CAVE

Cerebral artery–vein segmentation in digital subtraction angiography

Su, Ruisheng; van der Sluijs, P. Matthijs; Chen, Yuan; Cornelissen, Sandra; van den Broek, Ruben; van Zwam, Wim H.; van der Lugt, Aad; Niessen, Wiro J.; Ruijters, Danny; van Walsum, Theo

Citation (APA)
Su, R., van der Sluijs, P. M., Chen, Y., Cornelissen, S., van den Broek, R., van Zwam, W. H., van der Lugt, A., Niessen, W. J., Ruijters, D., & van Walsum, T. (2024). CAVE: Cerebral artery–vein segmentation in digital subtraction angiography. *Computerized Medical Imaging and Graphics*, *115*, Article 102392. https://doi.org/10.1016/j.compmedimag.2024.102392
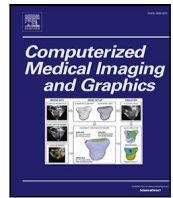
Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

CAVE: Cerebral artery–vein segmentation in digital subtraction angiography☆

Ruisheng Su [a,*], P. Matthijs van der Sluijs [a], Yuan Chen [b], Sandra Cornelissen [a], Ruben van den Broek [a], Wim H. van Zwam [c], Aad van der Lugt [a], Wiro J. Niessen [a,d], Danny Ruijters [e], Theo van Walsum [a]

[a] *Biomedical Imaging Group Rotterdam, Department of Radiology & Nuclear Medicine, Erasmus MC, University Medical Center Rotterdam, The Netherlands*
[b] *Department of Radiology & Nuclear Medicine, UMass Chan Medical School, Worcester, USA*
[c] *Department of Radiology & Nuclear Medicine, Maastricht UMC, Cardiovascular Research Institute Maastricht, The Netherlands*
[d] *Imaging Physics, Applied Sciences, Delft University of Technology, The Netherlands*
[e] *Philips Healthcare, Best, The Netherlands*

ARTICLE INFO

ABSTRACT

Cerebral X-ray digital subtraction angiography (DSA) is a widely used imaging technique in patients with neurovascular disease, allowing for vessel and flow visualization with high spatio-temporal resolution. Automatic artery–vein segmentation in DSA plays a fundamental role in vascular analysis with quantitative biomarker extraction, facilitating a wide range of clinical applications. The widely adopted U-Net applied on static DSA frames often struggles with disentangling vessels from subtraction artifacts. Further, it falls short in effectively separating arteries and veins as it disregards the temporal perspectives inherent in DSA. To address these limitations, we propose to simultaneously leverage spatial vasculature and temporal cerebral flow characteristics to segment arteries and veins in DSA. The proposed network, coined CAVE, encodes a 2D+time DSA series using spatial modules, aggregates all the features using temporal modules, and decodes it into 2D segmentation maps. On a large multi-center clinical dataset, CAVE achieves a vessel segmentation Dice of 0.84 ($\pm$0.04) and an artery–vein segmentation Dice of 0.79 ($\pm$0.06). CAVE surpasses traditional Frangi-based k-means clustering (P < 0.001) and U-Net (P < 0.001) by a significant margin, demonstrating the advantages of harvesting spatio-temporal features. This study represents the first investigation into automatic artery–vein segmentation in DSA using deep learning. The code is publicly available at https://github.com/RuishengSu/CAVE_DSA.

1. Introduction

1.1. Clinical background

Cerebrovascular diseases are a major contributor to global mortality and long-term disability (Roth et al., 2020). These diseases encompass a range of conditions, including ischemic stroke due to vessel occlusion, stenosis, and aneurysms. In order to diagnose and treat such conditions, dynamic imaging of cerebral blood vessels is conducted using X-ray digital subtraction angiography (DSA). DSA provides a means of visualizing blood flow dynamics and changes in vasculature appearance over time (Fig. 1), thereby offering valuable information for diagnosis,

procedural navigation, therapeutic decision-making, and evaluation of treatment outcomes.

DSA images are conventionally examined visually by neuroradiologists and interventionalists, which could be laborious, subjective, qualitative, and vulnerable to error. Automatic segmentation of arteries and veins promises to assist in this assessment by highlighting and quantifying vascular changes, providing a foundation for a range of downstream clinical applications, including quantitative evaluation of endovascular thrombectomy, automatic emboli detection, and image guidance for real-time endovascular navigation. For example, automatic artery–vein segmentation can be valuable for providing a venous roadmap for navigation during transvenous procedures, such
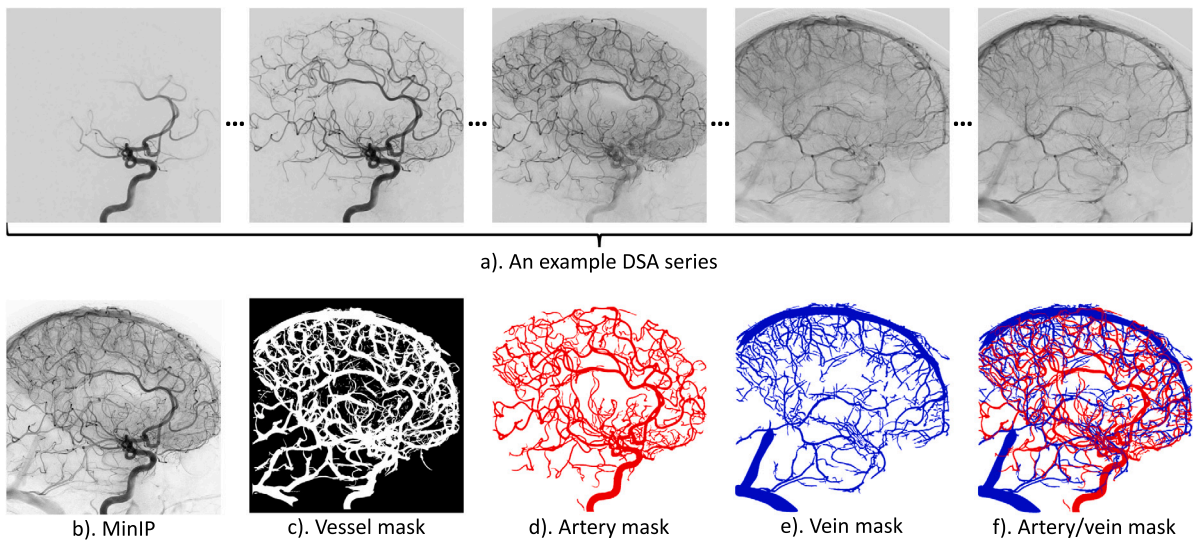
Fig. 1. Task illustration of artery–vein segmentation in DSA. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

as dural venous sinus stenting or transvenous embolization of dural arteriovenous fistula or arteriovenous malformations. The segmented arteries and veins provide a wealth of quantitative data that can be used to extract and analyze various blood flow-related biomarkers for peri-operative decision-making and post-operative prognosis.

DSA is a dynamic imaging technique that provides a visual representation of blood flow over time through a series of consecutive frames (Fig. 1). Although existing semantic segmentation networks could be utilized for end-to-end artery–vein segmentation, most methods only consider static frames. When addressing artery–vein segmentation in DSA series, which are 2D+time image sequences, the temporal dimension is relevant. We hypothesize that effectively incorporating spatio-temporal flow dynamics is key for achieving accurate artery–vein segmentation in DSA. The aim of this work, therefore is to harness the high-resolution contrast flow dynamics of DSA for improved artery–vein segmentation through spatio-temporal learning techniques.

