UNIVERSITY OF LEEDS

Vessel Tree Segmentation and Modality Agnostic Aneurysm Detection

Fengming Lin

University of Leeds

School of Computing

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy

February, 2025

Intellectual Property Statement

This candidate confirms that the work submitted is his own, except where work which has formed part of jointly authored publications has been included. The candidate confirms that appropriate credit has been given within the thesis, where reference has been made to the work of others. The contribution of the candidate and other authors to this work is explicitly indicated in the following.

I am the main author of all the following publications. I led the design of the study and wrote the manuscript, including analysis and discussion of the results. I also wrote the analysis code and carried out the data processing, statistical analysis, and experiments. The contribution of other authors was to help me design the study, provide data annotations, discuss the results, and review the manuscript.

- **Chapter 3:**

  Lin F, Xia Y, Song S, Ravikumar N, Frangi AF. High-throughput 3DRA segmentation of brain vasculature and aneurysms using deep learning. Computer Methods and Programs in Biomedicine. 2023 Mar 1;230:107355. [1]

- **Chapter 4:**

  Lin F, Xia Y, Ravikumar N, Liu Q, MacRaild M, Frangi AF. Adaptive semi-supervised segmentation of brain vessels with ambiguous labels. International Conference on Medical Image Computing and Computer-Assisted Intervention 2023 Oct 8 (pp. 106-116). Cham: Springer Nature Switzerland. [2]

- **Chapter 5:**

  Lin F, Xia Y, Deo Y, MacRaild M, Dou H, Liu Q, Wu K, Ravikumar N, Frangi AF. Unsupervised Domain Adaptation for Brain Vessel

Segmentation Through Transwarp Contrastive Learning. In2024 IEEE International Symposium on Biomedical Imaging (ISBI) 2024 May 27. IEEE. [3]

- **Chapter 6:**

  Lin F, Xia Y, MacRaild M, Deo Y, Dou H, Liu Q, Cheng N, Ravikumar N, Frangi AF. GS-EMA: Integrating Gradient Surgery Exponential Moving Average with Boundary-Aware Contrastive Learning for Enhanced Domain Generalization in Aneurysm Segmentation. In2024 IEEE International Symposium on Biomedical Imaging (ISBI) 2024 May 27. IEEE. [4]

Acknowledgements

I want to thank my main supervisor, Professor Alejandro Frangi, for allowing me to pursue a PhD degree and work alongside him at CISTIB. I am grateful for his enthusiastic leadership, continuous constructive criticism, and friendly advice, all of which have helped me become a better researcher. More importantly, I have learned from him the importance of having a deep passion for work and dedication to the field.

I am very grateful to all my lab mates over the last four years in Leeds for all the happy memories. Special thanks to my co-supervisors, Dr Yan Xia, Dr Nishant Ravikumar and Dr Arezoo Zakeri, for the numerous friendly and fruitful meetings and guidance.

I appreciate my collaborators and friends, Fangjun, Shuang, Yash, Qiongyao, Haoran, Ning, Nina, Kun, Xiang, Michael, Fergus and Ali, for your shared passion and for making my PhD experience more enjoyable.

I also thank our software engineers, Alejandro and Kattia, for all the technical support throughout different parts of the project.

Finally, I am deeply grateful to my parents and grandparents for their unflagging love and ongoing support.

Abstract

Segmentation of cerebral vessels and aneurysms is vital for diagnosing cerebrovascular conditions, developing treatment plans, and supporting in silico trials. The intricate nature of cerebrovascular anatomy and the small and irregular characteristics of aneurysms pose significant challenges for obtaining high-precision segmentation, which is vital for accurate clinical evaluations and planning of interventions.

The motivation stems from the necessity to address several critical challenges that compromise the efficiency of existing segmentation models. These challenges include class imbalance, where smaller structures, like aneurysms, are often under-represented in datasets, leading to suboptimal segmentation performance. Moreover, the insufficiency of labelled data constrains the potential of fully supervised models. In addition, the issue of domain shifts across various imaging modalities and patient populations can cause models trained on one dataset to perform inadequately on another. Finally, there is a need for models that can generalise effectively across diverse datasets from various clinical data sources, ensuring consistent performance regardless of the data source.

In response to these challenges, this thesis presents several novel contributions. It introduces a 3D patch-based multi-class model that effectively manages class imbalance and inter-class interference in vessel and aneurysm segmentation, employing advanced network elements and paired preprocessing and postprocessing techniques to enhance accuracy. A semi-supervised learning approach is also developed to leverage both labelled and unlabelled data, significantly improving segmentation consistency and continuity, particularly in scenarios with limited annotations. Furthermore, the thesis proposes a transwarp contrastive learning framework for unsupervised domain adaptation, allowing the model to handle domain shifts and perform robustly across different data modalities. Finally, a gradient-based domain generalisation method is introduced to ensure that segmentation models can generalise well across various imaging conditions, overcoming the challenges posed by data source variability.

Evaluation results demonstrate that the proposed methods significantly improve segmentation performance across multiple benchmarks. The 3D

patch-based multi-class model achieves a significant improvement in aneurysm segmentation, achieving a Dice score of 81.63% on the Aneurist-3DRA dataset, which is agnostic to data sources and aneurysm sizes. This performance notably surpasses the best state-of-the-art method, nnUNet, which achieves a Dice score of 66.86%, demonstrating the effectiveness of the proposed approach. The semi-supervised learning approach significantly enhances vessel segmentation accuracy while reducing annotation requirements, maintaining robust performance even with limited labeled data. On Aneurist-3DRA dataset with a resolution of 0.35 mm/pixel, this method achieves a surface distance of 0.2075 mm. The transwarp contrastive learning framework effectively mitigates domain shifts, ensuring consistent segmentation quality across diverse imaging modalities. On the MRA dataset, it achieves a Dice score of 72.65% for vessel segmentation, demonstrating its robustness in handling cross-domain variations. Finally, the gradient-based domain generalisation method enhances the ability of model to generalise to unseen datasets, reducing performance degradation when applied to new clinical settings. Trained on only a small subset of data from a single data center within the Aneurist-3DRA dataset, the model achieves a Dice score of 71.89%, demonstrating its strong generalisation capability across diverse clinical scenarios. Collectively, these contributions advance the field of cerebrovascular image segmentation by addressing key challenges and improving model adaptability and robustness.

Contents

List of Figures

List of Tables

Abbreviations

DL	Deep Learning
CNN	Convolutional Neural Network
3DRA	3D Rotational Angiography
MRA	Magnetic Resonance Angiography
MRI	Magnetic Resonance Imaging
CTA	Computerised Tomography Angiography
DSA	Digital Subtraction Angiography
UDA	Unsupervised Domain Adaptation
DG	Domain Generalisation
FDA	Fourier Domain Adaptation
EMA	Exponential Moving Average
HSDA	Homocentric Squares Domain Adaptation
EMA	Exponential Moving Average
GS	Gradient Surgery
CL	Contrastive Learning
VS	Volume Similarity
DSC	Dice Similarity Index
CE	Cross Entropy
DCE	Dice Cross Entropy
TP	True Positives
FP	False Positives
TN	True Negatives
FN	False Negatives
ROI	Region of Interest
SOTA	State Of The Art

Chapter 1

Introduction: Background, Motivation and Contribution

1.1 Background on Brain Vessel Structure, Imaging, and Disease

1.1.1 Brain Vessel Structure

Brain vessels [7] refer to the vascular system that supplies and drains brain tissue, including the arteries, veins, and capillaries. The primary function of the brain vascular system is to deliver oxygen and nutrients to brain tissue and to remove metabolic waste and carbon dioxide. The blood supply to the brain comes primarily from the internal carotid artery system and the vertebrobasilar artery system. In the internal carotid artery system, blood flows from the internal carotid artery (ICA) to the brain, then to the anterior cerebral artery (ACA), the middle cerebral artery (MCA), and others. These arteries supply different regions of the brain. In the vertebrobasilar artery system, blood flows from the basilar artery into the brain, then to the posterior cerebral artery (PCA), among others. The anterior cerebral arteries of the left and right sides of the internal carotid artery system are connected by the anterior communicating artery (AComA), and the internal carotid artery from the internal carotid artery system is connected to the posterior cerebral artery from the vertebrobasilar artery system via the posterior communicating artery (PComA).

The Circle of Willis [8] connects the main cerebral arteries (ACA, MCA, PCA) through the anterior and posterior communication arteries. The Circle of Willis forms a redundant blood supply system, allowing blood to be redistributed in case of arterial blockage or narrowing, thereby maintaining the blood supply to various brain regions.

Regarding the main functions of these vessels, the internal carotid artery and the basilar artery are responsible for delivering blood from the heart to the brain [9]. The anterior cerebral artery supplies the midline region of the frontal part of the brain and part of the parietal lobes. The middle cerebral artery supplies the lateral surface of the brain. The posterior cerebral artery supplies the occipital lobe and the medial aspect of the temporal lobe in the back of the brain. The anterior communicating artery and the posterior communicating artery primarily connect the arteries, providing collateral circulation to the brain to prevent ischemic damage caused by occlusion of a single artery [10].

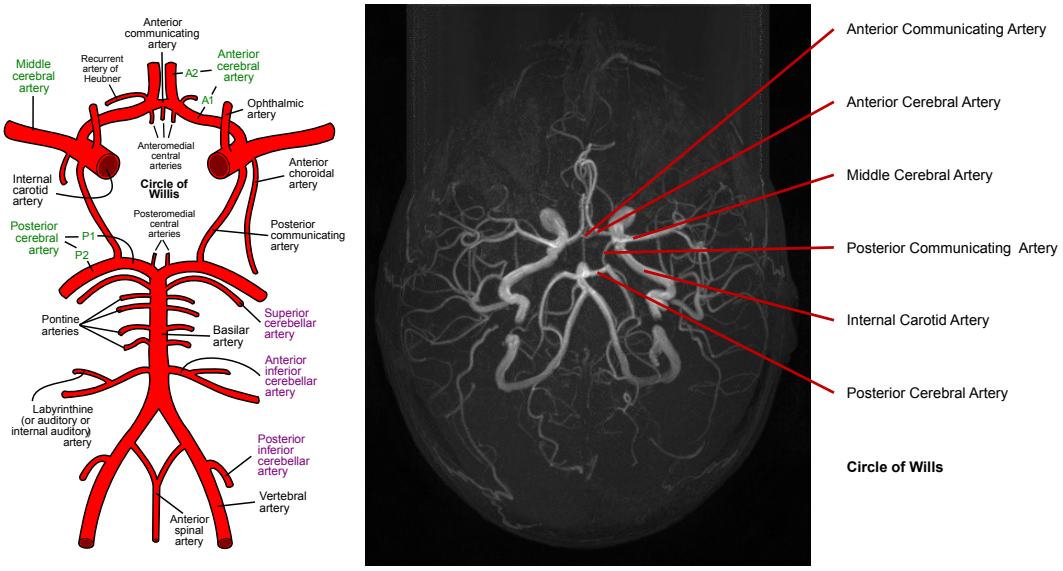

Figure 1.1: Anatomical and imaging visualisation of the Circle of Willis. The left image depicts a schematic of the Circle of Willis, showing key arteries like the Anterior Cerebral, Internal Carotid, and Posterior Cerebral Arteries. The right image is an MRA scan showing the Circle of Willis in a real human brain, with the same arteries labelled. This comparison highlights both the anatomical structure and its clinical significance in assessing blood flow and identifying potential vascular issues.

1.1.2 Aneurysm and Corresponding Analysis

A cerebral aneurysm [11, 12, 13, 14] is a localised abnormal bulge or weakening of the blood vessel wall [15]. There are various types of aneurysm [16], including the most common saccular aneurysm [17], which presents as a balloon-shaped bulge on one side of the blood vessel. Another common type is the fusiform aneurysm, where the entire circumference of the blood vessel wall expands in a spindle-like shape. When an aneurysm ruptures, it can cause severe brain haemorrhage or subarachnoid haemorrhage (SAH), which is known as a ruptured aneurysm [18, 19].

The causes of aneurysms [20] are complex and varied, including congenital factors such as poor development or structural abnormalities of the vessel wall and acquired factors such as atherosclerosis, hypertension, and infection. The occurrence of aneurysms is also influenced by various other factors, such as family history, smoking,

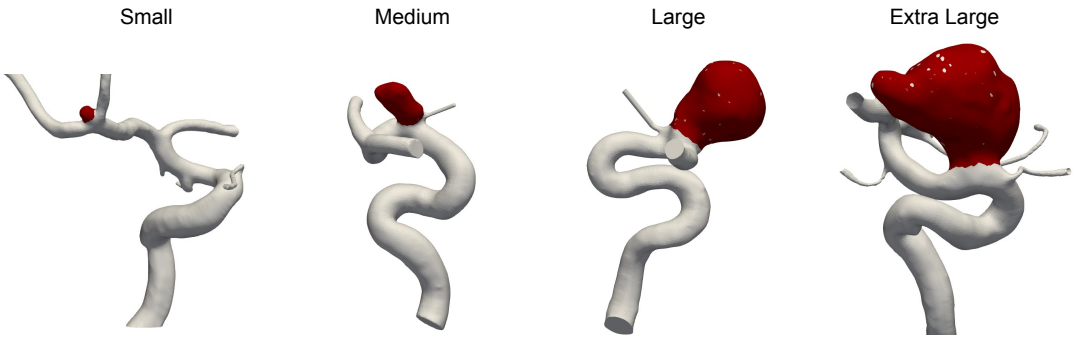

Figure 1.2: Variations in aneurysm sizes across different patients. This figure showcases different aneurysm sizes, including small, medium, large, and extra large, observed in various patients' cerebral arteries. Each model represents a distinct case, emphasising the variability in aneurysm dimensions. The red regions indicate the aneurysms, with larger sizes typically associated with higher risks of complications, such as rupture. The large variation in the size of various aneurysms reinforces the importance of tailoring the treatment strategy to the patient's specific situation.

hypertension, and demographic factors, such as age and gender (with women at higher risk than men).

The prevalence of aneurysms in the general population is approximately 3 % [21]. Many aneurysms are asymptomatic before rupture and are often discovered incidentally during health check-ups. Furthermore, non-ruptured aneurysms, due to their abnormal size and structure in Fig 1.2, may also cause symptoms by compressing surrounding tissues and causing local neurological dysfunction.

Quantitative and qualitative analyses in Fig. 1.3 are essential for evaluating and diagnosing aneurysms. Quantitative analysis includes measurements of the aneurysm's diameter, volume, shape, and growth rate. Typically, aneurysms can be classified by size [22]. Aneurysms smaller than 5 millimetres in diameter are classified as small aneurysms [23], which are the most common and have a lower risk of rupture. Aneurysms between 5 and 10 millimetres are medium aneurysms, requiring individualised treatment based on the patient's specific conditions (such as family history, hypertension, etc.). Aneurysms larger than 10 millimetres are classified as large aneurysms and carry

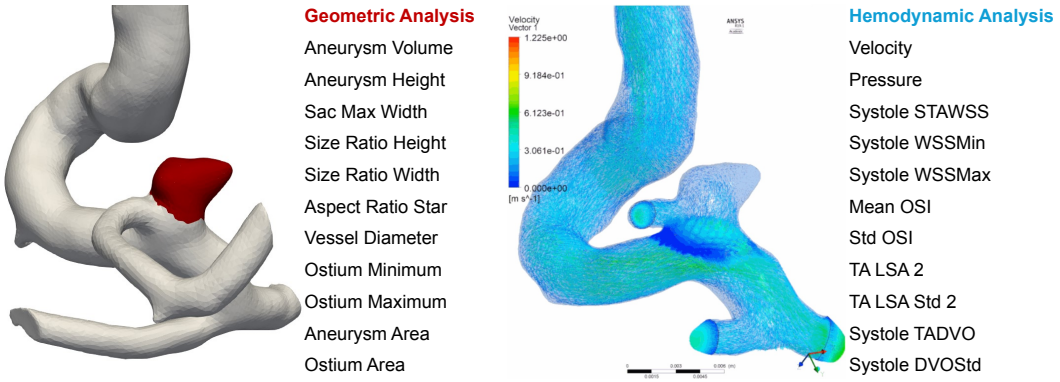

Figure 1.3: Quantitative factors for aneurysm analysis. This figure presents the key parameters used in the evaluation of cerebral aneurysms through geometric and hemodynamic analyses. On the left, the geometric analysis includes factors such as aneurysm volume, sac maximum width, and various size and aspect ratios, which help in assessing the aneurysm's physical dimensions and structural impact. On the right, the hemodynamic analysis examines factors like velocity, pressure, and wall shear stress, which are critical for understanding the blood flow dynamics within and around the aneurysm. These quantitative measures provide a comprehensive approach to assessing the risks associated with aneurysms and inform treatment decisions.

a higher risk of rupture, requiring a more aggressive diagnosis and treatment management, including possible surgical intervention. In particular, aneurysms larger than 25 millimetres are classified as giant aneurysms and almost always require immediate surgical treatment. These size-based classification standards help physicians assess the rupture risk of the aneurysm and formulate appropriate treatment plans.

Qualitative analysis [24] includes the location of the aneurysm, hemodynamic analysis, and other factors. Computational fluid dynamics (CFD) can evaluate factors such as internal blood flow velocity, pressure, shear stress, and other factors within the aneurysm. This information is crucial to understanding the formation of aneurysms, providing treatment recommendations, and simulating the performance of implanted devices. Managing aneurysms involves qualitative and quantitative analyses and depends on demographic factors such as the patient's age and health condition.

In Fig 1.3, Systole STAWSS (Spatially Time-Averaged Wall Shear Stress) refers

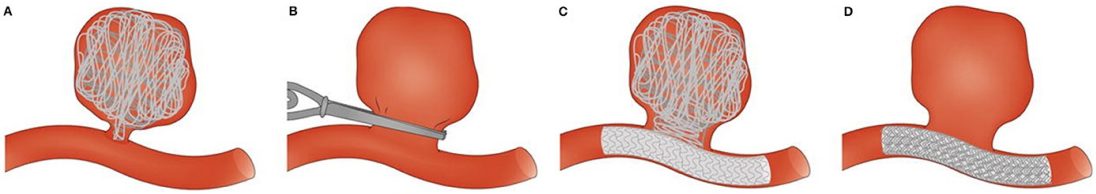

Figure 1.4: Diagrams of different aneurysm treatment options [5]. (A) Endovascular coiling of the aneurysm sac. (B) Surgical clipping of the aneurysm neck. (C) The endovascular treatment combines the use of coils and a stent. (D) Endovascular treatment with a flow diverter.

to the time-averaged shear stress on the arterial wall during the systolic phase of the cardiac cycle. Systole WSSMin (Minimum Wall Shear Stress) and Systole WSSMax (Maximum Wall Shear Stress) represent the minimum and maximum shear stress during systole. The mean OSI (Oscillatory Shear Index) indicates the average oscillatory shear index over time, whereas the Std OSI (Standard Deviation of Oscillatory Shear Index) represents the standard deviation of these oscillations. TA LSA 2 (Time-Averaged Local Shear Angle) refers to the time-averaged local shear angle. TA LSA Std 2 (Standard Deviation of Time-Averaged Local Shear Angle) is its standard deviation. Systole TADVO (Time-Averaged Directional Velocity Oscillation) represents the time-averaged directional velocity oscillation during systole. At the same time, Systole DVOStd (Standard Deviation of Directional Velocity Oscillation) indicates the standard deviation of this oscillation [25].

There are various methods for the treatment and management of aneurysms [26]. Monitoring with regular imaging studies is an option to track any changes in small asymptomatic aneurysms. Treatments aimed at managing hypertension and reducing blood lipid concentrations can help slow the progression of aneurysms. For aneurysms requiring surgical intervention, clipping and endovascular treatment are the two main methods. Clipping involves a craniotomy to place a clip in the neck of the aneurysm, preventing blood flow to it. Endovascular treatment involves using a catheter to insert embolic materials (such as coils or stents) into the aneurysm, similarly preventing blood flow into the aneurysm.

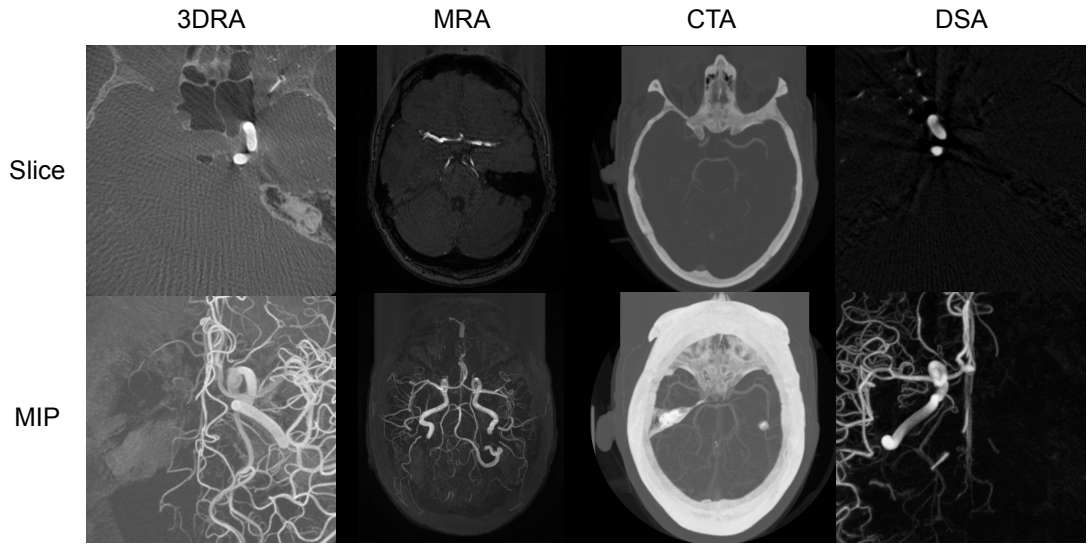

Figure 1.5: Comparative imaging modalities for brain vessel and aneurysm analysis. This figure illustrates the use of four different imaging modalities, including 3DRA, MRA, CTA, and DSA, to visualise brain vessels and aneurysms. The top row presents cross-sectional image slices that highlight the detailed anatomical structures captured by each modality. The bottom row shows maximum intensity projections (MIP), which emphasise the vascular structures and provide a clearer view of the aneurysms. Each modality offers unique advantages in resolution, contrast, and spatial details.

1.1.3 Brain Vessel Imaging Modalities

The diagnosis analysis and treatment planning for vascular disease utilise various imaging modalities, each with specific strengths and limitations. Fig. 1.5 illustrates four key techniques that provide different perspectives and levels of detail. This section discusses these modalities and outlines their unique advantages and clinical applications.

3D rotational angiography (3DRA) [27] is an advanced vascular imaging technique compared to other modalities. Generate high-resolution 3D vascular images by rotating a C-arm X-ray device and injecting a contrast agent. During surgery, doctors can constantly reposition and magnify aneurysms or other vascular structures, providing high-resolution local vascular structure information. The advantage of the 3DRA modality is its high resolution, making it particularly suitable for detecting complex

aneurysms and downstream image analysis tasks. However, 3DRA requires ionising radiation and contrast agents, which requires consideration of radiation exposure and the risks associated with patients who cannot use contrast agents.

Magnetic resonance angiography (MRA) [28] generates 3D vascular images through the interaction of magnetic fields and radio waves and is a unique sequence within Magnetic Resonance Imaging (MRI). The advantage of the MRA modality is that it is non-invasive and can be performed without the injection of contrast agents. Furthermore, MRA does not use ionising radiation, resulting in lower radiation exposure for patients, making it suitable for long-term follow-up and monitoring. However, the resolution of the MRA modality is relatively low, which is suitable for early screening but less effective for precise downstream quantitative and qualitative analysis.

Computed tomography angiography (CTA) [29] uses X-rays and computed tomography to obtain 3D vascular images by injecting contrast agents. CTA can quickly scan and obtain medium-resolution vascular images that clearly show the vascular anatomy. However, the high brightness of the bones in the CTA images can interfere with the quantitative and qualitative analyses. Also, CTA requires ionising radiation and contrast agents, so radiation exposure and contrast agent allergies must be considered.

Digital Subtraction Angiography (DSA) [30] is the gold standard for clinical diagnosis of aneurysms. DSA involves inserting a catheter and injecting a contrast agent to capture high-resolution, real-time dynamic images of blood vessels under X-ray. However, the resulting images are multi-angle 2D images, which makes DSA more suitable for clinical procedures rather than downstream analysis compared to other modalities.

1.1.4 InSilico Trials

In silico trials [6, 31] represent a sophisticated approach to medical research, employing advanced computer simulation and modelling techniques. These trials replicate the dynamics of human organs and tissues using high-performance computing and detailed physical models. They are particularly advantageous for evaluating the efficacy and safety of medical devices and therapeutic strategies. Compared to conventional in vivo (within a living organism) and in vitro (outside a living organism) experiments, in silico trials offer significant reductions in cost, duration, and invasiveness. They facilitate the rapid acquisition of precise and reproducible data from extensive population datasets. Through in silico trials, researchers can rigorously evaluate various hypotheses within

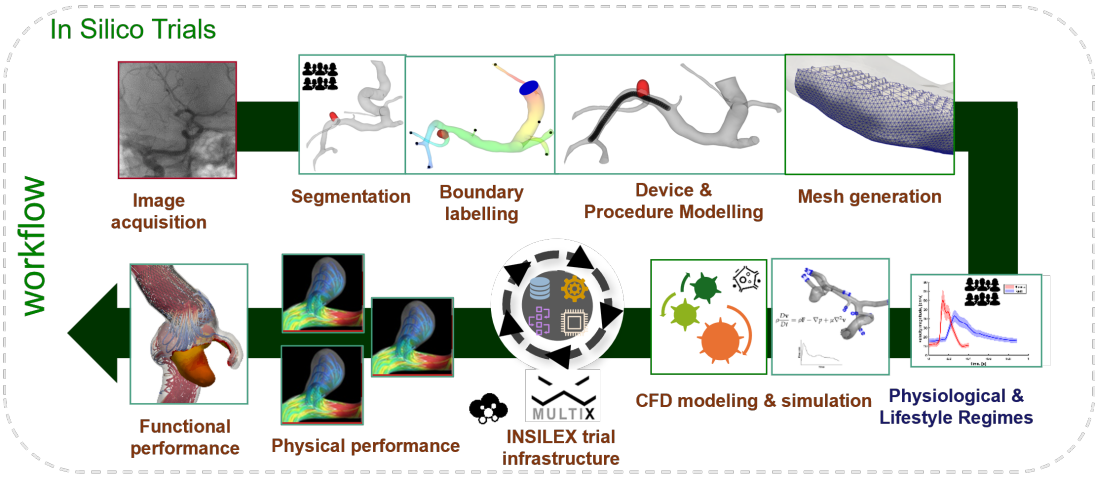

Figure 1.6: Workflow of in silico trials [6]. This illustration shows the key stages in the execution of in silico trials, beginning with the acquisition and segmentation of medical images to isolate vascular structures. The subsequent boundary labelling stage delimits the regions of interest, followed by the modelling of devices and procedures. Mesh generation is then used to render these models amenable for computational analysis, encompassing CFD modelling and simulation. The integration of physiological and lifestyle factors is critical for assessing device performance. The INSILEX trial infrastructure supports this workflow and facilitates precise and effective simulations for medical research.

a virtual framework, refine experimental designs, and mitigate risks and uncertainties associated with actual clinical trials.

In Fig. 1.6, in silico trials involve a comprehensive workflow that begins with image acquisition, where high-resolution vascular images are obtained using techniques such as 3DRA, MRA, CTA, and DSA. These images form the foundation for subsequent steps. Segmentation [1] uses automated algorithms to isolate vessel and aneurysm structures, generating accurate three-dimensional models. These models undergo boundary labelling to accurately define regions of interest, such as the aneurysm neck and vessel walls. Following this, device and procedure modelling simulates the interaction of medical devices, such as stents and coils, with vascular anatomy. The models are then

converted into computationally suitable mesh structures composed of three-dimensional elements representing the geometry of the vessels and aneurysms. Physiological and lifestyle regimes are integrated to simulate different conditions and assess the impact of lifestyle changes on vascular structures. CFD modelling and simulation [32, 33, 34] are used to evaluate hemodynamic characteristics, such as blood flow velocity, wall shear stress, and pressure distribution within vascular models. The INSILEX trial infrastructure supports these processes by integrating computational tools, data management systems, and collaborative platforms. Physical performance testing ensures the devices function correctly in the simulated environment, validated against real-world data. The final step, functional performance, assesses how well these devices perform under physiological conditions, completing the in silico trial process.

In silico trials offer significant advantages, including efficiency, non-invasiveness, and the ability to process large numbers of cases quickly, generating high-quality data for large-scale statistical analysis and personalised treatment planning. However, they also face challenges such as the need for specialised expertise, complex model construction, and substantial computational resources. The accuracy and verifiability of the models are crucial for the success of these trials. In the development of aneurysm treatment devices, in silico trials have been instrumental in optimising stent and coil designs by simulating their performance and predicting their impact on blood flow, thus reducing the risk and cost of clinical trials. As computing power and modelling techniques advance, the application of in silico trials is expected to expand in more medical fields, further enhancing personalised medicine. Integration with big data and artificial intelligence will improve the accuracy and efficiency of these simulations, providing more reliable support for clinical decision-making [35].

1.2 Challenges in Vascular Disease Analysis

The analysis of vascular disease in both clinical and research settings presents several challenges, including class imbalance [36, 37], insufficient labelling [38, 39], domain shifts [40, 41], and source-agnostic learning difficulties [42, 43]. These challenges arise from various factors, including data characteristics, technical limitations, and patient-individual differences. In the diagnosis of aneurysms, these problems are particularly pronounced. A single approach is often inadequate to address the diverse complexities encountered in practical applications. The following are the main challenges in vascu-

lature analysis, especially in aneurysm analysis, and how these challenges have led us to propose corresponding solutions.

1.2.1 Class Imbalanced Data

In brain vascular images, there is often a significant imbalance in the proportion of different anatomical structures and lesion areas.

For example, vascular structures typically account for less than 5% of brain tissue, and the size of aneurysms generally accounts for less than 1% of brain tissue. This imbalanced distribution leads to the dilemma: Traditional deep learning (DL) algorithms tend to focus on the dominant categories during training while neglecting the minority categories. However, the segmentation results we need are precisely these minority categories, such as aneurysms. If traditional deep learning-based medical image segmentation algorithms are used, the results often exhibit considerable over-segmentation. To address the class imbalance issue, we propose a method for class imbalanced segmentation in Chapter 3 to improve the model's segmentation performance on imbalanced data, particularly for small targets such as aneurysms.

1.2.2 Insufficient Labelled Data

High-quality vascular image data typically require manual annotation, which is a time-consuming and labour-intensive process. In medical imaging, the annotation process especially requires in-depth expertise from specialists. As a result, there is a large amount of unlabelled data that cannot be directly used for fully supervised training. Additionally, even when data are labelled, they are often partially annotated rather than fully annotated. This issue of insufficient labelling limits the performance of mainstream fully supervised deep learning models, as these models rely on large amounts of fully annotated data to learn accurate segmentation features. To address the issue of insufficient labelling, we propose a method for Semi-supervised Segmentation in Chapter 4. This method combines a small amount of labelled patch data with a large amount of unlabelled patch data, using semi-supervised techniques to enhance the model's segmentation performance. Semi-supervised learning can effectively utilise the unlabelled regions within partially annotated data and the entirely unlabelled data, reducing the dependency of deep learning on labelled data and significantly improving overall segmentation accuracy.

1.2.3 Domain Shift

The acquisition of vascular images can come from different modalities, and there is often a significant domain shift between these modalities. For example, the resolution, noise levels, and pixel value distributions of the images can vary. In practice, our datasets may come from multiple modalities, with annotations often available only in one modality and not in others. For instance, only the 3DRA modality data might be annotated, while the MRA modality data is not. In addition, these data are often not from the same batch of patients and are unmatched. When we train a model on one modality and then directly apply it to another, this domain shift between modalities can affect the model's performance on the target modality. So, how can we use these unannotated modality data to improve the model's performance on the target modality? To address this issue, in Chapter 5, we propose a new unsupervised domain adaptation algorithm that enables the model to adapt to datasets from different modalities, reducing the impact of domain shift, and thus enhancing the model's application scenarios.

1.2.4 Source Agnostic Data

The acquisition of vascular images can come from different devices, imaging parameters, and patient populations. Data sources are often agnostic. The image resolution, noise levels, and pixel value distributions can vary across source-agnostic data. When a model is trained on one dataset and then applied to new data, the domain shift between the inference data and training data can affect the model's performance. In Chapter 6, we propose a new Domain Generalization method aimed at training models that perform well on source-agnostic data. By incorporating new parameter update mechanisms during training, the model can learn to capture invariant features, enabling effective application to new unseen data.

1.3 Thesis Contribution and Overview

1.3.1 Thesis Contributions

In this thesis, we address several key challenges in the analysis of vascular diseases by proposing a series of significant innovative methods and techniques, both theoretically and practically. The main contributions of this thesis are as follows:

- **Class Imbalanced Segmentation**: To tackle the issue of class imbalance in brain vascular images, we designed a new segmentation algorithm that significantly improves the recognition of minority categories (aneurysms) while maintaining high precision. By adjusting the small target-aware network structure and utilising patch-based majority voting and self-refinement, our model performs exceptionally well on aneurysm data, overcoming the limitations of traditional methods in detecting small lesions.
- **Semi-supervised Segmentation**: To address the problem of insufficient labelling, we developed a semi-supervised learning method that combines a small number of labelled patches with a large number of unlabeled patches, effectively enhancing the performance of the segmentation model. The semi-supervised learning technique reduces the reliance on expensive annotated data and improves the robustness of the model.
- **Unsupervised Domain Adaptation**: To solve the issue of domain shift in vascular images, we proposed an unsupervised domain adaptation method, enabling the model to adapt to datasets from different modalities. With domain adaptation techniques, our model shows greater robustness in cross-dataset applications and can be effectively applied to the unlabeled target modality.
- **Domain Generalization**: To address the performance degradation of models when dealing with source-agnostic data, we explored domain generalization techniques aimed at training models that perform well across multiple different domains. Through multi-domain training and advanced learning paradigms, our model maintains high efficiency in completely unknown environments, effectively handling the diversity of devices, imaging parameters, and patient populations.

1.3.2 Thesis Overview

The structure of this thesis is as follows:

- **Chapter 1: Introduction**: This chapter introduces the research background and motivation, provides an overview of brain vessel structure and aneurysm basics, and discusses vascular imaging techniques and the application of in silico trials. It also outlines the main challenges in vascular disease analysis and proposes innovative methods to address these challenges.

- **Chapter 2: Literature Review on Deep Learning in Vessel Aneurysm Segmentation**: A comprehensive review of deep learning techniques in the field of vessel and aneurysm segmentation, covering existing data processing methods, network architectures, optimisation techniques, and advanced learning paradigms.
- **Chapter 3: Class Imbalanced Segmentation of Brain Vessel and Aneurysm**: This chapter details the segmentation method we designed to address the issue of class imbalance and validates its effectiveness and advantages on imbalanced data through experiments.
- **Chapter 4: Semi-supervised Segmentation**: This chapter introduces the semi-supervised learning method we proposed and demonstrates how it enhances the performance of the segmentation model by leveraging a small number of labelled patches and a large number of unlabeled patches.
- **Chapter 5: Domain Adaptation**: This chapter discusses the unsupervised domain adaptation technique we developed, explaining its role and effectiveness in handling domain shifts between different modalities.
- **Chapter 6: Domain Generalization**: This chapter explores domain generalization methods, showing how we trained models to maintain good performance on source-agnostic data and validating these methods through related experiments.
- **Chapter 7: Conclusions**: This chapter summarises the main findings of the thesis, discusses the limitations of the research, and proposes future research directions.

Chapter 2

Literature Review on Deep Learning in Medical Image Segmentation

Medical image segmentation is an important part of medical diagnosis and treatment. In recent years, the development of deep learning technology has brought new opportunities for the segmentation of medical images, and deep learning methods have achieved remarkable results in this field.

This engineering-focused review aims to systematically review the recent advances in deep learning in medical image segmenting, focussing on the foundation models, key modules, loss functions, optimisation algorithms, data processing techniques, some advanced learning paradigms, and future research directions. Through this review, we hope to provide a comprehensive reference for researchers and clinicians to help them better understand and apply deep learning techniques for medical image segmentation and to promote the development and progress of this area.

2.1 Introduction

2.1.1 Overview of Medical Image Segmentation

Medical image segmentation [44] distinguishes different structures, organs, and lesions from the background in a medical image. This process is one of the tasks in medical image analysis and an essential part of many downstream analysis tasks. Medical image segmentation gives doctors more information to help them better diagnose diseases and develop treatment plans. Traditional medical image segmentation methods are usually based on image processing techniques and machine learning algorithms, which typically require a lot of manual design and feature extraction. In recent years, the development of deep learning technology [45] has brought new opportunities for medical image segmentation, and deep learning methods have achieved remarkable results in medical image segmentation tasks. This chapter aims to provide a comprehensive understanding of medical image segmentation while laying the knowledge for subsequent discussions on vessel and aneurysm segmentation.

2.1.2 Challenges of Deep Learning in Medical Image Segmentation

Although deep learning performs well in medical image segmentation, it still faces many challenges. Below is a categorisation of the major challenges.

