Spatio-Temporal U-Net for Cerebral Artery and Vein Segmentation in Digital Subtraction Angiography

Ruisheng Su1, Matthijs van der Sluijs1, Sandra Cornelissen1, Pieter Jan van Doormaal1, Ruben van den Broeck1, Wim van Zwam2, Jeannette Hofmeijer3, Danny Ruijters4, Wiro Niessen1,5, Aad van der Lugt1, and Theo van Walsum1

1 1 Erasmus MC
2 2 Maastricht UMC
3 3 University of Twente
4 4 Philips Healthcare
5 5 TU Delft

Abstract. X-ray digital subtraction angiography (DSA) is widely used for vessel and/or flow visualization and interventional guidance during endovascular treatment of patients with a stroke or aneurysm. To assist in peri-operative decision making as well as post-operative prognosis, automatic DSA analysis algorithms are being developed to obtain relevant image-based information. Such analyses include detection of vascular disease, evaluation of perfusion based on time intensity curves (TIC), and quantitative biomarker extraction for automated treatment evaluation in endovascular thrombectomy. Methodologically, such vessel-based analysis tasks may be facilitated by automatic and accurate artery-vein segmentation algorithms. The present work describes to the best of our knowledge the first study that addresses automatic artery-vein segmentation in DSA using deep learning. We propose a novel spatio-temporal U-Net (ST U-Net) architecture which integrates convolutional gated recurrent units (ConvGRU) in the contracting branch of U-Net. The network encodes a 2D+t DSA series of variable length and decodes it into a 2D segmentation image. On a multi-center routinely acquired dataset, the proposed method significantly outperformed U-Net (P<0.001) and traditional Frangi-based K-means clustering (P<0.001). Particularly in artery-vein segmentation, ST U-Net achieved a Dice coefficient of 0.794, surpassing the existing state-of-the-art methods by a margin of 12%-20%. Code will be made publicly available upon acceptance.

Keywords: Deep Learning · X-Rays · ConvGRU · Stroke · Aneurysm · Vascular Malformations · Thrombectomy · Biomarkers.

1 Introduction

Cerebrovascular diseases remain a leading cause of death and long-term disability worldwide [15]. Such diseases include ischemic stroke due to vessel occlusion,
stenosis, and aneurysms. During diagnosis and neurovascular interventions, dynamic imaging of cerebral blood vessels is performed via X-ray digital subtraction angiography (DSA). Blood flow dynamics as well as changes in vasculature appearance over time can be visualized, providing vital information for diagnosis, procedural navigation, therapeutic decision making, and evaluation of treatment results.

So far, DSA images are visually inspected by neuroradiologists and interventionalists, which could be tedious, subjective, qualitative, and error-prone. Automatic artery-vein segmentation has the potential to aid this assessment by highlighting and quantifying vascular changes. This can serve as a basis for several downstream clinical applications, including quantitative evaluation of endovascular thrombectomy in stroke, automatic emboli detection, as well as image guidance in real time endovascular navigation. Besides, numerous quantitative blood flow related biomarkers could be extracted and analyzed from the segmented arteries and veins.

Vessel segmentation has been extensively studied in medical imaging for more than two decades, and in recent years this was boosted by the emergence of deep learning. Recent deep learning work has primarily centered around retinal [4,2] and lung imaging ([19],[16],[20]). In brain vascular imaging, several methods have been proposed on image modalities including MRA [13],[10],[11], CTA [17],[5] and DSA [18],[21]. Robben et al. proposed to segment cerebral vasculature in MRA images by leveraging anatomical information [13]. Fu et al. utilized a 3D CNN (ResU-Net3 model [3]) for bone and vessel segmentation in 3D CTA images [5]. Zhang et al. applied U-Net to segment vessels in cerebral 2D DSA frames [21]. Specifically on artery-vein segmentation, variants of the U-Net semantic segmentation architecture have been commonly adopted. Hemelings et al. applied U-Net to directly segment arteries and veins in fundus images [7]. Qin et al. designed a 3D U-Net for artery-vein segmentation in non-contrast lung CT images [12].
In the context of cerebral artery-vein segmentation, little work has been reported. Meijs et al. demonstrated the usage of a 3D U-Net in segmenting 4D CTA images [8]. Hilbert et al. also proposed a 3D U-Net model for arterial brain vessel segmentation in MRA images. To the best of our knowledge, no automatic artery-vein segmentation algorithm in DSA images has been presented so far. The recent related work by Zhang et al. segmented cerebral vessels in single 2D DSA frames without distinguishing arteries and veins [21]. It also demonstrated that U-Net could lead to false positives on skull edges or other motion artifacts whose appearance is similar to blood vessels.

