Medical Image Analysis 102 (2025) 103496

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Deep learning based coronary vessels segmentation in X-ray angiography using temporal information

Haorui He [a], Abhirup Banerjee [a,b,*], Robin P. Choudhury [b], Vicente Grau [a]

[a] Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford OX3 7DQ, United Kingdom
[b] Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford OX3 9DU, United Kingdom

ARTICLE INFO

Keywords:
Coronary vessels segmentation
Temporal information
X-ray coronary angiography
Nested encoder decoder
Vessel connectivity

ABSTRACT

Invasive coronary angiography (ICA) is the gold standard imaging modality during cardiac interventions. Accurate segmentation of coronary vessels in ICA is required for aiding diagnosis and creating treatment plans. Current automated algorithms for vessel segmentation face task-specific challenges, including motion artifacts and unevenly distributed contrast, as well as the general challenge inherent to X-ray imaging, which is the presence of shadows from overlapping organs in the background. To address these issues, we present Temporal Vessel Segmentation Network (TVS-Net) model that fuses sequential ICA information into a novel densely connected 3D encoder-2D decoder structure with a loss function based on elastic interaction. We develop our model using an ICA dataset comprising 323 samples, split into 173 for training, 82 for validation, and 68 for testing, with a relatively relaxed annotation protocol that produced coarse-grained samples, and achieve 83.4% Dice and 84.3% recall on the test dataset. We additionally perform an external evaluation over 60 images from a local hospital, achieving 78.5% Dice and 82.4% recall and outperforming the state-of-the-art approaches. We also conduct a detailed manual re-segmentation for evaluation only on a subset of the first dataset under strict annotation protocol, achieving a Dice score of 86.2% and recall of 86.3% and surpassing even the coarse-grained gold standard used in training. The results indicate our TVS-Net is effective for multi-frame ICA segmentation, highlights the network's generalizability and robustness across diverse settings, and showcases the feasibility of weak supervision in ICA segmentation.

1. Introduction

Invasive coronary angiography (ICA) is widely used in clinical practice for the diagnosis and interventions of coronary artery disease (CAD) (Lashgari et al., 2024). It serves as the gold standard for diagnosing coronary stenosis, while providing both qualitative and quantitative guidance for therapy (Kočka, 2015). The segmentation of coronary vessels enables clear visualization and calculation of quantitative clinical metrics such as fractional flow reserve (Tu et al., 2014). In ICA, the acquired vascular structures always overlap with background structures. Images are also affected by body movements and unevenly distributed contrast, resulting in segmentation artifacts such as misaligned boundaries and disconnected vessels (Blondel et al., 2006). In this work, we address these challenges and develop a completely automated approach to accurately and effectively delineate coronary vessels from sequences of consecutive ICA images.

Vessel segmentation has received substantial attention in the literature. Traditional methods include vessel enhancement and vessel tracking techniques (Frangi et al., 1998; Fazlali et al., 2015; Jerman et al., 2016), which often fail in the presence of large scale and contrast variations and mislabel other structures in the ICA such as the catheter or ribs. Tracking algorithms are more commonly applied on vessel centerlines using topological information and prior information, and are sensitive to ICA artifacts (Zou et al., 2009; Ambrosini et al., 2015). Sophisticated non-machine learning methods involving optimization and matrix decomposition have been proposed in Shin et al. (2016) and Xia et al. (2020). Although their results are comparable to the state-of-the-art (SOTA), problems such as large standard deviation during testing and incompleteness in topology caused by disconnected vessels are not solved.

Machine learning methods, especially neural networks, have become SOTA in most medical image analysis applications, including segmentation. They usually involve variants of U-Net (Ronneberger et al., 2015) or fully convolutional network (FCN) (Long et al., 2015). The context in which vascular segmentation has received the most attention is retinal imaging (Kamran et al., 2021; Tan et al., 2022). Maninis et al. (2016) showed automated methods can achieve retinal vessel

* Corresponding author at: Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford OX3 7DQ, United Kingdom.
E-mail address: abhirup.banerjee@eng.ox.ac.uk (A. Banerjee).

https://doi.org/10.1016/j.media.2025.103496
Received 13 July 2024; Received in revised form 28 December 2024; Accepted 2 February 2025
Available online 18 February 2025

delineation comparable to annotations by human experts. However, algorithms for retinal vessel segmentation are not directly applicable to ICA due to its unique challenges involving grayscale imaging, irregular vessel structures, significant variations in vessel diameters, and motion artifacts. Additionally, manual annotation of the entire vascular tree, especially the thin vessels, is time-consuming and difficult due to human errors, time constraints, and the inherent difficulty of accurately delineating small and low-contrast vessels. They often lead to coarse-grained annotations and make dataset size and quality key challenges for improving ICA segmentation performance. The second main challenge is the quality of ICA images, where inconsistent contrast often makes vessels indistinguishable from the background. Factors such as the timing of contrast injection, blood flow speed, patient movement, and variations in contrast dispersion within the coronary arteries contribute to this inconsistency, complicating accurate segmentation. The final challenge is related to maintaining the structural integrity of the generated vascular tree, wherein minimizing disconnected vessels and avoiding over or under-segmentation of vessel boundaries is crucial to reduce the risk of misclassifying stenosis. Hence, specific resolutions need to be designed.

1.1. Related work

Nasr-Esfahani et al. (2018) proposed a coronary vessel segmentation pipeline that includes an image enhancement module, a network for contextual feature extraction, and a network for probability feature extraction. Jiang et al. (2021) developed a multi-scale, multi-resolution method to handle the contrast distribution and large variations in the sizes of coronary vessels in the ICA. These methods with *ad hoc* components to boost performance are effective on main vessels but often fail to distinguish thin vessels and obtain inaccurate vessel boundaries, significantly affecting the stenosis identification. Accordingly, methods focusing on accurate single main vessel segmentation, right coronary artery (RCA) or left anterior descending (LAD) artery only, or classification of stenosis were developed in Au et al. (2018) and Yang et al. (2019). An advantage of these methods is that they only require the manual annotation of one branch per image, which is feasible in relatively large datasets, as the size of available annotated ICA datasets often constrains the coronary vessels segmentation performance. For example, in the approach of Nasr-Esfahani et al. (2018), the gold standard only consisted of 44 ICAs. In a recent study, a deep channel attention network for coronary vessels segmentation using a dataset of over 300 ICAs was proposed (Hao et al., 2020). A potential strategy is semi-supervised learning using methods such as the mean-teacher model, which can extract meaningful information from unlabeled images (He et al., 2022).

Motion artifacts, particularly from cardiac and respiratory movements, adversely affect the quality of the ICA. Accompanied by the wash-out effect of the contrast agent, they can make vessel identification difficult even for human experts. Early attempts at moving object segmentation can be traced to the field of computer vision. Fragkiadaki et al. (2015) proposed a dual-pathway neural network, which takes RGB video and its optical flow to detect objects in video sequences. A similar approach was used by Vlontzos and Mikolajczyk (2018), using a multi-stage U-Net architecture over low-level binary segmentation and optical flow. Hao et al. (2020) adopted a traditional U-Net with temporal fusion convolution and channel attention, but did not aim at maintaining the structural integrity of the vascular tree. Liang et al. (2021) designed Semi 3D U-Net that incorporates temporal feature extraction for coronary vessels segmentation from angiography videos, focusing on enhancing segmentation quality by leveraging both spatial and temporal features. Their method can only consider an odd number of temporal frames centered at the target frame. FCNs have been applied for three-frame ICA segmentation by Wan et al. (2021), using a tri-pathway FCN and influence matrix to decode the temporal information. These methods often face limitations, such as a lack of analysis within denser networks for effectively extrapolating both local and global information.

Similar approaches have been utilized in digital subtraction angiography (DSA) for cerebral vessel segmentation. Su et al. (2024) employed a spatio-temporal U-Net that integrates spatial and temporal features simultaneously through a Temporal Learning Module, allowing for cohesive decoding and learning of spatial–temporal dynamics. In contrast, Xie et al. (2024) proposed DSANet, which decouples the spatial and temporal feature extraction processes. This model leverages a Temporal Former module to capture temporal relationships and a Spatio-Temporal Fusion module to merge spatial and temporal information, resulting in a cascade decoding approach. However, in contrast to ICA images, DSA for cerebral vessels exhibits reduced complexity in terms of background structures and diminished motion artifacts.

The resolution of the challenge regarding disconnected vessels can aid in directly addressing the challenge of ICA quality, resulting in improved segmentation. Methods attempting to tackle disconnected vessels have been proposed for retinal image segmentation. Lin et al. (2022) created a model similar to the generative adversarial network (GAN) with a discrimination network to guide the segmentation network. Li et al. (2020) proposed IterNet, which iterates a mini U-Net multiple times with weight sharing and intra-model skip-connection. These are all implicit methods without clear explanations of the mechanisms through which vessel connectivity is improved. Lan et al. (2020) designed a loss function to close the gap between predicted vessel boundaries and the boundaries in the gold standard by fusing the disconnected parts together. Shit et al. (2021) proposed the ClDice loss to be used alongside the Dice loss, focusing on preserving the skeleton of segmented tubular structures and ensuring topological accuracy, specifically maintaining connected components and branch continuity. However, it is less sensitive to precise boundary delineation, which may impact performance in clinically significant regions such as stenotic areas. Oner et al. (2022) and Clough et al. (2022) introduced connectivity-optimizing loss functions aimed at retaining specific structural features, including linear continuity, separation of background regions, and global connectivity. While effective for preserving these aspects in network-like structures, these methods prioritize connectivity over boundary precision, potentially limiting their applicability to coronary vessels with fine-scale details and complex branching patterns.