1.2. Related work

Vessel segmentation is an extensively studied field in medical imaging for over two decades, with recent advancements propelled by deep learning. Recent literature on vessel segmentation (Moccia et al., 2018) primarily addresses retinal (Fraz et al., 2012; Chen et al., 2021a), lung (Sluimer et al., 2006; Van Rikxoort and Van Ginneken, 2013; Tan et al., 2021), and cardiovascular imaging (Huang et al., 2023a,b). These studies demonstrate the adaptability of segmentation methods across various anatomical structures and imaging modalities. In brain imaging, various methods have been proposed to segment cerebral vessels on different image modalities such as MRA (Phellan and Forkert, 2017; Phellan et al., 2017; Robben et al., 2016), CT angiography (CTA) (Fu et al., 2020; Su et al., 2020), and DSA (Liu et al., 2018; Su et al., 2021; Vepa et al., 2022; Zhang et al., 2020), each modality posing unique challenges and necessitating tailored approaches. The prevalent use of U-Net (Ronneberger et al., 2015) in these studies underscores its versatility. However, existing methods mainly rely on spatial vasculature features learned from static 2D DSA frames without considering the complete series. It has been shown that U-Net tends to generate false positives in the presence of subtraction artifacts that appear similar to blood vessels (Zhang et al., 2020).

Artery–vein segmentation has been primarily explored in fundus images (Hemelings et al., 2019) and non-contrast lung CT images (Qin et al., 2021) using U-Net. However, the segmentation of cerebral arteries and veins remains an under-explored area, with limited studies

on 4D CTA (Meijs et al., 2020), 3D MRA images (Hilbert et al., 2020), and 3D-DSA (Raz et al., 2021) using 3D U-Net models. These methods predominantly leverage spatial features (Hilbert et al., 2020) or manually crafted temporal parameters (Meijs et al., 2020) to differentiate between arteries and veins. In a previous study (Van Asperen et al., 2022), we investigated the automatic classification of arteries and veins using k-means clustering, contingent on vessel enhancement and binarization via the Frangi filter. The quality of vessel segmentation through the Frangi filter depends on the choice of the binarization threshold, a parameter that exhibits high variability across different DSA images. Nevertheless, to the best of our knowledge, no fully automatic and robust artery–vein segmentation algorithm has been developed for DSA images yet.

1.3. Contributions

This work presents CAVE, the first automatic method for artery–vein segmentation in DSA using deep learning, establishing a new benchmark. CAVE takes a 2D+time video sequence with variable length as input and produces 2D artery–vein segmentations as output. We leverage U-Net for spatial vasculature representation and temporal modules to learn temporal cerebral contrast flow characteristics simultaneously. On a multi-center clinical dataset, we demonstrate the utility of deep learning in cerebral artery–vein segmentation. CAVE may facilitate fast, accurate, and objective interpretation of cerebral vasculature in DSA, thus assisting endovascular interventions in clinical practice.

The remainder of this paper is organized as follows: Section 2 describes the proposed artery–vein segmentation method and Section 3 details the experimental dataset and annotation process. The extensive experiments and results are presented in Section 4, and further discussed in Section 5. Finally, Section 6 summarizes the conclusions of this study.

2. Methods

A cerebral digital subtraction angiography (DSA) series consists of a sequence of X-ray images captured post-contrast media injection, subtracted from the initial frame (pre-contrast image). This sequence presents the arrival of contrast media and the cerebral blood flow dynamics over time. The primary objective of this work is to develop an automated method for segmenting arteries and veins from a given DSA series, obtaining 2D artery–vein label masks as shown in Fig. 1.
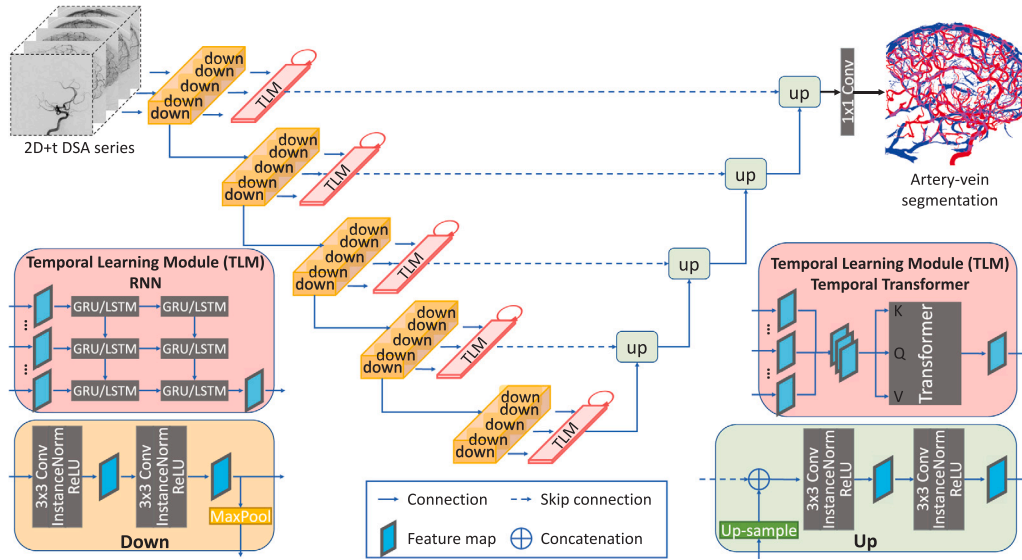
Fig. 2. Network architecture for artery–vein segmentation in DSA. Input is a 2D+time DSA series with variable series length. Output is a two-channel segmentation image with artery and vein represented by red and blue colors respectively. The temporal learning module (TLM) could be GRU, LSTM, or temporal transformer (TT). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Rather than relying on pixel-wise time-intensity curves (TICs) or vasculature appearance separately, we propose to simultaneously leverage spatial and temporal features to segment arteries and veins in an end-to-end way. The proposed network architecture (Fig. 2) takes DSA series of varying lengths as input. In the encoding path, each frame undergoes the same set of convolutional operations with shared weights, and the resulting features are temporally encoded to capture the spatio-temporal flow dynamics across all frames. These aggregated features are concatenated in the decoder path to generate high-resolution segmentation maps, resulting in a two-channel binary image that represents the segmented arteries and veins. The model is fully convolutional, thus allowing for input of varying sizes and temporal resolutions.

2.1. Spatial learning

The proposed CAVE (Fig. 2) employs a UNet-like architecture for spatial encoding and decoding of DSA frames, which comprises five down layers (yellow boxes) and five up layers (green boxes). Each layer utilizes double-convolution blocks with instance normalization and ReLU activation. Max pooling and bilinear upsampling are utilized for the contracting and expanding path respectively. The number of feature channels starts from 64 and doubles after each max pooling to offset spatial information loss and halves after each bilinear upsampling operation to reconstruct spatial resolution. In Fig. 2, it is only for illustration purposes that the down blocks appear to be replicated. In practice, all frames of the DSA series share the same down block for feature extraction, thereby avoiding an exponential increase in training parameters.

2.2. Temporal learning

We extend our approach beyond spatial representation learning by incorporating temporal learning techniques to improve the accuracy of artery–vein segmentation in DSA. Specifically, we explore three state-of-the-art techniques: convolutional GRU (ConvGRU) (Ballas et al., 2015), convolutional LSTM (ConvLSTM) (Shi et al., 2015), and temporal transformer (TT) based on (Vaswani et al., 2017). The convGRU and convLSTM architectures are designed to address the vanishing gradient problem (Hochreiter, 1998) in long sequences. The convGRU employs update and reset gates, while the convLSTM uses forget, input, and

output gates for a similar purpose. On the other hand, the transformer model leverages its attention mechanism to focus selectively on various segments of the input sequence, which could be particularly beneficial for understanding temporal relationships in the application of artery–vein segmentation in DSA. These techniques allow for the processing of input series of varying lengths and encoding them into fixed-sized representations, thereby enabling temporal aggregation of information. We apply this feature extraction and temporal aggregation process in all five layers, which operate on multi-scale image features.

