Uncertainty Modelling for End-to-End 3D Reconstruction of Coronary Arteries from 2D X-Ray Angiography

by

Rémi MARTIN

THESIS PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE IN PARTIAL FULFILLMENT FOR THE DEGREE OF DOCTOR OF PHILOSOPHY Ph.D.

MONTREAL, OCTOBER 30, 2022

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE UNIVERSITÉ DU QUÉBEC



This Creative Commons license allows readers to download this work and share it with others as long as the author is credited. The content of this work cannot be modified in any way or used commercially.

BOARD OF EXAMINERS

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Mr. Luc Duong, Thesis supervisor Department of Software and Information Technology Engineering, École de Technologie Supérieure

Mr. Mohamed Cheriet, Chair, Board of Examiners Department of Software and Information Technology Engineering, École de Technologie Supérieure

Mrs. Sylvie Ratté, Member of the Jury Department of Software and Information Technology Engineering, École de Technologie Supérieure

Mr. David Rivest-Hénault, External Examiner National Research Council Canada

THIS THESIS WAS PRESENTED AND DEFENDED

IN THE PRESENCE OF A BOARD OF EXAMINERS AND THE PUBLIC

ON OCTOBER 7, 2022

AT ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor, Prof. Luc Duong. His patience and dedication helped me develop my aptitudes in research by working on fascinating projects. His methodological guidance aided me in my PhD while nurturing my independence and creativity. Thanks you for your patience and your relentless help.

My gratitude also goes to the members of my jury, Sylvie Ratté, Mohamed Cheriet and David Rivest-Hénault for their valuable input on my research.

Thank you to Dr Joaquim Miró, his secretary and his catheterization team. They welcomed me during catheterization interventions, which greatly helped me to understand the various medical challenges and their needs. They also provided me with valuable advices and comments about the clinical side of my research project.

I also thank Prof. Theo Van Walsum, Dr. Paul Segars and Dr. Ehsan Samei for their time. Their valuable comments and input on my research greatly helped me improve the quality of my research work.

My gratitude extends to the technical staff of ETS, Patrice Dion and Samuel Docquier who helped me with many technical difficulties that I have encountered.

My thanks also go to the members of the lab, both LIVE and LINCS members for their support, advices and for many scientific discussions that have helped me shape my research objectives.

I would also like to thank the team at Crossfit Wonderland, who have helped me in many ways. Apart from giving me the opportunity to extend my teaching skills, you've helped me go through difficult periods and accept myself the way I am. Particular thanks to my mentor, workout partner and friend, Christiane. We have spent a lot of time together and you've helped me more than you know. A very special thanks goes to my best friend Fariba. We have been through so much together. You've helped me go through difficult periods and were always open for fascinating scientific discussions. To our friendship, brunches and our broken shoulders!

Des remerciements tout particuliers se portent aux membres de ma famille et particulièrement mes parents, mes soeurs et mes grand-mères pour leur support et leur patience. Ces dernières années ont été la source de beaucoup de changements. Merci pour leur acceptation et leur soutien sans failles, je n'y serai pas arrivé sans vous.

Last but not least, thanks to my friends and some of my previous flatmates for their support, particularly during the pandemic.

This research was made possible thanks to the financial support of the Fonds de recherche du Québec - Nature et technologies (FRQNT: 290478), NSERC Discovery grant, ETS Institutional funds PSIRE-I and NVIDIA Applied Research Accelerator Program.

Modélisation de l'incertitude dans le contexte de la reconstruction 3D des artères coronaires à partir d'angiographie à rayon X

Rémi MARTIN

RÉSUMÉ

Les maladies des artères coronaires sont l'une des principales causes de mortalité au Canada. Lorsque possible, les cardiologues préfèrent l'utilisation de cathétérisme, une intervention minimalement invasive, à des fins de traitement.

Lors d'une pose de stent en cathétérisme cardiaque, un cathéter est inséré dans le corps du patient suivi d'un stent qui est ensuite déplacé jusqu'à arriver à la structure vasculaire visée. Cette navigation est généralement faite par les cardiologues en se repérant sur une ou plusieurs projections 2D appelées angiographies. Les angiographies sont des radiographies sur lesquelles sont visibles les structures vasculaires grâce à l'injection d'agent de contraste par cathéter. Les difficultés rencontrées lors de telles interventions sont principalement liées à la technique d'imagerie utilisée. En effet, les angiographies sont une représentation en 2D de structures en 3D. Il est donc difficile d'estimer la taille et la profondeur réelle des artères. Ensuite, l'agent de contraste se dilue rapidement dans le sang. La fenêtre de temps pendant laquelle les structures vasculaires sont visibles est donc assez limitée. De même, les structures vasculaires bougent et se déforment en fonction des mouvements cardiaques et respiratoires. Tous ces artefacts rendent la navigation et la pose de stent difficile.

Notre objectif principal est de modéliser l'incertitude dans le contexte spécifique de la reconstruction 3D d'artères coronaires lors d'interventions de pose de stent en cardiologie. Cet objectif peut se décomposer en trois objectifs spécifiques : 1. Segmentation des artères coronaires sur des angiographies et fournir une mesure d'incertitude liée à la segmentation; 2. Effectuer la synthèse de données supplémentaires et paramétrables; 3. Effectuer la reconstruction 3D des artères segmentées à partir d'une vue unique.

Notre premier objectif spécifique se base sur l'utilisation de réseaux de neurones Bayésiens. En réalisant la segmentation de différentes structures vasculaires (aortes, artères coronaires, artères pulmonaires) à l'aide de convolutions et de couches entièrement connectées, nous souhaitons également utiliser les avantages de l'approche Bayésienne des probabilités afin de fournir une mesure d'incertitude. Cette mesure d'incertitude peut être décomposée en deux sous-catégories d'incertitudes : l'incertitude aléatoire et l'incertitude épistémique. La première représente les sources d'information incertaines directement présentes dans les images utilisées lors de l'apprentissage. La seconde représente ce que le modèle a appris, et plus particulièrement ce qu'il n'a pas appris. Fournir une mesure de ces incertitudes permet de comprendre les limites de notre apprentissage, tout en alertant des prédictions les plus incertaines.

Notre second objectif spécifique porte sur la synthèse d'angiographies d'artères coronaires. Nous avons pour cela utilisé un simulateur cardio-respiratoire pour générer des cinéangiographies paramétrables à partir de modèles 3D d'artères. Nous avons proposé un algorithme de transfert

de style basé sur une mesure de vascularité. Le transfert de style permet d'ajouter le style et le bruit particulier des vraies angiographies sur les simulations obtenues. Ainsi, notre méthode permet de synthétiser de nouvelles images d'artères coronaires à des fins d'apprentissage ou d'augmentation de données pour algorithmes d'apprentissage automatique.

Enfin, notre dernier objectif consiste à réaliser la reconstruction 3D d'artères coronaires à partir d'une seule image d'angiographie segmentée. Pour cela un modèle d'apprentissage est utilisé pour extraire les caractéristiques d'une image et de les associer à un graphe qui est graduellement déformé au cours du temps afin de former la reconstruction 3D finale. Utiliser une seule image permet d'éviter l'utilisation de méthodes de recalage temporel et d'adapter la méthode de reconstruction à plus de configurations de laboratoire de cathétérisation. Les modèles 3D ainsi obtenus sont des vues supplémentaires utilisables pendant les interventions. Ces vues contiennent moins d'ambiguïtés que des projections 2D qui demandent plus d'expérience de la part des cardiologues.

Mots-clés: Artères coronary, Reconstruction 3D, Segmentation avec incertitude, Incertitude, Synthèse d'artères coronaires

Uncertainty Modelling for End-to-End 3D Reconstruction of Coronary Arteries from 2D X-Ray Angiography

Rémi MARTIN

ABSTRACT

Coronary artery diseases are one of the main causes of death in Canada. Amongst those, stenoses of the coronary arteries are one of the most predominant. During a percutaneous intervention, a catheter is inserted through the femoral artery and guided towards to heart. During stenting procedures, a stent is guided alongside the catheter. Upon reaching the narrowed artery, the catheter is expanded to give the affected artery a more tubular shape.

X-ray angiography are currently the gold-standard imaging procedure for the guidance of catheters. These images are obtained using X-rays while injecting contrast agents in the patient's arteries. The main difficulty linked to the use of angiography is linked to the noise contained in the images and the ambiguities created by the projection of the 3D structures onto 2D images. As such, the outcome of the stenting procedures is intimately linked to the experience of the cardiologists. Using a 3D model of the arteries during percutaneous interventions could help alleviate the difficulty encountered during such interventions.

Hence, our main objective is to model uncertainty in the context of 3D reconstruction of coronary arteries during percutaneous interventions. Our contributions are three-fold: 1. Coronary artery segmentation on X-ray angiography using uncertainty metrics; 2. Fully customizable coronary artery angiography synthesis; 3. Monocular 3D reconstruction of coronary arteries using mesh deformation networks.

The first contribution is a novel method to segment coronary arteries in X-ray angiography. This is done using Bayesian Convolutional Neural Networks to also provide a pixel-wise measure of uncertainty regarding the yielded segmentation. This measure of uncertainty is then used alongside a fully-connected neural network designed to provide a threshold for the uncertainty values. The values above the threshold and then deemed too risky to use directly and flagged for the operator while the values below can safely be used for interventions.

Our second objective considers a new method to synthesize coronary artery X-ray angiography. We used a realistic cardio-respiratory simulator to generate fully customizable sequences of coronary arteries. We proposed a new loss function designed to work with CycleGAN to transfer the style of X-ray coronary arteries onto the simulated images. The new loss function is based on a vesselness measure that checks that the topology of the coronary arteries from the input image is respected in the stylized images. Our method allows for the generation of new images for learning purposes or data augmentation purposes by allowing the generation of out-of-distribution data.

For our last objective, we proposed the single-view 3D reconstruction of coronary arteries from a segmented X-ray angiography. To do so, we used two learning models. One model is trained

to extract visual and geometrical features from the segmented image. The other model uses the extracted features to adapt a mesh and gradually make it adapt to the shape of the object to reconstruct. Using a single image for the 3D reconstruction alleviates the need for temporal registration and allows the application of the method to multiple catheterization laboratory configurations. The obtained reconstruction can be used as an additional reference for the guidance of catheters during percutaneous interventions.

Keywords: Coronary arteries, Reconstruction 3D, Uncertainty-aware Segmentation, Uncertainty, Coronary artery synthesis

TABLE OF CONTENTS

INTRO	DUCTI	ON	1
01	Problem statement and motivations		
0.1	 Research objective and contributions 		3
0.2	Research		
CHAP	TER 1	LITERATURE REVIEW	7
1.1	Anatom	y of the heart	7
	1.1.1	Coronary arteries	8
	1.1.2	Stenosis	10
1.2	X-ray an	ngiography	12
1.3	Artery s	egmentation	15
1.4	Bayesia	n Neural Networks and uncertainties	16
	1.4.1	Uncertainty estimation	17
	1.4.2	Uses of uncertainty and uncertainty reduction	19
1.5	Generat	ive adversarial networks and style transfer	20
1.6	3D reco	nstruction and registration	21
	1.6.1	2D/3D registration	21
	1.6.2	General 3D reconstruction	22
	1.6.3	Artery reconstruction	
1.7	Summai	v	25
		· · · · · · · · · · · · · · · · · · ·	
CHAP	TER 2	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION	
CHAP	TER 2	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS	27
CHAP [*] 2.1	TER 2 Abstract	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS	27 27
CHAP ⁴ 2.1 2.2	TER 2 Abstract Introduc	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS	27 27 28
CHAP ⁴ 2.1 2.2 2.3	TER 2 Abstract Introduc Methode	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS	27 27 28 31
CHAP ⁴ 2.1 2.2 2.3	TER 2 Abstract Introduc Methodo 2.3.1	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS ction	27 27 28 31 32
CHAP 2.1 2.2 2.3	TER 2 Abstract Introduc Methode 2.3.1 2.3.2	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS	27 27 28 31 32 35
CHAP ⁴ 2.1 2.2 2.3	TER 2 Abstract Introduc Methode 2.3.1 2.3.2 2.3.3	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS	27 27 28 31 32 35 37
CHAP ⁴ 2.1 2.2 2.3	TER 2 Abstract Introduc Methode 2.3.1 2.3.2 2.3.3 Experim	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS	27 27 28 31 32 35 37 38
CHAP 2.1 2.2 2.3 2.4	TER 2 Abstract Introduc Methode 2.3.1 2.3.2 2.3.3 Experim 2.4.1	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS	27 27 28 31 32 35 37 38 38
CHAP ⁴ 2.1 2.2 2.3 2.4	TER 2 Abstract Introduc Methode 2.3.1 2.3.2 2.3.3 Experim 2.4.1 2.4.2	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS	27 27 28 31 32 35 37 38 38 43
CHAP 2.1 2.2 2.3 2.4	TER 2 Abstract Introduc Methode 2.3.1 2.3.2 2.3.3 Experim 2.4.1 2.4.2 2.4.3	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS	27 27 28 31 32 35 37 38 38 43 46
CHAP ⁴ 2.1 2.2 2.3 2.4	TER 2 Abstract Introduc Methode 2.3.1 2.3.2 2.3.3 Experim 2.4.1 2.4.2 2.4.3 Conclus	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS	27 27 28 31 32 35 37 38 43 43 46 48
 CHAP 2.1 2.2 2.3 2.4 2.5 	TER 2 Abstract Introduc Methode 2.3.1 2.3.2 2.3.3 Experim 2.4.1 2.4.2 2.4.3 Conclus	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS	27 27 28 31 32 35 37 38 38 43 46 48
 CHAP 2.1 2.2 2.3 2.4 2.5 CHAP 	TER 2 Abstract Introduc Methode 2.3.1 2.3.2 2.3.3 Experim 2.4.1 2.4.2 2.4.3 Conclus TER 3	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS	27 27 28 31 32 35 37 38 43 43 46 48
 CHAP 2.1 2.2 2.3 2.4 2.5 CHAP 	TER 2 Abstract Introduc Methode 2.3.1 2.3.2 2.3.3 Experim 2.4.1 2.4.2 2.4.3 Conclus TER 3	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS	27 27 28 31 32 35 37 38 38 43 46 48
 CHAP 2.1 2.2 2.3 2.4 2.5 CHAP 3.1 	TER 2 Abstract Introduc Methode 2.3.1 2.3.2 2.3.3 Experim 2.4.1 2.4.2 2.4.3 Conclus TER 3 Abstract	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS	27 27 28 31 32 35 37 38 38 43 46 48 51 51
 CHAP 2.1 2.2 2.3 2.4 2.5 CHAP 3.1 3.2 	TER 2 Abstract Introduc Methode 2.3.1 2.3.2 2.3.3 Experim 2.4.1 2.4.2 2.4.3 Conclus TER 3 Abstract Introduc	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS	27 27 28 31 32 35 37 38 43 43 48 48 51 51 52
 CHAP 2.1 2.2 2.3 2.4 2.5 CHAP 3.1 3.2 3.3 	TER 2 Abstract Introduc Methode 2.3.1 2.3.2 2.3.3 Experim 2.4.1 2.4.2 2.4.3 Conclus TER 3 Abstract Introduc Data	PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS	27 27 28 31 32 35 37 38 38 43 46 48 51 51 52 54

Page

	3.3.2 3.3.3	Patient data Pre-processing		
3.4	Method	lology		
	3.4.1	Loss function		
		3.4.1.1 GAN loss		
		3.4.1.2 Cycle consistency loss		
		3.4.1.3 Vesselness loss		
	3.4.2	Validation		
3.5	Results	Results and discussion		
	3.5.1	Training details		
	3.5.2	Evaluation		
3.6	Conclu	sion		
CHA	PTER 4	SINGLE-VIEW 3D RECONSTRUCTION	OF CORONARY	
		ARTERIES USING MESH DEFORMATION	NETWORKS 69	
4.1	Abstrac	xt		
4.2	Introdu	ction		
4.3	Method	ls		
	4.3.1	Data description and preprocessing		
		4.3.1.1 Geometric model dataset	74	
	4.3.2	Methodology		
	4.3.3	Model architecture		
		4.3.3.1 Perceptual features extraction		
		4.3.3.2 2D projection		
		4.3.3.3 Mesh deformation		
	4.3.4	Loss function		
		4.3.4.1 Chamfer loss function		
		4.3.4.2 Normal loss function		
		4.3.4.3 Laplacian regularization		
		4.3.4.4 Edge length regularization		
		4.3.4.5 Projection loss function		
4.4	Validat	ion		
	4.4.1	Metrics		
		4.4.1.1 3D distances		
		4 4 1 2 2D projections	85	
	442	Validation scheme	86	
		4 4 2 1 Geometric model data	86	
		4 4 2 2 Patient imaging data		
45	Results			
т.Ј	<u>4 5 1</u>	Traininσ		
	452	Geometric model data	۵۷ ۵۷	
	т. <i>3.2</i> 453	Patient imaging data	02 م2	
46	T.J.J Discuss	sion	92 06	
1.0		J1011		

4.7	Conclusion	99
CONC	LUSION AND RECOMMENDATIONS	01
BIBLI	DGRAPHY1	.05

LIST OF TABLES

		Page
Table 2.1	Summary of the different configurations used for the training of both the datasets used in this study: CamVid, NYUv2 (40-class version), and CHAOS	
Table 2.2	Comparison of accuracies and IoU on CamVid and NYUv2 for both the regular cross-entropy loss and our weighted version	40
Table 2.3	Values of the MCE and ECE on the three datasets considered in this study, for various calibration methods	41
Table 2.4	Quantitative results relative to the values obtained by the objective function on the test set of CamVid for both the training with the regular cross-entropy loss function and with our weighted cross- entropy. The "no threshold" columns present results for experiments in which we consider that every prediction is accurate (i.e. no pixel discarded). The numbers in the rows containing false positives and false negatives are expressed as a percentage over the entire number of pixels	
Table 2.5	Quantitative results relative to the values obtained by the objective function on the test set of NYUv2 for both the training with the regular cross-entropy loss function and with our weighted cross-entropy	
Table 2.6	Quantitative results relative to the values obtained by the objective function on the test set of CHAOS for both the training with the regular cross-entropy loss function and with our weighted cross-entropy	46
Table 2.7	Quantitative results for the regression of the threshold value on the test set of the CamVid and the NYUv2 datasets	47
Table 2.8	Quantitative results obtained for the regression of the threshold value on all three datasets	47
Table 3.1	Vessel overlap score (Dice coefficient) for different number of epochs used during the training of the model	62
Table 3.2	Vessel overlap score (Dice coefficient) and image quality (PSNR) for our method, Tmenova et al. (Tmenova, Martin & Duong, 2019)	66

Table 4.1	Summary of results for the geometric model dataset. Results are presented for both the left coronary artery (LCA) and right coronary artery (RCA) using the Dice coefficient and two 3D distances
Table 4.2	Results for the patient dataset both for the left coronary artery (LCA) and right coronary artery (RCA) using the Dice coefficient and a 2D

LIST OF FIGURES

Page

Figure 0.1	Overall methodology for the proposed research
Figure 1.1	Anatomy of the heart Taken from wikipedia.org (2017)
Figure 1.2	Schematic of the two distinct phases of the cardiac cycle: the diastole (left) and the systole (right) Taken from wikipedia.org (2017)
Figure 1.3	Position of the coronary arteries Taken from wikipedia.org (2022)
Figure 1.4	Stent positioning Taken from Wikipedia.org (2017) 11
Figure 1.5	Example of an X-ray angiography of coronary arteries
Figure 1.6	Catheterization laboratory in Sainte-Justine Hospital14
Figure 1.7	Schematic explaining the behavior of X-ray image acquisition
Figure 2.1	Entire pipeline of the proposed method. The discarded pixels in the last image are colored in black
Figure 2.2	Comparison of histograms of uncertainty, for the same image of the test set of the CamVid dataset, between the regular cross-entropy loss function (left) and our weighted cross-entropy loss function (right). The bars in blue represent the well-classified pixels, while the red ones correspond to wrong classifications. Displayed in log scale to ease the visualization
Figure 2.3	Comparison of the reliability diagrams for the uncalibrated (left) and calibrated (right) uncertainty distributions on the NYUv2 dataset 40
Figure 2.4	Comparison of uncertainty distribution for two images of the CHAOS dataset
Figure 2.5	Decomposition of the objective function (first column) into the number of false positives (second column) and false negatives (third column) for the unweighted (top row) and weighted (bottom row) versions of the loss function. The red line represents the optimal threshold value given by the minimum value of the objective function 43
Figure 2.6	Norm of the gradient of the objective function for the unweighted and weighted versions of the loss function

Figure 2.7	Visual results of the segmentation. The first image belongs to the CamVid dataset, the second one from the CHAOS dataset and the third one was extracted from NYUv2. The information is organized according to the following lines: original images, ground truth segmentation, predicted segmentation, correctness of the prediction (good predictions in yellow, wrong predictions in purple), thresholded segmentation using our method (the pixels discarded due to high uncertainty are shown in black) and thresholded segmentation with a naive threshold (half of the interval of uncertainty)
Figure 3.1	Examples of images from our datasets. The first row contains simulated images while the second row displays patient data images 54
Figure 3.2	Proposed method. Two mapping functions, G and F are used to respectively transfer the style from domain A to domain B and from B to A . A cycle consistency is used to insure that the transfer of style can be reverted back to the initial style. We propose a new vesselness consistency that compares the vessels in the input image with the vessels contained in the stylized output
Figure 3.3	Vessel overlap score (Dice coefficient) and image quality estimation (PSNR and SSIM) for different weighting of the vesselness loss used during the training of the model
Figure 3.4	Preliminary results of transfer of style from domain A to domain B. Columns organized as follows: input images, style transfer from input A to B, style transfer from stylized B to A (cycle consistency) 64
Figure 3.5	Preliminary results of transfer of style from domain A to domain B. Columns organized as follows: input images, style transfer from input B to A, style transfer from stylized B to A (cycle consistency) 65
Figure 4.1	Example of coronary arteries from the geometric model dataset. The first two lines are arteries extracted from the Rotterdam coronary artery algorithm evaluation challenge: left coronary arteries in first line, and right coronary arteries in second line. The last two lines contain coronary arteries simulated using the XCAT: left coronary arteries in the third line, and right coronary arteries in the fourth line 75
Figure 4.2	Visual representation of the proposed methodology. (a) During training phase, the mesh deformation network uses visual features extracted by the Perceptual feature extractor on an input image in order to gradually deform an initial shape of a coronary artery.

