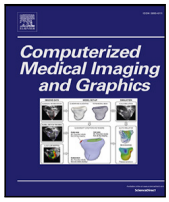
Computerized Medical Imaging and Graphics 115 (2024) 102393

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag



Cross-modality cerebrovascular segmentation based on pseudo-label generation via paired data

Zhanqiang Guo [a], Jianjiang Feng [a,*], Wangsheng Lu [b], Yin Yin [b], Guangming Yang [b], Jie Zhou [a]

[a] Department of Automation, BNRist, Tsinghua University, Beijing, China
[b] UnionStrong (Beijing) Technology Co.Ltd, Beijing, China

ARTICLE INFO

Keywords:
Unsupervised domain adaptation
Pseudo labels
Cerebral vessel segmentation
Registration

ABSTRACT

Accurate segmentation of cerebrovascular structures from Computed Tomography Angiography (CTA), Magnetic Resonance Angiography (MRA), and Digital Subtraction Angiography (DSA) is crucial for clinical diagnosis of cranial vascular diseases. Recent advancements in deep Convolution Neural Network (CNN) have significantly improved the segmentation process. However, training segmentation networks for all modalities requires extensive data labeling for each modality, which is often expensive and time-consuming. To circumvent this limitation, we introduce an approach to train cross-modality cerebrovascular segmentation network based on paired data from source and target domains. Our approach involves training a universal vessel segmentation network with manually labeled source domain data, which automatically produces initial labels for target domain training images. We improve the initial labels of target domain training images by fusing paired images, which are then used to refine the target domain segmentation network. A series of experimental arrangements is presented to assess the efficacy of our method in various practical application scenarios. The experiments conducted on an MRA-CTA dataset and a DSA-CTA dataset demonstrate that the proposed method is effective for cross-modality cerebrovascular segmentation and achieves state-of-the-art performance.

1. Introduction

Cerebrovascular diseases, such as stroke, pose a significant threat to human health worldwide due to their high morbidity and mortality rates (Yasugi et al., 2018; Shao et al., 2022). Alterations in the cerebral arterial system, including vessel stenosis and occlusion, typify stroke. The assessment of arterial vessel status is pivotal for diagnosing and managing cerebrovascular diseases. Such evaluations are instrumental as biomarkers in chronic cerebrovascular conditions that commonly precede stroke events, and the status of cerebrovascular health is a prognostic indicator for the likelihood of subsequent strokes (Gutierrez et al., 2015). In the clinical context, detailed knowledge of cerebrovascular conditions is indispensable. Patients suffering from cardiovascular diseases that result in stroke and myocardial infarction often necessitate comprehensive medical imaging across acute, subacute, and chronic phases, which serves as a vital source of data for clinical decision-making. These imaging techniques play a pivotal role in determining suitability for and monitoring responses to treatments, including thrombectomy and coiling (Birenbaum et al., 2011). For neurosurgeons, the manual analysis of brain scans is a labor-intensive process, particularly when tracing diminutive blood vessels in orthogonal views to elucidate vascular anatomy (Taher et al., 2020). Consequently, the precise and automated segmentation of cerebral vascular structures is critical for effective diagnosis and intervention in vascular diseases.

Computed Tomography Angiography (CTA), Digital Subtraction Angiography (DSA), and Magnetic Resonance Angiography (MRA) have emerged as important imaging technologies to provide extensive data on cerebral vessels, which have significantly advanced cerebrovascular research (Zhang et al., 2020a). CTA is a contrast-based, minimally invasive, and cost-efficient imaging modality, which has been clinically employed in the diagnosis of many vascular diseases (Fu et al., 2020). MRA, a non-contrast, non-invasive, and radiation-free technology, is based on blood flow or inflow angiography. However, the imaging quality of MRA is often reduced due to slow blood flow and slender vessels (Chen et al., 2018). DSA is regarded as the gold standard for the diagnosis of cerebrovascular diseases, since it offers high-resolution assessment to image intuitive vascular structure information (Barlinn and Alexandrov, 2011). Data from these three modalities are widely employed in image-based cranial and peripheral vascular disease diagnosis.

* Corresponding author.
E-mail address: jfeng@tsinghua.edu.cn (J. Feng).

https://doi.org/10.1016/j.compmedimag.2024.102393
Received 7 January 2024; Received in revised form 26 April 2024; Accepted 26 April 2024
Available online 1 May 2024

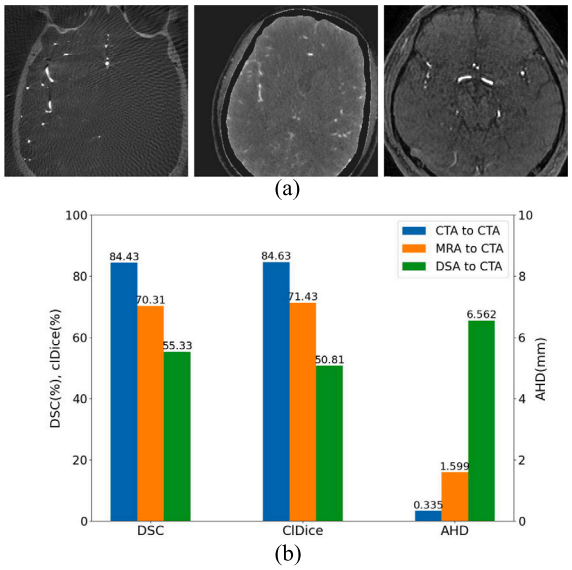


Fig. 1. Due to significant differences between different modalities as shown in (a), direct transfer of the vascular segmentation network results in poor performance (b). The same segmentation network (Çiçek et al., 2016) was trained using the training dataset of each of the three modalities. Subsequently, it was tested on the same CTA test dataset. The higher scores for DSC and $ClDice$ indicate better performance, and reverse for AHD.

Manual annotation of cerebral vessels is an arduous and complex task, characterized by a high degree of anatomical variation and intricate cerebral vascular structure (Almi'ani and Barkana, 2012). With the development of deep Convolution Neural Network (CNN), many automatic segmentation algorithms (Çiçek et al., 2016; Milletari et al., 2016; Qu et al., 2023; Chen et al., 2023) have been proposed to extract vessels with remarkable success. Nevertheless, these methods have a modality-specific limitation since they are trained and tested with data from the same modality. Their performance often declines significantly when tested on images of different modalities or even different devices within the same modality, which is called domain shift (Xu et al., 2018). A clear example of this challenge is presented in Fig. 1, where cerebrovascular images from three modalities (MRA, CTA, and DSA) exhibit significantly different appearances. Collecting annotations in the new domain is a straightforward solution, but it is expensive and time-consuming, even for professional experts.

Unsupervised Domain Adaptation (UDA) method presents an alternative approach to addressing this issue by transferring of knowledge from a labeled source domain to an unlabeled target domain, thereby eliminating the need for further manual annotation in the target domain (Toldo et al., 2020). Recently, most UDA methods have employed Generative Adversarial Networks (GANs) to minimize the differences in image appearances (Zhu et al., 2017; Xu et al., 2023) or latent features (Dou et al., 2019; Tzeng et al., 2017) between the source and target domains. However, for 3D medical images, most GAN-based UDA algorithms treat 3D voxels as a sequence of 2D slices in image synthesis due to hardware limitations and time-consuming training processes (Sun et al., 2020). Consequently, these methods are difficult to apply to 3D cerebrovascular images, where vascular features are not explicitly evident in 2D slice images, complicating the preservation of small vessel integrity and vascular network continuity.

In addition to GAN-based methods, pseudo label generation techniques are also employed in UDA (Chen et al., 2021; Wu et al., 2022). Typically, these methods use a network trained on source domain data to generate initial segmentation results for target domain data. These results are subsequently refined and utilized as pseudo labels to train the network of target domain. However, the performance of these methods is heavily dependent on the quality of the initial segmentation results of target domain images. Consequently, significant discrepancies between the source and target domains, typically in cerebrovascular images, can significantly affect the efficacy of these methods.

To tackle the challenges outlined above, we present a framework to achieve cross-modality cerebrovascular segmentation using paired images of different modalities acquired from the same patient. The central idea of our approach is to obtain high-quality pseudo-labels for the target domain images based on the source domain labels, a challenging task due to the significant differences between modalities and the potential variability in vessels across different imaging sessions. Our framework consists of three phases (as shown in Fig. 2). First, we enhance blood vessels by Hessian matrix-based filtering (Frangi et al., 1998) to reduce the discrepancy between two modalities and train the source domain segmentation network (f_s). Subsequently, given paired images, pseudo labels of target domain images are generated by applying a registration algorithm to align the source domain labels with the initial target domain segmentation results obtained via f_s. Finally, we train the target domain segmentation network (f_t) with the generated pseudo labels.

This paper is an extension of our conference paper (Guo et al., 2023). We propose an improved pseudo-label generation module to address the issue that over-segmentation process leads to thickened blood vessels in the prior framework. Furthermore, the experiments are more comprehensive, covering a wider range of application scenarios, which includes cases without/with labels in the target domain, and instances of multi-domain migration. The efficacy of each framework component, as well as its influence on final segmentation outcomes, is meticulously evaluated. Additionally, we have augmented the experimental databases and test images for more robust analysis. Our main contributions are summarized as follows:

- We present an unsupervised domain adaptation framework based on paired data. To the best of our knowledge, it is the first framework to address the UDA segmentation task of 3D vessel image. Leveraging characteristics of blood vessels, we employ vascular enhancement methods to reduce discrepancy between different domains, thereby facilitating more effective transfer learning processes. We also introduce two different methods to generate high-quality pseudo labels for target domain images.
- We present a series of experimental arrangements for assessing the efficacy of UDA method across various practical application scenarios, including target domain without or with labels, as well as transfer learning between multiple domains. Through meticulous experimentation, we have substantiated the efficiency of the proposed approach by comparing it with the widely utilized unsupervised domain transfer techniques, under a range of experimental settings.

2. Related work

2.1. Vessel segmentation

With the development of deep CNN, many automatic segmentation algorithms (Guo et al., 2021; Qu et al., 2023; Weng et al., 2023; Chen et al., 2023; Guo et al., 2024) have been developed to extract vessels from complex medical images. Ronneberger et al. (2015) proposed the U-Net, which has been widely used in the field of medical image segmentation. Focusing on MRA images, Zhang et al. (2020b) presented a cerebrovascular segmentation framework that removed redundant features and retained edge information in shallow features to obtain better cerebrovascular segmentation. Meng et al. (2020) proposed a Multiscale Dense CNN based on encoder–decoder architecture to automatically segment cerebral vessel in DSA images. And Fu et al. (2020) proposed a reconstruction system, supported by an optimized physiological anatomical-based, to automatically achieve CTA cerebrovascular reconstruction. However, the success of these methods is limited to training and testing with single modality data due to vast differences between data from different modalities.