Recurrent neural networks (RNN), such as GRU and LSTM, have been widely used in learning sequential information flow using a gate mechanism. In this work, we tailor ConvGRU and ConvLSTM to distinguish vessels from subtraction artifacts. GRU and LSTM models are employed to process sequential frame data, capturing temporal dependencies. The high-level architecture is shown in Fig. 2. The module takes the spatial features of size $B \times T \times C \times H \times W$ generated from the U-Net encoders of each frame as input, and produces aggregated feature maps of size $B \times C \times H \times W$ that align in dimensions with the corresponding decoder branch. B, T, C, H, W represents the batch size, frame length, channel size, image height, and image width respectively. In this work, the ConvGRU and ConvLSTM modules each consist of two convolutional RNN layers with a kernel size of 3×3 and with hidden dimensions equal to input dimensions. Such an end-to-end trained module is designed to learn complex spatio-temporal features. It is worth noting that this module allows a variable number of input frames.

Besides RNN, Transformer (Vaswani et al., 2017) has been shown effective in numerous deep learning tasks including image segmentation. We developed a temporal transformer module, as sketched in Fig. 2, to learn aggregate temporal features through temporal attention. The temporal transformer module, with its temporal attention capability, is utilized to selectively emphasize pertinent frames in a DSA series to enhance the differentiation between arteries, veins, and subtraction artifacts based on temporal flow dynamics in DSA. The module takes as input the UNet-encoded feature maps of all frames, and aggregates temporal features along the time dimension with a kernel of $1 \times 1 \times T$. The resulting feature map maintains the same size as a single-frame feature map, which is then concatenated with the decoding branch to construct the high-resolution segmentation map. In this work, the temporal transformer module consists of one multi-head attention layer with a kernel size of 1×1 and with hidden dimensions equal to input dimensions.

2.3. Loss function

As suggested in Isensee et al. (2021), we define the loss function for vessel segmentation ($\mathcal{L}_{vessel}(p, g)$) as a combination of a cross-entropy loss and a Dice loss (Kervadec and de Bruijne, 2023):

$$\mathcal{L}_{vessel}(p, g) = \mathcal{L}_{CE}(p, g) + \mathcal{L}_{DSC}(p, g), \tag{1}$$

where p and g denote the predicted class probability and the reference binary label respectively. For artery–vein segmentation, we similarly define the loss function $\mathcal{L}_{av}(p, g)$ as

$$\begin{aligned}\mathcal{L}_{av}(p, g) =& \mathcal{L}_{CE}(p_a, g_a) + \mathcal{L}_{CE}(p_v, g_v) \\ &+ \mathcal{L}_{DSC}(p_a, g_a) + \mathcal{L}_{DSC}(p_v, g_v). \end{aligned} \tag{2}$$

The subscript "a" and "v" denote artery and vein respectively. More specifically, the cross-entropy loss $\mathcal{L}_{CE}(p, g)$, defined as

$$\mathcal{L}_{CE}(p, g) = -g \log(p) + (1 - g) \log(1 - p), \tag{3}$$

measures the difference between the reference labels and predicted class probabilities. The Dice loss $\mathcal{L}_{DSC}(p_a, g_a)$ and $\mathcal{L}_{DSC}(p_v, g_v)$ are defined as

$$\mathcal{L}_{DSC}(p, g) = 1 - \frac{2 \sum_{i \in \Omega} g^{(i)} p^{(i)}}{\sum_{i \in \Omega} \left[g^{(i)} + p^{(i)} \right]}, \tag{4}$$

where Ω denotes the subset of the image space where g is positive.

3. Data

The study uses data from the MR CLEAN Registry (Jansen et al., 2018). It is an observational cohort study that included patients with acute ischemic stroke from sixteen centers in the Netherlands between March 2014 and December 2018. From this multi-center clinical registry, we manually segmented arteries and veins on 97 DSA series from different patients with either anterior–posterior (AP) or lateral views. The DSA series were acquired using various imaging systems (e.g., Philips, GE, and Siemens). The size of the individual frames of the acquired DSA series is 1024×1024 pixels. The series have different lengths, ranging from 10 to 50 frames, and varying temporal resolutions ranging from 0.5 to 4 frames per second.

The artery–vein annotations were created using an in-house developed tool in MeVisLab (Heckel et al., 2009) by four trained clinical students. To reduce inter-observer variability, the annotations were further refined by another trained student. An experienced radiologist was available for consultation during annotation. As visualized in Fig. 1, the annotation results in two segmentation images, with arterial annotation in red and venous annotation in blue. Overlapping pixels have both labels as the two annotations are independent. These annotations serve as the reference standard in this work.

4. Experiments and results

4.1. Implementation details

The proposed methods were implemented in PyTorch (Paszke et al., 2019), all trained on an NVIDIA RTX A40 GPU. As pre-processing, we resized all frames to 512×512 pixels, linearly 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 $\in [-5\%, 5\%]$ range, scaling $\in [-5\%, 5\%]$ range, and rotation $\in [-10°, 10°]$ range, were randomly applied during training with a probability of 0.5 for each.

We randomly split the dataset into training, validation, and testing set with a ratio of 50%-20%-30% on the patient level. This resulted in 52 DSA series for training, 19 for validation, and 26 for testing. The models were trained using RMSprop optimization (Hinton et al., 2012) and a ReduceLROnPlateau (Paszke et al., 2019) scheduler with a patience of 10 epochs, a decay factor of 0.5, and an initial learning rate of 1×10^{-5}. An early stopping strategy was applied with a patience of 50 epochs and a maximum of 1000 epochs.

4.2. Baselines

To benchmark the performance of CAVE and comprehensively assess the added value of simultaneous spatio-temporal learning in distinguishing vessels from subtraction artifacts, we implemented three representative solutions, and evaluated the CAVE against those.

Frangi+k-means. We implemented a conventional two-step artery–vein classification method (Van Asperen et al., 2022) that combines the Frangi filter and k-means clustering. First, the Frangi filter is applied to the static minimum intensity map (MinIP) of an input DSA, followed by fixed thresholding to obtain a binary vessel mask. Subsequently, all the vessel pixels within the mask are classified into arteries or veins. This clustering is achieved through k-means clustering with $k = 2$ using a set of intensity-based features derived from the time-intensity curve (TIC) of each pixel. These features include the area under the TIC curve (AUC), peak intensity, time to peak (TTP), arrival time, peak width, variance, standard deviation, and maximum uptake slope of the TIC. To ensure uniformity in the clustering process and mitigate the influence of varying magnitudes, all features are normalized to a range ranging from 0 to 1. Finally, the cluster with the lowest average TTP is identified as arterial, while the one with the highest average TTP is designated as venous.

U-Net semantic segmentation. Apart from classical machine learning, a standard U-Net could be directly utilized for identifying arteries and veins leveraging the spatial vascular features. We implemented a deep learning baseline using U-Net (Ronneberger et al., 2015; Zhang et al., 2020). The architecture is similar to the architecture in Fig. 2 with the contracting path replaced by stand-alone down blocks. To have all vessels present in the spatial dimensions, this U-Net approach uses the MinIP image of all frames of a DSA series as input. Consequently, U-Net purely relies on spatial vasculature features to differentiate between arteries and veins without considering the temporal contrast flow features.

U-Net+k-means. To utilize both spatial and temporal features for artery–vein segmentation, we further developed a two-stage baseline approach that cascades U-Net and k-means clustering that sequentially leverages spatial and temporal information. The U-Net encodes a MinIP image and produces a binary vessel segmentation. The vessel pixels are subsequently clustered into arteries and veins via k-means clustering on the time-intensity curves.

