TRAINING CROSS-MODALITY CEREBROVASCULAR SEGMENTATION NETWORKS WITH PAIRED IMAGES

Zhanqiang Guo[1,2], *Jianjiang Feng*[1,2(✉)], *Wangsheng Lu*[3], *Yin Yin*[3], *Guangming Yang*[3], *and Jie Zhou*[1,2]

[1] Department of Automation, Tsinghua University, Beijing, China
[2] Beijing National Research Center for Information Science and Technology, Beijing, China
[3] UnionStrong (Beijing) Technology Co.Ltd, Beijing, China

ABSTRACT

Automatic cerebral vessel segmentation from Computed Tomography Angiography (CTA) and Magnetic Resonance Angiography (MRA) is an important task in clinical diagnosis of cranial vascular diseases. With the success of deep convolution neural network, cerebrovascular segmentation has made great progress. But in most of previous work, training the segmentation network in different modalities usually requires a large amount of data to be labeled for each modality, which is expensive and time-consuming. In this paper, we proposed a framework to realize cross-modality cerebrovascular segmentation between CTA and MRA with paired data. We first train the source domain segmentation network, through which the initial segmentation results of the target domain images can be obtained. Then pseudo-label of the target domain is generated by registering the initial segmentation of the target domain with the paired source domain label, and it is used to train the target domain segmentation network. Experiments demonstrate the effectiveness of our proposed framework in cross-modality segmentation.

Index Terms— Cerebrovascular Segmentation, Cross-modality, Transfer Learning, Pseudo Labels

1. INTRODUCTION

Segmenting cerebral vascular from Computed Tomography Angiography (CTA) and Magnetic Resonance Angiography (MRA) is a crucial step in the diagnosis of cerebrovascular diseases. With the development of deep Convolution Neural Network (CNN), many automatic segmentation algorithms [1, 2] are proposed to extract cerebrovascular network and obtain promising performance. However, these deep models are trained and tested with the data of the same modality and their performances often decline significantly when tested on images from different acquisition protocols or modalities, which is called domain shift [3]. Manual annotation of vessels in a new domain is the most naive way to address this challenge, but it is expensive and time-consuming for the task of segmenting cerebral vessels.

Another way to solve this problem is Unsupervised Domain Adaptation (UDA), which transfers knowledge from the source domain with labels to the target domain without labels, and does not require additional manual annotation in the target domain [4]. Many algorithms [5, 6] have been proposed based on generative adversarial networks (GANs) to minimize the discrepancy between the source and target domains [7, 8]. Chen et al. [9] presented Synergistic Image and Feature Alignment (SIFA) to effectively adapt a segmentation network to an unlabeled target domain for cardiac substructure segmentation and abdominal multi-organ segmentation. Pei et al. [10] implemented domain adaptation on both feature and image levels with disentangled domain-invariant features and domain-specific features. Most of UDA algorithms based on GANs treated the 3D voxels as a sequence of 2D slices in image synthesis due to the limitations of hardware device and time-consuming training process [11]. But these methods are difficult to migrate to 3D cerebrovascular images, because the vascular features are not obvious in 2D slicers, and it is difficult to ensure the integrity of small vessels and the continuity of vascular networks.

Besides GAN based methods, the pseudo label generation methods are also used in UDA. These methods usually utilize the network trained by the source domain data to get the initial segmentation results of the target domain data, and then adjust them to serve as the pseudo labels of the target domain images. Chen et al. [12] introduced two complementary pixel-level and class-level denoising schemes with uncertainty estimation to reduce noisy pseudo labels, and achieved good performance in optic disc and cup segmentation. And an uncertainty-based filtering was proposed to select high-quality pseudo labels during the training of target domain [13]. However, the performances of these methods depend heavily on the quality of the initial target domain segmentation results. With great discrepancy between the source and target domain, their performance will be seriously affected.