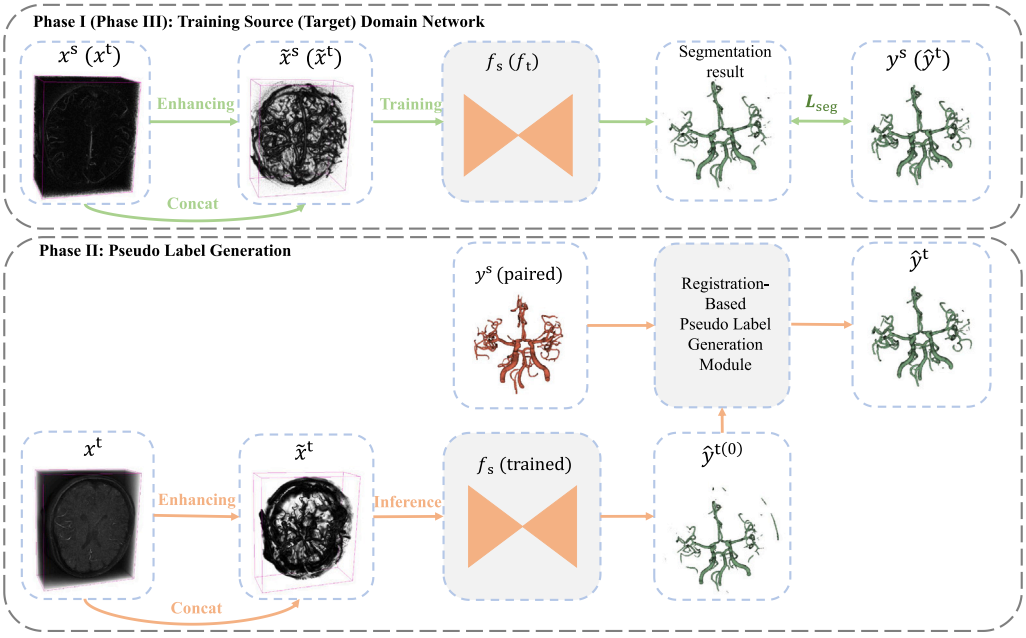


Fig. 2. The proposed cross-modality training framework consists of three phases. First, the source domain network f_s is trained with labeled images (x^s, y^s) from the source domain. Thereafter, the segmentation outcome $\hat{y}^{t(0)}$ of the target domain training image x^t is acquired using f_s (trained), and then transformed into the pseudo label $\hat{y}^t$ by registering it with paired label y^s (paired) from the source domain. Lastly, the target domain segmentation network f_t is trained with the images x^t from the target domain and their corresponding pseudo labels $\hat{y}^t$.

2.2. UDA based on GAN

Previous work has demonstrated that GAN has great potential in image style transfer (Zhu et al., 2017; Azadi et al., 2018), image generation (Li et al., 2023; Sharan et al., 2022) and other computer vision tasks (Al Khalil et al., 2021; Guerreiro et al., 2023). Recently, it has been widely used in UDA task to align the distributions of different domains due to its great superiority in capturing data distribution. Achieving domain adaptation based on GAN mainly focuses on aligning image appearance, feature or a combination of both.

The alignment of image appearance converts the original images of the source domain into the same style as those of target domain, inspired by great success in image-to-image translation (Hoffman et al., 2018; Lei et al., 2021). The target-style images, inheriting labels from the source domain, are then used to train the target domain segmentation network, which is tested with real images from the target domain (Bousmalis et al., 2017). Huo et al. (2018) proposed to translate source images to target domain by cycle generative adversarial networks to achieve cross-modality segmentation, and combined the generation and segmentation networks to implement end-to-end processing. Additionally, aligning the appearance of latent feature focuses on learning the invariant features across different domains by minimizing their distance metric (Dou et al., 2019; Tzeng et al., 2017). Sankaranarayanan et al. (2018) performed adversarial learning in latent feature space to achieve a compact embedding.

Recently, studies are increasingly exploring combined image and feature appearance alignment to more effectively mitigate the domain shift (Han et al., 2021; Pei et al., 2021). For example, Chen et al. (2020) presented Synergistic Image and Feature Alignment (SIFA) to effectively adapt a segmentation network to an unlabeled target domain for cardiac substructure and abdominal multi-organ segmentation. However, these GAN-based approaches for domain adaptation segmentation are limited to 2D images, and even for 3D medical images, they operate on 2D slices due to the complexity and instability of 3D GAN structures. Unfortunately, limited studies focus on utilizing 3D GAN for unsupervised domain adaptation in segmentation. Zhang et al. (2018) proposed a 3D image synthesis approach using large amounts of CT data to achieve realistic heart images. However, this approach is infeasible for most medical image segmentation tasks due to the lack of available data. Yao et al. (2022) developed a three-step framework incorporating a quartet self-attention module to achieve 3D UDA segmentation in brain structures and multi-organ abdominal segmentation. The requirement for image details in vascular segmentation, however, exacerbates the difficulty of 3D vessel image synthesis.

2.3. UDA based on pseudo label

Besides GAN-based methods, pseudo-label generation methods are also used in UDA (Xing et al., 2019; Huynh et al., 2022). These methods typically use networks trained on source domain data to generate the initial results for target domain training data, which are then adjusted to serve as the pseudo-labels of the target domain images. Chen et al. (2021) introduced a complementary pixel-level and class-level denoising scheme with uncertainty estimation to reduce noisy pseudo-labels, and achieved good performance in optic disc and cup segmentation. Furthermore, an uncertainty-based filtering was proposed to select high-quality pseudo-labels during the training of target domain segmentation network (Wu et al., 2022). However, the effectiveness of these methods is heavily reliant on the quality of their initial target domain segmentation outcomes. When significant differences exist between the source and target domains, their performance is severely compromised. This issue is particularly evident in cerebrovascular UDA segmentation, where data from different modalities exhibit substantial disparities, as illustrated in Fig. 1. Therefore, we propose a framework that utilizes paired data to address these issues. Specifically, we employ a Hessian matrix-based vascular enhancement technique to minimize differences between the source and target domains, and subsequently employ image registration to obtain more precise and reliable pseudo-labels.

3. Method

Let $D_s = \{(x_i^s)\}_{i=1}^N$ and $D_t = \{(x_i^t)\}_{i=1}^N$ denote the N paired images from source and target domains, respectively. Images from D_s possess either manual labels or automatic segmentation labels, with the acquisition method detailed below. We aim to exploit D_s and D_t

to improve the performance of the segmentation model within the target domain. As illustrated in Fig. 2, our framework encompasses three phases. Initially, the segmentation network f_s is trained utilizing data from the source domain, which encompass paired data (if D_s has manual labels) or annotated unpaired data. Additionally, segmentation labels for D_s are obtained if manual labels are absent. Subsequently, the preliminary segmentation results of the target domain data D_t are derived from f_s. The Iterative Closest Point (ICP)-based method or the improved VoxelMorph (Balakrishnan et al., 2019) (ImVM) method is utilized to obtain pseudo-labels for the target domain data. Finally, in the third stage, these pseudo labels are used to train the target domain segmentation network, whose structure is identical to that of the source domain network but with different parameters.

3.1. Vessel enhancement and source domain segmentation network (f_s)

As mentioned above, the images from different domains are of great difference due to the different imaging principles (Zhang et al., 2020a), resulting in significant performance deterioration when segmentation networks trained on source domain images D_s are tested on target domain images D_t. To alleviate this challenge, we propose to employ Frangi filtering (Frangi et al., 1998) for vessel enhancement ($x \rightarrow \tilde{x}$) prior to inputting the images into the segmentation network. We computed statistics on the gray-values of the same blood vessel and its surroundings in the paired images from source and target domains, denoted as P_{x^s}, P_{x^t}; similarly, the gray value distribution at the same position in the enhanced images is denoted as $P_{\tilde{x}^s}$, $P_{\tilde{x}^t}$. We calculated the KL divergence and the earth mover's distance (EMD) of the two modal gray distributions before and after vessel enhancement to measure the distance between two distributions. The KL divergence and EMD of the two modal gray distributions are 1.001 and 8.266 before vessel enhancement, and 0.157 and 4.672 afterward, indicating a reduced discrepancy between the source and target domains.

The source domain segmentation network (f_s) is trained using both original images and their corresponding enhanced versions as inputs. The network architecture is customized from the basic 3D U-Net network (Çiçek et al., 2016), with 3D convolution, max pooling, up-sampling and a shortcut connection that bridges the down-sampling and up-sampling paths. (The detailed design of the network structure is shown in Section 3.5 and Fig. 4.) In fact, most popular segmentation networks can also serve as backbones for our proposed UDA framework. Dice loss is used during the training of the segmentation network.

3.2. ICP-based pseudo-label generation

Given the target domain image x_i^t, the segmentation network f_s trained during the first stage can produce the probability map of vessel segmentation. Different threshold values are selected to generate initial segmentation result $\hat{y}_i^{t(0)}$ and over-segmentation result $\hat{y}_i^{t(1)}$:

$$\hat{y}_i^{t(0)} = (f_s(x_i^t) > \alpha_1), \tag{1}$$

$$\hat{y}_i^{t(1)} = (f_s(x_i^t) > \alpha_2). \tag{2}$$

(x_i^s, y_i^s) represents the source domain image and label paired with the target domain image x_i^t, and y_i^s is the manual label or automatic segmentation label, as mentioned above. We regard the foreground points in the initial segmentation outcome $\hat{y}_i^{t(0)}$ and the source domain paired label y_i^s as point clouds and utilize the ICP algorithm to acquire the registration result ($y_i^{s\rightarrow t}$). Due to the accuracy limitation in registration, differences in imaging modalities and variations in blood vessels, the registration result may not integrate precisely with authentic blood vessels of x_i^t in detail. However, it discloses probable blood vessel locations. Therefore, we exploit the trustworthy blood vessels from the over-segmentation result depending on the registration outcome to serve as pseudo label during the fine-tuning procedure, as illustrated in Fig. 3(a). Specifically, at the point p, the value of the pseudo-label is calculated as follows:

$$\hat{y}_i^t[p] = \begin{cases} 1 & \hat{y}_i^{t(1)}[p] = 1 \quad \text{and} \quad d(p, y_i^{s\rightarrow t}) < \epsilon \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where $d(p, y_i^{s\rightarrow t})$ is the minimum distance from point p to the point set of $y_i^{s\rightarrow t}$, and ϵ is the threshold of distance. The point p is selected as foreground in pseudo label when two conditions are satisfied: (i) the value at point p in the over-segmentation result is 1; (ii) the distance from point p to the registered resulting image $y_i^{s\rightarrow t}$ is close enough. The acquired pseudo-label is reliable, as the over-segmentation result typically encompasses most blood vessels.