More recent works have predominantly focused on adaptive or advanced architectures. The nnUNet framework, as proposed by Isensee et al. (2020), has become a widely adopted baseline for medical image segmentation tasks, due to the self-adapting function for generalizing across different datasets. Zhao et al. (2023) employed graph-based node similarity comparisons for skeleton segmentation using a graph attention network. This approach focuses on leveraging node relationships to enhance segmentation accuracy. He et al. (2024) fused the graph attention network and convolutional neural network to learn global geometric information during coronary vessels segmentation. Furthermore, Ruan and Xiang (2024) introduced the Vision Mamba UNet (VM-UNet), which utilizes State Space Models (SSMs) to capture extensive contextual information and long-range interactions with linear computational complexity. These methods are highly effective in various contexts, though not specifically designed for the unique challenges of ICA sequences.

1.2. Contributions

An analysis of the literature, as well as our own results on the public 323 ICA short sequence samples used in Hao et al. (2020), makes it clear that motion artifacts from the second challenge are the primary cause of the limited accuracy of coronary vessel segmentation in ICA. Body movements can shift clearly delineated vessels or parts of the vessels in earlier frames to low-contrast areas in the target frame, leading to inaccurate vessel boundaries and potential disconnections.

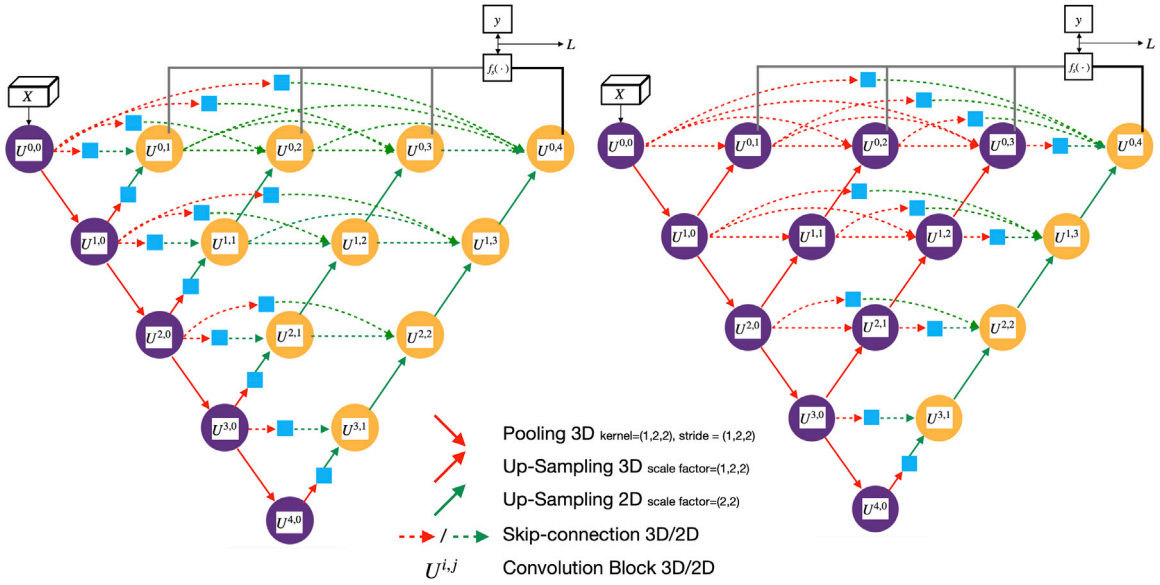

Fig. 1. The proposed TVS-Net model (left) and TVS-Net+ (right) for vessels segmentation from ICA sequences using temporal information. 3D and 2D blocks are represented in purple and orange, respectively, with all kernel sizes and strides shown in the boxes. The gray output paths at the top indicate deep supervision.

This disruption in structural integrity ultimately exacerbates the issues associated with vessel disconnection. Human annotators typically mitigate this issue by visualizing the ICA sequence, using relationships between consecutive frames to resolve overlapping vessels. This suggests that machine learning methods would also benefit from observing multiple ICA frames, mimicking human annotator's behavior. Thus, we hypothesize that introducing temporal information is fundamental for solving motion-related challenges without complex motion modeling. We introduce a novel architecture, the Temporal Vessel Segmentation Network (TVS-Net), where time is considered as the third dimension in the model, enabling effective extraction of spatio-temporal features to address motion artifacts. Our main contributions can be summarized in four aspects:

- We develop a new 3D (2D+T) framework that simultaneously extracts features from multiple consecutive ICA frames to segment the target frame. Our framework combines a novel densely nested 3D encoder that expands through additional convolutional nodes in its mid-layers and a highly connected 2D decoder. This dual-pathway design uses UNet++ as its backbone and ensures robust spatial–temporal feature extraction and precise spatial recognition, resulting in high-fidelity segmentation.
- We incorporate a connectivity-preserving loss function (Lan et al., 2020) to maintain vascular structural integrity and employ specific skeletonized metrics for evaluating structural accuracy. This approach is fundamental for generating precise vessel skeletons for 3D reconstruction.
- Our TVS-Net outperforms comparable approaches across varied data sources and annotation protocols, achieving 83.4% Dice and 84.3% recall on a public dataset. It achieves a higher recall of 86.3% on a refined subset with fine-grained annotations, notably surpassing the original dataset's manual annotations.
- TVS-Net attains 78.5% Dice and 82.4% recall on an external dataset comprising 60 ICAs, outperforming all SOTA methods. These results highlight the robustness and generalizability of our approach.

2. Materials and methods

This section introduces the input data, describes in detail the neural network architecture and loss function, and explains post-processing, evaluation methods, evaluation metrics, as well as the experimental settings and comparison methods. We denote X as the ICA sequence, with T being the total number of frames. x_t is frame number t, $x_{t=0}$ is the selected ICA frame for manual labeling, and y is gold standard segmentation, where $X \in [0,255]^{T\times H\times W}$ and $y \in \{0,1\}^{H\times W}$ with C, T, H, and W representing the number of channels, number of temporal frames, and height and width of the frame, respectively. A full dataset is $D = \{(X_n, y_n)\}_{n=1}^{N}$ with N being the total number of cases. The segmentation output of the network is $f_s(\cdot)$.

2.1. Study population

For training and inference of the proposed vessels segmentation algorithm, we use the database D_1 from Hao et al. (2020), acquired from Renji Hospital of Shanghai Jiao Tong University (SJTU). The raw data contains 120 sequences. These sequences were resized to 512 × 512, along with Poisson denoising (Cerciello et al., 2012). 323 frames were selected from the sequences for annotation. This publicly available dataset contained 323 short sequences, each containing two frames before and one after each annotated frame. Hence, our dataset contains $T = 4$ frames for every annotation with $X_n = \{x_{-2}, x_{-1}, x_0, x_1\}_n$. Hao et al. (2020) experimentally demonstrated that $T = 4$ provides the best generalizability in performance compared to $T = 2$, 3, and 5. For direct comparison, we divide the 323 samples into train, validation, and test sets as $N_{train} = 173$, $N_{val} = 82$, and $N_{test} = 68$, respectively, as in Hao et al. (2020).

We additionally acquired 60 ICA sequences from 39 patients who were admitted at the John Radcliffe (JR) Hospital, Oxford University Hospitals NHS Foundation Trust, with suspected coronary stenosis and provided informed consent. Each sequence consists of four frames of size 512 × 512, with the third frame of each sequence manually annotated by an expert. We label this dataset as D_2, and all images are used for external evaluation, making its $N_{test} = 60$.

2.2. TVS-Net and TVS-Net+

The usage of an encoder–decoder structure and skip-connections have been proven effective in U-Net (Ronneberger et al., 2015). To fully exploit this framework, U-Net++ (Zhou et al., 2018) fills the space between the encoder and decoder with extra skip-connected nodes, generating a denser network for multi-scale structure concatenation.

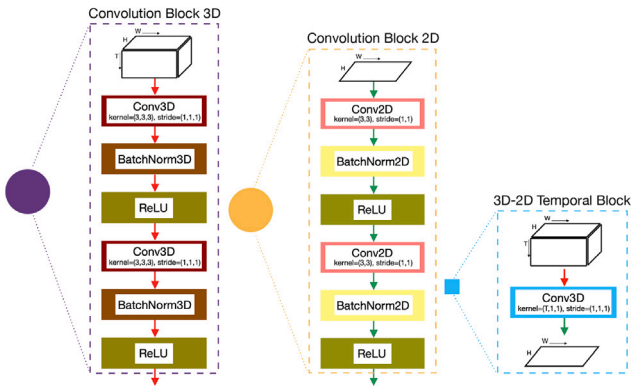

Fig. 2. Details of the convolution blocks and temporal block with the same legend for colored path in Fig. 1.

