Multi-Domain Brain Vessel Segmentation Through Feature Disentanglement

Francesco **Galati*** [1], Daniele **Falcetta*** [1], Rosa Cortese [2], Ferran Prados [3,4,5], Ninon Burgos [6], Maria A. Zuluaga [1,7]

1 Data Science Department, EURECOM, Sophia Antipolis, FR
2 Department of Medicine, Surgery and Neuroscience, University of Siena, IT
3 e-Health Center, Universitat Oberta de Catalunya, Barcelona, ES
4 Centre for Medical Image Computing (CMIC), Department of Medical Physics and Biomedical Engineering, University College London, UK
5 Queen Square MS Centre, Department of Neuroinflammation, UCL Queen Square Institute of Neurology, Faculty of Brain Sciences, University College London, UK
6 Sorbonne Université, Paris Brain Institute, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, Paris, FR
7 School of Biomedical Engineering & Imaging Sciences, King's College London, UK
* Joint first authors

Abstract

The intricate morphology of brain vessels poses significant challenges for automatic segmentation models, which usually focus on a single imaging modality. However, accurately treating brain-related conditions requires a comprehensive understanding of the cerebrovascular tree, regardless of the specific acquisition procedure. Our framework effectively segments brain arteries and veins in various datasets through image-to-image translation while avoiding domain-specific model design and data harmonization between the source and the target domain. This is accomplished by employing disentanglement techniques to independently manipulate different image properties, allowing them to move from one domain to another in a label-preserving manner. Specifically, we focus on manipulating vessel appearances during adaptation while preserving spatial information, such as shapes and locations, which are crucial for correct segmentation. Our evaluation effectively bridges large and varied domain gaps across medical centers, image modalities, and vessel types. Additionally, we conduct ablation studies on the optimal number of required annotations and other architectural choices. The results highlight our framework's robustness and versatility, demonstrating the potential of domain adaptation methodologies to perform cerebrovascular image segmentation in multiple scenarios accurately. Our code is available at https://github.com/i-vesseg/MultiVesSeg.

1. Introduction

Segmenting the cerebrovascular tree is crucial for accurately diagnosing and treating several brain-related conditions. The complex and intricate morphology of brain vessels requires the usage of multiple imaging modalities. Each modality has specific properties targeting a vessel type: angiographies focus on visualizing the arteries in the brain, while venographies primarily examine the veins. This variety of imaging modalities, combined with the different acquisition protocols and scanners utilized in clinical centers, poses challenges for automatic segmentation models, which struggle to generalize across different domains, i.e., varying centers, modalities, or vessel types (arteries or veins). When trained on a single source domain, models may become susceptible to distribution shifts, i.e., their performance may
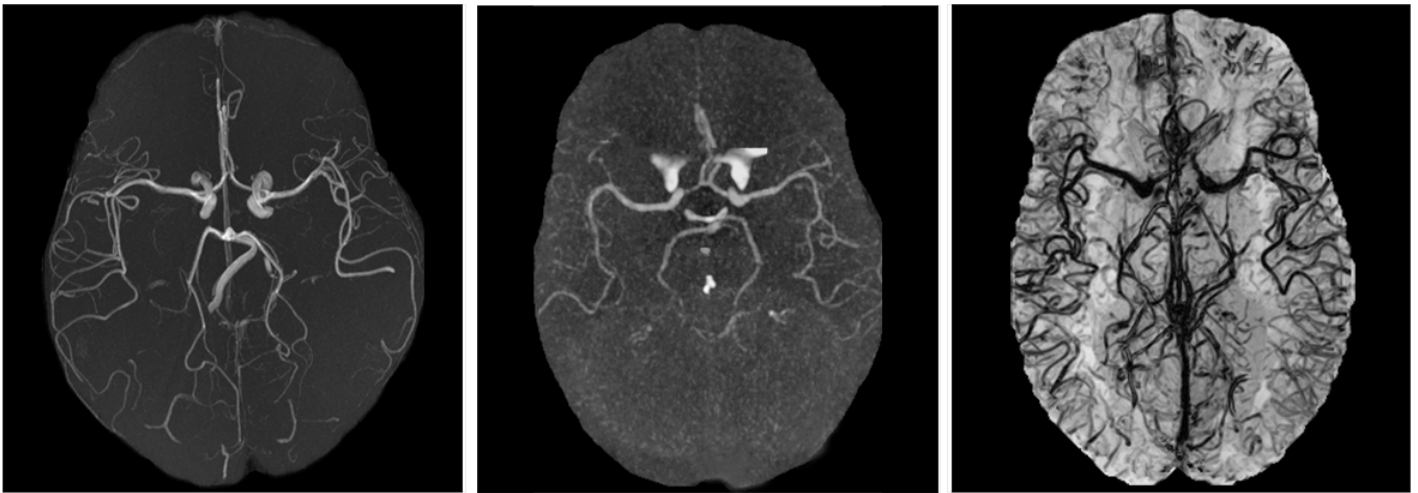
Figure 1: Maximum intensity projection (MIP) of a magnetic resonance angiography (left), MIP of a computed tomography angiography (center), and minimum intensity projection (mIP) of a magnetic resonance venography (right). All images are skull-stripped and viewed from the axial perspective.

decline when transitioning from one domain to another. At the same time, developing, deploying, and maintaining a segmentation model for each domain is impractical, as collecting medical images is costly, and data annotation is laborious and demands a high level of expertise.

Distribution shift has been approached from various perspectives, depending on the amount of labeled data available. Among the different approaches, domain adaptation (DA) aims to transfer predictive knowledge from a source domain with abundant labeled data to a target domain with limited or no labeled data (Guan and Liu, 2022). Despite numerous successful attempts to apply domain adaptation techniques in medical imaging, no current work has focused on brain vessel segmentation. Two factors may explain this. First, regardless of the number of modalities being examined, vessel segmentation remains challenging due to the relatively small size of vessels within a large image volume (Dang et al., 2022), which can easily be merged with the background during adaptation (Tsai et al., 2018). Second, the domain gap between existing modalities can vary widely.

The differences between the source and target data, referred to as domain-specific properties, can significantly impact the overall image appearance, encompassing various volume-related properties (e.g., spatial resolution, pixel spacing, image intensity range, contrast, or noise level). Moreover, differences can also manifest at a more localized level, impacting specific objects of interest: when examining vessel-related details in a brain scan, domain-specific variations can affect the vessels' intensities (e.g., vessels may be dark or bright), textures (e.g., vessels may have smooth or irregular surfaces), locations (e.g., vessels may

be central or peripheral), shapes (e.g., vessels may be thick or thin), and densities (e.g., vessels may be more or less numerous).

Figure 1 illustrates the visual disparity between a magnetic resonance angiography (MRA), a computed tomography angiography (CTA), and a magnetic resonance venography (MRV). This disparity can vary between different modalities. For example, arteries in MRA and CTA mainly differ in intensity distribution, as in the former, they stand out due to their high-intensity values, while in the latter, they blend with extracerebral tissue, making them harder to distinguish. Instead, the MRA-to-MRV domain gap also includes dissimilarities in the locations, shapes, and densities of the cerebral vasculature: although there is a correlation between the morphology of arteries and veins, the former are less numerous, occupy deeper positions within the brain tissue, and generally have larger sizes.

The larger the dissimilarities between source and target domains are, the more challenging it becomes to establish an image translation between the domains that facilitates effective segmentation. Indeed, translations from the target domain to the source cannot be performed fully, i.e., adapting all domain-specific properties to mimic the appearance of source images, as this would also involve vessel-related properties such as shapes, positions, and densities, which must be kept unchanged not to affect the final segmentation. To perform translations in a label-preserving way, i.e., generating hybrids between the source and the target domains, which only modify the necessary domain-specific properties to improve segmentation, we believe that additional disentangling mechanisms are required. To the best of our knowledge, these mechanisms are currently lacking,

which would facilitate adapting vessels from diverse origins to a standard labeled source domain for segmentation.

In this work, we aim to develop a 3D brain vessel segmentation tool that can be applied out-of-the-box across any new target domain, avoiding domain-specific model design, which often demands significant expertise and is highly dependent on dataset properties (Isensee et al., 2021). To this end, we build on image-to-image translation and semantic segmentation techniques to formulate a semi-supervised domain adaptation framework that learns a *disentangled representation* of image properties, allowing us to handle them independently. The disentangled representation enables us to introduce a novel mechanism of *label preservation*. This mechanism identifies and translates only a subset of all the domain-specific features, discarding the ones that may compromise spatial information, such as vessel shapes and locations, which is crucial for correct segmentation. As a result, our method eliminates the need for ad-hoc pre-processing steps commonly employed to homogenize image appearance across modalities, a process known in the literature as data harmonization (Eshaghzadeh Torbati et al., 2021), thereby enhancing the flexibility of our model and simplifying domain adaptation across different domains. To assess model performance when enlarging the domain gap, we conduct evaluations in three increasingly complex scenarios: multi-center MRA, MRA-to-CTA, and MRA-to-MRV adaptation for vessel segmentation. Finally, we investigate the properties of the proposed framework through extensive ablation studies focusing on determining the optimal number of source and target annotations, assessing the efficacy of disentanglement, and testing different architectural choices that may impact the performance of our model.

While disentanglement techniques have been employed in broader computer vision contexts as noted by prior works such as DRANet (Lee et al., 2021) and DRL-STNet (Lin et al., 2024), our approach introduces novel contributions specifically tailored to cerebrovascular segmentation. Unlike previous methods that primarily focus on disentangling content and style for general anatomical structures, our framework uniquely separates vessel-specific properties (intensities, textures, shapes, locations, and densities) from volume-related properties. This fine-grained disentanglement is crucial for cerebrovascular structures due to their intricate morphology and the complexity of domain gaps between vessel imaging modalities. Our method tackles the challenging problem of bridging the gap between fundamentally different vascular structures (arteries and veins) with distinct anatomical characteristics, advancing beyond existing image-to-image translation methods like XA-Sim2Real (Zhang et al., 2024) that typically address appearance transfer between similar vessel types. By leveraging path length regularization in a StyleGAN-based

architecture (Karras et al., 2020), we create a semantically organized latent space where different directions correspond to controllable aspects of variation, enabling independent manipulation of vessel intensities and textures while preserving spatial arrangements and geometrical properties. Our approach is further distinguished by its two-phase training strategy that limits adversarial learning to only the initial phase, ensuring stable convergence when navigating extreme domain shifts. Additionally, our integrated label-synthesis branch enforces label-preserving translations focused on vessel geometries without requiring a separate segmentation module, reducing architectural design complexity. Our comprehensive evaluation across progressively wider domain gaps (multi-center MRA, MRA-to-CTA, and MRA-to-MRV) demonstrates the framework's effectiveness in scenarios that haven't been previously attempted in vessel segmentation domain adaptation.

This paper is an extension of preliminary work (Galati et al., 2023), significantly advancing our previous research through several key contributions.

- We have reformulated our framework to function as an out-of-the-box tool, eliminating the need for domain-specific design choices or ad-hoc data harmonization between source and target domains.

- We introduce a *label preservation mechanism* that ensures image-to-image translation preserves vessels shapes and locations when navigating wide domain gaps.

- We extend our evaluation framework to include two additional clinically relevant scenarios: multi-center MRA adaptation and MRA-to-CTA adaptation, alongside our previous MRA-to-MRV scenario, allowing systematic analysis across progressively wider domain gaps.

- We conduct new ablation studies investigating key aspects of our framework, including the impact of architectural choices, the minimum number of annotated samples required for effective adaptation, and the contribution of feature disentanglement to segmentation accuracy.