3.3. ImVM-based pseudo-label generation

To ensure comprehensive inclusion of all blood vessels, a minuscule threshold is necessary in Eq. (4), yet this procedure results in thicker blood vessels in the over-segmentation results, leading to thicker-than-normal blood vessels in the pseudo-labels. Thus, we offer an alternative fine registration method that allows for the use of the registration results combined with initial segmentation outcomes as pseudo-labels, denoted as ImVM. As shown in Fig. 3(b), we use VoxelMorph (Balakrishnan et al., 2019) as the backbone of the registration framework. In accordance with our task, we make two crucial changes to the framework. First, due to the distinct differences in data gray level distributions across different imaging modalities, the performance of VoxelMorph could be impeded. To counteract this, vascular enhancement images are used for supervision as opposed to the original images. Second, due to the lack of annotated data for a given modality in our task, the segmentation result acquired from Phase I replaces the annotation for supervision.

The paired source domain vascular enhancement image $\tilde{x}_i^s$ and target domain vascular enhancement image $\tilde{x}_i^t$ are fed into the feature extraction network f_{vm} to obtain the displacement field u. The deformed image and label are represented by $\tilde{x}_i^s \circ u$ and $y_i^s \circ u$, respectively. The registration results are supervised via MSE loss and Dice loss:

$$L_{sim} = \text{MSE}(\tilde{x}_i^t, \tilde{x}_i^s \circ u) + \lambda_1 \text{Dice}(\hat{y}_i^{t(0)}, y_i^s \circ u). \tag{4}$$

In addition, similar to the work of Balakrishnan et al. (2019), we supervise the spatial gradients of displacement u to ensure its continuity:

$$L_{smooth} = \sum_{p\in\Omega} \|\nabla u(p)\|^2, \tag{5}$$

where Ω is the collection of all points in the image. The overall objective loss function is as follows:

$$L_{final} = L_{sim} + \lambda_2 L_{smooth}. \tag{6}$$

Upon performing inference, the trained network uses paired vascular enhancement images as input to obtain the corresponding deformation field u. The warping label $\hat{y}_i^{t(w)}$ is obtained by applying the deformation field u to the source domain label, mathematically expressed as:

$$\hat{y}_i^{t(w)} = y_i^s \circ u, \quad u = f_{vm}(\tilde{x}_i^s, \tilde{x}_i^t). \tag{7}$$

Given the constraints associated with registration performance, we amalgamate the initial segmentation outcome $\hat{y}_i^{t(0)}$ with the warping result $\hat{y}_i^{t(w)}$ to yield pseudo-labels for the target domain images. Specifically, we consider each interconnected domain c in $\hat{y}_i^{t(0)}$ to be a blood vessel if its concurrence proportion with $\hat{y}_i^{t(w)}$ exceeds a predefined threshold β. The pseudo-label is calculated as follows:

$$\hat{y}_i^t = \hat{y}_i^{t(w)} \cup C, \quad C = \{c | c \in \hat{y}_i^{t(0)}, \frac{|c \cap \hat{y}_i^{t(w)}|}{|c|} > \beta\}. \tag{8}$$

In the ImVM-based module, pseudo-label generation primarily relies on initial segmentation result of target domain ($\hat{y}_i^{t(0)}$) and deformed

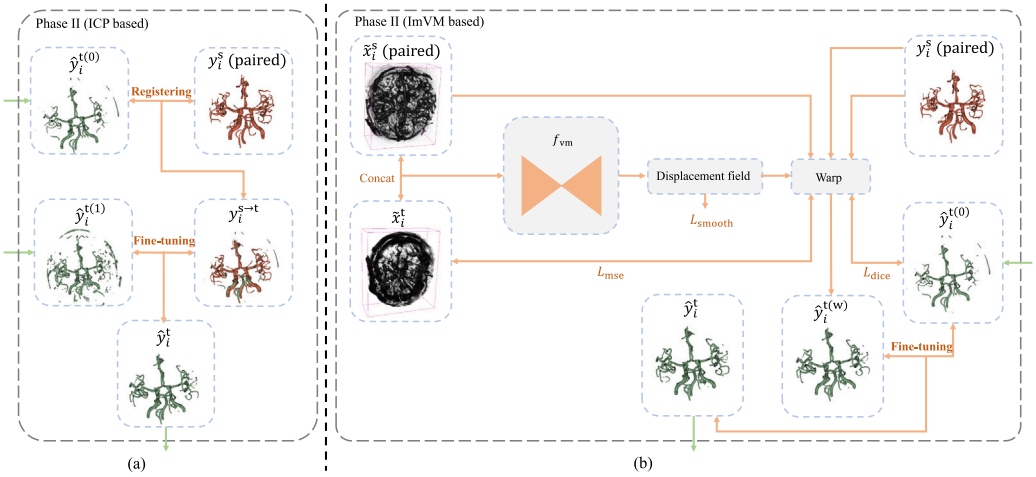


Fig. 3. Pseudo-label generation module based on ICP (a) and Pseudo-label generation module based on ImVM (b). In (a), the initial segmentation result $\hat{y}_i^{t(0)}$ from stage I is aligned with the paired source domain label y_i^s (paired) using the ICP algorithm. The over-segmentation result $\hat{y}_i^{t(1)}$ is then fine-tuned using the registration results $y_i^{s\to t}$ to derive the desired pseudo label $\hat{y}_i^t$. In (b), the paired vascular enhancement images ($\tilde{x}_i^s$ (paired), $\tilde{x}_i^t$) are fed into the registration network to obtain a deformation field. The paired source domain label y_i^s (paired) is then deformed and fine-tuned to produce the pseudo label $\hat{y}_i^t$.

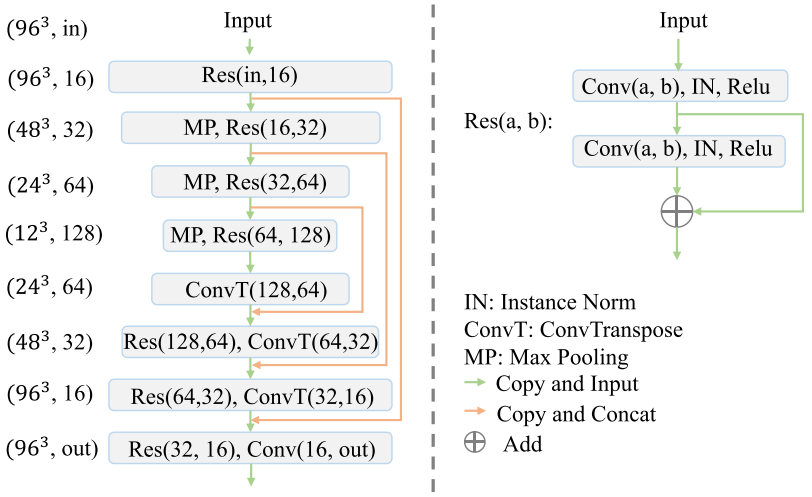


Fig. 4. The structure of network f_s, f_m and f_{vm}. The size and channel of features are displayed on the left side. The step and kernel size in Max Pooling are both set to 2. In the convolution layer labeled as Conv(a, b), 'a' and 'b' represent the number of input and output feature channels, respectively.

source domain label ($\hat{y}_i^{t(w)}$), foregoing the use of over-segmentation result ($\hat{y}_i^{t(1)}$) in the ICP-based module. This approach ensures that the thickness of blood vessels in the pseudo-label closely resembles that of real blood vessels, consequently enhancing the precision of the pseudo label.

3.4. Target domain segmentation network (f_t)

The target domain images and generated pseudo labels $\{(x_i^t, \hat{y}_i^t)\}_{i=1}^N$ are used to train the target domain segmentation network f_t, which has the same structure as the source domain segmentation network f_s. The parameters of f_s obtained during Phase I are implemented for the initialization of f_t. During the inference stage, the target domain image x^t and the enhanced image $\tilde{x}^t$ are fed into f_t to obtain the definitive segmentation outcome.

3.5. Network configurations and implementation details

3.5.1. Network configurations

The source domain segmentation network f_s, target domain segmentation network f_t, and feature extraction network f_{vm} have the same structure, which is customized from the basic 3D U-Net network (Çiçek et al., 2016). The detailed structure of the network is shown in Fig. 4.

3.5.2. Implementation details

Our segmentation network is implemented with PyTorch framework. During the training phase, a NVIDIA GeForce GTX 3090 GPU is used to train the network. We employ adaptive moment estimation (Adam) with an initial learning rate of 0.001. Additionally, a poly learning rate policy (Zhao et al., 2017) with power 0.9 is used. The hyperparameter ϵ is set to 4 pixels. The threshold values α_1 and α_2 are 0.5 and 0.03 respectively, and both the balancing weight λ_1 and λ_2 are 0.01. The predefined threshold β is 0.4. Our code is available at: https://github.com/gzq17/Cross-Modality-Vessel-Segmentation.

4. Experiments and results

4.1. Dataset and evaluation metrics

4.1.1. Dataset

The data utilized in this study were retrospectively collected from Xuanwu Hospital of Capital Medical University, China. The collected

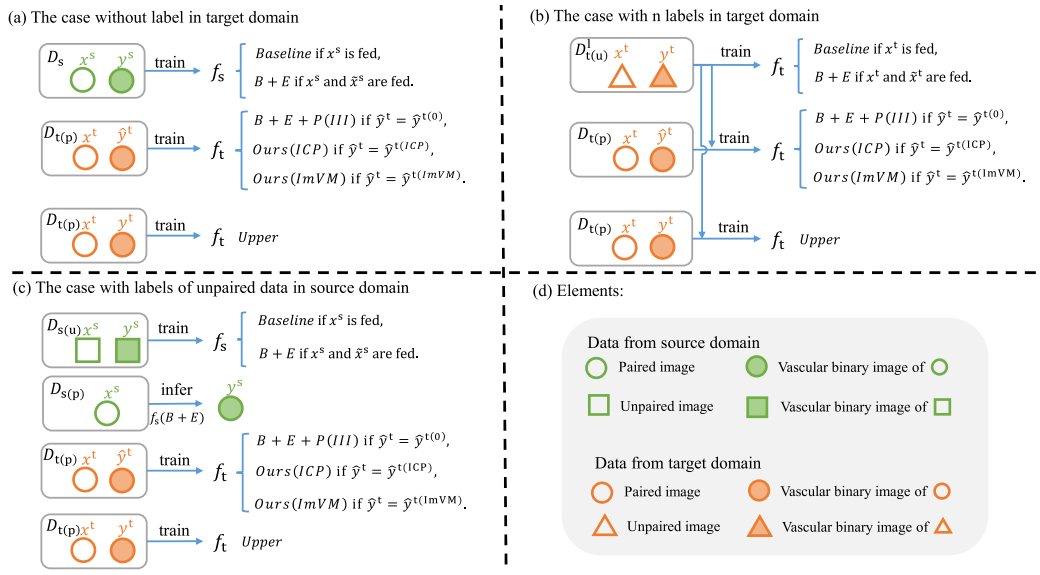


Fig. 5. Experimental settings in several experiments. The difference between *Baseline* and $B+E$ is that the former solely incorporates the original image as input, while the latter extends the input data by additionally integrating the vascular enhancement image. $\hat{y}^{t(0)}$ denotes the blood vessel segmentation outcome derived from the trained network of $B+E$ for paired target domain images. $\hat{y}^{t(ICP)}$ and $\hat{y}^{t(ImVM)}$ represent the pseudo labels obtained using the ICP-based module and the ImVM-based module, respectively.

data include four categories: paired MRA and CTA data from the same patient, paired DSA and CTA data from the same patient, unpaired CTA data, and unpaired MRA data. They form two multi-modal datasets: (i) MRA-CTA, comprising paired MRA and CTA data of 21 patients, unpaired MRA images of 65 patients and unpaired CTA images of 27 patients. (ii) DSA-CTA, which consists of paired DSA and CTA data of 26 patients and unpaired CTA images of 27 patients. The unpaired CTA data in two datasets are the same, and the CTA data have undergone skull-stripping to remove the bright skull regions that affect segmentation (Najm et al., 2019). All images are labeled with blood vessels and have a voxel spacing of $0.5\times0.5\times0.75$ mm^3. During annotation, we use Frangi filtering (Frangi et al., 1998) to generate the initial segmentation results for the vessels. Then, fine corrections are made by two radiologists to obtain the final vascular label. This study got ethical approval of Xuanwu Hospital of Capital Medical University (2020009) for using the clinically collected datasets.

