultant de l'alignement manuel des points de repères, fusion bimodale. Les voies biliaires 3D sont projetées sur l'image de fluoroscopie.

la bili-IRM brute qui contient du bruit. Essamlali et al. [2023] proposent de segmenter les voies biliaires automatiquement à partir de l'image de bili-IRM pour extraire seulement le signal utile.

Le recalage 3D-2D en CPRE est également envisagé par Zhang et al. [2023], cette fois-ci entre l'image de tomodensitométrie (CT) pré-opératoire et la fluoroscopie. Un premier test a été réalisé sur deux fantômes de voies biliaires. Puis le recalage est effectué lors de 18 CPRE. Les os, le foie, la vésicule et les voies biliaires sont segmentés à partir de l'image CT avec un U-Net 3D. Le recalage se fait à partir des os (vertèbres et côtes) et des lignes centrales des voies biliaires (figure 2.2). Le temps de recalage est d'environ une minute. Une étape est ajoutée pour enlever le bruit dû au mouvement de respiration du patient. Une erreur moyenne de recalage de 0,46mm±0,13mm et une erreur d'alignement, c'est-à-dire après compensation du mouvement de la respiration, de 0,64mm±0,24 sont obtenues.

Dans ces deux approches les images 3D sont projetées sur l'image de fluoroscopie. Les choix d'IRM par rapport au CT ne sont, très souvent, pas justifiés dans ces travaux. Dans les hôpitaux avec lesquels nous collaborons, l'IRM est plus souvent utilisée en clinique en pré-opératoire. C'est donc vers cette modalité que le projet MAAGIE s'est tourné pour l'étape de fusion 2D/3D.

Finalement, une dernière approche de réalité augmentée pour la CPRE, développée par Yang et al. [2024], utilise le recalage entre l'image de tomodensitométrie du patient et l'extrémité magnétique du fil guide. Une correction du mouvement dû à la respiration est

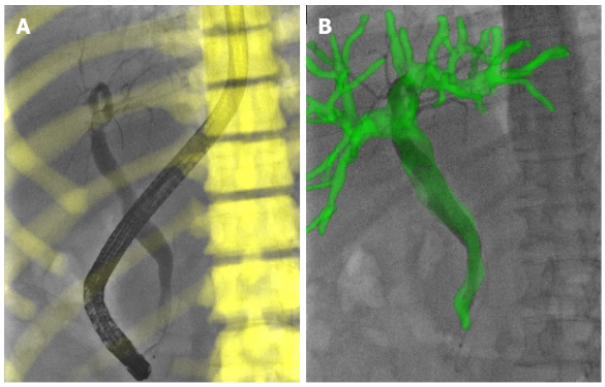


Figure 2.2 – D'après Zhang et al. [2023]. A) Image de fluoroscopie. Les os sont utilisés comme repères pour le recalage. B) Projection du modèle biliaire sur l'image de fluoroscopie.

appliquée. Une projection de la caméra de réalité augmentée sur le torse du patient permet de s'affranchir des rayons X et du produit de contraste en affichant l'extrémité du cathéter et les éléments de l'image de tomodensitométrie. L'erreur moyenne d'alignement, qui évalue la correction du mouvement, est de 0,75±0,17mm sur les fantômes et 2,79±0,46mm sur les patients. L'erreur de position du fil guide durant les expériences de navigation sur le fantôme est de 0,98±0,15mm. Enfin, ce que les auteurs appellent erreur de fusion moyenne, qui mesure l'erreur lors de la projection des points sur le patient, est de 1,67±0,34mm. Cette méthode a pour intérêt de supprimer les radiations et l'utilisation de produit de contraste. Cependant l'emploi d'un localisateur magnétique et d'instruments spécifiques engendrent un encombrement, un sur-coût et une complexification, non négligeables au regard de l'intervention.

Les erreurs de recalage obtenues par ces méthodes sont acceptables au regard des diamètres des voies biliaires. Dans les deux première méthodes, le recalage ne prend pas en compte la position du fil guide dans les VB mais seulement des structures anatomiques. Or les instruments et la position du patient déforment les voies biliaires qui ne seront donc pas exactement comme sur l'imagerie 3D. Ces raisons peuvent engendrer une erreur de localisation du fil guide dans les voies biliaires et mener à l'échec de la procédure.

Afin de surmonter ces contraintes, nous proposons de détecter le fil guide directement sur les images de fluoroscopie et de fournir ces informations à un algorithme de fusion et de localisation du fil guide dans les VB, développé à l'ISIR par Clément Alberge et Philippe Gauthier. Ainsi, la dernière section aborde l'étape de suivi automatique des instruments qui est nécessaire pour le recalage avec cet algorithme.

2.3 Segmentation du fil guide sur les images de fluoroscopie par des méthodes d'intelligence artificielle

Cette section se décline en deux parties, une première partie générale sur la segmentation d'images biomédicales, puis une partie qui se concentre sur la segmentation du fil guide pour les interventions médicales endovasculaires, car aucun travail n'a été fait sur la CPRE.

2.3.1 Segmentation d'images biomédicales

La segmentation de régions d'intérêt en imagerie biomédicale est une tâche récurrente qui consiste à identifier et indiquer quels pixels appartiennent à une région sur une image. La connaissance de la position et du contour d'un objet permet aux utilisateurs d'effectuer des mesures ou de détecter des lésions. Elle peut se faire manuellement, ce qui est extrêmement fastidieux et peu reproductible car il existe une forte variabilité intra- et inter-annotateurs. Pour cette raison une large communauté de scientifiques s'attelle, depuis plusieurs décennies, à développer des méthodes de segmentation automatique [Alzahrani and Boufama, 2021]. Une des méthodes communément utilisées est l'apprentissage supervisé avec des réseaux de neurones. Cette section dresse un panorama des méthodes récentes et fondées sur l'emploi de réseaux de neurones en segmentation d'images biomédicales.

On parle d'apprentissage supervisé lorsque l'on fournit à l'algorithme d'apprentissage des données annotées. En segmentation d'images biomédicales cela se traduit par donner en entrée d'un réseau des images et les segmentations de référence correspondantes. Ces segmentations de référence indiquent au réseau quels pixels constituent l'objet à segmenter et permettent la comparaison avec la prédiction du réseau pour l'apprentissage et l'évaluation finale.

U-Net

En 2015, l'avènement de U-Net [Ronneberger et al., 2015] marque un tournant décisif dans le domaine de la segmentation d'images biomédicales. Ronneberger et al. ont développé une architecture originale de réseau de neurones convolutifs qui s'imposera largement dans le domaine biomédical pour ses performances.

Les auteurs introduisent non seulement une architecture de réseau en U (figure 2.3) mais également des connexions entre la partie d'encodage et de décodage qui améliorent la localisation des structures à détecter. Chaque étage contient deux étapes de convolution, l'application d'une fonction d'activation et une étape d'agrégation. La fonction d'activation choisie est la fonction dite « ReLU » pour unité linéaire rectifiée [1]. De plus, une étape d'augmentation des données, c'est-à-dire de transformations faites sur les données en entrée du réseau pour en augmenter artificiellement le nombre et la variété, est ajoutée afin de compenser le manque d'images annotées disponibles, ce qui est souvent le cas dans les applications médicales.

nnU-Net

À la suite du succès de U-Net, « No-new UNet » ou nn-UNet, développé par Isensee et al. [2021], utilise le réseau de neurones convolutifs U-Net [Ronneberger et al., 2015] couplé à des règles heuristiques qui fixent automatiquement les paramètres permettant d'atteindre des performances qui sont actuellement au meilleur niveau de l'état de l'art en segmentation d'images biomédicales pour une grande diversité d'images et de régions d'intérêt à segmenter. Cette méthode est illustrée à la figure 2.4.

Paramètres fixés Certains paramètres de la méthode sont fixes quelles que soient les données en entrée. Ils ont été choisis par les auteurs comme configuration robuste. Ils

1. $ReLU(x) = max(0, x)$

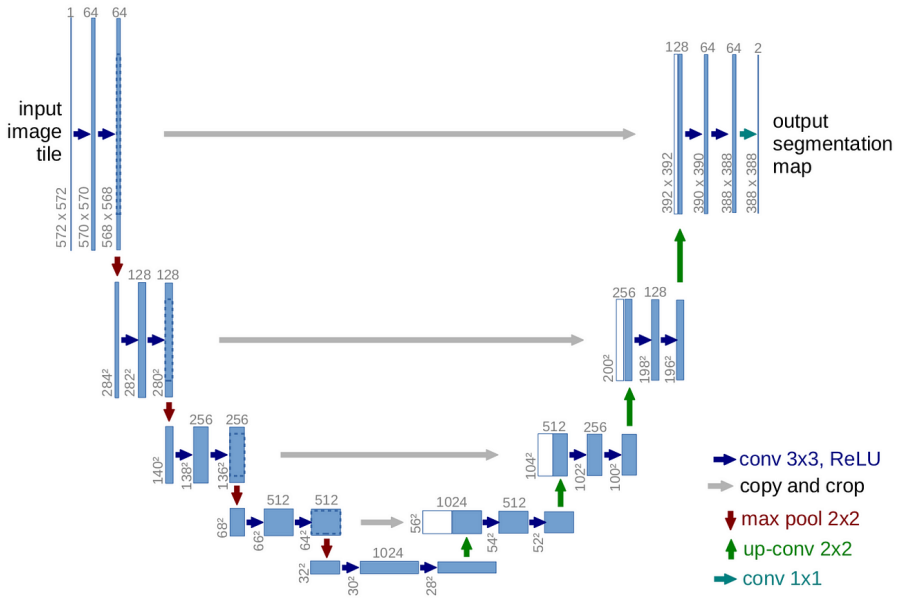


Figure 2.3 – Illustration de l'architecture du réseau U-Net par Ronneberger et al. [2015]

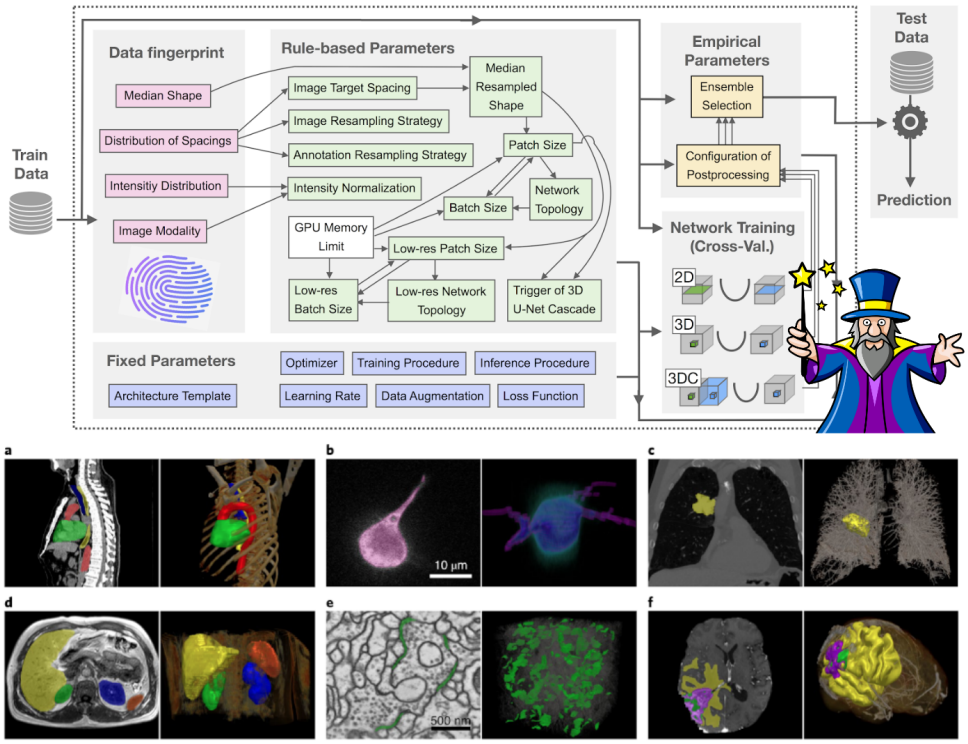


Figure 2.4 – Schéma illustratif de la méthode nn-UNet [Isensee et al., 2021].

concernent l'architecture des réseaux U-Net, le schéma d'optimisation et l'augmentation des données principalement.

- **Architecture du réseau**

  L'architecture globale des réseaux nnU-Net en deux et trois dimensions est la même que celle décrite par Ronneberger et al. [2015]. Malgré le maintien de la charpente de U-Net, certaines modifications sont apportées au niveau des couches. Tout d'abord,

le réseau normalise les instances, c'est-à-dire un échantillon du jeu d'entraînement plutôt que sur l'ensemble du lot, ce qui est plus adapté aux petites tailles de lots. Les activations ReLU présentes dans U-Net sont remplacées par des *leaky ReLU*[2] avec une pente de $-0,01$. Ainsi, chaque résolution est constituée de deux blocs, eux-mêmes constitués d'une convolution, d'une normalisation et d'une activation par *leaky ReLU*.

- **Entraînement et optimisation**

  La fonction de coût est la somme des fonctions de Dice et d'entropie croisée (elles seront détaillées au chapitre 5) pour chaque résolution. Le fait de calculer la fonction de coût pour chaque résolution s'appelle *supervision profonde* et permet d'améliorer l'entraînement. La méthode d'optimisation est une descente de gradient stochastique avec un *momentum* de Nesterov $\mu = 0,99$. Le taux d'apprentissage est fixé à 10^{-2} au début de l'entraînement. Puis il décroît à un taux de $(1 - \text{époque}/\text{époque}_{max})^{0,9}$ au cours de l'entraînement, époque désignant le nombre d'itérations complètes sur le jeu de données d'entraînement. Le nombre d'époques est fixé à 1000 pour tous les entraînements.

- **Augmentation des données**

  Une des étapes les plus importantes de la méthode nnU-Net est celle d'augmentation des données pour limiter l'effet du sur-apprentissage sur un jeu de données limité. En effet, moins un jeu de données est diversifié, plus la probabilité que le modèle apprenne des caractéristiques spécifiques au jeu de données est importante. Malgré de très bonnes performances sur le jeu de données d'entraînement, ce modèle ne sera pas capable de généraliser les caractéristiques apprises pour prédire les segmentations sur d'autres jeux de données. Une des stratégies pour diversifier le jeu de données est l'augmentation de données. Elle consiste à appliquer des transformations de toutes sortes sur le jeu de données initial pour le diversifier artificiellement. Dans nnU-Net plusieurs transformations sont appliquées avec des lois de probabilité uniformes :

  - **Spatiales**. Il s'agit de rotations, de changement d'échelle et d'effet miroir.
  - **Intensité**. Les intensités des pixels peuvent être modifiées ainsi que le contraste.
  - **Bruit**. Du bruit gaussien de moyenne nulle et d'écart-type $\sigma = 0,1$ est ajouté avec une probabilité de $0,15$.
  - **Flou**. Du flou gaussien est également appliqué avec une probabilité de $0,2$ avec un écart-type σ choisi uniformément dans l'intervalle $[0,5; 1,5]$.

Paramètres inférés

Le reste des paramètres est inféré grâce à des règles heuristiques et à l'extraction d'une empreinte pour le jeu de données. Cette empreinte va contenir les caractéristiques nécessaires à déterminer ces paramètres qui ne dépendent par conséquent que du jeu de données et pas des utilisateurs.

2. $LeakyReLU(x) = max(ax, x)$ avec $a = -0,01$ ici.

Empreinte du jeu de données Le jeu de données est automatiquement analysé par nnU-Net pour produire une empreinte. Celle-ci prend en compte les paramètres tels que la taille des images, la dimension des pixels, le nombre de classes et le nombre d'images d'entraînement. Les propriétés suivantes sont également prises en compte ; la moyenne, l'écart-type et les cinquième et quatre-vingt quinzième centiles des intensités des régions d'intérêt. Ces informations combinées à des règles heuristiques donnent les paramètres inférés pour le pré-traitement et le choix de la topologie du réseau. Détaillons ces paramètres :

- **Restriction à la zone anatomique** : Cette étape rogne les parties extérieures de l'image qui sont nulles pour ne conserver que la région utile. Sur les images IRM de la tête notamment, la boite crânienne est entourée d'une grande région vide. Cette étape permet de réduire le temps d'entraînement.

- **Normalisation en intensité** : Il s'agit de la seule étape de pré-traitement qui varie en fonction de la modalité d'imagerie. En effet, les images de tomodensitométrie sont exprimées en unités Hounsfield (HU) qui sont constantes pour un tissu donné. Cela veut dire que le foie du patient X aura une intensité du même ordre de grandeur que celui du patient Y. Cela n'est pas vérifié pour les autres modalités d'imagerie, comme l'IRM par exemple. Pour le premier cas la normalisation consiste à soustraire la moyenne des intensités de l'ensemble des données puis à diviser par l'écart-type. Dans notre cas, la normalisation consiste à soustraire pour chaque pixel la moyenne des intensités de l'image et à les diviser par l'écart-type des intensités de l'image (Z-score) (voir 5.3.1).

- **Ré-échantillonnage de l'espacement des voxels.** Cette étape n'existe que pour les images en 3D.

- **Topologie du réseau, tailles des imagettes et des lots**. Les tailles de lot et des imagettes sont intimement liées et il s'agit de trouver un compromis pour ne pas dépasser la limite des ressources de calcul. Ainsi nnU-Net privilégie une grande taille d'imagettes quitte à réduire la taille des lots et à avoir un gradient plus bruité. À partir de ces tailles, le nombre de sous-échantillonnages et les tailles de noyaux pour la convolution sont calculés automatiquement.

Paramètres empiriques

Sélection de la meilleure configuration U-Net Quatre configurations de réseau U-Net sont possibles, 2D, 3D avec résolution maximale ou faible et une cascade entre un U-Net 3D de faible résolution suivi d'un à la résolution maximale. Les performances de chaque configuration peuvent être fusionnées en prenant la moyenne des probabilités de sortie, c'est-à-dire les scores de la fonction *Softmax*.

Post-traitement La dernière étape consiste à conserver uniquement la plus grande composante connexe en traitant toutes les classes comme une seule et à voir si le score de Dice augmente. Si tel est le cas, la plus grande composante connexe est conservée et la même procédure est appliquée à chaque classe individuellement.

Discussion Cette méthode se hisse sur le podium pour 9 compétitions sur 10 [Ma, 2021] sur des bases de données publiques dans le cadre de la conférence MICCAI 2020 et 5 sur 7 pour MICCAI 2021 [Isensee et al., 2021] grâce à sa capacité de généralisation. Elle est par ailleurs développée pour être facilement utilisable pour un public non informaticien mais sensibilisé. En effet, il n'y a pas de code à écrire mais seulement deux étapes : mettre ses données au format prescrit et exécuter des commandes dans un terminal.

Cependant, du fait de son développement pour être un outil clé en main et généralisable, il est parfois fastidieux de modifier les paramètres de la méthode pour pouvoir en tester d'autres, comme la fonction de coût par exemple. De plus, cette méthode nécessite des ressources et un temps de calcul conséquents. Enfin, la plupart des jeux de données sur lesquels elle a été testée sont adaptés à la segmentation d'objets de taille plus importante que celle du fil guide. Il n'est pas garanti que la méthode fonctionne pour des éléments très fins comme le fil guide avec un jeu de données limité.

Plus récemment, la méthode *Segment Anything in medical images ou MedSAM*[3](segmenter n'importe quoi dans les images médicales) [Ma et al., 2024], inspirée de *Segment anything* de Kirillov et al. [2023] (figure 2.5), a été entraînée sur plus d'un million d'images médicales comportant dix modalités différentes. Le modèle prend en entrée une image et une information de l'utilisateur (points, boîte englobante, texte). Les entrées sont encodées parallèlement par un réseau de type *transformer* Dosovitskiy et al. [2021], puis un décodeur fournit un masque de segmentation de l'objet pointé par l'utilisateur sur l'image. Cette méthode se veut plus généralisable que les méthodes actuelles car elle est entraînée sur une large base de données, contrairement à nn-UNet qui donne de très bonnes performances par tâche mais qui n'est pas entraîné de manière globale pour prendre en entrée n'importe quelle image. Contrairement à beaucoup de méthodes par réseaux de neurones, MedSAM n'est pas entièrement automatique et nécessite une interaction avec l'utilisateur pour réaliser la tâche de segmentation. Les auteurs justifient ce choix face à la diversité de problèmes de segmentation dont regorge le domaine médical, ce qui rend la méthode plus modulaire et donc utilisable pour un plus grand nombre de problèmes. Une des limites exposées dans l'article est la sous-représentation de certaines modalités car la base de données contient majoritairement des images IRM, de tomodensitométrie et d'endoscopie, et relativement peu d'images angiographiques.

De là, une moindre performance de l'algorithme est observée pour segmenter des structures telles que les vaisseaux. Le problème de détection de fil guide se situe dans les limites de la méthode car il n'y a pas de fluoroscopie dans la base de données d'entraînement et les fils guides représentent une partie infime de l'image. La figure 2.6 montre un essai de MedSAM sur un image de notre base de données. Les auteurs préconisent de choisir une petite boîte englobante car le réseau a été entraîné ainsi. Nous observons que le réseau prédit plutôt un objet large comme une tumeur et ne reconnaît absolument pas le fil guide.

Ainsi, malgré une forte recrudescence de méthodes développées en segmentation d'images médicales, celles utilisant U-Net et nnU-Net se sont imposées comme méthodes phares.

La suite du chapitre s'intéresse à la détection automatique du fil guide sur les images de fluoroscopie.

3. https://github.com/bowang-lab/MedSAM

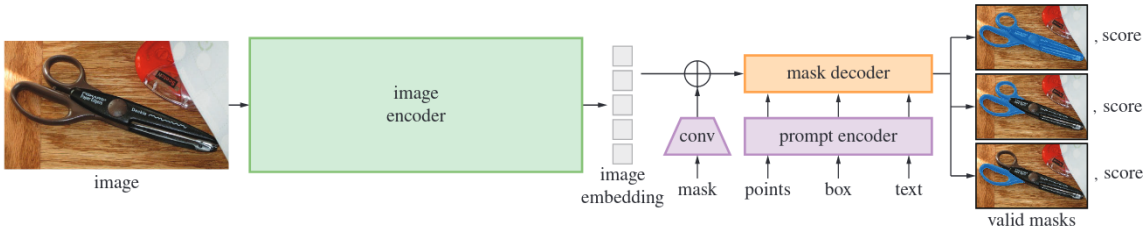


Figure 2.5 – Schéma du modèle SAM des auteurs Kirillov et al. [2023].

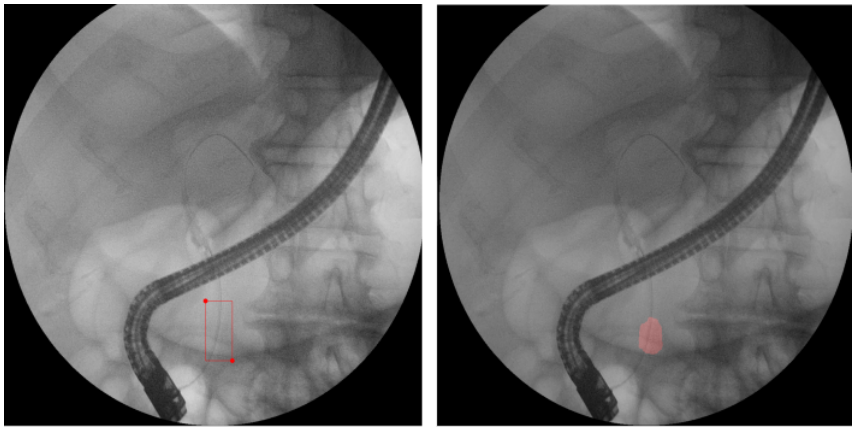


Figure 2.6 – Application de MedSAM sur les cholangiogrammes. À gauche, sélection de la boîte englobante où se trouve l'objet à segmenter c'est-à-dire le fil guide. À droite, la prédiction par le modèle SAM.

2.3.2 Segmentation du fil guide

Notre objectif est de connaître la position du fil guide dans l'image pour aider à sa localisation dans les voies biliaires. Nous nous intéressons donc à la segmentation du fil guide sur les images de fluoroscopie. Le cas de détection ou segmentation du fil guide dans le cas de la CPRE n'est cependant pas traité dans la littérature. Par conséquent nous nous sommes intéressés aux applications de même type dans le domaine cardio-vasculaire qui est, selon nous, le plus proche tout en conservant des différences importantes notamment au niveau du diamètre du fil guide et du contraste des images. En effet la plupart des cathéters utilisés dans le domaine endo-vasculaire ont un diamètre de $1,6$ mm alors que celui du fil guide pour la CPRE est de $0,89$ mm. Enfin, les configurations du fil guide dans le cas de la CPRE sont plus diverses que lors d'interventions vasculaires, du fait de la diversité anatomique des voies biliaires.

Nous détaillons dans cette section les méthodes qui utilisent des réseaux de neurones convolutifs et notamment le réseau U-Net [Ronneberger et al., 2015].

Ambrosini et al. [2017] s'intéressent à la détection de cathéters et de fils guides sur les séquences fluoroscopiques 2D acquises lors de procédures de cathétérisme hépatique, comme illustré dans la figure 2.7. Pour cela, ils utilisent un réseau U-Net [Ronneberger et al., 2015] couplé à un algorithme de connexion de branches sur des séquences d'images. Le réseau est entraîné avec 182 séquences et testé avec 55 séquences, chacune composée de quatre images. La prédiction en sortie du réseau est seuillée. Puis le squelette de cette prédiction est extrait. Enfin un algorithme reconnecte les branches de squelette entre elles pour contrer le fort taux de faux négatifs. Le critère de connexion est une distance maximale à ne pas dépasser entre deux points. Les auteurs évaluent la détection

en calculant l'erreur médiane en mm par rapport à la ligne centrale ainsi que par rapport à l'extrémité du cathéter. En effet, les distances entre les points de la ligne centrale de la prédiction et la référence sont mesurées puis celle entre l'extrémité du cathéter et de la prédiction. L'erreur médiane par rapport à la ligne centrale du cathéter est de 0,2 mm et de 0,9 mm par rapport à l'extrémité du cathéter. Dans notre cas, cet algorithme n'est pas utilisable directement, car nous ne disposons pas de séquences de quatre images à la suite contenant le fil guide.

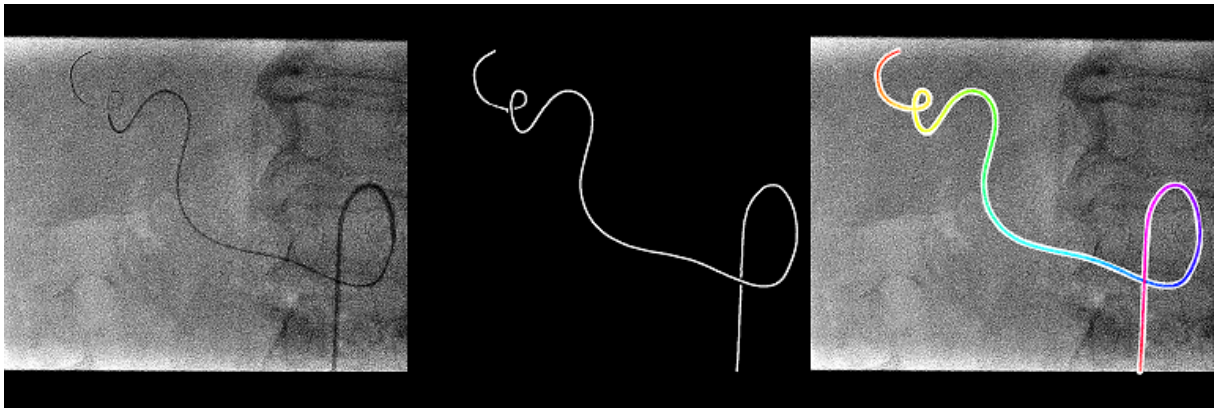


Figure 2.7 – Exemple d'image de fluoroscopie provenant du github associé à [Ambrosini et al., 2017] : `https://github.com/pambros/CNN-2D-X-Ray-Catheter-Detection`.

Wu et al. [2018] conçoivent une méthode pour segmenter l'extrémité du fil guide sur des images de fluoroscopie afin d'améliorer la navigation pour l'angioplastie coronaire. La méthode est composée de deux étapes. Une première étape consiste à détecter une boîte englobante dans l'image de fluoroscopie. Une deuxième étape segmente l'extrémité du fil guide dans la boîte englobante issue de la détection. L'étape de détection permet de réduire les faux positifs engendrés par les éléments du fond qui ressemblent au fil guide comme des électrodes ou les vertèbres. Le réseau de détection est un modèle Faster R-CNN [Ren et al., 2015] avec une ossature (*backbone*) de ResNet-50 [He et al., 2016]. Quant à la segmentation, le modèle est un modèle Deeplab [Chen et al., 2018] avec une ossature Res-Net-50 [He et al., 2016]. La probabilité d'appartenance de chaque pixel au fil guide en sortie du réseau de détection est multipliée à celle obtenue en sortie du réseau de segmentation, et le tout est seuillé pour extraire la prédiction finale. Les réseaux ont été pré-entraînés respectivement sur le jeu de données COCO [Lin et al., 2014] et PASCAL VOC [Everingham et al., 2010]. De plus les auteurs utilisent une technique d'augmentation de données et une fonction de Dice couplée à l'entropie croisée comme fonction de coût. Les images de taille 1080×1080 sont issues de 22 séquences dont 15 utilisées pour l'entraînement et 7 pour l'évaluation. Les auteurs évaluent la segmentation grâce au score de Dice ainsi que l'erreur de détection de l'extrémité en pixels qui sont respectivement de 0,939 et 0,532 pixels. Deux autres mesures d'évaluation sont introduites, le taux de fausses détections et le taux de détections manquantes, respectivement de 0,8% et de 9,9% . Le temps d'inférence est de 4 à 5 images par seconde.

En 2019, Subramanian et al. [2019] utilisent également un U-Net couplé à une information a priori spatiale sur plus de 10 000 radiographies de thorax pour détecter le cathéter veineux central et classer les images selon leur type. Deux réseaux U-Net segmentent respectivement le cathéter et les éléments anatomiques du thorax comme les poumons, les clavicules, le cœur, etc. Les auteurs utilisent ces informations ainsi que l'information a priori spatiale pour extraire des caractéristiques qui seront utilisées pour la classifica-

tion avec l'algorithme des forêts aléatoires. L'information a priori spatiale est obtenue en moyennant les annotations des cathéters par classe pour obtenir une carte de distribution spatiale (figure 2.8). Les auteurs obtiennent une mesure d'exactitude (*accuracy*) de 85,2% et une précision de 91,6% pour la tâche de segmentation des cathéters.

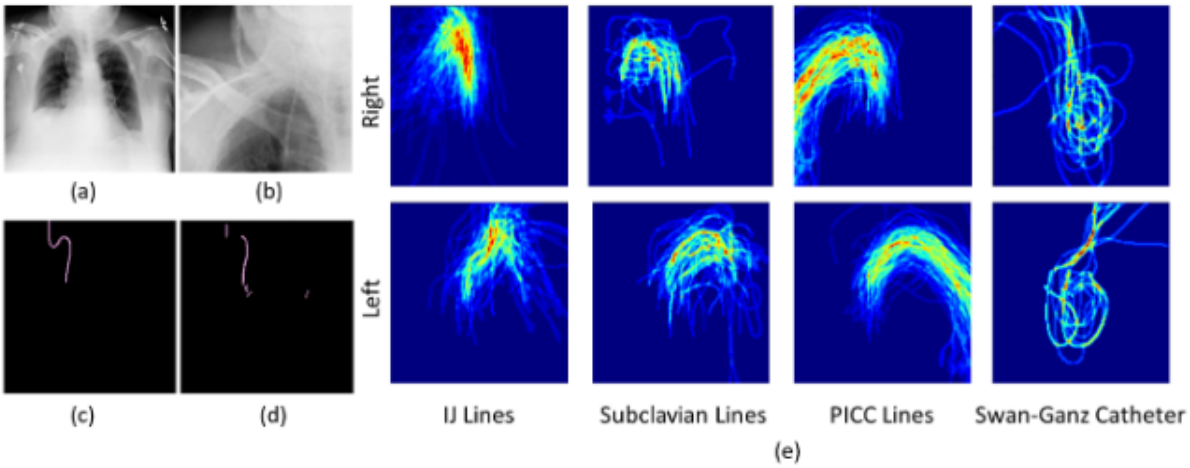


FIGURE 2.8 – Image tirée de [Subramanian et al., 2019]. Les images correspondent à : (a) une radiographie de thorax avec un cathéter, (b) un zoom sur la région du cathéter dans l'image, (c) l'annotation manuelle du cathéter sur l'image (a), et (d) la prédiction du réseau U-Net. Enfin, l'image (e) représente les superpositions pixel à pixel des annotations manuelles pour chaque type de cathéter, utilisées pour la construction des informations a priori spatiales.

En 2020, Nguyen et al. [2020] introduisent une méthode d'apprentissage profond pour les interventions endovasculaires en temps réel, appelée FW-Net (*Flow-guided Warping Net*). Cette méthode de bout en bout contient trois modules, un module de segmentation de type encodeur-décodeur, un module de calcul du flot optique et une fonction de déformation, liant les deux pour l'apprentissage de la continuité temporelle. Les auteurs utilisent des images issues d'une simulation d'intervention. Ils entraînent le réseau sur une base de données conséquente de 24 000 images et le testent sur 1000 images. En ce qui concerne le module de segmentation, la fonction de coût utilisée est une entropie croisée pondérée. Le module du flot optique prend la même image d'entrée que le module de segmentation et l'image d'après pour en extraire les caractéristiques de mouvement calculées par un réseau Flow-Net [Dosovitskiy et al., 2015]. Enfin la fonction de distorsion (*warping*) lie le résultat des deux modules précédents pour prédire la segmentation du cathéter au temps $t + 1$. Les auteurs obtiennent un score de Dice de 0,821 pour la segmentation de cathéter.

