COSTA: A Multi-Center TOF-MRA Dataset and a Style Self-Consistency Network for Cerebrovascular Segmentation

Lei Mou, Jinghui Lin, Yifan Zhao, *Senior Member, IEEE*, Yonghuai Liu, Shaodong Ma, Jiong Zhang, *Member, IEEE*, Wenhao Lv, Tao Zhou, Jiang Liu, *Senior Member, IEEE*, Alejandro F. Frangi, *Fellow, IEEE*, and Yitian Zhao, *Member, IEEE*

***Abstract*—Time-of-flight magnetic resonance angiography (TOF-MRA) is the least invasive and ionizing radiation-free approach for cerebrovascular imaging, but variations in imaging artifacts across different clinical centers and imaging vendors result in inter-site and inter-vendor heterogeneity, making its accurate and robust cerebrovascular segmentation challenging. Moreover, the limited availability and quality of annotated data pose further challenges for segmentation methods to generalize well to unseen datasets. In this paper, we construct the largest and most diverse TOF-MRA dataset (COSTA) from 8 individual imaging centers, with all the volumes manually annotated. Then we propose a novel network for cerebrovascular segmentation, namely CESAR, with the ability to tackle feature granularity and image style heterogeneity issues. Specifically, a coarse-to-fine architecture is implemented to refine cerebrovascular segmentation in an iterative manner. An automatic feature selection module is proposed to selectively fuse global long-range dependencies and local contextual information of cerebrovascular structures. A style self-consistency loss is then introduced to explicitly align diverse styles of TOF-MRA images to a standardized one. Extensive experimental results on the COSTA dataset demonstrate the effectiveness of our CESAR network against state-of-the-art methods. We have made 6 subsets of COSTA with the source code online available, in order to promote relevant research in the community.**

***Index Terms*—Multi-center and multi-vector, TOF-MRA, heterogeneity, style self-consistency, cerebrovascular segmentation.**

I. Introduction

THE human cerebrovascular system is a complex network of blood vessels that spans the entire brain [1]. Cerebrovascular structures play an important role in brain health, being not only essential for the delivery of oxygen and nutrients, but also central to the mechanisms for both cerebrovascular and systemic diseases, such as aneurysms [2], arteriovenous malformations [3], moyamoya disease [4], and Alzheimer's disease [5]. Accurately measuring morphological changes in the cerebrovascular system, especially in the cerebral arteries, is essential for early detection of and intervention in cerebrovascular and systemic diseases [6].

In clinical practice, 3D time-of-flight magnetic resonance angiography (TOF-MRA) is now routinely available to diagnose abnormalities in the cerebrovascular system, as it is the least invasive and ionizing radiation-free approach, in contrast to cerebrovascular digital subtraction angiography (DSA) and computed tomography angiography (CTA) [7], [8]. Therefore, the development and validation of precise and dependable methodologies for cerebrovascular segmentation are indispensable.

To obtain the topology of the cerebrovascular system, a number of cerebrovascular segmentation methods have been developed. These methods can be classified into two major categories: unsupervised methods requiring manually designed descriptors; and segmentation methods driven by models and data, which may be fully-supervised, semi-supervised,

Received 29 April 2024; revised 26 June 2024; accepted 3 July 2024. Date of publication 16 July 2024; date of current version 2 December 2024. This work was supported in part by the National Science Foundation Program of China under Grant 62272444, Grant 62371442, Grant 62103398, and Grant 62302488; in part by Zhejiang Provincial Natural Science Foundation of China under Grant LR22F020008, Grant LQ23F010007, Grant LR24F010002, and Grant LZ23F010002; in part by the Key Research and Development Program of Zhejiang Province under Grant 2024C03281; in part by Ningbo 2025 Science and Technology (S&T) Megaprojects under Grant 2022Z127 and Grant 2021Z134; and in part by the Key Project of Ningbo Public Welfare Science and Technology under Grant 2023S012 and Grant 2021S107. *(Corresponding authors: Yitian Zhao; Jinghui Lin.)*

Lei Mou, Shaodong Ma, Jiong Zhang, and Yitian Zhao are with the Laboratory of Advanced Theranostic Materials and Technology, Ningbo Institute of Materials Technology and Engineering, Chinese Academy of Sciences, Ningbo 315201, China, and also with Ningbo Cixi Institute of Biomedical Engineering, Ningbo 315300, China (e-mail: moulei@nimte.ac.cn; mashaodong@nimte.ac.cn; jiong.zhang@ieee.org; yitian.zhao@nimte.ac.cn).

Jinghui Lin and Wenhao Lv are with The First Affiliated Hospital of Ningbo University, Ningbo 315010, China (e-mail: fyylinjinghui@nbu.edu.cn; wenhaolv007@163.com).

Yifan Zhao is with the EPSRC Centre for Innovative Manufacturing in Through-Life Engineering Services, Cranfield University, MK43 0AL Cranfield, U.K. (e-mail: yifan.zhao@cranfield.ac.uk).

Yonghuai Liu is with the Department of Computer Science, Edge Hill University, L39 4QP Ormskirk, U.K. (e-mail: yonghuai.liu@edgehill.ac.uk).

Tao Zhou is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: taozhou@njust.edu.cn).

Jiang Liu is with the Research Institute of Trustworthy Autonomous Systems and Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: liuj@sustech.edu.cn).

Alejandro F. Frangi is with the Department of Computer Science, School of Health Sciences, University of Manchester, M13 9PL Manchester, U.K. (e-mail: alejandro.frangi@manchester.ac.uk).

Digital Object Identifier 10.1109/TMI.2024.3424976

or weakly-supervised. The first category of methods involves using manually designed feature descriptors or filters to segment the cerebrovascular structures. Examples of such methods include Hessian matrix-based filters [9], stochastic models [10], level-sets [11], the weighted symmetric filter [12], and fuzzy c-means clustering [13]. However, due to their inherent limitations, these methods suffer from suboptimal segmentation accuracy and efficiency. The second category is predominantly composed of learning-based methods for cerebrovascular segmentation, comprising both fully supervised methods [14], [15], [16], [17], [18], [19], [20], [21], [22] and semi/weakly supervised methods [23], [24], [25], [26].

Although many methods have been proposed to achieve precise segmentation of the cerebrovascular system, several challenges remain. On one hand, most existing methods [14], [23] have utilized TOF-MRA images acquired from a single-center/vendor to train a specific segmentation model: this has significantly limited their generalizability to unseen datasets [27], [28], [29], [30]. Moreover, differences in acquisition devices or protocols can cause disparities in scanning setup, magnetic field strengths, voxel sizes, and scanning resolutions, leading to shifts in data representations and domains across various data centers. Varying voxel spacing and resolutions may affect the boundaries and accuracy of the segmented cerebrovascular structures. Differences in manufacturers and magnetic field strengths can introduce variations in image quality, contrast, and intensity distributions. The domain shifts [31] in cross-domain MRA images captured by different devices limit their potential for enhancing the generalization of segmentation models when relying exclusively on a single-domain dataset. On the other hand, the scarcity of manually annotated high-quality public-access datasets hinders the precision of cerebrovascular segmentation in TOF-MRA images. Existing TOF-MRA datasets with cerebrovascular annotations used in some studies [21], [32] are publicly available but subject to different requirements, focuses and protocols. As depicted in Fig. 1, samples from the TubeTK[1] and IXI Dataset[2] reveal limitations in annotations: the annotation from [33] inadequately covers vascular voxels, resulting in noise and sparse labeling of cerebral arteries (green arrow), and [32] omits annotation of certain small vessel segments (yellow arrow). Therefore, establishing a comprehensive cross-domain TOF-MRA dataset with high standard unified cerebrovascular annotations is crucial to facilitate the development and validation of robust and effective segmentation models.

In this paper, we curated a dataset containing diverse TOF-MRA images from various centers and vendors. Furthermore, to address the challenges posed by the variability of TOF-MRA images and the intricate nature of cerebrovascular features, we propose a novel segmentation method called CESAR tailored for multi-center, multi-vendor TOF-MRA imaging. CESAR consists of three key components: a coarse-to-fine (C2F) backbone network, an automated global and local feature selection (AutoGLFS) module, and a style self-consistency loss, as depicted in Fig. 3. The C2F backbone encompasses two networks: CoarseNet, which generates an initial cerebrovascular map, and RefineNet, which improves segmentation quality by capturing global long-range dependencies and the local context of the cerebrovascular system. Additionally, the AutoGLFS module is designed to enhance segmentation accuracy by selecting informative features from both global and local perspectives. The style self-consistency loss is proposed to bolster model generalization by explicitly aligning the semantic styles of features from RefineNet and CoarseNet. To sum up, this study makes the following contributions:

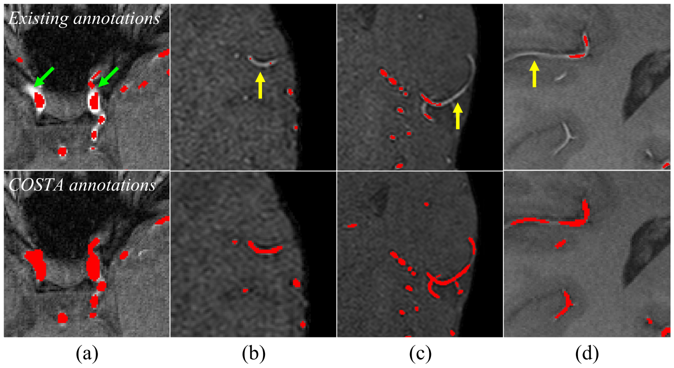

Fig. 1. Illustrative samples of TOF-MRA images and their cerebrovascular annotations (top) from two publicly available datasets - TubeTK (a) and IXI Dataset (b-d) - suggest that existing cerebrovascular annotations do not capture all the vessels of interest. The second row shows the manual annotations of the constructed COSTA.

- For the first time in the cerebrovascular segmentation field, we establish a large and publicly available TOF-MRA **C**erebr**O**vascular **S**egmen**TA**tion dataset (**COSTA**) with high-quality annotations, covering multiple centers and vendors. These manual annotations are available at https://github.com/iMED-Lab/COSTA
- We propose a novel CErebrovaSculAR segmentation network (CESAR) adaptable to heterogeneous multi-center TOF-MRA images. Our method leverages a style self-consistency loss to align heterogeneous styles to a uniform one in the latent space, which normalizes the styles of high-level features, thereby improving the performance of the model in segmenting cerebrovascular structures across different data centers.
- We propose an automatic global-local feature selection module to address the challenge of heterogeneous feature granularity in integrating global and local information. The proposed AutoGLFS selectively fuses long-range dependencies from the Transformer layer and local context from the CNN layer, thus enabling effective integration of global-local features to improve the segmentation performance.
- We perform a comprehensive evaluation of cerebrovascular segmentation methods, both quantitatively and qualitatively. Experimental results demonstrate the effectiveness of the CESAR against the other state-of-the-art methods.

