Biomedical Signal Processing and Control 82 (2023) 104510

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

Centerline-supervision multi-task learning network for coronary angiography segmentation

Yuanxiu Zhang [a,1], Yufeng Gao [a,1], Guangquan Zhou [a], Jianan He [a], Jun Xia [b], Guoyi Peng [d], Xiaojian Lou [e], Shoujun Zhou [c,*], Hui Tang [a,*], Yang Chen [a,f]

[a] Laboratory of Image Science and Technology, Key Laboratory of Computer Network and Information Integration, Southeast University, Nanjing, Jiangsu, China
[b] Department of radiology, Shenzhen Second People's Hospital, Shenzhen, China
[c] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China
[d] Foshan Stomatology Hospital & School of Medicine, Foshan University, Foshan, China
[e] PLA NO 906 Hospital & School of Medicine, Ningbo University, Ningbo, China
[f] Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, Southeast University, Nanjing, China

ARTICLE INFO

Keywords:
Deep learning
Attention mechanism
Multi-task learning
Vessel segmentation

ABSTRACT

With convolutional neural networks' remarkable performance in computer vision, more and more studies are applying deep learning to vessel image segmentation tasks. This work focuses on the task of **coronary X-ray angiography segmentation**, which is critical in the diagnosis of cardiovascular disorders. For vascular segmentation in coronary X-ray angiography images, we propose a novel **deep learning model** based on the UNet backbone. We first equip a **channel attention module** in skip-connections to improve pixel-wise segmentation accuracy by emphasizing the effective channel in low-level features. **A centerline auxiliary supervision module** is also introduced at the network's end to provide prior knowledge of vessel connectivity and thick vessels, utilizing the existing binary segmentation annotations efficaciously. Consequently, the network can devote more attention to pixels that ensure vessel tree connectivity and high confidence. Extensive experiments demonstrate the effectiveness of the two modules in improving the performance of the model. We compared our results to the recently proposed networks and revealed that these two modules can be added to other U-shaped networks to enhance performance. In our experiments, our method produced the best results in terms of sensitivity and dice score, with 82.48 and 85.28, respectively.

1. Introduction

Coronary heart disease (CHD) is one of the most common cardiovascular diseases worldwide [1]. And digital subtraction angiography (DSA) [2–5] plays a vital role in clinical diagnosis of cardiovascular disorders, such as detection of stenosis and plaque. However, low contrast and motion artifacts caused by the beating heart degrade the quality of coronary X-ray angiography images, impacting the doctor's clinical diagnosis process. Accordingly, researchers intend to employ the image segmentation method to extract the whole vascular tree to increase diagnosis efficiency.

X-ray angiography highlights vessel structures by injecting the contrast agent into the patient's vessels and then imaging under X-ray. X-ray angiography images often contain severe noise and artifacts caused by heartbeat. Simultaneously, due to uneven distribution of contrast agent in the vessels and interference of the surrounding background tissue, the vessels obtained after imaging also show uneven gray distribution, making it difficult to distinguish the vessels from the tissues. As a result of these factors, X-ray angiography images are more complex to process. For the challenges outlined above in the X-ray angiography image's vascular segmentation task, traditional approaches frequently require additional pre-processing operations. Deep learning approaches have emerged as a end-to-end solution for image segmentation tasks. Through numerous convolutions, nonlinear operations, and backpropagation learning, the model may generate a more extensive feature space and learn pixel features on vessels more directly. Many experimental results suggest that deep learning methods can save a lot of time and effort in artificial feature design processes, as well as perform well in image segmentation tasks [6–9]. Data and annotations, which take a lot of people and effort, are significant barriers to applying deep learning to new activities. As a result, an increasing number of academics are turning their attention to unsupervised and semi-supervised learning, as well as making full use of the information

* Corresponding authors.
E-mail addresses: sj.zhou@siat.ac.cn (S. Zhou), corinna@seu.edu.cn (H. Tang).
[1] Yuanxiu Zhang and Yufeng Gao contributed equally to this work.

https://doi.org/10.1016/j.bspc.2022.104510
Received 31 August 2022; Received in revised form 18 November 2022; Accepted 18 December 2022
Available online 27 December 2022

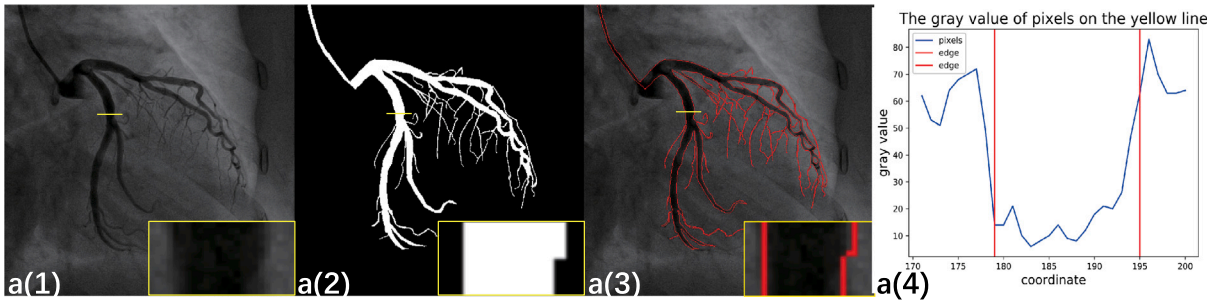

Fig. 1. a(1) the X-ray angiography image, a(2) the corresponding segmentation label, a(3) the X-ray angiography image with red boundary mark, a(4) the gray value of pixels on the yellow line, x-axis represents the location of the pixels, y-axis represents the gray value, and red lines represent the location of the boundary.

provided by existing annotations. Since ambiguity of vessel boundary and inaccuracy of manual marking, there are some uncertain pixels on the label, which will impair the model's performance. Fig. 1 shows four images, in which we can discover that the vessel's boundaries are ambiguous in a(1), (2), and (3) and the grey-scale distribution of pixels perpendicular to the direction of vessels in a(4). The red lines in a(4) reflect the location of the label border. The standard of labeling at the boundary is inconsistent, causing the category of pixels on the boundary to be unknown. Therefore, it is also crucial how to use annotations information accurately and effectively. In this paper, we proposed a UNet-based network, including the skip channel attention module and the centerline multi-task supervision module, which has been demonstrated to extract curvilinear structures from coronary X-ray angiography images efficiently through our experiments. General segmentation networks typically solely use binary segmentation labels for supervised learning. The centerline annotations obtained from segmentation annotations contain essential information that can be used to improve model segmentation performance, according to our analysis of the vessel segmentation problem. Thus, we designed a multi-task supervision module to efficiently employ annotation information to optimize the model. This paper makes the following contributions: (1) We added the channel attention module in skip-connections to filter the low-level features passed from the shallow layer to the deep layer in the network and demonstrated its effectiveness in enhancing the pixel-wise segmentation accuracy. (2) We took full advantage of vessel annotations: centerline annotations are derived from the binary annotations of vessels and adopted as an auxiliary supervision task of the model, which provides more priori knowledge for the model. Through several experiments, we discovered its efficiency on the X-ray angiography dataset. (3) We compared the results of multiple deep learning models on the coronary X-ray angiography dataset. In addition, our method exceptionally well in a variety of tasks. The remainder of this paper is laid out as follows. In Section 2, we present a brief review of related literature. In Section 3, we clarify the proposed method in depth. In Section 4, we present some experimental findings to show that the newly added modules are effective in stimulating the performance of the job, as well as a comparison of several other deep learning segmentation approaches based on the same dataset. Then, in Section 5, we discuss and draw some conclusions.