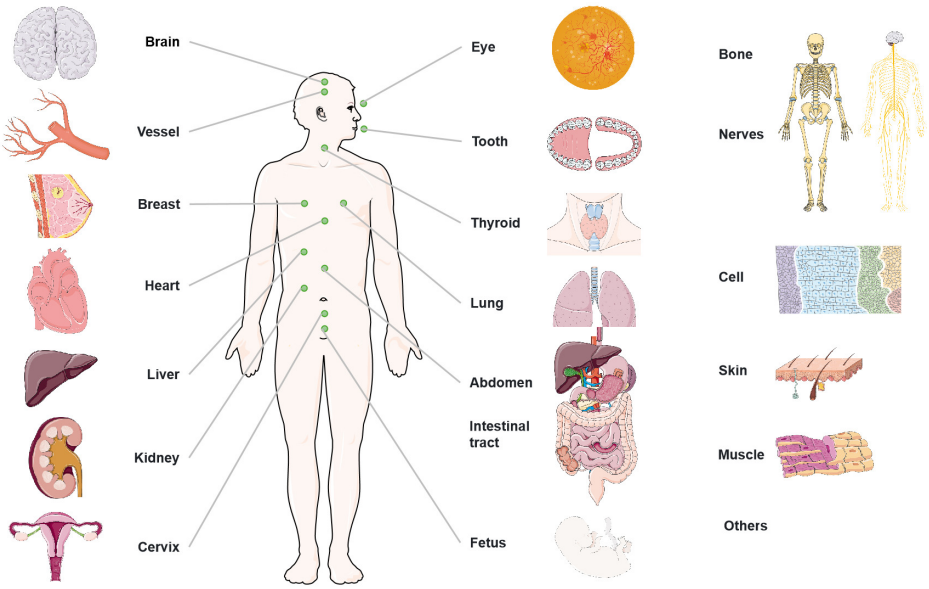

Figure 2.1: Medical Image Segmentation. This figure is used from or adapted from pictures provided by Servier Medical Art, licensed under a Creative Commons Attribution 4.0 Unported License.

Challenges of data

The complexity and variability of medical image data is one of the major challenges for deep learning models. Anatomical structures and lesion features vary significantly from patient to patient, and image quality varies across imaging devices and imaging conditions. In addition, noise and artefacts are often present in medical images, which can affect the performance of modelling training and prediction. Multimodality of data is also a problem; for example, the integration of data from different modalities, such as MRI and CT, requires complex processing and calibration. In addition, the category imbalance of medical image data is significant, where the target region of interest (e.g., tumour or lesion) occupies only a small portion of the image while normal tissues take up the majority of the image, resulting in a tendency for the model to ignore small targets during the training process. Missing and incomplete data can also adversely affect model training and performance. Finally, the large volume of high-resolution 3D medical image data leads to high computational resource requirements for training

deep learning models, which poses a great challenge for practical applications.

Challenges of labelling

High-quality medical image annotation requires in-depth knowledge and a lot of time from experts, which leads to the high cost of annotating datasets. In addition, the problem of inconsistent annotation quality is also common and different annotators can produce subjective errors, leading to inconsistent annotation results. The labelling process is not only time-consuming and laborious but also prone to errors, especially in complex medical images, which can further affect the optimization direction and final performance of the model. For some special lesions or rare diseases, it is even more difficult to obtain high-quality labelled data, which limits the application and promotion of deep learning models in these fields.

Challenges of model robustness and generalization ability

The robustness and generalisability of deep learning models are critical in clinical applications. Although models perform well when trained on specific datasets, they often perform poorly on different datasets or under different imaging conditions, indicating that the cross-modality and cross-domain generalization ability of models needs to be improved. In addition, the model needs to have the ability to handle noise, artefacts, and incomplete data to ensure stability and reliability in real clinical settings. Interpretability and reliability of the models are also key issues, especially in the medical field, where the decision-making process of the models needs to be transparent and interpretable to gain the trust and recognition of clinicians. Improving the performance of deep learning models in data from different sources and conditions is an important direction of current research.

By overcoming these challenges, the application of deep learning in medical image segmentation will become more extensive and efficient, providing more accurate and reliable support for clinical diagnosis and treatment.

2.1.3 Objectives of This Literature Review

This review aims to systematically review recent advances in deep learning in medical image segmentation, focussing on current challenges and solutions. By combining and analysing the existing research results, this chapter will summarise the current status of

deep learning applications in medical image segmentation and introduce the main deep learning models and techniques. In this review, we discuss data preprocessing techniques, including normalisation and data augmentation methods. We also evaluate the performance of different deep learning models and modules, exploring their advantages, disadvantages, and applicable scenarios. Additionally, we introduce the latest advanced techniques, such as semisupervised learning, transfer learning, and multimodal learning, and analyse their potential applications and future prospects in medical image segmentation. Finally, we summarise the limitations and challenges of current research and propose future research directions along with possible solutions. Through this review, we hope to provide a comprehensive reference for researchers and clinicians to help them better understand and apply deep learning techniques for medical image segmentation and to promote the development and progress of this field.

2.2 Dataset and Processing

A simple model can effectively accomplish segmentation tasks when powerful processing techniques are applied. In contrast, even a powerful foundation model may not meet basic segmentation requirements without adequate data processing. nnU-Net [46] exemplifies this by improving segmentation outcomes through advanced data processing strategies rather than modifying the network architecture. By employing various combinations of data processing techniques, nnU-Net has achieved remarkable success in numerous image segmentation challenges.

2.2.1 Preprocessing

Table 2.1 summarises various preprocessing techniques designed to address common issues in medical image segmentation, such as inconsistent image sizes, varying intensities, noise and artefacts, and insufficient data. These techniques ensure that the input data is standardised and robust, thus improving the performance and reliability of the segmentation models.

Table 2.1: Preprocessing Techniques in Medical Image Segmentation

Problem to Address	**Techniques**
Inconsistent Image Sizes	Resampling (Patch-based Learning)
	Interpolation
	Resizing
	Padding
Varying Intensities	Histogram Clipping
	Intensity Normalisation
	Intensity Standardisation
	Intensity Scaling
Background, Noise, Artifacts	RoI Cropping
	Gaussian Filtering
	Median Filtering
	Non-Local Means Filtering
	Noise Addition
	Artifact Reduction
Insufficient Data	Data Augmentation

Medical images often originate from various devices and scanning protocols, resulting in varied image sizes. Several preprocessing techniques are employed to address this issue. Resampling (Patch-Based Learning) adjusts image resolution by resampling to a consistent patch size, ensuring uniform input data for training. Interpolation modifies pixel values to a standard grid, standardising images in size and resolution. Resizing scales images to a fixed size, standardising input dimensions for the model. The padding adds borders to images to achieve a consistent size, ensuring no information is lost at the edges during resizing.

Differences in image intensities can arise due to varying imaging devices and protocols. Techniques to normalise these intensities include Histogram Clipping, which excludes extreme values and only keeps the main content (usually 5 % to 95%, depending on the dataset and modality) on a histogram map to reduce intensity variations. Intensity Normalisation scales pixel values to a specific range (usually 0 to 1). Intensity standardisation converts pixel values to a standard distribution (usually zero mean and unit variance), ensuring consistent contrast. Intensity scaling adjusts intensity values

based on predefined scales, harmonising intensity levels across images.

Medical images often contain noise and artefacts that can interfere with model training and predictions. Techniques to address these issues include RoI Cropping, which focuses on relevant areas by cropping the region of interest (RoI) and removing unnecessary background. Gaussian Filtering applies a Gaussian blur to reduce high-frequency noise, smoothing the image. Median filtering uses the median of pixel values in a neighbourhood to remove noise while preserving edges. Non-Local Means Filtering reduces noise while preserving details by comparing similar patches in the image. Noise addition adds synthetic noise to images during training to improve model robustness. Artefact reduction involves techniques specifically designed to remove artefacts from images, enhancing image quality.

Limited data can hinder model performance in medical image segmentation. Data augmentation techniques generate additional training samples to improve model generalization by introducing variability in the training data. Various augmentation techniques include geometric transformations (rotation, translation, scaling, shearing, flipping, random cropping), intensity transformations (brightness adjustment, contrast adjustment, gamma correction), spatial transformations (grid distortion, elastic deformation), noise addition (Gaussian noise, salt-and-pepper noise), colour transformations (hue adjustment, saturation adjustment) and other advanced methods (random erasing, mixup). Each of these techniques aims to enhance the diversity and robustness of the training data, ensuring that the models can handle a wide range of real-world (clinical) imaging scenarios.

These preprocessing techniques are essential for preparing medical images for deep learning models, ensuring that data is standardised, robust, and ready for effective model training and evaluation.

2.2.2 Postprocessing

Postprocessing techniques can refine and optimise the output from deep learning models. These techniques address various issues, such as noise reduction, edge refinement, and structural coherence, thereby enhancing the overall accuracy and clinical utility of the segmentation results.

Connected component analysis is commonly used to remove small, irrelevant regions by identifying and retaining only the largest connected components, thus eliminating

noise. Conditional processing based on Structure Size similarly removes small objects based on size criteria to ensure only relevant anatomical structures remain. Techniques like Conditional Random Fields (CRF) and Markov Random Fields (MRF) enhance edge continuity, reduce noise, and improve the overall smoothness and consistency of segmentation results by considering the spatial relationships between pixels. Anisotropic diffusion smooths segmentation output while preserving critical edges, effectively reducing noise and artefacts without blurring significant structural details. The hole-filling addresses small gaps in the segmentation results, creating more continuous and complete segmentations, which is essential for accurate medical analysis. Morphological operations, including erosion, dilation, opening and closing, are crucial for boundary smoothing, noise removal, and segmentation refinement. These operations ensure the structural integrity of segmented regions. Statistical Shape Models utilise prior knowledge about anatomical shapes to correct deviations and ensure segmentation accuracy, making the segmentation more reliable for clinical use.

By integrating these postprocessing techniques, the segmentation results become more robust, accurate, and suitable for practical clinical applications. Even when the segmentation output from deep learning models is less than ideal, postprocessing can significantly improve the results. For example, in aneurysm segmentation, deep learning models may struggle to locate aneurysms accurately. Using hybrid majority voting [1] postprocessing can help generate multiple potential segmentation masks. This approach increases the likelihood of identifying true aneurysms, shifting the focus from avoiding false negatives to ensuring true positives are not missed. This transformation emphasises thorough detection, ensuring critical areas are not overlooked.

Table 2.2: Postprocessing techniques and problems they address.

Technique	**Problem to Addressed**
Connected Component Analysis	Removal of small, irrelevant regions
Conditional Processing Based on Structure Size	Removal of small objects based on size
Conditional Random Fields (CRF)	Improving edge continuity and removing noise
Markov Random Fields (MRF)	Enhancing smoothness and consistency
Anisotropic Diffusion	Smoothing while preserving edges
Hole Filling	Filling small holes in segmentation results
Morphological Operations	Smoothing boundaries, refining segmentation
Statistical Shape Models	Ensuring anatomical shape accuracy

2.2.3 Vessel and Aneurysm Dataset

This section presents Table 2.3, which details various datasets related to brain vessels and aneurysms. The table includes 13 public and private datasets related to different modalities. Most of the datasets can be accessed by registering their challenge, or some of them can be downloaded directly.

Each row represents a specific dataset and provides detailed information about it. For instance, the Aneurist dataset contains 223 3DRA images, all of which have both vessel and aneurysm annotations. Another example is the CROWN dataset, which includes 300 MRA images but lacks vessel and aneurysm annotation information.

The purpose of this table is to offer a concise overview, helping researchers quickly understand and compare the features and annotations of different datasets, thus aiding in the selection of the most appropriate dataset for their research needs.

Dataset	Modality	Image	Vessel Label	Aneurysm Label
Aneurist [47]	3DRA	223	223	223
SHINY-ICARUS [48]	3DRA	35	35	-
Leeds General Infirmary	3DRA	163	-	-
Aneurisk [49]	3DRA	99	-	-
CROWN [50]	MRA	300	-	-
SMILE-UHURA [51]	MRA	14	14	-
Aneurist [47]	MRA	207	-	-
ADAM [52]	MRA	93	-	93
TubeTK [53]	MRA	109	-	-
Lausanne [54]	MRA	284	-	284
Aneurist [47]	CTA	198	-	-
Aneurist-UPF [47]	DSA	46	46	46
CADA [55]	DSA	109	-	109
IntrA [56]	MRA	-	103	-

Table 2.3: Brain Vessel Aneurysm Dataset.

2.3 Foundation Models for Medical Image Segmentation

2.3.1 CNN

Convolutional Neural Networks (CNNs) [57] are a class of deep learning models that excel at processing data with grid-like topologies, such as images. The core principle of CNNs involves the convolution operation, where filters or kernels are applied to the input data to create feature maps. This process captures spatial hierarchies and patterns in the data using layers of convolution, pooling, and nonlinear activations. For image segmentation tasks, CNNs are particularly effective because they can learn to recognise and delineate the boundaries of different regions within an image.

A prominent example of a CNN used for segmentation is the U-Net [58]. The architecture of U-Net in Figure 2.2 is characterised by its U-shaped structure, comprising an encoder (downsampling path) and a decoder (upsampling path), connected by skip connections that facilitate the transfer of high-resolution features from the encoder to the decoder. This design allows U-Net to localise and segment regions within an image precisely. Table 2.4 summarises various types of U-Net, including backbone block enhancement, bottleneck enhancement, skip connection enhancement, and learning structure enhancement.

(1) Backbone Block Enhancement

In Table 2.4 section 1, Backbone block enhancement refers to various advancements and modifications made to the traditional U-Net feature, which extracts blocks in both the encoder and decoder to improve its performance. These enhancements involve integrating additional blocks, layers, or mechanisms into the U-Net framework to address specific challenges and improve accuracy, efficiency, and feature extraction capabilities.

The original U-Net [58] was designed for general biomedical image segmentation and is known for its simplicity and effectiveness, particularly on small datasets. Building upon this, V-Net [59] was developed to handle 3D volumetric data, making it suitable for general 3D medical image segmentation. H-DenseUNet [60] introduced multiscale feature fusion, which is particularly effective for gastric cancer image segmentation. Thereafter, several enhancements were introduced: BCDU-Net [61] incorporated bidirectional, densely connected convolutions to improve breast cancer image segmentation; GP-Unet [62] employed gated-propagation for liver image segmentation; SUNet

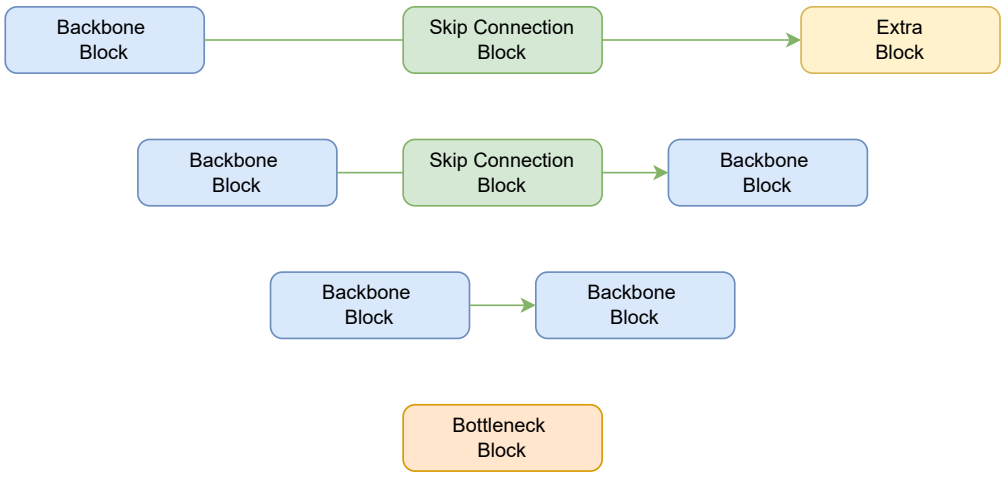

Figure 2.2: Variants of U-Net for Medical Image Segmentation. This figure illustrates various enhancements of the U-Net architecture categorised by modifications in backbone networks, bottleneck layers, skip connections and overall network structures. These advancements aim to improve segmentation performance by incorporating different design strategies and leveraging novel computational techniques.

[63] provided efficient segmentation for ultrasound images; and DUNet [64] utilised a deformable block for retinal vessel segmentation. Dense Multi-path U-Net [65] integrated a multi-path block for stroke lesion segmentation, and Stacked Dense U-Nets [66] applied dual transformation for accurate prostate segmentation. Additionally, Prostate U-net [67] achieved high accuracy in prostate segmentation, while LADDERNET [68] enhanced multi-level feature integration for fundus image segmentation. Further enhancements included USE-Net [69], which utilised squeeze-and-excitation blocks for prostate zonal segmentation, and AnatomyNet [70], which provided efficient bone segmentation with high accuracy for head and neck anatomy. Finally, nnU-Net [46] offered automated configuration for diverse medical image segmentation tasks, establishing itself as a highly adaptable model for various applications.

(2) Bottleneck Enhancement

In Table 2.4 section 2, bottleneck enhancement refers to improvements made to the latent space in bottleneck layers within the deepest U-Net architecture. These enhancements typically involve incorporating advanced attention mechanisms and multiscale feature extraction techniques to improve the network's ability to capture and utilise relevant features.

Attention U-Net [71] introduces a deep attention-aware network designed specifically for pancreas segmentation, allowing the model to focus on relevant areas of the input image. SA-UNet [72] employs spatial attention for retinal vessel segmentation, enhancing the model's ability to learn spatial dependencies within the image. R2U-Net [73] integrates a recurrent attention mechanism for skin lesion image segmentation, improving the model's focus on lesion areas. RA-UNet [74] extends the attention mechanism to liver and tumour segmentation, providing a deep attention-aware network that enhances segmentation accuracy. MA-Net [75] utilises multiscale attention for liver and tumour segmentation, capturing features at multiple scales to improve performance. FRCU-Net [76] incorporates fusion and relation calibration for general medical image segmentation, ensuring that features from different layers are effectively combined. MDU-Net [77] applies a compound attention mechanism for skin lesion segmentation, enhancing feature extraction and improving segmentation accuracy. JCS [78] combines joint classification and segmentation for organs at risk, providing a comprehensive approach that improves both tasks simultaneously.

(3) Skip Connection Enhancement

In Table 2.4 section 3, skip connection enhancement refers to improvements and modifications made to the skip connections within the U-Net architecture to enhance performance in medical image segmentation tasks. These enhancements typically involve incorporating advanced connectivity mechanisms and multiscale feature extraction techniques to improve the network's ability to capture and utilise relevant features effectively.

U-Net++ [79] introduces nested U-Nets, which allow for improved feature fusion and better gradient flow across the network, making it suitable for general medical image segmentation tasks. Projective Skip-Connections [80] enhance connectivity for better gradient flow, providing an efficient way to handle general image segmentation tasks.

Table 2.4: Comparison of Different Segmentation Methods based on CNN

(1) Backbone Block Enhancement			
Method	**Year**	**Applications**	**Highlight**
U-Net[58]	2015	General	Simple, effective for small datasets
V-Net[59]	2016	General	3D volumetric data
H-DenseUNet[60]	2018	Gastric cancer	Multi-scale feature fusion
BCDU-Net[61]	2019	Breast cancer	Bidirectional, densely connected
GP-Unet[62]	2020	Liver	Gated-Propagation
SUNet[63]	2019	Ultrasound	Efficient segmentation of ultrasound images
DUNet[64]	2019	Retinal vessel	Deformable block
Dense Multi-path U-Net[65]	2021	Stroke Lesion	Multi-path block
Stacked Dense U-Nets[66]	2019	Prostate	Dual Transformation
Prostate U-net[67]	2019	Prostate	High accuracy in prostate segmentation
LADDERNET[68]	2018	Fundus	Multi-level feature integration
USE-Net[69]	2018	Prostate zonal	Squeeze-and-Excitation blocks
AnatomyNet[70]	2019	Head and neck anatomy	Efficient bone segmentation, high accuracy
nnU-Net[46]	2021	General	Automated configuration for diverse tasks
(2) Bottleneck Enhancement			
Method	**Year**	**Applications**	**Highlight**
Attention u-net[71]	2018	Pancreas	Deep attention-aware network
SA-UNet[72]	2019	Retinal vessel	Spatial attention
R2U-Net[73]	2019	Skin lesion image	Recurrent attention mechanism
RA-UNet[74]	2020	Liver and tumor	Deep attention-aware network
MA-Net[75]	2020	Liver and tumor	Multi-scale attention
FRCU-Net[76]	2020	General	Fusion and relation calibration
MDU-Net[77]	2021	Skin lesion	Compound attention mechanism
JCS[78]	2021	Organs at risk	Joint classification and segmentation
(3) Skip Connection Enhancement			
Method	**Year**	**Applications**	**Highlight**
U-Net++[79]	2018	General	Nested U-Nets
Projective Skip-Connections[80]	2018	General	Enhanced connectivity for better gradient flow
Attention UNet++[81]	2020	Liver	Nested attention mechanisms
U-Net 3+[82]	2020	General	Enhanced multi-scale feature fusion
(4) Extra Structure Enhancement			
Method	**Year**	**Applications**	**Highlight**
Cascaded Unet[83]	2018	Brain tumor	High segmentation accuracy for glioma
Probabilistic U-Net[84]	2018	General	Probabilistic modeling, uncertainty estimation
Hierarchical Probabilistic U-Net[85]	2019	General	Hierarchical modeling, uncertainty estimation
Polar Transformation U-Net[86]	2019	Retinal	Polar transformation, multi-scale features
MRF-UNet[87]	2020	Brain tumor	Markov random fields
Bayesian Skip Net[88]	2017	General	Bayesian inference, uncertainty estimation
VAE U-Net[89]	2018	Brain	Variational autoencoder
PMBR U-Net[90]	2019	Lung nodule	Probability maps, recurrent connections
Path Aggregation U-Net[91]	2021	Brain tumor	Deep supervision
Teacher-student network[92]	2021	Abdominal	Knowledge distillation

Attention UNet++ [81] employs nested attention mechanisms specifically designed for liver segmentation, enabling the network to focus on relevant areas of the input image. U-Net 3+ [82] enhances multiscale feature fusion, making it highly effective for general medical image segmentation applications.

(4) Extra Structural Enhancement

In Table 2.4 section 4, learning structure enhancement demonstrates the improvement based on the combination of the unet model and other task knowledge. These enhancements typically involve incorporating advanced probabilistic models, hierarchical structures, and specialised transformations to improve the network's ability to learn and generalise from the data effectively.

Cascaded Unet [83] is designed for brain tumour segmentation and achieves high segmentation accuracy for gliomas by employing a cascaded approach. Probabilistic U-Net [84] introduces probabilistic modelling and uncertainty estimation, making it suitable for general medical image segmentation tasks. Hierarchical Probabilistic U-Net [85] extends this concept with hierarchical modelling to further improve uncertainty estimation. Polar Transformation U-Net [86] applies polar transformations to enhance multiscale feature extraction, making it effective for retinal image segmentation. MRF-UNet [87] incorporates Markov random fields for brain tumour segmentation, enhancing the network's ability to model spatial dependencies. Bayesian Skip Net [88] uses Bayesian inference for general medical image segmentation, providing robust uncertainty estimation. VAE U-Net [89] combines the U-Net architecture with a variational autoencoder for brain image segmentation, enhancing feature learning and representation. PMBR U-Net [90] leverages probability maps and recurrent connections for lung nodule segmentation, improving segmentation accuracy. Path Aggregation U-Net [91] employs deep supervision to enhance brain tumour segmentation. Finally, the Teacher-student network [92] applies knowledge distillation techniques to improve abdominal image segmentation.

2.3.2 Transformer

Transformer-based models in Table 2.5 have gained significant traction in the field of medical image segmentation due to their ability to capture long-range dependencies and contextual information effectively. These models can be broadly categorised into

three main groups: Transformer-Enhanced U-Net Variants, Standalone Transformer Backbone Designs, and Hybrid and Specialized Transformer Networks.

Transformer-Enhanced U-Net Variants

Transformer-enhanced U-Net variants integrate transformer architectures into the traditional U-Net framework to enhance feature learning and segmentation accuracy. Notable examples include TransU-Net[93], which incorporates enhanced contextual feature learning, and CoTr[94], which utilises a deformable transformer architecture for abdominal organ segmentation. UTNet[95] integrates a transformer encoder block, making it suitable for cardiac segmentation tasks, while Transclaw u-net[96] combines convolution with a transformer in the encoder to handle multi-organ segmentation. TransFuse[97] and TransBTSV2[98] improve feature fusion and brain tumour segmentation through transformer CNN feature fusion and enhanced swin transformer, respectively. RTNet[99] focuses on retinal segmentation by leveraging relation and global transformer blocks, whereas AFter-UNet[100] and UNETR[101] offer robust feature representation through axial fusion transformers and robust feature representation, respectively. MTU-Net[102] employs multiscale transformers for skin lesion segmentation, and UCTransNet[103] utilises multi-head attention in skip connections for multi-organ and cell nucleus segmentation. Swin-UNet[104], DS-TransUNet[105], and Swin UNETR[106] applies sliding window and self-supervised learning for multi-organ segmentation. Recent developments include LeViT-UNet[107], which employs faster encoders, and TransAttUnet[108], which incorporates multi-level attention mechanisms. PAG-TransYnet[109] introduces a hybrid dual pyramid transformer-CNN, further improving segmentation performance.

Standalone Transformer Backbone Design

Standalone Transformer Backbone Designs replace traditional convolutional networks with pure transformer architectures. ViT[110] and nnFormer[110] exemplify this approach by using pure vision transformers without CNNs and incorporating local and global volume-based self-attention mechanisms. Medical-Transformer[111] employs a Local-Global training strategy to enhance segmentation accuracy, while MISSFormer[112] models local context and multiscale features. UNETR[101] applies pure vision transformers for brain tumour and spleen segmentation, providing robust performance without

CNNs. HT-Net[113] introduces a hierarchical context-attention transformer, and D-former[114] utilises dilated transformers to improve segmentation accuracy. MedNeXt[115] leverages a transformer-driven ConvNeXt architecture, offering enhanced performance across various medical image segmentation tasks.

Hybrid and Specialized Transformer Networks

Hybrid and Specialised Transformer Networks combine transformers with other architectures or tailor them for specific applications. Multi-Branch Hybrid Transformer[116] integrates multiple body-edge branches for corneal endothelial cell segmentation, while Multi-Compound Transformer[117] learns cross-scale dependencies of different pixels for cell nuclei instance segmentation. Cross-Teaching Transformer[118] employs semi-supervised learning for cardiac segmentation, and Spine-transformers[119] use learnable positional embeddings for spine segmentation. Boundary-aware Transformer[120] focuses on extracting local details for skin lesion segmentation. Dor polyp segmentation task, Polyp-PVT[121] fuses different-level features and ColonFormer[122] provide enhanced multiscale feature extraction and hierarchical multi-level network. SE-Transformers[123] incorporate squeeze and expansion mechanisms for general medical image segmentation tasks. ST-GAN[124] combines GANs with transformers for cardiac segmentation.

2.3.3 Mamba

Mamba[125] is a novel sequence model architecture that enhances traditional state space models through the use of Selective State Space Models (SSMs). Unlike previous models, Mamba adjusts SSM parameters in an input-dependent manner, allowing the model to selectively transmit or forget information based on the current data, thus addressing the limitations faced by earlier models in handling discrete and information-dense data such as text. Although this approach precludes the use of efficient convolution calculations, researchers have developed a hardware-aware parallel algorithm to run in a recursive manner, enabling Mamba to achieve inference speeds five times faster than traditional Transformers and linear scaling with sequence length. Mamba's unique features include its rapid processing capability, selective SSM layers, and hardware-friendly design inspired by FlashAttention[126], making it a significant advancement in the field of machine learning. This innovative architecture is a simple generalization of the S4 architecture (Structured State Spaces for Sequence Modeling) previously developed by

Table 2.5: Comparison of Different Segmentation Methods based on Transformer

(1) Transformer-Enhanced U-Net Variants			
Method	**Year**	**Applications**	**Highlight**
TransU-Net[93]	2021	General	Enhanced contextual feature learning
CoTr[94]	2021	Abdominal organ	Deformable transformer architecture
UTNet[95]	2021	Cardiac	Transformer encoder block
Transclaw u-net[96]	2021	Multi-organ	Combines convolution with transformer in encoder
TransFuse[97]	2021	General	Transformer CNN feature fusion
TransBTSV2[98]	2022	Brain tumor	Enhanced swin transformer
RTNet[99]	2022	Retinal	Relation and global transformer block
AFter-UNet[100]	2022	General	Axial fusion transformer
UNETR[101]	2022	General	Robust feature representation
MTU-Net[102]	2022	Skin lesions	Multi-scale transformers
UCTransNet[103]	2022	Multi-organ, cell nucleus	Multi-head attention in skip connection
Swin-UNet[104]	2022	General	Enhanced multi-scale feature extraction
DS-TransUNet[105]	2022	General	Transformer interactive fusion module
Swin UNETR[106]	2022	Multi organs	Self-supervised learning
LeViT-UNet[107]	2023	General	Faster encoders
TransAttUnet[108]	2023	General	Multi-level attention
PAG-TransYnet[109]	2024	General	Hybrid dual pyramid transformer-CNN
(2) Standalone Transformer Backbone Design			
Method	**Year**	**Applications**	**Highlight**
ViT[110]	2021	General	Pure vision transformer without CNN
nnFormer[110]	2021	General	local and global volume-based self-attention mechanism
Medical-Transformer[111]	2021	General	Local-Global training strategy
MISSFormer[112]	2022	General	modeling local context, multi-scale features
UNETR[101]	2022	Brain tumor, spleen	Pure vision transformer without CNN
HT-Net[113]	2022	General	Hierarchical context-attention transformer
D-former[114]	2023	General	Dilated Transformer
MedNeXt[115]	2023	General	Transformer-driven ConvNeXt
(3) Hybrid and Specialized Transformer Networks			
Method	**Year**	**Applications**	**Highlight**
Multi-Branch Hybrid Transformer[116]	2021	Corneal endothelial cell	Hybrid Transformer with multiple body-edge branches
Multi-Compound Transformer[117]	2021	Cell nuclei instance	Learning the cross-scale dependencies of different pixels
Cross-Teaching Transformer[118]	2022	Cardiac	Semi-supervised learning
Spine-transformers[119]	2022	Spine	Learnable positional embeddings
Boundary-aware Transformer[120]	2021	Skin lesion	extract local details
Polyp-PVT[121]	2021	Polyp	Fuse different-level features
SE-Transformers[123]	2021	General	Squeeze and expansion
ColonFormer[122]	2022	Colon polyp	Hierarchical multi-level network
ST-GAN[124]	2023	Cardiac	GAN combined with Transformer

Albert Gu, allowing Mamba to rival or surpass Transformers in language modelling.

The Mamba-UNet family comprises various segmentation methods tailored for different medical imaging tasks, each with distinct enhancements and benefits. U-Mamba leverages residual blocks followed by the SSM-based Mamba block, making it suitable for abdominal organ segmentation [127]. Mamba-UNet integrates the Mamba architecture within a U-Net structure, excelling in abdominal organs and cardiac segmentation [128]. H-vmunet employs a high-order 2D-selective-scan, enhancing performance in general segmentation tasks [129]. Mamba-HUNet introduces a Hierarchical Upsampling Network, specifically beneficial for Multiple Sclerosis lesion segmentation [129]. P-Mamba incorporates a Perona Malik diffusion block for echocardiographic applications [130], while ProMamba utilises a prompt mechanism for polyp segmentation [131]. TM-UNet features a Triplet SSM, applicable in general segmentation tasks [132]. Semi-Mamba-UNet integrates self-supervised pixel-level contrastive learning for cardiac segmentation [133].

Swin-UMamba combines SwinUNet with a visual state space block for general segmentation tasks [134]. UltraLight VM-UNet processes features in parallel Vision Mamba, enhancing skin lesion segmentation [135]. VM-UNet adopts a pure SSM-based model, suitable for general applications [136], while VM-UNET-V2 infuses semantics and detail, excelling in general tasks [137]. Weak-Mamba-UNet utilises scribble-based annotations, beneficial for cardiac segmentation [138]. LMa-UNet employs large Mamba kernels for abdominal organ segmentation [139]. LightM-UNet focuses on long-range spatial dependencies through a residual pure Mamba fashion [140]. SegMamba excels in whole-volume feature modelling, particularly for brain tumour segmentation [141]. Lastly, T-Mamba uses shared positional encoding and frequency-based features for tooth segmentation [142].

In Table 2.6, these methods collectively showcase the versatility and advanced capabilities of the Mamba-UNet variants across a wide range of medical imaging tasks, each tailored to address specific challenges and enhance segmentation performance in their respective domains.

Method	Year	Application	Highlight
U-Mamba[127]	2024	Abdominal organ	Residual blocks followed by the SSM-based mamba block
Mamba-UNet[128]	2024	Abdominal organ, cardiac	Mamba in unet structure
H-vmunet[129]	2024	General	High-order 2D-selective-scan
Mamba-HUNet[129]	2024	Multiple Sclerosis lesion	Hierarchical Upsampling Network
P-Mamba[130]	2024	Echocardiographic	Perona Malik diffusion block
ProMamba[131]	2024	Polyp	Prompt
TM-UNet[132]	2024	General	Triplet SSM
Semi-Mamba-UNet[133]	2024	Cardiac	Self-supervised pixel-level contrastive learning
Swin-UMamba[134]	2024	General	SwinUNet with visual state space block
UltraLight VM-UNet[135]	2024	skin lesion	Process features in parallel Vision Mamba
VM-UNet[136]	2024	General	Pure SSM-based model
VM-UNET-V2[137]	2024	General	Semantics and detail infusion
Weak-Mamba-UNet[138]	2024	Cardiac	Scribble-based annotations
LMa-UNet[139]	2024	Abdominal organ	large Mamba kernels
LightM-UNet[140]	2024	Chest	Residual pure Mamba fashion for long-range spatial dependencies
SegMamba[141]	2024	Brain tumor	Excels in whole volume feature modeling
T-Mamba[142]	2024	Tooth	Shared positional encoding, frequency-based features

Table 2.6: Comparison of Different Segmentation Methods based on Mamba.

2.4 Learning Paradigms

2.4.1 Fully-supervised Learning

Principles

Fully-supervised learning in medical image segmentation involves training models using a comprehensive dataset where each image is paired with a corresponding ground truth segmentation mask. This paradigm leverages the availability of detailed annotations to learn direct mappings from input images to their respective segmentations.

Problem Formulation

In the fully-supervised learning framework, we aim to learn a mapping function $f : X \rightarrow Y$ where X represents the input images and Y denotes the ground truth segmentation masks. Given a training dataset $\{(x_i, y_i)\}_{i=1}^{N}$, where $x_i \in X$ and $y_i \in Y$, the objective is to minimise the discrepancy between the predicted segmentation $\hat{y}_i = f(x_i; \theta)$ and the ground truth y_i. Formally, this can be expressed as the minimisation of a loss function L_{fully}:

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} L(f(x_i; \theta), y_i)$$

Where θ denotes the parameters of the model. Commonly used loss functions in fully supervised learning for medical image segmentation are introduced below. The performance of fully-supervised models is evaluated using metrics such as the Dice coefficient, Intersection over Union, etc. The details of evaluation metrics are described in Section 2.5. Fully-supervised learning relies on large, annotated datasets to train models that can accurately segment medical images. The effectiveness of these models is highly dependent on the quality and quantity of the labelled data. With enough data, the robustness and generalization of the network will be qualitatively enhanced.

Techniques in Fully-Supervised Learning

Fully-supervised learning primarily relies on labelled data to train models. One of the key techniques in fully-supervised learning is the use of loss functions [143]. These functions measure the discrepancy between the predicted outputs and the actual labels, guiding the optimisation process to improve model performance. Various loss functions are tailored to handle specific challenges, such as class imbalance, boundary accuracy, and hard-to-classify examples. By carefully choosing and combining these loss functions, fully-supervised learning can achieve high accuracy and robustness in a wide range of applications.

(1) Pixel-based Losses

True positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) are commonly used terms in binary classification problems to evaluate the performance of a model. TP means the number of correctly predicted positive samples. FP means the number of incorrectly predicted positive samples. TN means the number of correctly predicted negative samples. FN means the number of missed positive samples (i.e., incorrectly predicted as negative).

Dice loss [144] measures the overlap between the predicted segmentation and the ground truth. It is particularly suitable for imbalanced datasets as it directly optimises the Dice coefficient, a measure of similarity. The loss function is defined as:

$$\text{Dice Loss} = 1 - \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{2.1}$$

This loss function balances precision and recall, ensuring that small structures are adequately segmented.

Tversky loss [145] is a generalization of dice loss, which allows the weighting of false positives and false negatives. It is defined as:

$$\text{Tversky Loss} = 1 - \frac{\text{TP}}{\text{TP} + \alpha\text{FP} + \beta\text{FN}} \tag{2.2}$$

where α and β control the penalties for FP and FN, respectively. This flexibility makes it suitable for highly imbalanced datasets.

Surface-to-surface loss [143] measures the average distance between the surfaces of the predicted segmentation and the ground truth. This loss provides a robust metric for assessing the quality of the segmentation, particularly in 3D medical imaging tasks. By focusing on the surfaces, it ensures that the predicted segmentation closely aligns with the true anatomical boundaries.

$$\text{Surface-to-Surface Loss} = \frac{1}{N}\sum_{i=1}^{N}\left(\min_{y\in\partial G} d(x_i, y) + \min_{x\in\partial P} d(y_i, x)\right) \tag{2.3}$$

where $N = |\partial P| + |\partial G|$, and ∂P and ∂G represent the sets of surface points of the predicted segmentation and the ground truth, respectively, and $d(x, y)$ denotes the Euclidean distance between points x and y.

(2) Distribution-based Losses

Cross-entropy (CE) loss [143] measures the performance of a classification model whose output is a probability value between 0 and 1.

$$\text{CE Loss} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)] \tag{2.4}$$

where y_i is the ground truth label and p_i is the predicted probability. CE Loss is widely used in segmentation tasks for its effectiveness in handling multi-class problems.

Focal loss [**?**] down-weights the contribution of easy examples and focuses more on hard examples. It is particularly useful for addressing class imbalance.

$$\text{Focal Loss} = -\frac{1}{N}\sum_{i=1}^{N}\alpha_i(1 - p_i)^{\gamma}\log(p_i) \tag{2.5}$$

where N is the total number of samples, y_i is the ground truth label, p_i is the predicted probability for the positive class, α_i is a weighting factor for class imbalance, and γ is the focusing parameter that adjusts the down-weighting effect on easy examples.

(3) Composite Losses

DiceCE loss [143] combines Dice Loss and Cross-Entropy Loss to improve performance, particularly on imbalanced datasets. Leveraging the strengths of both losses ensures accurate class probabilities and good overlap with the ground truth.