2. Related Work

2.1 Multi-modal brain vessel segmentation

The segmentation of the 3D cerebrovascular vessels has been widely explored in the literature (Chen et al., 2023), encompassing different modalities and vessel types (Wilson and Noble, 1999; Bériault et al., 2015; Meijs et al., 2017). Nonetheless, only a few works address multiple domains. In Passat et al. (2007), morphological operators simultaneously capture blood signals from paired time-of-flight MRA and T1-weighted MR sequences. In Zuluaga et al. (2015), a multi-scale tensor voting framework accounts for a voxel's

scale and vicinity in paired CTA and 3D phase-contrast MR images. Despite developing a unified artery segmentation algorithm across image modalities, both studies (Passat et al., 2007; Zuluaga et al., 2015) require modality-specific initialization and parameter tuning. Recently, Tetteh et al. (2020) introduced an angiography segmentation model using 2D orthogonal cross-hair filters and a novel loss function for class imbalance with false-positive rate correction. After pre-training on synthetic data, the model is fine-tuned to segment human MRA data and CTA microscopy scans of rat brains, requiring pixel-wise annotations of each imaging modality. Dang et al. (2022) propose a weak patch-based deep learning approach for artery and vein segmentation from two MR sequences. A common limitation of these two methods is that they require separate training (or fine-tuning) for each domain. Chen et al. (2018) try to bypass domain-specific manual annotation by leveraging a paired dataset of MRA-CTA scans to generate annotations via registration, thresholding, and size filtering. However, the method faces limitations arising from misalignment between arteries after registration and the difficulty of acquiring paired datasets.

2.2 Domain Adaptation

Unsupervised domain adaptation (UDA) has been applied to the segmentation of various organs, such as liver (Hong et al., 2022a), lung (Li et al., 2022), heart (Chen et al., 2019; Wu and Zhuang, 2021), abdominal structures (Hong et al., 2022b), and brain substructures (Billot et al., 2023). The adaptation from the source to the target distribution can occur at different levels: input-level (or image-level) (Yao et al., 2022a), feature-level (Wu et al., 2022), output-level (Billot et al., 2023), or as a combination of two of the previous categories (Chen et al., 2019). In recent years, unsupervised image-alignment methodologies have surged, driven by the advancements in neural style transfer (Gatys et al., 2015) and image-to-image translation (Zhu et al., 2017), allowing for the extraction and combination of image content and style. Many image-alignment approaches, however, involve intricate architectures with multiple components (Ning et al., 2021) and heavily depend on adversarial training (Chen et al., 2019). Due to these factors, their behavior is known to be often unstable and difficult to interpret. To better guide the learning process, researchers have recently been redirecting their attention from fully unsupervised to semi-supervised DA for medical segmentation (Liu et al., 2022; Gu et al., 2023), including limited target annotations into the training set.

In the specific context of DA for vessel segmentation, Peng et al. (2022) leverages two segmentation models, each tailored to particular ophthalmic imaging modalities, operating within a UDA learning module to enhance the ac-

curacy of 2D retinal vessel segmentation. Gu et al. (2023) introduce a semi-supervised DA method designed for 2D cross-anatomy segmentation of coronary arteries and retinal vessels, integrating domain-specific batch normalization and cross-domain contrastive learning into a self-ensembling mean-teacher framework. Despite achieving promising results, the former technique might not be well-suited for large domain gaps due to the substantial disparity between brain arteries and veins. At the same time, the latter might face challenges due to the three-dimensional intricacies of the cerebrovascular structure.

2.3 Domain Generalization

The main limitation of DA is the requirement for repeated training with each novel target domain. Domain generalization enables a good performance across a wide range of target domains without the need to retrain (Zhou et al., 2023). Among these, Lyu et al. (2022) propose a data augmentation strategy for retinal vessel, optic disc and optic cup, and lesion segmentation that uses a model-agnostic augmentation policy to generate novel domains and maximize their diversity. Hu et al. (2024) introduce a novel domain generalization method integrating a Hessian-based vector field and self-attention mechanism to enhance tubular shape feature representation alongside a unique data augmentation preserving vessel structures while altering image style. Alternatively, foundation models like SAM (Kirillov et al., 2023) and SEEM (Zou et al., 2023) have recently shown robust performance using vast training data and test-time prompts such as points, bounding boxes, masks, or text that guide the segmentation tasks. However, their practical use is hindered by the need for fine-tuning, interactive filtering of extraneous predictions, and high computational costs (Ma et al., 2024; Huang et al., 2024). Thus, they are practically hard to adapt when annotations are rare, and datasets are small, such as in biomedical applications. In contrast, in-context learning methods adapt to new tasks without additional training by incorporating task demonstrations as inputs. Among these, UniverSeg (Butoi et al., 2023) has shown promise in medical image segmentation tasks by prompting support sets of image-label pairs, outperforming few-shot baseline methods. Although UniverSeg considers retinal vessels and has demonstrated its capability to generalize to unseen anatomies, we argue that requiring UniverSeg to bridge large domain gaps, such as the one from retinal to brain vessels, without incorporating any adaptation mechanism might be highly demanding.

3. Method

Let $\mathcal{S}$ represent the source domain, and $\mathcal{T}$ represent a target domain. Our framework relies on three datasets:
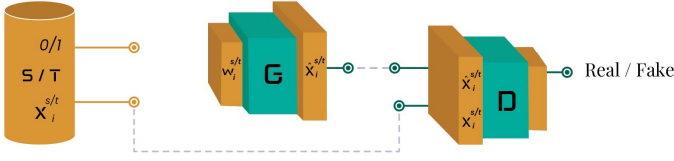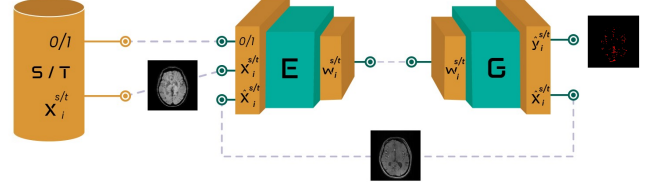
Phase 1:   **Phase 2:**

Figure 2: During the two-phase training algorithm, images x_i from domains $\mathcal{S}$ and $\mathcal{T}$ are input into our model consisting of the generator G, discriminator D, and encoder E. The training process is split into two distinct phases. In Phase 1 (left), G undergoes adversarial training with D to build a unified latent space that is both disentangled and semantically rich. In Phase 2 (right), the encoder E is trained for label-preserving image-to-image translation, while G is refined to generate segmentation masks $\hat{y}_i^t$ and $\hat{y}_i^s$.

$S = \{x_i^s, y_i^s\}_{i=1}^N$, a set of N labeled images from $\mathcal{S}$; $T_U = \{x_i^t\}_{i=1}^M$, a set of M unlabeled images from $\mathcal{T}$; and $T_L = \{x_i^{lt}, y_i^{lt}\}_{i=1}^m$, a target labeled dataset with $m \ll M$ annotated images, also from $\mathcal{T}$. We denote $T = T_U \cup \{x_i^{lt} \mid x_i^{lt} \in T_L\}_{i=1}^m$ as the set of all target samples, excluding the labels y_i^{lt}.

Our framework comprises a generator (G) and an encoder (E) accomplishing distinct tasks (Figure 2). The generator learns to generate realistic brain images, $\hat{x}$, by identifying the features from the source and target domains and representing them both within a unified and *disentangled* latent space $\mathcal{W}$. This representation allows our model to independently manipulate domain-specific features, enabling it to bridge broad domain gaps and compensate for the absence of data harmonization between the source and target at pre-processing. The encoder leverages the information in $\mathcal{W}$ to learn image-to-image translation in a *label-preserving* manner, i.e., focusing only on features that do not compromise spatial information. This is achieved using *cycle-consistency* and segmentation losses that enforce E to maintain the labels aligned in both domains.

Both G and E are trained in separate phases. This two-phase training strategy represents a crucial architectural decision that significantly impacts both training stability and segmentation performance. Splitting the training process into two distinct phases limits the adversarial training solely to the first phase, where an external discriminator D is incorporated to distinguish between real and fake images. By isolating adversarial training to Phase 1, we avoid the well-known instability issues associated with continuous adversarial learning throughout the entire adaptation process. The discriminator's role is deliberately confined to the initial phase, where it helps establish a disentangled latent space capable of representing both domains.

Once this foundation is established, removing the discriminator in Phase 2 allows the model to freely generate hybrid translations between domains without being penalized for producing outputs that do not strictly adhere to either domain's distribution. Excluding D from the second

phase, when the network learns image-to-image translation, prevents penalization of hybrid translations. This is particularly important when performing cross-domain translation between arteries and veins, which may require intermediate representations that would otherwise be rejected by a discriminator trained to distinguish between pure domain samples. Furthermore, this approach reduces architectural design complexity and prevents the adversarial component from interfering with the label-preserving mechanisms that operate during the image-to-image translation in Phase 2.

3.1 Feature Disentanglement

In Phase 1 (Figure 2 left), G is trained to establish an association between latent vectors w randomly sampled from $\mathcal{W}$ and the corresponding generated brain images, $\hat{x}$, which aim to resemble images from $\mathcal{S}$ or $\mathcal{T}$. To this end, we rely on adversarial learning with the aid of an external discriminator D. D acts as a binary classifier distinguishing between real and fake samples. In response, G aims to fool the discriminator by retrieving images that mimic the original ones from S and T. The parameters of G and D are optimized with the following loss function:

$$\mathcal{L}_{tot} = \mathcal{L}_{adv}(G, D) + \mathcal{L}_{R_1}(D) + \mathcal{L}_{pl}(G), \quad (1)$$

where $\mathcal{L}_{adv}$ is the non-saturating loss (Goodfellow et al., 2014), $\mathcal{L}_{R_1}$ is the R_1 regularization Mescheder et al. (2018), and $\mathcal{L}_{pl}$ is the path length regularization (Karras et al., 2020).

The regularization brought by $\mathcal{L}_{pl}$ transforms $\mathcal{W}$ into a disentangled latent space where different directions consistently correspond to individual, controllable aspects of variation in the generated images. This regularization encourages smooth transitions in the generated outputs when traversing the latent space, penalizing rapid changes by calculating $\mathcal{L}_{pl} = \mathbb{E}[(\|\nabla_w G(w)\|_2 - a)^2]$, where a is a running average of the gradient norm computed over previous iterations. As a result, the latent space becomes more uniform and semantically organized, facilitating the independent

manipulation of domain-specific features.

At the end of Phase 1, $\mathcal{W}$ can be queried to summarize the characteristics of both $\mathcal{S}$ and $\mathcal{T}$ in a shared and unwarped representation. Accordingly, this representation integrates the distinctive features from each domain, i.e., the domain-specific features.

In Phase 2 (Figure 2 right), E is trained. When fed with an image x_i^s from S, E learns to discover two corresponding latent representations, i.e., w_i^s and w_i^t, which are alternated by inputting an additional binary flag d, which functions as a domain selector that conditions the encoder's behavior: when $d = 0$, the encoder maps the input to the source domain's region in the latent space $\mathcal{W}$, and when $d = 1$, it maps to the target domain's region. This mechanism enables the model to perform both intra-domain reconstruction and cross-domain translation using a single encoder architecture, significantly reducing architectural design complexity compared to approaches that use separate encoders for each domain. The latent vectors guide G, which in this phase acts as a static decoder (with frozen parameters), to retrieve the source reconstruction $\hat{x}_i^s$, within the same domain.

$$\hat{x}_i^s = G(w_i^s) = G(E(x_i^s \mid d = 0)), \tag{2}$$

or the source-to-target translation $\hat{x}_i^t$ to the opposite domain,

$$\hat{x}_i^t = G(w_i^t) = G(E(x_i^s \mid d = 1)). \tag{3}$$

When learning w_i^s and w_i^t, E must encode the domain-specific features that recall the characteristics of either the source or target domain. *Disentanglement* ensures that all image properties, whether related to the whole volume (e.g., pixel spacing or image contrast) or specific to vessels (e.g., intensities, textures, shapes, locations, and densities of vessels), are individually represented within $\mathcal{W}$, facilitating E in establishing mappings between images at flexible semantic levels.