4.3. Evaluation metrics

We assess these methods on a hold-out test set composed of 26 DSA sequences from various patients from two perspectives: vessel segmentation and artery–vein segmentation (Table 1). We report the Dice coefficients, accuracy, sensitivity, and specificity for vessel segmentation, respectively defined below.

$$Dice = \frac{2 \, TP}{2 \, TP + FP + FN}, \tag{5}$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \tag{6}$$

$$Sens = \frac{TP}{TP + FN}, \tag{7}$$

$$Spec = \frac{TN}{TN + FP}. \tag{8}$$

Table 1

Performance of CAVE and other existing methods in vessel and artery–vein segmentation on the test set. Acc: accuracy, Sens: sensitivity, Spec: specificity, A-Dice: Artery Dice (Eq. (9)), V-Dice: Vein Dice (Eq. (10)), M-Dice: Multi-class Dice (Eq. (11)). The accuracy, sensitivity, and specificity of artery–vein segmentation are over both artery and vein classes.

Method	Vessel segmentation				Artery–vein segmentation					
	Acc	Sens	Spec	Dice	Acc	Sens	Spec	A-Dice	V-Dice	M-Dice
Frangi + k-means	0.83 ±0.048	0.69 ±0.13	0.89 ±0.088	0.70 ±0.094	0.82 ±0.050	0.63 ±0.11	0.89 ±0.083	0.65 ±0.077	0.54 ±0.12	0.60 ±0.084
U-Net	0.89 ±0.032	0.79 ±0.071	0.93 ±0.029	0.80 ±0.050	0.87 ±0.041	0.68 ±0.078	0.94 ±0.028	0.70 ±0.082	0.63 ±0.070	0.67 ±0.060
U-Net + k-means	0.89 ±0.032	0.79 ±0.071	0.93 ±0.029	0.80 ±0.050	0.87 ±0.035	0.73 ±0.071	0.93 ±0.028	0.73 ±0.060	0.65 ±0.067	0.69 ±0.050
CAVE (Temporal Transformer)	0.91 ±0.029	0.83 ±0.064	0.94 ±0.025	0.84 ±0.041	0.90 ±0.032	0.79 ±0.071	0.95 ±0.024	0.82 ±0.054	0.74 ±0.071	0.78 ±0.048
CAVE (ConvLSTM)	0.91 ±0.025	0.83 ±0.060	0.94 ±0.025	0.83 ±0.048	0.90 ±0.031	0.80 ±0.075	0.95 ±0.022	0.82 ±0.062	0.75 ±0.058	0.79 ±0.052
CAVE (ConvGRU)	0.91 ±0.027	0.84 ±0.062	0.94 ±0.022	0.84 ±0.039	0.90 ±0.035	0.79 ±0.078	0.95 ±0.024	0.82 ±0.077	0.76 ±0.054	0.79 ±0.057

For artery–vein segmentation, we additionally evaluate the artery Dice (*A-Dice*), vein Dice (*V-Dice*), and multi-class Dice (*M-Dice*). We define *A-Dice* and vein Dice *V-Dice* as

$$A\text{-}Dice = \frac{2\,TP_a}{2\,TP_a + FP_a + FN_a} \quad , \tag{9}$$

and

$$V\text{-}Dice = \frac{2\,TP_v}{2\,TP_v + FP_v + FN_v} \quad , \tag{10}$$

where the subscript a and v denote artery and vein respectively. TP, FP, and FN are true positive, false positive, and false negative respectively. We define the multi-class Dice (*M-Dice*) as

$$M\text{-}Dice = \frac{2\,(TP_a + TP_v)}{2\,(TP_a + TP_v) + FP_a + FP_v + FN_a + FN_v}. \tag{11}$$

Besides, we compute statistical significance using the paired Wilcoxon test on the Dice coefficients.

4.4. Quantitative analysis

In Table 1, we present the results of CAVE and other representative solutions on the test set in terms of vessel segmentation and artery–vein segmentation. Regarding vessel segmentation, both U-Net and CAVE significantly outperform the Frangi+k-means approach in terms of Dice coefficient with a margin of 10% (P<0.001). Notably, CAVE substantially surpasses U-Net by 3.6% (P = 0.023) and the Frangi+k-means approach by 14% (P<0.001), demonstrating the effectiveness of incorporating simultaneous spatio-temporal vascular and flow representation for separating vessels from subtraction artifacts or other static instruments.

With respect to artery–vein segmentation, the advantage of spatio-temporal learning is even more prominent as shown in Table 1. While Frangi+k-means, by relying solely on pixel-wise temporal characteristics, achieves a multi-class Dice coefficient of 0.60 ($\pm$0.084), spatial feature-based deep learning (U-Net) obtains a higher Dice coefficient of 0.67 ($\pm$0.060). The U-Net+k-means approach, which leverages spatial and temporal features in two sequential stages, achieves a Dice coefficient of 0.69 ($\pm$0.050). In contrast, by integrating both vasculature appearance and flow dynamics in an end-to-end spatio-temporal deep learning model, CAVE yields a significantly higher multi-class Dice coefficient of 0.79 ($\pm$0.057) compared to U-Net (P<0.001) and Frangi+k-means (P<0.001). Overall, CAVE demonstrates significant improvements in artery–vein segmentation by incorporating the temporal contrast flow aspects of DSA in end-to-end learning. We observe no statistically significant differences among ConvGRU, ConvLSTM, and the temporal transformer.

In terms of computational efficiency, CAVE takes on average 0.57 ± 0.2 s to segment a complete DSA series on an NVIDIA 2080 Ti GPU with 11 GB of memory.

Table 2

Performance of CAVE with respect to various temporal resolutions in frames per second (fps).

fps	1	0.5	0.25
M-Dice	0.79 ± 0.056	0.76 ± 0.060	0.64 ± 0.12

4.5. Impact of temporal resolution

We investigate the influence of temporal characteristics, specifically the number of frames of a DSA series and the temporal frequency, on artery–vein segmentation performance. As shown in Fig. 5, we do not find a notable correlation between the average Dice coefficient and the frame count of DSA series in the test set, evidenced by a Pearson correlation coefficient of 0.05 (P = 0.81). This could be attributed to the fact that all DSA series have the same temporal frequency and are complete in our dataset. In contrast, we observe a notable reduction in segmentation performance when the DSA series are temporally down-sampled for testing, as shown in Table 2. This suggests that the proposed CAVE method effectively utilizes temporal flow characteristics to enhance segmentation performance.

4.6. Qualitative analysis

To provide a comprehensive comparison between the proposed CAVE and U-Net, we present representative visualizations and qualitative evaluations. CAVE is particularly effective in distinguishing cerebral vessels from the subtraction artifacts, instruments, and subtraction artifacts, thanks to its ability to learn temporal contrast flow dynamics. In Fig. 3, we provide visual comparisons of vessel segmentation results using U-Net and CAVE. The error maps, shown in columns c and e, use orange to indicate false positives and light blue to indicate false negatives. We identify two scenarios. First, in regions #1 and #3, where U-Net mistakenly recognizes subtraction artifacts and static instruments as vessels (column c), CAVE correctly avoids such misclassifications (column e). Furthermore, CAVE successfully identifies venous vessels, as shown in regions #2 and #4, even when they are surrounded by subtraction artifacts based on their temporal characteristics, whereas U-Net fails to detect them in the presence of noise in the background.

