SEMI-SUPERVISED CEREBROVASCULAR SEGMENTATION USING TOF-MRA IMAGES BASED ON LABEL REFINEMENT AND CONSISTENCY REGULARIZATION

Haibin Huang[1,2], Yue Cui[1,2], Mingxia Shi[1], Shan Yu[1,2]
[1]Laboratory of Brain Atlas and Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Sciences, Beijing, China, [2]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China

ABSTRACT

Accurate segmentation of cerebrovascular structures is crucial for scientific research and clinical applications. However, manual labeling of the whole brain's sophisticated and complex vasculature network is costly and limited, and could potentially compromise the performance and generalizability of supervised model which solely relies on high-quality labels. Semi-supervised strategies have been investigated to effectively take advantage of abundant unlabeled data. In this study, we propose a novel confident learning-based mean-teacher framework (CL-MT), which integrates noisy label refinement to alleviate the adverse effects of label noise and consistency regularization tailored for noisy labeled regions to learn useful representations from unlabeled data. In addition, we propose a backbone model UST-Net, which incorporates convolution and Transformer in both the encoder and decoder. This architecture enables the model to capture long-range dependencies at various scales. Comprehensive experiments demonstrated that our model outperformed state-of-the-art supervised and semi-supervised methods and can be generalized to diverse human and non-human primate datasets.

Index Terms— Cerebrovascular segmentation, TOF-MRA, semi-supervised learning, confident learning, Transformer

1. INTRODUCTION

Cerebrovascular system plays a vital role in delivering blood, oxygen and nutrients to the brain. Abnormalities within this system, such as developmental defects, vascular stenosis, duplications, and deficiencies, can lead to various cerebrovascular diseases that impact brain functionality [1].

Non-invasive medical imaging techniques, such as time-of-flight magnetic resonance angiography (TOF-MRA), can capture the variation of cerebral vessels without contrast agents. Deep learning techniques such as 3D U-Net models have shown promising results in TOF-MRA based cerebrovascular segmentation task [2-4]. Nevertheless, supervised learning heavily relies on large amounts of high-quality labeled data. Manual labeling of the whole brain's sophisticated and complex vasculature network is costly and limited, and thus could potentially compromise the performance and generalizability of supervised model for vessel segmentation.

In contrast, semi-supervised learning takes the advantage of a large number of unlabeled data, focusing on two main approaches: pseudo-labeling and consistency regularization. Pseudo-labeling involves generating pseudo labels, which are combined with original labeled data for training [5]. However, given limited TOF-MRA labels, ensuring the reliability of voxel-wise pseudo labels is particular challenging, especially when overfitting occurs, model may confidently generate incorrect pseudo labels. Consistency regularization, often used in mean-teacher model, encourages consistency between student and teacher model's outputs [6]. For instance, some studies proposed that consistency loss should calculate on reliable predictions [7]. GCS introduced reconstruction consistency to generate input images to ensure robustness [8]. In addition, some studies have investigated consistency regularization from the perspectives of co-training [9] and interpolation consistency [10]. However, little attention has been paid to consistency target selection, which often has an impact on model's performance [7].

In this paper, we propose a novel semi-supervised model called confident learning-based mean-teacher framework (CL-MT). Our framework builds upon mean-teacher architecture, along with synergistic noisy label refinement and ambiguity-guided consistency regularization modules. We adapt the confident learning to refine noisy labels [11], thereby reducing mislabeled voxels and providing more accurate supervision for training. Furthermore, we preserved predictions on noisy regions characterized by confident learning for consistency loss, which encouraged perturbed stability in these ambiguous-yet-informative regions, driving the model to learn useful representations from unlabeled data. We conducted comprehensive experiments on heterogeneous multi-center datasets comprising images from different scanners for both humans and non-human primates. Compared to state-of-the-art supervised and semi-supervised methods, our model achieved superior segmentation performance and generalizability.

2. METHODOLOGY

2.1. Model overview

The proposed cerebrovascular segmentation model CL-MT was designed to robustly learn from a small amount of high-quality labeled data (Set-HQ) and a large amount of low-quality noisy labeled data (Set-LQ), as illustrated in **Fig. 1**.

