



Universiteit  
Leiden  
The Netherlands

## Deep learning for vascular segmentation and tissue characterization in CT images

Zhang, X.

### Citation

Zhang, X. (2026, January 7). *Deep learning for vascular segmentation and tissue characterization in CT images*. Retrieved from <https://hdl.handle.net/1887/4286096>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/4286096>

**Note:** To cite this publication please use the final published version (if applicable).

# Deep Learning for Vascular Segmentation and Tissue Characterization in CT Images

Xiaotong Zhang

## Colophon

### About the cover:

The cover was created with the assistance of ChatGPT, using a denoising diffusion process of a 3D hepatic vascular tree as a reference for image generation. The front side of the cover depicts an anatomical representation of the hepatic portal venous system, highlighting the intricate vascular network within the liver. The back side can be interpreted as projections of the liver from different perspectives.

Deep Learning for Vascular Segmentation and Tissue Characterization in CT Images  
Xiaotong Zhang

ISBN: 978-94-6522-990-4

Thesis layout & cover designed by Xiaotong Zhang

Printed by Ridderprint, the Netherlands

© 2025 Xiaotong Zhang, Leiden, the Netherlands

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the copyright owner.

# Deep Learning for Vascular Segmentation and Tissue Characterization in CT Images

## Proefschrift

ter verkrijging van  
de graad van doctor aan de Universiteit Leiden,  
op gezag van rector magnificus prof.dr.ir. H. Bijl,  
volgens besluit van het college voor promoties  
te verdedigen op dinsdag 7 januari 2026  
klokke 10:00 uur

door

Xiaotong Zhang  
geboren te Jining, Shandong, China  
in 1995

Promotor: Prof. dr. ir. B. P. F. Lelieveldt

Co-advisor: Dr. ir. J. Dijkstra  
Dr. ir. A. Broersen

Leden promotiecommissie: Prof. dr. P. H. A. Quax  
Prof. dr. ir. C. I. Sánchez  
*University of Amsterdam, Amsterdam*  
Dr. ir. T. van Walsum  
*Erasmus University Medical Center, Rotterdam*  
Dr. M. C. Burgmans

The research in this thesis was performed at the Division of Image Processing (LKEB),  
Department of Radiology of Leiden University Medical Center, The Netherlands.

Financial support for the publication of this thesis was kindly provided by:  
LKEB  
Library of Leiden University

# Contents

<b>1</b>	<b>General Introduction</b>	<b>1</b>
1.1	Vessel CTA . . . . .	1
1.2	Coronary plaque detection . . . . .	2
1.3	Liver anatomy . . . . .	4
1.4	Preoperative planning for liver surgery . . . . .	5
1.5	Deep learning in CTA images . . . . .	6
1.6	Thesis outline . . . . .	7
	References . . . . .	9
<b>2</b>	<b>Cross-sectional angle prediction of lipid-rich and calcified tissue on computed tomography angiography images</b>	<b>15</b>
2.1	Introduction . . . . .	18
2.2	Materials . . . . .	19
	2.2.1 Data acquisition . . . . .	19
	2.2.2 Ground truth . . . . .	19
	2.2.3 Data distribution . . . . .	21
2.3	Methods . . . . .	21
	2.3.1 Prediction workflow . . . . .	22
	2.3.2 Training of the networks . . . . .	25
	2.3.3 Evaluation method . . . . .	26
2.4	Results and evaluations . . . . .	26
	2.4.1 Plaque localization . . . . .	26
	2.4.2 Tissue-type distribution . . . . .	27
	2.4.3 Prediction evaluation . . . . .	31
2.5	Discussion and conclusion . . . . .	32
	References . . . . .	35
<b>3</b>	<b>Continuous and complete liver vessel segmentation with graph-attention guided diffusion</b>	<b>37</b>

3.1	Introduction . . . . .	40
3.2	Related Work . . . . .	41
3.2.1	Liver vessel segmentation . . . . .	41
3.2.2	Diffusion models for medical image segmentation . . . . .	42
3.2.3	Graph-based methods for medical image segmentation . . . . .	42
3.3	Methods . . . . .	43
3.4	Diffusion conditioning models . . . . .	43
3.4.1	Vanilla conditioning diffusion model . . . . .	43
3.4.2	Dynamic Conditioning model . . . . .	44
3.4.3	Multiscale graph-attention conditioning model . . . . .	44
3.4.4	Overall diffusion model conditioning . . . . .	47
3.4.5	Overall loss function . . . . .	47
3.5	Experiments and results . . . . .	47
3.5.1	Datasets . . . . .	47
3.5.2	Data pre- and post-processing . . . . .	48
3.5.3	Evaluation metrics . . . . .	48
3.5.4	Experimental setting . . . . .	49
3.5.5	Quantitative comparison . . . . .	49
3.5.6	Qualitative comparison . . . . .	53
3.5.7	Model ablation study . . . . .	57
3.6	Discussion . . . . .	60
3.6.1	Limitation . . . . .	60
3.7	Conclusion . . . . .	62
	References . . . . .	63
<b>4</b>	<b>Top-k maximum intensity projection priors for 3D liver vessel segmentation</b>	<b>67</b>
4.1	Introduction . . . . .	70
4.2	Methods . . . . .	71
4.2.1	Ground truth transformation . . . . .	73
4.2.2	Top-k MIP transformation . . . . .	73
4.2.3	Top-k MIP conditioning latent diffusion model . . . . .	74
4.2.4	Post-processing for artifact suppression . . . . .	74
4.3	Experiments and results . . . . .	75
4.3.1	Experimental setting . . . . .	75
4.3.2	Quantitative and qualitative comparison . . . . .	75
4.4	Discussion . . . . .	77
4.4.1	Effect of artifact suppression . . . . .	77
4.4.2	Limitations and possible improvements . . . . .	77
4.5	Conclusion . . . . .	78

References . . . . .	79
<b>5 Point-based Couinaud segmentation without liver vessel priors</b>	<b>81</b>
5.1 Introduction . . . . .	84
5.2 Related work . . . . .	85
5.2.1 Couinaud segmentation . . . . .	85
5.2.2 Dynamic graph reasoning . . . . .	85
5.3 Methods . . . . .	85
5.3.1 Relation to adaptive graph CNN . . . . .	87
5.3.2 Implicit anatomic contextual knowledge . . . . .	87
5.4 Experiments and results . . . . .	88
5.4.1 Datasets . . . . .	88
5.4.2 Experimental setting . . . . .	88
5.4.3 Evaluation metrics . . . . .	89
5.4.4 Quantitative comparison . . . . .	89
5.4.5 Qualitative comparison . . . . .	89
5.4.6 Model ablation study . . . . .	92
5.5 Discussion . . . . .	92
5.6 Conclusion . . . . .	95
References . . . . .	97
<b>6 Summary and future work</b>	<b>101</b>
6.1 Summary . . . . .	101
6.2 Discussion and future work . . . . .	102
6.3 General conclusions . . . . .	104
References . . . . .	105
<b>7 Samenvatting en toekomstig onderzoek</b>	<b>107</b>
7.1 Samenvatting . . . . .	107
7.2 Discussie en toekomstig onderzoek . . . . .	109
7.3 Algemene conclusies . . . . .	111
Referenties . . . . .	112
<b>List of publications</b>	<b>113</b>
<b>Acknowledgements</b>	<b>115</b>
<b>Curriculum Vitae</b>	<b>117</b>



# 1

## General Introduction

With the rapid advances in CT acquisition technology over the past decade, CT Angiography has now become commonplace in diagnostic and interventional radiology. CTA is primarily used to depict complex vasculature, providing detailed insights in disease processes such as atherosclerosis. In addition, CTA also provides valuable insights into individual anatomical features of patient vasculature, and as such may help in planning interventional procedures. However, the 3D nature and complexity of vascular anatomy necessitates analysis and visualization methods that assist the radiologist in the interpretation of large CTA datasets. Such methods should work automatically and reliably, such that the time expenditure for the radiologist to analyze CTA data is minimized.

A large body of prior research has been performed on automating the analysis of CT Angiography data. These methods are still suboptimal in their segmentation performance and robustness, and a fully automatic and reliable method remains a technical challenge. However, the recent emergence of AI methods such as deep learning holds promise for further automating this analysis, and for improving robustness of this automation. This thesis focuses on enhancing CTA image analysis and interpretation using the latest deep learning technologies. The overarching aim of this thesis is to overcome existing technical limitations in the automated analysis of CTA imaging data by developing advanced deep learning methods for two CTA application domains: coronary computed tomographic angiography (CCTA) and CTA of liver vasculature. In the following sections, we first introduce CT Angiography in more detail, followed by an introduction on the clinical context of the two applications addressed in this thesis and a detailed outline of the thesis.

### 1.1 Vessel CTA

CTA is one of the commonly used imaging modalities for diagnosing vascular diseases, offering a minimally invasive approach that reduces patient risk[1, 2]. Spatial resolution, temporal resolution, and volume coverage are three key image quality metrics of concern in CTA data acquisition[3]. To achieve faster image acquisition and greater scan volume coverage, multidetector-row computed tomography (MDCT)[4] was developed, which has greatly facilitated the clinical use of CTA. In addition to the

requirement for advanced CT techniques, contrast medium enhancement is a crucial component of CTA acquisition. A contrast medium with a high iodine concentration is considered ideal for clinical use [3]. Following intravenous injection, the injected contrast medium travels through the arm veins to the right heart chambers, passes through the lungs, and then enters the left heart chambers before reaching the arterial system[1]. After being distributed throughout the organs via the vascular system, the contrast medium returns to the right heart in a process known as recirculation [1]. The vascular lumen can be distinguished from the vessel wall based on the enhancement properties of the contrast medium.

## 1.2 Coronary plaque detection

Coronary Artery Disease (CAD), the leading cause of cardiovascular mortality and morbidity worldwide[5, 6, 7], is a pathological process of atherosclerotic plaque accumulation within the epicardial coronary arteries[8]. Atherosclerotic plaques can be categorized as either stable or vulnerable plaques and can be represented by various phenotypes. Stable plaque phenotypes contain larger calcium deposits which is significantly associated with suboptimal results of coronary interventions[8]. Vulnerable plaque phenotypes include a large lipid-rich necrotic core, thin fibrous cap, and spotty calcifications. An increased size in the lipid-rich necrotic core is highly correlated with an increased risk of Major Adverse Cardiovascular Event (MACE)[8], and the thin-fibrous caps that cover the lipid-rich cores may lead to plaque rupture[9]. Vulnerable plaques with these characterizations are high-risk plaques (HRP) shown in Fig. 1.1 and can develop into culprit lesions associated with acute coronary syndrome (ACS)[10]. Large observational studies show that patients with vulnerable plaques are five times more likely to experience MACE than patients without[8]. Thus, identifying high-risk features is clinically significant for individuals at high risk for future cardiovascular events.

Coronary plaque imaging modalities can be broadly classified as either invasive or noninvasive. Invasive imaging modalities shown in Fig. 1.2, such as intravascular ultrasound (IVUS), optical coherence tomography (OCT) and near infrared spectroscopy (NIRS), are considered the gold standard for plaque quantification and characterization[11]. IVUS is normally combined with post-processing methods, such as virtual histology intravascular ultrasound (VH-IVUS), to assess plaque composition. VH-IVUS can distinguish between fibrous, fibrofatty, necrotic core, and dense calcium[11], but its limited efficacy in identifying vulnerable plaques and low resolution limit its effectiveness[13, 14]. Although OCT can provide higher-resolution images than IVUS, assessing lipid cores with OCT may lead to up to 20% false positives because macrophages and calcifications have similar signal representations compared to lipid cores in OCT images[15]. Additionally, OCT's limited tissue penetration makes

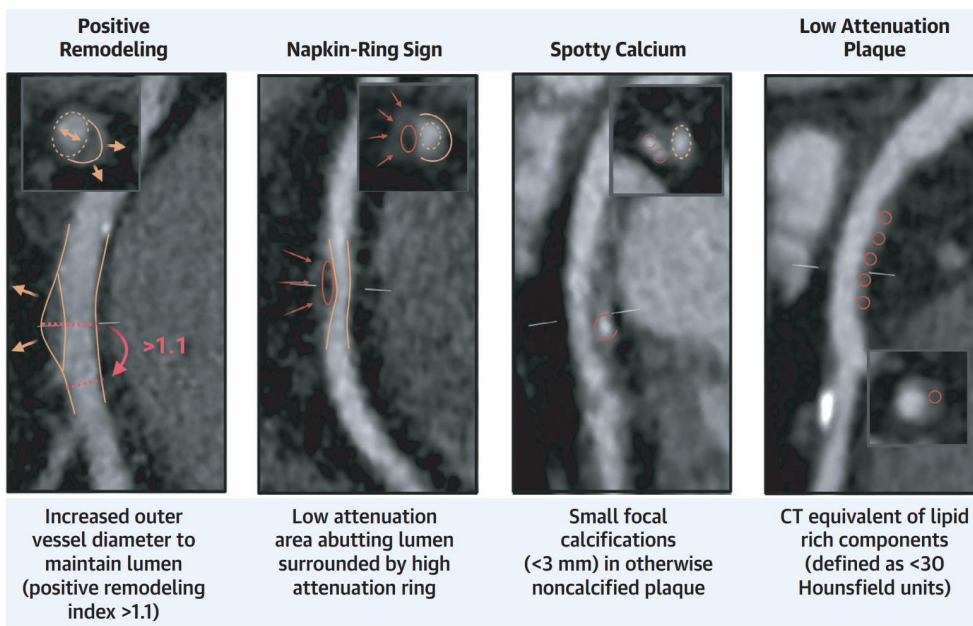


Figure 1.1: High-risk plaque features. (Figures reproduced from [11] under the CC BY-NC license)

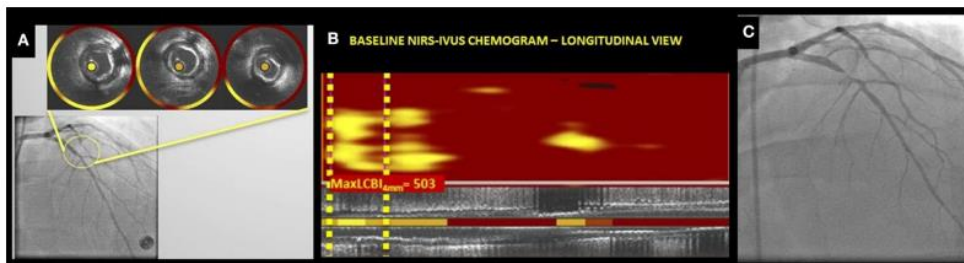


Figure 1.2: NIRS-IVUS imaging modality. (A) IVUS cross-sections and NIRS-IVUS rings, (B) NIRS-IVUS chemogram in longitudinal view, (C) vessel angiogram. (Figures reproduced from [12] under CC BY license.)

volumetric quantification of lipid plaque burdens difficult because the abluminal border of lipid pools cannot be detected.[8]. NIRS can accurately identify lipid areas based on their spectral properties. However, it cannot distinguish non-lipid components[8, 16], such as fibrous or calcified tissues, nor can it provide anatomical information about vessels. To overcome the limitations of IVUS and take advantage of NIRS, a combined NIRS-IVUS imaging modality shown in Fig. 1.2 has been proposed. This modality accurately detects lipid-rich plaque and visualizes lumen and plaque structures[13, 17]. However, all these invasive modalities increase the risk of harm to the patient.

Recent developments in noninvasive cardiac imaging modalities, such as computed tomography angiography (CTA), enable precise plaque phenotyping with minimal invasiveness. Coronary CTA can provide quantification of calcified, non-calcified, and low-attenuation plaque, as verified by intravascular imaging and histology[18, 19]. Noninvasive CTA can identify various high-risk features of vulnerable plaque, such as areas of low attenuation ( $<30$  Hounsfield units (HU)) on CT scans and napkin-ring signs, which are low-attenuation central areas surrounded by peripheral hyperattenuation[8]. Thus, CTA is a safe and efficient alternative for plaque characterization compared to invasive imaging modalities. However, HU-based plaque components differentiation is unreliable due to the significant overlap between the HU range of lipid-rich or fibrous plaque and that of the surrounding myocardial tissue[20, 21]. Some unclear cases still require more accurate invasive assessment[8], and a reduction in the dependence on HU ranges when differentiating coronary plaque components based on CTA images.

### **1.3 Liver anatomy**

Advances in hepatic surgery are significantly facilitated by exploring intrahepatic anatomy. Claude Couinaud (1922-2008) demonstrated that hepatic functional anatomy depends on vascular and biliary relationships rather than external surface anatomy[23, 24]. As shown in Fig. 1.3, the liver can be divided into eight functionally independent segments based on the distribution of the hepatic and portal veins in the liver[25], each segment has its own vascular inflow and outflow. The three hepatic veins divide the liver into four sections: the right lateral section, the right medial section, the left medial section, and the left lateral section. The planes containing the portal pedicles further divide the right lateral, right medial, and left lateral sections into upper and lower segments, resulting in segments II to VIII. In addition, segment I (the caudate lobe) lies posteriorly and functions independently from the other segments, receiving branches from both the right and left portal veins. These functionally independent segments provide critical anatomical guidance for liver surgery, enabling precise preoperative planning. Further details on preoperative planning for liver surgery are discussed in the following section.

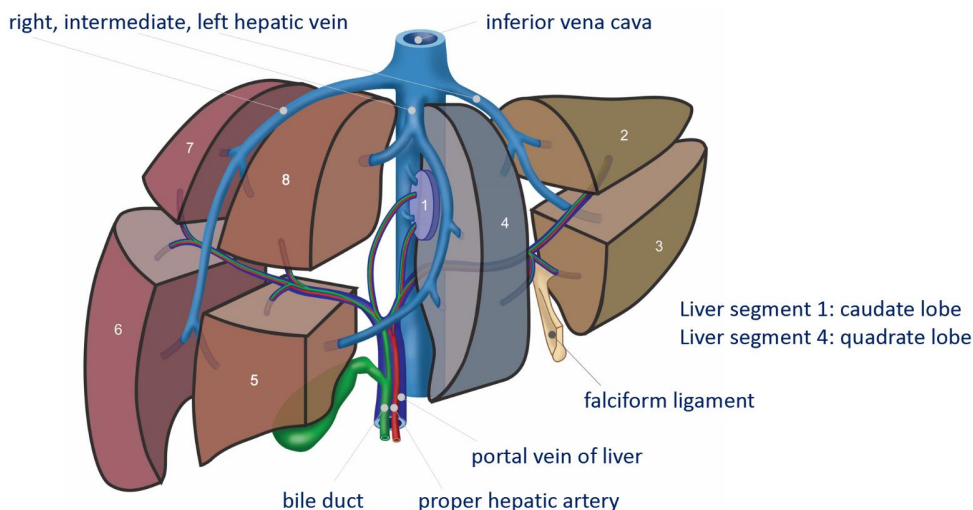


Figure 1.3: Segmental anatomy of liver organ. Eight functionally independent segments are divided based on the distribution of liver veins. (Figures reproduced from [22] under CC BY-NC-SA license.)

#### 1.4 Preoperative planning for liver surgery

Clinical treatments for Hepatocellular carcinoma (HCC) and colorectal cancer (CRC) are highly dependent on accurate preoperative planning [27, 28]. There are two key tasks in the preoperative planning of liver surgery. One involves identifying the relative positions of other organs, vessels, and tumor locations. The other involves planning the cutting path for liver resection or the puncture path for percutaneous radiofrequency ablation (RFA). Currently, computer-aided diagnosis systems are normally used for preoperative planning and surgery risk assessment. Designing an appropriate computer-assisted diagnosis system requires 3D visualization of liver anatomy as shown in Fig. 1.4, including main liver vessels, tumors, and self-contained liver segments, etc. For instance, during liver resection surgery, image segmentation of the liver and its internal structures is necessary for navigating the procedure and ensuring that the remaining liver percentage after resection is sufficient [27, 29]. Through preoperative planning, surgeons can identify the liver segment at risk to guide the resection of the tumor and the surrounding tissue [30]. The tumor cannot be resected by making an arbitrary cut because the hepatic or portal vein could be cut. Accidentally cutting the main hepatic vein could negatively affect blood drainage from the cancerous part of the liver, and cutting the portal vein could cut off the supply path from the other parts of the liver [31, 32, 33]. As with resection treatment, accurate preoperative planning is necessary for RFA treatment to minimize the risk of main

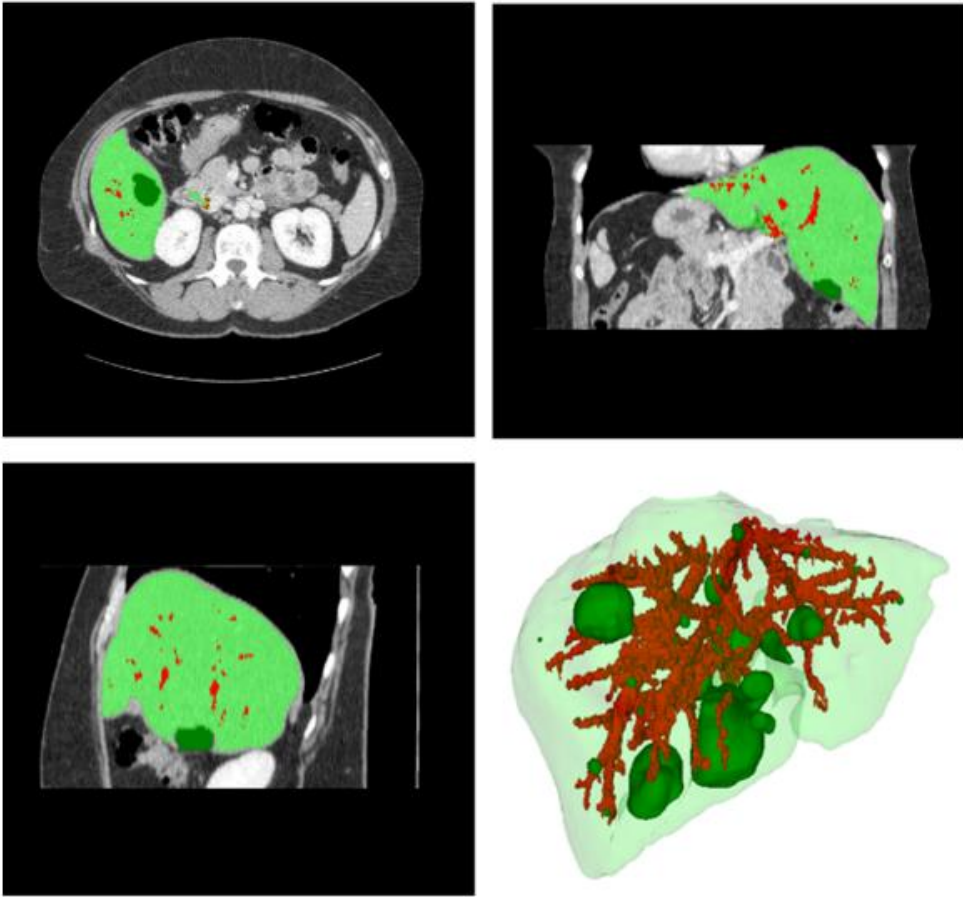


Figure 1.4: Spatial locations of liver vessels and tumors. (Figures reproduced from [26] under CC BY-NC-ND.)

vessel penetration and ensure proper needle placement[28, 34, 35]. Therefore, in order to perform precise surgical navigation and minimize the risk of cutting a major vessel, it is essential to divide the liver into functionally independent segments and accurately segment the liver vessels. However, accurately segmenting liver vessels and functional segments is challenging. The main challenges for accurate liver vessel segmentation are the complex vascular structures, thin branches, and small vessels. Inaccurate liver vessel segmentation can negatively impact the segmentation of the functional liver segments.

### 1.5 Deep learning in CTA images

Coronary CTA is currently the most frequently used noninvasive method for diagnosing coronary artery disease (CAD), providing information on stenosis quantifica-

tion, plaque classification, and plaque component characterization[36, 37, 38, 39]. However, CTA image interpretation remains a subjective task that depends on the experience of clinical experts[36]. To make CTA-based diagnosis more reliable and reproducible, convolutional neural networks (CNNs) have proven to be powerful tools for feature-based automated image analysis. CNN-based methods for CTA image analysis can be roughly divided into morphological analysis, including automatic coronary segmentation and centerline extraction[40, 41, 42], as well as functional analysis, such as coronary stenosis quantification and plaque characterization[43, 44, 45]. Despite the great success of deep learning in these tasks, noninvasive CTA still cannot replace invasive modalities for diagnosis, especially in ACS patients with ambiguous intravascular features on CTA images, such as low-attenuation lipid-rich tissue and other non-calcified plaque components mentioned in Fig. 1.1. Thus, to reduce the reliance on invasive modalities for diagnosis, exploring deep learning methods to learn the feature mapping between noninvasive CTA and invasive imaging has significant clinical value.

In addition to deep learning methods specifically developed for CTA images, many advanced techniques designed for general medical image analysis across other imaging modalities can also be adapted to the CTA domain. Medical image segmentation is one of the most widely studied tasks in this field. It plays a crucial role in medical image analysis by aiming to identify pixel-level associations with specific organs or lesions [46]. Numerous studies have demonstrated that CNN-based deep learning methods achieve accurate medical image segmentation with reduced sensitivity to image quality and significantly outperform traditional mathematical models and conventional machine learning techniques[47]. In addition to the well-known CNN-based methods, many advanced models from the field of computer vision, such as graph convolutional network (GCN)[48], point-based convolutional network (PCNN)[49], generative adversarial network (GAN)[50], and denoising diffusion probabilistic model (DDPM)[51], can be effectively adapted for downstream medical image segmentation tasks, such as the widely studied multi-organ segmentation[52, 53, 54, 55], etc. However, due to the inherently three-dimensional nature of most medical imaging modalities, extensive model optimization is often required to preserve anatomical completeness and continuity, unlike segmentation tasks in natural image processing.

## 1.6 Thesis outline

The overarching goal of this thesis is to overcome existing technical limitations and develop advanced deep learning methods for coronary plaque characterization, liver vessel segmentation, and Couinaud segmentation. All of the methods proposed in the thesis were developed using CTA modalities. This thesis is guided by the following three specific aims:

1. To investigate whether coronary plaque can be accurately detected and characterized using CTA images, thereby eliminating the need for invasive diagnostic procedures.
2. To explore the potential of 2D deep learning models to improve the continuity of 3D liver vessel segmentation.
3. To examine whether Couinaud liver segments can be delineated without relying on vessel priors, using deep learning-based approaches.

Specific Aim 1 is addressed in Chapter 2, which presents two methods for the detection and characterization of lipid-rich and calcified plaques in coronary CTA images. The two methods learn to map plaque features between gold-standard yet risky invasive modalities and the safe, non-invasive CTA modality. Unlike existing deep learning methods for plaque detection, we first propose predicting the plaque occurrence angle in the cross-sectional view to achieve a finer visualization of the angular distribution of coronary plaque.

Specific Aim 2 is addressed in Chapter 3 and Chapter 4. Chapter 3 presents a novel graph-attention guided diffusion model for continuous and complete liver vessel segmentation in CT images. We use the graph anatomy to constrain the diffusion model, which generates liver vessel segmentation with good 3D connectivity.

Chapter 4 presents a novel top-k maximum intensity projection (MIP) prior for continuous liver vessel segmentation in CT images. We mimic 3D segmentation as a CT reconstruction process. We use a latent diffusion model to learn the mapping between the top-k maximum intensity projections (MIPs) and the integral projections (IPs) of the ground truth liver vessel masks. Liver vessel segmentation can be reconstructed using the filtered back projection method based on the generated IPs.

Chapter 5 addresses thesis aim 3 by presenting a vessel-prior-free graph reasoning point net for Couinaud segmentation in CT images. Unlike the liver vessel-dependent delineation used in clinical practice, the graph reasoning module learns implicit graph embeddings instead of using explicit liver vessels for Couinaud segmentation.

Chapter 6 discusses the studies in this thesis in the context of the envisioned clinical applications and provides insights for further related research.

## References

- [1] A. Napoli, D. Fleischmann, F. P. Chan, et al. “Computed tomography angiography: state-of-the-art imaging using multidetector-row technology”. In: *Journal of computer assisted tomography* 28 (2004), S32–S45.
- [2] D. Fleischmann, A. S. Chin, L. Molvin, et al. “Computed tomography angiography: a review and technical update”. In: *Radiologic Clinics* 54.1 (2016), pages 1–12.
- [3] K. K. Kumamaru, B. E. Hoppel, R. T. Mather, and F. J. Rybicki. “CT angiography: current technology and clinical use”. In: *Radiologic clinics of North America* 48.2 (2010), page 213.
- [4] D. Fleischmann, G. D. Rubin, D. S. Paik, et al. “Stair-step artifacts with single versus multiple detector-row helical CT”. In: *Radiology* 216.1 (2000), pages 185–196.
- [5] N. J. Montarello, K. Singh, A. Sinhal, et al. “Assessing the impact of colchicine on coronary plaque phenotype after myocardial infarction with optical coherence tomography: rationale and design of the COCOMO-ACS study”. In: *Cardiovascular Drugs and Therapy* 36.6 (2022), pages 1175–1186.
- [6] A. S. Go, D. Mozaffarian, V. L. Roger, et al. “Heart disease and stroke statistics—2014 update: a report from the American Heart Association”. In: *circulation* 129.3 (2014), e28–e292.
- [7] E. J. Benjamin, P. Muntner, A. Alonso, et al. “Heart disease and stroke statistics—2019 update: a report from the American Heart Association”. In: *Circulation* 139.10 (2019), e56–e528.
- [8] A. Milzi, A. Landi, R. Dettori, et al. “Multimodal Intravascular Imaging of the Vulnerable Coronary Plaque”. In: *Echocardiography* 41.12 (2024), e70035.
- [9] I. Gianopoulos and S. S. Daskalopoulou. “Macrophage profiling in atherosclerosis: understanding the unstable plaque”. In: *Basic Research in Cardiology* 119.1 (2024), pages 35–56.
- [10] R. Wolny, J. Geers, K. Grodecki, et al. “Noninvasive atherosclerotic phenotyping: the next frontier into understanding the pathobiology of coronary artery disease”. In: *Current atherosclerosis reports* 26.7 (2024), pages 305–315.
- [11] L. P. Dawson, M. Lum, N. Nerleker, et al. “Coronary atherosclerotic plaque regression: JACC state-of-the-art review”. In: *Journal of the American College of Cardiology* 79.1 (2022), pages 66–82.
- [12] K. Kuku, M. Singh, Y. Ozaki, et al. “Near-infrared spectroscopy intravascular ultrasound imaging: state of the art”. In: *Frontiers in Cardiovascular Medicine* 7 (2020), page 107.
- [13] J. Pu, G. S. Mintz, E. S. Brilakis, et al. “In vivo characterization of coronary plaques: novel findings from comparing greyscale and virtual histology intravascular ultrasound and near-infrared spectroscopy”. In: *European Heart Journal* 33.3 (2012), pages 372–383.

- [14] T. W. Johnson, L. Räber, C. Di Mario, et al. “Clinical use of intracoronary imaging. Part 2: Acute coronary syndromes, ambiguous coronary angiography findings, and guiding interventional decision-making: An expert consensus document of the European Association of Percutaneous Cardiovascular Interventions: Endorsed by the Chinese Society of Cardiology, the Hong Kong Society of Transcatheter Endocardiovascular Therapeutics (HKSTENT) and the Cardiac Society of Australia and New Zealand”. In: *European Heart Journal* 40.31 (2019), pages 2566–2584.
- [15] L. Di Vito, F. Imola, L. Gatto, et al. “Limitations of OCT in identifying and quantifying lipid components: an in vivo comparison study with IVUS-NIRS.” In: *EuroIntervention: journal of EuroPCR in collaboration with the Working Group on Interventional Cardiology of the European Society of Cardiology* 13.3 (2017), pages 303–311.
- [16] C. Gardner, H. Tan, E. Hull, et al. “Detection of lipid core coronary plaques in autopsy specimens with a novel catheter-based near-infrared spectroscopy system”. In: *JACC: Cardiovascular Imaging* 1.5 (2008), pages 638–648.
- [17] S.-J. Kang, G. S. Mintz, J. Pu, et al. “Combined IVUS and NIRS detection of fibroatheromas: histopathological validation in human coronary arteries”. In: *JACC: Cardiovascular Imaging* 8.2 (2015), pages 184–194.
- [18] A. Lin, M. Kolossváry, I. Igum, et al. “Artificial intelligence: improving the efficiency of cardiovascular imaging”. In: *Expert review of medical devices* 17.6 (2020), pages 565–577.
- [19] M. Kolossváry, C. N. De Cecco, G. Feuchtner, and P. Maurovich-Horvat. “Advanced atherosclerosis imaging by CT: radiomics, machine learning and deep learning”. In: *Journal of Cardiovascular Computed Tomography* 13.5 (2019), pages 274–280.
- [20] B. Szilveszter, C. Celeng, and P. Maurovich-Horvat. “Plaque assessment by coronary CT”. In: *The International Journal of Cardiovascular Imaging* 32 (2016), pages 161–172.
- [21] M. Kolossváry, B. Szilveszter, B. Merkely, and P. Maurovich-Horvat. “Plaque imaging with CT—a comprehensive review on coronary CT angiography based risk assessment”. In: *Cardiovascular Diagnosis and Therapy* 7.5 (2017), page 489.
- [22] Leiden University Medical Center. *Leiden Drawing of Liver Segments and Vascularisation (English labels)*. <https://anatomytool.org/content/leiden-drawing-liver-segments-and-vascularisation-english-labels>. Licensed under CC BY 4.0. n.d.
- [23] S. R. Abdel-Misih and M. Bloomston. “Liver anatomy”. In: *Surgical Clinics* 90.4 (2010), pages 643–653.
- [24] F. Sutherland and J. Harris. “Claude Couinaud: a passion for the liver”. In: *Archives of surgery* 137.11 (2002), pages 1305–1310.
- [25] J. E. Skandalakis, L. J. Skandalakis, P. N. Skandalakis, and P. Mirilas. “Hepatic surgical anatomy”. In: *Surgical Clinics* 84.2 (2004), pages 413–435.
- [26] D. A. Oliveira, R. Q. Feitosa, and M. M. Correia. “Segmentation of liver, its vessels and lesions from CT images for surgical planning”. In: *Biomedical Engineering Online* 10 (2011), pages 1–23.