DiceFocal loss [143] combines Dice Loss with Focal Loss to focus more on hard-to-classify examples. It enhances the segmentation quality by addressing class imbalance and difficult cases simultaneously.

TopK loss [143] focuses on the top K most difficult pixels to classify, improving performance in challenging areas. By giving more weight to the hardest examples, it enhances the model's ability to handle difficult cases. The TopK Loss is defined as:

$$\text{TopK Loss} = \frac{1}{K} \sum_{i \in \text{TopK}} L_i \tag{2.6}$$

where L_i represents the loss for each pixel i (i.e., Dice loss), and TopK denotes the set of K pixels with the highest individual losses. This formulation ensures that the model pays more attention to the hardest-to-classify pixels, thus improving overall segmentation performance, especially in difficult areas.

Extended logarithmic (EL) loss [146] is designed to handle extreme class imbalance by applying a logarithmic transformation to the predicted probabilities, enhancing the distinction between different classes. The loss function is defined as:

$$\text{EL Loss} = -\frac{1}{N} \sum_{i=1}^{N} \left(y_i \log(1 + p_i) + (1 - y_i) \log(1 + (1 - p_i)) \right) \tag{2.7}$$

Here, N is the number of samples, y_i is the ground truth label for sample i, and p_i is the predicted probability for the positive class for sample i. By incorporating the logarithmic transformation, ELL increases the penalty for misclassifying the minority class, thus improving the model's ability to distinguish between classes in highly imbalanced datasets.

2.4.2 Semi-supervised Learning

Principles

Semi-supervised learning in medical image segmentation [38] is designed to utilise both labelled and unlabelled data to improve the performance of segmentation models. Given

the high cost and time associated with obtaining fully annotated medical images, semi-supervised learning methods are increasingly valuable. These methods leverage a small labelled dataset along with a larger unlabelled dataset to enhance the learning process, often employing techniques such as consistency regularisation, pseudo-labelling, and teacher-student frameworks.

Problem Formulation

In the semi-supervised learning framework, we aim to learn a mapping function $f : X \rightarrow Y$ where X represents the input images and Y denotes the ground truth segmentation masks. Given a training dataset consisting of labelled data $\{(x_i^l, y_i^l)\}_{i=1}^{N_l}$ and unlabelled data $\{x_i^u\}_{i=1}^{N_u}$, the objective is to leverage both datasets to improve segmentation performance. Formally, this can be expressed as the minimisation of a combined loss function L that includes supervised loss on labelled data L_s and unsupervised loss on unlabelled data L_u:

$$\min_{\theta} \left[\frac{1}{N_l} \sum_{i=1}^{N_l} L_s(f(x_i^l; \theta), y_i^l) + \lambda \frac{1}{N_u} \sum_{i=1}^{N_u} L_u(f(x_i^u; \theta)) \right]$$

Where θ denotes the parameters of the model and λ is a weighting factor that balances the contribution of the supervised and unsupervised losses.

Techniques in Semi-supervised Learning

Pseudo-labelling [147] involves initially training a model on the labelled dataset to generate pseudo-labels for the unlabelled data. These pseudo-labels are treated as ground truth in subsequent training iterations, effectively expanding the training set and iteratively refining the model's performance.

The teacher-student framework [148] comprises two models: a teacher and a student. The teacher model, trained on labelled data, generates pseudo-labels for the unlabelled data. The student model is then trained using both labelled and pseudo-labelled data. Kullback-Leibler Divergence loss is often used to align the predictions of the student model with those of the teacher model, ensuring effective learning from the teacher and improving the student's performance.

Uncertainty analysis [149] involves using uncertainty estimates to weight the loss function, focusing on areas with high uncertainty. By prioritising these uncertain re-

gions, the model can learn to handle challenging cases more effectively, thereby improving its robustness.

Confidence learning [150] penalises the model based on the confidence of its predictions. By encouraging high-confidence correct predictions, this technique helps in improving the reliability and trustworthiness of the segmentation outputs.

Consistency regularisation [151] ensures that the model generates stable predictions under various augmentations or perturbations of the input data. This technique involves applying transformations such as rotation, scaling, or noise addition to both labelled and unlabelled data and encouraging the model to maintain consistent output despite these changes.

2.4.3 Unsupervised Learning

Principles

Unsupervised learning in medical image segmentation [152] focuses on training models without using labelled data. This approach is advantageous when labelled datasets are scarce or unavailable. The primary objective is to discover patterns and structures within the data that can facilitate segmentation without explicit supervision. Techniques such as clustering, autoencoders, and generative models are commonly employed in this paradigm.

Self-supervised learning is a specific type of unsupervised learning where the model learns to segment images using automatically generated labels or pseudo-labels. This approach leverages abundant unlabelled data by creating supervisory signals from the data itself, thus eliminating the need for large annotated datasets. Self-supervised methods typically involve pretext tasks designed to train the model to understand the underlying structure of the images.

Problem Formulation

In the unsupervised learning framework, the goal is to learn a mapping function $f : X \rightarrow Y$ where X represents the input images and Y represents the latent segmentation maps. Given a training dataset $\{x_i\}_{i=1}^{N}$ where $x_i \in X$, the objective is to identify underlying patterns and structures in the data that correspond to meaningful segmentations. Formally, this can be expressed as the minimisation of an unsupervised

loss function L_u:

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} L_u(f(x_i; \theta))$$

Where θ denotes the parameters of the model. The loss function L_u is designed to capture the intrinsic structure of the data without relying on labelled examples.

In the context of self-supervised learning, the goal is to learn a robust representation $f : X \to Z$ from the input images X to a latent space Z using pretext tasks. These tasks are designed to generate pseudo-labels that serve as the ground truth for training. Given a dataset $\{x_i\}_{i=1}^{N}$ where $x_i \in X$, the objective is to learn a representation that can be transferred to downstream segmentation tasks. Formally, this can be expressed as the minimisation of a self-supervised loss function L_{ss}:

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} L_{ss}(f(x_i; \theta), g(x_i))$$

Where θ denotes the parameters of the model and g is a function that generates pseudo-labels from the data itself.

Techniques

Autoencoders are neural networks trained to reconstruct their input data. By learning to compress and decompress the input images, autoencoders capture significant features that can be used for segmentation. Variational autoencoders (VAE) [153] extend this idea by incorporating probabilistic modelling, allowing the model to generate new samples from the learnt latent space, which can facilitate the discovery of meaningful segmentations.

Generative adversarial networks (GANs) [154] consist of a generator and a discriminator network. The generator creates synthetic images, while the discriminator distinguishes between real and synthetic images. When applied to segmentation, the generator can produce segmentation masks that the discriminator evaluates. This adversarial process encourages the generation of realistic segmentation masks.

Contrastive learning [155] involves learning representations by comparing similar and dissimilar pairs of images. The model is trained to bring the representations of similar pairs closer and push apart those of dissimilar pairs. Methods like SimCLR[156] and MoCo [157] are popular in the unsupervised representation learning domain.

By combining various techniques within the unsupervised learning paradigm, including self-supervised learning as a specific approach, it is possible to effectively leverage unlabelled data to improve medical image segmentation, reducing the dependence on large annotated datasets and enhancing the robustness of the models.

2.4.4 Hybrid Learning

Principles

Hybrid learning combines multiple learning paradigms to leverage their complementary strengths and improve model performance. In medical image segmentation, hybrid learning methods often integrate aspects of supervised, semi-supervised, and unsupervised learning, as well as techniques like transfer learning and reinforcement learning. The goal is to create more robust and generalisable models by utilising various sources of information and different training strategies.

Techniques in Hybrid Learning

Transfer learning [158] involves using a pre-trained model on a large dataset (e.g., ImageNet) and fine-tuning it on a smaller, domain-specific dataset. This approach leverages the knowledge gained from the large dataset to improve performance on the target task. Transfer learning is particularly useful in medical image segmentation, where labelled data is often scarce.

Domain adaptation [159] techniques aim to generalise a model trained on a source domain to perform well on a target domain with different data distributions. Methods like domain adversarial neural networks (DANN) and maximum mean discrepancy (MMD) help align the feature distributions between the source and target domains. This is crucial for medical imaging, where data from different scanners or protocols can vary significantly.

Domain generalization [42] techniques extend the concept of domain adaptation by enabling models to generalise to unseen domains. Approaches like learning invariant representations and using meta-learning strategies help models perform well across various domains without requiring access to target domain data during training.

Incremental learning[160], also known as continuous learning, involves updating a model continuously as new data becomes available. Techniques such as memory-augmented networks and rehearsal methods help the model retain knowledge of previ-

ously seen data while adapting to new tasks. This is particularly important in medical imaging, where new data and evolving protocols are common.

Multi-task learning [161] involves training a single model to perform multiple related tasks simultaneously. In the context of medical image segmentation, this could mean combining segmentation with tasks such as classification, detection, or anatomical landmark identification. By sharing representations across tasks, multi-task learning can improve the performance and efficiency of the model. Techniques such as hard parameter sharing and task-specific layers are commonly used in multi-task learning frameworks.

2.5 Evaluation and Analysis

2.5.1 Quantitative Evaluation

Quantitative evaluation in medical image segmentation involves using specific metrics to objectively measure the performance of segmentation algorithms. These metrics provide a numerical assessment of how well the predicted segmentation aligns with the ground truth. Table 2.7 are some common evaluation metrics along with their formulas:

Table 2.7: Evaluation Metrics and Their Formulas

Metric	Formula
Dice Similarity Coefficient	$\frac{2\times TP}{2\times TP+FP+FN}$
Jaccard Coefficient	$\frac{TP}{TP+FP+FN}$
Volumetric Similarity	$1-\frac{\lvert FN-FP\rvert}{2\times TP+FN+FP}$
Hausdorff Distance	$\max\left\{\max_{p\in S}\min_{p'\in S'}\lVert p-p'\rVert, \max_{p'\in S'}\min_{p\in S}\lVert p'-p\rVert\right\}$
Average Hausdorff Distance	$\frac{1}{\lvert S\rvert}\sum_{p\in S}\min_{p'\in S'}\lVert p-p'\rVert$
Balanced Average Hausdorff Distance	$\frac{1}{2}\left(\mathrm{AHD}(S,S')+\mathrm{AHD}(S',S)\right)$
Mean (Average) Surface Distance	$\frac{1}{\lvert S\rvert}\sum_{p\in S}\min_{p'\in S'}\lVert p-p'\rVert$
Median Surface Distance	$\mathrm{median}_{p\in S}\left(\min_{p'\in S'}\lVert p-p'\rVert\right)$
95th Percentile Hausdorff Distance	$\mathrm{percentile}_{95}\left(\min_{p'\in S'}\lVert p-p'\rVert, \min_{p\in S}\lVert p-p'\rVert\right)$
Standard Deviation of Surface Distance	$\sqrt{\frac{1}{\lvert S\rvert}\sum_{p\in S}\left(\min_{p'\in S'}\lVert p-p'\rVert-\mathrm{MSD}(S,S')\right)^2}$
Sensitivity (Recall, True Positive Rate)	$\frac{TP}{TP+FN}$
Specificity (True Negative Rate)	$\frac{TN}{TN+FP}$
Precision	$\frac{TP}{TP+FP}$
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Balanced Accuracy	$\frac{1}{2}\left(\frac{TP}{TP+FN}+\frac{TN}{TN+FP}\right)$
F1 Score (F-Measure)	$\frac{2\times\mathrm{Precision}\times\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}}$
Mutual Information	$\sum_{x\in X}\sum_{y\in Y}p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right)$
Cohen Kappa	$\frac{p_o-p_e}{1-p_e}$
Rand Index	$\frac{TP+TN}{TP+TN+FP+FN}$
Adjusted Rand Index	$\frac{\mathrm{RI}-E[\mathrm{RI}]}{\max(\mathrm{RI})-E[\mathrm{RI}]}$
Interclass Correlation	$\frac{\sigma_b^2}{\sigma_b^2+\sigma_w^2}$
False Positive Rate (Fallout)	$\frac{FP}{FP+TN}$
False Negative Rate	$\frac{FN}{FN+TP}$

2.5.2 Qualitative Analysis

Qualitative analysis in Table 2.8 involves visual inspection of the segmentation results to assess the accuracy and quality of the predictions. This can include various techniques such as overlaying segmentation masks on the original images, using attention maps, and employing other visualisation tools to intuitively understand the performance of the model. Here are the key components of qualitative analysis in medical image

segmentation.

Overlaying Segmentation Masks [162]: One of the most straightforward methods for qualitative analysis is to overlay the predicted segmentation masks on the original medical images. This helps in visually assessing how well the segmented regions align with the actual structures in the images. By examining these overlays, researchers and clinicians can quickly identify areas where the model performs well and areas where it may need improvement.

Attention Maps [163]: Attention maps are another powerful tool for qualitative analysis. These maps highlight the regions of the image that the model focuses on when making its predictions. By visualising attention maps, one can gain insights into the decision-making process of the model, understand which parts of the image are considered most important, and identify any potential biases or errors in the model's focus. This is particularly useful in complex segmentation tasks where understanding the model's attention can lead to better model interpretation and refinement.

Saliency Maps [164]: Saliency maps indicate the importance of each pixel in the image for the model's prediction. These maps help in understanding which parts of the image contribute most to the final segmentation output. Saliency maps are useful for identifying key features and ensuring that the model correctly identifies relevant anatomical structures.

Gradient-weighted Class Activation Mapping (Grad-CAM) [165]: Grad-CAM is a technique that uses the gradients of any target concept flowing into the final convolutional layer to produce a coarse localisation map, highlighting the important regions in the image. This method is particularly useful for understanding deep learning models, as it provides a visual explanation for the model's decision.

Error Maps [166]: Error maps can be generated to highlight the differences between the predicted segmentation and the ground truth. These maps can show false positives, false negatives, and areas of disagreement, providing a clear visualisation of the model's performance and areas needing improvement. Error maps can be used on both point-based mesh data and pixel-based image data.

3D Visualisations [167]: For volumetric data such as MRI or CT scans, 3D visualisations of the segmented regions can provide a more comprehensive understanding of the model's performance. By examining the 3D structures, researchers can assess the spatial accuracy and continuity of the segmentation across different slices.

Table 2.8: Qualitative Analysis Techniques for Medical Image Segmentation

Technique	Highlight
Overlaying Masks	Overlaying predicted masks on original images to check alignment and accuracy.
Attention Maps	Visualises the model's focus regions to understand its decision-making process.
Saliency Maps	Indicates the importance of each pixel for model prediction, identifying key features.
Grad-CAM	Uses gradients to create localisation maps highlighting important image regions
Error Maps	Differences between predictions and GT, showing FP, negatives, and disagreements.
3D Visualisations	Views of segmented volumetric structures, assessing spatial accuracy and continuity.

By incorporating these visualisation techniques, qualitative analysis can provide valuable insights into the performance and reliability of medical image segmentation models, complementing the quantitative metrics and enhancing the overall evaluation process.

2.5.3 Statistical Analysis

Statistical analysis in medical image segmentation is crucial for validating the robustness and reliability of segmentation models. It involves the application of various statistical techniques to interpret, compare and understand the performance metrics and outcomes of the segmentation models. Table 2.9 shows key components and methodologies for conducting comprehensive statistical analysis.

Table 2.9: Evaluation Metrics and Their Formulas

Statistical Approach	**Detailed Methods**	
Descriptive Statistics	Mean	
	Standard Deviation	
	Median	
Inferential Statistics	Hypothesis Testing	T-Test
		ANOVA
		Chi-Square Test
	Confidence Intervals	
	Regression Analysis	
Correlation Analysis	Pearson Correlation	
	Spearman Rank Correlation	
Agreement Metrics	Cohen Kappa	
	Intraclass Correlation Coefficient	

Descriptive statistics summarise the basic features of the dataset and the results from the segmentation models. Commonly used descriptive statistics include measures of central tendency (mean, median, mode) and measures of variability (standard deviation, variance, range). These statistics help in understanding the distribution and spread of performance metrics like the Dice coefficient, Jaccard index, and Hausdorff distance. Specifically, the mean value provides the average value of the performance metrics across all samples. Standard deviation indicates the amount of variation or dispersion from the mean. The median value represents the middle value in the distribution, offering a measure that is less sensitive to outliers.

Inferential statistics allow researchers to make inferences about the population from which the samples are drawn. This includes hypothesis testing, confidence intervals, and regression analysis. Hypothesis testing is used to determine if there are significant differences between the performances of different models or between the predicted and ground truth segmentations. Common tests include t-tests, analysis of variance (ANOVA), and chi-square tests. Specifically, the t-test assesses whether the means of the two groups are statistically different from each other. ANOVA is used when comparing the means of more than two groups. The chi-square test evaluates whether there is a significant association between categorical variables. Confidence intervals provide a

range of values that likely contain the true population parameter. It helps understand the precision of the estimated performance metrics. Regression analysis examines the relationship between dependent and independent variables. In segmentation, it can be used to understand how various factors (e.g., image quality and patient demographics) impact segmentation performance.

Correlation analysis measures the strength and direction of the relationship between two variables. In the context of medical image segmentation, it can be used to explore the relationships between different performance metrics or between segmentation performance and other clinical variables. Pearson correlation measures the linear relationship between two continuous variables. Spearman rank correlation assesses the monotonic relationship between two ranked variables.

Agreement metrics evaluate the concordance between the predictions made by different segmentation models or between the model predictions and ground truth. Cohen Kappa measures the agreement between two raters or models, correcting for agreement occurring by chance. Besides, the intraclass correlation coefficient (ICC) assesses the reliability of measurements or ratings, which is suitable for evaluating consistency among multiple raters or segmentation methods.

Receiver Operating Characteristic (ROC) curves and the Area Under the Curve (AUC) provide a graphical representation of the model's diagnostic ability. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1 specificity) at various threshold settings. AUC represents the degree or measure of separability, indicating how well the model can distinguish between classes.

By implementing these statistical analysis techniques, researchers can gain a comprehensive understanding of the performance and reliability of their medical image segmentation models, ensuring that the models are robust, accurate, and clinically applicable.

Chapter 3

Class Imbalanced Segmentation

Background and Objectives: Automatic segmentation of the cerebral vasculature and aneurysms facilitates incidental detection of aneurysms. Assessing aneurysm rupture risk assists with pre-operative treatment planning and enables in-silico investigation of cerebral hemodynamics within and in the vicinity of aneurysms. However, ensuring precise and robust segmentation of cerebral vessels and aneurysms in neuroimaging modalities such as three-dimensional rotational angiography (3DRA) is challenging. The vasculature constitutes a small proportion of the image volume, resulting in a significant class imbalance (relative to surrounding brain tissue). Additionally, aneurysms and vessels have similar image/appearance characteristics, making distinguishing the aneurysm sac from the vessel lumen challenging.

Methods: We propose a novel multi-class convolutional neural network to tackle these challenges and facilitate the automatic segmentation of cerebral vessels and aneurysms in 3DRA images. The proposed model is trained and evaluated on an internal multi-centre dataset and an external publicly available challenge dataset.

Results: On the internal clinical dataset, our method consistently outperformed several state-of-the-art approaches for vessel and aneurysm segmentation, achieving an average Dice score of 0.81 (0.15 higher than nnU-Net) and an average surface-to-surface error of 0.20 mm (less than the in-plane resolution (0.35 mm/pixel)) for aneurysm segmentation; and an average Dice score of 0.91 and average surface-to-surface error of 0.25 mm for vessel segmentation. In 223 cases of a clinical dataset, our method accurately segmented 190 aneurysm cases.

Conclusions: The proposed approach can help address class imbalance problems and inter-class interference problems in multi-class segmentation. Besides, this method performs consistently on clinical datasets from four different sources, and the results generated are qualified for hemodynamic simulation. Code available at https://github.com/cistib/vessel-aneurysm-segmentation.

3.1 Introduction

Cerebral aneurysms are pathological protrusions of cerebral arterial walls (see Fig. 3.1, for example), and their rupture is the leading cause of subarachnoid haemorrhage in patients. Three-dimensional X-ray rotational angiography (3DRA) imaging is commonly used to visualize and characterize cerebral vessels and aneurysms through the reconstruction of tomographic slices of a region of interest like computed tomography

angiography (CTA), using single-plane radiographic equipment [168, 169]. In contrast with 3D magnetic resonance angiography (MRA) imaging and CTA, 3DRA provides images of higher spatial resolution and improved soft-tissue contrast [170, 171, 172], capturing fine vascular structures and enabling precise characterization of aneurysm morphology.

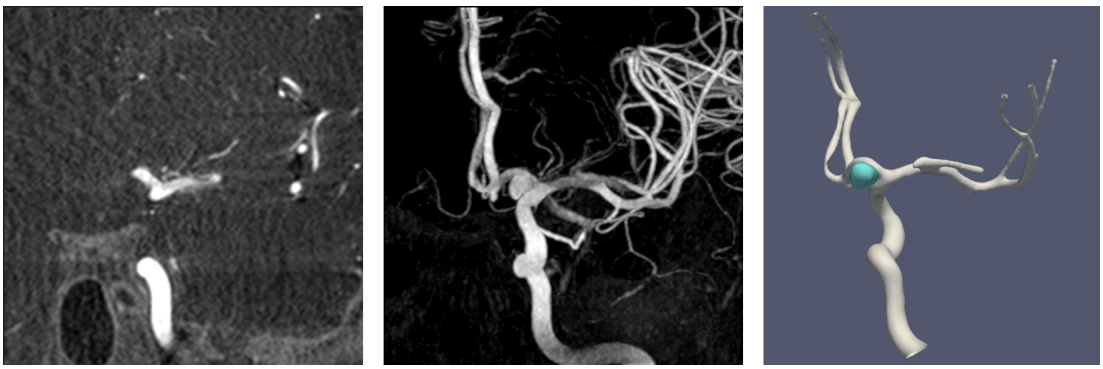

Figure 3.1: From left to right: 2D slice from a reconstructed 3DRA image; the maximum intensity projection (MIP) of the 3DRA image; and a 3D simulation-ready mesh of a cerebral aneurysm (blue) and the major vessels (white) in its vicinity reconstructed from its corresponding main vessel segmentation.

An accurate, automated, and reproducible cerebral vessel and aneurysm segmentation technique would facilitate various computational imaging and clinical applications. Segmentation of cerebrovasculature has found its use in pre-operative planning of invasive procedures [173], delivering image-guided therapies/treatments [174], and assessing cerebral hemodynamics through computational fluid dynamics (CFD) simulations [175]. Similarly, detection and segmentation of cerebral aneurysms are valuable as they facilitate incidental identification and quantitative characterization of aneurysm morphology [176]. The latter is especially useful as previous studies have shown that the size and shape of a cerebral aneurysm are essential biomarkers for evaluating rupture risk [177]. Precise characterization of aneurysm location and morphology is necessary for selecting a suitable/approved treatment strategy, pre-operative intervention planning, and post-operative assessment and monitoring. While cerebral aneurysms and their surrounding vessels can be detected and segmented manually, this process is time-consuming due to the high dimensionality of 3D image volumes. It is subject to inter- and intra-observer variability. Suppose the computer-assisted model can automatically detect the location

of the aneurysm and characterize its morphological properties such as neck diameter, aneurysm size, etc. In that case, these characteristics can support the clinical decision-making process. For instance, depending on the size and location of the aneurysm, the clinician would use coils or flow diverters to treat the aneurysm. Similarly, to decide the type of medical device for treating the aneurysm, the clinician relies on the information on whether there is a bifurcation around the aneurysm. On the one hand, the direct clinical needs involve automatically characterizing the location and morphological properties of the aneurysm, which could inform the best course of treatment for the aneurysm. On the other hand, to motivate clinical needs, *in-silico trials* [6] can help identify the best operational regimes for the use of certain devices and better inform the safety and efficacy of medical devices in clinical trials. In order to be able to scale up *in-silico trials* to large-scale, both qualified vessel and aneurysm segmentation are required to derive accurate geometrical and computational models.

Several previous studies have proposed automatic and semiautomatic techniques for cerebral vessel and aneurysm segmentation in 3D imaging modalities such as MRA, CTA, and 3DRA. Early work in the field relied on classical methods such as geodesic active regions for segmenting vessels and aneurysms. Recent approaches have focused on data-driven supervised learning-based methods due to the tremendous success of convolutional neural networks (CNNs) at detecting and segmenting objects/regions in images. As in several other domains, the segmentation performance afforded by deep learning-based approaches for cerebral vessels and aneurysms far exceeds that of classical methods. For example, in a recent study [178], the authors proposed DeepVesselNet, a CNN designed to segment cerebral vessels in MRA images. Here, 2D orthogonal cross-hair filters (convolutions) were used to preserve details of fine vascular structures in the learned features while incorporating 3D contextual information. As vessels constitute a small fraction of the overall image volume, the segmentation task suffers from a significant class imbalance between the foreground (vessel) and background (surrounding brain tissue) classes. This was addressed by training DeepVesselNet with a class-balanced cross-entropy loss function that minimizes the false-positive rate. Similarly, to incorporate 3D contextual information and improve the accuracy of segmenting fine vessels in digital subtraction angiography, Patel et al. [179] used DeepMedic, a powerful segmentation approach proposed in a previous study [180]. To improve the performance of deep learning on small object segmentation and obtain annotated training data at a

rapid pace, Vessel-CAPTCHA [181] proposes a novel annotation-efficient deep learning vessel segmentation framework. The framework only requires weak patch-level labels to discriminate between the vessel and non-vessel 2D patches in the training set. This framework can effectively segment vessels including both main and fine branches. Unlike the original U-Net architecture [58], DeepMedic is a 3D CNN with two parallel encoder pathways that learn features at different image resolutions to capture contextual information while keeping the computational cost low. 3D image patches centered at the same location in the image are used as inputs to the two pathways. The image is downsampled to a third of its original size for the second pathway. Several U-Net based approaches have been proposed for segmenting cerebral vessels and aneurysms. The method proposed by Livne et al. [182] is trained to segment cerebral vessels using 2D patches extracted from MRA images [183] with a U-Net. The method proposed by Shahzad et al. [184] segments ruptured intracranial aneurysms resulting in subarachnoid hemorrhage in CTA images using DeepMedic. Zhou et al. [185] proposed U-Net++ as a new framework for image segmentation to further improve segmentation performance. The redesigned skip connections in U-Net++ aggregate features across multiple scales within decoder sub-networks, leading to a highly flexible feature fusion scheme. Attention modules have been widely used in vessel segmentation networks to weigh the importance of relevant but under-represented structures/features. However, stand-alone segmentation networks trained and applied to imaging data without using appropriate pre- and post-processing steps typically lack robustness when segmenting fine structures (such as vessels and aneurysms) in the presence of significant class imbalance and variability in image appearance and soft-tissue contrast (typical of imaging data acquired across multiple centres). Therefore, to facilitate robust segmentation of diverse imaging data with imbalanced classes, Isensee et al. [46] proposed nnU-Net, which can automatically configure itself, including pre-processing, network architecture, and training and post-processing for any new task in the biomedical domain. The nnU-Net improves the robustness of the model by learning fixed, rule-based and empirical parameters.

The methods discussed thus far achieved state-of-the-art segmentation performance for cerebrovascular structures. However, several challenges remain to enable precise and robust characterization of cerebral vessels and aneurysms in 3D, namely, effectively dealing with the severe class imbalance, the difficulty of distinguishing between

the aneurysm and vessel lumen, and the lack of a robust deep learning framework for segmenting cerebral vessels and aneurysms for multi-center studies. Firstly, the aneurysm region often constitutes less than 1 % of the overall image volume. The features extracted from the small regions, such as fine vessels and aneurysms, may not be effectively propagated through a series of convolutional and up-/down-sampling layers by conventional CNN-based networks. Secondly, vessel interference is the main reason for over-segmentation due to the similarity of closed vessels in patch boundaries to aneurysms in morphology. Last, multi-center imaging data varies considerably in image appearance and spatial resolution due to different scanners and image acquisition protocols across different institutions. All these factors make precise segmentation of cerebral vessels and aneurysms challenging. Addressing these challenges is the main focus of this study and here are the contributions:

1) A dual-class segmentation network is proposed for the automatic segmentation of cerebral vessels and aneurysms in 3DRA images. To deal with the class imbalance inherent in such a segmentation task, especially for aneurysms, we proposed a cascaded transformer block at the end of the encoder to highlight aneurysm features. Multi-view blocks are designed to receive continuous features in a lower feature dimension. Learnable downsample blocks are proposed at the end of every encoder block to prevent small features from being washed out during down-sampling. Wide blocks are designed to extract high-level features in multi-dimensions.

2) For the inter-class interference challenge, we designed the multi-class network with weighted Dice loss and set aneurysms as a subclass of vessels. The semantic guidance from vessel features reduces the interference of brain tissue and skull with aneurysms and can significantly improve aneurysm segmentation performance.

3) To further enhance the aneurysm segmentation performance, we designed a post-processing pipeline including majority voting and self-refinement which can predict accurate aneurysm localization and boundary.

4) For hemodynamics simulation analysis, to the best of our knowledge, previous methods have segmented vessels or aneurysms individually, and most experiments have been validated using image-based evaluation metrics like Dice. Whether the independent outputs of these segmentation methods are suitable for vessel and aneurysm simulation is still unknown. Our method allows the simultaneous segmentation of shape-consistent vessels and aneurysms. More importantly, after generating the mesh

from the image-based output, a mesh-based surface-to-surface error evaluation was performed to verify that the output is suitable for simulation (surface-to-surface error 0.20 mm for aneurysm segmentation and 0.25 mm for vessel segmentation, less than the in-plane resolution 0.35 mm/pixel). The automatic segmentation pipeline can bridge the gap between clinical data and hemodynamic simulation input.

We comprehensively evaluated the proposed approach across two 3DRA datasets: an in-house multi-centre @neurIS dataset [186] and a publicly available cerebral aneurysm detection and analysis dataset (CADA) [187].

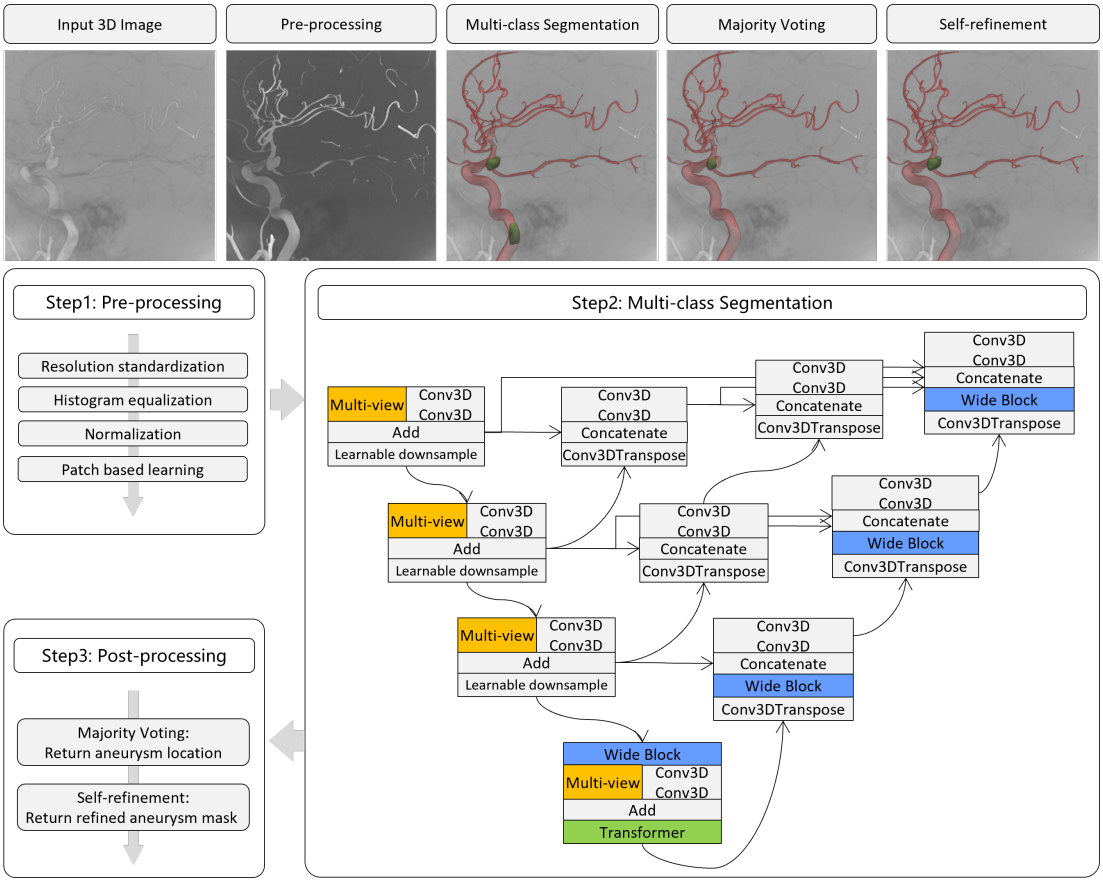

Figure 3.2: Schematic of the proposed cerebral vessels and aneurysm segmentation pipeline comprising three steps. Step 1: shows the pre-processing operations applied to the original 3DRA images; Step 2: shows the architecture of the proposed multi-class CNN-based segmentation model; Step 3: shows the post-processing operations applied to segmented images based on majority voting and self-refinement.

3.2 Methodology

This section describes the 3D multi-class cerebral vessel and aneurysm segmentation network proposed in this study and the overall pipeline developed to ensure robust and reproducible segmentation performance. The proposed multi-class CNN incorporates several architectural components dedicated to preserving fine structural details in-plane across multiple orthogonal planes, and ensures consistency in 3D for the vascular structures of interest. The proposed pipeline comprises three steps: pre-processing, multi-class segmentation, and post-processing in Fig. 3.2. Details of each step of the pipeline and the architectural components developed for the multi-class segmentation network are discussed in subsequent subsections.

3.2.1 Pre-processing

The first step in our segmentation pipeline focuses on processing the original 3DRA image volumes to generate 3D image patches suitable for training the multi-class segmentation network. Before extracting patches from the image volumes, we use a sequence of operations to reduce the variability across patients' images and stabilize subsequent segmentation network training. As the spatial resolution varies across patients' images in the @neurIST database, we standardized the resolution of all image volumes by resampling them to a fixed voxel size of $0.35 \times 0.35 \times 0.35\ mm$. We also applied histogram equalization to the resampled image volumes to reduce differences in tissue contrast across patients' images. Then, we normalized voxel intensities in all images to $[0, 1]$. Vascular structures are only partially labeled within the @neurIST dataset, with ground-truth segmentations available just for the major cerebral arteries and their branches in the vicinity of the aneurysm. Hence, we cropped each image volume using a bounding box encapsulating the labeled structures to reduce label noise/confounding information when training the segmentation network. Finally, we extracted 3D patches of size $64 \times 64 \times 64$ voxels from the cropped volumes, discarding all patches that had no associated labels for the vessels or aneurysm, and used the remaining patches for training our segmentation network. Each patch is labeled as follows: 0 for background, 1 for vessels, and 2 for aneurysms, ensuring that only relevant regions contribute to model training.

3.2.2 Multi-class Segmentation Network

The backbone of our multi-class segmentation network's architecture is U-Net++ [185], a deeply supervised encoder-decoder network with nested dense connections across convolution blocks nestled between the encoder and decoder paths (as illustrated in Fig. 3.2). The nested dense skip pathways help aggregate multiscale features at each convolution block in the decoder from all convolution blocks at the same resolution level or below (relative to the former) in the encoder. These dense connections help alleviate the restrictive behavior of skip connections (allowing only for the same-scale fusion of learned features) used in the original U-Net architecture and its variants. They enable rich semantic multiscale features from different encoder blocks to be used by each decoder block to generate segmentations. Additionally, we propose several additional feature extraction modules (discussed in subsequent sections) integrated into this backbone U-Net++ architecture to increase the network's sensitivity to detect vessels and aneurysms.

Cascaded Transformer

Cerebral vessels and aneurysms constitute less than 6% of the overall image volume in 3DRA images [188]. This leads to a significant class imbalance between the foreground (vessels and aneurysms) and background (surrounding tissues) classes. We designed a cascaded transformer block to address this challenge, which adaptively increases the network's attention on vessels and aneurysms [189]. As shown in Fig. 3.3., we first exploited this module at the end of the encoder to integrate local features with their global dependencies along the spatial and channel dimensions in parallel paths. The spatial attention module on the top left selectively highlights the locations that comprise vessels and aneurysms by a weighted sum of the features from all locations. Meanwhile, the channel attention module on the top right enhances the interdependence between different channels through a sequence of permutation and dot product operations acting channel-wise on the input feature maps. Then, we exploited three multi-head attention modules cascaded with multilayer perception (MLP) in latent space to learn features with a wider spatial context further. Since the model learns the feature order in latent space rather than by spatial position, we removed the position embedding layer here to reduce artificial interference and provide more room for learning. Finally, we add the features from different attention stages, increasing training stability and the

weight of key features specific to vessels and aneurysms to enhance overall segmentation performance.