Figure 4.7 Visual results for the reconstruction of coronary arteries from the test split of our patient imaging dataset. Each line contains a different reconstruction and is organized as follows: X-ray angiography image used as input for the reconstruction, followed by the reconstruction result oriented on the same view, a second X-ray angiography image from a different angle with the reconstruction rotated for that view94

Figure 4.3

Figure 4.4

Figure 4.5

Figure 4.6

Figure 4.8 A violin plot of the distribution of reprojection errors for all the patients in patient imaging dataset is displayed on the left of the figure. The results are separated for the left (LCA) and right coronary arteries (RCA), and a comparison is made with results obtained with the original Pixel2Mesh model. The distribution of the reprojection errors for all our patient imaging data is given on

LIST OF ABBREVIATIONS

BNN	Bayesian Neural Network
CAD	Coronary Artery Disease
CD	Chamfer Distance
CNN	Convolutional Neural Network
СТ	Computed Tomography
EMD	Earth Mover Distance
FCN	Fully Connected Network
GAN	Generative Adversarial Network
GCN	Graph Convolutional Network
IoU	Intersection over Union
KL	Kullback–Leibler
LAO	Left Anterior Oblique
LCA	Left Coronary Artery
MCE	Maximum Calibration Error
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
NIR	Near InfraRed
NN	Neural Network
OCT	Optical Coherence Tomography

XXII

PCI	Percutaneous Cardiac Intervention
PSNR	Peak Signal-to-Noise Ratio
RAO	Right Anterior Oblique
RCA	Right Coronary Artery
RGB	Red Green Blue
RNN	Recurrent Neural Network
SCD	Symmetric Chamfer Distance
SSIM	Structural Similarity Index Measure
VAE	Variational AutoEncoder
VSS	Vesselness
3D	Three dimensional
2D	Two dimensional

INTRODUCTION

Coronary artery diseases are one of the top causes of death in the world (Kaptoge *et al.*, 2019). To treat those diseases, cardiologists favor, when possible, minimally-invasive interventions such as cardiac catheterization. During a cardiac catheterization procedure, a guide wire is inserted into the vessels of the patient from his femoral artery (Bergersen, Foerster, Marshall & Meadows, 2008). A catheter is then inserted around the guide wire from the same area. At this point, two options are generally available to the cardiologist. The first one consists in guiding an inflatable balloon to the narrow part of the diseased artery by following the guide wire. The balloon is then inflated to dilate the artery. The second option involves the use of a stent. The stent is crimped around the balloon which is inserted in the same fashion as described previously. The main difference is that when the balloon will be sufficiently inflated, the stent will also expand and serve as a scaffolding keeping the vessel in a dilated state. The current imaging gold standard used for the guidance of percutaneous interventions is X-ray angiography. It consists of injecting some contrast agent into the soft vascular structures of the patient in order to render them visible on X-ray images. The main challenges of such interventions are linked to the stent itself. It is indeed a challenging task to evaluate with precision the appropriate diameter of the stent to use (Jordi Oller, Gundelwein, Miro & Duong, 2016; Gundelwein et al., 2018). Providing cardiologists with an optimal and accurate choice of stent optimized for a particular patient would greatly ease a stenting intervention and enhance its accuracy. However, even with optimal prior information, such intervention is still difficult to carry out and some problems remain. First of all, the contrast agent is quickly diluted in the blood of the patient, thus the window during which the arteries are visible on the angiography is limited. Second, because of their elastic properties and the blood pressure, the arteries move and are distorted throughout the cardiac cycle. Those deformations make it difficult for a cardiologist to precisely reproduce a given stent placement. Finally, angiography images present various ambiguities linked to the 2D projection of 3D structures. Indeed, the projection can create overlap between parts of the

arteries, and occlusion caused by the rips appearing before the arteries. These ambiguities are one of the reasons why the interpretation of angiography can be challenging. Therefore, a 3D online representation of the arteries could make cardiac navigation guidance easier, require less experience from the operator, and reduce the exposure of the patient to X-rays and contrast agents while also diminishing the number of cases requiring a new intervention. Numerous 3D reconstruction techniques have been proposed during the last decades (Wang *et al.*, 2018b; Habert, Dahdah & Cheriet, 2012; Herman & Meyer, 1993) but, to our knowledge, they tend to either not be adapted to the reconstruction of long and thin structures and require heavy calibration that is not necessarily practical in an intervention setting where the cameras do not have a fixed position in space. Additionally, the results provided by the current 3D reconstructions. In a domain such as medical imaging, this can be a limiting factor to the use of those results in an online intervention setting.

0.1 **Problem statement and motivations**

In the last 10 years or so the scientific community saw the rise of neural networks and deep learning frameworks. Such learning models are widely used in various domains such as autonomous driving vehicles (Valiente, Zaman, Ozer & Fallah, 2019), automated production (Yue, Ping & Lanxin, 2018), face recognition (He, Wu, Sun & Tan, 2018), etc. Recently, in the medical imaging field, solutions involving the use of deep learning frameworks have been proposed (Martin, Miró & Duong, 2019). However, even though deep learning models yield satisfying accuracies, the results they provide are not straightforward to always explain. Moreover, it is difficult to estimate how confident a model is about its predictions.

Percutaneous navigation guidance is, to date, limited to the use of 2D images. Because of various ambiguities caused by the 2D projection of 3D structures (occlusions, overlaps, etc.) and limitations linked to the use of X-ray imaging and a contrast agent (rapid dilution, poor

contrast, etc.), this navigation can become difficult and require a lot of experience from the cardiologist. We believe that 3D cardiac navigational guidance, however, could bring a solution to the depth-related problems encountered when solely using 2D projections of 3D structures. To this end, a 3D model must first be acquired. Current reconstruction methods are either not adapted to long and thin tubular structures or tend to use color information that is not available in medical images.

0.2 Research objective and contributions

The **main objective** of this thesis is to provide uncertainty modeling in the end-to-end 3D reconstruction of coronary arteries from X-ray angiography.

To guide cardiologists during percutaneous interventions, a 3D model of the arteries of interest must be calculated or extracted from pre-operative modalities. Recently, the literature saw a large rise in the number of different 3D reconstruction techniques available for various applications. Most of those methods require the use of prior segmentation which is still, to our knowledge, an open problem for arteries in X-ray imaging. Additionally, registration procedures require those segmentation masks as well. Moreover, the segmentation obtained by state-of-the-art convolutional neural network models rarely comes with a reliable confidence metric. Recently, robust Bayesian networks brought to the hands of the research community a way to evaluate the uncertainty of predictions. To date, this measure is only used for performance quantification and is not integrated into an automatic process designed to increase the performance of the network. Additionally, we observed that the amount of uncertainty yielded by pixel-wise segmentation of vascular vessels in X-ray angiography (Martin *et al.*, 2019) does not directly correlate to the accuracy of the given segmentation. Hence our first objective is the following:

Objective #1: Pixel-wise confidence estimation for segmentation in Bayesian Convolutional Neural Network (CNN) (contribution #1, Chapter 2).

First, we propose an epistemic uncertainty calibration technique designed to be used with CNN.

Epistemic uncertainty is a measure of confidence that one can have in the prediction of a network. We aim to calibrate the epistemic uncertainty to, not only increase our level of confidence about the results but also increase the accuracy of the segmentation. This objective is validated on different datasets: two composed of natural images and one containing 3D brain images. Those datasets have been chosen because they contain a large variety of different scenes, texture patterns, and colors because they contain multiple classes to segment, and finally because of their wide use in the literature on Bayesian networks.

The second step of this objective proposes a classification technique designed to assess the maximum level of acceptable uncertainty for a given segmentation. To do so, a neural network is trained to perform image-specific thresholding of the yielded uncertainty values. Uncertainty values deemed too high are then flagged and given to an operator, alongside the segmentations. Uncertainty can then be propagated to other tasks, such as 3D reconstruction to estimate the reliability of the reconstruction given the uncertainty in the input images.

Studies in the automatic processing of medical images, often lack available annotated data. This is due to multiple constraints: ethical and privacy issues, experience required to perform the annotation of the data, and the intra- and inter-rater variability. Furthermore, rare cases are under-represented in such datasets. These lead to models that either can't learn enough of the target task or generalize poorly to slightly different datasets and rare cases. Data augmentation techniques have been proposed to solve this issue. These solutions range from simple geometric transformations to more sophisticated techniques that involve the use of deep neural networks (Tmenova *et al.*, 2019; Hou *et al.*, 2017). However, none of those techniques, to the best of our knowledge, proposes the end-to-end synthesis of coronary artery angiography.

Objective #2: Unsupervised synthesis of realistic coronary artery X-ray angiography (contribution #2, Chapter 3).

First, a realistic cardio-respiratory simulator is used. It allows setting a large variety of parameters that will customize the shape and movement of the arteries on the resulting X-ray angiography.

Then a vesselness-aware adaption of the CycleGAN network is used to transfer the image-specific style from the X-ray angiography to the obtained simulations. Synthesized X-ray angiography can then be used to increase the accuracy of the segmentation task (objective #1) and the 3D reconstruction (objective #2) with a better representation of unseen rare cases.

Percutaneous cardiac navigation is a challenging task as it is guided by 2D projections of 3D structures. The acquisition process leads to many ambiguities including overlapping, occlusions, etc. These ambiguities hinder the interpretability of the X-ray angiography and make the navigation greatly dependent on the experience of the operator. We believe that a 3D visualization of the coronary arteries during percutaneous interventions could ease the navigation of catheters and hence improve the outcome of such operations while reducing the number of cases requiring a follow-up additional intervention. Current reconstruction techniques rely on the use of biplane or rotational X-ray angiography acquisition which require costly synchronization and motion compensation operations between views. Therefore our third objective is the following:

Objective #3: Single-view 3D reconstruction of coronary arteries using mesh deformation networks (contribution #3, Chapter 4).

The reconstruction is performed using a mesh deformation network. This particular version of CNN is designed to directly adapt the shape of a mesh defined as a graph. A graph representation of the arteries allows for the reconstruction of them as meshes, as opposed to voxel reconstruction, and hence offers the expression of relations between different parts of the arteries at a better memory efficiency (Kato, Ushiku & Harada, 2018). The model is trained using phantoms of coronary arteries generated using the XCAT realistic cardiovascular simulator (Segars, Sturgeon, Mendonca, Grimes & Tsui, 2010) and tested on real patient data.

Figure 0.1 gives an overview of the research objectives described above, and the links between them.



Figure 0.1 Overall methodology for the proposed research

CHAPTER 1

LITERATURE REVIEW

1.1 Anatomy of the heart

The heart is the body's engine. It is located in the thoracic region, the chest, right between the two lungs. This organ is, in fact, a hollow muscle that contracts and releases to feed the entire body with oxygen. As can be seen in Figure 1.1, the heart is composed of two ventricles: the left (9) and the right (10) ventricles. Each of those ventricles has a specific physiologic particularity that is tightly linked to the cardiac cycle. From the right ventricle arises the main pulmonary branch that is later split into two branches: the left and the right pulmonary arteries (5). Each of the pulmonary arteries reduces its size along its path to the lungs and bifurcates further and further to create millions of capillaries. The aorta (4) comes out of the left ventricle. It rises in direction of the head, to then adopt a U-shape that brings it to split in each leg.



Figure 1.1 Anatomy of the heart Taken from wikipedia.org (2017)

The cardiac cycle is composed of two phases: the diastole and the systole. Each of those two phases corresponds to a specific movement of the myocardium, the muscle composing the heart. During the systole (Figure 1.2), the right ventricle expels deoxygenated blood into the pulmonary arteries, using the main pulmonary trunk. Exactly at the same time, the right ventricle sends oxygenated blood, coming from the lungs, toward the rest of the body. The diastole (Figure 1.2) corresponds to the phase during which the muscles composing the heart will release their contraction. During that time, both ventricles will passively fill in with blood. Those different movements, corresponding to the two phases of the cardiac cycle, are responsible for the blood flow and hence the blood pressure inside the arteries.



Figure 1.2 Schematic of the two distinct phases of the cardiac cycle: the diastole (left) and the systole (right) Taken from wikipedia.org (2017)

1.1.1 Coronary arteries

The wall of an artery is composed of the following three layers ordered from the innermost part of the artery to the outermost (see (SEER Training, 2022)):

- Intima: mainly composed of endothelial cells. It is covered by a lumen that is directly in contact with the blood flow.
- Media: the largest layer. It is mostly composed of elastin, collagen, and muscle cells.
- Adventitia: mainly composed of collagen.

Because the media is much larger than the other layers, the elasticity of the entire artery is often modeled solely based on that particular layer. Indeed, by different contractions and dilations, the muscle cells of the media control the diameter of the lumen. The wall is elastic and adapts to the heartbeats and, hence, to the pressure created by the blood. The elasticity of an artery depends on the age of the patient and the diseases that might affect the artery.



Figure 1.3 Position of the coronary arteries Taken from wikipedia.org (2022)

In this thesis, we focused on the coronary arteries (pictured in red in Figure 1.3). They can be classified as left coronary arteries (LCA) and right coronary arteries (RCA). They wrap the heart muscle and supply it with blood filled with oxygen and nutrients.

1.1.2 Stenosis

Stenosis is a permanent narrowing of the diameter of a blood vessel. This pathology is mainly encountered in adult populations. The main impact of stenosis is the sudden change in blood pressure that it creates inside the vessel. Furthermore, the larger the vessel is, the stronger the effect on the blood pressure is. If left untreated, stenosis can, in the long term, create more respiratory problems and endocardium infections might appear.

Stenoses not only affect vessels but also heart valves (Levine, Coyne & Colvin, 2015). A valve's stenosis occurs when one or more leaflets become thicker and can no longer be completely open. Stenosis on a valve affects the blood flow similarly to stenoses on vessels.

To treat stenosis, cardiologists usually perform catheterization interventions. During cardiac catheterization, a long and thin hollow structure called a catheter is inserted inside the patient's body. The area of insertion can vary depending on the patient and the vessel that requires an intervention. Generally, the access point is on the femoral artery next to the patient's groin (Bergersen *et al.*, 2008). The dimensions of the catheter are chosen before the intervention depending on several factors such as the size of the internal cardiac structures of the patient and also on the type of tools that will be used to treat the stenosis. When possible, the treatment of stenosis, also called angioplasty, consists of dilating the narrowed structure of the vessel using a balloon that is inserted alongside the catheter and then inflated. An illustration of the process is given in Figure 1.4.

For cases when dilating the artery with a balloon is not enough to maintain it in the expanded configuration, cardiologists use a stent. Stents ensure longer-lasting results and diminish the number of required follow-up interventions. The procedures stay essentially the same except that a stent is crimped around the balloon. When the balloon is exactly positioned in the narrowed part of the artery, the cardiologist inflates it. The stent will be expended by the balloon and will maintain the artery in the dilated position, even after the deflation of the balloon. This procedure is illustrated in Figure 1.4.



Figure 1.4 Stent positioning Taken from Wikipedia.org (2017)

The objective of properly-performed stent placement is to obtain a pressure gradient as close to zero as possible between the entrance and the exit of the stent. The success of this objective mostly depends on the choice of the stent, and in particular its diameter, and also on the exact positioning of the stent. A null gradient is obtained in an ideal case scenario in which the diameter of the previously stenosed part of the vessel is equal to the diameter of the rest of the vessel. The choice of the stent and materials is usually done by the cardiologist relying on his experience and is motivated by various measurements performed during and before the intervention. Recently, some studies focused on proposing techniques capable of calculating the optimal stent to use during a stenting intervention (Jordi Oller et al., 2016; Gundelwein, Miró & Duong, 2015). This calculation is done specifically for one patient and is based on the geometry of the patient's arteries and different measures of pressure. The 3D model of the patient's arteries is commonly obtained several months prior to the intervention which can be a problem when dealing with a pediatric population. Moreover, such methods relying on blood flow simulations are generally time-consuming and therefore cannot be used during real-time interventional guidance. An ideal method would include the use of inline 3D reconstruction using angiography and real-time pressure measurements.

1.2 X-ray angiography

X-ray angiography (example in Figure 1.5) is a well-known medical imaging modality routinely used for cardiac percutaneous interventions. This modality consists in using X-rays to generate images of blood vessels. However, because vessels are composed of soft tissues they are not directly visible solely using X-rays. To visualize the vessels, cardiologists make use of contrast agents injected with a catheter. This contrast agent is non-ionic, non-osmolar, and has a low viscosity. Those characteristics allow for a minimum number of side effects to which the patient is exposed. The amount of contrast agent used depends on a large variety of factors, among which are: the size of the patient, the location of the catheter, the size of the target arteries, etc (Bergersen *et al.*, 2008). The acquisition phenomenon is pictured in Figure 1.7 in which the matrix in the center represents the patient's body. This body is composed of different segments, each having a different absorbance factor. When an X-ray passes through a segment of the body, a certain percentage of this X-ray is absorbed according to the segment's absorbance factor. When the X-rays leave the body, an image is created by a receptor according to the photons remaining in each X-ray. The more X-rays are remaining, the more the image will be composed of light colors. The more photons are absorbed, the more the image will become darker.

An example of a catheterization laboratory is given in Figure 1.6. The general setup of such an intervention room is composed of the intervention table, one or two pairs of X-ray generators, and receptors and screens used to visualize the obtained images.

Different angiography acquisition techniques exist (Grech, Debono, Xuereb, Fenech & Grech, 2012). Three of them are explained in the next paragraphs: monoplane, rotational, and biplane. Each of them has its particularities both in terms of behavior and required material.

Monoplane angiography is the most general way of acquiring angiography. It is performed with a single pair of X-ray emitters and receivers. However, because angiography produces two-dimensional images of three-dimensional objects, having only one view renders them more difficult to understand mostly because of the overlapping of internal structures. Hence, the monoplane angiography acquisition technique requires a more thorough pre-interventional

preparation to choose the orientation of the captor. The experience of the cardiologist along with his knowledge of the human anatomy help to fully understand what is being displayed in monoplane angiography. Even while limiting the visualization of the overall organ structure, this acquisition technique has the advantage to limit the amount of radiation and contrast agent used during the intervention. Furthermore, using only one captor also allows for gaining some space in the catheterization laboratory compared to the other existing acquisition techniques.



Figure 1.5 Example of an X-ray angiography of coronary arteries

The second angiography technique is called biplane X-ray angiography. It consists in taking two X-ray images of the same structure at two different angles at the same time. To do so, two pairs of radiographic sources and detectors are placed around the patient. The advantage of such a system is that it provides quickly and easily two different views without moving the imaging

system and without using more contrast agents than a monoplane acquisition. Similarly, as for monoplane X-ray angiography, it is important to properly choose the positions of the sources and detectors. Indeed, overlaps can occur and hide important areas of the region of interest. For example, stenosis can be hidden by surrounding structures (bones, organs, etc).



Figure 1.6 Catheterization laboratory in Sainte-Justine Hospital

Rotational angiography is the third and last type of angiography acquisition that will be presented in this subsection. It is, compared to the other modalities, a rather new acquisition technique. It consists in using a single radiographic source that is rotated around the patient. During the rotation, several images are acquired at a precise acquisition frequency. The advantage compared to the other acquisition techniques is that multiple views of the region of interest are captured. This allows for a better understanding of the overall vessel structure that can serve as compensation for various occlusions that might occur on some of the views. However, using these acquisition techniques increases the exposure of the patient to X-rays and contrast agents. This additional exposure can be dangerous for the patient both in terms of a higher X-ray dose, and also due to a bigger quantity of contrast agents that can damage the kidneys (Caro, Trindade & McGregor, 1991).


Figure 1.7 Schematic explaining the behavior of X-ray image acquisition

1.3 Artery segmentation

Vessel segmentation from X-ray angiography consists of the identification of vessels in X-ray angiography. It is a prolific area of study. The most popular approach is based upon the analysis of the Hessian matrix of the X-ray images - known as vesselness filters, to enhance elongated vessel-like structures (Frangi, Niessen, Vincken & Viergever, 1998). Since then, multiple variations and combinations of the vesselness filter have been used. M'hiri *et al.* (2016) use a graph-based method combined with the vesselness information to perform the segmentation of the coronary arteries. Kerkeni, Benabdallah, Manzanera & Bedoui (2016) performed a multiscale search based on vesselness and direction information to alleviate a loss of segmentation in regions of low vesselness values.

More recently, deep neural networks have been employed for the segmentation of coronary arteries (Fan *et al.*, 2018). Liang *et al.* (2021) proposed a 3D convolutional network for the segmentation of arteries by leveraging temporal information contained in sequences of X-ray angiography. Given enough data used during the training, these methods offer great accuracy but provide little to no explanation about the segmentation process and results.

However, most of those studies focus on the segmentation of only one type of artery, and none, to our knowledge, tries to generalize the learning to multiple different types. Furthermore, no measure of confidence is provided with the obtained segmentation, making their use uncertain in a medical context.

1.4 Bayesian Neural Networks and uncertainties

With the increase in computer performance, machine learning, and particularly Deep Learning, have been the actors of significant advances in artificial intelligence in the past decades. For various tasks, CNN has proven itself to be the current state-of-the-art method. Without requiring much prior knowledge, it allows for solving various computer vision problems if a sufficient amount of data is used. However, even though it can provide satisfying results, the feedback provided by neural networks is limited. Indeed, understanding what is learned by each neuron or even by each layer is of high complexity. It is thus difficult to estimate the confidence of a prediction. Some incidents induced by failures in machine learning predictions have recently been reported. One of the most notorious examples occurred in May 2016 when an autonomous-driving car crashed because its vision system failed by classifying a trailer as a bright sky (National Highway Traffic Safety Administration, 2017). This event resulted in the death of its driver. To avoid these types of incidents, solutions involving the use of Bayesian deep learning have been developed.

Bayesian probabilities, provide two main types of uncertainty that one can model: aleatoric and epistemic uncertainties (Der Kiureghian & Ditlevsen, 2009). The first one represents the uncertainty already contained in the input data. It can be considered as some underlying noise corrupting the data. The sources of noise can be varied: high illumination rendering the image difficult to understand, occlusions, etc. The second type of uncertainty is directly linked to the model. Epistemic uncertainty represents what has, and particularly what has not been learned by the model. For example, feeding a picture of a bird to an algorithm that is trained to recognize images of cats and dogs would lead to high epistemic uncertainty. The combination of both types of uncertainties is called predictive uncertainty. This uncertainty is used to assess a high

level of confidence in regions where the task performed well and can warn us of noisy regions that cannot be explained. Calibrating the predictive uncertainty (i.e. increasing the confidence that we can have in our model's prediction) then results in trying to alter both of the previously mentioned uncertainties. Because aleatoric uncertainty comes from the data acquisition process, our model cannot learn how to calibrate it. One way to affect it would be to change the process of data acquisition, which might not be feasible for most applications. Hence, to calibrate the general predictive uncertainty, one would have to focus on the epistemic uncertainty. Epistemic uncertainty, on the other hand, can be calibrated by incorporating new knowledge. This can be done in multiple ways: using more data, trying to tune the model's parameters, or even reviewing the formulation of the problem. Unfortunately, acquiring more data is always a challenging task in most domains, such as medical imaging. Furthermore, data augmentation does not appear like a plausible option in areas where rotations, cropping, or even lightning changes do not make sense regarding the real-world problem being solved. Finally, because of their black-box behavior, Neural networks provide low flexibility regarding how the problem is formulated.