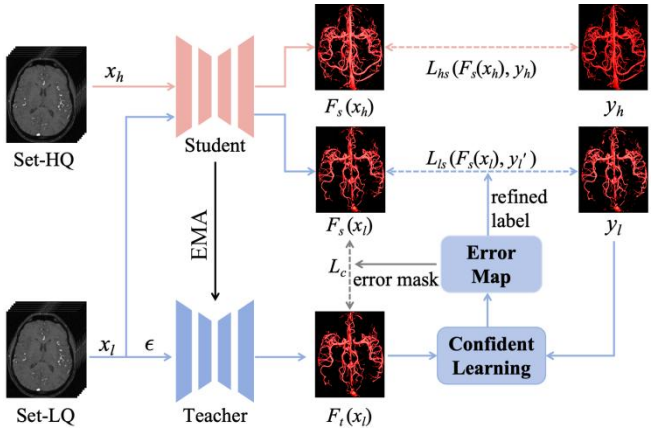

Fig. 1. The architecture of proposed confident learning-based mean-teacher (CL-MT) framework.

The CL-MT framework integrates a label refinement module for noisy labels in Set-LQ and a consistency regularization module for perturbed Set-LQ data. The student model is optimized by minimizing the supervised loss L_{hs} on Set-HQ, calibrated supervised loss L_{ls} on Set-LQ for refined labels, and ambiguity-guided consistency loss L_c between the predictions of student and teacher models within error mask.

2.2 Confident learning-based noisy label refinement

To alleviate potential misleading guidance from low-quality noisy labels, we adapted a confident learning-based label refinement module which aims to use the noisy labels while minimizing the impact of label noise. Based on the assumption of the classification noise process [12], label noise should be minimized in accordance with class probability. Confidence learning [11], originally developed for image-level classification tasks, has demonstrated its effectiveness in categorizing and quantifying labeling errors. In our context, we adapted confidence learning for teacher model's output and use it to guide the student model. Specifically, given a voxel x with observed noisy label y_l and latent true label y_l^*, the out-of-sample predicted probabilities p can be obtained via the teacher model. If voxel x with $y_l = i$ has high $p_j(x)$, i.e., $p_j(x) >= t_j$, the true label y_l^* of x should be j instead of i. Here, the threshold t_j can be obtained by calculating expected predicted probabilities $p_j(x)$ of the voxels labeled with j. t_j can be formulated as:

$$t_j = \frac{1}{|\mathbf{X}_{y_l=j}|} \sum_{x \in \mathbf{X}_{y_l=j}} p_j(x) \tag{1}$$

According to this threshold, the confidence joint matrix is defined by:

$$\mathbf{C}_{y_l,y_l^*}[i][j] := \left|\hat{\mathbf{X}}_{y_l=i,y_l^*=j}\right| \tag{2}$$

where

$$\hat{\mathbf{X}}_{y_l=i,y_l^*=j} := \left\{ x \in \mathbf{X}_{y_l=i} : p_j(x) \geq t_j, j = \underset{c \in \mathcal{C}_m : p_c(x) \geq t_c}{\arg\max}\ p_c(x) \right\} \tag{3}$$

With the confidence joint matrix, we can estimate the joint distribution matrix by:

$$\mathbf{Q}_{y_l,y_l^*}[i][j] = \frac{\frac{\mathbf{C}_{y_l,y_l^*}[i][j]}{\sum_{j \in M} \mathbf{C}_{y_l,y_l^*}[i][j]} \cdot |\mathbf{X}_{y_l=i}|}{\sum_{i \in M, j \in M} \left(\frac{\mathbf{C}_{y_l,y_l^*}[i][j]}{\sum_{j \in M} \mathbf{C}_{y_l,y_l^*}[i][j]} \cdot |\mathbf{X}_{y_l=i}| \right)} \tag{4}$$