In the experiments, we randomly crop patches of size $96\times96\times96$ to be fed into the network during training. It should be noted that DSA data contain only one vessel branch. As such, when the experiment of DSA-CTA dataset is conducted, the paired DSA and CTA images are separated into two halves along the middle sagittal plane, retaining the half that contains the same vessels as those in the DSA images. During testing, the CTA data annotation does not include the branch of the basilar artery, as this branch is not included in the DSA dataset.

4.1.2. Evaluation metrics

We evaluate the results based on the following metrics: (i) Dice Similarity Coefficient (DSC), (ii) ClDice (Shit et al., 2021), taking into account vascular connectivity, (iii) Average Hausdorff Distance (AHD), which takes voxel localization into consideration. Furthermore, to indicate the statistical significance of improvements of the proposed method, we also present the p-values for DSC using a paired t-test with each comparison method.

4.2. Experimental settings

In this paper, we report the evaluation of the proposed method in various cases in addition to the basic UDA experiment, where the target domain has no labels. The experiments are carried out in the cases where the target domain features a small and a large number of annotations. The efficacy of the proposed method is assessed using both CTA and MRA data as the target domain on the MRA-CTA dataset. Validation is also carried out on the DSA-CTA dataset, where CTA alone is utilized as the target domain. DSA data are excluded as the target domain because the performance of the unsupervised DSA segmentation schemes is acceptable due to the high quality of the DSA data. For clarity, the target domain image set $D_t=\{(x_i^t)\}_{i=1}^M$ is categorized into paired and unpaired dataset, represented by $D_{t(p)}=\{(x_i^{t(p)})\}_{i=1}^N$ and $D_{t(u)}=\{(x_i^{t(u)})\}_{i=1}^T$, respectively ($D_t=D_{t(p)}\cup D_{t(u)}$, $M=N+T$). Except for the investigations conducted in Section 4.9, the labeled images in the source domain have paired data in the target domain, that is $D_s=\{(x_i^s,y_i^s)\}_{i=1}^N$, in all remaining experiments. Via the proposed method, the pseudo-labels of target domain paired images ($D_{t(p)}$) can be generated from the source domain D_s.

4.2.1. The case without label in target domain

We train source domain segmentation network f_s with D_s. The pseudo labels $\{\hat{y}_i^{t(p)}\}_{i=1}^N$ of $D_{t(p)}$ are generated by f_s during Phase II. Subsequently, the target domain network f_t is trained with $\{(x_i^{t(p)},\hat{y}_i^{t(p)})\}_{i=1}^N$. The unpaired images of target domain ($D_{t(u)}$) are used to test the network.

To evaluate the effectiveness of our proposed framework, we reproduced several popular UDA algorithms, including methods based on GAN (SIFA (Chen et al., 2020), DDSeg (Pei et al., 2021)) and methods based on pseudo-label generation (DPL (Chen et al., 2021), FPL (Wu et al., 2022)).

Furthermore, comprehensive studies are conducted to evaluate the efficacy of each module within our proposed framework. The specific experimental settings are shown in Fig. 5(a). It is noteworthy that the segmentation network trained with the target domain images and corresponding labels serves as the upper limit to gauge the performance of the unsupervised domain adaptation segmentation methods (*Upper*).

4.2.2. The case with n labels in target domain

Assuming the presence of n labeled images $D_{t(u)}^l=\{(x_i^{t(u)},y_i^{t(u)})\}_{i=1}^n$ ($D_{t(u)}^l\subset D_{t(u)}$, $n<T$) in the target domain, the network f_t is trained with $D_{t(u)}^l$. This differs from other experiments where the network f_s is trained with labeled data from the source domain. Subsequently, the network outputs initial segmentation results for the paired target domain images ($D_{t(p)}=\{(x_i^{t(p)})\}_{i=1}^N$), whose pseudo labels $\{\hat{y}_i^{t(p)}\}_{i=1}^N$ are obtained in Phase II. Finally, the target domain segmentation network

Table 1
Segmentation performance on MRA-CTA dataset when there is no label in the target domain, with the best performance highlighted in bold. The p of $DSC(p)$ represents the p-value calculated by the t-test, and $*$ indicates the statistical difference between $Ours(ImVM)$ and other methods.

Method	$CTA2MRA$			$MRA2CTA$		
	DSC (%, ↑)	ClDice (%, ↑)	AHD (mm, ↓)	DSC (%, ↑)	ClDice (%, ↑)	AHD (mm, ↓)
Upper	86.60 (***)	89.30	0.445	84.43 (***)	84.63	0.335
SIFA (Chen et al., 2020)	63.85 (***)	70.35	0.629	56.35 (***)	55.63	0.950
DDSeg (Pei et al., 2021)	53.15 (***)	61.09	2.706	59.15 (***)	56.32	3.037
DPL (Chen et al., 2021)	74.27 (***)	74.88	1.361	66.75 (***)	68.80	1.757
FPL (Wu et al., 2022)	75.04 (***)	70.09	1.513	72.59 (*)	74.81	1.262
Baseline	72.81 (***)	65.56	1.234	70.31 (**)	71.43	1.599
B+E	72.94 (***)	64.09	1.052	71.16 (**)	72.08	1.613
B+E+P(III)	72.84 (***)	78.08	0.679	69.82 (***)	77.46	1.771
Ours(ICP)	77.59 (*)	**81.45**	0.600	72.22 (***)	80.76	0.653
Ours(ImVM)	**77.88** (–)	78.76	**0.532**	**75.04** (–)	**82.58**	**0.518**

$*$: $p < 0.05$, $**$: $p < 0.01$, $***$: $p < 0.001$.

Table 2
Segmentation performance on DSA-CTA dataset when there is no label in the target domain, with the best performance highlighted in bold. The p of $DSC(p)$ is consistent with Table 1.

Method	$DSA2CTA$		
	DSC (%, ↑)	ClDice (%, ↑)	AHD (mm, ↓)
Upper	75.93 (***)	73.48	1.514
SIFA (Chen et al., 2020)	58.67 (***)	55.62	2.394
DDSeg (Pei et al., 2021)	49.04 (***)	39.94	5.340
DPL (Chen et al., 2021)	46.64 (***)	46.68	7.056
FPL (Wu et al., 2022)	54.57 (***)	52.38	6.552
Baseline	55.33 (***)	50.81	6.562
B+E	56.17 (***)	53.14	5.891
B+E+P(III)	54.02 (***)	51.72	6.797
Ours(ICP)	65.03 (**)	62.50	2.590
Ours(ImVM)	**69.90** (–)	**63.85**	**1.734**

is trained using n images with real labels from $D^{l}_{t(u)}$, combined with paired images that have generated pseudo labels. The remainder of the unpaired dataset ($D_{t(u)} \setminus D^{l}_{t(u)}$) is reserved for testing purposes.

We carry out a series of comparative experiments to evaluate the efficiency of our proposed method in the cases where there are n labels in the target domain. And the experimental settings are shown in Fig. 5(b).

4.3. Target domain without label

4.3.1. State-of-the-art comparison

As the experimental results are summarized in Tables 1 and 2, the proposed framework improves the vessel segmentation accuracy in terms of DSC, ClDice and AHD when compared with both GAN-based methods and pseudo-label-based methods. Notably, the GAN-based approaches yield suboptimal results relative to the $Baseline$. This is due to the inconspicuous characteristics of vascular features in the 2D slices, which lead to significant artifacts and noise in the results. Moreover, the pseudo-label-based approaches show trivial improvement in comparison to the $Baseline$, because the pseudo-label generation based on uncertainty selection is difficult to solve the wrong segmentation of vascular-like structures across different domains. These findings are consistent with the results in Fig. 6. The proposed framework leverages paired source domain images to generate high-quality pseudo-labels for vascular structures. Additionally, vascular enhancement images are integrated into both the segmentation and pseudo-label generation stages to counteract the inherent domain differences that can degrade performance. Therefore, the proposed approach enhances the accuracy and robustness of target domain segmentation results in medical images with vascular structures.

Table 3
Comparison of the number of parameters and inference time.

Methods	Param (M, ↓)	Time (s, ↓)
SIFA (Chen et al., 2020)	43.7	10.84
DDSeg (Pei et al., 2021)	47.9	16.56
DPL (Chen et al., 2021)	8.08	3.18
FPL (Wu et al., 2022)	8.08	3.18
Ours(ICP/ImVM)	8.08	3.18

4.3.2. Inference speed and memory

To underscore the broader benefits of our proposed methodology, we have detailed the model's parameter count and the average inference time per image in Table 3. It is important to highlight that to ensure fair comparison, the pseudo-label generation methods (DPF, FPL) utilize the same backbone as our approach, ensuring that both the parameters and inference times are aligned across these models. However, as delineated in Section 4.3.1, the segmentation efficacy of these methods significantly trails that of our model. Furthermore, compared to GAN-based methods (SIFA, DDSeg), our approach not only demonstrates superior segmentation performance but also exhibits advantages in terms of reduced parameter count and faster inference time, thus reaffirming the strengths of the proposed framework.