II. COSTA Dataset

In this study, we curated a comprehensive cerebrovascular segmentation dataset, named **COSTA** by amalgamating eight

[1] https://public.kitware.com/Wiki/TubeTK/Data
[2] http://brain-development.org/ixi-dataset/

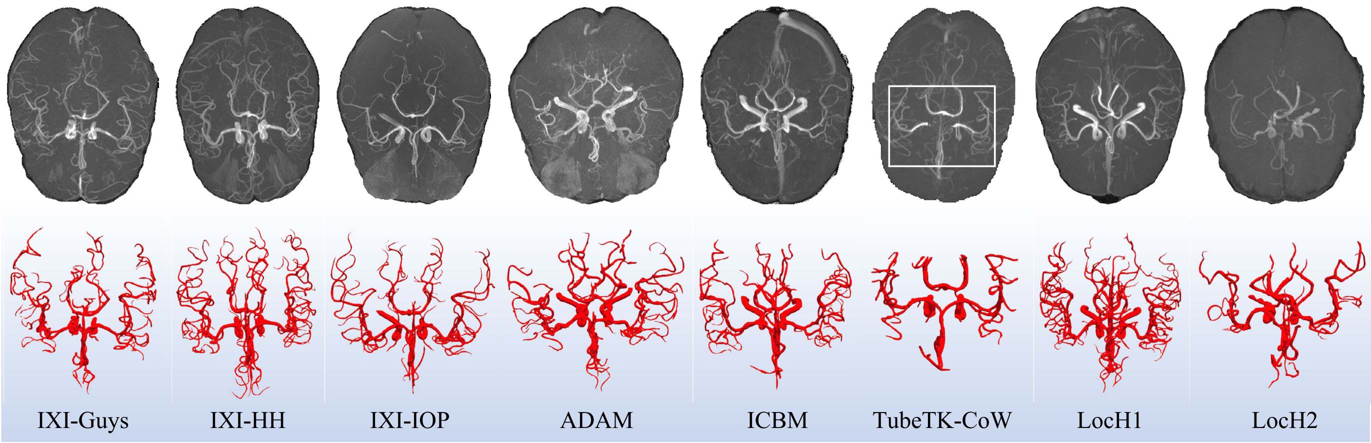

Fig. 2. Illustrative examples from the COSTA dataset. The first row shows the maximum intensity projection (MIP) of TOF-MRA images obtained from various centers, while the second row displays the corresponding ground truth represented as mesh surfaces. The contrast of the MIPs was modified to enhance the visualization of cerebrovascular structures.

TABLE I
A SUMMARY OF THE COSTA DATABASE

Dataset	Voxel spacing (mm^3)	Resolution	Manufacturer	Magnetic field strength	# Volume
IXI-Guys	$0.47 \times 0.47 \times 0.80$	$512 \times 512 \times 100$	Philips Medical Systems Gyroscan Intera 1.5T	1.5 Tesla	60
IXI-HH	$0.47 \times 0.47 \times 0.80$	$512 \times 512 \times 100$	Philips Medical Systems Intera 3T	3.0 Tesla	60
IXI-IOP	$0.26 \times 0.26 \times 0.80$	$1024 \times 1024 \times 92$	General Electric (scanner not available)	1.5 Tesla	50
ADAM	$0.20 \times 0.20 \times 0.40$ to $0.20 \times 0.20 \times 0.40$	$256 \times 256 \times 100$ to $560 \times 560 \times 140$	Not available	3.0 Tesla	107
ICBM	$0.62 \times 0.62 \times 0.62$	$288 \times 288 \times 200$	Siemens TRIO 3T	3.0 Tesla	50
TubeTK-CoW	$0.51 \times 0.51 \times 0.80$	$448 \times 448 \times 128$	Siemens Allegra head-only 3T MR system	3.0 Tesla	50
LocH1	$0.34 \times 0.34 \times 0.80$ to $0.39 \times 0.39 \times 0.90$	$406 \times 512 \times 186$ to $500 \times 640 \times 121$	Siemens HealthCare GmbH	3.0 Tesla	27
LocH2	$0.39 \times 0.39 \times 0.50$ to $0.43 \times 0.43 \times 1.00$	$442 \times 512 \times 192$	United Imaging Healthcare (scanner not available)	3.0 Tesla	19

datasets from both public and private sources, as illustrated in Fig. 2. COSTA encompasses 423 TOF-MRA volumes collected from eight distinct data centers, each utilizing MRI scanners from four different vendors. The eight datasets included in COSTA (Table I) exhibit significant diversity, particularly in terms of acquisition devices, scanning setup, magnetic field strengths, voxel sizes, and scanning resolutions. Details for each dataset are provided below.

- **IXI-Guys** dataset consists of 60 MR images of healthy subjects randomly selected from the IXI Dataset, all acquired at Guy's Hospital in London, UK. Each subject underwent TOF-MRA imaging using a 1.5T MRI scanner (Philips Medical System Gyroscan Intera) with the following parameters: TR/TE = 20/6.9 ms, flip angle = 25°. All images in the dataset share the same dimensions of $512 \times 512 \times 100$ voxels and have a voxel spacing of $0.47 \times 0.47 \times 0.80$ mm^3.
- **IXI-HH** dataset includes MRA images of 60 healthy volunteers randomly chosen from the IXI Dataset. These images were obtained using a Philips 3.0T MRI system at Hammersmith Hospital in London, UK, with imaging parameters set at TR/TE = 16.72/5.75 ms and a flip angle of 16°. All the images in this dataset share the same dimensions of $512 \times 512 \times 100$ voxels, with a voxel spacing of $0.47 \times 0.47 \times 0.80$ mm^3.
- **IXI-IOP** dataset comprises 50 MR images of healthy subjects randomly selected from the IXI Dataset, originally obtained from the Institute of Psychiatry. These images were acquired using a General Electric (GE) 1.5T MRI scanner

for TOF-MRA imaging, but specific scan parameters are unavailable. The dataset features images with dimensions of 1024 × 1024 × 92 and a voxel spacing of 0.26 × 0.26 × 0.80 mm^3.

- **ADAM** dataset was obtained from the online Aneurysm Detection and Segmentation Challenge[3] and comprises 107 TOF-MRA images. All the images were acquired using a 3.0T MRI scanner with a TR/TE of 22/3.45 ms. Notably, the dataset exhibits a range of physical resolutions and voxel sizes, varying from 0.20×0.20×0.40 mm^3 to 0.59×0.59×1.00 mm^3, and from 256 × 256 × 100 voxels to 560 × 560 × 140 voxels, respectively. Of the subjects in this dataset, 85 were diagnosed with aneurysms, while the remaining 22 were healthy individuals.
- **ICBM** dataset includes 50 TOF-MRA images of healthy volunteers, evenly split between 25 females and 25 males, ranging in age from 19 to 64 years. These images were sourced from the Magnetic Resonance Angiography Atlas Dataset[4] and were captured using a 3.0T Siemens MRI scanner. They exhibit a consistent physical resolution of 0.62 × 0.62 × 0.62 mm^3 and possess dimensions of 288 × 320 × 200 voxels.
- **TubeTK-CoW** dataset is a subset of the TubeTK. It includes 50 TOF-MRA images from 25 male and 25 female healthy volunteers, obtained with a standardized protocol on a SIEMENS 3.0T MRI scanner. The images have a voxel spacing of 0.51 × 0.51 × 0.80 mm^3 and dimensions of 448 × 448 × 128 voxels.
- **LocH1** dataset comprises 27 TOF-MRA images gathered from a local hospital, all acquired with a 3.0T SIEMENS MRI scanner. The images exhibit a range of physical resolutions, varying from 0.34×0.34×0.80 mm^3 to 0.39×0.39×0.90 mm^3, and voxel dimensions that span from 406×512×186 to 500×640 × 121. Among the subjects, 2 of the 27 had Moyamoya disease, while the others were healthy individuals..
- **LocH2** dataset includes 19 TOF-MRA volumes of healthy subjects from a different local hospital. These subjects all underwent TOF-MRA imaging using a 3T United Imaging Healthcare (UIH) MRI scanner. The dataset features a variety of physical resolutions, with ranges from 0.39 × 0.39 × 0.50 mm^3 to 0.43 × 0.43 × 1.00 mm^3, and consistent voxel dimensions of 442 × 512 × 192.

Among the 8 datasets, only TubeTK-CoW had manual annotations, as utilized in prior research [17]. For the remaining 7 datasets, we employed a semi-automated approach for voxel-level annotation of cerebral arteries. (1) We begin by generating initial cerebrovascular masks (M_{init}) for each TOF-MRA image using the Local Thresholding plug-in within the open-source 3D Slicer platform[5] [34]. (2) To refine the accuracy of M_{init}, a radiologist with more than five years of experience meticulously adjusts the masks on a slice-by-slice basis, resulting in M_{refine}. (3) Two additional radiologists with more than a decade of experience then conduct a thorough double-check of M_{refine}, correcting any inaccuracies to produce consensus masks (M_{double}). (4) The manual masks (M_{double}) and corresponding TOF-MRA images are split into two parts, M_1 and M_2, which serve as training data for the CS^2-Net [17] segmentation model (CS^2-Net$_1$ and CS^2-Net$_2$). (5) We apply the trained CS^2-Net$_1$ to segment images within M_2 and CS^2-Net$_2$ to segment images within M_1, resulting in segmentation masks M_{seg1} and M_{seg2}, respectively. (6) We merge M_{seg1} and M_{seg2} to obtain M_{seg}, which serves as the new M_{init}. This process is then iterated, involving Steps (2), (3), (4), and (5), until further annotation becomes unnecessary for cerebrovascular structures below the $1mm$ threshold, as determined by both the radiologists in Step (3). Ultimately, the consensus masks (M_{double}) obtained through these steps represent the ground truth for cerebrovascular structures in the COSTA dataset. The comprehensive process entailed approximately 2.0∼2.5 hours per image for its semi-automated annotation. Finally, a 5-fold cross-validation approach was utilized to divide the dataset into training (280 images), validation (71 images), and test (72 images) sets. All the private images in LocH1 and LocH2 were acquired with regulatory approvals and patient consents as appropriate, following the Declaration of Helsinki.

III. Proposed Method

The proposed CESAR comprises three main components: a coarse-to-fine backbone (C2F); an automatic global and local feature selection module (AutoGLFS); and a style self-consistency loss (SSL). In the C2F process, the CoarseNet initially processes the intensity histogram-standardized TOF-MRA image to produce a coarse cerebrovascular map. This map is then combined with the raw TOF-MRA image and finally fed into the RefineNet for cerebrovascular segmentation refinement.

A. Coarse-to-Fine Backbone

The CESAR framework employs a coarse-to-fine (C2F) backbone, drawing on refined feature extraction techniques [35], [36], [37], illustrated in Fig. 3. This architecture comprises two core components: the CoarseNet and the RefineNet. The CoarseNet, a derivative of the U-Net model [38], integrates an encoder-decoder structure, each equipped with four CNN-based feature extraction blocks. These blocks consist of two sequential convolutional layers with instance normalization and ReLU activation, followed by max-pooling layers to expand receptive fields while reducing spatial resolution. Additionally, there is a bridge layer between the encoder and decoder, which is similar in structure to the feature extraction block and facilitates feature transition and enhancement.