where M denotes the number of classes. Next, we can select the noise X_{err} with lowest confidence by using the joint distribution matrix and the prune-by-class strategy:

$$X_{\text{err}} = N \cdot \sum_{j \in C: j \neq i} \left(Q_{y_l,y_l^*}[i][j] \right) \tag{5}$$

where N denotes the number of voxels in the image X, '1' in X_{err} means the voxel is wrongly labeled. By utilizing error map X_{err}, the refined label y_l' is calculated by:

$$y_l' = y_l + X_{err} \cdot (-1)^{y_l} \tag{6}$$

2.3. Ambiguity-guided consistency regularization

Consistency regularization is a commonly used strategy that aims to encourage consistent predictions for the same input under different perturbations. However, research on the selection of consistency targets is still limited. One related study is UA-MT [7], which selects reliable predictions with lower uncertainty from teacher model when calculating the consistency loss. In contrast, AC-MT [13] argues that low-uncertainty regions contain less valuable information and have less prediction volatility, while high-uncertainty regions contain more informative clues. Encouraging consensus in these ambiguous-yet-informative regions may be worthwhile.

In our model, the ideas of AC-MT were adopted. We take the foreground of error map X_{err} as the consistency targets. Therefore, the ambiguity-guided consistency loss L_c can be calculated as follows:

$$\mathcal{L}_c(f_t, f_s, E) = \frac{\sum_v \left[E_v \sum_c \left(\| f_{t,v,c} - f_{s,v,c} \|^2 \right) \right]}{\sum_v E_v} \tag{7}$$

where E_v denotes the binary indicator at the vth voxel. $f_{t,v,c}$ and $f_{s,v,c}$ represent the predictions from teacher and student models for each class $c \in C$ at the vth voxel, respectively.

2.4. Segmentation backbone

We proposed a backbone network called U-shaped hybrid Swin Transformer Network (UST-Net) as the student and teacher model. This network architecture combines convolution and Swin Transformer [14], as shown in **Fig. 2**.

We incorporate the Transformer module into each layer of the encoder and decoder (except for the first layer), which enables the network to capture long-range dependencies across multiple scales without pre-training.

Specifically, we pass the down-sampled features through a convolution block and then Swin Transformer block to capture global contextual features. In decoder, we incorporate a skip attention block to better recover fine-grained details [15]. Then, the outputs of the attention block are passed through a convolution block to maintain a symmetric structure with the encoder, as shown in **Fig. 2**.

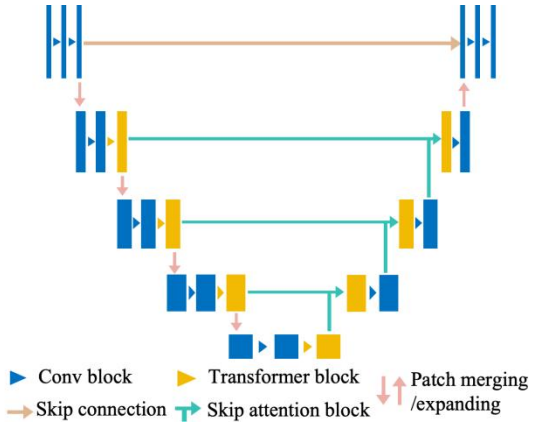

Fig. 2. The architecture of U-shaped hybrid Swin Transformer Network UST-Net.

2.5. Loss function

The overall loss of the model consists of a supervised loss L_{hs}, a calibrated supervised loss L_{ls}, and a consistency loss L_c, calculated as follows:

$$\mathcal{L} = \mathcal{L}_{hs} + \lambda(\mathcal{L}_c + \beta\mathcal{L}_{ls}) \quad (8)$$

where L_{hs} is the combination of cross-entropy loss and Dice loss. L_{ls} shares the same form as L_{hs} but applied to refined labels, and L_c is calculated by mean square error. The parameter β is set to an empirical value of 5. λ is a ramp-up trade-off weight commonly scheduled by the time-dependent Gaussian function.