The 2D U-Net++ has been applied for single frame whole vascular tree segmentation and semantic segmentation in Zhao et al. (2020). In order to incorporate and analyze temporal information, we introduce in this work the **TVS-Net** and **TVS-Net**+ model, presented in Fig. 1. Since our dataset consists of multiple ICA sequences with one annotated target frame for each sequence, in order to analyze the 3D input and 2D output structures simultaneously, the network incorporates 3D and 2D convolution blocks to build a 3D encoder (purple) and a 2D decoder (orange) that are densely fused together by temporal feature extraction. A 3D convolution block contains convolution, batch normalization, and Rectified Linear Unit (ReLU) with the kernel size for convolution being $(3, 3, 3)$ as shown in Fig. 2. A similar pipeline is applied in the 2D convolution block.

Specifically, the 3D encoder captures both spatial and temporal features from ICA sequences, treating time as a third dimension to enhance vessel segmentation. The 3D–2D temporal block then fuses these features across frames, enabling the 2D decoder to produce accurate vessel masks informed by the dynamics of the sequence. This design helps mitigate noise from overlapping vessels, organs, and variations in contrast by using temporal shifts to distinguish target vessels and reduce artifacts.

2.2.1. Model variants

The convolution node in the network is represented as $U^{i,j}$, with i representing the number of pooling or down-sampling operations from $U^{0,0}$ and j representing the number of received skip-connections of the node. In a 5-level pyramidal structure, $i, j \in \{0, 1, 2, 3, 4\}$. The number of channels in the output of each node equals 2^{i+5}. With the 6 interlinked nodes $(U^{0,1}, U^{0,2}, U^{0,3}, U^{1,1}, U^{1,2}, U^{2,1})$, the connection from $U^{0,0}$ to any node can be seen as a dense convolution block. For example, node $U^{1,2}$ receives two skip-connection outputs and one up-sampling output concatenated together, where these outputs are generated in different convolution layers and pyramid levels. This narrowing of the semantic gap between the encoder and decoder facilitates the optimization process of the network. The network is structured with 3D convolution blocks in the encoder and 2D convolution blocks in the decoder, which are accompanied by temporal feature extraction that always happens right after the encoder. We compare two model variants for the 6 interlinked nodes as the 3D encoder can expand through these nodes. The first variant, **TVS-Net**, is shown in Fig. 1, where those 6 nodes belong to the dense 2D decoder can be mathematically expressed as:

$$U^{i,j} = \begin{cases} V(U^{i-1,j}) & \text{if } j = 0, \\ v([F(U^{i,0}), S(F(U^{i+1,j-1}))]) & \text{if } j = 1, \\ v([[U^{i,k}]_{k=1}^{j-1}, F(U^{i,0}), S(U^{i+1,j-1})]) & \text{if } j > 1 \end{cases} \tag{1}$$

where functions $V(\cdot)$, $v(\cdot)$, $S(\cdot)$, $F(\cdot)$, and $[\cdot]$ represent the 3D and 2D convolution blocks, up-sampling layer, temporal feature block, and concatenation layer respectively. To elaborate, for nodes with $j = 0$, solely 3D convolution operations are performed. When $j = 1$, 2D convolution is conducted over the concatenation of the outputs of the 3D–2D temporal blocks at the same level and at the lower level. For nodes where $j > 1$, 2D convolution is executed over the concatenation of the output of the 3D–2D temporal block at the current level, the upsampled output of the 2D node from the lower level, and all 2D nodes that are skip-connected to that node. In the second variant named **TVS-Net**+, the architecture is modified by substituting the original 6 nodes in the mid-layers with 3D convolutional blocks, thereby creating a denser 3D encoder for spatial–temporal feature extraction. This alteration also entails rearranging the temporal feature blocks since these blocks follow the 3D convolution. Similar to Eq. (1), we express it as:

$$U^{i,j} = \begin{cases} V(U^{i-1,j}) & j = 0 \\ V([[U^{i,k}]_{k=0}^{j-1}, S(U^{i+1,j-1})]) & i + j < 4, j > 0 \\ v([[F(U^{i,k})]_{k=0}^{j-1}, S(U^{i+1,j-1})]) & i + j = 4, j > 0 \end{cases} \tag{2}$$

with the same notations. Both variants exhibit highly interconnected 3D encoder-2D decoder architectures. However, the first variant features a denser 2D decoder, while the second variant emphasizes a denser 3D encoder. These configurations impart distinct biases, with the first variant optimizing for spatial–temporal analysis and the second variant enhancing spatial recognition.

2.2.2. Temporal feature extraction

Time is treated as the third dimension in our model. As shown in the 3D–2D temporal block in Fig. 2, we first apply a 3D convolution with kernel size $(T, 1, 1)$. Then we perform dimensional compression, i.e. squeezing, on the produced feature map in its time axis, resulting in a 2D temporally fused output, allowing our model to simultaneously analyze spatial features while taking into account the temporal dynamics inherent in ICA sequences. Mathematically, this is expressed as:

$$F = Squeeze_T(U^{i,j} \circledast K) \tag{3}$$

where K is the learnable kernel with size $(4, 1, 1)$ in our case. This temporal extraction enhances accurate vessel segmentation by enabling consistent identification of vascular structures across frames. As a result, vessels that may appear faded or moved into low-contrast regions in the annotated frame can still be accurately detected based on their visibility in other frames without explicitly modeling vessel motion.

2.2.3. Deep supervision

We incorporate deep supervision (Lee et al., 2015) in our framework, represented by the gray paths at the top of the network in Fig. 1. We first calculate the loss after applying the sigmoid function on the outputs of blocks $U^{0,1}, U^{0,2}$, $U^{0,3}$, and $U^{0,4}$. We derive the final deep supervision loss as the average of the four loss values, serving as an additional regularization to diminish test error and expedite loss convergence.

2.3. Energy loss function

As the loss function in our proposed architecture, we minimize a function originally proposed in Xiang et al. (2006), inspired by the elastic energy of dislocations in crystals. The energy system of any single curve $\gamma\,(x(s), y(s), z(s))$ in 3D space (x, y, z) consists of two key functions (Lan et al., 2020):

$$\vec{w}(x, y, z) = -\frac{1}{4\pi} \int_{\gamma} \frac{\vec{r} \times d\vec{l}}{|\vec{r}|^3} \text{ and } E = \frac{1}{8\pi} \int_{\gamma} \int_{\gamma'} \frac{d\vec{l} \cdot d\vec{l'}}{|\vec{r}|} \tag{4}$$

where $d\vec{l} = \vec{\tau}\delta(\gamma)dxdydz$ is a differential part of the curve, $\delta(\gamma)$ is a delta function of the curve γ that is always perpendicular to γ, and $\vec{\tau}$ is the unit tangent vector of γ. $\vec{w} = (0, 0, w_3(x, y, z))$ is a function in the XYZ-plane, and $\vec{r} = (x-x(s), y-y(s), z-z(s))$ is a vector between a point in space to a point on the curve. More specifically, $\vec{w}(x, y, z)$ in Eq. (4) represents the dynamics of the system under given constraints, whereas

E represents the total elastic energy of the system. This system can be applied to vessel segmentation, as it is applicable to a collection of curves. If we define the stationary vessel boundary of the gold standard as γ_1 and the moving boundary of the prediction as γ_2 (moving because the location of the predicted boundary changes during training), γ in Eq. (4) can be substituted with $\gamma_1 \cup \gamma_2$. Specifically, γ_1 is smoothed to G_s using a Gaussian function, and γ_2 is represented as $H(\phi)$ by applying a regularized Heaviside function to the level set ϕ of the prediction $f_s(\cdot)$. By substituting $\vec{\tau}$ and $\vec{w}$, the dynamic equation of energy minimization is expressed as:

$$v(x,y) = -\frac{1}{4\pi}\int_{\Re^2} \frac{\vec{r}\cdot\nabla(G_s + \alpha H(\phi))}{|\vec{r}|^3} dx dy \tag{5}$$

where α is a hyper-parameter. If we define $G_s + \alpha H(\phi) = T(x,y)$ for convenience, the energy of our object system can be derived from Eq. (4):

$$E = \frac{1}{8\pi}\int_{\Re^3} dx dy \int_{\Re^3} \frac{\nabla T(x,y)\cdot\nabla T(x',y')}{|\vec{r}|} dx' dy'. \tag{6}$$

Here, Eq. (6) is the loss function L for this energy system, with Eq. (5) the corresponding gradient. Unlike previous studies that applied this loss function to single-frame ICA images, our method leverages the loss function not only to capture spatio-temporal dependencies but also to preserve connectivity across consecutive frames. To boost the optimization efficiency, Eqs. (5) and (6) can be further simplified by transforming them into 2D Fourier space (Lan et al., 2020).

2.4. Half tensor training

Since our proposed architecture is a dense quasi-3D network, we apply half tensor training with mixed precision, in order to achieve the optimal segmentation performance with a smaller model size. In our method, only the loss calculation is performed at single precision in 32-bit floating point (FP32), whereas all other components operate in 16-bit floating point (FP16). In particular, only the gold standard segmentation masks y_n are kept in FP32; the segmentation output $f_s(\cdot)$ is converted to FP32 from FP16 during the loss calculation.