1.4.1 Uncertainty estimation

Bayesian Neural Network is a Bayesian twist of more classical neural networks. Their key advantage is that they allow for the use of prior and the modeling of uncertainty. The main idea in such networks is to update the initial belief on the parameters p(w) when data D arrived to estimate the posterior probability, p(w|D). It is formulated as follows using Bayes' theorem:

$$p(w|D) = \frac{p(D|w)p(w)}{p(D)}$$
(1.1)

In practice, such calculation is intractable and is thus not feasible in terms of time complexity. Consequently, different approximations have been developed allowing us to estimate the posterior probability. Kononenko (1989) introduced some early work on Bayesian neural networks. Their method consisted of the application of Bayesian inferences in the computation of the activation of each neuron of the network. In this paper, the authors want to update the state of a neuron of the network V_{mn} given the state *S* after activation. The calculation of the posterior probability

 $P(V_{mn}|S)$ is made possible with the use of Bayes' theorem. This calculation inherently infers the new state of the neuron. It is performed using the following rules (Kononenko, 1989):

$$V'_{mn} = \begin{cases} 1 & \text{if } A(V_{mn}|S) > P(V_{mn}) \\ V_{mn} & \text{if } A(V_{mn}|S) = P(V_{mn}) \\ 0 & \text{if } A(V_{mn}|S) < P(V_{mn}) \end{cases}$$
(1.2)

where $A(V_{mn}|S)$ is used instead of the posterior $P(V_{mn}|S)$ which, given the assumptions used in the paper, doesn't lead to real probabilities (according to Coxx's axioms). Because of the use of Bayesian probabilities, this type of neural network has the advantage of allowing one to extract uncertainties and add prior knowledge to the computation of the solution. However, the method doesn't provide strict probabilities which can raise concerns regarding the quality of the uncertainties obtained. Moreover, the study is limited to the use within neural networks.

More recently, Gal & Ghahramani (2016) showed that the approximation of the Bayesian inference in neural networks and CNNs (i.e estimating the posterior probability) could be done by performing dropout on some units of the layers. In the study, the weights of the network, ω , are defined as follows:

$$\omega = \{W_i\}_{i=1}^L \tag{1.3}$$

In the previous equation, W_i represents the weights placed on the units located on the layer *i*. Because of the intractability of the posterior probability, the authors define a distribution $q(\omega)$ to help calculate an estimate. This distribution is defined by:

$$W_i = M_i.diag([z_{i,j}]_{i=1}^{K_i})$$
 (1.4)

where M_i is a matrix containing the variational parameters, z_i is a Bernoulli distribution and K_i is the size of the layer *i* (Gal & Ghahramani, 2016). The following step consists in minimizing the Kullback–Leibler (KL) distance between $q(\omega)$ and the posterior $p(\omega|X, Y)$. More recently, similar ideas have been developed to perform Bayesian estimation in neural networks. They rely

mostly on changes in the distribution of the dropout probability. Kingma, Salimans & Welling (2015) changed the Bernoulli distribution by a Gaussian. Using continuous relaxation and a custom optimization function, Concrete Dropout (Gal, Hron & Kendall, 2017) allows for the modeling of a Bayesian network will offering tunable dropout probabilities. R-Dropout (Zhang, Wang, Lu, Wang & Yin, 2017) is an extension of the previous works which focuses on the use of dropout along with convolutional layers. All these studies provide satisfying Bayesian estimates of neural networks while allowing to model uncertainties and enhancing the prediction results.

Once the training is finished, the remaining task is to estimate the uncertainty. More details on this matter are given in the methodology section.

1.4.2 Uses of uncertainty and uncertainty reduction

Uncertainties raised from probabilistic models have been used in several ways to either assess the certainty of the predictions or even, in certain cases, to enhance the results of the prediction.

In the context of multi-task learning, Kendall, Gal & Cipolla (2018) used the aleatoric uncertainty to optimally estimate the weight given to each task-related loss function in the global loss function. They showed that this method not only finds appropriate weights for the loss functions but it also provides an increase in performance for each of the given tasks.

Early approaches to epistemic uncertainty reduction (Caicedo & Zárate, 2011) focused on general Bayesian problems. Using a modified version of the Hop Skip Jump method (Stapleton, 1977), the authors achieved significant epistemic uncertainty reduction on the results yielded by Bayesian Inference. Because it uses the likelihood of a model, this technique is not directly applicable to Neural Networks. Indeed Bayesian Neural Networks only provide an estimate of the posterior probability and do not focus on the intractable likelihood.

In Neural Networks, to our knowledge, no method currently focuses on how to calibrate this uncertainty. However new studies focusing on the introduction of Bayesian convolutional neural

networks (as presented in section 1.4.1) pave the way for the implementation of more confident deep models.

1.5 Generative adversarial networks and style transfer

Generative networks are an easy and convenient way to generate realistically-looking new samples of a given dataset. The most popular current generative networks include variational autoencoders (VAE) and generative adversarial networks (GAN).

GANs (Goodfellow *et al.*, 2014) are composed of two neural networks that compete with each other: a generator and a discriminator. The discriminator is used to determine if a new sample is similar to the input dataset. The generator's task is to generate new samples that look realistic to the discriminator.

Style transfer is an application of GANs that is used to change the style of a set of images to closely match the style of another set of images. CycleGAN (Zhu, Park, Isola & Efros, 2017) leverages the idea of cycle consistency to ensure that we can transfer the style from one domain to another, and still be able to return to the original domain. In2I (Perera, Abavisani & Patel, 2018) is an extension of CycleGAN that proposes to add an input image of a different modality to provide more information concerning the structures of interest.

Style transfer methods have been applied to different medical imaging problems. Sun *et al.* (2022) used a multi-discriminator version of CycleGAN to generate pseudo-CT images from MRI. Wang, Lei, Curran, Liu & Yang (2021) used an attention-based CycleGAN to transfer style between different MRI modalities. CycleGAN has also been used as a means of denoising low-dose cardiac CT (Gu, Yang, Ye & Yang, 2021).

No direct application to arteries, except previous work (Tmenova *et al.*, 2019) which considered the transfer of style from real coronary angiography X-ray images to simulations. Using CycleGAN, we found out that the style could properly be transferred to XCAT simulations and

provide realistically looking images. However, parts of the arteries were lost in the stylization, hence limiting the synthesis of new realistic data for medical imaging purposes.

1.6 3D reconstruction and registration

In this section, we first describe 2D/3D registration techniques. We then present different state-of-the-art reconstruction techniques used both for general computer vision and then more specifically in the context of artery 3D reconstruction.

1.6.1 2D/3D registration

2D/3D registration methods are used to find the transformation that aligns a 3D model to its 2D representation (Unberath *et al.*, 2021).

While rigid registration will not modify the shape of the 3D model, but only translate and rotate it, nonrigid registration also considers the deformation between the original 3D shape and the projection. Several 2D/3D registration methods have been proposed for coronary arteries (Matl, Brosig, Baust, Navab & Demirci, 2017).

Some registration methods of vascular structures are based on the use of a priori topological information (Groher, Zikic & Navab, 2009). Others (Pinheiro, Kybic & Fua, 2016) represented the problem as tree search exploration. Rivest-Henault, Sundar & Cheriet (2012) proposed a two-steps method for nonrigid registration. First, an affine transformation is used to align the 3D centerline with the vessel contained in the X-ray angiography. Then a nonrigid registration is performed to account for any modification in the shape of the arteries. Recently, Yoon, Yoon & Lee (2021) proposed a nonrigid registration of the centerline of a pre-operative computed tomography angiography of coronary arteries onto X-ray angiography. The method is based on 3D cage deformation techniques and stochastic optimizations. While the registration error is low (2.0 mm), no information about the lumen of the artery is registered.

One of the limits of such methods is the availability of recent pre-operative models. Indeed, preoperative acquisitions are often performed months before interventions. In pediatric populations, in particular, large anatomical changes may hence be observed between the two imaging modalities. To palliate the lack or unavailability of pre-operative models, 3D reconstruction methods have emerged that can then be registered onto upcoming X-ray angiography images.

1.6.2 General 3D reconstruction

This subsection presents different techniques designed for 3D reconstruction. 3D reconstruction is the process of obtaining the three-dimensional shape of an object from one of multiple two-dimensional projections. It is an underconstrained problem (Li, Liu & Tang, 2007) even in the case of multi-view reconstruction.

State-of-the-art techniques function using a varying number of views: while some studies are specifically designed to work using a single view (Pan, Han, Chen, Tang & Jia, 2019), others take benefit of multiple views of the same object to obtain more 3D information (Xie, Yao, Sun, Zhou & Zhang, 2019). Xie *et al.* (2019) showed that additional views could help reach more accurate reconstructions to a certain number where the accuracy starts to plateau. One of the limitations of multi-view reconstruction is it often requires, especially in the medical imaging field, to capture both views exactly at the same time. One alternative to this limitation is to include in the reconstruction process a movement-correction registration or cardiac gating.

Voxel-based techniques have been very popular in the literature of the last decade. Space carving (Kutulakos & Seitz, 2000) was one of the first studies to propose the 3D reconstruction of general objects from multiple views without a priori geometric information. While this method provides good results, one of its drawbacks is that it requires a rather large number of input views that may not be available in an interventional setting for example.

Gwak, Choy, Chandraker, Garg & Savarese (2017) proposed the use of a recurrent network designed to perform 3D reconstruction from one to more views. When multiple views are used, latent results from each view are combined in the same fashion as recurrent neural networks. The

combination is then used to generate a 3D model. While this method presents good results, some of its drawbacks are that it relies on learning shapes from very specific categories that present very strong inter-category similarities while also taking as input pictures solely containing the object to reconstruct (i.e. without background). Furthermore, RNN-based solutions have been proven to not provide consistent results given differently ordered images while also having memory loss problems (Xie *et al.*, 2019).

Xie *et al.* (2019) recently proposed a method capable of doing 3D reconstruction with a varying number of views (from 1 to 20 shown in the experiments). The method is based on an encoder-decoder architecture that performs the reconstruction on each view. All the obtained models are then combined in a "context-aware fusion" before being refined. Their results show that such a method scales its accuracy with the number of used views up to a plateau in which additional views do not provide more accuracy. Other types of autoencoders have been employed for volumetric reconstruction

Using voxel representation for 3D reconstruction has several advantages, most of them linked to the ease of representation of the data. Indeed, it is convenient to use matrix representation for both the input images and the desired 3D model.

The drawback of the voxel-based reconstruction method is that they rely on a compromise between the accuracy of reconstruction and processing time. Small voxels offer great accuracy, while bigger voxels will provide a coarser reconstruction. However, the smaller the voxels are, the more time is needed to compute the reconstruction (Martin, Vachon, Miró & Duong, 2017). Further, the required memory is also important as the size of the computed matrix scales with the number of voxels.

Recent mesh-based reconstruction techniques have been proposed based on learning algorithms work that modifies an initial shape according to projections of the object to reconstruct (Pan *et al.*, 2019). Pixel2Mesh (Wang *et al.*, 2018b) proposed the 3D reconstruction of various objects (household objects, planes, etc) from a single image. The model leverages the knowledge acquired during training about the object to reconstruct and its relation with their 2D projection.

Mesh-based reconstruction allows increasing the accuracy of the reconstruction at no additional computational cost.

1.6.3 Artery reconstruction

This subsection describes different state-of-the-art 3D reconstruction techniques used nowadays for the reconstruction of vascular structures. Most of those studies focus on the reconstruction of coronary arteries (Brost, Behar, Rinaldi, Housden & Rhode, 2017).

Graph-based methods are really popular to perform 3D reconstruction. Unberath, Achenbach, Fahrig & Maier (2016) developed a graph-cuts extension by specializing in the task of coronary artery reconstruction. By minimizing an energy function based on the expansion of the graph-cuts algorithm, the authors obtain a depth label. This label is then used to deduce the 3D position of a point corresponding to identified 2D points on some images. By applying this method to multiple points of the centerlines of the 2D images, a 3D centerline is computed. A volume is then constructed around the extracted 3D centerline using the diameter of the artery identified on the angiography images. This method provides accurate results with an error margin of 0.5 cm. However, it requires the use of rotational angiography acquisition. As mentioned in previous sections, this acquisition technique can be dangerous for the patient and requires a large area around the intervention table to properly rotate around the patient. Indeed, a large number of images needs to be fed to the algorithm to obtain satisfying results.

A graph-based 3D reconstruction technique of large vascular structures was proposed using biplane angiography acquisition (Martin *et al.*, 2017). In other terms, using only two different views of the same anatomical structure, the study aimed to evaluate the degree of achievable reconstruction accuracy. Some limits are drawn by this study. First, the calibration of the cameras used in the biplane acquisition is a key factor impacting the accuracy of the reconstruction. Additionally, the initial segmentation used for the reconstruction is also constraining the accuracy of the reconstruction. The more accurate the segmentation is, the more accurate the reconstruction will be. Furthermore, the use of only two views and no prior information

to perform 3D reconstruction is a limiting factor for the accuracy of the obtained model. Adding prior knowledge such as the tube-like aspect of the artery would improve the results. As an alternative, more views could be used, but this would mean exposing the patient to additional doses of contrast agents and X-rays. Finally, no temporal information is used for the reconstruction. Hence, the reconstruction is done for each pair of angiography at each time step, instead of deducing a deformation pattern based on the cardiac and respiratory movements.

In the current literature, problems are encountered when the pair of images used are neither isocentric nor orthogonal. To counter this challenge, Kunio *et al.* (2017) proposed a novel reconstruction technique for the centerline of a vessel. However, this study only considers the centerline, and more efforts have to be put to extend it to a full 3D reconstruction of the shape of the artery.

1.7 Summary

X-ray angiographies are X-ray images in which a contrast agent is inserted into the blood structures of the patient to make the soft tissues visible in the image. The main difficulties encountered during such interventions are linked to the imaging quality that hence affects their interpretability. We hence believe that real-time 3D representations of the coronary arteries could ease percutaneous intervention guidance.

Segmentation of the coronary artery is still to date an open research problem. Various solutions have been proposed over the years. While those methods provide good segmentation accuracies, they often lack interpretability and a measure of confidence regarding the segmentation. In medical image analysis errors in segmentation of a few millimeters can be fatal. Hence, providing a measure of uncertainty in addition to the segmentation could greatly facilitate the inclusion of such techniques in catheterization processes.

Current 3D reconstruction methods for coronary arteries use the information provided by either biplane or rotational X-ray angiography. Although these methods create reconstructions with good accuracy, they often require the use of synchronization between views and require specific

equipment in the catheterization laboratories. Using a single view for reconstruction while leveraging the knowledge acquired about the general shape of an artery during training could provide accurate reconstruction as shown with single-view natural image reconstruction.

Learning-based medical image analysis often lacks available annotated data. Additionally, available datasets do not always fully represent the entire spectrum of possible data with rare cases being under-represented or not represented at all. Data augmentation techniques have been proposed to counter the problem of scarcity of data. These methods however fail to provide automatic annotation and generate out-of-distribution data. On the other hand, cardio-respiratory simulators have been developed to generate fully-customizable anatomies. Although the simulators offer realistically generated anatomies, the resulting X-ray images do not seem real as they do not contain X-ray angiography-specific noise and artifacts. Hence style-transfer methods could help enhance the realism of the provided simulations.

CHAPTER 2

PIXEL-WISE CONFIDENCE ESTIMATION FOR SEGMENTATION IN BAYESIAN CONVOLUTIONAL NEURAL NETWORKS

Rémi Martin^a, Luc Duong^a

^a Department of Software and Information Technology Engineering, École de Technologie Supérieure, 1100 Notre-Dame Ouest, Montréal, Québec, Canada H3C 1K3

Revisions submitted, October 2022

2.1 Abstract

Bayesian Convolutional Neural Networks represent an emerging state-of-the-art computer vision framework. This particular version of CNN allows the modeling of two types of uncertainties: epistemic uncertainty and aleatoric uncertainty. These uncertainties permit to evaluate the degree of confidence in predictions yielded by a model. Nevertheless, few studies automatically integrate uncertainty in an end-to-end prediction pipeline. Our research proposes a novel way to assess the degree of confidence in yielded predictions. To that end, a model is trained using the cross-entropy loss function. Thereafter, the learning of this model is resumed with a pixel weighting dynamic calculation designed to reduce the uncertainty of well-classified pixels and penalize wrong classifications. From the epistemic uncertainty measures provided by these two trainings, a histogram is calculated. A final neural network model is used to determine an interval of confidence for the predictions. This interval defines the pixels to be considered with caution at test time, depending on the uncertainty they yield. Validation is performed and shows that the uncertainty yielded by our sample weighting provides a better confidence interval than the regular, unweighted, cross-entropy loss function. Furthermore, our expected calibration error averaged over all datasets (0.07) is lower than available methods (0.1 for the state-of-the-art, 0.18 without calibration). Furthermore, the proposed uncertainty-based thresholding provides better accuracy than baseline uncertainty thresholding, while also minimizing the number of confident errors.

2.2 Introduction

The last decades have seen marked jumps in computer performance, machine learning, and particularly, Deep Learning, which have in turn led to significant advances in artificial intelligence (Litjens *et al.*, 2017). For computer vision tasks, Convolutional Neural Networks (CNN) are currently the basis of state-of-the-art architectures. Without explicitly requiring much prior knowledge, it allows solving various computer vision problems simply using a sufficient amount of data. Examples include medical image analysis (Amirkhani, Barshooi & Ebrahimi, 2022), object tracking (Barshooi & Amirkhani, 2022), autonomous driving car (Aladem & Rawashdeh, 2020), and so on. However, even though neural networks provide satisfactory results, the feedback they provide is limited. Indeed, understanding what is learned by each layer or even by each neuron is challenging. As a result, it is difficult to estimate the confidence of a prediction. A neural network can become highly confident in false predictions for various reasons, including wrong ground-truth annotation, a lack of data, etc.

Some incidents induced by wrong predictions in machine learning predictions have recently been reported (Guynn, 2015; National Highway Traffic Safety Administration, 2017). One of the most notorious examples occurred in May 2016 when an autonomous car crashed because its vision system failed by classifying a trailer as a bright sky (National Highway Traffic Safety Administration, 2017). To understand the leading factors of such incidents, solutions involving the use of uncertainty estimation have been developed (Konyushkova, Sznitman & Fua, 2019; Kohli & Torr, 2008), with the most popular revolving around the use of Bayesian probabilities.

Bayesian probabilities divides uncertainty into two main types that can be modelled: aleatoric and epistemic uncertainties (Der Kiureghian & Ditlevsen, 2009). The first one represents the uncertainty already contained in input data. It can be considered as underlying noise corrupting the data. Noise sources can be of various forms, including various lighting effects (e.g. reflection, shadows) making affected parts of the image difficult to understand, occlusions, etc. The second type of uncertainty, epistemic uncertainty, is directly linked to the model, and, represents what has, and particularly what has not been learned by the model. For example, feeding a picture

of a bird to an algorithm that is trained to distinguish between images of cats and dogs would lead to high epistemic uncertainty. The combination of both types of uncertainties is called predictive uncertainty (Gal & Ghahramani, 2016). This uncertainty is used to assess a high level of confidence in regions where the task is performed well, and can warn us about noisy regions that cannot be explained away.

Bayesian Neural Networks (BNN) (Kononenko, 1989) represent a convenient way to introduce uncertainty estimation into Neural Networks (NN). BNN are versions of NN in which prior distributions are placed on the weights and biases of the networks, instead of fixed values (Neal, 2012). Recently, Gal & Ghahramani (2016) showed that Bayesian variational inference in deep neural networks can be approximated by performing a Bernoulli dropout. Extending research has considered dynamic dropout probabilities (Gal *et al.*, 2017), as well as different distributions for the dropout probability (Zhang *et al.*, 2017; Kendall & Gal, 2017).

In an ideal situation, the predictive uncertainty should be low for most pixels, with the exception of wrongly predicted pixels. However, this ideal is rarely achieved, due mostly to a lack of knowledge on the target domain. Furthermore, Sensoy, Kaplan & Kandemir (2018) showed that, because of the frequentist approach of regular Neural Networks and the use of Softmax, the networks fail to infer the predictive distribution variance and have a tendency to inflate the probability of the predicted class. This inflation is a contributing factor leading to distributions of uncertainty far from the ideal described above.

Calibrating the confidence of a deep model's predictions thus results in trying to correlate its distribution of confidence to the accuracy of the predictions. For the distribution of uncertainty, this means a negative correlation with the accuracy of the predictions. Because aleatoric uncertainty derives the data acquisition process (Der Kiureghian & Ditlevsen, 2009), our model cannot learn how to alter its distribution. One way to modify this distribution involves working on the data acquisition process, which may not be feasible in many applications. To alter the general predictive uncertainty, the focus must therefore be placed on the epistemic uncertainty. However, the epistemic uncertainty distribution is directly linked to the knowledge gathered

by the network. Hence, an alteration of this uncertainty can be achieved in multiple ways: using more data, trying to tune the model's parameters, or even reviewing the formulation of the problem. Unfortunately, acquiring more annotated data is a costly task in many domains, including medical imaging and self-driving vehicles. Furthermore, regular data augmentation techniques do not appear to be a plausible option in domains where rotations, cropping, or even changes in lighting do not make sense regarding the real-world problem at hand or do not provide enough additional knowledge. Because of their black-box behavior, neural networks provide close to no explanation regarding how the problem is formulated.

To avoid highly impactful incidents in domains such as autonomous-driving cars and medical imaging, it is important to limit the number of wrong predictions yielded with low uncertainty. In other words, we favor the rejection of accurately predicted samples when there is high uncertainty as a means of reducing the chances of making wrong predictions with low uncertainty.



Figure 2.1 Entire pipeline of the proposed method. The discarded pixels in the last image are colored in black

Current measures of uncertainty do not directly reflect the correctness of the predictions. Indeed, good predictions can be yielded with high uncertainty as well as wrong predictions can come with low uncertainty. Some existing uncertainty or confidence calibration methods have been

proposed (Cheng, Zhang, Yuan & Sun, 2019; Guo, Pleiss, Sun & Weinberger, 2017; Kurz *et al.*, 2022). These are either however not adapted to pixel-wise classification or are directly used to modulate the predicted class probabilities without necessarily reducing the number of wrong predictions with low uncertainty. Similar studies have been proposed for the calibration of uncertainty (or confidence) in the specific context of semantic segmentation. Popular approaches include: ensembling methods (Mehrtash, Wells, Tempany, Abolmaesumi & Kapur, 2020), temperature scaling (Ding, Liu, Xiong & Shi, 2020), histogram binning (Fingscheidt, Gottschalk & Houben, 2022). None of these methods however focuses on the discarding of highly risky pixels out of an image-specific interval of confidence. Our research therefore focuses on defining a calibrated measure of pixel-wise uncertainty for semantic segmentation in Bayesian CNN. This calibration method allows us to implicitly define different penalties for wrong predictions with low uncertainty, and good predictions with low uncertainty. This measure of uncertainty is then used to define an image-specific interval of confidence outside of which predictions should be used with caution, or discarded. Figure 2.1 provides an overview of the entire pipeline of the proposed methodology.

The contributions of this work are the following:

- 1. The calibration of epistemic uncertainty for CNN-based pixel-wise semantic segmentation;
- 2. The proposal of an energy function designed to minimize the number of wrong predictions with low uncertainty and the number of good predictions with high uncertainty;
- 3. The definition of an image-specific interval of confidence.

The manuscript is organized as follows. The methodology is described in section 2.3, starting with the calibration of uncertainty in section (2.3.1), the definition of the interval of confidence (2.3.2), and the evaluation metrics (2.3.3). Results are presented in section 2.4.

2.3 Methodology

The method described below uses a Bayesian CNN (i.e. a CNN with distributions modeled over its weights) for pixel-wise classification in natural images. We define X as the input of the

network (i.e. images), y as the ground truth corresponding to the input images, and $\hat{y} = f(X)$ as the predictions of the network f.

Following Gal & Ghahramani (2016)'s framework, we ensure that the networks that we use contain dropout before every weight layers while also activating dropout at test time.