The RefineNet also comprises an encoder, decoder, and a bridge layer. Building upon the success of Transformers in computer vision [39], our RefineNet extends its capabilities by incorporating a shifted window-based hierarchical Transformer (SwinT) [40] as an additional encoder branch. The limitations in accurately segmenting cerebrovascular structures within TOF-MRA images arise from multiple factors, including the inherent fragility and sparse representation of local cerebrovascular structures. Additionally, the intricate and

[3] https://adam.isi.uu.nl

[4] https://www.nitrc.org/projects/icbmmra

[5] https://www.slicer.org/

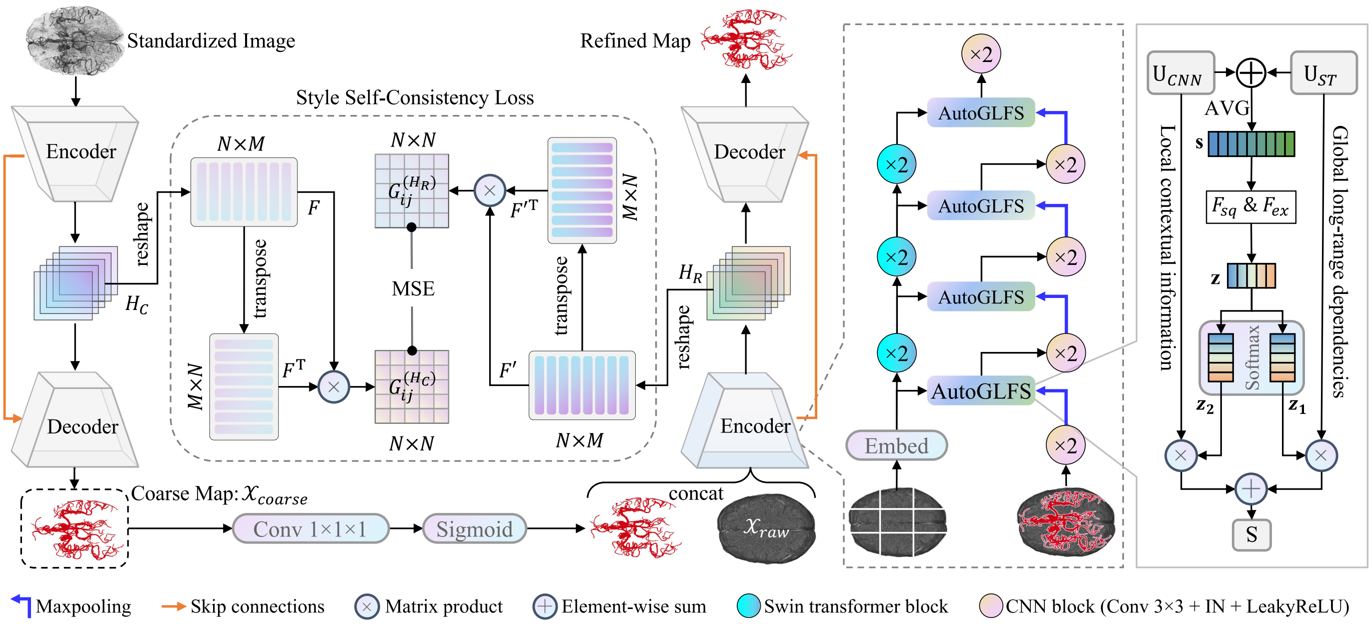

Fig. 3. Schematic of CESAR highlighting its key components: the coarse-to-fine backbone (C2F), style self-consistency loss (SSL), and automatic global and local feature selection module (AutoGLFS). The C2F framework allows fine-grained segmentation, while the SSL boosts adaptability to multi-centric data. The AutoGLFS selectively fuses the long-range dependencies and local contexts.

varied morphology of cerebrovascular networks, particularly the substantial regional differences in smaller vessel configurations, poses a challenge for existing methods. To enhance the precision of cerebrovascular segmentation, this study proposes the incorporation of SwinT and CNN modules within the hierarchical encoder framework of RefineNet. This integration is designed to specifically extract long-range dependencies inherent in cerebrovascular structures, while also capturing their intricate local discriminative features.

B. Automatic Global Local Feature Selection (AutoGLFS)

In the SwinT encoder of RefineNet, we utilize an embedding block alongside three SwinT blocks for hierarchical feature extraction. The embedding block employs a patch partition layer to divide the raw image ($\mathcal{X}_{raw} \in \mathbb{R}^{H\times W\times D\times C}$, where H, W, D and C represent its height, width, depth, and channels respectively) into non-overlapping patches, which are then linearly projected into a dimension S. Subsequently, three 3D SwinT blocks [41], are employed to perform feature extraction on these 3D patches. In the CNN encoder of RefineNet, we first convert the 2-channel (one channel for background, one for cerebrovascular structures) coarse segmentation map obtained from CoarseNet into a 1-channel cerebrovascular map ($\mathcal{X}_{coarse} \in \mathbb{R}^{H\times W\times D\times 1}$), by means of a $1\times 1\times 1$ kernel convolutional layer with an *Sigmoid* activation function. We then concatenate $\mathcal{X}_{coarse}$ and $\mathcal{X}_{raw}$ to create a 2-channel input for the CNN encoder. The proposed RefineNet integrates both the SwinT encoder and a CNN encoder, each comprising N ($N = 4$ in this paper) feature-encoding stages. However, effectively leveraging the global long-range dependencies extracted by the SwinT encoder and the local contextual semantics extracted by the CNN encoder poses a significant challenge.

To address this challenge, inspired by [42], we propose an AutoGLFS module for the automatic selection and fusion of discriminative cerebrovascular features from both global and local perspectives. Specifically, in the n^{th} SwinT encoder stage, we employ a $1\times 1\times 1$ convolutional layer to adjust the channel of the SwinT feature map ($\mathbf{U}_{ST}^{(n)}$) to match that of the CNN feature map ($\mathbf{U}_{CNN}^{(n)}$). Subsequently, an element-wise summation operation combines $\mathbf{U}_{ST}^{(n)}$ and $\mathbf{U}_{CNN}^{(n}$), yielding a feature union $\mathbf{U}^{(n)}$ with dimensions of $\frac{H}{2^n}\times\frac{W}{2^n}\times\frac{D}{2^n}\times C_{CNN}^{(n)}$. Here, $C_{CNN}^{(n)}$ represents the number of channels in $\mathbf{U}_{CNN}^{(n)}$. The feature union $\mathbf{U}^{(n)}$ is subjected to global average pooling to generate channel-level statistics regarding global spatial information, denoted as $\mathbf{s} \in \mathbb{R}^{C_{CNN}^{(n)}}$, i.e.,

$$\mathbf{s}_c = \frac{2^n}{H\times W\times D}\sum_{i}^{H/2^n}\sum_{j}^{W/2^n}\sum_{k}^{D/2^n}\mathbf{U}_c^{(n)}(i,j,k),$$
$$\mathbf{s} = \left[\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_c, \cdots, \mathbf{s}_{C_{CNN}^{(n)}}\right], \tag{1}$$

where $\mathbf{s}_c$ denotes the c^{th} element of $\mathbf{s}$. Learning of global-local features is enhanced by explicitly modeling channel interdependencies by performing Squeeze-and-Excitation [43] operations on $\mathbf{s}$: $\mathbf{z} = \mathcal{F}_{ex}(\sigma(\mathcal{F}_{sq}(\mathbf{s})))$, where $\mathcal{F}_{sq}$ is a fully connected layer to compact $\mathbf{s}$ to $\mathbb{R}^{d\times 1}$, and σ is a ReLU activation. $\mathcal{F}_{ex}$ is another fully connected layer that transforms the activated compact vector $\mathbf{s} \in \mathbb{R}^{d\times 1}$ to excitation vector $\mathbf{z} \in \mathbb{R}^{2C_{CNN}^{(n)}\times 1}$.

Subsequently, $\mathbf{z}$ is reshaped as $\mathbb{R}^{2\times C_{CNN}^{(n)}}$ and split into two vectors, $\mathbf{z}_1 \in \mathbb{R}^{C_{CNN}^{(n)}}$ and $\mathbf{z}_2 \in \mathbb{R}^{C_{CNN}^{(n)}}$, after applying Softmax to its first dimension. The final adaptively selected feature maps $\mathbf{S}$ are obtained by performing a channel-wise multiplication between attention vectors and feature maps, respectively, followed by an element-wise summation of the

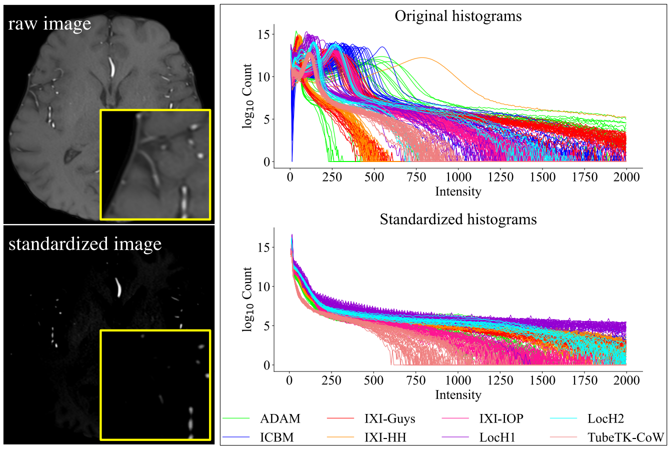

Fig. 4. Examples comparing raw images to standardized images. The right column represents the style and intensity histogram comparisons of the images before and after standardization, respectively.

two vectors:

$$\mathbf{S} = \mathbf{z}_1 \cdot \mathbf{U}_{ST}^{(n)} + \mathbf{z}_2 \cdot \mathbf{U}_{CNN}^{(n)}. \tag{2}$$

Compared to existing feature fusion methods that fuse features of the same granularity, the proposed AutoGLFS improves feature representation and enhances cerebrovascular segmentation performance by selectively integrating long-range dependencies and local contextual details across varying feature granularities.

C. Style Self-Consistency Loss

As depicted in Fig. 4, the histogram-standardized TOF-MRA images exhibit a consistent style, with major cerebrovascular features appearing as bright voxels against a relatively dark background. However, compared to raw TOF-MRA images, the standardized ones tend to lose small cerebrovascular structures (as indicated in the yellow box). As illustrated in Fig. 4, the standardized images exhibit a notably more consistent distribution of histogram variations in contrast to the raw, unprocessed ones.

Therefore, to fully exploit the advantages of both standardized and raw images while minimizing their limitations, we propose a style self-consistency loss ($\mathcal{L}_{SSC}$) for explicit style alignment from a raw image with high style variations to a standardized one with detail loss. Specifically, let the high-level features generated by the bridge layers in CoarseNet and RefineNet be denoted as H_C and H_R, respectively. Inspired by [44], we design a feature space to capture texture information as the ***style*** representations of TOF-MRA images, which were obtained by calculating the correlations between the filter responses taken over the spatial extent of the feature maps. Thus, the filter responses can be stored in a matrix $F \in \mathbb{R}^{N \times M}$. Here N is the number of filters, while M is equal to the product of the height, width, and depth of the feature maps. F_{ij} denotes the activation of the i^{th} filter at position j. Thus, the correlations of the filter responses are given by the Gram matrix $\mathbf{G}$ [45], where $\mathbf{G}_{ij}^{(H_C)}$ and $\mathbf{G}_{ij}^{(H_R)}$ are the inner product between the i-th and j-th vectorized feature maps in H_C and H_R, respectively: i.e.,

$$\mathbf{G}_{ij}^{(H_C)} = \sum_k F_{ik}^{(H_C)} F_{jk}^{(H_C)},$$
$$\mathbf{G}_{ij}^{(H_R)} = \sum_k F_{ik}^{(H_R)} F_{jk}^{(H_R)}. \tag{3}$$

The Gram matrix derived from the neural network's high-level features effectively encapsulates the image's style [44]. To address the influence of style heterogeneity on model performance, aligning varied styles to standardized representations within the high-level feature space becomes imperative for ensuring consistent feature style extraction. To align the style of the refined high-level features to that of the coarse high-level features, we optimize style matching by minimizing the mean square distance between the Gram matrices $\mathbf{G}^{(H_C)}$ and $\mathbf{G}^{(H_R)}$ of the refined high-level features and the coarse high-level features:

$$\mathcal{L}_{ssc}(H_R, H_C) = \frac{1}{4N^2M^2} \sum_{ij} \left(\mathbf{G}_{ij}^{(H_C)} - \mathbf{G}_{ij}^{(H_R)} \right)^2. \tag{4}$$

The style self-consistency loss is pivotal in transforming variable styles into a unified paradigm within the latent space, thereby enhancing both model convergence and training efficiency. This is particularly crucial for heterogeneous multi-center, multi-vendor datasets, where imposing consistency constraints on data styles markedly strengthens the model's adaptability to style heterogeneity.