Gherardini et al. [2020] s'intéressent également à la segmentation de cathéter en fluoroscopie. Pour cela, ils proposent de tirer parti du transfert d'apprentissage grâce à des données synthétiques et un réseau U-Net. Les auteurs modifient légèrement la structure du réseau U-Net pour le rendre moins lourd, en ôtant la deuxième convolution à chaque niveau, réduisant la taille d'entrée de 1024×1024 à 256×256. La méthode est ensuite évaluée sur trois bases de données, une base contenant 9000 images synthétiques, une base de 2000 images de fantômes et enfin une de procédures *in-vivo* contenant 1207 images. Les images synthétiques sont générées à partir d'arrière plan des images *in-vivo* et de cathéter générés automatiquement auxquels des transformations ont été appliquées. Le réseau est ensuite entraîné sur les bases synthétiques et fantômes séparément et ajusté (*fine-tuned*) sur trois sous-ensembles de la base d'images *in-vivo*. De plus, deux fonctions de coût sont

testées séparément, la fonction de coût de Dice et l'entropie croisée. La première fonction de coût obtient des meilleures performances que la deuxième pour toutes les expériences menées. Le transfert d'apprentissage permet d'obtenir les scores de Dice suivants pour chaque sous-ensemble de la base de donnée d'images *in-vivo* : 0,55, 0,26, 0,17. Ces résultats, qui traduisent bien la difficulté qu'il y a à segmenter le fil guide dans ce type d'images, sont proches de ceux obtenus avec la méthode d'Ambrosini et al. [2017] sur les mêmes données mais bénéficient d'une réduction de temps d'inférence grâce à l'allègement du nombre de paramètres du réseau.

En 2021, Zhou et al. [2021] développent FAR-Net. Cette méthode s'appuie sur une architecture d'auto-encodeur pour la segmentation, couplée à l'algorithme de détection de coins de Shi and Tomasi [1994] pour la détection de l'extrémité du fil guide. Le modèle a été entraîné sur 1494 images et testé sur 360 images. Le modèle atteint un score de Dice de 0,938 et une erreur moyenne de 0,596 pixels.

Une autre approche récente [Tang et al., 2023] consiste en une segmentation de l'extrémité (point) du cathéter grâce à un réseau « léger » et un mécanisme d'auto-distillation couplé à un post-traitement par filtrage bayésien qui corrige la segmentation grâce aux informations de la séquence. L'architecture utilisée est celle d'un encodeur-décodeur composé de quatre unités d'encodage de Mobile-NetV3 [Howard et al., 2019] et une unité de décodage. Le mécanisme d'auto-distillation repose sur la fusion des canaux de chaque module d'encodage. L'entraînement se fait grâce à deux fonctions de coût combinées, la fonction de Dice pour la segmentation et la norme L_2 pour le mécanisme d'auto-distillation. Une étape de post-traitement vise à réduire les fausses détections dues au manque de contexte en appliquant un filtrage bayésien sur la séquence d'images pour ajouter une dimension temporelle à la détection. Cette étape s'appuie notamment sur le calcul du flot optique. En tout 173 séquences d'angiogrammes, contenant 9884 images de taille 512 en niveaux de gris sont utilisées. Les images sont divisées manuellement en trois ensembles, d'entraînement, de validation et de test, comportant respectivement 7885, 1072 et 927 images. Les auteurs comparent leur réseau avec cinq réseaux de segmentation et trois décodeurs. Leur architecture a pour avantage d'être légère et par conséquent adaptée au matériel informatique des hôpitaux. L'évaluation du réseau se fait avec le score de Dice qui est de 77,65% pour 1,2M de paramètres et un temps d'inférence de 8,72ms. Une deuxième méthode d'évaluation est le taux de fausses détections qui est calculé sur l'image et sur la séquence. La méthode proposée permet d'obtenir un taux de fausses détections de 2,36 pour les images et de 2,08 pour les séquences.

Auteurs	Méthode	Intervention	Tâche	Données	Résultats
Ambrosini et al. [2017]	U-Net + Connexion des branches	Cathétérisme hépatique	Segmentation totale du cathéter	182/55 séquences	$E_{cl} = 0,2mm$ $E_{ext} = 0,9mm$
Wu et al. [2018]	Cascade Faster-RCNN + DeepLab	Angioplastie coronaire	Segmentation de l'extrémité du cathéter	15/7 séquences	$Dice = 0,939$ $E_{ext} = 0,532$px
[Subramanian et al., 2019]	U-Net + a priori de forme	Vérification de la position du cathéter veineux central	Détection et classification de la position du cathéter	> 10000	$Acc = 85,2\%$ $Prec = 91,6$
[Nguyen et al., 2020]	Encodeur-Décodeur + Flot optique	Simulation de cathétérisation aortique	Segmentation et suivi de cathéter	24000/1000	$Dice = 0,821$
[Gherardini et al., 2020]	U-Net + Transfert d'apprentissage	Simulation et Implantation percutanée de Valve aortique	Segmentation totale du cathéter	9000/2000/1207	Dice 0,55 ; 0,26 ; 0,17
[Zhou et al., 2021]	FAR-Net + Détection de coins	Anévrismes et Angioplastie coronaire	Segmentation de l'extrémité du fil guide	1494/360	$Dice = 0,938$ $E_{moy} = 0,596px$
[Tang et al., 2023]	Réseau léger + Filtrage Bayesien	Angioplastie coronaire	Suivi extrémité	8957/927	$Dice = 0,78$ $TFD = 2,36$

Table 2.1 – Tableau récapitulatif des méthodes de l'état de l'art en segmentation et détection de cathéters pour les interventions dans le domaine endovasculaire. E_{cl} : Erreur par rapport aux lignes centrales, E_{ext} : Erreur par rapport à l'extrémité, Acc : Exactitude, Prec : Précision et TFD : le taux de fausses détection défini dans Tang et al. [2023].

2.4 Discussion et conclusion

L'état de l'art en segmentation ou détection d'instruments tels que les cathéters ou les fils guides sur des images fluoroscopiques ou radiographiques s'appuie principalement sur la méthode U-Net couplée à une seconde méthode [Ambrosini et al., 2017, Subramanian et al., 2019, Gherardini et al., 2020, Nguyen et al., 2020]. D'autres approches, comme celle de Zhou et al. [2021], développent elles aussi des nouvelles architectures, qui restent largement inspirées de celle de U-Net mais avec des modifications architecturales complexes. Ainsi, nous en déduisons qu'un simple réseau U-Net ne peut segmenter à lui seul des instruments aussi fins et peu présents dans l'image.

Nous remarquons également une diversité dans l'objectif à atteindre en fonction des besoins des interventions. Certains se concentrent sur la segmentation d'une ou plusieurs parties, d'autres sur la détection de l'extrémité. La diversité dans les tâches entraîne une diversité des formats et du nombre des données d'entrée, et également de l'évaluation de ces tâches. Ainsi il est difficile de comparer toutes ces méthodes entre elles. Les cathéters en imagerie endovasculaire ont généralement un diamètre plus élevé que ceux utilisés en CPRE, ce qui rend ces méthodes peu applicables directement aux problèmes traités dans cette thèse. Elles restent néanmoins une source d'inspiration.

Ce chapitre nous montre qu'il existe un intérêt au suivi d'instruments par ordinateur lors d'interventions endoscopiques pour faciliter la navigation endoluminale et ainsi limiter le risque pour les patients. Il montre également qu'il n'y a pas de méthodes de suivi adaptées pour la CPRE. Comme le montre la multitude d'articles publiés autour de la segmentation de cathéters, synthétisée dans la table 2.1, cette étape d'identification des cathéters est nécessaire pour un futur recalage 2D/3D et le développement d'une aide à la localisation efficace des instruments en per-opératoire. La section 2.3.2 montre qu'un réseau U-Net simple ne suffit pas à segmenter le fil guide sans y ajouter d'autres méthodes. De même, MedSAM appliqué en inférence ne réussit pas à segmenter le fil guide. Les méthodes générales telles que nnU-Net sont-elles capables de bonnes performances sur des structures si fines ?

Dans la suite de ce manuscrit nous allons appliquer nnU-Net à nos bases de données pour la segmentation du fil guide. Ce résultat constituera notre référence. Nous comparerons ensuite les résultats obtenus avec un réseau U-Net et différentes fonctions de coût.

CHAPITRE 3

Construction de la base de données

Un élément essentiel en apprentissage automatique supervisé est la qualité des données en entrée, qui sont des images et des annotations. De plus, les algorithmes d'apprentissage statistique ont besoin d'une quantité et d'une variété importante de données pour être capables de généraliser, c'est-à-dire pour que le modèle prédise un résultat avec la même fiabilité sur des données qui n'auront pas été vues lors de l'apprentissage. Cette généralisation se fait donc mieux lorsqu'il y a beaucoup de données et qu'elles représentent la diversité de la population. Or dans notre contexte, nous disposons de peu de données exploitables (table 3.2), et il faudra en tenir compte lors du développement des méthodes et de la mise en place des algorithmes. L'annotation de ces données prend également du temps. Ce chapitre décrit la collecte de données, la construction de deux bases de données provenant de centres différents, l'étape de pré-traitement et d'annotation de ces dernières.

3.1 Collecte des données

Dans ce travail, nous avons utilisé des images de fluoroscopie 2D, issues de l'intervention de CPRE, provenant de deux centres experts d'endoscopie digestive des hôpitaux Saint-Antoine (SA) à Paris et Henri-Mondor (HM) à Créteil. Les images de cinquante patients de l'hôpital Saint-Antoine ont été acquises par un bras radioscopique Philips BV Pulsera puis stockées au format *bitmap* (BMP), encodées sur 8 bits. Ainsi les images sont en niveaux de gris, dans l'intervalle $[0, 255]$. Les images de quinze patients de l'hôpital Henri-Mondor ont été acquises à l'aide d'un bras radioscopique de marque Siemens CIOS α et enregistrées au format DICOM. Ces données ont été collectées et utilisées selon les protocoles soumis et validés par le comité d'éthique de Sorbonne Université le 30 janvier 2023 (No. CER-2023). Les formats, noms et tailles des images n'étant pas homogènes, la première étape a été de créer une base de données cohérente permettant de réaliser des expériences.

3.2 Création de la base de données

Pour chaque patient, les endoscopistes acquièrent les mêmes étapes clés de l'intervention. Ainsi pour un patient il y aura généralement au moins une image pour chaque prise de vue qui suit (figure 3.1 et en annexe A.1) :

- 0 : la coupole (dôme entre le foie et les poumons),
- 1 : l'hypochondre droit,
- 2 : l'endoscope,
- 3 : l'endoscope et le fil guide,
- 4 : l'endoscope, le fil guide et les voies biliaires opacifiées.

3.2.1 Nomenclature des images

Une première contribution a été d'organiser les images brutes en base de données. La première étape a consisté à renommer les images avec un système commun qui ait du sens. Ainsi, chaque image a été renommée selon le code suivant :

- Numéro de la base de données (00 pour Saint-Antoine et 01 pour Henri-Mondor)
- Numéro du patient (qui est arbitraire et ne correspond pas à une identification de l'hôpital)
- Clé d'identification de l'étape allant de 0 à 4
- Numéro de l'image (car il peut y avoir plusieurs images d'une même étape).

Par conséquent le troisième cholangiogramme provenant du patient numéro 26 de SA avec seulement l'endoscope visible aura pour dénomination "0026203.bmp". Toutes les images ne sont pas exploitables pour la tâche de segmentation des instruments. Le code d'identification sert à faciliter le tri des images par type pour sélectionner seulement celles utiles pour la tâche requise. Le numéro de patient est également un élément important à garder pour la partition en différents ensembles lors de l'entraînement pour pouvoir séparer facilement les patients sans qu'un patient ne se trouve dans deux ensembles en même temps.

3.2.2 Pré-traitement

Les données brutes ont été pré-traitées pour harmoniser la base de données. Certaines différences d'acquisition entre les deux centres ont conduit à des pré-traitements spécifiques pour chaque source d'images.

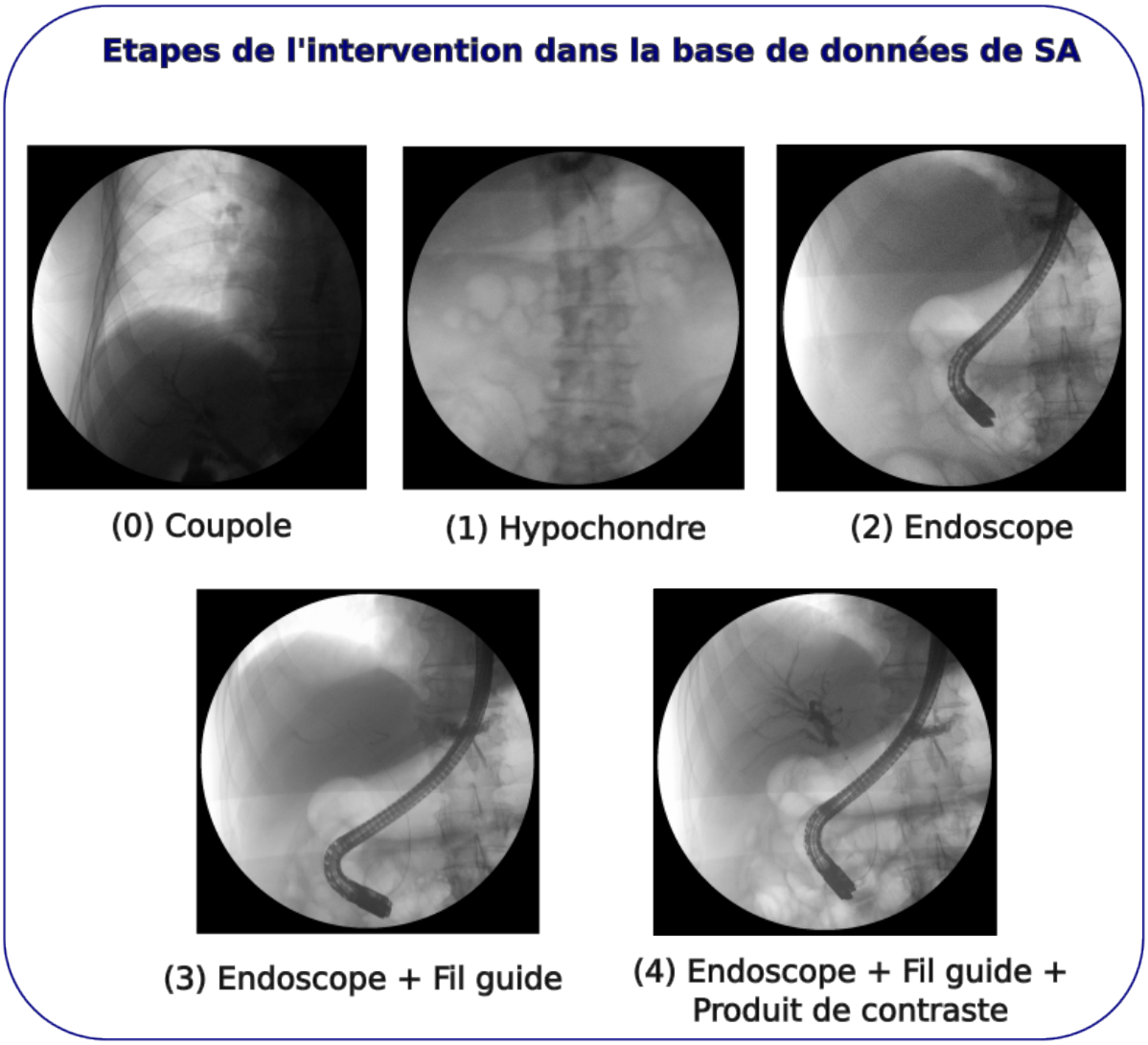


FIGURE 3.1 – Illustration des différentes étapes jalonnant l'intervention dans la base de données de l'hôpital Saint-Antoine.

Saint-Antoine (SA)

Plusieurs irrégularités sur les images justifient la nécessité d'un pré-traitement. Les données brutes n'avaient pas toutes la même taille. Par ailleurs des éléments non désirables comme un drapeau, l'heure et une bande de couleur étaient présents sur l'image (figure 3.2). Pour remédier à ces problèmes, un pré-traitement est appliqué aux cholangiogrammes. Les images sont traitées afin qu'elles aient toutes une taille finale de 1024×1024 pixels. Premièrement, l'image est découpée à la taille 997×997 et multipliée à un disque binaire pour en extraire la partie d'intérêt. Les images avec un côté inférieur à 997 sont étoffées d'une ou plusieurs couches de 0 jusqu'à atteindre cette taille. Finalement, les images 997×997 atteignent une taille de 1024×1024 par ajout de couches successives de 0 sur les bords droit et bas. Ce choix est fait pour éviter l'interpolation qui serait appliquée par une fonction de changement de taille classique [1].

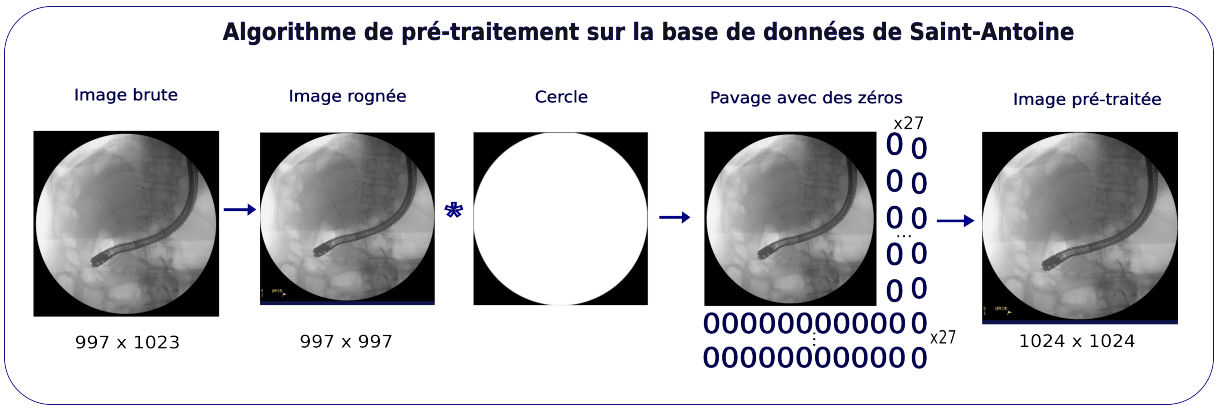


Figure 3.2 – Étape de pré-traitement sur les images de la base de données de l'hôpital Saint-Antoine. Cette étape enlève les éléments en bas à gauche et re-dimensionne toutes les images au format 1024×1024 *grâce à un pavage de zéros en bordures droite et basse de l'image.*

Henri-Mondor (HM)

Les images provenant de l'hôpital Henri-Mondor sont dans un format différent, qui est le format DICOM. Premièrement les images ont été anonymisées avec la bibliothèque Python *dicom-anonymizer*, pour effacer les méta données sensibles. Ensuite, les images ont été renommées selon le code mentionné dans la section 3.2.1. Les images natives ont une taille de 1536×1536 et elles ont été converties au format BMP et redimensionnées à la taille 1024×1024 par sous échantillonnage avec la méthode *resize* de la bibliothèque Python PILLOW. Par défaut le sous-échantillonnage se fait avec la méthode d'interpolation d'intensité des voisins.

Normalisation des intensités

Les intensités des images des deux bases de données sont normalisées entre 0 et 1 au cours de l'entraînement, juste avant l'étape d'augmentation des données dans les expériences avec U-Net (chapitre 5). Concernant nnU-Net, la normalisation appliquée est décrite dans la section 2.3.1.

1. Le code est disponible ici : https://gitlab.lip6.fr/gamartin/these/-/blob/main/preprocessing_sa.py

3.3 Caractéristiques des images

Résolution Le bras de fluoroscopie 2D étant mobile, la distance entre le patient et le récepteur peut varier d'une intervention à l'autre. Par conséquent, le champ de vue varie. Ainsi la résolution, que l'on définit ici par le nombre de pixels pour un champ de vue donné, n'est pas connue pour l'ensemble des images de la base de données. Il est tout de même possible de s'en approcher grâce au diamètre de l'endoscope qui est constant et mesure $11,3$ mm. Nous mesurons alors la distance en pixels entre deux points d'une section d'endoscope (figure 3.3) pour connaître la taille des pixels en millimètres. Nous le faisons pour dix images de chaque base de données et moyennons les résultats pour connaître l'ordre de grandeur de la résolution. Les résultats sont présentés dans le tableau 3.1. Nous remarquons que les images de la base HM ont une meilleure résolution spatiale car un pixel représente une plus petite distance en millimètres que pour la base SA.

Base de données	**Taille image (px)**	**Diamètre moyen endoscope (px)**	**Diamètre endoscope (mm)**	**Champ de vue moyen (mm)**	**Résolution moyenne (mm)**
SA	1024	50,8	11,3	232,8	0,23
HM	1536	83	11,3	209,9	0,14

Table 3.1 – Tailles en pixels et millimètres de l'endoscope et taille des images en pixels. De ces informations, nous déduisons le champ de vue qui est la taille de l'image en mm et la résolution qui représente la taille d'un pixel en mm.

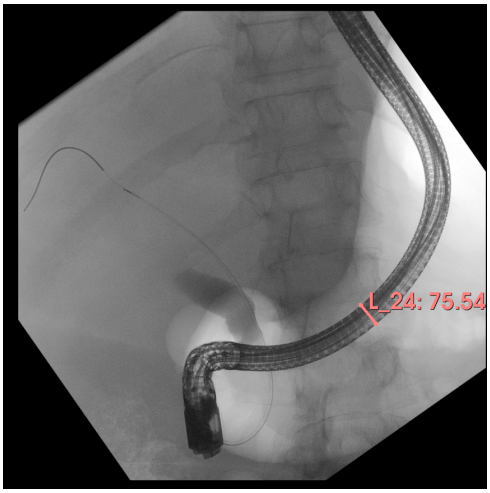
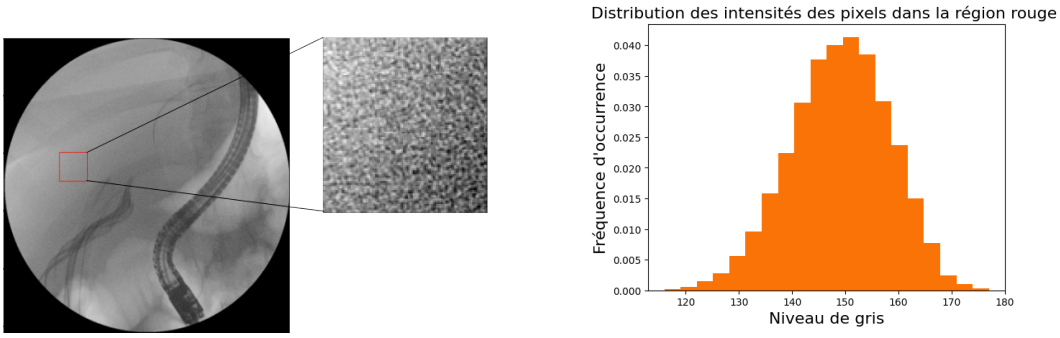


Figure 3.3 – Mesure du diamètre de l'endoscope en pixels (section rouge) sur une image de la base HM grâce au logiciel 3D Slicer.

Bruit Le bruit a été analysé en observant la distribution des intensités des pixels dans des régions homogènes de trois images de chaque base de données. Les régions homogènes sont des carrés de côté 100 pixels où l'image représente une même structure. On suppose qu'une région homogène parfaite aurait tous les pixels à la même intensité et une variance nulle. L'allure de la distribution sur la figure 3.4b ressemble à une gaussienne, ce qui signifie que la valeur des pixels s'écarte de plus ou moins σ par rapport à la moyenne. En mesurant précisément la moyenne μ et l'écart-type σ dans cette région, on obtient $\mu_{SA} = 136$ et $\sigma_{SA} = 9$. Par ailleurs le rapport signal sur bruit (SSB) est défini par :

$$SSB = \frac{\mu}{\sigma} \tag{3.1}$$

On retrouve ici un SSB_{SA} de 16. De la même façon pour la base HM on mesure $\mu_{HM} = 125$, $\sigma_{HM} = 7$ et un $SSB_{HM} = 18$. Nous concluons que les images de la base de données de Saint-Antoine sont légèrement plus bruitées que celle de la base de données Henri-Mondor.



(a) Extraction d'une région homogène dans l'image

(b) Distribution des intensités des pixels dans la région choisie

FIGURE 3.4 – *Étapes de l'analyse de bruit. Exemple avec une image de la base de données SA.*

Diversité des images et potentielles difficultés La difficulté majeure pour la tâche de segmentation du fil est la finesse de son diamètre qui vaut $0,035$ pouce ce qui équivaut à $0,889$mm. De plus, le fil guide n'est pas uniformément radio-opaque. Son extrémité est plus souple et hydrophile et de fait elle apparaît plus contrastée sur les cholangiogrammes. A contrario, le reste du fil plus rigide et hydrophobe est moins contrasté. Par conséquent, il est parfois difficile de percevoir un fil guide continu sur les images. La présence d'autres instruments tels que des électrodes, des ballons, des prothèses sont des obstacles à la bonne détection du fil guide, car ces structures sont soit plus contrastées, soit similaires. Enfin la visualisation du fil guide est parfois limitée par la présence de produit de contraste résiduel ou par certaines structures comme l'estomac qui génèrent un fort contraste ou le bruit généré par le mouvement. Des exemples sont illustrés dans la figure 3.5.

3.4 Annotation des images

Les images ont finalement été annotées manuellement grâce au logiciel libre d'annotation d'images par polygones, *label me*[2]. Ce logiciel permet de tracer des polygones et d'extraire les masques des objets détourés au format PNG. Ainsi les masques contenant l'endoscope et le fil guide ont été extraits. Le détourage de l'endoscope est plus aisé car il a un diamètre plus élevé. A contrario, il est plus difficile de détourer le fil guide qui est très fin. De plus, le fait de devoir tracer un polygone et non pas une rangée de points n'est pas bien adapté et empêche de maintenir un diamètre constant (figure 3.6). Malgré cette difficulté d'annotation, l'outil utilisé était le plus adapté parmi ceux envisagés. Une dilatation par un élément structurant diamant de rayon de taille trois pour SA (respectivement un

2. https://github.com/wkentaro/labelme

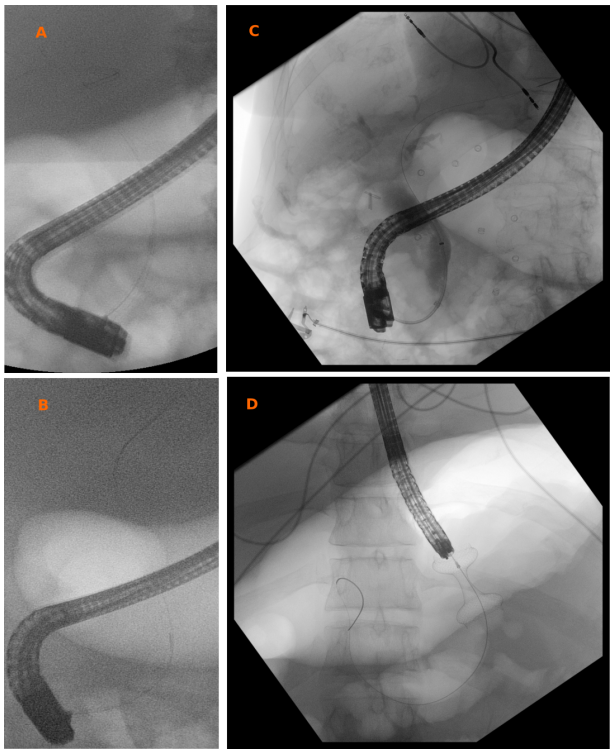


FIGURE 3.5 – Exemples d'images plus difficiles à segmenter. A : Bruit dû au mouvement, B : L'intensité de l'estomac masque le fil guide, C : Présence d'électrodes et produit de contraste, D : Présence d'électrodes et d'une prothèse.

pour HM) a été appliquée au fil guide pour corriger ce défaut d'annotation [3]. Cependant quelque discontinuités demeurent. En outre, le fil guide n'a pas été interpolé visuellement lorsqu'il n'était pas visible mais qu'on pouvait deviner son passage. Ces annotations ont ensuite été vérifiées et corrigées par deux endoscopistes experts.

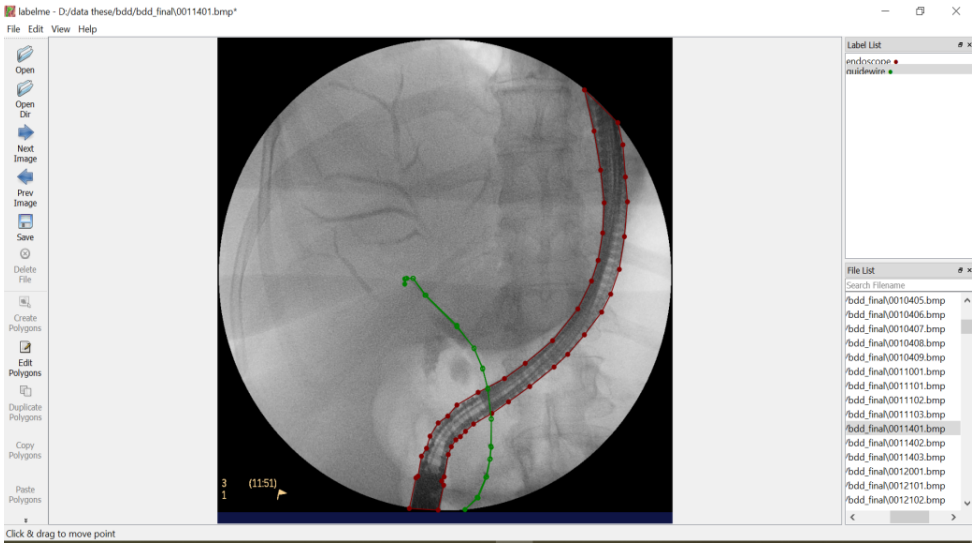


FIGURE 3.6 – Illustration de l'annotation des instruments par polygones avec le logiciel label me.

3. Le code est disponible ici : https://gitlab.lip6.fr/gamartin/these/-/blob/main/preproc_gw.py

Base de données	**Nombre de patients**	**0**	**1**	**2**	**3**	**4**	**Total**
SA	50	77	77	53	43	37	287
HM	15	2	14	22	22	54	114
Total	75	79	91	75	65	91	401

Table 3.2 – Nombre d'images après pré-traitement par base de données et par catégorie (seulement les images annotées pour les catégories 2 à 4).

3.5 Conclusion

Dans ce chapitre, nous avons décrit la construction de deux bases de données de la collecte au pré-traitement. Les images sont multicentriques, elles présentent des différences de bruit principalement mais aussi de résolution et de contraste. Le nombre d'images exploitables pour la segmentation de l'endoscope et du fil guide reste très limité comparé à la difficulté de la tâche et aux jeux de données des méthodes décrites dans l'état-de-l'art (chapitre 2). Notons cependant que l'acquisition d'images continue et augmentera avec le temps.

CHAPITRE 4

Méthodes d'évaluation de la segmentation de structures fines

Nous avons montré précédemment que la segmentation était une tâche courante en vision par ordinateur et notamment en imagerie médicale. Pour pouvoir utiliser un modèle de segmentation, et le valider, il est nécessaire de l'évaluer. Pour cela, il existe plusieurs méthodes quantitatives. Certaines méthodes sont utilisées systématiquement en segmentation d'images biomédicales, comme le score de Dice. Or Reinke et al. [2024] montrent que les méthodes d'évaluation sont souvent mal choisies car elles ne sont pas adaptées au problème à résoudre.

Dans ce chapitre nous nous interrogeons sur la pertinence des approches systématiques pour l'évaluation de notre modèle et de ses résultats. En effet, dans le cas du fil guide, plusieurs éléments semblent nécessiter une méthode d'évaluation spécifique. Il est important de prendre en compte dans l'évaluation d'une part le fort déséquilibre de classes, et d'autre part la structure fine et allongée dans l'évaluation. De surcroît la taille réduite de l'échantillon de données est également une caractéristique importante du problème. Dans le cas du fil guide, nous cherchons à obtenir une segmentation idéalement connexe. Par conséquent nous voulons choisir une méthode d'évaluation qui est capable de mesurer cette propriété.

Dans ce chapitre nous détaillons les méthodes d'évaluation usuelles et montrons qu'elles ne sont pas suffisantes pour l'évaluation de structures fines. Nous nous intéressons à des méthodes plus spécifiques, notamment celles utilisées pour la segmentation de vaisseaux. Enfin nous discutons de l'évaluation de la segmentation d'un point de vue plus clinique, plus compréhensible par les médecins.

Les indices que l'on présente peuvent être pour la plupart utilisés autant pour l'évaluation d'un modèle que pour servir de critère d'optimisation de son apprentissage. Dans la suite, nous analysons ces indices essentiellement du point de vue de l'évaluation d'un modèle en production.