3. EXPERIMENTS AND RESULTS

3.1. Datasets and preprocessing

To improve model's generalizability, heterogeneous MRA datasets were used for model training. A total of 113 high-quality manually labeled MRA images from three datasets (MIDAS, Forrest and IXI) formed the Set-HQ, and 210 low-quality labeled MRA images from six datasets formed the Set-LQ (MIDAS, Forrest, IXI, BraVa, CHUV and ADAM). The low-quality labels for Set-LQ were generated using FFCM-MRF, which is a statistical model-based method for cerebrovascular segmentation [16]. In addition, three human and three macaque independent datasets were used to evaluate the generalizability of our model (BraVa, CHUV, ADAM, Maca-WH, Maca-BJ and Maca-7T). All independent datasets were labeled by trained annotators. The MR scanning parameters and the number of subjects for each dataset are shown in **Table 1**.

The preprocessing of MRA images included 3dAutoBox for image cropping and N4BiasFieldCorrection for bias field correction. Note that we retained the skulls of each subject to preserve essential features for skull and small vessel identification in the present study.

Table 1. Dataset information.

Species	Dataset	H / L / I data	Scanner	Resolution (mm)
Humans	MIDAS [17]	54 / 55 / -	3T Siemens	0.51×0.51×0.80
	Forrest [1]	14 / 6 / -	7T Siemens	0.30×0.30×0.30
	IXI [2]	45 / 28 / -	1.5T Siemens	0.26×0.26×0.80
	BraVa [18]	- / 51 / 5	3T Philips	0.62×0.62×0.62
	CHUV [19]	- / 55 / 5	3T Siemens	0.47×0.47×0.70
	ADAM [20]	- / 15 / 5	1.5T GE	0.31×0.31×0.70
Macaques	Maca-WH [16]	- / - / 36	3T Siemens	0.50×0.50×0.50
	Maca-BJ [16]	- / - / 4	3T Siemens	0.21×0.21×0.32
	Maca-7T [16]	- / - / 4	7T Siemens	0.19×0.19×0.30

H = Set-HQ, L = Set-LQ, I = Independent dataset.

3.2. Implementation details

The proposed model was validated using 4-fold cross-validation. The training process consisted of 100 epochs, with 250 batches in each epoch. The batch size was set to 4, including 2 patches from Set-HQ and 2 patches from Set-LQ. The learning rate was set to 0.01. To avoid overfitting, various data augmentation techniques including rotation, scaling, gaussian noise and mirroring were used. The segmentation results were evaluated using five metrics: Dice score, precision, recall, average Hausdorff distance (AHD) and 95th percentile Hausdorff distance (95HD).

3.3. Comparisons of UST-Net with other models

We conducted supervised experiments to compare UST-Net with other widely used models, including 3D U-Net [21], V-Net [22], UNet++ [23], UNETR [24], and Swin UNETR [25]. As shown in **Table 2**, UST-Net demonstrated superior performance compared with other backbone models, with high Dice and recall scores and low AHD. UNETR and Swin UNETR only use Transformer in the encoder. In contrast, UST-Net incorporates Transformer and convolutions in both the encoder and decoder, suggesting that the proposed model may be a good identifier of vessels.

3.4. Comparisons with other semi-supervised strategies

We further conducted experiments to compare our semi-supervised strategy with several state-of-the-art semi-

[1] https://openneuro.org/datasets/ds000113/versions/1.3.0
[2] https://brain-development.org/ixi-dataset/

supervised strategies, including UA-MT [7], RC-MT [26], CPS [27], ICT [10] and DCT [9] based on the UST-Net backbone. As shown in **Table 3**, our method outperforms these methods in terms of all metrics except precision, and shows better performance than baseline. Results of UA-MT and RC-MT are poorer compared with the baseline model.

Table 2. Quantitative comparisons of backbone models.