DSA a modality with a powerful spatio-temporal visualization of blood flow dynamics in sequential frames (Figure 1). While existing vessel segmentation networks could be directly adopted for end-to-end artery-vein segmentation, most methods use static images. When addressing artery-vein segmentation in DSA series, which are 2D+t image sequences, the temporal dimension is relevant. We hypothesize that effective embedding of spatio-temporal flow dynamics is key towards accurate artery-vein segmentation in DSA.

In this work, we present the first automatic method for artery-vein segmentation in DSA using deep learning, and thereby establishing a new benchmark. Methodologically, we propose a novel spatio-temporal U-Net (ST U-Net) architecture which takes a 2D+t video sequence with variable length as input and a single 2D artery-vein segmentation as output. We model the temporal blood flow characteristics using convolutional gated recurrent units (ConvGRU). From a clinical perspective, we demonstrate the feasibility of deep learning-based artery-vein segmentation, via simultaneously leveraging vasculature and flow dynamics. We believe that the presented algorithm will facilitate fast, accurate, and objective cerebral vasculature interpretation in DSA, thus assisting endovascular interventions in clinical practice.

2 Methods

A DSA series includes a sequence of X-ray images subtracted by the first frame, demonstrating the cerebral blood flow over time. The goal of this work is to automatically segment the arteries and veins from an input DSA series, obtaining a 2D vessel mask as shown in Fig. 1.

2.1 Baseline methods

Frangi+K-means Traditional methods can be applied to this task. In this work, we implemented a recently proposed conventional two-step artery-vein classification method which combines Frangi filter and K-means unsupervised learning [1]. First, Frangi filter is applied on the static minimum intensity map (MinIP) of an input DSA series, followed by fixed thresholding to obtain a binary vessel mask. Subsequently, all the vessel pixels are classified into either arteries or veins using k-means clustering based on the time intensity curve (TIC) of each pixel.
Fig. 2: Network architecture of the proposed ST U-Net for artery-vein segmentation in DSA. Input is a 2D+t DSA series with variable series length. Output is a 2D segmentation image with white, red, and blue colors, indicating background, artery, and vein respectively.

**U-Net semantic segmentation** Apart from classical machine learning, standard U-Net could be directly used for segmenting arteries and veins. We implemented a deep learning baseline using U-Net [14,21]. The architecture is similar to the architecture in Fig. 2 with the contracting path replaced by stand-alone "down" blocks.

**U-Net + K-means** We further proposed a competitive baseline which cascades U-Net and K-means unsupervised learning. The U-Net takes a static MinIP image as input and produces a binary vessel segmentation. The vessel pixels are then clustered into arteries and veins via K-means.

### 2.2 Spatio-temporal U-Net

Rather than relying on pixel-wise TICs or vasculature appearance separately, we propose to simultaneously leverage spatial and temporal features to segment arteries and veins. Figure 2 shows a high-level overview of the proposed network architecture. Similar to U-Net, we follow the encoder-decoder architecture with four down layers (yellow boxes in Fig. 2) and four up layers (green boxes in Fig. 2). The number of channels, starting from 64, is doubled with max pooling in each down layer and halved with upsampling in each up layer. Compared to U-Net, an essential application-tailored modification is that the network takes 2D+t DSA series with variable series length rather than a single 2D image as input. All frames of the DSA series are fed into the same down block for feature extraction, followed by a convolutional GRU network for feature aggregation.
along the time axis. With the aim of learning low-level time intensity curve (TIC) characteristics, we use two layers of GRUs with a small convolutional kernel size of $1 \times 1$ or $3 \times 3$. Note that all down blocks in the same yellow cubic (Fig. 2) share the same weights even though they are shown separately in parallel. As a result, series with variable series length are encoded into a fixed-sized representation. This feature extraction and temporal aggregation process is repeated in all four layers, thus applied on multi-scale image features. The network does not include any fully connected layers; hence it is not bounded to fixed input image size.