2.5. Post processing

2.5.1. Skeletonization

In order to evaluate the performance of the proposed method for generating accurate vessel skeletons, the gold standard segmentation y_n and predicted segmentation $f_s(\cdot)$ are both skeletonized. We apply the skeletonization method $\sigma(\cdot)$ (Zhang and Suen, 1984), which includes clockwise pixel assessment of a 3×3 mask iterated through the image.

2.5.2. Denoising

In order to eliminate spurious components, we identify connected components in the predicted segmentation $f_s(\cdot)$ and remove those smaller than a specified threshold.

2.6. Evaluation methods

2.6.1. Conventional metrics

To evaluate segmentation performance, we use four conventional metrics: Area Under the Precision–Recall Curve (AUPRC), Dice coefficient, Recall, and Precision. This allows for direct comparison with SVS-Net (Hao et al., 2020) on the same 4-frame ICA sequence dataset.

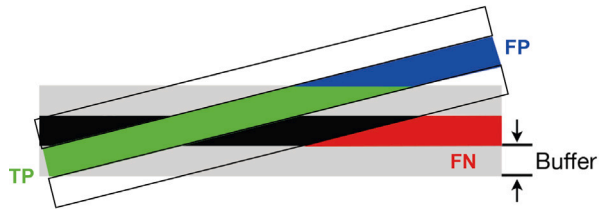

Fig. 3. Skeletonization metrics for vessel centerline with 1-pixel width. True Positives (TP) are green, whereas False Positives (FP) and False Negatives (FN) are blue and red, respectively.

Table 1
Comparison of different architecture variants.

Model	AUPRC	Dice (%)	Recall (%)	Precision (%)
TVS-Net	**0.8717**	**82.65 ± 2.56**	82.20 ± 4.68	**83.11 ± 5.41**
TVS-Net+	0.8653	82.28 ± 2.57	**82.85 ± 5.73**	82.31 ± 6.14

All values represent mean ± standard deviation (SD).

2.6.2. Skeletonization metrics

The skeletonization performance is evaluated using two metrics, originally introduced for quantifying geometry preservation in curvilinear structures (Youssef et al., 2015). The metrics, termed as completeness (C_r) and correctness (C_p), represent buffered versions of Recall and Precision on the skeleton, respectively. Over the vessel skeleton, True/False Positives/Negatives are defined as illustrated in Fig. 3. We add a 1-pixel buffer on each side of the skeleton, resulting in a 3 pixels wide buffered skeleton that remains within the boundaries of thin vessels, while providing robust detection of structural issues in thicker vessels. To obtain TP and FP, we buffer $\sigma(y_n)$ only to compare with the un-buffered $\sigma(f_s(\cdot))$. The part within the buffered $\sigma(y_n)$ is considered as TP and vice versa for FP. Equivalently, we obtain FN by solely buffering $\sigma(f_s(\cdot))$. The buffering of skeletons is performed by dilation of the single-pixel structure.

2.7. Experimental settings

The training of the network is performed on an NVIDIA Quadro RTX 8000 GPU, with all frames augmented on-the-fly by flipping, changing saturation, and changing contrast. We also apply rotation by 90° randomly for 0–3 times with probability 0.7. Trainings in full tensor are optimized with Adam using a learning rate of 10^{-5}, $\beta_1 = 0.85$, weight decay of 10^{-5}, and all other hyper-parameters set to the default in Pytorch. Trainings in half tensor are optimized with Stochastic gradient descent (SGD), to avoid problems in Adam with values exceeding the dynamic range of FP16. To support SGD, we apply a cosine annealing scheduler to gradually decrease the learning rate from 10^{-6} to 10^{-7}; besides, the weight decay is decreased to 10^{-6} and momentum is set to 0.9. For all trainings, the value of α in the loss function of Eq. (6) is set to 0.35. Hyperparameters are obtained by grid search. We perform training over 2000 epochs to ensure convergence of the loss function.

2.8. Comparison methods

To evaluate the performance of the proposed TVS-Net framework, we compare 6 SOTA models that have been applied in vessel segmentation, namely UNet (Ronneberger et al., 2015), RA-UNet (Chen et al., 2020), UNet++ (Zhou et al., 2018), nnUNet (Isensee et al., 2020), VMUNet (Ruan and Xiang, 2024), and SVS-Net (Hao et al., 2020) on datasets D_1 (SJTU) and D_2 (JR), with SVS-Net being the SOTA temporal ICA segmentation method. Specifically, SVS-Net employs the same temporal input data but does not utilize deep supervision, whereas the other methods neither use temporal input data nor deep supervision. The loss functions used for training are also consistent with their original implementations: UNet, UNet++, and SVS-Net are trained using

Table 2
Comparison of TVS-Net with SOTA methods on test dataset of D_1.

Method	AUPRC	Dice (%)	Recall (%)	Precision (%)	Completeness (%)	Correctness (%)
UNet	0.8169	79.34 ± 4.07***	79.31 ± 7.99***	79.37 ± 6.29***	76.08 ± 8.59***	71.96 ± 8.05***
RA-UNet	0.8109	72.13 ± 9.51***	62.22 ± 12.0***	*85.80 ± 9.35*	61.57 ± 12.2***	*78.47 ± 6.34*
UNet++	0.8687	81.54 ± 3.57***	81.13 ± 6.59***	81.96 ± 5.79***	80.99 ± 8.49***	75.02 ± 7.19**
nnUNet	0.8679	81.64 ± 4.05**	83.41 ± 8.74	79.94 ± 9.01*	79.72 ± 9.17*	73.32 ± 8.51*
VMUNet	*0.8839*	81.63 ± 2.85***	*83.68 ± 4.53**	79.67 ± 5.29***	*83.03 ± 5.27***	74.72 ± 5.69**
SVS-Net	0.8621	*83.23 ± 4.49*	79.41 ± 7.46***	**87.43 ± 5.24**	77.32 ± 7.54***	**82.71 ± 6.97**
TVS-Net	**0.8899**	**83.41 ± 2.45**	**84.31 ± 4.62**	82.52 ± 6.04	**84.83 ± 5.74**	76.55 ± 7.29
TVS-Net w Dice w/o DS	0.8826	82.49 ± 2.81*	84.60 ± 4.45	80.48 ± 5.74*	81.92 ± 5.04**	75.63 ± 7.02
TVS-Net w/o DS	0.8847	83.27 ± 2.57	83.49 ± 4.41*	83.05 ± 5.65	83.52 ± 5.54	76.87 ± 7.03

All values represent mean ± SD.
* p-value < 0.05, based on one-tailed Wilcoxon signed-rank test against the TVS-Net.
** p-value < 0.01, based on one-tailed Wilcoxon signed-rank test against the TVS-Net.
*** p-value < 0.001, based on one-tailed Wilcoxon signed-rank test against the TVS-Net.

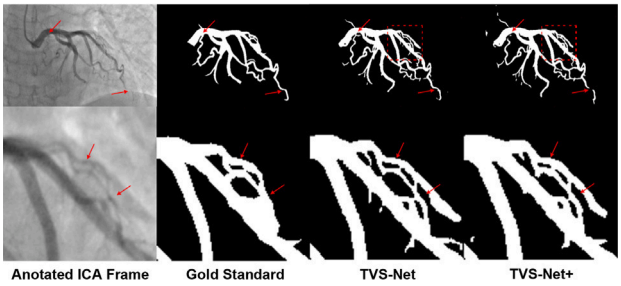

Fig. 4. Qualitative evaluation of segmentation performance using TVS-Net and TVS-Net+. The bottom row is a zoomed-in version of the red square in the top row.

Dice loss, while nnUNet and VMUNet utilize a combination of Dice and cross-entropy losses. Since removing the 3D–2D temporal blocks effectively transforms TVS-Net to a UNet++, this makes UNet++ a suitable comparison model for evaluating the efficacy of temporal feature extraction in TVS-Net. For external evaluation on the D_2 dataset, we directly apply the model trained on D_1. All quantitative metrics that require binarization of the generated segmentation apply the threshold of 0.5 (equivalent to 127 as pixel intensity). The optimal batch size for TVS-Net is determined to be 6, which is followed for all subsequent experiments.

3. Results

In this section, we present extensive quantitative and qualitative comparisons. First, we compare the performance of TVS-Net against its variant, TVS-Net+, to evaluate architectural effectiveness. Next, we benchmark TVS-Net against SOTA methods, highlighting its superior performance. Following this, we assess its generalizability on an out-of-distribution (OOD) dataset. Additionally, we evaluate the impact of our loss function and deep supervision setting on segmentation performance. Finally, we conduct a comprehensive study on the re-segmented gold standard.

3.1. TVS-Net vs. TVS-Net+

To compare the effectiveness of temporal information for multi-frame ICA segmentation, we train with the same hyper-parameters in two proposed network architectures: TVS-Net and TVS-Net+. Due to the large network size for TVS-Net+, the batch size for both networks is limited to 4 in this experiment. As presented in Table 1, the TVS-Net produces the best results with the highest AUPRC, Dice, and precision along with lower SD.