Methods	Dice (%) ↑	Precision (%) ↑	Recall (%) ↑	AHD ↓	95HD ↓
3D UNet	85.54 (3.75)	88.58 (5.14)	83.21 (6.42)	0.48 (0.28)	55.27 (23.10)
VNet	85.20 (3.92)	88.52 (4.49)	82.66 (7.09)	0.53 (0.37)	53.79 (21.17)
UNet++	85.70 (3.68)	**88.92 (4.79)**	83.19 (6.36)	0.48 (0.29)	55.50 (24.98)
UNETR	83.52 (3.81)	87.55 (5.04)	80.41 (6.89)	0.67 (0.43)	61.02 (22.94)
Swin UNETR	85.19 (3.72)	88.50 (5.00)	82.64 (6.50)	0.48 (0.30)	**51.70 (23.82)**
UST-Net	**85.84 (3.68)**	88.86 (5.02)	**83.51 (6.27)**	**0.46 (0.28)**	52.85 (24.33)

Table 3. Quantitative comparisons of semi-supervised strategies.

Methods	Dice (%) ↑	Precision (%) ↑	Recall (%) ↑	AHD ↓	95HD ↓
UA-MT	84.52 (3.80)	89.37 (4.90)	80.67 (6.67)	0.52 (0.34)	64.86 (25.95)
RC-MT	84.82 (3.72)	89.16 (4.76)	81.31 (6.10)	0.51 (0.34)	61.76 (25.12)
CPS	85.90 (3.90)	**89.42 (4.61)**	83.12 (6.67)	0.46 (0.32)	54.21 (26.89)
ICT	85.84 (3.96)	**89.42 (4.56)**	82.98 (6.57)	0.46 (0.33)	51.77 (22.44)
DCT	86.05 (3.95)	89.34 (4.57)	83.48 (6.70)	0.45 (0.33)	52.50 (25.15)
CL-MT	**86.69 (3.84)**	89.02 (5.32)	**85.00 (6.46)**	**0.43 (0.37)**	**49.53 (22.73)**
Baseline	85.84 (3.68)	88.86 (5.02)	83.51 (6.27)	0.46 (0.28)	52.85 (24.33)

3.5. Comparisons with state-of-the-art methods

To further validate the performance of the proposed model, we implemented state-of-the-art methods including the supervised methods ER-Net [4], CS^2-Net [3], nnU-Net [7] and the semi-supervised method GCS [8]. Among all the comparison methods, our method achieved the best performance in terms of all metrics except precision, as shown in **Table 4**.

Qualitative evaluations of each method are shown in **Fig. 3**. Results demonstrated that our method can accurately segment small vessels with low contrast and high curvature. Moreover, our method gains more continuous vessels near the intracranial borders of the image. These qualitative evaluations verify the effectiveness and superiority of our method.

Table 4. Quantitative comparisons with state-of-the-art methods.

Methods	Dice (%) ↑	Precision (%) ↑	Recall (%) ↑	AHD ↓	95HD ↓
nnU-Net	85.90 (3.59)	**89.63 (5.02)**	82.97 (6.34)	0.45 (0.28)	50.82 (25.83)
GCS	70.75 (6.51)	85.39 (9.90)	62.40 (11.75)	2.56 (1.49)	65.40 (26.38)
ER-Net	81.83 (5.33)	89.13 (5.39)	76.16 (7.50)	0.54 (0.43)	49.98 (26.48)
CS^2-Net	80.28 (5.82)	88.21 (5.85)	74.33 (8.64)	0.65 (0.49)	50.74 (28.49)
CL-MT	**86.69 (3.84)**	89.02 (5.32)	**85.00 (6.46)**	**0.43 (0.37)**	**49.53 (22.73)**

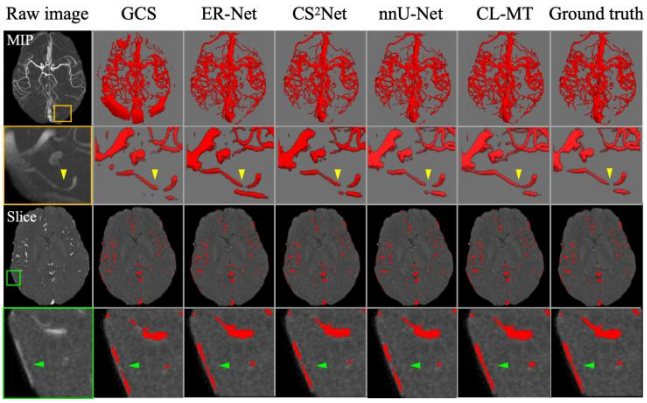

Fig. 3. Qualitative comparisons of cerebrovascular segmentation results on MIDAS dataset. These images have been skull-stripped solely for visualization purpose.