3 Materials and Experiments

3.1 Data and annotations

The dataset was obtained from an ongoing prospective observational multi-center registry, consisting of 3232 patients with acute ischemic stroke who underwent endovascular thrombectomy (EVT). The included DSA series were post-EVT acquisitions by various imaging systems (e.g., Philips, GE, and Siemens). Considering the manual annotation effort, we randomly selected a subset of 110 DSA series from 90 patients for annotation. During annotation, 13 DSA series were excluded according to the following criteria: 1) arterial and venous phase being present, 2) no severe motion artifacts, and 3) adequate contrast in brain and vessels. This resulted in 97 DSA series for training and evaluating the methods in this study. The size of the individual images in the DSA series is $1024 \times 1024$ pixels. The series have a variable length, ranging from 10 to 50 frames, and also have various temporal resolutions (0.5-4 frames per second).

Artery-vein annotation in DSA was performed using an in-house developed tool in MeVisLab [6]. All images were first annotated by four trained students working in pairs to avoid interobserver variability and then further refined by another trained student. An experienced radiologist was available for consultancy during annotation when in doubt. As visualized in Fig. 1d, pixels were annotated either as arterial pixels, shown in red, or venous pixels in blue. In case arteries and veins overlap, a pixel could have both labels. These annotations serve as the reference standard in this work.

3.2 Implementation details

The proposed methods were implemented in PyTorch [9], all trained on an NVIDIA RTX A40 GPU. As pre-processing, we resized all frames to $512 \times 512$ pixels, resampled the temporal resolution to 1 fps, and normalized the intensity values to $[0, 255]$. The corresponding mask images were also resized to the same resolution. In addition, data augmentation techniques, i.e., horizontal flipping, translation $\in [-5\%, 5\%]$, scaling $\in [-5\%, 5\%]$, and rotation $\in [-10^{\circ}, 10^{\circ}]$, were randomly applied during training with a probability of 0.5 for each.

In the experiments, we randomly split the entire dataset into training, validation, and testing set with a ratio of 50%-20%-30% on the patient level. During
Table 1: Performance comparison between the proposed ST U-Net and state-of-the-art methods in both vessel segmentation and artery-vein segmentation on the test set. KNN: K-means, Acc: accuracy, Sens: sensitivity, Spec: specificity, A-Dice: Artery Dice, V-Dice: Vein Dice, M-Dice: Multi-class Dice.

| Method       | Vessel segmentation | Artery-vein segmentation |
|--------------|---------------------|--------------------------|
|              | Acc     Sens     Spec | Dice  Acc     Sens     Spec | A-Dice | V-Dice | M-Dice |
| Frangi + K-means [1] | 0.830 ± 0.048 | 0.685 ± 0.127 | 0.887 ± 0.094 | 0.816 ± 0.050 | 0.891 ± 0.083 | 0.647 ± 0.077 | 0.539 ± 0.115 | 0.597 ± 0.084 |
| U-Net [14,21] | 0.890 ± 0.032 | 0.790 ± 0.071 | 0.933 ± 0.029 | 0.865 ± 0.050 | 0.941 ± 0.041 | 0.700 ± 0.028 | 0.631 ± 0.082 | 0.670 ± 0.060 |
| U-Net + K-means | -   -   -   -   -   - | 0.873 ± 0.035 | 0.725 ± 0.071 | 0.934 ± 0.028 | 0.730 ± 0.060 | 0.648 ± 0.067 | 0.692 ± 0.050 |
| Proposed      | 0.906 ± 0.031 | 0.825 ± 0.076 | 0.943 ± 0.022 | 0.834 ± 0.047 | 0.904 ± 0.035 | 0.793 ± 0.078 | 0.950 ± 0.024 | 0.824 ± 0.077 | 0.762 ± 0.054 | 0.794 ± 0.057 |

Training of all models, the sum of cross-entropy and Dice loss was minimized using RMSprop optimization and a ReduceLROnPlateau scheduler with a patience of 10 epochs and a decay factor of 0.5. An early stopping strategy was applied with a patience of 50 epochs and a maximum of 1000 epochs. The optimal initial learning rate was found to be $1 \times 10^{-5}$ for all models via a grid search, except for ST U-Net in vessel segmentation ($1 \times 10^{-4}$). In ConvGRU, a kernel size of $1 \times 1$ and $3 \times 3$ were optimal for vessel segmentation and artery-vein segmentation respectively.