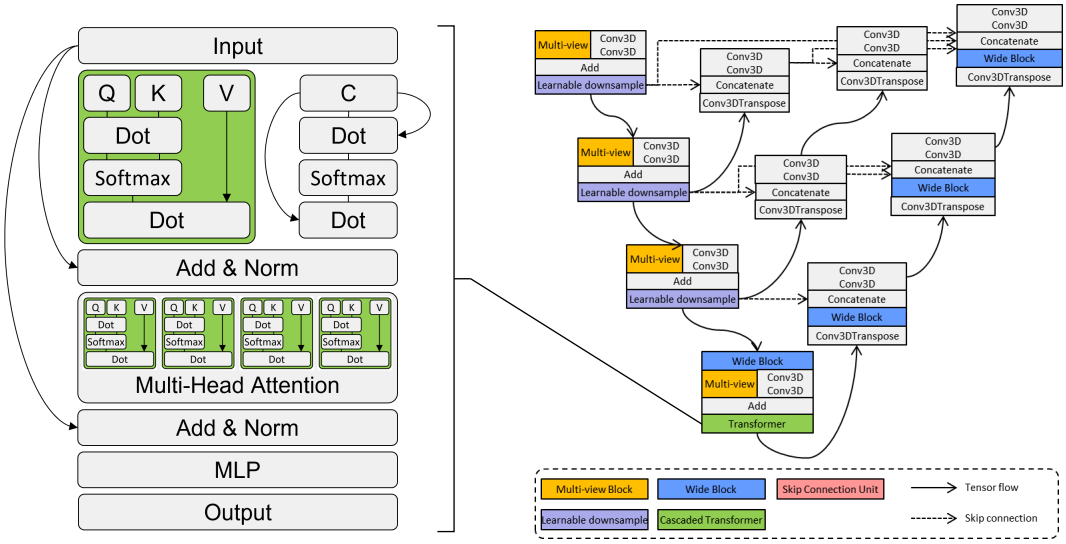

Figure 3.3: Cascaded Transformer block. Given the small proportion of aneurysms in the dataset, the attention mechanism enhances the weighting of aneurysm regions, effectively addressing class imbalance by improving their representation in the segmentation process.

Multi-view Block

Learning representative features of fine vessels is challenging as they may be as few as two pixels in diameter in Fig. 3.13. Because of noise points in the low-level features, convolution with a kernel size of three has difficulty distinguishing noise from such subtle features with a diameter of less than three. These noise-like features are either filtered out or over-enhanced by using only small kernel 3D convolution filters. But in large receptive field, this continuous and uniformly slender feature will be completely different from the noise point feature. We propose a multi-view convolution block in Fig. 3.2.2 composed of three branches to extract 2D features in the larger receptive field along orthogonal planes of the 3D image volume to tackle this challenge. This enables feature learning and orthogonal views, which are subsequently aggregated, to highlight the slender features specific to fine vessels. As shown in Fig. 3.2.2, in the second and third views, fine vessel features are less prominent due to their small proportion, whereas

in the first view, the fine vessels exhibit a distinct elongated distribution, making them more clearly captured by the neural network. This represents finer vascular structures than conventional 3D convolutions. Besides, we are also interested in preserving the 3D structure and morphology of vessels and aneurysms, which is somewhat lost using just 2D orthogonal convolutions. Hence, we also use a 3D convolution layer alongside each multi-view block in our network, and add the features learned by the former and latter. This combination of the multi-view block and a 3D convolution layer ensures that 3D contextual features are learned and aggregated with detailed features of fine vascular structures. The structure of the multi-view block is shown in Fig. 3.2.2.

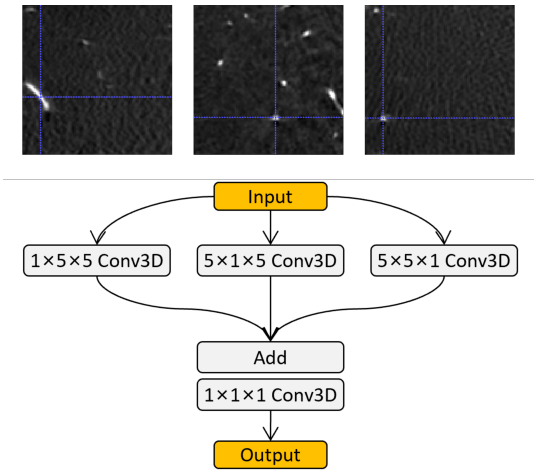

Figure 3.4: Multi-view block. Extract features from three orthogonal views using 2D convolutions in large kernels. The fine vessel feature is not obvious in the second and third views, but the slender feature is easily captured in the first view. A $5 \times 1 \times 5$ kernel is used because, at the final downsampling layer, the feature map size is $8 \times 8 \times 8$. The kernel size is chosen to be ≤ 8 to effectively capture spatial features while maintaining sufficient receptive field coverage within the downsampled resolution.

Learnable Downsample Block

Down-sampling of learned feature maps through pooling operations is essential in CNNs to increase the receptive field size of the network and enable learning of hierarchical features while keeping model complexity (i.e. the number of learnable parameters) reasonable to reduce overfitting and ensure computational tractability. During down-

sampling, weak features of small vessels and aneurysms are easily ignored/lost if standard pooling operations are used (e.g. max-pooling). We designed a learnable downsample block to retain weak features to compensate for this. This block halves the size of the input feature maps along three parallel branches comprising two stridden convolution branches and a max-pooling branch, as illustrated in Fig. 3.5 (left). These three branches dissociate the spatial and channel information through a 3D convolution layer (with a kernel size of 3) for the former and three parallel 2D convolutional layers (multi-view block) of factorized asymmetric 3D convolutions for the latter (as shown in Fig. 3.5). This enables effective learning of spatial and channel-wise features alongside down-sampling the feature maps in each path. The max-pooling path downsamples the input feature maps and aggregates them across channels using a 3D convolution layer with a kernel size of one. The resulting feature maps from all three branches are concatenated, resulting in downsampled feature maps that have preserved weak features across multiple scales and dimensions.

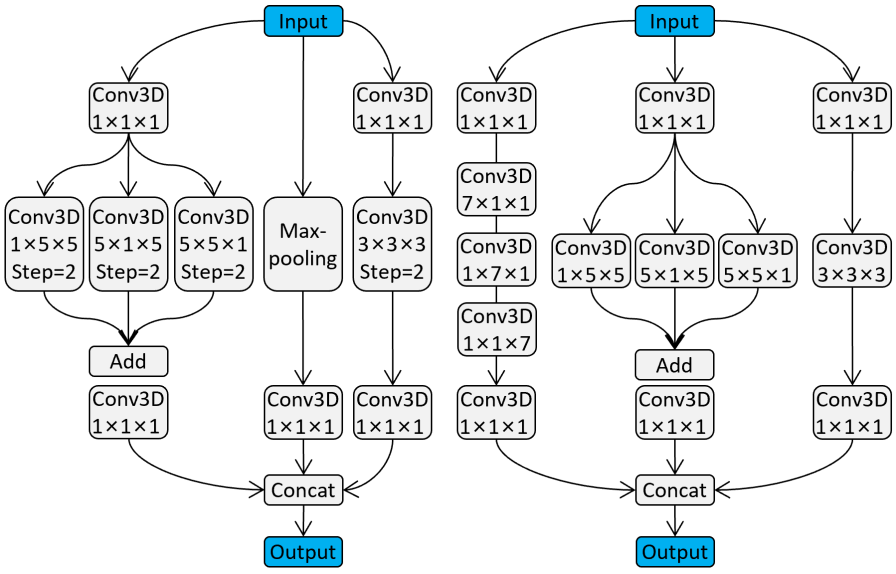

Figure 3.5: Learnable downsample block (left) and wide block (right).

Wide Block

To further increase the receptive field size of the network, with a limited increase in model complexity, and learn features with a wider spatial context, we designed the wide

block. This module comprises three branches in Fig. 3.5 (right). One branch aggregates the feature maps across channels and learns local features with a 3D convolution of kernel size three. The other two branches utilize asymmetric 3D convolutions applied in parallel and serially with a large kernel size to learn multiscale features along all dimensions. The resulting feature maps from each branch are finally concatenated and provided as input to subsequent network layers.

Loss Function

The aneurysm regions account for only a small part of the brain image, and a critical imbalance exists in the distribution of the positive and negative samples. Thus, following the generalized dice loss proposed in [190], in this work, we used the weighted dice loss (Eq. 3.1) that weights inversely proportional to labels area to predict better labels with general small regions, i.e., the aneurysms in our case. w_a and w_v are the weights of the aneurysm and vessel volume in Eq. (3.2). Dice is calculated in Eq. (3.3). Because of the class imbalance issue, we compensated for this by multiplying the Dice of aneurysm segmentation, whose proportion is small by the larger weights, and the vessel segmentation, whose proportion is big by the smaller weights.

$$Loss = (1 - w_a) * Dice_a + (1 - w_v) * Dice_v \tag{3.1}$$

$$w_a = \frac{V_a}{V_a + V_v}, w_v = \frac{V_v}{V_a + V_v} \tag{3.2}$$

where V_a is the volume of the aneurysm, and V_v is the volume of the vessel. We use the volume proportion to balance class imbalance.

3.2.3 Post-processing

Majority voting

The multi-class segmentation network processes patch data. In the segmentation result of a single patch, a vessel near the patch edge has a closed geometry and is, therefore, incorrectly identified as an aneurysm. After ensembling the predictions of patches, there will be over-segmentation of aneurysms like the fourth column of Fig. 3.2. As visualized in Fig. 3.6 and Fig. 3.7, the 2D image is divided into four groups of patches represented in red, yellow, green and blue boxes. Similarly, the 3D image is divided into eight groups

of patches. Each group of patches is fed into the multi-class segmentation model for prediction. For actual aneurysms in 3.6, all eight groups had positive prediction results (eight votes). For fake aneurysms (vessels near the patch edge mentioned above) in Fig. 3.7, only four groups (pred 2, 4, 6, 7) had positive prediction results (four votes), the remaining four groups (pred 0, 1, 3, 5) did not predict the controversial area to be an aneurysm. Therefore, after ensembling the results of eight groups, we keep only the predicted area with the highest number of votes and use this area as the aneurysm prediction of majority voting. The proposed approach decomposes each 3DRA image volume into eight groups of patches and uses these to train the multi-class segmentation network. The starting points of the eight groups are the eight vertices of the 3D image after zero-padding, which is to ensure the eight groups do not overlap completely. Patch-based learning allows semantic features to be learned from the 3DRA images in their native resolution with a limited degree of down-sampling throughout the network, not afforded by methods that learn features directly from the original image volumes due to GPU memory constraints.

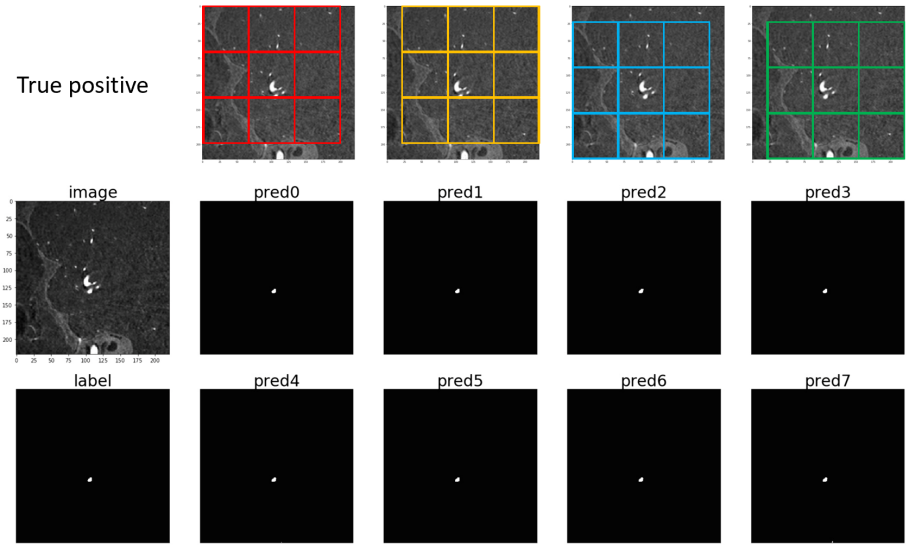

Figure 3.6: Actual aneurysm majority voting result. 2D image have four patch groups, 3D image will have eight patch groups. The different colours represent the different patch selection strategies. The label shows here is an aneurysm and the prediction of eight different groups is correct.

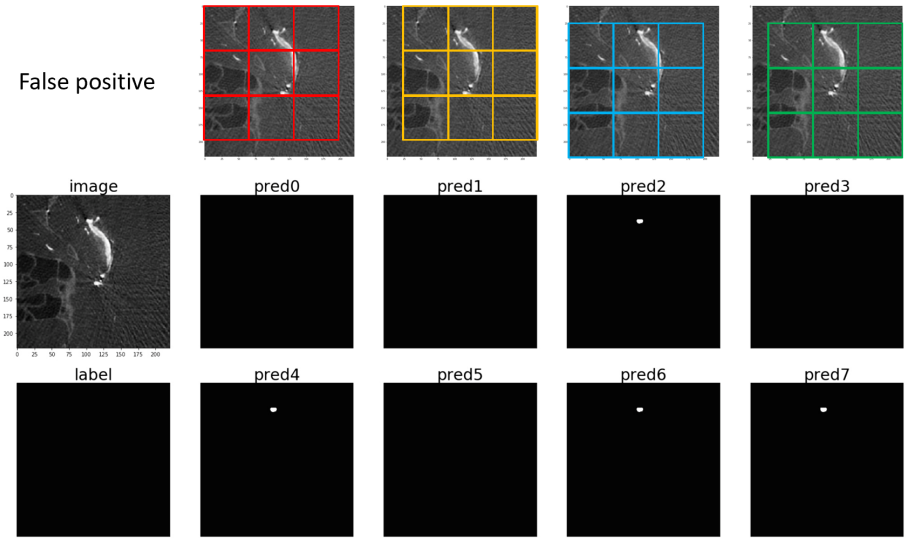

Figure 3.7: Fake aneurysm majority voting result. The different colours represent the different patch selection strategies. The label shows there is no aneurysm. The predictions of pred0, 1, 3, and 5 are correct, and pred2, 4, 6, and 7 are incorrect, which will cause over-segmentation.

Self-refinement

The fourth column of Fig. 3.2 shows that the aneurysm may still be inaccurately over-/undersegmented. Therefore, we refine the segmentation results by selecting one patch centered on the aneurysm prediction (after majority voting) and feeding this patch into the multi-class segmentation model to predict the aneurysm's shape. If there are multiple independent aneurysms, patches will also be selected multiple times. During this process, the predictions close to the edge of the patch are ignored to prevent interference from vessels at the edges here. Finally, we can obtain aneurysm predictions with correct locations and accurate boundaries in the fifth column of Fig. 3.2.

By combining majority voting with self-refinement, we can prevent vessels at the edge of the patch from being predicted as aneurysms, effectively suppressing aneurysm over-segmentation.

3.3 Experimental Setup

3.3.1 Datasets

The proposed method is trained and validated on 3DRA images from 223 patients acquired from the @neurIST project [186]. These images were acquired across four different centers with different scanners and imaging protocols. There are significant variations in image appearance and resolution across image data from different centers, as shown in Fig. 3.8. Image data from the @neurIST database were split patient-wise into training, validation, and test sets using a ratio of 7 : 1 : 2, respectively, and five-fold cross-validation experiments were conducted to thoroughly evaluate the segmentation performance of the proposed approach and the state-of-the-art methods. The test sets in different cross-validation experiments traverse the entire data set. We also trained and evaluated our approach on a publicly available dataset, CADA [187], which comprises 3DRA images of patients with medium and large cerebral aneurysms. These images were acquired as part of the Cerebral Aneurysm Detection and Analysis challenge, hosted at the international conference on medical image computing and computer-assisted interventions in 2020. The training data released as part of the CADA challenge comprised 109 3DRA images with 127 annotated aneurysms. We also split the labeled data into training, validation, and test sets by patient in a 7 : 1 : 2 ratio. When preparing the training data, we randomly extracted 3D patches around the aneurysm. Negative patches that did not contain aneurysms were not selected for training. When training the segmentation model, there were 904 patches extracted from the cropped volumes, with 716 patches being used as the training set and the rest patches being used as the validation set to monitor the training process. In addition, we applied data augmentation for these patches including left and right 90-degree rotation.

3.3.2 Network Training

The proposed multi-class segmentation network was trained using the Adam optimizer [191] with a learning rate of 0.0003. The model converged after ten epochs, and the validation loss was minimized around 30 epochs. All experiments were conducted on an NVIDIA 1080Ti GPU with 11 GB memory. The batch size was kept consistent across all experiments and was set to four. The best model is chosen according to the

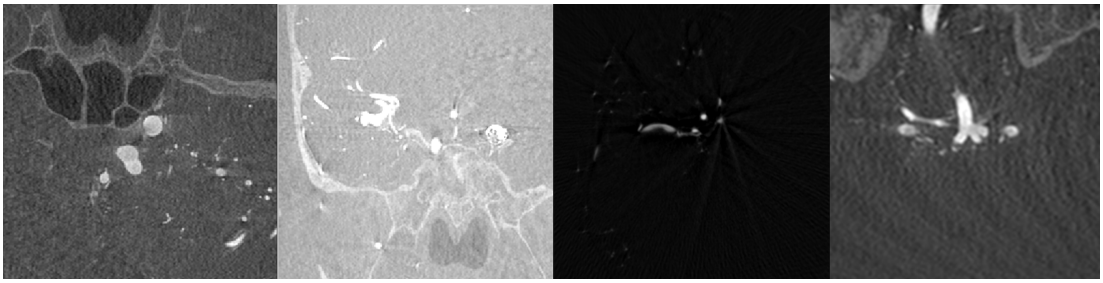

Figure 3.8: Examples of images collected in the @neurIST dataset: 2D visualization of data from different four sources showed great differences in pixel distribution and aneurysm size.

validation loss. We trained the network of Step2 in Fig. 3.2 with the above setting and shared the best weight to majority voting and self-refinement. Table. 3.1 provides the comparison of model complexity in terms of model parameters and training time/epoch, between the proposed method and the benchmarked networks. Table. 3.2 shows the hyper-parameter configurations for the benchmarked networks and our method. The loss functions of U-Net, Dual Attention Net, U-Net++, and 3DResU-Net in benchmark methods are consistent with the proposed method, i.e., the weighted dice loss. The loss function of DeepVesselNet and nnU-Net is Dice loss + CrossEntropy. We kept their original loss functions since DeepVesselNet and nnU-Net are self-contained, highly encapsulated frameworks with dedicated pre-processing and post-processing strategies, and changing their loss functions may affect their overall performance.

Table 3.1: Comparison of model complexity in terms of model parameters and training time/epoch between the proposed method and the benchmark networks.

Model complexity	Total params	Training time
U-Net++	1,857,939	78 s/epoch
Dual Attention Net	1,633,747	60 s/epoch
DeepVesselNet	1,608,147	57 s/epoch
Vessel-CAPTCHA	16,337,666	161 s/epoch
Ours	16,812,195	182 s/epoch

Table 3.2: Summary of the hyper-parameter configurations for the benchmarked networks and the proposed method.

Hyper-parameters	U-Net++	Dual Attention Net	DeepVesselNet	3DResU-Net	Vessel-CAPTCHA	nnU-Net	Ours
Optimizer	Adam	Adam	Adam	Adam	Adam	SGD	Adam
Learning Rate	0.0003	0.0003	0.0003	0.0003	0.0003	0.001 Decay	0.0003
Epochs	40	40	40	40	40	200	40
Batch Size	4	4	4	4	64	4	4
Patch Size	64	64	64	64	64	64	64
Dimension	3D	3D	3D	3D	2D	3D	3D

3.3.3 Evaluation Metrics

The proposed multi-class segmentation network is used to segment both cerebral vessels and aneurysms, denoted as y_{pred}. Several evaluation metrics are used to evaluate the similarity of the masks predicted for vessels and aneurysms individually regarding ground-truth masks (denoted as y_{true}). These include the Dice (or F_1) score, Jaccard (JAC) index, and the volume similarity (VS) index.

The Dice similarity index [192] measures overlap between y_{pred} and y_{true} and is calculated:

$$Dice = \frac{2 \times TP}{FP + 2 \times TP + FN}. \tag{3.3}$$

The Jaccard index [192] is computed as the intersection over the union of two sets and measures the similarity and diversity between two sets. It is computed:

$$JAC = \frac{TP}{FP + TP + FN}. \tag{3.4}$$

The VS index [192] measures the similarity between segmented regions of interest volumes in the predicted and ground-truth masks. It represents the absolute volume difference divided by the sum of the compared volumes.

$$VS = 1 - \frac{abs(FN - FP)}{FP + 2 \times TP + FN}. \tag{3.5}$$

The surface-to-surface distance error metrics estimate the error between the ground-truth surfaces S (the surfaces are obtained from the marching cube algorithm, the computation is based on vertices on the surface), and the segmentation prediction surfaces S'. The distance between a point p_i on surface S and the surface S' is given by the minimum of the Euclidean norm. And we compare the similarity between

the predicted and ground-truth vessel and aneurysm geometries by generating surface mesh-based representations of these structures from their corresponding masks.

$$d(p_i, S') = \min_{p' \in S'} \|p_i - p'\|_2 \tag{3.6}$$

Doing this for all N points in the ground-truth surface S gives the average surface-to-surface distance error:

$$d(S, S') = \frac{1}{N} \sum_{i=1}^{N} d(p_i, S') \tag{3.7}$$

Overall, our surface-to-surface error is computed by measuring vertex-to-vertex distances, instead of face-to-face or vertex-to-face distances.

As highlighted, the ground-truth masks available for the @neurIST dataset are only partially labeled, i.e. vessel masks cover only the major artery branches near the aneurysm rather than the entire vascular tree visible within the 3DRA image field-of-view. On the other hand, as our segmentation framework is trained patch-wise to retain fine vascular details, during inference, our approach can segment the entire vascular tree using the learned representations for identifying vascular structures within the image volume. This results in a large proportion of correctly identified pixels as vessels, for which no ground-truth labels exist. Thus, the surface-to-surface distance may be more appropriate in this scenario as it only computes the distance errors for the GT labeled region.

Across all comparative evaluations conducted comparing the proposed segmentation framework with state-of-the-art approaches and in the ablation study evaluating the impact of each module included in the proposed multi-class network, we assess the statistical significance of the obtained segmentation results using paired-sample Student's t-tests.

3.4 Results

3.4.1 Visual Comparison

Fig. 3.9 the 3D renderings of obtained segmentation results captured in the entire field of view in 3DRA images. These data samples were randomly selected from different data collection centers. As can be seen, the proposed method can capture much more abundant vascular structures in the images, which were mislabelled as the background

in manual annotations. The segmented results preserve the continuity and topology of the vascular trees and are visually comparable to the annotated regions.

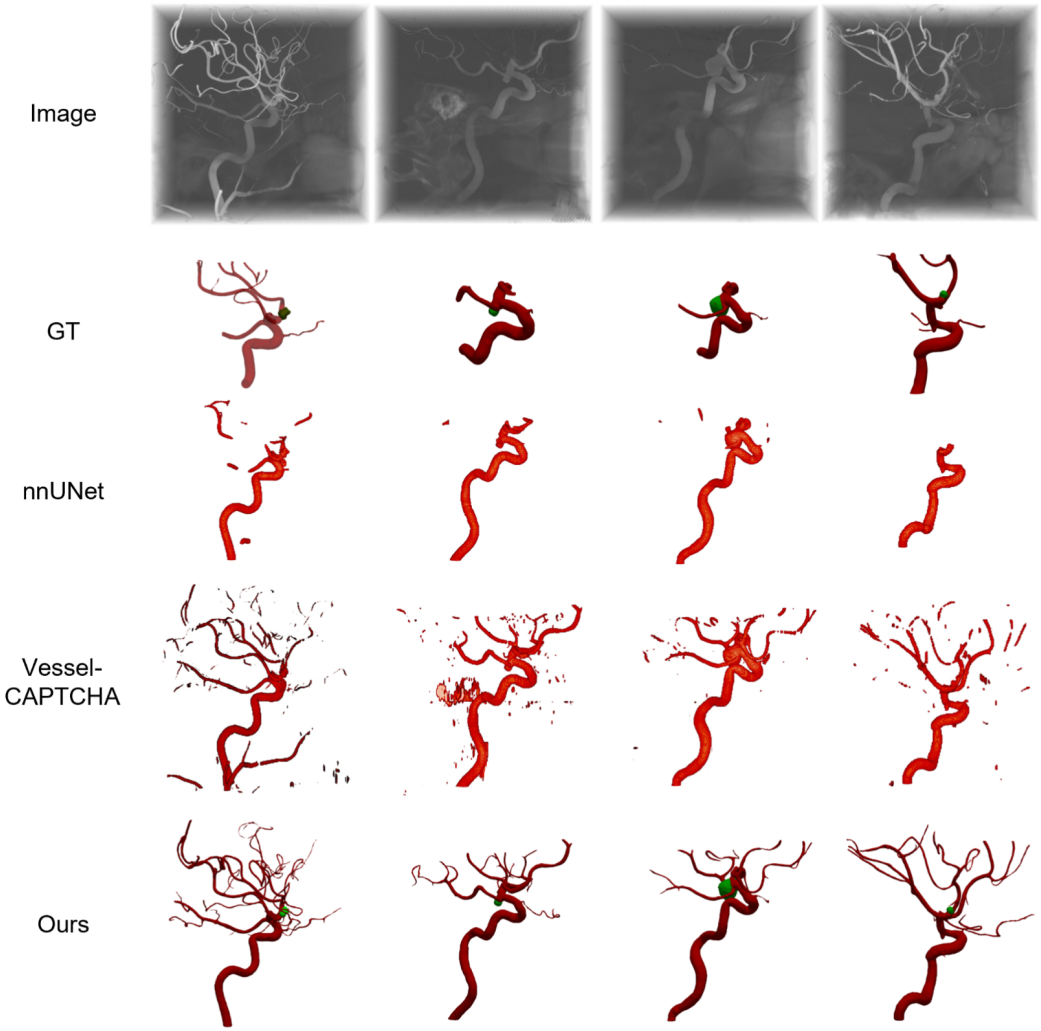

Figure 3.9: 3D renderings of obtained segmentations. These data samples were selected from different centers.

The surface meshes generated using the vessel and aneurysm segmentations predicted at different stages of our segmentation pipeline are shown in Fig. 3.10. These figures demonstrate the utility of the proposed post-processing steps to suppress false-positive predictions for the aneurysm and refine the same. Cropped vessels share similarities in topology and appearance with aneurysms near patch boundaries. Therefore,

initial segmentation using the proposed multi-class segmentation network (step 2 in Fig. 3.2) is prone to incorrectly labeling tortuous vessels and vessels near patch boundaries as aneurysms (example in the third column of Fig. 3.10). These false-positive predictions for aneurysms are artifacts of patch-based learning due to the limited spatial context available to the network during feature learning and can be effectively reduced using majority voting (described as part of the post-processing step earlier in section 3.2.3-C). The resulting aneurysm segmentation following the suppression of false positives by majority voting (see the fourth column of Fig. 3.10) is used to provide aneurysm location and extract patches in the neighborhood, which are fed back into the multi-class network to refine the segmentation near the aneurysm (called' self-refinement). The improvement in aneurysm segmentation accuracy afforded by these two post-processing steps involving majority voting and self-refinement is also highlighted for the test set in Fig. 3.10.

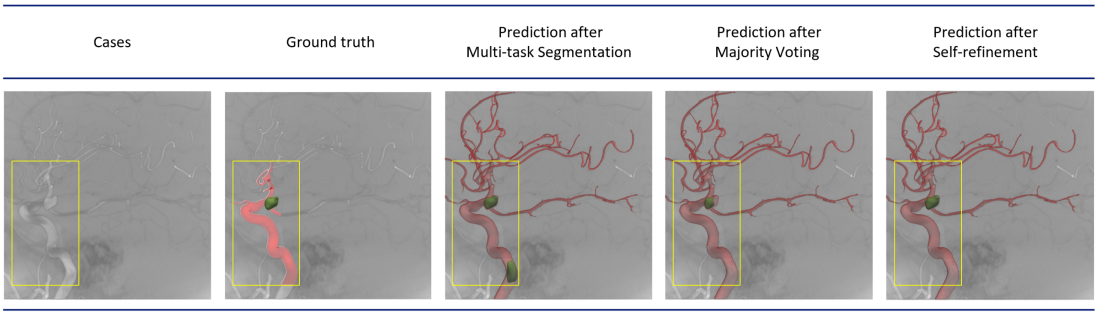

Figure 3.10: 3D renderings of obtained segmentations after different steps.

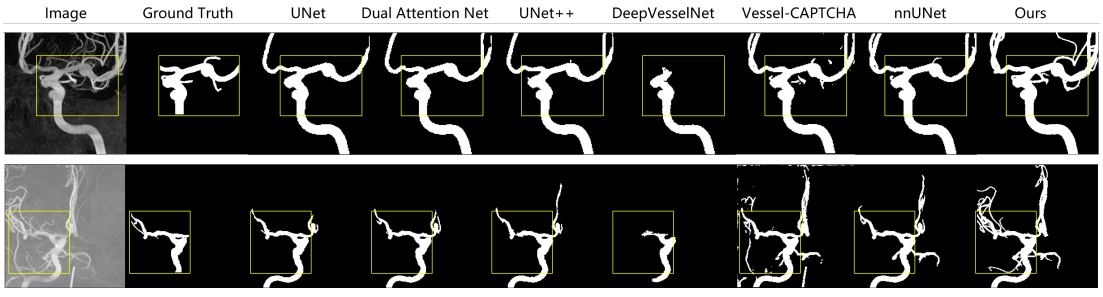

Figure 3.11: Maximum intensity projection for vessel segmentation. The yellow box is the golden standard area, where all quantitative evaluations are carried out. Our method captures more fine vascular structures than its state-of-the-art counterparts.

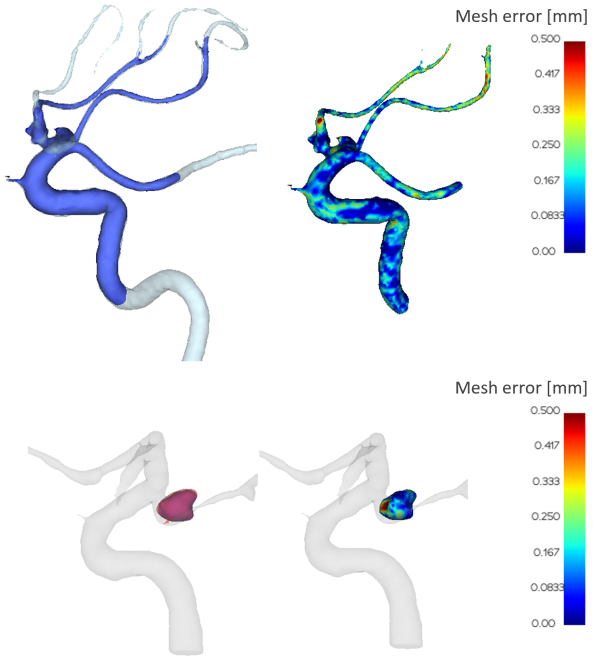

Figure 3.12: Surface-to-surface error. This error is calculated based on vertex-to-vertex distance. Upper left: vessel overlap of ground-truth (blue) and prediction (translucid white). Upper right: surface-to-surface vessel error. Bottom left: aneurysm overlap of ground truth (red) and prediction (orange). Bottom right: aneurysm surface-to-surface error.

Visual comparisons of the maximum intensity projections of segmentations predicted using our approach, and those predicted by state-of-the-art techniques for two samples from the @neurIST test set are presented in Fig. 3.11. These figures highlight our approach's ability to preserve fine vascular structures greater than its state-of-the-art counterparts. In the @neurIST dataset, ground-truth masks are available only for large vessels near the aneurysm in the second column of Fig. 3.11. Therefore, for every case, after getting the final segmentation result, we cropped the prediction into a mask with the same size and position as the ground truth. Then the cropped mask is quantitatively evaluated with ground truth. Due to the lack of annotation of small vessels, the image-based assessment does not provide the most reasonable evaluation of the segmentation effect. In addition to evaluating segmentation quality using image-based metrics such as Dice, we computed the surface-to-surface distance error between the

predicted and ground-truth meshes by reconstructing the former just within the field of view of the corresponding ground-truth mask. The surface-to-surface error metric also provides information regarding the spatial distribution of errors across the anatomical structures of interest, i.e., mapping the vertex-wise nearest neighbor distances between the predicted and ground-truth meshes onto each former vertex (as shown in Fig. 5.1). Evaluation of surface-to-surface errors in this manner thus provides spatial context to where segmentation errors are incurred and help quantify localized errors, complementing other global image-based metrics (such as Dice) used to evaluate segmentation performance.

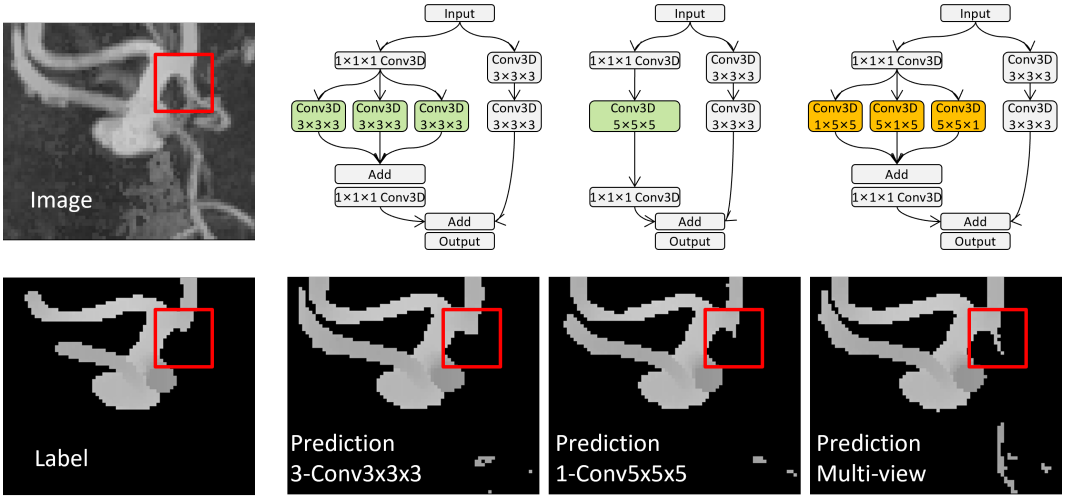

Figure 3.13: Comparison of maximum intensity projection between multi-view block (yellow), three conv3D-3×3×3 layers, and one conv3D-5×5×5 layer (green). The multi-view block can predict more small vessels even not annotated in the label.

Each block of our encoder is dual-path, one for 3D convolution and the other for 2D convolution in multi-view blocks. In Fig. 3.13, to verify that the role of the multi-view block is to provide additional fine features to the 3D convolution backbone pathway rather than adding more parameters, we replaced the three orthogonal 2D convolution layers in the multi-view block with three 3D convolution layers in kernel size 3 or one 3D convolution layer in kernel size 5. The multi-view block has fewer parameters than the other two settings but can mine additional features that are different from the 3D convolution features. Comparing the predictions in the red box, the multi-view block can predict more small vessels even those not annotated on the label. We have also

performed an ablation study with and without the cascaded transformer and visualized the attention maps overlaid on the original images, as shown in Fig. 3.14. As can be seen by comparing each case in the same column, with the cascaded transformer, the model reduces focus on irrelevant context structures like vessel bend and bifurcation. Besides, the aneurysm necks are better identified with the attention of the cascaded transformer.

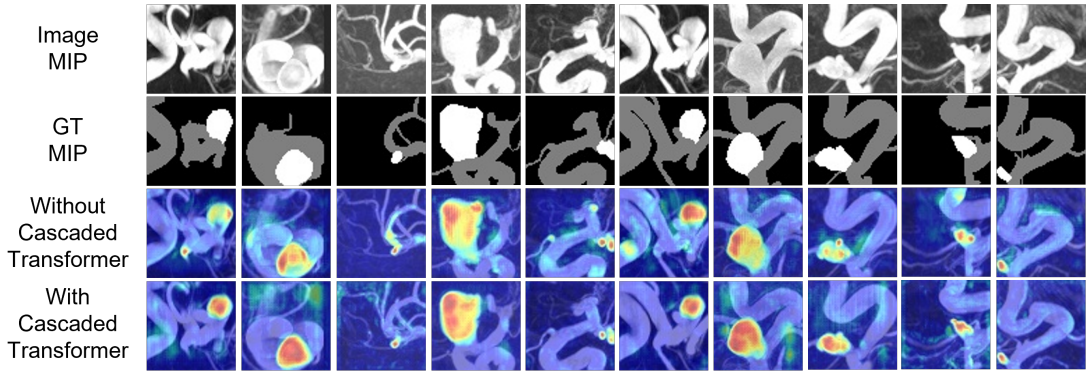

Figure 3.14: Visual comparison of attention maps with and without cascaded transformer. The first, second, third, and fourth rows are the image MIPs, vessel and aneurysm segmentation ground truth, attention maps without the cascaded transformer, and attention maps with the cascaded transformer, respectively. As can be seen by comparing each case in the same column, with the cascaded transformer, the model reduces focus on irrelevant context structures like vessel bend and bifurcation.

3.4.2 Quantitative Evaluation

Quantitative metrics summarize our approach's segmentation performance and the state-of-the-art, namely, U-Net++, Dual Attention Net, DeepVesselNet, and nnU-Net, across all test samples from the 5-fold cross-validation experiments conducted using the @neurist dataset is presented in Table 3.3. All the quantitative evaluation results were calculated within a bounding box encapsulating the GT labeled region (yellow box in Fig. 3.11). These results indicate that our method consistently outperforms all competing methods in the Dice score, Jaccard index, VS index, and the average surface-to-surface error across the 5-fold cross-validation experiments conducted. The statistical significance of the obtained segmentation results was evaluated (using paired-sample Student's t-tests), revealing that our approach achieved significant improvements over

Table 3.3: Compare with state-of-the-art on whole @neurIST dataset. 5-fold cross-validation experiments were conducted for each method. The results were calculated in the golden standard area. Our method outperforms other methods on the main evaluation metrics of segmentation and is statistically significant on most items.