- [27] O. I. Alirri and A. A. A. Rahni. “Survey on liver tumour resection planning system: steps, techniques, and parameters”. In: *Journal of Digital Imaging* 33.2 (2020), pages 304–323.
- [28] M. Luo, H. Jiang, and T. Shi. “Multi-stage puncture path planning algorithm of ablation needles for percutaneous radiofrequency ablation of liver tumors”. In: *Computers in Biology and Medicine* 145 (2022), page 105506.
- [29] H. E. Blum. “Hepatocellular carcinoma: therapy and prevention”. In: *World journal of gastroenterology* 11.47 (2005), page 7391.
- [30] P.-A. Clavien and Y. Fong. *Malignant liver tumors: current and emerging therapies*. Jones & Bartlett Learning, 2004.
- [31] J. G. Fortner and L. H. Blumgart. “A historic perspective of liver surgery for tumors at the end of the millennium”. In: *Journal of the American College of Surgeons* 193.2 (2001), pages 210–222.
- [32] O. A. Catalano, A. H. Singh, R. N. Uppot, et al. “Vascular and biliary variants in the liver: implications for liver surgery”. In: *Radiographics* 28.2 (2008), pages 359–378.
- [33] G. Stapleton, R. Hickman, and J. Terblanche. “Blood supply of the right and left hepatic ducts”. In: *British journal of surgery* 85.2 (1998), pages 202–207.
- [34] C. Baegert, C. Villard, P. Schreck, and L. Soler. “Multi-criteria trajectory planning for hepatic radiofrequency ablation”. In: *International conference on medical image computing and computer-assisted intervention*. Springer. 2007, pages 676–684.
- [35] C. Baegert, C. Villard, P. Schreck, et al. “Trajectory optimization for the planning of percutaneous radiofrequency ablation of hepatic tumors”. In: *Computer Aided Surgery* 12.2 (2007), pages 82–90.
- [36] D. Han, J. Liu, Z. Sun, et al. “Deep learning analysis in coronary computed tomographic angiography imaging for the assessment of patients with coronary artery stenosis”. In: *Computer Methods and Programs in Biomedicine* 196 (2020), page 105651.
- [37] M. Zreik, R. W. Van Hamersvelt, N. Khalili, et al. “Deep learning analysis of coronary arteries in cardiac CT angiography for detection of patients requiring invasive coronary angiography”. In: *IEEE transactions on medical imaging* 39.5 (2019), pages 1545–1557.
- [38] Y. Hong, F. Commandeur, S. Cadet, et al. “Deep learning-based stenosis quantification from coronary CT angiography”. In: *Proceedings of Spie—the International Society for Optical Engineering*. Volume 10949. 2019, page 109492I.
- [39] A. Lin, N. Manral, P. McElhinney, et al. “Deep learning-enabled coronary CT angiography for plaque and stenosis quantification and cardiac risk prediction: an international multicentre study”. In: *The Lancet Digital Health* 4.4 (2022), e256–e265.
- [40] W. Huang, L. Huang, Z. Lin, et al. “Coronary artery segmentation by deep learning neural networks on computed tomographic coronary angiographic images”. In: *2018 40th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE. 2018, pages 608–611.

- [41] H. Cui, Y. Xia, Y. Zhang, and L. Zhong. “Validation of right coronary artery lumen area from cardiac computed tomography against intravascular ultrasound”. In: *Machine Vision and Applications* 29.8 (2018), pages 1287–1298.
- [42] A. Dorobaniu, V. Ogreen, and R. Brad. “Coronary centerline extraction from ccta using 3d-unet”. In: *Future Internet* 13.4 (2021), page 101.
- [43] A. R. van Rosendael, G. Maliakal, K. K. Kolli, et al. “Maximization of the usage of coronary CTA derived plaque information using a machine learning based algorithm to improve risk stratification; insights from the CONFIRM registry”. In: *Journal of cardiovascular computed tomography* 12.3 (2018), pages 204–209.
- [44] M. Zreik, R. W. Van Hamersvelt, J. M. Wolterink, et al. “A recurrent CNN for automatic detection and classification of coronary artery plaque and stenosis in coronary CT angiography”. In: *IEEE transactions on medical imaging* 38.7 (2018), pages 1588–1598.
- [45] A. Gudigar, S. Nayak, J. Samanth, et al. “Recent trends in artificial intelligence-assisted coronary atherosclerotic plaque characterization”. In: *International journal of environmental research and public health* 18.19 (2021), page 10003.
- [46] S. Sushanki, A. K. Bhandari, and A. K. Singh. “A review on computational methods for breast cancer detection in ultrasound images using multi-image modalities”. In: *Archives of Computational Methods in Engineering* 31.3 (2024), pages 1277–1296.
- [47] A. K. Upadhyay and A. K. Bhandari. “Advances in deep learning models for resolving medical image segmentation data scarcity problem: A topical review”. In: *Archives of Computational Methods in Engineering* 31.3 (2024), pages 1701–1719.
- [48] S. Zhang, H. Tong, J. Xu, and R. Maciejewski. “Graph convolutional networks: a comprehensive review”. In: *Computational Social Networks* 6.1 (2019), pages 1–23.
- [49] A. Boulch. “ConvPoint: Continuous convolutions for point cloud processing”. In: *Computers & Graphics* 88 (2020), pages 24–34.
- [50] A. Aggarwal, M. Mittal, and G. Battineni. “Generative adversarial network: An overview of theory and applications”. In: *International Journal of Information Management Data Insights* 1.1 (2021), page 100004.
- [51] J. Ho, A. Jain, and P. Abbeel. “Denoising diffusion probabilistic models”. In: *Advances in neural information processing systems* 33 (2020), pages 6840–6851.
- [52] F. Tian, Z. Tian, Z. Chen, et al. “Surface-GCN: Learning interaction experience for organ segmentation in 3D medical images”. In: *Medical Physics* 50.8 (2023), pages 5030–5044.
- [53] D. Li, Y. Wei, and R. Zhu. “A comparative study on point cloud down-sampling strategies for deep learning-based crop organ segmentation”. In: *Plant Methods* 19.1 (2023), page 124.
- [54] L. P. Maguluri, K. Chouhan, R. Balamurali, et al. “Adversarial deep learning for improved abdominal organ segmentation in CT scans”. In: *Multimedia Tools and Applications* 83.35 (2024), pages 82107–82129.

- [55] Y. Wang, Z. Wan, Y. Qiu, and Z. Wang. “Devil is in Details: Locality-Aware 3D Abdominal CT Volume Generation for Self-Supervised Organ Segmentation”. In: *Proceedings of the 32nd ACM International Conference on Multimedia*. 2024, pages 10640–10648.



# 2

## Cross-sectional angle prediction of lipid-rich and calcified tissue on computed tomography angiography images

*This chapter was adapted from:*

**Zhang, X.**, Broersen, A., Sokooti, H., Ramasamy, A., Kitslaar, P., Parasa, R., Karaduman, M., Mohammed, A.S.A.J., Bourantas, C.V. and Dijkstra, J., **Cross-sectional angle prediction of lipid-rich and calcified tissue on computed tomography angiography images.** (2024) International Journal of Computer Assisted Radiology and Surgery, 19(5), pp.971-981, 2024.



## Abstract

The assessment of vulnerable plaque characteristics and distribution is important to stratify cardiovascular risk in a patient. Computed tomography angiography (CTA) offer a promising alternative to invasive imaging but is limited by the fact that the range of Hounsfield units (HU) in lipid-rich areas overlap with the HU range in fibrotic tissue and that the HU range of calcified plaques overlap with the contrast within the contrast filled lumen. This chapter is to investigate if lipid-rich and calcified plaques can be detected more accurately on cross-sectional CTA images using deep learning methodology. Two deep learning (DL) approaches are proposed, a 2.5D Dense U-Net and 2.5D Mask-RCNN, which separately perform the cross-sectional plaque detection in the Cartesian and polar domain. The spread-out view is used to evaluate and show the prediction result of the plaque regions. The accuracy and F1-score are calculated on a lesion level for the DL and conventional plaque detection methods. For the lipid-rich plaques, the median and mean values of the F1-score calculated by the two proposed DL methods on 91 lesions were approximately 6 and 3 times higher than those of the conventional method. For the calcified plaques, the F1-score of the proposed methods were comparable to those of the conventional method. The median F1-score of the Dense U-Net based method was 3% higher than that of the conventional method. The two methods proposed in this paper contribute to finer cross-sectional predictions of lipid-rich and calcified plaques compared to studies focusing only on longitudinal prediction. The angular prediction performance of the proposed methods outperforms the convincing conventional method for lipid-rich plaque and is comparable for calcified plaque.

## 2.1 Introduction

Coronary artery disease (CAD) is one of the primary causes of death in the world. Recent studies have demonstrated that outcomes in patients with CAD depends not only on lesion severity but also on its composition. Lesion with a lipid rich phenotype are at a higher risk to progress and cause major adverse cardiovascular events such as death or an acute coronary syndrome (ACS)[1, 2, 3, 4]. In order to accurately diagnose the severity of CAD and characterize plaque phenotype, image processing techniques used for atherosclerotic plaque detection should have the capacity to describe plaque composition and location in as much detail as possible.

In general, invasive imaging modalities, such as intravascular ultrasound (IVUS)[5], optical coherence tomography (OCT)[6] and near-infrared spectroscopy (NIRS)[7], are considered as the gold standard for characterizing plaque composition which closely matches the histopathology of the plaque. However, these approaches are invasive, are associated with risk and can be used only in patients with symptomatic obstructive CAD. Non-invasive modalities such as computed tomography angiography (CTA) have been widely used for atherosclerotic plaque detection as they enable complete assessment of the coronary artery tree and can assess coronary artery pathology[8, 9]. However, due to the significant attenuation overlap between lipid-rich plaques and fibrotic or even the surrounding myocardial tissue, it is a challenging task for the cardiologist to distinguish lipid-rich plaques from normal tissue and mixed plaques from calcified plaques on CTA images solely based on the Hounsfield units (HU) range[10].

To avoid tedious manually crafted and complex feature extractions, many[3, 8, 11, 12] deep learning (DL)-based methods have been proposed to classify atherosclerotic plaques on IVUS, OCT and CTA images. For the CTA-based methods, Zreik et al.[11] proposed a method that has a recurrent convolutional neural network (CNN) architecture to classify calcified and non-calcified plaques and evaluate stenosis degree on multi-planar reformatted (MPR) images . Liu et al.[12] utilized the straightened MPR volume as the input of a 3D CNN and predicted calcified, non-calcified and mixed plaque regions on a 3D CT volume. However, both methods provide a coarse plaque range prediction on the axial direction of vessel instead of a detailed prediction on the cross-sectional view.

Considering the limitations of existing methods[3, 8, 11, 12] for atherosclerosis plaque classification and the need to accurately classify plaque phenotypes we introduce alternative DL approaches for the prediction of plaque types, that were trained and tested using NIRS-IVUS pull-backs as the gold standard that were matched with CTA images. In the following part of the paper, we use the term *plaque* to represent the lipid-rich or calcified tissue.

Inspired by the classical architectures of the 2.5D U-Net[13], DenseNet[14] and the Mask R-CNN[15], two different methods for CT cross-sectional plaque classification are presented in this paper. This paper mainly contributes to three points. 1) The method is independent of the EEM detection. 2) Plaque angle prediction: this study focuses on the cross-sectional plaque distribution instead of longitudinal detection. 3) Refined visualization: the spread-out view from NIRS is used to evaluate and visualize the angle-wise plaque distribution.

## 2.2 Materials

### 2.2.1 Data acquisition

We analyzed 176 cardiac vessels from 64 patients who were prospectively recruited in the study, which aimed to examine the efficacy of CTA in assessing plaque pathology and physiology using NIRS-IVUS imaging as reference standard. In total 64 patients with obstructive CAD and typical angina symptoms were recruited and underwent CTA before being listed for percutaneous coronary intervention or functional assessment - depending on the clinical indication - and NIRS-IVUS imaging. The CTA was obtained by a dual-source CT scanner (Somatom Force, Siemens Healthineers, Forchheim, Germany). CTA images were reconstructed by highest strength model-based iterative reconstruction (ADMIRE 5) with a slice thickness of 0.50 mm and increments of 0.30 mm. NIRS-IVUS was performed in the major epicardial vessels and their large side branches with a diameter of more than 2 mm by a 2.4F Makoto™ NIRS-IVUS 35-65MHz Imaging System (Infraredx, Burlington, USA), that was withdrawn at a constant pull back speed 0.5 mm/s.

### 2.2.2 Ground truth

To train and test the DL methods using NIRS-IVUS images we co-registered this data with the CTA images; the registration process[16] shown in Fig. 2.1 and the plaque estimations of NIRS-IVUS were matched to the CT domain. The accuracy of the registration is closely related to the quality of the training labels, which affect the model performance directly. Several anatomical information that is seen in both modalities was used for this process including the origin of the side branches, location of the plaque and the presence of large calcific spots that are clearly visible in both datasets.

The plaque locations on IVUS were represented as pie-shaped labels as shown in Fig. 2.1, since NIRS only provides the angles with a high lipid plaque probability and calcium plaques block the ultrasound signal so the actual depth of calcium is not defined precisely.

The regions with probability larger than 0.6[17] on the original NIRS chemogram are considered as lipid-rich plaque. According to this threshold, pie-shaped binary

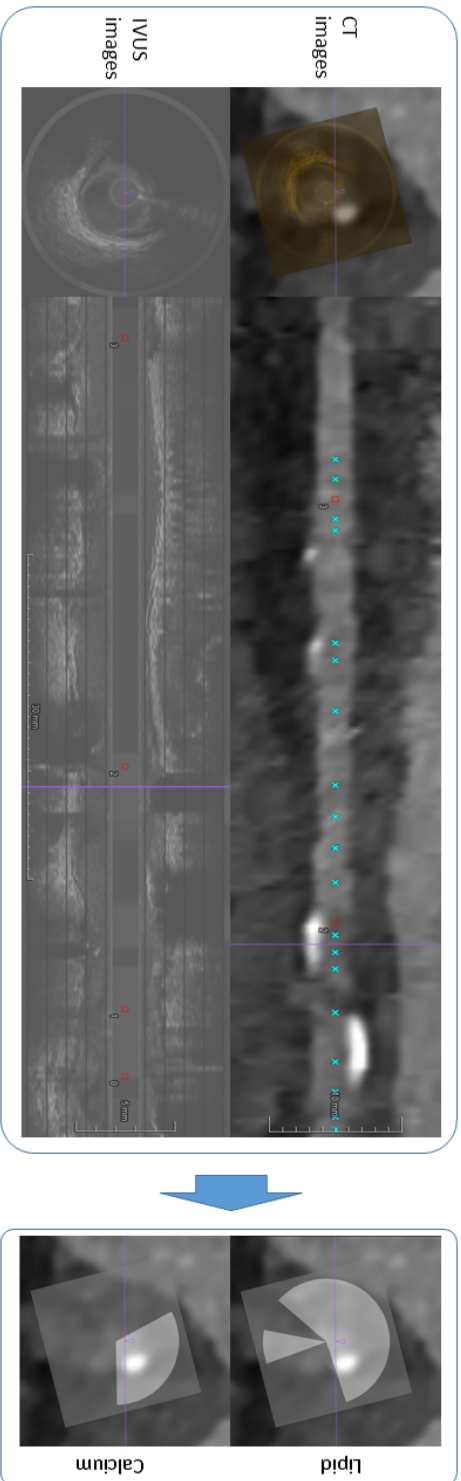


Figure 2.1: Vessel registration process to map IVUS images onto CT images.

masks originated from catheter center are generated. However, in most occasions the catheter center does not match to the lumen center, hence translations and rotations are required when co-registering the catheter centered mask from the IVUS domain to the CTA image domain. The calcium masks were drawn manually by a clinical expert. The registered binary mask labels of lipid-rich and calcified plaques on CTA image are shown in Fig. 2.2(d) and Fig. 2.2(e) respectively.

Registered masks are transformed to boxes in the polar view to fit the architecture of Mask R-CNN shown in Fig. 2.4.

The widely used semi-automated plaque analysis software (QAngioCT Research Edition 3.1, Medis Medical Imaging Systems BV, the Netherlands)[18] was used to segment the CTA data and classify plaque composition in the conventional approach. CTA analysis was performed by an experienced operator with an established reproducibility blindly to the NIRS-IVUS images. We will use conventional method to indicate QAngio CT in the following part.

### 2.2.3 Data distribution

After excluding the vessels with metal stents, there are 49 patients (44 for training and 5 for validation) in the training set and 15 patients in the independent test set. In the 2.5D Dense U-Net experiment, 18468 (lipid:1931, calcified:3297, mixed:1979, normal:11261), 2290 (lipid:392, calcified:247, mixed:171, normal:1480) and 7839 (lipid:851, calcified:1240, mixed:670, normal:5078) CTA frames separately from 44, 5 and 15 patients were used for training, validation and testing.

In the 2.5D Mask R-CNN experiment, CTA frames with plaque angle less than 10 degrees were excluded, and CTA frames without lipid or calcified tissue were also excluded in the training. 6640 (lipid:1768, calcified:3019, mixed:1853), 789 (lipid:372, calcified:246, mixed:171) and 7818 (lipid:833, calcified:1237, mixed:667, normal:5081) CTA frames separately from 44, 5 and 15 patients were used for training, validation and testing.

## 2.3 Methods

The 2.5D Dense U-Net-based method and the 2.5D Mask R-CNN-based method have different advantages in plaque prediction. The cross-sectional CTA images can be used directly as input for the Dense U-Net to learn global features to *distinguish plaques* but with a separate post-processing step shown in Fig. 2.3 for angular analysis. The Mask R-CNN does not need the post-processing because it is performed in polar view and the angular information can be obtained directly.

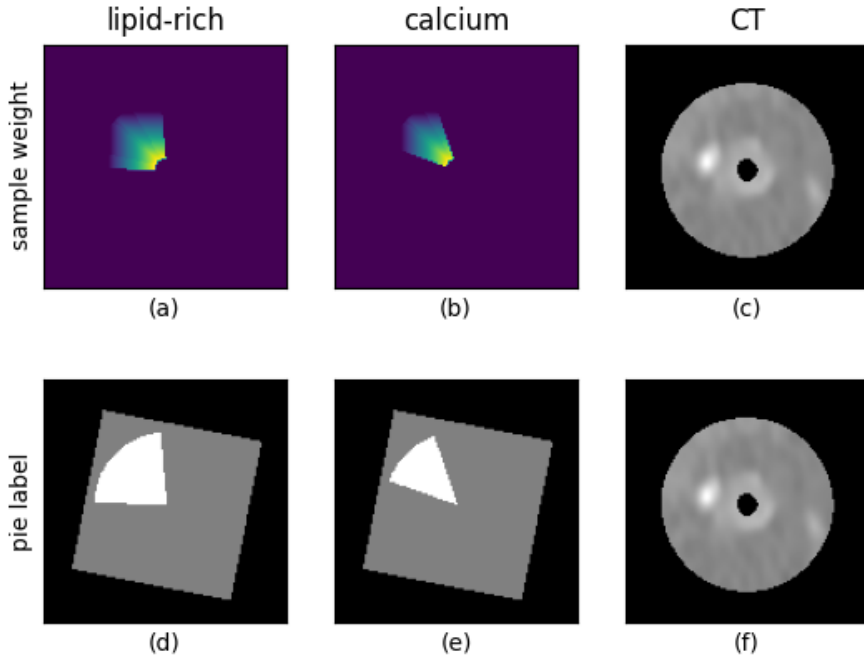


Figure 2.2: Binary labels and sample weights used for 2.5D Dense U-Net. (a)(b): pixel-wise gradient sample weights for lipid-rich plaques and calcified plaques; (d)(e): binary pie-shaped labels for lipid-rich and calcified plaques; (c)(f) cross-sectional CT image.

### 2.3.1 Prediction workflow

#### Dense U-Net

The pipeline of the 2.5D Dense U-Net is shown in Fig. 2.3. At the end of the pipeline of the U-Net, the prediction of lipid-rich and calcified plaques is obtained from two individual branches. A pixel-wise gradient sample weight is used to restrict the loss function during the training process. A special post-processing is applied to generate a spread-out plot to evaluate the model performance for the prediction of the plaque distribution.

Weighted binary cross entropy loss is defined in Eq.(2.1) to optimize separated branches in the 2.5D Dense U-Net. In Eq.(2.1),  $predict$  represents the predicted segmentation shown in Fig. 2.3,  $target$  represents the pie-shaped binary label shown in Fig. 2.2.  $weight_{sample}$  represents the pixel-wise gradient sample weight shown in Fig. 2.2. The pixel value on the sample weight map is inversely proportional to the distance between this pixel and the lumen center.  $Reg_{kernel}$  is a  $l_2$  norm regularization term which be used as a penalty term of the convolutional kernel to avoid overfitting

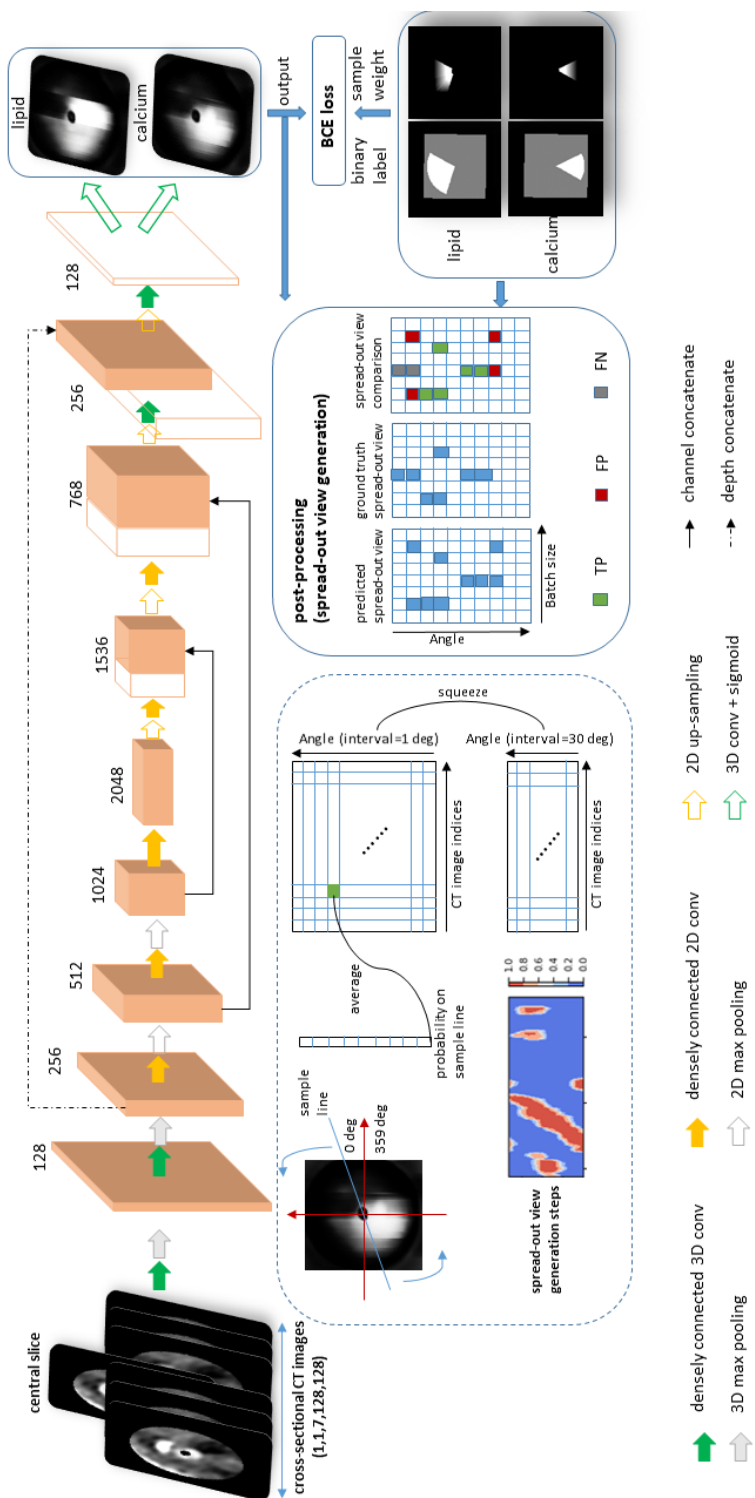


Figure 2.3: Pipeline of 2.5D Dense U-Net based method.

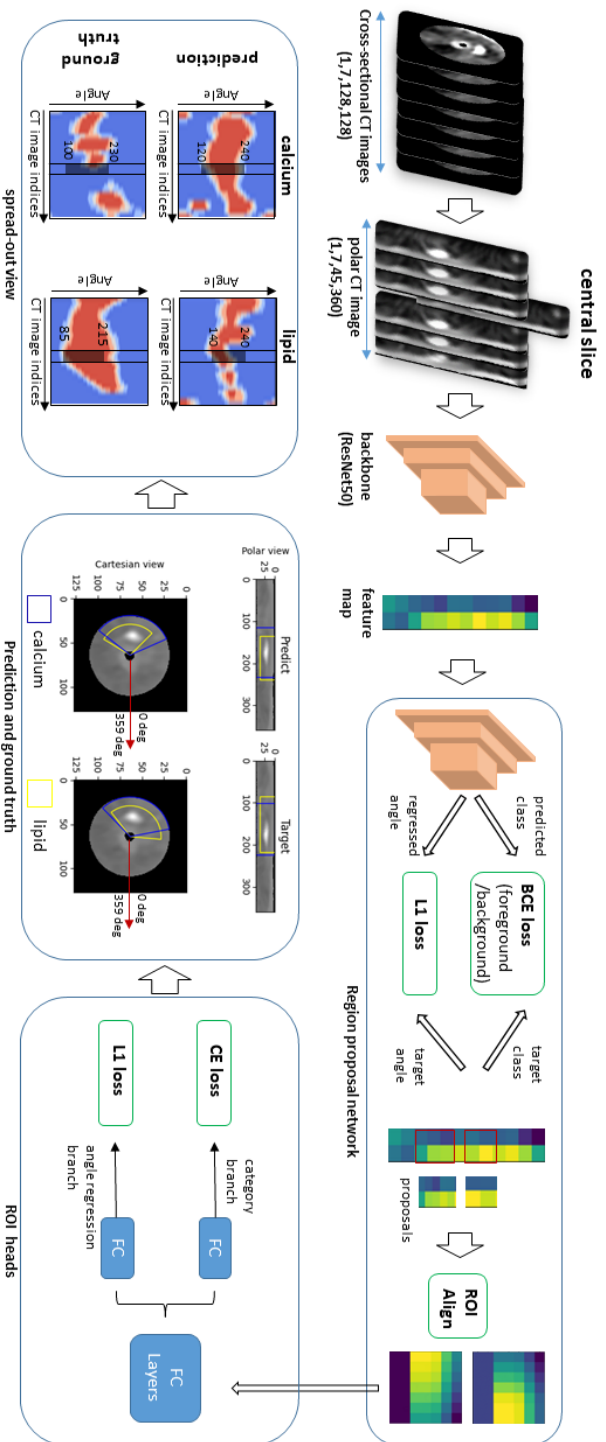


Figure 2.4: Pipeline of 2.5D Mask R-CNN based method.

during the training.

$$\begin{aligned} Loss_{DenseUNet} = & BCE(predict, target) \times weight_{sample} + Reg_{kernel} \\ & weight_{sample} \propto lumedistance \end{aligned} \quad (2.1)$$

### Mask R-CNN

The pipeline of the 2.5D Mask R-CNN is shown in Fig. 2.4. In this study, we adopt the classical Mask R-CNN architecture with a ResNet50[19] as a backbone. To take advantage of the context information along the longitudinal direction of the vessels, 2.5D polar CT images are used as the input to the network. The loss is  $loss_{classification} + loss_{regression} + loss_{segmentation}$  in the classical Mask R-CNN. There are no specific segmentation labels for plaques. Therefore, the segmentation branch in the classical Mask R-CNN[15] architecture is discarded in our architecture.

The loss function of 2.5D Mask R-CNN is a multi-task loss, combining a binary cross entropy loss with a  $l_1$  loss, shown as Eq.(2.2).  $loss_{cls_{bf}}$  and  $loss_{reg_{bf}}$  represents the classification loss and angle regression loss for distinguishing foreground and background,  $loss_{cls}$  and  $loss_{reg}$  represents further classification and regression loss used to separate different kinds of plaques.  $predict_{bf}^{class}$ ,  $predict_{bf}^{angle}$ ,  $predict^{class}$  and  $predict^{angle}$  are the predicted category and regressed angle.  $target_{bf}^{class}$ ,  $target_{bf}^{angle}$ ,  $target^{class}$  and  $target^{angle}$  are the target category and target angle.

$$\begin{aligned} Loss_{MaskRCNN} = & loss_{cls_{bf}} + loss_{reg_{bf}} + loss_{cls} + loss_{reg} \\ = & BCE(predict_{bf}^{class}, target_{bf}^{class}) \\ & + L1(predict_{bf}^{angle}, target_{bf}^{angle}) \\ & + BCE(predict^{class}, target^{class}) \\ & + L1(predict^{angle}, target^{angle}) \end{aligned} \quad (2.2)$$

#### 2.3.2 Training of the networks

All the experiments in this study used Adam optimizer with a learning rate of 0.0001 to train 50 epochs with a batch size of 32. The 2.5D data block for the dense U-Net was randomly rotated from 0 to 359 degrees around the image center which could increase the variation in plaque occurring angle. The size of the input data block is  $128 \times 128 \times 7$  for the 2.5D Dense U-Net and  $45 \times 360 \times 7$  for the 2.5D Mask R-CNN.

To avoid overfitting, a 50% random dropout was applied to the final layer of the encoder and a  $l_2$  norm regularization was introduced to each convolutional layer in the 2.5D Dense U-Net. A data ratio of *lipid:calcium:mixed:normal* tissue was set as 3:1:1:5 per batch.

The masking branch in the classical Mask R-CNN network was discarded since accurate segmentation labels of different plaque types do not exist, and only the

branches for classification and regression were maintained. Bounding box initialization was performed just on the angle direction (x-axis of the polar image) because no ground truth information along the depth direction (y-axis of the polar image) exists.

### 2.3.3 Evaluation method

In this patient level study, a lesion is defined as a certain region in the vessel that contains an elevated ( $\geq 40\%$ ) plaque burden. 91 lesions were extracted from the total 48 test vessels of 15 patients according to the expert's knowledge and all the metrics were calculated on the lesion-level. *Accuracy* and *F1-score*, calculated from the spread-out view, are used to evaluate the prediction. IOU, Dice and Hausdorff distance, which are commonly used in the pixel-based segmentation, are not applicable in this study. Because these metrics are calculated in the image domain but there are no pixel-level segmentation labels. The spread-out view shows the plaque distribution on the cross-sectional CTA images and can give a comprehensive overview about the continuity of the plaques in a vessel. The spread-out view generation process is shown on the left-bottom part in Fig. 2.3. If the coordinate of one element on the spread-out view is  $(\alpha, n)$ , then the value of this element represents the probability that there is a plaque at  $\alpha$  degree on CT image  $n$ .

Violin plots combined with box plots are used to show the distribution of spread-out view-based *accuracy* and *F1-score* of the lesions. The spread-out view based *accuracy* and *F1-score* are defined by  $Accuracy(SpreadOut_{pred}, SpreadOut_{gt})$  and  $F1(SpreadOut_{pred}, SpreadOut_{gt})$  that the inputs are predicted spread-out view and ground truth spread-out view derived from NIRS-IVUS estimations.

## 2.4 Results and evaluations

### 2.4.1 Plaque localization

The first and second row in Fig.2.5 show the gradient-weighted class activation mapping (Grad-CAM)[20] generated with the 2.5D Dense U-Net predictions and the corresponding pie-shaped label for two cross-sections located in a LCx (left circumflex artery) and a LAD (left anterior descending artery) respectively. Fig.2.6 shows the bounding boxes generated with the 2.5D Mask R-CNN predictions.

The 2.5D Mask R-CNN uses bounding boxes to localize the plaque regions as shown in Fig.2.6. The blue boxes represent calcified plaques and the yellow boxes represent lipid-rich plaques. We show two different views in Fig.2.6, a polar view and the Cartesian view. For the polar view, the x-axis shows the angle coordinate from 0 degree to 359 degrees and y-axis shows the radius from 0 mm to 5.625 mm. The plaque detection is performed on the polar CT images and the predicted rectangular bounding boxes are mapped on the Cartesian view as pie-shaped bounding boxes for a more comprehensive visualization. The x-coordinate of left and right side of the

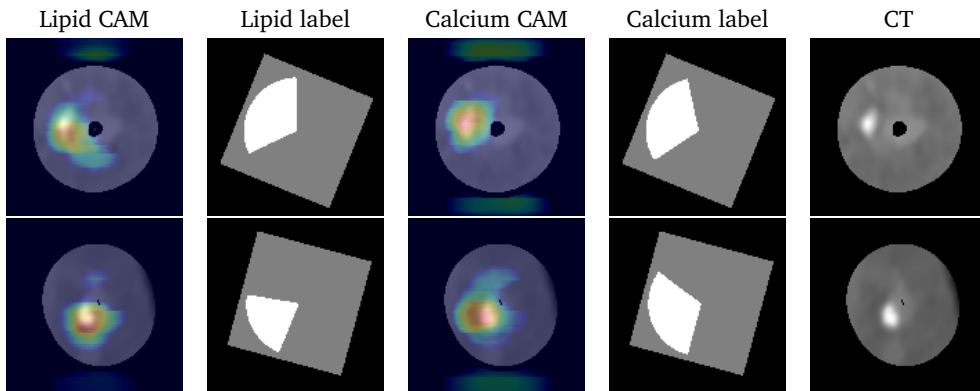


Figure 2.5: Cross-sectional plaque localization predicted by 2.5D Dense U-Net. From left to right there is the Grad-CAM view of the lipid-rich plaque, pie-shaped label of the lipid-rich plaque, Grad-CAM view of the calcified plaque, and pie-shaped label of calcified plaque as well as the CT image from the LCx (1st row) and from the LAD (2nd row).

rectangular bounding box directly represent the start and end angle, hence there is no need to perform a post-processing step as for the 2.5D Dense U-Net.

In the following section, we show the spread-out views based on the predicted cross-sectional plaque distributions and use these spread-out views to calculate metrics for the evaluation of the trained model.