4.1 Cas limites

Dans ce chapitre nous nous intéressons à plusieurs méthodes d'évaluation existantes. Pour illustrer leurs performances nous allons les tester sur des cas limites (figure 4.1) :

- une prédiction vide
- une prédiction de toute l'image
- une prédiction aléatoire
- une prédiction partielle
- une prédiction partielle avec des faux positifs

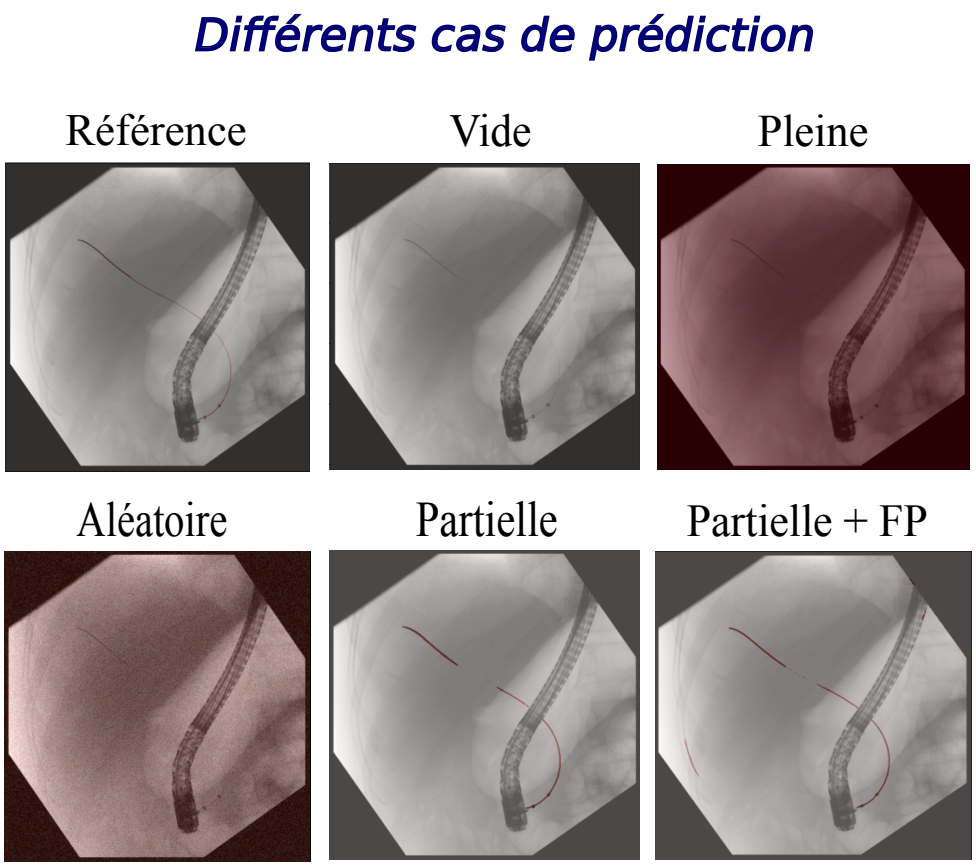


FIGURE 4.1 – Illustration de différents cas de prédictions qui seront utilisés pour montrer les limites et avantages de chaque méthode d'évaluation.

4.2 Méthodes communes pour l'évaluation de modèles en segmentation automatique

La plupart des méthodes d'évaluation citées dans cette section sont bâties à partir des éléments de la matrice de confusion (table 4.1) entre les classes prédites et les classes réelles de chaque pixel, soient : les vrais positifs (VP), les vrais négatifs (VN), les faux positifs (FP) et les faux négatifs (FN) (figure 4.2). Ces méthodes d'évaluation sont très utilisées en classification mais peuvent également être utilisées en segmentation car la segmentation peut être vue comme une classification binaire au niveau du pixel. Cette section regroupe une liste de critères d'évaluation détaillés par Müller et al. [2022] à propos des méthodes communes d'évaluation en segmentation d'images médicales.

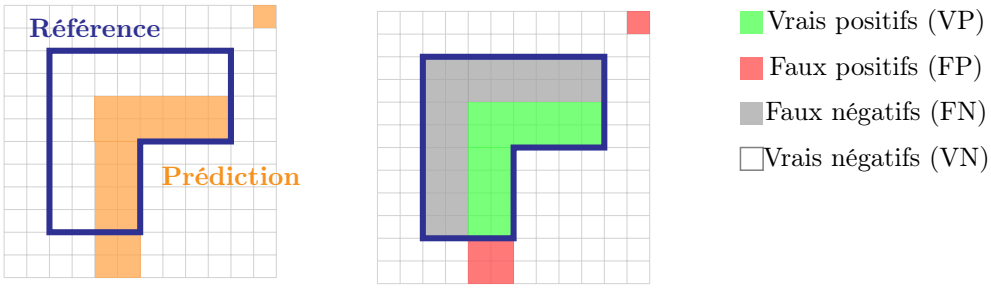


FIGURE 4.2 – Illustration des différentes classifications de pixels prédits en fonction de la référence.

		Prédiction	
		Vrai	Faux
Référence	Vrai	VP = 20	FN = 28
	Faux	FP = 5	VN = 91

TABLE 4.1 – Matrice de confusion correspondant à la figure 4.2.

La justesse (*Accuracy*)

La justesse (ou *Accuracy*) représente le nombre de prédictions correctes, positives ou négatives, soit le nombre de pixels bien classés sur le nombre total de pixels :

$$Accuracy = \frac{VP + VN}{VP + VN + FN + FP} \tag{4.1}$$

Elle n'est pas adaptée dans le cas d'un déséquilibre de classes car elle prend en compte les vrais négatifs. Ainsi, dans le cas du fil guide, leur nombre sera obligatoirement d'un ordre de grandeur disproportionné par rapport à celui des vrais positifs et des faux négatifs. Par conséquent, on peut avoir un bon score de justesse en trouvant peu de vrais positifs. Par exemple dans le cas de la prédiction vide, nous obtenons $0,99$. Comme les pixels du fil guide représentent un faible pourcentage de l'image la justesse n'est pratiquement pas affectée lorsqu'il n'est pas détecté. Elle peut ainsi être augmentée artificiellement en agrandissant le support de l'image. Par conséquent, elle n'est pas une bonne candidate pour l'évaluation de la segmentation en général et dans notre cas d'étude en particulier.

La sensibilité ou le rappel

Elle représente le taux de vrais positifs, ou le taux d'éléments pertinents :

$$Sensibilité = Rappel = \frac{VP}{VP + FN} \tag{4.2}$$

Dans notre cas il s'agirait du nombre de pixels du fil guide bien prédits sur le nombre total de pixels du fil guide. Par exemple dans le cas de la segmentation pleine, nous avons une sensibilité de $1,0$, ce qui indique qu'on a bien détecté tous les pixels du fil guide. Cependant avec ce critère nous n'avons pas d'information sur les faux positifs, c'est-à-dire les pixels de l'environnement classés par erreur comme appartenant au fil guide, lesquels sont très nombreux dans le cas de cet exemple.

La précision

La précision quantifie le nombre de vrais positifs sur l'ensemble des éléments positifs :

$$\textit{Précision} = \frac{VP}{VP + FP} \tag{4.3}$$

Dans notre cas d'étude il s'agirait du nombre de pixels du fil guide bien prédits sur le nombre de pixels prédits positivement comme appartenant au fil guide. Plus la précision est élevée et plus cela signifie que les pixels prédits sont effectivement du fil guide et pas des faux positifs. Pour les trois premiers cas limites, elle vaut 0. Par ailleurs, la segmentation partielle obtient une meilleure précision que celle avec faux positifs ($0,73 > 0,64$).

La précision et le rappel sont deux valeurs qui se complètent et il faut que les deux soient élevées pour que la segmentation soit considérée correcte. Elles sont utilisées ensemble dans le score de Dice (voir section 4.3.1).

La spécificité

Elle représente le taux de vrais négatifs :

$$\textit{Spécificité} = \frac{VN}{VN + FP} \tag{4.4}$$

Dans le cas déséquilibré, cette valeur n'est pas intéressante car elle concerne les vrais négatifs et tendra vers la valeur 1 systématiquement.

4.3 Méthodes communes pour l'évaluation de segmentations d'images médicales

En segmentation d'images biomédicales, certaines évaluations sont systématiques et permettent de comparer les résultats à ceux de l'état de l'art, lorsque ceux-ci ont été obtenus sur les mêmes jeux de données lorsque c'est possible. Ces critères d'évaluation sont pour la plupart dérivés des méthodes vues dans la section précédente.

4.3.1 Score de Dice

L'indice de similarité de Dice, de Sørensen-Dice ou encore F_1-score est une moyenne harmonique entre la précision et le rappel. Il mesure le chevauchement entre la prédiction P et la référence R. Il est défini de la manière suivante :

$$DSC = 2 \cdot \frac{\textit{Précision} \cdot \textit{Rappel}}{\textit{Précision} + \textit{Rappel}} = \frac{2 \cdot VP}{2 \cdot VP + FP + FN} = \frac{2|P \cap R|}{|P| + |R|} \tag{4.5}$$

Discussion Il s'agit de la méthode d'évaluation la plus utilisée en segmentation. Cependant, le score de Dice présente certaines limites. Premièrement, comme illustré sur la figure 4.3, il est sensible à la taille des structures. Plus les structures à segmenter sont petites et plus une erreur de segmentation se ressentira dans le score [Reinke et al., 2024, Zhao et al., 2020]. De plus, le score de Dice ne prend pas en compte la forme de l'objet à

segmenter (figure 4.4). En particulier pour une structure fine et allongée, une segmentation légèrement décalée par rapport à la structure par exemple sera très mal évaluée.

Sur les exemples d'avant, nous obtenons des scores de Dice de 0 pour les segmentations vide, totale et aléatoire. Dans le cas de la segmentation partielle, le score de Dice est plus élevé que dans celui de la segmentation partielle avec des FP, 78% contre 73%, ce qui fait sens car les FP sont au dénominateur et font donc baisser le score de Dice.

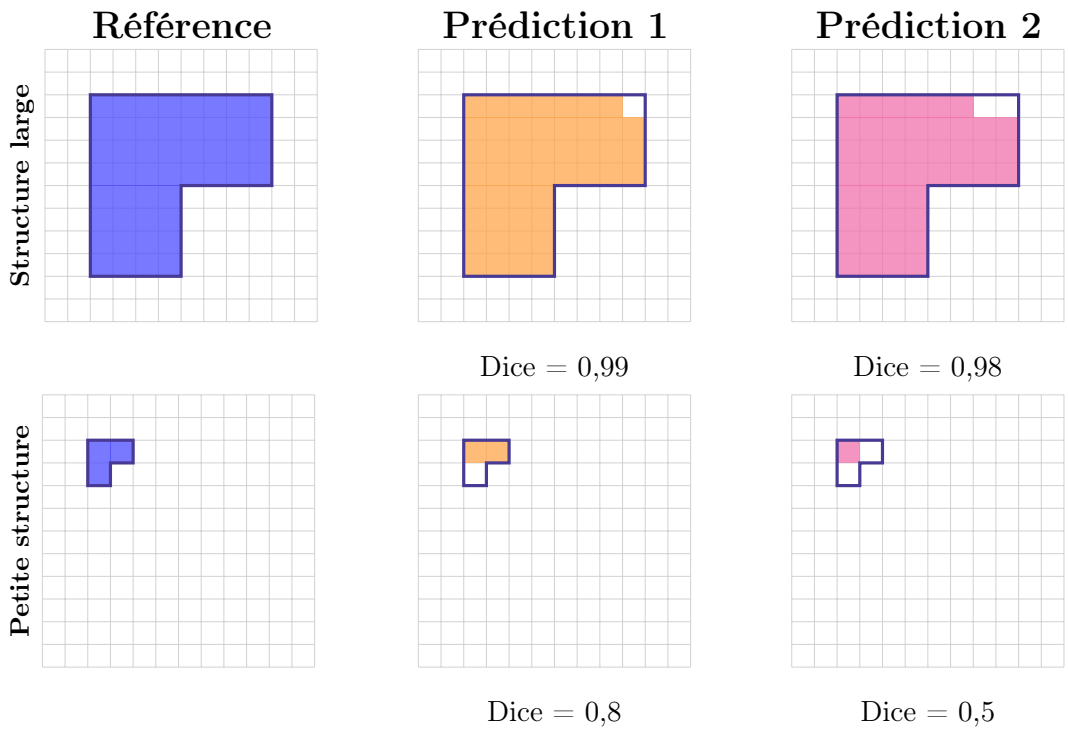


FIGURE 4.3 – Sur cette figure, nous comparons le score de Dice entre deux prédictions. La prédiction 1 omet un pixel et la prédiction 2, deux pixels. On observe que si la structure de référence est assez large, le score de Dice varie peu entre les deux prédictions. Au contraire, dans le cas où la structure à segmenter est petite (trois pixels ici), le score de Dice varie énormément entre les deux prédictions. Schéma inspiré de [Reinke et al., 2024].

Le score de Dice est en fait un cas particulier du F_β-score où β vaut 1 et qui s'écrit :

$$F_\beta = (1+\beta^2)\frac{\textit{Précision} \times \textit{Rappel}}{(\beta^2 \times \textit{Précision}) + \textit{Rappel}} \tag{4.6}$$

avec $\beta \in \mathbb{R}^+$.

Un autre cas particulier est **l'intersection sur l'union (IoU)** ou indice de Jaccard.

$$IoU = \frac{VP}{VP + FP + FN} = \frac{|P \cap R|}{|P \cup R|} \tag{4.7}$$

NB : Il existe une relation fonctionnelle entre l'IoU et le score de Dice, ce qui fait que ces deux mesures sont redondantes.

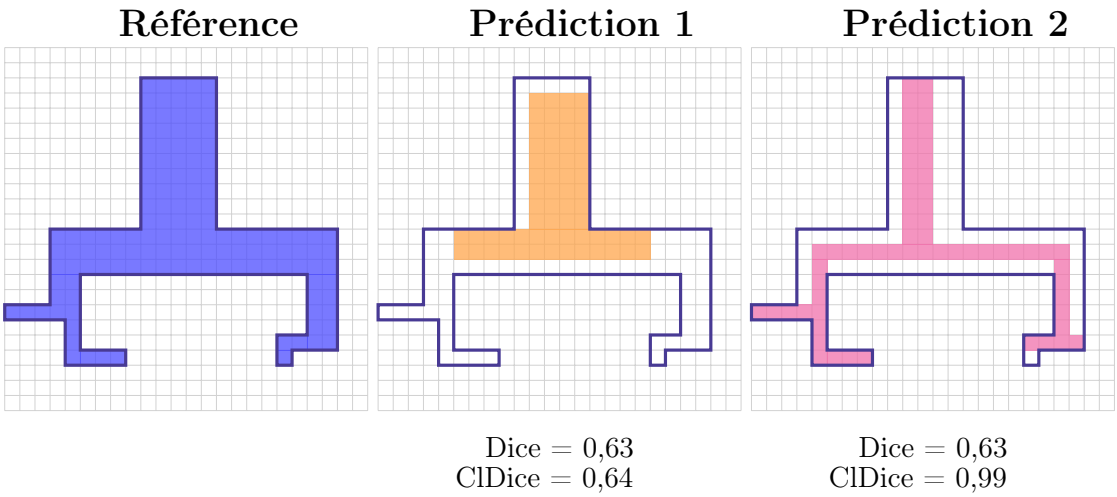


FIGURE 4.4 – Sur cette figure, nous comparons le score de Dice obtenu entre deux prédictions d'un objet ayant une structure complexe. La prédiction 1 est constituée de 62 pixels qui sont des VP mais dont la structure est différente de la référence car elle prédit plutôt les parties plus larges. La prédiction 2 est constituée de 62 pixels également mais répartis le long de chaque ramification par rapport à la structure de référence. Ainsi, les deux prédictions vont avoir le même score de Dice, mais le ClDice de la seconde prédiction sera nettement meilleur. Schéma inspiré de [Reinke et al., 2024].

4.3.2 Distances

Les distances entre deux ensembles sont également des indicateurs de la qualité de la segmentation.

La distance de Hausdorff

La distance de Hausdorff permet de comparer deux ensembles fermés, bornés notés comme précédemment P pour la prédiction et R pour la référence (notons que P et R sont discrets dans ce cas). On définit d'abord $h(P,R)$ et $h(R,P)$ (figure 4.5) par :

$$\begin{cases} h(P,R) &= \max\limits_{p\in P}\left\{\min\limits_{r\in R}\{d(p,r)\}\right\} \\ h(R,P) &= \max\limits_{r\in R}\left\{\min\limits_{p\in P}\{d(p,r)\}\right\} \end{cases} \tag{4.8}$$

avec d la distance euclidienne.

Cette fonction n'est pas symétrique, et la distance de Hausdorff est définie comme le maximum entre les deux :

$$HD(R,P) = \max(h(P,R), h(R,P)) \tag{4.9}$$

De plus elle est sensible aux données aberrantes, il est donc fréquent de ne prendre que le 95e centile pour éliminer cet inconvénient.

Dans le cadre des exemples nous obtenons des valeurs de distance de Hausdorff de 1084 pixels pour la segmentation vide, 667 pixels pour les segmentations totale et aléatoire, 70 pixels pour la segmentation partielle sans FP et 470 pixels pour celle avec des FP. Ici

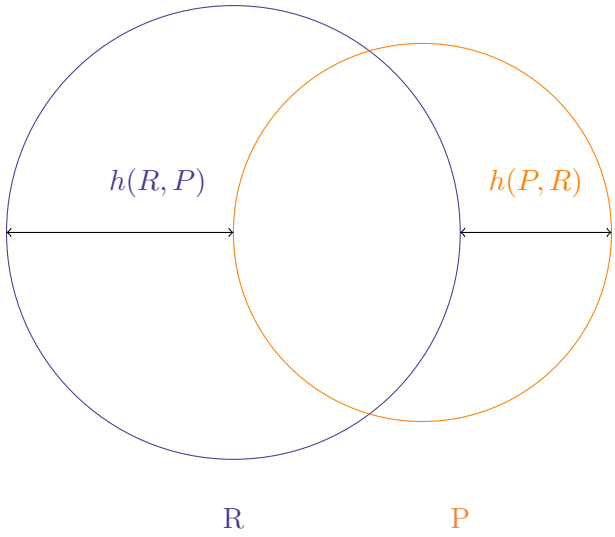


FIGURE 4.5 – Illustration de la fonction h en fonction de deux ensembles P et R. La distance de Hausdorff est le maximum entre ces deux quantités.

aussi la méthode d'évaluation pénalise les FP surtout s'ils sont éloignés de la référence comme sur la segmentation partielle avec FP.

La distance moyenne

La distance moyenne permet d'avoir une indication plus globale sur l'erreur de segmentation.

$$d_{moy}(P, R) = \frac{1}{N} \sum_{p \in P} \min_{r \in R} d(p, r) \tag{4.10}$$

avec $d(p, r)$ une fonction de distance comme la distance euclidienne.

NB : Notons que le terme distance ici est un abus de langage car cette quantité n'est pas une distance au sens mathématique, dans le sens où elle ne vérifie pas les conditions de séparation ni d'inégalité triangulaire.

4.3.3 Courbes

La **courbe ROC** (Receiver Operating Characteristic) évalue la performance d'un classifieur pour différents seuils de réglages. On porte en abscisse le taux de faux positifs et en ordonnée le taux de vrais positifs en fonction du seuil. En pratique on mesure l'aire sous la courbe (AUC) qui se calcule de la manière suivante :

$$AUC = 1 - \frac{1}{2}\left(\frac{FP}{FP + VN} + \frac{FN}{FN + VP}\right) \tag{4.11}$$

Une aire sous la courbe de $0,5$ indique un classifieur aléatoire. Un bon classifieur a une aire sous la courbe proche de 1.

4.3.4 Limites des méthodes usuelles

Les méthodes d'évaluation décrites jusque là sont utilisées fréquemment pour quantifier la performance des algorithmes de segmentation et notamment en imagerie médicale. Cependant il existe un sous-problème qui est la segmentation de structures fines pour lequel ces méthodes sont inadaptées. En effet plusieurs aspects peuvent poser problème comme le déséquilibre de classe déjà évoqué précédemment car très fréquent en imagerie médicale. Ce problème peut être atténué en ne prenant pas en compte les vrais négatifs dans le calcul de l'évaluation. Un autre problème est celui de la topologie de l'objet à segmenter dont on veut qu'elle soit maintenue. Aucune des méthodes présentées ne permet de garantir que la topologie reste la même entre la référence et la prédiction. Or cette information est essentielle notamment dans le cas du fil guide où l'on veut que la topologie soit préservée.

4.4 Méthodes spécifiques pour l'évaluation du structures fines en segmentation d'images médicales

4.4.1 Mesures de chevauchement

Centerline Dice ou ClDice

Le *ClDice* a été introduit par Shit et al. [2021] pour évaluer la segmentation de structures tubulaires en réseau comme les vaisseaux sanguins. Il prend en compte la ligne centrale de la structure à segmenter, c'est-à-dire son squelette. Le principe est illustré figure 4.6.

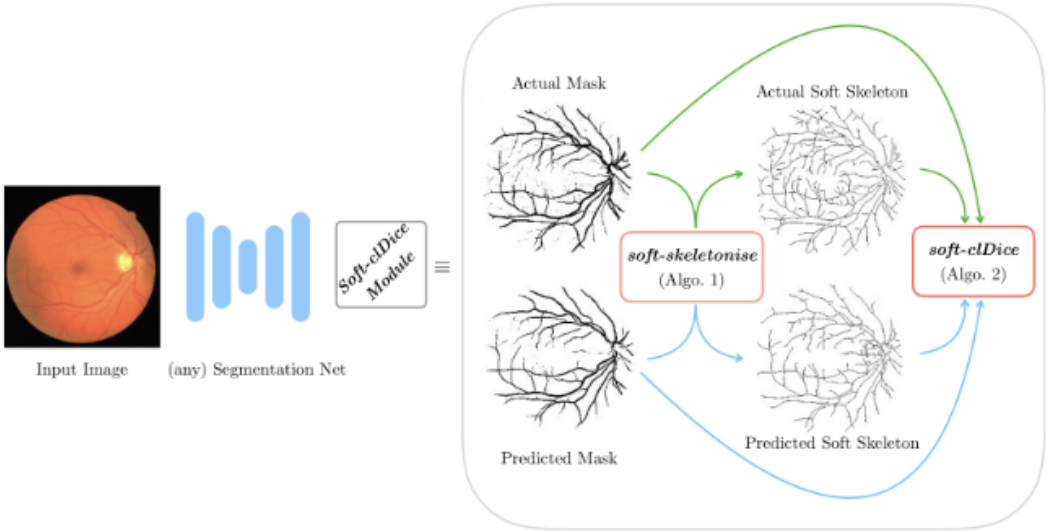


FIGURE 4.6 – Calcul du ClDice. Les squelettes S_R et S_V sont extraits à partir des masques de référence V_R et prédit V_P respectivement. Le ClDice est calculé à partir de ces quatre images. Source : Shit et al. [2021]

Dans l'article, deux nouvelles mesures sont définies, la précision topologique (T_{prec})

et la sensibilité topologique (T_{sens}) qui prennent en compte le squelette. Le ClDice est la moyenne harmonique de ces deux nouvelles valeurs.

Les auteurs introduisent les notations suivantes, V_R est le masque binaire de référence et V_P celui prédit. De même, S_R est le squelette de V_R et S_P celui de V_P (figure 4.7).

La précision topologique est définie de la manière suivante :

$$T_{prec}(S_P, V_R) = \frac{|S_P \cap V_R|}{|S_P|} \tag{4.12}$$

Elle est sensible aux faux positifs. Dans l'exemple de la figure 4.7, $T_{prec}(S_P, V_R)$ vaut 1 car le squelette de la prédiction est compris dans la référence.

La sensibilité topologique est définie de façon similaire :

$$T_{sens}(S_R, V_P) = \frac{|S_R \cap V_P|}{|S_R|} \tag{4.13}$$

Elle est sensible aux faux négatifs. Toujours dans l'exemple de la figure 4.7, $T_{sens}(S_R, V_P) = \frac{24}{51} = 0,47$.

Enfin le ClDice est défini par la moyenne harmonique de ces deux quantités :

$$clDice(V_R, V_P) = 2 \times \frac{T_{prec}(S_P, V_R) \times T_{sens}(S_R, V_P)}{T_{prec}(S_P, V_R) + T_{sens}(S_R, V_P)} \tag{4.14}$$

On obtient finalement un ClDice pour l'exemple de la figure 4.7 de $ClDice = 2 \times \frac{1 \times 0,47}{1+0,47} = 0,64$.

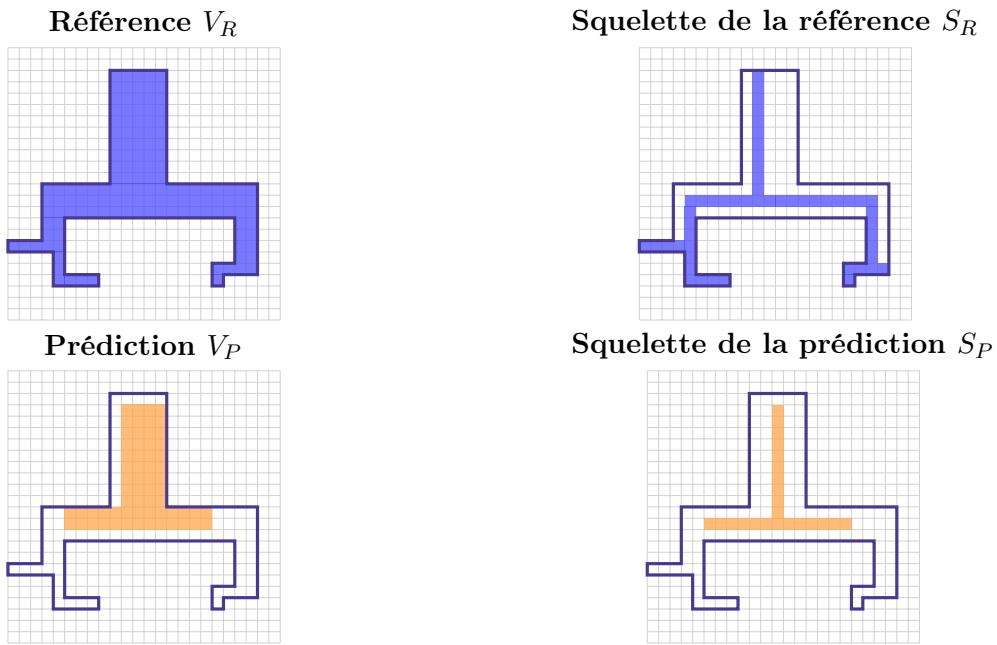


FIGURE 4.7 – Calcul du ClDice. Pour calculer la précision au sens du Cldice, on compte le nombre de pixels compris à la fois dans V_R et S_P puis on divise par le nombre de pixels de S_P. De même pour la sensibilité, on compte le nombre de pixels qui sont dans V_P et S_R et on divise par le nombre de pixels dans S_R. Sur cette figure, $|V_R| = 135$, $|V_P| = 62$, $|S_R| = 51$ et $|S_P| = 23$ pixels.

En regardant sur les exemples précédents, le score de ClDice n'est pas défini pour la segmentation vide car on ne peut pas diviser par 0, il est nul pour les segmentations totale

et aléatoire et il est meilleur pour la segmentation partielle par rapport à celle avec des FP $(0,90 > 0,80)$.

Indice de Tversky

Cet indice généralise le score de Dice en pondérant la précision et le rappel par des coefficients α et β au dénominateur :

$$Tv(P,R) = \frac{|P \cap R|}{|P \cap R| + \alpha|P \setminus R| + \beta|R \setminus P|} = \frac{VP}{VP + \alpha FP + \beta FN} \tag{4.15}$$

Cela permet d'insister sur le poids des faux positifs ou des faux négatifs (respectivement). Il est utilisé dans le cas de jeux de données déséquilibrés comme c'est souvent le cas en imagerie médicale [Salehi et al., 2017]. Dans le cas du fil guide, on choisit $\alpha = 0,1$ et $\beta = 0,9$ ainsi on pénalise les faux négatifs. En effet, nous désirons détecter la plus grande proportion plus de fil guide possible quitte à avoir plus de faux positifs.

Reprenons nos exemples, nous obtenons des scores de Tversky nuls ou quasi nuls dans les trois premiers cas, et un score légèrement meilleur pour la segmentation sans FP par rapport à celle avec les FP $(0,80 > 0,78)$ car ce sont les FN qui sont pénalisés $(\beta = 0,9)$ plus que les FP.

Méthode standardisée pour l'évaluation d'algorithmes d'extraction de la ligne centrale des artères coronaires [Schaap et al., 2009]

Dans cet article Schaap et al. [2009] développent une méthode d'évaluation standardisée d'algorithme d'extraction de ligne centrale pour la segmentation des artères coronaires. Nous présentons cette méthode ici car elle pourrait être utilisée pour d'autres structures fines.

L'auteur commence par définir la notion de vrais positifs, faux positifs et faux négatifs de la manière suivante (cf figure 4.8, les notations de l'article ont été simplifiées) :

- **TPR :** Un point de la ligne centrale de référence est classé dans cette catégorie si le chemin prédit est compris dans le rayon du vaisseau annoté (partie grise).
- **FN :** Au contraire, il est classé faux négatif dans le cas où un des points est en dehors du rayon annoté.
- **TPM :** Un point de la ligne centrale prédite est classé vrai positif si les points de la ligne centrale de référence sont compris dans une plus petite distance que le rayon du vaisseau annoté.
- **FP :** Il est classé comme faux positif sinon.

A l'aide de ces notations, les auteurs définissent deux mesures de chevauchement :

- **Mesure de chevauchement.** Cette mesure de similarité est proche du coefficient de Dice :

$$OV = \frac{|TPM| + |TPR|}{|TPM| + |TPR| + |FN| + |FP|} \tag{4.16}$$

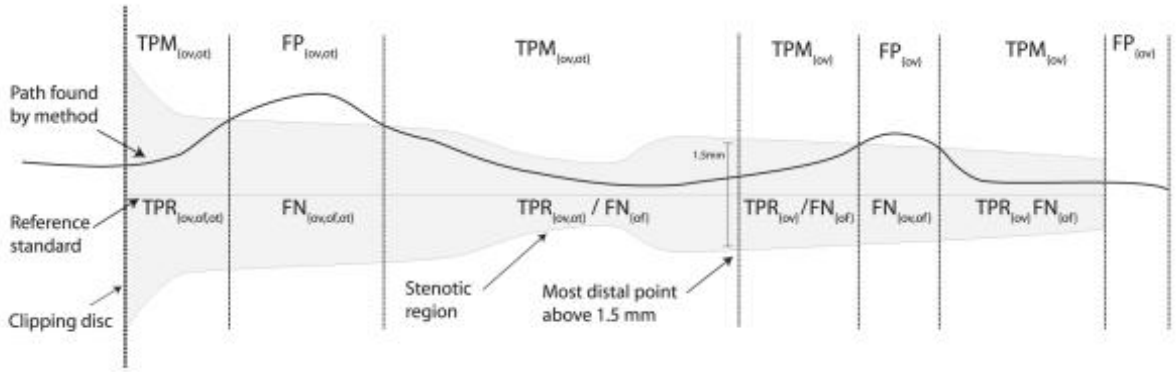


FIGURE 4.8 – Illustration des termes utilisés dans les mesures de chevauchement [Schaap et al., 2009]. En gris clair sont représentés la ligne centrale de référence ainsi que le rayon annoté de l'artère. Puis en plus foncé est représentée la ligne centrale prédite par l'algorithme. Les termes au-dessus de la ligne centrale prédite lui sont associés tandis que ceux en dessous sont associés à la ligne centrale de référence.

- **Mesure de chevauchement jusqu'à la première erreur :** Caractérise le niveau d'extraction d'une artère coronaire avant de faire une erreur :

$$OF = \frac{|TPR_{of}|}{|TPR_{of}| + |FN_{of}|} \tag{4.17}$$

avec TPR_{of}, FN_{of} calculés de la même manière que précédemment mais en s'arrêtant au premier FN rencontré.

En pratique cette mesure n'est pas très utilisée. Mais elle peut être intéressante si on l'associe aussi à la longueur détectée avant l'erreur.

4.4.2 Coefficient de Corrélation de Matthews (MCC)

Le MCC prend en compte chaque élément de la matrice de confusion tout comme la justesse mais de manière plus atténuée :

$$MCC = \frac{VP \times VN - FP \times FN}{\sqrt{(VP + FP)(VP + FN)(VN + FP)(VN + FN)}} \tag{4.18}$$

Il va alors être mieux adapté aux jeux de données déséquilibrés. Il est utilisé en complément de l'indice de Dice par Lamy et al. [2020] pour la segmentation de vaisseaux du foie, car il a l'avantage de prendre en compte les vrais négatifs. Il varie entre -1 et $+1$. Un MCC de 0 indique un classifieur aléatoire. En pratique nous utiliserons un MCC normalisé (avec des valeurs dans $[0, 1]$). Un score de MCC élevé traduit une bonne classification de chaque élément. Dans l'étude des cas limites, les trois premiers cas nous donnent un MCC normalisé de 0,5 donc une mauvaise segmentation. Dans le cas de la segmentation partielle nous obtenons des MCC de 0,89, et 0,87 pour la segmentation partielle avec des faux positifs. Le MCC est donc meilleur lorsqu'il n'y a pas de faux positifs.

4.4.3 Évaluer la topologie

La conservation de la topologie des structures à segmenter est un élément indispensable, surtout dans le cas de structures fines telles que le fil guide ou les vaisseaux. Il

existe des invariants topologiques comme les nombres de Betti. Ceux-ci permettent de caractériser le nombre de composantes connexes, de « trous » et autres caractéristiques.