3.2 Label Preservation Mechanism

In the specific context of brain vessels, domain-specific and domain-invariant features may vary depending on the source and targeted modalities. For example, images from two different angiographic techniques (e.g., Time-of-Flight MRA and CTA) will share a set of domain-specific and invariant features, as they both focus on the brain arteries, but this will not hold if the target domain is a venography, since vein geometry and location differ largely from that of the arteries. In this scenario, conventional approaches tend to modify domain-specific features indiscriminately, including those related to vessel geometry and positioning. Altering these attributes compromises segmentation accuracy, as the resulting translations no longer align with the original annotations.

To address this limitation, we introduce a new label preservation mechanism, a crucial component of our framework, which ensures information integrity during domain adaptation. This mechanism enables the learning of domain-invariant and domain-specific features across different domains by enforcing consistency between segmentation masks before and after translation. More specifically, in Phase 2, we integrate image segmentation into our framework by expanding the generator with an additional label-synthesis branch G_{lsb} (Zhang et al., 2021) (Figure 2 right). This branch is designed to output semantic segmentation masks that align with the generated images: while G renders the source reconstruction $\hat{x}_i^s$ and the source-to-target translation $\hat{x}_i^t$, its label-synthesis branch predicts the associated segmentation maps $\hat{y}_i^s$ and $\hat{y}_i^t$. With this branch, we avoid using a separate segmentation module, thus decreasing architectural design complexity. It consists of three fully connected layers attached to the feature vectors of G, which are optimized in isolation while freezing all the other parameters inside the generator. To carry out this optimization, segmentation losses $\mathcal{L}_\text{S}$ are computed for both $\hat{y}_i^s$ and $\hat{y}_i^t$ based on the same reference annotation y_i^s. The segmentation loss is calculated as:

$$\mathcal{L}_\text{S} = \mathcal{L}_\text{dice} + \mathcal{L}_\text{ce}, \tag{4}$$

where $\mathcal{L}_\text{ce}$ is the cross-entropy loss, and $\mathcal{L}_\text{dice}$ is the Dice loss defined as:

$$\mathcal{L}_\text{dice} = -\frac{2 \cdot \text{TP} + \epsilon}{2 \cdot \text{TP} + \text{FP} + \text{FN} + \epsilon}, \tag{5}$$

where TP, FP, and FN represent true positives, false positives, and false negatives respectively, and $\epsilon = 10^{-5}$ is a small constant to prevent division by zero.

Requiring the model to output the same segmentation masks post-reconstruction and post-translation is crucial to guarantee that labels are preserved during both processes. This requirement backpropagates to E and ensures that w_i^s and w_i^t share the necessary domain-specific features to preserve the position and shapes of objects, particularly vessels (as it is our object of interest), which are consequently excluded from the translation process. For example, transforming pixel spacing (i.e., a domain-specific feature) in case it differs between $\mathcal{S}$ and $\mathcal{T}$ may increase or decrease the overall image scale, thus negatively affecting the segmentation. For this reason, E will avoid translating pixel spacing. Here, disentanglement proves beneficial, helping the model separate domain-specific features to automatically identify those contributing to improving the segmentation while discarding compromising ones. Consequently, the model can modify vessel intensities or textures while preserving their spatial arrangement and geometrical properties. The resulting outputs are thus hybrids between $\mathcal{S}$ and $\mathcal{T}$, intended to facilitate the segmentation process.

This automatic alignment of the two domains allows us to discard data harmonization during pre-processing, which is often a domain-specific and time-consuming task.

3.3 Cycle Consistency

Up to this point, we have described the forward pass carried out during Phase 2 by E and G from a source image x_i^s to its source reconstruction $\hat{x}_i^s$ and source-to-target translation $\hat{x}_i^t$, with corresponding predictions $\hat{y}_i^s$ and $\hat{y}_i^t$. This data flow remains equivalent when working with an input image x_i^t from the target domain, resulting in one target reconstruction and one target-to-source translation. However, segmentation losses are only computed when annotations are available, i.e., in T_L. In both domains, feeding the model with opposite values of d in succession corresponds to performing two complementary translations that neutralize each other's effects, as the first changes the input image's domain, while the second brings it back to its original one. This cyclical behavior can be exploited to enable the computation of cycle-consistency reconstruction losses alongside the intra-domain reconstruction losses, i.e., each translation is immediately followed by its inverse to enforce image fidelity. This approach reduces architectural complexity compared to traditional cycle-based methods, which typically necessitate two encoder-decoder pairs (Zhu et al., 2017). However, both the generator and encoder must be used twice during each cycle. Consequently, when the cycle is completed, the losses are propagated into G and E only once, taking into account the most recent pass. We compute both intra-domain and cycle-consistency reconstruction losses $\mathcal{L}_R$ as:

$$\mathcal{L}_R = \mathcal{L}_{MSE} + \mathcal{L}_{LPIPS}, \tag{6}$$

where $\mathcal{L}_{MSE}$ is the mean squared error loss calculated between the input and the reconstructed image, and $\mathcal{L}_{LPIPS}$ is the perceptual similarity metric that measures differences in feature space as defined in Zhang et al. (2018).

Figure 3 illustrates the steps involved in Phase 2. Moreover, Algorithm 1 provides a formal description of the complete procedure, complementing the visual representation in Figures 2 and 3.

3.4 Inference

Given a new image x_{new}^t, the model generates its reconstruction in $\mathcal{T}$, i.e., $\hat{x}_{new}^t$, and its translation in $\mathcal{S}$, i.e., $\hat{x}_{new}^s$. Simultaneously, G_{lsb} retrieves the segmentation masks $\hat{y}_{new}^t$ and $\hat{y}_{new}^s$, corresponding respectively to $\hat{x}_{new}^t$ and $\hat{x}_{new}^s$. Since both predictions contain valuable information about vessel segmentation, the final segmentation mask is obtained by averaging $\hat{y}_{new}^t$ and $\hat{y}_{new}^s$ before the last argmax operation. Notably, the model only performs reconstruction

---

Algorithm 1: Two-Phase Training for Multi-Domain Vessel Segmentation

Input: Source dataset $S = \{(x_i^s, y_i^s)\}_{i=1}^N$, target datasets $T_U = \{x_i^t\}_{i=1}^M$, $T_L = \{(x_i^{lt}, y_i^{lt})\}_{i=1}^m$

Output: Trained generator G and encoder E for vessel segmentation

Phase 1: Disentangled latent space learning;
Initialize generator G and discriminator D;
Train G and D adversarially for 250,000 iterations using $\mathcal{L}_{tot} = \mathcal{L}_{adv}(G, D) + \mathcal{L}_{R_1}(D) + \mathcal{L}_{pl}(G)$;

Phase 2: Label-preserving image-to-image translation;
Initialize encoder E and freeze parameters of G except for label-synthesis branch (G_{lsb});

for *iteration = 1 to 20000* **do**
  Sample batch from $S \cup T_U \cup T_L$;
  for *each source image $x_i^s \in S$* **do**
    // Source reconstruction ($d = 0$) and source-to-target translation ($d = 1$);
    $w_i^s \leftarrow E(x_i^s|d = 0)$, $w_i^t \leftarrow E(x_i^s|d = 1)$;
    $\hat{x}_i^s \leftarrow G(w_i^s)$, $\hat{y}_i^s \leftarrow G_{lsb}(w_i^s)$;
    $\hat{x}_i^t \leftarrow G(w_i^t)$, $\hat{y}_i^t \leftarrow G_{lsb}(w_i^t)$;
    Compute $\mathcal{L}_R(x_i^s, \hat{x}_i^s)$ and $\mathcal{L}_S(y_i^s, \hat{y}_i^s, \hat{y}_i^t)$;
  end
  for *each target image $x_i^t \in T_U \cup T_L$* **do**
    // Target reconstruction ($d = 1$) and target-to-source translation ($d = 0$);
    $w_i^t \leftarrow E(x_i^t|d = 1)$, $w_i^s \leftarrow E(x_i^t|d = 0)$;
    $\hat{x}_i^t \leftarrow G(w_i^t)$, $\hat{y}_i^t \leftarrow G_{lsb}(w_i^t)$;
    $\hat{x}_i^s \leftarrow G(w_i^s)$, $\hat{y}_i^s \leftarrow G_{lsb}(w_i^s)$;
    Compute $\mathcal{L}_R(x_i^t, \hat{x}_i^t)$;
    if $x_i^t \in T_L$ **then**
      Compute $\mathcal{L}_S(y_i^{lt}, \hat{y}_i^t, \hat{y}_i^s)$;
    end
  end
  Update E by gradient descent on $\mathcal{L}_R$ and $\mathcal{L}_S$;
  Update G_{lsb} by gradient descent on $\mathcal{L}_S$;
end
return *trained G and E*

---

when used with a source image x_{new}^s. The translation capability, which involves generating $\hat{x}_{new}^t$ and the corresponding $\hat{y}_{new}^t$, is not used, since our main goal is the segmentation of the target domain.

4. Experiments and Results

4.1 Experimental Setup

In this section, we detail our experimental methodology for evaluating vessel segmentation across diverse medical
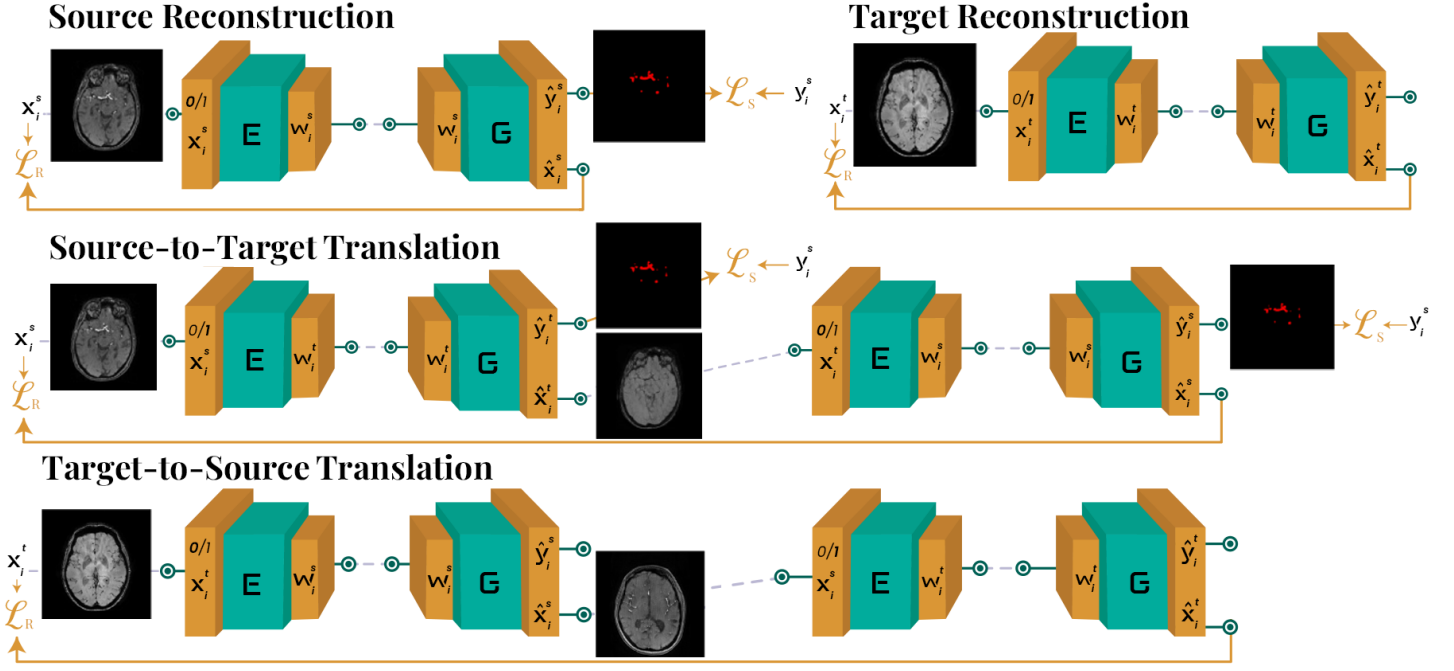
Figure 3: In Phase 2 of our training algorithm, we perform both source and target reconstructions (first row, source domain on the left and target domain on the right) and source-to-target and target-to-source translations (second and third rows). The backpropagation of $\mathcal{L}_R$ exclusively updates the weights of E, while $\mathcal{L}_S$ influences both E and G.

imaging modalities. We first describe the datasets used in our experiments, including their characteristics and the preprocessing steps employed. We then provide implementation details of our framework, highlighting the network architectures and training procedures. Finally, we outline the evaluation metrics used.