Fig. 4 presents two examples illustrating the performance of U-Net and CAVE in artery–vein segmentation. Compared to manual annotations (column a), CAVE (column e) shows fewer errors than U-Net (column c). While U-Net performs well in recognizing large vessels such as proximal arteries and the superior sagittal sinus (SSS), it struggles to correctly classify distal arteries and veins, as highlighted in regions #1 and #2. In contrast, CAVE achieves accurate classification in both small
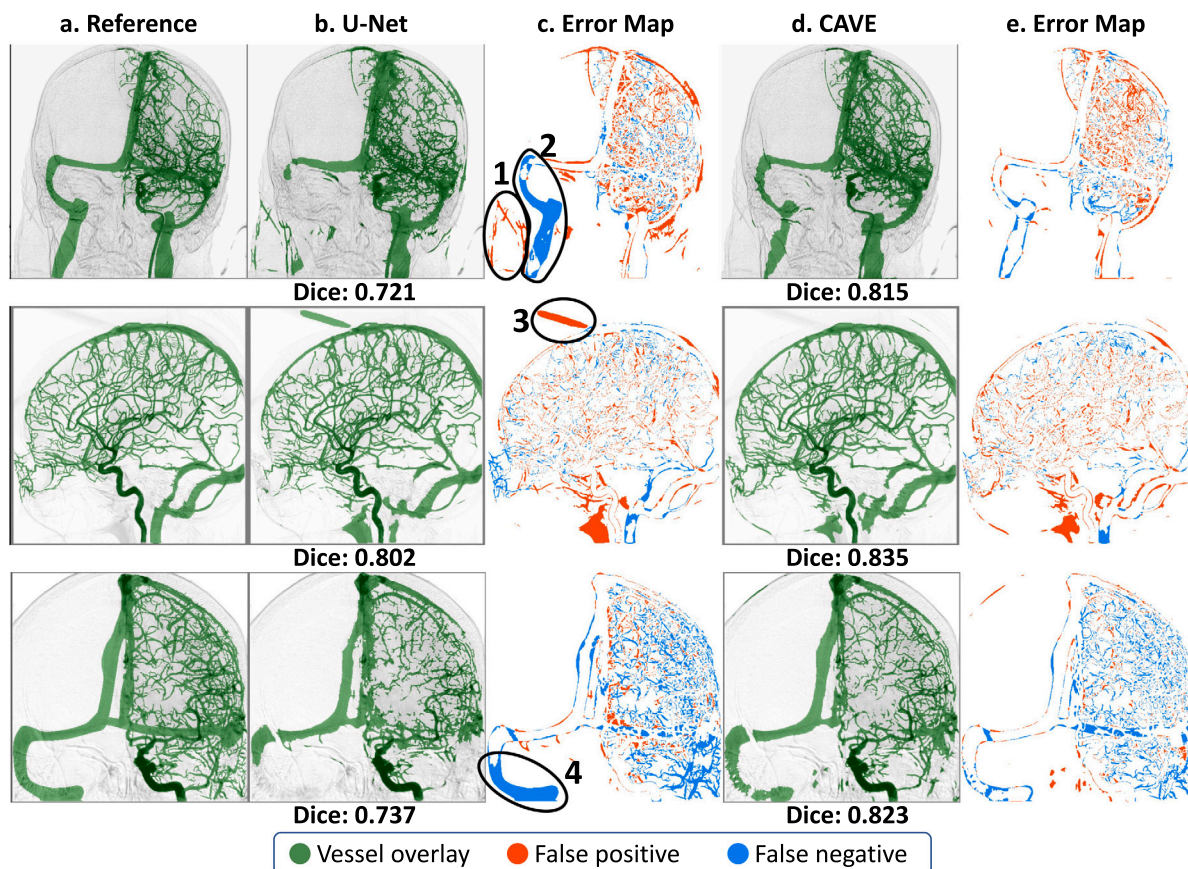
Fig. 3. Example visualizations of cerebral vessel segmentation results of U-Net and the proposed CAVE. Column a: manual annotation of vessels overlaid on the MinIP image; column b: segmentation output of U-Net; column c: U-Net error map with false positives (orange) and false negatives(light blue); column d: segmentation output of CAVE; column e: CAVE error map with false positives (orange) and false negatives (light blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and large vessels, demonstrating its effectiveness in capturing spatial and temporal features. Additional visual comparisons of these methods are available in the appendix.

5. Discussion

In this work, we proposed a fully automated deep learning-based method for artery–vein segmentation in cerebral DSA images. Through quantitative and qualitative comparisons, we have demonstrated the added value of simultaneous spatio-temporal learning (CAVE) against spatial learning (U-Net), temporal learning (k-means), and sequential two-stage spatial and temporal learning (U-Net + k-means).

This application-tailored solution innovatively leverages the distinctive temporal contrast flow characteristics in DSA to precisely identify cerebral arteries and veins while alleviating misclassifications of subtraction artifacts and other surgical instruments. The primary focus of the analyses revolves around evaluating the inherent benefits of harnessing temporal information. For quantitative assessments, we employ the foundational U-Net as our benchmark. With the rapidly evolving landscape of deep network architectures, the U-Net may not be positioned among the most recent models in the current literature. Recent advancements in state-of-the-art networks, such as UNet++ (Zhou et al., 2018), PointRend (Kirillov et al., 2020), nnU-Net (Isensee et al., 2021), TransUNet (Chen et al., 2021b), or CTransNet (Wang et al., 2022) may outperform U-Net, when leveraging purely spatial features from MinIP images. Nevertheless, adapting the baseline spatial learning module and capitalizing on the abundant temporal flow dynamics are orthogonal and mutually reinforcing efforts to improve performance.

CAVE is designed to take complete DSA series with variable lengths and output an overall 2D artery–vein segmentation map for an input DSA series. The explored TLM modules are flexible in terms of input dimensions. These modules are capable of handling DSA series with variable lengths, while alternative methods such as 3D U-Net, direct feature concatenating, or temporal convolutional network (TCN) (Lea et al., 2016) typically cannot. It necessitates temporal sliding or temporal padding/cropping to utilize models that require fixed frame length. Such models would also require temporarily varying annotations to train.

In comparison with the hold-out strategy, K-fold cross-validation may provide a more robust assessment of the described methods in this manuscript. We tested the proposed CAVE (with ConvGRU) method on five random data splits and obtained M-Dice coefficients of 0.79 (± 0.057), 0.79 (± 0.052), 0.78 (± 0.043), 0.78 (± 0.049), and 0.78 (± 0.057). No significant performance differences (P = 0.4, Kruskal–Wallis test) were observed among these splits. This indicates that the hold-out method allows for reliable performance estimates. Considering this observation, alongside the substantial computational demands of cross-validation, we opted for the hold-out strategy in subsequent performance evaluations.