Dans le cas du fil guide nous nous intéressons principalement aux composantes connexes, le cas de « trous » n'apparaissant que rarement lors de la formation de boucles dans les voies biliaires.

Nous décidons d'évaluer la différence entre le nombre de composantes connexes dans la référence avec leur nombre dans la prédiction et de sommer cette différence absolue sur l'ensemble de test.

On définit la différence du nombre de composantes connexes $Diff_{CC}$ telle que :

$$Diff_{CC} = \frac{1}{N}\sum_{i=1}^{N} |\beta_{0,\mathcal{R}} - \beta_{0,\mathcal{P}}| \tag{4.19}$$

avec N le nombre d'images dans l'ensemble de test, $\beta_{0,\mathcal{R}}$ (respectivement $\beta_{0,\mathcal{P}}$) le nombre de composantes connexes dans la référence (respectivement dans la prédiction).

Plus la valeur de $Diff_{CC}$ est faible et plus le nombre de composantes connexes est le même entre la référence et la prédiction. À l'inverse, une valeur élevée témoigne d'une discontinuité du fil guide ou d'un nombre élevé de faux positifs. Par exemple la segmentation vide obtient un score de 2 qui est le nombre de composantes connexes dans la référence. La segmentation totale a un score de 1 car elle est composée d'une composante connexe, il en reste donc une lorsqu'on fait la différence avec la référence. La segmentation aléatoire a un très grand nombre de composantes connexes (3484). La segmentation partielle a trois composantes connexes donc une seule de plus que la référence. Enfin, les FP augmentent le nombre de composantes connexes de la prédiction et le dernier cas a une différence qui vaut 32.

Récemment, Rougé et al. [2025] ont proposé le ccDice qui est un score similaire à celui de Dice prenant en compte les composantes connexes. Ce score n'est pas étudié dans ce travail mais pourrait l'être dans les perspectives.

4.4.4 Mesures utilisées pour la segmentation de vaisseaux

Mesure de similarité morphologique [La Barbera, 2022]

La Barbera [2022] décrit une nouvelle fonction de coût qui utilise les filtres de *vesselness* pour la segmentation de vaisseaux et qui seront décrits dans le prochain chapitre. La fonction de coût contient un terme dit de similarité morphologique qui est également utilisé comme fonction d'évaluation de la segmentation.

La matrice hessienne H est calculée pour chaque pixel de r_m de R et chaque pixel p_m de P la référence et la prédiction après avoir appliqué $5-q$ noyaux gaussiens d'écart-type σ_i. Les valeurs propres de chaque matrice hessienne sont ordonnées dans une matrice Λ selon les vecteurs directeurs associés. Pour chaque élément de la matrice Λ de la prédiction, on cherche le terme associé dans la matrice Λ de la référence, c'est-à dire qu'on associe les valeurs propres pour que l'angle entre les vecteurs propres de la référence et de la prédiction soit minimal. Enfin ce calcul est fait pour chaque voxel, qui sont au nombre de $M_{\tilde{R}}$. Ainsi on retrouve le terme de similarité morphologique en trois dimensions :

$$\mathrm{MsLoss}_q(P_q, R) = \frac{1}{3M_{\tilde{R}}}\sum_{m=1}^{M_{\tilde{R}}}\sum_{o=1}^{3}(\Lambda_o(H(\sum_{i=1}^{5-q}(g_{\sigma_i} * p_m))) - \Lambda_o(H(\sum_{i=1}^{5-q}(g_{\sigma_i} * r_m))))^2 \tag{4.20}$$

avec $M_{\tilde{R}}$ le nombre de voxels de la dilatation de R avec un élément structurant carré de taille $3 \times 3 \times 3$ et Λ_o l'élément o de la matrice Λ.

Et nous l'adaptons en deux dimensions :

$$\mathrm{MsLoss}_q(P_q, R) = \frac{1}{2M_{\tilde{R}}} \sum_{m=1}^{M_{\tilde{R}}} \sum_{o=1}^{2} (\Lambda_o(H(\sum_{i=1}^{5-q}(g_{\sigma_i} * p_m))) - \Lambda_o(H(\sum_{i=1}^{5-q}(g_{\sigma_i} * r_m))))^2 \quad (4.21)$$

4.5 Évaluation clinique : erreur de détection des points clés

Il est également possible d'évaluer la détection de l'extrémité du fil guide car il s'agit de la partie la plus importante. La segmentation n'est alors plus la tâche évaluée. Une mesure pertinente serait la détection du point d'extrémité, en évaluant d'une part si l'extrémité est bien présente sur la prédiction et d'autre part à quelle distance avec une marge de confiance.

4.6 Conclusion et tableau récapitulatif

Nous montrons dans ce chapitre que les méthodes classiques d'évaluation de la segmentation ne sont pas forcément adaptées dans le cas de la segmentation de structures fines. Nous dressons alors un panorama d'autres méthodes qui semblent plus adaptées pour compléter l'évaluation et donner plus d'informations à la fois aux personnes développant le modèle et à celles qui l'utilisent. Le tableau 4.2 en est un récapitulatif.

Ces méthodes d'évaluation sont indépendantes des modèles d'apprentissage utilisés. Cependant, dans un cas où une méthode d'évaluation donne un score non satisfaisant pour la tâche, nous pourrions imaginer utiliser cette méthode dans l'optimisation du modèle.

Méthode	**Vide**	**Totale**	**Aléatoire**	**Partielle**	**Partielle+FP**	**Avantages**	**Inconvénients**
VP	0	3227	1577	2349	2743		
VN	1 045 349	0	523 373	1045349	1043828		
FP	0	1 045 349	521976	0	1521		
FN	3227	0	165	878	484		
Accuracy	0,99	0,0	0,49	0,99	0,99		Pas adaptée au déséquilibre de classe.
Sensibilité	0,0	1,0	0,51	0,83	0,85		
Précision	0,0	0,0	0,0	0,73	0,64		
Spécificité	1,0	0,0	0,50	0,99	0,99		
Dice	0	0,0	0,0	0,78	0,73	Ne prend pas en compte les VN.	Ne prend pas en compte la forme de la structure. Dépend de la taille de la structure.
Distance de Hausdorff	1084	667,5	667,3	70	470	Permet de comparer la distance entre les FP de deux prédictions.	Mesure pessimiste, très sensible aux points aberrants (FP très éloignés).
ClDice	-	0,0	0,0	0,90	0,80	Prend en compte la structure de l'objet à segmenter contrairement au score de Dice.	
Tversky ($\alpha = 0,1$,$\beta = 0,9$)	0,0	0,01	0,01	0,80	0,78	Permet de pondérer les FP et les FN en fonction de notre objectif.	Il faut fixer des paramètres α et β qui correspondent à ce que l'on veut évaluer. Il faut faire un compromis entre les deux.
MCC_{norm}	0,5	0,5	0,5	0,89	0,87	Prend en compte tous les éléments de la matrice de confusion.	
$Diff_{CC}$	2	1	3484	1	32	Indice sur l'écart du nombre de composantes connexes dans la prédiction.	

Table 4.2 – Tableau récapitulatif des méthodes et de leurs avantages ou inconvénients pour l'évaluation de résultats de segmentation en imagerie médicale (exemples de la figure 4.1). Les colonnes correspondent donc aux prédictions vide, totale, aléatoire, partielle et partielle avec des FP.

CHAPITRE 5

Segmentation du fil guide

Ce chapitre porte sur la détection et la segmentation du fil guide dans les cholangiogrammes, qui est l'objectif principal de la thèse. Cette segmentation sur les images de fluoroscopie est une étape intermédiaire qui indique de manière automatique à l'algorithme de fusion où se trouve le fil guide. Une segmentation complète n'est pas nécessaire pour que l'algorithme de fusion fonctionne mais certains points clés doivent être détectés, notamment l'extrémité, et le point d'entrée dans les voies biliaires car la prédiction de localisation s'appuie sur ces points. Néanmoins, l'étape de segmentation demeure difficile car le fil guide est une structure très fine qui représente moins de 1% de l'image. Par conséquent, il existe un fort déséquilibre de classes dans les données, car la plupart des pixels ne font pas partie du fil guide. Enfin l'ensemble de données est très petit comparé à ceux utilisés dans l'état de l'art qui réunissent souvent plus de mille images pour l'entraînement. Ces arguments ne sont pas en faveur de l'utilisation d'un réseau de neurones. Cependant l'utilisation de méthodes classiques uniquement, par exemple les filtres de *vesselness*, ne permettent pas une segmentation du fil guide satisfaisante car ils le rehaussent sur l'image sans l'isoler des autres structures (Annexe : figure B.1). Par ailleurs, l'état de l'art établi dans le chapitre 2 montre qu'un réseau U-Net seul ne permet pas de résoudre cette tâche, même avec un grand nombre de données. Notre approche propose d'utiliser un réseau U-Net comme architecture de base et d'expérimenter différentes fonctions de coût. La démarche consiste à ne pas alourdir le réseau avec des modules qui augmenteraient les ressources de calcul et de temps. Nous cherchons plutôt à simplifier le processus d'apprentissage en choisissant des fonctions de coût spécifiques, capables de contraindre le réseau à apprendre ce type de structures sans augmenter le nombre de paramètres et donc le temps de calcul.

La suite du chapitre dresse un panorama critique des fonctions de coût utilisées en segmentation et qui seront utilisées dans les expériences. Après une description détaillée des données, le plan d'expérience est expliqué. Finalement les résultats qualitatifs et quantitatifs sont présentés et analysés.

5.1 État de l'art sur les fonctions de coût pour la segmentation de structures fines et le déséquilibre de classes

La fonction de coût est un élément essentiel de l'apprentissage par ordinateur car c'est cette fonction que l'on cherche à optimiser, c'est-à-dire que l'on va chercher à s'approcher du minimum global de cette fonction pour trouver les poids du modèle. En apprentissage supervisé, la fonction de coût évalue la dissimilarité entre la prédiction en sortie du réseau et la segmentation de référence. Cette évaluation sert à corriger les poids du réseaux via un mécanisme de rétro-propagation. La segmentation de référence correspond à la segmentation manuelle faite par des experts du domaine. Dans un processus d'apprentissage, le réseau va d'abord prédire au hasard puis l'erreur calculée va orienter l'apprentissage jusqu'à converger vers une prédiction fiable.

Différents types de fonctions de coût ont été utilisés en imagerie biomédicale :

- **L'entropie-croisée** est la plus classique. Elle sera utilisée comme référence, couplée à la fonction de Dice.
- **La fonction de Dice** [Milletari et al., 2016], la plus utilisée en segmentation d'images biomédicales, mesure le recouvrement entre la prédiction et la référence.
- **La fonction de Tversky** [Salehi et al., 2017], qui permet de pondérer différemment les faux positifs et les faux négatifs.
- **La fonction *Focal*** [Lin et al., 2017] adaptée aux déséquilibres de classes extrêmes.
- **La fonction *ClDice*** [Shit et al., 2021] qui prend en compte les lignes centrales des objets à segmenter.
- **La fonction de *Vesselness*** [La Barbera et al., 2023] contraint le réseau à prédire des structures fines et allongées en utilisant les valeurs du filtre de Frangi et al. [1998].

La discussion qui suit complète celle faite pour les mesures d'évaluation au chapitre 4.

5.1.1 Notations

Pour la suite du chapitre nous noterons p_i la prédiction du réseau de neurones au pixel i appartenant à l'image de prédiction $\mathcal{P}$: cette valeur représente la probabilité du pixel d'appartenir à la classe c, elle appartient à l'intervalle $]0, 1[$ car la dernière couche d'activation du réseau est une fonction sigmoïde. De même, r_i est la valeur du i-ème pixel de la segmentation de référence $\mathcal{R}$ qui est binaire. Enfin, N représente le nombre de pixels et C le nombre de classes. Dans le cas de la segmentation binaire du fil guide, C est égal à 2.

5.1.2 Entropie croisée

L'entropie croisée a d'abord été très utilisée comme fonction de coût pour les tâches de classification. C'est également une des premières fonctions objectif utilisées en segmentation, notamment par Ronneberger et al. [2015]. La tâche de segmentation peut être

vue comme une classification à l'échelle du pixel. L'entropie croisée peut résoudre des tâches binaires mais elle est généralisable pour un nombre C de classes. Ces deux types de fonctions de coût sont décrites ci-après d'après [Ma et al., 2021a].

Entropie croisée binaire (ECB) L'entropie croisée binaire, que nous notons $\mathcal{L}_{\mathcal{ECB}}$, mesure la somme des dissimilarités entre les valeurs de chaque pixel de $\mathcal{R}$ et $\mathcal{P}$. Le i-ème pixel de la segmentation de référence a pour valeur r_i qui peut prendre les valeurs 0 ou 1 uniquement, selon que le pixel appartient au fil guide ou non. Les valeurs de p_i sont comprises dans l'intervalle $]0, 1[$. Cette fonction s'écrit :

$$\mathcal{L}_{ECB}(\mathcal{P}, \mathcal{R}) = -\frac{1}{N}\sum_{i=1}^{N}[r_i \cdot \log(p_i) + (1 - r_i) \cdot \log(1 - p_i)] \tag{5.1}$$

Entropie croisée généralisée (ECG) L'entropie croisée binaire se généralise pour la classification multi-classes par l'entropie croisée généralisée (ECG) notée $\mathcal{L}_{ECG}$:

$$\mathcal{L}_{ECG}(\mathcal{P}, \mathcal{R}) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{C} r_{i,k} \cdot \log(p_{i,k}) \tag{5.2}$$

où pour la classe k, $r_{i,k}$ vaut 1 si le pixel i appartient à la classe k et 0 sinon, et $p_{i,k}$ est la probabilité que le i-ème pixel appartienne à la classe k.

Discussion L'entropie croisée permet au modèle de bien généraliser lorsque l'hypothèse d'égale importance entre les classes est respectée. En revanche, lorsque le jeu de données est déséquilibré, un fort biais est introduit lors de l'apprentissage. Il existe plusieurs stratégies pour contrer ce déséquilibre, en associant des poids plus importants aux classes les moins fréquentes comme par exemple dans [Ronneberger et al., 2015].

5.1.3 Fonction de coût de Dice

Le score de Dice est une mesure de similarité introduite comme une fonction de coût par Milletari et al. [2016] pour lutter contre le déséquilibre de classes en segmentation, c'est-à-dire lorsque l'objet à segmenter est très petit par rapport au reste de l'image. En effet, seuls les pixels de l'objet sont pris en compte dans le calcul. Le score de Dice tel qu'il est défini dans la section 4.3.1 n'est pas utilisable en l'état comme une fonction de coût car il n'est pas différentiable. Milletari et al. [2016] proposent une formulation qu'ils nomment *Soft Dice*. Le fonction de coût de Dice associée, notée $\mathcal{L}_{Dice}$ calculée sur la prédiction $\mathcal{P}$ et la segmentation de référence $\mathcal{R}$ est définie par :

$$\mathcal{L}_{Dice}(\mathcal{P}, \mathcal{R}) = 1 - \frac{2\sum_{i=1}^{N} P_i R_i}{\sum_{i=1}^{N} P_i^2 + \sum_{i=1}^{N} R_i^2} \tag{5.3}$$

où N est le nombre de pixels, P_i le pixel i de $\mathcal{P}$ et R_i le pixel i de $\mathcal{R}$. Les termes au carré au dénominateur sont justifiés par les auteurs comme permettant une convergence plus rapide du réseau de neurones.

Discussion Plusieurs limites du score de Dice sont mises en avant par Zhao et al. [2020]. La première limite identifiée est que la distance séparant la référence et la prédiction n'est

pas prise en compte. Ainsi si on compare deux prédictions composées uniquement de faux positifs de la même taille mais avec une distance différente à la référence, le score de Dice sera le même, ce qui signifie que le score de Dice ne discrimine pas les faux positifs en fonction de leur distance. Une autre limite, très importante, est que lorsque deux régions d'intérêt sont présentes sur l'image, l'une ayant une surface beaucoup plus importante que l'autre, le score de Dice ne sera pas très sensible au fait que la plus petite région soit détectée. De plus si la région d'intérêt est petite, la même erreur en nombre de pixels va être plus coûteuse que dans le cas d'une plus grande région d'intérêt. D'une manière générale, le score de Dice est plus sensible pour les structures de petite surface comme le fil guide.

Combinaison de fonctions de coût Une première idée pour diminuer les inconvénients des deux premières fonctions de coût présentées est de les combiner. C'est d'ailleurs la combinaison des fonctions de coût de Dice et d'entropie croisée qui est utilisée comme paramètre par défaut dans la méthode nnU-Net [Isensee et al., 2021].

5.1.4 Fonction de coût de Tversky

Cette nouvelle fonction de coût, introduite par Salehi et al. [2017] est dérivée du score de Tversky décrit dans la section 4.4.1. Elle est définie par :

$$\mathcal{L}_{Tversky}(\mathcal{P},\mathcal{R}) = 1 - \frac{\sum_{i=1}^{N} p_i r_i}{\sum_{i=1}^{N} p_i r_i + \alpha \sum_{i=1}^{N} p_i(1-r_i) + \beta \sum_{i=1}^{N}(1-p_i) r_i} \tag{5.4}$$

Les notations sont les mêmes que précédemment. Les paramètres α et β sont à fixer par les utilisateurs. En général, ils sont choisis de sorte que $\alpha+\beta = 1$. Le terme α pondère la précision en accordant plus d'importance aux faux positifs, c'est-à-dire quand r_i vaut 0 et p_i est élevée, tandis que le terme β pondère le rappel au travers des faux négatifs, qui se traduisent par r_i égal à 1 et une valeur p_i faible.

Discussion Cette fonction de coût a pour avantage de pondérer les faux positifs ou faux négatifs en fonction de la tâche à effectuer. Néanmoins, il faut mener des expériences pour déterminer quels paramètres α et β sont les plus appropriés.

5.1.5 Fonction de coût Focal

La fonction de coût *Focal* [Lin et al., 2017] dérive directement de la fonction d'entropie croisée à laquelle on ajoute le terme de pondération $(1-p_i)^\gamma$. Le principe repose sur le fait qu'un exemple facile aura moins de poids dans la fonction de coût qu'un exemple difficile. Dans un exemple difficile la probabilité p_i sera faible, le terme de pondération sera alors important. À l'inverse un cas simple aura une probabilité p_i proche de 1. Le terme de pondération avec la puissance γ sera alors très faible par rapport aux exemples faciles. Ainsi un exemple facile bien classé fera moins diminuer la fonction de coût qu'un exemple difficile bien classé.

La fonction est définie par :

$$\mathcal{L}_{Focal}(\mathcal{P},\mathcal{R}) = -\frac{1}{N}\sum_{i=1}^{N}[(1-p_i)^\gamma r_i \cdot \log(p_i) + (-p_i)^\gamma (1-r_i)\cdot \log(1-p_i)] \tag{5.5}$$

avec $\gamma \geq 0$ un paramètre à déterminer. Notons que pour $\gamma = 0$, on retrouve la fonction d'entropie croisée.

Elle peut être étendue pour plus de deux classes comme suit :

$$\mathcal{L}_{Focal}(\mathcal{P}, \mathcal{R}) = -\frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{C} (1 - p_{i,k})^{\gamma} r_{i,k} \cdot \log(p_{i,k}) \tag{5.6}$$

Discussion La fonction de coût *Focal* nuance la fonction d'entropie croisée en attribuant des poids plus élevés aux cas difficiles. Elle est ainsi plus adaptée pour les déséquilibres de classes.

5.1.6 Fonction de coût *ClDice*

La fonction de coût de ClDice est dérivée du score ClDice décrit dans la section 4.4.1. Elle s'écrit :

$$\mathcal{L}_{ClDice}(\mathcal{P}, \mathcal{R}) = (1 - \alpha)\mathcal{L}_{Dice}(\mathcal{P}, \mathcal{R}) + \alpha(1 - ClDice(\mathcal{P}, \mathcal{R})) \tag{5.7}$$

avec α compris entre 0 et $0,5$. Elle favorise les prédictions ayant le nombre de composantes connexes attendu.

5.1.7 Du filtre à la fonction de coût pour les structures tubulaires

Une fonction de coût spécifique pour les structures tubulaires a été développée par La Barbera et al. [2023] pour segmenter les vaisseaux sanguins sur des images par tomodensitométrie d'enfants. Son principe repose sur les filtres de *vesselness* comme par exemple celui de Frangi et al. [1998], qui ont été conçus pour rehausser les structures tubulaires, fines et allongées sur les images, comme les vaisseaux ou les neurones. L'injecter ainsi dans la fonction de coût a pour but de contraindre le réseau de neurones à prédire ce type de structures.

D'autres filtres, comme les filtres morphologiques d'ouvertures et fermetures par chemins comme ceux proposés par Talbot and Appleton [2007] et Merveille et al. [2018] pourraient également être envisagés mais ne sont pas utilisés dans ce travail.

Filtres de *vesselness*

Prenons l'exemple du filtre de Frangi et al. [1998] qui s'appuie sur les valeurs propres de la matrice hessienne pour rehausser les vaisseaux en 3D et 2D sur les images d'angiographie. La matrice hessienne décrit les dérivées partielles secondes dans chaque direction de l'espace. Soit f une fonction continue et deux fois dérivable définie sur $\mathbb{R}^3$ telle que $f(x, y, z)$ est l'intensité au voxel (x, y, z) d'une image $\mathcal{I}$. La matrice hessienne $H(f)$ de f au voxel (x, y, z) est définie ainsi :

$$H(f) = \begin{bmatrix} h_{x,x} & h_{x,y} & h_{x,z} \\ h_{y,x} & h_{y,y} & h_{y,z} \\ h_{z,x} & h_{z,y} & h_{z,z} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial y \partial z} \\ \frac{\partial^2 f}{\partial z \partial x} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} \end{bmatrix} \tag{5.8}$$

où toutes les dérivées sont prises au point (x, y, z).

En pratique, l'image est d'abord convoluée par un noyau gaussien g_σ d'écart-type σ pour assurer les conditions de continuité et de dérivabilité de f. Les vecteurs propres notés e_1, e_2 et e_3, orthonormés et associés aux valeurs propres λ_1, λ_2 et λ_3 sont ensuite calculés. Les valeurs propres sont ordonnées telles que $|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$. L'analyse des valeurs et vecteurs propres indique les directions principales de courbure locale du vaisseau. La première hypothèse émise consiste à considérer un vaisseau comme une structure allongée apparaissant dans des niveaux de gris clairs sur un fond plus foncé. Cela se traduit par les propriétés suivantes sur les valeurs propres :

- $|\lambda_1| \approx 0$
- $\lambda_2 \approx \lambda_3 \ll 0$

Concernant les vecteurs propres, e_1 est associé à l'axe de la structure tubulaire et e_2, e_3 sont orthogonaux à cet axe et entre eux. Ils sont associés aux deux axes d'ellipse de la section de la structure tubulaire.

Le filtre de Frangi et al. [1998] met en avant les structures ayant ces propriétés avec un terme noté F pour chaque voxel, calculé par :

$$F = \begin{cases} (1 - \exp(\frac{-R_a^2}{2\alpha^2})) \exp(\frac{-R_b^2}{2\beta^2})(1 - \exp(\frac{-S^2}{2\gamma^2})) & \text{si } \lambda_2 \text{ et } \lambda_3 \leq 0 \\ 0 \quad \text{sinon} \end{cases} \tag{5.9}$$

avec

- $R_a = |\lambda_2|/|\lambda_3|$, une mesure de distinction entre les structures plates et les lignes
- $R_b = |\lambda_1|/\sqrt{|\lambda_2\lambda_3|}$, terme qui mesure la dissimilarité par rapport à une structure de type boule (*blob*)
- $S = \sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}$.

La valeur de F sera d'autant plus élevée que les propriétés sont bien satisfaites. Les paramètres α, β, γ sont des seuils à fixer par l'utilisateur pour gérer la sensibilité du filtre.

En deux dimensions, il n'y a plus que deux valeurs propres. Une structure tubulaire qui est claire sur un fond foncé aura comme propriétés $|\lambda_1| \approx 0$ et $\lambda_2 \ll 0$. La réponse du filtre vaut alors :

$$F = \begin{cases} (\exp(\frac{-R_b^2}{2\beta^2})(1 - \exp(\frac{-S^2}{2\gamma^2})), & \text{si } \lambda_2 \leq 0 \\ 0 \text{ sinon} \end{cases} \tag{5.10}$$

avec cette fois $R_b = |\lambda_1|/|\lambda_2|$ et $S = \sqrt{\lambda_1^2 + \lambda_2^2}$.

Notons que dans le cas du fil guide, l'objet est foncé sur un fond plus clair, il faut alors prendre l'inverse de l'image.

Fonction de coût associée au filtre de Frangi

Le fonction de coût se décompose en deux termes, le premier qui calcule la valeur du filtre de Frangi (notée Fv) et le deuxième qui calcule la similarité morphologique (notée Ms) entre la référence $\mathcal{R}$ et la prédiction $\mathcal{P}$, et qui seront détaillés dans la suite. Cette méthode a initialement été conçue pour des images 3D mais est adaptée aux images 2D dans la section suivante.

Lors de l'apprentissage, la valeur de la fonction de coût est calculée à partir de la référence et de la prédiction à l'époque t. La première étape consiste en une convolution avec un noyau gaussien g_{σ_i} dont l'écart-type σ peut varier pour sélectionner les structures en fonction de leurs tailles. Il s'agit d'une approche multi-échelles. Cette convolution permet le calcul de la matrice hessienne H aux pixels de la référence (r_m) et de la prédiction (p_m).

Le terme de similarité morphologique (Ms) est présent pour s'assurer que les valeurs propres sont associées dans l'ordre exact entre la référence et la prédiction lors du calcul de la *vesselness*. En effet dans les cas où la prédiction d'un vaisseau a une direction différente de celle de la référence, il faut s'assurer que la fonction de coût compare les bonnes valeurs propres entre elles, c'est-à-dire en comparant des valeurs propres dont les vecteurs propres ont des directions proches et pas les deux plus petites valeurs propres. Les valeurs propres sont ordonnées dans la matrice Λ de sorte à minimiser l'angle entre les vecteurs propres de la référence et de la prédiction. La figure 5.1 illustre ce principe.

Le terme de similarité morphologique est défini comme suit :

$$\mathrm{Ms}(\mathcal{P},\mathcal{R}) = \frac{1}{2M_{\tilde{\mathcal{R}}}} \sum_{m=1}^{M_{\tilde{\mathcal{R}}}} \sum_{o=1}^{2} (\Lambda_o(H(\sum_{j=1}^{5-q}(g_{\sigma_j} * p_m))) - \Lambda_o(H(\sum_{i=1}^{5-q}(g_{\sigma_i} * r_m))))^2 \tag{5.11}$$

avec $M_{\tilde{\mathcal{R}}}$ le nombre de pixels de la dilatation de $\mathcal{R}$ par un élément structurant de taille 3. L'écart-type du noyau gaussien σ_j varie à plusieurs échelles selon j qui est compris entre 1 et $5-q$. Le choix du paramètre q va dépendre du nombre d'échelles auxquelles les objets à segmenter appartiennent. Le terme Λ_o désigne l'élément o de la matrice Λ.

Le terme de *vesselness* de Frangi (Fv) calcule la valeur du filtre de Frangi pour les pixels de la prédiction et de la référence (en prenant sa valeur maximale) :

$$\mathrm{Fv}(\mathcal{P},\mathcal{R}) = \frac{1}{M_{\hat{F}}} \sum_{m=1}^{M_{\hat{F}}} (1 - \frac{F(H(\sum_{i=1}^{5-q}(g_{\sigma_i} * p_m))}{F_{\max}}) \tag{5.12}$$

avec $M_{\hat{F}}$ le nombre de pixels dans la référence et F_{max} la valeur maximale du filtre de vesselness dans la référence.

La fonction de coût pour les structures tubulaires (Ts) réunit ces deux termes en les additionnant et en les pondérant par les coefficients w_{ms} et w_{fv} qui sont fixés par l'utilisateur :

$$\mathrm{Ts}(\mathcal{P},\mathcal{R}) = \frac{1}{L} \sum_{l=1}^{L} (w_{ms}\,\mathrm{Ms}(\mathcal{P}_l,\mathcal{R}_l) + w_{fv}\,\mathrm{Fv}(\mathcal{P}_l,\mathcal{R}_l)) \tag{5.13}$$

avec L la taille de lot, P_l la prédiction de l'image l du lot, R_l la segmentation de référence correspondante. Ce terme est ajouté à la fonction de Dice par exemple pour améliorer la segmentation.

Discussion Pour des vaisseaux, les diamètres varient en fonction de la profondeur dans l'arbre vasculaire. Dans notre cas le fil guide a un diamètre constant. Il suffit donc de sélectionner une seule échelle de σ soit $q = 4$. L'inconvénient majeur repose sur le choix des hyperparamètres comme les coefficients w_{fv} et w_{ms}.

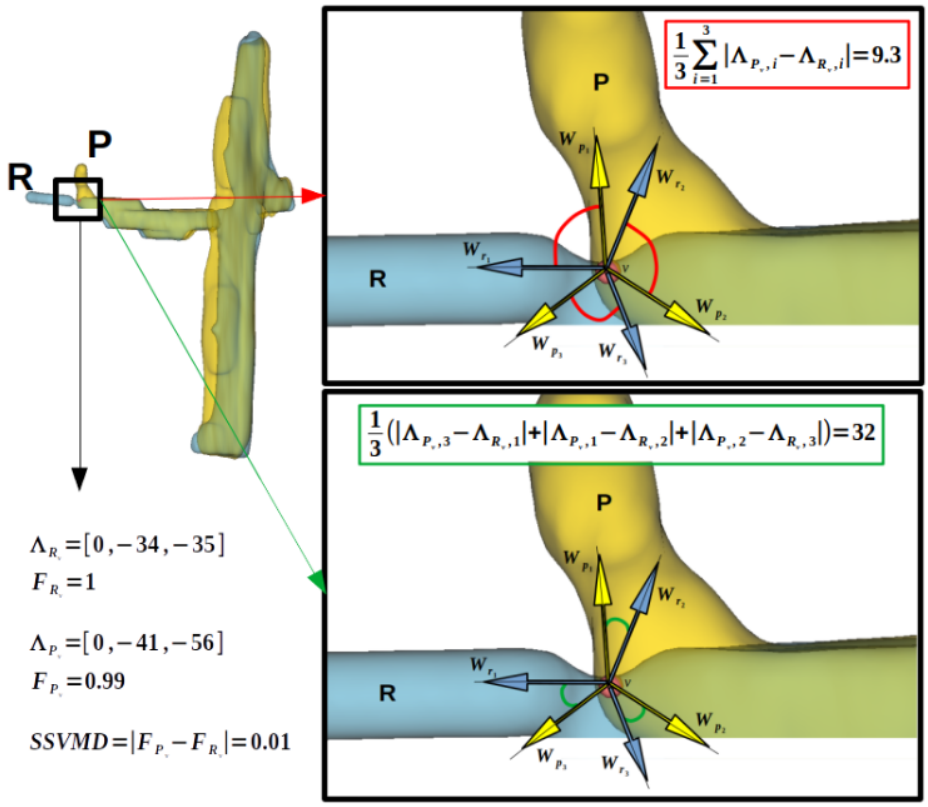


Figure 5.1 – La figure, provenant de [La Barbera et al., 2023], illustre l'influence d'un mauvais appariement des valeurs propres entre elles sur le score de SSVMD (somme des carrés de la différence des valeurs de vesselness [Cao et al., 2010]). Dans le cas où la prédiction et la référence ne seraient pas orientées dans la même direction, le terme Ms permet d'associer les valeurs propres de la prédiction et de la référence selon l'angle minimal entre les vecteurs propres.

5.1.8 Contours actifs et chemins minimaux

Dans cette section nous introduisons les modèles de contours actifs et de chemins minimaux et leur usage dans des fonctions du coût.

Contours actifs

Les modèles des contours actifs ou *snakes* ont été définis par Kass et al. [1988]. Ils reposent sur la minimisation de l'énergie d'une courbe initialisée par les utilisateurs (ou à l'issue d'un pré-traitement) à proximité des contours de l'objet. L'énergie de la courbe comporte deux termes : interne et externe.

Soit $\mathcal{C}$ la courbe paramétrée par v dans une image I en $2D$ et Ω son domaine de définition :

$$\begin{aligned} \mathcal{C} : \Omega &\longrightarrow \mathbb{R}^2 \\ v &\longmapsto (x(v), y(v)), \end{aligned} \tag{5.14}$$

avec x et y les coordonnées des points du *snake*.

L'énergie E d'une telle courbe peut s'écrire :

$$E : \mathcal{A} \to \mathbb{R} \tag{5.15}$$
$$\mathcal{C} \mapsto E(\mathcal{C}) = \underbrace{\int_\Omega \frac{w_1(v)}{2}|\mathcal{C}'(v)|^2 + \frac{w_2(v)}{2}|\mathcal{C}''(v)|^2 dv}_{E_{interne}} - \underbrace{\lambda \int_\Omega |\nabla I(\mathcal{C}(v))| dv}_{E_{externe}},$$

avec

- $\mathcal{C}'$ and $\mathcal{C}''$ les dérivées première et seconde de $\mathcal{C}$ en fonction de v. Ces deux premiers termes sont pondérés par les coefficients réels w_1 et w_2 qui représentent respectivement l'élasticité et la rigidité, et régularisent la courbe de façon à ce qu'elle soit plus lisse. Ces termes définissent l'énergie interne de la courbe.

- λ est un coefficient réel constant, et ∇I le gradient de l'intensité de l'image I qui va attirer la courbe vers le contour de l'objet.