#### 2.4.2 Tissue-type distribution

Spread-out views of four arteries are shown in Fig.2.7 and Fig.2.8. In the 2.5D Dense U-Net experiment, prediction thresholds for both lipid-rich and calcified plaques were 0.6. This means that only predictions with a probability higher than 0.6 are used in the final results. In the 2.5D Mask R-CNN experiment, the threshold of the lipid-rich plaques is 0.3 and that of calcified plaque is 0.5 which means that the predicted bounding boxes with a higher probability are used in the results during the evaluation. These thresholds are empirically determined and a trade-off between spread-out view based FN and FP. If the threshold is high, the predicted results with moderate probability will be rejected. The FN will increase and the FP will decrease. On the other hand, a low threshold allows more predicted results with low confidence which may underestimate predictions.

In Fig.2.7 and 2.8, the two DL-based methods and the conventional method are compared with the NIRS-IVUS estimations treated as a ground truth in the third row. The areas with lipid-rich plaques are scattered on the spread-out views from the Mask R-CNN. The lipid-rich areas between neighboring CTA images on the spread-out views

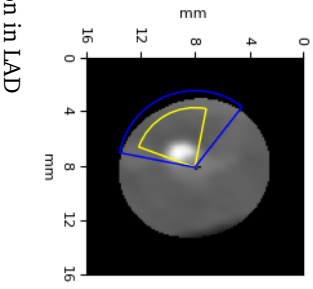
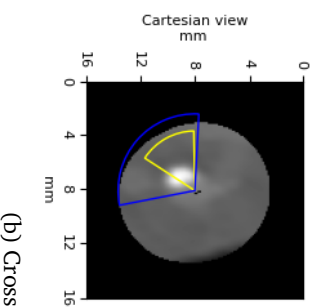
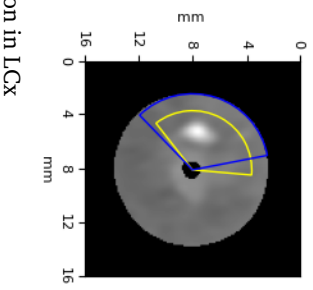
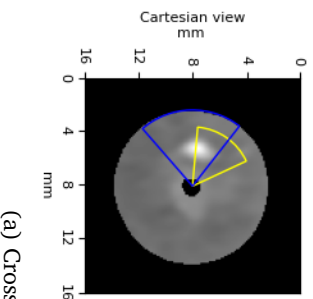
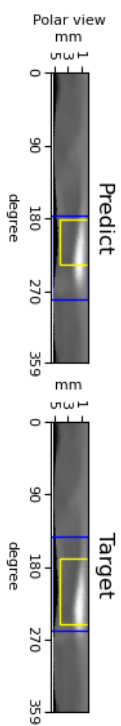
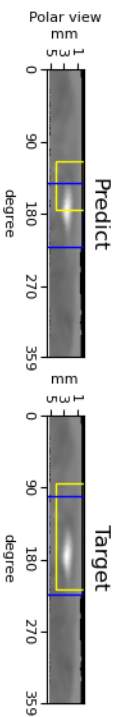


Figure 2.6: Cross-sectional plaque localization predicted by the 2.5D Mask R-CNN. (a) Predicted and target bounding boxes indicating lipid-rich (yellow) and calcified (blue) plaque position on a cross-section in the LGCx; (b) Predicted and target bounding boxes indicating lipid-rich (yellow) and calcified (blue) plaque position on a cross-section in the LAD. The difference in radius between lipid-rich and calcified plaques is only for the purpose of visualization.

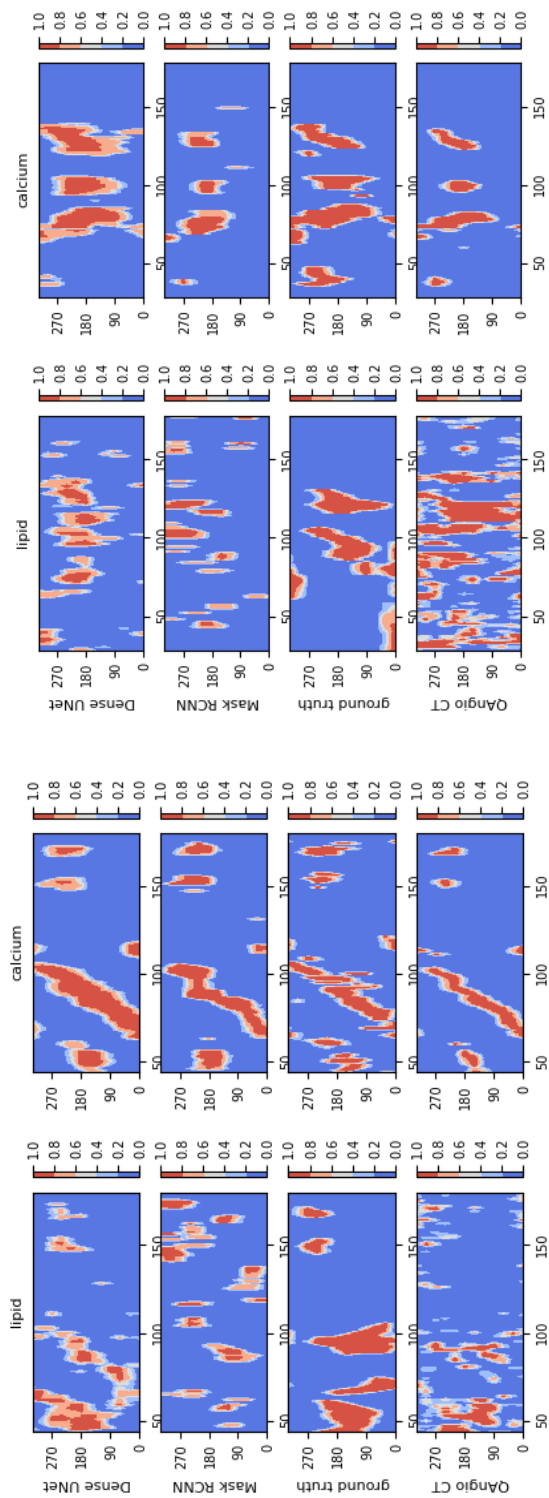


Figure 2.7: Spread-out view of lipid-rich and calcified plaques of two arteries in the test dataset. Predicted by Dense U-Net and Mask R-CNN, respectively. X-axis: indices of CT image. Y-axis: plaque occurring angle from  $0^\circ$  to  $359^\circ$ . Color bar: probability of plaque appearance from 0 to 1. (left:(a), right:(b))

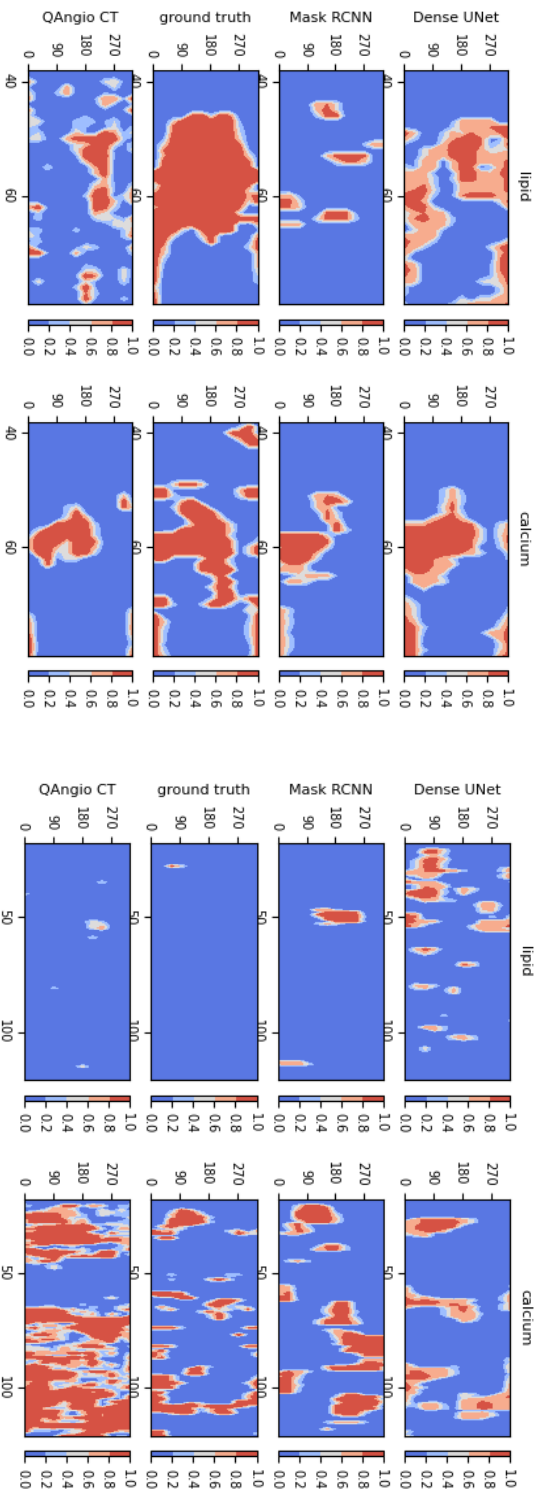


Figure 2.8: Spread-out view of lipid-rich and calcified plaques of two arteries in the test dataset. Predicted by NIRS-IVUS and QAngioCT respectively. X-axis: indices of CT image. Y-axis: plaque occurring angle from  $0^\circ$  to  $359^\circ$ . Color bar: probability of plaque appearance from 0 to 1. (left: (a), right: (b))

predicted by the Dense U-Net are more continuous than both the Mask R-CNN and the conventional methods.

The lumen intensity of the vessel shown in Fig.2.8(b) is higher than the others. Therefore, the threshold (350 HU) used to distinguish calcified plaques from the lumen area in the conventional method is invalid and causes meaningless predictions at the bottom-right spread-out view in Fig.2.8(b). The areas with calcified plaques are visually similar among the predicted spread-out views, ground truth and the conventional method, except the high lumen intensity case shown in Fig.2.8(b).

### 2.4.3 Prediction evaluation

To evaluate the similarity between predicted spread-out views and ground truth spread-out views numerically, the *accuracy* and *F1-score* were calculated on the spread-out views of the calcified and lipid-rich plaques. The results are shown in Fig.2.9. The blue, yellow, green and red violin plots represent the different methods used for plaque prediction, QAngioCT, 2.5D Dense U-Net, 2.5D Mask R-CNN and the classical 2.5D U-Net.

The median value of the *Accuracy* of the *calcified* plaques in the box plot of Fig.2.9(a) is similar among different methods. The distribution of values in the Dense U-Net and Mask R-CNN is more aggregated than the conventional method and classical 2.5D U-Net.

The Dense U-Net shows the highest median value in the *F1-score* distribution of *calcified* plaques shown in Fig.2.9(c). The median of the conventional method and Mask R-CNN is comparable to each other but higher than classical 2.5D U-Net.

Moreover, the Dense U-Net performs better in both median and density plots when comparing the *Accuracy* metric distribution of *lipid* plaques as shown in Fig.2.9(b).

The median value of the *F1-score* is comparable between the Dense U-Net and Mask R-CNN estimations for the *lipid* plaques and both are around 20% in Fig.2.9(d). The Mask R-CNN has a more aggregated distribution.

The numerical evaluation, median and mean value of *Accuracy* and *F1-score*, are shown in Table 2.1 and Table 2.2 for the *lipid-rich* and *calcified* plaques respectively, in 91 lesions.

For the prediction of lipid-rich plaques in Table 2.1, the 2.5D Dense U-Net had the highest median (0.18) and mean (0.21) value for the *F1-score* among all studied methods. The performance of the 2.5D Mask R-CNN was lower with a median and mean value of 0.17 and 0.18 respectively. The *F1-score* of both methods proposed in this paper exceeded the score of the conventional method and classical 2.5 U-Net, with their median and mean values being around 6 and 3 times higher than those of the conventional method.

In Table 2.2, the difference between the mean *F1-score* of the 2.5D Dense U-Net

and the conventional method was only 0.01, but the median values of the *F1-score* for the 2.5D Dense U-Net was 0.03 higher than that of the conventional method. The results of the 2.5D Mask R-CNN were lower than the 2.5D Dense U-Net but comparable with the conventional method. The median of the *F1-score* is still 0.14 higher than the classical 2.5D U-Net. Compared to the ground truth, the characterization of calcified plaque of the conventional method is accurate except for the cases with high intensity lumen (HU>350) shown in Fig.2.8(b).

Table 2.1: Median and average of spread-out view based Accuracy and F1-score of lipid-rich plaque predictions in 91 lesions. (\* represents the proposed pipeline, † represents the conventional method, the metrics of our proposed pipeline are in bold type.)

Methods	Accuracy median(mean)	F1-Score median(mean)
2.5D Dense U-Net (with grad-weight)*	<b>0.88 (0.86)</b>	<b>0.18 (0.21)</b>
2.5D Mask R-CNN (angle regression)*	<b>0.82 (0.82)</b>	<b>0.17 (0.18)</b>
2.5D U-Net[13] (with grad-weight)	0.90 (0.88)	0.02 (0.11)
QAngioCT†[18]	0.91 (0.86)	0.03 (0.08)

Table 2.2: Median and average of spread-out view based Accuracy and F1-score of calcified plaque predictions in 91 lesions. (\* represents the proposed pipeline, † represents the conventional method, the metrics of our proposed pipeline are in bold type.)

Methods	Accuracy median(mean)	F1-Score median(mean)
2.5D Dense U-Net (with grad-weight)*	<b>0.88 (0.87)</b>	<b>0.44 (0.38)</b>
2.5D Mask R-CNN (angle regression)*	<b>0.83 (0.83)</b>	<b>0.41 (0.37)</b>
2.5D U-Net[13] (with grad-weight)	0.86 (0.84)	0.27 (0.24)
QAngioCT†[18]	0.88 (0.85)	0.41 (0.37)

## 2.5 Discussion and conclusion

Two different methods for plaque angle prediction on cross-sectional CT images are proposed. We first propose to use NIRS-IVUS plaque labels to train the network to predict the cross-sectional distribution of lipid and calcified plaques in the coronary CTA arteries. It is an angle-level multi-target classification instead of an image-level binary classification in this study.

The 2.5D Mask R-CNN pipeline has fewer steps than the 2.5D Dense U-Net, but the metrics of the predicted results shown in Table 2.1 and Table 2.2 indicate that Dense U-Net has a slight advantage for the prediction of both lipid-rich plaques and calcified

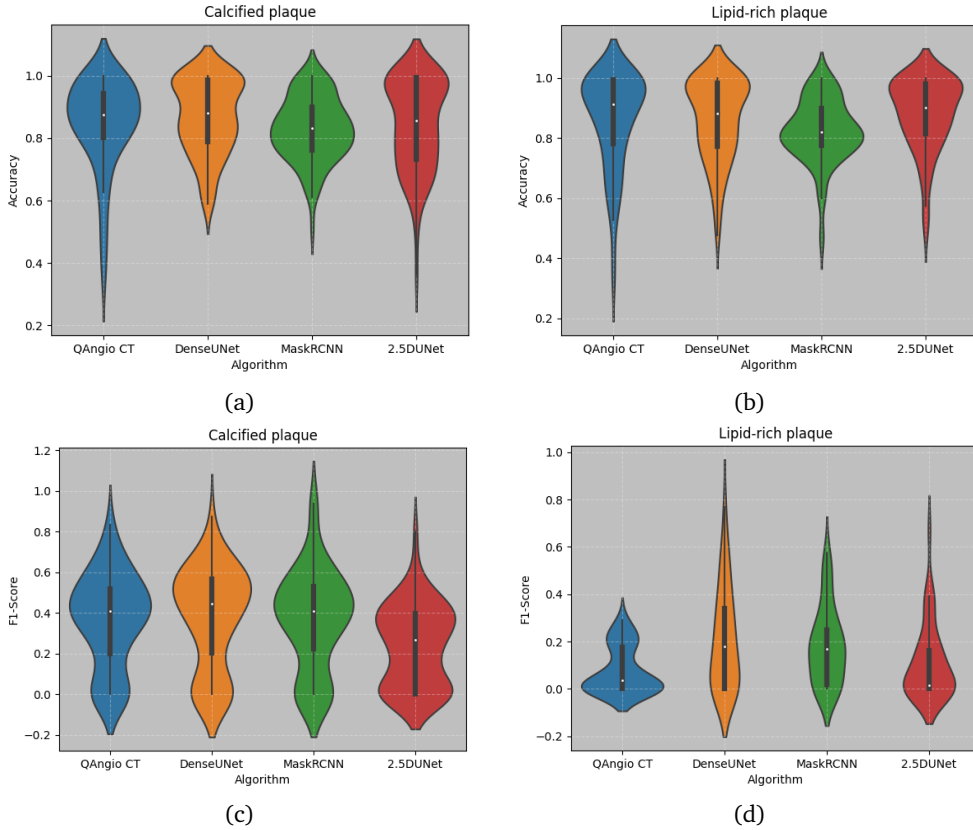


Figure 2.9: Distribution of the spread-out view based accuracy and F1-scores for 91 lesions. Blue, yellow, green and red violin plots represent the distribution of the conventional method, 2.5D Dense U-Net, 2.5D Mask R-CNN and classical 2.5D U-Net.

plaques.

Except for the high intensity lumen case shown in Fig.2.8(b), the conventional method can generate confident calcium spread-out views based on fixed HU thresholds (<350 HU for calcified plaques). Small calcified regions cannot be detected if the intensity of the lumen is higher.

The calcified areas with a high probability are clustered on the calcium spots in Fig.2.5. The Dense U-Net based method tends to over-estimate the angle range of the calcium plaques compared to the Mask R-CNN as shown in Fig.2.7(a).

The comparison between the different methods indicates that the proposed methods have capabilities to improve the detection performance of lipid-rich plaques. The median values of the *F1-score* of the proposed methods are 6 times higher than that of a conventional method which needs the EEM detection.

The high confidence areas of lipid-rich plaques are distributed around calcium spots. The difference between the Grad-CAM of the lipid-rich plaques and calcified plaques demonstrates that although the predicted plaque regions overlap, the focal areas of the network are different for different kinds of plaques.

The results in this paper suggest that it is promising to distinguish different kinds of plaques on CT images based on deep learning networks. However, it is uncertain whether all the lipid-rich areas can be detected on the cross-sectional CTA images using DL methods. The overlap between the lipid-rich tissue and normal tissue makes it difficult for DL models to identify the lipid-rich features on CTA images. The predictions of calcified plaques by the 2.5D Dense U-Net resemble the ground truth, which indicates that the registrations are accurate to guide the network to extract useful features. Moreover, comparing the violin plots that belong to the different methods, it is clear that the Dense U-Net and Mask R-CNN based predictions of calcified plaques are similar or better than a conventional method.

The main limitation in this paper is that the dataset used for training is relative small, only 176 arteries from 64 patients are involved in this study. Another limitation is that some areas with lipid-rich tissue are hard to distinguish from surrounding tissues and normal vessel wall. Furthermore, the resolution of CT images is relatively low compared to NIRS-IVUS and blooming artifacts caused by calcified plaques make it hard to distinguish calcified plaques from the contrast in the lumen. The next generation CT scanners such as photon-counting CT may overcome these limitations by its higher resolution and stronger artifact suppression.

In conclusion, this study presents two methods to predict lipid-rich and calcified plaques on the cross-sectional CT images using DL approaches. To the best of our knowledge, this is the first study to explore the mapping relationship between lipid-rich plaques on CT images and that on NIRS-IVUS images. Though the plaque detection is a challenging topic based on the currently used CT technology, the two methods proposed in this paper contribute to finer predictions of lipid-rich and calcified plaques. Especially for the lipid-rich plaques, the metrics calculated on the spread-out views are improved compared to the conventional method.

## References

- [1] M.-J. Bertrand, P. Lavoie-L' Allier, and J.-C. Tardif. "Near-infrared spectroscopy (NIRS): a novel tool for intravascular coronary imaging". In: *Developments in Near-Infrared Spectroscopy* (2017), page 25.
- [2] F. Saremi and S. Achenbach. "Coronary plaque characterization using CT". In: *American Journal of Roentgenology* 204.3 (2015), W249–W260.
- [3] A. Gudigar, S. Nayak, J. Samanth, et al. "Recent trends in artificial intelligence-assisted coronary atherosclerotic plaque characterization". In: *International Journal of Environmental Research and Public Health* 18.19 (2021), page 10003.
- [4] A. V. Finn, M. Nakano, J. Narula, et al. "Concept of vulnerable/unstable plaque". In: *Arteriosclerosis, Thrombosis, and Vascular Biology* 30.7 (2010), pages 1282–1292.
- [5] S. Sathyanarayana, S. Carlier, W. Li, and L. Thomas. "Characterisation of atherosclerotic plaque by spectral similarity of radiofrequency intravascular ultrasound signals". In: *EuroIntervention* 5.1 (2009), pages 133–139.
- [6] G. J. Tearney, H. Yabushita, S. L. Houser, et al. "Quantification of macrophage content in atherosclerotic plaques by optical coherence tomography". In: *Circulation* 107.1 (2003), pages 113–119.
- [7] A.-S. Schuurman, M. Vroegindewey, I. Kardys, et al. "Near-infrared spectroscopy-derived lipid core burden index predicts adverse cardiovascular outcome in patients with coronary artery disease during long-term follow-up". In: *European Heart Journal* 39.4 (2018), pages 295–302.
- [8] H. Liu, A. Wingert, J. Wang, et al. "Extraction of coronary atherosclerotic plaques from computed tomography imaging: a review of recent methods". In: *Frontiers in Cardiovascular Medicine* 8 (2021), page 597568.
- [9] M. Daghem, R. Bing, Z. A. Fayad, and M. R. Dweck. "Noninvasive imaging to assess atherosclerotic plaque composition and disease activity: coronary and carotid applications". In: *JACC Cardiovascular Imaging* 13.4 (2020), pages 1055–1068.
- [10] B. Szilveszter, C. Celeng, and P. Maurovich-Horvat. "plaque assessment by coronary CT". In: *International Journal of Cardiovascular Imaging* 32.1 (2016), pages 161–72.
- [11] M. Zreik, R. W. van Hamersvelt, J. M. Wolterink, et al. "A recurrent cnn for automatic detection and classification of coronary artery plaque and stenosis in coronary CT angiography". In: *IEEE Transactions on Medical Imaging* 38.7 (2019), pages 1588–1598.
- [12] J. Liu, C. Jin, J. Feng, et al. "A vessel-focused 3D convolutional network for automatic segmentation and classification of coronary artery plaques in cardiac CTA". In: *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges*. Volume 11395. 2019, pages 131–141.
- [13] M. H. Vu, G. Grimbergen, T. Nyholm, and T. Löfstedt. "Evaluation of multislice inputs to convolutional neural networks for medical image segmentation". In: *Medical Physics* 47.12 (2020), pages 6216–6231.

- [14] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. “Densely connected convolutional networks”. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pages 2261–2269.
- [15] K. He, G. Gkioxari, P. Dollár, and R. Girshick. “Mask R-CNN”. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017, pages 2980–2988.
- [16] M. A. De Graaf, A. Broersen, P. H. Kitslaar, et al. “Automatic quantification and characterization of coronary atherosclerosis with computed tomography coronary angiography: cross-correlation with intravascular ultrasound virtual histology”. In: *International Journal of Cardiovascular Imaging* 29.5 (2013), pages 1177–1190.
- [17] A. O’Brien, A. LaCombe, A. Stickland, and R. D. Madder. “Intracoronary near-infrared spectroscopy: an overview of the technology, histologic validation, and clinical applications”. In: *Global Cardiology Science & Practice* 2016.2 (2016), e201618.
- [18] H.-B. Park, B. K. Lee, S. Shin, et al. “Clinical feasibility of 3D automated coronary atherosclerotic plaque quantification algorithm on coronary computed tomography angiography: comparison with intravascular ultrasound”. In: *European Radiology* 25.10 (2015), pages 3073–83.
- [19] K. He, X. Zhang, S. Ren, and J. Sun. “Deep residual learning for image recognition”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pages 770–778.
- [20] R. R. Selvaraju, M. Cogswell, A. Das, et al. “Grad-CAM: visual explanations from deep networks via gradient-based localization”. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017, pages 618–626.

# 3

## Continuous and complete liver vessel segmentation with graph-attention guided diffusion

*This chapter was adapted from:*

**Zhang, X., Broersen, A., van Erp, G., Pintea, S.L. and Dijkstra, J., Continuous and complete liver vessel segmentation with graph-attention guided diffusion. (2025) Knowledge-Based Systems, 331, Article 114686.**



## Abstract

Improving connectivity and completeness are the most challenging aspects of liver vessel segmentation, especially for small vessels. These challenges require both learning the continuous vessel geometry and focusing on small vessel detection. However, current methods do not explicitly address these two aspects and cannot generalize well when constrained by inconsistent annotations. Here, we take advantage of the generalization of the diffusion model and explicitly integrate connectivity and completeness in our diffusion-based segmentation model. Specifically, we use a graph-attention module that adds knowledge about vessel geometry. Additionally, we perform the graph-attention at multiple-scales, thus focusing on small liver vessels. Our method outperforms eight state-of-the-art medical segmentation methods on two public datasets: *3D-ircadb-01* and *LiVS*.

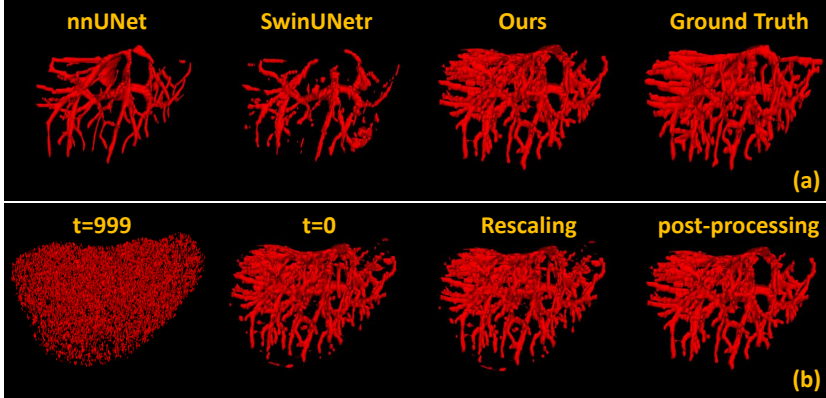


Figure 3.1: (a) Vessel trees of *nnUNet* [1], *Swin UNETR* [2], our proposed model and the ground truth. (b) Vessel trees predicted by our method across different steps ( $t=0$  is the final diffusion iteration). Rescaling is the process of recovering the resolution of resized CT images to that of the original CT images.

### 3.1 Introduction

Liver cancer is the fourth leading cause of death according to statistics on cancer-related mortality [3]. Furthermore, the liver is a frequent site for metastasis of various primary tumors, such as gastrointestinal tumors, breast cancer, lung cancer, and melanoma [4]. Both primary and secondary liver cancer have multiple treatment options, including surgery, and various interventional oncology liver treatments. The preoperative planning of these treatments can be improved with accurate segmentation of the liver vessels [5, 6]. In preoperative planning of liver tumor resection [6], visualizing the spatial location between liver vessels and tumors in a 3D view is essential to reduce surgical risk. Liver vessel segmentation is used primarily to ensure that the main vessels are not located near the planned resection region, thus reducing bleeding. Liver vessels also indicate the boundaries for the Couinaud classification [6]. Furthermore, it can assist in targeting the correct tumor nutrient supply vessel to decrease the recurrence rate in embolic therapies [7]. Hence, accurate liver vessel segmentation is essential in liver tumor surgery. However, acquiring automatic liver vessel segmentation is challenging due to the complex anatomy.

Automatic liver vessel segmentation is performed on computed tomography (CT) images. Traditionally, methods relied on image filtering [8, 9], active contour models [10, 11], or tracking methods [12, 13]. Currently, the leading methods on liver vessel segmentation are based on deep artificial networks [14, 15, 1, 2]. *nnUNet* [1] can extract features automatically from the CT images. However, *nnUNet* cannot ensure vessel continuity, as shown in Fig. 3.1(a). Attention-based methods

[16, 17] as used in *Swin UNETR* [2] improve vessel continuity. Yet, *Swin UNETR* struggles with horizontally distributed vessel, as seen in Fig. 3.1(a). In addition, the performance of the above Convolutional Neural Network (CNN) based methods have the generalization problem, especially when the label annotation styles across data are different [18]. Thus, ensuring vessel continuity in all directions, localizing small vessels, and improving model generalization for the inconsistent annotation style remains challenging.

Here, we address these challenges by proposing a model that makes continuous, and complete predictions, and generalizes well to different annotation styles. The inherent anatomical structure of the vascular tree inspires our network design. We are motivated by the assumption that introducing an explicit vascular graph can help the model better capture irregular and long-range vessel connectivity, thus adding continuity. Additionally, we use multiple scales in our graph structure, thus enabling the detection of small vessels, adding completeness to the model. Moreover, prior work tends to overlook the underlying data distribution, resulting in a strong dependence on annotation quality. To achieve accurate segmentation despite imperfect annotations, we rely on a diffusion model to learn the underlying data distribution, and mitigate the dependence on annotation quality.

Concretely, our model starts from a 2D diffusion model [19, 20]. We opt for a 2D rather than a 3D diffusion model to reduce computational requirements. To ensure vessel continuity, we add graph-attention layers [21] into the diffusion model. Because the graph is sparse, we compensate for this by integrating neighboring features on the graph in a local ensemble module [22]. The local ensemble module ensures a smooth transition between different nodes [22]. Secondly, to segment small vessels, we extract features at multiple scales in the nodes of the graph. The effectiveness of these components is shown in Fig. 3.1(b).

Our contributions are: (i) explicitly incorporating vessel continuity by adding graph-attention conditioning to a diffusion model for liver vessel segmentation; (ii) explicitly focusing on small vessels by relying on multi-scale graph-features when conditioning the diffusion model; (iii) continuous and complete vessel segmentations on two public datasets *3D-ircadb-01* [23] and *LiVS* [24], compared to existing work.

## 3.2 Related Work

### 3.2.1 Liver vessel segmentation

Liver vessel segmentation currently relies on *CNN* and attention methods. *CNN* methods either rely on *FCN* (fully convolutional networks) [15, 25], or follow the *UNet* architecture [14, 26]. Attention methods can be grouped into self-attention [16, 27] and graph-attention [28, 17] methods. *UNet* [29] and its variants are effective for medical segmentation. However, they struggle with imbalanced data [30],

such as between the vascular and the liver region. Moreover, their performance is limited by the receptive-field size of the convolutions [31]. While the data imbalance can be mitigated by improving the *Dice* loss[14], learning long-range dependencies remains a challenge [2]. To address this, *Swin-transformer* [32] uses a transformer architecture, and thus its receptive-field size covers the full input resolution. Similarly, graph-attention networks [33] capture long-range dependencies by computing node attention-coefficients. Augmenting the *UNet* by *Swin-transformer* [16], or using the graph-attention to assist the *UNet* training[17] is also effective in practice. In addition to methods specifically developed for liver vessel segmentation, several works on retinal vessel segmentation [34], skin lesion segmentation [35], and general medical image segmentation [36, 37] also incorporate self-attention mechanisms or U-Net variants to improve segmentation accuracy. Here, we also combine the capabilities of *CNNs* and attention mechanisms. Moreover, we jointly enforce continuity of the vessel-tree segmentations and focus on small vessels.

### 3.2.2 Diffusion models for medical image segmentation

Diffusion [19] methods showcase promising results for medical image segmentation. These diffusion methods are either non-dynamic conditioning [38] or dynamic conditioning [39, 40]. The non-dynamic models concatenate medical images to the input, and do not adapt this conditioning information over time. Their performance [38, 41] is comparable to (or lower than) *nnUNet* [1], which is the standard medical segmentation baseline. On the other hand, dynamic conditioning methods [40] use an extra encoder to generate time-dependent conditioning information. Similarly to the non-dynamic conditioning models, their accuracy is limited. More recently, *HiDiff* [42] and *MedSegDiff* [39] using a hybrid constrained method, obtains state-of-the-art results. *MedSegDiff* predicts a weighted combination of a diffusion segmentation and an auxiliary segmentation from the conditioning branch. *HiDiff* [42] also relies on a prior segmentation as one of the conditioning inputs to guide the diffusion model. This combination weakens the contribution of the diffusion model. Here, we build on a 2D dynamic conditioning diffusion model constrained by a hybrid loss, yet we only predict the diffusion segmentation.

### 3.2.3 Graph-based methods for medical image segmentation

Graphs are one of the most intuitive way to represent complex anatomical structures. Graph-based methods can segment medical tree structures [43, 44], but also non-structural medical data [45, 46]. Although, the graphs add long-range dependencies in *CNNs*, they tend to miss small branches [47]. This is due to the sparsity of the nodes, causing loss of information. Here, we also rely on graphs to add connectivity for vessel segmentation. To avoid missing small vessels, we use the local ensemble module of *LIIF* [22] which smoothes the feature between nodes. Additionally, we use

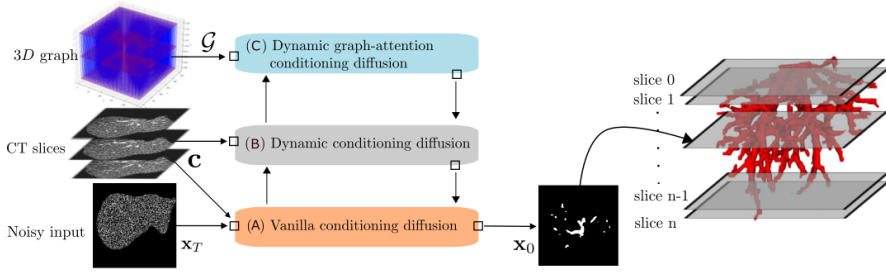


Figure 3.2: **Overview of our model:** (A) A vanilla diffusion model starting from noisy images  $\mathbf{x}_T$ , and predicting segmentation masks  $\mathbf{x}_0$  (in orange); (B) A dynamic conditioning model, conditioned on three CT slices  $\mathbf{c}$  (in gray); and (C) A multiscale graph-attention conditioning model, starting from a graph structure  $\mathcal{G}$  (in blue).

multiscale graph features to focus on small vessels.

### 3.3 Methods

#### 3.4 Diffusion conditioning models

Our model contains three components, as show in Fig. 3.2: (A) The vanilla conditioning diffusion model over three CT slices, for conditioning; (B) Dynamic conditioning diffusion, starting from the same CT slices but using a separate encoder; and (C) Conditioning diffusion model with multiscale graph-attention guidance.