2. Related work

2.1. Segmentation methods

In earlier studies, there have an increasing number of researches on X-ray angiography image segmentation. Since the vessels exhibit prominent tubular characteristics, some researchers focused on this and presented a vessel enhancement approach based on the hessian matrix due to its eigenvectors related to vascular structure. Based on the enhanced images, the foreground vessels and background can be separated more easily [10–13]. Based upon the related researches of the hessian matrix, some scholars have introduced a multidirectional filter bank, which causes the feature map to have a greater filter response along the direction of the vessels [14–16]. The multi-directional filter bank is intended to reduce background noise and enhance vessel response. Because X-ray angiography images are a series of images in time, some studies attempted to maximize the information in the time dimension by using matrix decomposition to remove the background and obtain a more exact blood vessel tree [17–19]. However, considering matrix decomposition is an iterative process, it takes a long time to decompose high-dimensional matrices. Some tracking methods [20–22] built the segmentation pipeline using prior information from structures such as centerlines or surfaces. Long et al. [6] proposed a fully convolutional network (FCN) for semantic segmentation, which discards the fully connected layers in the traditional network architecture and replaces it with a deconvolution layer to enable the end-to-end image segmentation. FCN's accomplishment demonstrates that the end-to-end network can be used to address the image segmentation problem with good results. On this basis, Ronneberger et al. [7] suggested a U-shaped network architecture (UNet) with an encoding path, a decoding path, and skip-connection paths for transmitting the feature map from encoding path to decoding path, which performs well on various medical applications. Many useful structures have been proposed to increase the performance of deep learning models, such as normalization [23–26], residual block [27,28], dense block [29,30], inception block [31,32], dilated convolution [33], feature aggregation module [34], spatiotemporal super-resolution framework [35–37], encoder–decoder architectures [38,39] and so on. The current researches on 2D vessel segmentation methods are more focused on the retinal vessel segmentation. In order to integrate global, local, and channel features, Mou et al. [8] established channel and spatial attention network (CS-Net), which contains the residual encoder module and the channel and spatial attention module (CSAM). Gu et al. [9] adopted dense atrous convolution (DAC) block and residual multi-kernel pooling (RMP) to encode multi-scale context features, which named Context Encoder Network (CE-Net). Meanwhile, some scholars are also working on designing some novel loss functions to improve segmentation outcomes. Navarro et al. [40] added two auxiliary tasks, a distance map and a contour map, at the end of the network for higher precision segmentation. Simultaneously, a growing number of deep learning models [41–48] are being applied to volume data segmentation tasks.

2.2. Attention mechanism

The attention method was first employed for machine translation jobs by Bahdanau et al. [49]. In machine translation tasks, different words have different importance in the same sentence due to context content. As a result, different words should have different weights, calculated by the attention module. And now attention mechanism is a common and useful strategy in machine learning, which has been widely used in the fields of computer vision, knowledge graph, and natural language processing [50,51]. Like the human cognitive process, the

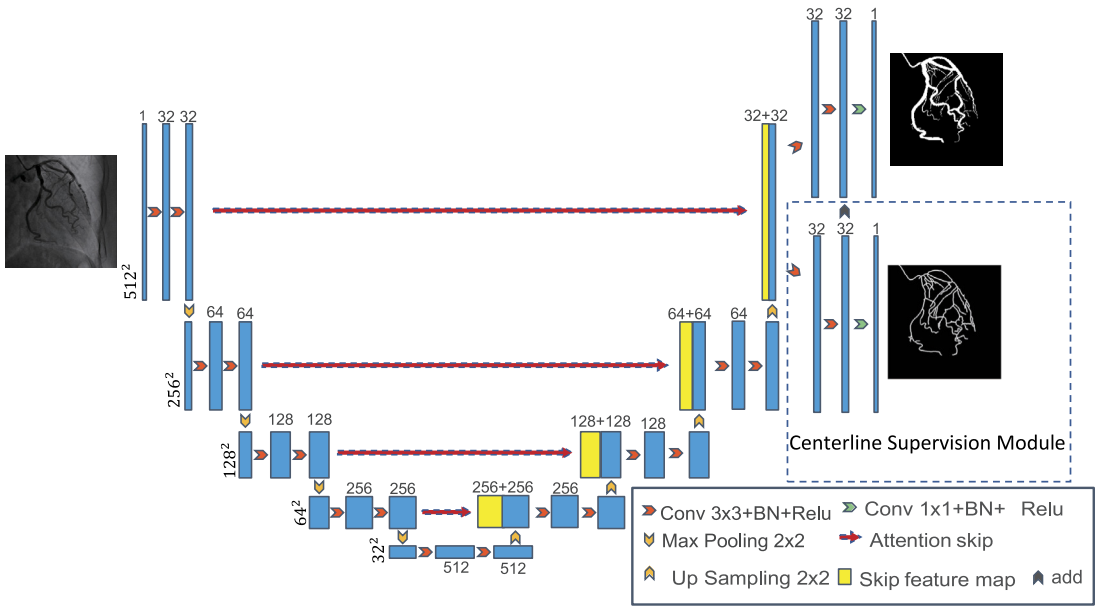

Fig. 2. Illustration of the proposed centerline-supervision multi-task learning network for coronary angiography segmentation. While retaining the backbone of U-Net, two modules, channel attention in skip-connections and centerline auxiliary supervision module at the end, have been added to this network.

model focuses more on local information from the global information by adding the attention mechanism. According to different dimensions, the soft attention mechanism can be divided into three categories: spatial attention [52], channel attention [53] and mixed attention [54]. Oktay et al. [55] utilized a spatial attention mechanism that combines an encoding feature map and a decoding feature map to solve the pancreas segmentation challenge. The attention mechanism enables the deep learning networks to pay attention to sufficient information from a vast number of features.

Attention mechanism is a brain signal processing mechanism unique to human vision. Human vision can identify the target area that needs to be focused on-also known as the focus of attention-by quickly scanning the entire image, and then devote more attention resources to this area to gather more specific information about the target while suppressing irrelevant information. This is a way for human to quickly filter out high-value information from a big volume of information using their limited attentional resources. It is a survival strategy developed over the course of human evolution. The human visual attention system significantly increases the speed and precision of processing visual data.

In essence, the attention mechanism in deep learning is similar to the selective visual attention mechanism in humans. The main purpose is to narrow down the information from a wide pool of knowledge to that which is more important to the task at hand. The weight symbolizes the significance of the information, and the weight coefficient computation reflects the focusing process. Regarding the precise calculation method of the Attention mechanism, it can be divided into two steps: the first step is to determine the information's weight coefficient, and the second step is to carry out the information's weighted summation in accordance with the weight coefficient.

2.3. Multi-task learning

Rich Caruana first proposed multi-task learning in 1997 [56]. Multi-task Learning is an approach to inductive transfer that improves generalization by using the domain information in the training signals of related tasks as an inductive bias. According to recent studies, multi-task learning can train two target tasks or improve the main task's performance by adding additional tasks. Navarro et al. [40] used distance map and contour map as complementary tasks in multi-organ segmentation. The addition of auxiliary learning tasks can add some prior knowledge into the network. In the multi-organ segmentation task, learning the distance and contour map can enforce shape-prior job leveraging the existing target labels. The proposed model retains the backbone of UNet [7]: an encoding path, a decoding path, and skip-connections, since this model performs better and is more lightweight in trails. And on this basis, we made the following improvements: channel attention module in skip-connections and centerline auxiliary supervision module at the end of the network, as indicated in Figs. 2 and 3.