To address the challenges mentioned above, we propose a framework to achieve cross-modality cerebrovascular segmentation based on paired images, which are different modalities from the same patient. The key of our idea is to obtain

precise pseudo-labels of images in the target domain based on the source domain labels, which is not easy due to the great differences between the two modalities and the potentially large time gap between the two imaging, during which variability in vessels may occur. Our framework consists of three phases. First, to reduce the discrepancy between two modalities, we enhance blood vessels by Hessian matrix based filtering [14] and train the Source Domain Segmentation Network (SDSN). Second, the initial labels of target domain images are obtained through the SDSN. The pseudo-labels of target domain are generated by fusing the initial labels of target domain images with the source domain labels. Finally, with the pseudo labels, we train the Target Domain Segmentation Network (TSDN). Experiments show the effectiveness of our proposed framework.

2. METHODOLOGY

Let $D_s = \{(x_i^s, y_i^s)\}_{i=1}^N$ denote the N images and their labels from source domain, and $D_t = \{(x_i^t)\}_{i=1}^N$ donate the paired images from target domain without labeling. We aim to exploit D_s and D_t to improve the model performance on the target domain. As shown in Fig. 1, our framework consists of three stages: training source domain segmentation network with source domain data, generating target domain pseudo labels, and training target domain segmentation network, whose structure is same as source domain network but has different parameters.

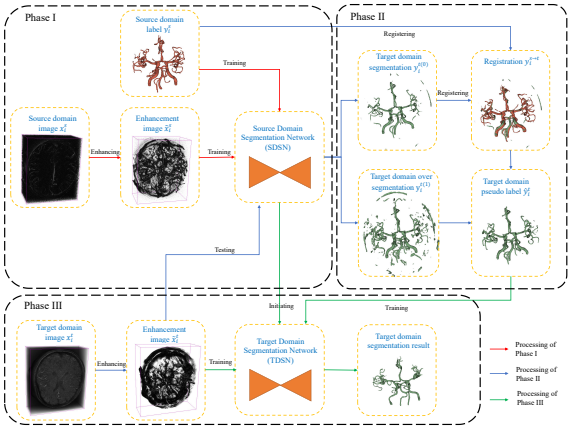

Fig. 1. Illustration of the proposed training procedure.

2.1. Source Domain Segmentation Network (SDSN)

As mentioned in section 1, CTA and MRA images are of great difference due to the different imaging principles [15]. The performances of segmentation network decline significantly when trained with the source domain images x^s and tested on target domain images x^t. In order to alleviate this problem, vessel enhancement by Frangi filtering [14] is done

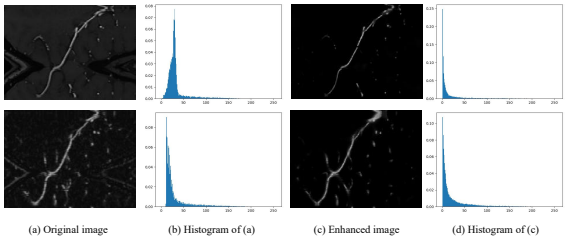

Fig. 2. Images and gray-value histograms of a corresponding region of two different modalities before and after enhancement. KL divergences of the two modal gray distributions are 1.001 and 0.172 before and after enhancement.

$(x \to \tilde{x})$ before feeding the images to the segmentation network. Fig. 2 illustrates the change in the gray-value distribution of the images after enhancement. We computed statistics on the gray-values of the same blood vessel and its surroundings, and calculated the KL divergence of the two modal gray distributions in the paired images before and after vessel enhancement. The result indicates that the discrepancy between the source and target domains is reduced.

The structure of the segmentation network used in our framework is the basic 3D U-Net [1], in fact, any popular segmentation network can be used for our proposed UDA method. Dice loss is used during training the segmentation network.