The calculation of the epistemic uncertainty, written U in this paper, follows the formulation given in Eq. 2.1:

$$U(\hat{y}) = \frac{1}{T} \sum_{t}^{T} H(\hat{y_{t}})$$
(2.1)

In this equation, T refers to the number of Monte Carlo samplings performed for the estimation of the epistemic uncertainty, and H is Shannon's entropy. The activation of the dropout at test time along with the T Monte Carlo predictions allow for randomness between predictions of the same sample, and hence, provide an estimate of the epistemic uncertainty (Gal & Ghahramani, 2016).

2.3.1 Epistemic uncertainty calibration

To ease the calculation of our confidence interval, we first counter the inflating effect of the Softmax activation function. To this end, we calibrate the distribution of the epistemic uncertainty yielded by the model. The idea here is that this uncertainty should span the entire interval of possible values $[0; -\log(1/C)]$, where C represents the number of classes. Furthermore, the distribution of uncertainty should ideally reflect the accuracy of the classification to properly recognize potentially-wrong predictions.

This is done by initially training the model using the regular cross-entropy loss function to obtain the desired level of accuracy. As a second step, we resume the training with the same loss function multiplied by new pixel-specific weights. This separation into two distinct trainings is in fact very similar to the behavior of transfer learning. It ensures that the network will first learn weights and features for the task at hand, and then slightly adapt them to modify the distribution of epistemic uncertainty.

In an ideal case scenario (i.e a perfect calibration), we would want the network to be confident low uncertainty - about well-classified predictions while yielding high uncertainty for wrong classifications. This scenario is implemented by the network in two steps. Firstly, the distribution of probabilities for the classification of one pixel must reflect the confidence of the network. Ideally, when the network has enough information, the predicted class should have a probability close to 1, and the pixel should thus yield an uncertainty close to 0 (see Eq. 2.1). However, if the knowledge accumulated by the network to make correct predictions regarding a specific sample is insufficient, the probability for each of the classes should be approximately the same, therefore leading to high uncertainty. Secondly, the network must be coherent during multiple Monte-Carlo predictions.

The calculation of the pixel-specific weights is based on the amount of epistemic uncertainty yielded by the model for each of the pixels. For wrongly classified pixels, we want to increase the yielded uncertainty while reducing the value for well-classified pixels.

To do so, we first define a label for well-classified pixels, $w_L = \{w_L^i | i = 1..N\}$ such that:

$$w_L^i = \begin{cases} 1 \text{ for } y^i = \hat{y}^i \\ 0 \text{ for } y^i \neq \hat{y}^i \end{cases}$$
(2.2)

where *i* represents the index of the currently considered pixel, with a total of *N* pixels. The label for wrongly classified pixels, $w_R = \{w_R^i | i = 1..N\}$ is defined similarly, with:

$$w_R^i = 1 - w_L^i$$
 (2.3)

To increase the epistemic uncertainty of wrongly classified pixels, we penalize them based on how far they are from the maximum possible uncertainty. This maximum is expressed as $-\log(1/C)$ as it is the maximum achievable epistemic uncertainty for a classification in *C* classes. The weight for wrongly classified pixels therefore becomes:

$$\mathcal{L}_{w_R}(y,\hat{y}) = \mathcal{L}_{cross-entropy}(y,\hat{y}) \times w_R \times (-\log(\frac{1}{C}) - U(\hat{y}))$$
(2.4)

where $U(\hat{y})$ is the epistemic uncertainty of the predictions and $\mathcal{L}_{cross-entropy}$ represents the cross-entropy loss function. In this equation, w_R is used to only penalize wrongly classified pixels.

To decrease the epistemic uncertainty of the correct classifications, we want to penalize them in the loss function by how far they are from the minimum achievable uncertainty, i.e. 0. This is expressed as:

$$\mathcal{L}_{w_L}(y, \hat{y}) = \mathcal{L}_{cross-entropy}(y, \hat{y}) \times w_L \times U(\hat{y})$$
(2.5)

where w_L is used to only affect well-classified data.

Hence, the weighted loss function is expressed in the following way:

$$\mathcal{L}(y,\hat{y}) = \mathcal{L}_{cross-entropy}(y,\hat{y}) + \alpha \mathcal{L}_{w_R}(y,\hat{y}) + \beta \mathcal{L}_{w_L}(y,\hat{y})$$
(2.6)

The regular cross-entropy loss is added to the two weighted losses to maintain coherence in the predictions. α and β constants are respectively used to define the importance given to the penalty of wrong and correct classifications.

Because correctly estimating the epistemic uncertainty is a time-consuming process, due to the Monte Carlo sampling (predicting T times and averaging for each sample of the entire dataset), we decided to use only the entropy of the class probabilities as a measure of uncertainty during the calculation of the loss. In the results section, we will show that while that formulation does not totally emulate the epistemic uncertainty, it is a good approximation that can successfully be used to change the distribution of the uncertainty. During the training process, the epistemic uncertainty is thus calculated using the approximation presented in Eq. 2.7:

$$U(\hat{y}^{i}) \approx \tilde{U}(\hat{y}^{i}) = H(\hat{y}^{i}) = -\sum_{c}^{C} \hat{y}_{c}^{i} \log(\hat{y}_{c}^{i})$$
(2.7)

where y_c^i represents the probability for pixel *i* to belong in class *c*. At test time, however, Eq. 2.1 is still used to report the full epistemic uncertainty of the results.

2.3.2 Confidence interval calculation

The second step of the methodology consists in calculating an interval of calibrated uncertainty where the pixels included can be used with a low risk of error. This is done through a neural network trained for regression.

Table 2.1Summary of the different configurations used for the training of both the
datasets used in this study: CamVid, NYUv2 (40-class version), and CHAOS

				Number of epochs	
Dataset	Model	Learning rate	Batch size	1st training	2nd training
CamVid	FCN DenseNet	1e-3 w/ RMSProp	3	150	89
NYUv2	Refinenet	1e-5 w/ Adam	16	100	50
CHAOS	FCN DenseNet	1e-3 w/ RMSProp	6	150	75

Because each image yields highly different uncertainty distributions, we want to make our training image-specific and not based on the entire dataset. In other words, the goal is as follows: through the training set we want to learn how to take the uncertainty distribution of an image as input, and find a threshold value that minimizes the number of wrongly-classified pixels on the left (low uncertainty) and the number of well-classified pixels on the right (high uncertainty).

First, we took each image of our dataset, computed the predictions, and from the logits, deduced the epistemic uncertainty (not the approximation) using our weighting alongside the categorical cross-entropy. For each image, we grouped the uncertainty values into a histogram. This approximation of the distribution by a histogram makes the computation less expensive by reducing the dimensionality of the input and reduces the degree to which the regression is affected by outliers.

To prepare the ground truth for our regression network, we first defined an objective function to minimize. For each image, we wish to find the threshold value l, in terms of uncertainty, that minimizes the number of good predictions with high uncertainty and the number of wrong predictions with low uncertainty. This objective function hence depends on the separated distributions of uncertainty for well- and wrongly-classified pixels. This function is defined as

follows:

$$E(l) = \sum_{m}^{M} \gamma \frac{1}{1 + e^{\alpha(m-l)}} + \frac{1}{1 + e^{-\alpha(m-l)}}$$
(2.8)

where *M* is the entire histogram of uncertainty for one image and *m* represents one bin of that histogram. The constant α represents the steepness of the sigmoid functions, while γ controls the balance between well-classified pixels above the threshold value (referred to as false negative or FN) and wrongly-classified pixels below that value (referred to as false positive or FP).

We iterate through the different possible threshold values ranging from $[0; -\log(1/C)]$ to obtain the ideal threshold value, called *l* in Eq. 2.8. Because that range is infinite, we make sure we use the iterative steps that match the size of the different bins of the histograms, to allow the computation.

This objective function cannot be applied directly at test time as we do not know which pixels have been correctly classified by our initial network. Thus, the histogram values are fed into a neural network, along with the ground truth threshold values. This network learns how to use the general histogram of an image (containing the uncertainties of both w_L and w_R) to threshold it according to the objective function defined above. The network is composed of 6 dense layers respectively of size: 2048, 512, 128, 32, and 8. It is trained with the logcosh loss function and optimized using the Adam optimizer with default parameters. The loss function has been chosen for its relative similarity with the least squared error, while also preventing the loss to be too affected by outliers and errors of predictions.

Because the first network (the one solving the semantic segmentation task) returns different uncertainty distributions for the training set and the other data sets, respectively, we decided to train the latest network using a mix of training and validation sets to avoid overfitting on the training set. This will ensure that our network does not get specialized on the thresholding of the initial training set, and will hopefully ensure a better generalization.

2.3.3 Metrics

The proposed methodology is validated using various metrics, aimed to show the usefulness of the method for the calibration as well as for the discarding of highly uncertain pixels.

For the calibration of the uncertainty distributions, the results are mainly evaluated using three metrics: the histograms of uncertainty for well- and wrongly classified pixels, Reliability Diagrams (Niculescu-Mizil & Caruana, 2005), Expected Calibration Error (ECE), and Maximum Calibration Error (MCE) (Naeini, Cooper & Hauskrecht, 2015).

Reliability diagrams present the correlation between the confidence of predictions (x-axis) and their accuracy (y-axis). The closer the bars of this diagram are compared to a diagonal line, the better the calibration. The same idea is conveyed numerically by the MCE measure. It is defined as follows:

$$MCE = \max_{m \in M} |acc(\hat{Y}_m) - m|$$
(2.9)

M represents different bins of the confidence distribution calculated as $Conf(\hat{y}) = 1 - (U(\hat{y})/max(U))$. $acc(\hat{Y}_m)$ is the average accuracy of the predictions with confidence $m \in M$. Similarly, the ECE is defined as follows:

$$ECE = \sum_{m \in M} \frac{|m|}{n} |acc(\hat{Y}_m) - m|$$
(2.10)

The discarding of pixels is evaluated by comparing the number of false positives and false negatives of our methodology with different baselines.

Additionally, the overall performance of the semantic segmentation is reported using measures of accuracy and of overlapping (i.e. Intersection-over-Union, abbreviated IoU). This measure is also known as Jaccard index or Tanimoto index.

2.4 Experiments

The experiments presented in this section were performed on two different CNN models: Refinenet (Lin, Milan, Shen & Reid, 2017) and FCN Densenet (Jégou, Drozdzal, Vazquez, Romero & Bengio, 2017) of respectively 23M and 18M parameters. These networks were chosen for the high accuracy they offer in classification tasks and because they use the Softmax activation function as well as dropout layers. Along with the previously mentioned models, we used two different natural image datasets: CamVid (Brostow, Fauqueur & Cipolla, 2009; Brostow, Shotton, Fauqueur & Cipolla, 2008) and NYUv2 (Nathan Silberman,Derek Hoiem & Fergus, 2012). Both datasets are used for semantic segmentation. The first one contains images of streets acquired by cameras mounted on cars. The second one is a dataset of interiors. A full description of the experimental configurations is presented in Table 2.1. A third dataset has been used: CHAOS (Kavur, Selver, Dicle, Barış & Gezer, 2019), a medical imaging dataset containing MRI and CT images from 40 different patients. The task associated with the dataset is to segment multiple organs (liver, kidneys, and spleen). For this research, we chose only one modality of this dataset: the T2-weighted MRI images, as it provides sufficient but challenging contrast for the organs to segment.

Our entire methodology takes on average a day to train for each dataset on a Titan X GPU. Our custom calibrating loss function converges in about 50 to 60 epochs, depending on the dataset.

2.4.1 Calibration of uncertainty

The experiments on both datasets were conducted using a value of $\alpha = 4$ and $\beta = 1$ in Eq. 2.6. These values have been selected experimentally. The training history of both the regular loss and our weighting of the categorical cross-entropy, averaged out over 5 different trainings each, first shows that the convergence of the loss is not affected by the weighting method. It seems that both the weighted and the unweighted losses converge in about the same number of epochs. Second, as expected, the weighted loss function values are globally higher than the unweighted loss values due to the added penalties.



Figure 2.2 Comparison of histograms of uncertainty, for the same image of the test set of the CamVid dataset, between the regular cross-entropy loss function (left) and our weighted cross-entropy loss function (right). The bars in blue represent the well-classified pixels, while the red ones correspond to wrong classifications. Displayed in log scale to ease the visualization

Table 2.2 depicts a comparison of image-wise accuracies between the regular cross-entropy and our weighted loss on both datasets. As can be seen, our weighting of the regular cross-entropy loss function does not suffer from a loss of accuracy, and instead offers a slight increase of accuracy: 0.4% for CamVid and 2.3% for NYUv2. Hence, we see that calibrating the distribution of uncertainty does not negatively affect the accuracy of the model. In other words, there is no need for a compromise to meet the two objectives: predicting correctly and having a confidence closely correlated to the accuracy. Identical observations can be made on the NYUv2 and CHAOS datasets. This increase in accuracy is in line, if not greater, with results found by Kendall & Gal (2017) on CamVid (67.1 IoU) and Cheng *et al.* (2019) on different datasets.

A comparison of the histogram of uncertainty between the unweighted loss (left histogram) and of our custom weighting of the cross-entropy loss function (right histogram) is given in Figure

2.2 for CamVid. The two histograms displayed in the figure correspond to the same image, with the two versions of the loss. First, it can be seen that in the regular cross-entropy loss function the distribution of uncertainty does not span the range of possible values $[0; -\log(1/C)]$. As seen in Sensoy *et al.* (2018), this is mostly due to the tendency of the Softmax activation function to inflate the probability of the most probable class.

Table 2.2 Comparison of accuracies and IoU on CamVid and NYUv2 for both the regular cross-entropy loss and our weighted version

Dataset	Cross-entropy	Accuracy	IoU
ComVid	Regular	88.5%	65.2
Calliviu	Weighted	88.9%	66.1
NYUv2	Regular	63.1%	38.1
	Weighted	65.4%	38.7
CHAOS	Regular	94.3%	89.3
	Weighted	95.7%	92.4



Figure 2.3 Comparison of the reliability diagrams for the uncalibrated (left) and calibrated (right) uncertainty distributions on the NYUv2 dataset

Second, the major density of wrongly classified samples is located towards the end of the spectrum for our custom weighting, which is not the case for the unweighted loss that yields uncertainty distributions that tend to overlap more between the well- and wrongly-classified pixels. Finally, the amount of right classifications with very low uncertainty (close to 0) is

greater in our weighted loss. However, our weighted cross-entropy fails to reduce the number of well-classified uncertain samples. These uncertain good predictions are not necessarily a bad sign as they could be the result of a lack of knowledge regarding the classified areas or may even be due to randomness.

Dataset	Calibration	MCE	ECE
CamVid	Uncalibrated	0.48	0.23
CamVid	Mehrtash et al. (2020)	0.15	0.10
CamVid	Ours	0.11	0.06
NYUv2	Uncalibrated	0.19	0.12
NYUv2	Mehrtash et al. (2020)	0.17	0.10
NYUv2	Ours	0.09	0.06
CHAOS	Uncalibrated	0.40	0.18
CHAOS	Mehrtash et al. (2020)	0.14	0.09
CHAOS	Ours	0.15	0.08

Table 2.3 Values of the MCE and ECE on the three datasets considered in this study, for various calibration methods

Reporting uncertainty seems to be important in cases of both good and wrong predictions. Indeed, good but uncertain predictions are no better than wrong confident predictions. Hence, it seems that our weighting of the cross-entropy loss function properly offers a similar, if not slightly greater, accuracy while, also offering a better calibration of the uncertainty distribution, that is a distribution more correlated with the correctness of the predictions. The same observations can be drawn from the distributions of the uncertainty of the two other datasets.

Figure 2.3 presents reliability diagrams: one for the uncalibrated setting (image on the left), and one for the calibrated setting (right). The reliability diagrams are displayed for the test split of the NYUv2 dataset. Similar results are obtained for the two other datasets. As can be seen, the confidence distribution is now closer to a straight line of slope 1. It means that the calibrated confidence is more directly correlated to the accuracy of the predictions (i.e., the uncertainty is inversely correlated to the accuracy). The performance of the calibration is also shown in Figure 2.4 which displays visual results regarding the difference in uncertainty distribution with and

without calibration. Two scenarios are presented: one image that presents several segmentation problems for both training and one where the segmentation performs well. In this figure, we can visually see that the calibrated uncertainty correlates more with the correctness of the prediction.



Figure 2.4 Comparison of uncertainty distribution for two images of the CHAOS dataset

This observation is further confirmed by Table. 2.3 that presents results of the MCE and ECE metrics. The table shows that the proposed calibration process properly offers, for most datasets, lower MCE and ECE values than the state-of-the-art method (Mehrtash *et al.*, 2020) meaning that both the mean and maximum errors of uncertainty calibration are reduced. Additionally,

our method provides a reduction of both MCE and ECE compared to Kendall & Gal (2017)'s uncertainty calculation on the CamVid dataset (Uncalibrated version in the Table).



Figure 2.5 Decomposition of the objective function (first column) into the number of false positives (second column) and false negatives (third column) for the unweighted (top row) and weighted (bottom row) versions of the loss function. The red line represents the optimal threshold value given by the minimum value of the objective function

In this subsection, we showed an improvement in uncertainty calibration using our weighting version of the cross-entropy loss function. This better calibration is a first step towards identifying wrong predictions.

2.4.2 Uncertainty classification

The results presented in this section were calculated while considering a value of $\gamma = 0.5$. This value was chosen, for both datasets, to slightly favour a reduction of false positives, while also accounting for the imbalance between the number of correct and wrong classifications.



Figure 2.6 Norm of the gradient of the objective function for the unweighted and weighted versions of the loss function

The first column of Figure 2.5 gives an overview of our objective function for different thresholding values both for the original cross-entropy loss (first line) and our weighted version (second line). Additionally, the figure provides a decomposition of the objective function into the number of false positives (second column) and false negatives (third column). What can be seen is that, for both losses the minimum value (see y-axis) of the objective function is almost the same. However, the biggest difference is that because our distribution of uncertainty spans the entire range of possible values, an error in the regression of the threshold value (more details in the upcoming section) will have a lesser impact on our weighted loss than on the original loss. This is further confirmed by Figure 2.6 which displays the norm of the gradient of the objective function for both scenarios. It can be seen in this figure that in the case of the regular weighted loss, the gradient is globally much smoother than the gradient of the unweighted loss, with an almost constant increase of the gradient and fewer bumps making it less prone to be affected by errors of regression of the threshold value.

Tables 2.4, 2.5, and 2.6 present some numerical results concerning our objective function designed to find a ground-truth threshold value for the histograms of uncertainty. We can see that our proposed weighted function properly reduces the number of wrong predictions with low

uncertainty (referred to as false positives in the table) that we make as compared to the regular cross-entropy loss function with the same γ (see Eq. 2.8).

Table 2.4 Quantitative results relative to the values obtained by the objective function on the test set of CamVid for both the training with the regular cross-entropy loss function and with our weighted cross-entropy. The "no threshold" columns present results for experiments in which we consider that every prediction is accurate (i.e. no pixel discarded). The numbers in the rows containing false positives and false negatives are expressed as a percentage over the entire number of pixels

	Threshold accuracy	Percentage FP	Percentage FN
Cross-entropy objective function	78.1%	4.0%	17.9%
Cross-entropy no threshold	88.5%	11.5%	0.0%
Weighted cross-entropy objective function	78.4%	2.8%	18.8%
Weighted cross-entropy no threshold	88.9%	11.1%	0.0%

Table 2.5Quantitative results relative to the values obtained by the objective functionon the test set of NYUv2 for both the training with the regular cross-entropy lossfunction and with our weighted cross-entropy

	Threshold accuracy	Percentage FP	Percentage FN
Cross-entropy objective function	71.3%	5.2%	23.5%
Cross-entropy no threshold	63.1%	36.9%	0.0%
Weighted cross-entropy objective function	71.6%	3.4%	25.0%
Weighted cross-entropy no threshold	65.4%	34.6%	0.0%

However, the number of false negatives (i.e. good predictions with high uncertainty) is higher with our proposed weighting, which is consistent with the original assumptions made during the design of the loss function, which boils down to a preference to avoid making wrong predictions with low uncertainty. The table also presents a comparison of our objective function with results that include no discarding process, which is considering every single prediction as well-predicted (i.e. no pixel discarded). Because the accuracy of the initial classification is high on CamVid, classifying all predictions as true predictions yields a high threshold accuracy. However, we can see that the total number of false positives, i.e., the number of wrong predictions with low uncertainty, is much higher if all predictions are considered accurate.

	Threshold accuracy	Percentage FP	Percentage FN
Cross-entropy objective function	92.3%	1.2%	6.5%
Cross-entropy no threshold	94.3%	5.6%	0.0%
Weighted Cross-entropy objective function	93.2%	0.7%	6.7%
Weighted cross-entropy no threshold	95.7%	4.3%	0.0%

Table 2.6Quantitative results relative to the values obtained by the objective functionon the test set of CHAOS for both the training with the regular cross-entropy lossfunction and with our weighted cross-entropy

2.4.3 Histogram thresholding

Table 2.7 provides quantitative results for our threshold-value regression. As can be seen, the predicted threshold values are close to the values returned by our objective function: an average error of 0.12 for CamVid, 0.08 for NYUv2, and 0.07 for CHAOS. Table 2.8 shows that this error in the regression of the threshold values leads to a decrease of 2% of false positives on average, along with a simultaneous increase of false negatives, which means that the regression of the threshold value can avoid having a large number of wrong predictions with low uncertainty but at the cost of a larger number of discarded pixels. Tuning the parameters of the objective function (i.e. γ) and the model (i.e. α and β in the loss function) is one way to adjust that balance between false negatives and false positives depending on the number of wrong predictions with low uncertainty one can afford to make on the targeted task.

Figure 2.7 presents visual results of the entire pipeline of the proposed method: from the original images, to the semantic segmentation and then the discarding of highly uncertain pixels. As

can be seen in the fourth and fifth rows, the pixels discarded due to high uncertainty seem to be located in regions containing the highest densities of wrong classifications. Most of these regions correspond to boundaries between objects, or appear to be very small objects. This result is coherent with previous findings (Cheng *et al.*, 2019). This high density of uncertainty in boundaries and small objects is mostly due to the inability of the network to distinguish between small structures and likely due to a lack of knowledge regarding some objects. More pixels are discarded by the confidence interval in NYUv2 and in CamVid than in CHAOS. This is due to the fact the first two datasets contain a greater number of small objects and a greater variety of classes, leading to lower accuracy in the segmentation of the images. This accuracy is reflected in the uncertainty of the pixels, as defined as our objective, leading to more discarded pixels. The last line of the figure pictures results with naive thresholding: rejection of all pixels with uncertainty in the higher half of the uncertainty distribution. This line shows that our discarding method offers a smaller rate of rejection of pixels than a naive approach that tends to reject more well-classified pixels.