3.6. Generalization experiments

To evaluate the generalizability of our method, we conducted experiments on independent human and macaque datasets. As shown in **Fig. 4**, our method achieved high Dice scores on the ADAM, BraVa, CHUV, Maca-7T and Maca-WH datasets, with large improvements on the Maca-7T dataset. These results showed that our method exhibits good generalizability independent of scanning sites and species.

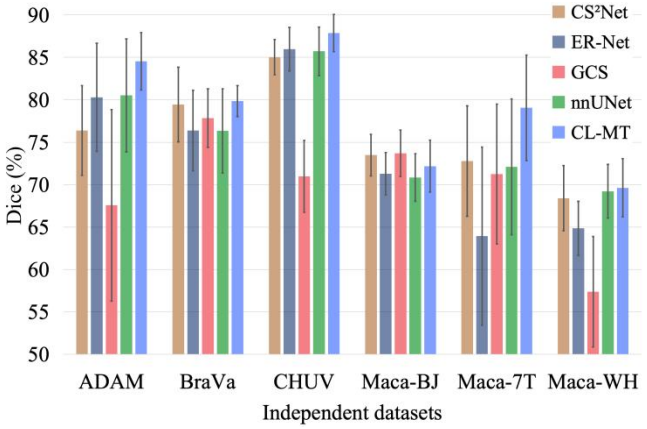

Fig. 4. Quantitative evaluations of cerebrovascular segmentation results using independent datasets.

4. CONCLUSION

In this work, we proposed a novel semi-supervised model for cerebrovascular segmentation using a small amount of high-quality labeled data and abundant low-quality noisy labeled data. We demonstrated that our model is highly accurate and reliable compared to state-of-the-art methods. Furthermore, experiments on human and macaque independent datasets demonstrated that the proposed model could generalize to heterogeneous human and non-human primate datasets, independent of variations in field strengths, scanner platforms and acquisition parameters.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using open access human subject data and private macaque subject data. Ethical approval was granted by the Institutional Review Board/Ethics Committee of Chinese Academy of Sciences Institute of Automation.

6. ACKNOWLEDGMENTS

This work was supported by STI 2030 - Major Projects (No. 2021ZD0200402), National Science Foundation of China (No. 82371486) and Beijing Natural Science Foundation (No. 4222057). The authors declare no conflicts of interest.

7. REFERENCES

[1] J. Gutierrez, V. Guzman, F. Khasiyev, et al., "Brain arterial dilatation and the risk of Alzheimer's disease," Alzheimer's & Dementia, vol. 15, no. 5, pp. 666-674, 2019.

[2] F. Isensee, P. F. Jaeger, S. A. A. Kohl, et al., "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation," Nature Methods, vol. 18, no. 2, pp. 203-211, 2021.

[3] L. Mou, Y. Zhao, H. Fu, et al., "CS2-Net: Deep learning segmentation of curvilinear structures in medical imaging," Medical Image Analysis, vol. 67, pp. 101874, 2021.

[4] L. Xia, H. Zhang, Y. Wu, et al., "3D vessel-like structure segmentation in medical images by an edge-reinforced network," Medical Image Analysis, vol. 82, pp. 102581, 2022.

[5] Q. Wu, Y. Chen, N. Huang, et al.m, "Weakly-supervised cerebrovascular segmentation network with shape prior and model indicator," in Proceedings of the International Conference on Multimedia Retrieval, 2022, pp. 668-676.

[6] A. Tarvainen, H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," Advances in Neural Information Processing Systems, vol. 30, 2017.