4.3.3. Ablation study

The result of ablation experiment presented in Table 1 shows that the inclusion of vascular enhancement images can effectively reduce the influence of domain differences and enhance the accuracy of segmentation results, as evidenced by the comparison between $Baseline$ and $B+E$. Additionally, the performance of $B+E+Phase(III)$ is deteriorating in contrast to $B+E$. This degradation is potentially attributable to the utilization of noisy labels, impeding the network's ability to learn the correct information. Our proposed approach yields significant improvements in comparison to the $B + E + Phase(III)$, indicating the reliability and effectiveness of our approach in generating high-quality pseudo-label for vascular structures. Moreover, the result using ImVM-based module outperforms that of ICP-based method, which could be attributed to the fact that the pseudo-label generated from over-segmentation result leads to thicker blood vessels. This can also be verified by the MIP results in Fig. 6 (the results of $Ours(ICP)$ include more false positives).

4.4. Comparison of different machine learning methods as backbone

In this subsection, we compare our framework with more deep learning-based fully supervised segmentation methods. Our scheme can leverage most fully supervised methods as backbones, thereby facilitating an analysis of their impacts on our proposed framework. Specifically, we integrate 3D U-Net (Çiçek et al., 2016), DVN (Tetteh et al., 2020), and nnU-Net (Isensee et al., 2021) as backbone

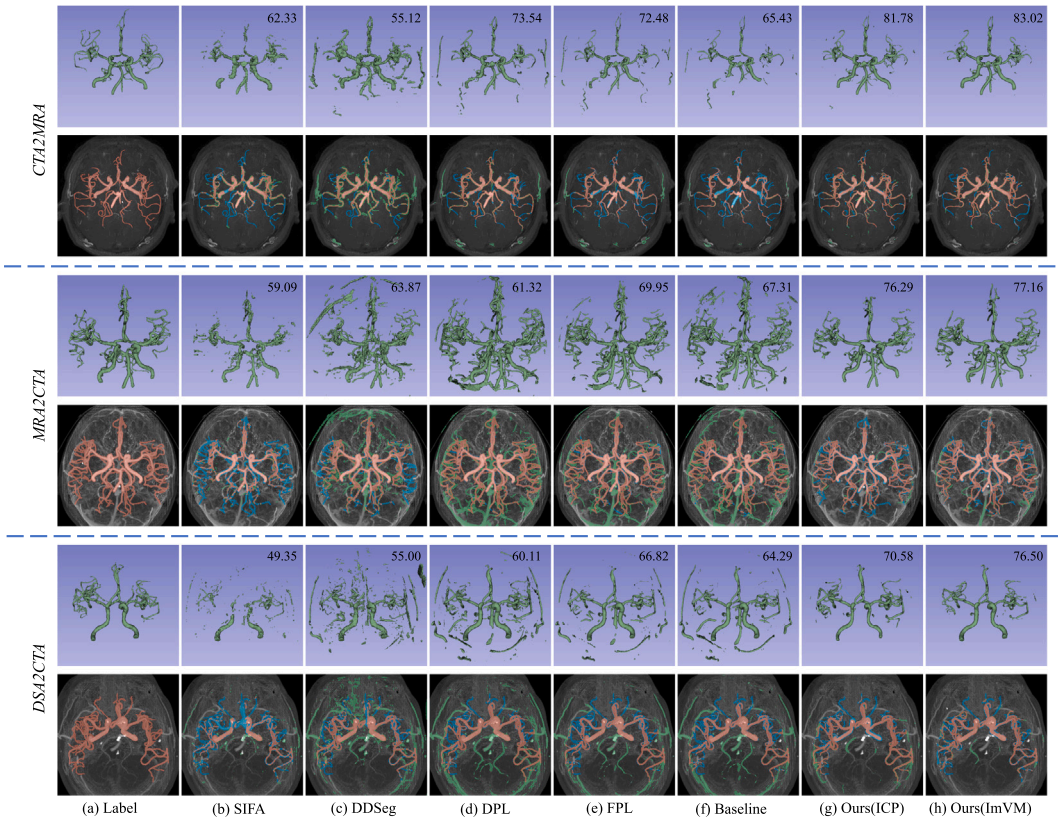


Fig. 6. Segmentation results on three testing images from three experiments in the case where there is no label in target domain. The segmentation outcomes are presented for each image in a two-row format, displaying the 3D results and Maximum Intensity Projection (MIP) results in the transverse plane. The Dice score (%) is incorporated in the upper right corner of the 3D result image. In the MIP results, true positives, false negatives, and false positives are represented by red, blue, and green pixels, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

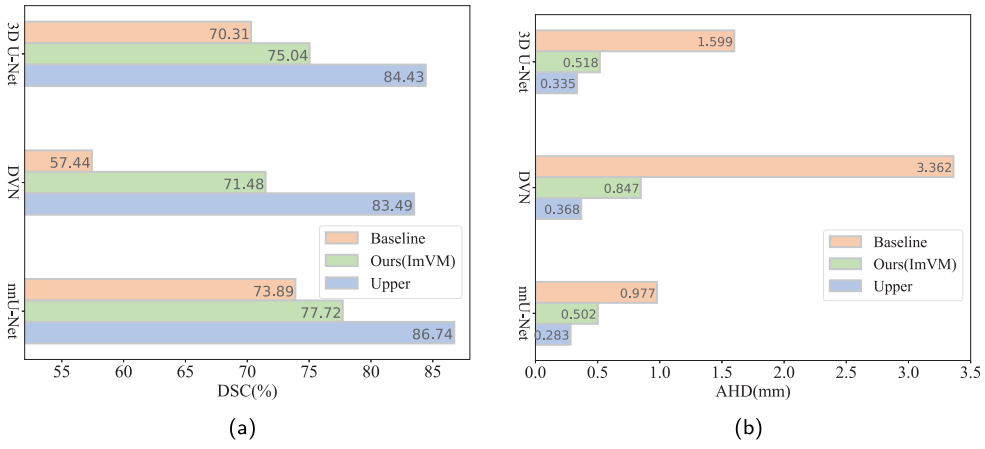


Fig. 7. The segmentation results in the $MRA2CTA$ experiment with different backbones.

Table 4
The number of parameters and inference time of different backbones.

The backbone of method	Param (M, ↓)	Time (s, ↓)
3D U-Net (Çiçek et al., 2016)	8.08	3.18
DVN (Tetteh et al., 2020)	4.69	2.83
nnU-Net (Isensee et al., 2021)	30.78	9.32

architectures of our method and assess their performance in three contexts: direct application for migration (*Baseline*), utilization within our framework (*Ours*($ImVM$)), and direct training with the labeled data from target domain ($Upper$). The results in the $MRA2CTA$ experiment, delineated in Fig. 7, illustrate the performances across different backbones, while Table 4 details the computational requirements, including the number of parameters and the inference time for each backbone. Our method consistently outperforms the *Baseline* across various backbones, with a particularly marked improvement when employing DVN. This enhancement is likely attributable to DVN's heightened sensitivity to domain variances, which adversely affects the performance of *Baseline*. Furthermore, a horizontal comparison across different backbones reveals that nnU-Net outperforms the others, aligning with expectations due to nnU-Net's ability to fully exploit the dataset characteristics. However, this comes at the cost of increased

Table 5
Segmentation performance when there are a few labels in the target domain ($n = 2$), with the best performance shown in bold. The p of $DSC(p)$ is consistent with Table 1.

Method	$CTA2MRA$			$MRA2CTA$			$DSA2CTA$		
	DSC (%, ↑)	ClDice (%, ↑)	AHD (mm, ↓)	DSC (%, ↑)	ClDice (%, ↑)	AHD (mm, ↓)	DSC (%, ↑)	ClDice (%, ↑)	AHD (mm, ↓)
Upper	86.68 (∗∗∗)	89.98	0.429	85.48 (∗∗)	85.09	0.324	83.62 (∗∗)	82.33	1.113
Baseline	77.94 (∗∗∗)	80.65	0.637	72.86 (∗∗∗)	79.02	1.755	76.65 (∗∗)	78.41	2.884
B+E	82.45 (∗∗∗)	81.98	0.928	77.17 (∗∗∗)	75.98	1.404	77.84 (∗)	78.70	2.514
B+E+P(III)	82.69 (∗)	83.61	0.620	78.92 (∗∗)	78.78	1.332	77.85 (∗∗∗)	78.21	2.749
Ours(ICP)	83.92(0.096)	**86.18**	0.479	78.85 (∗∗)	80.27	0.628	79.85 (∗∗∗)	**80.50**	1.345
Ours(ImVM)	**84.06** (–)	84.05	**0.445**	**82.33** (–)	**83.00**	**0.393**	**82.36** (–)	79.90	**0.773**

Table 6
Segmentation performance when there are many labels in the target domain ($n = 15$), with the best result shown in bold. The p of $DSC(p)$ is consistent with Table 1.

Method	$CTA2MRA$			$MRA2CTA$			$DSA2CTA$		
	DSC (%, ↑)	ClDice (%, ↑)	AHD (mm, ↓)	DSC (%, ↑)	ClDice (%, ↑)	AHD (mm, ↓)	DSC (%, ↑)	ClDice (%, ↑)	AHD (mm, ↓)
Upper	86.91 (∗∗∗)	90.24	0.300	86.06(0.093)	85.96	0.272	85.87(0.106)	85.99	0.672
Baseline	85.25 (∗∗∗)	88.67	0.314	82.35 (∗)	83.87	1.129	83.40 (∗∗∗)	83.33	0.899
B+E	85.38 (∗∗∗)	88.92	0.367	82.74 (∗)	83.95	0.902	83.21 (∗∗)	84.10	0.852
B+E+P(III)	86.14 (∗∗)	89.33	0.323	83.57 (∗)	84.88	0.769	82.51 (∗)	83.28	1.772
Ours(ICP)	85.67 (∗∗∗)	89.27	0.378	84.67(0.074)	85.56	0.287	85.51 (∗)	85.84	0.599
Ours(ImVM)	**86.35** (–)	**89.68**	**0.285**	**85.33** (–)	**86.28**	**0.276**	**86.05** (–)	**86.16**	**0.534**

computational load, as evidenced by its larger parameter count and extended inference time presented in Table 4.

4.5. Target domain with n labels

Fully utilizing labeled data from the source domain to improve segmentation in the target domain is crucial, especially when limited labeled data are available in the target domain. Table 5 displays the outcomes acquired by each method when two target domain images have labels ($n = 2$). It is evident that our proposed method significantly improves the segmentation results across all three experimental categories in comparison to the $Baseline$. The use of the ImVM-based module within our framework shows noteworthy improvement of DSC by 6.12%, 9.47%, and 5.71% for $CTA2MRA$, $MRA2CTA$, and $DSA2CTA$, respectively. These outcomes indicate the robust performance of our proposed framework when dealing with scarcity of labeled data in the target domain.