Table 2.7 Quantitative results for the regression of the threshold value on the test set of the CamVid and the NYUv2 datasets

	Prediction error		
	(out of -log(1/C))		
	Min Mean Max		
CamVid	2.61×10^{-6}	0.12	0.32
NYUv2	4.27×10^{-5}	0.08	0.44
CHAOS	6.13×10^{-5}	0.07	0.23

Table 2.8	Quantitative results obtained for the
regression	n of the threshold value on all three
	datasets

	CamVid	NYUv2	CHAOS
Threshold accuracy	75.4%	64.2%	92.0%
Percentage FP	0.8%	1.2%	0.3%
Percentage FN	23.8%	34.6%	7.7%

2.5 Conclusion

This article presents a novel method for using epistemic uncertainty in the context of pixel-wise segmentation using Bayesian CNN. Our method consists in estimating an interval of confidence outside of which pixel-wise segmentation results should be considered with caution. The results obtained show that the proposed loss is capable of calibrating the uncertainty distribution by enhancing the correlation between the uncertainty distribution and the correctness of the predictions. Additionally, our discarding method produces a low number of wrong predictions with low uncertainty, compared to current methods designed to handle the uncertainty of CNN predictions. We also showed that discarding highly uncertain pixels, instead of modulating the risks of using the obtained segmentations. The experiments are conducted on three different datasets (2 composed of natural images, 1 with medical images) which shows a good insensibility to multiple domains of application differently impacted by unhandled wrong predictions. Future works will focus on including attention information in the calculation of the confidence interval.



Figure 2.7 Visual results of the segmentation. The first image belongs to the CamVid dataset, the second one from the CHAOS dataset and the third one was extracted from NYUv2. The information is organized according to the following lines: original images, ground truth segmentation, predicted segmentation, correctness of the prediction (good predictions in yellow, wrong predictions in purple), thresholded segmentation using our method (the pixels discarded due to high uncertainty are shown in black) and thresholded segmentation with a naive threshold (half of the interval of uncertainty)
CHAPTER 3

UNSUPERVISED SYNTHESIS OF REALISTIC CORONARY ARTERY X-RAY ANGIOGRAPHY

Rémi Martin^a, Paul Segars^b, Ehsan Samei^b, Joaquim Miró^c, Luc Duong^a

 ^a Department of Software and Information Technology Engineering, École de Technologie Supérieure, 1100 Notre-Dame Ouest, Montréal, Canada
 ^b Center for Virtual Imaging Trials, Carl E. Ravin Advanced Imaging Laboratories, Department

of Radiology, Duke University Medical Center, Durham, United States

^c Department of Pediatrics, CHU Sainte-Justine, Montréal, Canada

Paper submitted for publication, July 2022

3.1 Abstract

Purpose Medical image analysis suffers from a sparsity of annotated data necessary in learningbased models. Cardiorespiratory simulators have been developed to counter the lack of data. However, the resulting data often lacks of realism. Hence, the aim of the proposed method is to synthesize realistic and fully customizable X-ray angiography of coronary arteries for the training of learning-based biomedical tasks.

Methods 3D models of coronary arteries are generated with a fully-customizable realistic cardio-respiratory simulator. The transfer of X-ray angiography style to simulator-generated images is performed using a new vesselness-specific adaptation of the CycleGAN model. The CycleGAN model is paired with a vesselness-based loss function that is designed as a vessel-specific structural integrity constraint.

Results Validation is performed both on the style and on the preservation of the shape of the arteries of the images. The results show a PSNR of 14.125, a SSIM of 0.898 and, a Dice overlap of 89.5

Conclusion We proposed a novel fluoroscopy-based style transfer method for the enhancement of realism of simulated coronary artery X-ray angiography. The results show that the proposed

model is capable of accurately transferring the style of X-ray X-ray angiography to the simulations while keeping the integrity of the structures of interest (i.e. the topology of the coronary arteries).

3.2 Introduction

The scarcity of data and mainly of annotated data is a major problem in machine-learning based applications. This problem is even more predominant in medical imaging as problems of ethic and privacy impair the sharing of data between research entities. Additionally, rare phenomenon are often misrepresented in datasets, leading to unbalanced classes or skewed representations of the data.

Data synthesis is a technique that not only allows to perform data augmentation, but also enables the generation of fully customizable data and automatically annotated data. Examples of data synthesis are numerous. They include the generation of rare classes or rare examples such as the synthesis of traffic light (Wang, Ma & Yang, 2022), the generation of images of rare eye diseases (Yoo, Choi, Kim, Ryu & Kim, 2021), or the synthesis of realistic anterior segment optical coherence tomography (Zheng et al., 2021). Domain transfer between different imaging modalities are also performed with the generation of angiography images from retinal fundus photographs (Tavakkoli, Kamran, Hossain & Zuckerbrod, 2020). However, most of those methods 1) generate the samples from their representation of the input data, hence still under-representing rare cases, and 2) don't add further real domain constraints regarding the generated structures. Those problems affect the use of current synthesis methods for medical image generation. Simulators offer solutions to synthesize realistic medical images. The XCAT is a cardio-respiratory simulator capable of generating various anatomical structures, such as coronary arteries, while simulating their respective movements (e.g. cardiac and respiratory). The obtained images however lack realism as they are noise-free and do not mimic the real medical image style.

Style transfer consists in transferring the style of one set of images to another. Examples include changing all images of cats to images of dogs, applying Picasso's style to images and so on. Two

different kinds of style transfer are explored in the literature: paired and impaired style transfer. Paired style transfer considers that there is an explicit one-to-one mapping from every image of set A to an image of set B. Unpaired style transfer does not learn from explicitly defined pairing between images of one set to the other set but rather tries to find a transformation that satisfies transfer conditions. Paired image-to-image translation is however difficult to implement as it requires explicit pairing which may not be available depending on the application. As such, unpaired style transfer offers more flexibility on that regard and is considered in the current study.

CycleGAN (Zhu *et al.*, 2017) aims to perform unpaired image translation by adding a cycle consistency constraint to the training process. That constraint ensures that going from a domain A to a domain B also allows to go back from domain B to domain A without loss of information compared to the input image. In2I (Perera *et al.*, 2018) extends the work done with CycleGAN (Zhu *et al.*, 2017) by adding an input image of a different modality along with a new encoder to guide the GAN network into properly transferring the style from domain A to domain B. That additional input image brings more information about what is contained into the image (e.g. RBG + NIR image as input).

Concerning the direct application of style transfer to coronary artery X-ray angiography, there is, to our knowledge, no other study than Tmenova *et al.* (2019). This study is a direct application of the CycleGAN network to the image-to-image translation between simulations and X-ray X-ray angiography. The results show that X-ray angiography style can properly be transferred to simulations. However, we can observe that this method lacks of structural integrity for the topology of the coronary arteries (i.e. parts of the arteries are lost during the transfer).

As seen with the results obtained in our previous study (Tmenova *et al.*, 2019), performing style transfer onto images of coronary arteries is difficult. Indeed, parts of the arteries tend to vanish in the synthesized images. This is mainly due to the elongated and thin nature of the shape of the coronary arteries. To counter this problem, we propose to use a vesselness enhancement of the arteries that guides the GAN into identifying the main structures of interest.

Hence, in this work, we propose a framework that synthesizes coronary artery X-ray angiography from fully customizable 3D cardiorespiratory simulations. The simulated coronary artery topologies are first projected onto noise-free and contrast-enhanced images. A vesselness-aware GAN is then used to transfer the style of real coronary X-ray angiography onto the projected simulations.

The manuscript is organized as follows. The first section (section 3.3) describes the data used in our research plus the pre-processing applied to it. Section 3.4 contains description of the model and the loss functions used in the proposed method. Section 3.5 presents the results and the discussion relative to our findings.

3.3 Data

Our data is composed of two sets of images: simulated data, containing images acquired from a realistic cardiorespiratory simulator (Segars *et al.*, 2010; Segars *et al.*, 2013, 2015); and patient data composed of real coronary artery X-ray angiography (described in the sections below).



Figure 3.1 Examples of images from our datasets. The first row contains simulated images while the second row displays patient data images

Examples of our datasets are pictured in Figure 3.1 and a description of those datasets is given in the subsections below. As can be seen in this figure, the two datasets are quite different. On the one hand, the images in the patient data are more noisy due to the use of X-ray. The boundaries between the various organs, bones and other structures are often blurry. The arteries are not pictured with a constant pixel intensity due to the rapid dilution of the contrast agent in the blood. Finally these images contain wires, catheters, stents and other clinical objects unseen in simulations. On the other hand, simulated images are a lot smoother, they display sharp edges between organs and do not contain any clinical objects. The projection process does not simulate the dilution and propagation of contrast agent and hence display the entire artery on the rendered images. Furthermore, ribs are not as apparent in the simulation as they would be in real X-ray angiography. Finally, the simulations do not account for the non-uniform X-ray absorptivity in each organs.

3.3.1 Simulated data

Simulations have been performed using the XCAT cardiorespiratory simulator. This simulator proposes a large variety of parameters that affect the simulated anatomies. It is for examples possible to indicate if the patient is healthy, change what motions affects the geometry of the arteries (i.e. respiratory, cardiac, both), the duration of a cardiac cycle etc. The simulator can also be used to project the simulated anatomies onto images. By controlling the absorption coefficient of each organ, it is hence possible to create angiography-like images with visible arteries. The projection process is also fully customizable, with parameters such as the position of the cameras, the focal length and so on.

While the simulator offers accurate coronary artery geometries, the projection process does not generate realistic X-ray angiography images. Indeed, the resulting images lack of noise, contrast agent dilution and other artefacts that compose a real X-ray angiography image.

In overall 54 different anatomies have been used. The simulations have been conducted using both cardio and respiratory motions on a total length of 2 cardiac cycles. We simulated 67 geometries per anatomy, accounting for a total of simulated geometries.

Each of the simulated geometry has been then been projected in order to obtain simulated X-ray images. The projections have been performed on anteroposterior views only. The projection process has been applied on both left and right coronary arteries. We hence obtained a total of 7236 images.

3.3.2 Patient data

The patient data comprises images acquired from pediatric patients in the Sainte-Justine hospital. The set of patients contains both male and female of various ages ranging from a few months old to 18 years old. The set is composed of a total of 8524 images collected from 46 patients.

3.3.3 Pre-processing

Each of the sets listed above (i.e. simulated data and patient data) have been separated into different subsets. These subsets correspond to different angulation of the C-arm during the acquisition of the images. This separation insures that the model is learning a proper mapping between similar views of the arteries in both domains.

Our final subsets are the following: simulated right coronary arterty with anteroposterior (RCAAP) view (3618 images from 54 patients), patient RCAAP (1282 images from 30 patients), simulated left coronary artery with anteroposterior (LCAAP) view (3618 images from 54 patients) and patient LCAAP (1514 images from 34 patients).

Each of these subsets has then been separated in training and test sets with the following respective percentages: 75% and 25%, insuring no common patients between the sets.

3.4 Methodology

Our method aims to transfer the style of X-ray angiography onto images created from simulations of phantoms of coronary arteries. The process is explained in Figure 3.2.



Figure 3.2 Proposed method. Two mapping functions, *G* and *F* are used to respectively transfer the style from domain *A* to domain *B* and from *B* to *A*. A cycle consistency is used to insure that the transfer of style can be reverted back to the initial style. We propose a new vesselness consistency that compares the vessels in the input image with the vessels contained in the stylized output

We feed the network with two styles, each forming a dataset. Dataset A is composed of the simulations described in section 3.3.1 and dataset B is composed of the real patient X-ray angiography (section 3.3.2). The model works in an unsupervised manner, meaning that no

ground-truth stylization is provided for the input images. Instead a cycle-consistency constraint is added to ensure the coherence of the generated images. To do so, the model transfers the style from one domain to the other and then back to the first one. The resulting image should be nearly identical to the input.

A Generative Adversarial Network (GAN) (Goodfellow *et al.*, 2014) is used in order to learn mappings from the domains listed above. The mapping function G is used to transfer images from domain A to domain B, while the mapping function F is used to transfer images from domain B to domain A. The mapping functions are implemented using generators. Full implementation details are explained in CycleGAN's paper (Zhu *et al.*, 2017).

A discriminator is used for each domain. Each discriminator is used to recognize fake samples from real ones for a given domain. These guide the generators to produce more realistic images.

3.4.1 Loss function

The loss function used in our methodology is composed of multiple terms: a GAN loss, a Cycle consistency loss (Zhu *et al.*, 2017), and our proposed vesselness loss term.

The final loss function is expressed as follows:

$$\mathcal{L}(X, Y, G, F, D) = \mathcal{L}_{GAN}(G, D, X, Y) + \alpha_1 \mathcal{L}_{cvc}(X, G, F) + \alpha_2 \mathcal{L}_{vss}(X, G, F)$$
(3.1)

Wehre X and Y are images of both sets, G and F represents the two generators for domain transfer and D is the discriminator.

3.4.1.1 GAN loss

The GAN loss drives the method to learn 1) good reconstruction of input samples, and 2) a discriminator for both style domains. The GAN loss (Goodfellow *et al.*, 2014) is expressed as

follows:

$$\mathcal{L}_{GAN}(G, D, X, Y) = E_{y \sim p_{data}(y)} [\log D_Y(y)] + E_{x \sim p_{data}(x)} [\log(1 - D_Y(G(x)))] + E_{x \sim p_{data}(x)} [\log D - X(x)] + E_{y \sim p_{data}(y)} [\log(1 - D_X(G(y)))]$$
(3.2)

Where D_Y and D_X are the respective discriminators for newly generated images from Y and X domains.

3.4.1.2 Cycle consistency loss

The cycle consistency loss (Zhu *et al.*, 2017) is used to reduce the space of possible generations during the training of the unpaired image to image translation. This is done by checking that the produced stylization can be reverted to the original style by applying the inverse mapping function. The resulting back stylization is then compared to the input image.

The cycle-consistency loss is defined with the following equation as proposed by CycleGAN (Zhu *et al.*, 2017).

$$L_{cyc}(G,F) = E_{x \sim p_{data}(x)}[\|F(G(x)) - x\|_1] + E_{y \sim p_{data}(y)}[\|F(G(y)) - y\|_1]$$
(3.3)

Where *G* and *F* are two functions that map the different image domains *X* and *Y*, and $x \in X$ and $y \in Y$ are different image samples.

3.4.1.3 Vesselness loss

The vesselness loss compares the objects of interests (i.e. the arteries) in both the stylized and the original images. This is to ensure that the entire topology of the artery is kept during the stylization process. It checks that no part of the artery disappears, and no extra parts are added while applying the style to the simulated images.

To do so, we use a vesselness filter that highlights the thin and tubular shapes in an image (Frangi *et al.*, 1998). It uses the eigenvalues of the Hessian matrix. VSS(A) the vesselness values corresponding to image A.

The VSS values from the original and the stylized images are then binarized and compared. This is to ensure that the shape of the artery matches between the original and stylized images.

The binarization is performed using two fixed threshold: one for the patient images and one for the simulated images.

The comparison is performed using the Dice coefficient that measures the overlap between two binary images. It is used with the binarized images and expressed as follows:

$$DICE(X_i, Y_i) = \frac{2|X_i \cap Y_i|}{|X_i| + |Y_i|}$$
(3.4)

Where X_i and Y_i are image from datasets X and Y.

Hence the vesselness loss is expressed using the following equation:

$$\mathcal{L}_{vss}(X, Y, G) = \sum_{i} DICE \left(\sigma(VSS(x_i), \sigma(VSS(G(x_i)))) + \sum_{j} DICE \left(\sigma(VSS(y_j), \sigma(VSS(G(y_j)))) \right)$$
(3.5)

Where $\sigma(.)$ is a thresholding function.

3.4.2 Validation

The validation is performed in multiple steps. A first qualitative study is performed. Then a quantitative evaluation is done.

The qualitative evaluation is performed in two steps. We first discuss the results for the transfer of style in both directions (A to B and B to A). Then we revert the obtained stylized image to it's original style domain (back transformation) and compare it to the original input image.

The quantitative evaluation is performed using two types of metrics. First, the quality of the generated images is checked, then in a second step, we check the integrity of the stylized arteries.

For the quality of the generated images, we compare the peak signal-to-noise ratio (PSNR) of the stylization compared to a set of images of the target dataset.

The PSNR is defined as follows:

$$PSNR() = 10 \times \log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$
(3.6)

Where MAX_I represents the maximum value of an image I, and MSE is the mean squared error.

The preservation of the arteries between the input images and the generations is evaluated in two steps. The arteries are first manually segmented in both sets of images. Then we calculate the overlap between the two segmentations. This is done using the Dice coefficient defined in 3.4.

Using those two metrics the results of our proposed method are compared to results obtained with CycleGAN (Zhu *et al.*, 2017) and IN2I (Perera *et al.*, 2018). The comparison with CycleGAN is performed using fine-tuned results obtained in our previous study (Tmenova *et al.*, 2019). We performed multiple training of the IN2I algorithm and report here the best results that we obtained.

3.5 Results and discussion

3.5.1 Training details

Two training processes have been performed, one for left coronary arteries and one for right coronary arteries.

We used a learning rate of 0.0002. Based on results from our previous work (Tmenova *et al.*, 2019), we experimentally set λ at 20 for RCA and 30 for LCA. This parameters from the CycleGAN models controls the importance of the cycle-consistency loss.

Number of epochs	10	20	30	40	50	60
Vessel overlap (ours)	84	85	87	89	88	86

Table 3.1Vessel overlap score (Dice coefficient)for different number of epochs used during the
training of the model

Table 3.1 provides Dice scores for the overlap of the vessel between the input image and the resulting transfer of style at different epochs used during the training. The best overlap appears to be after training the model for 40 epochs, giving an overlapping score of 89%. Training for more epochs does not increase the accuracy of the style transfer but reather slowly decreases it over time.



Figure 3.3 Vessel overlap score (Dice coefficient) and image quality estimation (PSNR and SSIM) for different weighting of the vesselness loss used during the training of the model

Figure 3.3 shows the influence of the weight assigned to the vesselness loss on the results, as measured with the vessel overlapping and the quality of the image. In terms of preservation

of the arteries, we can observe that the best weight is 35. The image quality, as indicated by the PSNR and SSIM values, does not seem to fluctuate a lot and hence does not seem greatly affected by the weight of the vesselness loss. This indicates that we can improve the preservation of arteries during the transfer of style without scarifying image quality. As a result, we ran the following experiments with a vesselness weight of 35 for both RCA and LCA datasets.

3.5.2 Evaluation

Figure 3.4 shows the results obtained from the transfer of style from domain A (simulations) to domain B (patient data). The first two lines display the results obtained for the left coronary arteries, while the last two represent results for the right coronary arteries.

As can be seen, the model is able to transfer the style of X-ray angiography to images extracted from the XCAT cardiorespiratory simulator without altering the shapes of the organs and other anatomic structures.

We can additionally observe that the pixel intensities of the arteries is slightly altered. The pixel intensity of the arteries become gradually less dark along the arteries. This is due to the parameters used in the vesselness function and also to the dilution of the contrast agent observable in the set of X-ray angiography. Indeed, a lower response to the vesselness filter seems to transfer to a lower pixel intensity in the stylized images.

Moreover, the style transfer creates artifacts around the spine. Indeed dark elongated structures are added at locations of dark intensity pixels in the original XCAT simulation. This is due to the presence of catheters and other interventional tools in our set of X-ray angiography. These artifacts seem to match the location of catheters and guide wires that would be observable in patient angiography. No clear difference in quality of results is observable between LCA and RCA indicating that the model leverages the knowledge accumulated for both types of arteries, without being impacted by the shape of the artery.



Figure 3.4 Preliminary results of transfer of style from domain A to domain B. Columns organized as follows: input images, style transfer from input A to B, style transfer from stylized B to A (cycle consistency)

Figure 3.5 provides a visual of the results obtained for the transfer from domain B to A. The results are organized as follows: LCA is displayed in the first two lines, RCA in the last two. As can be seen, all the main structures and organs are still visible in the stylized image.



Figure 3.5 Preliminary results of transfer of style from domain A to domain B. Columns organized as follows: input images, style transfer from input B to A, style transfer from stylized B to A (cycle consistency)

Although the style of the background is mainly respected, we can observe that the thinnest arteries tend to lose intensity. This is mainly due to the choice of sigmas, σ , used in the

vesselness loss. Indeed, the thin vessels are still present in the stylized images but are displayed with a very light intensity resembling the one in the background.

For both directions of style transfer, reverting the results to their original style provides images visually nearly identical to the input image. This further indicates that the first-way transfer of style does not lose the thin arteries, but displays them in a very light intensity. We can however observe that the reverted images are slightly blurrier.

Table 3.2 gives a comparison of our results with our previous work (Tmenova *et al.*, 2019) and an additional different method (In2I) (Perera *et al.*, 2018). As can be seen, both image quality (PSNR) and artery preservation (Dice) are higher for our method Our experiments using IN2I (Perera *et al.*, 2018) suffered from mode collapse and although the quality of the images was high (high PSNR), the artery preservation was quite low (Dice coefficient of 12%). Results indicated that the model was capable of learning the general shape of the coronary arteries by generating new artery-shaped structures but without considering the shape of the original arteries to stylize. Moreover, from the difference in Dice coefficient between our proposed method and our previous study (Tmenova *et al.*, 2019), we can deduce that the vesselness-based loss contributes to increasing the preservation of arteries during the stylization process.

	Our method	Tmenova et al.	Perera et al.
Vessel overlap	89.5	59.9	12
PSNR	14.125	23.541	13.123
SSIM	0.898	0.849	0.873

Table 3.2Vessel overlap score (Dice coefficient) andimage quality (PSNR) for our method, Tmenova et al.(Tmenova et al., 2019)

3.6 Conclusion

We proposed a novel unsupervised method to synthesize realistic coronary artery X-ray angiography. This is done by transferring the style of X-ray angiography onto simulated images of coronary arteries. The simulations are performed with a realistic cardio-respiratory simulator

that offers a large range of customization. To preserve the coronary artery topology, we introduce a vesselness-based loss function that compares the arteries of the input to those in the stylized image. The results indicate that our method is able to transfer the style from one domain to another (14.125 PSNR) while maintaining the structure of the arteries (89.5% overlap between input and stylized image). Furthermore, our method outperforms the current state-of-the-art methods both in terms of image quality and vessel overlapping. The limitations of the methods are intimately linked to those of the vesselness filter. Further work could consider the use of different vesselness segmentations such as deep learning methods.

CHAPTER 4

SINGLE-VIEW 3D RECONSTRUCTION OF CORONARY ARTERIES USING MESH DEFORMATION NETWORKS

Rémi Martin^a , Theo van Walsum^b , Paul Segars^c , Ehsan Samei^c , Joaquim Miró^d , Luc Duong^a

 ^a Département de Génie Mécanique, École de Technologie Supérieure, 1100 Notre-Dame Ouest, Montréal, Canada
 ^b Department of Radiology and Nuclear Medicine, Erasmus MC, Rotterdam, The Netherlands
 ^c Center for Virtual Imaging Trials, Carl E. Ravin Advanced Imaging Laboratories, Department of Radiology, Duke University Medical Center, Durham, United States
 ^d Department of Pediatrics, CHU Sainte-Justine, Montréal, Canada

Paper submitted for publication, October 2021

4.1 Abstract

Despite X-ray angiography being the current gold standard for coronary artery imaging during percutaneous cardiac interventions (PCI), its interpretation is challenging due to its 2D representation of 3D structures. Using 3D representations of coronary arteries could improve PCI. However, pre-operative 3D acquisitions are hardly usable in a clinical setting as anatomical changes may occur between the time of acquisition and the time of the intervention, especially in pediatric populations. 3D reconstruction from X-ray angiography could palliate these problems. Current coronary artery 3D reconstruction methods rely on the use of biplane or rotational X-ray angiography. These acquisition modes however require prior synchronization and correspondence methods between the views. Therefore, this study proposes a learning-based single-view 3D reconstruction of coronary arteries. The reconstruction is performed using a mesh deformation network, paired with a new projection loss function term that learns a general 3D topology of coronary arteries from multiple 2D projections (i.e. using rotational acquisition). The model is trained using a geometric model from a realistic cardio-respiratory simulator and 3D patient data. Inference is performed using a single X-ray angiography image. The Chamfer distance between the reconstruction and the ground-truth is evaluated at 1.61*mm* for the

geometric model dataset. Validation on the patient imaging dataset yields an average reprojection error of 0.12*mm* which is within acceptable clinical ranges. We show that learning-based reconstruction is promising for coronary arteries. The results indicate that using four views during the training of the model provides the best results for learning-based 3D reconstruction of our datasets of coronary arteries.