The performance of these two networks can be visually assessed in Fig. 4. From the red arrows in the first row, it can be seen that the TVS-Net not only yields a good delineation of thick vessels but also accurately defines small bifurcations and the distal part of the coronary vascular tree, pivotal for preserving the vascular topology. This, along with the computational efficiency and performance improvement, suggests the relative superiority of the proposed TVS-Net model in our study.

3.2. Evaluation on public dataset D_1

Fig. 5 presents the segmentation performance on the test set of D_1, showcasing complex views of the Left Circumflex (LCx) and Right Coronary Artery (RCA). It is evident that TVS-Net surpasses other SOTA methods by effectively preserving vascular structures in the main and distal branches. In the particularly narrow vessel areas of the first two images for LCx, it identifies vessels absent from the gold standard annotations, accurately capturing the width of each branch. Moreover, the significant performance disparity in the third image for RCA is notable, primarily because dataset D_1 comprises only about 25% of RCA images. This limitation restricts weaker frameworks from effectively transferring learned features from the other images. Extreme samples can be observed in the output of UNet and RA-UNet as they fail to generalize on RCA with catastrophic disconnection and missingness of major branches.

Quantitatively, as shown in Table 2, TVS-Net achieves the highest scores in AUPRC, Dice, recall, and C_r. It surpasses the Dice score of the SOTA VMUNet by 2.2% and the recall score of the SOTA temporal framework SVS-Net by 6.2%. Additionally, it is the only framework that does not show a reduction from recall to C_r, indicating there is no major disconnection or shift on the vessel skeleton, further demonstrating the high structural integrity of its generated segmentations.

3.3. Evaluation on out-of-distribution dataset D_2

For additional quantitative evaluation of the generalizability of the proposed TVS-Net, we apply the trained model from the dataset D_1 on the out-of-distribution (OOD) dataset D_2 acquired from the Oxford JR Hospital. As visible from the results presented in Table 3, TVS-Net achieves superior performance in terms of all metrics, namely AUPRC, Dice, recall, precision, C_r and C_p, with performance improvements of 4.0%, 5.0%, 2.0%, 0.3%, 4.6%, and 7.0% respectively, compared to the second best SOTA.

In Fig. 7, the Precision–Recall (PR) curve of TVS-Net lies above all other PR curves, demonstrating a dominant generalizability for all binarization thresholds. Qualitatively, a similar observation is evident in Fig. 6. All other frameworks produce noisy or incomplete vascular structures in the first three RCA images. Most importantly, even SOTA SVS-Net and VMUNet cannot delineate the distal branches in the first and third images. While SOTA nnUNet manages to detect some of the finer vessels, it struggles with false positives, mistakenly identifying the

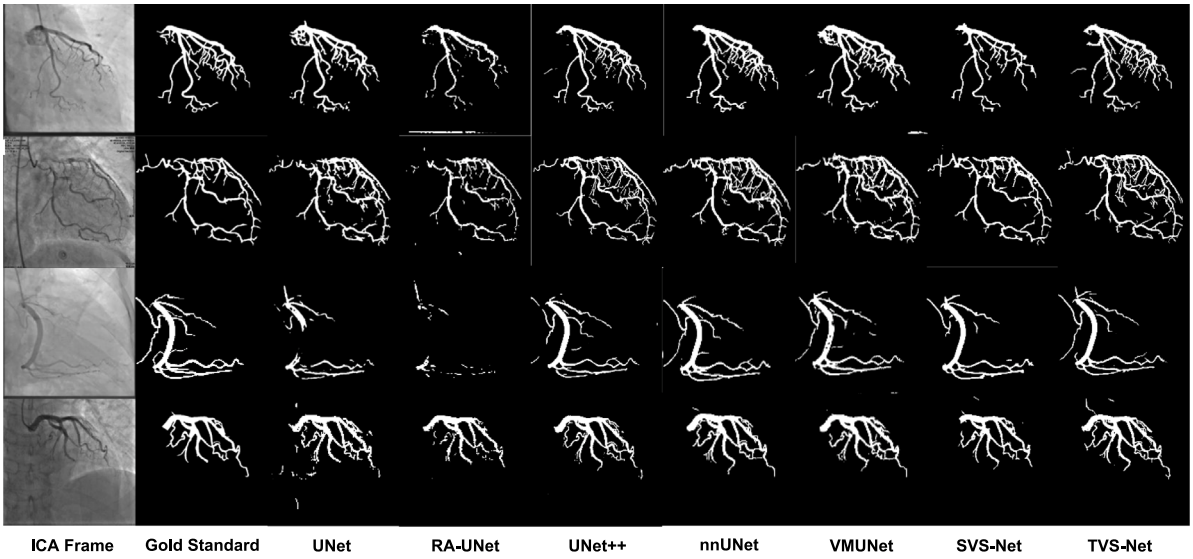

Fig. 5. Qualitative evaluation of segmentation performance of SOTA methods and TVS-Net on dataset D_1.

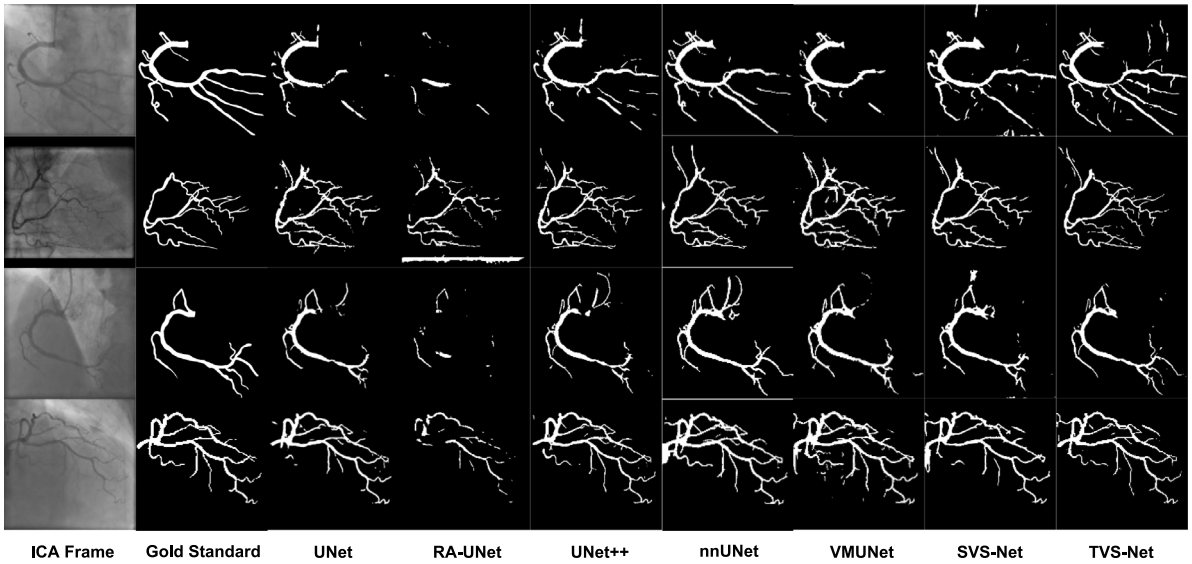

Fig. 6. Qualitative evaluation of segmentation performance of SOTA methods and TVS-Net on OOD dataset D_2.

Table 3
Comparison of TVS-Net with SOTA methods on OOD dataset D_2.

Method	AUPRC	Dice (%)	Recall (%)	Precision (%)	Completeness (%)	Correctness (%)
UNet	0.7207	67.88 ± 18.1***	67.86 ± 21.6***	67.90 ± 9.33***	58.88 ± 20.1***	68.49 ± 15.2***
RA-UNet	0.6746	51.18 ± 18.5***	39.12 ± 18.0***	73.98 ± 14.6	39.43 ± 17.9***	73.59 ± 15.3*
UNet++	0.8030	*75.71 ± 10.4***	75.65 ± 13.3***	*75.77 ± 10.0*	78.62 ± 14.7***	69.52 ± 12.4***
nnUNet	*0.8189*	74.81 ± 8.99***	79.03 ± 9.56***	71.01 ± 11.1***	81.81 ± 15.1***	69.23 ± 12.7***
VMUNet	0.8109	74.69 ± 9.07***	*80.74 ± 12.2**	69.48 ± 10.1***	*83.90 ± 13.1****	*73.68 ± 11.1****
SVS-Net	0.7642	73.64 ± 9.72***	76.44 ± 11.9***	71.03 ± 11.2***	71.79 ± 12.5***	67.59 ± 13.4***
TVS-Net	**0.8437**	**78.49 ± 9.74**	**82.41 ± 11.8**	**75.79 ± 9.86**	**86.53 ± 13.5**	**77.78 ± 11.7**

All values represent mean ± SD.
* p-value < 0.05, based on one-tailed Wilcoxon signed-rank test against the TVS-Net.
*** p-value < 0.001, based on one-tailed Wilcoxon signed-rank test against the TVS-Net.

catheter as a vessel in the third image. Hence, the high performance improvement and accurate segmentation quality together demonstrate the TVS-Net's ability to generalize in OOD real-world scenarios.