[7] L. Yu, S. Wang, X. Li, et al.m, "Uncertainty-aware self-ensembling model for semi-supervised 3D left atrium segmentation," in Medical Image Computing and Computer Assisted Intervention–MICCAI, 2019, pp. 605-613.

[8] C. Chen, K. Zhou, Z. Wang, et al., "Generative consistency for semi-supervised cerebrovascular segmentation from TOF-MRA," IEEE Transactions on Medical Imaging, vol. 42, no. 2, pp. 346-353, 2022.

[9] S. Qiao, W. Shen, Z. Zhang, et al.m, "Deep co-training for semi-supervised image recognition," in Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 135-152.

[10] V. Verma, K. Kawaguchi, A. Lamb, et al., "Interpolation consistency training for semi-supervised learning," Neural Networks, vol. 145, pp. 90-106, 2022.

[11] C. Northcutt, L. Jiang, I. Chuang, "Confident learning: Estimating uncertainty in dataset labels," Journal of Artificial Intelligence Research, vol. 70, pp. 1373-1411, 2021.

[12] D. Angluin, P. Laird, "Learning from noisy examples," Machine Learning, vol. 2, pp. 343-370, 1988.

[13] Z. Xu, Y. Wang, D. Lu, et al., "Ambiguity-selective consistency regularization for mean-teacher semi-supervised medical image segmentation," Medical Image Analysis, vol. 88, pp. 102880, 2023.

[14] Z. Liu, Y. Lin, Y. Cao, et al.m, "Swin transformer: Hierarchical vision transformer using shifted windows," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012-10022.

[15] H.-Y. Zhou, J. Guo, Y. Zhang, et al., "nnFormer: volumetric medical image segmentation via a 3D transformer," IEEE Transactions on Image Processing, vol. 32, pp. 4036-4045, 2023.

[16] Y. Cui, H. Huang, J. Liu, et al., "FFCM-MRF: An accurate and generalizable cerebrovascular segmentation pipeline for humans and rhesus monkeys based on TOF-MRA," *bioRxiv*, 2023.11.07.566142, 2023.

[17] E. Bullitt, D. Zeng, G. Gerig, et al., "Vessel tortuosity and brain tumor malignancy: a blinded study1," Academic Radiology, vol. 12, no. 10, pp. 1232-1240, 2005.

[18] S. N. Wright, P. Kochunov, F. Mut, et al., "Digital reconstruction and morphometric analysis of human brain arterial vasculature from magnetic resonance angiography," NeuroImage, vol. 82, pp. 170-181, 2013.

[19] T. Di Noto, G. Marie, S. Tourbier, et al., "Towards automated brain aneurysm detection in TOF-MRA: Open data, weak labels, and anatomical knowledge," Neuroinformatics, vol. 21, no. 1, pp. 21-34, 2023.

[20] K. M. Timmins, I. C. van der Schaaf, E. Bennink, et al., "Comparing methods of detecting and segmenting unruptured intracranial aneurysms on TOF-MRAS: the ADAM challenge," NeuroImage, vol. 238, pp. 118216, 2021.

[21] O. Ronneberger, P. Fischer, T. Broxm, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention–MICCAI, 2015, pp. 234-241.

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

[23] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, et al., "Unet++: Redesigning skip connections to exploit multiscale features in image segmentation," IEEE Transactions on Medical Imaging, vol. 39, no. 6, pp. 1856-1867, 2019.

[24] A. Hatamizadeh, Y. Tang, V. Nath, et al.m, "Unetr: Transformers for 3d medical image segmentation," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 574-584.

[25] A. Hatamizadeh, V. Nath, Y. Tang, et al.m, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," in International MICCAI Brainlesion Workshop, 2021, pp. 272-284.

[26] L. Xie, Z. Chen, X. Sheng, et al., "Semi-supervised region-connectivity-based cerebrovascular segmentation for time-of-flight magnetic resonance angiography image," Computers in Biology and Medicine, vol. 149, pp. 105972, 2022.

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