4.2 Introduction

Coronary artery disease (CAD) is one of the most common types of heart disease worldwide, with a prevalence of 5-8% (McPherson & Tybjaerg-Hansen, 2016). CAD is manifested by a narrowing or total occlusion of the coronary arteries, which hinders the supply of blood and nutrients to the heart. CAD may lead to heart attacks and heart failures when left untreated.

Percutaneous coronary interventions (PCI) are usually preferred over open heart surgery for the treatment of CADs (Dawkins *et al.*, 2005; Çimen, Gooya, Grass & Frangi, 2016). PCI generally involve less risk and cause fewer secondary post-intervention effects (Roos-Hesselink *et al.*, 2006). In PCI, a catheter is inserted in the patient's vessels. This catheter is then navigated towards the narrowed part of the arteries to dilate it and to place a stent. The current gold standard imaging technique is X-ray angiography, a sequence acquisition performed while injecting a radiopaque contrast agent in the vessel. Medical imaging interpretation is challenging and relies heavily on the experience of the cardiologist. Some of the associated difficulties are: radiographic noise in images, vessel overlapping, and the quick dilution of the contrast agent (Galassi *et al.*, 2018). Furthermore, cardiac and respiratory movements are additional factors that make X-ray angiography difficult to understand in a real-time intervention setting. As a result, the interpretation of X-ray angiography is subject to high intra- and inter-variability as shown by Garcia *et al.* (2013).

Different types of X-ray angiography acquisitions can be performed: monocular, rotational and biplane(Green, Chen, Messenger, Groves & Carroll, 2004; Klein *et al.*, 2011). Monocular X-ray angiography is done using a single pair of an X-ray source and detector mounted on a

C-arm. Meanwhile, rotational acquisitions are performed using a single C-arm rotating around the patient in order to acquire time sequences of the organ of interest across multiple views, and biplane acquisitions are carried out using two C-arms oriented at two different angles to semi-simultaneously acquire two different images.

Multiple views of the same coronary artery structure offer more information than a single view, especially in overlapping parts of the coronary arteries. Acquiring multiple views simplifies the interpretation (Çimen *et al.*, 2016), but however requires additional steps. Rotational acquisitions usually alter the intervention processes and may require longer X-ray exposure. Biplane and rotational acquisitions do not guarantee that all the images are acquired at the point in the cardiac cycle, hence requiring motion compensation (Azizmohammadi, Martin, Miro & Duong, 2019) or modifications of the acquisition procedures (e.g., breath holding). Finally, calibration procedures must be employed in order to match points between images (Vachon, Miró & Duong, 2017). Calibration is especially needed as the intrinsic and extrinsic parameters of the cameras may be noisy or truncated (Cong *et al.*, 2015). Single-plane acquisitions are favored as they offer a less cumbersome, and more affordable, alternative to their biplane and rotational counterparts (Banerjee *et al.*, 2019).

Recent studies have shown the benefits of using pre-operative 3D acquisitions of the coronary arteries instead of X-ray angiography. 3D models alleviate the problem of foreshortening (Cong *et al.*, 2015) and help in the choice of a stent for a specific patient (Gundelwein *et al.*, 2018). 3D models of the arteries are however hard to obtain in catheterization laboratories. Besides, the overlay of pre-operative 3D imaging, such as CT or MR, during PCI is challenging as images are often acquired several months prior to the intervention, and may thus not reflect anatomical changes. Furthermore, these medical images are acquired in a different reference frame and may need to be registered to the patient in the intervention settings.

Alternatively, 3D reconstruction techniques can be used to obtain 3D models from routinely acquired X-ray angiography images. These techniques may offer the advantages of using 3D models while reducing the drawbacks of using pre-operative models. They are mostly based on

two or more views of the structure of interest. Coronary artery reconstruction is carried out using either a model-based approach or tomographic reconstruction (Çimen *et al.*, 2016). Model-based reconstruction recovers a binary representation of the coronary arteries while tomographic reconstruction focuses on obtaining the X-ray attenuation coefficient of the reconstructed volume (Çimen *et al.*, 2016).

Graph methods have been investigated for the model-based 3D reconstruction of coronary arteries. Our previous work (Martin *et al.*, 2017) performed the reconstruction of large vascular structures using random walks on a 3D graph. The graph method is guided by back-projection onto biplane X-ray angiography, without considering any motion between the two images. Similar approaches using graph-cuts methods have been proposed (Liao, Luc, Sun & Kirchberg, 2010). Zheng, Meiying & Jian (2010) adopted a forward projection approach designed to deform an initial shape to adapt it to the artery observable on input images. Their method utilizes deformable snake curves paired with manual pairing of landmarks between views. Banerjee *et al.* (2019) introduced a 3D coronary tree reconstruction method paired with a rigid and non-rigid motion correction. Using two different asynchronized projections, their reconstruction of the artery's centerline is performed by 3D point cloud intersection and 2D wrapping, representing the motion. Wang *et al.* (2020) proposed to use biplane angiographic imaging paired with OCT imaging. They manually paired the 2D centerlines of angiographic images to create the 3D centerline which was then augmented with a lumen whose diameters along the centerline were extracted from the OCT imaging.

Multi-view reconstruction techniques generally offer results with good accuracy, given that proper calibration and motion compensation is employed. They may, however, affect regular intervention protocols in many aspects either by interrupting the intervention or by modifying its protocols (Çimen *et al.*, 2016; Banerjee *et al.*, 2019). Single-view acquisitions are less cumbersome and more widely used during interventions (Banerjee *et al.*, 2019). Single-view reconstruction constitutes a challenging problem that requires solving ambiguities about aspects of the object that do not appear on the input image. This is particularly true for coronary arteries, in which, to our knowledge, no study has focused on the use of a single view.

Single-view reconstruction has recently been proposed for natural images (Pan *et al.*, 2019; Wang *et al.*, 2018b). In these studies, the reconstruction of various objects, such as tables, monitors or lamps, is performed using neural networks. During training, the models learn a relationship between the 2D and 3D information provided. These models build a base prior knowledge about the general structure of the objects to reconstruct. Wang *et al.* (2018b) proposed a reconstruction based on the deformation of an initial shape represented as a graph. Using features extracted from an input image, the nodes of the graph are gradually moved in order to obtain the final reconstruction. Alternatively, Mescheder, Oechsle, Niemeyer, Nowozin & Geiger (2019) proposed a reconstruction method based on occupancy networks using natural images. The method is based on a new multi-resolution representation of 3D shapes combined with learning-based methods. Nevertheless, these methods are not designed for the tortuous, irregular and thin shape of coronary arteries or for the large anatomic variability among patients (Khwansang & Chentanez, 2019). Additionally, single-view reconstruction methods are mostly based on neural networks, which require a large amount of data to be trained.

Since the availability of annotated data is a prevailing problem in medical imaging, multiple solutions have emerged ranging from data augmentation techniques to solutions that involve the generation of new data (Oulbacha & Kadoury, 2020; Tmenova *et al.*, 2019). These methods are, however, not designed to produce different 3D topologies of coronary arteries, and no study has focused on showing the benefits of such data augmentation methods in coronary arteries.

In this study, we propose a new learning-based method designed to perform 3D reconstruction of coronary arteries from single-view X-ray angiography. Using a new loss function that addresses the ambiguities arising in 3D reconstruction, we show that using single-view X-ray angiography provides a good initial reconstruction that can be refined over time. To alleviate the scarcity of data, we train the method using solely geometric model data acquired from a cardiorespiratory simulator and from coronary artery centerlines augmented with a lumen of an arbitrary diameter.

Hence, our research contributes to:

- the learning-based single-view 3D reconstruction of coronary arteries from X-ray angiography;
- the proposal of a projection loss function term designed to penalize incorrect 3D-2D correspondences;
- the learning of the general topology of the coronary arteries using multiple views during the training process.

This manuscript is organized as follows. Section 4.3 describes the data, along with the preprocessing applied to it (4.3.1), the model architecture (4.3.3) and the loss function (4.3.4) used for the 3D reconstruction of coronary arteries. Section 4.4 describes the metrics and validation strategies used to validate the results of our method, while the study results are presented and analyzed in section 4.5.

4.3 Methods

4.3.1 Data description and preprocessing

Two types of data were used in this study: geometric model data and patient imaging data. The geometric model dataset was mainly composed of 3D coronary artery simulations and of 3D centerlines extracted from the Rotterdam coronary artery algorithm evaluation framework (Schaap *et al.*, 2009). Patient imaging data consisted of patient X-ray angiography. Examples of 3D coronary arteries from our geometric model dataset are available in Figure 4.1.

4.3.1.1 Geometric model dataset

The geometric model dataset was composed of coronary artery models generated from computational phantoms as well as from patient CTA. The phantom models were acquired using a series of 4D computational XCAT phantoms (Segars *et al.*, 2010, 2013, 2015). A total of 54 phantoms, modeled from different patients were used, including male and female adults and pediatric patients of varying ages. Each phantom included both the left and the right coronary arteries, as well as the cardiac and respiratory motions. The phantoms were set up to generate time-changing 3D models of the coronaries, creating 50 time-varying 3D models for each phantom, which each one moving relative to the different cardiac phases and breathing patterns.



Figure 4.1 Example of coronary arteries from the geometric model dataset. The first two lines are arteries extracted from the Rotterdam coronary artery algorithm evaluation challenge: left coronary arteries in first line, and right coronary arteries in second line. The last two lines contain coronary arteries simulated using the XCAT: left coronary arteries in the third line, and right coronary arteries in the fourth line

Given the above, the geometric model dataset consisted of 2700 3D geometric models of left coronary arteries and 2700 models of right coronary arteries. The left coronary models consisted of the left circumflex artery and the left anterior descending artery, while the right coronary artery models consisted of the right coronary artery and the acute marginal segment.

The second half of the geometric model dataset was built from centerlines acquired through the Rotterdam coronary artery algorithm evaluation framework (Schaap *et al.*, 2009). The centerlines contain only the main branch of the artery, with no bifurcation. Both the right and left coronary arteries are represented. A total of 48 centerlines extracted from different patients, are available. All the centerlines were scaled and augmented with a lumen in order to generate different 3D models of the arteries. The diameters for the lumen were extracted from average diameters value reported in Dodge Jr, Brown, Bolson & Dodge (1992).

All the 3D shapes from the geometric model dataset were normalized in scale and orientation, and centered around the origin. As a result, all points of the geometric model were comprised in the [-1; 1] range for each axis. The models were then projected onto multiple views, creating segmentation masks. For each model, we used a total of 12 views spanning regularly spaced angles from 90° right anterior oblique (RAO) to 90° left anterior oblique (LAO), with a constant angle increment.

4.3.1.1.1 Patient imaging dataset

The patient imaging dataset consisted of coronary artery X-ray angiography images from a total of 14 patients acquired at Sainte-Justine Hospital. Permission to conduct this study was granted by the institutional review board of CHU Sainte-Justine. X-ray angiography images were manually segmented, cropped and centered to contain only the artery to reconstruct at the center of the image.

Both sets (geometric model dataset and patient imaging dataset) were split into training, testing and validation sets, respectively with the following percentages: 75%, 15%, 10%. There was no overlap of patients between the different sets.

4.3.2 Methodology

Our method aims to perform 3D reconstruction of coronary arteries from a single image. Figure 4.2 gives an overview of the general methodology. We use a network that progressively deforms

an initial shape, represented as a graph, to match a given input image as closely as possible. The network is composed of two parts designed to work together: a convolutional neural network (CNN) and a graph convolutional network (GCN). The CNN is used to extract visual features from an input image at different levels of detail, with each level corresponding to a different depth of the network. The GCN is used to deform an initial shape according to the perceptual features extracted from an input image. This deformation is performed at three different stages using the different levels of detail of the features extracted from the CNN. Each deformation stage corresponds to a level of detail, and is paired with an upsampling of the vertices of the graph. The training (Figure 4.2, part 1) is guided by a direct comparison of the reconstruction to a ground truth 3D model, and by projecting the reconstruction onto additional views of the coronary arteries to reconstruct. For inference (Figure 4.2, part 2), the reconstruction is performed using only one input image.

The input image is a segmentation mask of the coronary artery to reconstruct. Using a segmentation mask, rather than an X-ray angiography image allows for a seamless integration of both geometric model data and patient imaging data while limiting the addition of bias due to potentially erroneous automatic segmentation. The CNN is thus limited to extracting spatial and topological features that best represent the shape of the artery.

The initial shape is represented as a graph G = (V, E, F), with vertices V, edges E that connect the vertices, and features F associated with the vertices. Two vertices are connected if they belong to the same triangular face in the mesh. At each stage, the GCN projects the vertices of the shape onto the perceptual features extracted by the CNN. The combination of the projection with the associated features of the vertices is then used by the GCN to deform the 3D shape by updating the graph features. Indeed, only the vertices are moved during the deformation process, while the edges are left untouched. This displacement of the vertices is performed while ensuring coherence with the edges of the graph: proximity of connected edges, length of edges, normals of the nodes.



Figure 4.2 Visual representation of the proposed methodology. (a) During training phase, the mesh deformation network uses visual features extracted by the Perceptual feature extractor on an input image in order to gradually deform an initial shape of a coronary artery. The reconstruction process is trained with a 3D comparison to the ground truth and a 2D comparison to three additional views. (b) During the testing phase, only one input image is submitted to the trained 3D reconstruction network to perform the 3D reconstruction of the coronary arteries

When the initial shape is deformed through all the stages (levels of detail), the faces are added to the vertices of the graph to form the final reconstructed mesh.

4.3.3 Model architecture

4.3.3.1 Perceptual features extraction

The deformation of the initial shape is guided by perceptual features, which represent the shape of the coronary artery to reconstruct, including its boundaries, the curvatures and the branching points. The perceptual features extraction is performed using a VGG-16 network without the classification layers. The input of the network is a 224x224 image. Different layers of convolution are applied to the input image with a small filter size of (3x3), and pooling operations are performed after two consecutive convolutions. The coarse to fine features extracted by the network have the following dimensions:

- 1^{*st*} features of size (56, 56, 64)
- 2^{nd} features of size (28, 28, 128)
- 3^{rd} features of size (14, 14, 256)

The progressive extraction of low- to high-level features helps to gradually refine the reconstruction.

4.3.3.2 2D projection

A linear 2D projection of the initial shape is performed in our method using the camera's intrinsic parameters. The projection is used to associate the perceptual features extracted from the input image with the vertices of the 3D shape.

Each image of our dataset is associated with a 3D mesh that is rotated and centered accordingly. This greatly simplifies the projection process as it removes the need for any calculation of rotation or translation of the camera. Only the focal length, f is considered here for the initial projections. For the geometric model dataset, an initial value of 224 has been chosen in order to best mimic the positioning of coronary arteries in X-ray angiography acquisitions.

The projection process is expressed as follows:

$$\mathbf{x}' = C\mathbf{x} = \begin{bmatrix} \mathbf{x}'_0 \\ \mathbf{x}'_1 \\ \mathbf{x}'_2 \end{bmatrix}$$
(4.1)

where \mathbf{x} is a 3D point, \mathbf{x}' is its 2D projection, and *C* is the matrix of intrinsic parameters defined in Eq. 4.2.

$$C = \begin{bmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{bmatrix}$$
(4.2)

with f defined as the focal length and $[c_x; c_y]$ the optical center.

The final Cartesian coordinates are expressed as:

$$\overline{\mathbf{x}} = \begin{bmatrix} \frac{\mathbf{x}'_0}{\mathbf{x}'_2} \\ \frac{\mathbf{x}'_1}{\mathbf{x}'_2} \end{bmatrix}$$
(4.3)

4.3.3.3 Mesh deformation

The deformation of the mesh is done by combining its projection with the features extracted from the image by the CNN. The combination is performed using mesh deformation blocks which update vertex locations (i.e. vertex features). The deformation is performed by the GCN. It comprises three main operations: graph projection, graph convolution and graph pooling.

4.3.3.3.1 Graph projection

The initial shape is gradually deformed according to the features extracted from the input image. The shape is projected multiple times throughout the depths of the network. This projection is performed using Eq. 4.1. The vertices are then matched with the visual features extracted at their 2D projected location on the image.

4.3.3.3.2 Graph convolution

The convolution operation of the GCN is defined using the principles described by Bronstein, Bruna, LeCun, Szlam & Vandergheynst (2017). It is used to update the features $\mathbf{f}_i \in F$ associated with each vertex $\mathbf{v}_i \in V$ of a graph. It is defined as follows (Wang *et al.*, 2018a):

$$\mathbf{f}'_i = \mathbf{w}_0 \mathbf{f}_i + \sum_{j \in N(\mathbf{v}_i)} \mathbf{w}_1 \mathbf{f}_j$$
(4.4)

where the \mathbf{f}'_i represent the newly updated features of \mathbf{v}_i , \mathbf{w}_0 and \mathbf{w}_1 are weight matrices and $N(\mathbf{v}_i)$ is a set that contains all the neighbors of \mathbf{v}_i .

In our case, the feature matrix F is a concatenation of the 3D coordinates of the vertex v_i , information about the 3D shape (edges and normals) and the matching perceptual features of the vertex.

4.3.3.3.3 Graph pooling

Graph pooling is an operation that performs an up-sampling of the vertices of the graph. It is done by subdividing each triangle in the mesh into four new triangles. To this end, each edge of the original triangle is divided into two, and new faces are created with the new vertices.

4.3.4 Loss function

The final loss function used to optimize the training of the 3D reconstruction is composed of two loss functions and two regularisations: a Chamfer loss function term (Eq. 4.5), a normal loss function term, a Laplacian regularization and an edge length regularization. Additionally, we propose an extra term in the loss function based on the reprojection of the reconstruction onto the input image and additional views.

4.3.4.1 Chamfer loss function

The Chamfer distance is used in our method to express the 3D displacement between the location of the vertices of the reconstructed mesh and the vertices from the groundtruth. The Chamfer distance thus only considers the point cloud aspect of the graph, without the connections between

the nodes. It is defined as follows:

$$\mathcal{L}_{CD}(T,I) = \frac{1}{|T|} \sum_{t \in T} DF_I(t)$$
(4.5)

with *T* and *I* being two point clouds and *DF* a distance function between every point of one point cloud ($t \in T$) to the closest point in the other point cloud ($i \in I$).

The Chamfer distance is used in a symmetric setting in order to check the distance from the reconstruction to the ground truth and vice versa. This ensures that no part of the ground truth is omitted in the reconstruction and that no extra parts are added. Hence, the symmetric Chamfer distance, *SDC*, is expressed as follows:

$$\mathcal{L}_{SDC}(T,I) = \mathcal{L}_{SDC}(I,T) = \alpha \mathcal{L}_{CD}(T,I) + \beta \mathcal{L}_{CD}(I,T)$$
(4.6)

where α and β are two constants used to favor one direction of the distance or the other.

4.3.4.2 Normal loss function

This loss function term ensures that the normal associated with an edge remains consistent with the normal from the ground truth. To this end, a condition is added to allow each edge to be perpendicular to its associated normal from the ground truth (Wang *et al.*, 2018b). Applying this condition ensures that adding the faces at the end of the reconstruction process will result in a mesh whose normals are consistent with respect to the ground truth.

4.3.4.3 Laplacian regularization

This regularization is used to maintain a relative proximity between neighboring vertices during deformation (Wang *et al.*, 2018b). It ensures that vertices that are connected by an edge stay close together. This is important at the end of the reconstruction method when the faces are added to limit the amount of faces intersection.

4.3.4.4 Edge length regularization

This regularization is used to prevent outliers in the vertices by penalizing long edges during the training process. It is performed by averaging the length of each edge of the reconstruction.

4.3.4.5 **Projection loss function**

The proposed term consists in projecting the vertices of the reconstruction onto the input and additional views taken from different angles. The loss function term first ensures that no extra part will be added to the arteries (false positives). Indeed, because the arteries are small and contain bifurcations very close to the main artery, 3D distances fail to capture vertices that would fall to the background in 2D. Second, using multiple views will favor tubular shapes as extra vertices, usually near the intersection of two bifurcations, will be penalized when projected.

As a result, we define a projection loss function term that penalizes wrong 3D/2D correspondences. We first define A and B, two sets of points represented as matrices composed of 2D points. A contains projections of the reconstruction, while B contains ground truth 2D points. The intersection $A \cap B$ is thus defined as the number of points in A that appear in B, including duplicate points.

$$V = AB^T - diag^{-1}(AA^T I)$$
(4.7)

In this equations, $diag^{-1}$ represents the operation that extracts the diagonal of a matrix as a vector. *V* is defined as a vector of length n, which contains a 0 value, where an element of *A* can be found in *B*. *I* is the identity matrix. A vector *D* is then defined, containing the minimal value of each element, *i*, of *V*, such that:

$$D_i = \min V_i \tag{4.8}$$

Finally, the intersection between A and B is expressed using the following equation:

$$A \cap B = 1 - \frac{D}{D + \epsilon} \tag{4.9}$$

where ϵ is a very small value added to avoid divisions by zero.

From this, we define a new similarity metric.

$$S(A,B) = \frac{2*|A \cap B|}{2*|A|}$$
(4.10)

with |A| > |B|. This condition ensures that the value of the loss function term will be comprised in the [0; 1] range. This modification, compared to the traditional Dice coefficient, is due to an important imbalance of number of points between *A* and *B* as the number of predicted vertices is much lower than the number of vertices contained in the ground truth.

Hence, the final formulation of the projection loss function term for one view is as follows:

$$\mathcal{L}_{S}(A, B) = 1 - S(A, B)$$
 (4.11)

In order to compare the reconstruction to multiple views of the same ground truth structure, it is necessary to first rotate the sets of 3D points before projecting them. Our projection loss function term, for the similarity of multiple views (\mathcal{L}_{MS}), then becomes an addition of all the individual projections:

$$\mathcal{L}_{MS}(A,B) = \sum_{\nu \in V} \mathcal{L}_S(A_\nu, B_\nu) = \sum_{\nu \in V} 1 - S(A_\nu, B_\nu)$$
(4.12)

where V represents the set of all desired views.

4.4 Validation

The validation was performed in two steps. A first validation was performed on the geometric model data in order to assess the performance of the training and define hyperparameters and measures of performance on a test set with arteries presenting similar topologies as in the training set. Second, a quantitative and a qualitative validation were performed on patient imaging data. This second dataset allowed us to express the degree of generalization from a training performed

using solely the geometric model data, and provides insight into the clinical performance of the model.

4.4.1 Metrics

4.4.1.1 3D distances

The Chamfer distance is defined in Eq. 4.5. The second distance used in this paper is the Earth Mover distance (EMD), and is defined as follows:

$$EMD(X', X) = \min_{\phi: X' \to X} \sum_{\mathbf{x}' \in X'} \|\mathbf{x} - \phi(\mathbf{x}')\|$$
(4.13)

where X and X' are two sets of points and ϕ is a function that transforms X' into X with minimal cost.