3. Material and methods

3.1. Channel attention skip

In our work, channel attention [53] is adopted in the skip-connections to emphasize the features passed from the encoding path to the decoding path. The whole calculation process can see Fig. 3. We can define the form of the attention skip module (AS) as: $\mathbf{X} \rightarrow \mathbf{Y}, \mathbf{X} \in \mathbb{R}^{C\times H\times W}, \mathbf{Y} \in \mathbb{R}^{C\times H\times W}$, where $\mathbf{X}$ and $\mathbf{Y}$ represent the feature map of input and output, respectively. And $\mathbf{X} = [x_1, x_2, \ldots, x_i, \ldots, x_C], \mathbf{Y} = [y_1, y_2, \ldots, y_i, \ldots, y_C], x_i, y_i \in \mathbb{R}^{H\times W}$. The first step is to use the global average pooling (**Global Pool** Module in Fig. 3) to integrate the spatial information on each channel.

$$z_i = \frac{1}{H \times W} \sum_{m=1}^{H} \sum_{n=1}^{W} x_i(m, n) \tag{1}$$

where $\mathbf{Z} \in \mathbb{R}^{C\times 1}, \mathbf{Z} = [z_1, z_2, \ldots, z_i, \ldots, z_C]$ represents the feature map after global average pooling. In order to capture the channel-wise dependencies with a small amount of calculations and parameters, this module uses two fully connected (**FC** Module in Fig. 3) layers in the next.

$$\boldsymbol{S} = g(\boldsymbol{W_2}\sigma(\boldsymbol{W_1}\boldsymbol{Z})) \tag{2}$$

where $\sigma(\cdot)$ refers to ReLu function, $g(\cdot)$ refers to Sigmoid function, and $\boldsymbol{W_1} \in \mathbb{R}^{\frac{C}{r}\times C}, \boldsymbol{W_2} \in \mathbb{R}^{C\times\frac{C}{r}}$ (r is the reduction ratio) represent the parameters in fully connected layers. And $\boldsymbol{S} \in \mathbb{R}^{C\times 1}$ is the scale factor calculated from the global features for $\boldsymbol{X}$, then $\boldsymbol{Y}$ can be calculated by

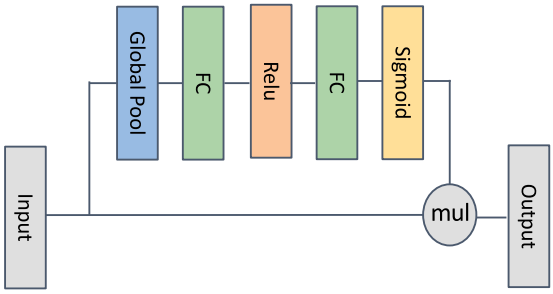

Fig. 3. Illustration of the attention skip module. For each skip-connection layer, the channel attention scale mask is calculated by global average pooling and fully connected layers. Then the encoding feature map is scaled to obtain the output.

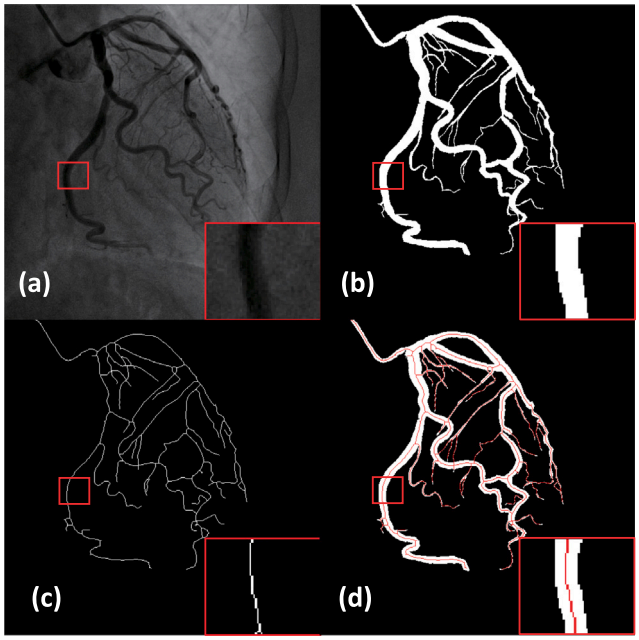

Fig. 4. An example of (a). X-ray angiography image, (b). vessel annotations, (c). centerline image, (d). fusion image of vessel and centerline, white pixels represent vessels and red pixels represent centerline.

channel-wise multiplication between the input feature map $\boldsymbol{X}$ and the scalar $\boldsymbol{S}$.

$$\boldsymbol{Y} = \boldsymbol{S} \cdot \boldsymbol{X} \tag{3}$$

The excellent performance of the U-Net [7] depends mostly on the skip-connections because they can combine the low-level features with the high-level features, which is in line with the characteristics of medical images. Inspired by [53,57], the residual block similar to the skip-connections has better performance after adding the channel attention module. Therefore, we hope that the effective low-level feature map can be enhanced in the skip-connections to better match the high-level features. In summary, the channel attention module compresses features through the global pooling layer, and then combines the weight information in each channel through the fully connected layer to generate a weight mask. And the effectiveness of the attention skip is shown in Section 4.

3.2. Centerline auxiliary task

A problematic point in vessel segmentation is uneven vessel thickness. For the pixel-wise segmentation task, the goal is to classify every pixel correctly, so the segmentation result is more biased towards the optimal pixel classification and ignores the segmentation target's shape characteristics. In addition, for vessel segmentation tasks, the confidence between pixels is often different. The confidence of pixels on the boundary is lower than that on the central axis due to the labeling error. In our work, inspired by multi-task learning, we use the centerline auxiliary supervision module to enhance thin vessels' filter response during training.

Centerline Annotations: Inspired by the method [58], we analyze the role of centerline in 2D vessel segmentation. To show the characteristics of the centerline, we list four images in Fig. 4, from left to right are X-ray angiography image, vessel annotation, centerline image, and fusion image where white pixels represent vessel and red pixels represent centerline. We summarized the four characteristics of centerline: **(1)** The set of centerline points is the smallest point set that can express the topology and connectivity of vessels. This means that accurate prediction of centerline can ensure the integrity of the vessel tree structure. **(2)** The thickness of vessels in X-ray angiography images varies greatly, and the thin vessels that are more difficult to segment are almost equivalent to the corresponding centerlines, as shown in Fig. 4(d). Therefore, learning the centerline as an auxiliary task can be considered as selective learning for difficult samples. **(3)** The foreground pixels represented by the centerline are positive samples with higher confidence. Due to manual labeling error, the vessel edge pixels category is often uncertain, and the pixels on centerline have a higher degree of confidence than the edge pixels. This means that using the centerline as the pixel-wise segmentation task's target can keep the difference between positive and negative samples and make the model learn the difference between the vessel pixels and non-vessel pixels more fully. **(4)** The centerline annotations no longer require additional manpower but can be easily obtained from the vessel annotations by distance transformation. Therefore, we believe that using the centerline annotations for training is a more efficient use of the segment annotations.