2.2. Target Domain Pseudo Labels

The enhancement image $(\tilde{x}_i^t)$ from target domain is fed into the SDSN. By choosing two threshold values, the initial segmentation result and over-segmentation result can be obtained, denoted as $y_i^{t(0)}$, $y_i^{t(1)}$. Converting binary image to 3D point cloud, we perform Iterative Closest Point (ICP) [16] algorithm to align source domain paired label y_i^s and segmentation result $y_i^{t(0)}$, obtaining $y_i^{s\to t}$. But the registration result $y_i^{s\to t}$ may not be perfectly matched to real blood vessels of images from target domain in detail due to the timing of paired images, differences in imaging modalities, and variation in blood vessels. Therefore, in our method, the pseudo label $\hat{y}_i^t$ is generated by finding the vessel labeling closest to the blood vessel in $y_i^{s\to t}$ as the annotation of the target image x_i^t in over-segmentation result $y_i^{t(1)}$. At the point p, the value of the pseudo-label is calculated as follows:

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

where $d(p, y_i^{s\to t})$ is the minimum distance from point p to the point set of $y_i^{s\to t}$, and ϵ is the threshold of distance. This formula indicates that the point p is considered to be foreground when the over-segmentation result $y_i^{t(1)}$ is 1 and the distance between this point and the registered resulting image $y_i^{s\to t}$ is

Table 1. Segmentation performance when there is no label in the target domain.

	$MRAtoCTA$			$CTAtoMRA$		
Metrics	**DSC(%)**	**clDice(%)**	**AHD(mm)**	**DSC(%)**	**clDice(%)**	**AHD(mm)**
Upper limit	77.03±7.12	83.40±4.52	0.521±0.355	88.33±3.19	89.51±3.41	0.263±0.170
SIFA [9]	25.67±11.83	25.04±13.87	4.515±2.910	38.13±12.36	34.11±11.60	3.118±1.704
DDSeg [10]	38.80±7.31	40.45±7.25	6.238±2.619	58.17±7.19	57.34±8.99	3.022±1.579
DPL [12]	49.34±9.08	50.26±9.71	5.361±2.011	59.36±12.10	56.81±10.10	2.668±1.529
FPL [13]	66.10±8.12	69.04±6.44	3.029±1.578	40.78±14.85	45.46±12.01	3.182±1.472
Origin-net	62.59±7.25	68.24±6.37	2.488±1.489	40.72±14.91	44.37±14.23	2.808±1.780
SDSN	65.36±6.34	72.07±5.92	2.132±1.549	56.18±6.65	59.98±8.39	2.349±1.282
TDSN	**72.87±3.34**	**75.46±4.93**	**0.792±0.329**	**72.25±4.95**	**72.87±7.32**	**0.831±0.389**

close enough. In fact, since the over-segmentation results can cover most of the blood vessels, the pseudo-labels obtained by our method are reliable.

2.3. Target Domain Segmentation Network (TDSN)

The segmentation network structure used in this stage is the same as that of stage I. We initialize TDSN with the network parameters trained in phase I. The target domain images with pseudo labels $\{(\tilde{x}_i^t, \hat{y}_i^t)\}_{i=1}^N$ are used to train the network. During testing, the vessel enhanced image $\tilde{x}_i^t$ of the target domain testing image x_i^t is fed into the trained target domain segmentation network to obtain the final segmentation result.

3. EXPERIMENTS AND RESULTS

3.1. Data and Implementation Details

Implementation Details. Our segmentation network was implemented with PyTorch framework, and trained on a NVIDIA GeForce GTX 3090 GPU. The adaptive moment estimation (Adam) with initial learning rate 0.001 was employed during training. In addition, a poly learning rate policy [17] with power 0.9 was used, and the batch size was set to 8, the maximum epoch was 10000. The hyperparameter ϵ was set to 4 pixels. The threshold values of $y_i^{t(0)}$ and $y_i^{t(1)}$ were 0.5 and 0.03 respectively.