Method	U-Net++[185]		Dual Attention[189]		DeepVesselNet[178]		nnU-Net[46]		Ours	
Anatomical Structure	Vessel	Aneu	Vessel	Aneu	Vessel	Aneu	Vessel	Aneurysm	Vessel	Aneu
Dice	0.8982± 0.1091	0.5980± 0.3398	0.8893± 0.1380	0.5949± 0.3372	0.8918± 0.0968	0.1558± 0.2347	0.8687± 0.1197	0.6686± 0.3190	**0.9125**± 0.0759	**0.8163**± 0.2672
Jaccard Index	0.8271± 0.1225	0.5035± 0.3216	0.8184± 0.1453	0.4993± 0.3205	0.8147± 0.1174	0.1083± 0.1899	0.7820± 0.1371	0.5745± 0.3058	**0.8455**± 0.0935	**0.7486**± 0.2655
Volume Similarity	0.9461± 0.0999	0.6579± 0.3372	0.9453± 0.0951	0.6634± 0.3338	0.9455± 0.0740	0.2619± 0.2767	0.9131± 0.1074	0.7321± 0.3072	**0.9525**± 0.0489	**0.8693**± 0.2520
Surface-to-surface error (mm)	0.3441± 0.4422	1.1398± 2.3495	0.3518± 0.4668	1.0623± 2.3829	0.4227± 0.7354	4.3694± 7.4209	0.8903± 1.2450	1.0611± 2.9100	**0.2586**± 0.3066	**0.2021**± 0.1790
p-value (Dice)	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	/	/
p-value (Surf)	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	< 0.05	/	/

Table 3.4: Ablation study. These models are in the same pre-processing and post-processing method (proposed). The experiments remove different modules separately. Due to the dataset being partially labeled, the Dice values cannot measure unlabeled fine vessels and small aneurysms, so we use surface-to-surface error to measure the performance on labeled parts.

	Surface-to-surface error (mm)					
Model	Model 1	Model 2	Model 3	Model 4	Model 5	Ours
Absent module	Nested Block	Multi-view	Transformer	L-Down	Wide	
Modules	U-Net Multi-view Transformer L-Down Wide	U-Net++ Transformer L-Down Wide	U-Net++ Multi-view L-Down Wide	U-Net++ Multi-view Transformer Wide	U-Net++ Multi-view Transformer L-Down 	U-Net++ Multi-view Transformer L-Down Wide
Vessel	0.3127 ± 0.2462	0.3329 ± 0.3466	0.3327 ± 0.3258	0.3286 ± 0.2876	0.3137 ± 0.2536	$\mathbf{0.2586 \pm 0.3066}$
Aneurysm	0.4630 ± 1.0255	1.0878 ± 4.2235	0.3658 ± 0.7354	0.2738 ± 0.3327	0.3921 ± 0.6521	$\mathbf{0.2021 \pm 0.1790}$

the state-of-the-art in terms of Dice and the average surface-to-surface error metrics for both vessels and aneurysms. For vessel segmentation, all methods achieved a Dice score higher than 0.85, indicating that all methods (including ours) were well suited to this task. On the other hand, aneurysm segmentation was more challenging as the target region often constitutes less than 1 % of the overall image volume. We found that the state-of-the-art methods investigated in this study failed to perform adequately on this task.The nnU-Net achieved the best results among the state-of-the-art methods are dedicated pre-processing and post-processing approaches, with a Dice of 0.67 and surface-to-surface error of 1.06 mm. Our approach provided the best aneurysm segmentation performance, achieving a 15 % and 0.86 mm improvement over nnU-Net in terms of aneurysm Dice and surface-to-surface error.

Tables 3.4 summarizes the results from the ablation study conducted to evaluate the impact of each module included in the proposed multi-class segmentation network on the quality of the predicted vessel and aneurysm segmentation, respectively. The ablation studies which remove modules separately were conducted to verify whether each module contributed positively to the final segmentation performance and determine the importance of different modules. Due to the dataset being partially labeled, the Dice values can not measure unlabeled fine vessels and small aneurysms, so we use surface-to-surface error to measure the performance on labeled parts. Model 1 to Model 5 represent the cases where one module is removed from the final model. The results comparison between ours with Model 1 to Model 5 show that the surface-to-surface error is increased no matter which module is discarded, proving that each adopted technique contributes to the improved accuracy of vessel segmentation. The final model yields the smallest surface-to-surface error in aneurysm segmentation. The absence of certain modules can lead to completely incorrect segmentation results, thus reducing overall segmentation performance on a test set, so each module is essential for aneurysm and vessel segmentation.

We also compared single-class segmentation networks (trained individually for aneurysm and vessel segmentation) with the multi-class network, where identical network architectures were used except for the output layer. Results from this comparison are presented in Table 3.5, which highlights the added advantage of multi-class learning relative to training independent networks for segmenting each structure individually. Our multi-class network significantly outperforms the single-class network for aneurysm

average surface-to-surface error (0.2021 vs 0.7051). However, for vessel segmentation, the multi-class network only provided a marginal improvement over the single-class network. Since aneurysms grow on blood vessels, the aneurysm part is also labeled as part of the vessel during training, the learning of vascular features will have a positive effect on the extraction of aneurysm features.

Table 3.5: Comparison between single-class and multi-class. The pre-processing and post-processing of the experiments are the same.

Training type	Single-class		Multi-class	
Anatomical Structure	Aneurysm	Vessel	Aneurysm	Vessel
Surface-to-surface error (mm)	0.7051± 0.9031	0.3024± 0.1641	0.2021± 0.1790	0.2586± 0.3066

The proposed approach was also trained and evaluated on the public cerebral aneurysm segmentation (CADA-AS) challenge dataset [187]. The segmentation performance of our approach was compared against the best-performing methods in this challenge, 3DResU-Net. When comparing these methods, instead of using $128 \times 128 \times 128$ patch size described in the challenge paper, an input patch size of $64 \times 64 \times 64$ was used to analyze and segment the images due to limited computing resources of 1080ti GPU card. Results summarizing aneurysm segmentation performance on test data from the CADA-AS challenge are presented in Table 3.6. Since the data is fully labelled, we add the Dice Similarity Index and surface-to-surface error. These results indicate that our approach outperforms 3DResU-Net in terms of both metrics.

Table 3.6: Compare our method on the cerebral aneurysm segmentation (CADA-AS) dataset with its champion methods. All models were retrained and tested on this single-class dataset.

CADA-AS	3DResU-Net	Ours
Dice Similarity Index	0.7464± 0.1379	0.8737 ± 0.0747
Surface-to-surface error (mm)	0.4102± 0.3924	0.3817 ± 0.3984

A key aspect of quantitative analysis of cerebral aneurysms, either in assessing cereb-

ral hemodynamics or aneurysm rupture risk, is the precise characterization of their morphological properties. Hence, besides evaluating aneurysm segmentation quality using standard image-based and mesh-based metrics, we compared our approach with state-of-the-art segmentation approaches to preserve each segmented aneurysm's maximum diameter and volume relative to the ground truth. The Bland-Altman plots in Fig. 3.15, summarize the average errors between predicted and ground-truth measurements for aneurysm maximum diameter and volume and their corresponding 95 % confidence intervals for each method investigated. These results indicate no apparent bias in our model's ability to preserve critical morphological characteristics of aneurysms, unlike DeepVesselNet, for example. Additionally, the average errors incurred by our approach regarding the ground-truth measurements are consistently lower than all competing approaches.

Table 3.7: False-positive prediction rate for the aneurysm with or without post-processing.

Post-processing	False-Positive Rate	Dice Similarity Index
None	65/223	0.6432 ± 0.3333
Ours	23/223	0.8163 ± 0.2672

In addition, Table. 3.7 has included the false-positive rate and Dice score for the aneurysm segmentation before and after applying the post-processing technique. As can be seen, the proposed post-processing can fix 42 over-segmented aneurysm cases while improving the overall Dice for aneurysm by 15 %. Table. 3.8 illustrates the average prediction time for one volumetric image generated by the proposed framework in different steps. The proposed method uses an average of 1.5 minutes to process an image volume for vessel and aneurysm segmentation.

Table 3.8: Average prediction time for one image volume.

Step	Time
Step1: Preprocessing	5 s
Step2: Multi-class Segmentation	45 s
Step3: Post-processing	46 s
Total	96 s

3.5 Discussion

This chapter presents a multi-class convolution neural network and a 3D patch-based pipeline for cerebrovascular and aneurysm segmentation on 3DRA images. Vessel and aneurysm segmentation in 3DRA is very challenging due to the small percentage of vessels and aneurysms and the interference of the skull. Compared to standardized, homogeneous data, clinical data exhibit more noise, heterogeneous, diverse appearance and resolution, making this task more challenging. With severe class imbalance, automatic segmentation methods struggle to extract the complete contextual and local information from images. To alleviate those issues, our proposed network has a transformer block sensitive to small-scale features, multi-view blocks sensitive to continuous features, the learnable downsample block that prevents subtle features from being lost, and wide blocks that expand local perceptual fields. Besides, the dedicated post-processing methods of majority voting and self-refinement can effectively suppress the over-segmentation of the clinical aneurysm, enhancing the entire pipeline's clinical robustness.

3.5.1 Class Imbalance

In whole brain tissue, the percentage of blood vessels is less than 6% [188], and the proportion of aneurysms is less than that of vessels. Hence, U-Net++ and Dual Attention Net are potent models that retain rich intermediate features and focus more on critical target information. However, these models can only capture the main vessels and obvious aneurysms. Still, these models lose subtle features during the convolution operation, viz., small aneurysms whose diameter is less than 10 mm. Our method designed multi-view blocks and wide blocks for the class imbalance problem by extracting additional information to complement the backbone network feature. These blocks can extract continuous information through cascade and parallel low-dimensional convolution layers with large kernels. We also exploited a transformer block at the end of the encoder to enlarge the proportion of target features. In Table 3.4, after adding different new modules, the segmentation of both aneurysms and vessels is improved, which also verifies that the proposed modules positively affect the final segmentation results. In addition, in order to reduce the proportion of negative samples and thus make the training converge, when preparing the training patches, we select the patches around the aneurysm as input instead of feeding all patches obtained after cropping

the data to a model. This can increase the proportion of aneurysms and vessels in a single patch.

3.5.2 Inter-class Separability

Aneurysm segmentation usually suffers from inter-class interference. Due to the prior knowledge that all aneurysms grow on vessels while extracting the vessel features, the deep network also extracts and enhances more potential aneurysm features near the vessels. Without proper guidance of vessel features in multi-class, the single-class model may segment brain tissue or noise area into aneurysms. The multi-class learning brings a huge boost to aneurysm segmentation. From Table 3.5, multi-class aneurysm segmentation surface-to-surface error improved by 0.5 mm over single-class.

3.5.3 Inter-institutional Data Variability

Our data were obtained from four institutions. While data from different sources are all from the same modality, viz. 3DRA images, there are large differences in image appearance, intensity distribution, resolution and aneurysm size (Fig. 3.8). This poses a great challenge to the robustness of the model. Automatic models such as nnU-Net and Deep Vessel Net use pre-processing and post-processing methods like patch-based learning and Gaussian standardization. These methods perform well on challenging datasets with good pre-processing. However, these segmentation methods lead to aneurysm over-segmentation in clinical data. To improve clinical robustness, we propose majority voting, which returns the prediction most like an aneurysm to find the exact aneurysm localization. Then, self-refinement further corrects the aneurysm contour details. Through our experiment, we found the necessity of the post-processing step in our current framework setting because 1) it helps mitigate the over-segmentation of aneurysms due to the morphological similarity between aneurysms and vessel endings at the patch boundaries; 2) it avoids to use of larger patch sizes that would further aggregate the data imbalance and result in performance degradation in aneurysm segmentation.

The fourth column in Fig. 3.2 top shows majority voting returns the maximum prediction probability and effectively suppresses over-segmentation. The later self-refinement makes the aneurysm details more accurate. To validate the robustness of the pipeline, besides clinical data, we also evaluated our method on the CADA-AS

competition dataset (c.f. Table 3.6). Our segmentation method yielded results close to the championship method [187]. For in silico clinical trials, the accuracy of aneurysm localization is as important as Dice and Surface-to-surface error. Table 3.9 shows the success rate of aneurysm prediction for each algorithm. Only cases with a surface-to-surface error of less than 1 mm were defined as success cases, which means that such cases have accurate localization and segmentation. In 223 clinical cases from four different data centers, our method yielded accurate aneurysm prediction in 190 cases. However, among other methods, the best is nnU-Net which only got 173 success cases. In addition, our method outperforms these comparative methods in terms of clinical indicators such as aneurysm diameter and volume. The Bland-Altman plots in Fig. 3.15 demonstrate that our method yields a difference of -0.03 ± 0.54 mm and -4.4 ± 71.1 mm^3 in aneurysm diameter and volume with ground truth, which is the smallest (best) compared to other methods.

Table 3.9: Compare our method on aneurysm segmentation success rate (surface-to-surface error smaller than 1 mm) with SOTA methods.

Methods	Aneurysm segmentation success rate
U-Net++	150 / 223
Dual Attention Net	152 / 223
DeepVesselNet	144 / 223
nnU-Net	173 / 223
Ours	190 / 223

Although our method achieved improved performance for automatic segmentation of vessels and aneurysms, due to the limitation of incomplete labeling of 3DRA datasets, the wrongly labeled background pixels for missing vessels could interfere with the overall training process. Thus, future work would involve leveraging semi-supervised schemes to enhance the learning of unlabelled parts, e.g., relabelling the missing annotations during the training process by introducing pseudo-labels. Meanwhile, since there are still a large number of unlabelled 3DRA image data in our clinical dataset, the joint training of the labeled 3DRA data and unlabeled data under a semi-supervised setting is also a worthy direction of research.

There is still room for improvement in 3DRA vessel and aneurysm segmentation. One key limitation of current methods is the constraint on computational resources,

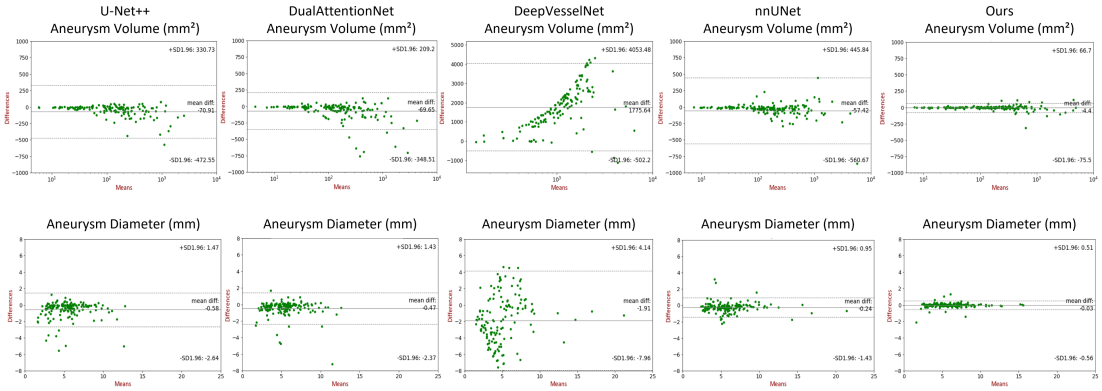

Figure 3.15: Bland-Altman plots: Mean and difference of aneurysm radius and volume between ground-truth and predictions. Our method has a more compact distribution. In terms of clinical criteria, our predictions are much closer to the ground truth.

which restricts the patch size to $64 \times 64 \times 64$. This patch-based approach introduces challenges, particularly when vessels near patch boundaries interfere with aneurysm segmentation due to limited contextual information. If computational resources allow, incorporating full-size 3D images into the segmentation pipeline (rather than relying on patches) could significantly enhance performance. A global view of the entire vascular structure would reduce boundary effects, improve the model ability to differentiate aneurysms from vessels, and ultimately lead to a more accurate and robust segmentation. Hence, transitioning from patch-based processing to full-volume segmentation could be a promising direction for future advancements.

3.6 Conclusion

This work proposed a 3D patch-based multi-class model for vessel and aneurysm segmentation on 3DRA images. The proposed approach addressed class imbalance problems and inter-class interference problems in multi-class segmentation. Experimental results showed that the proposed method outperformed several popular state-of-the-art approaches for tackling similar challenges, such as U-Net++, DeepVesselNet, and nnU-Net. This work aims to alleviate class imbalance and inter-class interference, which are common and challenging problems in cerebrovascular and aneurysm segmentation. The deliberately designed network architectures such as the cascaded transformer, multi-view block, and wide block as well as the proposed post-processing strategies of the

majority voting and self-refinement contribute positively to mining vascular and aneurysm features through the proposed end-to-end trainable network. The aforementioned issues are also present in brain MRA and CTA when it comes to cerebrovascular and aneurysm segmentation. The proposed model is generic and can be applied to mitigate the issues of class imbalance and inter-class interference in brain MRA and CTA, promising to facilitate accurate clinical analyses. The systematic evaluation of the model performance on other modalities would be the scope of future work.

Chapter 4

Semi-supervised Segmentation

Accurate segmentation of brain vessels is crucial for cerebrovascular disease diagnosis and treatment. However, existing methods face challenges in capturing small vessels and handling datasets that are partially or ambiguously annotated. In this chapter, we propose an adaptive semi-supervised approach to address these challenges. Our approach incorporates innovative techniques including progressive semi-supervised learning, adaptative training strategy, and boundary enhancement. Experimental results on 3DRA datasets demonstrate the superiority of our method in terms of mesh-based segmentation metrics. By leveraging the partially and ambiguously labelled data, which only annotates the main vessels, our method achieves impressive segmentation performance on mislabelled fine vessels, showcasing its potential for clinical applications.

4.1 Introduction

Accurate segmentation of cerebral vessels is clinically significant as it provides crucial anatomical information for the diagnosis and assessment of cerebrovascular diseases [193]. Furthermore, segmenting small vessels is essential as they play important roles in brain function and pathological processes. Accurate segmentation of small vessels provides comprehensive morphological information about the vascular network, facilitating patient-specific modeling of cerebral hemodynamics, which can be used to better understand pathologies, plan interventions, and design treatment devices [194, 166, 195, 196].

However, the task of accurate vessel segmentation is challenging due to several reasons. Firstly, the small proportion of vessels in brain tissue makes segmentation difficult, particularly for small arterioles [197, 198]. To address this, convolution-based methods [46] designed for medical imaging are enhanced in segmentation capability through experienced pre-processing and post-processing techniques. Further, transformer-based methods [104, 93] have been proposed to leverage fully supervised learning to explore the features of small targets in-depth. Secondly, clinical vessel annotations are focused only on regions surrounding pathologies such as aneurysms [199, 200], and only the main vessels are labelled, leaving out fine vessels. This ambiguously labelled data negatively impacts the performance of fully supervised learning approaches. Semi-supervised learning methods with pseudo-labeling techniques [201, 202] have been proposed to overcome this limitation. Thirdly, clinical images often exhibit high noise levels [186], and there are significant variations in pixel distribution across different imaging cen-

ters. Traditional semi-supervised methods[203, 204, 205, 206] using pseudo-labeling [201] tend to overly incentivize the confidence of the model in vessel segmentation, leading to excessive over-segmentation [149].

Therefore, we propose the adaptive semi-supervised model in Fig.4.1, aiming to address the challenge of partially annotated intracranial vessel segmentation. The model employs a Teacher-Student structure, with the Swin-UNet [104] serving as the backbone network. We partition the partially annotated data into labelled patches and unlabelled patches, which are fed into the teacher and student networks, respectively. Instead of the conventional approach of directly supervising the student network in knowledge distillation, the teacher network learns vessel knowledge from the labelled patches and teaches it to the student network. Additionally, the teacher network's output is used as refined pseudo-labels for further learning by the student network. The key innovations are as follows:

- We introduce the adaptive semi-supervised model, utilizing a progressive semi-supervised learning strategy. Ground truth is used to teach the teacher network, and the teacher network, in turn, instructs the student network, leading to incremental improvements in segmentation performance.

- We propose addressing the challenges associated with semi-supervised learning through unsupervised domain adaptation techniques. This enables the adaptation of knowledge from labelled patches to unlabelled patches without any domain shift.

- We introduce the Fourier high-frequency boundary loss. Except for Dice and Cross-Entropy loss, we extract the high-frequency boundary features using the Fourier transform and calculate their mean squared error.

- We introduced a data augmentation technique called adaptive histogram attention (AHA) to address the variations in pixel distribution within clinical data. AHA enables the model to better focus on discriminating between other brain tissues and vessels, facilitating the extraction of vessel structural features.

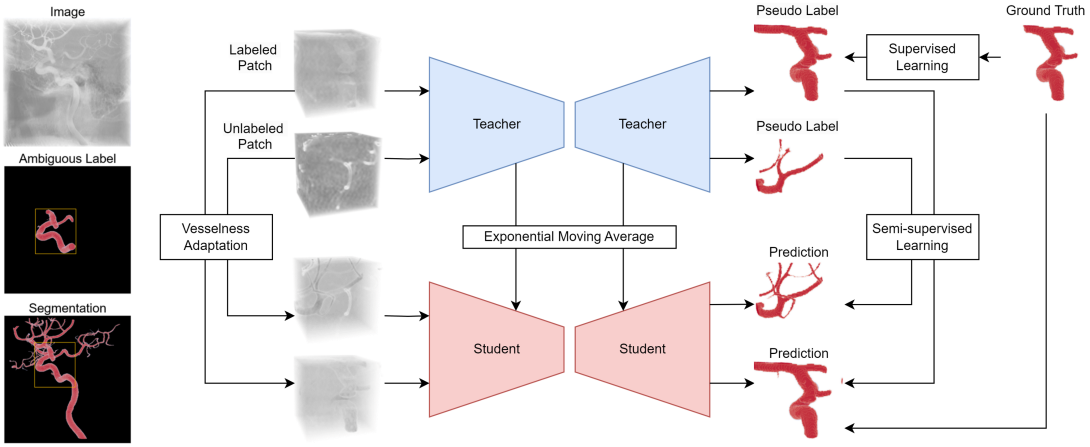

Figure 4.1: Schematic of the proposed adaptive semi-supervised model.

4.2 Methodology

4.2.1 Preprocessing

The purpose of preprocessing is not only to remove noise but also to facilitate the model in extracting size, structural, and generalization features specific to the vessels.

Resolution standardization: To tackle the resolution inconsistencies in clinical data, all data is standardized to a spacing of 0.35mm/pixel. This allows the model to learn the size/shape features in cases with initially different resolutions.

Adaptive Histogram Attention: The distribution patterns in 3DRA data histograms show that vessels are typically in the higher pixel value range, brain tissues in the middle, and backgrounds in the lower range. While deep learning models can easily distinguish between the background and vessels, they may struggle with distinguishing between brain tissues and vessels, leading to over-segmentation. AHA tackles this by identifying areas of abrupt shifts in the histogram, using them for normalization, effectively eliminating the background from the histogram, thereby enabling the model to focus on distinguishing between vessels and other brain tissues. This method emphasizes the extraction of structural features rather than mere threshold-based features.

Patch Grouping: We extract overlapped 3D patches from both the annotated and unannotated regions. These two groups of patches are subsequently fed into the teacher and student networks, respectively. This approach enables the model to learn local generalized vessel features rather than specific fitting features of individual cases.

4.2.2 Problem Formulation

In this chapter, we have datasets sampled from two groups. The labelled group contains labelled patches $\mathcal{D}_l = \left\{\left(x_i^l, y_i^l\right)\right\}_{i=1}^{N_l}$, and the unlabled group contains unlabelled patches $\mathcal{D}_u = \{(x_i^u)\}_{i=1}^{N_u}$. We use adaptive histogram attention to get labelled and unlabelled vessel-like patches $\hat{\mathcal{D}}_l = \left\{\left(\hat{x}_i^l\right)\right\}_{i=1}^{N_l}$ and $\hat{\mathcal{D}}_u = \{(\hat{x}_i^u)\}_{i=1}^{N_u}$. Our model consists of teacher and student networks. We update the weights in the student network (encoder F_s, decoder G_s) as an exponential moving average (EMA) of weights in the teacher network (encoder F_t, decoder G_t) to ensemble the information in different training steps. The prediction of teacher network on labelled and unlabelled patches are denoted as $p_i^l = G_t\left(F_t\left(x_i^l\right)\right)$ and $p_i^u = G_t\left(F_t\left(x_i^u\right)\right)$. We also denote the prediction of the student network on labelled and unlabelled vessel-like patches as $\hat{p}_i^l = G_s\left(F_s\left(\hat{x}_i^l\right)\right)$ and $\hat{p}_i^u = G_s\left(F_s\left(\hat{x}_i^u\right)\right)$. Our goal is to learn a task-specific student network using F_s and G_s to accurately predict labels on test data from the unlabelled patches.

4.2.3 Supervised Learning

In the teacher network, labelled patches $\mathcal{D}_l$ are passed through the CNN-based feature extractor F_t, which are then passed through the task-specific segmentation generator G_t to minimize the supervised loss $\mathcal{L}_{full\_sup}$ which includes cross-entropy $\mathcal{L}_{CE}$, Dice similarity coefficient loss $\mathcal{L}_{DSC}$, and boundary loss $\mathcal{L}_B$.

$$\mathcal{L}_{full\_sup} = \mathcal{L}_{CE} + \mathcal{L}_{DSC} + \mathcal{L}_B \tag{4.1}$$

$$\mathcal{L}_{CE} = -\frac{1}{N_l}\sum_{i=1}^{N_l} y_i^l \log\left(p_i^l\right) \tag{4.2}$$

$$\mathcal{L}_{DSC} = \frac{1}{N_l}\sum_{i=1}^{N_l}\left(1 - \frac{2\left|p_i^l \cap y_i^l\right|}{|p_i^l| + |y_i^l|}\right) \tag{4.3}$$

$$\mathcal{L}_B = \frac{1}{N_l}\sum_{i=1}^{N_l}\left(H\left(p_i^l\right) - H\left(y_i^l\right)\right)^2 \tag{4.4}$$

$$H\left(p_i^l\right) = \mathcal{F}^{-1}\left(\mathcal{F}\left(p_i^l\right) \cdot \mathbf{1}_{mask}\right) \tag{4.5}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the Fourier transform [207] and the Fourier inverse transform respectively. The high-frequency filter mask $\mathbf{1}_{mask}$ with value zero in the middle and value one on the edge has the same shape as p_i^l.

4.2.4 Semi-supervised Learning

To perform alignment at the instance level, the adaptive vessel-like labelled and unlabelled patches are passed through the teacher network to get segmentation prediction p_i^l and p_i^u. Meanwhile, we input labelled and unlabelled patches to HSDA to get new patches, and they are passed through the student network to get segmentation prediction $\hat{p}_i^l$ and $\hat{p}_i^u$. Next, we employ the mean square error (MSE) [208] and cosine similarity [209] as defined in Eq 5.3 and Eq 4.8 to reduce the discrepancy between the two predictions and thus increase the vessel invariance of the student model.

$$\mathcal{L}_{semi\_sup} = \mathcal{L}_{mse} + \mathcal{L}_{sim} \tag{4.6}$$

$$\mathcal{L}_{mse} = \frac{1}{N_l}\sum_{i=1}^{N_l}\left(p_i^l - \hat{p}_i^l\right)^2 + \frac{1}{N_u}\sum_{i=1}^{N_u}\left(\hat{p}_i^u - p_i^u\right)^2 \tag{4.7}$$

$$\mathcal{L}_{sim} = \frac{1}{N_l}\sum_{i=1}^{N_l} h\left(p_i^l, \hat{p}_i^l\right) + \frac{1}{N_u}\sum_{i=1}^{N_u} h\left(\hat{p}_i^u, p_i^u\right) \tag{4.8}$$

$$h(u, v) = \exp\left(\frac{u^T v}{\|u\|_2 \|v\|_2}\right) \tag{4.9}$$

The weight ratio between fully supervised and semi-supervised losses is 4:1. We prioritize the fully supervised loss to ensure training robustness and prevent the network from becoming overly confident and introducing noise during the initialization stage.

4.3 Experiments and Results

4.3.1 Datasets

In our experiments, we utilized 3D rotational angiography (3DRA) modality dataset Aneurist [186], which comprises 223 partially annotated 3D brain vessel images. These images were acquired from four different centers using different scanners and imaging protocols. As shown in Fig. 4.2, there are significant variations in image appearance and resolution across the data from different centers. we trained our models using full-size images and partially and ambiguously annotated labels. Due to the incomplete annotations, quantitative analysis was performed within the bounding box of annotated regions, while qualitative analysis was conducted across the entire image.

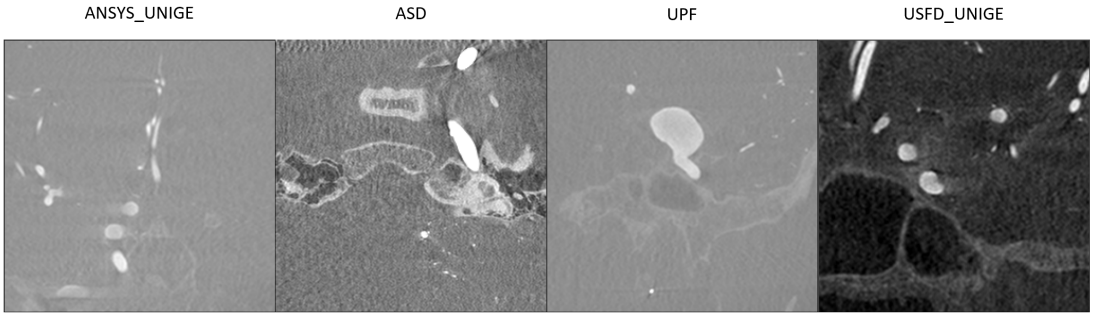

Figure 4.2: Examples of 3DRA images in grayscale collected from AneurIST dataset: 2D visualization of data from four different sources showed great differences in pixel distribution and noise levels.

4.3.2 Experimental Setup

The experiments were conducted using a high-performance computing setup. We utilized an NVIDIA GeForce RTX 3090 GPU with 24GB of VRAM. The experimental system was equipped with a high-capacity RAM of 128GB, enabling the handling of large datasets and memory-intensive tasks.

Our proposed adaptive semi-supervised model was implemented based on the Swin-UNet architecture [104], serving as the backbone of both teacher and student networks. During training, we employed a batch size of 1 and utilized patch-based learning with a patch size of [128, 128, 128]. The models were trained for 100 epochs, during which the optimization was performed using the Adam optimizer. We employed data augmentation techniques, such as random rotations and flip, to enhance model generalization. The learning rate was initially set to 0.001, and a learning rate decay strategy was applied, reducing the learning rate by a factor of 0.1 every ten epochs. The parameters of the teacher network are updated normally, and the parameters of the student network are updated according to the exponential moving average (EMA) [210]. To achieve the effect of the EMA during the training process, we employed the no gradient decorator to ensure that gradient calculations are not performed during the EMA process. Prior to optimizing the parameters, we updated the parameters of the teacher network by invoking the EMA function in Eq 5.10. In this function, the weight factor *decay* is calculated based on the iteration count and the initial decay rate of 0.999 in Eq.5.11. The student network parameters are updated by applying the weight factor *decay*, thereby

incorporating the knowledge learned by the teacher network gradually.

$$\mathcal{W}_{\text{stu}} = decay \times \mathcal{W}_{\text{stu}} + (1 - decay) \times \mathcal{W}_{\text{tea}} \tag{4.10}$$

$$decay = \min\left(1 - \frac{1}{iteration \times 10 + 1}, decay\right) \tag{4.11}$$

The image data from these databases were split on a patient-wise basis into training, validation, and test sets using a ratio of 7:1:2, respectively. To ensure a thorough evaluation of the segmentation performance, we performed five-fold cross-validation experiments, with the test sets in different cross-validation folds covering the entire dataset.

We use the well-trained student network for inference.

4.3.3 Evaluation Metrics

We utilized six evaluation metrics to assess the segmentation performance of our method: Dice similarity coefficient (DSC): Measures the overlap between predicted and ground truth segmentations. Sensitivity: Calculates the proportion of correctly identified positive instances. Precision: Quantifies the accuracy of positive predictions. Specificity: Measures the ability to correctly identify negative instances. Jaccard index (Jac): Evaluates the overall agreement between predicted and ground truth segmentations. Volume similarity (VS): Measures the similarity of segmented volume with the ground truth.

However, due to the ambiguous annotation of the dataset, most of the fine vessels are not labelled. This can lead to situations where segmentation results with higher accuracy actually have lower DSC. To provide a more comprehensive evaluation of segmentation performance, we employed surface-to-surface distance error (Surface Error) metrics to measure segmentation accuracy based on mesh representations. Furthermore, in our qualitative analysis, we employ visualization techniques to further evaluate the segmentation results, including the degree of over-segmentation and the accuracy of fine vessel segmentation.

The surface error metrics estimate the error between the ground-truth surfaces S, and the segmentation prediction surfaces S'. The distance between a point p_i on surface S and the nearest point on surface S' is given by the minimum of the Euclidean norm. And we compare the similarity between the prediction and ground-truth by generating surface mesh-based representations of these structures from their corresponding masks

in Eq.4.12. Doing this for all N points in the ground-truth surface S gives the average surface-to-surface distance error in Eq.4.13. The p-value is calculated based on surface error.

$$d(p_i, S') = \min_{p' \in S'} \|p_i - p'\|_2 \tag{4.12}$$

$$d(S, S') = \frac{1}{N} \sum_{i=1}^{N} d(p_i, S') \tag{4.13}$$

The surface-to-surface error is computed by measuring vertex-to-vertex distances, instead of face-to-face or vertex-to-face distances.

4.3.4 Qualitative Results and Analysis

In Fig. 4.3, we compared the proposed method with state-of-the-art approaches on four different data sources. Our method demonstrated superior performance in segmenting fine vessels without introducing excessive over-segmentation noise, especially in datasets with high levels of noise, such as ANSYS, ASD, and UPF. Notably, the nnUNet [46] was greatly affected by ambiguous labels and could only segment major vessels. The Swin-UNet [104], utilizing the swin-transformer structure for feature extraction, outperformed convolution in nnUNet by extracting a larger number of vessel branches. VASeg [1], employing majority voting and thresholding techniques, achieved a better recovery of fine vessels. CPS [201], due to the utilization of semi-supervised cross pseudo-supervision, exhibited increased segmentation uncertainty and introduced excessive noise when handling datasets with higher noise levels. Because of our semi-supervised model's emphasis on training robustness, the fully supervised loss is assigned a higher weight compared to the semi-supervised loss. As a result, the network becomes more conservative in its predictions. When dealing with datasets containing lesser noise, the model does not fully unleash its predictive capabilities. Instead, it tends to be more cautious and restrained in making predictions to ensure reliability. In contrast, our method showcased the ability to segment a significant number of fine vessels while maintaining robustness and avoiding the introduction of excessive noise.

4.3.5 Quantitative Results and Analysis

Due to the uncertain and ambiguous nature of our dataset annotations, where only the main vessels near the aneurysm are labelled, we utilized mesh-based evaluation metrics

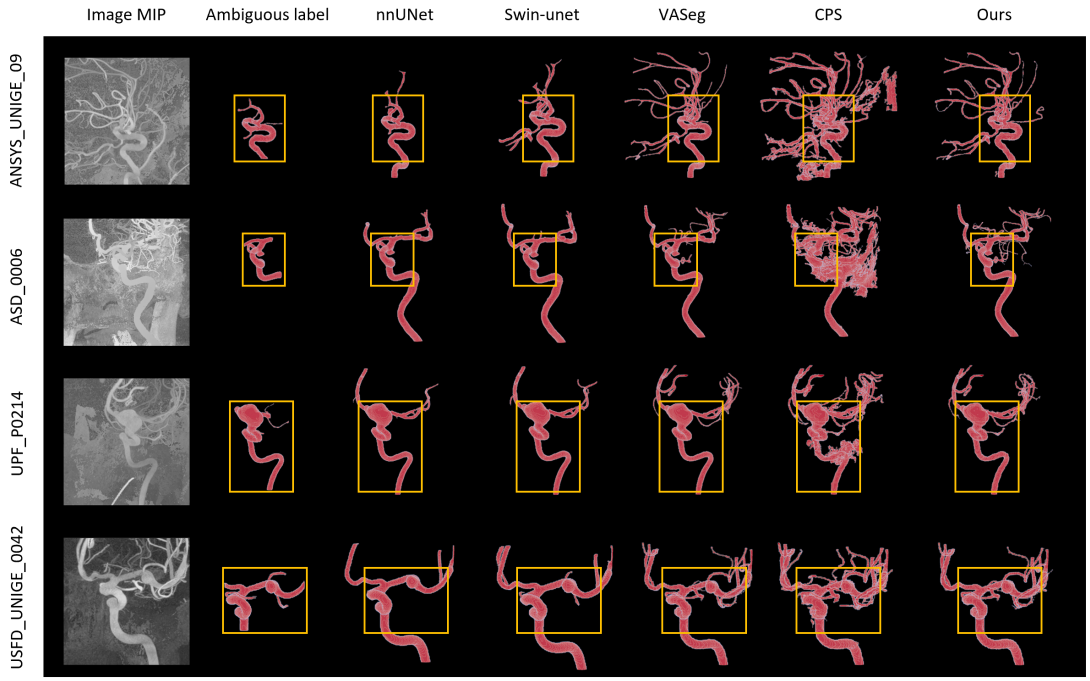

Figure 4.3: Comparison with State-of-the-Art Methods on four different data sources. The yellow box is the golden standard area where all quantitative evaluations are carried out.

as our primary performance measure, while pixel-based evaluation metrics such as the DSC were used as supplementary reference metrics.

In Table. 4.1, we compared our method with several approaches, including the convolution-based fully supervised method nnUNet, the transformer-based fully supervised method Swin-UNet, the VASet method that addresses ambiguous label issues through preprocessing and postprocessing, and the traditional semi-supervised method CPS using cross pseudo-supervision. Due to the incomplete annotation, pixel-based metrics such as DSC and VS cannot accurately measure the segmentation accuracy. Segmenting more unannotated vessels may lead to a decrease in DSC and similar metrics. Therefore, we employ mesh-based surface error as a more reliable metric for evaluation. By comparing the surface error, we found that our method achieved the highest accuracy in vessel surface segmentation, with an average mesh error of 0.20mm (0.35mm/pixel). Additionally, it is worth noting that our method achieves high sensitivity, second only to CPS. This indicates that our method successfully identifies a larger

proportion of positive instances, meaning that it effectively captures the majority of the annotated vessels.

Table 4.1: Compare with state-of-the-art on whole Aneurist dataset. When annotations are incomplete, a higher Dice score may indicate worse performance. Thus, the mesh-based surface error serves as the primary evaluation metric, while pixel-based metrics such as the DSC are used as supplementary evaluation criteria due to incomplete annotation. Quantitative analysis was performed within the annotated regions instead of the full image.