##### 3.4.1 Vanilla conditioning diffusion model

**Diffusion model.** Conditioning diffusion models extend the Denoising Diffusion Probabilistic Models (DDPM) [19]. DDPM is composed of a forward process and a reverse process.

The forward process gradually adds Gaussian noise to the inputs  $\mathbf{x}_0$  over a number of  $T$  timesteps. In our case  $\mathbf{x}_0$  is the ground truth vessel segmentation mask. The variance of the Gaussian noise is modeled by  $\beta_t$ , which is typically a linear function of  $t$ . Thus, the distribution of the noisy vessel mask  $\mathbf{x}_t$  given the ground truth  $\mathbf{x}_0$ , is:

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1 - \bar{\alpha}_t)I), \quad (3.1)$$

where  $\bar{\alpha}_t = \prod_{s=1}^t (1 - \beta_s)$ . In the forward process, we obtain the noisy vessel mask  $\mathbf{x}_t$  from  $\mathbf{x}_0$  as a linear combination with the noise  $\epsilon_t \sim \mathcal{N}(0, I)$ :

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon_t. \quad (3.2)$$

In the reverse process, we train a model  $p_\theta$  with parameters  $\theta$ , to iteratively denoise an input noisy image  $\mathbf{x}_T$ . This aims to recover the clean segmentation mask  $\mathbf{x}_0$ . The

model  $p_\theta$  follows a Gaussian distribution:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t)). \quad (3.3)$$

The mean  $\mu_\theta(\mathbf{x}_t, t)$  and variance  $\Sigma_\theta(\mathbf{x}_t, t)$  of the reverse process are functions of a noise model  $\epsilon_\theta(\mathbf{x}_t, t)$ . The noise  $\epsilon_\theta$  is typically modelled by a *UNet* [29, 14]. This *UNet* noise model  $\epsilon_\theta$  is trained by minimizing the difference between the estimated noise  $\epsilon_\theta$  and the true noise  $\epsilon_t$  at a number of sampled timesteps  $t \sim [1, T]$ :

$$L_{\text{den}}(\mathbf{x}_0, \theta) = \mathbb{E}_{t \sim [1, T], \mathbf{x}_0, \epsilon_t} \|\epsilon_t - \epsilon_\theta(\mathbf{x}_t, t)\|^2. \quad (3.4)$$

**Conditioning diffusion models.** Following [38], we add a conditioning to the DDPM model that is independent of the timestep  $t$ . We use three consecutive CT slices,  $\mathbf{c}$ , as conditioning for our DDPM model. We evaluate the vanilla conditioning model in the experiments.

### 3.4.2 Dynamic Conditioning model

Vanilla conditioning cannot adapt the condition across the noise levels (time steps). To address this, at every diffusion timestep  $t$ , we embed the CT slices,  $\mathbf{c}$ , by using the *GenericUNet* encoder from *nnUNet* [1], giving rise to  $\mathbf{f}_c^t$ . To obtain time-dependent conditioning,  $\mathbf{f}_c^t$ , at each timestep  $t$  into the bottleneck of the *vanilla conditioning diffusion model*, as shown in Fig. 3.3(B) (solid downward arrow). Additionally, we use group convolutions in the conditioning, thus keeping the features per slice separate. To let the CT embedding  $\mathbf{f}_c^t$  adapt over time, we merge the noisy features of the *vanilla conditioning model* into the *dynamic conditioning model*, at the appropriate depth (in Fig. 3.3(B) with dashed arrows).

### 3.4.3 Multiscale graph-attention conditioning model

The *vanilla conditioning* and the *dynamic conditioning* both use the denoising loss in Eq. (3.4). To make use of the geometric structure of vessels, we first map the 3D vessel tree into a graph  $\mathcal{G}$ . Subsequently, we use graph-attention [21] to add this geometric structure as a condition into the diffusion model.

**Vessel graph construction.** We construct a 3D vessel graph  $\mathcal{G}=(\mathbf{V}, \mathbf{E})$ , where the nodes  $\mathbf{V}$  are locations along the vessel, and the edges  $\mathbf{E}$  indicate vascular connectivity. We start from the full volume  $[D \times H \times W]$  of a ground truth vessel tree, and we split it into non-overlapping sub-volumes  $[d \times h \times w]$ . Each node  $\mathbf{v} \in \mathbf{V}$  corresponds to a sub-volume and is the average of the voxel coordinates along the vessel region. If there is no vascular annotation in a sub-volume, we use the central voxel as the node. The graph edges,  $\mathbf{e} \in \mathbf{E}$ , are the geodesic distances between nodes, as in the Vessel Graph Network of Shin et al. [44]. Nodes with a small distance, but belonging to different vessel

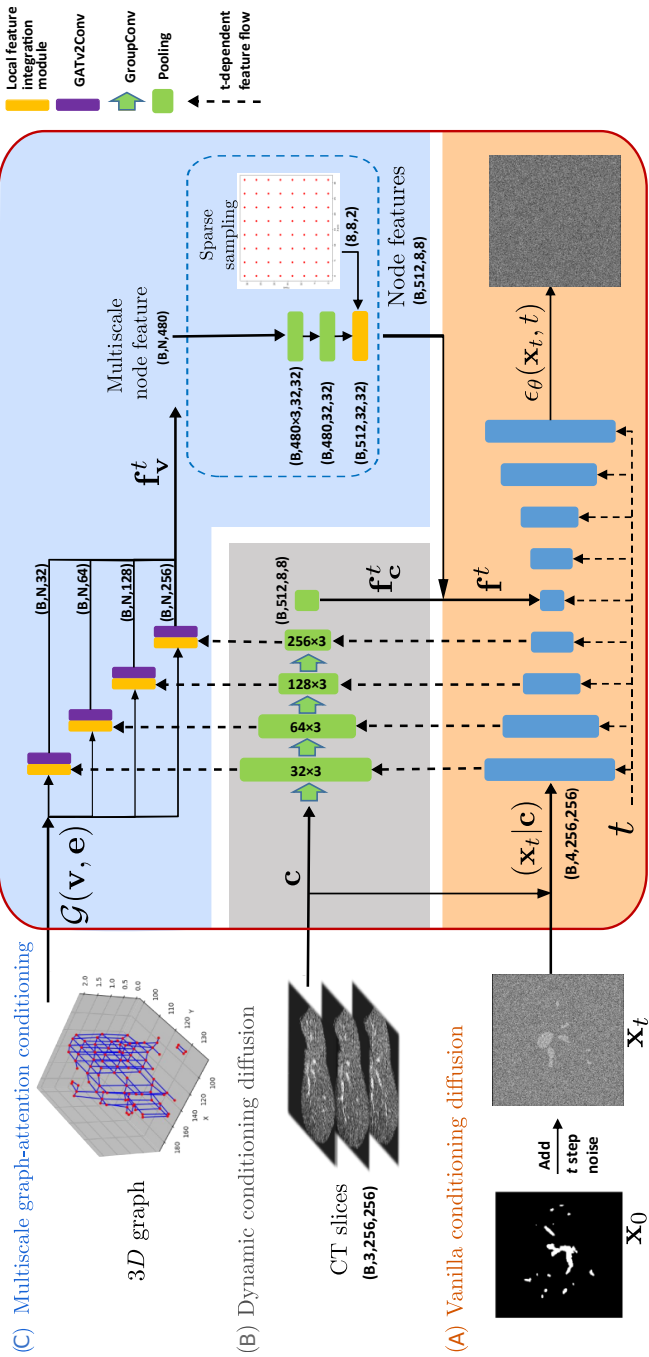


Figure 3.3: **Our network architecture:** (A) vanilla conditioning diffusion (orange); (B) dynamic conditioning diffusion (gray); (C) multiscale graph-attention conditioning (blue). These components interact through the vertical dashed/solid arrows. The dashed upwards arrows adapt the conditioning over time. The solid downwards arrows add the conditioning features: CT slices  $c$ , CT slice embeddings  $f^t_c$  and graph embeddings  $f^t_v$ .

branches, should not be connected. For this, we use the binary vessel label as a speed function to calculate the travel time from one node to another, as in Li et al. [17].

**Graph training and inference.** During training, we construct the graph  $\mathcal{G}$  using the ground truth vessel masks, as in Fig. 3.3(C). We only display the foreground nodes, corresponding to the location of the vessels, and leave the background nodes transparent. During inference, we do not have access to the ground truth vessel masks. Therefore, we input to the *multiscale graph-attention conditioning model* a fully-connected graph, as in Fig. 3.2(C). This is a viable choice, because during training, the graph helps adapt the weights of the component (C). Specifically, via the graph-attention, it extracts informative features from component (B). At inference, these attention-weights are trained, and can be applied on the new input features coming from component (B). In the ablation studies we show the effectiveness of using a fully-connected graph at inference.

**Multiscale graph-attention.** We use the CT-slice embeddings,  $\mathbf{f}_c^t$ , from the *dynamic conditioning model* to extract node features at each timestep  $t$  (dashed arrows in Fig. 3.3(C)). We process the node features via a graph-attention layer GATv2 [21] and a local feature integration module (LIIF) [22], to obtain node attention-coefficients. These node attention-coefficients are concatenated over the different scales (network depths), coming from the *dynamic conditioning model*, giving rise to multiscale node attention-coefficients:  $\mathbf{f}_v^t$  at each node  $\mathbf{v}$  and timestep  $t$ .

The nodes in the vessel graph  $\mathcal{G}$  are sparse (only  $32 \times 32$  nodes for one CT slice). To compensate for this sparsity, we use the local features integration module LIIF [22], which is popular for image super-resolution. We extend LIIF from 2D to 3D, and apply it on our CT-embeddings  $\mathbf{f}_c^t$ . Specifically, for a CT-slice embedding  $\mathbf{f}_c^{t,i}$  at timestep  $t$  and location  $i$  corresponding to a graph node  $\mathbf{v}_i$ , we use the graph neighboring locations  $\mathbf{v}_{ne(i)}$  to define a new embedding:

$$\hat{\mathbf{f}}_c^{(t,i)} = \sum_{ne(i)} \frac{\mathcal{S}(\mathbf{v}_i, \mathbf{v}_{ne(i)})}{\overline{\mathcal{S}}} \text{LFI}\left(\mathbf{f}_c^{(t,ne(i))}, \mathbf{v}_i - \mathbf{v}_{ne(i)}\right), \quad (3.5)$$

where our neighboring locations  $ne(i)$  vary across  $x, y$ , and  $z$  directions, rather than just  $x, y$ ; and  $\mathcal{S}(\cdot, \cdot)$  computes the area a 3D cube between the graph node  $\mathbf{v}_i$  and its neighbor  $\mathbf{v}_{ne(i)}$ ;  $\overline{\mathcal{S}} = \sum_{ne(i)} \mathcal{S}(\mathbf{v}_i, \mathbf{v}_{ne(i)})$ ;  $\mathbf{f}_c^{(t,ne(i))} = \text{GridSample}(\mathbf{f}_c^t, ne(i))$  [48]; and the local feature integration module is  $\text{LFI}\left(\mathbf{f}_c^{(t,ne(i))}, \mathbf{v}_i - \mathbf{v}_{ne(i)}\right) = \text{Conv2d}\left(\text{Cat}\left(\mathbf{f}_c^{(t,ne(i))}, \mathbf{v}_i - \mathbf{v}_{ne(i)}\right)\right)$ . Inside the function  $\text{LFI}(\cdot, \cdot)$  [22], we extract the neighboring features of each node based on their 3D coordinates  $\mathbf{v}_{ne(i)}$ , and embed the relative position differences  $(\mathbf{v}_i - \mathbf{v}_{ne(i)})$ , which represent the spatial relationships among neighborhood features.

Subsequently, we use new CT-slice embeddings  $\hat{\mathbf{f}}_c^{(t,i)}$  to extract node features via a

graph-attention layer GATv2 [21], for every two neighboring locations  $i, j$ :

$$\mathbf{f}_v^{(t,i,j)} = \frac{\exp \left[ \mathbf{a}^\top \text{LeakyReLU} \left( \mathbf{W}(\hat{\mathbf{f}}_c^{(t,i)} + \hat{\mathbf{f}}_c^{(t,j)}) \right) \right]}{\sum_{j'} \exp \left[ \mathbf{a}^\top \text{LeakyReLU} \left( \mathbf{W}(\hat{\mathbf{f}}_c^{(t,i)} + \hat{\mathbf{f}}_c^{(t,j')}) \right) \right]}. \quad (3.6)$$

During training, we want to learn which CT embeddings correspond to the foreground vessels and which not. Therefore, we process the node features via a convolutional layer with sigmoid activation, to obtain  $\hat{\mathbf{f}}_v^t = \text{sigmoid}(\text{Conv}(\mathbf{f}_v^t))$ , and optimize the model parameters  $\theta$  using a binary cross-entropy loss:

$$L_{\text{graph}}(\mathcal{G}, \theta) = - \sum_{v \in \mathcal{G}} \sum_t \log(\hat{\mathbf{f}}_v^t). \quad (3.7)$$

### 3.4.4 Overall diffusion model conditioning

We use the multiscale graph-attention embeddings  $\mathbf{f}_v^t$  together with the CT embeddings,  $\mathbf{f}_c^t$ , to condition the reverse diffusion process. Therefore, Eq. (3.3) defining the reverse process, becomes:

$$p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t | \mathbf{f}^t, t), \Sigma_\theta(\mathbf{x}_t | \mathbf{f}^t, t)), \quad (3.8)$$

$$\text{where } \mathbf{f}^t = \mathbf{f}_c^t + \mathbf{f}_v^t. \quad (3.9)$$

### 3.4.5 Overall loss function

Our overall loss function, used to fit the model parameters  $\theta$ , is a combination of the denoising loss in Eq. (3.4) and graph loss in Eq. (3.7):

$$L_{\text{total}}(\mathbf{x}_0, \mathbf{c}, \mathcal{G}, \theta) = L_{\text{den}}(\mathbf{x}_0, \mathbf{c}, \mathcal{G}, \theta) + L_{\text{graph}}(\mathcal{G}, \theta), \quad (3.10)$$

where the denoising loss  $L_{\text{den}}$  relies on the reverse diffusion process in Eq. (3.9).

## 3.5 Experiments and results

### 3.5.1 Datasets

We use two public datasets: *3D-ircadb-01* [23] and *LiVS* [24], as detailed in Tab. 3.1. *3D-ircadb-01* contains 20 cases, while *LiVS* contains 532 cases. In the *3D-ircadb-01* dataset every slice is annotated, but some small vessels are not annotated [26]. In the *LiVS* dataset only a subset of randomly chosen slices are annotated.

CT scans with thick slices ( $\geq 2.5$  mm for *3D-ircadb-01* and  $\geq 5$  mm for *LiVS*) are rare, forming outliers. Thus, for *LiVS* we exclude these cases from training and test. For the relatively small *3D-ircadb-01* dataset, we only exclude the cases with thick slices from the testset. To effectively evaluate on the small *3D-ircadb-01* dataset, we asked a clinical expert to score the completeness of the annotated vessels. After thickness exclusion, cases {04,06,08,11,16} were marked as complete. Therefore, we use these for testing. Specifically, we perform leave-one-out cross-validation on these

five cases, where at every fold we train on 19 cases and test on 1 test sample, from the list above. We average the results over all folds. For *LiVS* we report average metrics over 3-fold cross-validation.

### 3.5.2 Data pre- and post-processing

For the *3D-ircadb-01* [23] dataset, we first crop the liver region and resize the cropped CT slices to  $256 \times 256$  px. The liver masks exclude the *vena cava*. We clip the intensity of the CT slices to  $[0, 400]$  HU (Hounsfield units). The CT slices are already cropped, resized, and clipped in the *LiVS* [24] dataset. Because not all *LiVS* slices are annotated, we use ITK-SNAP[49] to interpolate the annotations. In our method, we sample the central slice of the 2.5D block only from the set of CT slices with ground truth annotations.

During inference, we rescale the diffusion predictions back to the physical resolution of the original CT image of  $512 \times 512$  px. During post-processing, we remove disconnected noisy spots, with a volume less than 1% of the largest connected region, using connected region analysis [50]. When the vessel annotations are continuous in the longitudinal direction, we find post-processing [14] more effective to obtain the final vessels. For discontinuous annotations, ensemble inference with different seeds, is more effective.

### 3.5.3 Evaluation metrics

For all our experiments we report: Dice similarity coefficient (*DSC*), voxel-wise sensitivity (*Sen*), voxel-wise specificity (*Spe*) [51], centerline Dice (*clDice*) [52], and a custom connected region-wise connectivity (*Con*) following Gegúndez-Arias *et al.* [53]. The *Con* metric is the ratio of the total number of connected regions, in the predicted tree  $\mathcal{T}$  and the total number of connected regions in the ground truth vessel

Table 3.1: Dataset overview. In *3D-ircadb-01* [23] every CT slice is annotated, but some small vessels are missing with an inconsistent annotation style. *LiVS* [24] contains more CT volumes and has stable annotation style, but only a subset of the slices are annotated.

	<i>3D-ircadb-01</i>	<i>LiVS</i>
Available scans	20	532
Used scans	20	303
Exclusion	completeness score	#annotations (< 30 slices)
	CT thickness $\geq 2.5$ mm	CT thickness $\geq 2.5$ mm
Pixel spacing	0.57 mm – 0.87mm	0.51 mm – 0.98 mm
Slice thickness	1.00 mm – 4.00mm	0.62 mm – 5.00 mm
Continuous annotation	Yes	No
High-contrast tumors	No	Yes
Annotation consistency	Low	High

tree  $\mathcal{T}^*$ :

$$\text{Con}(\mathcal{T}, \mathcal{T}^*) = \frac{|\text{comp}(\mathcal{T})|}{|\text{comp}(\mathcal{T}^*)|} \geq 1, \quad (3.11)$$

where  $\text{comp}(\cdot)$  computes the connected components. We only consider connected regions with a volume greater than  $120 \text{ mm}^3$ , as in Huang *et al.* [14]. We exclude over-connected segmentations, where  $\text{Con} < 1.0$ .

### 3.5.4 Experimental setting

**Baseline models.** We compare our model with seven state-of-the-art medical segmentation methods, including three diffusion-based methods (*HiDiff*[42], *MedSegDiff*[39], *EnsemDiff*[38]), two self-attention methods *MERIT*[54] and *Swin UNETR* [2], one self-configuring method *nnUNet* [1] and one specific liver vessel segmentation method [24]. The *HiDiff*, *MERIT*, *MedSegDiff* and *EnsemDiff* use inputs of size  $[3 \times 256 \times 256]$ , while *Swin UNETR* and *nnUNet* are 3D methods starting from the cropped CT slices as input. *Swin UNETR*, *EnsemDiff* and *MedSegDiff* require ensembles. For *Swin UNETR*, we train 5 models and ensemble their segmented results. For *EnsemDiff* and *MedSegDiff*, we ensemble 5 diffusion results using different random seeds, but with a single training. Our method does not use ensembles when training with continuous annotations on *3D-ircadb-01*, but uses  $5 \times$  ensembles for discontinuous annotations on *LiVS*.

**Implementation details.** We perform all our experiments on 1 NVIDIA RTX A6000 GPU with 48 GB memory. We use the *AdamW* optimizer with an initial learning rate of  $1 \times 10^{-4}$  and a batch size of 10. We input 4 channels: namely, three CT-slices and one noisy ground truth at each time step  $t$  during training, and we use a random Gaussian noise channel during inference. Our model converges within 160k iterations. We train the *EnsemDiff* and *MedSegDiff* models following their official implementation for 60k[38] and 100k[39], respectively. We also train the *MERIT*[54] and *HiDiff*[42] using their official implementations and recommended configurations. On both the *3D-ircadb-01* and *LiVS* datasets, we use the standard DDPM sampling scheme [20] with 1000 denoising steps, during inference, for all the diffusion-based experiments. For our model, the CT block size is  $[3 \times 256 \times 256]$  and the vessel graph  $\mathcal{G}$  consists of nodes of size  $(N, 3)$  and edges of size  $(E, 2)$ . Where we set the number of graph nodes  $N$  is  $32 \times 32$  per CT slice.

### 3.5.5 Quantitative comparison

**Quantitative evaluation on *3D-ircadb-01*.** Tab. 3.2 provides the numerical evaluation of our proposed model, *HiDiff* [42], *MERIT* [54], *TransUNet* [55], *MedSegDiff* [39], *EnsemDiff* [38], *Swin UNETR* [2] and *nnUNet* [1]. We group the methods per network representation: 3D (using 3D convolutions) or 2.5D (using 2D convolutions

over 2.5D CT slices). The 3D non-diffusion methods such as *nnUNet* and *Swin UNETR* are comparable to or being outperformed by *MedSegDiff* and *HiDiff*, but are better than *EnsemDiff*. Intuitively, diffusion-based methods with dynamic conditioning like *MedSegDiff* and *HiDiff* perform better than the non-dynamic conditioning methods, like *EnsemDiff*. Although, *HiDiff*, *EnsemDiff* and *MedSegDiff* and our proposed model are built on diffusion models, our method still exceeds them in terms of *DSC* and *Sen* scores. Our model exceeds the best-performing baselines for individual metrics by 1.49% and 6.61% in *DSC* and *Sen* scores, as shown in Tab. 3.2. These improvements are due to the graph-attention conditioning, adding continuity and completeness to the vessel predictions. Moreover, our method has the lowest standard deviation for *DSC* and *Sen* compared to the other methods. This indicates that our model tends to make more stable predictions.

The *Spe* metric evaluates the degree of false positive predictions for a segmented liver vessel tree. *Swin UNETR*, *TransUNet*, *EnsemDiff*, *HiDiff* and *MedSegDiff* obtain comparable *Spe* scores. However, our method obtains a slightly lower score than the others. The lower *Spe* scores could be explained by the the missing small-vessel annotations. Interestingly, in Tab. 3.2, *MERIT* and *nnUNet* have the highest averaged *Spe* but with the lowest averaged *Sen*, which is a trade-off between segmentation accuracy and completeness.

Tab. 3.2 also reports vessel connectivity, in the *clDice* [52] and *Con* metrics. *Con* (in Eq. (3.11)) should ideally be as close as possible to 1. Additionally, we also report in the brackets the number of connected regions of the segmented vessel tree and the ground truth. Our method achieves the highest *clDice* score compared to the baselines, with the exception of *TransUNet*. The *clDice* scores of both our method and *TransUNet* are  $\approx 75\%$ , which indicates that our method can best fit the centerlines of the segmented liver vessel tree, following the ground truth. *nnUNet* achieves the third highest *clDice* score. However, the *clDice* calculation depends on the erosion centerline, which can introduce bias. *nnUNet* has the lowest *Sen* score, which indicates incomplete segmentations. The *Con* score of our method is the closest value to 1 compared to the other methods, which is indicative of the continuous predictions in our model. Interestingly, *MedSegDiff* which is also based on a diffusion model, obtains the worst connectivity score. This may be due to less precise auxiliary segmentations used in *MedSegDiff*.

**Quantitative evaluation on *LIVS*.** Fig. 3.5 (left) compares our method with *HiDiff* [42], *MERIT* [54], *MedSegDiff* [39], *TransUNet* [55] and *EnsemDiff* [38], *Swin UNETR* [2], *nnUNet* [1] and Gao et al. [24]. The ground truth of the liver vessels is discontinuous in the longitudinal direction, as shown in Fig. 3.6(j). On *LIVS* we cannot report *clDice* and *Con* scores, because of the discontinuous annotations. In

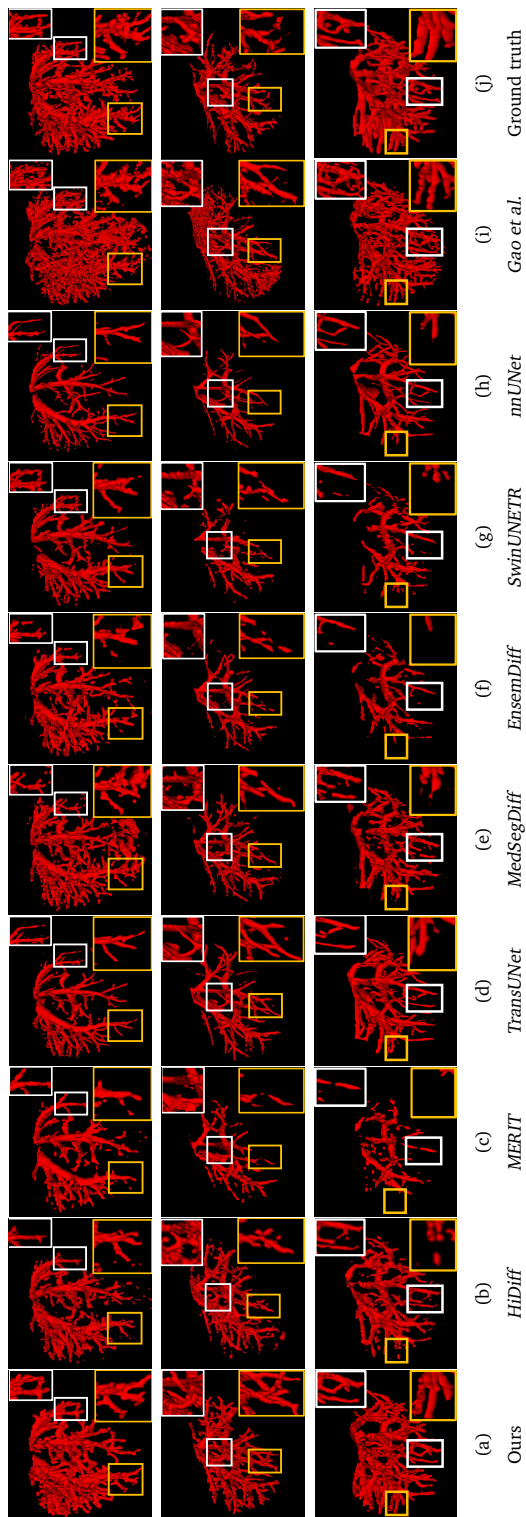


Figure 3.4: **Visualizations on the 3D-ircadb-01 [23] dataset.** (a) The liver vessel tree segmented by our proposed model; (b) – (i) are the segmentations from baselines: *HiDiff*[42], *MERIT*[54], *TransUNet*[55], *MedSegDiff*[39], *EnsemDiff*[38], *Swin UNETR* [2], *nnUNet* [1], and [24]; (j) The ground truth. Yellow/white boxes highlight segmentation completeness and continuity.

Table 3.2: **Results on the 3D-ircadb-01 [23] dataset.** We compare our model with: *HiDiff*[42], *MERIT*[54], *TransUNet*[55], *MedSegDiff*[39], *EnsemDiff*[38], *Swin UNETR* [2], *nnUNet* [1] and [24]. Our method is the best in terms of *DSC*, *clDice*, *Sen* and *Con* scores, but worse in *Spe* scores. Interestingly, *nnUNet* has the highest *Spe* score and the lowest *Sen* score, which may be due to a trade-off between detailed and accurate segmentation.

	Type	<i>DSC</i> (%)	<i>clDice</i> (%)	<i>Sen</i> (%)	<i>Spe</i> (%)	<i>Con</i> ( $\rightarrow$ 1)
Gao et al.[24]	2D	60.19 $\pm$ 4.69	56.48 $\pm$ 11.74	64.98 $\pm$ 4.48	99.80 $\pm$ 0.10	8.73 (96/11)
nnUNet[1]	3D	60.54 $\pm$ 6.86	71.60 $\pm$ 4.37	44.76 $\pm$ 8.09	<b>99.99</b> $\pm$ 0.00	2.36 (26/11)
Swin UNETR[2]	3D	55.61 $\pm$ 9.80	62.75 $\pm$ 7.91	42.30 $\pm$ 11.46	99.98 $\pm$ 0.01	5.45 (60/11)
TransUNet[55]	3D	69.77 $\pm$ 3.73	<b>75.83</b> $\pm$ 5.50	58.63 $\pm$ 11.70	99.97 $\pm$ 0.01	3.73 (41/11)
EnsemDiff[38]	2.5D	55.07 $\pm$ 9.70	60.99 $\pm$ 9.64	40.45 $\pm$ 10.60	99.98 $\pm$ 0.02	7.18 (79/11)
MedSegDiff[39]	2.5D	59.67 $\pm$ 7.74	66.02 $\pm$ 7.87	47.42 $\pm$ 10.43	99.95 $\pm$ 0.05	8.64 (95/11)
<i>MERIT</i> [54]	2.5D	49.50 $\pm$ 9.57	53.00 $\pm$ 8.79	34.30 $\pm$ 8.81	99.99 $\pm$ 0.01	5.50 (59/11)
<i>HiDiff</i> [42]	2.5D	63.32 $\pm$ 4.01	60.24 $\pm$ 2.71	54.45 $\pm$ 7.25	99.94 $\pm$ 0.01	6.70 (72/11)
Ours	2.5D	<b>71.26</b> $\pm$ 1.93	74.61 $\pm$ 1.21	<b>71.59</b> $\pm$ 4.07	99.89 $\pm$ 0.04	<b>1.09</b> (12/11)

Fig. 3.5, our method has the best *DSC* and *Sen* scores. Especially in terms of *Sen*, our method performs better than the baselines, indicating a more complete vessel segmentation. In terms of *DSC* scores, we are comparable to *nnUNet* because the *DSC* scores are susceptible to outliers, as shown in Fig. 3.5 (right). The methods of *HiDiff* and Gao *et al.* have comparable *Sen* scores to *nnUNet*, but its *DSC* score is lower, because of having the lowest *Spe* score of all methods. *SwinUNetr* has the lowest *DSC* scores and *MERIT* has the lowest *Sen* scores. As with the 3D-ircadb-01 dataset, our improvements are due to the graph-attention conditioning. Although *HiDiff* and *MedSegDiff* have lower scores than our model in terms of *DSC* and *Sen*, it outperforms the *EnsemDiff* model. This may be due to the dynamic conditioning method used in *HiDiff* and *MedSegDiff*. The standard deviations of our predictions and predictions of *HiDiff* and *MedSegDiff* are comparable, and they are both lower than the standard deviation of *EnsemDiff*. This could indicate that the dynamic conditioning diffusion model (ours, *HiDiff* and *MedSegDiff*) can provide more stable results than the vanilla conditioning model (*EnsemDiff*).

However, these diffusion-based methods are less efficient than deterministic methods during inference, as shown in the Fig. 3.5. Our model is slower than *MedSegDiff* and *EnsemDiff* when the inference batch size is set to 50 slices. This slowdown is due to the additional cost of graph attention and local feature integration (LFI) in our model. The graph attention introduces anatomic structure into the model, improving accuracy, but also increases inference time compared to other diffusion baselines. Similarly, LFI compensates for the sparse node representation, but the grid sample layer used in LFI results in heavier computation.

	Repr. type	DSC (%)	Sen (%)	Spe (%)	sec. /slice	steps
Gao et al.[24]	2D	73.14 ± 11.44	76.84 ± 11.57	99.63 ± 0.34	< 0.1	-
<i>nnUNet</i> [1]	3D	81.39 ± 5.04	76.06 ± 7.48	99.91 ± 0.05	< 0.1	-
<i>Swin UNETR</i> [2]	3D	65.54 ± 6.65	62.71 ± 10.31	99.76 ± 0.13	< 0.1	-
<i>TransUNet</i> [55]	3D	78.21 ± 6.17	75.11 ± 7.58	99.86 ± 0.08	< 0.1	-
<i>EnsemDiff</i> [38]	2.5D	70.00 ± 9.20	56.27 ± 10.81	<b>99.97 ± 0.03</b>	20.67	1000
<i>MedSegDiff</i> [39]	2.5D	76.85 ± 7.03	65.96 ± 9.43	99.96 ± 0.04	26.92	1000
<i>MERIT</i> [54]	2.5D	69.54 ± 6.72	56.63 ± 8.43	99.95 ± 0.04	< 0.1	-
<i>HiDiff</i> [42]	2.5D	70.53 ± 6.66	78.32 ± 5.49	99.63 ± 0.12	0.25	10
Ours	2.5D	<b>81.41 ± 6.64</b>	<b>81.35 ± 6.93</b>	99.84 ± 0.12	43.21	1000

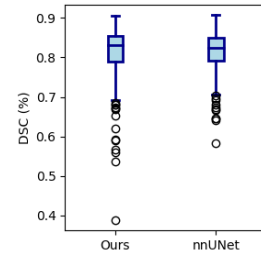


Figure 3.5: **Results on the LiVS [24] dataset.** *Left:* The performance of our method compared with *HiDiff* [42], *MERIT* [54], *TransUNet* [55], *MedSegDiff* [39], *EnsemDiff* [38], *Swin UNETR* [2], *nnUNet* [1] and Gao et al. [24]. *ciDice* and *Con* are not applicable for the *LiVS* dataset with discontinuous annotations. Our method outperforms others in terms of *Sen* scores. *Spe* of our model is slightly lower than *EnsemDiff*, which is a trade-off between completeness (*Sen*) and accuracy (*Spe*) of vessel segmentation. *Right:* Our *DSC* score is comparable with *nnUNet* because high-contrast tumors in the liver cause more outliers (shown in the box plot), negatively affecting the averaged *DSC* score. Deterministic methods are more efficient than generative methods in inference.

### 3.5.6 Qualitative comparison

**Qualitative evaluation on 3D-ircadb-01.** In Fig. 3.4 we provide a qualitative evaluation on three test cases from the *3D-ircadb-01* dataset. Compared to the other methods, the appearance of our predicted vessel-tree segmentation (first column) is the most similar to the ground truth (last column). The vascular structures marked by the yellow and white boxes in Fig. 3.4 are almost completely segmented by our method, while the other methods oversegment or miss parts of the vessel tree. Especially the segmentation results of *MERIT* [54], *Swin UNETR* [2] and *nnUNet* [1], are visibly sparse. These results relate to the low *DSC* and *Sen* scores of *MERIT*, *Swin UNETR* and *nnUNet* in Tab. 3.2. Although the vessel structures of *EnsemDiff* [38] (sixth column) and *MedSegDiff* [39] (fifth column) are denser than those of the non-diffusion methods, they provide more discontinuous vessel masks at the extremities of the tree (*i.e.* for smaller vessels).

In Fig. 3.4 we also highlight the connectivity of the distal liver vessel branches in the yellow boxes. The visualizations show that our segmentation is as continuous as the ground truth. This corresponds to a *Con* score closer to 1 in Tab. 3.2. Comparing the vessel branches of *nnUNet* with the other methods for the distal vessels, we see that these are precise, being exceeded only by our method. This is again consistent with the *Con* score of *nnUNet* in Tab. 3.2 – the second best score.