D. Loss Function

Finally, we combine the deep supervision loss function with the proposed $\mathcal{L}ssc$ to jointly optimize the parameters of the proposed cerebrovascular segmentation model CESAR. The loss function of each deep supervision output is the sum of cross-entropy (L_{CE}) and Dice (L_{Dice}) loss. Mathematically, the overall loss function is defined as:

$$\mathcal{L} = \mathcal{L}_{coarse}^{(ds)} + \mathcal{L}_{refine}^{(ds)} + \mathcal{L}_{ssc}, \tag{5}$$

where $\mathcal{L}_{coarse}^{(ds)}$ and $\mathcal{L}_{refine}^{(ds)}$ represent the deep supervision loss of CoarseNet and RefineNet, respectively, i.e.,

$$\mathcal{L}_{coarse}^{(ds)} = \mathcal{L}_{refine}^{(ds)} = \sum_{i=1}^{S} \omega_i \cdot \left(L_{CE}^{(i)} + L_{Dice}^{(i)} \right), \tag{6}$$

where S represents the number of deep supervision layers, while ω_i represents the weight of the i^{th} layer. Note that we follow [46] to set the weights of the two supervised layers with the lowest resolution to zero. The weight ω_i is governed by the relationship $\omega_i = \frac{1}{2^{i-1}}\omega_1$, where $\sum_{i=1}^{S} \omega_i = 1$ and ω_1 is initialized to 1.

IV. Experimental Setup and Results

A. Data Preprocessing

1) Intensity Histogram Standardization: Cerebrovascular TOF-MRA datasets, acquired from a range of scanners by different vendors using diverse protocols at various sites, may exhibit heterogeneous intensity levels for the same tissue type. To address this concern, we employed intensity histogram standardization [47] to harmonize the intensity profiles.

2) Dynamic Voxel Spacing Resampling: Voxel spacing refers to the physical resolution of voxels, which can vary between images from different scanners and protocols. When the spacing deviates significantly from the target spacing during resampling with fixed target spacing, it can result in jagged vascular surfaces and the loss of small cerebrovascular structures. To cope with this issue, we propose a dynamic voxel spacing resampling (DyR) approach. Firstly, we determine the reference target spacing $\mathbf{T}_\delta$ using the method from [46], where δ denotes the axis index. We then use a threshold ε to ascertain the resampling spacing for each case. Specifically, for a given axis spacing ($\mathbf{O}_\delta$), we set its final target spacing ($\hat{\mathbf{T}}_\delta$) to $\mathbf{O}_\delta$ if $(1-\varepsilon)\mathbf{T}_\delta \leqslant \mathbf{O}_\delta \leqslant (1+\varepsilon)\mathbf{T}_\delta$, and to $(1-\varepsilon)\mathbf{T}_\delta$ if $\mathbf{O}_\delta \leqslant (1-\varepsilon)\mathbf{T}_\delta$; otherwise, we set it to $(1+\varepsilon)\mathbf{T}_\delta$. To ensure isotropy in the dynamically resampled image and avoid disproportionate axis spacing, we set $\varepsilon = 1/3$ in this paper.

B. Implementation Details

The proposed CESAR is implemented using PyTorch 1.19.0 on an Ubuntu 20.04 LTS operating system, powered by a single GPU (NVIDIA GeForce RTX 3090) with CUDA 11.6. We employed mini-batch stochastic gradient descent (SGD) and an initial learning rate of 0.01 to optimize the model parameters. The model was trained for 2000 epochs, with 250 mini-batch iterations for one epoch, and the learning rate was decayed according to the poly [48] strategy. Training the proposed CESAR model for one fold at 2000 epochs takes approximately 2.6 days. To improve the robustness of the training process, we used the pre-trained weights provided in [41] to initialize the Swin encoder. Due to GPU memory limitations, all the training images were randomly cropped to patches of 128 × 128 × 64 voxels before being fed to the network, where the batch size was set to 2.

To enhance reproducibility, we placed great emphasis on utilizing the nnUNet framework [46] for constructing, training, and validating the proposed model. To exclude interference by the skull and extracranial regions on cerebrovascular segmentation, skull stripping of all TOF-MRA images was performed using BET (Brain Extraction Tool) [49]. During inference, a sliding window approach was used with a window size of 128 × 128 × 64 voxels and overlapping ratio of 0.5 for the prediction of all the test TOF-MRA images. In the inference process of CESAR, RefineNet relies on the initial cerebrovascular probability map generated by CoarseNet as an additional input. Therefore, CESAR remains a unified system consisting of both CoarseNet and RefineNet components.

C. Evaluation Metrics

To assess the overall segmentation performance of cerebrovascular structures, we follow [15] to use the DICE similarity coefficient. Moreover, to evaluate the performance of cerebrovascular boundary segmentation, we also introduce 95% Hausdorff distance (HD95) and average symmetric surface distance (ASD). We define P_S and G_S as the sets of predicted cerebrovascular surface points and ground truth surface points, respectively. ASD represents the average shortest distance between the points in P_S and G_S, which is given by

$$\mathrm{ASD} = \frac{1}{|P_S| + |G_S|}\left(\sum_{p\in P_S}\min_{g\in G_S} d\left(p, g\right) + \sum_{g\in G_S}\min_{p\in P_S} d\left(g, p\right)\right),$$

where $d\left(\cdot,\cdot\right)$ is the Euclidean distance. The Hausdorff distance can be represented as:

$$\mathrm{HD} = \max\left\{\max_{p\in P_S}\min_{g\in G_S} d\left(p, g\right), \max_{g\in G_S}\min_{p\in P_S} d\left(g, p\right)\right\}.$$

Note that the HD95 is the 95^{th} percentile of the HD. The lower ASD or HD95 score the better segmentation performance. Additionally, to evaluate the topological integrity of the segmented cerebrovascular structures, we introduce the centerline Dice (clDice) coefficient [50] as the fourth evaluation metric.

D. Cerebrovascular Segmentation Results

To demonstrate the superior performance of the proposed method, we handpicked two commonly used 3D medical image segmentation baseline methods, namely 3D-UNet [51] and V-Net [52], for comparison on the COSTA dataset. To perform a thorough evaluation of the efficacy of the proposed method for cerebrovascular segmentation, we meticulously selected some state-of-the-art fully supervised methods that are specifically tailored for this task, including CS^{2}Net [17], ER-Net [20], ISA-Model [22], and the multi-task deep CNN along with a topology-aware loss function method (denoted as MtTopo) [54]. Furthermore, we conducted a performance comparison between the baseline methods SwinUNETR [53] and nnUNet [46] utilized by the proposed method to validate its exceptional performance. A recently proposed algorithm, the fast fuzzy c-means clustering and Markov random field (FFCM-MRF) optimization method [13], based on traditional feature descriptors, has also been included for comparison.

Fig. 5 illustrates the overall average performance in terms of ASD, HD95, DICE and clDice of the proposed CESAR versus the compared methods. From an overall perspective, CESAR surpasses all the methods for cerebrovascular segmentation. Specifically, CESAR achieved the lowest ASD and HD95 scores of 0.142 and 0.780, respectively. In terms of DICE and clDice, CESAR reached 0.962 and 0.962, outperforming the runner-up nnUNet (0.941 and 0.955, respectively).

Table II shows a comprehensive overview of all the methods in this study across healthy and unhealthy groups. As demonstrated in Table II, our proposed method achieves the best cerebrovascular segmentation performance in a multi-center and multi-vendor dataset scenario in healthy and unhealthy groups, respectively, outperforming all the other methods in terms of accuracy with the highest DICE scores of 0.962 and 0.956 for the healthy and unhealthy groups, respectively. Additionally, the proposed method has the lowest HD95 value (0.784 and 0.740 for healthy and unhealthy groups, respectively) among all the compared methods, indicating the smallest difference between the ground truth and the predicted segmentation boundaries. Considering the presence of lesions in unhealthy data as a challenge for cerebrovascular segmentation, the proposed method maintains optimal segmentation

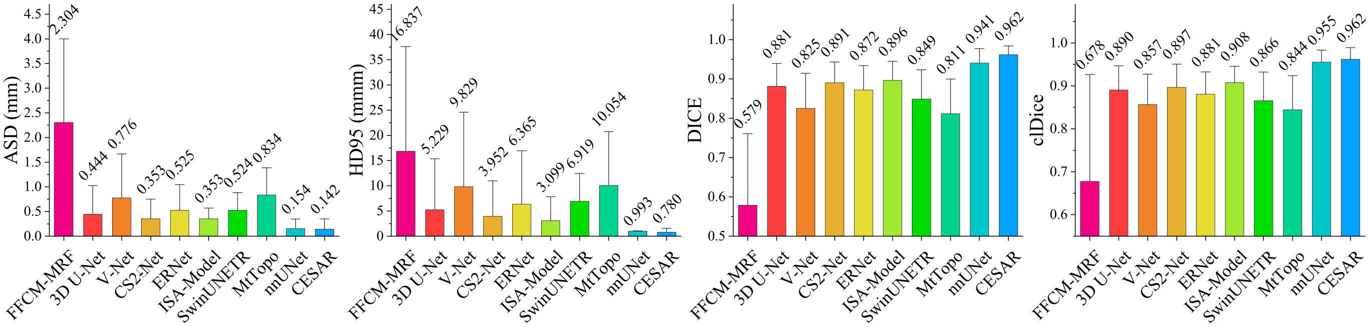

Fig. 5. Overall performance in terms of ASD, HD95, DICE and clDice of different methods on the COSTA database.

TABLE II
PERFORMANCE OF DIFFERENT METHODS FOR CEREBROVASCULAR SEGMENTATION IN HEALTHY AND UNHEALTHY GROUPS

Methods	**ASD** (*mm*) ↓		**HD95** (*mm*) ↓		**DICE** ↑		**clDice** ↑	
	Healthy	Unhealthy	Healthy	Unhealthy	Healthy	Unhealthy	Healthy	Unhealthy
FFCM-MRF [13]	2.227±1.499	3.154±1.648	16.604±8.082	19.402±6.562	0.581±0.114	0.556±0.066	0.686±0.124	0.581±0.100
3D U-Net [51]	0.440±0.287	0.488±0.281	5.209±4.230	5.453±3.674	0.883±0.060	0.866±0.039	0.891±0.046	0.880±0.048
V-Net [52]	0.793±0.363	0.592±0.304	10.212±5.404	5.626±4.319	0.823±0.067	0.856±0.041	0.855±0.046	0.873±0.040
CS^2-Net [17]	0.355±0.193	0.338±0.182	4.058±2.980	2.776±2.536	0.891±0.061	0.884±0.026	0.896±0.045	0.902±0.031
ER-Net [20]	0.523±0.252	0.544±0.289	6.464±3.958	5.278±3.660	0.874±0.059	0.852±0.043	0.882±0.046	0.864±0.040
ISA-Model [22]	0.342±0.185	0.477±0.227	2.870±3.361	5.623±3.868	0.895±0.035	0.911±0.024	0.909±0.023	0.897±0.028
SwinUNETR [53]	0.530±0.173	0.456±0.278	7.119±2.175	4.714±3.224	0.849±0.042	0.849±0.053	0.864±0.033	0.887±0.043
MtTopo [54]	0.834±0.311	0.837±0.251	10.201±5.209	8.443±2.785	0.810±0.043	0.826±0.018	0.846±0.038	0.828±0.022
nnUNet [46]	0.155±0.102	**0.155±0.073**	1.013±1.680	0.774±0.809	0.942±0.020	0.920±0.018	0.956±0.015	0.949±0.018
CESAR	**0.140±0.100**	0.166±0.064	**0.784±1.567**	**0.740±0.685**	**0.963±0.015**	**0.947±0.022**	**0.962±0.013**	**0.952±0.015**

performance, despite all the methods exhibiting a similar declining trend in terms of DICE and clDice.