4.1.1 Datasets

Our experiments use the following datasets.

OASIS-3 (LaMontagne et al., 2019). We randomly select 49 time-of-flight (TOF) MRA volumes. These volumes have a median grid size of $576 \times 768 \times 232$ voxels and a median voxel size of $0.30 \times 0.30 \times 0.60$ mm. Our selection encompasses 27 cognitively normal subjects and 10 patients at different stages of cognitive decline, all adults ranging in age from 42 to 95 years.

IXI[1]. We sample 50 TOF MRA volumes, with a median grid size of $359 \times 481 \times 100$ voxels and a median voxel size of $0.47 \times 0.47 \times 0.80$ mm. All images were acquired from healthy subjects spanning an age range of 20 to 86 years.

TopCoW (Yang et al., 2024). We use the 40 CTA volumes within the first release of the dataset. The volumes exhibit a median grid size of $290 \times 366 \times 211$ voxels and a median voxel size of $0.46 \times 0.46 \times 0.70$ mm. The patients within this cohort were all in the process of recovering from disorders related to strokes.

Susceptibility-weighted images (SWI). We use a private

dataset consisting of 28 SWI venographies from retrospective studies previously conducted at UCL Queen Square Institute of Neurology, Queen Square MS Centre, University College London. The images have a median grid size of $480 \times 480 \times 288$ voxels and a median voxel size of $0.50 \times 0.50 \times 0.50$ mm and include adult subjects showing no visible lesions on SWI.

For IXI, we use the vessel annotations provided in Falcetta et al. (2025). For OASIS-3 and SWI, all image volumes were annotated by two experts (RC, MAZ) to obtain vessel masks. For TopCoW, we used the masks included in the dataset, which include annotations only of the vessels constituting the circle of Willis (CoW). Brain masks were obtained using SynthStrip (Hoopes et al., 2022). For TopCoW, we generated brain annotations through a registration and resampling procedure initiated from the pairwise MRA.

The datasets undergo separate pre-processing without any inter-domain harmonization. First, all volumes are resampled using bicubic interpolation to fix a uniform spacing, calculated as the dataset's median value, with minor increments made if the images do not fit into a volume of 512^3 voxels. Next, each volume is rescaled based on its mean and standard deviation, and then clipped between the 0.1 and 99.9 percentiles and normalized in the range $[-1, +1]$. The segmentation masks undergo one-hot encoding, resulting in a three-dimensional label: one dimension for the brain, one for the vessels, and an additional one for the background.

---

1. https://brain-development.org/ixi-dataset

Table 1: Source domain performance on OASIS-3

	Dice	Precision	Recall	clDice
Vessels	73.7 ± 2.8	66.9 ± 4.8	82.5 ± 3.7	76.9 ± 5.2

4.1.2 Implementation Details

Our framework is implemented in PyTorch 1.9.1. Phase 1 and Phase 2 use batches of four images each, and run for 250k and 20k iterations, respectively. In addition, a preliminary phase of 15k iterations is conducted before Phase 2, to pre-train the model using only source data. After training, the models with the best validation performance on $\mathcal{S}$ and $\mathcal{T}$ are selected for the final evaluation. The generator G and discriminator D are based on StyleGAN2 (Karras et al., 2020), while the label-synthesis branch is adapted from DatasetGAN (Zhang et al., 2021). As in (Richardson et al., 2021), the encoder E maps input images into the extended latent space $\mathcal{W}+$ of StyleGAN2, using a ResNet backbone inspired by Yao et al. (2022b). Building upon this backbone, multiple outputs are branched out: one for latent code prediction and the other for feature tensor prediction. These branches are then connected to G through a dynamic skip connection module Yang et al. (2022), which filters the residual information to establish fine-level content correspondences. Our framework is configured for multi-class segmentation to extract both brain tissue and vessels simultaneously. The segmentation masks include three classes: background, brain, and vessels. All code and experiments can be accessed on github.com/i-vesseg/MultiVesSeg.

4.1.3 Evaluation setup

We evaluate models based on their segmentation performance on the target datasets using test splits. We assess performance using the Dice coefficient (Dice), the centerlineDice (clDice) (Shit et al., 2021), Average Symmetric Surface Distance (ASSD) (Yeghiazaryan and Voiculescu, 2018), precision, and recall.

4.2 Ablation Studies

We study how the performance of our model is impacted by the number of available annotated images in both the target and source domains, as well as by various architectural choices. Given the substantial domain gap between angiographies and venographies, which depict two different vessel types, we utilize TOF MRA images from OASIS-3 as the source domain $\mathcal{S}$ and SWI images as the target domain $\mathcal{T}$ to analyze the behavior of our model in this particularly complex scenario.

4.2.1 Intra-domain Performance

We first assess the performance of our method in intra-domain vessel segmentation. In Phase 1, we include a source dataset (S) of $N = 35$ source volumes and a target dataset (T) of $|T| = M + m = 20$ target volumes into our training set. As this phase is entirely unsupervised, the division between the unlabeled and labeled target sets T_U and T_L does not have any impact. Subsequently, we pre-train the encoder E and the segmentation branch of G using only the source data (left half of the first row in Figure 3), ignoring source-to-target translation. For evaluation, we split equally the remaining 14 TOF MRAs between validation and testing, following a 70-15-15 ratio. The results on the testing set are presented in Table 1, demonstrating that our method's performance is comparable to state-of-the-art approaches for brain artery segmentation (Livne et al., 2019; Dang et al., 2022).

4.2.2 Impact of Target Annotations

We investigate the model's sensitivity to the number of annotated target images. Using a fixed number of source images $(N = 35)$, we gradually increase the number of annotated target images into the training set (T_L). We begin with $m = 0$ and progress to $m = 1$ and $m = 3$ midpoint slices, extracted from three distinct volumes. This sequence concludes with the inclusion of the full three volumes into T_L. The remaining volumes are used without annotations $(M = 17)$. Four images are set aside for validation, and another four are kept for testing.

Figure 4 (left) reports vessel segmentation performance. As expected, the performance improves as the number of available annotated samples increases. In particular, there is a performance boost observed during the transition from $m = 0$ to $m = 1$ slice, marking the shift from an unsupervised DA scenario to a semi-supervised one. However, as the number of labeled slices increases from $m = 3$ to cover three whole volumes $(m = 831$ slices in total), the extent of this performance improvement gradually diminishes, suggesting a trend toward saturation. This indicates that while the model benefits from additional annotated target images, it already exhibits good behavior when only a few target labels are available. Based on this finding and the trade-off between annotation effort and performance, we set $m = 3$ for all subsequent experiments in this study.

4.2.3 Impact of Source Annotations

We investigate the scenario where the number of available source images varies $(N = [0, 10, 20, 35])$ while the number of annotated target images is fixed $(m = 3$ slices). Source and target validation and testing sets are the same as in Sections 4.2.1 and 4.2.2. Figure 4 (right) summarizes the obtained results. We start considering a few-shot segmen-
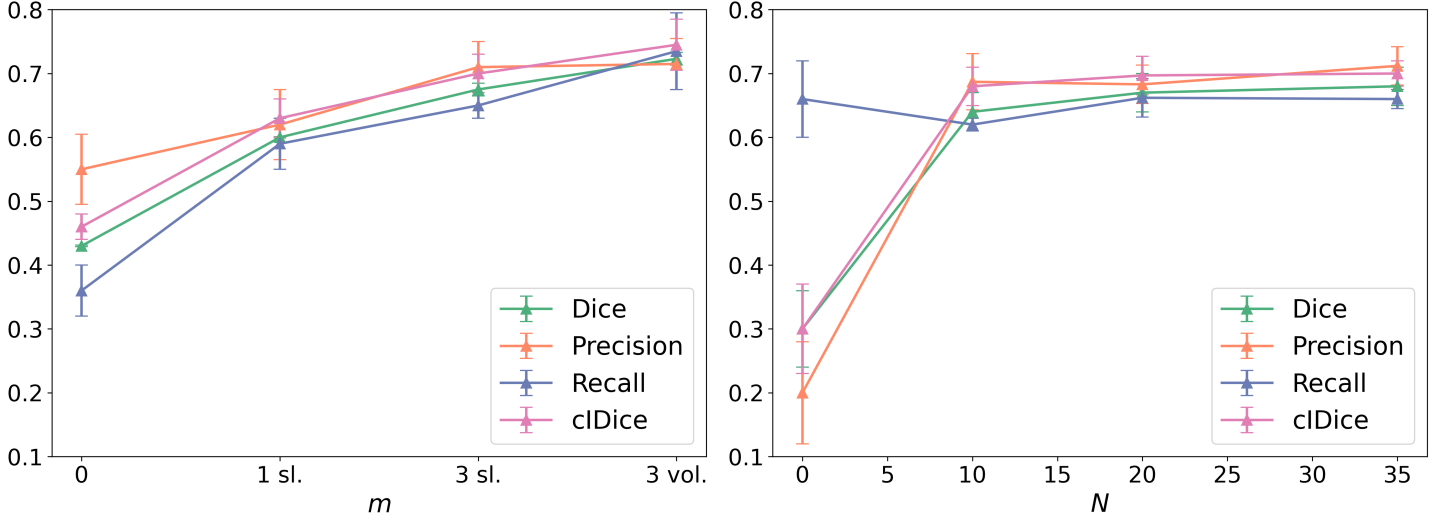
Figure 4: Vessel segmentation performance with varying target annotations m (left) and source annotations N (right). Vertical error bars represent the standard deviation across the testing set.

tation scenario, where minimal annotations are employed to train segmentation in $\mathcal{T}$, and there is no contribution from $\mathcal{S}$ (i.e., $N = 0$). In this case, our model performs exclusively target reconstruction (right half of the first row in Figure 3), calculating the segmentation loss only for T_L. A sharp performance increase is observed when passing from $N = 0$ to $N = 10$ source volumes, i.e. when we activate source reconstruction and inter-domain translations. After, there is a modest increase of only 3.4% in Dice when moving from $N = 10$ to $N = 35$. Despite the improvements becoming more gradual, overall, the contribution from the labeled source proves to be crucial for achieving satisfactory results. To maximize segmentation performance and make full use of our available data resources, we employed all $N = 35$ source volumes for our subsequent experiments.

4.2.4 Architecture Elements

We perform an ablation study to assess how the different elements in our architecture affect the segmentation accuracy. Specifically, we examine the effects of the following features, which we notice to have the most significant impact on the results: residual connections (Res.), to enhance information flow between E and G; domain-specific batch normalization (DSBN), to normalize feature maps separately for the two domains; balanced data sampling (BDS), to ensure that each batch contains two samples from S, one from T_L and one from T_U; and intensity inversion (Inv.), to flip the intensity values of the input images, thus mitigating the disparity between domains capturing vessels in dark and bright appearances respectively.