In this study, we observe that, despite the differences in network design, the TLM modules (i.e., ConvGRU, ConvLSTM, and Temporal Transformer) perform comparably in the application of artery–vein segmentation on our dataset. This could be attributed to their shared capability to capture temporal dependencies. While the temporal transformer is often considered superior (Vaswani et al., 2017; Devlin et al., 2018; Dai et al., 2019), its effectiveness may not consistently extend to all scenarios. In this context, temporal information is relevant. All
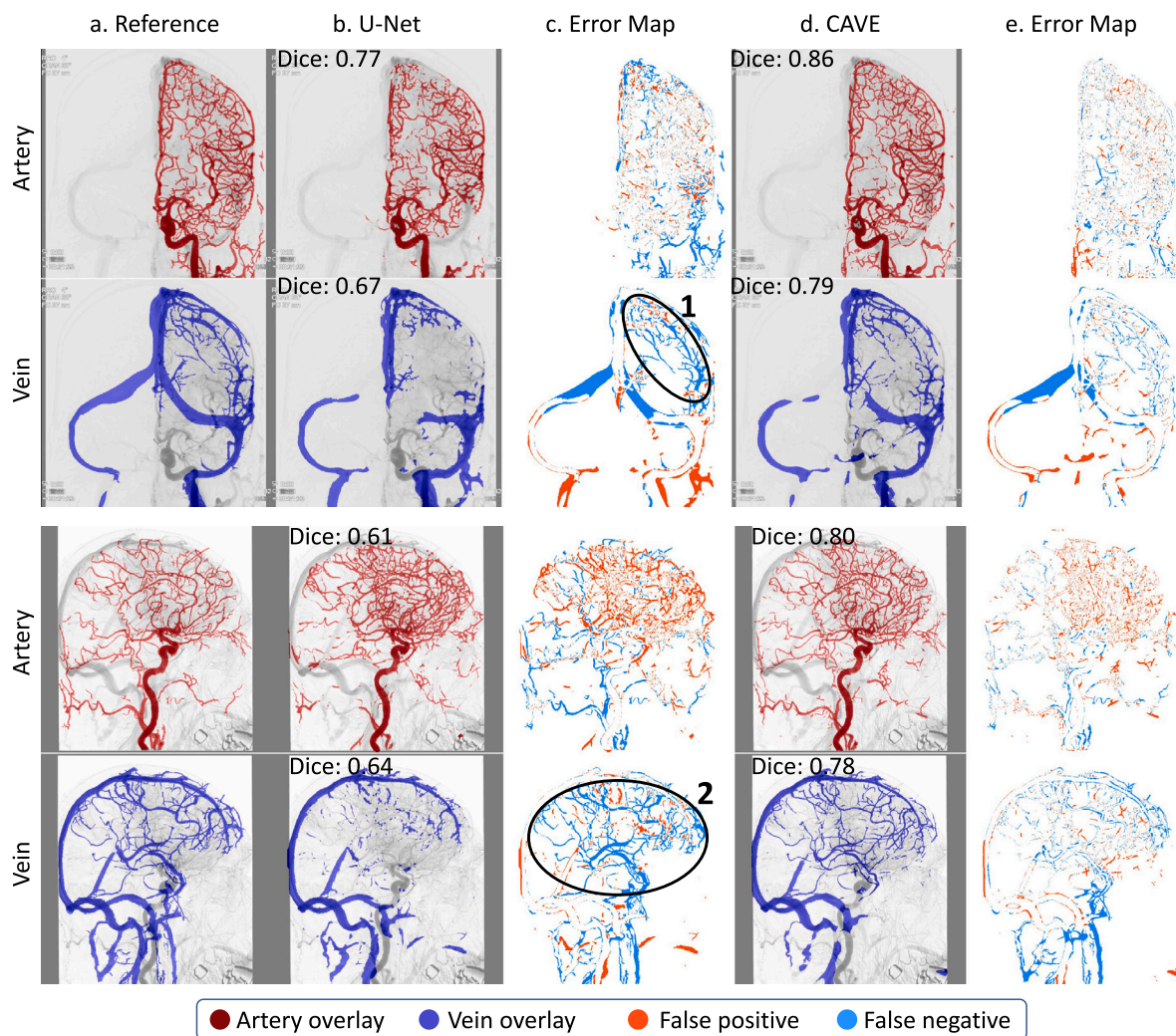
Fig. 4. Two example visualizations (in rows) of artery–vein segmentation. Column a: manual annotation of arteries (red) and veins (blue) overlaid on the MinIP image; column b: segmentation output of U-Net; column c: U-Net error map with false positives (orange) and false negatives(light blue); column d: segmentation output of CAVE; column e: CAVE error map with false positives (orange) and false negatives (light blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
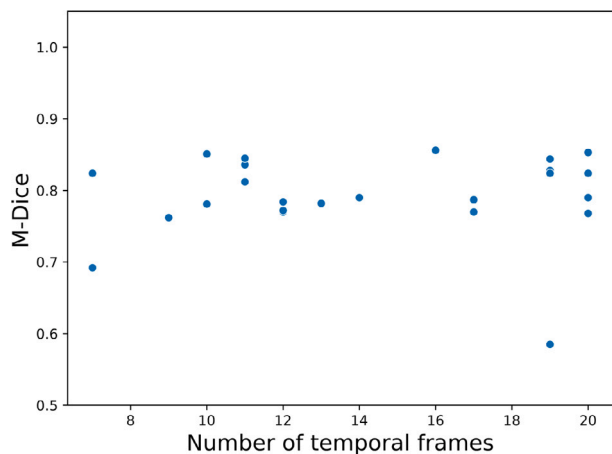

Fig. 5. Association between the segmentation Dice and the number of frames of DSA series in the test set.

This aligns with a broader understanding that the practical impact of architectural advantages can vary greatly depending on the specifics of the task and data.

We note that the segmentation Dice scores of veins are on average lower than those of arteries across all methods. This observation can be attributed to several factors: (1) venous vessels typically exhibit relatively lighter intensities and more ambiguous boundaries, making them challenging to be accurately predicted; (2) the quality of vein annotations may also be comparatively lower due to increased annotation difficulties; (3) there tends to be a scarcity of venous phase frames, especially in cases of incomplete DSA series when image acquisition is prematurely stopped by the operator.

Building upon this work, future research may investigate the influence of patient motion on those spatio-temporal techniques for artery–vein segmentation. From the clinical perspective, there is significant value in conducting thorough validations to assess the clinical generalizability of CAVE.

6. Conclusion

We have presented CAVE, a deep learning-based cerebral vessel and artery–vein segmentation method in digital subtraction angiography. It integrates both spatial vascular appearance and temporal
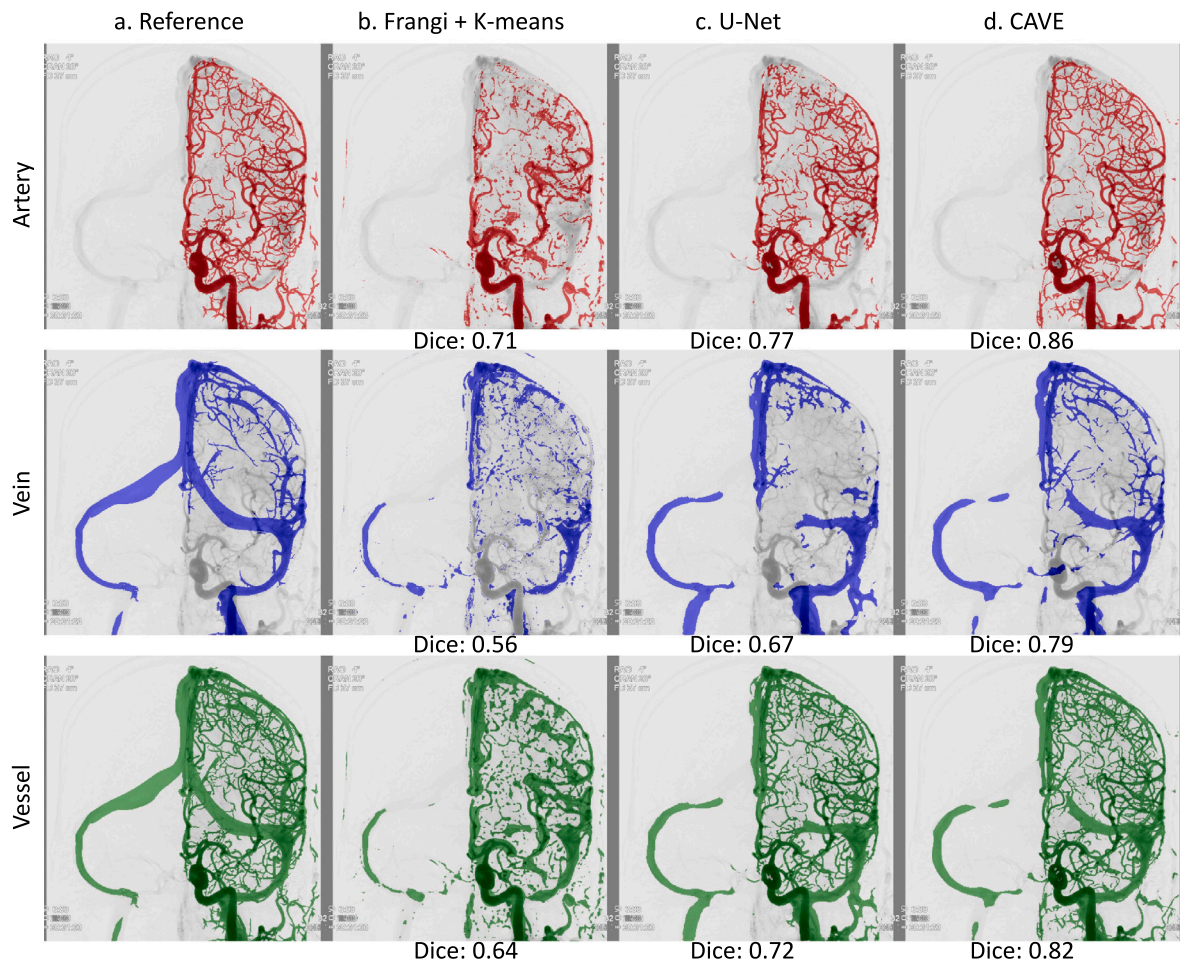
modules do the job of extracting temporal patterns, whereas the exact network design and configuration do not make a notable difference.