Multi-task Learning: And in vessel segmentation task, we identify the point set of foreground pixels in centerline as $\boldsymbol{P}_{cl}$ and the foreground pixels in vessels as $\boldsymbol{P}_{v}$. Obviously, the relationship between the two sets is $\boldsymbol{P}_{cl} \subset \boldsymbol{P}_{v}$. Therefore, the centerline supervision task can be regarded as a subtask in vessel segmentation task and can share the network's low-level features. And we found in experiments that, for example, the parallel multi-task training method in [40] will affect the accuracy of the main task due to the existence of ambiguity information between different tasks. Fig. 5(a) illustrates parallel supervision and each supervision task has an independent prediction path, resulting in different tasks with different performance. Because the centerline prediction task is a subtask of the segmentation task, we add the centerline feature map to the segmentation task to highlight the response of centerline and surrounding pixels and also ensure the importance of the main task, which named feature map fusion, see Fig. 5(b), (c). And we list the experiment results in Section 4, which demonstrates that adding feature map fusion operations is beneficial to improve the performance of the main task.

3.3. Network architecture and loss function

Network Architecture: The overall architecture of our network is shown in Fig. 2. The backbone is U-Net [7], including encoding path, decoding path, and skip-connections. On this foundation, we include the centerline supervision auxiliary module and the channel attention module. There are four downsampling blocks in the encoding path, each of which includes two convolutional layers followed by an attention skip and a max-pooling. And in the decoding path, there are four corresponding upsampling blocks. Each convolutional layer includes a 3×3 convolution, a normalization layer, and a Relu activation function. And in skip-connections, the feature map is calculated by channel attention module. After the last upsampling layer, we add a

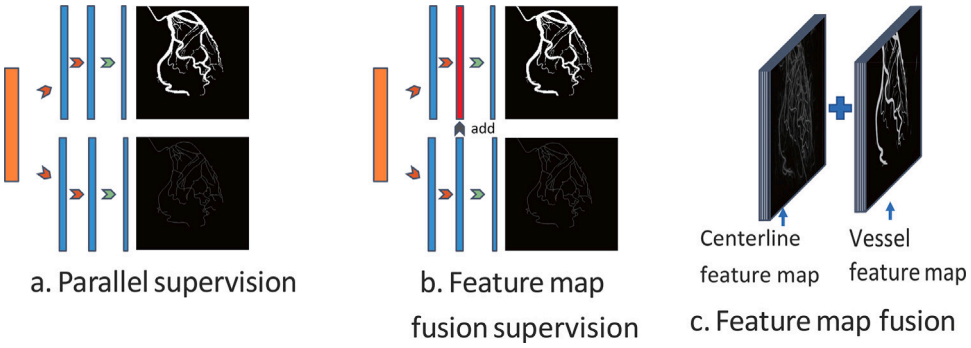

Fig. 5. Illustration of different multi-tasks supervision methods: (a). parallel supervision, each task has an independent prediction path. (b). feature map fusion supervision, the sub-task feature map is added to the main task feature map (the red block). (c). illustration of feature map fusion. In (a) and (b), the orange block represents the last upsampling layer in the U-shaped network.

Table 1
Comparison of the results of U-Net, U-Net with channel attention skip and Attention U-Net.

	Para	Evaluation 1					Evaluation 2			
		Auc	Acc.	Sen.	Sp.	Dice	Acc.	Sen.	Sp.	Dice
UNet	8.62M	98.08	98.06	81.62	99.17	84.66	98.07	79.89	99.28	84.47
UNet+AS	8.66M	**98.13**	**98.13**	**82.27**	**99.21**	**85.13**	**98.13**	80.45	**99.32**	84.84
Att UNet [55]	8.72M	98.11	98.11	82.24	99.18	85.09	98.11	**80.60**	99.29	**84.86**

prediction path for centerline task and a feature map fusion operation. In order to show the role of the two modules (attention skip and centerline auxiliary supervision), we also add these two modules to other backbones (CS-Net [8] and CE-Net [9]) to verify the effectiveness of the modules on the improvement of segmentation performance, as shown in Section 4.

Loss Function: We adopt two different loss functions on these two tasks. Firstly, we use $\mathcal{L}_{seg}$ to represent the vessel segmentation loss. $\mathcal{L}_{seg}$ is a combination of dice loss $\mathcal{L}_{dice}$ and crossentropy loss $\mathcal{L}_{CE}$ which has been proven to have better segmentation performance in [42], defined as:

$$\mathcal{L}_{seg} = \mathcal{L}_{dice} + \mathcal{L}_{CE} \tag{4}$$

Since the centerline prediction is an extreme sample imbalance task, we take focal loss [59] as the loss function of the centerline supervision task. Given the probability $p(x)$ of a pixel in location x to belong to the foreground class and the ground truth by $g(x)$. $\mathcal{L}_{cl}$ is defined as:

$$\begin{aligned}\mathcal{L}_{cl} = \frac{1}{N}\sum_{x} &-\alpha(1-p(x))^{\gamma} g(x) log(p(x)) \\ &-(1-\alpha)p(x)^{\gamma}(1-g(x))log(1-p(x))\end{aligned} \tag{5}$$

where N represents the total number of pixels, α, γ are the hyperparameters. Focal loss can control the proportional relationship between positive and negative samples through parameter α, and parameter γ is used to adjust the loss of difficult samples. Therefore, focal loss is more suitable for the centerline prediction task with extremely imbalance samples and more difficult samples. We also compared the differences between different loss functions in the centerline task in Section 4.

And the total loss $\mathcal{L}$ can be defined as:

$$\mathcal{L} = w_{seg}\mathcal{L}_{seg} + w_{cl}\mathcal{L}_{cl} \tag{6}$$

where w_{seg} and w_{cl} respectively represent the weights corresponding to the loss of the two tasks.

4. Experimentals and results

4.1. Experimental details

Dataset: We obtained 300 X-ray angiography images with expert annotations from Shenzhen Second People's Hospital, half of which are left coronary arteries (LCA), and the other half are right coronary arteries (RCA). And centerline labels are obtained by skeletonizing the segmentation labels. We normalized the images to improve the contrast because the gray distribution is uneven. Since our dataset is not large, data augmentation is indispensable. We adopted the data augmentation method similar to nnunet [42] in the training set of all experiments, including elastic deformation, rotation, gaussian noise transformation, gamma transformation, etc. And we adopted 4-fold cross-validation method, all results are the average across all folds.

Implementation Details: We used PyTorch with the NVIDIA GPU TITAN X to implement all the experiments in this paper. The initial learning rate was set at 0.003, weight decay 0.00001 and plateau learning rate decay strategy is adopted in our experiments. We chose adaptive moment estimation (Adam) optimization. The maximum epoch is 300. And in loss function, α is 0.5 and γ is 2, w_{seg} and w_{cl} are both 1.0. And r in channel attention module is 8.