Table 2 displays the configurations obtained by deactivating each assessed component. Residual connections appear to exert the most influence on the model's function-

Table 2: Architectural Choices

Res.	DSBN	BDS	Inv.	Dice	clDice
✓	✓	✓	✓	**72.2 ± 2.5**	**75.4 ± 3.3**
✗	✓	✓	✓	14.4 ± 3.4	17.0 ± 3.4
✓	✗	✓	✓	69.3 ± 2.8	73.7 ± 3.2
✓	✓	✗	✓	71.2 ± 2.4	74.4 ± 3.3
✓	✓	✓	✗	71.8 ± 3.0	74.3 ± 3.3

ing, causing a substantial drop in Dice from 72.2% to 14.4%. Residual connections emerge as indispensable components, serving to preserve spatial information during reconstruction and facilitating the network's manipulation of low-level semantic attributes (Yao et al., 2022b). Domain-specific batch normalization causes a drop of 2.9%; balanced data sampling 1.1% and intensity inversion brings a negligible effect of 0.4% in the Dice. Notably, intensity inversion is specific to MRA-to-MRV, thereby falling within the definition of data harmonization between the source and target domains. Proving that this inversion does not impact the performance reinforces the hypothesis that our method does not necessitate domain-specific pre-processing to address the domain gap. However, this is true only in the semi-supervised setting: after conducting an additional experiment with $m = 0$, we notice a significant Dice score drop from 40.9% to 0.1% when intensity inversion is not used. This underlines the need for some form of guidance in establishing connections between vessels across TOF MRA and SWI modalities. This guidance could come in the form of labeled examples or intensity harmonization, but it represents an essential requirement for the correct functioning of our model.

4.3 Comparison with State-of-the-Art Methods

We compare the best results obtained through our ablation studies with a 2D U-Net (Ronneberger et al., 2015) as a baseline and seven DA state-of-the-art methods. For the baseline, we implement a fully-supervised approach with limited target annotations, training only on the few target samples in T_L (m=3 slices to maintain methodological consistency). This implementation[2] uses a single input channel and three output classes (background, brain, vessels). The architecture consists of four encoding/decoding stages with feature maps ranging from 64 to 512. We optimize using Adam (lr=2e-4) with a combined BCE and Dice loss function, processing 512×512 2D slices and employing validation-based early stopping. The seven DA SOTA methods are: CycleGAN (Zhu et al., 2017), Synergistic Image and Feature Adaptation (SIFA) (Chen et al., 2019), SynthSeg (Billot et al., 2023), UniverSeg (Butoi et al., 2023), Automatic Augmentation for Domain Generalization (AADG) (Lyu et al., 2022), DCDA (Peng et al., 2022), and Contrastive Semi-supervised learning for Cross Anatomy Domain Adaptation (CS-CADA) (Gu et al., 2023). In particular:

1) CycleGAN is a well-established method to perform unpaired image-to-image translation on natural images. After translating data from $\mathcal{T}$, we feed the results into a 2D U-Net previously trained on S, as CycleGAN does not provide segmentation.

2) SIFA is an UDA technique based on image-to-image translation for multi-class medical segmentation, therefore trained using both S and T, without utilizing any target label y_i^{lt};

3) SynthSeg is a UDA 3D output-level alignment method based on synthetic data generation for brain synthesis and segmentation. It is trained with masks y_i^s from S, determining the best checkpoint based on the target performance;

4) UniverSeg is a foundation model focusing on unseen medical segmentation tasks without additional training;

5) AADG is a multi-source domain generalization framework based on data manipulation of retinal vessel images. To leverage training from multiple source domains, the network is trained using all datasets except the target;

6 and 7) DCDA and CS-CADA are, respectively, unsupervised and semi-supervised DA methods designed for retinal vessel segmentation and 2D coronary artery segmentation. We train these methods using both S and T, including labels from T_L for CS-CADA.

Using OASIS-3 as a source domain, we conduct experiments in three distinct domain adaptation scenarios of increasing difficulty to ensure a broader perspective, comparing our model's performance in adapting to the following shifts:

---

2. https://github.com/wolny/pytorch-3dunet

1) Multi-center (MC) MRA, where MRAs are used as $\mathcal{S}$ and $\mathcal{T}$, but from different centers. Thirty-six unlabeled volumes (T_U) from IXI enter the training set; seven are kept for validation and seven for testing;

2) MRA-to-CTA, where the target domain is CTAs from TopCoW, including 28 volumes without annotations (T_U) for training, six for validation and six for testing; and

3) MRA-to-MRV, with SWIs used as $\mathcal{T}$, of which 20 volumes deprived of labels (T_U) are included in the training set, four in the validation set and four in the testing set.

In MC MRA and MRA-to-MRV, we extract three midpoint slices from T_U to form T_L. In MRA-to-CTA, we allocate three entire volumes for T_L since they only have CoW annotations. For the source dataset, 35 are allocated for training (S), while seven volumes are used for validation and testing.

For the sake of fairness, we evaluate both cross-modality vessel and brain segmentation since most of the methods (e.g., SIFA, SynthSeg, AADG, and UniverSeg) have been developed for segmenting large objects, such as the brain. To ensure consistent comparison, all methods were retrained for brain and vessel segmentation with two exceptions: UniverSeg was used as originally designed for zero-shot segmentation on unseen medical tasks without retraining, and SynthSeg utilized its original pre-trained model for brain tissue segmentation only, as it was specifically designed for this task. Notably, while the pre-trained SynthSeg model performs well with magnetic resonance images (both MRA and MRV), it fails to segment brain tissue in CTA images due to the fundamental differences in image contrast mechanisms. Table 3 summarizes the obtained results, and Figure 5 displays a visual comparison of the results across MC MRA, MRA-to-CTA, and MRA-to-MRV.

In general, most methods have a very good performance on the brain task, but there is a clear difficulty in segmenting the vessels. This becomes particularly visible in the MRA-to-MRV scenario: both the U-Net baseline, trained with full supervision on the reduced dataset T_L, and the considered state-of-the-art methods in domain adaptation and generalization struggle to segment veins. The performance degradation trend from MC MRA and MRA-to-CTA to MRA-to-MRV highlights the challenge posed by increasing domain gaps. UniverSeg fails to segment vessels across all scenarios but demonstrates satisfactory brain segmentation performance despite not requiring additional training. CS-CADA, the only other SSDA model besides ours, provides poor results overall, likely because it originally relies on a larger annotated target set than T_L.

Notably, our proposed method achieves high performance in the target domain for both brain and vessel segmentation. In particular, it bridges even the widest domain gaps, successfully segmenting veins using only three annotated target slices and leveraging information from the

Table 3: Segmentation performance of different methods in the target domains. UniverSeg (both tasks) and SynthSeg (brain only) use the original pre-trained models.

MC MRA

		U-Net	CycleGAN	SIFA	SynthSeg	UniverSeg*	AADG	DCDA	CS-CADA	Ours
Dice	Vessels	65.6 ± 4.4	30.5 ± 3.3	53.7 ± 1.9	41.7 ± 5.4	7.3 ± 2.6	45.3 ± 2.9	12.0 ± 2.8	43.3 ± 5.0	**69.9 ± 2.3**
	Brain	95.1 ± 1.6	81.7 ± 2.9	89.4 ± 4.0	93.6 ± 3.3*	95.4 ± 1.2	97.7 ± 0.3	90.5 ± 1.2	69.3 ± 6.3	**97.8 ± 0.2**
Precision	Vessels	62.6 ± 8.2	35.3 ± 5.1	56.5 ± 4.1	61.5 ± 7.4	5.4 ± 3.1	**78.2 ± 4.1**	23.1 ± 4.8	54.7 ± 15.1	70.0 ± 4.0
	Brain	92.5 ± 3.0	86.8 ± 1.7	89.6 ± 1.0	**99.2 ± 0.4***	93.9 ± 2.3	98.8 ± 0.4	87.3 ± 1.0	81.7 ± 7.1	98.0 ± 0.5
Recall	Vessels	**70.3 ± 6.7**	27.4 ± 4.4	51.7 ± 3.6	31.7 ± 4.9	14.7 ± 4.9	32.0 ± 3.1	8.2 ± 2.1	37.7 ± 4.7	70.2 ± 4.7
	Brain	**98.0 ± 0.5**	77.4 ± 5.3	89.5 ± 7.6	88.9 ± 5.9*	97.0 ± 0.5	96.6 ± 0.7	94.1 ± 2.9	60.5 ± 6.8	97.7 ± 0.5
clDice	Vessels	68.7 ± 6.5	25.4 ± 3.0	51.5 ± 3.0	41.0 ± 6.4	8.1 ± 2.1	35.8 ± 2.6	8.4 ± 2.3	40.2 ± 5.4	**76.8 ± 2.9**
ASSD	Vessels	3.39 ± 1.48	10.18 ± 3.36	3.38 ± 1.44	4.17 ± 0.86	15.59 ± 1.91	4.49 ± 1.42	11.59 ± 2.95	6.10 ± 1.90	**1.55 ± 0.57**

MRA-to-CTA

		U-Net	CycleGAN	SIFA	SynthSeg	UniverSeg*	AADG	DCDA	CS-CADA	Ours
Dice	Vessels	70.5 ± 3.0	33.1 ± 4.3	60.9 ± 2.9	55.1 ± 23.6	11.5 ± 9.0	5.8 ± 6.0	0.0 ± 0.0	0.0 ± 0.0	**74.5 ± 4.2**
	Brain	95.8 ± 1.7	93.1 ± 1.6	94.4 ± 1.9	5.1 ± 7.0*	95.9 ± 0.8	94.8 ± 1.6	91.2 ± 3.5	85.6 ± 10.4	**96.6 ± 1.1**
Precision	Vessels	72.7 ± 13.5	27.1 ± 5.2	63.5 ± 8.3	52.2 ± 15.6	38.9 ± 19.9	22.5 ± 22.3	0.0 ± 0.0	0.0 ± 0.0	**73.3 ± 13.3**
	Brain	94.2 ± 3.1	94.9 ± 1.4	94.1 ± 3.3	49.1 ± 49.1*	93.5 ± 1.5	91.5 ± 3.4	95.1 ± 1.1	95.2 ± 0.9	**96.2 ± 1.8**
Recall	Vessels	72.1 ± 10.5	43.1 ± 2.5	59.1 ± 2.6	63.2 ± 27.7	6.9 ± 5.7	3.4 ± 3.5	0.0 ± 0.0	0.0 ± 0.0	**78.8 ± 8.5**
	Brain	97.5 ± 0.4	91.4 ± 2.2	94.8 ± 2.0	2.8 ± 3.8*	98.5 ± 1.2	**98.5 ± 1.5**	87.8 ± 6.3	79.2 ± 15.2	97.1 ± 1.1
clDice	Vessels	72.5 ± 6.7	39.7 ± 5.2	68.9 ± 6.3	63.5 ± 29.0	18.0 ± 9.0	nan ± nan	nan ± nan	nan ± nan	**78.0 ± 8.7**
ASSD	Vessels	2.40 ± 0.99	3.64 ± 0.53	2.12 ± 0.56	4.01 ± 5.82	12.24 ± 3.10	13.92 ± 6.99	12.97 ± 2.11	nan ± nan	**1.46 ± 0.49**