L'objectif de la méthode est de minimiser E pour obtenir la meilleure segmentation, c'est-à-dire une courbe sur le contour qui est un minimum local de E. Le principal inconvénient de cette méthode est la difficulté d'initialisation de la courbe et le choix des paramètres. De plus, ces modèles déformables ne permettent pas facilement de changement de topologie entre la courbe initiale et la courbe finale.

Contours actifs géodésiques

Caselles et al. [1997] ont amélioré le modèle précédent en introduisant celui des contours actifs géodésiques. L'idée est de définir une métrique en fonction du contenu de l'image et de trouver la courbe de longueur minimale dans cet espace.

La première modification est de considérer l'abscisse curviligne pour s'affranchir du terme de dérivée seconde pondéré par w_2. L'énergie vaut à présent :

$$E(C) = \frac{w_1}{2}\int_\Omega |\mathcal{C}'(v)|^2 dv - \lambda \int_\Omega |\nabla I(\mathcal{C}(v)| dv \tag{5.16}$$

La seconde modification par rapport aux modèles de contours actifs est l'introduction d'une fonction d'indication des bords, strictement décroissante, g_I :

$$g_I = \frac{1}{1 + |\nabla(G_\sigma * I)|^2} \tag{5.17}$$

avec G_σ un noyau gaussien d'écart-type σ, ∇ l'opérateur gradient.

La nouvelle énergie devient :

$$E(C) = \frac{w_1}{2}\int_\Omega |\mathcal{C}'(v)|^2 dv + \lambda \int_\Omega g_I(|\nabla I(\mathcal{C}(v)|)^2 dv \tag{5.18}$$

Finalement, une fonction de niveau ϕ est introduite de sorte que la courbe corresponde au niveau zéro de la fonction ϕ. Cette nouvelle forme permet de s'affranchir de la paramétrisation qui est arbitraire, et fait évoluer ϕ à la place. Cette étape résout le problème de différence de topologie au cours de l'évolution car le contour final ne dépend pas de l'initialisation.

Chemins minimaux

L'approche par chemins minimaux introduite par Cohen and Kimmel [1997] propose de trouver le minimum global de l'énergie du contour actif entre deux points p_0 et p_1, réduisant ainsi les inconvénients des deux méthodes citées plus haut. Pour cela, ils modifient l'énergie des *snakes* en ajoutant le terme interne dans le terme de potentiel externe. Ainsi, le chemin minimal entre p_0 et p_1 dans l'ensemble $\mathcal{A}_{p_0,p_1}$ de tous les chemins entre ces points est celui qui minimise E qui s'écrit :

$$\begin{aligned} E :& \mathcal{A}_{p_0,p_1} \to \mathbb{R} \\ \mathcal{C} \mapsto E(\mathcal{C}) &= \int_\Omega w|\frac{\partial \mathcal{C}}{\partial v}(v)|^2 + P(\mathcal{C}(v))dv \\ &= wL(\mathcal{C}) + \int_\Omega P(\mathcal{C}(v))dv \\ &= \int_\Omega \tilde{P}(\mathcal{C}(v))dv, \end{aligned} \tag{5.19}$$

L'énergie est ainsi calculée pour tous les chemins démarrant de p_0 en appliquant la méthode de « marche rapide »(*fast marching*) de Sethian [1996]. Enfin, l'algorithme de descente de gradient appliqué depuis le point d'arrivée, désigné p_1, sur l'ensemble des chemins possibles noté $\mathcal{A}_{p_0,p_1}$ extrait le chemin d'énergie minimale. Le potentiel P est défini en fonction du contenu de l'image. En pratique, v est fixé comme le paramètre de longueur d'arc ce qui implique $|\mathcal{C}'(v)| = |\frac{\partial \mathcal{C}}{\partial v}(v)| = 1$. Par conséquent, l'énergie se simplifie par le seul potentiel $\tilde{P}(p) = w + P(p)$. Le résultat dépend fortement du choix de potentiel. Il est généralement choisi comme fonction des intensités des pixels, par exemple la différence absolue entre les intensités de l'image et celle au pixel p_0. Cohen and Kimmel [1997] utilisent également les filtres de *vesselness*, mentionnés ci-dessus, comme potentiel pour les structures allongées.

Cette méthode, contrairement aux modèles de contours actifs, assure d'atteindre le minimum global de la fonction. Cependant l'algorithme est très sensible au choix des points de départ et d'arrivée qui sont souvent définis manuellement.

Fonctions de coût à partir des modèles de contours actifs

En 2019, Chen et al. [2019] introduisent les contours actifs comme fonction de coût dans un réseau de neurones pour la segmentation d'images biomédicales. Leur méthode, nommée $AC\mathcal{L}oss$, s'appuie sur l'équation de Bresson et al. [2007] qui tire profit de la variation totale pour accélérer le processus de minimisation d'énergie des contours actifs :

$$\begin{aligned} E(u, c_1, c_2, \lambda) &= TV_g(u) + \lambda \int_\Omega r_1(x, c_1, c_2)udx \\ &= \int_0^{Length(C)} g|\nabla u|ds + \lambda \int_\Omega ((c_1 - I(x))^2 - (c_2 - I(x))^2)udx \end{aligned} \tag{5.20}$$

où le premier terme est le terme de longueur d'arc et le second le terme de région, u est une fonction indicatrice qui vaut 1 dans C et 0 ailleurs, c_1 et c_2 sont les valeurs moyennes de l'image à l'intérieur et à l'extérieur de la courbe fermée.

La fonction de coût $\mathcal{L}_{AC}$ qui s'appuie sur la précédente équation d'énergie s'exprime

alors par :

$$\mathcal{L}_{AC} = Length + \lambda.Region \tag{5.21}$$

Les auteurs ont comparé leur fonction de coût avec celles de Dice et d'entropie croisée pour les architectures de réseaux U-Net [Ronneberger et al., 2015] et Dense-Net [Huang et al., 2017] sur le jeu de données ACDC [Bernard et al., 2018] contenant des images IRM du coeur.

Ils obtiennent la meilleure distance de Hausdorff et le meilleur score de Dice avec Dense-Net combiné à leur fonction de coût. De plus, l'impact du paramètre λ a été testé en le faisant varier. Les résultats obtenus pour différentes valeurs sont similaires et aucune valeur ne semble diminuer la performance. Les auteurs en déduisent une certaine robustesse de la méthode en fonction de λ.

En 2020, Ma et al. [2021b] ont injecté les contours actifs géodésiques dans la fonction de coût combinée ($\mathcal{L}_{GAC}$) avec la fonction de coût de Dice et la norme L_1. Le réseau prédit une fonction de niveau ϕ qui doit minimiser l'énergie des contours actifs.

$$\mathcal{L} = \mathcal{L}_{Dice} + L_1 + \alpha\mathcal{L}_{GAC} \tag{5.22}$$

La fonction $\mathcal{L}_{GAC}$ est définie par :

$$\mathcal{L}_{GAC}(\phi) = \int_{\Omega} g_I \delta_\epsilon(\phi)|\nabla\phi|dx \tag{5.23}$$

avec g_I la même fonction que mentionné précédemment et ϕ la fonction de niveaux définie par :

$$\phi(x) = \begin{cases} -\inf_{y\in\partial S} \|x-y\|_2, & x \in S_{\text{in}} \\ 0, & x \in \partial S \\ \inf_{y\in\partial S} \|x-y\|_2, & x \in S_{\text{out}} \end{cases} \tag{5.24}$$

Ici $||.||_2$ représente la distance euclidienne, S_{in}, S_{out} et ∂S sont respectivement l'intérieur, l'extérieur et le bord de l'objet.

Le terme δ_ϵ est défini par :

$$\delta_\epsilon(\phi) = \frac{\epsilon}{\pi(\phi^2 + \epsilon^2)} \tag{5.25}$$

qui est une version lisse et dérivable de la fonction de Dirac.

Ici le réseau prédit une fonction de niveau, qui est transformée en carte de probabilités grâce à une fonction de Heaviside afin de calculer le score de Dice avec la référence.

5.1.9 Conclusion

Dans cette section nous nous sommes intéressés à différentes fonctions de coût qui pourraient améliorer la segmentation de structures fines car elles prennent en compte le déséquilibre de classe (Tversky, Focal) ou bien parce qu'elles sont adaptées aux structures allongées (*Vesselness*, ClDice). Nous nous sommes également intéressés aux approches par contours actifs et chemins minimaux car ceux-ci permettent la segmentation de contours, c'est-à-dire de courbes fines. Plusieurs méthodes par contours actifs ont été ainsi introduites dans des fonctions de coût. Nous nous en inspirons pour l'introduire dans une fonction de coût qui utilise la *vesselness* comme terme d'attraction à la place du gradient

et où nous cherchons à minimiser une énergie de courbe. Cette méthode permettrait de s'affranchir de l'initialisation qui est l'inconvénient majeur des méthodes par contours actifs et d'ajouter une contrainte dans la fonction de coût. Ainsi dans la suite du chapitre nous comparons ces fonctions de coût à celle la plus utilisée en segmentation d'images biomédicales, la fonction de coût de Dice (parfois combinée à l'entropie croisée) et cherchons à montrer l'impact de ces dernières sur les résultats de la segmentation.

5.2 Données

Pour mettre au point notre modèle de segmentation, les données ont été sélectionnées dans les bases décrites dans le chapitre 3. Nous avons pris soin de ne prendre que les images contenant le fil guide sans ou avec très peu d'opacification (c'est-à-dire les images ayant pour code correspondant le chiffre trois dans la base de données) car cette dernière trouble la visualisation du fil guide sur les cholangiogrammes. Ce choix restreint énormément le nombre d'images pour l'entraînement du réseau de neurones. Les images de chaque base de données ont été séparées en trois sous-ensembles, d'entraînement, de validation et de test dans les proportions décrites par le tableau 5.1. La création de sous-ensembles est soumise à une contrainte majeure. Étant donné qu'il y a plusieurs images par patient et peu de patients, les choix de découpe en sous-ensembles sont limités. Non seulement ils doivent être faits manuellement pour qu'aucune image d'un même patient ne soit utilisée à la fois pour l'entraînement et l'évaluation, ce qui fausserait cette dernière, mais le nombre d'images par patient n'est pas constant, ce qui signifie que la séparation peut être déséquilibrée. Par exemple un patient pour lequel beaucoup d'images sont disponibles ne va pas pouvoir être mis dans l'ensemble de validation, car on risquerait alors de ne pas avoir assez d'images pour l'entraînement. En effet, si on met une des images du patient dans l'ensemble de validation, toutes ses autres images doivent y être également. De plus, cette redondance des images limite la capacité de généralisation du réseau de neurones. Le tableau 5.1 indique également que la base de données de SA est plus diverse que celle de HM.

Base de données	**Nombre de patients**	**Nombre d'images**	**Entraînement**	**Validation**	**Test**
Saint-Antoine	36	48	31	6	11
Mondor	15	73	50	13	10

Table 5.1 – Répartition du nombre de patients et d'images pour chaque base de données selon les ensembles d'entraînement, de validation et de test.

5.3 Méthode proposée

La méthode consiste à entraîner un réseau U-Net sur les données en variant les fonctions de coût pour étudier leur effet sur les résultats. Un réseau nnU-Net est également entraîné comme méthode de référence car il s'agit d'un U-Net optimisé. Les entraînements sont réalisés sur chaque base de données individuellement.

5.3.1 Méthode de référence

Pour comparer nos résultats, nous utilisons une méthode de référence qui est nnU-Net [Isensee et al., 2021]. Cette méthode est décrite dans la section 2.3.1. La fonction de coût employée est la fonction de Dice combinée à l'entropie croisée. Pour utiliser nnU-Net, il est nécessaire de mettre les données au format prescrit. Des plis pour la validation croisée sont automatiquement créés mais il a fallu les modifier manuellement car un même patient peut avoir plusieurs images dans le jeu de données. Par conséquent, le partage des ensembles d'entraînement et de test ne peut se faire aléatoirement sans risque de contaminer l'ensemble de test. Après cette étape, nnU-Net pré-traite les données. L'entraînement d'un réseau U-Net 2D est effectué pour chaque pli de la validation croisée. Enfin, le modèle (i.e. les poids après entraînement) infère les résultats sur l'ensemble de test. Les entraînements ont été faits sur le cluster *Convergence*[1] muni de cartes graphiques NVIDIA A100 de 80Go.

5.3.2 Fonction de coût combinant contour actif géodésique et filtre de Frangi

Nous proposons d'ajouter un terme dans la fonction de coût qui minimise l'énergie de la courbe à partir d'une métrique issue de l'image. Nous remplaçons le terme g_I par un opérateur calculant la réponse du filtre de Frangi notée F pour plusieurs échelles σ :

$$\mathcal{L} = \mathcal{L}_{Dice} + \omega \mathcal{L}_{FGAC} \tag{5.26}$$

$$\mathcal{L}_{FGAC} = E(u) = \int_{\Omega} W_I |\nabla u| ds \tag{5.27}$$

avec u la prédiction, et

$$W_I = 1 - \frac{F(H(\sum_i (G_{\sigma_i} * I)))}{F_{max}} \tag{5.28}$$

En pratique nous utilisons l'implémentation en Pytorch[2] de la réponse au filtre de Frangi développée par Sirazitdinov et al. [2022] qu'ils utilisent dans une fonction de coût dans un réseau antagoniste génératif.

5.3.3 Paramètres de l'expérience

Pour chaque expérience, seule la fonction de coût varie, dans l'ensemble présenté et discuté dans la section 5.1. L'architecture du réseau et les autres paramètres d'entraînement sont fixes.

Architecture du réseau Le réseau choisi est un réseau U-Net implémenté dans l'environnement MONAI [Cardoso et al., 2022]. Ce choix est fait car MONAI est un environnement dont le code source est partagé. Il implémente les structures essentielles en apprentissage par réseau de neurones et est spécialisé pour l'imagerie médicale. L'architecture du réseau U-Net la plus récente, est décrite par Kerfoot [2019] et utilise des connexions résiduelles.

1. Cluster du LIP6
2. https://github.com/ilyas-sid/SoftFrangiFilter2D

Le réseau prend en entrée des imagettes en niveaux de gris de taille 512×512 et l'imagette annotée correspondante. Il prédit ensuite un masque qui va être comparé à la référence.

La partie descendante du réseau a cinq niveaux. Le premier extrait 64 filtres caractéristiques de l'image et le plus profond en extrait 1024. Enfin la dernière couche possède une activation de type sigmoïde avec un seuil de 0,1 pour détecter plus de détails. La profondeur choisie est un compromis, un réseau trop profond perdrait l'information du fil guide et un réseau pas assez profond n'apprendrait pas.

Entraînement Les paramètres suivants sont fixés pour toutes les expériences. La taille de lot est de huit, ce qui correspond à un compromis entre le temps d'exécution et le bruit du gradient. La méthode d'optimisation est la méthode Adam [Kingma and Ba, 2014]. Enfin le taux d'apprentissage vaut 10^{-4}. Les temps d'entraînement sont relevés pour une carte graphique NVIDIA RTX A3000 munie de CUDA 12.0.

Augmentation des données L'augmentation des données est appliquée au cours de l'entraînement. Une étude approfondie (annexe C) a été menée pour trouver les transformations pertinentes. Ces transformations ont été modifiées par rapport à l'article publié à la conférence ISBI sur lequel s'appuie ce chapitre. Par conséquent les résultats sont différents de ceux de l'article. Nous appliquons d'abord une mise à l'échelle des intensités entre 0 et 1 sur tout les images, puis les transformations suivantes :

- un retournement (effet miroir),
- une rotation,
- un agrandissement,
- un recadrage aléatoire,
- une transformation affine aléatoire.

Transfert d'apprentissage Nous avons d'abord entraîné les réseaux sur chaque base de données avec des poids initiaux aléatoires. Puis nous avons transféré les poids de l'entraînement réalisé sur une base de données à une autre base de données pour réduire le temps d'apprentissage. Les résultats se sont avérés concluants lors de l'utilisation des poids du modèle entraîné sur la base SA comme poids initiaux du modèle HM. Ce transfert d'apprentissage permet de réduire le nombre d'époques à 200 au lieu de 500.

Paramètres des fonctions de coût Les fonctions de coût détaillées dans la section 5.1 ont des paramètres différents, qui doivent être fixés pour les expériences. Nous avons ainsi fixé :

- Les paramètres α et β de la fonction de coût de Tversky à 0,1 et 0,9 respectivement.
- Le paramètre γ de la fonction Focal à 2. La fonction Focal est combinée avec la fonction de Dice avec des poids de 1.
- La fonction de coût de Vesselness est combinée avec la fonction de Dice et pondérée par $w_{Dice} = 1$, $w_{ms} = 0,3$ et $w_{fv} = 0,01$. Nous fixons une valeur de $\sigma_{max=5}$.

- Pour l'expérience avec le ClDice, l'entraînement des 200 premières époques pour SA (respectivement 100 pour HM) est réalisé avec la fonction de Dice seule. Au-delà, l'entraînement se fait avec les deux fonctions de coût pondérées par 0,5. Le nombre d'itérations dans la fonction de ClDice est fixé à 1 à cause de l'épaisseur de fil guide. Le paramètre de lissage vaut également 1.

- Pour l'expérience avec FGAC nous procédons comme pour ClDice concernant les époques. Le terme de Dice et de FGAC sont additionnés avec un coefficient 1. Pour le calcul de la réponse du filtre de Frangi, nous choisissons $\sigma = [0,5; 1; 2; 3]$.

Post-traitement Les prédictions du réseau de neurones présentent une quantité élevée de faux positifs éloignés du fil guide. Nous proposons un post-traitement simple, avec des opérations de morphologie mathématique. Nous appliquons une reconstruction morphologique après une ouverture binaire par un élément structurant de forme diamant et de rayon 4 pour la base de données SA (respectivement 2 pour HM). Nous choisissons ces valeurs limites car au-delà le fil guide disparaît, il fait en effet 9 pixels de largeur pour SA et 5 pour HM.

5.4 Résultats

Pour chaque base de données nous présentons les résultats qualitatifs et quantitatifs avant et après post-traitement.

5.4.1 Saint-Antoine

Résultats avant post-traitement

Nous évaluons les modèles obtenus pour chaque pli de la validation croisée sur un ensemble de test. Pour chaque modèle nous moyennons les résultats obtenus sur chaque image du test, ce qui nous donne une moyenne par modèle pour chaque méthode d'évaluation.

Puis les résultats obtenus pour chaque critère sur l'ensemble de test sont agrégés par moyenne pour les trois plis de validation. Nous estimons ainsi un résultat moyen pour chaque fonction de coût. Nous reportons ces résultats par fonction de coût avant post-traitement dans le tableau 5.2.

Nous observons d'abord dans le tableau 5.2 que la performance de nnU-Net est très supérieure pour toutes les mesures réalisées. Cependant entraîner un réseau nnU-Net sur trois plis a duré 33 heures sur une GPU plus performante que pour l'entraînement d'un U-Net simple qui ne dépasse pas deux heures, pour des performances convenables. Nous remarquons également que les valeurs de distance de Hausdorff sont particulièrement élevées (tableau 5.2). En effet, les méthodes hors nnU-Net génèrent beaucoup de faux positifs, visibles sur les figures 5.2 et 5.3. Notons que les FP ne sont pas aléatoires mais correspondent souvent à des structures fines sur l'image, comme le bord, les vertèbres ou les voies biliaires. Nous appliquons un post-traitement pour les éliminer en tentant de conserver les parties du fil guide bien détectées.

Method	Dice	ClDice	Tversky	MCC	HD	$Diff_{cc}$
nnU-Net	**78** $\pm$**4**	**94** $\pm$**5**	**74** $\pm$**5**	**89** $\pm$**2**	**25** $\pm$**28**	**1**
Dice / Dice + CE	62,7 $\pm 2,8$	$70,0 \pm 3,8$	$61,97 \pm 1,9$	$81,8 \pm 0,01$	470 $\pm$5	14 $\pm$2
Tversky	$66,6 \pm 2,7$	$72,9 \pm 3,4$	**67,9** $\pm$**2,0**	$83,6 \pm 0,01$	533 $\pm$98	19 $\pm$10
Dice + Focal	$61,6 \pm 2,9$	$67,7 \pm 4,1$	$61,6 \pm 1,3$	$81,2 \pm 0,01$	479 $\pm$66	18 $\pm$6
Dice + Vessel (eq 5.13)	$62,5 \pm 2,9$	$69,2 \pm 3,7$	$62,2 \pm 2,3$	$81,7 \pm 0,01$	461 $\pm$59	13 $\pm$3
ClDice	$63,2 \pm 4,4$	$69,2 \pm 6,5$	$62,6 \pm 3,3$	$81,9 \pm 2,2$	**411** $\pm$**79**	13 $\pm$7
Dice + FGAC	**66,8** $\pm$**2,3**	**74** $\pm$**3,2**	$65,6 \pm 1,2$	**83,7** $\pm$**0,01**	438 $\pm$62	**10** $\pm$**3**

Table 5.2 – Résultats quantitatifs pour la base de données SA avant post-traitement. Le tableau présente la moyenne et l'écart-type pour chaque méthode d'évaluation sur l'ensemble des plis de validation croisée sur l'ensemble de test des 11 images. Les valeurs en gras représentent les meilleures valeurs pour chaque score.

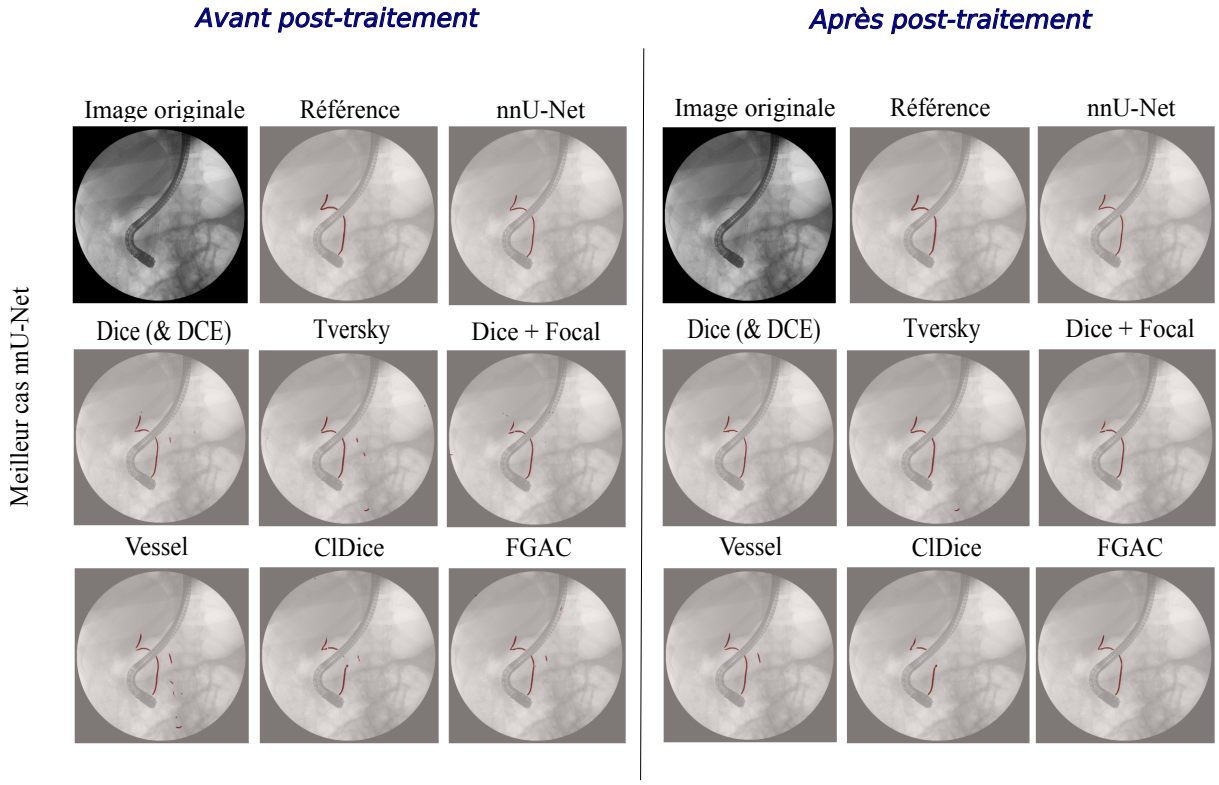


Figure 5.2 – Comparaison qualitative des prédictions en fonction des différentes fonctions de coût sur une image de test de la base de données SA obtenant le meilleur score de Dice avec nnUNet avant et après post-traitement. La référence et les prédictions sont superposées à l'image originale et affichées en rouge.

Résultats après post-traitement

Le post-traitement permet une amélioration globale des divers scores en réduisant le nombre de faux positifs (table 5.3). Cela est observable sur la nette réduction de la distance de Hausdorff mais aussi du nombre de composantes connexes. Le ClDice montre que nous réussissons à conserver la structure du fil guide dans la prédiction. Le score de Tversky est calculé pour des paramètres $\alpha = 0,3$ et $\beta = 0,7$, ce qui signifie que nous pondérons d'avantage les faux négatifs au dénominateur. Par conséquent plus le nombre

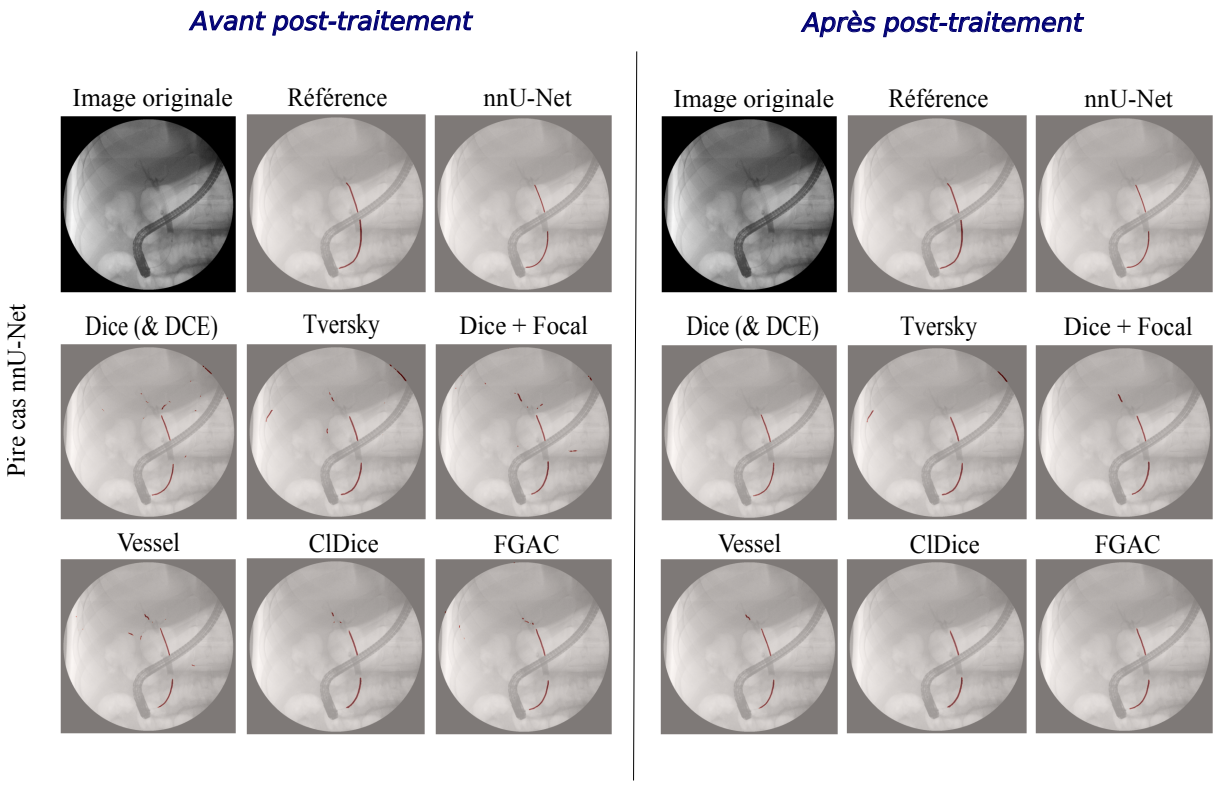


FIGURE 5.3 – Comparaison qualitative des prédictions en fonction des différentes fonctions de coût sur une image de test de la base de données SA obtenant le pire Dice Score avec nn-UNet avant et après post-traitement.

de FN est faible et plus l'indice de Tversky est élevé. Nous en déduisons que nous arrivons à détecter un fil guide assez complet. Sur la figure 5.3, qui illustre la pire prédiction dans l'échantillon de test pour la méthode nnU-Net, nous observons que la segmentation du fil guide s'interrompt à l'endroit où il y a du produit de contraste. Or le produit de contraste est de la même intensité que le fil guide. Même un filtre de *vesselness* ne pourrait pas le rehausser dans ce cas. Le réseau prédit les objets fins et contrastés mais ne parvient pas à incorporer l'information de continuité du fil. Le réseau a également du mal à prédire les pixels du fil à la sortie de l'endoscope pour certains patients. Enfin, nous pouvons mentionner le MCC normalisé qui est supérieur à 80% dans tous les cas. Le MCC évalue chaque élément de la matrice de confusion et prend en compte les FP et les FN au numérateur contrairement à la justesse. Ce qui donne un indice sur la qualité de la segmentation.

Nous retenons que les fonctions de coût de Tversky et FGAC sont celles qui obtiennent les meilleurs scores. De plus les écart-types des scores pour la fonction FGAC sont plus faibles que ceux pour la fonction de Tversky mis à part pour la distance de Hausdorff (figure 5.4). Nous en déduisons que la méthode est plus stable sur les trois plis de validation croisée.

5.4.2 Henri-Mondor

Résultats avant post-traitement

Qualitativement les résultats sur la base de données HM semblent meilleurs que ceux sur la base SA, sur plusieurs aspects : le fil détecté est plus lisse et plus continu, il y a moins

Method	Dice	ClDice	Tversky	MCC	HD	$Diff_{cc}$
nnU-Net	**78** ± 4	**94** ± 5	**74** ± 5	**89** ± 2	**25** ± 28	**1**
Dice / Dice + CE	$65,1$ $\pm 2,6$	77 $\pm 2,4$	$60,3$ $\pm 3,0$	$83,7$ $\pm 0,01$	178 ± 49	$\mathbf{0,7}$ $\pm 0,1$
Tversky	$\mathbf{71,1}$ $\pm \mathbf{2,5}^*$	$80,9$ $\pm 2,9$	$\mathbf{70,6}$ $\pm \mathbf{1,7}$	$\mathbf{85,9}$ $\pm \mathbf{0,01}$	214 ± 18	$1,9$ $\pm 0,7$
Dice + Focal	$65,1$ $\pm 3,5$	$74,7$ $\pm 4,5$	$63,5$ $\pm 1,7$	$83,0$ $\pm 0,02$	245 ± 74	3 $\pm 1,3$
Dice + Vessel	$68,9$ $\pm 3,6$	$80,6$ $\pm 4,1$	$65,0$ $\pm 3,0$	$85,3$ $\pm 0,02$	$\mathbf{131}$ $\pm \mathbf{36}$	$0,9$ $\pm 0,4$
ClDice	$65,9$ $\pm 3,7$	$74,7$ $\pm 4,6$	$63,9$ $\pm 2,9$	$83,3$ $\pm 1,9$	199 ± 74	$2,5$ $\pm 1,3$
Dice + FGAC	$70,9$ $\pm 0,8^*$	$\mathbf{81,9}$ $\pm \mathbf{0,9}^*$	$67,8$ $\pm 0,2$	$\mathbf{85,9}$ $\pm \mathbf{0,01}$	145 ± 36	$1,4$ $\pm 0,3$

Table 5.3 – Résultats quantitatifs pour la base de données SA après post-traitement. Le tableau présente la moyenne et l'écart-type pour chaque méthode d'évaluation sur l'ensemble des plis de validation croisée sur l'ensemble de test des 11 images. Les valeurs en gras représentent la meilleure valeur de la colonne pour chaque score et les symboles $*$ en exposant signifient que la valeur est significativement différente de celle obtenue avec la fonction de coût de Dice avec un test de Student.

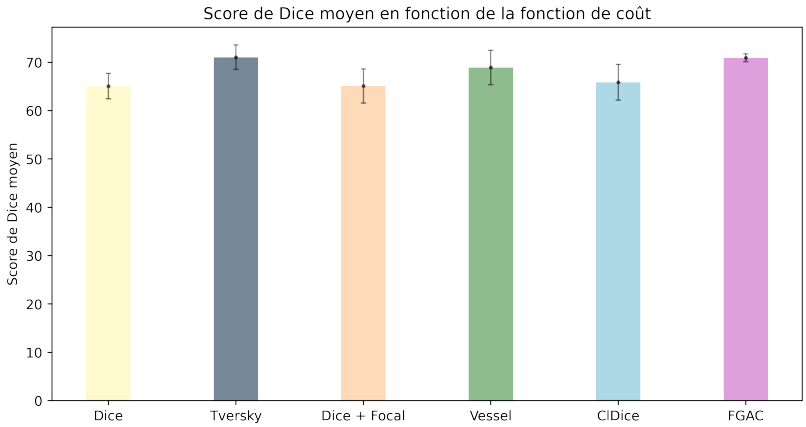


Figure 5.4 – Score de Dice moyen après post-traitement sur la base de données SA.

de faux positifs. Cependant la première image de l'ensemble de test abaisse la performance globale du modèle, sauf pour nnU-Net. Nous cherchons alors à déterminer pourquoi cette image (figure 5.5) représente un cas plus difficile. L'hypothèse est d'abord que le fil est à peine sorti de l'endoscope, il est donc plus court que sur les images d'entraînement. La seconde est le manque de contraste dû à la présence de produit de contraste sur une grande surface du fil guide présent et la présence d'un cathéter.