### 3.3 Evaluation metrics

Evaluation was performed on a separate test set with 26 DSA series from different patients. We present the performance of various methods from two perspectives: vessel segmentation and artery-vein segmentation (Table 1). We report the Dice coefficients, accuracy, sensitivity, and specificity. For artery-vein segmentation, we additionally analyze the artery Dice (A-Dice), vein Dice (V-Dice) and multi-class Dice (M-Dice), respectively defined as

$$A-Dice = \frac{2TP_a}{2TP_a + FP_a + FN_a}, \quad V-Dice = \frac{2TP_v}{2TP_v + FP_v + FN_v}, \quad (1)$$

$$M-Dice = \frac{2(TP_a + TP_v)}{2(TP_a + TP_v) + FP_a + FP_v + FN_a + FN_v}, \quad (2)$$

where the subscript a and v denote artery and vein respectively. Besides, we determine the statistical significance of performance differences between methods using the paired Wilcoxon test on the Dice coefficient.
4 Results and Discussion

Quantitative analysis In Table 1 we show the performance of the proposed method and existing methods in terms of vessel segmentation and artery-vein segmentation on the test set. For vessel segmentation, both the U-Net and the ST U-Net showed significantly superior performance over the Frangi+K-means approach in terms of Dice coefficient with $P < 0.001$. Notably, the proposed method significantly surpassed U-Net ($P=0.023$) and the Frangi+K-means approach ($P < 0.001$), demonstrating the value of spatio-temporal feature representation in recognizing vessels with contrast flow.

With respect to artery-vein segmentation, the advantage of spatio-temporal learning was even more prominent as shown in Table 1. When relying on pixel-wise temporal characteristics, a multi-class Dice coefficient of 0.597 ($\pm 0.084$) was achieved. Spatial feature-based deep learning reached a higher Dice coefficient of 0.670 ($\pm 0.060$). The U-Net+K-means approach learns both spatial and temporal features, albeit in two separate stages, achieving a Dice coefficient of 0.692 ($\pm 0.050$). In contrast, fusing both vasculature appearance and flow dynamics in ST U-Net boosted the artery-vein segmentation performance. The multi-class Dice coefficient of 0.794 ($\pm 0.057$) was significantly higher than U-Net ($P < 0.001$) and Frangi+K-means ($P < 0.001$).

Qualitative analysis To further illustrate the differences between the proposed architecture and a U-Net on a MinIP, we present visualizations and qualitative comparisons of the methods. First, the proposed ST U-Net helped in distinguishing skull, instruments and other subtraction artifacts from cerebral vessels based on temporal contrast flow dynamics. Fig. 3 shows three comparisons of the
vessel segmentation results between U-Net and the proposed ST U-Net. In the error maps (i.e., column c and e), orange indicates false positive while light blue indicates false negative. Region #1 and #3 highlight two scenarios where U-Net recognized subtraction artifacts and static instrument as vessels (column c), which were correctly avoided in ST U-Net (column e). Furthermore, ST U-Net was able to discover venous vessels (region #2 and #4) surrounded by motion artifacts based on the temporal characteristics while U-Net missed those due to the noisy background.

Fig. 4 illustrates the performance of U-Net and ST U-Net with respect to artery-vein segmentation with two examples. Comparing the errors maps of U-Net (column c) and ST U-Net (column e) with respect to manual annotations (column a), less false positives (orange) and false negatives (light blue) are observed in the proposed ST U-Net (column e). In particular, although U-Net succeeds in recognizing large vessels, e.g., proximal arteries and superior sagit-
tal sinus (SSS), it shows limited capability in correctly segmenting small arteries/veins. Such examples are highlighted in region #1 and #2 in Fig. 4.

5 Conclusion

We have presented a deep learning-based cerebral artery-vein segmentation method in DSA. The novel and application-tailored spatio-temporal U-Net takes 2D+t videos as input and outputs 2D multi-class segmentation images. The benefit of spatio-temporal learning has been underpinned via quantitative and qualitative analyses. The promising performance of the proposed method reveals its potential value in various clinical applications for vessel-based quantitative analyses.