When the target domain already possesses a significant number of labeled data, our proposed approach has proven to still effectively utilize the labels from the source domain to further enhance the segmentation outcomes in target domain. Our methodology achieves remarkable outcomes as indicated in Table 6, where the results when there are 15 labeled images ($n = 15$) in the target domain are shown. It can be seen that the labels from the source domain can be effectively used, irrespective of whether the ICP-based module or the ImVM-based method is employed to generate pseudo-labels. Direct utilization of initial segmentation results as pseudo-labels combined with the existing labeled data to train the target domain segmentation network ($B + E + P(III)$) shows little effect. This is attributed to the difficulty of removing noise in the pseudo-labels resulting from the same network structure. Conversely, our method effectively optimizes the initial segmentation outcomes by utilizing paired labels from source domain to generate high-quality pseudo-labels, ultimately contributing to superior segmentation results in the target domain.

4.6. The effect of the number of labeled images in target domain

In this study, we examine the influence of the number of labeled images in the target domain on segmentation results, in conjunction with the experimentation outlined in earlier sections. Fig. 8 demonstrates the performance of the $Baseline$, $Ours(ICP)$, $Ours(ImVM)$, and $Upper$ while controlling for equivalent test data on MRA-CTA dataset ($MRA2CTA$). As expected, the performance of both the $Baseline$ and theoretical upper limit ($Upper$) increases with n. The performance utilizing the ImVM-based module outperforms that of the ICP-based method due to its ability to generate higher-quality pseudo-labels and to better address the issue of thicker blood vessels in the pseudo-label observed with the latter. In addition, while our proposed framework displays smaller improvements as n increases compared to the $Baseline$, this is to be anticipated given the limited learning capacity of the network, leading to a faster rate of improvement for $Baseline$ in comparison to the upper limit (the performance of network no longer increases once the number of training images reaches a certain quantity). Another notable finding is that as n continues to increase, our proposed method approaches the upper limit, showcasing the potential for our framework to achieve improved results with less annotated data.

4.7. The effect of the quality of generated pseudo-label

4.7.1. Analysis of the pseudo-label generation parameter

In the ImVM-based module, the generation of pseudo labels is predominantly predicated on the initial segmentation outcome and the deformed source domain label through registration. The combination of these outcomes, as outlined by Eq. (8), yields pseudo labels, with β serving as a primary parameter governing the reliance on segmentation results. Fig. 9 shows the impact of β on the generated pseudo labels and the final segmentation results in the $CTA2MRA$ experiment. The comparison of Fig. 9(a) and Fig. 9(b) demonstrates that variations in β lead to similar trends in the quality of the pseudo labels and the performance of segmentation results, which is intuitive as the segmentation network is trained using the generated pseudo labels. Additionally, a large β places greater reliance on the deformed source domain annotation to generate pseudo labels. However, due to registration errors, there may be an offset between the generated pseudo labels and the real labels, as well as vascular discontinuity, ultimately leading to suboptimal segmentation results. Conversely, a small β incorporates segmentation results directly with the registration outcomes as foreground labels, introducing noise from the initial segmentation results and consequently reducing performance. An appropriate β ($\beta = 0.4$ in our experiments) can effectively merge the registration results with reliable regions from the segmentation outcomes to produce high-quality pseudo labels, subsequently leading to enhanced segmentation outcomes in the target domain.

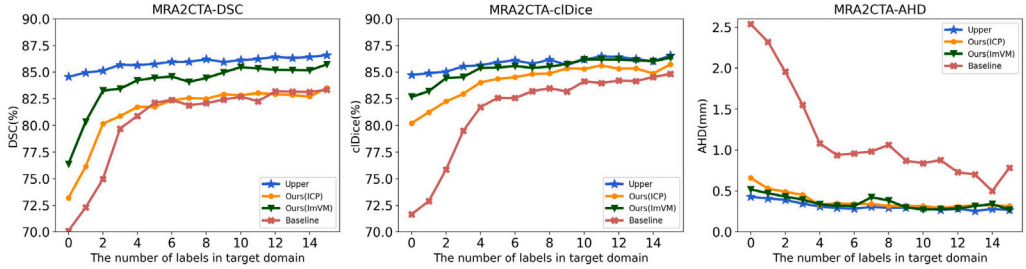


Fig. 8. Performance of each method versus the number of labeled images (n) in the target domain on MRA-CTA dataset with the CTA as target domain ($MRA2CTA$).

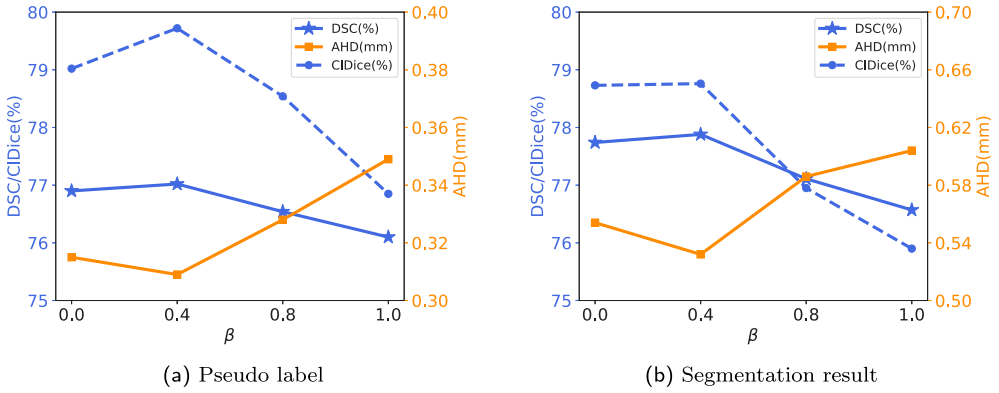


Fig. 9. The impact of parameter β on generated pseudo labels and the segmentation results.

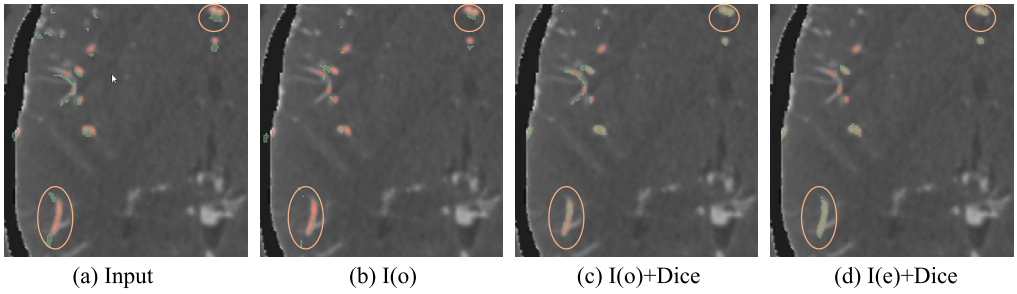


Fig. 10. Registration results in the local area of the 2D slice. Figure (a) shows the unaligned real label and initial segmentation result. Figure (b–d) show the registration results, where the real labels are represented in red and the registration results in green. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 7
Registration and segmentation performance in the $DSA2CTA$ experiment. The results are evaluated by DSC and AHD, and the best performance is highlighted in bold. The p of $DSC(p)$ is consistent with Table 1.

Method	Reg (Pseudo-label)		Seg (Result)	
	DSC(%, ↑)	AHD (mm, ↓)	DSC (%, ↑)	AHD (mm, ↓)
ICP-based	62.31 (***)	2.639	65.03 (**)	2.590
I(o)	46.07 (***)	2.160	49.92 (***)	2.178
I(o)+Dice	65.66 (***)	1.490	65.25 (***)	1.918
I(e)+Dice	**69.73** (–)	**1.154**	**69.90** (–)	**1.734**

4.7.2. Analysis of registration result

Table 7 presents the outcomes of the ablation experiment on the pseudo-label generation module, where the registration performance of the ICP-based method is evaluated using generated pseudo labels, and highlights the influence of registration outcomes on the target domain segmentation results. Notably, in our framework that incorporates the ImVM module, the registration outcomes are used to generate pseudo labels, thus making the quality of the resulting pseudo labels reliant upon the performance of the registration process. Among them, $I(o)$ utilizes only MSE for supervision and uses the original image as input. $I(o)+Dice$ builds on $I(o)$ and supplements MSE with Dice loss, comparing the source domain label to the initial segmentation outcome of the target domain image. $I(e)+Dice$ treats vascular enhancement maps as inputs and ensures the supervision consistency with $I(o)+Dice$.

Table 7 presents compelling evidence that a positive correlation exists between the quality of the generated pseudo labels and segmentation performance in the target domain, a finding that aligns with common intuition. Furthermore, Fig. 10 provides detailed insights into the registration outcomes within specific areas of the 2D slice. With the ImVM-based module, more precise registration results can be obtained, which can translate into an enhancement of the segmentation outcomes in the target domain when employed to generate pseudo labels.

4.8. Comparison of results between MRA2CTA and DSA2CTA

In the experiment of $DSA2CTA$ and $MRA2CTA$, the same CTA testing dataset is used, with the exception of the basilar artery in the $DSA2CTA$ experiment due to its absence in the DSA dataset. A detailed examination of the vascular extraction outcomes derived from CTA data for both experiments is illustrated in Fig. 11.

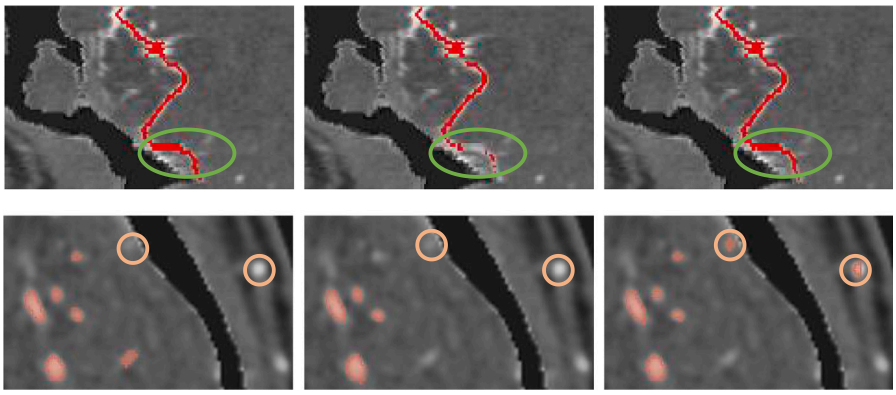


Fig. 11. The segmentation results of one testing CTA sample. (a) shows the corresponding labels, while (b) and (c) show the results of $Our(ImVM)$ method in the $MRA2CTA$ and $DSA2CTA$ experiments ($n = 15$), respectively. The first row shows a blood vessel straightened with Curved Projection Reformation (CPR) on the plane. The second row displays a partial view of a 2D slicer.

Table 8
Segmentation performance of MRA2CTA when there are labels of unpaired data in source domain. The best performance is highlighted in bold. The p of $DSC(p)$ is consistent with Table 1.