Evaluation Metrics: To compare performance on vessel segmentation, we compute the evaluation metrics, namely sensitivity (Sen), accuracy (Acc), specificity (Sp) and dice similariy coefficient (DSC), are defined as

$$\begin{aligned} Sen &= \frac{TP}{TP+FN} \qquad Acc = \frac{TP+TN}{N} \\ Sp &= \frac{TN}{TN+FP} \qquad DSC = \frac{2TP}{2TP+FP+FN} \end{aligned} \tag{7}$$

where TP, TN, FP, FN represent the number of true positives, true negatives, false positives and false negatives, respectively, and $N = TP + FP + TN + FN$. In addition, we also adopted the area under receiver operation characteristic curve (AUC) to measure segmentation performance. The calculation of evaluation metrics directly on network output results can reflect the advantages and disadvantages of the network in pixel-wise segmentation results. To measure the connectivity of vessels in the segmentation results, we extract the maximum connected domain of the network output to ensure that the obtained vessel tree is complete and connected, reflecting whether there is rupture in the segmentation result. Therefore, the comparison of our experimental results includes two aspects: **evaluation 1,** calculate the evaluation metrics on the output of the network (threshold is 0.5 in binarization), **evaluation 2,** calculate the evaluation metrics on the maximum connected domain (threshold is 0.5 in binarization).

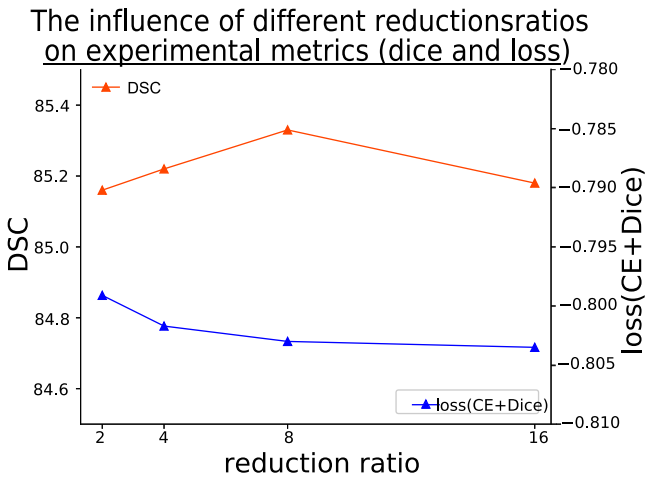

Fig. 6. Illustration of the influence of different reduction ratios(2, 4, 8, 16) on DSC and training loss. The red curve represents the dice similarity coefficient on validation set and the blue curve represents the loss in training phase.

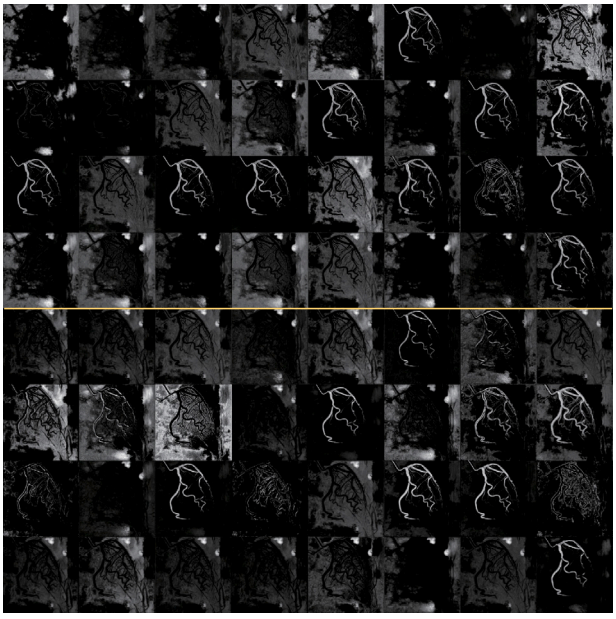

Fig. 7. Illustration of the feature maps for feature fusion. The feature maps above the yellow line are obtained from the segmentation branch, and the feature maps below the yellow line are obtained from the centerline prediction branch.

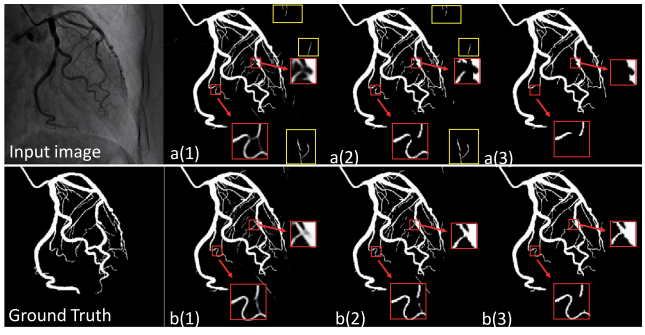

Fig. 8. Illustration of evaluation 1 and evaluation 2. The two images in the first column respectively represent the input images and the ground truth. *a* and *b* represent the results of UNet and our proposed method, respectively. And 1, 2, 3 represent the model's output logits, binary results, and maximum connected domain in binary results, respectively.

4.2. Attention skip (AS) in vessel segmentation

Firstly, in order to explore the role of the attention mechanism in the skip structure in U-Net network, we try to add channel attention (U-Net+AS) and spatial attention (Att U-Net [55]) to the U-Net backbone, as shown in Table 1. We can see that the addition of two attention modules can improve the U-Net's performance when the network model parameters are almost unchanged. The sensitivity and dice score are improved (sensitivity increased by about 0.6, the dice score increased by about 0.5) after adding the attention mechanism, which shows that these two networks can segment more true vessel pixels than base U-Net. The channel attention module performs better in evaluation 1, which shows that the addition of the channel attention mechanism improves the model's accuracy for pixel-wise segmentation. On the other hand, attention U-Net (spatial attention module) prefers to emphasize spatial information, so the connectivity of binary results is more complete and the metric performance is better in evaluation 2. From the results of these three experiments, it can be concluded that adding the attention module to the skip-connections connecting the low and high feature maps in the U-shaped network, whether it is channel attention or spatial attention, can improve pixel-wise segmentation performance. And then we conducted some experiments for hyperparameters r which represents the reduction ratio in attention module. As shown in Fig. 6, we observed DSC and loss value by setting r to 2, 4, 8, 16. The DSC (red curve) reaches the highest value when r is 8, in the meanwhile the loss (blue curve) reaches a low level. Therefore, we set r to 8 in our model.

4.3. Centerline supervision (CL) in vessel segmentation

We then integrated the centerline supervision module with the network that added channel attention in skip-connections. We list the results in Table 2. The third row (UNet+AS, CL(parallel)) represents the model similar to Navarro's [40]. This model has two diverging parallel prediction branches at the end of the last up convolution. And the fourth row (UNet+AS, CL(fusion)) represents the model that includes feature map fusion between the two prediction branches. We can see that the third-row result is almost the same as the second-row result, which means that parallel branches do not improve the performance much. Since the centerline prediction task can be considered a subtask of the vessel segmentation task, the feature maps' high-response regions should be similar to that of vessels, shown in Fig. 7. Therefore, we add centerline feature maps to the vessel segmentation task to enhance the thin vessel's response. The fusion centerline supervision model (the fourth row) improves performance simultaneously on two evaluations based on the channel attention model. The centerline auxiliary supervision module can be considered as a prior spatial attention mechanism. Given the centerline as a sub-task, the network can pay more attention to the important pixels on the vessel skeleton, which are often the pixels that ensure the connectivity and topology of the vessels. Secondly, in the centerline auxiliary supervision task module, we tried a variety of loss functions. As shown in Table 3, we only changed the loss function of the centerline task while keeping the network structure (attention skip and centerline auxiliary supervision) unchanged, and conducted a series of comparison experiments. The four rows from top to bottom of the table represent the results of experiments with different loss functions, including cross-entropy (CE), soft dice (Dice), the combined loss of cross-entropy and dice (CE+Dice), and focal loss. The cross-entropy loss function is a classic loss function in machine learning, which is often used to calculate the pixel-level loss in the field of image segmentation. And soft dice loss [43] is proposed to solve the problem of uneven foreground and background categories, but due to the gradient's instability when it propagates in the direction, it is often used in conjunction with cross-entropy loss [42]. Based on cross-entropy loss, focal loss [59] is proposed to improve difficult samples' learning speed in target detection applications. The foreground pixels in centerline images account for a very small proportion, that is, the

Table 2
Comparison of the results of centerline auxiliary supervision. The third and fourth rows represent the network with the addition of the centerline supervision module. The CL module in the third model contains two parallel prediction paths, while the feature map fusion is added to the module in the fourth model.