MRA-to-MRV

		U-Net	CycleGAN	SIFA	SynthSeg	UniverSeg*	AADG	DCDA	CS-CADA	Ours
Dice	Vessels	29.1 ± 4.9	5.1 ± 0.3	0.8 ± 0.5	10.9 ± 1.9	3.6 ± 1.1	2.0 ± 1.2	0.0 ± 0.0	0.4 ± 0.2	**67.5 ± 1.7**
	Brain	83.0 ± 2.5	75.0 ± 0.8	91.4 ± 1.9	97.4 ± 0.1*	83.5 ± 2.5	96.7 ± 0.5	75.6 ± 1.1	25.7 ± 2.5	**97.8 ± 0.2**
Precision	Vessels	18.2 ± 4.0	11.5 ± 0.8	2.6 ± 1.1	51.4 ± 5.7	6.5 ± 2.4	1.3 ± 0.6	2.4 ± 2.1	0.8 ± 0.4	**71.1 ± 4.4**
	Brain	71.3 ± 3.7	62.6 ± 0.7	97.8 ± 0.2	96.9 ± 0.5*	72.8 ± 3.8	97.4 ± 0.3	62.0 ± 1.6	32.5 ± 3.8	**97.8 ± 0.4**
Recall	Vessels	**76.0 ± 5.5**	3.3 ± 0.3	0.5 ± 0.3	6.1 ± 1.2	2.5 ± 0.7	6.3 ± 4.0	0.0 ± 0.0	0.3 ± 0.1	64.4 ± 2.1
	Brain	**99.5 ± 0.2**	93.6 ± 1.2	85.9 ± 3.5	97.9 ± 0.5*	98.1 ± 0.2	96.0 ± 0.9	97.0 ± 0.5	21.3 ± 1.8	97.8 ± 0.5
clDice	Vessels	33.5 ± 5.9	4.1 ± 0.3	0.6 ± 0.4	10.4 ± 1.9	2.7 ± 0.9	1.8 ± 1.1	nan ± nan	0.4 ± 0.2	**69.9 ± 2.7**
ASSD	Vessels	9.30 ± 0.53	10.54 ± 0.29	15.04 ± 6.98	14.41 ± 1.41	20.32 ± 2.52	10.46 ± 6.46	60.66 ± 5.62	18.73 ± 0.96	**1.42 ± 0.20**

associated arteries in the source modality. This demonstrates the model's capability to cope with the differences between arteries and veins, which are not limited to their low-level attributes, such as intensities and textures, but also encompass higher-level aspects, including their location and shape.

Our better performance in cross-modality vessel segmentation, especially in comparison to other techniques following similar translation paradigms, can be explained by the key elements introduced in our framework. Compared to CycleGAN, which requires separate networks for bidirectional translations and performs image-to-image translation independent of segmentation, our approach integrates translation and segmentation through our label-synthesis branch, directly enforcing label preservation. Unlike SIFA, which aligns feature distributions without distinguishing between vessel-related and volume-related features, our disentanglement mechanism specifically isolates vessel properties that can be safely translated without compromising spatial information. While CS-CADA introduces contrastive learning for domain adaptation, it cannot selectively translate domain-specific features that preserve vessel positions and

morphology. Similarly, DCDA's disentangling approach focuses primarily on style transfer for retinal vessels, which is insufficient for handling the complex domain shifts between cerebrovascular structures with different anatomical characteristics. Finally, our two-phase training strategy isolates adversarial learning to only the initial phase, providing more stable convergence compared to methods like CycleGAN and SIFA which rely on continuous adversarial training throughout the adaptation process. These architectural choices collectively enable our framework to effectively bridge the challenging domain gaps in cerebrovascular segmentation while maintaining accurate vessel delineation.

Figure 6 provides 3D visualizations of our segmentation results across the three domain adaptation scenarios, demonstrating the preservation of vessel topology and spatial continuity. The 3D renderings show that our method maintains the complex branching patterns and connectivity of the cerebrovascular tree across all tested scenarios. Notably, even in the most challenging MRA-to-MRV scenario, where the method must bridge the gap between arterial and venous structures with distinct anatomical characteristics, these visualizations reveal successful preservation of the
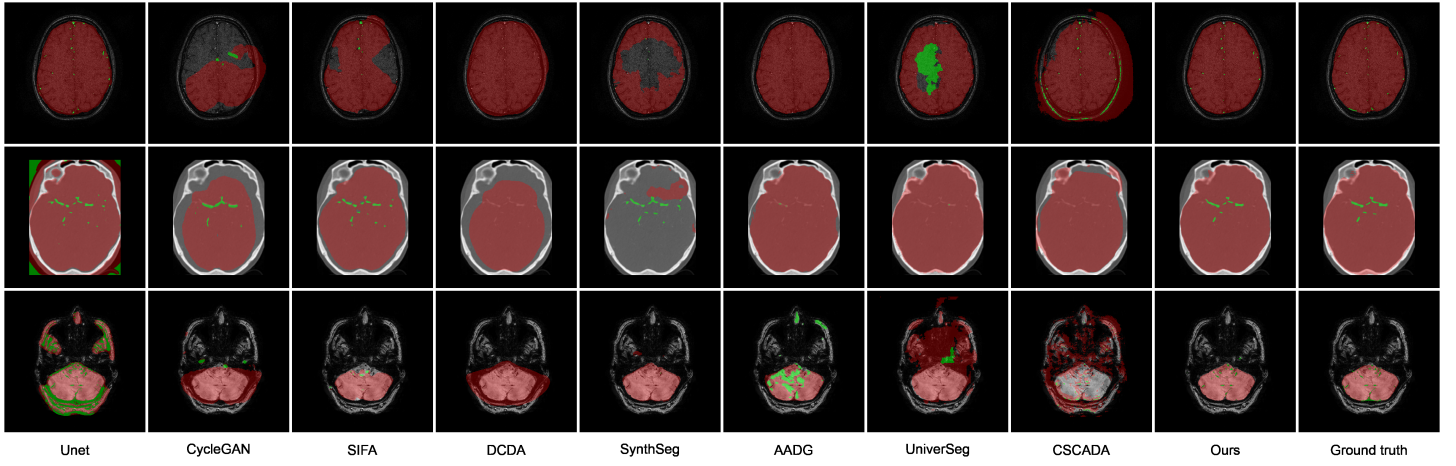
Figure 5: Comparison of the segmentation results for brain and vessels in the target MRA, CTA, and SWI images using different methods. Red indicates brain masks, while green represents vessels. The rows display slices at varying levels: top, middle, and bottom.
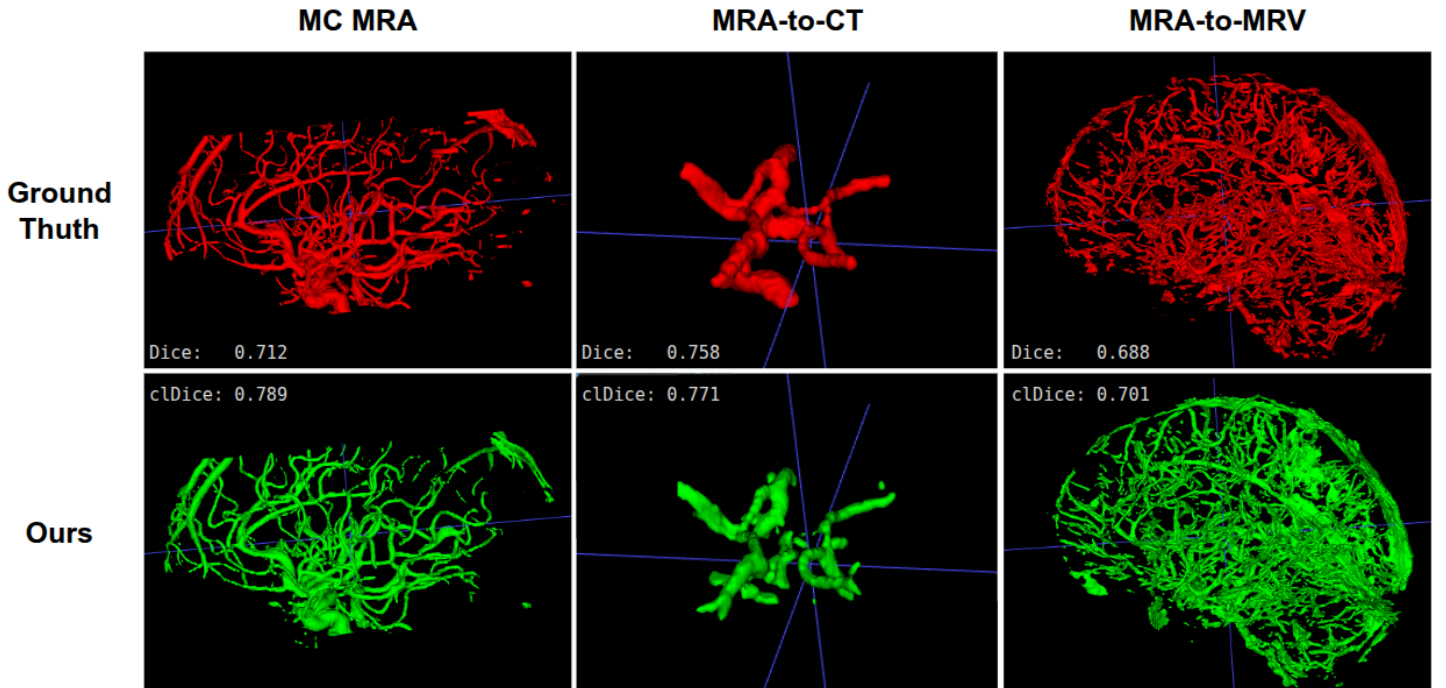

Figure 6: 3D visualization of vessel segmentation results across the three domain adaptation scenarios. Ground truth (red) and our method's predictions (green) are shown as 3D renderings with corresponding Dice and clDice scores.

overall vascular architecture with high fidelity to the ground truth topology.

4.4 Qualitative Analysis

Adopting the path length regularization (Karras et al., 2020) has an important role in the domain adaptation process as it allows the disentangling of the latent space $\mathcal{W}$, enabling inter-domain translations that can handle independently volume-related image properties, such as overall spatial information and appearance, and vessel-related properties, such as their intensities, textures, shapes, locations, and densities. This allows preserving the target content while mimicking the appearance of a source image as it is better recognized by the segmentation branch. Keeping vessel position and shape unchanged, despite these being domain-specific features, is a key property to guarantee correct segmentation. By relying on the aforementioned capabilities, we have gathered evidence of the ability to separate the vessel-related features by visually inspecting the target-to-source translations generated by our model compared to other image-level alignment methods.