Methods	nnUNet	Swin-unet	VASeg	CPS	Ours
Sensitivity	0.9196 ± 0.0637	0.8967 ± 0.1219	0.9572 ± 0.0510	0.9872 ± 0.0189	0.9793 ± 0.0183
Precision	0.9300 ± 0.0345	0.8689 ± 0.0487	0.8792 ± 0.0761	0.6636 ± 0.1238	0.8018 ± 0.0877
Specificity	0.9956 ± 0.0028	0.9932 ± 0.0032	0.9934 ± 0.0044	0.9740 ± 0.0129	0.9881 ± 0.0056
Jac	0.8605 ± 0.0680	0.7889 ± 0.1091	0.8440 ± 0.0763	0.6572 ± 0.1209	0.7873 ± 0.0817
VS	0.9728 ± 0.0263	0.9424 ± 0.0649	0.9450 ± 0.0509	0.7968 ± 0.0996	0.8977 ± 0.0602
DSC	0.9236 ± 0.0408	0.8772 ± 0.0787	0.9134 ± 0.0479	0.7862 ± 0.0960	0.8786 ± 0.0536
Surface Error	0.8903 ± 1.2450	0.6801 ± 1.6093	0.2586 ± 0.3066	0.3068 ± 0.1210	$\mathbf{0.2075 \pm 0.0640}$
p-value	<0.05	<0.05	<0.05	<0.05	/

Table.4.2 presents the results of the ablation study conducted to analyze the impact of different components in our method. We used the fully supervised Swin-UNet as the baseline, trained solely using the teacher network with Dice Cross Entropy loss. In the second set of experiments, we introduced the Fourier boundary loss to the fully supervised loss. The inclusion of this loss led to a noticeable improvement in surface error, indicating enhanced boundary delineation. In the third set of experiments, we employed vessel adaptation by feeding the data into the student network. We also incorporated the semi-supervised loss to train the student network. The results showed a significant increase in the number of predicted vessels, as evidenced by the improved sensitivity. Additionally, the surface error achieved a level of 0.20mm, indicating precise vessel segmentation at the boundary. After transitioning from fully supervised to semi-supervised learning, the sensitivity increased from 0.92 to 0.97, demonstrating the validity of our hypothesis to utilize the teacher network's predictions to complement ambiguous labels and jointly supervise the student network's output. As a result, the network predicted more segmentation regions that are likely to be vessels.

Table 4.2: Ablation study.

Modules	Supervised Loss (Dice + CE)	Supervised Loss Boundary Loss	Supervised Loss Boundary Loss Semi-supervised Loss
Sensitivity	0.8967 ± 0.1219	0.9236 ± 0.0600	0.9793 ± 0.0183
Precision	0.8689 ± 0.0487	0.9116 ± 0.0577	0.8018 ± 0.0877
Specificity	0.9932 ± 0.0032	0.9954 ± 0.0035	0.9881 ± 0.0056
Jac	0.7889 ± 0.1091	0.8456 ± 0.0652	0.7873 ± 0.0817
VS	0.9424 ± 0.0649	0.9636 ± 0.0440	0.8977 ± 0.0602
DSC	0.8772 ± 0.0787	0.9149 ± 0.0417	0.8786 ± 0.0536
Surface Error	0.6801 ± 1.6093	0.3483 ± 0.3111	$\mathbf{0.2075 \pm 0.0640}$
p-value	<0.05	<0.05	/

4.4 Conclusion

In summary, our semi-supervised model brings forward innovative techniques for cerebral vessel segmentation. With semi-supervised learning and domain adaptation-like strategies, Fourier high-frequency boundary loss, and adaptive histogram attention, we achieve better segmentation accuracy and robustness on whole vessels, paving the way for clinical uses such as treatment planning. However, our model might underperform on lower noise datasets due to our focus on robustness. Future research will explore contrastive learning to improve performance on low-noise datasets. We also plan to expand the state-of-the-art (SOTA) comparison to include not only architecture-based comparisons but also models performing similar tasks in the literature [211, 212].

Chapter 5

Unsupervised Domain Adaptation

Unsupervised domain adaptation (UDA) aims to align the labelled source distribution with the unlabelled target distribution to obtain domain-invariant predictive models. Since cross-modality medical data exhibit significant intra and inter-domain shifts and most are unlabelled, UDA is more important while challenging in medical image analysis. This chapter proposes a simple yet potent contrastive learning framework for UDA to narrow the inter-domain gap between labelled source and unlabelled target distribution. Our method is validated in cerebral vessel data sets. Experimental results show that our approach can learn latent features from labelled 3DRA modality data and improve vessel segmentation performance in unlabelled MRA modality data.

5.1 Introduction

Cerebral vessel segmentation provides invaluable assistance in precise diagnosis, surgical planning, monitoring disease progression, and evaluating treatment results. In silico trials [34] using segmentation-derived vascular models enable the simulation of biological behaviours and support drug development efforts, ultimately improving the efficacy of personalised treatment strategies. Traditional manual segmentation methods, while highly accurate, are labour intensive and time consuming. This has motivated the exploration of automated segmentation techniques [1] as a more efficient alternative. However, the transition to computerised methods poses a new challenge: the reliance on fully labelled data for training.

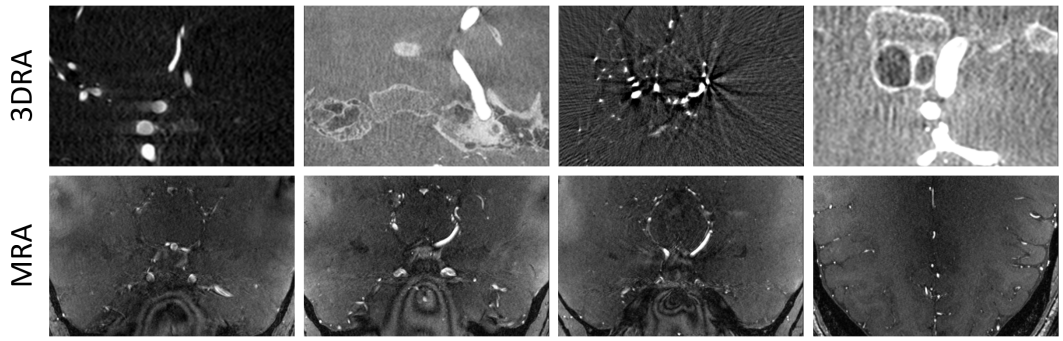

Figure 5.1: Visualization of 3DRA and MRA data reveals significant intra- and inter-domain shifts.

In scenarios where clinical vessel annotations in 3D Rotational Angiography (3DRA) [213] modalities are ambiguous, semi-supervised learning methods [2] with teacher-

student structure leveraging both labelled and unlabeled patches have been proposed to address the issue. However, when the challenge escalates further, with the network fed unlabelled magnetic resonance angiography (MRA) [214, 215] data as input, semi-supervised methods become inadequate for supervision. To address the challenge of data of different modes without labelled data, unsupervised domain adaptation (UDA) techniques, such as FDA [216], DAFormer [217], HRDA [218], MIC [219], and MSCDA [220], have emerged. These methods transfer knowledge from well-annotated source domains to unlabelled target domains.

These methods address the domain shift between the source and target data. However, the domain shift is considerable, even from inner source data in the context of cerebral vascular images. Therefore, the network needs stronger content feature extraction capabilities while reducing its reliance on style features. Fig.5.1 shows that 3DRA and MRA exhibit a substantial domain shift, with a significant domain shift even within the 3DRA modality when clinical data from different data centres are scanned. Mitigating domain shifts for segmentation across other modalities is more challenging due to the intricate nature of blood vessels, individual variabilities, and inherent noise and artefacts in imaging techniques.

Our work introduces a unique method integrating insights from unsupervised domain adaptation, semi-supervised learning, and contrastive learning to address significant domain shifts in cerebral vascular images. The key innovations are as follows:

- This work presents an innovative symmetric adaptation network tailored for cross-modality segmentation of brain vasculature. This represents the first study to apply UDA techniques to the segmentation of the cerebral vessels from 3DRA to MRA.
- We introduce transwarp contrastive learning, a method that investigates features in the time and frequency domains within the latent space to achieve the alignment of content and style.
- Furthermore, we design a new homocentric squares Fourier domain adaptation to handle cross-domain nuisance variability without explicit feature alignment.

5.2 Methodology

This section presents our novel unsupervised domain adaptation approach (see Fig.5.2) to learning instance-specific and domain-invariant representations.

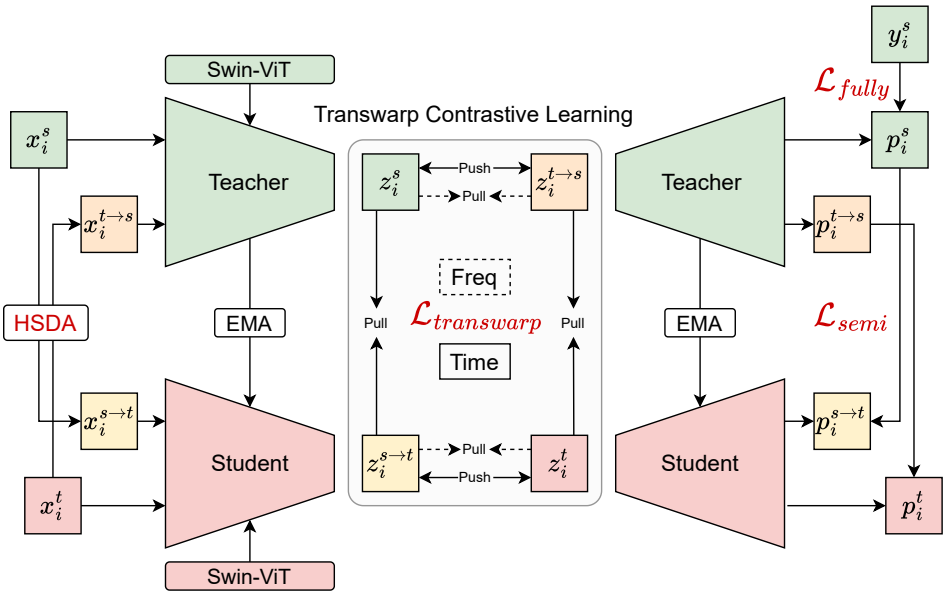

Figure 5.2: Schematic of the proposed method. The method utilizes a composite loss function incorporating fully supervised, semi-supervised, and transwarp Contrastive Learning.

5.2.1 Method Overview and Problem Formulation

For image style transfer, FDA has demonstrated that low-frequency components of images represent style features [216]. Therefore, we utilise the low-frequency components of latent features to extract style features from different inputs. This novel paradigm in transwarp contrastive learning utilises a student-teacher network architecture, incorporating both content and style features using Fourier transform with a low-frequency mask, with the ultimate aim of narrowing the gap caused by different data modalities and magnifying the invariant feature extraction capabilities of the model.

As shown in Fig. 5.2, the student receives two inputs: labelled source domain data (x_i^s) and unlabelled target domain data undergone style transfer ($x_i^{t\to s}$). On the

contrary, the teacher processes unlabelled target domain data (x_i^t) and style-transferred labelled source domain data ($x_i^{s\to t}$). At the same time, we undertake the extraction of both content and style features. From the teacher and student and teacher networks, the characteristics of the content in the time domain are defined as z_i^s, $z_i^{t\to s}$, z_i^t and $z_i^{s\to t}$. In parallel, we extract style features in the frequency domain, capturing the Fourier low-frequency attributes. These style features are expressed as s_i^s, $s_i^{t\to s}$, s_i^t, and $s_i^{s\to t}$.

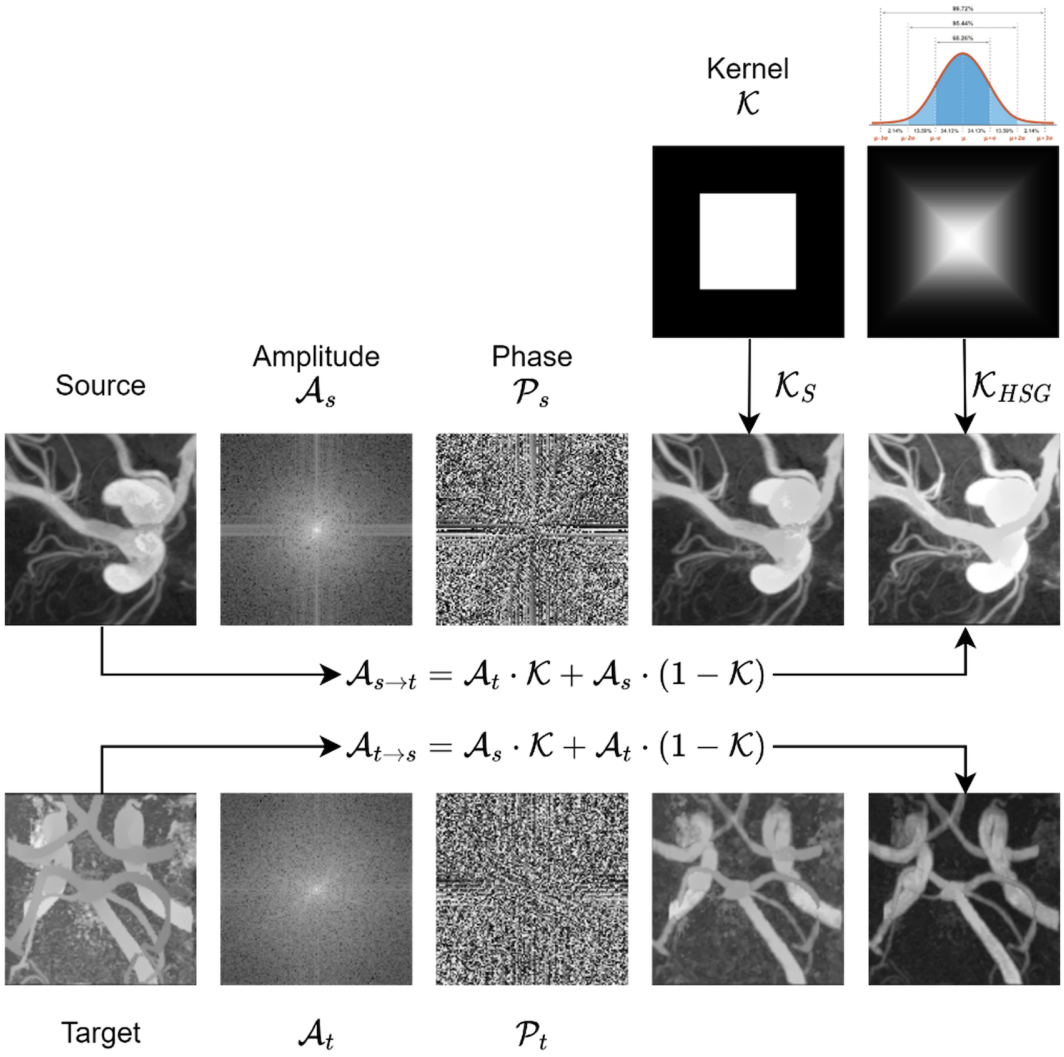

Figure 5.3: Homocentric squares Gaussian kernel $\mathcal{K}_{HSG}$ for image adaptation on 3DRA (source) and MRA (target) vessel patch.

The prediction of the teacher network on the source domain and the target on the source domain is denoted as p_i^s and $p_i^{t\to s}$. We also denote the prediction of the student network in the target domain and the source in the target domain as p_i^t and $p_i^{s\to t}$. Our goal is to learn a task-specific network using a labelled source data set $\{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ and an unlabelled target data set $\{(x_i^t)\}_{i=1}^{N_t}$ to predict labels on test data from the target domain accurately.

5.2.2 Homocentric Squares Domain Adaptation

The purpose of pre-processing is to remove noise and facilitate the model extracting features specific to the vessels. FDA proposes that the style of an image can be migrated by removing the low-frequency amplitude between the target image and the source image [216]. However, as observed in [216] [221], the generated images consist of incoherent dark patches caused by abrupt changes in amplitude around the rectangular mask. Also, considering that the distribution of the spectrum is square instead of circles, we propose a 3D homocentric square Gaussian mask in Fig. 5.3 to exchange the target and source amplitude to make a smooth transition in their style.

Let $\mathcal{F}_A(\cdot)$ and $\mathcal{F}_P(\cdot)$ be the amplitude and phase spectrum in frequency space of a 3D image, and $\mathcal{F}^{-1}$ indicate the inverse Fourier transform. We define a 3D homocentric square Gaussian mask (in Fig. 5.3) $\mathcal{K}_{HSG}$ of the same size as $\mathcal{F}_A$, with σ being the standard deviation. Given two patches x_i^s and x_i^t, our proposed homocentric squares domain adaptation (HSDA) can be formulated as:

$$x_i^{s\to t} = \mathcal{F}^{-1}[\mathcal{F}_P(x_i^s), \mathcal{F}_A(x_i^t) \cdot \mathcal{K}_{HSG} + \mathcal{F}_A(x_i^s) \cdot (1 - \mathcal{K}_{HSG})] \tag{5.1}$$

where the target patch is randomly selected from the target dataset.

5.2.3 Fully-supervised Learning

In the student network, labelled source samples $\{(x_i^s, y_i^s)\}_{i=1}^{N_s}$ are passed through the task-specific segmentation network to minimise the supervised loss $\mathcal{L}_{fully}$, which includes Dice similarity coefficient loss and cross-entropy:

$$\mathcal{L}_{fully} = \frac{1}{N_s} \sum_{i=1}^{N_s} \left(1 - \frac{2\,|p_i^s \cap y_i^s|}{|p_i^s| + |y_i^s|} - y_i^s \log\left(p_i^s\right) \right) \tag{5.2}$$

5.2.4 Semi-supervised Learning

We employ semi-supervised learning, as the source data is labelled while the target data lacks labels. This involves using the predictions of the teacher network as pseudo-labels to supervise the output of the student network.

To perform alignment at the instance level, for the source patches, there are two predictions p_i^s and $p_i^{s\to t}$. For the target patches, there are two predictions p_i^t and $p_i^{t\to s}$. First, for the teacher network, the model's ability to segment the source style is improved in fully supervised learning, where p_i^s is supervised from y_i^s. After style migration, the teacher network also improves the segmentation performance of $p_i^{t\to s}$. We then use the output of the teacher network as a pseudo-label to supervise the output of the student network. Specifically, we employ the mean square error defined in Eq. 5.3 to reduce the discrepancy between the two predictions of the same patch.

$$\mathcal{L}_{semi} = \frac{1}{N_s}\sum_{i=1}^{N_s}\left(p_i^s - p_i^{s\to t}\right)^2 + \frac{1}{N_t}\sum_{i=1}^{N_t}\left(p_i^{t\to s} - p_i^t\right)^2 \tag{5.3}$$

5.2.5 Transwarp Contrastive Learning

Our objective in the time domain (content) is to enable the network to learn instance-specific features.

In our framework, we define the latent vector z in Fig.5.2 as a combination of two components:

- c: The content latent features capture instance-specific features from input data. It represents the shared structural information across different modalities.
- s: The style latent features encode modality-specific features that distinguish different domains.

First, we aim to achieve positive content pairs (see Eq. 5.5) by reducing the distance between content features derived from similar patches. The cosine similarity is computed to pull s_i^s closer to $s_i^{s\to t}$ and $s_i^{t\to s}$ closer to s_i^t. As these pairs originate from identical patches and have only undergone processing through different networks, their latent features should closely align. On the contrary, we set negative content pairs (see Eq. 5.6), derived from different modalities and patches, to maximise their inherent distance. This involves using the cosine similarity to push apart c_i^s from $c_i^{t\to s}$ and similarly, distance $c_i^{s\to t}$ from c_i^t.

Venturing into the frequency domain (style), we aim for the network to become predominantly agnostic to modality regarding feature style. The fundamental objective is for the network to encapsulate the inherent vasculature data style, surpassing the intricacies of specific patches or modalities. The post-encoding outcome should display a more consistent style representation irrespective of the data's source, whether from various patches or modalities.

To achieve this goal, we take advantage of the low-frequency component of the latent feature as the style feature, aiming to ensure stylistic consistency across all segments. To accomplish this, we calculate the upbeat style pairs as outlined in Eq. 5.7, aligning s_i^s, $s_i^{s\to t}$, $s_i^{t\to s}$, and s_i^t to be near each other. Furthermore, to preserve the integrity of the latent feature and mitigate excessive distortion, we restrict the number of encoder blocks to three, allowing for just two layers of max-pooling.

$$h(u,v) = \frac{u^T v}{\|u\|_2 \|v\|_2} \tag{5.4}$$

$$pos_i^c = h\left(z_i^s, z_i^{s\to t}\right) + h\left(z_i^{t\to s}, z_i^t\right) \tag{5.5}$$

$$neg_i^c = h\left(z_i^s, z_i^{t\to s}\right) + h\left(z_i^{s\to t}, z_i^t\right) \tag{5.6}$$

$$pos_i^s = h\left(s_i^s, s_i^{t\to s}\right) + h\left(s_i^s, s_i^t\right) + h\left(s_i^{s\to t}, s_i^t\right) + h\left(s_i^{s\to t}, s_i^{t\to s}\right) \tag{5.7}$$

$$\mathcal{L}_{transwarp} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{(e^{pos_i^c} + e^{pos_i^s})/\tau}{e^{pos_i^c} + e^{pos_i^s} + e^{neg_i^c}} \tag{5.8}$$

In summary, transwarp contrastive learning leverages Fourier transformation to extract temporal components as content features while utilising low-frequency components as style features. This approach enhances contrast in content aspects while reducing contrast in style aspects and enhancing network adaptation capabilities across domains.

5.2.6 Overall Framework and Training Objective

The weight ratio of fully supervised loss is higher than that of semi-supervised loss and contrastive loss of the transwarp (8:1:1). We prioritise the fully supervised loss

to ensure training robustness and prevent the network from becoming overly confident and introducing noise during the initialisation stage.

$$\mathcal{L} = \lambda_1 \cdot \mathcal{L}_{fully} + \lambda_2 \cdot \mathcal{L}_{semi} + \lambda_3 \cdot \mathcal{L}_{transwarp} \tag{5.9}$$

5.3 Experiments and Results

5.3.1 Datasets

Our study trained the network using labelled 3DRA and unlabelled MRA from the Aneurist data set and tested it with the SMILE data set.

Aneurist-3DRA [186]: This dataset includes 223 partially annotated 3D brain vessel images from four different centres, each using unique scanners and imaging protocols. This dataset is the same as the one used in Chapter 3; however, while Chapter 3 utilizes the full dataset for training, here we train with only 1/4 of the data. This subset is sourced from a single data center among the four data centers in Aneurist-3DRA.

Aneurist-MRA [186]: Consisting of 207 high-quality clinical MRA data, this dataset lacks vascular segmentation labels and is not paired with Aneurist-3DRA.

SMILE-UHURA [222]: This dataset offers 14 fully labelled cerebral vessel cases in MRA. ToF angiography images were initially pre-segmented automatically and underwent extensive manual refinement to ensure accuracy.

5.3.2 Experimental Setup

Our experiments used an NVIDIA GeForce RTX 3090 GPU (24GB VRAM) with 128GB RAM. We use Swin-UNet [104] for both teacher and student networks, using patch-based learning with patch size [128, 128, 128]. We aim to have 100 training epochs with optimisation using the Adam optimiser. Data augmentation techniques, including random rotational adjustments and flipping, were applied. The initial learning rate is 0.001 and decreases by a factor of 0.1 every ten epochs. Teacher network parameters θ_{tea} received standard updates, while student network parameters θ_{stu} were updated using the exponential moving average approach in Eq. 5.10.

$$\theta_{stu} = \alpha \cdot \theta_{stu} + (1 - \alpha) \cdot \theta_{tea} \tag{5.10}$$

$$\alpha = \min\left(1 - \frac{1}{iter + 1}, decay\right) \tag{5.11}$$

5.3.3 Quantitative Results

In Table 5.1, we present a summary of segmentation performance compared to state-of-the-art (SOTA) techniques, which encompasses various scenarios: (1) Source-only learning on the target image, (2) Different UDA methods, and (3) Fully supervised learning with labelled target data. MSCDA is designed for medical images in UDA methods, and DAFormer, MIC, and HRDA are natural image UDA methods. Our approach excels in critical metrics, including the Dice similarity coefficient (DSC), Sensitivity (Sen), Jaccard index (Jac) and Volume similarity (VS), highlighting its effectiveness in detecting positive instances, achieving overall agreement, and ensuring volume matching accuracy. While we outperform other domain adaptation methods and the source-only learning baseline, there is room for further enhancement in our unsupervised approach compared to fully supervised methods.

Table 5.1: Comparison of Segmentation Performance with UDA SOTAs and different training strategy.* indicates $p < 0.05$ in t-test.

Methods	DSC (%) ↑	Sen (%) ↑	Jac (%) ↑	VS (%) ↑
$\mathcal{S} \rightarrow \mathcal{T}$[1]	31.48 ± 6.76	18.89 ± 5.00	18.88 ± 5.00	31.52 ± 6.75
FDA [216]	61.84 ± 7.08	46.29 ± 8.48	45.16 ± 7.77	64.88 ± 8.47
MSCDA[220]	41.18 ± 4.70	27.57 ± 4.96	26.04 ± 3.84	49.12 ± 8.69
DAFormer[217]	57.75 ± 6.35	42.84 ± 8.07	40.89 ± 6.52	63.37 ± 9.70
MIC[219]	67.16 ± 2.02	59.07 ± 7.16	50.59 ± 2.27	84.18 ± 9.49
HRDA[218]	68.35 ± 2.74	60.03 ± 8.57	51.98 ± 3.14	83.31 ± 9.68
Ours	72.65 ± 6.65 *	64.75 ± 8.06 *	57.46 ± 7.80 *	85.47 ± 9.65 *
$\mathcal{T} \rightarrow \mathcal{T}$[104]	79.76 ± 1.92	74.61 ± 7.77	66.37 ± 2.69	90.06 ± 5.74

Table 5.2 presents an ablation study that builds upon the foundation of the FDA and progressively integrates various components into our approach. As we observe improvements in the experimental results, it becomes evident that each component in our method contributes positively to the outcome.

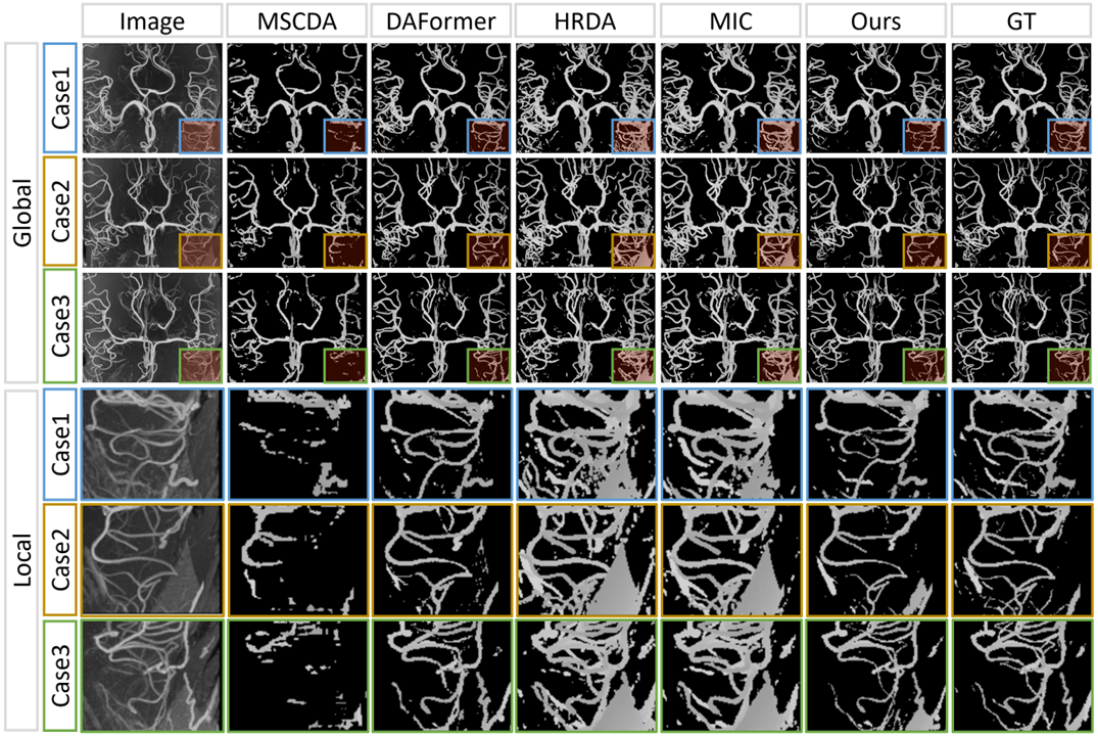

Figure 5.4: Visualisation comparison on MIP maps. Ours shows less over-segmentation on local area.

Table 5.2: Ablation Study: Gradual Addition of Components from Top to Bottom.

Components	DSC (%) ↑	Sen (%) ↑	Jac (%) ↑	VS (%) ↑
$\mathcal{L}_{fully}$	61.84 ± 7.08	46.29 ± 8.48	45.16 ± 7.77	64.88 ± 8.47
$\mathcal{L}_{fully} + \mathcal{L}_{semi}$	64.60 ± 7.36	49.08 ± 8.75	48.00 ± 8.17	67.48 ± 8.42
$\mathcal{L}_{fully} + \mathcal{L}_{semi} + \mathcal{L}_{transwarp}$	67.55 ± 6.81	52.75 ± 8.65	50.95 ± 7.67	72.16 ± 8.47
$\mathcal{L}_{fully} + \mathcal{L}_{semi} + \mathcal{L}_{transwarp}$ + HSDA (Ours)	72.65± 6.65	64.75± 8.06	57.46± 7.80	85.47± 9.65

5.3.4 Visual Inspection

In addition to conducting quantitative comparisons, we also performed visual comparisons of results with similar numerical values in Table 5.1. In Fig. 5.4, we visually compared maximum intensity projection (MIP) maps between our method and the top-performing algorithm MIC and HRDA from UDA-SOTAs. It is evident from the image that within the red bounding box, MIC and HRDA tends to exhibit over-segmentation. Its capacity to distinguish between vessels and brain tissue noise is significantly inferior to our approach.

5.4 Conclusion

In conclusion, our novel transwarp contrastive learning framework represents a significant advancement in the field of UDA in the segmentation of brain vessels. We have effectively managed the challenges posed by diverse data modalities by utilising content pairs from the time domain and style pairs from the frequency domain. This innovative approach not only facilitates knowledge transfer from source domains with domain shift but also ensures exceptional precision in feature extraction for application in target domains. Future work will focus on enhancing domain generalization capabilities.

Chapter 6

Domain Generalization

The automated segmentation of cerebral aneurysms is pivotal for accurate diagnosis and treatment planning. Confronted with significant domain shifts and class imbalance in 3D Rotational Angiography (3DRA) data from various medical institutions, the task becomes challenging. These shifts include differences in image appearance, intensity distribution, resolution, and aneurysm size, all of which complicate the segmentation process. To tackle these issues, we propose a novel domain generalization strategy that employs gradient surgery exponential moving average (GS-EMA) optimization technique coupled with boundary-aware contrastive learning (BACL). Our approach is distinct in its ability to adapt to new, unseen domains by learning domain-invariant features, thereby improving the robustness and accuracy of aneurysm segmentation across diverse clinical datasets. The results demonstrate that our proposed approach can extract more domain-invariant features, minimizing over-segmentation and capturing more complete aneurysm structures.

6.1 Introduction

The accurate segmentation of cerebral aneurysms is vital for diagnosing and treating patients effectively. This process is not just about detecting aneurysms early; it involves precise measurements of their size and shape, which are critical for formulating treatment plans [34, 31]. However, the variability in imaging data quality (see Fig.6.1) from different medical centers presents a significant challenge, complicating the segmentation process.

This variability necessitates a domain generalization (DG) approach, where a model trained on data from multiple sources can adapt to new, unseen domains. The diversity of multi-source data makes DG a daunting challenge in medical imaging, pushing the need for models that generalize well across different medical centers and data types.

Unlike traditional DG approaches such as domain alignment [223], data augmentation [224], ensemble learning [225], self-supervised learning [226], disentangled representation learning [227], and others, our method takes a different approach. We enhance domain generalization by leveraging gradient surgery exponential moving average (GS-EMA), offering an innovative solution to address DG challenges.

In deep learning, EMA is a frequently used technique for parameter averaging in models, aimed at enhancing the generalization performance and stability of the model. In a teacher-student [2] network setup, the teacher network undergoes a process of

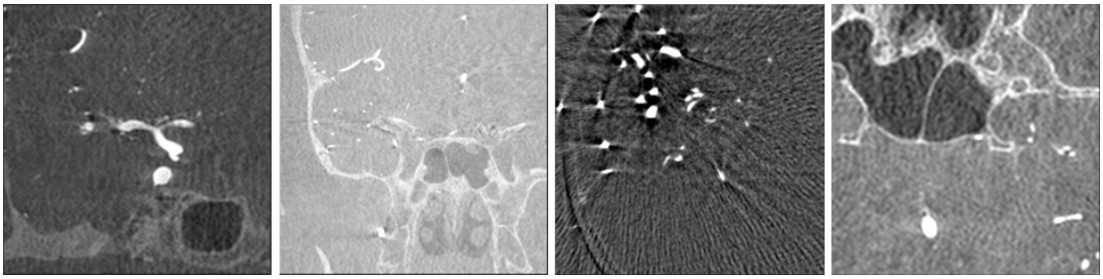

Figure 6.1: Illustration of the variability in imaging data quality from different medical centers.

parameter smoothing, driven by the student network. However, initially, there are no specific conditions set for this transfer. Consequently, all parameters learned by the student network, whether they are domain-invariant or domain-specific, are updated into the teacher network at some rate. This approach poses a challenge as it fails to distinguish between domain-invariant and domain-specific parameters. To address this issue, we introduce the concept of gradient surgery.

Deep neural networks are trained using gradient descent, where gradients guide the optimization process across the landscape defined by the loss function and training data. The gradient surgery framework [228, 229] aims to resolve conflicts arising in multi-task learning. The conflicting gradients are typically averaged to obtain a final gradient for parameter updates. GSMorph [230] propose alternative methods like normal vector projection to derive the ultimate gradient for parameter updates. Instead of devising a new projection method as suggested by others, we approach the problem by analyzing the relationships between gradients to determine whether EMA parameter updates should occur.

Additionally, there is a class imbalance problem in 3D data segmentation due to the small proportion of aneurysms. After multiple downsampling steps, these small features are more likely to be overlooked in the latent space. To tackle this, we introduce the concept of boundary-awareness to traditional contrastive learning [231].

Contributions: Our study introduces innovative techniques that enhance model adaptability. We integrate gradient surgery with EMA updates, strengthening the ability of model to learn domain-invariant features. This novel approach promises to elevate the performance of DG tasks in medical imaging, ensuring that our model can generalize effectively to new datasets and medical centers. Additionally, we pioneer the

use of boundary-aware contrastive learning, enabling our model to discern small target features especially for cerebral aneurysms.

6.2 Methodology

Fig. 6.2 depicts our neural network architecture dedicated for domain generalization tasks. It initiates with 3D source images, which undergoes image transformation to produce target images. Once both the source and target images are obtained, they are separately fed into the encoders of the student and teacher networks.

After acquiring the latent space features, a boundary-aware contrastive learning loss is computed. The central notion here is to amalgamate the same instance subjected to diverse transformations, while distancing different instances, aiming to grasp instance-aware representations. This contrastive learning differs from transformation predictions, as it strives to attain transformation-invariant representations. The latent space features are then decoded to yield predictions, which are supervised using ground truth.

Within the student network, the green arrow signifies fully supervised learning for the source images, and the yellow arrow represents the same for the target images. By analyzing the gradient relationship between these two losses, a novel GS-EMA strategy is devised to update the parameters of teacher network. If the gradient angle between the losses is less than 90 degrees, it indicates that the network has learned domain-invariant features, prompting an EMA update. Conversely, if the gradient angle exceeds 90 degrees, no EMA update is performed, as this suggests the network has grasped domain-specific features, which is not conducive to domain generalization tasks. Ultimately, after several updates, a teacher network enriched with more domain-invariant features is achieved, readying it for domain generalization tasks.

6.2.1 Problem Definition and Data Transformation

Let $\mathcal{X}$ be the input (image) space and $\mathcal{Y}$ be the segmentation (label) space, a domain is defined as a joint distribution P_{XY} on $\mathcal{X} \times \mathcal{Y}$. In the context of DG, we have access to K similar but distinct source domains $\left\{\left(x_s^k, y_s^k\right)\right\}_{k=1}^{K}$, each associated with a joint distribution P_{XY}^k. Note that $P_{XY}^i \neq P_{XY}^j$ with $i \neq j$ and $i, j \in \{1, ..., K\}$. The goal of DG is to learn a predictive model using only source domain data such that the prediction error on an unseen target domain is minimized.