Fig. 6. Example visualization of artery, vein, and vessel segmentation. Column a: manual annotation of arteries (red), veins (blue), and vessels (green) overlaid on the MinIP image; column b: segmentation output of the Frangi + *k*-means approach; column c: segmentation output of U-Net; column d: segmentation output of CAVE.

contrast flow dynamics in a unified end-to-end framework, producing high-quality multi-class segmentations from cerebral DSA series with variable lengths. Experimental results on a multi-center clinical dataset demonstrate that CAVE significantly outperforms existing methods. Qualitative analyses further underpin the advantages of CAVE in distinguishing vessels from subtraction artifacts. CAVE has the potential to facilitate vessel-based quantitative analyses for clinical diagnosis, prognosis, and treatment planning in endovascular interventions.

CRediT authorship contribution statement

Ruisheng Su: Writing – original draft, Visualization, Validation, Software, Methodology, Data curation, Conceptualization. **P. Matthijs van der Sluijs:** Writing – review & editing, Validation, Data curation. **Yuan Chen:** Writing – review & editing, Visualization, Validation, Methodology, Formal analysis, Conceptualization. **Sandra Cornelissen:** Writing – review & editing, Validation, Supervision, Conceptualization. **Ruben van den Broek:** Writing – review & editing, Visualization, Formal analysis, Data curation. **Wim H. van Zwam:** Writing – review & editing, Validation, Supervision, Funding acquisition. **Aad van der Lugt:** Writing – review & editing, Visualization, Validation, Resources, Project administration, Funding acquisition. **Wiro J. Niessen:** Writing – review & editing, Visualization, Validation, Supervision. **Danny Ruijters:** Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition, Conceptualization. **Theo van Walsum:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: - Danny Ruijters is an employee of Philips Healthcare. - Wiro J. Niessen is founder, scientific lead, and shareholder of Quantb B. - Aad van der Lugt has received research support from GE HealthCare, Siemens Healthineers, Philips Healthcare, Stryker, Penumbra, Cerenovus, Medtronic, and Thrombolytic Sciences inc (all paid to institution).

Data availability

For data sharing policies, please refer to Jansen et al. (2018)

Appendix. Extended qualitative comparisons
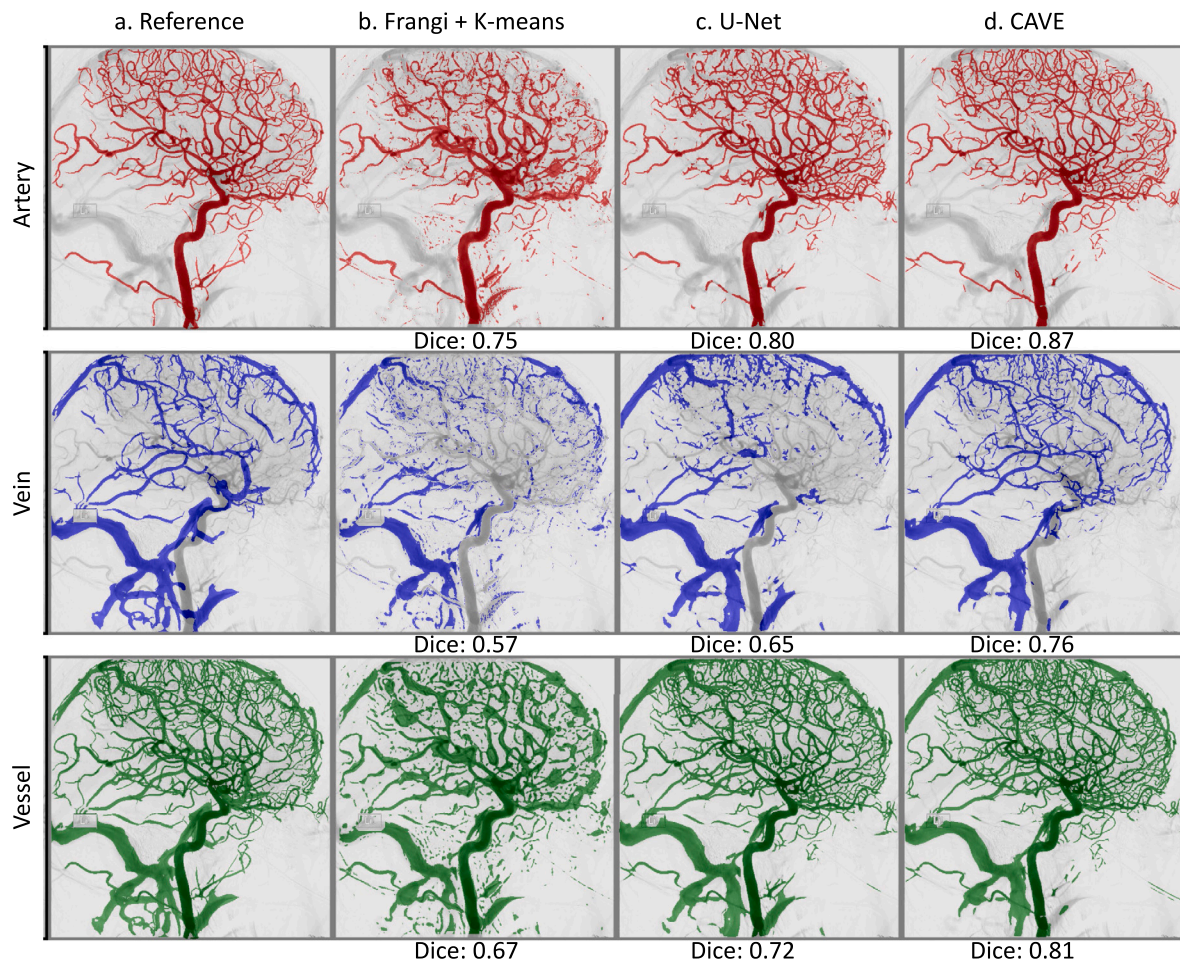
See Figs. 6 and 7.

Fig. 7. Example visualization of artery, vein, and vessel segmentation. Column a: manual annotation of arteries (red), veins (blue), and vessels (green) overlaid on the MinIP image; column b: segmentation output of the Frangi + *k*-means approach; column c: segmentation output of U-Net; column d: segmentation output of CAVE.

References

Ballas, N., Yao, L., Pal, C., Courville, A., 2015. Delving deeper into convolutional networks for learning video representations. arXiv preprint arXiv:1511.06432.

Chen, C., Chuah, J.H., Raza, A., Wang, Y., 2021a. Retinal vessel segmentation using deep learning: A review. IEEE Access.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y., 2021b. Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306.

Dai, Z., Yang, Z., Yang, Y., Carbonell, J., Le, Q.V., Salakhutdinov, R., 2019. Transformer-xl: Attentive language models beyond a fixed-length context. arXiv preprint arXiv:1901.02860.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Fraz, M.M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A.R., Owen, C.G., Barman, S.A., 2012. Blood vessel segmentation methodologies in retinal images–a survey. Comput. Methods Programs Biomed. 108 (1), 407–433.