D'un point de vue quantitatif, dans le tableau 5.4 nous remarquons que les fonctions de coût différentes n'ont pas des scores assez différents pour être significativement mieux par rapport à la fonction de coût de Dice.

Résultats après post-traitement

Le post-traitement permet de nettement réduire les valeurs de la distance de Hausdorff mais au détriment des vrais positifs dans certains cas (table 5.5). Il est néanmoins possible d'appliquer le post-traitement sous contrôle de l'expert médical pour éviter cela. Le meilleur modèle après post-traitement est celui avec la fonction de coût de Vesselness combinée à la fonction de coût de Dice. Les performances sur l'ensemble de données SA

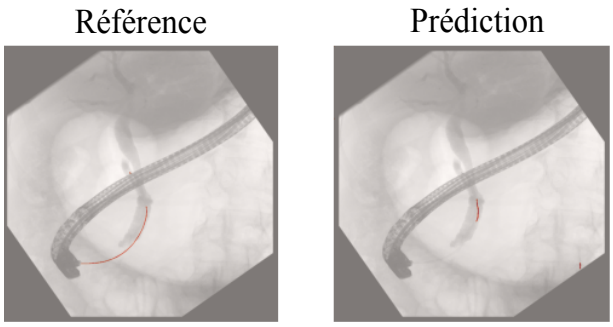


FIGURE 5.5 – Illustration de l'image qui est systématiquement mal segmentée. Nous remarquons qu'il y a effectivement une grande quantité de produit de contraste et que le fil n'est pas déployé à la même longueur que sur les autres images.

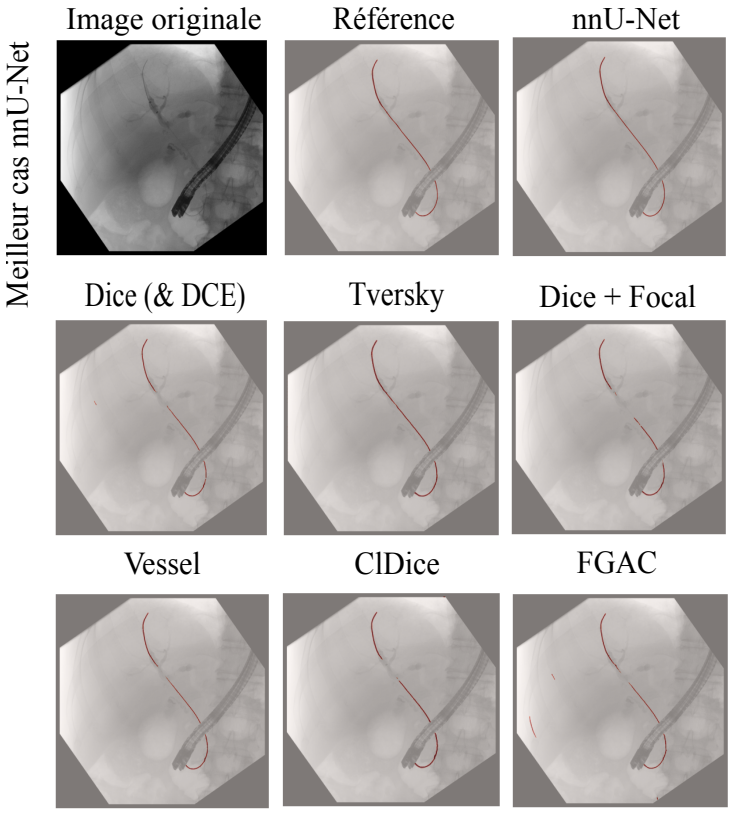


FIGURE 5.6 – Comparaison qualitative des prédictions en fonction des différentes fonctions de coût sur une image de test de la base de données HM obtenant le meilleur score de Dice avec nn-UNet.

sont meilleures que sur la base de données HM notamment à cause de la première image qui abaisse le score global, car on s'attendrait à de meilleurs résultats sur la base de données HM.

Coût temporel d'exécution

Le tableau 5.6 (respectivement 5.7) indique le temps d'entraînement et d'inférence pour chaque méthode expérimentée sur la base de donnée SA (respectivement HM).

Pour la base de données SA, nnU-Net prend 30 fois plus de temps à être entraîné que pour un U-Net de MONAI avec les fonctions de Dice, Tversky et Focal et 15 fois plus

Method	Dice	ClDice	Tversky	MCC	HD	$Diff_{cc}$
nnU-Net	**79** $\pm$**5**	**90** $\pm$**5**	**79** $\pm$**5**	**89** $\pm$**3**	**92** $\pm$**215**	**2**
Dice / Dice + CE	$64,5$ $\pm 0,7$	$73,2$ $\pm 0,7$	$63,5$ $\pm 0,4$	$82,3$ $\pm 0,3$	366 $\pm$97	$8,6$ $\pm$3
Tversky	$60,0$ $\pm 3,3$	$70,9$ $\pm 3,8$	$63,8$ $\pm 2,1$	$80,5$ $\pm 1,4$	432 $\pm$49	$12,5$ $\pm$3
Dice + Focal	$59,6$ $\pm 2,5$	$68,6$ $\pm 2,1$	$60,5$ $\pm 2,3$	$80,0$ $\pm 1,2$	458 $\pm$30	$8,8$ $\pm$2
Dice + Vessel	$\mathbf{65,1}$ $\pm \mathbf{0,5}$	$73,2$ $\pm 0,8$	$\mathbf{64,7}$ $\pm \mathbf{0,4}$	$\mathbf{82,6}$ $\pm \mathbf{0,2}$	396 $\pm$42	$11,6$ $\pm 2,2$
ClDice	$60,6$ $\pm 1,2$	$\mathbf{76,0}$ $\pm \mathbf{2,5}$	$62,3$ $\pm 1,3$	$80,4$ $\pm 0,6$	437 $\pm$20	$8,2$ $\pm 1,9$
Dice + FGAC	$62,0$ $\pm 0,9$	$70,2$ $\pm 0,7$	$61,4$ $\pm 1,7$	$81,1$ $\pm 0,36$	402 $\pm$26	**6** $\pm$**3,5**

Table 5.4 – Résultats quantitatifs pour la base de données HM avant post-traitement. Le tableau présente la moyenne et l'écart-type pour chaque méthode d'évaluation sur l'ensemble des plis de validation croisée sur l'ensemble de test des 10 images.

Method	Dice	ClDice	Tversky	MCC	HD	$Diff_{cc}$
nnU-Net	**79** $\pm$**5**	**90** $\pm$**5**	**79** $\pm$**5**	**89** $\pm$**3**	**92** $\pm$**215**	**2**
Dice / Dice + CE	$60,9$ $\pm 2,9$	$70,9$ $\pm 2,3$	$58,3$ $\pm 3,6$	$80,8$ $\pm 1,6$	182 $\pm$61	$1,6$ $\pm 0,4$
Tversky	$60,8$ $\pm 2,8$	$72,6$ $\pm 3,1$	$\mathbf{64,2}$ $\pm \mathbf{3,6}$	$80,7$ $\pm 1,2$	257 $\pm$152	$3,2$ $\pm 0,8$
Dice + Focal	$59,4$ $\pm 2,8$	$71,4$ $\pm 5,4$	$60,1$ $\pm 2,6$	$79,9$ $\pm 1,3$	227 $\pm$40	$3,1$ $\pm 0,3$
Dice + Vessel	$\mathbf{64,3}$ $\pm \mathbf{0,3}$	$75,5$ $\pm 3,4$	$63,1$ $\pm 0,3$	$\mathbf{82,3}$ $\pm \mathbf{0,2}$	**170** $\pm$**44**	$\mathbf{1,5}$ $\pm \mathbf{0,5}$
ClDice	$60,7$ $\pm 1,2$	$\mathbf{76,4}$ $\pm \mathbf{2,4}$	$62,1$ $\pm 1,3$	$80,4$ $\pm 0,6$	347 $\pm$38	$4,5$ $\pm 1,3$
Dice + FGAC	$61,6$ $\pm 2,1$	$75,6$ $\pm 2,5$	$60,3$ $\pm 2,8$	$80,8$ $\pm 1,0$	219 $\pm$51	$3,5$ $\pm 1,9$

Table 5.5 – Résultats quantitatifs pour la base de données HM après post-traitement. Le tableau présente la moyenne et l'écart-type pour chaque méthode d'évaluation sur l'ensemble des plis de validation croisée sur l'ensemble de test des 10 images.

de temps que la fonction de coût *Vesselness*, le tout pour des performances proches. De même en inférence nnU-Net met 9 fois plus de temps que les méthodes avec U-Net de MONAI. Notons que nous n'avons pas compté le temps de post-traitement. Il serait bon de l'ajouter dans des analyses plus approfondies. Cependant le post-traitement effectué ne prend pas beaucoup de temps à être exécuté. De même, pour la base de données HM les temps d'entraînement et d'inférence sont nettement inférieurs à ceux de nnU-Net. Ils sont également inférieurs à ceux obtenus sur la base de données SA car nous avons réduit le nombre d'époques grâce au transfert d'apprentissage.

Nous notons également que certaines fonctions de coût sont plus longues à calculer que d'autres, c'est le cas pour la fonction de coût de *Vesselness*. Cela s'explique car elle prend en compte les valeurs propres et leur ordonnancement. Les fonctions de coût de ClDice et de FGAC se situent au milieu en temps de calcul. Cela s'explique car elles nécessitent des calculs intermédiaires comme le calcul du squelette ou de la valeur du filtre de Frangi pour différentes valeurs de σ.

5.5 Reconstruction par chemins minimaux

Nous avons détaillé l'approche par chemins minimaux dans la section 5.1.8. Nous privilégions cette approche pour tenter de reconstruire le fil guide à partir d'une prédiction incomplète et des intensités dans l'image originale. Nous avons commencé par appliquer la méthode semi-automatique là où la prédiction s'interrompt. Le résultat dépend d'un paramètre ϵ qui régularise le chemin.

Dans un premier temps nous avons expérimenté sur quelles distances et avec quel ϵ

Méthode	Nombre d'époques	Temps d'entraînement (h)	Temps d'inférence (s)
nnU-Net	1000	33	5,6
Dice / Dice + CE	500	1	0,63
Tversky		1	
Dice + Focal		1	
Dice + Vessel		2,2	
ClDice		1,5	
Dice + FGAC		1,5	

Table 5.6 – Temps d'entraînement sur trois plis et temps d'inférence pour une image sur la base de données SA en fonction de la méthode. Matériel pour nnU-Net : carte graphique NVIDIA A100 de 80Go. Matériel pour MONAI : carte graphique NVIDIA RTX A3000.

Méthode	Nombre d'époques	Temps d'entraînement (h)	Temps d'inférence (s)
nnU-Net	1000	52	9,4
Dice / Dice + CE	200	0,75	0,98
Tversky		0,75	
Dice + Focal		0,75	
Dice + Vessel		1,5	
ClDice		1,5	
Dice + FGAC		1	

Table 5.7 – Temps d'entraînement sur trois plis et temps d'inférence pour une image sur la base de données HM en fonction de la méthode. Matériel pour nnU-Net : carte graphique NVIDIA A100 de 80Go. Matériel pour MONAI : carte graphique NVIDIA RTX A3000.

nous pouvions obtenir des résultats en initialisant manuellement les points de départ et d'arrivée (figure 5.7). Sur cet exemple la méthode fonctionne pour toutes les valeurs du paramètre ϵ, qui sont très faibles car c'est la réponse du filtre de Frangi qui a été utilisée comme potentiel, et pour différentes distances. Cependant, il faut être très précis lors de l'initialisation des points, car un léger décalage peut compromettre le résultat.

Notons que l'utilisation des chemins minimaux est particulièrement utile pour les segments du fil guide pris en compte dans l'algorithme de localisation.

Dans un deuxième temps nous avons cherché à reconstruire le fil guide en initialisant les points au niveau des extrémités des composantes connexes ordonnées de la prédiction. Puis nous avons appliqué la méthode pour les extrémités deux à deux. Nous obtenons ainsi le fil guide reconstruit illustré par la figure 5.8, en agrégeant les parties obtenues avec la prédiction incomplète.

Une autre méthode a été décrite par [Cohen and Kimmel, 1997] qui à partir d'un ensemble de composantes connexes trouve les points selles de la carte d'énergie et part de ces points spécifiques comme point d'arrivée pour appliquer la descente de gradient. Malheureusement, nous n'avons pas réussi à implémenter cette méthode, mais elle reste dans nos perspectives d'amélioration de ce travail.

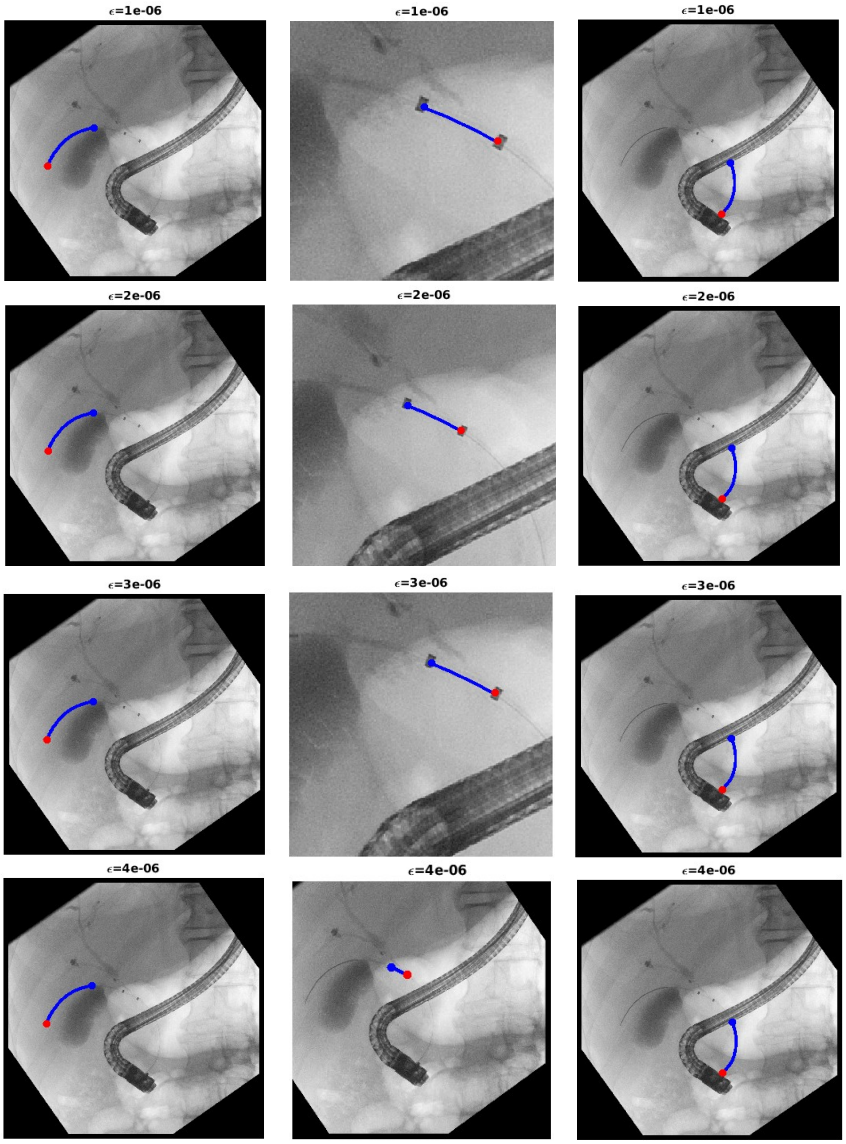


Figure 5.7 – La figure illustre la méthode des chemins minimaux appliquée pour différentes valeurs de ϵ (lignes) et différentes partie du fil guide (colonnes) sur une image de la base de données HM.

5.6 Conclusion

Dans ce chapitre nous avons d'abord comparé les résultats de la segmentation du fil guide avec la méthode nnU-Net et ceux obtenus avec un réseau U-Net plus simple et différentes fonctions de coût et ce sur deux bases de données différentes. Nous avons également proposé une nouvelle fonction de coût qui s'appuie sur les méthodes de contours actifs géodésiques et sur le filtre de Frangi. Cette méthode obtient de meilleurs résultats qu'avec la fonction de Dice, tout comme la fonction de coût de Tversky sur la base de données SA. De plus les faibles écarts-types obtenus pour chaque score indiquent une certaine robustesse de la méthode.

Nous obtenons des résultats proches de ceux de nnU-Net, pour un coût bien moindre en temps de calcul. Nous obtenons sur la base de données SA par exemple, le meilleur score de Dice de $\mathbf{71,1 \pm 2,5}$ avec la fonction de coût de Tversky comparé à $\mathbf{78 \pm 4}$ avec nnU-Net, et ce en entraînant un réseau 33 fois moins de temps. De même pour les autres

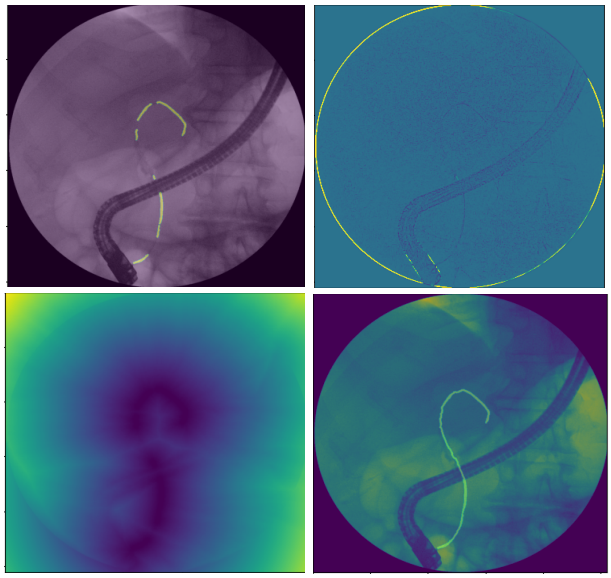


FIGURE 5.8 – Reconstruction du fil guide à partir de la prédiction d'un réseau de neurones avec la méthode des chemins minimaux. De haut en bas et de gauche à droite : les composantes connexes de la prédiction sans les FP, le potentiel calculé à partir de la valeur du filtre de Frangi (les pixels du fil guide sont plus sombres en intensité que ceux du fond), la carte d'énergie calculée à partir des composantes connexes et du potentiel, le fil guide reconstruit morceau par morceau.

critères d'évaluation comme le ClDice, $\mathbf{80,9 \pm 2,9}$ contre $\mathbf{94 \pm 5}$ ou le MCC avec un score de $\mathbf{85,9 \pm 0,01}$ contre $\mathbf{89 \pm 2}$.

Les résultats obtenus sur la base de données HM pour différentes fonctions de coût sont également proches de ceux obtenus avec nnU-Net pour moins de ressources utilisées, notamment grâce au transfert d'apprentissage.

Ces résultats sont aussi à mettre en regard des résultats obtenus dans l'état de l'art par les méthodes de segmentation de cathéter dans le domaine endo-vasculaire. Nous nous comparons aux méthodes qui veulent détecter l'intégralité du fil guide et qui calculent le score de Dice comme Gherardini et al. [2020] qui obtiennent un score de Dice de 55 en entraînant le réseau sur 9000 images.

Nous remarquons cependant que certaines parties importantes peuvent ne pas être segmentées comme le démarrage du fil guide à sa sortie de l'endoscope notamment pour la base de données SA.

Nous avons alors proposé d'utiliser la reconstruction par chemins minimaux pour compléter cette segmentation. La reconstruction donne de bons résultats sur les images testées mais il faudrait l'appliquer à la totalité de l'ensemble de test pour confirmer ces résultats.

Pour conclure, nous avons montré qu'il était possible de segmenter le fil guide sur deux bases de données contenant peu d'images avec des résultats au-dessus de l'état de l'art et proches de ceux obtenus avec nnU-Net avec des méthodes exploitant moins de ressources de calcul. Nous nous intéressons maintenant à l'apport de la segmentation de l'endoscope pour l'amélioration de celle du fil guide. Ces travaux sont traités dans le chapitre 6.

CHAPITRE 6

Apport de la segmentation de l'endoscope

Dans ce chapitre nous nous intéressons à l'endoscope. Plus particulièrement nous voulons savoir si la connaissance de sa position dans l'image peut nous aider à améliorer la segmentation du fil guide. En effet, le fil guide est toujours connexe à l'endoscope car il en sort. De plus, le fil guide passe quasi-systématiquement derrière l'endoscope sur les images. Nous faisons l'hypothèse que l'endoscope est également plus aisé à segmenter du fait de sa surface plus étendue et de son plus grand contraste avec le reste de l'image.

Nous décrivons d'abord la segmentation de l'endoscope seul avec un réseau U-Net pour chacune des bases de données décrites précédemment. Nous déterminons ensuite si un réseau de neurones est capable de modéliser la relation spatiale entre l'endoscope et le fil guide pour mieux segmenter ce dernier. Nous appliquons d'abord un U-Net multi-objets pour segmenter l'endoscope et le fil guide sur la base de données SA. Puis nous appliquons la méthode IterNet [Li et al., 2020] sur les même données. Cette partie a bénéficié du stage de Yuxuan Xi que j'ai encadré [Xi, 2024].

6.1 Segmentation de l'endoscope seul

6.1.1 Données

Pour les données nous réunissons les images contenant l'endoscope uniquement et l'endoscope avec le fil guide (codes 2 et 3 respectivement décrits dans le chapitre 3), ce qui augmente le nombre d'images par rapport à l'expérience décrite dans le chapitre 5. Nous avons à présent 90 images pour la base de données SA et 94 pour celle de HM.

6.1.2 Méthode

La méthode consiste à entraîner un réseau U-Net sur les données, avec une fonction de coût usuelle qui est la combinaison de la fonction de Dice et de l'entropie croisée.

Architecture du réseau Nous choisissons la même architecture que dans le chapitre précédent. Le réseau prend en entrée des imagettes en niveaux de gris de taille 512×512 et l'imagette annotée correspondante. La partie descendante du réseau a cinq niveaux. Le

premier extrait 64 filtres caractéristiques de l'image et le plus profond en extrait 1024. Enfin la dernière couche possède une activation de type sigmoïde avec un seuil de $0,5$.

Entraînement Les paramètres suivants sont fixés pour toutes les expériences. La taille de lot est de quatre. La méthode d'optimisation est la méthode Adam [Kingma and Ba, 2014] avec un taux d'apprentissage valant 10^{-3}. Les temps d'entraînement sont relevés pour une carte graphique NVIDIA RTX A3000 munie de CUDA 12.0. Nous réalisons une validation croisée sur les trois plis d'entraînement et de validation.

Augmentation des données Nous augmentons les données artificiellement pour cet entraînement en appliquant :

- un retournement (effet miroir),
- une rotation,
- un agrandissement,
- un recadrage aléatoire,
- une transformation affine aléatoire.

Transfert d'apprentissage Pour l'entraînement de la base de données HM nous utilisons les poids du modèle entraîné sur la base de données SA comme poids initiaux.

Post-traitement Pour la base de données de SA, le réseau segmente l'endoscope et détecte le contour du cercle dans lequel est comprise la fluoroscopie. Pour s'en défaire nous gardons uniquement la plus grande composante connexe.

Pour la base de données de HM nous gardons la plus grande composante connexe et appliquons une fermeture morphologique pour réduire certains trous dans la structure.

6.1.3 Résultats

Nous divisons les images en un ensemble de test et un ensemble d'entraînement (tableau 6.1). L'ensemble d'entraînement est ensuite séparé en trois plis de validation croisée, contenant chacun un ensemble d'entraînement et un de validation.

Base de données	Entraînement	Validation	Test	Total
Saint-Antoine	56	21	13	90
Mondor	51	23	20	94

Table 6.1 – Nombre d'images selon les centres et répartition dans les ensembles d'entraînement, de validation et de test.

Les mesures de Dice et de distance de Hausdorff présentées dans le tableau 6.2 montrent que même avec une base de données limitée, l'endoscope est bien segmenté par le réseau U-Net et le post-traitement. Les résultats sur les images de la base HM sont meilleurs grâce au transfert d'apprentissage. Le post-traitement permet de réduire le nombre de faux positifs, ce qui se ressent sur la valeur de la distance de Hausdorff. Cependant cette

dernière reste assez élevée pour la base de données SA, bien qu'assez stable (écart-type de 4,5 pixels) en comparaison avec la base de données HM dont la distance de Hausdorff est plus faible mais avec un écart-type de 30 pixels.

Base de données	**Dice**	**HD** (px)	**Temps d'entraînement (min)**	**Temps d'inférence (s)**
Saint-Antoine	$90,95 \pm 0,38$	$101 \pm 4,5$	40	7
Mondor	$97,73 \pm 0,17$	40 ± 30	35	7

Table 6.2 – Scores moyens obtenus par base de données sur les trois plis de la validation croisée, après post-traitement. Le temps d'entraînement est la durée moyenne par pli de validation croisée.

La figure 6.1 montre que le post-traitement est très efficace si l'endoscope est détecté comme une seule composante connexe, ce qui arrive dans 11 cas sur 13 dans l'ensemble de test de la base de données SA. La deuxième ligne de la figure montre que dans le cas inverse, le post-traitement élimine la moitié de l'endoscope car nous ne conservons que la plus grande composante connexe. Cette erreur de segmentation peut être expliquée par un contraste plus faible entre l'endoscope et la colonne vertébrale qu'entre l'estomac et l'endoscope. Ici, la partie pertinente, c'est-à-dire l'extrémité de l'endoscope, est conservée mais le cas inverse aurait pu se produire en fonction des changements de contraste dans le reste de l'image.

Une sélection manuelle des composantes connexes à conserver permettrait de résoudre ce problème, au prix d'une action minime de l'utilisateur.

Pour les images de la base de données HM, la figure 6.2 montre que l'endoscope est globalement bien détecté (première ligne), mais il arrive que l'extrémité le soit moins (deuxième ligne). Ainsi malgré un score de Dice plus élevé en moyenne pour les images de la base HM, l'extrémité est mieux segmentée dans plus de cas de la base de données SA (13 cas sur 13 comparé à 18 cas sur 20 pour HM).

6.1.4 Conclusion

Dans cette partie, nous avons proposé une méthode de segmentation de l'endoscope à partir d'un réseau de neurones classique, auquel nous avons ajouté des transformations pour l'augmentation des données et un post-traitement appropriés. Les résultats montrent que la tâche de segmentation de l'endoscope est plus aisée que celle de segmentation du fil guide. Avec peu de données, nous parvenons à avoir une segmentation obtenant un score de Dice de $90,95\%$ pour la base de données SA et $97,73\%$ pour la base de données HM. Ces résultats sont proches de ceux de Huang et al. [2021] qui obtenaient une intersection sur l'union de $95,85\%$ en s'appuyant sur 1381 images d'endoscope. Les auteurs n'indiquent malheureusement pas l'écart-type obtenu. Nous montrons ainsi qu'il n'y a sûrement pas besoin d'autant d'images d'entraînement pour cette tâche. Cependant, ajouter quelques images à notre ensemble d'entraînement permettrait sans doute d'améliorer les résultats.

6.2 Segmentation multi-objets

Dans cette section nous nous intéressons au problème de segmentation conjointe de l'endoscope et du fil guide. Nous voulons tester l'hypothèse qu'un réseau de neurones exploite implicitement la relation spatiale entre le fil guide et l'endoscope pour mieux

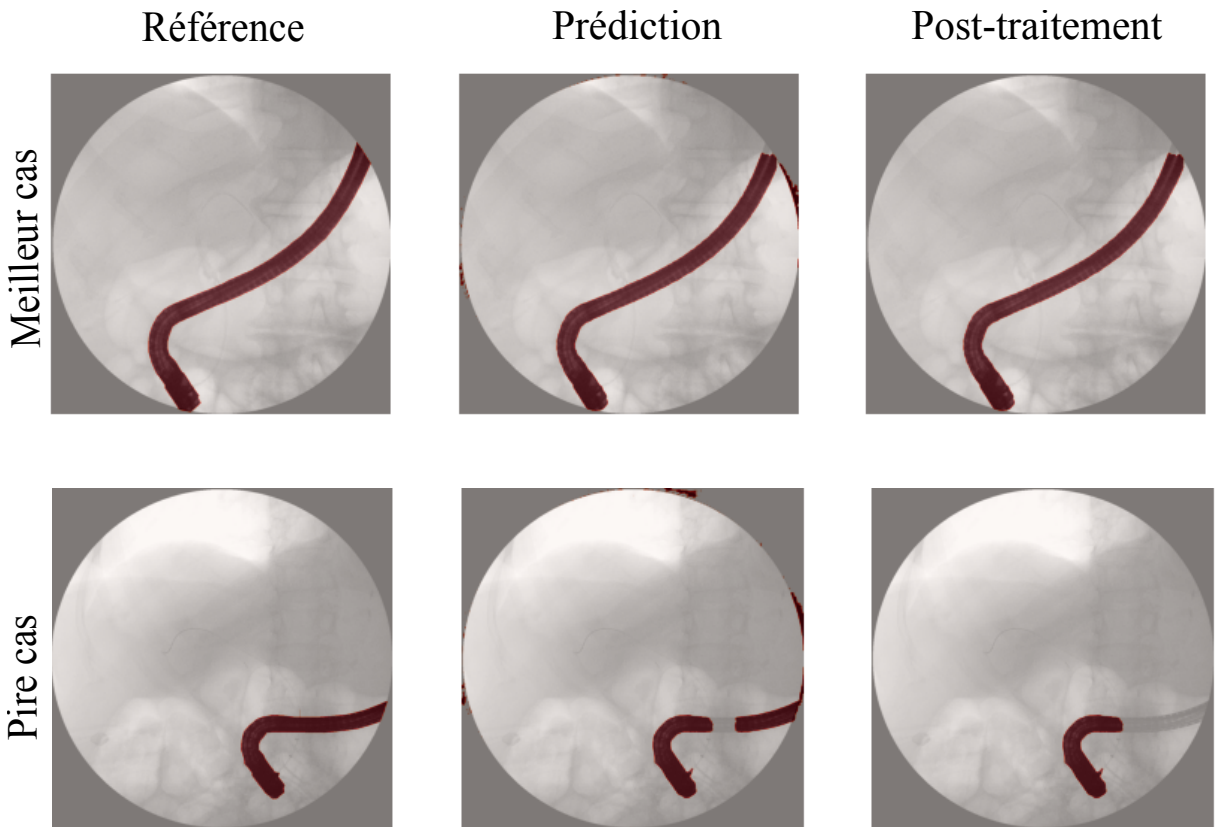


FIGURE 6.1 – Résultats qualitatifs de segmentation de l'endoscope sur la base de données SA. La première colonne présente la segmentation de référence sur l'image originale, la deuxième, la prédiction en sortie de réseau U-Net et la troisième la prédiction après post-traitement. La première ligne présente le meilleur cas dans l'ensemble de test et la seconde le pire cas dans l'ensemble de test au sens du score de Dice.

segmenter le fil guide. Pour cela nous appliquons d'abord un réseau U-Net multi-objets avec deux fonctions de coût, en seconde approche, nous appliquons un réseau Iter-Net spécialisé dans la segmentation de vaisseaux. Plusieurs expériences ont été réalisées au cours du stage de Xi [2024]. Cependant nous ne présenterons ici que celles obtenant les meilleurs résultats. Notons, que nous avons testé la segmentation des deux objets en tant qu'entité unique comme méthode de référence mais comme attendu la surface du fil guide est si négligeable par rapport à celle de l'endoscope que le réseau ne détecte pas bien le fil.

6.2.1 Données

Les expériences n'ont été réalisées que sur la base de données SA avec les images comprenant les deux instruments (n=48). Les données pour les deux expériences qui suivent sont les mêmes que celles du chapitre 5.