To enhance the model adaptability to previously unseen data domains, we use data

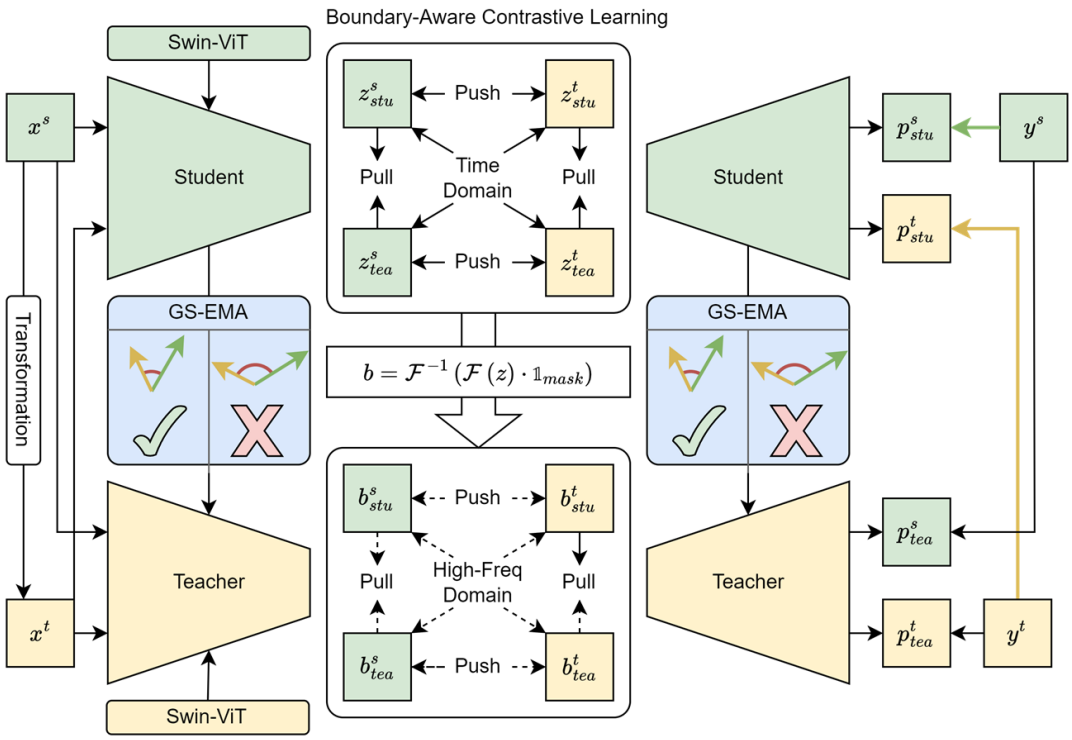

Figure 6.2: Schematic of the proposed model.

transformations to simulate the distribution of the target domain data. The simulated target data is represented as $\{(x_k^t, y_k^t)\}_{k=1}^K$. The process of data transformation encompasses several key steps, including geometric transformations, intensity alterations, noise injection and smoothing, histogram shifting, as well as bias field correction. These operations collectively aim to generate diverse target data, empowering the model with enhanced generalization capabilities to adapt to various data sources and target domains.

6.2.2 Gradient Surgery Exponential Moving Average

In a teacher-student network setup, when the student network is tasked with learning from data originating from different domains, we calculate distinct losses for each domain in Eq. 6.2. This allows us to obtain gradient information specific to each domain. Our fundamental hypothesis is that when the angle between gradients from different domains is less than 90 degrees, it suggests that the student network has effectively

learned how to extract domain-invariant features. In such cases, we employ EMA to update the parameters of student network, subsequently transferring these parameters to the teacher network. This transfer is performed to better capture universal features.

However, when the angle between gradients from different domains exceeds 90 degrees, it indicates that the student network is primarily focused on learning domain-specific features. In such scenarios, we abstain from utilizing EMA for parameter updates and refrain from transmitting these parameters to the teacher network. This strategic approach ensures that the student network can efficiently discriminate between features originating from different domains, enabling it to adapt effectively to the challenges of multi-task learning.

$$\mathcal{L}_{DCE}(p, y) = \frac{1}{N} \sum_{i=1}^{N} \left(1 - \frac{2\,|p \cap y|}{|p| + |y|} - y \log p\right) \tag{6.1}$$

$$\mathcal{L}_{stu}^{src} = \mathcal{L}_{DCE}(p_{stu}^{s}, y^{s}),\ \mathcal{L}_{stu}^{trg} = \mathcal{L}_{DCE}(p_{stu}^{t}, y^{t}) \tag{6.2}$$

Data: Student network parameters θ_{stu}; Teacher network parameters θ_{tea}; Loss on source data in student network $\mathcal{L}_{src}$; Loss on target data in student network $\mathcal{L}_{trg}$; EMA decay coefficient α.

Result: Decide whether updated θ_{tea} with EMA from θ_{stu}.

for each mini-batch **do**

- $\nabla\mathcal{L}_{stu}^{src} \rightarrow g_{src}$;
- $\nabla\mathcal{L}_{stu}^{trg} \rightarrow g_{trg}$;
- **if** $\langle g_{src}, g_{trg} \rangle \leq 0$ **then**
  - $\theta'_{tea} = \theta_{tea} \cdot \alpha + (1 - \alpha) \cdot \theta_{stu}$;
- **else if** $\langle g_{src}, g_{trg} \rangle > 0$ **then**
  - $\theta'_{tea} = \theta_{tea}$;
- **end**
- Update θ_{tea} with θ'_{tea};
- Update θ_{stu} as needed;

end

Algorithm 1: Gradient Surgery Exponential Moving Average

6.2.3 Boundary Aware Contrastive Learning

In our study, we tackle the challenge of uneven distribution of classes in the segmentation of aneurysms by proposing a unique contrastive learning approach that operates within a teacher-student network configuration. This method enhances the distinction between matching (positive) and non-matching (negative) sample pairs by employing a Fourier transformation strategy, which is particularly adept at isolating high-frequency elements that delineate boundaries. Transitioning from volume-based to boundary-based analysis ensures that the presence of small aneurysms is not disproportionately low compared to larger vessels.

Both the student and teacher branches receive two distinct sets of data: the original data from the source domain, represented as x^s, and the corresponding transformed data x^t. Consequently, the latent feature representations from the student network are symbolized as z^s_{stu} and z^t_{stu}, while those from the teacher network are signified as z^s_{tea} and z^t_{tea}.

Advancing further, we harness the power of Fourier transformation paired with a high-frequency filter in Eq. 6.3 to extract features that are cognizant of the boundaries within the data. These extracted features from both student and teacher networks are represented as b^s_{stu}, b^t_{stu}, b^s_{tea}, and b^t_{tea} respectively. Our primary objective within this feature space is to cultivate instance-specific representations that are closely aligned when the same instance is encoded differently, while simultaneously ensuring a clear demarcation between distinct instances, irrespective of the encoder used.

To clarify the relationships within our contrastive learning framework, we delineate the instances processed through different encoders as positive pairs when they originate from the same instance. This includes pairs like z^s_{stu} with z^s_{tea}, and z^t_{stu} with z^t_{tea}. In contrast, negative pairs consist of different instances that have been encoded either by the same or by different encoders, such as z^s_{stu} with z^t_{stu}, and z^s_{tea} with z^t_{tea}, as well as cross-encoder pairs like z^s_{stu} with z^t_{tea}, and z^s_{tea} with z^t_{stu}. These delineations form the basis of our contrastive learning process.

Moving forward, we apply a Fourier transformation to the volume features to construct an amplitude map, which is crucial for identifying the salient high-frequency components that highlight boundaries in Eq.6.3. A specialized square mask is then utilized to isolate this high-frequency information in Eq.6.3. Here, the Fourier transform and its inverse are denoted by $\mathcal{F}$ and $\mathcal{F}^{-1}$ respectively. The mask $\mathbb{1}_{mask}$, with

value zero in its center and one at the periphery, has the same shape as z.

For boundary features, positive pairs are formed by analogous instances across the student and teacher networks, such as b^s_{stu} with b^s_{tea}, and b^t_{stu} with $b^{s\to t}_{tea}$. Conversely, negative pairs are created by combining features from distinct instances, which may be within the same network or across both, exemplified by pairs such as b^s_{stu} with b^t_{stu}, and b^s_{tea} with b^t_{tea}, as well as inter-network pairs like b^s_{stu} with b^t_{tea}, and b^s_{tea} with b^t_{stu}.

$$b = \mathcal{F}^{-1}\left(\mathcal{F}\left(z\right) \cdot \mathbb{1}_{mask}\right) \tag{6.3}$$

$$h(u, v) = \frac{u^T v}{\|u\|_2 \|v\|_2} \tag{6.4}$$

$$\mathcal{L}_c = -\log \frac{\sum_{i=1}^{N_p} e^{\left(h\left(u_i^+, v_i^+\right)\right)}}{\sum_{i=1}^{N_p} e^{\left(h\left(u_i^+, v_i^+\right)\right)} + \sum_{j=1}^{N_n} e^{\left(h\left(u_j^+, v_j^-\right)\right)}} \tag{6.5}$$

To quantify the similarity of these pairs, we compute the cosine similarity for each within both the time and frequency domains in Eq.6.4 and Eq.6.5. The similarity for positive pairs is expressed as $h(u_i^+, v_i^+)$ where i spans all positive pair indices, and the similarity for negative pairs is articulated as $h(u_j^+, v_j^-)$ where j represents the indices of all negative pairs. Here, N_p stands for the count of positive pairings, and N_n corresponds to the count of negative pairings.

The contrastive learning loss for these high-frequency boundary pairs is then calculated using the same equation as for the volumetric pairs. By summing up the volumetric contrastive learning loss with the boundary contrastive learning loss, we derive a comprehensive boundary-aware contrastive learning loss. This loss function is designed to finely tune our model to discriminate between the nuanced features of aneurysms, enhancing its segmentation performance.

6.2.4 Overall Framework and Training Objective

The loss function consists of two parts. $\mathcal{L}_{DCE}$ fully supervises the four outputs of the teacher and student networks. BACL includes volume contrast $\mathcal{L}^z_c$ and boundary contrast $\mathcal{L}^b_c$. The ratio of λ_1 to λ_2 is set at 0.25:0.5.

$$\mathcal{L} = \lambda_1 \cdot (\mathcal{L}^{src}_{stu} + \mathcal{L}^{trg}_{stu} + \mathcal{L}^{src}_{tea} + \mathcal{L}^{trg}_{tea}) + \lambda_2 \cdot (\mathcal{L}^z_c + \mathcal{L}^b_c) \tag{6.6}$$

We chose the ratio of 0.25:0.5 because the two parts of the loss function contain a different number of components. The first term, $\mathcal{L}_{DCE}$, consists of four loss components: $\mathcal{L}^{src}_{stu}$, $\mathcal{L}^{trg}_{stu}$, $\mathcal{L}^{src}_{tea}$, and $\mathcal{L}^{trg}_{tea}$, while the second term, BACL, consists of only two

components: $\mathcal{L}_c^z$ and $\mathcal{L}_c^b$. To balance the contributions of these terms, we normalized their relative weight based on their component ratio of 4:2, setting λ_1 to 0.25 and λ_2 to 0.5. This ensures that each individual loss term has a comparable influence on the optimization process and prevents any part from dominating due to differences in the number of components.

6.3 Experiments and Results

6.3.1 Experimental Setting

Dataset: We tested our method with 3DRA images from 223 patients from the @neurIST dataset [186]. These images were collected from four distinct medical institutions, each employing varied scanning equipment and imaging protocols. Consequently, this dataset exhibits a broad various in both visual characteristics and resolution. The data diversity can evaluate the robustness and adaptability of our proposed GS-EMA method.

Implementation details: Our study was conducted on a NVIDIA RTX 3090 GPU. We utilized the Swin-UNet [104] architecture for both the student and teacher networks in our framework. The training was set to 100 epochs. To determine whether to apply EMA updates, we experimented with setting the EMA coefficient α to either 0.9999 or 0.9. We started with an initial learning rate of 0.001 and adjusted it downwards by multiplying by 0.1 after every ten epochs. The code will be publicly available soon.

6.3.2 Quantitative Results

Table.6.1 includes comparison with state-of-the-art (SOTA) methods and two ablation studies. Our model outperforms traditional segmentation approaches like nnUNet [46] and VASeg focussing on aneurysms [1], as well as domain-generalising methods for medical image segmentation, including CMDG [232] and FedDG [224]. The ablation study highlights that our GS-EMA algorithm, which regulates EMA updates with gradient relation, surpasses regular and non-EMA methods in segmenting aneurysms. It also indicates superior results for BACL when integrating volume (BACL-V) and boundary (BACL-B) learning, compared to using either alone.

	DSC (%) ↑	Sen (%) ↑	Jac (%) ↑	VS (%) ↑
nnUNet [46]	59.61	57.51	47.38	70.91
VASeg [1]	60.28	54.47	49.82	67.91
FedDG [224]	64.50	64.31	54.26	74.73
CMDG [232]	65.01	64.10	54.11	73.38
Ours	**71.89**	**70.88**	**62.36**	**80.00**
no EMA	61.52	55.64	50.91	69.01
EMA	64.71	62.86	54.03	72.64
GS-EMA	**68.49**	**72.79**	**58.40**	**76.63**
BACL-V	68.49	72.79	58.40	76.63
BACL-B	70.62	**75.14**	60.54	78.22
BACL	**71.89**	70.88	**62.36**	**80.00**

Table 6.1: Quantitative results including compare with SOTAs and ablation studies. Critical metrics includes the Dice similarity coefficient (DSC), Sensitivity (Sen), Jaccard index (Jac) and Volume similarity (VS).

6.3.3 Visual Inspection

Fig. 6.3 offers a visual comparison of aneurysm segmentation between our method and SOTAs. It is evident from the comparison that our approach is less prone to over-segmentation while also being able to segment aneurysms more completely.

Fig. 6.4 shows a t-SNE comparison of latent features using EMA and GS-EMA. The larger overlap achieved by GS-EMA indicates a stronger capability of the model to extract domain-invariant features.

6.4 Conclusion

In summary, our study introduces an effective GS-EMA algorithm and a boundary-aware contrastive learning technique for aneurysm segmentation. These methods outperform existing approaches by minimising over-segmentations and capturing more complete aneurysm structures. For future work, we plan to apply our GS-EMA technique to a wider array of medical imaging datasets for further validation and enhancement.

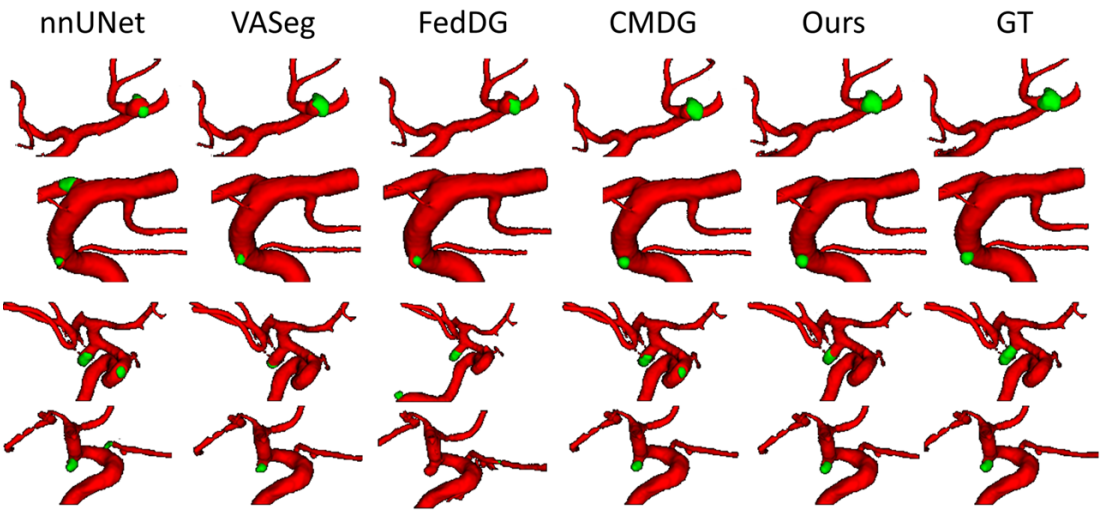

Figure 6.3: Comparative visualization of SOTAs and ours method on aneurysm segmentation.

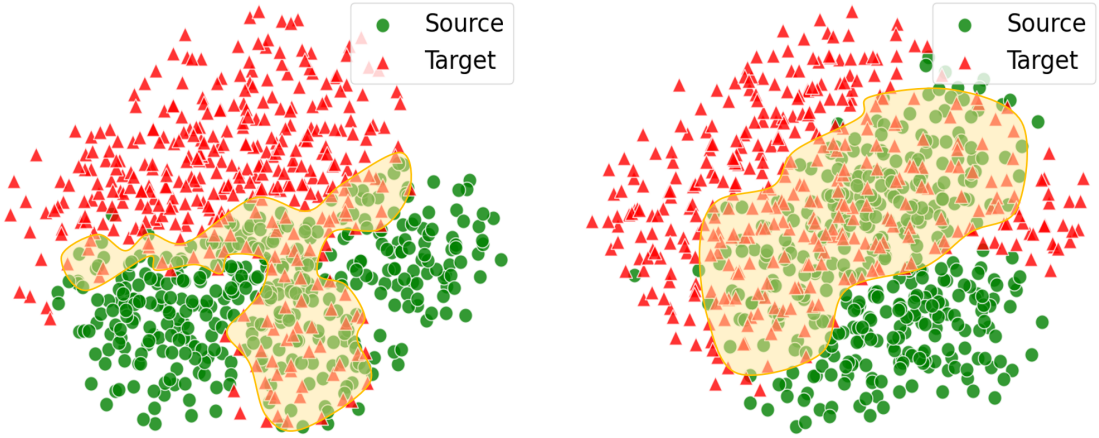

Figure 6.4: The t-SNE visualization of latent features from EMA (left) and GS-EMA (right).

Chapter 7

Conclusions

7.1 Summary and Achievements

The motivation of the thesis stems from the critical need to improve the robustness of brain vessels and aneurysm segmentation. Challenges such as class imbalance, insufficiently labelled data, domain shifts, and data source agnosticism hinder the performance of existing models. This thesis helps to address these challenges by developing and integrating advanced class-imbalance learning, semi-supervised learning, domain adaptation, and domain generalization techniques to enhance segmentation robustness. The primary contributions of this work can be summarised as follows.

Chapter 2 offers an engineering-focused review of medical image segmentation, including data processing techniques, foundational models, learning paradigms, and evaluation methods. This tutorial-style review serves as an invaluable resource, offering direct reference points for practitioners in medical image segmentation. It provides a solid foundation for novice and experienced researchers, aiding in their understanding and implementation of advanced segmentation techniques.

In Chapter 3, we introduced a 3D patch-based multi-class model for segmenting vessels and aneurysms in 3DRA images, addressing challenges such as class imbalance and inter-class interference through innovations in network structure and data processing strategy. To enhance the extraction and representation of small target features, the model incorporates architectural innovations like cascaded transformers, multiview blocks, and wide blocks. For the common aneurysm over-segmentation problem, we proposed combined post-processing strategies called majority voting self-refinement. Our method demonstrates significant advances compared to existing methods, particularly in aneurysm segmentation.

In Chapter 4, we presented a semi-supervised model for cerebral vessel segmentation to tackle the problem of partial and incomplete labelling. The proposed pseudo-label-based semi-supervised network, along with novel Fourier high-frequency boundary loss and adaptive histogram attention, can significantly improve the consistency in whole vessel segmentation, especially can improve fine vessel continuity. The results indicated that the semi-supervised approach outperformed fully supervised methods, especially when labelled data were scarce.

In Chapter 5, we introduced a novel transwarp contrastive learning framework for unsupervised domain adaptation in brain vessel segmentation. This framework effectively addresses the challenges posed by the diverse data modalities using content pairs

from the time domain and style pairs from the frequency domain. Additionally, the proposed Fourier transformed homocentric square domain adaptation filter is effective for vascular image style transfer, enabling the segmentation model to adapt to new domains without requiring labelled data from the target domain. This innovative approach not only facilitates knowledge transfer from source domains with domain shift but also ensures exceptional precision in feature extraction for applications in target domains.

In Chapter 6, we introduced a domain generalization framework for aneurysm segmentation. This framework proposed an effective gradient surgery exponential moving average algorithm to achieve feature disentanglement and employed novel boundary-aware contrastive learning for self-supervised learning. These innovative methods surpass existing approaches by minimising over-segmentations and capturing more complete aneurysm structures. The experimental results demonstrated that the proposed methods could generalise well to unseen domains, maintaining high aneurysm segmentation accuracy.

The contributions of this thesis include the development of innovative segmentation methods, the integration of advanced learning paradigms, and the thorough evaluation of these methods in various challenging scenarios. The integration of class-imbalance learning, semi-supervised learning, domain adaptation, and domain generalization techniques culminated in a highly effective segmentation framework. This framework not only addressed existing challenges in brain vessel and aneurysm segmentation, but also set new benchmarks in terms of precision and robustness. The outcomes of this research have significant implications for improving the accuracy and clinical utility of vessel and aneurysm segmentation models.

Beyond these applications in vessel and aneurysm images, the proposed approaches are well suited for segmenting vessel-like structures in other domains, such as retinal vessel analysis and pulmonary airway segmentation. Additionally, its capability to enhance small target detection makes it applicable to tasks like pulmonary nodule detection.

7.2 Limitations and Future Research Directions

Although this thesis presents significant advances in the segmentation of brain vessels and aneurysms, several limitations remain. Addressing these limitations provides a

foundation for future research directions to further enhance the efficacy and applicability of the proposed methods.

7.2.1 Deep Learning-based Haemodynamic Analysis

One limitation of the current approach is the lack of integration with haemodynamic analysis. Traditional segmentation methods often fail to capture the dynamic and physical properties of blood flow within vessels and aneurysms. The absence of physics-informed constraints in our deep learning models means that the predictions may not fully align with the underlying haemodynamic principles.

To address this, future work could explore the integration of physics-informed neural networks (PINNs) with deep learning models to better capture these physical properties and constraints. In this framework, segmentation and PINNs would be performed synchronously, where the segmentation results serve as direct input to the PINNs for haemodynamic analysis. By incorporating domain-specific knowledge through PINNs, this approach could improve the physical consistency of segmentation results while simultaneously enhancing the accuracy of haemodynamic simulations.

7.2.2 Domain Incremental Learning

Another limitation is the model's ability to adapt to new domains without forgetting previously learnt knowledge. Current domain adaptation and generalization techniques are effective but do not fully address the challenge of continuous learning in diverse data sets and imaging modalities. Developing models that can incrementally learn and adapt to new domains is crucial for their long-term applicability and robustness. Future research in domain incremental learning could enable models to continuously improve and generalise across diverse datasets and imaging modalities, making them more robust and versatile.

7.2.3 Virtual Population and Generative Models

The issue of insufficient medical image data remains a significant challenge. Despite using advanced data augmentation techniques, the diversity and volume of available training data are often limited. This limitation affects the model's performance and generalization capabilities. Addressing this issue through generative models could be a promising direction. Creating a virtual population that simulates diverse anatomical

variations and pathologies can provide abundant and varied training data, significantly enhancing the model's performance and generalization capabilities. Future work could focus on developing sophisticated generative models to simulate realistic and diverse medical images for training purposes.

While the proposed methods demonstrate significant improvements in brain vessel and aneurysm segmentation, addressing these limitations will be critical for further advancements. Integrating deep learning with haemodynamic analysis, developing domain incremental learning techniques, and leveraging generative models to create virtual populations are promising future research directions that can enhance the accuracy, robustness, and applicability of segmentation models in clinical practice.

References

[1] F. Lin, Y. Xia, S. Song, N. Ravikumar, and A. F. Frangi, "High-throughput 3DRA segmentation of brain vasculature and aneurysms using deep learning," Computer Methods and Programs in Biomedicine, p. 107355, 2023.

[2] F. Lin, Y. Xia, N. Ravikumar, Q. Liu, M. MacRaild, and A. F. Frangi, "Adaptive semi-supervised segmentation of brain vessels with ambiguous labels," in International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 106–116, Springer, 2023.

[3] F. Lin, Y. Xia, Y. Deo, M. MacRaild, H. Dou, Q. Liu, K. Wu, N. Ravikumar, and A. F. Frangi, "Unsupervised domain adaptation for brain vessel segmentation through transwarp contrastive learning," in 2024 IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1–5, IEEE, 2024.

[4] F. Lin, Y. Xia, M. MacRaild, Y. Deo, H. Dou, Q. Liu, N. Cheng, N. Ravikumar, and A. F. Frangi, "GS-EMA: Integrating gradient surgery exponential moving average with boundary-aware contrastive learning for enhanced domain generalization in aneurysm segmentation," in 2024 IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1–5, 2024.

[5] R. D. Perrone, A. M. Malek, and T. Watnick, "Vascular complications in autosomal dominant polycystic kidney disease," Nature Reviews Nephrology, vol. 11, no. 10, pp. 589–598, 2015.

[6] A. Sarrami-Foroushani, T. Lassila, M. MacRaild, J. Asquith, K. C. Roes, J. V. Byrne, and A. F. Frangi, "In-silico trial of intracranial flow diverters replicates and expands insights from conventional clinical trials," Nature communications, vol. 12, no. 1, p. 3861, 2021.

[7] S. Naydin, B. Marquez, and K. M. Liebman, Vascular Anatomy of the Brain, pp. 3–29. Cham: Springer International Publishing, 2022.

[8] E. J. Roth, Circle of Willis, pp. 797–798. Cham: Springer International Publishing, 2018.

[9] A. A. Zeal and A. L. Rhoton, "Microsurgical anatomy of the posterior cerebral artery," Journal of neurosurgery, vol. 48, no. 4, pp. 534–559, 1978.

[10] K. Menshawi, J. P. Mohr, and J. Gutierrez, "A functional perspective on the embryology and anatomy of the cerebral blood supply," Journal of stroke, vol. 17, no. 2, p. 144, 2015.

[11] J. L. Brisman, J. K. Song, and D. W. Newell, "Cerebral aneurysms," New England journal of medicine, vol. 355, no. 9, pp. 928–939, 2006.

[12] W. I. Schievink, "Intracranial aneurysms," New England Journal of Medicine, vol. 336, no. 1, pp. 28–40, 1997.

[13] A. Keedy, "An overview of intracranial aneurysms," McGill Journal of Medicine: MJM, vol. 9, no. 2, p. 141, 2006.

[14] R. D. Brown and J. P. Broderick, "Unruptured intracranial aneurysms: epidemiology, natural history, management options, and familial screening," The Lancet Neurology, vol. 13, no. 4, pp. 393–404, 2014.

[15] J. B. Hartman, H. Watase, J. Sun, D. S. Hippe, L. Kim, M. Levitt, L. Sekhar, N. Balu, T. Hatsukami, C. Yuan, et al., "Intracranial aneurysms at higher clinical risk for rupture demonstrate increased wall enhancement and thinning on multicontrast 3d vessel wall mri," The British journal of radiology, vol. 92, no. 1096, p. 20180950, 2019.

[16] S. Tada and J. M. Tarbell, "Interstitial flow through the internal elastic lamina affects shear stress on arterial smooth muscle cells," American Journal of Physiology-Heart and Circulatory Physiology, vol. 278, no. 5, pp. H1589–H1597, 2000.

[17] V. S. Fennell, M. Y. S. Kalani, G. Atwal, N. L. Martirosyan, and R. F. Spetzler, "Biology of saccular cerebral aneurysms: a review of current understanding and future directions," Frontiers in surgery, vol. 3, p. 43, 2016.

[18] J. Xiang, S. K. Natarajan, M. Tremmel, D. Ma, J. Mocco, L. N. Hopkins, A. H. Siddiqui, E. I. Levy, and H. Meng, "Hemodynamic–morphologic discriminants for intracranial aneurysm rupture," Stroke, vol. 42, no. 1, pp. 144–152, 2011.

[19] N. K. Yoon, S. McNally, P. Taussky, and M. S. Park, "Imaging of cerebral aneurysms: a clinical perspective," Neurovascular Imaging, vol. 2, pp. 1–7, 2016.

[20] A. O. Caglayan and M. Dundar, "Inherited diseases and syndromes leading to aortic aneurysms and dissections," European Journal of Cardio-thoracic surgery, vol. 35, no. 6, pp. 931–940, 2009.

[21] B.-s. Kim, "Unruptured intracranial aneurysm: screening, prevalence and risk factors," Neurointervention, vol. 16, no. 3, p. 201, 2021.

[22] J. R. Cebral, M. A. Castro, J. E. Burgess, R. S. Pergolizzi, M. J. Sheridan, and C. M. Putman, "Characterization of cerebral aneurysms for assessing risk of rupture by using patient-specific computational hemodynamics models," American Journal of Neuroradiology, vol. 26, no. 10, pp. 2550–2559, 2005.

[23] F. Ikawa, A. Morita, S. Tominari, T. Nakayama, Y. Shiokawa, I. Date, K. Nozaki, S. Miyamoto, T. Kayama, and H. Arai, "Rupture risk of small unruptured cerebral aneurysms," Journal of neurosurgery, vol. 132, no. 1, pp. 69–78, 2019.

[24] S. Lang, P. Hoelter, A. Birkhold, M. Schmidt, J. Endres, C. Strother, A. Doerfler, and H. Luecking, "Quantitative and qualitative comparison of 4d-dsa with 3d-dsa using computational fluid dynamics simulations in cerebral aneurysms," American Journal of Neuroradiology, vol. 40, no. 9, pp. 1505–1510, 2019.

[25] N. Mu, Z. Lyu, M. Rezaeitaleshmahalleh, J. Tang, and J. Jiang, "An attention residual u-net with differential preprocessing and geometric postprocessing: Learning how to segment vasculature including intracranial aneurysms," Medical image analysis, vol. 84, p. 102697, 2023.

[26] K. Lee, J. Zhang, V. Nguyen, J. Han, J. Johnson, R. Kirollos, and M. Teo, "The evolution of intracranial aneurysm treatment techniques and future directions," Neurosurgical Review, vol. 45, pp. 1–25, Feb 2022.

[27] W. J. van Rooij, M. Sprengers, A. N. de Gast, J. Peluso, and M. Sluzewski, "3d rotational angiography: the new gold standard in the detection of additional intracranial aneurysms," American Journal of Neuroradiology, vol. 29, no. 5, pp. 976–979, 2008.

[28] M. P. Hartung, T. M. Grist, and C. J. François, "Magnetic resonance angiography: current status and future directions," Journal of Cardiovascular Magnetic Resonance, vol. 13, no. 1, p. 19, 2011.

[29] J. K. Min, L. J. Shaw, and D. S. Berman, "The present state of coronary computed tomography angiography: a process in evolution," Journal of the American College of Cardiology, vol. 55, no. 10, pp. 957–965, 2010.

[30] W. R. Brody, "Digital subtraction angiography," IEEE Transactions on Nuclear Science, vol. 29, no. 3, pp. 1176–1180, 1982.

[31] F. Pappalardo, G. Russo, F. M. Tshinanu, and M. Viceconti, "In silico clinical trials: concepts and early adoptions," Briefings in bioinformatics, vol. 20, no. 5, pp. 1699–1708, 2019.

[32] M. MacRaild, A. Sarrami-Foroushani, T. Lassila, and A. F. Frangi, "Accelerated simulation methodologies for computational vascular flow modelling," Journal of the Royal Society Interface, vol. 21, no. 211, p. 20230565, 2024.

[33] M. Macraild, A. Sarrami-Foroushani, T. Lassila, and A. Frangi, "Reduced order modelling of intracranial aneurysm flow using proper orthogonal decomposition and neural networks," International Journal for Numerical Methods in Biomedical Engineering, 2024.

[34] Q. Liu, A. Sarrami-Foroushani, Y. Wang, M. MacRaild, C. Kelly, F. Lin, Y. Xia, S. Song, N. Ravikumar, T. Patankar, Z. Taylor, T. Lassila, and A. F. Frang, "Hemodynamics of thrombus formation in intracranial aneurysms: an in-silico observational study," APL Bioengineering, 2023.

[35] L. Marques, B. Costa, M. Pereira, A. Silva, J. Santos, L. Saldanha, I. Silva, P. Magalhães, S. Schmidt, and N. Vale, "Advancing precision medicine: A review of innovative in silico approaches for drug development, clinical pharmacology and personalized healthcare," Pharmaceutics, vol. 16, no. 3, p. 332, 2024.

[36] K. Ghosh, C. Bellinger, R. Corizzo, P. Branco, B. Krawczyk, and N. Japkowicz, "The class imbalance problem in deep learning," Machine Learning, pp. 1–57, 2022.

[37] P. Billion Polak, J. D. Prusa, and T. M. Khoshgoftaar, "Low-shot learning and class imbalance: a survey," Journal of Big Data, vol. 11, no. 1, p. 1, 2024.

[38] K. Han, V. S. Sheng, Y. Song, Y. Liu, C. Qiu, S. Ma, and Z. Liu, "Deep semi-supervised learning for medical image segmentation: A review," Expert Systems with Applications, p. 123052, 2024.

[39] R. Jiao, Y. Zhang, L. Ding, B. Xue, J. Zhang, R. Cai, and C. Jin, "Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation," Computers in Biology and Medicine, p. 107840, 2023.

[40] S. Tan, X. Peng, and K. Saenko, "Class-imbalanced domain adaptation: An empirical odyssey," in Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16, pp. 585–602, Springer, 2020.

[41] S. Kumari and P. Singh, "Deep learning for unsupervised domain adaptation in medical imaging: Recent advancements and future perspectives," Computers in Biology and Medicine, vol. 170, p. 107912, 2024.

[42] K. Zhou, Z. Liu, Y. Qiao, T. Xiang, and C. C. Loy, "Domain generalization: A survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 4, pp. 4396–4415, 2022.

[43] J. Wang, C. Lan, C. Liu, Y. Ouyang, T. Qin, W. Lu, Y. Chen, W. Zeng, and S. Y. Philip, "Generalizing to unseen domains: A survey on domain generalization," IEEE transactions on knowledge and data engineering, vol. 35, no. 8, pp. 8052–8072, 2022.

[44] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, "Medical image segmentation using deep learning: A survey," IET image processing, vol. 16, no. 5, pp. 1243–1267, 2022.

[45] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," nature, vol. 521, no. 7553, pp. 436–444, 2015.

[46] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," Nature Methods, vol. 18, no. 2, pp. 203–211, 2021.

[47] M. Villa-Uriol, G. Berti, D. Hose, A. Marzo, A. Chiarini, J. Penrose, J. Pozo, J. Schmidt, P. Singh, R. Lycett, et al., "@ neurist complex information processing toolchain for the integrated management of cerebral aneurysms," Interface Focus, vol. 1, no. 3, pp. 308–319, 2011.

[48] C. Garcia et al., "Shiny-icarus challenge." https://www.synapse.org/Synapse:syn45774070/wiki/619953, 2023.

[49] "Aneurisk dataset." http://ecm2.mathcs.emory.edu/aneuriskweb/repository.

[50] "CROWN: Circle of Willis Intracranial Artery Classification and Quantification Challenge." https://crown.isi.uu.nl/, 2023.

[51] S. Chatterjee, H. Mattern, F. Dubost, S. Schreiber, A. Nurnberger, and O. Speck, "Smile-uhura challenge 2023." https://doi.org/10.7303/syn47164761, 2023.

[52] K. M. Timmins, I. C. van der Schaaf, E. Bennink, Y. M. Ruigrok, X. An, M. Baumgartner, P. Bourdon, R. De Feo, T. Di Noto, F. Dubost, et al., "Comparing methods of detecting and segmenting unruptured intracranial aneurysms on tof-mras: the adam challenge," Neuroimage, vol. 238, p. 118216, 2021.

[53] E. Bullitt, D. Zeng, G. Gerig, S. Aylward, S. Joshi, J. K. Smith, W. Lin, and M. G. Ewend, "Vessel tortuosity and brain tumor malignancy: a blinded study1," Academic radiology, vol. 12, no. 10, pp. 1232–1240, 2005.

[54] T. D. Noto, G. Marie, S. Tourbier, Y. AlemÃ¡n-GÃ³mez, O. Esteban, G. Saliou, M. B.Cuadra, P. Hagmann, andJ. Richiardi, ""lausanne$_t$of-mra$_a$neurysm$_c$ohort",″ 2022.

[55] M. Ivantsits, L. Goubergrits, J.-M. Kuhnigk, M. Huellebrand, J. Bruening, T. Kossen, B. Pfahringer, J. Schaller, A. Spuler, T. Kuehne, et al., "Detection and analysis of cerebral aneurysms based on x-ray rotational angiography-the cada 2020 challenge," Medical image analysis, vol. 77, p. 102333, 2022.

[56] X. Yang, D. Xia, T. Kin, and T. Igarashi, "Intra: 3d intracranial aneurysm dataset for deep learning," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2656–2666, 2020.

[57] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," IEEE transactions on neural networks and learning systems, vol. 33, no. 12, pp. 6999–7019, 2021.

[58] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, pp. 234–241, Springer, 2015.

[59] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in 2016 Fourth International Conference on 3D Vision (3DV), pp. 565–571, IEEE, 2016.

[60] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-denseunet: Hybrid densely connected unet for liver and tumor segmentation from ct volumes," IEEE Transactions on Medical Imaging, vol. 37, no. 12, pp. 2663–2674, 2018.

[61] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-directional convlstm u-net with densely connected convolutions," IEEE Access, vol. 7, pp. 115227–115238, 2019.

[62] Q. Yan, B. Wang, D. Wu, et al., "A novel approach to automatic segmentation of liver tumor from ct images using gated-propagation u-net," IEEE Access, vol. 8, pp. 15017–15026, 2020.

[63] H. Zhang, F. Xie, Y. Zhuang, et al., "Sunet: Structured unet for ultrasound image segmentation," in Medical Image Computing and Computer Assisted Intervention–MICCAI 2019, pp. 45–53, Springer, 2019.

[64] Q. Jin, Z. Meng, T. Pham, Q. Chen, L. Wei, and R. Su, "Dunet: A deformable network for retinal vessel segmentation," Knowledge-Based Systems, vol. 178, pp. 149–162, 2019.

[65] B. Hu, Z. Shu, Y. Cui, et al., "Dense multi-path u-net for ischemic stroke lesion segmentation," Frontiers in Neurology, vol. 12, p. 629515, 2021.

[66] H. Zhang, F. Xie, Y. Zhuang, et al., "Stacked dense u-nets with dual transformation for robust mr prostate segmentation," in Medical Image Computing and Computer Assisted Intervention–MICCAI 2019, pp. 203–211, Springer, 2019.

[67] L. Yu, H. Chen, Q. Dou, et al., "Automatic 3d prostate segmentation in mr images using deep feature fusion network," IEEE Access, vol. 7, pp. 45991–46003, 2019.

[68] J. Zhuang, "Laddernet: Multi-path networks based on u-net for medical image segmentation," arXiv preprint arXiv:1810.07810, 2018.

[69] A. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel squeeze excitation in fully convolutional networks," in Medical Image Computing and Computer Assisted Intervention–MICCAI 2018, pp. 421–429, Springer, 2018.