Method	$MRA2CTA(Unpaired\ Labels)$		
	DSC (%, ↑)	ClDice (%, ↑)	AHD (mm, ↓)
Upper	84.43 (***)	84.63	0.335
SIFA (Chen et al., 2020)	58.25 (***)	52.36	1.516
DDSeg (Pei et al., 2021)	60.59 (***)	59.60	1.949
DPL (Chen et al., 2021)	64.01 (***)	66.27	1.715
FPL (Wu et al., 2022)	68.76 (***)	74.47	1.207
Baseline	65.70 (***)	71.11	1.582
B+E	67.31 (***)	65.05	2.146
B+E+P(III)	70.95 (***)	72.00	1.441
Ours(ICP)	72.74 (***)	79.29	0.694
Ours(ImVM)	**75.01** (–)	**80.85**	**0.539**

The findings indicate that employing DSA annotated data for training CTA in vascular segmentation enables the delineation of more intricate blood vessels than utilizing MRA data. This is attributed to the capability of DSA in imaging thinner blood vessels. Consequently, using DSA for segmentation training in the CTA domain incorporates a higher number of blood vessels in the outcomes, albeit potentially generating increased vascular-like noise within the vasculature, which is shown in the orange circle in the second row of Fig. 11.

4.9. The case with labels of unpaired data in source domain

In another practical application scenario, the presence of labeled data (denoted as $D_{s(u)} = \{(x_i^{s(u)}, y_i^{s(u)})\}_{i=1}^{K}$) in the source domain lacks corresponding paired data in the target domain. To evaluate the performance of the proposed method in this situation, $D_{s(u)}$ is employed to train the source domain segmentation network f_s, subsequently yielding segmentation results, which serve as labels, for paired data from source domain. The proposed method is then implemented. The ablation experimental design utilized in this experiment is shown in Fig. 5(c). Notably, the pseudo labels are generated using paired source domain labels, which are derived from the network trained on the $B+E$ model.

In this experiment, we set $K = 20$. Table 8 presents the outcomes when MRA is considered as the source domain and CTA as the target domain. Despite the absence of paired labeled data in the source domain, the proposed method continues to deliver commendable results. The experimental findings suggest a reduction in the reliance on paired data in our proposed framework. Furthermore, comparing these results with those in Table 1 reveals a slight decline in performance ($Ours(ICP)$ and $Ours(ImVM)$) due to the use of source domain paired data labels generated through the network, which inherently harbor certain inaccuracies.

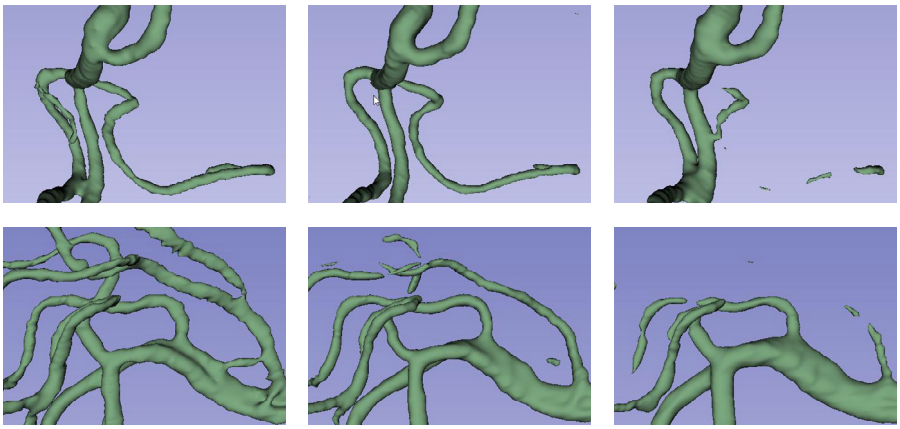
4.10. The result of DSA2CTA2MRA

In this study, we evaluate our proposed approach for three-domain transfer segmentation that combines the MRA-CTA and DSA-CTA datasets. Specifically, we assume the availability of annotated data in DSA, paired data in DSA-CTA and MRA-CTA, but the unavailability of paired data between DSA and MRA. By employing our proposed framework twice, we successfully achieve transfer segmentation from DSA to MRA. In the first step, we use DSA-CTA data to train the segmentation network for the target domain (CTA). Next, we utilize the segmentation results of the paired CTA data in the MRA-CTA dataset as the labels for the CTA data, before proceeding with transfer segmentation from the CTA domain to the MRA domain. Our approach demonstrates effective transfer learning between various medical imaging modalities, highlighting the potential use of modalities with greater commonality or quantity as a bridge for achieving transfer segmentation across modalities. Notably, when implementing the comparative methods, we refrain from utilizing CTA data as a bridge since paired data is not a prerequisite for these approaches. Similar to the $DSA2CTA$ experiment, the testing MRA dataset has also excluded the basilar artery. In this experiment, the setting of the ablation experiment is consistent with Fig. 5(a), with the sole difference being it involves two migrations. Additionally, the present study's experimental scope is limited by dataset availability, although it is reasonable to expect that the method employed could also be adapted for any cases with two-step migration, such as the case of $DSA2MRA2CTA$. The use of DSA as either the second or third modality is unlikely to be of significant worth, given the already high level of segmentation accuracy in DSA.

Table 9 presents the outcomes of the conducted experiment. In agreement with earlier experimental findings, GAN-based UDA techniques exhibit suboptimal performance. Pseudo-label generation-based

Table 9
Segmentation performance of DSA2CTA2MRA experiment when there is no label in the target domain. The best performance is highlighted in bold. The p of $DSC(p)$ is consistent with Table 1.

Method	$DSA2CTA2MRA$		
	DSC (%, ↑)	ClDice (%, ↑)	AHD (mm, ↓)
Upper	87.73 (∗∗∗)	90.66	0.334
SIFA (Chen et al., 2020)	55.69 (∗∗∗)	53.62	3.254
DDSeg (Pei et al., 2021)	51.03 (∗∗∗)	42.19	5.942
DPL (Chen et al., 2021)	56.21 (∗∗∗)	56.86	5.404
FPL (Wu et al., 2022)	57.63 (∗∗∗)	56.24	4.900
Baseline	56.43 (∗∗∗)	56.44	5.498
B+E	60.79 (∗∗∗)	57.89	4.401
B+E+P(III)	62.47 (∗∗∗)	58.97	3.944
Ours(ICP)	**73.44** (∗∗)	**73.75**	1.066
Ours(ImVM)	72.12 (–)	72.93	**0.958**



(a) Label (b) *CT2MRA* (c) *DSA2CTA2MRA*

Fig. 12. The segmentation results of two testing MRA samples. For both the $CTA2MRA$ and $DSA2CTA2MRA$ experiments, identical test data are employed. The accompanying figure exhibits the local 3D visualization results for the same region of the identical test image.

methods demonstrate a marginal enhancement over the *Baseline*. Notably, the proposed framework surpasses the comparative approaches, even after two domain migrations, highlighting its efficacy and potential for multiple domain transfers. A further observation reveals that the performance of the ICP-based module marginally exceeds that of the ImVM-based module. This discrepancy may be attributed to the heightened errors introduced by the latter's dual fine registrations, while the former employs over-segmentation results derived from the DSA training model on MRA data during pseudo-label generation. Moreover, the testing images utilized in the $DSA2CTA2MRA$ and $CTA2MRA$ ($n = 0$) experiments remain consistent, barring the exclusion of the basilar artery evaluation in the former. Fig. 12 juxtaposes the outcomes obtained from both experiments on identical data, revealing that employing CTA annotated data for training the MRA segmentation model yields superior results compared to $DSA2CTA2MRA$. This observation aligns with the expected outcomes, as in the $DSA2CTA2MRA$ experiment, the CTA labels used in the second step (CTA2MRA experiment) are generated through the first step (DSA2CTA experiment), which inherently contain errors.

5. Discussion

In this study, we introduce an unsupervised domain adaptation approach that leverages paired data. We employ a Hessian matrix-based vessel enhancement method on images to reduce domain discrepancy. Subsequently, a registration method is employed to create high-quality pseudo-labels for training the segmentation network on the target domain. The findings presented in Table 1 illustrate that the proposed method significantly outperforms existing domain transfer segmentation approaches on vascular datasets.

The popular GAN-based UDA methods (Hoffman et al., 2018; Han et al., 2021; Pei et al., 2021; Chen et al., 2020) have shown disappointing outcomes in the task of 3D cerebral vascular segmentation due to the limited salience of blood vessel features in the 2D sections incorporated in these approaches. In light of this, we present an alternative method that eschews GAN-based feature alignment in favor of a multi-stage strategy for blood vessel segmentation in 3D space, which yields impressive results. Moreover, our proposed approach effectively capitalizes on the paired data to reduce the adverse influence of strong dependency on initial segmentation results, a common issue in some current pseudo-label generation methods (Chen et al., 2021; Wu et al., 2022).

Furthermore, we validate our method in the cases where there are various numbers of labels in the target domain, as presented in Tables 5 and 6. The outcomes show that our approach is not only appropriate for unsupervised domain migration but also performs well when some labels are present in the target domain. In addition, as demonstrated in Fig. 8, the performance of our method approaches that of using all real labels when the number of labels in the target domain is relatively large, indicating the potential of our method to achieve better performance with fewer labeled images. By examining Fig. 8 and the findings in Section 4.9, it can be inferred that when labeled data is available in the source domain, our proposed framework can produce acceptable segmentation outcomes. This is achieved by utilizing data paired with the source domain that is available in the target domain, while eliminating or significantly reducing labeling requirements in the target domain.

We recognize there are some limitations to our proposed approach. Specifically, the training process requires paired data, which constrains our method's applicability to a certain extent. Additionally, the training

phase is segmented into three stages rather than an end-to-end procedure, though the resultant inference is indeed end-to-end. This division of the training phase may constitute a limitation to the effectiveness of our approach. Thus, as future work, we aim to explore methods for optimizing the training process to streamline the training phase and mitigate these limitations.

6. Conclusion

In this paper, we propose an unsupervised domain adaptation methodology that hinges on paired data. The crux of our approach is the generation of reliable pseudo labels for the target domain using paired images from source and target domains. During each training stage, we incorporate vascular enhancement images in conjunction with the original images to mitigate the influence of domain shifts. The experimental results demonstrate that our proposed method effectively achieves unsupervised domain adaptation of segmenting blood vessels. Importantly, we establish the efficacy of our approach in the presence of labeled data in the target domain, thereby offering the prospect of improved performance with reduced labeled data. Future studies will focus on extending the proposed method to other vessel datasets.

CRediT authorship contribution statement

Zhanqiang Guo: Writing – original draft, Software, Methodology. **Jianjiang Feng:** Writing – review & editing, Methodology, Investigation, Funding acquisition, Conceptualization. **Wangsheng Lu:** Validation, Data curation. **Yin Yin:** Project administration, Methodology. **Guangming Yang:** Resources, Funding acquisition. **Jie Zhou:** Supervision, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This study got ethical approval of Xuanwu Hospital of Capital Medical University, China (2020009) for using the clinically collected cerebral MRA-CTA dataset and DSA-CTA dataset.