**Qualitative evaluation on LIVS.** In Fig. 3.6, we provide a qualitative comparison in 3D

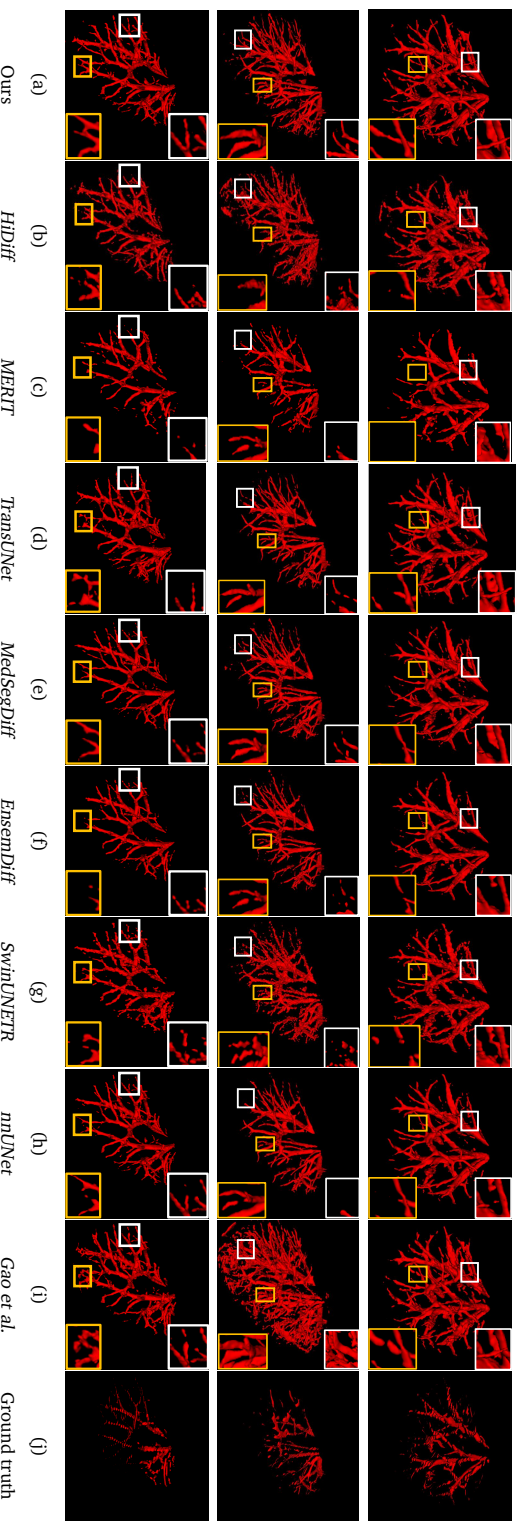
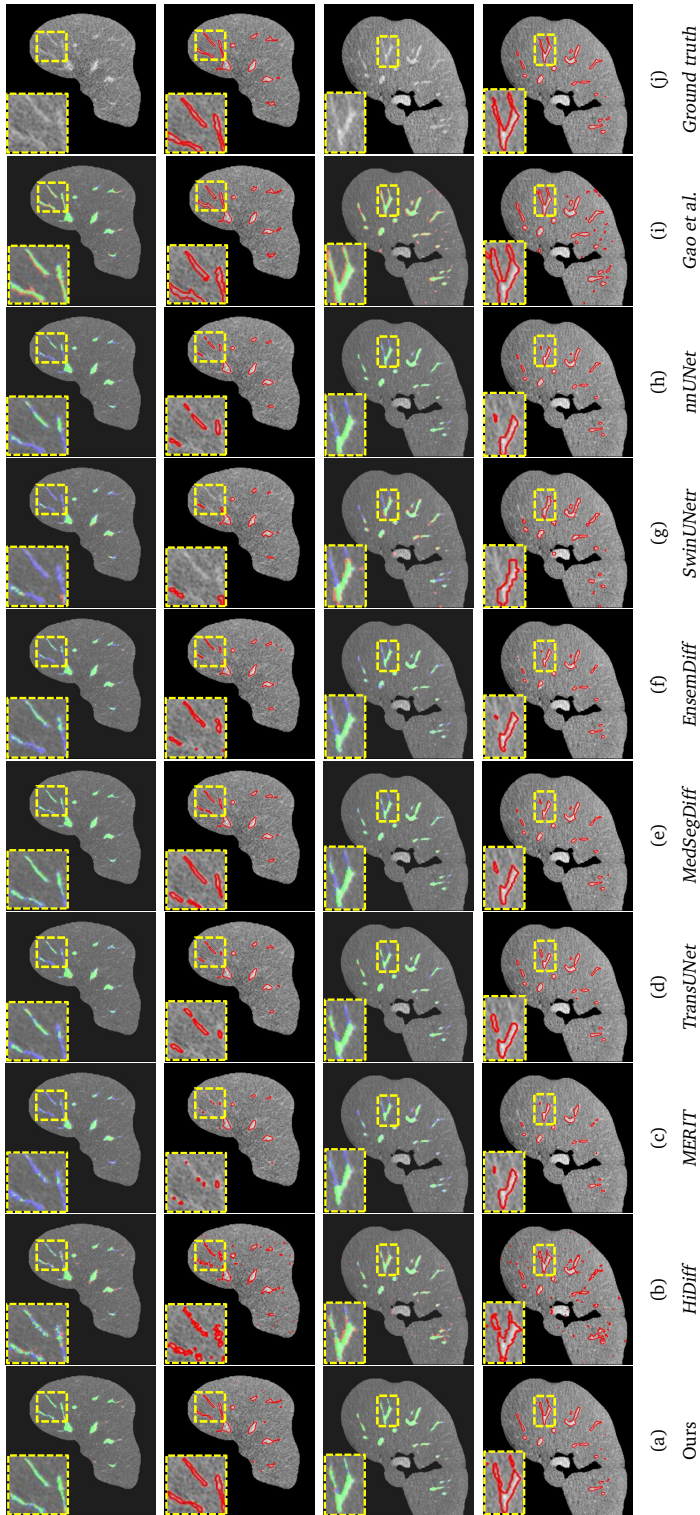


Figure 3.6: **Visualizations of the LiVS [24] dataset.** (a) The liver vessel tree segmented by our proposed model; (b)–(i) are the liver vessel tree segmented by the baselines: *HDiff* [42], *MERTT* [54], *TransUNet* [55], *MedSegDiff* [39], *EnsemDiff* [38], *SwinUNETR* [2], *mUNet* [1] and Gao et al. [24]; (j) The discontinuous (partially annotated) ground truth liver vessel tree. Yellow and white boxes show completeness and continuity between our model and baselines. Enlarged views shown in the corners.



(Part 1/2) Cross-sectional visualization examples on *LiVS* dataset (first two cases).

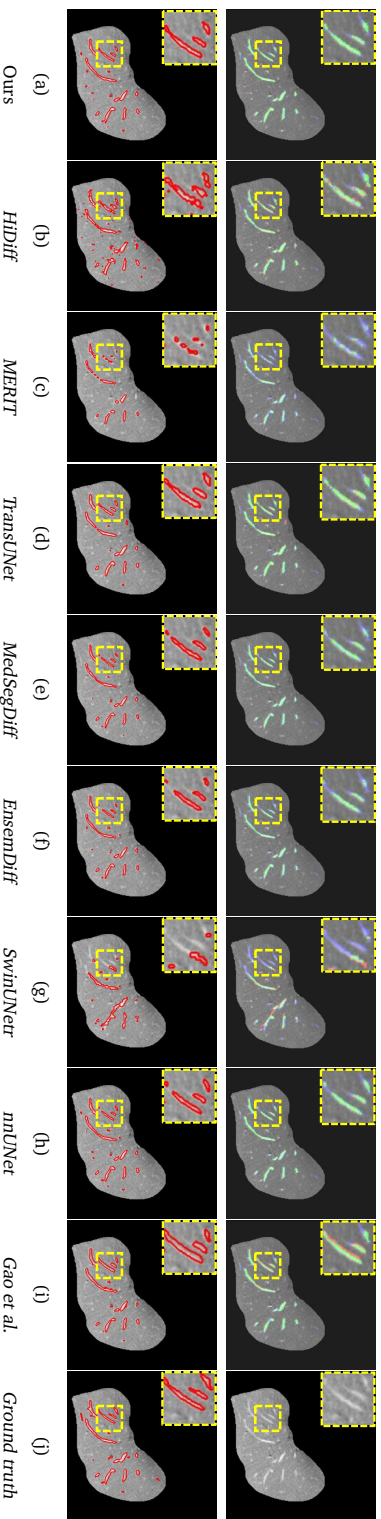


Figure 3.7: **(Part 2/2) Cross-sectional visualizations on LiVS [24] dataset.** We compare segmentation masks from our proposed model with: *HiDiff* [42], *MERTT* [54], *TransUNet* [55], *MedSegDiff* [39], *EnsemDiff* [38], *Swin UNETR* [2], *nnUNet* [1] and Gao et al. [24]. The first and second rows are respectively overlaid masks and contour comparisons. Green indicates true positives, red false positives, and blue false negatives. Differences are highlighted in yellow boxes. Our method better preserves fine vessel connectivity.

Table 3.3: **Ablation study on 3D-ircadb-01 dataset.** We ablate the effect of each individual component presented in Fig. 3.3 and the effect of the post-processing: (A) vanilla conditioning model [38]; (B) dynamic conditioning model; (C) multiscale graph-attention conditioning; and (D) post-processing. Interestingly, the (A) *vanilla model* and the (A) *vanilla model* combined with (B) *dynamic conditioning* perform best in terms of *Spe* scores. This is because predicting less structures (low *Sen* scores) entails fewer false positives. However, the overall combination performs best in *DSC* score.

(A) Vanilla conditioning	(B) Dynamic conditioning	(C) Dynamic multiscale graph-attention	(D) Post-processing	DSC(%)	Sen(%)	Spe(%)
✓	×	×	×	55.07 ± 9.70	40.45 ± 10.6	99.98 ± 0.02
✓	✓	×	×	59.00 ± 5.43	44.22 ± 6.19	<b>99.98 ± 0.01</b>
✓	✓	✓	×	62.69 ± 3.93	<b>79.80 ± 6.33</b>	99.70 ± 0.17
✓	✓	✓	✓	<b>71.26 ± 1.93</b>	71.59 ± 4.07	99.89 ± 0.04

on three test cases between our model and the eight baselines. Although it is hard to directly compare the segmentations of the different methods, due to the discontinuous ground truth (shown in Fig. 3.6(j)), inter-method comparison can still be informative. The yellow and white boxes in Fig. 3.6 show that most baselines predict fractured small vessels for the distal branch, while our segmented vessel structures are denser and more complete.

We compare vessel completeness in Fig. 3.7, where we show a qualitative comparison for the slices of the three cases in Fig. 3.6, in a 2D cross-sectional view. The first, third and fifth rows show the overlaid segmentation masks of the liver vessel mask of different methods, when compared to the ground truth mask. Green, red and blue colors represent true positive, false positive and false negative, respectively. The blue areas of the overlaid masks from columns (b) to (i) reflect the missing contours of the baselines. Our model has fewer blue regions, indicating that our segmentation is more complete for both the large and small vessel structures. The red contours in the second, forth and final rows of Fig. 3.7 outline the boundary of the predicted liver vessel segmentation of different methods. The contours in the last column (j) correspond to the boundary of ground truth vessel mask. We enlarged the regions using the yellow boxes, to highlight differences in the vessel completeness. Comparing the contours of the baselines from column (b) to (i) with the ground truth, tiny vessel blobs are not outlined. For the baseline (i), the contours reflect a heavy oversegmentation. Our model benefits from the long-range feature dependency encoded in the multi-scale graph-attention conditioning.

### 3.5.7 Model ablation study

**Effect of different model components and post-processing.** For completeness, we perform the ablation studies on both the *3D-ircadb-01* and *LIVS* datasets. Because the

Table 3.4: **Ablation study on *LiVS* dataset.** We ablate the effect of each individual component presented in Fig. 3.3 and the effect of the post-processing: (A) vanilla conditioning model [38]; (B) dynamic conditioning model; (C) multiscale graph-attention conditioning; and (D) post-processing. Inference ensembling is used due to the discontinuous annotations of the dataset. Post-processing is not mandatory for this kind of dataset.

(A) Vanilla conditioning	(B) Dynamic conditioning	(C) Dynamic multiscale graph-attention	(D) Post-processing	DSC(%)	Sen(%)	Spe(%)
✓	×	×	×	70.00 ± 9.20	56.27 ± 10.81	<b>99.97 ± 0.03</b>
✓	✓	×	×	73.39 ± 7.55	60.96 ± 9.51	99.96 ± 0.03
✓	✓	✓	×	<b>81.41 ± 6.64</b>	<b>81.35 ± 6.93</b>	99.84 ± 0.12
✓	✓	✓	✓	81.04 ± 6.50	78.02 ± 7.19	99.88 ± 0.11

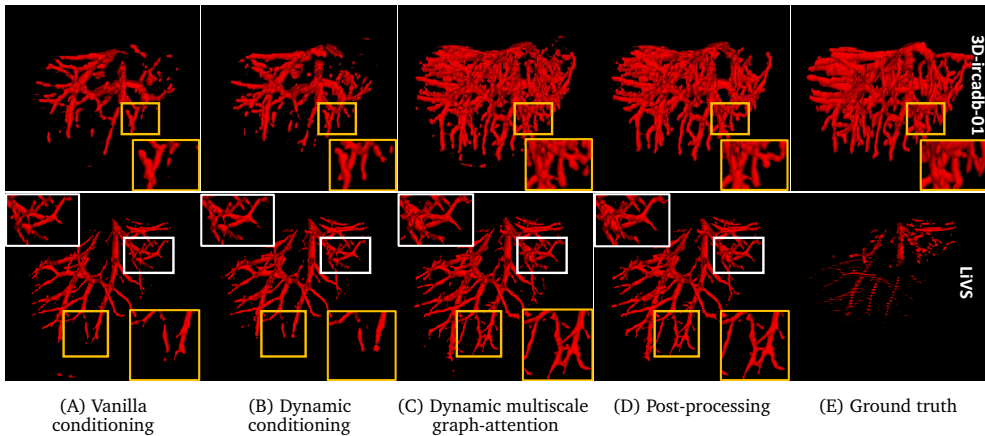


Figure 3.8: **Visualization of the per-component ablation study for the *3D-IRCadb-01* dataset in Tab. 3.3 and the *LiVS* dataset in Tab. 3.4.** Each column shows the cumulative effect of adding components (A) to (D), compared against the ground truth (E). (A) Vanilla conditioning; (B) Dynamic conditioning; (C) Dynamic multiscale graph-attention; (D) Post-processing; (E) Ground truth.

three components of our model : (i) vanilla conditioning, (ii) dynamic conditioning and (iii) multiscale graph-attention, are highly interconnected (see Fig. 3.3), we cannot evaluate them independently. To explore the influence of different conditioning levels on the performance of liver vessel segmentation, we perform an additive ablation study in Tab. 3.3 and Tab. 3.4, and visualize the ablation in Fig. 3.8. We start from the (A) *vanilla conditioning model* and subsequently add new conditioning components to it. Noteworthy, the (A) *vanilla model* (in the first row) and the combination of the vanilla model with the (B) *dynamic conditioning model* (in the second row) achieve the highest *Spe* scores. This is due to these models predicting fewer structures in the segmentation masks (as shown by the low *Sen* score and in Fig. 3.8), and therefore

Table 3.5: **Ablation study testing the effect of the number of graph nodes, on the 3D-IRCADB-01 and LiVS datasets.** No post-processing or inference ensembling was used in this ablation study on node number. As the number of graph nodes increase, the computations (GFLOPs) increase. Using  $32^2 \times 3$  provides a good tradeoff between accuracy and computations, while balancing *Sen* and *Spe*.

#nodes per batch	3D-ircadb-01			LiVS			GFLOPs per batch
	DSC(%)	Sen(%)	Spe(%)	DSC(%)	Sen(%)	Spe(%)	
$16^2 \times 3$	61.69 ± 4.03	76.99 ± 7.95	99.69 ± 0.22	66.13 ± 10.46	88.02 ± 4.32	99.29 ± 0.29	533.95
$32^2 \times 3$	62.69 ± 3.93	79.80 ± 6.33	99.70 ± 0.17	71.90 ± 8.85	87.07 ± 4.71	99.51 ± 0.20	565.20
$64^2 \times 3$	62.71 ± 4.88	54.80 ± 9.00	99.92 ± 0.04	75.57 ± 5.72	67.87 ± 8.05	99.90 ± 0.05	690.17

having fewer false positives. The final model combining all three components: (A) *vanilla conditioning*, (B) *dynamic conditioning* and (C) *multiscale graph-attention* has the highest *DSC*, and *Sen* scores. The graph-attention conditioning contributes considerably to the performance of our model. Post-processing is an effective way to remove noisy predictions, and to improve the *DSC* score as shown in Tab. 3.3, but it can negatively affect the *Sen*. However, for datasets with discontinuous annotations (such as *LiVS*), post-processing is not mandatory because inference ensembles can also remove noisy predictions. Overall, we conclude that all components have a beneficial effect on the vessel segmentation scores.

**The effect of different graph node numbers.** We perform ablation studies testing the effect of the number of graph nodes, on *3D-ircad-01* and *LiVS* datasets. To isolate the effect of the number of graph nodes, we do not apply post-processing or inference ensembling, in this ablation study. In addition to the node numbers of  $32 \times 32 \times 3$  per batch used in our implementation, we also report results for configurations of  $16 \times 16 \times 3$  and  $64 \times 64 \times 3$ , as shown in Tab. 3.5. Although our model with a configuration of  $64 \times 64 \times 3$  achieves the highest *DSC* and *Spe* scores on both datasets, it yields the lowest *Sen* score and incurs higher computational complexity in terms of GFLOPs. The lower *Sen* score of our model with the  $64 \times 64 \times 3$  configuration indicates relatively incomplete vessel segmentation, making it insufficient to meet the requirement for segmentation completeness. The sparser node configuration of  $16 \times 16 \times 3$  results in a lower *Spe* score, indicating a higher number of false positives in the segmentation. To balance *Sen* and *Spe* while considering computational complexity, the  $32 \times 32 \times 3$  configuration used in our implementation represents a good tradeoff.

**Graph-attention in inference.** In inference, we input a fully-connected graph in the component (C), as shown in Fig. 3.2. While the node coordinates are uniformly distributed in the graph, the edge weights between nodes are adapted by the trained graph-attention layer. Fig. 3.9 visualizes the learned edge weights for an input fully-connected graph. In Fig. 3.9, the cross-sectional view shows that the vessel area and

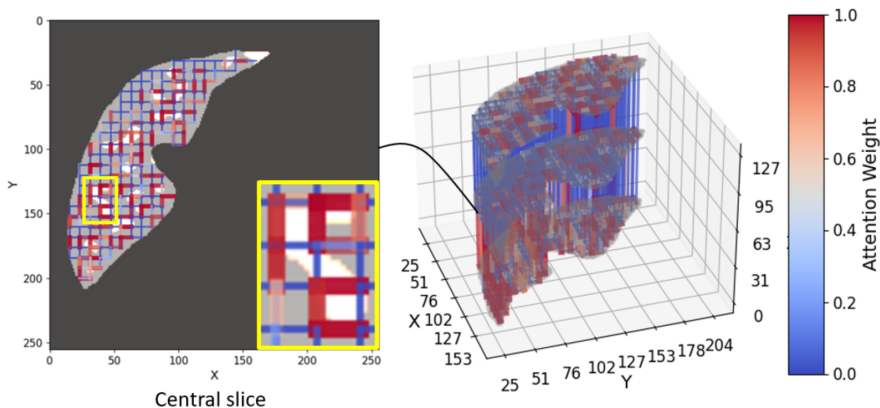


Figure 3.9: **Edge attention weights of a fully-connected graph in inference.** Blue/red represent low/high attention weights, respectively. We also show the box enlarged in the bottom right corner. The vessel area and its neighborhood attract more edge attention, thus demonstrating the utility of inputting a fully-connected graph in inference.

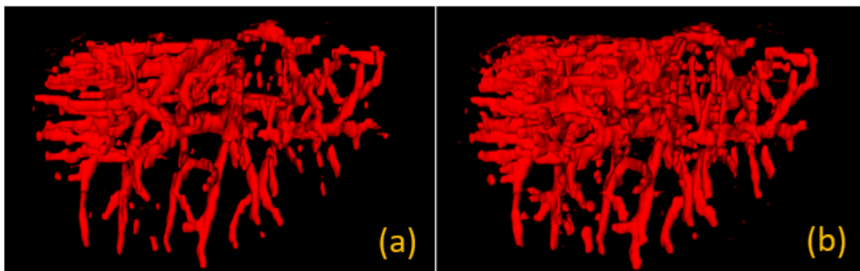


Figure 3.10: **Inference difference between using an empty graph, without edges (a) and a fully-connected graph (b) as the input.** A fully-connected graph leads to more dense and continuous predictions.

its neighborhood have higher-magnitude edge-attention. Additionally, in Fig. 3.10 we compare the difference in predictions when inputting an empty graph, without edges (a) and a fully-connected graph (b). Using an empty graph leads to sparser predictions. This analysis demonstrates that graph attention layers are still effective, even with a uniform graph as input.

## 3.6 Discussion

### 3.6.1 Limitation

Fig. 3.11 shows the limitation of our model and contains examples with the worst DSC scores for the *LiVS* and *3D-ircadb-01* datasets. On these worst cases, our

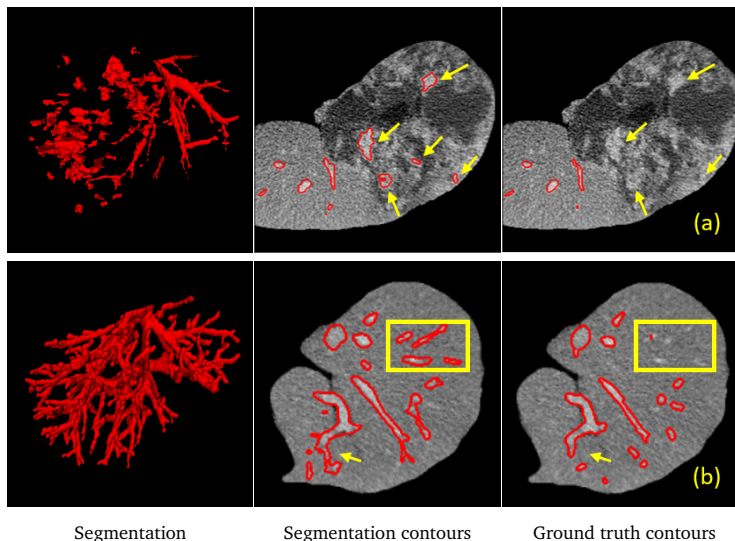


Figure 3.11: **Analysis of model limitations.** (a) First row: Visualization of a failure case on the *LiVS* dataset. (b) Second row: Visualization of a failure case on the *3D-ircadb-01* dataset. In case (a) the failure is due to the contrast marked by yellow arrows around the tumors being misclassified as vessel structures. In case (b) the failure is due to the unlabeled vessels [14, 26] (highlighted in the yellow box), causing low *DSC* scores. The model also makes a mistake by predicting segmentation masks where there should not be (over-segmentation), as shown by the yellow arrow. This is caused by the inconsistency in annotations in the *3D-ircadb-01* dataset [23].

proposed model inaccurately predicts vessel segmentations when there is a large tumor surrounded by contrast-rich regions, as seen in Fig. 3.11(a). This failure is reflected by the outliers of the *DSC* scores of the *LiVS* dataset in Fig. 3.5. All tumors in *3D-ircadb-01* are characterized as low-intensity regions, and are not surrounded by contrast, so this failure appears only on the *LiVS* dataset. Fig. 3.11(b) shows a worst performing example on the *3D-ircadb-01* dataset. The low *DSC* score is due to missing annotations [14, 26]. The yellow box highlights a case where vessels were not annotated. Our model correctly predicts this, while still being penalized in the *DSC* scores. Using yellow arrows we indicate regions where our model over-predicts structure that is not truly present. The inconsistency in the annotation quality in the *3D-ircadb-01* dataset[23] is an additional challenge in the training of the model. This may lead to that the model learns to focus on the wrong information when predicting segmentations.

### 3.7 Conclusion

In this study, we focus on liver vessel segmentation from CT volumes. To this end, we propose to augment conditional diffusion models with geometric graph-structure computed at multiple resolutions. The role of the graph structure is to ensure connectivity in the segmentation, across neighboring slices in the CT volume. Moreover, we use multi-scale features in the graph, thus allowing the model to focus on small vessels, that otherwise would be missed. Our proposed model achieves state-of-the-art results, in terms of *DSC*, *Sen* and vessel connectivity on two standard benchmarks: *3D-ircadb-01* [23] and *LiVS* [24], when compared to baselines such as *HiDiff* [42], *MERIT* [54], *TransUNet* [55], *MedSegDiff* [39], *EnsemDiff* [38], *Swin UNETR* [2], *nnUNet* [1] and Gao et al. [24].

## References

- [1] F. Isensee, P. F. Jaeger, S. A. Kohl, et al. “nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation”. In: *Nature methods* 18.2 (2021), pages 203–211.
- [2] A. Hatamizadeh, V. Nath, Y. Tang, et al. “Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images”. In: *International MICCAI Brainlesion Workshop*. Sept. 2021, pages 272–284.
- [3] X. Li, P. Ramadori, D. Pfister, et al. “The immunological and metabolic landscape in primary and metastatic liver cancer”. In: *Nature Reviews Cancer* 21.9 (2021), pages 541–557.
- [4] G. Disibio and S. W. French. “Metastatic patterns of cancers: results from a large autopsy study”. In: *Archives of pathology & laboratory medicine* 132.6 (2008), pages 931–939.
- [5] J. Llovet, R. Kelley, and A. Villanueva. “Hepatocellular carcinoma”. In: *Nat Rev Dis Primers* 7.6 (2021).
- [6] O. Alirr and A. Rahni. “Survey on liver tumour resection planning system: steps, techniques, and parameters”. In: *Journal of Digital Imaging* 33.2 (2020), pages 304–323.
- [7] H. Huang. “Influence of blood vessel on the thermal lesion formation during radiofrequency ablation for liver tumors”. In: *Medical physics* 40.7 (2013), page 073303.
- [8] S. Survarachakan, E. Pelanis, Z. Khan, et al. “Effects of enhancement on deep learning based hepatic vessel segmentation”. In: *Electronics* 10.10 (2021), page 1165.
- [9] J. Lamy, O. Merveille, B. Kerautret, and N. Passat. “A Benchmark Framework for Multi-region Analysis of Vesselness Filters”. In: *IEEE Transactions on Medical Imaging* 41.12 (2022), pages 3649–3662.
- [10] Y. Cheng, X. Hu, J. Wang, et al. “Accurate vessel segmentation with constrained b-snake”. In: *IEEE Transactions on Image Processing* 24.8 (2015), pages 2440–2455.
- [11] M. Chung, J. Lee, J. W. Chung, and Y.-G. Shin. “Accurate liver vessel segmentation via active contour model with dense vessel candidates”. In: *Computer methods and programs in biomedicine* 166 (2018), pages 61–75.
- [12] O. Friman, M. Hindennach, C. Kühnel, and H. Peitgen. “Multiple hypothesis template tracking of small 3D vessel structures”. In: *Medical image analysis* 14.2 (2010), pages 160–171.
- [13] S. Cetin and G. Unal. “A higher-order tensor vessel tractography for segmentation of vascular structures”. In: *IEEE Transactions on Medical Imaging* 34.10 (2015), pages 2172–2185.
- [14] Q. Huang, J. Sun, H. Ding, et al. “Robust liver vessel extraction using 3D U-Net with variant dice loss function”. In: *Computers in Biology and Medicine* 101 (2018), pages 153–162.

- [15] T. Kitrungsrotsakul, X. Han, Y. Iwamoto, et al. “VesselNet: A deep convolutional neural network with multi pathways for robust hepatic vessel segmentation”. In: *Computerized Medical Imaging and Graphics* 75 (2019), pages 74–83.
- [16] M. Wu, Y. Qian, X. Liao, et al. “Hepatic vessel segmentation based on 3D swin-transformer with inductive biased multi-head self-attention”. In: *BMC Medical Imaging* 23.1 (2023), pages 1–14.
- [17] R. Li, Y. Huang, H. Chen, et al. “3D graph-connectivity constrained network for hepatic vessel segmentation”. In: *IEEE Journal of Biomedical and Health Informatics* 26.3 (2021), pages 1251–1262.
- [18] B. Nichyporuk, J. Cardinell, J. Szeto, et al. “Rethinking generalization: The impact of annotation style on medical image segmentation”. In: *Journal of Machine Learning for Biomedical Imaging* (2022).
- [19] J. Ho, A. Jain, and P. Abbeel. “Denoising diffusion probabilistic models”. In: *Advances in neural information processing systems*. Volume 33. 2020, pages 6840–6851.
- [20] A. Nichol and P. Dhariwal. “Improved denoising diffusion probabilistic models”. In: *International Conference on Machine Learning*. PMLR, July 2021, pages 8162–8171.
- [21] S. Brody, U. Alon, and E. Yahav. “How Attentive are Graph Attention Networks?” In: *International Conference on Learning Representations*. 2022.
- [22] Y. Chen, S. Liu, and X. Wang. “Learning continuous image representation with local implicit image function”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, pages 8628–8638.
- [23] L. Soler, A. Hostettler, V. Agnus, et al. “3D image reconstruction for comparison of algorithm database: A patient specific anatomical and medical image database”. In: ().
- [24] Z. Gao, Q. Zong, Y. Wang, et al. “Laplacian salience-gated feature pyramid network for accurate liver vessel segmentation”. In: *IEEE Transactions on Medical Imaging* (May 2023).
- [25] B. Ibragimov, D. Toesca, D. Chang, et al. “Combining deep learning with anatomical analysis for segmentation of the portal vein for liver SBRT planning”. In: *Physics in Medicine & Biology* 62.23 (2017), page 8943.
- [26] Q. Yan, B. Wang, W. Zhang, et al. “Attention-guided deep neural network with multi-scale feature fusion for liver vessel segmentation”. In: *IEEE Journal of Biomedical and Health Informatics* 25.7 (2020), pages 2629–2642.
- [27] X. Wang, X. Zhang, G. Wang, et al. “TransFusionNet: Semantic and Spatial Features Fusion Framework for Liver Tumor and Vessel Segmentation Under JetsonTX2”. In: *IEEE Journal of Biomedical and Health Informatics* 27.3 (2022), pages 1173–1184.
- [28] D. Zhang, S. Liu, S. Chaganti, et al. “Graph attention network based pruning for reconstructing 3D liver vessel morphology from contrasted CT images”. In: *CoRR* (2020).

- [29] O. Ronneberger, P. Fischer, and T. Brox. “U-net: Convolutional networks for biomedical image segmentation”. In: *Medical Image Computing and Computer-Assisted Intervention*. Springer International Publishing, 2015, pages 234–241.
- [30] R. Zhao, B. Qian, X. Zhang, et al. “Rethinking dice loss for medical image segmentation”. In: *2020 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2020, pages 851–860.
- [31] Y. Liu, J. Yu, and Y. Han. “Understanding the effective receptive field in semantic image segmentation”. In: *Multimedia Tools and Applications* 77 (2018), pages 22159–22171.
- [32] Z. Liu, Y. Lin, Y. Cao, et al. “Swin transformer: Hierarchical vision transformer using shifted windows”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pages 10012–10022.
- [33] P. Velickovi, G. Cucurull, A. Casanova, et al. “Graph Attention Networks”. In: *International Conference on Learning Representations* (2018).
- [34] X. Zhang, Q. Zhu, T. Hu, et al. “Joint high-resolution feature learning and vessel-shape aware convolutions for efficient vessel segmentation”. In: *Computers in Biology and Medicine* 191 (2025), page 109982.
- [35] D. Dai, C. Dong, S. Xu, et al. “Ms RED: A novel multi-scale residual encoding and decoding network for skin lesion segmentation”. In: *Medical image analysis* 75 (2022), page 102293.
- [36] Q. Yan, S. Liu, S. Xu, et al. “3D medical image segmentation using parallel transformers”. In: *Pattern Recognition* 138 (2023), page 109432.
- [37] D. Dai, C. Dong, Q. Yan, et al. “I2u-net: A dual-path u-net with rich information interaction for medical image segmentation”. In: *Medical Image Analysis* 97 (2024), page 103241.
- [38] J. Wolleb, R. Sandkühler, F. Bieder, et al. “Diffusion models for implicit image segmentation ensembles”. In: *International Conference on Medical Imaging with Deep Learning*. PMLR, Dec. 2022, pages 1336–1348.
- [39] J. Wu, R. FU, H. Fang, et al. “MedSegDiff: Medical Image Segmentation with Diffusion Probabilistic Model”. In: *Medical Imaging with Deep Learning*. 2023.
- [40] Z. Xing, L. Wan, H. Fu, et al. “Diff-UNet: A Diffusion Embedded Network for Volumetric Segmentation”. In: *CoRR* (2023).
- [41] F. Bieder, J. Wolleb, A. Durrer, et al. “Memory-Efficient 3D Denoising Diffusion Models for Medical Image Processing”. In: *Medical Imaging with Deep Learning*. 2023.
- [42] T. Chen, C. Wang, Z. Chen, et al. “HiDiff: hybrid diffusion framework for medical image segmentation”. In: *IEEE Transactions on Medical Imaging* (2024).
- [43] R. Selvan, T. Kipf, M. Welling, et al. “Graph refinement based airway extraction using mean-field networks and graph neural networks”. In: *Medical image analysis* 64 (2020), page 101751.
- [44] S. Shin, S. Lee, I. Yun, and K. Lee. “Deep vessel segmentation by learning graphical connectivity”. In: *Medical Image Analysis* 58 (2019), page 101556.