3.4. Efficacy of energy loss function

To evaluate the performance gain from the energy loss function L in Eq. (6), we train TVS-Net with Dice loss and the energy loss separately, without deep supervision in both cases. This approach allows for a

Table 4
Performance evaluation on the new gold standard with 10 re-segmented samples.

Method	AUPRC	Dice (%)	Recall (%)	Precision (%)	Completeness (%)	Correctness (%)
UNet	0.8307	82.11 ± 2.08**	81.90 ± 5.82**	82.32 ± 5.41*	73.07 ± 7.99**	78.28 ± 6.75**
RA-UNet	0.8132	72.10 ± 9.33**	61.47 ± 11.0**	**87.17 ± 4.16**	56.15 ± 10.9**	84.91 ± 4.40
UNet++	0.8839	*84.65* ± 3.19*	83.07 ± 5.42*	86.02 ± 5.22	*82.91* ± *5.53*	85.13 ± 4.65
nnUNet	*0.8889*	83.42 ± 2.41**	84.25 ± 4.61	82.61 ± 3.03**	82.84 ± 3.85	85.68 ± 4.31
VMUNet	0.8877	82.12 ± 1.92**	*85.36* ± *4.32*	79.11 ± 5.95**	82.87 ± 4.56	*85.93* ± *4.83*
SVS-Net	0.8338	79.84 ± 4.25*	74.38 ± 8.49**	86.14 ± 4.23	66.38 ± 7.89**	85.74 ± 4.39
TVS-Net	**0.9018**	**86.20 ± 1.96**	**86.26 ± 5.24**	*86.16* ± *3.27*	**84.00 ± 5.25**	**87.55 ± 6.17**
Old gold standard	N/A	87.66 ± 5.30	85.44 ± 7.95	90.01 ± 4.05	76.81 ± 11.24	91.35 ± 5.08

All values represent mean ± SD.
* p-value < 0.05, based on one-tailed Wilcoxon signed-rank test against the TVS-Net.
** p-value < 0.01, based on one-tailed Wilcoxon signed-rank test against the TVS-Net.

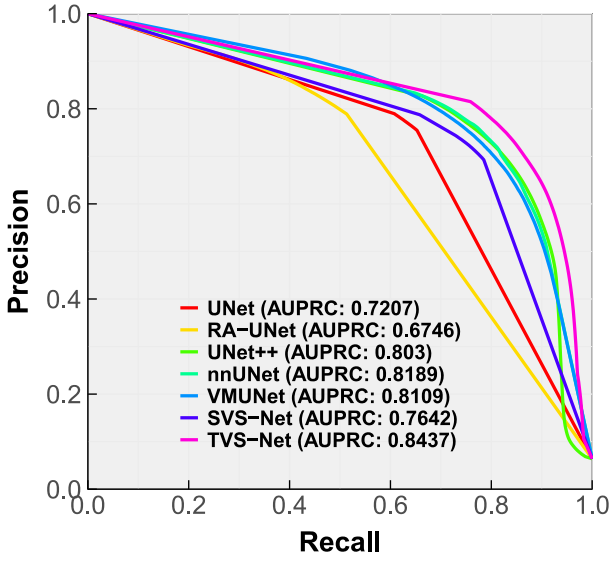

Fig. 7. Precision–Recall curve for evaluation on OOD dataset D_2.

clearer evaluation of the individual contribution of each component. From the last two rows of Table 2 and Fig. 8, we find that the energy loss function improves Dice by 1%. The recall with Dice loss is slightly higher, likely due to over-segmentation of the vessel boundaries, as seen in the zoomed area of the first row in Fig. 8 near the arrows. This over-segmentation is also evident from the lower precision value by the Dice loss. Additionally, the Dice loss causes disconnections on the main branch of the first image in Fig. 8, which not only affect the geometry but also yield 1.95% lower C_r and 1.64% lower C_p for skeletonized vessels (Table 2). Furthermore, we observe an increase from recall to C_r when applying the energy loss, indicating that the vessel skeleton and boundaries are more consistently preserved with this loss function. Notably, the elastic loss better preserves the vascular geometry even in false positives (which could potentially be true positives), as highlighted in the boxed area of the second row in Fig. 8.

3.5. Effectiveness of deep supervision

The advantage of deep supervision (DS), introduced in Section 2.2.3, is presented in Table 2, comparing TVS-Net to TVS-Net w/o DS. With the use of deep supervision, both Dice and recall metrics increase. A more pronounced trend is found for skeletonized vessels, where the TVS-Net with DS yields higher C_r, showing that DS significantly aids in preserving vessel connectivity.

For qualitative evaluation of the effects of deep supervision, we observe in Fig. 9 that for the segmentation of both vascular trees and skeletons, the deep supervision has enabled the identification of previously missed vessels (red FN in left panel) with accurate delineation (green TP in right panel). Hence, the utilization of deep supervision enhances the detection of missed vessels and rectifies vessel disconnection, with FP occurring only at the distal end of some vessels.

3.6. Comparison on fine-detailed segmentation

Performing a thorough manual annotation of all vessels in an ICA image is very demanding; thus, gold standard annotations are typically coarse-grained (incomplete). The zoomed-in region in Fig. 4 highlights this limitation, e.g., at the location indicated by the lower red arrow. This affects the reliability of evaluation metrics. Hence, we further select a subset of 10 samples from the test dataset and perform a more comprehensive manual re-segmentation by an expert, aiming at accurately delineating all the vessels in the images regardless of their sizes. To prevent bias in the selection, we rank the test set images based on the Dice scores of their segmentation using TVS-Net, and select those at the $0, 10, 20, 30, 40, 50, 60, 75, 85, 100$th percentiles. We consider this newly segmented dataset as the new "fine-grained" gold standard and compare it with the "coarse-grained" gold standard. Three example cases with the minimum (0th), median (50th), and maximum (100th) Dice scores are shown in Fig. 10. From the results, it is clear that the level of incompleteness in the initial annotations varies significantly between samples, which has an important effect on the calculated metrics.

Using these 10 re-segmented samples as the new gold standard, we perform a direct evaluation using the TVS-Net trained on the old gold standard, as presented in Table 4. It can be observed that our proposed TVS-Net improves over the SOTA temporal segmentation method SVS-Net by 7.95% in Dice and 16% in recall. Interestingly, our automated segmentation results are even higher in terms of recall compared to the old gold standard it trained on, whereas the difference in terms of Dice metric is only 1.5%. Our proposed method also produces the highest C_r of 84.0% with comparable C_p of 87.55% and precision of 86.16%, demonstrating that our method produces the least disconnection and shift. The same trend is observed qualitatively in Fig. 11. TVS-Net also outperforms the SOTA single frame segmentation method VMUNet by 4.97% in Dice, further demonstrating the effectiveness of incorporating temporal information.

4. Discussions and conclusion

In this paper, we develop a novel deep learning framework with a densely connected 3D encoder-2D decoder, named TVS-Net, that utilizes multiple frames of ICA sequences for generating accurate coronary vessels segmentation. The architecture integrates temporal convolution blocks for fusing the image sequence information and a unique energy loss for enhancing topology preservation onto a dense framework featuring deep supervision. This framework can be combined with vascular post-processing techniques, such as the method proposed by Qiu et al. (2023), to further enhance segmentation quality. To enable direct comparison between our method and the closest SOTA temporal

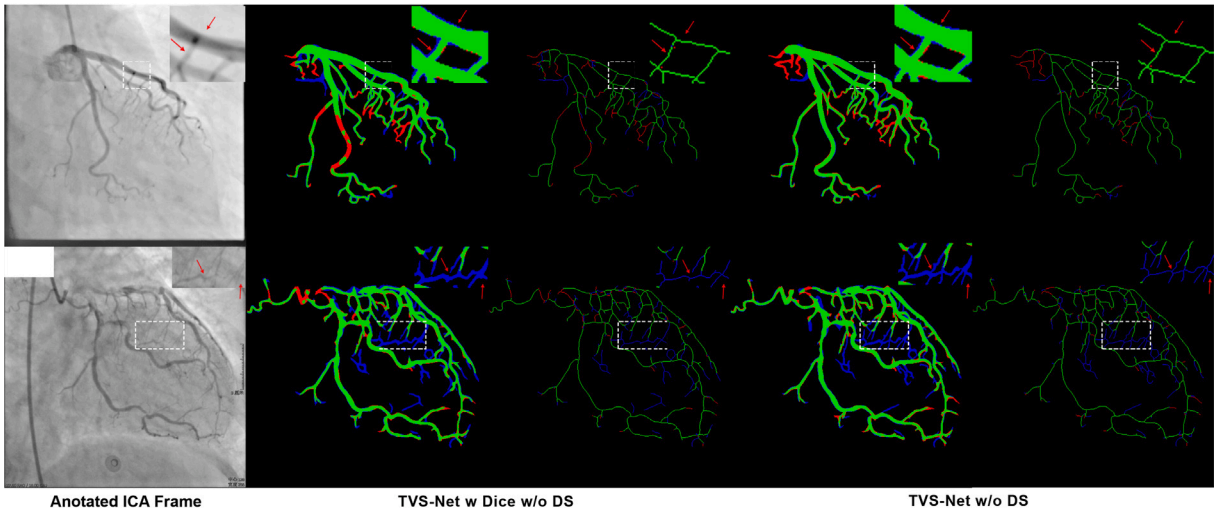

Fig. 8. Qualitative evaluation of segmentation and skeletonization performance of TVS-Net with Dice loss and energy loss, without deep supervision. The color codes for TP, FP, and FN are the same as in Fig. 3.