References

Al Khalil, Y., Amirrajab, S., Pluim, J., Breeuwer, M., 2021. Late fusion U-net with GAN-based augmentation for generalizable cardiac MRI segmentation. In: International Workshop on Statistical Atlases and Computational Models of the Heart. Springer, pp. 360–373.

Almi'ani, M.M., Barkana, B.D., 2012. Automatic segmentation algorithm for brain MRA images. In: IEEE Long Island Systems, Applications and Technology Conference. pp. 1–5.

Azadi, S., Fisher, M., Kim, V.G., Wang, Z., Shechtman, E., Darrell, T., 2018. Multi-content GAN for few-shot font style transfer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7564–7573.

Balakrishnan, G., Zhao, A., Sabuncu, M.R., Guttag, J., Dalca, A.V., 2019. Voxelmorph: a learning framework for deformable medical image registration. IEEE Trans. Med. Imag 38 (8), 1788–1800.

Barlinn, K., Alexandrov, A.V., 2011. Vascular imaging in stroke: comparative analysis. Neurotherapeutics 8 (3), 340–348.

Birenbaum, D., Bancroft, L.W., Felsberg, G.J., 2011. Imaging in acute stroke. West. J. Emerg. Med 12 (1), 67–76.

Bousmalis, K., Silberman, N., Dohan, D., Erhan, D., Krishnan, D., 2017. Unsupervised pixel-level domain adaptation with generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3722–3731.

Chen, C., Dou, Q., Chen, H., Qin, J., Heng, P.A., 2020. Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation. IEEE Trans. Med. Imag 39 (7), 2494–2505.

Chen, C., Liu, Q., Jin, Y., Dou, Q., Heng, P.A., 2021. Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 225–235.

Chen, X., Lu, Y., Bai, J., Yin, Y., Cao, K., Li, Y., Chen, H., Song, Q., Wu, J., 2018. Train a 3D U-Net to segment cranial vasculature in CTA volume without manual annotation. In: International Symposium on Biomedical Imaging. IEEE, pp. 559–563.

Chen, C., Zhou, K., Wang, Z., Zhang, Q., Xiao, R., 2023. All answers are in the images: A review of deep learning for cerebrovascular segmentation. Comput. Med. Imag. Graph 107, 102229.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 424–432.

Dou, Q., Ouyang, C., Chen, C., Chen, H., Glocker, B., Zhuang, X., Heng, P.A., 2019. Pnp-adanet: Plug-and-play adversarial domain adaptation network at unpaired cross-modality cardiac segmentation. IEEE Access 7, 99065–99076.

Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A., 1998. Multiscale vessel enhancement filtering. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 130–137.

Fu, F., Wei, J., Zhang, M., Yu, F., Xiao, Y., Rong, D., Shan, Y., et al., 2020. Rapid vessel segmentation and reconstruction of head and neck angiograms using 3D convolutional neural network. Nat. Commun 11 (1), 1–12.

Guerreiro, J., Tomás, P., Garcia, N., Aidos, H., 2023. Super-resolution of magnetic resonance images using generative adversarial networks. Comput. Med. Imag. Graph 108, 102280.

Guo, Z., Feng, J., Lu, W., Yin, Y., Yang, G., Zhou, J., 2023. Training cross-modality cerebrovascular segmentation networks with paired images. In: International Symposium on Biomedical Imaging. IEEE, pp. 1–5.

Guo, Z., Tan, Z., Feng, J., Zhou, J., 2024. 3D vascular segmentation supervised by 2D annotation of maximum intensity projection. IEEE Trans. Med. Imag. Early Access.

Guo, X., Xiao, R., Lu, Y., Chen, C., Yan, F., Zhou, K., He, W., Wang, Z., 2021. Cerebrovascular segmentation from TOF-MRA based on multiple-U-net with focal loss function. Comput. Methods Programs Biomed. 202, 105998.

Gutierrez, J., Cheung, K., Bagci, A., Rundek, T., Alperin, N., Sacco, R.L., Wright, C.B., et al., 2015. Brain arterial diameters as a risk factor for vascular events. J. Am. Heart Assoc 4 (8), e002289.

Han, X., Qi, L., Yu, Q., Zhou, Z., Zheng, Y., Shi, Y., Gao, Y., 2021. Deep symmetric adaptation network for cross-modality medical image segmentation. IEEE Trans. Med. Imag. 41 (1), 121–132.

Hoffman, J., Tzeng, E., Park, T., Zhu, J.Y., Isola, P., Saenko, K., Efros, A., Darrell, T., 2018. Cycle-consistent adversarial domain adaptation. In: International Conference on Machine Learning. pp. 1989–1998.

Huo, Y., Xu, Z., Moon, H., Bao, S., Assad, A., Moyo, T.K., Savona, M.R., Abramson, R.G., Landman, B.A., 2018. Synseg-net: Synthetic segmentation without target modality ground truth. IEEE Trans. Med. Imag 38 (4), 1016–1025.

Huynh, D., Kuen, J., Lin, Z., Gu, J., Elhamifar, E., 2022. Open-vocabulary instance segmentation via robust cross-modal pseudo-labeling. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7020–7031.

Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H., 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 18 (2), 203–211.

Lei, H., Liu, W., Xie, H., Zhao, B., Yue, G., Lei, B., 2021. Unsupervised domain adaptation based image synthesis and feature alignment for joint optic disc and cup segmentation. IEEE J. Biomed. Health Inform 26 (1), 90–102.

Li, J., Chen, T., Qian, X., 2023. Generalizable pancreas segmentation modeling in CT imaging via meta-learning and latent-space feature flow generation. IEEE J. Biomed. Health Inform 27 (1), 374–385.

Meng, C., Sun, K., Guan, S., Wang, Q., Zong, R., Liu, L., 2020. Multiscale dense convolutional neural network for DSA cerebrovascular segmentation. Neurocomputing 373, 123–134.

Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision. 3DV, pp. 565–571.

Najm, M., Kuang, H., Federico, A., Jogiat, U., Goyal, M., Hill, M.D., Demchuk, A., Menon, B.K., Qiu, W., 2019. Automated brain extraction from head CT and CTA images using convex optimization with shape propagation. Comput. Methods Programs Biomed. 176, 1–8.

Pei, C., Wu, F., Huang, L., Zhuang, X., 2021. Disentangle domain features for cross-modality cardiac image segmentation. Med. Image Anal. 71, 102078.

Qu, Z., Zhuo, L., Cao, J., Li, X., Yin, H., Wang, Z., 2023. Tp-net: Two-path network for retinal vessel segmentation. IEEE J. Biomed. Health Inform 27 (4), 1979–1990.

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

Sankaranarayanan, S., Balaji, Y., Jain, A., Lim, S.N., Chellappa, R., 2018. Learning from synthetic data: Addressing domain shift for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3752–3761.

Shao, S., Wang, T., Mumtaz, A., Song, C., Yao, C., 2022. Predicting cardiovascular and cerebrovascular events based on instantaneous high-order singular entropy and deep belief network. IEEE J. Biomed. Health Inform 27 (4), 1670–1680.

Sharan, L., Romano, G., Koehler, S., Kelm, H., Karck, M., DeSimone, R., Engelhardt, S., 2022. Mutually improved endoscopic image synthesis and landmark detection in unpaired image-to-image translation. IEEE J. Biomed. Health Inform 27 (1), 374–385.

Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.P.W., Bauer, U., Menze, B.H., 2021. ClDice-a novel topology-preserving loss function for tubular structure segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 16560–16569.

Sun, Y., Yuan, P., Sun, Y., 2020. MM-GAN: 3D MRI data augmentation for medical image segmentation via generative adversarial networks. In: IEEE International Conference on Knowledge Graph. ICKG, IEEE, pp. 227–234.

Taher, F., Soliman, A., Kandil, H., Mahmoud, A., Shalaby, A., Gimel'farb, G., El-Baz, A., 2020. Accurate segmentation of cerebrovasculature from TOF-mra images using appearance descriptors. IEEE Access 8, 96139–96149.

Tetteh, G., Efremov, V., Forkert, N.D., Schneider, M., Kirschke, J., Weber, B., Zimmer, C., et al., 2020. Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-D angiographic volumes. Front. Neurosci. 14, 592352.

Toldo, M., Maracani, A., Michieli, U., Zanuttigh, P., 2020. Unsupervised domain adaptation in semantic segmentation: a review. Technologies 8 (2), 35.

Tzeng, E., Hoffman, J., Saenko, K., Darrell, T., 2017. Adversarial discriminative domain adaptation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7167–7176.

Weng, W., Ding, H., Bai, J., Zhou, W., Wang, G., 2023. Cerebrovascular segmentation in phase-contrast magnetic resonance angiography by a radon projection composition network. Comput. Med. Imag. Graph 107, 102228.

Wu, J., Gu, R., Dong, G., Wang, G., Zhang, S., 2022. FPL-UDA: Filtered pseudo label-based unsupervised cross-modality adaptation for vestibular schwannoma segmentation. In: International Symposium on Biomedical Imaging. IEEE, pp. 1–5.

Xing, F., Bennett, T., Ghosh, D., 2019. Adversarial domain adaptation and pseudo-labeling for cross-modality microscopy image quantification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 740–749.

Xu, X., Chen, Y., Wu, J., Lu, J., Ye, Y., Huang, Y., Dou, X., et al., 2023. A novel one-to-multiple unsupervised domain adaptation framework for abdominal organ segmentation. Med. Image Anal 88, 102873.

Xu, R., Chen, Z., Zuo, W., Yan, J., Lin, L., 2018. Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3964–3973.

Yao, K., Su, Z., Huang, K., Yang, X., Sun, J., Hussain, A., Coenen, F., 2022. A novel 3D unsupervised domain adaptation framework for cross-modality medical image segmentation. IEEE J. Biomed. Health Inform 26 (10), 4976–4986.

Yasugi, M., Hossain, B., Nii, M., Kobashi, S., 2018. Relationship between cerebral aneurysm development and cerebral artery shape. J. Adv. Comput. Intell. Intell. Inform 22 (2), 249–255.

Zhang, B., Liu, S., Zhou, S., Yang, J., Wang, C., Li, N., Wu, Z., Xia, J., 2020a. Cerebrovascular segmentation from TOF-MRA using model-and data-driven method via sparse labels. Neurocomputing 380, 162–179.

Zhang, H., Xia, L., Song, R., Yang, J., Hao, H., Liu, J., Zhao, Y., 2020b. Cerebrovascular segmentation in MRA via reverse edge attention network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 66–75.

Zhang, Z., Yang, L., Zheng, Y., 2018. Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9242–9251.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2881–2890.

Zhu, J.Y., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2223–2232.