- [45] Z. Tian, X. Li, Y. Zheng, et al. “Graph-convolutional-network-based interactive prostate segmentation in MR images”. In: *Medical physics* 47.9 (2020), pages 4164–4176.
- [46] R. Soberanis-Mukul, N. Navab, and S. Albarqouni. “Uncertainty-based graph convolutional networks for organ segmentation refinement”. In: *Medical Imaging with Deep Learning*. PMLR. 2020, pages 755–769.
- [47] L. Liu, J. Wolterink, C. Brune, and R. Veldhuis. “Anatomy-aided deep learning for medical image segmentation: a review”. In: *Physics in Medicine & Biology* 66.11 (2021), 11TR01.
- [48] M. Jaderberg, K. Simonyan, A. Zisserman, et al. “Spatial transformer networks”. In: *Advances in neural information processing systems* 28 (2015).
- [49] P. Yushkevich, J. Piven, H. Hazlett, et al. “User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability”. In: *Neuroimage* 31.3 (July 2006), pages 1116–1128.
- [50] J. Weaver. “Centrosymmetric (cross-symmetric) matrices, their basic properties, eigenvalues, and eigenvectors”. In: *The American Mathematical Monthly* 92.10 (1985), pages 711–717.
- [51] D. Müller, I. Soto-Rey, and F. Kramer. “Towards a guideline for evaluation metrics in medical image segmentation”. In: *BMC Research Notes* 15.1 (2022), page 210.
- [52] S. Shit, J. Paetzold, A. Sekuboyina, et al. “cDice-a novel topology-preserving loss function for tubular structure segmentation”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, pages 16560–16569.
- [53] M. Gegúndez-Arias, A. Aquino, J. Bravo, and D. Marín. “A function for quality evaluation of retinal vessel segmentations”. In: *IEEE Transactions on Medical Imaging* 31.2 (2011), pages 231–239.
- [54] M. M. Rahman and R. Marculescu. “Multi-scale hierarchical vision transformer with cascaded attention decoding for medical image segmentation”. In: *Medical Imaging with Deep Learning*. PMLR. 2024, pages 1526–1544.
- [55] J. Chen, J. Mei, X. Li, et al. “TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers”. In: *Medical Image Analysis* 97 (2024), page 103280.

# 4

## Top-k maximum intensity projection priors for 3D liver vessel segmentation

*This chapter was adapted from:*

**Zhang, X., Broersen, A., van Erp, G., Pinteá, S.L. and Dijkstra, J., Top-K Maximum Intensity Projection Priors for 3D Liver Vessel Segmentation.** (2025) In 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI), pp. 1-5. IEEE.



## Abstract

Liver-vessel segmentation is an essential task in the pre-operative planning of liver resection. State-of-the-art  $2D$  or  $3D$  convolution-based methods focus on liver vessel segmentation on  $2D$  CT cross-sectional views, which do not take into account the global liver-vessel topology. To maintain this global vessel topology, we rely on the underlying physics used in the CT reconstruction process, and apply this to liver-vessel segmentation. Concretely, we introduce the concept of *top-k maximum intensity projections*, which mimics the CT reconstruction by replacing the integral along each projection direction, with keeping the top-k maxima along each projection direction. We use these top-k maximum projections to condition a diffusion model and generate  $3D$  liver-vessel trees. We evaluate our  $3D$  liver-vessel segmentation on the *3D-ircadb-01* dataset, and achieve the highest *Dice* coefficient, intersection-over-union (*IoU*), and *Sensitivity* scores compared to prior work.

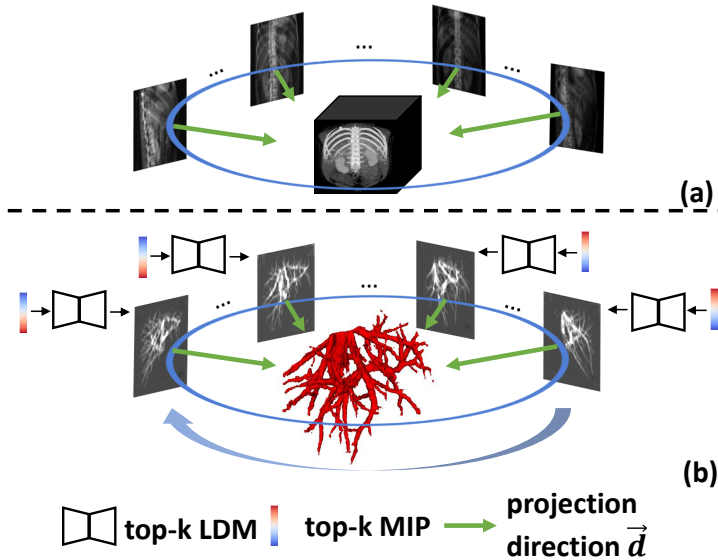


Figure 4.1: (a) **Standard CT reconstruction**: Given the integral projections and projections directions  $\vec{d}$ , it reconstructs the underlying 3D object by back-projecting the integral along each direction; (b) **Our proposed top-k MIP**: Given the projections directions  $\vec{d}$ , our model reconstructs the 3D liver-vessel tree by computing the top-k maximum value of the CT scan along each direction and inputting this into a latent diffusion model.

## 4.1 Introduction

Liver resection is the standard surgical treatment option for primary and secondary liver cancer [1]. In the preoperative planning of the tumor resection, it is essential to minimize perioperative blood loss [1]. Due to the complex anatomy of liver vessels, it is challenging to obtain accurate liver-vessel segmentations automatically.

Liver-vessel segmentation is currently done with convolutional neural networks (CNN). Prior work relies on 2D CNN [2, 3] or 3D CNN [4, 5] methods to perform segmentation on CT images. However, these methods use sub-volumes (cropped in three dimensions) of the CT images. These sub-volumes approaches do not take the global topology of the liver-vessel tree into account which could lead to discontinuous and incomplete vessel tree predictions.

To address these shortcomings, prior work [6, 7, 8] makes use of *maximum intensity projection* (MIP) [9]. Given a 3D volume and a set of directions, MIP computes the maximum voxel value along each direction. Therefore, MIP encodes the global vessel topology on 2D projections, and is characterized by high signal-to-noise ratio, while enhancing local vessel probability [9]. However, existing MIP-based methods cannot perform accurately while solely relying on the MIP information. Due to the lack of 3D information, MIP is only an adjunct for segmentation in existing

*MIP*-based methods, which must be combined with a 3D U-Net model [6, 7, 8] or a simulated 3D filter [10]. Thus, they need to further encode 3D spatial information from the CT cross-sectional view by using 3D U-Net models [6] or intersecting 2D filters to simulate 3D convolutions [10]. Therefore, these methods fail to take full advantage of the global topological information of the *MIP*. Moreover, prior *MIP*-based methods focus on the cerebrovascular [10], coronary [6, 8], and retinal vessel [8] segmentation, whereas liver-vessels are characterized by more complex and dense 3D structures.

To incorporate the 3D liver-vessel structure, we introduce a *top-k maximum intensity projection (top-k MIP)*. Unlike *MIP*, which preserves only the maximum voxel value along any given direction, a *top-k MIP* preserves the top- $k$  maximum values along that direction. This allows us to encode fine-grained information regarding the underlying 3D structure of the vessel tree. This information would be missed by simply taking the maximum. Therefore, our model no longer needs to rely on additional 3D convolutions over the CT cross-sectional view to encode this 3D structure.

Fig. 4.1(a) shows the standard CT reconstruction process: given 2D integral projections and their projection directions  $\vec{d}$ , the CT voxels can be computed by back-projecting the integral along each direction, thus reconstructing the 3D object. Similarly, in Fig. 4.1(b) we start from integral projections of liver vessel tree and projection directions  $\vec{d}$ , the vessel tree projections are computed based on the top- $k$  maximum values along each direction, and we aim to reconstruct the 3D liver-vessel tree. To reconstruct the 3D liver-vessel tree use these *top-k MIPs* as conditions in a latent diffusion model (LDM) [11]. Overall, we make the following contributions: (1) we propose incorporating the 3D topology of the liver structure in a principled way by relying on the underlying physics used in CT reconstruction; (2) to this end, we propose a novel *top-k maximum intensity projection* that encodes the fine-grained 3D liver-vessel structure and combine this into a latent diffusion model; (3) finally, we demonstrate improved accuracy of our model on the 3D-ircadb-01 [12] dataset when compared to state-of-the-art methods such *nnUNet* [13], *SwinUNetr* [14], *EnsemDiff* [15] and *MedSegDiff* [16].

## 4.2 Methods

Our model, displayed in Fig. 4.2, represents the global 3D topological structure of the liver via the *top-k MIP* (denoted by  $\mathbf{c}$ ) of the CT cross-sectional view. We further encode *top-k MIP* (via the yellow block) into the latent condition,  $\mathbf{c}'$ . This  $\mathbf{c}'$  is used to condition the latent diffusion model (depicted in blue) which generates the integral projections of ground truth vessel tree by denoising a given noisy ground truth, in the latent space. The ground truth,  $\mathbf{x}_0$ , and the noisy input  $\mathbf{x}_t$  are represented as *integral projections* over directions  $\vec{d}$  of the true 3D vessel tree segmentation. These *integral*

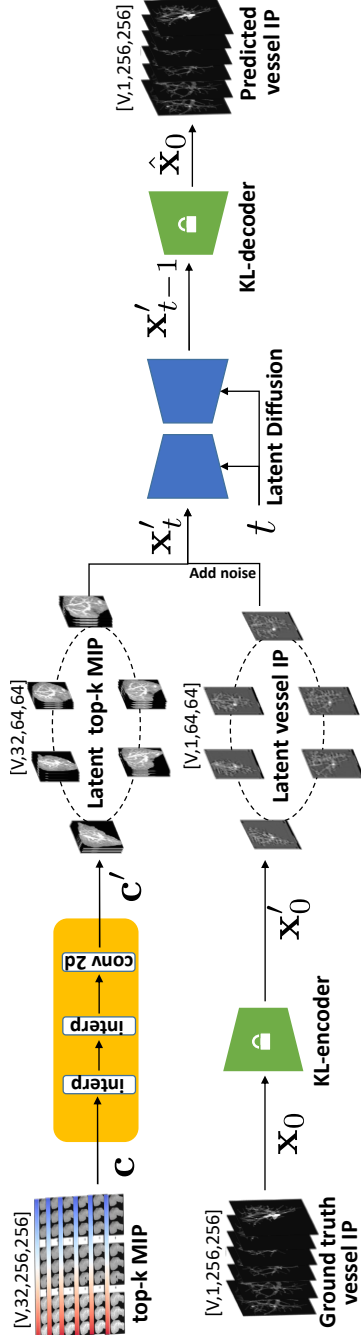


Figure 4.2: **Model outline.** Our model represents the global 3D topology of the liver by computing the *top-k MIP* over the CT volume,  $\mathbf{c}$ . Subsequently, it encodes these *top-k MIP* into a latent condition,  $\mathbf{c}'$  (orange). This  $\mathbf{c}'$  latent is used to condition the latent diffusion model (blue) which recovers the ground truth vessel tree  $\mathbf{x}_0$  from noisy inputs  $\mathbf{x}_t$ . We represent the ground truth via *integral projections (IP)* of the 3D ground truth vessel tree. The ground truth  $\mathbf{x}_0$  and the noisy input  $\mathbf{x}_t$  are encoded via a KL-encoder (green) to be used in the latent diffusion U-net [11]. We denote the different viewing directions by  $V$  in the batch size.

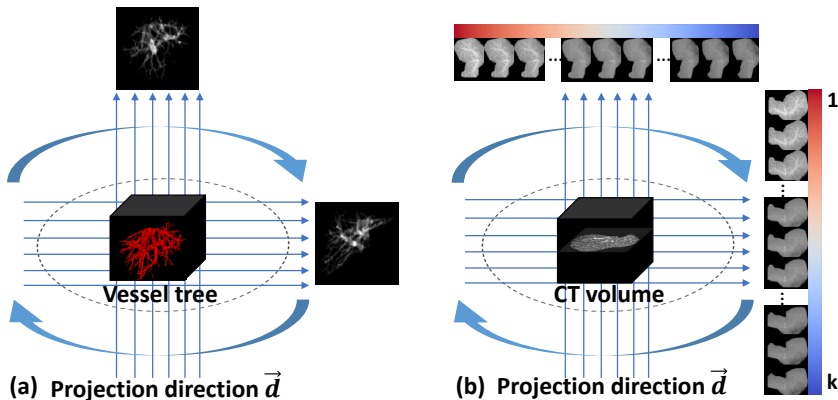


Figure 4.3: **(a) Ground truth IP:** We include all vessel tree slices in the *integral projection (IP)* and accumulate them along projection directions  $\vec{d}$ . **(b) Condition top-k MIP:** We include all CT slices in the *top-k maximum intensity projection (top-k MIP)* and keep the top-k maxima along projection directions  $\vec{d}$ ;

*projections (IP)* are further encoded/decoded via an autoencoder (depicted in green), to be used in the latent diffusion model.

#### 4.2.1 Ground truth transformation

For obtaining the ground truth *integral projection (IP)*, similar to the CT forward projection [17], we assume a bundle of parallel rays  $R^{\vec{d}}$  across different directions  $\vec{d}$ , penetrating the ground truth binary liver-vessel tree, as in Fig. 4.3(a). The *IP* is the sum of the attenuation coefficient  $\mu$  ( $\mu=1$  for binary liver-vessel trees) along a unit path  $\Delta l^{\vec{d}}$ , which is equivalent to the logarithm of the ratio between the ray intensity  $R^{\vec{d}}$  after the attenuation by the liver vessel tree, and the initial ray intensity  $R_0^{\vec{d}}$ :

$$\mathbf{x}_0 \stackrel{\text{def}}{=} \sum_{\Delta l^{\vec{d}}} \mu = -\ln \left( \frac{R^{\vec{d}}}{R_0^{\vec{d}}} \right). \quad (4.1)$$

#### 4.2.2 Top-k MIP transformation

For computing the condition  $\mathbf{c}$  to be input to the latent diffusion model we use *top-k MIP*. Different from the *integral attenuation* in Eq. (4.1), *top-k MIP* keeps the top-k maxima along a ray path  $\Delta l^{\vec{d}}$ , as shown in Fig. 4.3(b):

$$\mathbf{c} \stackrel{\text{def}}{=} \text{top-}k \left\{ \max \left( \Delta l_i^{\vec{d}} \mu, \Delta l_{i+1}^{\vec{d}} \mu, \dots, \Delta l_n^{\vec{d}} \mu \right) \right\}, \quad (4.2)$$

where  $n$  is the number of CT voxels located along a ray, and  $\Delta l_i^{\vec{d}} \mu$  is the value of the  $i$ -th CT voxel along the ray.

### 4.2.3 Top-k MIP conditioning latent diffusion model

We rely on a latent diffusion model [11] conditioned on the *top-k MIP*,  $\mathbf{c}$ , for recovering the denoised liver vessel *IPs*,  $\hat{\mathbf{x}}_0$ . We encode  $\mathbf{c}$  into a latent vector via interpolation and convolution (yellow block in Fig. 4.2), into  $\mathbf{c}'$ . Following [11], we also encode the *IP*,  $\mathbf{x}_0$ , via a pre-trained Kullback-Leibler (KL)-divergence auto-encoder (green blocks in Fig. 4.2) to obtain  $\mathbf{x}'_0$ .

Standard in the diffusion forward process, we iteratively add Gaussian noise  $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$  to  $\mathbf{x}'_0$ . We aim to denoise the noisy inputs  $\mathbf{x}'_t$ , for  $t \in \{T, \dots, 0\}$  by estimating the noise  $\epsilon_\theta$ , modelled as a U-Net [11] (depicted in blue in Fig. 4.2). For this, we optimize the parameters of the model,  $\theta$ , by minimizing the loss over time steps,  $t$ :

$$\mathcal{L}_t(\mathbf{x}'_0, \epsilon, \theta) = \|\epsilon_\theta(\mathbf{x}'_t | \mathbf{c}', t) - \epsilon\|_2^2. \quad (4.3)$$

Once the model is trained, given the encoded *top-k MIP* condition  $\mathbf{c}'$ , and the input Gaussian noise  $\mathbf{x}'_T$ , we gradually denoise the noisy input  $\mathbf{x}'_t$ ,  $t \in \{T, \dots, 0\}$  to the estimate of the encoded *IP*,  $\hat{\mathbf{x}}'_0$ . Given the estimate of the encoded *IP*,  $\hat{\mathbf{x}}'_0$ , we decode this via the pre-trained KL-divergence auto-encoder (green block in Fig. 4.2) into an estimate of the *IP*,  $\hat{\mathbf{x}}_0$ .

### 4.2.4 Post-processing for artifact suppression

Given the estimated *IPs* of the vessel tree,  $\hat{\mathbf{x}}_0$ , we reconstruct the 3D liver vessel tree via filtered back projection (FBP) [17]. Given that, minor inconsistency in the projection domain can cause severe stripe artifacts in the reconstructed images, we employ a simple optimization to suppress these artifacts. Specifically, assume  $\mathbf{T}$  initialized by CT image is a reconstructed liver vessel tree without artifact, and given a matrix  $\mathbf{A}^{\vec{d}}$  recording the voxel indices penetrated by rays under different projection directions  $\vec{d}$ , we impose a projection consistency. Concretely, to dilute the irrelevant background introduced by the initialization of  $\mathbf{T}$ , we enforce that the projection of the reconstructed tree  $\mathbf{T}$  on a certain projection direction encoded in  $\mathbf{A}^{\vec{d}}$  should be as close as possible to the estimated *IP* under that projection direction,  $\hat{\mathbf{x}}_0^{\vec{d}}$ :

$$\mathbf{T} = \arg \min_{\mathbf{T}} \sum_{\vec{d}} \|\mathbf{A}^{\vec{d}} \mathbf{T} - \hat{\mathbf{x}}_0^{\vec{d}}\|_2^2. \quad (4.4)$$

In practice, performing this optimization 10 times is sufficient to suppress the stripe artifacts. Thus, we obtain the binary liver-vessel tree  $\mathbf{T} \geq \text{percentile}(\mathbf{T}, 95)$ . Finally, we perform connected region analysis to cancel small spurious predictions surrounding the segmented vessel tree.

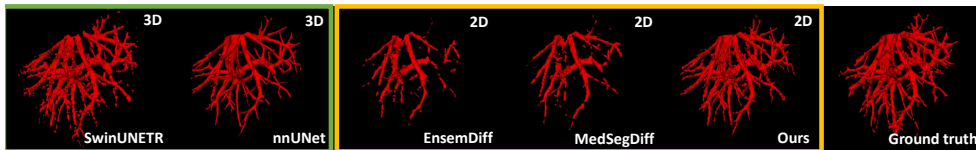


Figure 4.4: **Qualitative comparison on the 3D-ircadb-01 [12] dataset.** We highlight in green the predictions of the 3D methods and in yellow the predictions of the 2D methods. Our model makes more complete and continuous predictions, resembling the ground truth.

## 4.3 Experiments and results

### 4.3.1 Experimental setting

We evaluate our proposed model on the 3D-ircadb-01 [12] dataset, consisting of 20 CT scans. We exclude the vena cava from the region of interest, using liver masks to focus only the vessels within the liver. We resize the liver-masked ROIs to  $[256 \times 256 \times 256]$  for each CT scan, and project 180 viewing directions from 0 to 180 degrees for each CT scan, totaling 3600 projections.  $k$  is 32 when creating the *top-k MIP*.

Small liver vessel annotations are incomplete in the 3D-ircadb-01 dataset [5]. To mitigate this, we asked clinical experts to score the completeness of the annotations based on a 5-point criterion and picked cases (4, 6, 8, 11, 16) whose score  $\geq 4$  as the test set. We apply leave-one-out cross-validation, such that we have 19 cases (*i.e.* 3420 projections) for training and a single case (*i.e.* 180 projections) for testing, per fold.

We trained *nnUNet* [13] and *SwinUNETr* [14] using 3D CT sub-volumes as input, and using their default pre- and post-processing. For *MedSegDiff* [16] and *EnsemDiff* [15] we used 2D CT slices, and ensemble segmentation results 5 times. Our method does not require an ensemble. For all the experiments, we used an NVIDIA A100 (40GB) GPU.

### 4.3.2 Quantitative and qualitative comparison

We report voxel-wise Dice coefficient (*DSC*), Intersection-Over-Union (*IoU*), Sensitivity (*Sen*), Specificity (*Spe*) and centerline Dice coefficient (*clDice*) [18]. Table 4.1 shows the quantitative evaluation, while Fig. 4.4 shows a few prediction examples. Our method achieves the highest *DSC*, *IoU*, and *Sen* scores compared to the other baselines on the 3D-ircadb-01 dataset, demonstrating the added value of the proposed top-k MIP projections. *nnUNet* has a higher *clDice* score than our method. However, the vessel centerline was estimated from ground truth vessel annotations, following [18]. Therefore, incomplete and discontinuous voxel-wise ground truth could bias the centerline extraction in *clDice* [18]. Fig. 4.4 shows that both our method and *nnUNet* predict well-connected vessel structures compared to the ground truth, indicating that *clDice* might not be a sensible choice when comparing segments with inconsistent

Table 4.1: **Quantitative evaluation on the 3D-ircadb-01 [12] dataset.** Our method predicts complete liver-vessel segmentations, while having the highest *IoU*, *Sen* and *DSC* scores. The lower *cIDice* scores of our method compared to *mUNet* could be due to imprecise annotations. Finally, there is a trade-off between complete segmentation (higher *Sen*) or accurate segmentation (higher *Spe*).

	Conv type	View type	<i>DSC</i> (%)	<i>cIDice</i> (%)	<i>IoU</i> (%)	<i>Sen</i> (%)	<i>Spe</i> (%)	<i>FLOPs</i> (G)	<i>Params</i> (M)
<i>mUNet</i> [13]	3D	cross section	58.76 ± 9.89	<b>71.46</b> ± 5.67	42.31 ± 10.19	43.32 ± 11.18	<b>100</b> ± 0	2.90 × 10 <sup>3</sup>	30.7
Swin UNETR[14]	3D	cross section	57.80 ± 9.93	64.16 ± 7.10	41.31 ± 9.51	46.71 ± 13.21	99.96 ± 0.02	6.15 × 10 <sup>2</sup>	62.2
EnsemDiff[15]	2D	cross section	54.82 ± 9.64	60.61 ± 9.55	38.37 ± 9.21	40.05 ± 10.45	99.98 ± 0.02	9.96 × 10 <sup>2</sup>	113.7
MedSegDiff[16]	2D	cross section	59.59 ± 7.73	66.03 ± 8.05	42.85 ± 7.53	47.38 ± 10.44	99.95 ± 0.05	1.05 × 10 <sup>3</sup>	136.8
Ours	2D	projection	<b>64.25</b> ± 7.19	65.87 ± 8.16	<b>47.75</b> ± 8.00	<b>54.03</b> ± 11.23	99.96 ± 0.01	6.83 × 10 <sup>2</sup>	84.5

Table 4.2: Liver vessel *IPs* estimation with (w/) and without (w/o) artifact suppression (opt). Artifact suppression makes the vessel *IPs* more similar to the ground truth vessel *IPs*.

	PSNR( $\uparrow$ )	SSIM( $\uparrow$ )
w/o opt	$13.51 \pm 0.78$	$0.45 \pm 0.07$
w/ opt	<b><math>15.15 \pm 1.32</math></b>	<b><math>0.66 \pm 0.05</math></b>

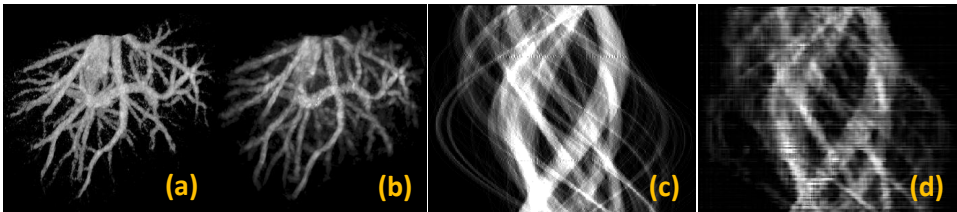


Figure 4.5: Vessel reconstruction with/without artifacts suppression. (a) reconstructed vessel tree with artifact suppression; (b) reconstructed vessel tree without artifact suppression; (c) full-view projection of a slice of the reconstructed vessel trees with artifacts suppression; (d) full-view projection without artifacts suppression. Artifact suppression improves the consistency between different projection views. The vertical axis in (c) and (d) is the projection view.

connectivity. Additionally, all methods in Table 4.1 have *Spe* scores close to 1, indicating that there are few false positives in liver vessel segmentation.

## 4.4 Discussion

### 4.4.1 Effect of artifact suppression

Our model generates liver vessel *IPs* based on the intensity of the *top-k MIP*. However, due to memory constraints, our current design limits the model from learning correlations between different projection directions. The optimization in Eq. (4.4) compensates for this shortcoming. This optimization suppresses the reconstruction artifact caused by projection inconsistency.

Table 4.2 tests the added value of this artifact suppression optimization. The *IPs* of the reconstructed vessel tree with artifact suppression have higher *PSNR* [19] and *SSIM* [19], indicating that the optimized *IPs* have a similar appearance to the ground truth. Full-view projections in Fig. 4.5(c)-(d) also show that the vessel *IPs* consistency between projections is improved by the artifact suppression.

### 4.4.2 Limitations and possible improvements

**Model limitations.** Our proposed model can handle most cases, where the contrast is within normal ranges, in the liver vessel area. However, segmenting extremely low-contrast vascular regions, as shown in the yellow box in Fig. 4.6 remains challenging for our model.

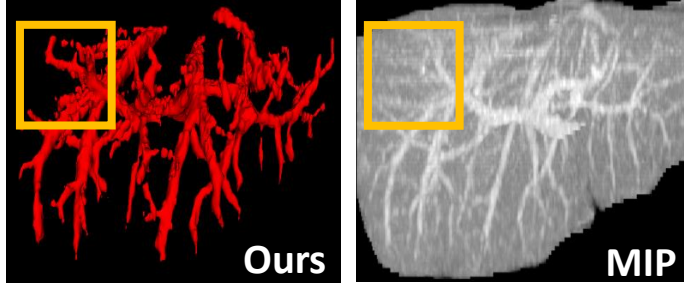


Figure 4.6: **Failure analysis.** Correctly segmenting low-contrast CT images is challenging for our method.

**Possible improvements.** Strengthening the correlation between projection views and learning a dynamic projection for the *top-k* MIP are the two improvement directions. Currently, each view is processed independently, and thus we cannot test using the standard MIP. In the standard MIP the depth information is collapsed to a single maximum value, and cannot be recovered. A model that considers the correlation between projection views, will make it possible to use of the standard MIP, instead of the *top-k* MIP. But now, the standard MIP, whose depth information is completely collapsed to a single maximum, is not applicable to the current method.

## 4.5 Conclusion

We propose a principled way of incorporating 3D liver-vessel topology in 2D diffusion models for liver-vessel segmentation. Accordingly, we draw inspiration from the physics involved in the 3D CT reconstruction. Concretely, we propose to condition a latent diffusion model on *top-k maximum intensity projections* of the CT cross-sectional view. Our proposed model achieves competitive results compared to the existing baselines, validating our approach.

## References

- [1] O. Alirr and A. Rahni. “Survey on liver tumour resection planning system: steps, techniques, and parameters”. In: *Journal of Digital Imaging* 33.2 (2020), pages 304–323.
- [2] T. Kitrungrotsakul, X. Han, Y. Iwamoto, et al. “VesselNet: A deep convolutional neural network with multi pathways for robust hepatic vessel segmentation”. In: *Computerized Medical Imaging and Graphics* 75 (2019), pages 74–83.
- [3] Z. Gao, Q. Zong, Y. Wang, et al. “Laplacian salience-gated feature pyramid network for accurate liver vessel segmentation”. In: *IEEE Transactions on Medical Imaging* 42.10 (2023), pages 3059–3068.
- [4] Q. Huang, J. Sun, H. Ding, et al. “Robust liver vessel extraction using 3D U-Net with variant dice loss function”. In: *Computers in biology and medicine* 101 (2018), pages 153–162.
- [5] Q. Yan, B. Wang, W. Zhang, et al. “Attention-guided deep neural network with multi-scale feature fusion for liver vessel segmentation”. In: *IEEE Journal of Biomedical and Health Informatics* 25.7 (2020), pages 2629–2642.
- [6] Z. Guo, Z. Tan, J. Feng, and J. Zhou. “3D vascular segmentation supervised by 2D annotation of maximum intensity projection”. In: *IEEE Transactions on Medical Imaging* (2024).
- [7] Y. Liu, H. Kwak, and I. Oh. “Cerebrovascular segmentation model based on spatial attention-guided 3D inception U-Net with multi-directional MIPs”. In: *Applied Sciences* 12.5 (2022), page 2288.
- [8] H. Chen, X. Wang, H. Li, and L. Wang. “3D vessel segmentation with limited guidance of 2D structure-agnostic vessel annotations”. In: *IEEE Journal of Biomedical and Health Informatics* (2024).
- [9] S. Napel, M. Marks, G. Rubin, et al. “CT angiography with spiral CT and maximum intensity projection”. In: *Radiology* 185.2 (1992), pages 607–610.
- [10] G. Tetteh, V. Efremov, N. Forkert, et al. “Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes”. In: *Frontiers in Neuroscience* 14 (2020), page 592352.
- [11] R. Rombach, A. Blattmann, D. Lorenz, et al. “High-resolution image synthesis with latent diffusion models”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, pages 10684–10695.
- [12] L. Soler, A. Hostettler, V. Agnus, et al. “3D image reconstruction for comparison of algorithm database: A patient specific anatomical and medical image database”. In: ().
- [13] F. Isensee, P. F. Jaeger, S. A. Kohl, et al. “nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation”. In: *Nature methods* 18.2 (2021), pages 203–211.

- [14] A. Hatamizadeh, V. Nath, Y. Tang, et al. “Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images”. In: *International MICCAI Brainlesion Workshop*. Sept. 2021, pages 272–284.
- [15] J. Wolleb, R. Sandkühler, F. Bieder, et al. “Diffusion models for implicit image segmentation ensembles”. In: *International Conference on Medical Imaging with Deep Learning*. PMLR, Dec. 2022, pages 1336–1348.
- [16] J. Wu, R. FU, H. Fang, et al. “MedSegDiff: Medical Image Segmentation with Diffusion Probabilistic Model”. In: *Medical Imaging with Deep Learning*. 2023.
- [17] R. Schofield, L. King, U. Tayal, et al. “Image reconstruction: Part 1–understanding filtered back projection, noise and image acquisition”. In: *Journal of cardiovascular computed tomography* 14.3 (2020), pages 219–225.
- [18] S. Shit, J. Paetzold, A. Sekuboyina, et al. “clDice-a novel topology-preserving loss function for tubular structure segmentation”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021, pages 16560–16569.
- [19] A. Hore and D. Ziou. “Image quality metrics: PSNR vs. SSIM”. In: *2010 20th international conference on pattern recognition*. IEEE. 2010, pages 2366–2369.

# 5

## Point-based Couinaud segmentation without liver vessel priors

*This chapter was adapted from:*

**Zhang, X., Broersen, A., van Erp, G., Pintea, S.L. and Dijkstra, J., Skip priors and add graph-based anatomical information, for point-based Couinaud segmentation.** (2025) In Reconstruction and Imaging Motion Estimation, and Graphs in Biomedical Image Analysis: First International Workshop, RIME 2025, and 7th International Workshop, GRAIL 2025, Daejeon, South Korea, September 27, 2025, Proceedings. Springer-Verlag, Berlin, Heidelberg, 131-140.



## Abstract

The preoperative planning of liver surgery relies on Couinaud segmentation to reduce the risk of bleeding and guide the resection procedure. Using 3D point-based representations, rather than voxelizing the CT volume, has the benefit of preserving the physical resolution of the CT. However, point-based representations need prior knowledge of the liver vessel structure, which is time consuming to acquire. Here, we propose a point-based method for Couinaud segmentation, without explicitly providing the liver vessel structure. To allow the model to learn the anatomical structure, without providing the liver vessel structure, we add a graph reasoning module on top of the point features. This adds implicit contextual information to the model, by learning affinities across CT-voxels. Our method has competitive results on the *MSD* and *LiTS* public datasets in Dice coefficient and average surface distance scores compared to four state-of-the-art point-based baselines.

## 5.1 Introduction

Primary liver cancer, the sixth most common cancer worldwide, is predominantly caused by HCC (Hepatocellular Carcinoma) [1]. Effective treatment for primary liver cancer relies on two main procedures: liver resection and radiofrequency ablation [1]. Both approaches depend on accurate Couinaud segmentation, to reduce the risk of main vessel puncture, and to guide the placement of ablation needles [2, 3]. Couinaud segmentation divides the liver into eight functionally independent segments. The right-, middle- and left-hepatic veins divide the liver into four sections. These sections are then further split by the horizontal plane defined by the portal vein. In clinical practice, as shown in Fig. 5.1, Couinaud segments are manually delineated based on the hepatic and portal veins.

Prior work for automatically Couinaud segmentation, voxelizes the liver CT to be used in 3D convolutional neural networks (CNNs) [4, 5, 6, 7]. More recently [8], computes point embeddings from sampled 3D points within the liver area. Using sample 3D points has the added value that they preserve the physical CT resolution without the need to resize or crop along the axial direction. Therefore, here we restrict our focus to point-based methods. While relying on 3D point-based representations, our proposed method does not need prior liver vessel information, unlike Zhang et al. [8]. Yet, without prior liver vessel knowledge, we lose anatomical information. To incorporate the anatomic structure, we add a graph reasoning module that learns affinities between the embeddings of the 3D points.

To summarize: (i) We propose a 3D point-based method for Couinaud segmentation that does not rely on anatomic prior knowledge of the liver vessel structure; (ii) We use graph reasoning to learn affinities between different points along the liver and, thus, implicitly learn anatomic structure; (iii) We evaluate our method on two public datasets: *MSD* and *LiTS* and show competitive accuracy when compared to four state-of-the-art point-based segmentation methods.

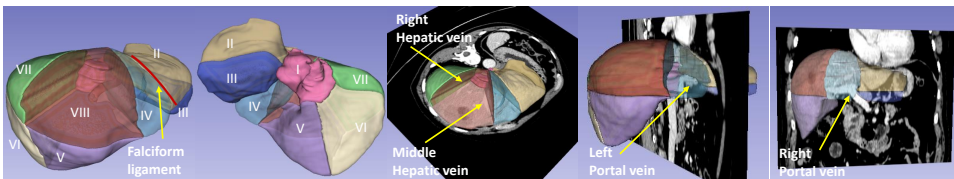


Figure 5.1: Couinaud segmentation is challenging because it requires prior knowledge of the liver vessels.