6.2.2 U-Net multi-objets

Nous appliquons encore une fois un réseau U-Net, dans lequel seule la couche de sortie du réseau va être modifiée. En effet, celle-ci fournit à présent une prédiction contenant trois canaux, et donc chaque canal contient la probabilité pour chaque pixel d'appartenir

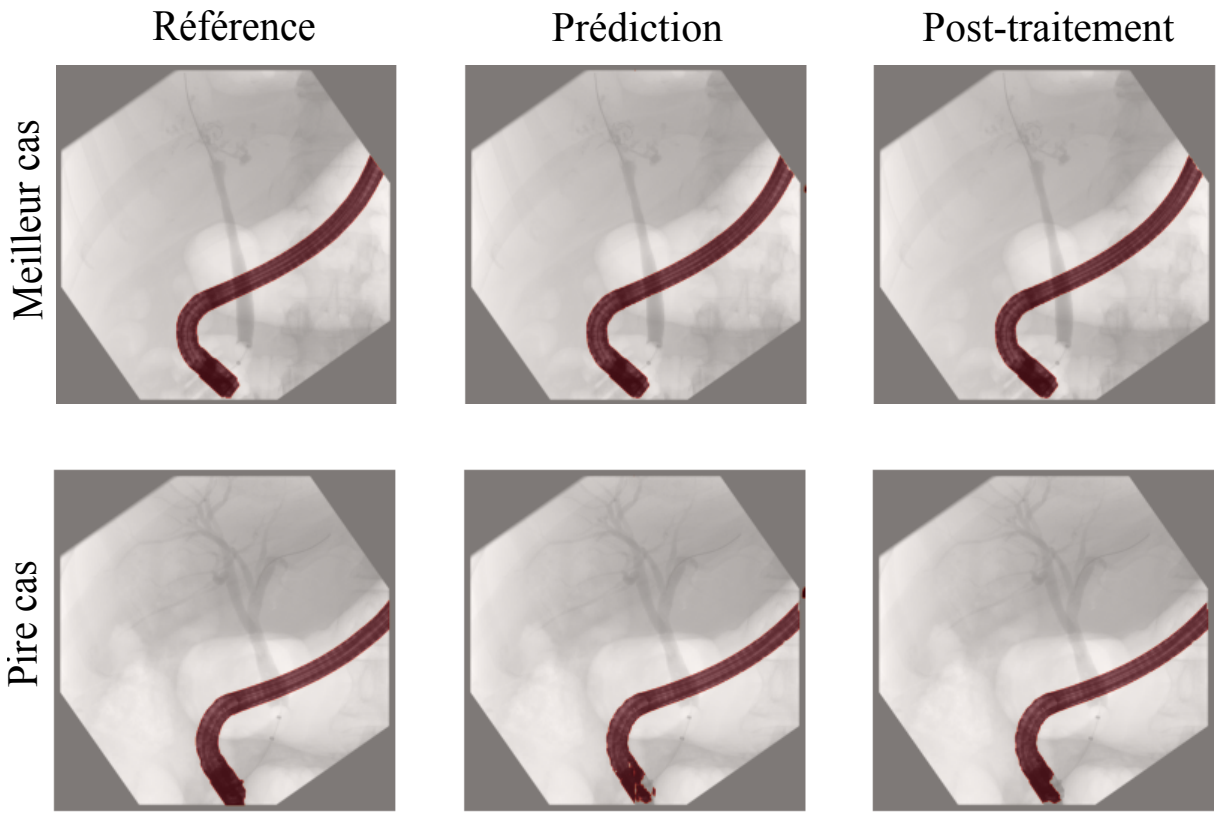


FIGURE 6.2 – Résultats qualitatifs de segmentation de l'endoscope sur la base de données HM. La première colonne présente la segmentation de référence sur l'image originale, la deuxième, la prédiction en sortie de réseau U-Net et la troisième la prédiction après post-traitement. La première ligne présente le meilleur cas dans l'ensemble de test et la seconde le pire cas dans l'ensemble de test au sens du score de Dice.

à un des instruments, l'endoscope, le fil guide ou au reste de l'image. Nous remplaçons la fonction sigmoïde dans la couche de sortie par une fonction *softmax*[1] adaptée au cas multi-objets. Les autres paramètres sont les mêmes que pour la segmentation de l'endoscope seul.

Nous testons ce réseau avec une fonction de coût de Dice combinée à l'entropie croisée et avec une fonction de coût de Tversky (avec $\alpha = 0,3$ et $\beta = 0,7$). Nous choisissons d'appliquer la fonction de Tversky car elle est plus adaptée pour la segmentation du fil guide comme montré à la section 5.4.1. En pondérant les FN par 0,7 nous espérons améliorer la continuité du fil.

6.2.3 IterNet

IterNet [Li et al., 2020] est un réseau de neurones développé pour la segmentation de vaisseaux sur des images de la rétine. Les vaisseaux sur ce type d'image ont des diamètres qui s'affinent au fil des ramifications et sont de moins en moins perceptibles. Or il est important de maintenir la connexité du réseau vasculaire pour simuler et calculer le flux par exemple. Le fil guide ayant une structure similaire aux vaisseaux, nous proposons d'utiliser Iter-Net pour réduire les faux négatifs et améliorer la connexité. La méthode

1. La fonction *softmax* s'écrit $Softmax(x_i) = \frac{exp(x_i)}{\sum_{j=1}^{k} exp(x_j)}$ pour un vecteur x de i-ème composante x_i en entrée et k le nombre de classes. Les valeurs de sortie sont comprises entre 0 et 1.

IterNet est une cascade de réseaux U-Net à la suite. Le premier est un réseau U-Net classique qui applique 5 niveaux de convolution. Il commence par extraire 32 cartes de caractéristiques jusqu'à 512 au plus profond niveau. Ce U-Net est suivi de $(K-1)$ réseaux U-Net légers appelés mini-UNets qui sont moins profonds et n'extraient que 256 cartes de caractéristiques au plus au niveau le plus profond. La particularité d'IterNet est que chaque mini-UNet qui suit prend en entrée la sortie de l'avant dernière couche du réseau précédent concaténée avec les cartes de sortie du premier étage de convolution du premier U-Net (figure 6.3). Le premier U-Net apprend les caractéristiques principales des images et les suivants affinent le résultat. Les poids entre les mini-Unets sont partagés pendant l'entraînement.

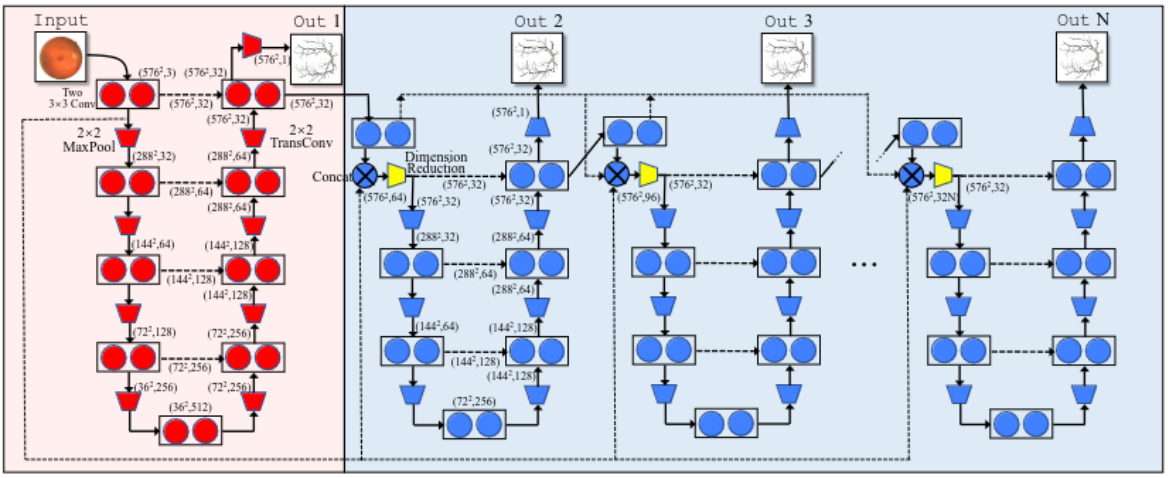


Figure 6.3 – Illustration de la méthode IterNet. Source : Li et al. [2020].

Pour cette expérience nous ne faisons qu'une seule itération, c'est-à-dire que nous utilisons un réseau U-Net suivi d'un seul mini-Unet qui extrait des cartes de caractéristiques allant de 32 à 256.

6.2.4 Résultats

Nous pouvons affirmer d'après le tableau 6.3, que la segmentation de l'endoscope est satisfaisante pour les trois méthodes, car les scores de Dice sont proches de 90%. De plus la distance de Hausdorff est très faible, ce qui est mieux que pour la segmentation de l'endoscope seul (tableau 6.2).

Endoscope	Avant post-traitement		Après Post-traitement	
	Dice	**HD**	**Dice**	**HD**
U-Net (DCE)	$88,99 \pm 0,3$	$\mathbf{428} \pm \mathbf{44}$	$90,3 \pm 0,4$	$18 \pm 1,2$
U-Net (TV)	$\mathbf{89,05} \pm \mathbf{0,19}$	437 ± 44	$\mathbf{90,7} \pm \mathbf{0,16}$	$\mathbf{16} \pm \mathbf{0,9}$
IterNet	$79,54 \pm 7,72$	512 ± 62	$89,64 \pm 0,6$	$44 \pm 4,5$

Table 6.3 – Résultats obtenus par la méthode de segmentation conjointe pour l'endoscope avant et après post-traitement. Les résultats en gras sont les meilleurs par score.

Les scores pour le fil guide (tableau 6.4) sont au contraire très en deçà de ceux obtenus au chapitre 5. Cela se confirme en observant les prédictions sur la figure 6.4.

Fil guide	Après post-traitement					
	Dice	**HD**	**ClDice**	**Tv**	**MCC**	**DiffCC**
U-Net (DCE)	$\mathbf{37,25}$ $\pm\mathbf{0,96}$	$\mathbf{300}$ $\pm\mathbf{36}$	$\mathbf{40,13}$ $\pm\mathbf{1,45}$	$39,64$ $\pm 1,45$	$\mathbf{69,2}$ $\pm\mathbf{0,38}$	$\mathbf{9}$ $\pm\mathbf{3,6}$
U-Net(TV)	$37,56$ $\pm 3,39$	294 ± 41	$41,45$ $\pm 4,58$	$\mathbf{37,75}$ $\pm\mathbf{0,7}$	$69,24$ $\pm 1,56$	8 $\pm 4,1$
IterNet	$34,58$ $\pm 10,02$	359 ± 85	$38,82$ $\pm 12,46$	$37,16$ $\pm 9,5$	$67,87$ ± 5	10 $\pm 5,1$

Table 6.4 – Résultats de la segmentation conjointe pour fil guide avec U-Net et IterNet.

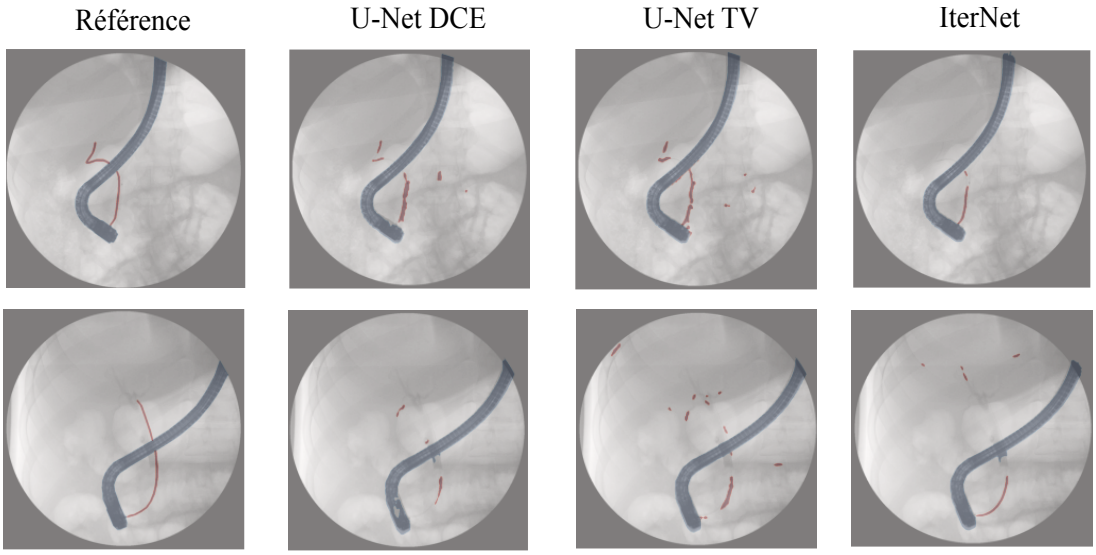


Figure 6.4 – Comparaison des prédictions après post-traitement pour les méthodes U-Net avec la fonction de coût de Dice combinée avec l'entropie croisée, U-Net avec la fonction de coût de Tversky et Iter-Net.

Nous remarquons également en analysant les prédictions d'IterNet que le départ du fil guide est bien segmenté dans tous les cas de l'ensemble de test (figure 6.5). Nous en déduisons que le réseau extrait sûrement une relation entre endoscope et fil guide mais qui ne bénéficie que localement à la segmentation du fil guide. Dans le chapitre 5, le départ du fil guide n'était pas aussi bien détecté sur la base de données SA.

6.3 Discussion

Dans Xi [2024], un modèle moyen est obtenu moyennant les pixels des prédictions obtenues pour chaque pli de validation croisée puis en seuillant le résultat. Les scores d'évaluation sont ainsi calculés sur cette assemblage moyen. Néanmoins, dans le chapitre précédent nous avons moyenné les scores obtenus pour chaque pli de la validation croisée. Nous reprenons la seconde approche pour comparer les résultats avec ceux du chapitre 5. La première approche reste valide. Elle a tendance à donner de meilleurs résultats globalement. Elle pourra être envisagée dans les perspectives car elle apporte plus de robustesse au résultat. En effet, si le réseau détecte le fil guide sur l'image avec les trois modèles issus de la validation croisée, alors on aura tendance à dire qu'il est plus probable que ce pixel appartiennent au fil guide. Au contraire s'il n'est détecté qu'une fois, il est probable que ce soit un faux positif. L'inconvénient majeur est que cette approche allonge sensiblement le temps d'inférence, et prend plus d'espace en mémoire car il faut charger les poids des

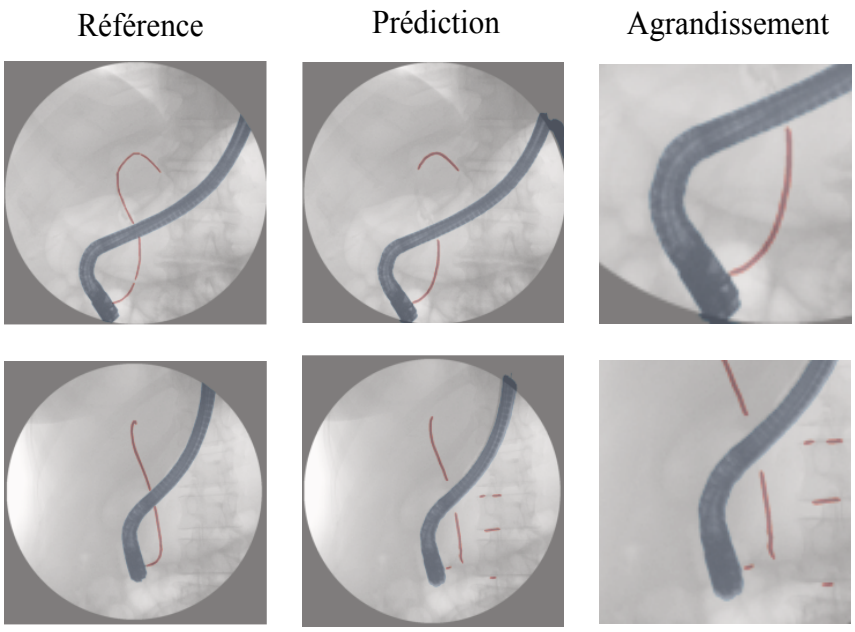


FIGURE 6.5 – Cette figure montre le départ du fil guide pour deux images dans l'ensemble de test avec la méthode IterNet.

trois modèles.

Notons cependant que les résultats pour le fil guide obtenus avec la méthode d'assemblage de modèles et les trois méthodes testées donne des résultats toujours inférieurs à ceux du chapitre 5. Ces résultats mériteraient une étude plus approfondie, avec plus de méthodes d'évaluation et un post-traitement similaire à celui de chapitre 5 car ils présentent beaucoup de faux positifs.

6.4 Conclusion

Nous avons montré dans un premier temps que la segmentation de l'endoscope seul était une tâche quasiment résolue avec un U-Net classique. Ce premier résultat nous permet de faire des mesures sur l'image par exemple, car le diamètre de l'endoscope est fixe. Les autres expériences montrent que la segmentation globale du fil guide n'est pas améliorée par la segmentation de l'endoscope simultanément. Cependant au niveau de la sortie de l'endoscope, le réseau IterNet détecte systématiquement le fil guide. C'est un résultat intéressant car cette partie du fil est utilisée par le logiciel de localisation du fil guide dans les voies biliaires. Nous pouvons ainsi envisager de combiner le résultat de IterNet avec le meilleur résultat obtenu dans le chapitre 5.

CHAPITRE 7

Conclusion et perspectives

7.1 Conclusion

Dans ce travail de thèse nous avons cherché à segmenter les instruments endoscopiques utilisés pour la CPRE sur les images de fluoroscopie 2D afin d'améliorer ce geste complexe. Nous disposions d'une faible quantité d'images, non annotées. Ainsi nous avons cherché à répondre à la question de la segmentation de structures fines et allongées sur une base de données réduite, et comment l'évaluer. Puis nous nous sommes intéressés à la segmentation conjointe de l'endoscope et du fil guide, dans l'optique d'améliorer la segmentation du fil guide.

Nous avons d'abord proposé une synthèse critique des travaux concernant la CPRE et la segmentation de cathéter en imagerie endo-vasculaire. Ainsi nous avons remarqué qu'il n'existait pas de méthode pour la segmentation du fil guide sur les images de CPRE. Nous avons noté également que la plupart des méthodes développées utilisaient l'architecture U-Net couplée à une autre méthode dépendant de l'application et nécessitant des jeux de données conséquents.

Dans le chapitre 3, nous décrivons les données utilisées pour alimenter l'entraînement de nos modèles. Nous montrons par la même occasion que la quantité de données dont nous disposons est très réduite notamment par rapport à celles utilisées dans les méthodes de l'état de l'art.

Dans le chapitre 4 nous avons montré que les méthodes habituelles d'évaluation des modèles de segmentation d'images médicales n'étaient pas suffisantes pour évaluer la segmentation de structures fines et allongées. Ainsi nous avons cherché d'autres mesures plus adaptées pour fournir une analyse plus détaillée des résultats afin d'avoir plus d'informations à propos des modèles. En particulier nous avons souligné l'apport de mesures telles que le ClDice, l'indice de Tversky, le coefficient de corrélation de Matthews et la différence du nombre de composantes connexes pour ce type de structures. Une des conclusions importante de cette analyse est qu'une interprétation détaillée des résultats fournis par un modèle ou algorithme nécessite souvent d'associer plusieurs mesures fournissant des informations complémentaires.

Nous nous sommes ensuite intéressés à la segmentation du fil guide (chapitre 5) dans le cas où peu de données sont disponibles. Pour cela nous avons appliqué la méthode

nnU-Net sur nos données. Les résultats de segmentation étaient très satisfaisants mais leur obtention nécessitait beaucoup de ressources. Nous avons alors exploré la possibilité d'utiliser un réseau U-Net avec MONAI et différentes fonctions de coût, plus ou moins spécialisées pour la segmentation de structures fines. Nous avons à cette occasion introduit une nouvelle fonction de coût qui s'appuie sur les filtres de *vesselness* et sur les modèles par contours actifs. Sur la base de données SA, nous obtenons des meilleurs résultats de segmentation avec les fonctions de coût de Tversky et FGAC qu'avec la fonction de coût de Dice, avec comme score de Dice de $\mathbf{71,1 \pm 2,5}$ avec la fonction de coût de Tversky comparé à $\mathbf{78 \pm 4}$ avec nnU-Net et ce en entraînant un réseau en 33 fois moins de temps. De même pour les autres critères d'évaluation comme le ClDice, $\mathbf{80,9 \pm 2,9}$ contre $\mathbf{94 \pm 5}$ ou le MCC avec un score de $\mathbf{85,9 \pm 0,01}$ contre $\mathbf{89 \pm 2}$. Sur la base de données HM, nous obtenons également des résultats de segmentation satisfaisants, avec comme meilleur modèle celui ayant pour fonction de coût Dice combinée à celle de Vesselness. Nous obtenons un score de Dice moyen de $\mathbf{64,3 \pm 0,3}$, un score de ClDice de $\mathbf{75,5 \pm 3,4}$, un score de Tversky de $\mathbf{63,1 \pm 0,3}$ et une distance de Hausdorff de $\mathbf{170 \pm 44}$ pixels.

Les résultats pour les deux bases de données sont meilleurs que la méthode de Gherardini et al. [2020] qui obtient comme meilleur score de Dice **55** pour la segmentation des cathéters entiers en entraînant sur 9000 images. Les autres scores obtenus dans l'état de l'art sont meilleurs mais ne concernent que l'extrémité du cathéter qui est la partie la plus simple a segmenter.

La dernière contribution du chapitre est la reconstruction par chemins minimaux à partir des composantes connexes prédites. La méthode n'a pas été testée sur toutes les images mais annonce des résultats prometteurs.

Enfin dans le chapitre 6, nous avons cherché à déterminer si une structure telle que l'endoscope pouvait aider à la segmentation d'une structure spatialement liée, c'est-à-dire le fil guide. Pour cela nous avons testé des méthodes d'apprentissage conjoint. Après avoir mené différentes expériences d'apprentissage multi-objets avec U-Net et IterNet nous en déduisons que ces méthodes ne permettent pas l'amélioration du fil guide. Cependant nous remarquons qu'IterNet réussit à bien segmenter le fil guide à la sortie de l'endoscope ce qui manquait dans le chapitre 5.

À notre connaissance, nous sommes les premiers à avoir répondu à la question de la segmentation du fil guide sur les images de fluoroscopie pour la CPRE. Il s'agit donc d'un travail exploratoire mais dont les résultats sont prometteurs et exploitables pour les étapes suivantes. Ainsi plusieurs perspectives se dessinent pour le continuer.

7.2 Perspectives

Plusieurs perspectives sont envisageables à la suite de ce travail. Nous les présentons de manière thématique. Ainsi nous détaillons les perspectives cliniques dans un premier temps, puis les perspectives méthodologiques.

7.2.1 Perspectives cliniques

Données

Dans le cadre de cette thèse nous avons annoté une partie des données reçues mais d'autres sont disponibles. Un premier travail serait de les pré-traiter. Ensuite on pourrait tester les différents modèles sur ces données. Enfin on pourrait ré-entraîner le meilleur

modèle et viser un modèle fonctionnel pour les médecins. Cette nouvelle étape s'accompagnerait d'une analyse plus complète pour identifier les cas limites où la segmentation ne fonctionne pas et comment y remédier.

Logiciel de localisation du fil guide dans les voies biliaires

Dans cette perspective nous faisons l'hypothèse que nous choisissons le meilleur modèle du chapitre 5 et que nous l'utilisons pour la suite. Nous avons obtenu des résultats préliminaires intéressants concernant l'utilisation de nos segmentations comme entrée du logiciel de localisation per-opératoire en cours de développement à l'ISIR. Cependant nous n'avons pour l'instant pas suffisamment de patients qui ont une fluoroscopie exploitable accompagnée d'une segmentation manuelle des voies biliaires en 3D pour pouvoir analyser plus systématiquement ces résultats.

Sur la figure 7.1, nous remarquons que la segmentation du fil guide issue de notre modèle du chapitre 5 est incomplète notamment à cause du produit de contraste présent dans l'image. Cependant elle est suffisante pour que le logiciel interpole les points nécessaires à son bon fonctionnement.

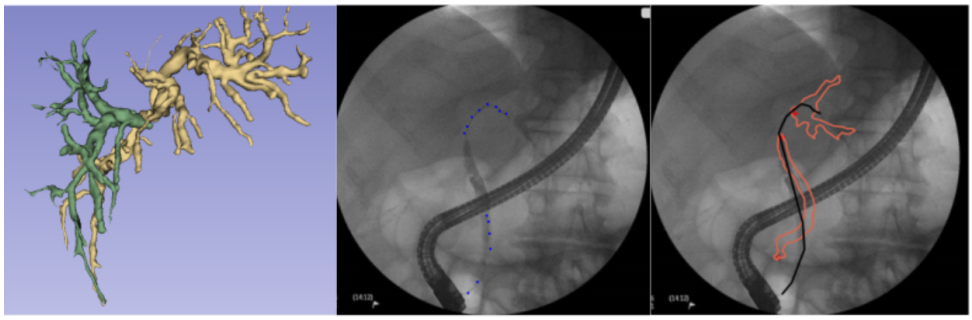


FIGURE 7.1 – Résultat préliminaire de l'application clinique. La première image est la segmentation manuelle des voies biliaires en 3D du patient, établie par les endoscopistes et les radiologues. La deuxième image représente l'échantillonnage des points de la prédiction automatique du fil guide selon la méthode du chapitre 5. La dernière image représente le fil guide interpolé à partir de ces points et la projection de la voie biliaire identifiée comme étant celle où le fil est localisé.

7.2.2 Perspectives méthodologiques

Dans les perspectives méthodologiques nous proposons des pistes d'amélioration ou de recherche qui nous semblent intéressantes pour la détection du fil guide.

Généralisation

Nous pourrions envisager dans un premier temps de mélanger les bases de données pour entraîner les modèles et avoir plus de généralisation. Nous avons déjà commencé à tester cette hypothèse. Nous obtenons des résultats proches de ceux trouvés individuellement pour chaque base mais avec un temps d'entraînement plus long avec 1 à 2 heures par pli de validation croisée contre 20 à 30 minutes par base de données et par pli.

Expériences sur la fonction de coût FGAC

Dans le chapitre 5, nous avons introduit une nouvelle fonction de coût qui améliore les résultats dans le cas de la base de données SA. Nous aimerions la tester plus amplement, sur d'autres jeux de données similaires pour déterminer sa généralisation potentielle.

Amélioration de la segmentation

Nous avons montré que la segmentation n'avait pas besoin d'être complète pour que l'algorithme de fusion puisse fonctionner. Nous aimerions cependant ajouter que certains points particuliers du fil guide jouent un rôle important dans l'algorithme de localisation, comme le point au niveau de l'entrée dans la voie biliaire principale, le point au niveau de l'entrée dans les voies biliaires secondaires, et l'extrémité du fil guide. Ainsi travailler avec une segmentation de référence non binaire donnerait par exemple plus de poids à ces points importants.

Une des pistes explorées lors d'un projet de première année de master que j'ai proposé était d'utiliser un algorithme de détection des voies biliaires dans des boîtes englobantes, puis de faire l'hypothèse que le fil guide est dans les voies biliaires, donc que sur la même image sans produit de contraste, le fil se trouve dans une boîte englobante identifiée. Le but est de réduire l'espace à segmenter et ainsi faciliter la tâche. Le problème majeur était le manque de données et le fait que la boîte englobante couvrait la quasi totalité de l'image si le fil guide était très loin dans les voies biliaires. Une perspective serait d'appliquer un algorithme de détection sur plus d'images, avec une boîte englobante par partie du fil guide. L'avantage est que les boîtes englobantes sont faciles à annoter par rapport aux voies biliaires.

Une dernière perspective serait d'introduire des connaissances a priori pendant l'entraînement. Du plus simple au plus compliqué on retrouve le point de démarrage du fil, sa longueur du fil ou le fait que le fil soit continu.

Dans ce travail nous n'avons pas traité les voies biliaires mais elles représentent un potentiel point de repère pour le recalage 2D-3D. De plus, l'arborescence des voies biliaires se prête facilement à la représentation sous forme de graphe. Une perspective de recherche serait d'extraire un graphe des voies biliaires et de l'utiliser pour le recalage puis pour la localisation du fil guide.

Pour conclure, ce travail représente une étape importante pour la suite du projet MAAGIE, où les résultats vont pouvoir être combinés à ceux de la segmentation sur les images de bili-IRM ainsi qu'au logiciel de localisation du fil guide pour être mis en place dans le cadre clinique et faciliter l'intervention de CPRE.

Publications

Conférences internationales avec comité de relecture :

Sur le sujet de thèse

1. G. Martin, A. Becq, I. Bloch, M. Camus, A. Pinna, and J. Szewczyk. Comparison of Loss Functions for Guidewire Automatic Detection on a Limited Dataset of Fluoroscopic Images for Endoscopic Retrograde Cholangiopancreatography. *International Symposium on Biomedical Imaging*, May 2024b

2. G. Martin, S. El-Madafri, A. Becq, J. Szewczyk, and I. Bloch. Instruments Segmentation in X-ray Fluoroscopic Images for Endoscopic Retrograde Cholangio Pancreatography. *Studies in Health Technology and Informatics*, 294 :133–134, May 2022. doi: 10.3233/shti220416. URL https://hal.sorbonne-universite.fr/hal-03684463. Publisher : IOS Press

Autres sujets

1. O. Thiery, G. Martin, I. Bloch, M. Dres, U. Saleem, and A. Pinna. Semi-Automatic A-Line Detection and Confidence Scoring in Lung Ultrasound. In *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 635–639, Oct. 2022. doi: 10.1109/BioCAS54905.2022.9948591. URL https://ieeexplore.ieee.org/document/9948591. ISSN : 2163-4025

Conférence nationale avec abstract

1. G. Martin, A. Becq, I. Bloch, M. Camus, A. Pinna, and J. Szewczyk. Comparaison de fonctions de coût pour la détection automatique de fil guide sur des images fluoroscopiques de cholangiopancréatographie rétrograde endoscopique. *Colloque Français d'Intelligence Artificielle en Imagerie Biomédicale*, 2024a

ANNEXE A

Base de données

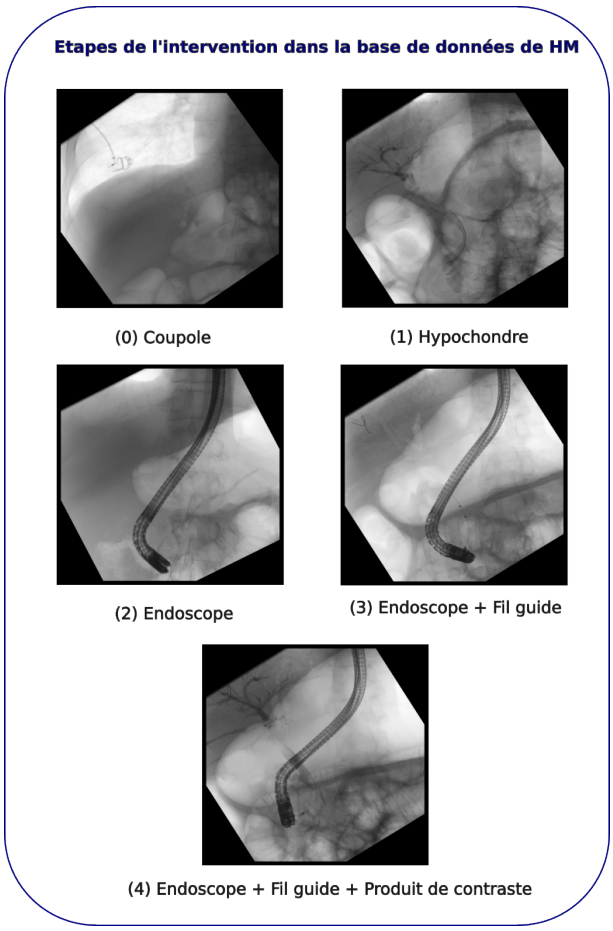


FIGURE A.1 – Illustration des différentes étapes jalonnant l'intervention dans la base de données de l'hôpital Henri Mondor.

ANNEXE B

Application d'un filtre de Frangi sur les données

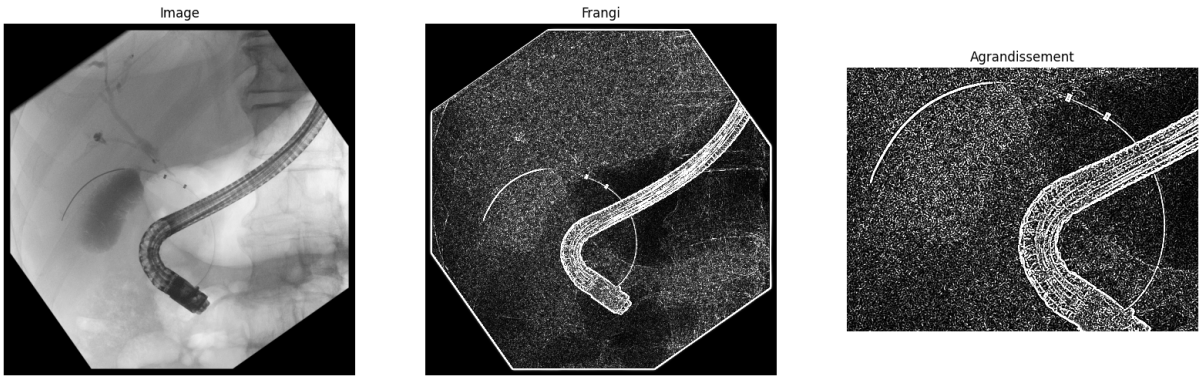


FIGURE B.1 – De gauche à droite : Une image du jeu de données d'Henri Mondor, Application d'un filtre de Frangi sur l'image, Agrandissement de la réponse du filtre dans la région du fil guide. Cette figure montre la capacité des filtres de vesselness à rehausser les structures fines, en effet le fil guide ressort. Elle montre aussi que ces filtres ne permettent pas d'extraire le fil guide sur cette image car d'autres structures comme l'endoscope ont des valeurs similaires.

ANNEXE C

Augmentation des données

L'augmentation des données est une stratégie connue en apprentissage par réseaux de neurones pour diversifier un jeu de données réduit. En effet, les réseaux de neurones nécessitent un nombre conséquent d'exemples divers pour être capables de généraliser une tâche. Pour augmenter les données artificiellement, il suffit d'appliquer des transformations spatiales, d'intensité ou de bruit. Dans la méthode nn-UNet, les transformations appliquées sont considérables, ce qui a pour effet d'allonger le temps d'entraînement et nécessiter plus de ressources de calculs. Toujours dans l'objectif de trouver un compromis entre la performance et les ressources, nous avons appliqué différentes combinaisons de transformations. Pour cela nous avons utilisé un U-Net avec la fonction de coût Dice et 300 époques. Nous avons essayé de reproduire les transformations décrites dans nn-UNet avec MONAI lorsque cela était possible.

Dans cette annexe, nous décrivons chaque transformation plus en détail et montrons lesquelles ont été retenues pour les expériences décrites dans le chapitre 5.

- A. Recadrage : Cette étape réduit la dimension des images à la taille 512×512 et diminue le temps d'entraînement. L'inconvénient est que le réseau apprend sur des parties de l'image sans contexte. De plus, en inférence le réseau va prédire plusieurs imagettes et il faudra les rassembler pour obtenir la prédiction finale.