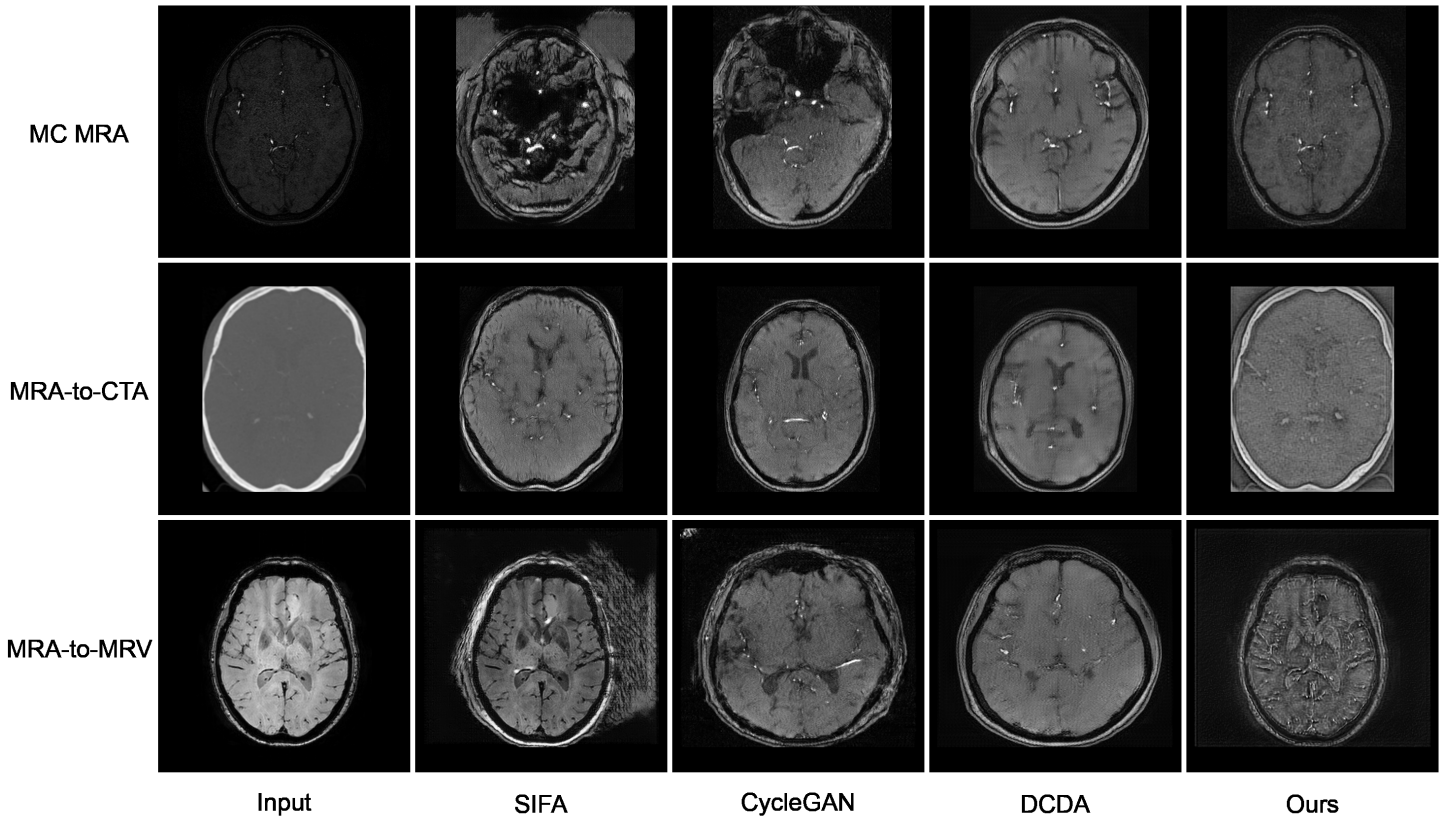In Figure 7, we display three cases of translation: one

Figure 7: Target-to-source translations produced by the different image-level alignment methods.

from MC MRA (first row), another from MRA-to-CTA (second row), and a third from MRA-to-MRV (third row). These examples highlight how the different models act on the vessel-related properties. In particular, we identified three problematic behaviors that compromise the accuracy of the final segmentation results. These behaviors involve the translation of label-altering features due to their domain-specific nature. Firstly, vessels undergo displacement, resulting in changes to their position and size. This occurs while resizing the whole brain to align with the pixel spacing of the source domain. Specifically, CycleGAN and DCDA tend to translate all domain-specific features without distinction, including, in fact, the pixel spacing, and thus leading to spatial misalignment between the source and target domains. We believe this problem arises because the segmentation loss does not influence the prior translation enough, which is exactly the case in CycleGAN, where translation and segmentation are completely separate. Secondly, vessels are observed merging with the background and vanishing. This is noticeable as the number of bright vessels in the translations is never greater than the vessels in the target domain. The phenomenon is particularly evident in SWI images, where veins are generally more abundant than arteries in TOFs. The third issue arises in SIFA, which initially appears to better preserve the positions and shapes of the brain and vessels during translation, despite generating some shadow artifacts around the skull in MC MRA

and MRA-to-MRV translations. However, most veins from SWIs are left untransformed and do not resemble arteries after translation. Only a few veins, likely those aligning well with the typical artery arrangement, transform into bright vessels. We attribute this behavior to the network's inability to link arteries and veins during translation without some form of guidance.

These findings align with what was observed for Figure 5, where problematic vessels are either omitted from the final segmentation or displaced. Also, this reinforces the importance of enforcing label-preserving translations in our problem. Notably, our model uniquely transforms dark vessels from the input (SWI) into bright vessels without relocating them or reducing their number to replicate the typical arrangement of arteries in TOF MRA images. This ability to selectively translate only some domain-specific features, particularly those unrelated to vessel size and position, enables our approach to adapt veins and arteries and retrieve accurate segmentations. Lastly, we emphasize that achieving a hyper-realistic translation of target volumes is not the central focus of our model. We acknowledge that our translations may not appear entirely source-like but rather appear as hybrid representations. Indeed, the ability of the network to translate input images aims exclusively to serve the segmentation process, which is the primary objective of the proposed method.

4.5 Computational Costs

We evaluate the computational efficiency of our approach. All experiments were conducted on two NVIDIA GeForce RTX 2080 Ti GPUs, each with 12GB of memory. For our framework, Phase 1 processes at approximately 0.42 iterations per second, requiring about 166 hours ($\approx$ 7 days) to complete 250,000 iterations. Phase 2 runs at approximately 0.20 iterations per second, taking about 28 hours to complete 20,000 iterations. This results in a total training time of approximately 194 hours ($\approx$ 8 days).

The computational cost of our approach stems primarily from the comprehensive disentanglement process in Phase 1, which builds a semantically rich latent space capable of independently representing various domain-specific features. This intensive investment enables our model to effectively handle challenging domain gaps without requiring domain-specific preprocessing steps that would otherwise add significant manual effort during data preparation. Additionally, Phase 2 involves complex image-to-image translations and cycle-consistency operations that contribute to the computational demands but are essential for preserving label integrity across domains.

Despite the high training costs, our framework offers a favorable trade-off between computational requirements and segmentation performance, particularly in challenging domain adaptation scenarios where other methods fail to produce acceptable results. The computational overhead can be justifiable when considering the performance improvements, especially when considering domains with a significant gap.

5. Discussion and Conclusion

In this paper, we presented an end-to-end semi-supervised domain adaptation framework designed as an out-of-the-box tool for segmenting arteries and veins in images from different centers and/or modalities. To this end, we opted for a minimal pre-processing strategy that avoids any data harmonization between source and target domains. While enhancing the versatility of our model, this comes at the cost of widening the domain gap between the two domains. Our investigations analyzed this trade-off, delving into the concepts and mechanisms crucial for the effective functioning of our model. To address the problem of domain shift arising from different medical centers, imaging modalities, and vessel types, we rely on the path length regularization (Karras et al., 2020), which allows for representing heterogeneous volumetric data in a unified and disentangled latent space. Consequently, we explored the potential of disentanglement, investigating the possibility of modifying selected domain-specific features to achieve inter-domain translation in a label-preserving manner. One important design choice in our framework is the adoption of a 2.5D approach rather than full 3D processing. While our method targets 3D vessel segmentation, we process volumetric data as a series of multi-channel 2D slices. This technique enables larger batch sizes and reduces memory requirements, making the method more accessible with limited computational resources. We acknowledge that this approach may affect the continuity of vessels across distant slices. However, our evaluation using topology-aware metrics like clDice demonstrates that the method effectively preserves the topological structure of the cerebrovascular tree. This practical balance between accuracy and efficiency is particularly relevant for clinical applications where computational resources may be constrained. In addition to assessing the efficacy of disentanglement, we conducted ablation studies to determine the optimal number of source and target annotations and to evaluate the influence of key architectural choices on performance. Finally, we compared our framework against other state-of-the-art domain adaptation and domain generalization methods. Our approach demonstrates superior performance, accurately segmenting 3D brain vessels primarily using annotations from arterial images, which are comparatively easier to obtain. The results exhibit promising performance in semi-supervised domain adaptation scenarios, overcoming the difficulties posed by large domain gaps, in particular between veins and arteries, and the intricate morphology of the cerebrovascular tree. Despite our accomplishments, we acknowledge the potential for improvement. First, we highlight the necessity of our model to repeat training for each new target domain, and we note that in-context learning, as offered by methods like UniverSeg, presents a viable alternative. Furthermore, our model requires guidance in the form of m target annotated 2D slices. Again, foundation models can prove beneficial: by pre-training on extensive collections of tree-like objects, segmentation models can acquire a broader representation of vessels. This approach facilitates linking vessels from distant modalities without relying on any additional guidance.

Acknowledgments

Ethical Standards

The work follows appropriate ethical standards in conducting research and writing the manuscript, following all applicable laws and regulations regarding treatment of animals or human subjects.

Conflicts of Interest

We declare we do not have any conflicts of interest

Data availability

All datasets utilized in this study are publicly available and can be accessed freely. However, access to some of these datasets requires prior registration or application for access due to the sensitive nature of the data or to comply with data protection regulations. Despite the public availability of the datasets, direct sharing of the data by the authors is not permissible due to copyright and licensing restrictions imposed by the data providers. We encourage interested researchers to obtain the data directly from the respective repositories and websites, adhering to the specified access procedures and usage policies. Where applicable, the authors will share the labels obtained from their analyses.

References

Benjamin Billot, Douglas N Greve, Oula Puonti, Axel Thielscher, Koen Van Leemput, Bruce Fischl, Adrian V Dalca, Juan Eugenio Iglesias, et al. Synthseg: Segmentation of brain MRI scans of any contrast and resolution without retraining. *Medical Image Analysis*, 86:102789, 2023.

Victor Ion Butoi, Jose Javier Gonzalez Ortiz, Tianyu Ma, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. Universeg: Universal medical image segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 21438–21451, 2023.

Silvain Bériault, Yiming Xiao, D. Louis Collins, and G. Bruce Pike. Automatic swi venography segmentation using conditional random fields. *IEEE Transactions on Medical Imaging*, 34(12):2478–2491, 2015.

Cheng Chen, Qi Dou, Hao Chen, Jing Qin, and Pheng-Ann Heng. Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(1):865–872, 2019.

Cheng Chen, Kangneng Zhou, Zhiliang Wang, Qian Zhang, and Ruoxiu Xiao. All answers are in the images: A review of deep learning for cerebrovascular segmentation. *Computerized Medical Imaging and Graphics*, 107:102229, 2023.

Xuhui Chen, Yi Lu, Junjie Bai, Youbing Yin, Kunlin Cao, Yuwei Li, Hanbo Chen, Qi Song, and Jun Wu. Train a 3d u-net to segment cranial vasculature in cta volume without manual annotation. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 559–563, 2018.

Vien Ngoc Dang, Francesco Galati, Rosa Cortese, Giuseppe Di Giacomo, Viola Marconetto, Prateek Mathur, Karim Lekadir, Marco Lorenzi, Ferran Prados, and Maria A. Zuluaga. Vessel-CAPTCHA: An efficient learning framework for vessel annotation and segmentation. *Medical Image Analysis*, 75:102263, 2022.

Mahbaneh Eshaghzadeh Torbati, Davneet S. Minhas, Ghasan Ahmad, Erin E. O'Connor, John Muschelli, Charles M. Laymon, Zixi Yang, Ann D. Cohen, Howard J. Aizenstein, William E. Klunk, Bradley T. Christian, Seong Jae Hwang, Ciprian M. Crainiceanu, and Dana L. Tudorascu. A multi-scanner neuroimaging data harmonization using ravel and combat. *NeuroImage*, 245: 118703, 2021.

Daniele Falcetta, Vincenzo Marciano, Kaiyuan Yang, Jon Cleary, Loïc Legris, Massimiliano Domenico Rizzaro, Ioannis Pitsiorlas, Hava Chaptoukaev, Benjamin Lemasson, Bjoern Menze, and Maria A. Zuluaga. VesselVerse: A Dataset and Collaborative Framework for Vessel Annotation . In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2025*, volume LNCS 15972. Springer Nature Switzerland, September 2025.