Fig. 6 illustrates cerebrovascular segmentation results of different methods over representative images from each center. The degree of over-segmentation increases as the color of the segmented meshes deviates from that of the ground truth, as represented by the highlighted color other than red in each row. The effectiveness of the proposed method in segmenting cerebrovascular is further validated through the visualization of the results illustrated in Fig. 6. Specifically, the cerebrovascular meshes segmented with the proposed method demonstrate a higher degree of color consistency with the ground truth meshes, as evidenced by the reduced number of highlighted colors in each row.

Fig. 7 presents a detailed comparison of the segmentation results among Swin UNETR, nnUNet, and the proposed method. Fig. 7(a) shows the segmentation outcomes for two healthy subjects, while Fig. 7(b) showcases results for subjects with aneurysms and Moyamoya disease. It can be seen from Fig. 7(a) that the proposed method excels in segmenting both thick vessels, such as the Circle of Willis, and thin vessels. In terms of artifact generation, Swin UNETR exhibits a relatively higher frequency, while nnUNet demonstrates a comparatively lower incidence. In contrast, the proposed CESAR method shows nearly negligible artifact generation. Specifically, the surface of the segmented Circle of Willis using the proposed method closely matches the ground truth, as indicated by the green highlighting in the first row. Moreover, when it comes to segmenting tiny vessels, the proposed method exhibits superior continuity, while nnUNet produces broken vessels. Although Swin UNETR generates segments with high continuity, they are noticeably thicker than the ground truth.

In the context of cerebrovascular segmentation in subjects with medical conditions, the proposed method maintains its superior performance. In the first row of Fig. 7(b), the proposed method accurately segments the surface of the aneurysm and the surrounding cerebrovascular structures. It achieves segmented boundaries that closely align with the ground truth. In contrast, nnUNet exhibits excessive under-segmentation (highlighted in red), while Swin UNETR demonstrates over-segmentation (highlighted in green). A similar trend is observed in the second row, which focuses on cerebrovascular segmentation in subjects with Moyamoya disease.

E. Ablation Study of Key Components

In this section, we perform an ablation study to assess each component's effectiveness. We start with nnUNet as the baseline, denoted as **M0**, and introduce a coarse-to-fine architecture based on M0 (**M1**). After constructing M1, we evaluate the impact of our preprocessing strategy by applying the DyR method to preprocess the training data (**M2**). We then incorporate AutoGLFS into M2 (**M3**). To verify the effectiveness of the style self-consistency loss, we reintegrate it into M2 once more (**M4**). Finally, we integrate the proposed style self-consistency loss function into M3 and M4 (**M5**) respectively to investigate its potential to further enhance performance.

The quantitative results presented in Table III demonstrate that the performance of the baseline method can be improved by utilizing a coarse-to-fine architecture, as evidenced by

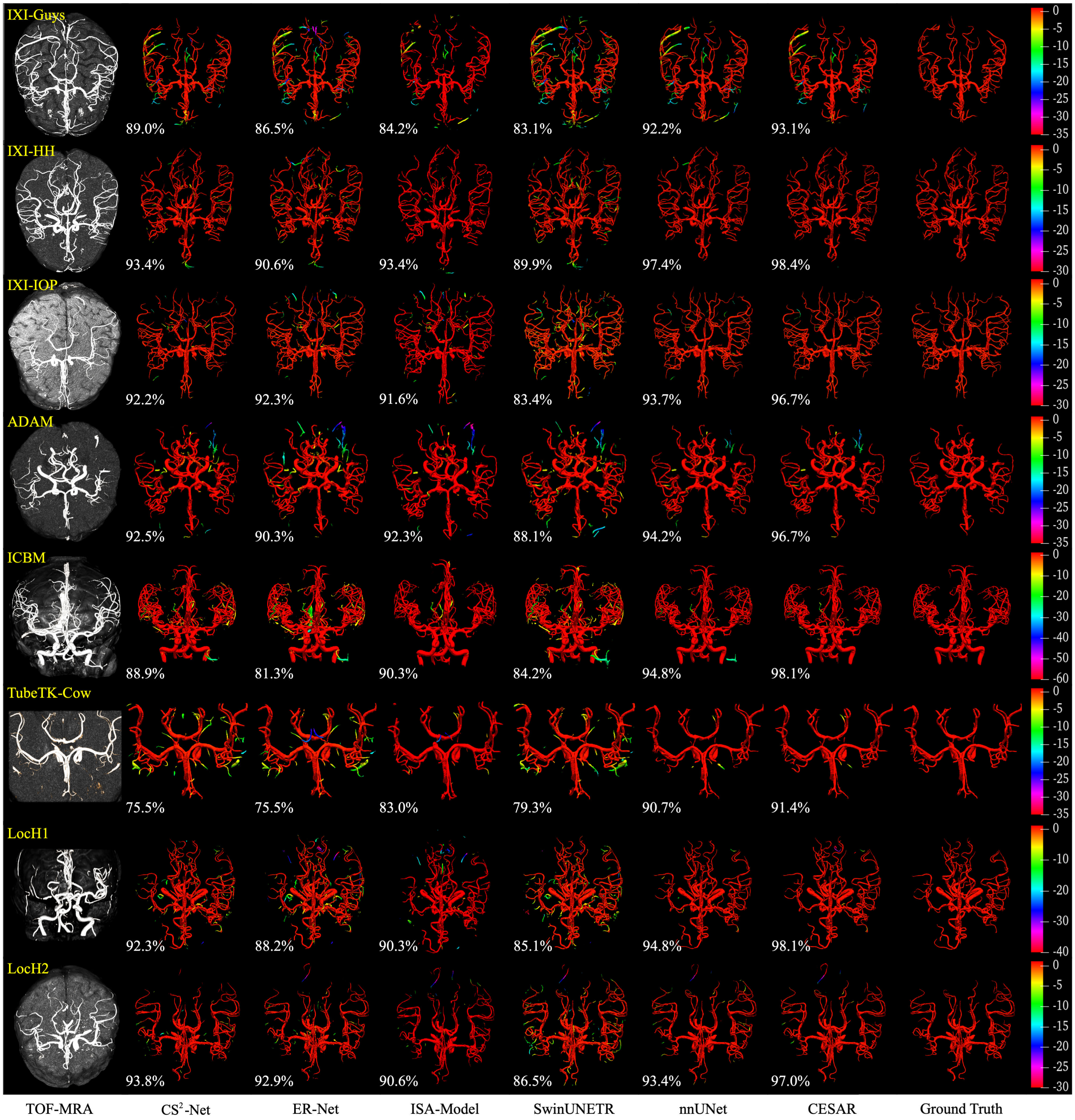

Fig. 6. The cerebrovascular segmentation results of images from various centers using different methods are illustrated in rows from top to bottom, accompanied by their respective DICE scores (%). All the segmentations are reconstructed as meshes with an identical smoothing factor and further presented as signed distance fields relative to the ground truth. The final column displays the signed distance field between the ground truth and itself, with the color proximity to the ground truth indicating the degree of similarity between the segmentation results and the actual values.

the clear improvements observed in the ASD and HD95 metrics. By incorporating DyR into M1 (i.e. M2), HD95 decreased by 6.1%, with a corresponding increase of 1.2% in DICE, despite an increase in ASD from 0.147 to 0.155. These results suggest that the proposed DyR method has the potential to improve model performance by retaining more semantic information related to the cerebrovascular features in the TOF-MRA images. By observing the quantitative results of M3 in Table III, it becomes apparent that the cerebrovascular surface segmentation performance is further improved by integrating AutoGLFS for global long-range dependencies and local semantic contextual selection. This improvement is evidenced by a decrease in ASD from 0.155 to 0.148 and HD95 from 0.908 to 0.816, which show that the HD95

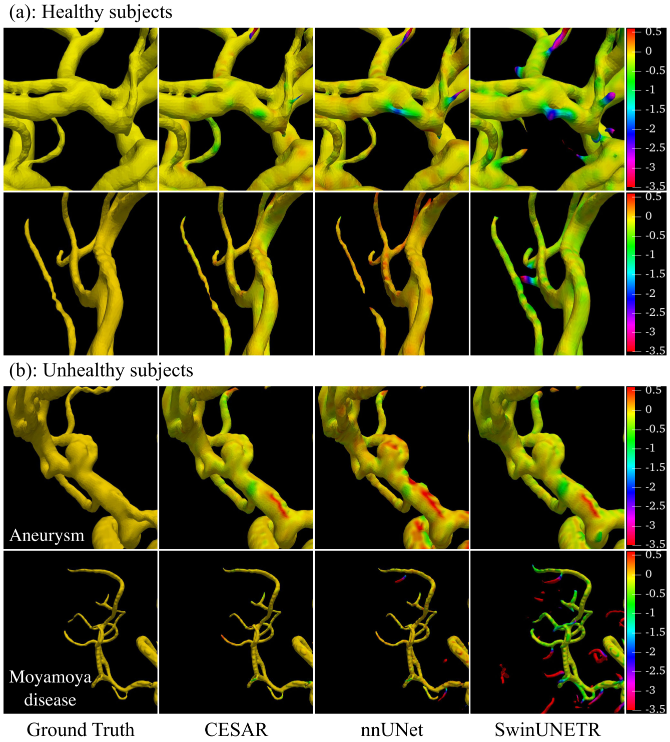

Fig. 7. Visualization of segmentation results of (a) health subjects and (b) subjects with aneurysm and Moyamoya disease. The color bar displays the signed distance field between the surface and ground truth of the segmentation outcomes.

TABLE III
ABLATION RESULTS OF VARIOUS COMPONENTS IN THE PROPOSED CESAR MODEL

	DyR	ASD ↓	HD95 ↓	DICE ↑	clDice ↑
M0	×	0.154±0.100	0.993±1.627	0.941±0.021	0.955±0.016
M1	×	0.147±0.107	0.969±1.629	0.945±0.025	0.958±0.018
M2	√	0.155±0.120	0.908±1.822	0.957±0.023	0.958±0.019
M3	√	0.148±0.108	0.816±1.599	0.956±0.018	0.959±0.015
M4	√	0.144±0.091	0.801±1.127	0.960±0.015	0.961±0.014
M5	√	**0.142±0.090**	**0.780±1.250**	**0.962±0.017**	**0.962±0.014**

value is reduced by nearly 9% compared to that of M2. The decreases in ASD and HD95 indicate that, on average, the model predictions are closer to the ground truth surface, with fewer significant discrepancies compared to the ground truth.

Upon examination of the quantitative results presented in Table III, it becomes apparent that although M3 exhibits no improvement in DICE compared to M2, it demonstrates enhancements in the other metrics such as ASD, HD95, and clDice. This validates the efficacy of the proposed AutoGLFS. Moreover, M4 showcases improvements across all the evaluation metrics relative to M2, thus further substantiating the effectiveness of the proposed style self-consistency loss. Ultimately, the amalgamation of AutoGLFS and style self-consistency loss (referred to as M5) leads to additional enhancements for cerebrovascular segmentation. These findings underscore the effectiveness and significance of both modules for enhancing the performance of cerebrovascular segmentation. Overall, the ablation results in Table III indicate that the proposed components are effective in enhancing the overall cerebrovascular segmentation performance of the model.