- B. Normalisation : Cette étape déduit à l'intensité de chaque pixel la moyenne d'intensité de l'image et divise le tout par l'écart-type des intensités dans l'image.

- C. Modification de l'échelle des intensités : Cette étape applique une transformation pour que chaque intensité soit comprise entre 0 et 1.

- D. Recadrage aléatoire : Il s'agit de la même étape que la première mais à des endroits aléatoires dans l'image.

- E. Effet miroir aléatoire : Retourne l'image selon l'axe de symétrie vertical avec une probabilité de $0,5$.

- F. Agrandissement aléatoire : Cette transformation agrandit l'image avec une probabilité de $0,16$, et d'un coefficient d'agrandissement compris entre $0,7$ et $1,4$.

- G. Transformation affine aléatoire : Applique une transformation affine avec une probabilité de $0,1$.
- H. Bruit gaussien : Applique un bruit gaussien de moyenne nulle et d'écart-type $0,1$ avec une probabilité de $0,15$.
- I. Flou gaussien : Applique un noyau gaussien sur l'image avec une probabilité de $0,2$.
- J. Rotation aléatoire : Applique une rotation de 90° avec une probabilité de $0,16$.
- K. Ajustement du contraste aléatoire : Modifie le contraste de l'image avec une probabilité de $0,15$ et un coefficient γ variant entre $0,7$ et $1,5$.

A	B	C	D	E	F	G	H	I	J	K	Dice (Entraînement)	Dice (Test)
x											9	2
x	x										23	21
x		x									29	30
		x	x								60	68
		x	x	x							54	62
		x	x		x						59	64
		x	x			x					60	65
		x	x				x				60	65
		x	x					x			58	66
		x	x						x		52	62
		x	x							x	2	1
		x	x	x		x	x	x	x		3	2
		x	x	x	x	x	x		x		3	0
		x	x	x	x	x			x		59	63

Table C.1 – Récapitulatif des différentes expériences menées pour déterminer les transformations d'augementation des données. A : Recadrage ; B : Normalisation, C : Échelle d'intensités ; D : RandCrop ; E RandFlip ; F RandZoom ; G RandAffine ; H Bruit gaussien ; I Flou gaussien ; J RandRotate ; K Randadjust.

Nous remarquons dans un premier temps que la normalisation allonge significativement le temps d'entraînement pour peu de bénéfice par rapport à la mise à l'échelle des intensités. Elle n'est donc pas retenue dans les méthodes d'augmentation des données. Nous testons ensuite chaque transformation avec la (C) Intensité entre 0 et 1 et la (D) le recadrage aléatoire qui sont nécessaires à l'apprentissage. Nous effectuons ensuite des combinaisons avec toutes les transformations sauf l'agrandissement, puis sans le flou gaussien. Enfin nous n'appliquons que les transformations spatiales. Le tableau C.1 illustre ces expériences.

Nous choisissons donc de n'appliquer que les transformations spatiales ce qui correspond à la dernière expérience.

Bibliographie

Y. Alzahrani and B. Boufama. Biomedical Image Segmentation : A Survey. *SN Computer Science*, 2(4) :310, May 2021. ISSN 2661-8907. doi: 10.1007/s42979-021-00704-7. URL https://doi.org/10.1007/s42979-021-00704-7.

P. Ambrosini, D. Ruijters, W.-J. Niessen, A. Moelker, and T. van Walsum. Continuous roadmapping in liver TACE procedures using 2D–3D catheter-based registration. *International Journal of Computer Assisted Radiology and Surgery*, 10(9) :1357–1370, Sept. 2015. ISSN 1861-6410, 1861-6429. doi: 10.1007/s11548-015-1218-x. URL http://link.springer.com/10.1007/s11548-015-1218-x.

P. Ambrosini, D. Ruijters, W.-J. Niessen, A. Moelker, and T. van Walsum. Fully Automatic and Real-Time Catheter Segmentation in X-Ray Fluoroscopy. In M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne, editors, *Medical Image Computing and Computer Assisted Intervention–MICCAI 2017*, volume Lecture Notes in Computer Science 10434, pages 577–585. Springer International Publishing, Cham, 2017. ISBN 978-3-319-66184-1 978-3-319-66185-8. doi: 10.1007/978-3-319-66185-8_65.

T. H. Baron, B. T. Petersen, K. Mergener, A. Chak, J. Cohen, S. E. Deal, B. Hoffman, B. C. Jacobson, J. L. Petrini, M. A. Safdi, D. O. Faigel, and I. M. Pike. Quality indicators for endoscopic retrograde cholangiopancreatography. *Gastrointestinal Endoscopy*, 63(4, Supplement) :S29–S34, Apr. 2006. ISSN 0016-5107. doi: 10.1016/j.gie.2006.02.019. URL https://www.sciencedirect.com/science/article/pii/S0016510706002823.

A. Becq. *Elaboration et mise en pratique préclinique d'outils robotiques d'assistance à la réalisation d'une cholangiopancréatographie rétrograde par voie endoscopique Projet MAGIE : Modèle d'Assistance au Geste Interventionnel en Endoscopie*. PhD thesis, Sorbonne Université, 2024.

A. Becq, J. Szewczyk, G. Salin, M. Chartier, U. Chaput, R. Leenhardt, X. Dray, L. Arrive, and M. Camus. ERCP 2.0 : Biliary 3D-reconstruction in patients with malignant hilar stricture. *Clinics and Research in Hepatology and Gastroenterology*, 47 (7) :102172, 2023. ISSN 2210-7401. doi: 10.1016/j.clinre.2023.102172. URL https://www.sciencedirect.com/science/article/pii/S2210740123000979.

O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. Gonzalez Ballester, G. Sanroma, S. Napel, S. Petersen, G. Tziritas, E. Grinias, M. Khened, V. A. Kollerathu, G. Krishnamurthi, M.-M. Rohé, X. Pennec, M. Sermesant, F. Isensee, P. Jäger, K. H. Maier-Hein, P. M. Full, I. Wolf, S. Engelhardt, C. F. Baumgartner, L. M. Koch, J. M. Wolterink, I. Išgum, Y. Jang, Y. Hong, J. Patravali, S. Jain, O. Humbert, and P.-M. Jodoin. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis : Is the problem solved ? *IEEE Transactions on Medical Imaging*, 37(11) :2514–2525, 2018. doi: 10.1109/TMI.2018.2837502.

S. Beucher and F. Meyer. The morphological approach to segmentation : the watershed transformation. In *Mathematical Morphology in Image Processing*, pages 433–481. CRC Press, 1993.

X. Bresson, S. Esedoḡlu, P. Vandergheynst, J.-P. Thiran, and S. Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and vision*, 28(2) :151–167, 2007.

K. Cao, K. Du, K. Ding, J. M. Reinhardt, and G. E. Christensen. Regularized nonrigid registration of lung ct images by preserving tissue volume and vesselness measure. *Grand Challenges in Medical Image Analysis*, pages 43–54, 2010.

M. J. Cardoso, W. Li, R. Brown, N. Ma, E. Kerfoot, Y. Wang, B. Murrey, A. Myronenko, C. Zhao, D. Yang, V. Nath, Y. He, Z. Xu, A. Hatamizadeh, A. Myronenko, W. Zhu, Y. Liu, M. Zheng, Y. Tang, I. Yang, M. Zephyr, B. Hashemian, S. Alle, M. Z. Darestani, C. Budd, M. Modat, T. Vercauteren, G. Wang, Y. Li, Y. Hu, Y. Fu, B. Gorman, H. Johnson, B. Genereaux, B. S. Erdal, V. Gupta, A. Diaz-Pinto, A. Dourson, L. Maier-Hein, P. F. Jaeger, M. Baumgartner, J. Kalpathy-Cramer, M. Flores, J. Kirby, L. A. D. Cooper, H. R. Roth, D. Xu, D. Bericat, R. Floca, S. K. Zhou, H. Shuaib, K. Farahani, K. H. Maier-Hein, S. Aylward, P. Dogra, S. Ourselin, and A. Feng. MONAI : An open-source framework for deep learning in healthcare, 2022. URL http://arxiv.org/abs/2211.02701. arXiv :2211.02701 [cs].

V. Caselles, R. Kimmel, and G. Sapiro. Geodesic Active Contours. *International Journal of Computer Vision*, 22(1) :61–79, Feb. 1997. ISSN 1573-1405. doi: 10.1023/A:1007979827043. URL https://doi.org/10.1023/A:1007979827043.

T. Chan and L. Vese. An Active Contour Model without Edges. In M. Nielsen, P. Johansen, O. F. Olsen, and J. Weickert, editors, *Scale-Space Theories in Computer Vision*, pages 141–151, Berlin, Heidelberg, 1999. Springer. ISBN 978-3-540-48236-9. doi: 10.1007/3-540-48236-9_13.

F. Cheevasuvit, H. Maitre, and D. Vidal-Madjar. A robust method for picture segmentation based on a split-and-merge procedure. *Computer Vision, Graphics, and Image Processing*, 34(3) :268–281, 1986.

L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. DeepLab : Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40 (4) :834–848, 2018. doi: 10.1109/TPAMI.2017.2699184.

X. Chen, B. M. Williams, S. R. Vallabhaneni, G. Czanner, R. Williams, and Y. Zheng. Learning active contour models for medical image segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 11632–11640, 2019.

L. D. Cohen and R. Kimmel. Global Minimum for Active Contour Models : A Minimal Path Approach. *International Journal of Computer Vision*, 24(1) :57–78, Aug. 1997. ISSN 1573-1405. doi: 10.1023/A:1007922224810. URL https://doi.org/10.1023/A:1007922224810.

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet : A Large-Scale Hierarchical Image Database. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

A. Dosovitskiy, P. Fischer, E. Ilg, P. Häusser, C. Hazirbas, V. Golkov, P. Van der Smagt, D. Cremers, and T. Brox. Flownet : Learning optical flow with convolutional networks. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2758–2766, 2015. doi: 10.1109/ICCV.2015.316.

A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. An image is worth 16x16 words : Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

A. Essamlali, V. Millot-Maysounabe, M. Chartier, G. Salin, A. Becq, L. Arrivé, M. D. Camus, J. Szewczyk, and I. Claude. Bile Duct Segmentation Methods Under 3D Slicer Applied to ERCP : Advantages and Disadvantages. *International Journal of Biomedical Engineering and Clinical Science*, 9(4) :66–74, Nov. 2023. ISSN 2472-1301. doi: 10.11648/j.ijbecs.20230904.11. URL https://www.sciencepg.com/article/10.11648/j.ijbecs.20230904.11.

M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2) :303–338, June 2010. ISSN 1573-1405. doi: 10.1007/s11263-009-0275-4. URL https://doi.org/10.1007/s11263-009-0275-4.

E. W. Forgy. Cluster analysis of multivariate data : efficiency versus interpretability of classifications. *Biometrics*, 21 :768–769, 1965.

A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever. Multiscale vessel enhancement filtering. In *Medical Image Computing and Computer Assisted Intervention–MICCAI*, volume LNCS 1496, pages 130–137. 1998. ISBN 978-3-540-65136-9 978-3-540-49563-5. doi: 10.1007/BFb0056195. URL http://link.springer.com/10.1007/BFb0056195.

S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6) :721–741, 1984. doi: 10.1109/TPAMI.1984.4767596.

M. Gherardini, E. Mazomenos, A. Menciassi, and D. Stoyanov. Catheter segmentation in X-ray fluoroscopy using synthetic data and transfer learning with light U-Nets. *Computer Methods and Programs in Biomedicine*, 192 :105420, 2020. ISSN 01692607. doi:

10.1016/j.cmpb.2020.105420. URL https://linkinghub.elsevier.com/retrieve/pii/S0169260719312301.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. doi: 10.1109/CVPR.2016.90.

A. Howard, M. Sandler, B. Chen, W. Wang, L.-C. Chen, M. Tan, G. Chu, V. Vasudevan, Y. Zhu, R. Pang, H. Adam, and Q. Le. Searching for MobileNetV3. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1314–1324, 2019. doi: 10.1109/ICCV.2019.00140.

G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger. Densely Connected Convolutional Networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2261–2269, Los Alamitos, CA, USA, jul 2017. IEEE Computer Society. doi: 10.1109/CVPR.2017.243. URL https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.243.

L. Huang, X. Lu, X. Huang, X. Zou, L. Wu, Z. Zhou, D. Wu, D. Tang, D. Chen, X. Wan, Z. Zhu, T. Deng, L. Shen, J. Liu, Y. Zhu, D. Gong, D. Chen, Y. Zhong, F. Liu, and H. Yu. Intelligent difficulty scoring and assistance system for endoscopic extraction of common bile duct stones based on deep learning : multicenter study. *Endoscopy*, 53 (5) :491–498, 2021. ISSN 0013-726X, 1438-8812. doi: 10.1055/a-1244-5698.

F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein. nnU-Net : a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2) :203–211, 2021. ISSN 1548-7105. doi: 10.1038/s41592-020-01008-z. URL https://www.nature.com/articles/s41592-020-01008-z.

N. Kanopoulos, N. Vasanthavada, and R. Baker. Design of an image edge detection filter using the Sobel operator. *IEEE Journal of Solid-State Circuits*, 23(2) :358–367, 1988. doi: 10.1109/4.996.

M. Kass, A. Witkin, and D. Terzopoulos. Snakes : Active contour models. *International Journal of Computer Vision*, 1(4) :321–331, Jan. 1988. ISSN 1573-1405. doi: 10.1007/BF00133570. URL https://doi.org/10.1007/BF00133570.

E. Kerfoot. Left-Ventricle Quantification Using Residual U-Net. In *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges*, pages 371–380, 2019. ISBN 978-3-030-12029-0.

T. Kim, J. Kim, H. S. Choi, E. S. Kim, B. Keum, Y. T. Jeen, H. S. Lee, H. J. Chun, S. Y. Han, D. U. Kim, S. Kwon, J. Choo, and J. M. Lee. Artificial intelligence-assisted analysis of endoscopic retrograde cholangiopancreatography image for identifying ampulla and difficulty of selective cannulation. *Scientific Reports*, 11(1) :8381, Apr. 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-87737-3. URL https://www.nature.com/articles/s41598-021-87737-3.

D. P. Kingma and J. Ba. Adam : A method for stochastic optimization. *arXiv preprint arXiv :1412.6980*, 2014.

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 *IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023.

A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. URL https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf.

G. La Barbera. *Learning anatomical digital twins in pediatric 3D imaging for renal cancer surgery*. PhD thesis, Institut Polytechnique de Paris, Nov. 2022. URL https://theses.hal.science/tel-03911159.

G. La Barbera, L. Rouet, H. Boussaid, A. Lubet, R. Kassir, S. Sarnacki, P. Gori, and I. Bloch. Tubular structures segmentation of pediatric abdominal-visceral ceCT images with renal tumors : Assessment, comparison and improvement. *Medical Image Analysis*, 90 :102986, 2023. ISSN 1361-8415. doi: 10.1016/j.media.2023.102986. URL https://www.sciencedirect.com/science/article/pii/S1361841523002463.

J. Lamy, O. Merveille, B. Kerautret, N. Passat, and A. Vacavant. Vesselness filters : A survey with benchmarks applied to liver imaging. In *International Conference on Pattern Recognition (ICPR)*, pages 3528–3535, 2020. doi: 10.1109/ICPR48806.2021.9412362. URL https://hal.archives-ouvertes.fr/hal-02544493.

L. Li, M. Verma, Y. Nakashima, H. Nagahara, and R. Kawasaki. IterNet : Retinal image segmentation utilizing structural redundancy in vessel networks. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 3645–3654, 2020. doi: 10.1109/WACV45572.2020.9093621.

T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dollàr, and C. L. Zitnick. Microsoft COCO : common objects in context. *CoRR*, abs/1405.0312, 2014. URL http://arxiv.org/abs/1405.0312.

T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollàr. Focal Loss for Dense Object Detection. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2980–2988, 2017. URL https://openaccess.thecvf.com/content_iccv_2017/html/Lin_Focal_Loss_for_ICCV_2017_paper.html.

J. Ma. Cutting-edge 3D Medical Image Segmentation Methods in 2020 : Are Happy Families All Alike ?, Jan. 2021. URL http://arxiv.org/abs/2101.00232. arXiv :2101.00232.

J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, and A. L. Martel. Loss Odyssey in medical image segmentation. *Medical Image Analysis*, 71 :102035, 2021a. ISSN 1361-8415. doi: 10.1016/j.media.2021.102035. URL https://www.sciencedirect.com/science/article/pii/S1361841521000815.

J. Ma, J. He, and X. Yang. Learning Geodesic Active Contours for Embedding Object Global Information in Segmentation CNNs. *IEEE Transactions on Medical Imaging*, 40(1), 2021b. ISSN 1558-254X. doi: 10.1109/TMI.2020.3022693. URL https://ieeexplore.ieee.org/document/9187860/citations?tabFilter=papers#citations.

J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang. Segment anything in medical images. *Nature Communications*, 15(1) :1–9, Jan. 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-44824-z. URL https://www.nature.com/articles/s41467-024-44824-z. Number : 1 Publisher : Nature Publishing Group.

D. Malka, J. Blanc, B. K., A. Bretagne-Bignon, R. Coriat, T. de Baère, C. de la Fouchardière, C. Dromain, J. Edeline, M. Gelli, S. Horn, V. Laurent-Croise, C. Lepage, A. Lièvre, A. Lopez, J. Meilleroux, S. Manfredi, C. Neuzillet, F. Prat, O. Rosmorduc, G. Roth, A. Sa Cunha, M. Sarabi, and L. Verlingue. *Thésaurus National de Cancérologie Digestive*. SNFGE, 2022. URL http://www.tncd.org.

G. Martin, S. El-Madafri, A. Becq, J. Szewczyk, and I. Bloch. Instruments Segmentation in X-ray Fluoroscopic Images for Endoscopic Retrograde Cholangio Pancreatography. *Studies in Health Technology and Informatics*, 294 :133–134, May 2022. doi: 10.3233/shti220416. URL https://hal.sorbonne-universite.fr/hal-03684463. Publisher : IOS Press.

G. Martin, A. Becq, I. Bloch, M. Camus, A. Pinna, and J. Szewczyk. Comparaison de fonctions de coût pour la détection automatique de fil guide sur des images fluoroscopiques de cholangiopancréatographie rétrograde endoscopique. *Colloque Français d'Intelligence Artificielle en Imagerie Biomédicale*, 2024a.

G. Martin, A. Becq, I. Bloch, M. Camus, A. Pinna, and J. Szewczyk. Comparison of Loss Functions for Guidewire Automatic Detection on a Limited Dataset of Fluoroscopic Images for Endoscopic Retrograde Cholangiopancreatography. *International Symposium on Biomedical Imaging*, May 2024b.

W. S. McCune, P. E. Shorb, and H. Moscovitz. Endoscopic cannulation of the ampulla of vater : a preliminary report. *Annals of Surgery*, 167(5) :752, 1968.

O. Merveille, H. Talbot, L. Najman, and N. Passat. Curvilinear Structure Analysis by Ranking the Orientation Responses of Path Operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(2) :304–317, Feb. 2018. ISSN 1939-3539. doi: 10.1109/TPAMI.2017.2672972. URL https://ieeexplore.ieee.org/document/7862284/?arnumber=7862284.

F. Milletari, N. Navab, and S.-A. Ahmadi. V-Net : Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In *Fourth International Conference on 3D Vision (3DV)*, pages 565–571, Oct. 2016. doi: 10.1109/3DV.2016.79. URL https://ieeexplore.ieee.org/abstract/document/7785132.

D. Müller, I. Soto-Rey, and F. Kramer. Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, 15(1) :210, June 2022. ISSN 1756-0500. doi: 10.1186/s13104-022-06096-y. URL https://doi.org/10.1186/s13104-022-06096-y.

A. Nguyen, D. Kundrat, G. Dagnino, W. Chi, M. E. M. K. Abdelaziz, Y. Guo, Y. Ma, T. M. Y. Kwok, C. Riga, and G.-Z. Yang. End-to-End Real-time Catheter Segmentation with Optical Flow-Guided Warping during Endovascular Intervention. In *IEEE International Conference on Robotics and Automation*, pages 9967–9973, 2020. doi: 10.1109/ICRA40945.2020.9197307. URL http://www.scopus.com/inward/record.url?scp=85092731403&partnerID=8YFLogxK.

N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1) :62–66, 1979. doi: 10.1109/TSMC.1979.4310076.

A. Reinke, M. D. Tizabi, M. Baumgartner, M. Eisenmann, D. Heckmann-Nötzel, A. E. Kavur, T. Rädsch, C. H. Sudre, L. Acion, M. Antonelli, T. Arbel, S. Bakas, A. Benis, F. Buettner, M. J. Cardoso, V. Cheplygina, J. Chen, E. Christodoulou, B. A. Cimini, K. Farahani, L. Ferrer, A. Galdran, B. van Ginneken, B. Glocker, P. Godau, D. A. Hashimoto, M. M. Hoffman, M. Huisman, F. Isensee, P. Jannin, C. E. Kahn, D. Kainmueller, B. Kainz, A. Karargyris, J. Kleesiek, F. Kofler, T. Kooi, A. Kopp-Schneider, M. Kozubek, A. Kreshuk, T. Kurc, B. A. Landman, G. Litjens, A. Madani, K. Maier-Hein, A. L. Martel, E. Meijering, B. Menze, K. G. M. Moons, H. Müller, B. Nichyporuk, F. Nickel, J. Petersen, S. M. Rafelski, N. Rajpoot, M. Reyes, M. A. Riegler, N. Rieke, J. Saez-Rodriguez, C. I. Sánchez, S. Shetty, R. M. Summers, A. A. Taha, A. Tiulpin, S. A. Tsaftaris, B. Van Calster, G. Varoquaux, Z. R. Yaniv, P. F. Jäger, and L. Maier-Hein. Understanding metric-related pitfalls in image analysis validation. *Nature Methods*, 21(2) :182–194, Feb. 2024. ISSN 1548-7105. doi: 10.1038/s41592-023-02150-0. URL https://www.nature.com/articles/s41592-023-02150-0.

S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN : Towards Real-Time Object Detection with Region Proposal Networks. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/file/14bfa6bb14875e45bba028a21ed38046-Paper.pdf.

M. Reuterwall, A. Waldthaler, J. Lubbe, N. Kadesjö, R. Pozzi Mucelli, M. Del Chiaro, M. Lohr, and U. Arnelo. Bimodal ERCP, a new way of seeing things. *Endoscopy International Open*, 8(3) :E368–E376, Mar. 2020. ISSN 2364-3722. doi: 10.1055/a-1070-8749.

O. Ronneberger, P. Fischer, and T. Brox. U-net : Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015.

F. Rosenblatt. The perceptron : a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65 6 :386–408, 1958. URL https://api.semanticscholar.org/CorpusID:12781225.

P. Rougé, O. Merveille, and N. Passat. ccDice : A Topology-Aware Dice Score Based on Connected Components. In C. Chen, Y. Singh, and X. Hu, editors, *Topology- and Graph-Informed Imaging Informatics*, pages 11–21, Cham, 2025. Springer Nature Switzerland. ISBN 9783031739675. doi: 10.1007/978-3-031-73967-5_2.

D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323 :533–536, 1986. URL https://api.semanticscholar.org/CorpusID:205001834.

S. S. M. Salehi, D. Erdogmus, and A. Gholipour. Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks. In Q. Wang, Y. Shi, H.-I. Suk, and K. Suzuki, editors, *Machine Learning in Medical Imaging*, pages 379–387, Cham, 2017. Springer International Publishing. ISBN 978-3-319-67389-9. doi: 10.1007/978-3-319-67389-9_44.

M. Schaap, C. T. Metz, T. van Walsum, A. G. van der Giessen, A. C. Weustink, N. R. Mollet, C. Bauer, H. Bogunović, C. Castro, X. Deng, E. Dikici, T. O'Donnell, M. Frenay, O. Friman, M. H. Hoyos, P. H. Kitslaar, K. Krissian, C. Kühnel, M. A. Luengo-Oroz, M. Orkisz, O. Smedby, M. Styner, A. Szymczak, H. Tek, C. Wang, S. K. Warfield, S. Zambal, Y. Zhang, G. P. Krestin, and W. J. Niessen. Standardized Evaluation Methodology and Reference Database for Evaluating Coronary Artery Centerline Extraction Algorithms. *Medical Image Analysis*, 13(5) :10.1016/j.media.2009.06.003, Oct. 2009. ISSN 1361-8415. doi: 10.1016/j.media.2009.06.003. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3843509/.

J. A. Sethian. A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences*, 93(4) :1591–1595, Feb. 1996. doi: 10.1073/pnas.93.4.1591. URL https://www.pnas.org/doi/abs/10.1073/pnas.93.4.1591.

J. Shi and Tomasi. Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600, 1994. doi: 10.1109/CVPR.1994.323794.

S. Shit, J. C. Paetzold, A. Sekuboyina, I. Ezhov, A. Unger, A. Zhylka, J. P. W. Pluim, U. Bauer, and B. H. Menze. ClDice - a Novel Topology-Preserving Loss Function for Tubular Structure Segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16555–16564, June 2021. doi: 10.1109/CVPR46437.2021.01629. ISSN : 2575-7075.

K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations (ICLR 2015)*, pages 1–14. Computational and Biological Learning Society, 2015.

I. Sirazitdinov, H. Schulz, A. Saalbach, S. Renisch, and D. V. Dylov. Tubular shape aware data generation for segmentation in medical imaging. *International Journal of Computer Assisted Radiology and Surgery*, 17(6) :1091–1099, 2022.

V. Subramanian, H. Wang, J. T. Wu, K. C. Wong, A. Sharma, and T. Syeda-Mahmood. Automated detection and type classification of central venous catheters in chest X-rays. In *Medical Image Computing and Computer Assisted Intervention*, pages 522–530. Springer, 2019.

H. Talbot and B. Appleton. Efficient complete and incomplete path openings and closings. *Image and Vision Computing*, 25(4) :416–425, Apr. 2007. ISSN 0262-8856. doi: 10.1016/j.imavis.2006.07.021. URL https://www.sciencedirect.com/science/article/pii/S0262885606002575.

H. Tang, H. K. Li, C. F. Yang, J.-L. Dillenseger, G. Coatrieux, J. Feng, S. J. Zhou, and Y. Chen. A multiple catheter tips tracking method in X-ray fluoroscopy images by a new lightweight segmentation network and Bayesian filtering. *The International Journal of Medical Robotics + Computer Assisted Surgery : MRCAS*, 19(6) :e2569, Dec. 2023. ISSN 1478-596X. doi: 10.1002/rcs.2569.

O. Thiery, G. Martin, I. Bloch, M. Dres, U. Saleem, and A. Pinna. Semi-Automatic A-Line Detection and Confidence Scoring in Lung Ultrasound. In *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 635–639, Oct. 2022. doi: 10.1109/BioCAS54905.

2022.9948591. URL https://ieeexplore.ieee.org/document/9948591. ISSN : 2163-4025.

Y.-D. Wu, X.-L. Xie, G.-B. Bian, Z.-G. Hou, X.-R. Cheng, S. Chen, S.-Q. Liu, and Q.-L. Wang. Automatic guidewire tip segmentation in 2D X-ray fluoroscopy using convolution neural networks. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7, 2018. ISBN 978-1-5090-6014-6. doi: 10.1109/IJCNN.2018.8489337. URL https://ieeexplore.ieee.org/document/8489337/.

Y. Xi. Instruments detection on 2D X-ray fluoroscopic images. Master's thesis, Sorbonne Université, 2024.

S. Yang, Y. Wang, D. Ai, H. Geng, D. Zhang, D. Xiao, H. Song, M. Li, and J. Yang. Augmented reality navigation system for biliary interventional procedures with dynamic respiratory motion correction. *IEEE Transactions on Biomedical Engineering*, 71(2) : 700–711, 2024. doi: 10.1109/TBME.2023.3316290.

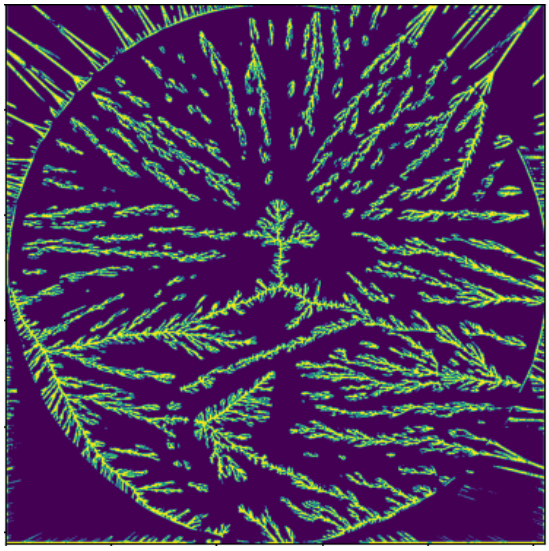
D. Zhang, Y. S, G. Hx, Y. Yj, D. Cj, Y. J, and L. My. Real-time continuous image guidance for endoscopic retrograde cholangiopancreatography based on 3D/2D registration and respiratory compensation. *World Journal of Gastroenterology*, 29(20), May 2023. ISSN 2219-2840. doi: 10.3748/wjg.v29.i20.3157. URL https://pubmed.ncbi.nlm.nih.gov/37346159/.

R. Zhao, B. Qian, X. Zhang, Y. Li, R. Wei, Y. Liu, and Y. Pan. Rethinking Dice loss for medical image segmentation. In *IEEE International Conference on Data Mining (ICDM)*, pages 851–860, 2020. doi: 10.1109/ICDM50108.2020.00094.

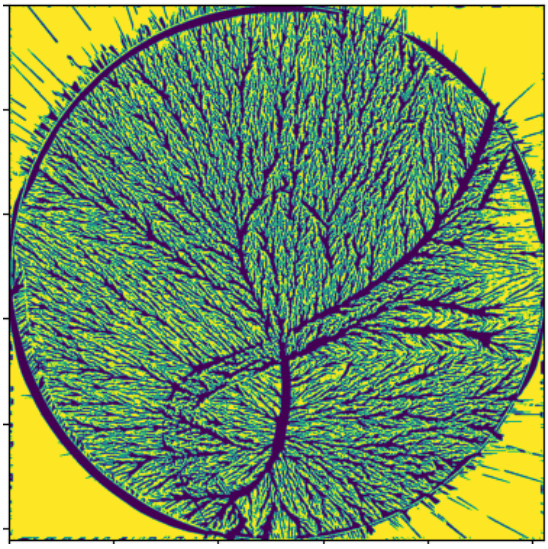
L. Zhou, C. Zhang, and M. Wu. D-LinkNet : LinkNet with pretrained encoder and dilated convolution for high resolution satellite imagery road extraction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 192–1924. IEEE, 2018. ISBN 978-1-5386-6100-0. doi: 10.1109/CVPRW.2018.00034.

Y.-J. Zhou, X.-L. Xie, X.-H. Zhou, S.-Q. Liu, G.-B. Bian, and Z.-G. Hou. A Real-Time Multifunctional Framework for Guidewire Morphological and Positional Analysis in Interventional X-Ray Fluoroscopy. *IEEE Transactions on Cognitive and Developmental Systems*, 13(3) :657–667, Sept. 2021. ISSN 2379-8920, 2379-8939. doi: 10.1109/TCDS.2020.3023952. URL https://ieeexplore.ieee.org/document/9197674/.

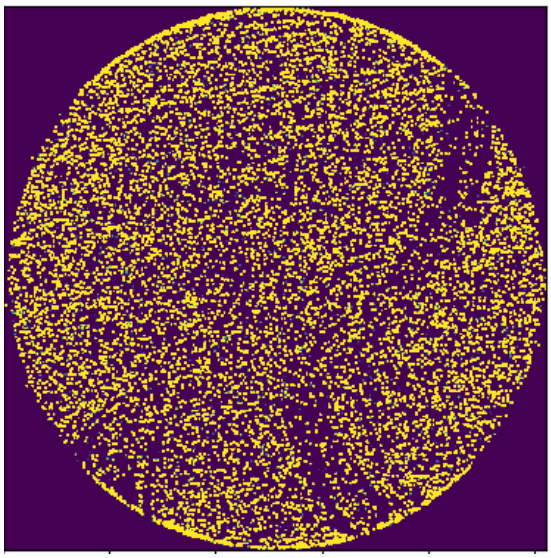
Il arrive que le résultat obtenu ne soit pas toujours celui attendu d'un point de vue scientifique... Il se révèle parfois être artistique. En voici un florilège :



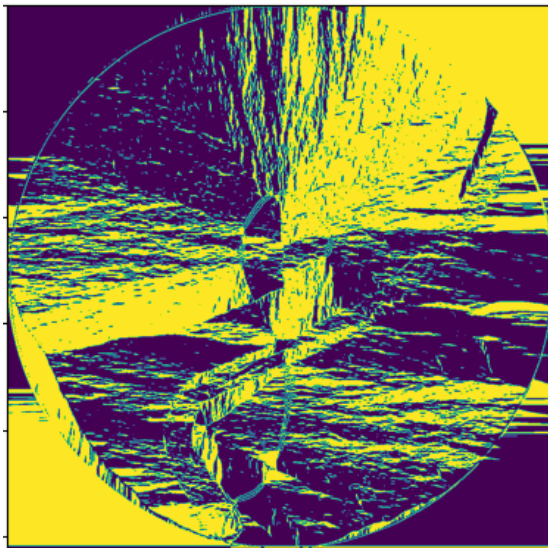
(a) Flocon



(b) Forêt



(c) Pointillisme



(d) A la façon de Kandinsky