Fu, F., Wei, J., Zhang, M., Yu, F., Xiao, Y., Rong, D., Shan, Y., Li, Y., Zhao, C., Liao, F., et al., 2020. Rapid vessel segmentation and reconstruction of head and neck angiograms using 3D convolutional neural network. Nat. Commun. 11 (1), 1–12.

Heckel, F., Schwier, M., Peitgen, H.-O., 2009. Object-oriented application development with MeVisLab and Python. Informatik 2009–Im Focus das Leben.

Hemelings, R., Elen, B., Stalmans, I., Van Keer, K., De Boever, P., Blaschko, M.B., 2019. Artery–vein segmentation in fundus images using a fully convolutional network. Comput. Med. Imaging Graph. 76, 101636.

Hilbert, A., Madai, V.I., Akay, E.M., Aydin, O.U., Behland, J., Sobesky, J., Galinovic, I., Khalil, A.A., Taha, A.A., Wuerfel, J., et al., 2020. BRAVE-NET: fully automated arterial brain vessel segmentation in patients with cerebrovascular disease. Front. Artif. Intell. 78.

Hinton, G., Srivastava, N., Swersky, K., 2012. Neural networks for machine learning—Lecture 6e—rmsprop: Divide the gradient by a running average of its recent magnitude. URL: https://www.cs.toronto.edu/tijmen/csc321/slides/lectureslideslec6.pdf.

Hochreiter, S., 1998. The vanishing gradient problem during learning recurrent neural nets and problem solutions. Internat. J. Uncertain. Fuzziness Knowledge-Based Systems 6 (02), 107–116.

Huang, Q., Jia, L., Ren, G., Wang, X., Liu, C., 2023a. Extraction of vascular wall in carotid ultrasound via a novel boundary-delineation network. Eng. Appl. Artif. Intell. 121, 106069.

Huang, Q., Zhao, L., Ren, G., Wang, X., Liu, C., Wang, W., 2023b. NAG-Net: Nested attention-guided learning for segmentation of carotid lumen-intima interface and media-adventitia interface. Comput. Biol. Med. 156, 106718.

Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H., 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. methods 18 (2), 203–211.

Jansen, I.G., Mulder, M.J., Goldhoorn, R.-J.B., 2018. Endovascular treatment for acute ischaemic stroke in routine clinical practice: prospective, observational cohort study (MR CLEAN registry). bmj 360.

Kervadec, H., de Bruijne, M., 2023. On the dice loss variants and sub-patching. In: Medical Imaging with Deep Learning, Short Paper Track.

Kirillov, A., Wu, Y., He, K., Girshick, R., 2020. Pointrend: Image segmentation as rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9799–9808.

Lea, C., Vidal, R., Reiter, A., Hager, G.D., 2016. Temporal convolutional networks: A unified approach to action segmentation. In: Computer Vision–ECCV 2016 Workshops: Amsterdam, the Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14. Springer, pp. 47–54.

Liu, B., Jiang, Q., Liu, W., Wang, M., Zhang, S., Zhang, X., Zhang, B., Yue, Z., 2018. A vessel segmentation method for serialized cerebralvascular DSA images based on spatial feature point set of rotating coordinate system. Comput. Methods Programs Biomed. 161, 55–72.

Meijs, M., Pegge, S.A., Vos, M.H., Patel, A., van de Leemput, S.C., Koschmieder, K., Prokop, M., Meijer, F.J., Manniesing, R., 2020. Cerebral artery and vein segmentation in four-dimensional CT angiography using convolutional neural networks. Radiol.: Artif. Intell. 2 (4), e190178.

Moccia, S., De Momi, E., El Hadji, S., Mattos, L.S., 2018. Blood vessel segmentation algorithms—review of methods, datasets and evaluation metrics. Comput. Methods Programs Biomed. 158, 71–91.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al., 2019. Pytorch: An imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 8026–8037.

Phellan, R., Forkert, N.D., 2017. Comparison of vessel enhancement algorithms applied to time-of-flight MRA images for cerebrovascular segmentation. Med. Phys. 44 (11), 5901–5915.

Phellan, R., Lindner, T., Helle, M., Falcao, A.X., Forkert, N.D., 2017. Automatic temporal segmentation of vessels of the brain using 4D ASL MRA images. IEEE Trans. Biomed. Eng. 65 (7), 1486–1494.

Qin, Y., Zheng, H., Gu, Y., Huang, X., Yang, J., Wang, L., Yao, F., Zhu, Y.-M., Yang, G.-Z., 2021. Learning tubule-sensitive CNNs for pulmonary airway and artery-vein segmentation in CT. IEEE Trans. Med. Imaging 40 (6), 1603–1617.

Raz, E., Shapiro, M., Mir, O., Nossek, E., Nelson, P., 2021. Arterial and venous 3D fusion AV-3D-DSA: a novel approach to cerebrovascular neuroimaging. Am. J. Neuroradiol. 42 (7), 1282–1284.

Robben, D., Türetken, E., Sunaert, S., Thijs, V., Wilms, G., Fua, P., Maes, F., Suetens, P., 2016. Simultaneous segmentation and anatomical labeling of the cerebral vasculature. Med. Image Anal. 32, 201–215.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

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., 2020. Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. J. Am. Coll. Cardiol. 76 (25), 2982–3021.

Shi, X., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.C., 2015. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. Adv. Neural Inf. Process. Syst. 2015, 802–810.

Sluimer, I., Schilham, A., Prokop, M., van Ginneken, B., 2006. Computer analysis of computed tomography scans of the lung: a survey. IEEE Trans. Med. Imaging 25 (4), 385–405. http://dx.doi.org/10.1109/TMI.2005.862753.

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., 2021. Autotici: Automatic brain tissue reperfusion scoring on 2D DSA images of acute ischemic stroke patients. IEEE Trans. Med. Imaging 40 (9), 2380–2391. http://dx.doi.org/10.1109/TMI.2021.3077113.

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., 2020. Automatic collateral scoring from 3D CTA images. IEEE Trans. Med. Imaging 39 (6), 2190–2200.

Tan, W., Zhou, L., Li, X., Yang, X., Chen, Y., Yang, J., 2021. Automated vessel segmentation in lung CT and CTA images via deep neural networks. J. X-ray Sci. Technol. (Preprint), 1–15.

Van Asperen, V., Van Den Berg, J., Lycklama, F., Marting, V., Cornelissen, S., Van Zwam, W.H., Hofmeijer, J., Van Der Lugt, A., Van Walsum, T., Van Der Sluijs, M., et al., 2022. Automatic artery/vein classification in 2D-DSA images of stroke patients. In: Medical Imaging 2022: Image-Guided Procedures, Robotic Interventions, and Modeling. Vol. 12034, SPIE, pp. 366–377.

Van Rikxoort, E.M., Van Ginneken, B., 2013. Automated segmentation of pulmonary structures in thoracic computed tomography scans: a review. Phys. Med. Biol. 58 (17), R187.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.

Vepa, A., Choi, A., Nakhaei, N., Lee, W., Stier, N., Vu, A., Jenkins, G., Yang, X., Shergill, M., Desphy, M., et al., 2022. Weakly-supervised convolutional neural networks for vessel segmentation in cerebral angiography. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 585–594.

Wang, H., Cao, P., Wang, J., Zaiane, O.R., 2022. Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 36, (3), pp. 2441–2449.

Zhang, M., Zhang, C., Wu, X., Cao, X., Young, G.S., Chen, H., Xu, X., 2020. A neural network approach to segment brain blood vessels in digital subtraction angiography. Comput. Methods Programs Biomed. 185, 105159.

Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J., 2018. Unet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4. Springer, pp. 3–11.