	Para	Evaluation 1					Evaluation 2			
		Auc	Acc.	Sen.	Sp.	Dice	Acc.	Sen.	Sp.	Dice
UNet	8.62M	98.08	98.06	81.62	99.17	84.66	98.07	79.89	99.28	84.47
UNet+AS	8.66M	98.13	98.13	82.27	**99.21**	85.13	98.13	80.45	**99.32**	84.84
UNet+AS+CL(parallel)	8.69M	98.20	98.12	82.28	99.18	85.10	98.12	80.55	99.30	84.85
UNet+AS+CL(fusion)	8.69M	**98.22**	**98.15**	**82.61**	99.20	**85.32**	**98.15**	**80.94**	99.31	**85.12**

Table 3
Comparison of the results of the different loss functions in the centerline auxiliary supervision module. Each row from top to bottom lists the results of using CE (cross-entropy) loss, Dice loss, CE+Dice loss, and focal loss in the centerline supervision module.

	Para	Evaluation 1					Evaluation 2			
		Auc	Acc.	Sen.	Sp.	Dice	Acc.	Sen.	Sp.	Dice
UNet+AS+CL(CE)	8.69M	98.04	98.12	81.92	**99.21**	84.97	98.12	80.29	**99.32**	84.82
UNet+AS+CL(Dice)	8.69M	98.17	98.10	82.05	99.18	84.92	98.10	80.19	99.31	84.69
UNet+AS+CL(CE+Dice)	8.69M	98.18	98.12	82.42	99.17	85.05	98.12	80.60	99.29	84.75
UNet+AS+CL(focal)	8.69M	**98.22**	**98.15**	**82.61**	99.20	**85.32**	**98.15**	**80.94**	99.31	**85.12**

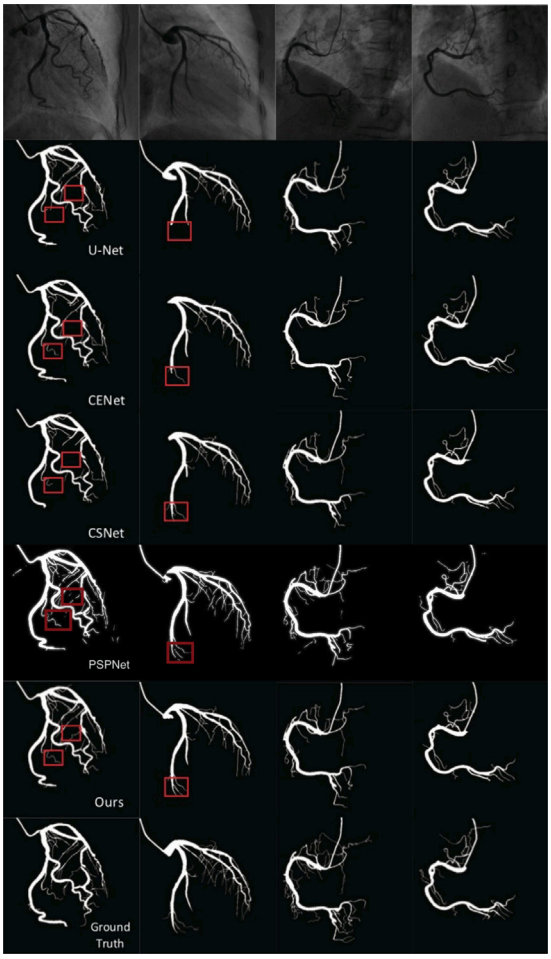

Fig. 9. Qualitative results: the four columns represent the four test data, and six rows represent X-ray angiography images, results of U-Net, CENet, CSNet, PSPNet, the proposed method, and ground truth from top to bottom.

centerline supervision task can be regarded as a task with a severe imbalance between positive and negative samples. The experimental results also show that using focal loss as the auxiliary task loss function has reached the highest in several assessment criteria.

4.4. Quantitative and qualitative results

We also conducted quantitative and qualitative assessments among multiple methods, including U-Net [7], attention U-Net [55], CENet [9], CSNet [8], PSPNet [46] and our proposed method. As the segmentation networks were newly proposed in recent years, CSNet, and CENet have better performance on fundus vessel image segmentation, so we also used them in the X-ray angiography image segmentation. Besides, PSPNet was recently used in X-ray angiography image segmentation and obtain good results. See Table 4, our proposed method has better performance in both evaluation methods, especially in sensitivity and dice score. And the backbone of our method is U-Net, but in the experiments, we found that these two modules are migratable, which means that the addition of these two modules in other U-shaped networks can also bring some performance improvements. In Table 4, the sixth row (CENet+AS, CL) and the seventh row (CSNet+AS, CL) represent the results of adding two modules to the CENet and CSNet models, respectively. The results show that the evaluation metrics based on the network with AS and Cl module are improved compared with the prototype networks. In qualitative assessment, firstly, we will explain why we use two evaluation metrics for evaluation. As shown in Fig. 8, the two images in the first column, respectively, represent the input images and the ground truth. *a* and *b* represent the results of UNet and our proposed method. And the images with index 1, 2, and 3 represent the output of the model, binary results, and maximum connected domain in binary results, respectively. Evaluation 1 is based on binary images (a(2) and b(2)), which means that some discrete vessel pixels will also be considered as true positives for statistics. However, these pixels cannot be directly used in the subsequent image analysis because these vessels are not an extension of the vessel tree. Therefore, we extracted the maximum connectivity from the binary images, which can be regarded as a complete vessel tree that can be used for subsequent processing. The areas marked by the red boxes are located inside the vessel tree, but due to the contrast agent dose or noise issues, the probability response is low in UNet results. Multi-task learning makes the model pay more attention to the classification of these pixels in the training process so that its probability response is improved, and the thin vessels can be retained on the vessel tree in the binary images (as shown in a(3) and b(3)). At the same time, it can also be seen from Fig. 8 that our method has a better suppression on non-vessel pixels (false positive samples, marked by yellow boxes) because the foreground pixels and non-vessel pixels in the centerline annotations

Table 4
Comparison of several methods in coronary angiography image segmentation, including U-Net, attention U-Net, CENet, CSNet, PSPNet and our proposed method. Besides, we also added AS and CL modules based on CSNet and CENet.