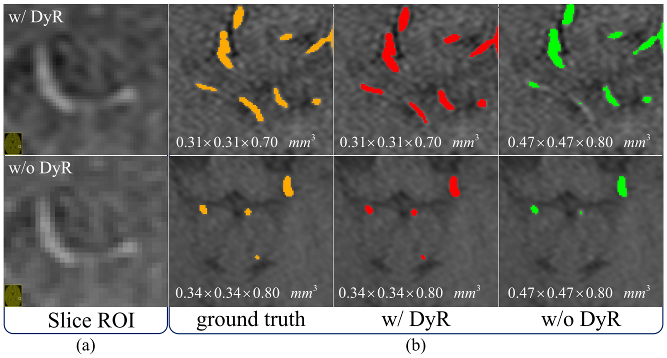

Fig. 8. Cerebrovascular segmentation results of TOF-MRA images with and without DyR, showing (a) ROI resampled with and without DyR, and (b) cerebrovascular segmentation results by CESAR in resampled images with and without DyR.

V. DISCUSSIONS

A. Effect of DyR on Segmentation Accuracy

Fixed-target spacing resampling in multi-source heterogeneous TOF-MRA images often leads to the loss of surface information in large vessels and small vessels with few voxels. To address this challenge, we propose a dynamic voxel spacing resampling (DyR) approach in this study. DyR effectively mitigates interpolation artifacts and the loss of significant information associated with fixed-target resampling. While fixed-spacing resampling in TOF-MRA images from different centers aims to standardize physical resolution, it often results in jagged vessel surface renderings, as shown in Fig. 8(a). In contrast, our DyR method offers flexibility by enabling dynamic image resampling within an effective range, ensuring accurate cerebrovascular surface representation without compromising image quality. Fig. 8(b) illustrates the effectiveness of the proposed CESAR for cerebrovascular segmentation on images resampled with and without DyR (fixed spacing). In addition, the comparison between M2 and M1 in Table III offers quantitative results for cerebrovascular segmentation with and without DyR applied to TOF-MRA images. The findings highlight the superior performance of the proposed model on images resampled with DyR (highlighted in red) compared to those without DyR (highlighted in green).

B. Boosting Generalizability via Training Set Expansion

To investigate the influence of diverse data centers and vendors on model performance, we conducted two sets of experiments, COSTA-I and COSTA-II. In COSTA-I, we used the IXI-HH dataset from COSTA for training and testing, reserving the other seven data centers for an unseen test set. In COSTA-II, we employed the IXI-HH, IXI-Guys, and IXI-IOP datasets for training and created an independent unseen test set from images of the remaining five data centers. Table IV summarizes the quantitative results from evaluating different methods in both experimental setups. According to the quantitative results on the test set within the model

TABLE IV

CEREBROVASCULAR SEGMENTATION RESULTS OF DIFFERENT METHODS ON DIFFERENT CONFIGURATIONS OF THE CONSTRUCTED COSTA DATABASE

	Methods	Testing on training centers				Testing on unseen centers			
		ASD ↓	HD95 ↓	DICE ↑	clDice ↑	ASD ↓	HD95 ↓	DICE ↑	clDice ↑
COSTA-I	FFCM-MRF [13]	1.591±0.462	13.705±3.812	0.643±0.036	0.731±0.051	2.446±1.630	17.463±8.463	0.566±0.117	0.667±0.133
	3D U-Net [51]	0.380±0.032	3.031±2.080	0.915±0.008	0.918±0.013	1.143±1.271	10.100±7.068	0.758±0.145	0.755±0.104
	V-Net [52]	0.507±0.523	7.208±1.827	0.894±0.017	0.875±0.020	2.857±1.966	19.313±8.348	0.618±0.216	0.578±0.233
	CS^2-Net [17]	0.257±0.377	3.050±1.999	0.918±0.021	0.894±0.025	1.988±1.674	16.849±9.030	0.715±0.185	0.694±0.180
	ER-Net [20]	0.375±0.404	4.859±2.725	0.909±0.020	0.899±0.024	3.915±6.717	23.623±16.12	0.541±0.285	0.496±0.341
	ISA-Model [22]	0.608±0.200	8.939±4.500	0.910±0.018	0.904±0.018	1.239±0.596	13.110±6.443	0.745±0.103	0.770±0.097
	SwinUNETR [53]	0.468±0.476	6.592±1.593	0.902±0.029	0.877±0.030	3.110±2.118	19.772±10.082	0.626±0.172	0.588±0.218
	MtTopo [54]	1.086±1.338	9.273±8.413	0.755±0.144	0.793±0.161	1.251±1.408	10.536±8.656	0.724±0.139	0.768±0.166
	nnUNet [46]	0.146±0.130	0.145±0.263	0.961±0.007	0.960±0.008	**0.643±0.399**	**6.583±6.177**	0.780±0.081	**0.878±0.057**
	CESAR	**0.119±0.104**	**0.078±0.175**	**0.964±0.007**	**0.963±0.006**	0.703±0.561	7.624±7.190	**0.792±0.080**	0.865±0.061
COSTA-II	FFCM-MRF [13]	2.051±1.147	16.832±5.801	0.596±0.095	0.694±0.123	2.530±1.780	16.842±9.551	0.563±0.122	0.663±0.127
	3D U-Net [51]	0.394±0.453	5.328±2.778	0.880±0.030	0.901±0.024	0.877±0.504	9.312±5.295	0.785±0.101	0.791±0.085
	V-Net [52]	0.705±0.768	9.553±3.330	0.848±0.021	0.862±0.023	1.374±0.821	13.856±5.702	0.751±0.108	0.749±0.090
	CS^2-Net [17]	0.395±0.172	5.440±2.837	0.903±0.134	0.897±0.021	1.174±0.921	11.447±6.521	0.767±0.099	0.759±0.083
	ER-Net [20]	0.355±0.158	4.718±2.637	0.906±0.016	0.905±0.019	1.108±0.921	10.614±6.460	0.762±0.093	0.759±0.080
	ISA-Model [22]	0.501±0.193	5.682±3.594	0.893±0.052	0.898±0.037	0.912±0.400	8.485±3.003	0.772±0.067	0.771±0.093
	SwinUNETR [53]	0.491±0.140	7.034±1.979	0.874±0.027	0.880±0.018	1.408±1.488	14.546±12.032	0.779±0.097	0.785±0.071
	MtTopo [54]	0.647±0.455	7.110±4.289	0.818±0.074	0.884±0.028	0.863±0.505	9.237±3.969	0.775±0.072	0.812±0.093
	nnUNet [46]	0.153±0.183	0.876±1.500	0.945±0.019	0.947±0.010	0.501±0.225	**4.747±3.552**	0.821±0.039	0.891±0.043
	CESAR	**0.148±0.187**	**0.761±1.678**	**0.961 ± 0.009**	**0.963±0.010**	**0.474±0.512**	5.014±5.331	**0.843±0.040**	**0.897±0.038**

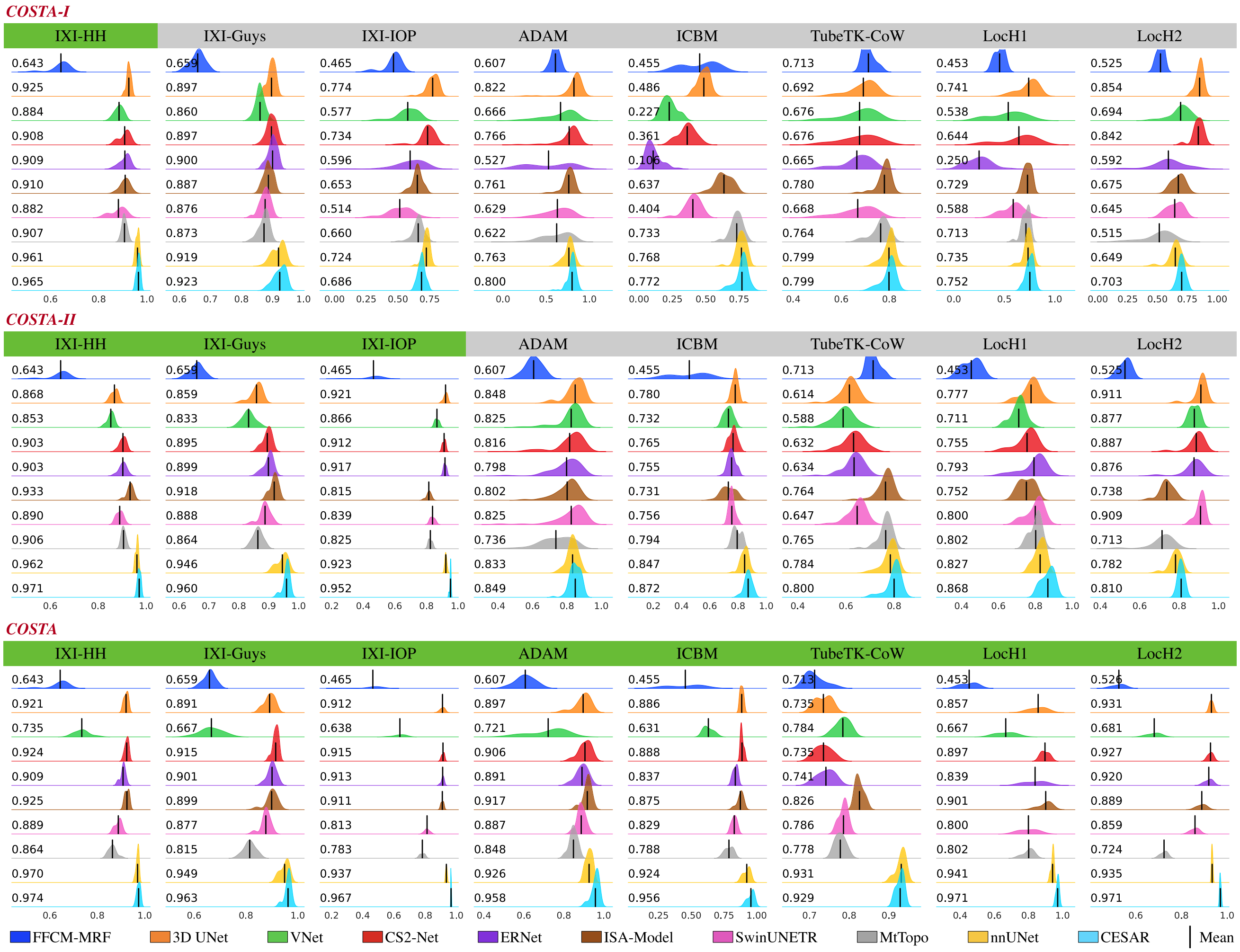

Fig. 9. The distribution of DICE scores for different methods trained at different centers and evaluated separately on both seen and unseen test centers. The performance of FFCM-MRF [13], 3D U-Net [51], V-Net [52], CS^2-Net [17], ER-Net [20], ISA-Model [22], SwinUNETR [53], MtTopo [54], nnUNet [46], and the proposed CESAR in terms of DICE is demonstrated in each set of configurations, from top to bottom. Datasets with green backgrounds denote training centers, while those with gray backgrounds signify unseen centers.

training centers, our method outperforms the others in both COSTA-I and COSTA-II configurations. However, the DyR strategy used in COSTA-I was ineffective since it involved images from the IXI-HH dataset with identical voxel spacing.

Nonetheless, our AutoGLFS and style self-consistency loss function exhibits the potential to enhance segmentation accuracy, as indicated by a higher DICE score of 0.792. In the COSTA-II study, the incorporation of additional training data necessitates the expansion of the test set, which subsequently includes a broader spectrum of multi-center and multi-vendor data characterized by inherent heterogeneity. This increase in data diversity and complexity introduces greater variability, posing additional challenges for the model. As a result, the model experiences a slight degradation in overall performance. This highlights the importance of developing robust methods capable of effectively handling the increased variability and complexity associated with heterogeneous datasets.