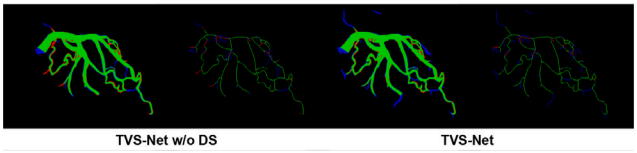

Fig. 9. Vessel segmentation and skeletonization performance of TVS-Net without and with deep supervision. Color codes are the same.

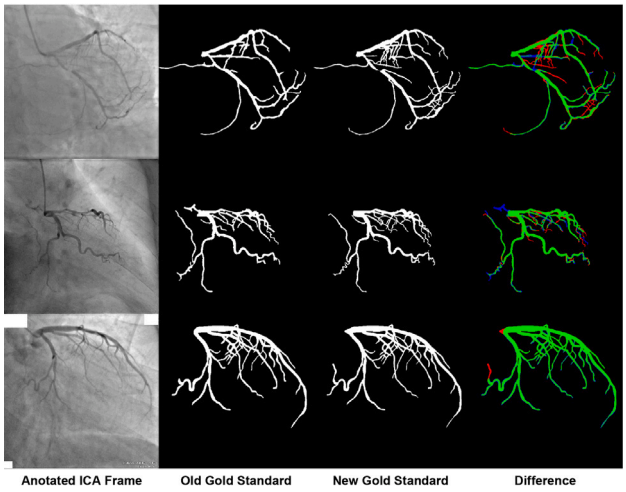

Fig. 10. Three re-segmented samples with minimum (81.61%), median (84.69%), and maximum (94.76%) Dice scores (top to bottom). Color codes are the same.

segmentation (Hao et al., 2020) and other SOTA methods, we train our framework on the same public dataset using the same data split. The experimental analysis not only demonstrates the advantages of a multi-frame approach but also illustrates the problem of incomplete annotations of vascular trees. The significant performance improvements observed by the TVS-Net over UNet++ underscore the effectiveness of integrating densely connected 3D–2D temporal blocks. We propose the use of skeletonization metrics and demonstrate the ability of our method to preserve thin vascular structures accurately. The combined use of spatio-temporal encoding, connectivity-preserving loss function, and deep supervision enables our proposed TVS-Net to capture a wide range of vessel appearances and motion patterns. This generalizability is further highlighted in our evaluation using data collected from a local hospital, where our method achieves the highest Dice, recall, precision, and AUPRC of 78.49%, 82.41%, 75.79% and 0.8437, respectively, despite the fact that our hyperparameters were optimized for the public dataset alone and no retraining was performed. The smaller size of the private dataset and the random selection of cases may contribute to greater variance and performance variability. Due to the OOD nature of the private data, all models show less confident predictions. However, the fact that our method excels in this context demonstrates its ability to adapt to different annotation protocols and data characteristics and highlights its strong potential for real-world deployment. Additionally, we proposed a variant of TVS-Net, named TVS-Net+, with the same ideology after expanding the 3D encoder. However, due to the larger size of TVS-Net and available GPU resources, its batch size is limited to 4, which impacts the network's ability to generalize effectively, thereby limiting the observed performance gains of TVS-Net+.

We re-segment 10 cases with a strict annotation protocol, including all visible vessels, and use it as the new (fine-grained) gold standard for the same evaluation pipeline. Trained by the original (coarse-grained) gold standard, our proposed TVS-Net achieves a recall of 86.26%, which is 0.96% higher than the original gold standard and 16% higher than the current SOTA method SVS-Net (Hao et al., 2020). The performance in accurately preserving vascular skeletons achieves 84.00% in C_r, improving on the original gold standard by 9.36% and SVS-Net by 26.54%. The qualitative evaluation illustrates the improvements, including the reduction in over-segmented vascular boundaries. Most interestingly, our extensive analyses demonstrate the feasibility of weak supervision with coarse-grained annotations for coronary vessels segmentation. This is evidenced by the superior delineation achieved by the TVS-Net model compared to its training gold standard — the coarse-grained annotations when evaluated against fine-grained annotations. It is important to note that the ICA datasets were created by selecting high-quality frames, limiting our ability to fully demonstrate the network's performance across low-quality frames, where manual segmentation is more challenging and often results in coarser gold standards. Consequently, by modulating the completeness level of manual annotations in the dataset, this framework can also facilitate the exploration of a time-performance trade-off between manual and automatic segmentations in annotation protocols, as well as elucidate the impact of partially segmented ground truth on final trained segmentation quality.

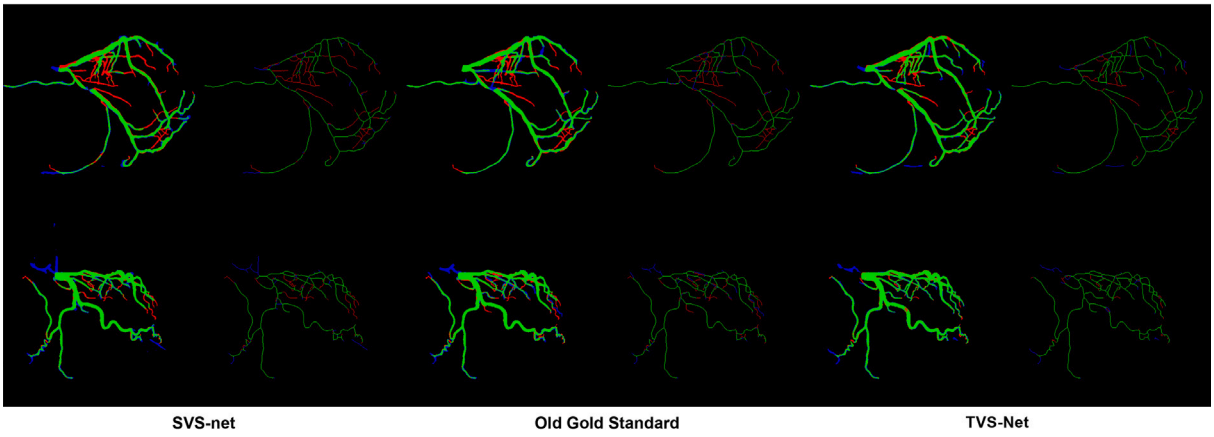

Fig. 11. Qualitative evaluation of segmentation and skeletonization on the new gold standard. The same color code is used here.

CRediT authorship contribution statement

Haorui He: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Abhirup Banerjee:** Writing – review & editing, Writing – original draft, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Robin P. Choudhury:** Writing – review & editing, Supervision, Resources, Funding acquisition, Data curation. **Vicente Grau:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge the use of the facilities and services of the Institute of Biomedical Engineering (IBME), Department of Engineering Science, University of Oxford, and the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out this work. http://dx.doi.org/10.5281/zenodo.22558

This work was supported in part by the British Heart Foundation (BHF) Project under Grant PG/20/21/35082, BHF Centre of Research Excellence, Oxford, NIHR Oxford Biomedical Research Centre, and the CompBioMed 2 Centre of Excellence in Computational Biomedicine (European Commission Horizon 2020 research and innovation program, grant agreement No. 823712). A. Banerjee is supported by the Royal Society University Research Fellowship (Grant No. URF\R1\221314).

Appendix A. Supplementary data

The electronic supplementary material available online includes additional details of the derivation of energy loss function, skeletonization method, as well as the extended explanations of refined segmentation.

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.media.2025.103496.

Data availability

Data will be made available on request.

References

Ambrosini, P., Smal, I., Ruijters, D., Niessen, W.J., Moelker, A., van Walsum, T., 2015. 3D catheter tip tracking in 2D X-Ray image sequences using a hidden Markov model and 3D rotational angiography. In: Augmented Environments for Computer-Assisted Interventions. pp. 38–49.

Au, B., Shaham, U., Dhruva, S., Bouras, G., Cristea, E., Lansky, A., et al., 2018. Automated characterization of stenosis in invasive coronary angiography images with convolutional neural networks. arXiv preprint arXiv:1807.10597.

Blondel, C., Malandain, G., Vaillant, R., Ayache, N., 2006. Reconstruction of coronary arteries from a single rotational X-ray projection sequence. IEEE Trans. Med. Imaging 25 (5), 653–663. http://dx.doi.org/10.1109/TMI.2006.873224.

Cerciello, T., Bifulco, P., Cesarelli, M., Fratini, A., 2012. A comparison of denoising methods for X-ray fluoroscopic images. Biomed. Signal Process. Control. 7 (6), 550–559.

Chen, X., Yao, L., Zhang, Y., 2020. Residual attention u-net for automated multi-class segmentation of covid-19 chest ct images. arXiv preprint arXiv:2004.05645.

Clough, J.R., Byrne, N., Oksuz, I., Zimmer, V.A., Schnabel, J.A., King, A.P., 2022. A topological loss function for deep-learning based image segmentation using persistent homology. IEEE Trans. Pattern Anal. Mach. Intell. 44 (12), 8766–8778. http://dx.doi.org/10.1109/TPAMI.2020.3013679.