The Chamfer distance roughly represents the average distance between the two point clouds while the earth mover distance considers the sum of all the distances. The distances are only applicable to geometric model data, as no ground truth 3D model was available for the patient imaging data.

4.4.1.2 2D projections

The reconstruction was also validated by projecting the 3D mesh obtained into input views. The first measure of reprojection accuracy is performed by calculating the Dice coefficient between the projection and the ground truth segmentation mask. This coefficient measures the similarity between two masks. The Dice coefficient, calculated using Eq. 4.14, gives a value ranging from 1, if the two masks (A and B) are completely identical, to 0, if none of their parts match. The reprojection coefficient is calculated both for the simulated and the real data.

$$Dice(A, B) = \frac{2 * |A \cap B|}{|A| + |B|}$$
(4.14)

The second metric used for the validation of the reprojection is the reprojection error. It is calculated by first manually extracting the centerline of the 3D reconstruction. This centerline is then projected onto the original 2D plane. Second, the centerline of the artery contained in the original image is calculated using a skeletonization on the segmentation mask. Finally, both 2D centerlines are compared. The comparison is done by calculating the Euclidean distance d(.,.) of every point x_1^i of one centerline X_1 to the closest point of the other one x_2^i and averaging them:

$$E(X_1, X_2) = \frac{1}{|X_1|} \sum_{\mathbf{x}_1^i \in X_1} d(\mathbf{x}_1^i, \mathbf{x}_2^i)$$
(4.15)

To ensure that the reprojection fully matches the ground truth image, we calculate a symmetric formulation of the reprojection error.

$$SE(X_1, X_2) = \frac{1}{2}(E(X_1, X_2) + E(X_2, X_1))$$
(4.16)

The Dice coefficient and the reprojection error are complementary as one validates the outer shape of the artery while the second considers the shape of the centerline.

4.4.2 Validation scheme

4.4.2.1 Geometric model data

Our method was first validated on the geometric model dataset. Using the two distances (EMD and Chamfer distance) and the Dice coefficient, we tested the ability of our network to reconstruct our 3D model test data. The results were used to assess the success of the training and to validate the benefits of the projection term in the loss function. To this end, the results obtained by our method were compared against Pixel2Mesh (Wang *et al.*, 2018b). Dice scores were calculated on both the original view used as input for the reconstruction and additional views unseen by the model.
Additionally, a qualitative evaluation was performed. The idea was to provide a visual that supports the previously mentioned metrics.

Finally, we evaluate the influence of the number of views on the accuracy of the reconstruction for the projection term in the loss function. To do so, we report the values of the metrics presented in the subsection above when the number of views is changed. This experiment was used to assess an ideal number of views to use. The results were then compared to those obtained with similar studies in the literature (Xie *et al.*, 2019).

4.4.2.2 Patient imaging data

As a second step, we evaluated the generalization of the training with geometric model data to patient imaging data. As this dataset contains only X-ray angiography images with no 3D ground-truth model, the validation was performed using the projection metrics presented in section 4.4.1. These values were used to evaluate the accuracy of the reconstruction as compared to the initial 2D projections. Additionally, they served as a good basis of comparison of accuracy against recent articles using learning-based networks for the reconstruction of coronary arteries (Tsompou *et al.*, 2020) and coronary artery trees (Banerjee *et al.*, 2019).

Furthermore, a qualitative evaluation was performed to give a visual indicator of the accuracy of the reconstruction. To that end, some reconstructed arteries from our patient imaging test dataset were displayed, along with the segmentation mask overlayed with their reprojection, and the initial X-ray angiography image.

4.5 Results

4.5.1 Training

The training was performed on both types of coronary arteries (LCA and RCA) using a learning rate value of $1e^{-4}$. We used a constant of 1 to balance the contribution of our proposed loss function term relative to the other loss function terms used.



Figure 4.3 Evolution of the values of the different losses for the validation set during the training. The horizontal axis represents the number of epochs. Vertical axes displayed in log scale

Figure 4.3 shows the evolution of all the term of the loss function used in this study. As can be seen, all of them converge and reach a plateau at around 50 epochs. This means that the proposed loss function is fully optimizable. Furthermore, as can be seen, low projection loss is achievable both on the training set and on the validation set, indicating that no over-fitting is produced. From the figure, we set a maximum of 50 for the results presented in the upcoming sections.

The first chart (top) of Figure 4.4 shows the evolution of the EMD distance when varying the number of views in the projection term of the loss function during training. 0 views refers to a training with no projection loss function term used. As can be seen, the EMD distance decreases each time a new view is added, until 4 views are used, after which the accuracy reaches a plateau.

The second chart (bottom) of Figure 4.4 shows the change of Dice score when varying the number of views used in the projection term of the loss function during training. The same observations can be drawn with respect to the 3D distance. The Dice score increases along with the number of views and reaches a plateau at around 4 views. Hence, using the projection term in the loss function with 4 equally spread projections (0° , 45° , 90° and 135°) seems to yield the best results with our datasets. Adding more views in the projection loss function term does not increase the accuracy of the reconstruction, and would further increase the training time, along with the memory consumption. Hence, all the results presented in the next subsections will be generated with our proposed method using 4 views for the projection term of the loss function.

Figure 4.5 displays features extracted by the network at different levels of detail, or depth, of the network. As can be seen, the network first considers the global shape of the artery, and then gradually focuses on different details contained in the image. These details help in refining the reconstruction throughout the different passes in the deformation network.



Figure 4.4 Results of Chamfer distance and EMD (on the left) and Dice score (on the right) for different numbers of views used in the projection loss function. Results calculated for the test split of our geometric model dataset. Values are normalized, and hence with no unit

4.5.2 Geometric model data

Table 4.1 provides the 3D (Chamfer and EMD) and 2D (Dice) reconstruction accuracy results from our 3D model and patient imaging datasets. The last two lines respectively represent the reconstruction performed by the initial Pixel2Mesh model and that with our method.



Figure 4.5 Features extracted by the network in the geometric model dataset. Each line represents one of the three levels of detail of the methodology

The average distance between the points of the reconstruction and the ground truth was around 1.6*mm*, which lies in an acceptable range for clinical use. Additionally, the Dice score was also high, with an average of 88.5%, which indicates a large overlap between the reconstruction and the ground truth.

No significant difference was observed between the Dice scores in the input view and the Dice scores in additional views. This indicates that our method learns a global topology of the coronary arteries and does not perform facade reconstruction, given that the training set is representative enough of topologies and shapes that would appear in the test set.

Table 4.1 also provides a comparison of results with Pixel2Mesh (Wang *et al.*, 2018b). As can be seen in the Table, the 3D accuracy of the reconstruction for the original model was very similar

to our results, with only a few tenths of a millimeter's difference. Additionally, the Dice score from our method was globally higher. A lower Dice score with an almost equal 3D distance is mostly due to the inability of the Chamfer distance to significantly discriminate between points falling between two branches of an artery. Indeed, in such cases, the distance to the ground truth would be small. However, when projected, those points fall to the background of the image, and are thus penalized, hence the lower overall Dice score.

Table 4.1 Summary of results for the geometric model dataset. Results are presented for both the left coronary artery (LCA) and right coronary artery (RCA) using the Dice coefficient and two 3D distances

		Our method	Pixel2Mesh
LCA	Dice (%)	88.21	79.18
		± 1.20	± 1.80
	Chamfer	1.64	1.75
		± 0.51	± 0.54
	EMD	48.79	49.12
		± 2.08	± 2.34
RCA	Dice (%)	89.32	78.57
		± 1.02	± 1.40
	Chamfer	1.58	1.67
		± 0.34	± 0.32
	EMD	45.21	46.21
		± 1.54	± 1.97

Table 4.2Results for the patient dataset both for the leftcoronary artery (LCA) and right coronary artery (RCA)using the Dice coefficient and a 2D distance

		Our method	Pixel2Mesh
LCA	Dice $(\%)$	58.63	50.31
	Dice (10)	± 3.11	± 3.11
	Mean reprojection	0.12	0.17
	error (mm)	± 0.04	± 0.02
RCA	Dicc (%)	62.45	50.90
	Dice (%)	± 1.98	± 3.04
	Mean reprojection	0.11	0.15
	error (mm)	± 0.07	± 0.06

Figure 4.6 provides a visualization of the results for both right and left coronary arteries from our simulated dataset. As can be seen, the reconstruction with our method (center column) appears visually close to the ground truth. Indeed, the general shape seems to be respected, and the bifurcation is properly reconstructed. The last 3D model represents the reconstruction with the original Pixel2Mesh. This example correlates with the projection loss function results (Table 4.1, "Our method"), and perfectly illustrates the motivation behind that term in the loss function. Without using the projection term in the loss function, we can see that the reconstruction contains artifacts at the points of bifurcations of the coronary arteries. This is due to additional points that are close in terms of distance to ground truth points, but that fall to the background when projected onto the input image. Additionally, the overall surface of the mesh is less noisy with the proposed loss function, and with less visible spikes.

4.5.3 Patient imaging data

Table 4.2 provides insight into the reconstruction error for the patient imaging data. Ground truth 3D models corresponding to X-ray angiography images are challenging to acquire. Hence, the evaluation on patient imaging data is mainly performed using the Dice coefficient and the reprojection error. Reconstruction of patient imaging data is challenging, as indicated by the lower accuracy observed than that obtained on geometric model data. The main reason is that the model has been trained using geometric model data, which may vary in terms of shape and topology of coronary arteries.

Additionally, a qualitative evaluation was performed for patient imaging data. Figure 4.7 presents results for the reconstruction of both right and left coronary arteries from our patient imaging dataset. As can be seen, the shape of the reconstructed arteries seems to match the corresponding view. This is further illustrated by the first and third columns of Figure 4.7, which overlay the reprojection of the reconstruction on top of both the X-ray angiography images used as input for the reconstruction method and a second X-ray angiography image from an additional view, after proper calibration. While the global shape seems to be well reconstructed, there are a few differences arising from the reprojection, and these are mainly located at bifurcations or

when the shape is very irregular. This is mainly due to the difference between the topology of the geometric model data used to train the network and the topologies of the patient imaging data. The major differences are located around the bifurcations. However, the projection on the second view shows that the topology of the reconstruction matches the observed arteries, up to a given scaling factor.



Figure 4.6 Visual results for the reconstruction of coronary arteries from our simulated dataset. The first line contains a right coronary artery (RCA), while the second one contains a left coronary artery (LCA). For each line, the ground-truth is on the left, the reconstruction with the projection term of the loss function is in the middle, and the reconstruction without projection loss function term is on the right (original Pixel2Mesh)

Furthermore, the reconstruction using patient imaging data may also contain some artifacts such as some random noise in the surface of the mesh or an increase in the diameter at various

parts of the artery. The diameter enlargement mostly occurs in the left coronary arteries, which generally present a broader range of diameters in our dataset. The above-mentioned artifacts are however reduced for the reconstructions with our proposed loss function, as indicated by the results in Table 4.2.



Figure 4.7 Visual results for the reconstruction of coronary arteries from the test split of our patient imaging dataset. Each line contains a different reconstruction and is organized as follows: X-ray angiography image used as input for the reconstruction, followed by the reconstruction result oriented on the same view, a second X-ray angiography image from a different angle with the reconstruction rotated for that view

Figure 4.8 gives additional insight into the performances of the proposed method. Indeed, our results globally outperform those obtained using the original Pixel2Mesh both in terms of mean and standard deviation for the reprojection error. Additionally, our results are less prone to large reprojection errors. The second plot displayed on the right of Figure 4.8 gives the average reprojection error for all 14 patients contained in the patient imaging dataset. The

figures show that the reprojection errors are globally consistent across all patients, without any major deviation from the mean reprojection error.

Furthermore, a comparison of our results with those of two different state-of-the-art methods was performed. The first method (Banerjee *et al.*, 2019) performs 3D reconstruction of arterial trees from multiple views, using epipolar geometry and movement compensation between the different views. The comparison with this method was done using the mean reprojection error between the 2D centerline of the input image and the projected 3D centerline of the reconstruction. Our mean reprojection error $(0.12 \pm 0.04mm)$ was slightly higher than theirs $(0.092 \pm 0.055mm)$. This was mostly due the fact that our reconstruction was performed using a single view, versus multiple views in Banerjee's method (Banerjee *et al.*, 2019). Using a single view limits the amount of 3D structural information that can be used for the reconstruction. Moreover, our mean reprojection error was calculated on our patient imaging test dataset while the training was performed on the geometric model dataset. These results thus show that the model is capable of generalizing the knowledge gathered on 3D models to unseen patient data, which vary greatly in terms of topology. Furthermore, our results indicate that a single view reconstruction can provide good results that could be refined over time.

Similar observations can be drawn with the comparison of the Dice coefficient with Tsompou et al.'s (Tsompou *et al.*, 2020). Our Dice coefficient (60%) was significantly lower than theirs Tsompou *et al.* (2020) (81.5%). Tsompou *et al.* (2020) higher accuracy in the reconstruction can be explained by the manual selection of corresponding points between the images used for the 3D reconstruction. As well, we can see that the difference in the Dice coefficient (-21.5%) was much greater than that of the difference of the reprojection error (+0.028mm). This means that the artery is globally well reconstructed in terms of spatial location and positions of the centerline. However, the lower Dice coefficient indicates that the lumen is reconstructed with less accuracy, and that it is either too large or too narrow.

Overall, the Dice coefficients obtained and the visual results indicate that the network was capable of generalizing the knowledge accumulated on geometric model data to patient imaging

data. However the level of generalization remains limited on topologies that are very different from those encountered in the training set. This mostly affects the shape of the lumen of the reconstructed arteries, as can be seen with the Dice coefficient. The generalization could be improved by adding more variety (mainly of shapes and bifurcations) in the training set.



Figure 4.8 A violin plot of the distribution of reprojection errors for all the patients in patient imaging dataset is displayed on the left of the figure. The results are separated for the left (LCA) and right coronary arteries (RCA), and a comparison is made with results obtained with the original Pixel2Mesh model. The distribution of the reprojection errors for all our patient imaging data is given on the right side of the figure

4.6 Discussion

Single-view 3D reconstruction is very challenging. The main challenges in 3D monocular reconstruction techniques are linked to the ambiguities and the missing information created by 3D-to-2D projections. X-ray angiography can contain vessel overlapping and foreshortening. Our method aims to counter these problems by using learning-based techniques. The network is trained in order to learn the general topology of the coronary arteries. Using a learning-based method has the advantage of not having to define hard constraints specific to coronary arteries and to the calibration of views. These constraints are instead learned given a specific training

set, which thus provides a convenient way to extend the method to different types of arteries. Thus, learning-based methods perform general shape reconstruction, up to a scale factor that can later be refined using calibration methods.

This study presents a novel learning-based single-view 3D reconstruction technique for coronary arteries. By leveraging the knowledge acquired on a training set composed of geometric model data, the level of generalization to patient imaging data is also shown. The geometric model data allows to train our network with both projections and 3D shapes. An additional interest in using geometric model data lies in the opportunity to augment the training data with additional views capable of solving ambiguities of 3D-to-2D projections. Furthermore, it is a convenient way of performing data augmentation when only limited patient imaging data is available.

After the training, only a single view is used for the reconstruction, along with the knowledge acquired through the training of the network. The use of segmentation masks of coronary arteries, rather than the original X-ray angiography images and simulations, allowed the network to learn to analyze input shapes, instead of the style and grain of the images. Furthermore, segmentation is an important step that limits the reconstruction to the desired structure of interest.

Performing the reconstruction with a single view preserves the original PCI protocols, unlike multi-view reconstruction, which generally leads to increased acquisition time, and which requires the C-arms to be moved. Finally, using a single view removes any need for motion compensation that would occur when using multiple angiographic views.

Learning-based methods are however limited when the X-ray angiography images of the test data are very different from those of the training data. This is attributable to the data sources involved: 3D model vs. patient imaging data. This accuracy discrepancy is particularly true when the ground truth contains different arteries from the ones contained in our trained data (different from right coronary artery, right acute marginal, left anterior descending and left circumflex). Furthermore, the difference is especially significant when the angulation of the view is different from what has been used during training. On 3D model test data, the reconstructions are both accurate, as compared to the 3D groundtruth, and their reprojection matches the available 2D

views. This indicates a good promise for the model when it comes to generalizing to new anatomies with topologies similar to those used during training. Limitations encountered are mostly seen when using patient imaging data. Indeed, our patient imaging dataset comprises highly irregular and complex topologies that are not fully represented in the training set. These limitations are manifested in the 3D reconstructions by a difference of lumen diameter between the reconstruction and the ground truth images.

One way to counter the limitations could be to acquire more patient imaging data with ground truth 3D models. A second option would be to generate additional data with different topologies. Finally, the single view reconstruction could be used as a preliminary step towards the final reconstruction and could be refined through time using subsequent acquisitions.

An evaluation of the proposed projection loss function was performed. By using the new loss function, the reconstruction accuracy increased both in 2D and 3D, indicating that it successfully guided the network towards learning a correspondence between 2D and 3D information. Furthermore, an evaluation of the effectiveness of increasing the number of views used in the projection term of the loss function was conducted, with rotations limited around one axis only. For completeness, both cranial and caudal angles could be considered. According to our experiments, the ideal number of views to use in the projection with our datasets was 4. After that value, there was no benefit either in terms of reconstruction error or in reprojection error. This result is consistent with previous findings in natural images (Xie *et al.*, 2019).

The proposed single-view reconstruction method presents many clinical advantages. First, it has the advantage of offering instant reconstruction at test time. Second, single-view reconstruction paired with instant reconstruction allows for the 3D reconstruction of sequences of angiography. Hence, the proposed method can be used in combination with a calibration method to efficiently integrate 3D information into PCI without altering the clinical workflows. Efficient use within catheterization laboratories would however require the acquisition of more data for training purposes, along with the integration of an automatic segmentation method for the network input. Future works could consider the addition of coronary arteries with abnormal shapes, such as a stenosis or an aneurysm, to the training set. This would allow a better representation of all possible patient imaging data outcomes, and improve the accuracy of the reconstruction. Furthermore, future studies could use a pre-operative 3D model of the arteries as the initial shape to be deformed in order to guide the reconstruction process and limit the solution to a 3D/2D registration. This registration could also be used to refine the single-view reconstruction using an additional view, where necessary. Finally, future works could also consider using stereo-based methods with a biplane dataset in order to create a patient imaging dataset with ground truth 3D models.

4.7 Conclusion

Deep learning methods are promising for learning 3D representations of coronary arteries from a single X-ray angiography view. They allow the learning of 3D-2D correspondence mapping, which limits the number of views required, as well as the need for synchronization between them. The use of 3D reconstruction makes 3D information of the coronary arteries available during PCI. We have shown that the use of four views is ideal for the learning-based 3D reconstruction of coronary arteries. The proposed approach paves the way for the introduction of seamless, on-the-fly, and sequential 3D reconstruction of coronary arteries into clinical workflows.

CONCLUSION AND RECOMMENDATIONS

The navigation of catheters during percutaneous cardiac interventions for the treatment of heart diseases is a challenging task that heavily relies on the experience of cardiologists. The main goal of our research was to propose an end-to-end 3D monocular reconstruction of coronary arteries from X-ray angiography, with uncertainty modeling, to overcome the limitations imposed by 2D navigational guidance. To address the aforementioned objective, we defined three specific objectives.

In Chapter 2, we introduced a new segmentation method for coronary arteries using deep learning methods with an evaluation and calibration of the uncertainty associated with the results. Then, in Chapter 3, we proposed a novel vesselness-based style transfer method to synthesize new coronary artery X-ray angiography. Finally, in Chapter 4, we evaluated a single-view 3D reconstruction method of coronary arteries. All these objectives are part of a pipeline designed to perform end-to-end 3D reconstruction of the coronary using a single-view X-ray angiography. Indeed, the 3D reconstruction (Chapter 4) is performed using a segmentation mask of the coronary arteries produced by an uncertainty-aware network (Chapter 2) whose performance and generalization potential can be increased using fully customizable and annotated data (Chapter 3).

The lack of annotations and representation in the learning model's data is a predominant problem in machine learning and mainly in medical image analysis. To enhance the accuracy and confidence of our reconstruction and segmentation methods, in Chapter 3, we proposed a novel way of synthesizing fully-customizable X-ray angiography using a realistic cardiovascular simulator alongside a style transfer method. This method provides new automatically segmented images of arteries whose topology varies according to different parameters such as the patient's sex, age, pathologies, etc. This not only increases the accuracy and confidence of the segmentation

network but also indirectly improves the ability of the 3D reconstruction network to generalize to different structures and coronary artery topologies.

In Chapter 4, we showed that single-view 3D reconstruction of coronary arteries could be performed by leveraging the knowledge acquired by a network regarding the general topology of coronary arteries. To our knowledge, this is the first study to propose the reconstruction of coronary arteries using a single view of X-ray angiography. Using a single view limits the need for temporal registration methods during the reconstruction process and adapts to single-plane and biplane angiography acquisitions. The results indicate that single-view reconstruction is a promising way to perform initial 3D reconstruction that could then be refined over time.

The proposed research paves the way for the adoption of safer end-to-end machine learning methods in automatic medical image analysis. The objective is to assist cardiologists during percutaneous interventions. Hence, the idea is to guide while also raising red flags regarding uncertain predictions which require the cardiologist's experience. It is an additional step in the direction of the explainability of learning models. While state-of-the-art models focus on trying to explain what models learn on a given dataset using attention maps or transformer models, our approach focuses on the calibration of uncertainty by correlating the confidence of learning models with the accuracy of their results. This approach is designed as assistance for cardiologists by automating confident cases while relying on their experience and knowledge of difficult and highly uncertain scenarios.

Future works could first consider the evaluation of the uncertainty yielded by the 3D reconstruction task and compare it to the uncertainty linked to the segmentation. This would improve the confidence in the end-to-end reconstruction technique. Furthermore, the uncertainty yield by the segmentation could be propagated through the reconstruction process. Indeed, we showed in a previous study (Martin *et al.*, 2019) that the highest values of uncertainty in the segmentation of a coronary artery were located at its edges. Uncertainty values could hence be used during the

projection process in the 3D reconstruction as additional visual features that would later impact the mesh deformation. Second, saliency or attention maps could be included in the uncertainty calculation to further penalize the lack of confidence in highly important parts of the images. Third, the 3D reconstruction could further be improved by performing the reconstruction over multiple frames of a cine-angiography and registering them together. Additionally, combining the 3D reconstruction with a motion prediction technique would be a great way to limit the usage of X-rays and contrast agents. This would indeed allow us to first acquire the geometry of the coronary arteries and then modify it using the predicted motion and deformation. Fourth, the proposed synthesis of coronary artery X-ray images could be evaluated as a data augmentation technique for the segmentation and 3D reconstruction methods. As discussed in Chapter 2, Epistemic uncertainty is highly affected by a lack of data. Providing additional annotated data could further improve the calibration of the uncertainty while also increasing the accuracy of the segmentation results. Finally, mixed reality would provide clinicians with a more immersive environment that could facilitate the guidance of catheters. It would allow displaying information and parts of the arteries that are normally not visible on 2D X-ray angiography.