Fig. 9 offers a comprehensive visualization of the DICE scores of various methods across all the data centers under both the COSTA-I and COSTA-II configurations. Green-highlighted bars correspond to training data centers, while gray bars represent unseen centers. Comparative analysis of DICE scores across different centers in COSTA-I reveals that the proposed CESAR shows only slight enhancements than existing state-of-the-art methods over IXI-Guys, ICBM, MIDSA-CoW, and LocH1 where training data is limited. However, under COSTA-II with an increased training dataset, CESAR outperforms the state-of-the-art methods on all the unseen data centers except for LocH2. Remarkably, CESAR achieves substantial improvements of 4.5%, 10.02%, and 11.61% in DICE scores for ADAM, ICBM, and LocH1 data centers, respectively.

The primary objective of assessing unseen centers is to gauge the models' generalization performance. In our investigation, all the methods showed a notable drop in performance when tested on specific COSTA-I. In COSTA-II, with a more extensive training dataset, we expect significant improvements in generalization across all the models, with CESAR showing the best performance. This underscores the role of diversified training samples in enabling models to better understand cerebrovascular structures in TOF-MRA images, enhancing their generalizability and stability. Additionally, joint analysis of quantitative results in Table II indicates that the performance of CESAR improves as more TOF-MRA images are included in the training set, reaching comparable levels as seen in COSTA-I and COSTA-II training centers. In summary, these findings emphasize that maximizing training data volume is an effective strategy for improving models' generalization capability.

Upon examination of Table IV, a discernible decline in the performance of the cerebrovascular segmentation model becomes evident when subjected to an increase in training centers. Theoretically, an augmented number of training centers, concomitant with an increase in training samples, typically augurs an elevation in model performance. However, within this study, the augmentation of training centers coincides with a proportional escalation in domain gaps within the dataset, signifying an amplification in data heterogeneity. This proliferation of domain gaps significantly undermines the model's generalizability, consequently leading to the observed decrease in performance. Nevertheless, in comparison to existing methods, our method remains adept at maintaining optimal performance, demonstrating superior resilience against domain gap occurrence, as corroborated by the findings in Tables II and IV.

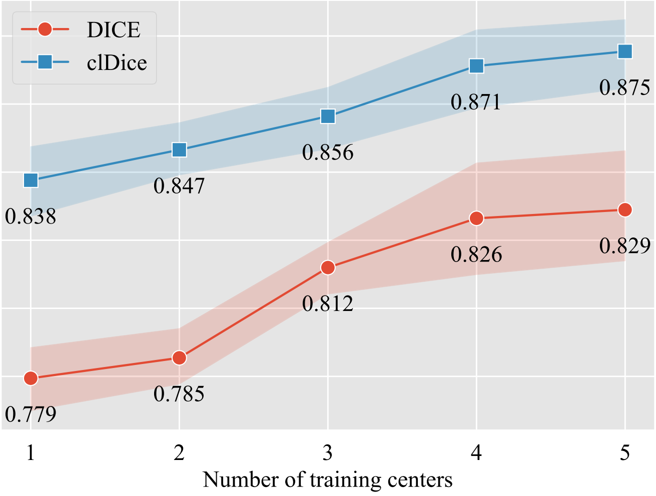

Fig. 10. The performance of CESAR on unseen centers, measured in terms of DICE and clDice, demonstrates a notable improvement with the incremental increase in the number of training centers.

To ascertain the efficacy of expanding training centers in further enhancing CESAR's generalization capability, we conducted supplementary validation experiments to assess its generalization performance relative to an escalating number of training centers (designated as T_N). Specifically, we curated three distinct subsets from the COSTA database—namely TubeTK-CoW, LocH1, and LocH2—as independent, unseen validation sets. Commencing with $T_N = 1$, we systematically augmented the number of training centers. Within this experimental framework, we executed 5-fold intra-centers cross-validation and K-fold inter-centers cross-validation, where $K = C_5^{T_N}$. All quantitative findings are depicted in Fig. 10. As demonstrated in Fig. 10, it is evident that the performance of CESAR on previously unseen centers demonstrates a gradual enhancement as the number of training centers is incrementally augmented. These quantitative findings suggest that with the expansion of training centers, CESAR exhibits an elevated standard of segmenting cerebral vasculature within the unseen center data, as evidenced by higher DICE scores obtained. Moreover, its topological integrity is also notably improved, as indicated by higher clDice scores. The aforementioned series of validations implies that enhancing model generalization through the expansion of the training set proves to be an effective strategy. Furthermore, it underscores that the proposed CESAR model exhibits superior capacity in learning from multicenter data despite existing domain gaps, thereby achieving state-of-the-art segmentation performance.

C. COSTA: Advancing Domain Generalization Catalyst

Insufficient training data can significantly impact a model's performance, necessitating the expansion of the dataset. However, data augmentation may not always be a feasible solution. On the other hand, our proposed method involves training a cerebrovascular segmentation model on a diverse dataset

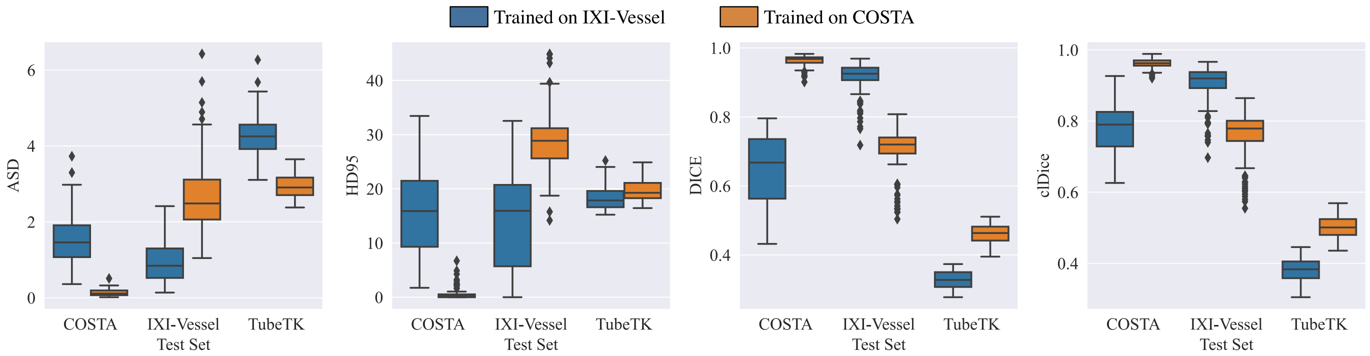

Fig. 11. The performance in terms of ASD, HD95, DICE, and clDice of the proposed method, trained on IXI-Vessel [32] and COSTA, respectively, and tested on the third-party independent unseen dataset, TubeTK-42 [33].

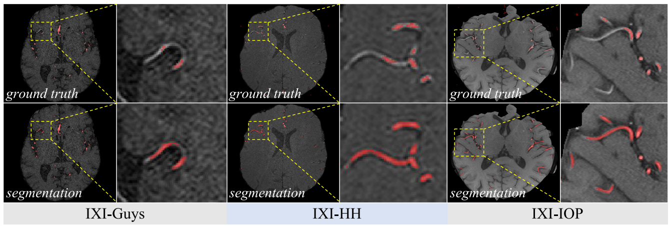

Fig. 12. Segmented cerebrovascular structures compared to the original ground truth for the CESAR model trained on the COSTA dataset, but tested on the IXI-Vessel [32] dataset.

from multiple centers and vendors, which has the potential to improve generalization. Given the limited availability of high-quality annotated images in our study, it is prudent to systematically explore and validate multi-center cerebrovascular segmentation techniques that leverage domain generalization and adaptation methods. Alternatively, developing innovative approaches for domain generalization and adaptation offers promise for enhancing the model's performance and increasing its effectiveness in real-world scenarios.

To better highlight the ability of the constructed COSTA dataset to facilitate domain generalization, we constructed a new set of validation scenarios in this section. Specifically, we integrated a larger publicly available dataset of cerebrovascular segmentation, IXI-Vessel [32], composed of TOF-MRA images, for model training (with a training and validation ratio of 4:1). We then juxtaposed its performance with that of a model trained solely on COSTA for cerebrovascular segmentation. The IXI-Vessel dataset encompasses 570 cases of cerebrovascular annotations, sourced from the IXI-Database, which comprises TOF-MRA images from three distinct data centers. Additionally, to scrutinize the generalization performance of cerebrovascular segmentation, we introduced the TubueTK-42 [33] dataset as an independent unseen test set for quantitative performance assessment. To this end, we first applied the model trained on the COSTA dataset to predict cerebrovascular structures in TOF-MRA images from the COSTA, IXI-Vessel, and TubeTK-42 datasets. Subsequently, we deployed the model trained exclusively on the IXI-Vessel dataset to make predictions on the same datasets. The quantitative results of these experiments are illustrated in Fig. 11.

From Fig. 11, it can be seen that the cerebrovascular segmentation performance of the model trained on the IXI-Vessel dataset is inferior to that of the model trained on the COSTA dataset across various metrics, including ASD, HD95, DICE, and clDice, particularly when evaluated on TubeTK-42. Two primary factors may account for this outcome. Firstly, although the COSTA dataset comprises fewer TOF-MRA samples (423 cases) compared to the IXI-Vessel dataset (570 cases), it benefits from a larger number of centers. Therefore, the COSTA dataset demonstrates increased levels of heterogeneity and diversity, effectively bolstering the resilience of models trained on COSTA against the detrimental impacts of data heterogeneity. Secondly, as detailed in Fig. 1, the labeling quality of the IXI-Vessel dataset is suboptimal, with some small cerebrovascular structures left unlabeled. Consequently, the model encounters challenges in adequately capturing these intricate vascular features, resulting in diminished generalization performance.

Upon examination of Fig. 11, it becomes apparent that the cerebrovascular segmentation performance of our model, trained on COSTA and tested on the IXI-Vessel dataset, is inferior to that of the model trained directly on IXI-Vessel. This discrepancy may stem from the fact that the manual annotations in the IXI-Vessel dataset did not encompass the small cerebrovascular structures, which the CESAR model was able to recognize. As demonstrated in Fig. 12, we presented the segmentation outcomes of CESAR across three distinct centers within the IXI-Vessel dataset alongside their corresponding manual annotations. Notably, CESAR exhibited superior segmentation of small cerebrovascular structures that were not accounted for in the manual annotations. Consequently, the CESAR model trained on the IXI-Vessel dataset exhibited suboptimal performance in terms of ASD, HD95, DICE, and clDice metrics.

In summary, the COSTA dataset constructed in this study demonstrates considerable potential to enhance model generalizability. Furthermore, to improve the model's generalization, there are two promising avenues to explore: 1) *Single-Source-Based Generalization*: This strategy entails training the model on one source and evaluating its performance on another source, with the assumption that the model can learn transferable features. Prior research [55] has demonstrated encouraging results using this strategy. 2) *Multi-Source-Based Generalization*: In this strategy, the model is simultaneously trained on multiple sources, employing joint optimization to

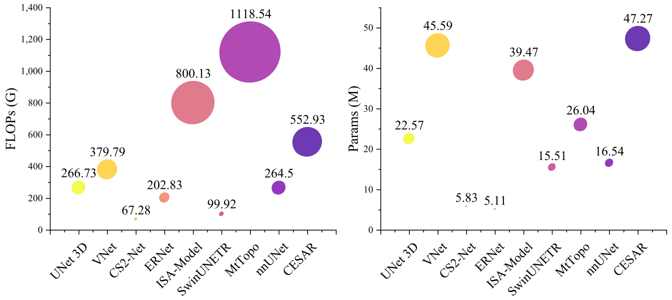

Fig. 13. Comparative quantitative analysis of network parameters and computational effort between the proposed method and other methods.

learn a shared representation capable of generalization across all the sources.