Fazlali, H.R., Karimi, N., Soroushmehr, S.M.R., Sinha, S., Samavi, S., Nallamothu, B., Najarian, K., 2015. Vessel region detection in coronary X-ray angiograms. In: 2015 IEEE International Conference on Image Processing. ICIP, IEEE, pp. 1493–1497.

Fragkiadaki, K., Arbelaez, P., Felsen, P., Malik, J., 2015. Learning to segment moving objects in videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4083–4090.

Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A., 1998. Multiscale vessel enhancement filtering. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 130–137.

Hao, D., Ding, S., Qiu, L., Lv, Y., Fei, B., Zhu, Y., Qin, B., 2020. Sequential vessel segmentation via deep channel attention network. Neural Netw. 128, 172–187.

He, H., Banerjee, A., Beetz, M., Choudhury, R.P., Grau, V., 2022. Semi-supervised coronary vessels segmentation from invasive coronary angiography with connectivity-preserving loss function. In: 2022 IEEE 19th International Symposium on Biomedical Imaging. ISBI, pp. 1–5. http://dx.doi.org/10.1109/ISBI52829.2022.9761695.

He, H., Banerjee, A., Choudhury, R.P., Grau, V., 2024. Automated coronary vessels segmentation in X-ray angiography using graph attention network. In: Statistical Atlases and Computational Models of the Heart. Regular and CMRxRecon Challenge Papers. Springer Nature Switzerland, Cham, pp. 209–219.

Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K., 2020. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods 18, 203–211.

Jerman, T., Pernuš, F., Likar, B., Špiclin, Ž., 2016. Enhancement of vascular structures in 3D and 2D angiographic images. IEEE Trans. Med. Imaging 35 (9), 2107–2118. http://dx.doi.org/10.1109/TMI.2016.2550102.

Jiang, Z., Ou, C., Qian, Y., Rehan, R., Yong, A., 2021. Coronary vessel segmentation using multiresolution and multiscale deep learning. Inform. Med. Unlocked 24, 100602.

Kamran, S.A., Hossain, K.F., Tavakkoli, A., Zuckerbrod, S.L., Sanders, K.M., Baker, S.A., 2021. RV-GAN: Segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 34–44.

Kočka, V., 2015. The coronary angiography – An old-timer in great shape. Cor Vasa 57 (6), e419–e424.

Lan, Y., Xiang, Y., Zhang, L., 2020. An elastic interaction-based loss function for medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. Springer International Publishing, Cham, pp. 755–764.

Lashgari, M., Choudhury, R.P., Banerjee, A., 2024. Patient-specific *In Silico* 3D coronary model in cardiac catheterisation laboratories. Front. Cardiovasc. Med. 11, 1398290.

Lee, C.-Y., Xie, S., Gallagher, P., Zhang, Z., Tu, Z., 2015. Deeply-supervised nets. In: Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics. PMLR, pp. 562–570.

Li, L., Verma, M., Nakashima, Y., Nagahara, H., Kawasaki, R., 2020. IterNet: Retinal image segmentation utilizing structural redundancy in vessel networks. In: 2020 IEEE Winter Conference on Applications of Computer Vision. WACV, pp. 3645–3654.

Liang, D., Wang, L., Han, D., Qiu, J., Yin, X., Yang, Z., Xing, J., Dong, J., Ma, Z., 2021. Semi 3D-TENet: Semi 3D network based on temporal information extraction for coronary artery segmentation from angiography video. Biomed. Signal Process. Control. 69, 102894. http://dx.doi.org/10.1016/j.bspc.2021.102894.

Lin, G., Bai, H., Zhao, J., Yun, Z., Chen, Y., Pang, S., Feng, Q., 2022. Improving sensitivity and connectivity of retinal vessel segmentation via error discrimination network. Med. Phys. 49 (7), 4494–4507.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3431–3440.

Maninis, K.-K., Pont-Tuset, J., Arbeláez, P., Van Gool, L., 2016. Deep retinal image understanding. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI. Springer International Publishing, Cham, pp. 140–148.

Nasr-Esfahani, E., Karimi, N., Jafari, M., Soroushmehr, S., Samavi, S., Nallamothu, B., Najarian, K., 2018. Segmentation of vessels in angiograms using convolutional neural networks. Biomed. Signal Process. Control. 40, 240–251.

Oner, D., Kozinski, M., Citraro, L., Dadap, N.C., Konings, A.G., Fua, P., 2022. Promoting connectivity of network-like structures by enforcing region separation. IEEE Trans. Pattern Anal. Mach. Intell. 44 (09), 5401–5413. http://dx.doi.org/10.1109/TPAMI.2021.3074366.

Qiu, Y., Li, Z., Wang, Y., Dong, P., Wu, D., Yang, X., Hong, Q., Shen, D., 2023. CorSegRec: A topology-preserving scheme for extracting fully-connected coronary arteries from CT angiography. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 670–680.

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.

Ruan, J., Xiang, S., 2024. VM-UNet: Vision mamba UNet for medical image segmentation. arXiv:2402.02491.

Shin, S.Y., Lee, S., Noh, K.J., Yun, I.D., Lee, K.M., 2016. Extraction of coronary vessels in fluoroscopic X-Ray sequences using vessel correspondence optimization. In: Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016. Springer International Publishing, Cham, pp. 308–316.

Shit, S., Paetzold, J.C., Sekuboyina, A., Ezhov, I., Unger, A., Zhylka, A., Pluim, J.W., Bauer, U., Menze, B.H., 2021. clDice - a novel topology-preserving loss function for tubular structure segmentation. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, IEEE Computer Society, pp. 16555–16564. http://dx.doi.org/10.1109/CVPR46437.2021.01629.

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. Comput. Med. Imaging Graph. 115, 102392. http://dx.doi.org/10.1016/j.compmedimag.2024.102392.

Tan, Y., Yang, K.-F., Zhao, S.-X., Li, Y.-J., 2022. Retinal vessel segmentation with skeletal prior and contrastive loss. IEEE Trans. Med. Imaging 41 (9), 2238–2251. http://dx.doi.org/10.1109/TMI.2022.3161681.

Tu, S., Barbato, E., Köszegi, Z., Yang, J., Sun, Z., Holm, N.R., et al., 2014. Fractional flow reserve calculation from 3-dimensional quantitative coronary angiography and TIMI frame count. JACC: Cardiovasc. Interv. 7 (7), 768–777.

Vlontzos, A., Mikolajczyk, K., 2018. Deep segmentation and registration in X-ray angiography video. arXiv preprint arXiv:1805.06406.

Wan, T., Chen, J., Zhang, Z., Li, D., Qin, Z., 2021. Automatic vessel segmentation in X-ray angiogram using spatio-temporal fully-convolutional neural network. Biomed. Signal Process. Control. 68, 102646.

Xia, S., Zhu, H., Liu, X., Gong, M., Huang, X., Xu, L., et al., 2020. Vessel segmentation of X-Ray coronary angiographic image sequence. IEEE Trans. Biomed. Eng. 67 (5), 1338–1348. http://dx.doi.org/10.1109/TBME.2019.2936460.

Xiang, Y., Chung, A.C., Ye, J., 2006. An active contour model for image segmentation based on elastic interaction. J. Comput. Phys. 219 (1), 455–476.

Xie, Q., Guo, M., Mou, L., Zhang, D., Chen, D., Shan, C., Zhao, Y., Su, R., Zhang, J., 2024. Dsca: A digital subtraction angiography sequence dataset and spatio-temporal model for cerebral artery segmentation. arXiv:2406.00341.

Yang, S., Kweon, J., Roh, J.-H., Lee, J.-H., Kang, H., Park, L.-J., et al., 2019. Deep learning segmentation of major vessels in X-ray coronary angiography. Sci. Rep. 9, 16897.

Youssef, R., Ricordeau, A., Sevestre-Ghalila, S., Benazza-Benyahya, A., 2015. Evaluation protocol of skeletonization applied to grayscale curvilinear structures. In: 2015 International Conference on Digital Image Computing: Techniques and Applications. DICTA, pp. 1–6. http://dx.doi.org/10.1109/DICTA.2015.7371256.

Zhang, T.Y., Suen, C.Y., 1984. A fast parallel algorithm for thinning digital patterns. Commun. ACM 27 (3), 236–239.

Zhao, C., Tang, H., Tang, J., Zhang, C., He, Z., Wang, Y.-P., et al., 2020. Semantic segmentation to extract coronary arteries in fluoroscopy angiograms. MedRxiv.

Zhao, C., Xu, Z., Hung, G.-U., Zhou, W., 2023. Eagmn: Coronary artery semantic labeling using edge attention graph matching network. Comput. Biol. Med. 166, 107469. http://dx.doi.org/10.1016/j.compbiomed.2023.107469.

Zhou, Z., Siddiquee, M.M.R., 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. Springer, pp. 3–11.

Zou, P., Chan, P., Rockett, P., 2009. A model-based consecutive scanline tracking method for extracting vascular networks from 2-D digital subtraction angiograms. IEEE Trans. Med. Imaging 28 (2), 241–249.