	Para	Evaluation 1					Evaluation 2			
		Auc	Acc.	Sen.	Sp.	Dice	Acc.	Sen.	Sp.	Dice
UNet [7]	8.62M	98.08	98.06	81.62	99.17	84.66	98.07	79.89	99.28	84.47
Att UNet [55]	8.72M	98.11	98.11	82.24	99.18	85.09	98.11	80.60	99.29	84.86
CENet [9]	28.9M	97.92	98.03	81.00	99.18	84.24	98.04	78.96	**99.32**	84.05
CSNet [8]	8.92M	98.15	98.08	81.72	99.18	84.69	98.07	79.31	**99.32**	84.34
PSPNet [46]	29.4M	98.19	98.03	81.67	99.18	84.73	98.06	79.02	**98.23**	84.48
CENet+AS+CL	29.0M	98.06	98.07	81.44	99.19	84.53	98.07	79.47	**99.32**	84.29
CSNet+AS+CL	8.99M	**98.23**	98.10	82.16	99.18	84.97	98.10	80.27	99.30	84.70
Proposed	8.69M	98.22	**98.15**	**82.61**	**99.20**	**85.32**	**98.15**	**80.94**	99.31	**85.12**

have a larger difference between classes. Therefore, we believe that the centerline auxiliary supervision module can be considered a spatial attention mechanism based on prior knowledge. And then, we list four coronary angiography images and their corresponding results in Fig. 9. These results are from UNet, CENet, CSNet, PSPNet and the proposed method. There are some apparent differences in the two images on the left, marked with red boxes. The two images on the right mainly show the difference in some thin vessels. We can see that these methods can roughly extract the vessel tree, and the main difference is the integrity of the vessels and the segmentation of the small thin vessels. Due to the effect of contrast agent dose, many vessels in X-ray angiography images show low contrast and are very difficult to segment. In comparison, our method can segment a complete vessel tree to a greater extent.

5. Discussions

Since medical images often contain clear and specific segmentation targets, low-level features play an important role in model inference. This is one of the reasons why UNet has excellent performance in medical image segmentation. The channel attention module in the skip-connections can be considered as filtering some redundant information in low-level features to improve the pixel-wise segmentation ability of the model in experiments. We do not change the feature map's content but filter the channel dimension's features to retain the information, which is more important for the network to make predictions. The common segmentation methods mainly use cross-entropy loss for the pixel-wise prediction, where each pixel has the same weight. Centerline supervision task can make the network more focused on the pixels, which have high confidence and maintain vessels connectivity. Therefore, we believe that the centerline labels are valuable, and its acquisition method is direct and straightforward, which can be considered a more effective use of the original segmentation label. However, there is still much future work in multi-task learning: coordinate the two tasks' learning progress and design a more suitable centerline loss or joint loss. Since the learning progress of different tasks is different, synchronizing the two tasks' learning can help the model converge more stably and efficiently. Besides, the joint loss can combine the characteristics of the two tasks with learning more deeply the relationship between different tasks.

6. Conclusions

In our work, we proposed an UNet-based network architecture, including two modules: the channel attention skip connection and the centerline auxiliary supervision. The channel attention skip module can be regarded as a general structure that can be used to improve the model's accuracy. And the centerline auxiliary supervision module is proposed based on the key point in the vessel segmentation task and the characteristics of the centerline itself. In the subsequent work, we will continue to study how to use the centerline information to achieve better results in the vessel segmentation task.

CRediT authorship contribution statement

Yuanxiu Zhang: Conceptualization, Methodology, Software. **Yufeng Gao:** Methodology, Writing – original draft. **Guangquan Zhou:** Data collection, Data curation. **Jianan He:** Data collection, Data curation. **Jun Xia:** Visualization, Investigation. **Guoyi Peng:** Software, Validation. **Xiaojian Lou:** Software, Validation. **Shoujun Zhou:** Writing – review & editing. **Hui Tang:** Writing – review & editing. **Yang Chen:** Software, Validation, Supervision.

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.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported in part by the State Key Project of Research and Development Plan, China under Grants 2022YFC2408500 and 2022YFC2401600, in part by the National Natural Science Foundation of China under Grant T2225025, in part by the Key Research and Development Programs in Jiangsu Province of China under Grant BE2021703 and BE2022768.

References

[1] E.J. Benjamin, S.S. Virani, C.W. Callaway, A.M. Chamberlain, A.R. Chang, S. Cheng, S.E. Chiuve, M. Cushman, F.N. Delling, R. Deo, et al., Heart disease and stroke statistics—2018 update: a report from the American Heart Association, Circulation (2018).

[2] S. Bash, J.P. Villablanca, R. Jahan, G. Duckwiler, M. Tillis, C. Kidwell, J. Saver, J. Sayre, Intracranial vascular stenosis and occlusive disease: evaluation with CT angiography, MR angiography, and digital subtraction angiography, Am. J. Neuroradiol. 26 (5) (2005) 1012–1021.

[3] S.E. Nissen, J. Elion, D. Booth, J. Evans, A. DeMaria, Value and limitations of computer analysis of digital subtraction angiography in the assessment of coronary flow reserve, Circulation 73 (3) (1986) 562–571.

[4] O. Wink, W.J. Niessen, M.A. Viergever, Fast delineation and visualization of vessels in 3-D angiographic images, IEEE Trans. Med. Imaging 19 (4) (2000) 337–346.

[5] T.-C. Huang, C.-K. Chang, C.-H. Liao, Y.-J. Ho, Quantification of blood flow in internal cerebral artery by optical flow method on digital subtraction angiography in comparison with time-of-flight magnetic resonance angiography, PLoS One 8 (1) (2013) e54678.

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

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

[8] L. Mou, Y. Zhao, L. Chen, J. Cheng, Z. Gu, H. Hao, H. Qi, Y. Zheng, A. Frangi, J. Liu, CS-Net: Channel and spatial attention network for curvilinear structure segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2019, pp. 721–730.

[9] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, J. Liu, CE-Net: context encoder network for 2D medical image segmentation, IEEE Trans. Med. Imaging 38 (10) (2019) 2281–2292.

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

[11] K. Krissian, G. Malandain, N. Ayache, R. Vaillant, Y. Trousset, Model based multiscale detection of 3D vessels, in: Proceedings. Workshop on Biomedical Image Analysis (Cat. No. 98EX162), IEEE, 1998, pp. 202–210.

[12] T. Jerman, F. Pernuš, B. Likar, Ž. Špiclin, Beyond Frangi: an improved multiscale vesselness filter, in: Medical Imaging 2015: Image Processing, Vol. 9413, International Society for Optics and Photonics, 2015, p. 94132A.

[13] A. Carballal, F.J. Novoa, C. Fernandez-Lozano, M. García-Guimaraes, G. Aldama-López, R. Calviño-Santos, J.M. Vazquez-Rodriguez, A. Pazos, Automatic multiscale vascular image segmentation algorithm for coronary angiography, Biomed. Signal Process. Control 46 (2018) 1–9.

[14] P.T. Truc, M.A. Khan, Y.-K. Lee, S. Lee, T.-S. Kim, Vessel enhancement filter using directional filter bank, Comput. Vis. Image Underst. 113 (1) (2009) 101–112.

[15] X. Xu, B. Liu, F. Zhou, Hessian-based vessel enhancement combined with directional filter banks and vessel similarity, in: 2013 ICME International Conference on Complex Medical Engineering, IEEE, 2013, pp. 80–84.

[16] T. Wan, X. Shang, W. Yang, J. Chen, D. Li, Z. Qin, Automated coronary artery tree segmentation in x-ray angiography using improved hessian based enhancement and statistical region merging, Comput. Methods Programs Biomed. 157 (2018) 179–190.

[17] H. Ma, A. Hoogendoorn, E. Regar, W.J. Niessen, T. van Walsum, Automatic online layer separation for vessel enhancement in X-ray angiograms for percutaneous coronary interventions, Med. Image Anal. 39 (2017) 145–161.