Francesco Galati, Daniele Falcetta, Rosa Cortese, Barbara Casolla, Ferran Prados, Ninon Burgos, and Maria A Zuluaga. A2V: A semi-supervised domain adaptation framework for brain vessel segmentation via two-phase training angiography-to-venography translation. In *34th British Machine Vision Conference*, pages 750–751, 2023.

Leon A Gatys, Alexander S Ecker, and Matthias Bethge. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*, 2015.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, volume 27, 2014.

Ran Gu, Jingyang Zhang, Guotai Wang, Wenhui Lei, Tao Song, Xiaofan Zhang, Kang Li, and Shaoting Zhang. Contrastive semi-supervised learning for domain adaptive segmentation across similar anatomical structures. *IEEE Transactions on Medical Imaging*, 42(1):245–256, 2023.

Hao Guan and Mingxia Liu. Domain adaptation for medical image analysis: A survey. *IEEE Transactions on Biomedical Engineering*, 69(3):1173–1185, 2022.

Jin Hong, Simon Chun-Ho Yu, and Weitian Chen. Unsupervised domain adaptation for cross-modality liver segmentation via joint adversarial learning and self-learning. *Applied Soft Computing*, 121:108729, 2022a.

Jin Hong, Yu-Dong Zhang, and Weitian Chen. Source-free unsupervised domain adaptation for cross-modality abdominal multi-organ segmentation. *Knowledge-Based Systems*, 250:109155, 2022b.

Andrew Hoopes, Jocelyn S. Mora, Adrian V. Dalca, Bruce Fischl, and Malte Hoffmann. SynthStrip: skull-stripping for any brain image. *NeuroImage*, 260:119474, 2022. ISSN 1053-8119.

Dewei Hu, Hao Li, Han Liu, and Ipek Oguz. Domain generalization for retinal vessel segmentation via hessian-based vector field. *Medical Image Analysis*, 95:103164, 2024.

Yuhao Huang, Xin Yang, Lian Liu, Han Zhou, Ao Chang, Xinrui Zhou, Rusi Chen, Junxuan Yu, Jiongquan Chen, Chaoyu Chen, et al. Segment anything model for medical images? *Medical Image Analysis*, 92:103061, 2024.

Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.

Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8110–8119, 2020.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision (ICCV)*, pages 4015–4026, 2023.

Pamela J. LaMontagne, Tammie LS. Benzinger, John C. Morris, Sarah Keefe, Russ Hornbeck, Chengjie Xiong, Elizabeth Grant, Jason Hassenstab, Krista Moulder, Andrei G. Vlassenko, Marcus E. Raichle, Carlos Cruchaga, and Daniel Marcus. Oasis-3: Longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. *medRxiv*, 2019.

Seunghun Lee, Sunghyun Cho, and Sunghoon Im. Dranet: Disentangling representation and adaptation networks for unsupervised cross-domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15252–15261, 2021.

Jiaxin Li, Houjin Chen, Yanfeng Li, Yahui Peng, Jia Sun, and Pan Pan. Cross-modality synthesis aiding lung tumor segmentation on multi-modal mri images. *Biomedical Signal Processing and Control*, 76:103655, 2022.

Hui Lin, Florian Schiffers, Santiago López-Tapia, Neda Tavakoli, Daniel Kim, and Aggelos K. Katsaggelos. Drlstnet: Unsupervised domain adaptation for cross-modality medical image segmentation via disentangled representation learning. In *MICCAI 2024 FLAIR Challenge Proceedings*, 2024.

Xiaofeng Liu, Fangxu Xing, Nadya Shusharina, Ruth Lim, C-C Jay Kuo, Georges El Fakhri, and Jonghye Woo. Act: Semi-supervised domain-adaptive medical image segmentation with asymmetric co-training. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pages 66–76, 2022.

Michelle Livne, Jana Rieger, Orhun Utku Aydin, Abdel Aziz Taha, Ela Marie Akay, Tabea Kossen, Jan Sobesky, John D. Kelleher, Kristian Hildebrand, Dietmar Frey, and Vince I. Madai. A u-net deep learning framework for high performance vessel segmentation in patients with cerebrovascular disease. *Frontiers in Neuroscience*, 13, 2019.

Junyan Lyu, Yiqi Zhang, Yijin Huang, Li Lin, Pujin Cheng, and Xiaoying Tang. Aadg: Automatic augmentation

for domain generalization on retinal image segmentation. *IEEE Transactions on Medical Imaging*, 41(12):3699–3711, 2022. .

Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15(1):654, 2024.

Midas Meijs, Ajay Patel, Sil C van de Leemput, Mathias Prokop, Ewoud J van Dijk, Frank-Erik de Leeuw, Frederick JA Meijer, Bram van Ginneken, and Rashindra Manniesing. Robust segmentation of the full cerebral vasculature in 4d ct of suspected stroke patients. *Scientific reports*, 7(1):15622, 2017.

Lars Mescheder, Andreas Geiger, and Sebastian Nowozin. Which training methods for GANs do actually converge? In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 3481–3490, 2018.

Munan Ning, Cheng Bian, Dong Wei, Shuang Yu, Chenglang Yuan, Yaohua Wang, Yang Guo, Kai Ma, and Yefeng Zheng. A new bidirectional unsupervised domain adaptation segmentation framework. In Aasa Feragen, Stefan Sommer, Julia Schnabel, and Mads Nielsen, editors, *Information Processing in Medical Imaging*, pages 492–503, Cham, 2021.

Nicolas Passat, Christian Ronse, Joseph Baruthio, J-P Armspach, and Jack Foucher. Watershed and multimodal data for brain vessel segmentation: Application to the superior sagittal sinus. *Image and Vision Computing*, 25(4):512–521, 2007.

Linkai Peng, Li Lin, Pujin Cheng, Ziqi Huang, and Xiaoying Tang. Unsupervised domain adaptation for cross-modality retinal vessel segmentation via disentangling representation style transfer and collaborative consistency learning. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5, 2022.

Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. Encoding in style: A stylegan encoder for image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2287–2296, 2021.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, 2015.

Suprosanna Shit, Johannes C. Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien P. W. Pluim, Ulrich Bauer, and Bjoern H. Menze. cldice - a novel topology-preserving loss function for tubular structure segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16560–16569, 2021.

Giles Tetteh, Velizar Efremov, Nils D. Forkert, Matthias Schneider, Jan Kirschke, Bruno Weber, Claus Zimmer, Marie Piraud, and Björn H. Menze. Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes. *Frontiers in Neuroscience*, 14, 2020.

Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7472–7481, 2018.

D.L. Wilson and J.A. Noble. An adaptive segmentation algorithm for time-of-flight mra data. *IEEE Transactions on Medical Imaging*, 18(10):938–945, 1999.

Fuping Wu and Xiahai Zhuang. Unsupervised domain adaptation with variational approximation for cardiac segmentation. *IEEE Transactions on Medical Imaging*, 40(12): 3555–3567, 2021.

Jianghao Wu, Ran Gu, Guiming Dong, Guotai Wang, and Shaoting Zhang. Fpl-uda: Filtered pseudo label-based unsupervised cross-modality adaptation for vestibular schwannoma segmentation. In *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5, 2022.

Kaiyuan Yang, Fabio Musio, Yihui Ma, Norman Juchler, Johannes C. Paetzold, Rami Al-Maskari, Luciano Höher, Hongwei Bran Li, Ibrahim Ethem Hamamci, Anjany Sekuboyina, Suprosanna Shit, Houjing Huang, Diana Waldmannstetter, Florian Kofler, Fernando Navarro, Martin Menten, Ivan Ezhov, Daniel Rueckert, Iris Vos, Ynte Ruigrok, Birgitta Velthuis, Hugo Kuijf, Julien Hämmerli, Catherine Wurster, Philippe Bijlenga, Laura Westphal, Jeroen Bisschop, Elisa Colombo, Hakim Baazaoui, Andrew Makmur, James Hallinan, Bene Wiestler, Jan S. Kirschke, Roland Wiest, Emmanuel Montagnon, Laurent Letourneau-Guillon, Adrian Galdran, Francesco Galati, Daniele Falcetta, Maria A. Zuluaga, Chaolong Lin, Haoran Zhao, Zehan Zhang, Sinyoung Ra, Jongyun Hwang, Hyunjin Park, Junqiang Chen, Marek Wodzinski, Henning Müller, Pengcheng Shi, Wei Liu, Ting Ma, Cansu Yalçin, Rachika E. Hamadache, Joaquim Salvi, Xavier Llado, Uma Maria Lal-Trehan Estrada, Valeriia Abramova, Luca Giancardo, Arnau Oliver, Jialu Liu, Haibin Huang,

Yue Cui, Zehang Lin, Yusheng Liu, Shunzhi Zhu, Tatsat R. Patel, Vincent M. Tutino, Maysam Orouskhani, Huayu Wang, Mahmud Mossa-Basha, Chengcheng Zhu, Maximilian R. Rokuss, Yannick Kirchhoff, Nico Disch, Julius Holzschuh, Fabian Isensee, Klaus Maier-Hein, Yuki Sato, Sven Hirsch, Susanne Wegener, and Bjoern Menze. Topcow: Benchmarking topology-aware anatomical segmentation of the circle of willis (cow) for cta and mra. *arXiv : 2312.17670*, 2024.

Shuai Yang, Liming Jiang, Ziwei Liu, and Chen Change Loy. Unsupervised image-to-image translation with generative prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18332–18341, 2022.

Kai Yao, Zixian Su, Kaizhu Huang, Xi Yang, Jie Sun, Amir Hussain, and Frans Coenen. A novel 3d unsupervised domain adaptation framework for cross-modality medical image segmentation. *IEEE Journal of Biomedical and Health Informatics*, 26(10):4976–4986, 2022a.

Xu Yao, Alasdair Newson, Yann Gousseau, and Pierre Hellier. A style-based gan encoder for high fidelity reconstruction of images and videos. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, editors, *Computer Vision – ECCV 2022*, pages 581–597, 2022b.

Varduhi Yeghiazaryan and Irina D. Voiculescu. Family of boundary overlap metrics for the evaluation of medical image segmentation. *Journal of Medical Imaging*, 5(1): 015006, 2018.

Baochang Zhang, Zichen Zhang, Shuting Liu, Shahrooz Faghihroohi, Heribert Schunkert, and Nassir Navab. XA-Sim2Real: Adaptive Representation Learning for Vessel Segmentation in X-ray Angiography . In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume LNCS 15006, pages 747–756, 2024.

Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–595, 2018.

Yuxuan Zhang, Huan Ling, Jun Gao, Kangxue Yin, Jean-Francois Lafleche, Adela Barriuso, Antonio Torralba, and Sanja Fidler. Datasetgan: Efficient labeled data factory with minimal human effort. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10145–10155, 2021.

Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. Domain generalization: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4396–4415, 2023.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2223–2232, 2017.

Xueyan Zou, Jianwei Yang, Hao Zhang, Feng Li, Linjie Li, Jianfeng Wang, Lijuan Wang, Jianfeng Gao, and Yong Jae Lee. Segment everything everywhere all at once. *Advances in neural information processing systems*, 36:19769–19782, 2023.

Maria A Zuluaga, Roman Rodionov, Mark Nowell, Sufyan Achhala, Gergely Zombori, Alex F Mendelson, M Jorge Cardoso, Anna Miserocchi, Andrew W McEvoy, John S Duncan, et al. Stability, structure and scale: improvements in multi-modal vessel extraction for seeg trajectory planning. *International journal of computer assisted radiology and surgery*, 10:1227–1237, 2015.