Data and metrics. A dataset provided by a local hospital was used to evaluate the proposed framework. The dataset includes paired CTA and MRA data of 21 patients, unpaired MRA images of 20 patients, unpaired CTA images of 20 patients. The public dataset, TubeTK[1], is not suitable for testing our experiments due to the large discrepancy between the MRA images from TubeTK and the MRA images from our dataset. We tested the fully supervised model, trained with the MRA images from our dataset, on TubeTK and got a Dice similarity coefficient (DSC) of 39.07%, indicating the significant gap between two MRA datasets due to different acquisition protocols. In the experiment, the voxel size of the two modality data was unified to $0.5\times0.5\times0.75$ mm^3, and patches of size $96\times96\times96$ were randomly cropped to feed into the segmentation network for training. Furthermore, to fully verify the effectiveness of our proposed method, we adopted 4-fold cross-validation on paired data, and tested on unpaired data using the training result of the first fold experiment. We evaluated the results based on DSC, clDice [2] and Average Hausdorffs Distance (AHD).

We conducted two types of experiments: the target domain without labeled data and the target domain with a few labeled data. In addition, in each type of experiments, we use MRA data and CTA data as the target domain, respectively, with $MRAtoCTA$ meaning using MRA images as the source domain, CTA images as the target domain, and $CTAtoMRA$ on the contrary.

3.2. Target Domain without Label

Let $\{(x_i^{s(p)}, y_i^{s(p)})\}_{i=1}^N$ denote the paired images and labels from source domain, $\{x_i^{t(p)}\}_{i=1}^N$ donate the paired images in target domain and $\{x_i^{t(u)}\}_{i=1}^M$ donate the unpaired data from the target domain. We train the SDSN with $D_{no}^{net1} = \{(\tilde{x}_i^{s(p)}, y_i^{s(p)})\}_{i=1}^n$, where $\tilde{x}_i$ is the enhancement image of x_i, and generate pseudo-labels $\{\hat{y}_i^{t(p)}\}_{i=1}^n$ for paired images from target domain. During experiment, we evaluate the generated pseudo-labels, and the DSC of the manual annotations with the generated MRA and CTA pseudo-labels are 72.67 and 74.16 respectively. Then the TDSN is trained with $\{(\tilde{x}_i^{t(p)}, \hat{y}_i^{t(p)})\}_{i=1}^n$ and tested with the remaining target domain images $\{\tilde{x}_i^{t(p)}\}_{i=n+1}^N$ (for the first fold experiment, tested with $\{\tilde{x}_i^{t(p)}\}_{i=n+1}^N \cup \{\tilde{x}_i^{t(u)}\}_{i=1}^M$). Meanwhile, we also directly use the original data for transfer learning, that is, train the network with $\{(x_i^{s(p)}, y_i^{s(p)})\}_{i=1}^n$ and test with $\{x_i^{t(p)}\}_{i=n+1}^N$, to illustrate the role of the enhancement step.

We compared our method with four state-of-the-art UDA methods: two methods based GAN (SIFA [9], DDSeg [10]) and two methods based pseudo label (DFL [12], FPL [13]). We also reported the results of fully supervised training (Upper limit) as the upper bound of UDA methods. Fig. 3 shows the results of the experiment. Due to the similar performance of these networks on main blood vessel branches, we show