VI. Conclusion and Future Works

In this work, we took the initiative to construct the COSTA database, a multi-center, multi-vendor TOF-MRA database, meticulously annotated for quality. However, segmentation methods can be affected by TOF-MRA imaging style, especially in multi-center, multi-vendor settings. To tackle this challenge, we propose CESAR, a deep learning-based cerebrovascular segmentation network. CESAR achieves state-of-the-art segmentation performance using a dataset of 423 TOF-MRA images from eight centers and four vendors, making it suitable for multi-center and multi-vendor scenarios.

However, our proposed method has some limitations, such as the prerequisite for skull stripping on TOF-MRA images before segmentation. In Fig. 13, it is evident that the proposed method exhibits the highest number of network parameters (47.27M[6]) compared to the other methods investigated in this study, consequently leading to relatively lower network efficiency. Moreover, the computational complexity of the proposed method is measured at 552.93G FLOPs, resulting in relatively time-consuming training and inference processes. Hence, the identified low network efficiency stands as a drawback of the proposed method.

In future research, we intend to augment our dataset with high-quality TOF-MRA images to build more robust cerebrovascular segmentation models. This will support the clinical deployment of CESAR and similar models. Additionally, we plan to explore solutions for CESAR's limitations, including reducing model parameters and computational complexity for larger patch resolution and optimizing training and inference speed. In summary, this study lays the groundwork for the development of more accurate and efficient cerebrovascular segmentation methods in clinical practice.

References

[1] M. Bernier, S. C. Cunnane, and K. Whittingstall, "The morphology of the human cerebrovascular system," *Hum. Brain Mapping*, vol. 39, no. 12, pp. 4962–4975, Dec. 2018.

[2] C. G. Drake, "Progress in cerebrovascular disease. Management of cerebral aneurysm," *Stroke*, vol. 12, no. 3, pp. 273–283, May 1981.

[3] I. G. Fleetwood and G. K. Steinberg, "Arteriovenous malformations," *Lancet*, vol. 359, no. 9309, pp. 863–873, 2002.

[4] K. Hasuo, F. Mihara, and T. Matsushima, "MRI and MR angiography in moyamoya disease," *J. Magn. Reson. Imag.*, vol. 8, no. 4, pp. 762–766, Jul. 1998.

[5] T. Hashimura, T. Kimura, and T. Miyakawa, "Morphological changes of blood vessels in the brain with Alzheimer's disease," *Psychiatry Clin. Neurosci.*, vol. 45, no. 3, pp. 661–665, Sep. 1991.

[6] J. Zhu, S. Teolis, N. Biassou, A. Tabb, P.-E. Jabin, and O. Lavi, "Tracking the adaptation and compensation processes of Patients' brain arterial network to an evolving glioblastoma," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 488–501, Jan. 2022.

[7] C. Majoie et al., "MR angiography at 3T versus digital subtraction angiography in the follow-up of intracranial aneurysms treated with detachable coils," *Amer. J. Neuroradiol.*, vol. 26, no. 6, pp. 1349–1356, 2005.

[8] A. Boulin and L. Pierot, "Follow-up of intracranial aneurysms treated with detachable coils: Comparison of gadolinium-enhanced 3D time-of-flight MR angiography and digital subtraction angiography," *Radiology*, vol. 219, no. 1, pp. 108–113, Apr. 2001.

[9] R. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 1998, pp. 130–137.

[10] M. S. Hassouna, A. A. Farag, S. Hushek, and T. Moriarty, "Cerebrovascular segmentation from TOF using stochastic models," *Med. Image Anal.*, vol. 10, no. 1, pp. 2–18, Feb. 2006.

[11] N. D. Forkert et al., "3D cerebrovascular segmentation combining fuzzy vessel enhancement and level-sets with anisotropic energy weights," *Magn. Reson. Imag.*, vol. 31, no. 2, pp. 262–271, Feb. 2013.

[12] Y. Zhao et al., "Automatic 2-D/3-D vessel enhancement in multiple modality images using a weighted symmetry filter," *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 438–450, Feb. 2018.

[13] Y. Cui et al., "FFCM-MRF: An accurate and generalizable cerebrovascular segmentation pipeline for humans and rhesus monkeys based on TOF-MRA," *Comput. Biol. Med.*, vol. 170, Mar. 2024, Art. no. 107996.

[14] X. Guo et al., "Cerebrovascular segmentation from TOF-MRA based on multiple-U-net with focal loss function," *Comput. Methods Programs Biomed.*, vol. 202, Apr. 2021, Art. no. 105998.

[15] F. Fu et al., "Rapid vessel segmentation and reconstruction of head and neck angiograms using 3D convolutional neural network," *Nature Commun.*, vol. 11, no. 1, pp. 1–12, Sep. 2020.

[16] Y. Wang et al., "JointVesselNet: Joint volume-projection convolutional embedding networks for 3D cerebrovascular segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 106–116.

[17] L. Mou et al., "CS2-Net: Deep learning segmentation of curvilinear structures in medical imaging," *Med. Image Anal.*, vol. 67, Jan. 2021, Art. no. 101874.

[18] G. Tetteh et al., "DeepVesselNet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-D angiographic volumes," *Frontiers Neurosci.*, vol. 14, p. 1285, Dec. 2020.

[19] A. Hilbert et al., "BRAVE-NET: Fully automated arterial brain vessel segmentation in patients with cerebrovascular disease," *Frontiers Artif. Intell.*, vol. 3, p. 78, Sep. 2020.

[20] L. Xia et al., "3D vessel-like structure segmentation in medical images by an edge-reinforced network," *Med. Image Anal.*, vol. 82, Nov. 2022, Art. no. 102581.

[21] Y. Chen, D. Jin, B. Guo, and X. Bai, "Attention-assisted adversarial model for cerebrovascular segmentation in 3D TOF-MRA volumes," *IEEE Trans. Med. Imag.*, vol. 41, no. 12, pp. 3520–3532, Dec. 2022.

[22] C. Chen, K. Zhou, T. Lu, H. Ning, and R. Xiao, "Integration- and separation-aware adversarial model for cerebrovascular segmentation from TOF-MRA," *Comput. Methods Programs Biomed.*, vol. 233, May 2023, Art. no. 107475.

[23] C. Chen, K. Zhou, Z. Wang, and R. Xiao, "Generative consistency for semi-supervised cerebrovascular segmentation from TOF-MRA," *IEEE Trans. Med. Imag.*, vol. 42, no. 2, pp. 346–353, Feb. 2023.

[24] S. Fan, Y. Bian, H. Chen, Y. Kang, Q. Yang, and T. Tan, "Unsupervised cerebrovascular segmentation of TOF-MRA images based on deep neural network and hidden Markov random field model," *Frontiers Neuroinform.*, vol. 13, p. 77, Jan. 2020.

[25] F. Zhao et al., "Semi-supervised cerebrovascular segmentation by hierarchical convolutional neural network," *IEEE Access*, vol. 6, pp. 67841–67852, 2018.

[6] Network parameter and computational complexity quantization were performed using the THOP (https://github.com/Lyken17/pytorch-OpCounter) library.

[26] Z. Chen et al., "Generative adversarial network based cerebrovascular segmentation for time-of-flight magnetic resonance angiography image," *Neurocomputing*, vol. 488, pp. 657–668, Jun. 2022.

[27] J. Su, H. Shen, L. Peng, and D. Hu, "Few-shot domain-adaptive anomaly detection for cross-site brain images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 3, pp. 1819–1835, Nov. 2021.

[28] C. You et al., "Incremental learning meets transfer learning: Application to multi-site prostate MRIsegmentation," in *Proc. Int. Workshop Distrib., Collaborative, Federated Learn.* Cham, Switzerland: Springer, 2022, pp. 3–16.

[29] C. You, Y. Zhou, R. Zhao, L. Staib, and J. S. Duncan, "SimCVD: Simple contrastive voxel-wise representation distillation for semi-supervised medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 41, no. 9, pp. 2228–2237, Sep. 2022.

[30] C. You et al., "Rethinking semi-supervised medical image segmentation: A variance-reduction perspective," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 36, 2024, pp. 1–38.

[31] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2962–2971.

[32] Ž. Bizjak, A. Chien, I. Burnik, and Ž. Špiclin, "Novel dataset and evaluation of state-of-the-art vessel segmentation methods," in *Proc. Med. Imag., Image Process.*, vol. 11596, Apr. 2022, p. 102.

[33] E. Bullitt et al., "Vessel tortuosity and brain tumor malignancy," *Acad. Radiol.*, vol. 12, no. 10, pp. 1232–1240, Oct. 2005.

[34] A. Fedorov et al., "3D slicer as an image computing platform for the quantitative imaging network," *Magn. Reson. Imag.*, vol. 30, no. 9, pp. 1323–1341, Nov. 2012.

[35] G. Pavlakos, X. Zhou, K. G. Derpanis, and K. Daniilidis, "Coarse-to-fine volumetric prediction for single-image 3D human pose," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7025–7034.

[36] J. Jin, J. Hou, J. Chen, H. Zeng, S. Kwong, and J. Yu, "Deep coarse-to-fine dense light field reconstruction with flexible sampling and geometry-aware fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 4, pp. 1819–1836, Apr. 2022.

[37] Y. Zheng, D. Huang, S. Liu, and Y. Wang, "Cross-domain object detection through coarse-to-fine feature adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 13766–13775.

[38] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[39] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[40] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10012–10022.

[41] Y. Tang et al., "Self-supervised pre-training of Swin transformers for 3D medical image analysis," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2022, pp. 20730–20740.

[42] X. Li, W. Wang, X. Hu, and J. Yang, "Selective kernel networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 510–519.

[43] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[44] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2414–2423.

[45] L. Gatys, A. S. Ecker, and M. Bethge, "Texture synthesis using convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–9.

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

[47] L. G. Nyul, J. K. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Trans. Med. Imag.*, vol. 19, no. 2, pp. 143–150, Feb. 2000.

[48] L. C. Chen, G. Papandreou, and I. Kokkinos, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Jun. 2017.

[49] M. Jenkinson et al., "BET2: MR-based estimation of brain, skull and scalp surfaces," in *Proc. 11th Annu. Meeting Org. Hum. Brain Mapping*. Toronto, ON, Canada, vol. 17, 2005, p. 167.

[50] S. Shit et al., "ClDice—A novel topology-preserving loss function for tubular structure segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 16555–16564.

[51] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Interv.* Cham, Switzerland: Springer, 2016, pp. 424–432.

[52] F. Milletari, N. Navab, and S. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (DV)*, Oct. 2016, pp. 565–571.

[53] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images," in *Proc. 7th Int. Brainlesion Glioma, Multiple Scler., Stroke Trauma. Brain Inj.*, 2022, pp. 272–284.

[54] S. Banerjee, D. Toumpanakis, A. K. Dhara, J. Wikström, and R. Strand, "Topology-aware learning for volumetric cerebrovascular segmentation," in *Proc. IEEE 19th Int. Symp. Biomed. Imag. (ISBI)*, Mar. 2022, pp. 1–4.

[55] C. Ouyang et al., "Causality-inspired single-source domain generalization for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 42, no. 4, pp. 1095–1106, Apr. 2023.