[70] Z. Zhu, Y. Xia, W. Shen, et al., "Anatomynet: Deep learning for fast and fully automated wholeâ€volume segmentation of head and neck anatomy," Medical Physics, vol. 46, no. 2, pp. 576–589, 2019.

[71] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, et al., "Attention u-net: Learning where to look for the pancreas," arXiv preprint arXiv:1804.03999, 2018.

[72] K. Guo, Q. Chen, K. Ma, X. Yang, J. Wu, and L. Yang, "Sa-unet: Spatial attention u-net for retinal vessel segmentation," IEEE Access, vol. 7, pp. 143656–143667, 2019.

[73] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual u-net for medical image segmentation," Journal of medical imaging, vol. 6, no. 1, pp. 014006–014006, 2019.

[74] Q. Jin, Z. Meng, C. Sun, H. Cui, and R. Su, "Ra-unet: A hybrid deep attention-aware network to extract liver and tumor in ct scans," Frontiers in Bioengineering and Biotechnology, vol. 8, p. 605132, 2020.

[75] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "Ma-net: A multi-scale attention network for liver and tumor segmentation," IEEE Transactions on Medical Imaging, vol. 39, no. 11, pp. 3461–3472, 2020.

[76] S. Wang, Y. Hu, and J. Yang, "Frcu-net: Fusion and relation calibration u-net for medical image segmentation," IEEE Access, vol. 8, pp. 120818–120828, 2020.

[77] L. Wang, X. Zhou, Y. Tao, et al., "Mdu-net: Multi-scale dense unet for skin lesion segmentation," IEEE Access, vol. 9, pp. 15822–15832, 2021.

[78] H. Lin, Q. Dou, E. Tsougenis, B. D. De Vos, G. Wang, L. Liu, X. Wang, P.-A. Heng, and B. Glocker, "Joint classification and segmentation of organs at risk in ct images using a hybrid convolutional neural network," IEEE Transactions on Medical Imaging, vol. 40, no. 7, pp. 1835–1845, 2021.

[79] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 3–11, Springer, 2018.

[80] M. Drozdzal, E. Vorontsov, G. Chartrand, S. Kadoury, and Y. Bengio, "The importance of skip connections in biomedical image segmentation," Medical Image Analysis, vol. 44, pp. 69–76, 2018.

[81] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "Attention unet++: A nested attention-aware u-net for liver ct segmentation," in IEEE Access, vol. 8, pp. 53946–53955, IEEE, 2020.

[82] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, S. Wang, and P.-A. Heng, "Unet 3+: A full-scale connected unet for medical image segmentation," arXiv preprint arXiv:2004.08790, 2020.

[83] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in Medical

Image Computing and Computer Assisted Intervention–MICCAI 2018, pp. 178–190, Springer, 2018.

[84] S. A. Kohl, B. Romera-Paredes, C. Meyer, J. De Fauw, J. R. Ledsam, K. H. Maier-Hein, S. Eslami, D. J. Rezende, and O. Ronneberger, "A probabilistic u-net for segmentation of ambiguous images," in Advances in Neural Information Processing Systems, pp. 6965–6975, 2018.

[85] C. F. Baumgartner, K. C. Tezcan, K. Chaitanya, A. M. Hötker, U. J. Muehlematter, S. T. Schindera, and E. Konukoglu, "Phiseg: Capturing uncertainty in medical image segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 119–127, Springer, 2019.

[86] S. Wang, B. Yin, A. Zhang, and L. Lu, "Polar-transformation multi-scale u-net for retinal vessel segmentation," IEEE Access, vol. 7, pp. 164834–164843, 2019.

[87] C. Ye, Y. Guo, and Y. Tang, "Mrf-unet: Incorporating markov random fields in u-net for brain tumor segmentation," IEEE Access, vol. 8, pp. 106823–106832, 2020.

[88] A. Kendall, Y. Gal, and R. Cipolla, "What uncertainties do we need in bayesian deep learning for computer vision?," in Advances in neural information processing systems, pp. 5574–5584, 2017.

[89] A. V. Dalca, G. Balakrishnan, J. Guttag, and M. R. Sabuncu, "Variational auto-encoders for medical image segmentation," IEEE transactions on medical imaging, vol. 38, no. 1, pp. 107–118, 2018.

[90] Z. Zhou, J. H. Shin, L. Zhang, S. R. Gurudu, M. B. Gotway, and J. Liang, "Lung nodule detection and segmentation using 3d neural networks trained on weakly labeled data," in Medical Imaging 2019: Image Processing, vol. 10949, p. 1094923, International Society for Optics and Photonics, 2019.

[91] F. Lin, Q. Wu, J. Liu, D. Wang, and X. Kong, "Path aggregation u-net model for brain tumor segmentation," Multimedia Tools and Applications, vol. 80, pp. 22951–22964, 2021.

[92] D. Qin, J.-J. Bu, Z. Liu, X. Shen, S. Zhou, J.-J. Gu, Z.-H. Wang, L. Wu, and H.-F. Dai, "Efficient medical image segmentation based on knowledge distillation," IEEE Transactions on Medical Imaging, vol. 40, no. 12, pp. 3820–3831, 2021.

[93] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "Transunet: Transformers make strong encoders for medical image segmentation," arXiv preprint arXiv:2102.04306, 2021.

[94] Y. Xie, J. Zhang, C. Shen, and Y. Xia, "Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation," in Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24, pp. 171–180, Springer, 2021.

[95] Y. Gao, M. Zhou, and D. N. Metaxas, "Utnet: a hybrid transformer architecture for medical image segmentation," in Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24, pp. 61–71, Springer, 2021.

[96] Y. Chang, H. Menghan, Z. Guangtao, and Z. Xiao-Ping, "Transclaw u-net: Claw u-net with transformers for medical image segmentation," arXiv preprint arXiv:2107.05188, 2021.

[97] Y. Zhang, H. Liu, and Q. Hu, "Transfuse: Fusing transformers and cnns for medical image segmentation," in Medical image computing and computer assisted intervention–MICCAI 2021: 24th international conference, Strasbourg, France, September 27–October 1, 2021, proceedings, Part I 24, pp. 14–24, Springer, 2021.

[98] Y. Jiang, Y. Zhang, X. Lin, J. Dong, T. Cheng, and J. Liang, "Swinbts: A method for 3d multimodal brain tumor segmentation using swin transformer," Brain sciences, vol. 12, no. 6, p. 797, 2022.

[99] S. Huang, J. Li, Y. Xiao, N. Shen, and T. Xu, "Rtnet: relation transformer network for diabetic retinopathy multi-lesion segmentation," IEEE Transactions on Medical Imaging, vol. 41, no. 6, pp. 1596–1607, 2022.

[100] X. Yan, H. Tang, S. Sun, H. Ma, D. Kong, and X. Xie, "After-unet: Axial fusion transformer unet for medical image segmentation," in Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 3971–3981, 2022.

[101] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, "Unetr: Transformers for 3d medical image segmentation," in Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 574–584, 2022.

[102] H. Wang, S. Xie, L. Lin, Y. Iwamoto, X.-H. Han, Y.-W. Chen, and R. Tong, "Mixed transformer u-net for medical image segmentation," in ICASSP 2022-2022 IEEE international conference on acoustics, speech and signal processing (ICASSP), pp. 2390–2394, IEEE, 2022.

[103] H. Wang, P. Cao, J. Wang, and O. R. Zaiane, "Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer," in Proceedings of the AAAI conference on artificial intelligence, vol. 36, pp. 2441–2449, 2022.

[104] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," in European conference on computer vision, pp. 205–218, Springer, 2022.

[105] A. Lin, B. Chen, J. Xu, Z. Zhang, G. Lu, and D. Zhang, "Ds-transunet: Dual swin transformer u-net for medical image segmentation," IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1–15, 2022.

[106] Y. Tang, D. Yang, W. Li, H. R. Roth, B. Landman, D. Xu, V. Nath, and A. Hatamizadeh, "Self-supervised pre-training of swin transformers for 3d medical image analysis," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 20730–20740, 2022.

[107] G. Xu, X. Zhang, X. He, and X. Wu, "Levit-unet: Make faster encoders with transformer for medical image segmentation," in Chinese Conference on Pattern Recognition and Computer Vision (PRCV), pp. 42–53, Springer, 2023.

[108] B. Chen, Y. Liu, Z. Zhang, G. Lu, and A. W. K. Kong, "Transattunet: Multi-level attention-guided u-net with transformer for medical image segmentation," IEEE Transactions on Emerging Topics in Computational Intelligence, 2023.

[109] F. Bougourzi, F. Dornaika, A. Taleb-Ahmed, and V. T. Hoang, "Rethinking attention gated with hybrid dual pyramid transformer-cnn for generalized segmentation in medical imaging," arXiv preprint arXiv:2404.18199, 2024.

[110] H.-Y. Zhou, J. Guo, Y. Zhang, L. Yu, L. Wang, and Y. Yu, "nnformer: Interleaved transformer for volumetric segmentation," arXiv preprint arXiv:2109.03201, 2021.

[111] J. M. J. Valanarasu, P. Oza, I. Hacihaliloglu, and V. M. Patel, "Medical transformer: Gated axial-attention for medical image segmentation," in Medical image computing and computer assisted intervention–MICCAI 2021: 24th international conference, Strasbourg, France, September 27–October 1, 2021, proceedings, part I 24, pp. 36–46, Springer, 2021.

[112] X. Huang, Z. Deng, D. Li, X. Yuan, and Y. Fu, "Missformer: An effective transformer for 2d medical image segmentation," IEEE Transactions on Medical Imaging, vol. 42, no. 5, pp. 1484–1494, 2022.

[113] M. Ma, H. Xia, Y. Tan, H. Li, and S. Song, "Ht-net: hierarchical context-attention transformer network for medical ct image segmentation," Applied Intelligence, pp. 1–14, 2022.

[114] Y. Wu, K. Liao, J. Chen, J. Wang, D. Z. Chen, H. Gao, and J. Wu, "D-former: A u-shaped dilated transformer for 3d medical image segmentation," Neural Computing and Applications, vol. 35, no. 2, pp. 1931–1944, 2023.

[115] S. Roy, G. Koehler, C. Ulrich, M. Baumgartner, J. Petersen, F. Isensee, P. F. Jaeger, and K. H. Maier-Hein, "Mednext: transformer-driven scaling of convnets for medical image segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 405–415, Springer, 2023.

[116] Y. Zhang, R. Higashita, H. Fu, Y. Xu, Y. Zhang, H. Liu, J. Zhang, and J. Liu, "A multi-branch hybrid transformer network for corneal endothelial cell segmentation," in Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France,

September 27–October 1, 2021, Proceedings, Part I 24, pp. 99–108, Springer, 2021.

[117] Y. Ji, R. Zhang, H. Wang, Z. Li, L. Wu, S. Zhang, and P. Luo, "Multi-compound transformer for accurate biomedical image segmentation," in Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24, pp. 326–336, Springer, 2021.

[118] X. Luo, M. Hu, T. Song, G. Wang, and S. Zhang, "Semi-supervised medical image segmentation via cross teaching between cnn and transformer," in International conference on medical imaging with deep learning, pp. 820–833, PMLR, 2022.

[119] R. Tao, W. Liu, and G. Zheng, "Spine-transformers: Vertebra labeling and segmentation in arbitrary field-of-view spine cts via 3d transformers," Medical Image Analysis, vol. 75, p. 102258, 2022.

[120] J. Wang, L. Wei, L. Wang, Q. Zhou, L. Zhu, and J. Qin, "Boundary-aware transformers for skin lesion segmentation," in Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24, pp. 206–216, Springer, 2021.

[121] B. Dong, W. Wang, D.-P. Fan, J. Li, H. Fu, and L. Shao, "Polyp-pvt: Polyp segmentation with pyramid vision transformers," arXiv preprint arXiv:2108.06932, 2021.

[122] N. T. Duc, N. T. Oanh, N. T. Thuy, T. M. Triet, and V. S. Dinh, "Colonformer: An efficient transformer based method for colon polyp segmentation," IEEE Access, vol. 10, pp. 80575–80586, 2022.

[123] S. Li, X. Sui, X. Luo, X. Xu, Y. Liu, and R. Goh, "Medical image segmentation using squeeze-and-expansion transformers," arXiv preprint arXiv:2105.09511, 2021.

[124] Y. Zhang, Y. Wang, L. Xu, Y. Yao, W. Qian, and L. Qi, "St-gan: A swin transformer-based generative adversarial network for unsupervised domain adaptation of cross-modality cardiac segmentation," IEEE Journal of Biomedical and Health Informatics, 2023.

[125] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," arXiv preprint arXiv:2312.00752, 2023.

[126] T. Dao, D. Fu, S. Ermon, A. Rudra, and C. Ré, "Flashattention: Fast and memory-efficient exact attention with io-awareness," Advances in Neural Information Processing Systems, vol. 35, pp. 16344–16359, 2022.

[127] J. Ma, F. Li, and B. Wang, "U-mamba: Enhancing long-range dependency for biomedical image segmentation," arXiv preprint arXiv:2401.04722, 2024.

[128] Z. Wang, J.-Q. Zheng, Y. Zhang, G. Cui, and L. Li, "Mamba-unet: Unet-like pure visual mamba for medical image segmentation," arXiv preprint arXiv:2402.05079, 2024.

[129] R. Wu, Y. Liu, P. Liang, and Q. Chang, "H-vmunet: High-order vision mamba unet for medical image segmentation," arXiv preprint arXiv:2403.13642, 2024.

[130] Z. Ye and T. Chen, "P-mamba: Marrying perona malik diffusion with mamba for efficient pediatric echocardiographic left ventricular segmentation," arXiv preprint arXiv:2402.08506, 2024.

[131] J. Xie, R. Liao, Z. Zhang, S. Yi, Y. Zhu, and G. Luo, "Promamba: Prompt-mamba for polyp segmentation," arXiv preprint arXiv:2403.13660, 2024.

[132] H. Tang, L. Cheng, G. Huang, Z. Tan, J. Lu, and K. Wu, "Rotate to scan: Unet-like mamba with triplet ssm module for medical image segmentation," arXiv preprint arXiv:2403.17701, 2024.

[133] Z. Wang and C. Ma, "Semi-mamba-unet: Pixel-level contrastive cross-supervised visual mamba-based unet for semi-supervised medical image segmentation," arXiv preprint arXiv:2402.07245, 2024.

[134] J. Liu, H. Yang, H.-Y. Zhou, Y. Xi, L. Yu, Y. Yu, Y. Liang, G. Shi, S. Zhang, H. Zheng, et al., "Swin-umamba: Mamba-based unet with imagenet-based pre-training," arXiv preprint arXiv:2402.03302, 2024.

[135] R. Wu, Y. Liu, P. Liang, and Q. Chang, "Ultralight vm-unet: Parallel vision mamba significantly reduces parameters for skin lesion segmentation," arXiv preprint arXiv:2403.20035, 2024.

[136] J. Ruan and S. Xiang, "Vm-unet: Vision mamba unet for medical image segmentation," arXiv preprint arXiv:2402.02491, 2024.

[137] M. Zhang, Y. Yu, L. Gu, T. Lin, and X. Tao, "Vm-unet-v2 rethinking vision mamba unet for medical image segmentation," arXiv preprint arXiv:2403.09157, 2024.

[138] Z. Wang and C. Ma, "Weak-mamba-unet: Visual mamba makes cnn and vit work better for scribble-based medical image segmentation," arXiv preprint arXiv:2402.10887, 2024.

[139] J. Wang, J. Chen, D. Chen, and J. Wu, "Large window-based mamba unet for medical image segmentation: Beyond convolution and self-attention," arXiv preprint arXiv:2403.07332, 2024.

[140] W. Liao, Y. Zhu, X. Wang, C. Pan, Y. Wang, and L. Ma, "Lightm-unet: Mamba assists in lightweight unet for medical image segmentation," arXiv preprint arXiv:2403.05246, 2024.

[141] Z. Xing, T. Ye, Y. Yang, G. Liu, and L. Zhu, "Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation," arXiv preprint arXiv:2401.13560, 2024.

[142] J. Hao, L. He, and K. F. Hung, "T-mamba: Frequency-enhanced gated long-range dependency for tooth 3d cbct segmentation," arXiv preprint arXiv:2404.01065, 2024.

[143] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, and A. L. Martel, "Loss odyssey in medical image segmentation," Medical Image Analysis, vol. 71, p. 102035, 2021.

[144] R. Zhao, B. Qian, X. Zhang, Y. Li, R. Wei, Y. Liu, and Y. Pan, "Rethinking dice loss for medical image segmentation," in 2020 IEEE International Conference on Data Mining (ICDM), pp. 851–860, IEEE, 2020.

[145] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, "Tversky loss function for image segmentation using 3d fully convolutional deep networks," in International workshop on machine learning in medical imaging, pp. 379–387, Springer, 2017.

[146] S. Jadon, "A survey of loss functions for semantic segmentation," in 2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB), pp. 1–7, IEEE, 2020.

[147] H. Basak and Z. Yin, "Pseudo-label guided contrastive learning for semi-supervised medical image segmentation," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 19786–19797, 2023.

[148] W. Cui, Y. Liu, Y. Li, M. Guo, Y. Li, X. Li, T. Wang, X. Zeng, and C. Ye, "Semi-supervised brain lesion segmentation with an adapted mean teacher model," in Information Processing in Medical Imaging: 26th International Conference, IPMI 2019, Hong Kong, China, June 2–7, 2019, Proceedings 26, pp. 554–565, Springer, 2019.

[149] M. N. Rizve, K. Duarte, Y. S. Rawat, and M. Shah, "In defense of pseudo-labeling: An uncertainty-aware pseudo-label selection framework for semi-supervised learning," arXiv preprint arXiv:2101.06329, 2021.

[150] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li, "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," Advances in neural information processing systems, vol. 33, pp. 596–608, 2020.

[151] Y. Fan, A. Kukleva, D. Dai, and B. Schiele, "Revisiting consistency regularization for semi-supervised learning," International Journal of Computer Vision, vol. 131, no. 3, pp. 626–643, 2023.

[152] K. Raza and N. K. Singh, "A tour of unsupervised deep learning for medical image analysis," Current Medical Imaging, vol. 17, no. 9, pp. 1059–1077, 2021.

[153] C. Baur, S. Denner, B. Wiestler, N. Navab, and S. Albarqouni, "Autoencoders for unsupervised anomaly segmentation in brain mr images: a comparative study," Medical Image Analysis, vol. 69, p. 101952, 2021.

[154] S. Xun, D. Li, H. Zhu, M. Chen, J. Wang, J. Li, M. Chen, B. Wu, H. Zhang, X. Chai, et al., "Generative adversarial networks in medical image segmentation: A review," Computers in biology and medicine, vol. 140, p. 105063, 2022.

[155] K. Chaitanya, E. Erdil, N. Karani, and E. Konukoglu, "Contrastive learning of global and local features for medical image segmentation with limited annotations," Advances in neural information processing systems, vol. 33, pp. 12546–12558, 2020.

[156] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in International conference on machine learning, pp. 1597–1607, PMLR, 2020.

[157] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 9729–9738, 2020.

[158] A. W. Salehi, S. Khan, G. Gupta, B. I. Alabduallah, A. Almjally, H. Alsolai, T. Siddiqui, and A. Mellit, "A study of cnn and transfer learning in medical imaging: Advantages, challenges, future scope," Sustainability, vol. 15, no. 7, p. 5930, 2023.

[159] Y. Fang, P.-T. Yap, W. Lin, H. Zhu, and M. Liu, "Source-free unsupervised domain adaptation: A survey," Neural Networks, p. 106230, 2024.

[160] G. M. Van de Ven, T. Tuytelaars, and A. S. Tolias, "Three types of incremental learning," Nature Machine Intelligence, vol. 4, no. 12, pp. 1185–1197, 2022.

[161] Y. Zhang and Q. Yang, "A survey on multi-task learning," IEEE transactions on knowledge and data engineering, vol. 34, no. 12, pp. 5586–5609, 2021.

[162] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in Proceedings of the IEEE international conference on computer vision, pp. 2961–2969, 2017.

[163] J. An and I. Joe, "Attention map-guided visual explanations for deep neural networks," Applied Sciences, vol. 12, no. 8, p. 3846, 2022.

[164] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim, "Sanity checks for saliency maps," Advances in neural information processing systems, vol. 31, 2018.

[165] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in Proceedings of the IEEE international conference on computer vision, pp. 618–626, 2017.

[166] O. U. Aydin, A. A. Taha, A. Hilbert, A. A. Khalil, I. Galinovic, J. B. Fiebach, D. Frey, and V. I. Madai, "An evaluation of performance measures for arterial brain vessel segmentation," BMC medical imaging, vol. 21, pp. 1–12, 2021.

[167] Y. Liu, H.-S. Kwak, and I.-S. Oh, "Cerebrovascular segmentation model based on spatial attention-guided 3d inception u-net with multi-directional mips," Applied Sciences, vol. 12, no. 5, p. 2288, 2022.

[168] M. V. Orlov, P. Hoffmeister, G. M. Chaudhry, I. Almasry, G. H. Gijsbers, T. Swack, and C. I. Haffajee, "Three-dimensional rotational angiography of the left atrium and esophagusâ€"a virtual computed tomography scan in the electrophysiology lab?," Heart Rhythm, vol. 4, no. 1, pp. 37–43, 2007.

[169] J. Ector, S. De Buck, D. Nuyens, T. Rossenbacker, W. Huybrechts, R. Gopal, F. Maes, and H. Heidbüchel, "Adenosine-induced ventricular asystole or rapid ventricular pacing to enhance three-dimensional rotational imaging during cardiac ablation procedures," Europace, vol. 11, no. 6, pp. 751–762, 2009.

[170] T. Sugahara, Y. Korogi, K. Nakashima, S. Hamatake, S. Honda, and M. Takahashi, "Comparison of 2d and 3d digital subtraction angiography in evaluation of intracranial aneurysms," American Journal of Neuroradiology, vol. 23, no. 9, pp. 1545–1552, 2002.

[171] R. Anxionnat, S. Bracard, X. Ducrocq, Y. Trousset, L. Launay, E. Kerrien, M. Braun, R. Vaillant, F. Scomazzoni, A. Lebedinsky, et al., "Intracranial aneurysms: clinical value of 3d digital subtraction angiography in the therapeutic decision and endovascular treatment," Radiology, vol. 218, no. 3, pp. 799–808, 2001.

[172] P. M. White, J. M. Wardlaw, and V. Easton, "Can noninvasive imaging accurately depict intracranial aneurysms? a systematic review," Radiology, vol. 217, no. 2, pp. 361–370, 2000.

[173] F. Taher, A. Soliman, H. Kandil, A. Mahmoud, A. Shalaby, G. Gimelfarb, and A. El-Baz, "Accurate segmentation of cerebrovasculature from tof-mra images using appearance descriptors," IEEE Access, vol. 8, pp. 96139–96149, 2020.

[174] M. Russ, R. O'Hara, S. S. Nagesh, M. Mokin, C. Jimenez, A. Siddiqui, D. Bednarek, S. Rudin, and C. Ionita, "Treatment planning for image-guided neurovascular interventions using patient-specific 3d printed phantoms," in Medical Imaging 2015: Biomedical Applications in Molecular, Structural, and Functional Imaging, vol. 9417, pp. 580–590, SPIE, 2015.

[175] J. R. Cebral, F. Mut, D. Sforza, R. Löhner, E. Scrivano, P. Lylyk, and C. Putman, "Clinical application of image-based cfd for cerebral aneurysms," International journal for numerical methods in biomedical engineering, vol. 27, no. 7, pp. 977–992, 2011.

[176] N. Chalouhi, B. L. Hoh, and D. Hasan, "Review of cerebral aneurysm formation, growth, and rupture," Stroke, vol. 44, no. 12, pp. 3613–3622, 2013.

[177] A. Firouzian, R. Manniesing, Z. H. Flach, R. Risselada, F. van Kooten, M. C. Sturkenboom, A. van der Lugt, and W. J. Niessen, "Intracranial aneurysm segmentation in 3d ct angiography: Method and quantitative validation with and without prior noise filtering," European journal of radiology, vol. 79, no. 2, pp. 299–304, 2011.

[178] G. Tetteh, V. Efremov, N. D. Forkert, M. Schneider, J. Kirschke, B. Weber, C. Zimmer, M. Piraud, and B. H. Menze, "Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes," Frontiers in Neuroscience, vol. 14, p. 592352, 2020.

[179] T. R. Patel, N. Paliwal, P. Jaiswal, M. Waqas, M. Mokin, A. H. Siddiqui, H. Meng, R. Rai, and V. Tutino, "Multi-resolution cnn for brain vessel segmentation from cerebrovascular images of intracranial aneurysm: a comparison of u-net and deepmedic," in Medical Imaging 2020: Computer-Aided Diagnosis, vol. 11314, pp. 677–685, SPIE, 2020.

[180] K. Kamnitsas, E. Ferrante, S. Parisot, C. Ledig, A. V. Nori, A. Criminisi, D. Rueckert, and B. Glocker, "Deepmedic for brain tumor segmentation," in

Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Second International Workshop, BrainLes 2016, with the Challenges on BRATS, ISLES and mTOP 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 2, pp. 138–149, Springer, 2016.

[181] V. N. Dang, F. Galati, R. Cortese, G. Di Giacomo, V. Marconetto, P. Mathur, K. Lekadir, M. Lorenzi, F. Prados, and M. A. Zuluaga, "Vessel-captcha: an efficient learning framework for vessel annotation and segmentation," Medical Image Analysis, vol. 75, p. 102263, 2022.

[182] M. Livne, J. Rieger, O. U. Aydin, A. A. Taha, E. M. Akay, T. Kossen, J. Sobesky, J. D. Kelleher, K. Hildebrand, D. Frey, et al., "A u-net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease," Frontiers in neuroscience, vol. 13, p. 422707, 2019.

[183] N. F. Dengler, V. I. Madai, J. Wuerfel, F. C. von Samson-Himmelstjerna, P. Dusek, T. Niendorf, J. Sobesky, and P. Vajkoczy, "Moyamoya vessel pathology imaged by ultra–high-field magnetic resonance imaging at 7.0 t," Journal of Stroke and Cerebrovascular Diseases, vol. 25, no. 6, pp. 1544–1551, 2016.

[184] R. Shahzad, L. Pennig, L. Goertz, F. Thiele, C. Kabbasch, M. Schlamann, B. Krischek, D. Maintz, M. Perkuhn, and J. Borggrefe, "Fully automated detection and segmentation of intracranial aneurysms in subarachnoid hemorrhage on cta using deep learning," Scientific Reports, vol. 10, no. 1, p. 21799, 2020.

[185] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," IEEE transactions on medical imaging, vol. 39, no. 6, pp. 1856–1867, 2019.

[186] S. Benkner, A. Arbona, G. Berti, A. Chiarini, R. Dunlop, G. Engelbrecht, A. F. Frangi, C. M. Friedrich, S. Hanser, P. Hasselmeyer, et al., "@neurist: infrastructure for advanced disease management through integration of heterogeneous data, computing, and complex processing services," IEEE transactions on information technology in biomedicine, vol. 14, no. 6, pp. 1365–1377, 2010.

[187] M. Ivantsits, L. Goubergrits, J.-M. Kuhnigk, M. Huellebrand, J. Brüning, T. Kossen, B. Pfahringer, J. Schaller, A. Spuler, T. Kuehne, et al., "Cerebral

aneurysm detection and analysis challenge 2020 (cada)," in Cerebral Aneurysm Detection and Analysis: First Challenge, CADA 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 1, pp. 3–17, Springer, 2021.

[188] W. A. Copen, M. H. Lev, and O. Rapalino, "Brain perfusion: computed tomography and magnetic resonance techniques," Handbook of clinical neurology, vol. 135, pp. 117–135, 2016.

[189] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 3146–3154, 2019.

[190] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. Jorge Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3, pp. 240–248, Springer, 2017.

[191] D. P. Kingma, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[192] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool," BMC medical imaging, vol. 15, pp. 1–28, 2015.

[193] J. Ma and Z. Nie, "Exploring large context for cerebral aneurysm segmentation," in Cerebral Aneurysm Detection and Analysis: First Challenge, CADA 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 1, pp. 68–72, Springer, 2021.

[194] A. Hilbert, V. I. Madai, E. M. Akay, O. U. Aydin, J. Behland, J. Sobesky, I. Galinovic, A. A. Khalil, A. A. Taha, J. Wuerfel, et al., "Brave-net: fully automated arterial brain vessel segmentation in patients with cerebrovascular disease," Frontiers in artificial intelligence, p. 78, 2020.

[195] P. Dai, H. Luo, H. Sheng, Y. Zhao, L. Li, J. Wu, Y. Zhao, and K. Suzuki, "A new approach to segment both main and peripheral retinal vessels based on gray-voting and gaussian mixture model," PloS one, vol. 10, no. 6, p. e0127748, 2015.

[196] M. Ciecholewski and M. Kassjański, "Computational methods for liver vessel segmentation in medical imaging: A review," Sensors, vol. 21, no. 6, p. 2027, 2021.

[197] F. Fu, J. Wei, M. Zhang, F. Yu, Y. Xiao, D. Rong, Y. Shan, Y. Li, C. Zhao, F. Liao, et al., "Rapid vessel segmentation and reconstruction of head and neck angiograms using 3d convolutional neural network," Nature communications, vol. 11, no. 1, p. 4829, 2020.

[198] M. W. Law and A. C. Chung, "Vessel and intracranial aneurysm segmentation using multi-range filters and local variances," in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007: 10th International Conference, Brisbane, Australia, October 29-November 2, 2007, Proceedings, Part I 10, pp. 866–874, Springer, 2007.

[199] T. Krings, D. M. Mandell, T.-R. Kiehl, S. Geibprasert, M. Tymianski, H. Alvarez, K. G. Terbrugge, and F.-J. Hans, "Intracranial aneurysms: from vessel wall pathology to therapeutic approach," Nature Reviews Neurology, vol. 7, no. 10, pp. 547–559, 2011.

[200] E. A. Samaniego, J. A. Roa, and D. Hasan, "Vessel wall imaging in intracranial aneurysms," Journal of neurointerventional surgery, vol. 11, no. 11, pp. 1105–1112, 2019.

[201] X. Chen, Y. Yuan, G. Zeng, and J. Wang, "Semi-supervised semantic segmentation with cross pseudo supervision," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2613–2622, 2021.

[202] S. Chatterjee, K. Prabhu, M. Pattadkal, G. Bortsova, C. Sarasaen, F. Dubost, H. Mattern, M. de Bruijne, O. Speck, and A. Nürnberger, "Ds6, deformation-aware semi-supervised learning: Application to small vessel segmentation with noisy training data," Journal of Imaging, vol. 8, no. 10, p. 259, 2022.

[203] D. Nie, Y. Gao, L. Wang, and D. Shen, "Asdnet: Attention based semi-supervised deep networks for medical image segmentation," in Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV 11, pp. 370–378, Springer, 2018.

[204] S. Chen, G. Bortsova, A. García-Uceda Juárez, G. Van Tulder, and M. De Bruijne, "Multi-task attention-based semi-supervised learning for medical image segmentation," in Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part III 22, pp. 457–465, Springer, 2019.

[205] X. Luo, J. Chen, T. Song, and G. Wang, "Semi-supervised medical image segmentation through dual-task consistency," in Proceedings of the AAAI conference on artificial intelligence, vol. 35, pp. 8801–8809, 2021.

[206] R. Jiao, Y. Zhang, L. Ding, R. Cai, and J. Zhang, "Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation," arXiv preprint arXiv:2207.14191, 2022.

[207] W. T. Cochran, J. W. Cooley, D. L. Favin, H. D. Helms, R. A. Kaenel, W. W. Lang, G. C. Maling, D. E. Nelson, C. M. Rader, and P. D. Welch, "What is the fast fourier transform?," Proceedings of the IEEE, vol. 55, no. 10, pp. 1664–1674, 1967.

[208] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance," Climate research, vol. 30, no. 1, pp. 79–82, 2005.

[209] A. R. Lahitani, A. E. Permanasari, and N. A. Setiawan, "Cosine similarity to determine similarity measure: Study case in online essay assessment," in 2016 4th International Conference on Cyber and IT Service Management, pp. 1–6, IEEE, 2016.

[210] G. Wang, X. Liu, C. Li, Z. Xu, J. Ruan, H. Zhu, T. Meng, K. Li, N. Huang, and S. Zhang, "A noise-robust framework for automatic segmentation of covid-19 pneumonia lesions from ct images," IEEE Transactions on Medical Imaging, vol. 39, no. 8, pp. 2653–2663, 2020.

[211] Z. Xu, Y. Wang, D. Lu, X. Luo, J. Yan, Y. Zheng, and R. K.-y. Tong, "Ambiguity-selective consistency regularization for mean-teacher semi-supervised medical image segmentation," Medical Image Analysis, vol. 88, p. 102880, 2023.

[212] F. Zhao, Y. Chen, F. Chen, X. He, X. Cao, Y. Hou, H. Yi, X. He, and J. Liang, "Semi-supervised cerebrovascular segmentation by hierarchical convolutional neural network," IEEE Access, vol. 6, pp. 67841–67852, 2018.

[213] H. Bogunović, J. M. Pozo, M. C. Villa-Uriol, C. B. Majoie, R. van den Berg, H. A. Gratama van Andel, J. M. Macho, J. Blasco, L. San Román, and A. F. Frangi, "Automated segmentation of cerebral vasculature with aneurysms in 3dra and tof-mra using geodesic active regions: an evaluation study," Medical physics, vol. 38, no. 1, pp. 210–222, 2011.

[214] Y. Deo, R. Bonazzola, H. Dou, Y. Xia, T. Wei, N. Ravikumar, A. F. Frangi, and T. Lassila, "Learned local attention maps for synthesising vessel segmentations from t2 mri," in International Workshop on Simulation and Synthesis in Medical Imaging, pp. 32–41, Springer, 2023.

[215] Y. Deo, H. Dou, N. Ravikumar, A. F. Frangi, and T. Lassila, "Shape-guided conditional latent diffusion models for synthesising brain vasculature," arXiv preprint arXiv:2308.06781, 2023.

[216] Y. Yang and S. Soatto, "Fda: Fourier domain adaptation for semantic segmentation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4085–4095, 2020.

[217] L. Hoyer, D. Dai, and L. Van Gool, "Daformer: Improving network architectures and training strategies for domain-adaptive semantic segmentation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9924–9935, 2022.

[218] L. Hoyer, D. Dai, and L. Van Gool, "Hrda: Context-aware high-resolution domain-adaptive semantic segmentation," in European Conference on Computer Vision, pp. 372–391, Springer, 2022.

[219] L. Hoyer, D. Dai, H. Wang, and L. Van Gool, "Mic: Masked image consistency for context-enhanced domain adaptation," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11721–11732, 2023.

[220] S. Kuang, H. C. Woodruff, R. Granzier, T. J. van Nijnatten, M. B. Lobbes, M. L. Smidt, P. Lambin, and S. Mehrkanoon, "Mscda: Multi-level semantic-guided contrast improves unsupervised domain adaptation for breast mri segmentation in small datasets," Neural Networks, vol. 165, pp. 119–134, 2023.

[221] H. Basak and Z. Yin, "Semi-supervised domain adaptive medical image segmentation through consistency regularized disentangled contrastive learning," arXiv preprint arXiv:2307.02798, 2023.

[222] S. Chatterjee, H. Mattern, M. Dörner, A. Sciarra, F. Dubost, H. Schnurre, R. Khatun, C.-C. Yu, T.-L. Hsieh, Y.-S. Tsai, et al., "Smile-uhura challenge–small vessel segmentation at mesoscopic scale from ultra-high resolution 7t magnetic resonance angiograms," arXiv preprint arXiv:2411.09593, 2024.

[223] W. Lu, J. Wang, H. Li, Y. Chen, and X. Xie, "Domain-invariant feature exploration for domain generalization," Transactions on Machine Learning Research, 2022.

[224] Q. Liu, C. Chen, J. Qin, Q. Dou, and P.-A. Heng, "Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1013–1023, 2021.

[225] S. Wang, L. Yu, K. Li, X. Yang, C.-W. Fu, and P.-A. Heng, "Dofe: Domain-oriented feature embedding for generalizable fundus image segmentation on unseen datasets," IEEE Transactions on Medical Imaging, vol. 39, no. 12, pp. 4237–4248, 2020.

[226] R. Krishnan, P. Rajpurkar, and E. J. Topol, "Self-supervised learning in medicine and healthcare," Nature Biomedical Engineering, vol. 6, no. 12, pp. 1346–1352, 2022.

[227] H. Zhang, Y.-F. Zhang, W. Liu, A. Weller, B. Schölkopf, and E. P. Xing, "Towards principled disentanglement for domain generalization," in Proceedings of the

IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8024–8034, 2022.

[228] T. Yu, S. Kumar, A. Gupta, S. Levine, K. Hausman, and C. Finn, "Gradient surgery for multi-task learning," Advances in Neural Information Processing Systems, vol. 33, pp. 5824–5836, 2020.

[229] L. Mansilla, R. Echeveste, D. H. Milone, and E. Ferrante, "Domain generalization via gradient surgery," in Proceedings of the IEEE/CVF international conference on computer vision, pp. 6630–6638, 2021.

[230] H. Dou, N. Bi, L. Han, Y. Huang, R. Mann, X. Yang, D. Ni, N. Ravikumar, A. F. Frangi, and Y. Huang, "Gsmorph: Gradient surgery for cine-mri cardiac deformable registration," in MICCAI 2023, (Cham), pp. 613–622, Springer Nature Switzerland, 2023.

[231] C. Yang, X. Guo, Z. Chen, and Y. Yuan, "Source free domain adaptation for medical image segmentation with fourier style mining," Medical Image Analysis, vol. 79, p. 102457, 2022.

[232] C. Ouyang, C. Chen, S. Li, Z. Li, C. Qin, W. Bai, and D. Rueckert, "Causality-inspired single-source domain generalization for medical image segmentation," IEEE Transactions on Medical Imaging, vol. 42, no. 4, pp. 1095–1106, 2022.