[1]https://public.kitware.com/Wiki/TubeTK/Data

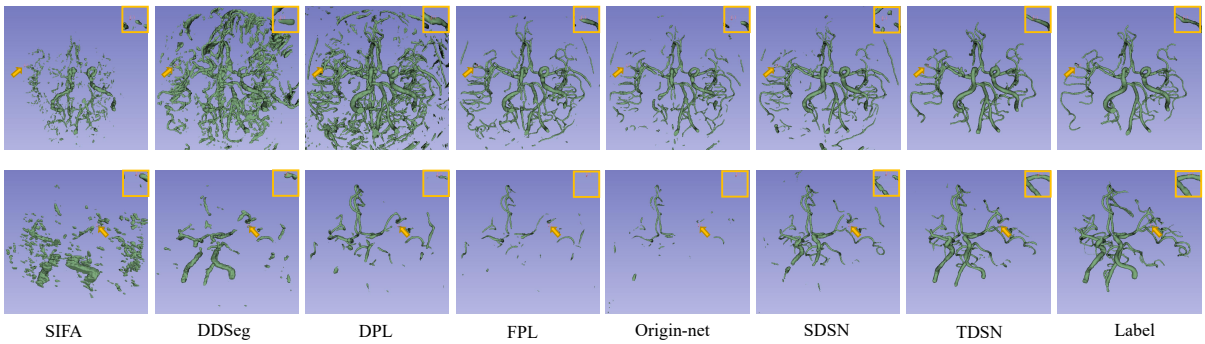

Fig. 3. Visualization of segmentation results. The first and second rows show the results of different methods on two examples of $MRAtoCTA$ and $CTAtoMRA$ respectively.

the close-ups of the vessels highlighted by yellow rows in the yellow boxes. As we can see, the results of methods based on GAN are greatly worse than those of our method. And DFL and FPL also show bad performances due to the great discrepancy between CTA and MRA. Meanwhile, TDSN produces better results in removing noise, preserving small vessels and maintaining vascular continuity than SDSN, which is trained with source domain images and tested on target domain images. These findings are consistent with the results in Table 1. Another observation from Fig. 3 and Table 1 is that the results of transfer learning using the images without enhancement (Origin-net) are worse than those of our framework, illustrating that vessel enhancement can reduce the discrepancy between domains.

3.3. Target Domain with a Few Labels

Excepting UDA tasks, it is also worth studying if we can exploit a few data with annotations from new domains encountered in a real application. In this experiment, we assume that there are two images with labels in the target domain for each fold experiment. In experiment, the image $x_{k_1}^{t(p)}$ and $x_{k_2}^{t(p)}$ has label $y_{k_1}^{t(p)}$ and $y_{k_2}^{t(p)}$. The data $D_{few}^{net1} = \{(\tilde{x}_i^{s(p)}, y_i^{s(p)})\}_{i=1}^{n} \cup \{(\tilde{x}_i^{t(p)}, y_i^{t(p)})\}_{i=k_1,k_2}$ is used to train SDSN, and generate pseudo-labels $\{\hat{y}_i^{t(p)}\}_{i=1,i\neq k_1,k_2}^{n}$ for paired images. Then we train the TDSN with $\{(\tilde{x}_i^{t(p)}, y_i^{t(p)})\}_{i=k_1,k_2} \cup \{(\tilde{x}_i^{t(p)}, \hat{y}_i^{t(p)})\}_{i=1,i\neq k_1,k_2}^{N}$ and test with images $\{\tilde{x}_i^{t(p)}\}_{i=n+1}^{N}$ (for the first fold experiment, test with $\{\tilde{x}_i^{t(p)}\}_{i=n+1}^{N} \cup \{\tilde{x}_i^{t(u)}\}_{i=1}^{M}$). In addition, in this experiment, we also test the segmentation network, denoted as Weak-net, trained with only a small amount of labeled data $\{(\tilde{x}_i^{t(p)}, y_i^{t(p)})\}_{i=k_1,k_2}$ in the target domain as a comparison.

Table 2 shows the metrics of this experiment. Comparing the results of SDSN and TDSN, we can get the same conclusion as the previous experiment. And our framework produces better results than the network trained with a few images from the target domain (Weak-net), demonstrating the effectiveness of our framework on transfer learning. And compared with Table 1, it is also in line with expectations for the increase in SDSN and TDSN due to the labeled data in the target domain.

Table 2. Segmentation performance when there are a few labels in the target domain.