BIBLIOGRAPHY

- Aladem, M. & Rawashdeh, S. A. (2020). A single-stream segmentation and depth prediction CNN for autonomous driving. *IEEE Intelligent Systems*, 36(4), 79–85.
- Amirkhani, A., Barshooi, A. H. & Ebrahimi, A. (2022). Enhancing the Robustness of Visual Object Tracking via Style Transfer. CMC-COMPUTERS MATERIALS & CONTINUA, 70(1), 981–997.
- Azizmohammadi, F., Martin, R., Miro, J. & Duong, L. (2019). Model-free cardiorespiratory motion prediction from X-ray angiography sequence with LSTM network. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 7014–7018.
- Banerjee, A., Galassi, F., Zacur, E., De Maria, G. L., Choudhury, R. P. & Grau, V. (2019). Point-cloud method for automated 3D coronary tree reconstruction from multiple nonsimultaneous angiographic projections. *IEEE transactions on medical imaging*, 39(4), 1278–1290.
- Barshooi, A. H. & Amirkhani, A. (2022). A novel data augmentation based on Gabor filter and convolutional deep learning for improving the classification of COVID-19 chest X-Ray images. *Biomedical Signal Processing and Control*, 72, 103326.
- Bergersen, L., Foerster, S., Marshall, A. C. & Meadows, J. (2008). *Congenital heart disease: the catheterization manual*. Springer Science & Business Media.
- Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A. & Vandergheynst, P. (2017). Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4), 18–42.
- Brost, A., Behar, J. M., Rinaldi, C. A., Housden, R. J. & Rhode, K. S. (2017). 3D Reconstruction of Coronary Veins from a Single X-Ray Fluoroscopic Image and Pre-operative MR. *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges: 7th International Workshop, STACOM 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers*, 10124, 66.
- Brostow, G. J., Shotton, J., Fauqueur, J. & Cipolla, R. (2008). Segmentation and Recognition Using Structure from Motion Point Clouds. *ECCV*(1), pp. 44-57.
- Brostow, G. J., Fauqueur, J. & Cipolla, R. (2009). Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters*, 30(2), 88–97.

- Caicedo, J. M. & Zárate, B. A. (2011). Reducing epistemic uncertainty using a model updating cognitive system. *Advances in Structural Engineering*, 14(1), 55–65.
- Caro, J., Trindade, E. & McGregor, M. (1991). The risks of death and of severe nonfatal reactions with high-vs low-osmolality contrast media: a meta-analysis. AJR. American journal of roentgenology, 156(4), 825–832.
- Cheng, F., Zhang, H., Yuan, D. & Sun, M. (2019). Leveraging semantic segmentation with learning-based confidence measure. *Neurocomputing*, 329, 21–31.
- Çimen, S., Gooya, A., Grass, M. & Frangi, A. F. (2016). Reconstruction of coronary arteries from X-ray angiography: A review. *Medical image analysis*, 32, 46–68.
- Cong, W., Yang, J., Ai, D., Chen, Y., Liu, Y. & Wang, Y. (2015). Quantitative analysis of deformable model-based 3-D reconstruction of coronary artery from multiple angiograms. *IEEE Transactions on Biomedical Engineering*, 62(8), 2079–2090.
- Dawkins, K., Gershlick, T., De Belder, M., Chauhan, A., Venn, G., Schofield, P., Smith, D., Watkins, J. & Gray, H. (2005). Percutaneous coronary intervention: recommendations for good practice and training. *Heart*, 91(suppl 6), vi1–vi27.
- Der Kiureghian, A. & Ditlevsen, O. (2009). Aleatory or epistemic? Does it matter? *Structural Safety*, 31(2), 105–112.
- Ding, Y., Liu, J., Xiong, J. & Shi, Y. (2020). Revisiting the evaluation of uncertainty estimation and its application to explore model complexity-uncertainty trade-off. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 4–5.
- Dodge Jr, J. T., Brown, B. G., Bolson, E. L. & Dodge, H. T. (1992). Lumen diameter of normal human coronary arteries. Influence of age, sex, anatomic variation, and left ventricular hypertrophy or dilation. *Circulation*, 86(1), 232–246.
- Fan, J., Yang, J., Wang, Y., Yang, S., Ai, D., Huang, Y., Song, H., Hao, A. & Wang, Y. (2018). Multichannel fully convolutional network for coronary artery segmentation in x-ray angiograms. *Ieee Access*, 6, 44635–44643.
- Fingscheidt, T., Gottschalk, H. & Houben, S. (2022). Deep Neural Networks and Data for Automated Driving: Robustness, Uncertainty Quantification, and Insights Towards Safety.

- Frangi, A. F., Niessen, W. J., Vincken, K. L. & Viergever, M. A. (1998). Multiscale vessel enhancement filtering. *International conference on medical image computing and computer-assisted intervention*, pp. 130–137.
- Gal, Y. & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. *international conference on machine learning*, pp. 1050–1059.
- Gal, Y., Hron, J. & Kendall, A. (2017). Concrete dropout. *Advances in Neural Information Processing Systems*, pp. 3584–3593.
- Galassi, F., Alkhalil, M., Lee, R., Martindale, P., Kharbanda, R. K., Channon, K. M., Grau, V. & Choudhury, R. P. (2018). 3D reconstruction of coronary arteries from 2D angiographic projections using non-uniform rational basis splines (NURBS) for accurate modelling of coronary stenoses. *PloS one*, 13(1), e0190650.
- Garcia, J. A., Eng, M., Hudson, P., Klein, A., Chen, S., Kim, M., Groves, B., Messenger, J., Wink, O. & Carroll, J. (2013). Impact of three dimensional in-room imaging (3DCA) in the facilitation of percutaneous coronary interventions. *Journal of Cardiology and Vascular Medicine*, 1(1), 1.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. & Bengio, Y. (2014). Generative adversarial nets. 2672–2680.
- Grech, M., Debono, J., Xuereb, R. G., Fenech, A. & Grech, V. (2012). A comparison between dual axis rotational coronary angiography and conventional coronary angiography. *Catheterization and Cardiovascular Interventions*, 80(4), 576–580.
- Green, N. E., Chen, S.-Y. J., Messenger, J. C., Groves, B. M. & Carroll, J. D. (2004). Threedimensional vascular angiography. *Current problems in cardiology*, 29(3), 104–142.
- Groher, M., Zikic, D. & Navab, N. (2009). Deformable 2D-3D registration of vascular structures in a one view scenario. *IEEE Transactions on medical imaging*, 28(6), 847–860.
- Gu, J., Yang, T. S., Ye, J. C. & Yang, D. H. (2021). CycleGAN denoising of extreme low-dose cardiac CT using wavelet-assisted noise disentanglement. *Medical Image Analysis*, 74, 102209.
- Gundelwein, L., Miró, J., Barlatay, F. G., Lapierre, C., Rohr, K. & Duong, L. (2018). Personalized stent design for congenital heart defects using pulsatile blood flow simulations. *Journal* of biomechanics, 81, 68–75.

- Gundelwein, L., Miró, J. & Duong, L. (2015). Automatic stent placement and stent size selection for preoperative planning of pulmonary artery intervention. *Joint MICCAI workshops on Computing and Visualisation for Intravascular Imaging and Computer-Assisted Stenting*.
- Guo, C., Pleiss, G., Sun, Y. & Weinberger, K. Q. (2017). On calibration of modern neural networks. *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 1321–1330.
- Guynn, J. (2015). Google photos labeled black people 'gorillas'.
- Gwak, J., Choy, C. B., Chandraker, M., Garg, A. & Savarese, S. (2017). Weakly supervised 3d reconstruction with adversarial constraint. 2017 International Conference on 3D Vision (3DV), pp. 263–272.
- Habert, S., Dahdah, N. & Cheriet, F. (2012). A novel method for an automatic 3D reconstruction of coronary arteries from angiographic images. *IEEE Information Science, Signal Processing and their Applications (ISSPA), 11th International Conference on*, 484–489.
- He, R., Wu, X., Sun, Z. & Tan, T. (2018). Wasserstein cnn: Learning invariant features for nir-vis face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 41(7), 1761–1773.
- Herman, G. T. & Meyer, L. B. (1993). Algebraic reconstruction techniques can be made computationally efficient (positron emission tomography application). *IEEE transactions on medical imaging*, 12(3), 600–609.
- Hou, L., Agarwal, A., Samaras, D., Kurc, T. M., Gupta, R. R. & Saltz, J. H. [Preprint at https://arxiv.org/abs/1712.05021]. (2017). Unsupervised histopathology image synthesis.
- Jégou, S., Drozdzal, M., Vazquez, D., Romero, A. & Bengio, Y. (2017). The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation. *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017 IEEE Conference on, pp. 1175–1183.
- Jordi Oller, X., Gundelwein, L., Miro, J. & Duong, L. (2016). Automatic stent fitting using CFD estimates of hemodynamic parameters for pulmonary artery stenosis.
- Kaptoge, S., Pennells, L., De Bacquer, D., Cooney, M. T., Kavousi, M., Stevens, G., Riley, L. M., Savin, S., Khan, T., Altay, S. et al. (2019). World Health Organization cardiovascular disease risk charts: revised models to estimate risk in 21 global regions. *The Lancet Global Health*, 7(10), e1332–e1345.
- Kato, H., Ushiku, Y. & Harada, T. (2018). Neural 3d mesh renderer. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3907–3916.

- Kavur, A. E., Selver, M. A., Dicle, O., Barış, M. & Gezer, N. S. (2019, apr). CHAOS Combined (CT-MR) Healthy Abdominal Organ Segmentation Challenge Data. Retrieved from: https: //doi.org/10.5281/zenodo.3431873.
- Kendall, A. & Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision? Advances in Neural Information Processing Systems, pp. 5580–5590.
- Kendall, A., Gal, Y. & Cipolla, R. (2018). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. 7482–7491.
- Kerkeni, A., Benabdallah, A., Manzanera, A. & Bedoui, M. H. (2016). A coronary artery segmentation method based on multiscale analysis and region growing. *Computerized Medical Imaging and Graphics*, 48, 49–61.
- Khwansang, N. & Chentanez, V. (2019). Anatomic variations of coronary arteries: origins, branching patterns, and abnormalities. *Asian Biomedicine*, 12(3), 117–123.
- Kingma, D. P., Salimans, T. & Welling, M. (2015). Variational dropout and the local reparameterization trick. Advances in Neural Information Processing Systems, pp. 2575– 2583.
- Klein, A. J., Garcia, J. A., Hudson, P. A., Kim, M. S., Messenger, J. C., Casserly, I. P., Wink, O., Hattler, B., Tsai, T. T., Chen, S. J. et al. (2011). Safety and efficacy of dual-axis rotational coronary angiography vs. standard coronary angiography. *Catheterization and Cardiovascular Interventions*, 77(6), 820–827.
- Kohli, P. & Torr, P. H. (2008). Measuring uncertainty in graph cut solutions. *Computer Vision and Image Understanding*, 112(1), 30–38.
- Kononenko, I. (1989). Bayesian neural networks. *Biological Cybernetics*, 61(5), 361–370.
- Konyushkova, K., Sznitman, R. & Fua, P. (2019). Geometry in active learning for binary and multi-class image segmentation. *Computer vision and image understanding*, 182, 1–16.
- Kunio, M., O'Brien, C. C., Lopes, A. C., Bailey, L., Lemos, P. A., Tearney, G. J. & Edelman, E. R. (2017). Vessel centerline reconstruction from non-isocentric and non-orthogonal paired monoplane angiographic images. *The International Journal of Cardiovascular Imaging*, 1–10.
- Kurz, A., Hauser, K., Mehrtens, H. A., Krieghoff-Henning, E., Hekler, A., Kather, J. N., Fröhling, S., von Kalle, C., Brinker, T. J. et al. (2022). Uncertainty estimation in medical image classification: systematic review. *JMIR Medical Informatics*, 10(8), e36427.

- Kutulakos, K. N. & Seitz, S. M. (2000). A theory of shape by space carving. *International journal of computer vision*, 38(3), 199–218.
- Levine, S., Coyne, B. & Colvin, L. (2015). *Clinical Exercise Electrocardiography* (ed. 1). Jones & Bartlett Publishers.
- Li, Z., Liu, J. & Tang, X. (2007). A closed-form solution to 3D reconstruction of piecewise planar objects from single images. 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–6.
- Liang, D., Wang, L., Han, D., Qiu, J., Yin, X., Yang, Z., Xing, J., Dong, J. & Ma, Z. (2021). Semi 3D-TENet: Semi 3D network based on temporal information extraction for coronary artery segmentation from angiography video. *Biomedical Signal Processing and Control*, 69, 102894.
- Liao, R., Luc, D., Sun, Y. & Kirchberg, K. (2010). 3-D reconstruction of the coronary artery tree from multiple views of a rotational X-ray angiography. *The international journal of cardiovascular imaging*, 26(7), 733–749.
- Lin, G., Milan, A., Shen, C. & Reid, I. (2017). Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1925–1934.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B. & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42, 60–88.
- Martin, R., Vachon, E., Miró, J. & Duong, L. (2017). 3D reconstruction of vascular structures using graph-based voxel coloring. *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, pp. 1032–1035.
- Martin, R., Miró, J. & Duong, L. (2019). Epistemic uncertainty modeling for vessel segmentation. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 5923–5927.
- Matl, S., Brosig, R., Baust, M., Navab, N. & Demirci, S. (2017). Vascular image registration techniques: A living review. *Medical image analysis*, 35, 1–17.
- McPherson, R. & Tybjaerg-Hansen, A. (2016). Genetics of coronary artery disease. *Circulation research*, 118(4), 564–578.

- Mehrtash, A., Wells, W. M., Tempany, C. M., Abolmaesumi, P. & Kapur, T. (2020). Confidence calibration and predictive uncertainty estimation for deep medical image segmentation. *IEEE transactions on medical imaging*, 39(12), 3868–3878.
- Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S. & Geiger, A. (2019). Occupancy networks: Learning 3d reconstruction in function space. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4460–4470.
- M'hiri, F., Duong, L., Desrosiers, C., Leye, M., Miró, J. & Cheriet, M. (2016). A graph-based approach for spatio-temporal segmentation of coronary arteries in X-ray angiographic sequences. *Computers in biology and medicine*, 79, 45–58.
- Milaire, J., Beauthier, J.-P. & Dhem, A. (2001). *Anatomie médicale: aspects fondamentaux et applications cliniques*. DE BOECK.
- Naeini, M. P., Cooper, G. F. & Hauskrecht, M. (2015). Obtaining Well Calibrated Probabilities Using Bayesian Binning. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, (AAAI'15), 2901–2907. Retrieved from: http://dl.acm.org/citation.cfm?id= 2888116.2888120.
- Nathan Silberman, Derek Hoiem, P. K. & Fergus, R. (2012). Indoor Segmentation and Support Inference from RGBD Images. *ECCV*.
- National Highway Traffic Safety Administration. (2017). Tesla Crash Preliminary Evaluation Report.
- Neal, R. M. (2012). Bayesian learning for neural networks. Springer Science & Business Media.
- Niculescu-Mizil, A. & Caruana, R. (2005). Predicting good probabilities with supervised learning. *Proceedings of the 22nd international conference on Machine learning*, pp. 625–632.
- Oulbacha, R. & Kadoury, S. (2020). MRI to CT Synthesis of the Lumbar Spine from a Pseudo-3D Cycle GAN. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 1784–1787.
- Pan, J., Han, X., Chen, W., Tang, J. & Jia, K. (2019). Deep Mesh Reconstruction from Single RGB Images via Topology Modification Networks. *Proceedings of the IEEE International Conference on Computer Vision*, pp. 9964–9973.
- Perera, P., Abavisani, M. & Patel, V. M. (2018). In2i: Unsupervised multi-image-to-image translation using generative adversarial networks. 2018 24th International Conference on Pattern Recognition (ICPR), pp. 140–146.

- Pinheiro, M. A., Kybic, J. & Fua, P. (2016). Geometric graph matching using monte carlo tree search. *IEEE transactions on pattern analysis and machine intelligence*, 39(11), 2171–2185.
- Rivest-Henault, D., Sundar, H. & Cheriet, M. (2012). Nonrigid 2D/3D registration of coronary artery models with live fluoroscopy for guidance of cardiac interventions. *IEEE Transactions on Medical Imaging*, 31(8), 1557–1572.
- Roos-Hesselink, J., Meijboom, F., Spitaels, S., VanDomburg, R., VanRijen, E., Utens, E., Bogers, A. & Simoons, M. (2006). Long-term outcome after surgery for pulmonary stenosis (a longitudinal study of 22–33 years). *European heart journal*, 27(4), 482–488.
- Schaap, M., Metz, C. T., van Walsum, T., van der Giessen, A. G., Weustink, A. C., Mollet, N. R., Bauer, C., Bogunović, H., Castro, C., Deng, X. et al. (2009). Standardized evaluation methodology and reference database for evaluating coronary artery centerline extraction algorithms. *Medical image analysis*, 13(5), 701–714.
- SEER Training. (2022). Classification and Structure of Blood Vessels [Format]. Retrieved from: https://training.seer.cancer.gov/anatomy/cardiovascular/blood/ classification.html.
- Segars, W. P., Norris, H., Sturgeon, G. M., Zhang, Y., Bond, J., Minhas, A., Tward, D. J., Ratnanather, J., Miller, M., Frush, D. et al. (2015). The development of a population of 4D pediatric XCAT phantoms for imaging research and optimization. *Medical physics*, 42(8), 4719–4726.
- Segars, W., Sturgeon, G., Mendonca, S., Grimes, J. & Tsui, B. M. (2010). 4D XCAT phantom for multimodality imaging research. *Medical physics*, 37(9), 4902–4915.
- Segars, W., Bond, J., Frush, J., Hon, S., Eckersley, C., Williams, C. H., Feng, J., Tward, D. J., Ratnanather, J., Miller, M. et al. (2013). Population of anatomically variable 4D XCAT adult phantoms for imaging research and optimization. *Medical physics*, 40(4), 043701.
- Sensoy, M., Kaplan, L. & Kandemir, M. (2018). Evidential deep learning to quantify classification uncertainty. Advances in Neural Information Processing Systems, pp. 3183–3193.
- Stapleton, W. (1977). Hop, Skip, Jump. Self & Society, 5(12), 396–396.
- Sun, H., Xi, Q., Fan, R., Sun, J., Xie, K., Ni, X. & Yang, J. (2022). Synthesis of pseudo-CT images from pelvic MRI images based on an MD-CycleGAN model for radiotherapy. *Physics in Medicine & Biology*, 67(3), 035006.

- Tavakkoli, A., Kamran, S. A., Hossain, K. F. & Zuckerbrod, S. L. (2020). A novel deep learning conditional generative adversarial network for producing angiography images from retinal fundus photographs. *Scientific Reports*, 10(1), 1–15.
- Tmenova, O., Martin, R. & Duong, L. (2019). CycleGAN for style transfer in X-ray angiography. *International journal of computer assisted radiology and surgery*, 14(10), 1785–1794.
- Tsompou, P. I., Andrikos, I. O., Karanasiou, G. S., Sakellarios, A. I., Tsigkas, N., Kigka, V. I., Kyriakidis, S., Michalis, L. K. & Fotiadis, D. I. (2020). Validation study of a novel method for the 3D reconstruction of coronary bifurcations. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 1576–1579.
- Unberath, M., Achenbach, S., Fahrig, R. & Maier, A. (2016). Exhaustive graph cut-based vasculature reconstruction. 1143–1146.
- Unberath, M., Gao, C., Hu, Y., Judish, M., Taylor, R. H., Armand, M. & Grupp, R. (2021). The impact of machine learning on 2d/3d registration for image-guided interventions: A systematic review and perspective. *Frontiers in Robotics and AI*, 8, 716007.
- Vachon, É., Miró, J. & Duong, L. (2017). Online C-arm calibration using a marked guide wire for 3D reconstruction of pulmonary arteries. *Medical Imaging 2017: Image-Guided Procedures, Robotic Interventions, and Modeling*, 10135, 101351V.
- Valiente, R., Zaman, M., Ozer, S. & Fallah, Y. P. (2019). Controlling Steering Angle for Cooperative Self-driving Vehicles utilizing CNN and LSTM-based Deep Networks. 2019 IEEE Intelligent Vehicles Symposium (IV), pp. 2423–2428.
- Wang, D., Ma, X. & Yang, X. [Preprint at https://arxiv.org/abs/2203.15006]. (2022). TL-GAN: Improving Traffic Light Recognition via Data Synthesis for Autonomous Driving.
- Wang, H., Wang, J., Wang, J., Zhao, M., Zhang, W., Zhang, F., Xie, X. & Guo, M. (2018a). Graphgan: Graph representation learning with generative adversarial nets. *Thirty-second* AAAI conference on artificial intelligence.
- Wang, J., Paritala, P. K., Mendieta, J. B., Komori, Y., Raffel, O. C., Gu, Y. & Li, Z. (2020). Optical coherence tomography-based patient-specific coronary artery reconstruction and fluid–structure interaction simulation. *Biomechanics and Modeling in Mechanobiology*, 19(1), 7–20.
- Wang, N., Zhang, Y., Li, Z., Fu, Y., Liu, W. & Jiang, Y.-G. (2018b). Pixel2mesh: Generating 3d mesh models from single rgb images. *Proceedings of the European Conference on Computer Vision (ECCV)*, 52–67.

- Wang, T., Lei, Y., Curran, W. J., Liu, T. & Yang, X. (2021). Contrast-enhanced MRI synthesis from non-contrast MRI using attention CycleGAN. *Medical Imaging 2021: Biomedical Applications in Molecular, Structural, and Functional Imaging*, 11600, 116001L.
- Xie, H., Yao, H., Sun, X., Zhou, S. & Zhang, S. (2019). Pix2Vox: Context-aware 3D reconstruction from single and multi-view images. *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2690–2698.
- Yoo, T. K., Choi, J. Y., Kim, H. K., Ryu, I. H. & Kim, J. K. (2021). Adopting low-shot deep learning for the detection of conjunctival melanoma using ocular surface images. *Computer Methods and Programs in Biomedicine*, 205, 106086.
- Yoon, S., Yoon, C. H. & Lee, D. (2021). Topological recovery for non-rigid 2D/3D registration of coronary artery models. *Computer Methods and Programs in Biomedicine*, 200, 105922.
- Yue, G., Ping, G. & Lanxin, L. (2018). An End-to-End model based on CNN-LSTM for Industrial Fault Diagnosis and Prognosis. 2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC), pp. 274–278.
- Zhang, P., Wang, D., Lu, H., Wang, H. & Yin, B. (2017). Learning uncertain convolutional features for accurate saliency detection. *Proceedings of the IEEE International Conference* on computer vision, pp. 212–221.
- Zheng, C., Bian, F., Li, L., Xie, X., Liu, H., Liang, J., Chen, X., Wang, Z., Qiao, T., Yang, J. et al. (2021). Assessment of generative adversarial networks for synthetic anterior segment optical coherence tomography images in closed-angle detection. *Translational Vision Science & Technology*, 10(4), 34–34.
- Zheng, S., Meiying, T. & Jian, S. (2010). Sequential reconstruction of vessel skeletons from X-ray coronary angiographic sequences. *Computerized Medical Imaging and Graphics*, 34(5), 333–345.
- Zhu, J.-Y., Park, T., Isola, P. & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE international conference* on computer vision, pp. 2223–2232.