## 5.2 Related work

### 5.2.1 Couinaud segmentation

Prior work on automatic Couinaud segmentation creates liver atlases [9, 10], divides the liver into voxels to be used in 3D convolutional neural networks (CNNs) [4, 5, 6, 7], or builds deep models on top of sampled 3D points [8, 11]. Atlas-based and partial CNN-based methods [4, 9, 10, 7] require manual landmarks along the hepatic veins, whereas the other prior-free CNN-based methods [5, 6] on CT images need to resize the CT volume to a fixed grid size which changes the physical resolution of the CT images. Point-based models [8, 11] address the limitation of voxelized methods, while still requiring prior liver vessel information. Here, we build on 3D point-based methods, while discarding the need for prior anatomical information, and learning this implicitly via dynamic graph reasoning.

### 5.2.2 Dynamic graph reasoning

Dynamic graph reasoning is widely used in both image-based [12, 13, 14, 15] as well as point-based semantic segmentation methods [16, 17, 18], to capture long-range dependencies. Most image-based methods [12, 13, 14] consider all position pairs when calculating affinities, resulting in high complexity. Unlike these methods, DGMN [15] proposed an adaptive sampling method that considers only limited positions. Similarly, point-based methods also suffer from high complexity of affinity calculation.  $K$ -nearest neighbor (k-NN) is typically used for the complexity reduction, as in [17, 18]. Alternatively, Ma et al. [16] propose to learn channel dependencies instead of dependencies between nodes to capture global contextual information while reducing the computations. Here, we take advantage of both k-NN in the point domain, and the adaptive sampling method in [15] to reduce computations.

## 5.3 Methods

Our model starts from a set of 3D points,  $\mathbf{p}$ , sampled from the liver region over the complete CT volume, and their associated intensities. We follow the design of adaptive graph CNN (AGCNN) [19], processing the points at four levels, as in Fig. 5.2. We extend AGCNN with the green blocks: the grid feature embeddings,  $f(\mathbf{p})$ , enhancing the contextual information; and graph reasoning model,  $\mathcal{G}_r(f(\mathbf{p}))$ , which dynamically learns affinities across points. At the last level, we interpolate the point embeddings,  $\mathbf{p}_4$ , and feed the result to an MLP with two layers ( $\{64, 8\}$ ). We predict the eight Couinaud segments in the liver region of the CT volume. We use the cross-entropy loss to train the model.

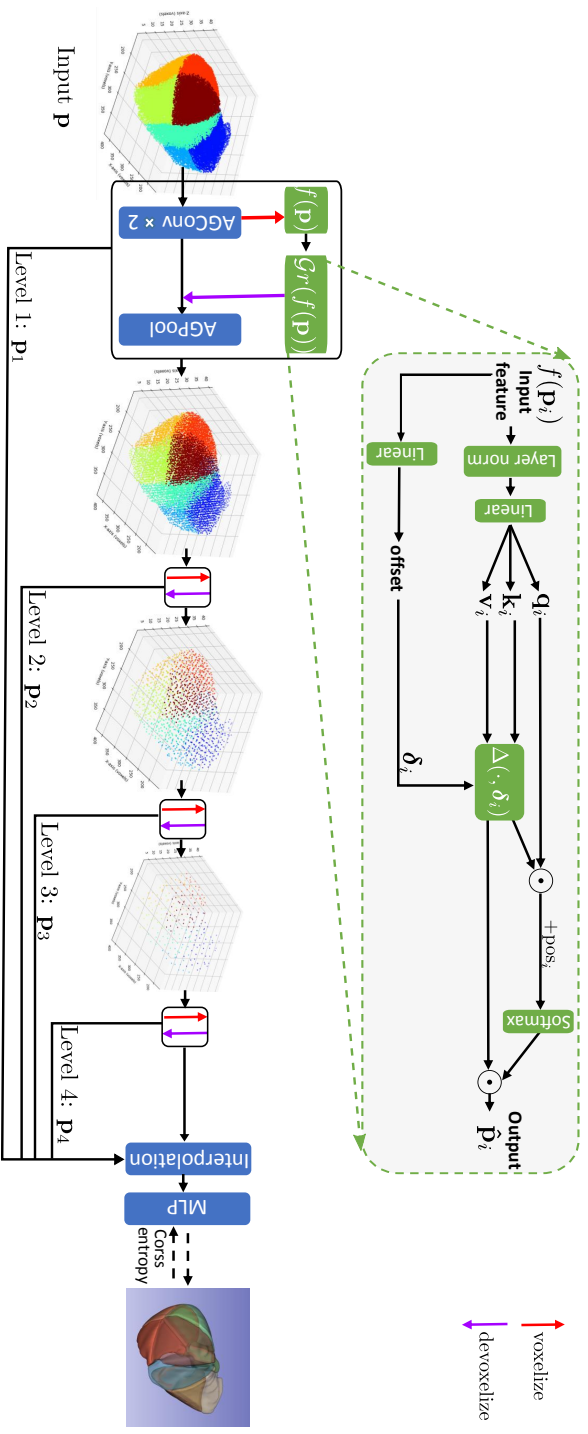


Figure 5.2: Network architecture. We build on the design of [19], and extend this with the green blocks: grid feature embeddings,  $f(p)$  adding contextual information; and graph reasoning model,  $\mathcal{G}^r(f(p))$ , learning dynamic affinities in neighborhood areas.

### 5.3.1 Relation to adaptive graph CNN

AGCNN starts from a set of 3D points,  $\mathbf{p}$ , and their associated CT intensities. Before training, AGCNN precomputes a set of  $K \ll N$  neighbors for each point, using a ball-query sampling scheme, with radius  $r$ , as in Fig. 5.3. At the first level the neighbors of point  $j$ ,  $\mathbf{p}^{ne(j)}$ , are taken from the initial set of points  $\mathbf{p}$ , while at lower resolution levels, the neighbors are taken from the previous level points,  $\mathbf{p}_{i-1}$ ,  $i \in \{2, 3, 4\}$ .

To add implicit anatomical vessel structure to AGCNN, we use a graph reasoning module  $\mathcal{G}$ . Specifically, at each level  $i$  we voxelize all  $N_i$  points,  $\mathbf{p}_i$  into a grid of size  $[M_i \times M_i \times M_i]$ , using the method of Liu *et al.* [20]. In this voxelized space, the graph reasoning module  $\mathcal{G}r(\cdot)$  dynamically learns which voxels are informative. And it learns to select  $3^3$  voxels with which to compute affinities. Intuitively, AGCNN computes affinities of points in the current level with the previous level, while our graph reasoning module computes affinities among the current level points.

### 5.3.2 Implicit anatomic contextual knowledge

**Grid feature embeddings.** The inputs to the model are a set of 3D points in  $\mathbb{R}^3$ , together with their corresponding CT intensity values. We revoxelize the point features and obtain the voxelized point embeddings,  $f(\mathbf{p}_i)$  at each level,  $i$ , by performing two 3D residual convolutions.

**Dynamic graph reasoning.** The grid-space  $[M_i \times M_i \times M_i]$  at each level  $i$  contains many voxels  $\left(\geq \left(\frac{32}{2^{i-1}}\right)^3\right)$  at different levels. This causes memory bottlenecks when calculating the affinity matrix. To address this, we draw inspiration from DGMM [15] and extend their method from 2D to 3D. Specifically, we want to consider only a subset of  $3^3$  voxels out of the  $[M_i \times M_i \times M_i]$  grid, when computing affinities. Following DGMM [15], we use self-attention [21] to define affinities between voxelized points. Given the voxelized point feature  $f(\mathbf{p}_i)$  at level  $i$ , we first project this to *query*  $\mathbf{q}_i$ , *key*  $\mathbf{k}_i$ ,

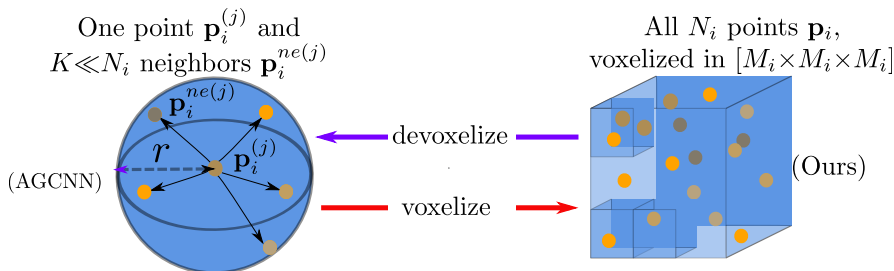


Figure 5.3: **Left:** AGCNN computes offline a set of  $K \ll N$  neighbors  $\mathbf{p}^{ne(j)}$  for each point  $\mathbf{p}^{(j)}$  to be used in the graph affinities. **Right:** Our approach voxelizes all  $N$  points,  $\mathbf{p}$ , into a grid  $[M \times M \times M]$ , and dynamically focuses on  $3^3$  voxels when learning affinities.

and value  $\mathbf{v}_i$ , via a shared linear layer. Additionally, we extend the offsets to learnable 3D offsets  $\boldsymbol{\delta}_i$  pointing to a set of  $3^3$  voxels. These  $3^3$  voxels, to which the offsets are pointing, should contain all the useful contextual information encoded by neighboring voxels. Similar to DGMN [15], we use a deformable unfold layer [22],  $\Delta(\cdot, \boldsymbol{\delta}_i)$ , to adapt the keys and values —  $\mathbf{k}_i, \mathbf{v}_i$ . The output of the graph reasoning is simply a 3D self-attention block over deformed keys and values:

$$\hat{\mathbf{p}}_i = \text{softmax}(\mathbf{q}_i \cdot \Delta(\mathbf{k}_i, \boldsymbol{\delta}_i) + \text{pos}_i) \cdot \Delta(\mathbf{v}_i, \boldsymbol{\delta}_i), \quad (5.1)$$

where  $\boldsymbol{\delta}_i = \text{Linear}(f(\mathbf{p}_i))$ ,  $\boldsymbol{\delta}_i \in \mathbb{R}^{3 \times 3^3}$ ,  $\text{pos}_i \in \mathbb{R}^{3^3}$  are the positional embeddings for the query  $\mathbf{q}_i$  of input features [23]. Finally, we devoxelize  $\hat{\mathbf{p}}_i$  back to point representations using the coordinate-based interpolation [20].

## 5.4 Experiments and results

### 5.4.1 Datasets

We evaluate our method on two public datasets: *MSD* [24] and *LiTS* [25]. Given that there are no Couinaud segment annotations in these two datasets, we use the annotations of Tian *et al.* [26] and Zhang *et al.* [8]. The *MSD* dataset has 192 annotated CT scans, the median in-plane resolution is 0.80 mm (min.: 0.57 mm, max.: 0.98 mm), and the median of interplanar resolutions is 5.00 mm (min.: 1.25 mm, max.: 7.50 mm). The *LiTS* dataset contains 131 annotated CT scans, with median in-plane and interplanar resolutions of 0.77 mm (min.: 0.56 mm, max.: 1.00 mm) and 1.00 mm (min.: 0.70 mm, max.: 5.00 mm), respectively.

### 5.4.2 Experimental setting

We reoriented all CT scans to left-posterior-inferior (LPI) and maintain the original CT image origin and spacing for all experiments. We divide both datasets into training/validation/test sets following the ratio 10/3/7. The CT values are truncated to the range of  $[-100, 300]$  Hounsfield units and then normalized to  $[0, 1]$ . We consider each voxel,  $\mathbf{v}$ , of a CT scan as a point, and compute the point coordinates as:  $\mathbf{p} = s\mathbf{d}\mathbf{v} + \mathbf{o}$  where  $s$ ,  $d$  and  $\mathbf{o}$  are the physical spacing, direction and origin parameters recorded in the CT. We also normalize the physical coordinates of the points to  $[0, 1]$ . We set the  $r$  in the ball-query sampling scheme (Fig. 5.3) to  $\frac{1}{2 \cdot 64}$  and  $\frac{1}{2 \cdot 32}$ , and the first-scale grid size is  $64^3$ , and  $32^3$  for the *MSD* and *LiTS* datasets, respectively. We randomly sample 10% points ( $\approx 50\text{K}$ ) for each training iteration. We evaluate all methods only on the liver region, by masking out other areas. We train our model with 400 epochs and report the final number. We use a stochastic gradient descent (SGD) optimizer with a momentum of 0.98 and a learning rate of 0.01. For all the experiments, we use an NVIDIA A100 (40GB) GPU. We consider point-based baselines: PointNet [27],

PointNet++ [28], AGCNN [19], and Zhang *et al.*'s [8] method, using their default settings for both training and inference.

### 5.4.3 Evaluation metrics

We evaluate all methods only on the liver region, by masking out other areas. We report *Dice* coefficient and average surface distance (*ASD*) in our evaluation. We use *Torchmetrics* [29] and *MONAI* [30] to calculate the two metrics. Surprisingly, these results are slightly different than reported in Zhang *et al.* [8]. We also report inference times and GFLOPs [31].

### 5.4.4 Quantitative comparison

Tab. 5.1 and Tab. 5.2 show the quantitative comparison in *Dice* coefficient and *ASD* scores. In Tab. 5.1, our method achieves the highest *Dice* score for each Couinaud segment. Both our method and *PointNet* have lower *ASD* scores compared to the other methods. *PointNet* has a low *ASD* score because, during inference, it uses a voting scheme which can reduce false positives. However, both *PointNet++* and *PointNet* have large segmentation errors for segment (I). This may be due to the segment having the lowest volume fraction in the liver.

In Tab. 5.2, our method has the highest *Dice* coefficient and the lowest *ASD* averaged over all segments. Segment (IV) is the left medial [32] section (light blue in Fig. 5.1). Segment (IV) is difficult to distinguish because of the large neighborhood area compared to the other segments. Segments (V) and (VIII) belong to the right anterior [32] section, which is adjacent to segment (IV). Thus, inaccurate segmentation of (IV) also negatively affects the accuracy of (V) and (VIII). Moreover, segment (I) is also difficult to segment due to its small volume (in pink in Fig. 5.1). This is exacerbated by the high axial resolution (*i.e.*  $\leq 1.0$  mm) of the *LiTS* dataset, which increases the voxel-wise class imbalance.

PointNet++, AGCNN, and our method use multi-scale point sampling. This results in longer inference times compared to PointNet and Zhang *et al.*, which do not employ multi-scale point sampling. The GFLOPs of AGCNN and our method depend on the radius used in the ball-query sampling.

### 5.4.5 Qualitative comparison

We visualize our Couinaud segmentation and the baseline segmentation on the *MSD* and *LiTS* datasets, as shown in Fig. 5.4 and Fig. 5.5. In Fig. 5.4, we show three different cases with varying axial spacing from 1.5 mm to 7.5 mm, in the *MSD* dataset. For the cases with lower axial resolution (5 mm and 7.5 mm), we show the plane in axial view, and use the red dotted bounding-box to highlight the boundaries between segments. We also show one case with relatively high axial resolution (1.5 mm) in a coronal view, in the last row of Fig. 5.4. All boxes are in the same location, in the

Table 5.1: **Quantitative evaluation on MSD**: We report results per segments (I – VIII), as well as the average. *PointNet* is competitive in *ASD*, however, its *Dice* coefficient is low over all segments. Our method achieves the highest average in *Dice* coefficient, and comparable average in *ASD* with *PointNet*, demonstrating the effectiveness of the added implicit knowledge. (We denote with \* the use of extra vessel-priors.)

	<i>MSD</i> (Dice %) ↓					<i>MSD</i> (ASD mm) ↓				
	<i>PointNet</i> [27]	<i>PointNet++</i> [28]	<i>AGCNN</i> [19]	Zhang <i>et al.</i> [8]	Ours	<i>PointNet</i> [27]	<i>PointNet++</i> [28]	<i>AGCNN</i> [19]	Zhang <i>et al.</i> [8]	Ours
(I)	62.26	72.26	63.96	80.45	<b>83.31</b>	4.04	5.01	2.76	2.96	<b>1.97</b>
(II)	81.53	80.27	77.79	82.39	<b>86.71</b>	1.67	<b>1.02</b>	6.19	4.75	1.89
(III)	69.80	72.75	64.55	75.06	<b>79.35</b>	3.31	<b>1.96</b>	5.78	3.30	2.51
(IV)	61.74	68.04	63.26	69.83	<b>73.26</b>	5.76	5.90	6.52	5.46	<b>4.12</b>
(V)	70.55	71.78	68.78	71.49	<b>75.89</b>	<b>3.56</b>	5.61	11.75	5.91	6.36
(VI)	75.88	75.29	69.91	72.51	<b>79.51</b>	<b>2.68</b>	4.57	8.20	5.23	5.15
(VII)	82.38	82.24	80.35	80.56	<b>85.11</b>	<b>2.47</b>	3.02	5.62	5.57	3.27
(VIII)	75.86	76.03	73.72	75.74	<b>80.16</b>	<b>4.10</b>	4.19	6.79	4.82	4.84
Avg	72.50	74.83	70.29	76.00	<b>80.41</b>	<b>3.45</b>	3.91	6.70	4.75	3.76
	Time (s) per case					<i>GfLOPs</i>				
	2.81	21.55	6.13	1.73	10.03	302.38	64.65	638.54	608.85	771.15

Table 5.2: **Quantitative evaluation on *LiTS***: Results per Couinaud segment (I – VIII) and the overall average. Zhang *et al.* [8] obtains higher scores on some segments, yet their method uses vessel priors. Our method, without prior vessel knowledge, achieves the highest average in *Dice* coefficient and the lowest average in *ASD*.

	<i>LiTS</i> (Dice %) $\uparrow$				<i>LiTS</i> (ASD mm) $\downarrow$					
	PointNet [27]	PointNet++ [28]	AGCNN [19]	Zhang <i>et al.</i> [8]	Ours	PointNet [27]	PointNet++ [28]	AGCNN [19]	Zhang <i>et al.</i> [8]	Ours
(I)	49.80	37.85	69.20	<b>73.64</b>	68.84	8.39	8.58	<b>4.36</b>	6.54	6.97
(II)	70.69	72.78	82.62	82.82	<b>86.17</b>	9.39	5.31	4.58	4.81	<b>3.18</b>
(III)	58.26	65.68	76.09	72.46	<b>80.82</b>	15.18	7.42	6.04	5.09	<b>3.18</b>
(IV)	53.87	70.19	73.88	<b>75.40</b>	75.24	8.46	9.69	10.00	<b>7.88</b>	8.38
(V)	80.46	80.51	78.97	81.92	<b>83.03</b>	<b>5.04</b>	5.85	7.25	6.49	6.20
(VI)	77.69	79.23	73.45	79.28	<b>79.29</b>	4.55	6.02	5.76	6.67	<b>3.78</b>
(VII)	79.92	81.71	80.42	<b>83.40</b>	82.79	<b>4.38</b>	4.93	6.75	4.63	4.60
(VIII)	77.20	79.29	77.53	80.02	<b>80.26</b>	6.45	6.67	7.71	<b>5.84</b>	7.71
Avg	68.49	70.90	76.52	78.62	<b>79.56</b>	7.73	6.81	6.56	6.00	<b>5.50</b>
Time (s) per case										
									<i>GFLOPs</i>	
	10.62	106.97	19.81	6.52	13.83	302.38	64.65	256.10	608.85	141.56

Table 5.3: **Model ablation study in Dice coefficient and average surface distance (ASD) on the MSD [24] and LiTS [25] datasets.** (a) AGCNN [19] baseline; (b) without graph reasoning,  $\mathcal{G}r(f(\mathbf{p}))$ , in our model; (c) without grid feature embeddings,  $f(\mathbf{p})$ , in our proposed model; (d) our proposed model. All our model components contribute to the final model’s scores.

	MSD		LiTS	
	Dice $\uparrow$ (%)	ASD $\downarrow$ (mm)	Dice $\uparrow$ (%)	ASD $\downarrow$ (mm)
(a)	70.29 ( $\pm$ 16.69)	6.70 ( $\pm$ 5.80)	76.52 ( $\pm$ 8.90)	6.56 ( $\pm$ 2.65)
(b)	78.02 ( $\pm$ 11.06)	4.00 ( $\pm$ 3.39)	77.46 ( $\pm$ 10.87)	6.58 ( $\pm$ 4.03)
(c)	68.98 ( $\pm$ 14.58)	7.73 ( $\pm$ 5.19)	72.67 ( $\pm$ 10.30)	9.81 ( $\pm$ 5.44)
(d)	<b>80.41</b> ( $\pm$ 10.74)	<b>3.76</b> ( $\pm$ 4.04)	<b>79.56</b> ( $\pm$ 6.95)	<b>5.50</b> ( $\pm$ 2.42)

corresponding CT image. We can see that the boundaries of our method are most similar to the ground truth boundaries (last column). On the last row, both our method and Zhang *et al.*’s [8] boundaries follow the box centerline. However, Zhang *et al.*’s [8] method does not correctly recognize segment (III) in this case, as seen in the right-upper corner. In Fig. 5.5, we show the boundary comparison for three cases in three views (axial, coronal and sagittal). Similar to the results in Fig. 5.4, the boundaries of our segments are located on the box centerline. The baselines boundaries are slightly offset, in this case, as seen in the first row in Fig. 5.5. In the second and third rows, all baselines except ours show varying degrees of misclassification.

#### 5.4.6 Model ablation study

To verify that all the components of our model contribute to the segmentation, we perform model ablations in Tab. 5.3. We consider four cases: (a) the baseline AGCNN [19]; (b) our method with grid feature embeddings,  $f(\mathbf{p})$ , but without graph reasoning,  $\mathcal{G}r(f(\mathbf{p}))$ ; (c) our model with graph reasoning,  $\mathcal{G}r(f(\mathbf{p}))$ , but without grid feature embeddings,  $f(\mathbf{p})$ ; (d) our complete model. Grid feature embeddings  $f(\mathbf{p})$  are an important bridge between features in 3D point-space,  $\mathbf{p}$ , and the graph reasoning module  $\mathcal{G}r(f(\mathbf{p}))$ . As seen for the setting (c) in Tab. 5.3, the graph reasoning module fails to work directly, on the voxelized point-features. In addition, the graph reasoning module in itself contributes to the model predictions, as seen in the setting (b) of Tab. 5.3. Removing  $\mathcal{G}r(f(\mathbf{p}))$  decreases the dice scores by  $\approx 2\%$  on both the MSD and LiTS datasets. On both the LiTS data and the MSD dataset, all model components prove useful.

### 5.5 Discussion

Although the proposed method does not need an explicit vessel prior as [8, 11], it still depends on the anatomical landmarks in the liver to learn graph structures as

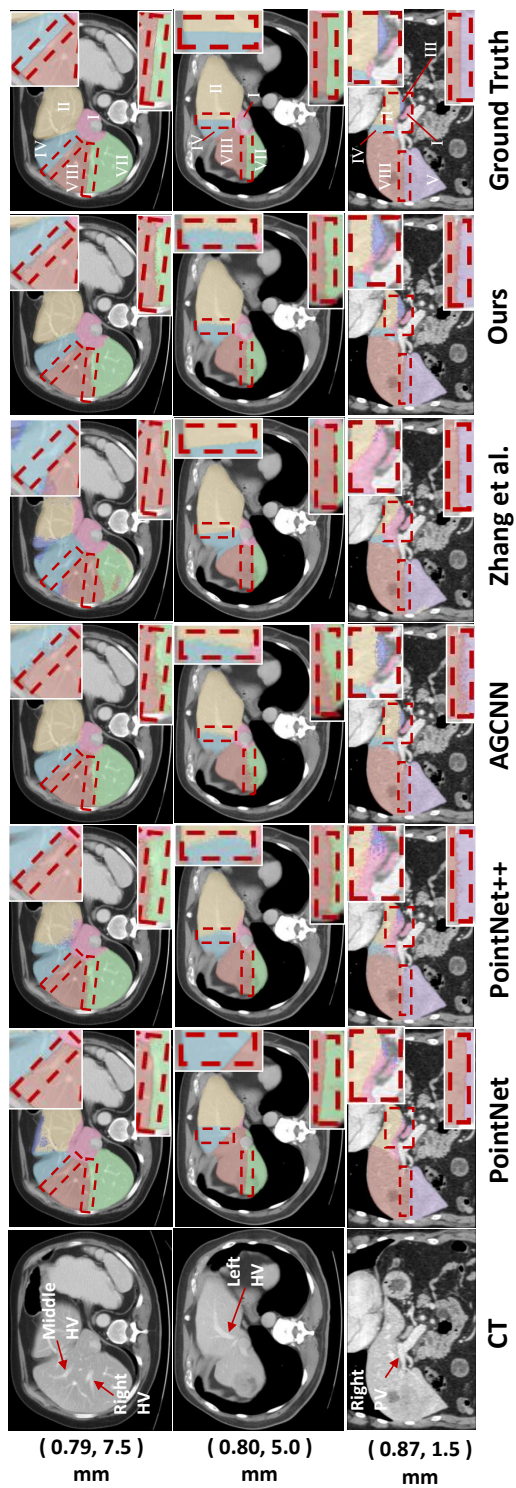


Figure 5.4: **Qualitative evaluation on MSD.** The anatomical landmarks [32] are marked with red arrows in the CT image, as a reference. The red dotted bounding-box highlights the boundary between segments. Our method predicts segment boundaries that are closest to the ground truth. We show to the left the in-plane and interplane resolutions. (HV: hepatic vein, PV: portal vein)

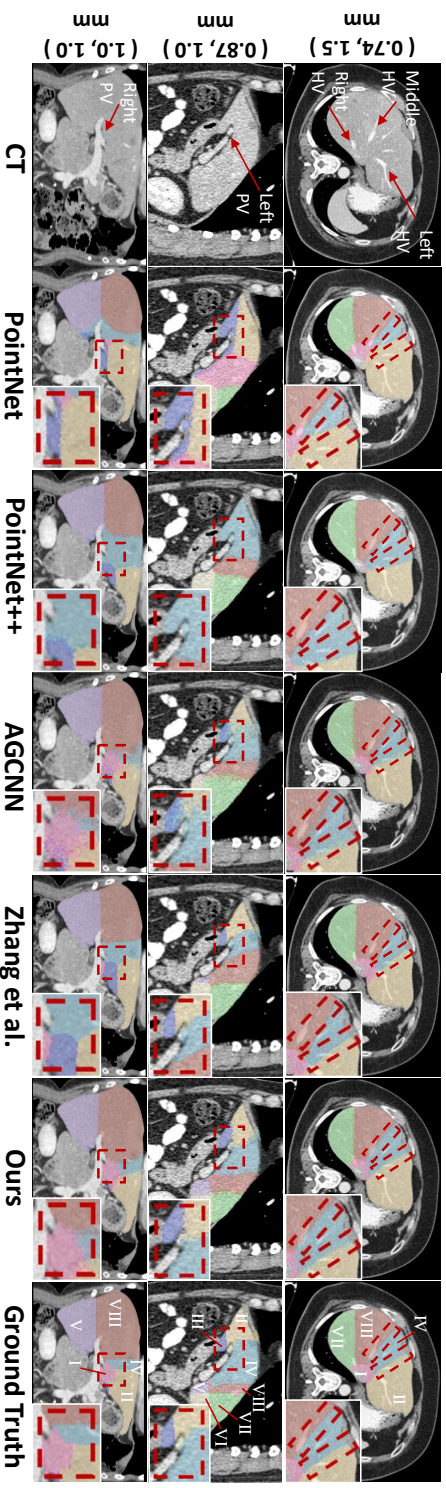


Figure 5.5: **Qualitative evaluation on LITS.** Three cases in axial, sagittal, and coronal views. The red arrows mark the anatomical landmarks [32], and the bounding-boxes highlight segment boundaries. Our method displays the most accurate boundaries.

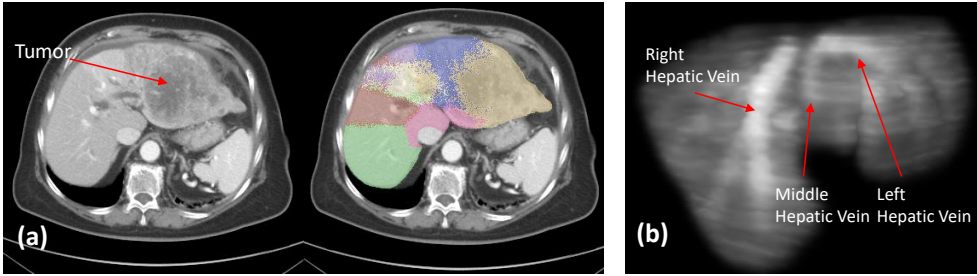


Figure 5.6: **Failure case analysis.** (a) The anatomical landmarks disappear in the tumor region marked by the red arrow. (b) Output of  $\mathcal{G}_r(f(\mathbf{p}))$  of a successful case. The model learns the graph affinities based on the anatomical landmarks in the CT images.

shown in Fig. 5.6 (b). Therefore, the proposed method cannot accurately discriminate segments in the liver when a large tumor appears, as shown in Fig. 5.6 (a). This is because the partial anatomical landmarks in the CT image are absent, such as the left hepatic vein and the left portal vein. If no anatomical landmarks are visible, the model divides the liver based on noise in the CT images, resulting in segmentations without well-defined borders. In addition, a relatively larger liver with a large tumor, as in Fig. 5.6, also causes inaccurate segmentation. This is because there is an obvious difference between the 3D coordinates of a normal-sized liver and those of a larger liver with a tumor. The proposed point-based method depends on the 3D coordinates when performing the inference, coordinates outliers may fail the discrimination of the Couinaud segment boundary.

## 5.6 Conclusion

We propose a 3D point-based method for Couinaud segmentation for CT images. Unlike prior work, we do not rely on anatomic prior knowledge of the vessel structure in the liver area. Our model learns implicitly contextual anatomic information, by learning affinities between voxels in the CT volume in a dynamic graph reasoning module. This implicit contextual information makes our model competitive compared to state-of-the-art, where we exceed PointNet, PointNet++, AGCNN and Zhang *et al.*'s [8] method on the *MSD* and *LiTS* public benchmarks. This approach voids the time-consuming definition of the vessel structure in the liver while still showing the better performance.



## References

- [1] J. M. Llovet, R. K. Kelley, A. Villanueva, et al. “Hepatocellular carcinoma (primer)”. In: *Nature Reviews: Disease Primers* 7.1 (2021), page 6.
- [2] O. I. Alirri and A. A. A. Rahni. “Survey on liver tumour resection planning system: steps, techniques, and parameters”. In: *Journal of Digital Imaging* 33.2 (2020), pages 304–323.
- [3] M. Luo, H. Jiang, and T. Shi. “Multi-stage puncture path planning algorithm of ablation needles for percutaneous radiofrequency ablation of liver tumors”. In: *Computers in Biology and Medicine* 145 (2022), page 105506.
- [4] Z. Arya, G. Ridgway, A. Jandor, and P. Aljabar. “Deep learning-based landmark localisation in the liver for Couinaud segmentation”. In: *Medical Image Understanding and Analysis: 25th Annual Conference, MIUA 2021, Oxford, United Kingdom, July 12–14, 2021, Proceedings 25*. Springer. 2021, pages 227–237.
- [5] X. Jia, C. Qian, Z. Yang, et al. “Boundary-aware dual attention guided liver segment segmentation model”. In: *KSII Transactions on Internet and Information Systems (TIIS)* 16.1 (2022), pages 16–37.
- [6] J. Tian, L. Liu, Z. Shi, and F. Xu. “Automatic couinaud segmentation from CT volumes on liver using GLC-UNet”. In: *International workshop on machine learning in medical imaging*. Springer. 2019, pages 274–282.
- [7] M. Wang, R. Jin, J. Lu, et al. “Automatic CT liver Couinaud segmentation based on key bifurcation detection with attentive residual hourglass-based cascaded network”. In: *Computers in Biology and Medicine* 144 (2022), page 105363.
- [8] X. Zhang, S. Ali, T. Liu, et al. “Robust and smooth Couinaud segmentation via anatomical structure-guided point-voxel network”. In: *Computers in Biology and Medicine* 182 (2024), page 109202.
- [9] D. Boltcheva, N. Passat, V. Agnus, et al. “Automatic anatomical segmentation of the liver by separation planes”. In: *Medical Imaging 2006: Visualization, Image-Guided Procedures, and Display*. Volume 6141. SPIE. 2006, pages 383–394.
- [10] S. Pla-Alemany, J. A. Romero, J. M. Santabàrbara, et al. “Automatic multi-atlas liver segmentation and Couinaud classification from CT volumes”. In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE. 2021, pages 2826–2829.
- [11] X. Zhang, Y. Liu, S. Ali, et al. “Anatomical-aware point-voxel network for couinaud segmentation in liver ct”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2023, pages 465–474.
- [12] Y. Chen, M. Rohrbach, Z. Yan, et al. “Graph-based global reasoning networks”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pages 433–442.
- [13] F. Manessi, A. Rozza, and M. Manzo. “Dynamic graph convolutional networks”. In: *Pattern Recognition* 97 (2020), page 107000.

- [14] X. Wang, R. Girshick, A. Gupta, and K. He. “Non-local neural networks”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, pages 7794–7803.
- [15] L. Zhang, M. Chen, A. Arnab, et al. “Dynamic Graph Message Passing Networks for Visual Recognition”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).
- [16] Y. Ma, Y. Guo, H. Liu, et al. “Global context reasoning for semantic segmentation of 3D point clouds”. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2020, pages 2931–2940.
- [17] Y. Wang, Y. Sun, Z. Liu, et al. “Dynamic graph cnn for learning on point clouds”. In: *ACM Transactions on Graphics (tog)* 38.5 (2019), pages 1–12.
- [18] W. Zhou, Q. Wang, W. Jin, et al. “Graph Transformer for 3D point clouds classification and semantic segmentation”. In: *Computers & Graphics* 124 (2024), page 104050.
- [19] M. Wei, Z. Wei, H. Zhou, et al. “AGConv: Adaptive graph convolution on 3D point clouds”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.8 (2023), pages 9374–9392.
- [20] Z. Liu, H. Tang, Y. Lin, and S. Han. “Point-voxel cnn for efficient 3d deep learning”. In: *Advances in neural information processing systems* 32 (2019).
- [21] Z. Liu, Y. Lin, Y. Cao, et al. “Swin transformer: Hierarchical vision transformer using shifted windows”. In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2021, pages 10012–10022.
- [22] X. Zhu, H. Hu, S. Lin, and J. Dai. “Deformable convnets v2: More deformable, better results”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2019, pages 9308–9316.
- [23] P. Shaw, J. Uszkoreit, and A. Vaswani. “Self-Attention with Relative Position Representations”. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. Edited by M. Walker, H. Ji, and A. Stent. New Orleans, Louisiana: Association for Computational Linguistics, June 2018, pages 464–468.
- [24] M. Antonelli, A. Reinke, S. Bakas, et al. “The medical segmentation decathlon”. In: *Nature communications* 13.1 (2022), page 4128.
- [25] P. Bilic, P. Christ, H. B. Li, et al. “The liver tumor segmentation benchmark (lits)”. In: *Medical Image Analysis* 84 (2023), page 102680.
- [26] J. Tian, L. Liu, Z. Shi, and F. Xu. “Automatic couinaud segmentation from CT volumes on liver using GLC-UNet”. In: *International workshop on machine learning in medical imaging*. Springer. 2019, pages 274–282.
- [27] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. “Pointnet: Deep learning on point sets for 3d classification and segmentation”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pages 652–660.