References

1. van Asperen, V., van den Berg, J., Lycklama, F., Marting, V., Cornelissen, S., van Zwam, W., Hofmeijer, J., van der Lugt, A., van Walsum, T., van der Sluijs, M., Su, R.: Automatic artery/vein classification in 2d DSA images of stroke patients. In: Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling. SPIE (Jan 2022). https://doi.org/10.1117/12.2606412
2. Chen, C., Chuah, J.H., Raza, A., Wang, Y.: Retinal vessel segmentation using deep learning: A review. IEEE Access (2021)
3. Diakogiannis, F.I., Waldner, F., Caccetta, P., Wu, C.: Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS Journal of Photogrammetry and Remote Sensing 162, 94–114 (2020)
4. Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G., Barman, S.A.: Blood vessel segmentation methodologies in retinal images—a survey. Computer methods and programs in biomedicine 108(1), 407–433 (2012)
5. Fu, F., Wei, J., Zhang, M., Yu, F., Xiao, Y., Rong, D., Shan, Y., Li, Y., Zhao, C., Liao, F., et al.: Rapid vessel segmentation and reconstruction of head and neck angiograms using 3d convolutional neural network. Nature communications 11(1), 1–12 (2020)
6. Heckel, F., Schwier, M., Peitgen, H.O.: Object-oriented application development with mevislab and python. Informatik 2009–Im Focus das Leben (2009)
7. Hemelings, R., Elen, B., Stalmans, I., Van Keer, K., De Boever, P., Blaschko, M.B.: Artery–vein segmentation in fundus images using a fully convolutional network. Computerized Medical Imaging and Graphics 108, 101636 (2019)
8. Meijs, M., Pegge, S.A., Vos, M.H., Patel, A., van de Leemput, S.C., Koschmieder, K., Prokop, M., Meijer, F.J., Manniesing, R.: Cerebral artery and vein segmentation in four-dimensional ct angiography using convolutional neural networks. Radiology: Artificial Intelligence 2(4), e190178 (2020)
9. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al.: Pytorch: An imperative style, high-performance deep learning library. In: Advances in neural information processing systems. pp. 8026–8037 (2019)
10. Phellan, R., Forkert, N.D.: Comparison of vessel enhancement algorithms applied to time-of-flight mra images for cerebrovascular segmentation. Medical physics 44(11), 5901–5915 (2017)
11. Phellan, R., Lindner, T., Helle, M., Falcao, A.X., Forkert, N.D.: Automatic temporal segmentation of vessels of the brain using 4d asl mra images. IEEE transactions on biomedical engineering 65(7), 1486–1494 (2017)
12. Qin, Y., Zheng, H., Gu, Y., Huang, X., Yang, J., Wang, L., Yao, F., Zhu, Y.M., Yang, G.Z.: Learning tubule-sensitive cnns for pulmonary airway and artery-vein segmentation in ct. IEEE Transactions on Medical Imaging 40(6), 1603–1617 (2021)
13. Robben, D., Türetken, E., Sunaert, S., Thijs, V., Wilms, G., Fua, P., Maes, F., Suetens, P.: Simultaneous segmentation and anatomical labeling of the cerebral vasculature. Medical image analysis 32, 201–215 (2016)
14. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
15. Roth, G.A., Mensah, G.A., Johnson, C.O., Addolorato, G., Ammirati, E., Baddour, L.M., Barengo, N.C., Beaton, A.Z., Benjamin, E.J., Benziger, C.P., et al.: Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the gbd 2019 study. Journal of the American College of Cardiology 76(25), 2982–3021 (2020)
16. Sluimer, I., Schilham, A., Prokop, M., van Ginneken, B.: Computer analysis of computed tomography scans of the lung: a survey. IEEE Transactions on Medical Imaging 25(4), 385–405 (2006). [https://doi.org/10.1109/TMI.2005.862753]
17. Su, J., Wolff, L., van Es, A.C.M., Van Zwam, W., Majoie, C., Dippel, D.W., Van Der Lugt, A., Niessen, W.J., Van Walsum, T.: Automatic collateral scoring from 3d cta images. IEEE transactions on medical imaging 39(6), 2190–2200 (2020)
18. Su, R., Cornelissen, S.A.P., van der Sluijs, M., van Es, A.C.G.M., van Zwam, W.H., Dippel, D.W.J., Lycklama, G., van Doormaal, P.J., Niessen, W.J., van der Lugt, A., van Walsum, T.: autotici: Automatic brain tissue reperfusion scoring on 2d disa images of acute ischemic stroke patients. IEEE Transactions on Medical Imaging 40(9), 2380–2391 (2021). [https://doi.org/10.1109/TMI.2021.3077113]
19. Tan, W., Zhou, L., Li, X., Yang, X., Chen, Y., Yang, J.: Automated vessel segmentation in lung ct and cta images via deep neural networks. Journal of X-ray science and technology (Preprint), 1–15 (2021)
20. Van Rikxoort, E.M., Van Ginneken, B.: Automated segmentation of pulmonary structures in thoracic computed tomography scans: a review. Physics in Medicine & Biology 58(17), R187 (2013)
21. Zhang, M., Zhang, C., Wu, X., Cao, X., Young, G.S., Chen, H., Xu, X.: A neural network approach to segment brain blood vessels in digital subtraction angiography. Computer methods and programs in biomedicine 185, 105159 (2020)