[18] M. Jin, R. Li, J. Jiang, B. Qin, Extracting contrast-filled vessels in X-ray angiography by graduated RPCA with motion coherency constraint, Pattern Recognit. 63 (2017) 653–666.

[19] S. Xia, H. Zhu, X. Liu, M. Gong, X. Huang, X. Lei, H. Zhang, J. Guo, Vessel segmentation of X-ray coronary angiographic image sequence, IEEE Trans. Biomed. Eng. (2019).

[20] R. Manniesing, M.A. Viergever, W.J. Niessen, Vessel axis tracking using topology constrained surface evolution, IEEE Trans. Med. Imaging 26 (3) (2007) 309–316.

[21] P. Zou, P. Chan, P. Rockett, A model-based consecutive scanline tracking method for extracting vascular networks from 2-D digital subtraction angiograms, IEEE Trans. Med. Imaging 28 (2) (2008) 241–249.

[22] Y. Chen, Y. Zhang, J. Yang, Q. Cao, G. Yang, J. Chen, H. Shu, L. Luo, J.-L. Coatrieux, Q. Feng, Curve-like structure extraction using minimal path propagation with backtracking, IEEE Trans. Image Process. 25 (2) (2015) 988–1003.

[23] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015, arXiv preprint arXiv:1502.03167.

[24] D. Ulyanov, A. Vedaldi, V. Lempitsky, Instance normalization: The missing ingredient for fast stylization, 2016, arXiv preprint arXiv:1607.08022.

[25] J.L. Ba, J.R. Kiros, G.E. Hinton, Layer normalization, 2016, arXiv preprint arXiv:1607.06450.

[26] Y. Wu, K. He, Group normalization, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 3–19.

[27] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[28] X. Tao, H. Dang, X. Zhou, X. Xu, D. Xiong, A lightweight network for accurate coronary artery segmentation using X-Ray angiograms, Front. Public Health 10 (2022) 892418.

[29] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[30] A. Song, L. Xu, L. Wang, B. Wang, X. Yang, B. Xu, B. Yang, S.E. Greenwald, Automatic coronary artery segmentation of CCTA images with an efficient feature-fusion-and-rectification 3D-UNet, IEEE J. Biomed. Health Inf. 26 (8) (2022) 4044–4055.

[31] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[32] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.

[33] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, A.L. Yuille, Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2017) 834–848.

[34] W. Liu, H. Yang, T. Tian, Z. Cao, X. Pan, W. Xu, Y. Jin, F. Gao, Full-resolution network and dual-threshold iteration for retinal vessel and coronary angiograph segmentation, IEEE J. Biomed. Health Inf. 26 (9) (2022) 4623–4634.

[35] B. Qin, H. Mao, Y. Liu, J. Zhao, Y. Lv, Y. Zhu, S. Ding, X. Chen, Robust PCA unrolling network for super-resolution vessel extraction in X-ray coronary angiography, IEEE Trans. Med. Imaging (2022) 1.

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

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

[38] J. Lourenço-Silva, M.N. Menezes, T. Rodrigues, B. Silva, F.J. Pinto, A.L. Oliveira, Encoder-decoder architectures for clinically relevant coronary artery segmentation, in: M.S. Bansal, I. Măndoiu, M. Moussa, M. Patterson, S. Rajasekaran, P. Skums, A. Zelikovsky (Eds.), Computational Advances in Bio and Medical Sciences, Vol. 13254, Springer International Publishing, Cham, 2022, pp. 63–78.

[39] T.J. Jun, J. Kweon, Y.-H. Kim, D. Kim, T-Net: Nested encoder–decoder architecture for the main vessel segmentation in coronary angiography, Neural Netw. 128 (2020) 216–233.

[40] F. Navarro, S. Shit, I. Ezhov, J. Paetzold, A. Gafita, J.C. Peeken, S.E. Combs, B.H. Menze, Shape-aware complementary-task learning for multi-organ segmentation, in: International Workshop on Machine Learning in Medical Imaging, Springer, 2019, pp. 620–627.

[41] J.C. Montoya, Y. Li, C. Strother, G.-H. Chen, Deep learning angiography (DLA): three-dimensional C-arm cone beam CT angiography generated from deep learning method using a convolutional neural network, in: Medical Imaging 2018: Physics of Medical Imaging, Vol. 10573, International Society for Optics and Photonics, 2018, p. 105731N.

[42] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P.F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, et al., Nnu-net: Self-adapting framework for u-net-based medical image segmentation, 2018, arXiv preprint arXiv:1809.10486.

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

[44] Y. Ma, Y. Hua, H. Deng, T. Song, H. Wang, Z. Xue, H. Cao, R. Ma, H. Guan, Self-supervised vessel segmentation via adversarial learning, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Montreal, QC, Canada, 2021, pp. 7516–7525.

[45] J. Park, J. Kweon, H. Bark, Y.I. Kim, I. Back, J. Chae, J.-H. Roh, D.-Y. Kang, P.H. Lee, J.-M. Ahn, S.-J. Kang, D.-W. Park, S.-W. Lee, C.W. Lee, S.-W. Park, S.-J. Park, Y.-H. Kim, Selective Ensemble Methods for Deep Learning Segmentation of Major Vessels in Invasive Coronary Angiography, Preprint, Radiology and Imaging, 2021.

[46] X. Zhu, Z. Cheng, S. Wang, X. Chen, G. Lu, Coronary angiography image segmentation based on PSPNet, Comput. Methods Programs Biomed. 200 (2021) 105897.

[47] J. Zhang, G. Wang, H. Xie, S. Zhang, N. Huang, S. Zhang, L. Gu, Weakly supervised vessel segmentation in X-ray angiograms by self-paced learning from noisy labels with suggestive annotation, Neurocomputing 417 (2020) 114–127.

[48] J.M. Wolterink, T. Leiner, I. Išgum, Graph convolutional networks for coronary artery segmentation in cardiac CT angiography, 2019, arXiv:1908.05343 [cs, eess]. arXiv:1908.05343.

[49] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, 2014, arXiv preprint arXiv:1409.0473.

[50] P. Anderson, X. He, C. Buehler, D. Teney, M. Johnson, S. Gould, L. Zhang, Bottom-up and top-down attention for image captioning and visual question answering, in: The IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2018.

[51] A. Ambartsoumian, F. Popowich, Self-attention: A better building block for sentiment analysis neural network classifiers, 2018, arXiv preprint arXiv:1812.07860.

[52] T. Xiao, Y. Xu, K. Yang, J. Zhang, Y. Peng, Z. Zhang, The application of two-level attention models in deep convolutional neural network for fine-grained image classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 842–850.

[53] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132–7141.

[54] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, X. Tang, Residual attention network for image classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3156–3164.

[55] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., Attention u-net: Learning where to look for the pancreas, 2018, arXiv preprint arXiv:1804.03999.

[56] R. Caruana, Multitask learning, Mach. Learn. 28 (1) (1997) 41–75.

[57] W. Zhu, Y. Huang, L. Zeng, X. Chen, Y. Liu, Z. Qian, N. Du, W. Fan, X. Xie, AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy, Med. Phys. 46 (2) (2019) 576–589.

[58] F. Lugauer, Y. Zheng, J. Hornegger, B.M. Kelm, Precise lumen segmentation in coronary computed tomography angiography, in: International MICCAI Workshop on Medical Computer Vision, Springer, 2014, pp. 137–147.

[59] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988.