- [28] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. “Pointnet++: Deep hierarchical feature learning on point sets in a metric space”. In: *Advances in neural information processing systems* 30 (2017).
- [29] N. S. Detlefsen, J. Borovec, J. Schock, et al. “Torchmetrics-measuring reproducibility in pytorch”. In: *Journal of Open Source Software* 7.70 (2022), page 4101.
- [30] M. J. Cardoso, W. Li, R. Brown, et al. “Monai: An open-source framework for deep learning in healthcare”. In: *arXiv preprint arXiv:2211.02701* (2022).
- [31] xiaoju ye. *calflops: a FLOPs and Params calculate tool for neural networks in pytorch framework*. 2023. URL: <https://github.com/MrYxJ/calculate-flops.pytorch>.
- [32] S. Rutkauskas, V. Gedrimas, J. Pundzius, et al. “Clinical and anatomical basis for the classification of the structural parts of liver”. In: *Medicina (Kaunas)* 42.2 (2006), pages 98–106.



# 6

## Summary and future work

Coronary plaque detection and intrahepatic anatomy analysis are two important topics in coronary artery disease diagnosis and preoperative planning for liver surgery based on CTA images. In this thesis, we aimed to develop novel deep learning-based approaches that address key challenges and substantially advance the state-of-the-art in both areas. In this chapter, we summarize the previous chapters and discuss potential research directions for future work.

### 6.1 Summary

We present deep learning solutions to address the three specific aims: non-invasive coronary plaque detection (Chapter 2), 2D model-based continuous liver vessel segmentation (Chapters 3 and 4), and lastly, vessel prior-free Couinaud liver segmentation (Chapter 5).

In **Chapter 2**, we focus on the detection and characterization of coronary plaque in CTA images, aiming to reduce the need for invasive imaging modalities in patients. The ground truth obtained from the invasive NIRS-IVUS modality was registered to the corresponding CTA images. The registered labels on the CTA images were used as targets for training the proposed deep learning models. We designed two convolutional neural network (CNN)-based methods to predict the plaque occurrence angle in the cross-sectional view. Both methods employ a 2.5D approach to introduce contextual information along the longitudinal direction of the coronary arteries. The first method is developed on top of a Dense U-Net, which is used to fit pie-shaped binary masks for both lipid-rich and calcified plaques. The predicted masks represent the plaque occurrence areas in the cross-sectional view. The angular range of the area is calculated through an additional post-processing step and visualized in a spread-out view, similar to the chemogram view used in the NIRS modality. Another method is developed on top of a Mask R-CNN architecture. Unlike the first method, this approach is applied to images in a polar coordinate system. The start and end angles of plaques can be directly predicted using bounding box regression. The proposed methods are trained and validated on an in-house dataset consisting of 64 patients, achieving a sixfold increase in the median F1-score compared to the conventional HU-based method at the lesion level for detecting lipid-rich plaques.

**Chapter 3** focuses on improving the continuity of the segmented liver vessel tree. We leverage both a graph attention mechanism and a diffusion model to facilitate continuous and complete liver vessel segmentation. To incorporate the vascular graph into the diffusion model, the ground truth vessel masks in each batch are transformed into graphs composed of nodes and edges, explicitly representing vascular connectivity. Edge weights can be learned through graph attention layers, which serve as a dynamic matrix representing the affinities between nodes distributed along the vessels. The weighted node features can be injected into the diffusion model as a condition to encourage the generation of continuous vessel segmentation. The proposed model is evaluated on two public datasets and outperforms other state-of-the-art medical segmentation methods in terms of Dice coefficient and connectivity metrics.

**Chapter 4** presents a top-k MIP prior, which preserves more depth information than the original MIP, for liver vessel segmentation. The proposed top-k MIP prior represents vessel connectivity through a sequence of 2D projection views, while preserving vessel diameter information using the top-k maxima of CT intensity. To learn the mapping between the top-k MIP and the integral projections of the ground truth vessel mask, we employ a 2D latent diffusion model. The top-k MIP prior is used as the condition for the latent diffusion model, which generates integral projections of the vessel tree. The generated integral projections can be reconstructed using the filtered back-projection method commonly employed in the CT reconstruction field. The evaluation results demonstrate that the proposed prior contributes to continuous vessel segmentation and achieves competitive performance compared to other medical segmentation baselines.

**Chapter 5** focuses on Couinaud liver segmentation without the use of explicit vessel guidance. To compensate for the absence of explicit vascular structures, we introduce 3D graph reasoning to dynamically learn intrahepatic anatomy. Additionally, to preserve the physical resolution of the CT images, we adopt a point-based architecture for this task. The learned implicit intrahepatic anatomy can serve as a replacement for explicit vessel priors, providing embeddings for the point-based model. The proposed point-based method with graph reasoning outperforms other point-based baselines in both Dice coefficient and surface distance, while eliminating the need for liver vessel priors in Couinaud segmentation.

## 6.2 Discussion and future work

The overarching aim of the thesis was to develop novel deep learning-based approaches that address key challenges and substantially advance the state-of-the-art in both areas. This was realized by developing automatic methods for non-invasive coronary plaque detection and intrahepatic anatomy analysis, including liver vessel segmentation and Couinaud liver segmentation. First, coronary plaque detection was explored

in Chapter 2. We develop two deep learning methods to detect and characterize coronary plaques in cross-sectional CTA images. Although the proposed methods outperform the conventional HU-based approach in quantitative metrics, several challenges remain in coronary plaque detection. The overlapping HU range between lipid-rich plaques and the surrounding myocardial tissue makes accurate distinction of the plaque area challenging. Calcified plaques with high intensities can also cause artifacts, resulting in low-attenuation shadows that resemble the appearance of lipid-rich plaques. Additionally, blooming artifacts caused by implanted metal stents present another challenge, as they can obscure plaque features. Current methods do not perform well on artifact-affected CTA images; therefore, coronary arteries with implanted stents were excluded from the experiments. Most of the existing challenges are closely related to CTA image quality—specifically, whether the features of different plaque types are clearly distinguishable. Although the results presented in Chapter 2 demonstrate that deep-learning methods have the potential to extract more implicit features from CTA images and achieve more accurate plaque detection than the conventional HU-based method, essential improvements will rely on advanced CT imaging techniques, such as photon-counting CT (PCCT) [1]. Developing deep-learning methods based on the PCCT modality will be an important direction for future research.

In Chapters 3,4 and 5, Intrahepatic anatomy analysis was investigated. Chapters 3 and 4 primarily address continuous liver vessel segmentation, while Chapter 5 focuses on Couinaud segmentation without using vessel priors. Both liver vessel segmentation and Couinaud segmentation are essential steps in the preoperative planning of liver surgery. However, the challenges of these two tasks differ: connectivity is the main concern for vessel segmentation, while accurate boundary determination is key for Couinaud segmentation. The two proposed methods in Chapters 3 and 4 start from different image views (cross-sectional and projection views) and employ diffusion model-based architectures to improve the connectivity of the segmented vessel tree. Unlike most existing segmentation methods, which are typically based on deterministic models, diffusion models can learn the underlying data distribution and intrinsic class characteristics. We utilize this advantageous property of the diffusion model in our vessel segmentation task. To better control the stochasticity of the diffusion model and guide the generation toward the desired output, we use graph attention as a condition in Chapter 3 and mimic the CT reconstruction process in Chapter 4. These chapters demonstrate that these strategies are all beneficial for enhancing vessel connectivity in 3D. The quantitative and qualitative results show that vessel connectivity can be significantly improved compared to the other advanced segmentation methods. However, current methods have two main limitations. First, the proposed models are supervised, and their performance can be adversely affected

by inaccurate vessel annotations. Second, inference with diffusion models is more computationally expensive compared to other deterministic segmentation methods. Thus, future work should aim to overcome these two limitations. Weakly supervised methods may help mitigate the negative impact of inaccurate vessel annotations. Additionally, developing more advanced mathematical frameworks for diffusion models could improve the efficiency of both training and inference.

Couinaud segmentation, the second focus of intrahepatic anatomy analysis, also faces the challenge of inaccurate vessel annotations, as conventional methods used in clinical practice rely heavily on vessel priors to determine the boundaries between liver segments. This challenge is addressed by our proposed prior-free method presented in Chapter 5. We adopt graph reasoning to learn implicit intrahepatic anatomy instead of relying on explicit liver vessels. Consequently, the proposed method is independent of vessel annotation quality and eliminates the need for a labor-intensive vessel annotation process. However, neither the proposed method nor the baseline methods that were compared can handle livers with large tumors effectively. This is because parts of the intrahepatic structures are deteriorated by the tumor, making it difficult for the neural network to capture anatomical features in the tumor-occupied regions. Therefore, future research could focus on strategies to compensate for missing anatomical structures. Enhancing the learning of long-range dependencies could be a potential solution, allowing the tumor-occupied regions lacking anatomical features to be associated with neighboring tumor-free regions.

All the chapters in this thesis are centered on the two main topics of coronary plaque detection and intrahepatic anatomy analysis based on CTA/CT images. Although all the proposed methods are developed based on the CT modality, they can potentially be applied to similar clinical tasks using other imaging modalities.

### **6.3 General conclusions**

In conclusion, this thesis proposes deep learning methods for two important clinical topics: noninvasive coronary plaque detection and intrahepatic anatomy analysis based on CTA/CT images. Specifically, we first explore the mapping of plaque features between invasive images and noninvasive CTA images in the cross-sectional view, achieving more accurate and finer angle-wise plaque detection. Moreover, we efficiently leverage the advantages of different models such as diffusion models, graph neural networks, and point-based models, and adapt them appropriately to facilitate continuous liver vessel segmentation and eliminate the need for vessel priors in Couinaud segmentation. All deep learning methods proposed in this thesis have been validated both quantitatively and qualitatively, demonstrating superior performance compared to other state-of-the-art baselines.

## References

- [1] A. Meloni, F. Frijia, D. Panetta, et al. “Photon-counting computed tomography (PCCT): technical background and cardio-vascular applications”. In: *Diagnostics* 13.4 (2023), page 645.



# 7

## Samenvatting en toekomstig onderzoek

De detectie van coronaire plaques en analyse van de anatomische structuren in de lever zijn twee belangrijke onderwerpen binnen de diagnostiek van coronair lijden en de preoperatieve planning van leverchirurgie op basis van CTA-beelden. Het doel in dit proefschrift was om nieuwe, deep learning-gebaseerde methoden te ontwikkelen die de cruciale uitdagingen aanpakken om de huidige technieken in beide domeinen aanzienlijk vooruit te helpen. In dit hoofdstuk vatten wij de voorgaande hoofdstukken samen en bespreken wij mogelijke onderzoeksrichtingen voor toekomstig werk.

### 7.1 Samenvatting

Wij presenteren deep learning-oplossingen om drie specifieke doelstellingen te beantwoorden: niet-invasieve detectie van coronaire plaques (Hoofdstuk 2), op een 2D-model gebaseerde, continue segmentatie van levertaten (Hoofdstukken 3 en 4), en ten slotte Couinaud-segmentatie van de lever zonder voorkennis van de vaatstructuur (Hoofdstuk 5).

In **Hoofdstuk 2** richten wij ons op de detectie en karakterisering van coronaire plaques in CTA-beelden, met als doel het gebruik van invasieve beeldvormingsmodaliteiten bij patiënten te verminderen. De referentiedata verkregen uit de invasieve NIRS-IVUS-modaliteit werd geregistreerd op de overeenkomstige CTA-beelden. De geregistreerde labels op de CTA-beelden werden gebruikt voor het trainen van de voorgestelde deep learning-modellen. Wij ontwierpen twee op convolutionele neurale netwerken (CNN's) gebaseerde methoden om de hoek van de plaque in het beeld met de dwarsdoorsnede te voorspellen. Beide methoden maken gebruik van een 2.5D-benadering om contextuele informatie langs de longitudinale richting van de kransslagaders te integreren. De eerste methode is ontwikkeld op basis van een Dense U-Net, waarmee taartpuntvormige binaire maskers voor zowel vetrijke als verkalkte plaques worden geplaatst. De voorspelde maskers representeren de gebieden met plaque in de dwarsdoorsnede. De hoeken van dit gebied worden berekend in een aanvullende post-processing stap en gevisualiseerd in een uitgevouwen weergave, vergelijkbaar met de chemogram weergave die wordt gebruikt in de NIRS-modaliteit. Een tweede methode is ontwikkeld op basis van een Mask R-CNN-architectuur. In tegenstelling tot de eerste methode wordt deze benadering toegepast op beelden in

een polair coördinatensysteem. De begin- en eindhoeken van plaques kunnen direct worden voorspeld via stapsgewijze regressie. De voorgestelde methoden zijn getraind en gevalideerd op een interne dataset bestaande uit 64 patiënten, en realiseerden een zesvoudige toename van de mediane F1-score in vergelijking met de conventionele HU-gebaseerde methode op laesieniveau voor de detectie van vetrijke plaques.

**Hoofdstuk 3** richt zich op het verbeteren van de continuïteit van de gesegmenteerde levervaatboom. Wij maken gebruik van zowel een aandachtsmechanisme gebaseerd op grafen als een diffusiemodel om continue en volledige segmentatie van levervaten te faciliteren. Om de vasculaire graaf in het diffusiemodel te integreren, worden de referentivaatmaskers in elke batch getransformeerd naar grafen bestaande uit knopen en lijnen, die de vasculaire connectiviteit expliciet representeren. De gewichten van de lijnen kunnen worden geleerd via aandachtslagen gebaseerd op grafen, die fungeren als een dynamische matrix die de affiniteiten tussen knopen langs de vaten weergeeft. De gewogen knoopkenmerken kunnen als condities worden geïnjecteerd in het diffusiemodel om de generatie van een continue vaatsegmentatie te bevorderen. Het voorgestelde model is geëvalueerd op twee publieke datasets en overtreft andere state-of-the-art medische segmentatiemethoden wat betreft Dice-coëfficiënt en connectiviteitsmaten.

**Hoofdstuk 4** introduceert een top-k MIP-prior, die meer diepte-informatie behoudt dan de oorspronkelijke MIP, voor de segmentatie van levervaten. De voorgestelde top-k MIP-prior representeert de vaatconnectiviteit via een reeks 2D-projectieweergaven. De diameterinformatie van de vaten blijft behouden door gebruik te maken van de top-k maxima van CT-intensiteiten. Om de connecties te leren tussen de top-k MIP en de integrale projecties van de referentivaatmaskers, maken wij gebruik van een 2D latent diffusiemodel. De top-k MIP-prior wordt toegepast als conditie voor het latent diffusiemodel, dat integrale projecties van de vaatboom genereert. De gegenereerde integrale projecties kunnen worden gereconstrueerd met behulp van de gefilterde terug-projectie methode, die veelvuldig wordt toegepast binnen het CT-reconstructiedomein. De evaluatieresultaten tonen aan dat de voorgestelde prior bijdraagt aan de continuïteit van de vaatsegmentatie en vergelijkbaar presteert met andere medische segmentatie-benchmarks.

**Hoofdstuk 5** richt zich op de Couinaud-segmentatie van de lever zonder gebruik te maken van expliciete voorkennis van de vaatstructuur. Ter compensatie van het ontbreken van expliciete vasculaire structuren introduceren wij 3D graafredenering om de anatomische structuren in de lever dynamisch te leren. Bovendien passen wij voor deze taak een punt-gebaseerde architectuur toe om de fysieke resolutie van de CT-beelden te behouden. De geleerde impliciete anatomische structuren in de lever kan fungeren als vervanging voor expliciete voorkennis van de vaten waarin het punt-gebaseerde model geëmbod kan worden. De voorgestelde punt-gebaseerde

methode met graafredenering overtreft andere punt-gebaseerde benchmarks, in zowel de Dice-coëfficiënten als de afstand tussen de oppervlakten, terwijl er geen voorkennis van de levervaten meer nodig is voor de Couinaud-segmentatie.

## 7.2 Discussie en toekomstig onderzoek

Het overkoepelende doel van dit proefschrift was het ontwikkelen van nieuwe deep learning-gebaseerde benaderingen die belangrijke uitdagingen aanpakken en de stand van de techniek op beide onderzoeksgebieden aanzienlijk vooruitbrengen. Dit werd gerealiseerd door het ontwikkelen van automatische methoden voor een niet-invasieve detectie van coronaire plaques en analyse van anatomische structuren in de lever, waaronder segmentatie van levervaten en Couinaud-segmentatie van de lever. Allereerst werd in Hoofdstuk 2 de detectie van coronaire plaques onderzocht. Wij ontwikkelden twee deep learning-methoden voor de detectie en karakterisering van coronaire plaques op dwarsdoorsnedebeelden van CTA-beelden. Hoewel de voorgestelde methoden de conventionele HU-gebaseerde aanpak overtreffen volgens kwantitatieve maatstaven, blijven er diverse uitdagingen bestaan bij de detectie van coronaire plaques. Het overlappende HU-bereik tussen vetrijke plaques en het omringende myocardweefsel maakt een nauwkeurige afbakening van het plaquegebied bijzonder uitdagend. Verkalkte plaques met hoge intensiteiten kunnen eveneens artefacten veroorzaken, wat resulteert in schaduwen met lage-intensiteitswaarden die sterk lijken op vetrijke plaques. Bovendien vormen heldere-artefacten, veroorzaakt door geïmplanteerde metalen stents, een extra uitdaging aangezien zij de plaquekenmerken kunnen verhullen. Huidige methoden presteren onvoldoende bij CTA-beelden die door dergelijke artefacten worden beïnvloed; daarom zijn coronaire arteriën met geïmplanteerde stents uitgesloten van de experimenten. De meeste van de bestaande uitdagingen hangen nauw samen met de beeldkwaliteit van een CTA. In het bijzonder of de kenmerken van verschillende plaquetypen duidelijk van elkaar te onderscheiden zijn. Hoewel de resultaten die in Hoofdstuk 2 worden gepresenteerd aantonen dat deep learning-methoden het potentieel hebben om meer impliciete kenmerken uit CTA-beelden te extraheren en daarmee een nauwkeurigere plaquedetectie te realiseren dan de conventionele HU-gebaseerde methode, zullen wezenlijke verbeteringen afhangen van geavanceerde CT-technieken, zoals foton-tellende CT (PCCT) [1]. Het ontwikkelen van deep learning-methoden op basis van de PCCT-modaliteit zal een belangrijke richting vormen voor toekomstig onderzoek.

In Hoofdstukken 3, 4 en 5 werd de analyse van de anatomische structuren in de lever onderzocht. Hoofdstukken 3 en 4 richten zich primair op de continue segmentatie van levervaten, terwijl Hoofdstuk 5 zich concentreert op de Couinaud-segmentatie zonder gebruik te maken van voorkennis van de vaatstructuur. Zowel de segmentatie van levervaten als Couinaud-segmentatie vormen essentiële stap-

pen in de preoperatieve planning van leverchirurgie. De uitdagingen van deze twee taken verschillen echter: de connectiviteit is belangrijk bij vaatsegmentatie, terwijl een nauwkeurige begrenzing essentieel is voor de Couinaud-segmentatie. De twee voorgestelde methoden in Hoofdstukken 3 en 4 gaan uit van verschillende beeldweergaven (dwarsdoorsneden en projectieweergaven) en maken gebruik van op diffusiemodellen gebaseerde architecturen om de connectiviteit van de gesegmenteerde vaatboom te verbeteren. In tegenstelling tot de meeste bestaande segmentatiemethoden, die doorgaans gebaseerd zijn op discriminerende modellen, kunnen diffusiemodellen de onderliggende dataverdeling en intrinsieke kenmerken leren. Deze voordelige eigenschap van het diffusiemodel benutten wij in onze taak voor vaatsegmentatie. Om de stochastiek van het diffusiemodel beter te beheersen en de gewenste te genereren output te sturen, maken wij in Hoofdstuk 3 gebruik van een aandachtsmechanisme gebaseerd op grafen als voorkennis en bootsen wij in Hoofdstuk 4 het CT-reconstructieproces na. Deze hoofdstukken tonen aan dat deze strategieën allemaal bijdragen aan het verbeteren van de vaatconnectiviteit in 3D. De kwantitatieve en kwalitatieve resultaten laten zien dat de vaatconnectiviteit aanzienlijk kan worden verbeterd in vergelijking met andere geavanceerde segmentatiemethoden. Huidige methoden kennen echter twee belangrijke beperkingen. Ten eerste zijn de voorgestelde modellen gesuperviseerd, waardoor hun prestaties nadelig kunnen worden beïnvloed door onnauwkeurige vaatannotaties. Ten tweede is de inferentie met diffusiemodellen reken technisch duurder in vergelijking met andere onderscheidende segmentatiemethoden. Toekomstig onderzoek zou zich daarom moeten richten op het wegnemen van deze twee beperkingen. Semi-gesuperviseerde methoden kunnen mogelijk helpen om de negatieve impact van onnauwkeurige vaatannotaties te verminderen. Daarnaast kan de ontwikkeling van meer geavanceerde wiskundige raamwerken voor diffusiemodellen bijdragen aan een hogere efficiëntie van zowel de training als in het toepassen.

Couinaud-segmentatie, de tweede focus van de analyse van de anatomische structuren in de lever, wordt eveneens geconfronteerd met de uitdaging van onnauwkeurige vaatannotaties, aangezien de conventionele methoden die in de klinische praktijk worden toegepast sterk afhankelijk zijn van voorkennis van de vaatstructuur om de grenzen tussen de leversegmenten te bepalen. Deze uitdaging wordt aangepakt door onze voorgestelde voorkennis-vrije methode die in Hoofdstuk 5 wordt gepresenteerd. Wij passen graafredenering toe om de impliciete anatomische structuren in de lever te leren, in plaats van te vertrouwen op expliciete kennis over de levervaten. Daardoor is de voorgestelde methode onafhankelijk van de kwaliteit van vaatannotaties en wordt de noodzaak van een arbeidsintensief annotatieproces voor vaten geëlimineerd. Geen van de voorgestelde of vergeleken baselines kan echter effectief omgaan met levers die grote tumoren bevatten. Dit komt doordat delen van de anatomische structuren in de

lever door de tumor zijn aangetast, waardoor het voor het neuraal netwerk moeilijk wordt anatomische kenmerken in de door de tumor bezette regio's te herkennen. Toekomstig onderzoek zou zich daarom kunnen richten op strategieën om ontbrekende anatomische structuren te compenseren. Een potentiële oplossing is om het leren van langeafstandsafhankelijkheden te versterken, doordat de regio's met tumoren zonder anatomische kenmerken kunnen worden geassocieerd met aangrenzende tumorvrije regio's.

Alle hoofdstukken in dit proefschrift richten zich op de twee hoofdthema's: de detectie van coronaire plaques en de analyse van anatomische structuren in de lever op basis van CTA/CT-beelden. Hoewel alle voorgestelde methoden zijn ontwikkeld op de CT-modaliteit, kunnen zij in potentie worden toegepast op vergelijkbare klinische taken met andere beeldvormingsmodaliteiten.

### 7.3 Algemene conclusies

In dit proefschrift worden deep learning-methoden geïntroduceerd in twee belangrijke klinische domeinen: niet-invasieve detectie van coronaire plaques en analyse van anatomische structuren in de lever op basis van CTA/CT-beelden. In het bijzonder verkennen wij eerst de connectie van plaquekenmerken tussen invasieve beelden en niet-invasieve CTA-beelden in dwarsdoorsnedeweergaven, waarmee een nauwkeurigere en fijnmazige op hoeken gebaseerde plaquedetectie wordt uitgevoerd. Daarnaast benutten wij op efficiënte wijze de voordelen van verschillende modellen, zoals diffusiemodellen, graaf gebaseerde neurale netwerken en punt-gebaseerde modellen, en passen deze doelgericht aan om continue segmentatie van levervaten te faciliteren. Dit neemt de noodzaak van voorkennis van de vaatstructuur bij Couinaud-segmentatie weg. Alle in dit proefschrift voorgestelde deep learning-methoden zijn zowel kwantitatief als kwalitatief gevalideerd en tonen superieure prestaties aan in vergelijking met andere toonaangevende methoden.

## Referenties

- [1] A. Meloni, F. Frijia, D. Panetta, et al. "Photon-counting computed tomography (PCCT): technical background and cardio-vascular applications". In: *Diagnostics* 13.4 (2023), page 645.

# List of publications

## Journal articles

**Xiaotong Zhang**, Alexander Broersen, Hessam Sokooti, Anantharaman Ramasamy, Pieter Kitslaar, Ramya Parasa, Medeni Karaduman, Ameer Soudeed Ali Jan Mohammed, Christos V. Bourantas, and Jouke Dijkstra. "Cross-sectional angle prediction of lipid-rich and calcified tissue on computed tomography angiography images." *International Journal of Computer Assisted Radiology and Surgery*, 19, no. 5 (2024): 971-981.

**Xiaotong Zhang**, Alexander Broersen, Gonnie van Erp, Silvia L. Pintea, and Jouke Dijkstra. "Continuous and complete liver vessel segmentation with graph-attention guided diffusion." *Knowledge-Based Systems*, 331 (2025), Article 114686.

Anantharaman Ramasamy, Hessam Sokooti, **Xiaotong Zhang**, Evangelia Tzorovili, Retesh Bajaj, Pieter Kitslaar, Alexander Broersen, Rajiv Amersey, Ajay Jain, Mick Ozkor, Johan H C Reiber, Jouke Dijkstra, Patrick W Serruys, James C Moon, Anthony Mathur, Andreas Baumbach, Ryo Torii, Francesca Pugliese, Christos V Bourantas. "Novel near-infrared spectroscopy-intravascular ultrasound-based deep-learning methodology for accurate coronary computed tomography plaque quantification and characterization." *European Heart Journal Open*, 3, no. 5 (2023): oead090.

Anantharaman Ramasamy, Ramya Parasa, Hessam Sokooti, **Xiaotong Zhang**, Ibrahim Halil Tanboga, Pieter Kitslaar, Alexander Broersen, Krishnaraj S Rathod, Rajiv Amersey, Ajay Jain, Mick Ozkor, Johan H C Reiber, Jouke Dijkstra, Patrick W Serruys, James C Moon, Anthony Mathur, Ryo Torii, Francesca Pugliese, Andreas Baumbach, Christos V Bourantas. "Computed tomography versus near-infrared spectroscopy for the assessment of coronary atherosclerosis." *EuroIntervention*, Dec 2;20(23) (2024):e1465-e1475.

Nathan Angelo Lecaros Yap, Anantharaman Ramasamy, Ibrahim Halil Tanboga, Xingwei He, Murat Cap, Retesh Bajaj, Medeni Karaduman, Ajay Jain, Pieter Kitslaar, Alexander Broersen, **Xiaotong Zhang**, Hessam Sokooti, Johan HC Reiber, Jouke Dijkstra, Mick Ozkor, Patrick W Serruys, James C Moon, Anthony Mathur, Andreas Baumbach, Ryo Torii, Francesca Pugliese, Christos V Bourantas. "Implications of coronary calcification on the assessment of plaque pathology: a comparison of computed tomography and multimodality intravascular imaging." *European Radiology*, 35, no. 4 (2025): 1745-1760.

Akash Sivananthan, Ibrahim Tanboga, Thamil Kumaran S K, Ameer Mohammed,

Anantharaman Ramasamy, Andreas Kalogeropoulos, Mick Ozkor, Gonul Zeren, Nathan A.L. Yap, Xingwei He, Patrick Serruys, James C Moon, Anthony Mathur, Alexander Broersen, Pieter Kitslaar, **Xiaotong Zhang**, Johan H.C. Reiber, Jouke Dijkstra, Francesca Pugliese, Andreas Baumbach, Ryo Torii, Christos V Bourantas. "Deep-learning analysis of computed tomography coronary angiography data enables more accurate computation of the shear stress distribution than conventional analysis by experts: A head-to-head comparison with near-infrared spectroscopy-intravascular ultrasound-based modelling." *Journal of Cardiovascular Computed Tomography*, Oct 14 (2025): S1934-5925(25)00445-9.

### International conference proceedings

**Xiaotong Zhang**, Alexander Broersen, Gonnie CM Van Erp, Silvia L. Pintea, and Jouke Dijkstra. "Top-K Maximum Intensity Projection Priors for 3D Liver Vessel Segmentation." In *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*, pp. 1-5. IEEE, 2025.

**Xiaotong Zhang**, Alexander Broersen, Gonnie CM Van Erp, Silvia L. Pintea, and Jouke Dijkstra. "Skip priors and add graph-based anatomical information, for point-based Couinaud segmentation." In *Reconstruction and Imaging Motion Estimation, and Graphs in Biomedical Image Analysis: First International Workshop, RIME 2025, and 7th International Workshop, GRAIL 2025, Daejeon, South Korea, September 27, 2025, Proceedings*. Springer-Verlag, Berlin, Heidelberg, 131-140.

### Journal abstract

Anantharaman Ramasamy, Hessam Sokooti, **Xiaotong Zhang**, Evangelia Tzorovili, Retesh Bajaj, Pieter Kitslaar, Rajiv Amersey et al. "TCT-250 A Novel Coronary Computed Tomography Angiography Deep-Learning Methodology for Coronary Atheroma Quantification and Characterization Trained Using Near-Infrared Spectroscopy-Intravascular Ultrasound." *Journal of the American College of Cardiology* 80, no. 12\_Supplement (2022): B98-B99.

Anantharaman Ramasamy, Hessam Sokooti, **Xiaotong Zhang**, Evangelia Tzorovili, Retesh Bajaj, Pieter Kitslaar, Alexander Broersen et al. "OP1 A novel coronary computed tomography angiography deep-learning methodology for coronary atheroma assessment trained using near-infrared spectroscopy-intravascular ultrasound." *Heart*, 108, no. Suppl 2 (2022): A1-A1.

# Acknowledgements

I still remember a writing topic from my primary school days: “How would people’s lives change by 2020?” At that time, I believed that life in 2020 – a year that seemed so far away – would be like what was described in science fiction. When 2020 finally arrived, life didn’t seem as fancy as I had imagined back in primary school, but the year turned out to be personally meaningful to me. I finished my master’s degree, moved from student life to working life, and later applied for a PhD – a chance to return to being a student once more. I am grateful to LKEB for accepting my application, which made this thesis possible.

I would like to thank my promoter, Prof. Boudewijn Lelieveldt, supervisor, Dr. Jouke Dijkstra, co-supervisor Dr. Alexander Broersen, and advisor Dr. Silvia L. Pintea for their supportive and open academic atmosphere, insightful guidance, and patience.

Boudewijn, thank you for your effective and supportive guidance throughout the entire process – from my first registration as a graduate student to my final thesis and defense.

Jouke, I am deeply grateful for the inspiring weekly discussions, your support for my ideas, and your patient and detailed guidance that continually motivated me in my research.

I truly appreciate Alex and Silvia for the detailed discussion on experimental design and implementation, their patient guidance on paper revisions and rebuttals, and their inspiring advice throughout my academic journey.

I would also like to thank the wonderful colleagues at LKEB for their kindness and companionship. Chang, it was such a pleasure being your neighbor and office mate over the past few years. Your cat brought me so much joy, and our relaxed conversations were always a comforting reminder of your kindness and support. Yanli, thank you for organizing so many wonderful gatherings and for making our trip to Iceland with Xiaowu such an unforgettable experience. Xiaowu, I am really glad I had you as a friendly senior when I started my PhD and always enjoyed our talks about research and future careers. Jingnan, thank you for organizing the summer boat event and for hosting countless *UNO* nights at your place. Yunjie, thank you for always bringing your delicious signature dishes to our gatherings. Qiuyu, I am really happy we got to explore some museums and enjoy meals together. Donghang, I am glad we got to attend MICCAI in South Korea together and enjoy the amazing local food. Ruochen, I enjoyed taking the computer vision course and working on the group project with you in Amsterdam. Li-Hsin, thank you for answering my questions about the defense process – it really saved me a lot of time. Faeze, I truly appreciated our discussions

during the weekly progress meetings as well as the casual conversations we shared. Also, many thanks to Bo, Chinmay, Mody, Vincent, Viktor, Laurens, Jenia, Gonnies, Ruixin, Alexander, Soumyadeep, Navid, Marius, Oleh, Rob, Berend, Els, Jeroen, Efe, Niels, Michèle, Patrick, Baldur, Denis, and Julian for their help and kindness.

I extend my sincere thanks to my dear friends. Xinyu, thank you for taking me to explore London so many times. Xinming, I am very happy that we reconnected in Leiden after finishing our undergraduate studies. Jingyi, I am really glad we spent the summer internship together and visited Shanghai Disneyland. Hongju, Sijia, Yuan, and Qian, thank you for our always-enjoyable conversations despite the seven-hour time difference.

I would like to thank my family for their constant love and support. I am especially grateful to my parents for their unconditional support, and to my aunt and cousin for their care and encouragement. You will always be my safe haven.

# Curriculum Vitae

Xiaotong Zhang was born in Jining, Shandong Province, China, in August 1995. She obtained her high school diploma in June 2013. In September 2013, she began her undergraduate studies in Biomedical Engineering at Shandong First Medical University and graduated in June 2017. In September 2017, she began her master's studies in Biomedical Engineering at Northeastern University and obtained her degree in January 2020. During her master's studies, she received the National Graduate Scholarship in 2019. After obtaining her master's degree, she worked in industry as an algorithm development engineer until July 2021. In September 2021, she began her PhD studies at Leiden University Medical Center in the Netherlands. Her PhD research focuses on the characterization of vascular plaques and vessel segmentation using deep learning techniques.