$MRAtoCTA$	**DSC(%)**	**clDice(%)**	**AHD(mm)**
Weak-net	65.80±6.93	74.54±5.23	1.605±1.651
SDSN	69.02±5.92	**79.21±4.92**	1.296±1.518
TDSN	**74.54±3.05**	78.34±3.89	**0.756±0.325**
$CTAtoMRA$	**DSC(%)**	**clDice(%)**	**AHD(mm)**
Weak-net	82.43±4.42	83.30±5.64	0.483±0.289
SDSN	77.21±4.55	79.67±5.19	0.706±0.436
TDSN	**84.54±3.23**	**86.69±4.27**	**0.319±0.177**

4. CONCLUSION

Cerebrovascular segmentation from CTA and MRA images is an important task in clinical diagnosis. When there are annotated data from one modality, it is an interesting and clinically valuable task to construct segmentation models for other modalities. In this paper, we propose a framework for cross-modality segmentation based on paired data. First, we perform Hessian matrix based vessel enhancement on the images of the source and target domains to reduce the discrepancy between the two domains. A segmentation network is trained with annotated data from source domain, through which the initial segmentation results of the target domain images can be obtained. Then, by registering with the source domain labels, the pseudo-labels of the target domain images are generated, which are used to train the target domain segmentation network. Experiments show that our method can achieve good results in the cross-modality segmentation task. It is of interest to apply our method to other vessel datasets (such as coronary, airway) and improve the proposed framework to accommodate unpaired data in 3D vessel cross-modality segmentation in the future.

5. COMPLIANCE WITH ETHICAL STANDARDS

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

6. REFERENCES

[1] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 424–432.

[2] Suprosanna Shit, Johannes C Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien PW Pluim, Ulrich Bauer, and Bjoern H Menze, "cldice-a novel topology-preserving loss function for tubular structure segmentation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 16560–16569.

[3] Ben Glocker, Robert Robinson, Daniel C Castro, Qi Dou, and Ender Konukoglu, "Machine learning with multi-site imaging data: An empirical study on the impact of scanner effects," *arXiv preprint arXiv:1910.04597*, 2019.

[4] Marco Toldo, Andrea Maracani, Umberto Michieli, and Pietro Zanuttigh, "Unsupervised domain adaptation in semantic segmentation: a review," *Technologies*, vol. 8, no. 2, pp. 35, 2020.

[5] Liang Han and Zhaozheng Yin, "Unsupervised network learning for cell segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021, pp. 282–292.

[6] Xiaokang Chen, Yuhui Yuan, Gang Zeng, and Jingdong Wang, "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.

[7] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell, "Cycada: Cycle-consistent adversarial domain adaptation," in *International Conference on Machine Learning*. PMLR, 2018, pp. 1989–1998.

[8] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2223–2232.

[9] Cheng Chen, Qi Dou, Hao Chen, Jing Qin, and Pheng Ann Heng, "Unsupervised bidirectional cross-modality adaptation via deeply synergistic image and feature alignment for medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2494–2505, 2020.

[10] Chenhao Pei, Fuping Wu, Liqin Huang, and Xiahai Zhuang, "Disentangle domain features for cross-modality cardiac image segmentation," *Medical Image Analysis*, vol. 71, pp. 102078, 2021.

[11] Yi Sun, Peisen Yuan, and Yuming Sun, "MM-GAN: 3D MRI data augmentation for medical image segmentation via generative adversarial networks," in *2020 IEEE International Conference on Knowledge Graph (ICKG)*. IEEE, 2020, pp. 227–234.

[12] Cheng Chen, Quande Liu, Yueming Jin, Qi Dou, and Pheng-Ann Heng, "Source-free domain adaptive fundus image segmentation with denoised pseudo-labeling," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2021, pp. 225–235.

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

[14] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever, "Multiscale vessel enhancement filtering," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 1998, pp. 130–137.

[15] Fan Fu, Jianyong Wei, Miao Zhang, Fan Yu, Yueting Xiao, Dongdong Rong, Yi Shan, Yan Li, Cheng Zhao, Fangzhou Liao, et al., "Rapid vessel segmentation and reconstruction of head and neck angiograms using 3D convolutional neural network," *Nature Communications*, vol. 11, no. 1, pp. 1–12, 2020.

[16] Paul J Besl and Neil D McKay, "Method for registration of 3-d shapes," in *Sensor fusion IV: Control Paradigms and Data Structures*. Spie, 1992, vol. 1611, pp. 586–606.

[17] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia, "Pyramid scene parsing network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2881–2890.