A dual path encoder-decoder network for placental vessel segmentation in fetoscopic surgery

Yunbo Rao[1,5*], Tian Tan[1], Shaoning Zeng[2], Zhanglin Chen[3], Jihong Sun[4]
[1] School of Information and Software Engineering, University of Electronic Science and Technology of China
[e-mail: raoyb@uestc.edu.cn, 202022090721@std.uestc.edu.cn]
[2] Yangtze Delta Region Institute, University of Electronic Science and Technology of China
[e-mail: zeng@csj.uestc.edu.cn]
[3] Chinese Academy of Sciences Shenzhen Institutes of Advanced Technology
[e-mail: zl.cheng@siat.ac.cns]
[4] Sir Run Run Shaw Hospital, Zhejiang University School of Medicine,
[e-mail: sunjihong@zju.edu.cn]
[5]Sichuan Provincial Engineering Research Center of Communication Technology for Intelligent IoT, University of Electronic Science and Technology of China
[e-mail: raoyb@uestc.edu.cn]
[*]Corresponding author: Yunbo Rao

Received November 22, 2022; revised October 21, 2023; accepted December 25, 2023; published January 31, 2024

Abstract

A fetoscope is an optical endoscope, which is often applied in fetoscopic laser photocoagulation to treat twin-to-twin transfusion syndrome. In an operation, the clinician needs to observe the abnormal placental vessels through the endoscope, so as to guide the operation. However, low-quality imaging and narrow field of view of the fetoscope increase the difficulty of the operation. Introducing an accurate placental vessel segmentation of fetoscopic images can assist the fetoscopic laser photocoagulation and help identify the abnormal vessels. This study proposes a method to solve the above problems. A novel encoder-decoder network with a dual-path structure is proposed to segment the placental vessels in fetoscopic images. In particular, we introduce a channel attention mechanism and a continuous convolution structure to obtain multi-scale features with their weights. Moreover, a switching connection is inserted between the corresponding blocks of the two paths to strengthen their relationship. According to the results of a set of blood vessel segmentation experiments conducted on a public fetoscopic image dataset, our method has achieved higher scores than the current mainstream segmentation methods, raising the dice similarity coefficient, intersection over union, and pixel accuracy by 5.80%, 8.39% and 0.62%, respectively.

Keywords: Fetoscopy, Deep Learning, Vessel Segmentation, Twin-to-Twin Transfusion Syndrome

http://doi.org/10.3837/tiis.2024.01.002 **ISSN : 1976-7277**

1. Introduction

During a fetoscopic laser photocoagulation, doctors need to observe the blood vessels through a fetoscope, to obtain information about the target vessels [1, 2]. However, due to the small size of the fetoscope, its field of view is too narrow to have good visibility [3]. Therefore, it is hard to observe the complete vascular tree at one time. What is worse, many small vessels are easily out of view from human eyes. Through image segmentation of fetoscopic images, the interesting objects like blood vessels can be extracted. In this way, one can quickly and accurately obtain the target information. In recent years, computer-assisted intervention technologies based on deep learning has become a research hotspot. In particular, deep learning methods like Fully Convolutional Networks (FCN) [4] and U-net [5] have been widely applied in the field of medical image segmentation.

However, it is still a challenge to realize the automatic segmentation of placental vessels in fetoscopic images due to the problems like poor imaging quality, low contrast, and large changes in vessel shape of the fetoscope. Due to the limitations of equipments and the complex intrauterine environment, various problems occur in the imaging. As depicted in **Fig. 1**, for example, excessive highlights (the first one), shadows (the second one), impurity tissue occlusion (the third one), and unclear vascular edge (the fourth one) are the common issues that prevent an effective automatic segmentation. In order to achieve a more accurate segmentation of blood vessels under the above situations, a variety of improvements had been proposed for the current deep networks, e.g., U-net [5], to achieve a better segmentation effect. In particular, Oktay et al. [6] introduced an attention mechanism on the basis of U-net to have an attention U-net. RU-net and R2U-net [7] replaced the convolutional layers in U-net with recurrent convolutional layers. W-net [8] was formed by bridging two U-nets. Zhou et al. [9] redesigned the skip connection of U-net, to proposed U-net++, and adopted the nested dense skip connection. KiU-net [10] combined Ki-net and U-net through a dual-path structure, so that the networks on both sides can obtain complementary information. An asymmetrical multitask attention U-net [11] also proposed by dividing the segmentation task into two stages. In this design, the latter stage used the hierarchical feature map of the previous stage. BiO-Net [12] made an inference from a recursive manner by reusing modules without introducing additional parameters. Liu Tao [13] designed a network based on graph reasoning, which added a prior knowledge on the shape constraints to improve the segmentation. However, we observed that U-net and all of its current upgraded versions were far from promising in the task of fetoscopic placental vessel segmentation.

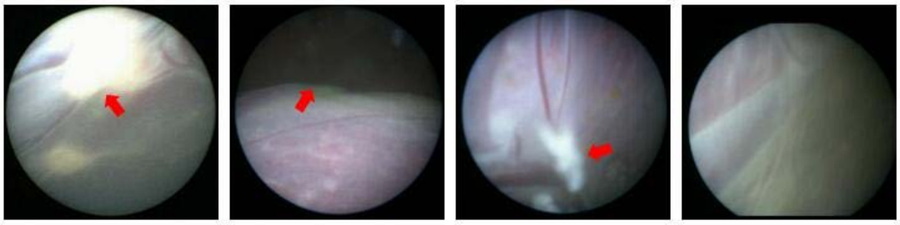

Fig. 1. Fetoscopic images from the Fetoscopy Placenta dataset [14]. The red arrows indicate three typical problems that affect the observation of blood vessels, such as excessive highlighting, shadows, and impurity tissue in the uterus. The last image shows an issue of unclear vascular edge.

In this paper, we design an encoder-decoder network with a dual-path structure to segment placental vessels in fetoscopic images, which is named as Dual-Path Fetoscopic network (DPF-net). In the encoder blocks, we introduce a multi-scale continuous convolution structure to obtain rich feature information. At the same time, each encoder block and decoder block has a channel attention mechanism to emphasize the more relevant channels. What is more, each block has a switching connection with the block corresponding to the other path, to fuse the two paths. In this way, the network is able to capture more helpful information. Our experimental results confirm our expectations as well.

The main contributions of this work are as follows:

(1) A novel dual-path encoder-decoder network is designed to segment the complex placental blood vessels in fetoscopic images to assist doctors to obtain the required information about blood vessels.

(2) The key part of the network consists of pairs of encoder blocks and decoder blocks with a channel attention mechanism and a multi-scale convolution is proposed, which is a new way to capture and extract high-level features.

(3) Our method achieves the best segmentation results on a public Fetoscopic Placenta dataset [14]. The obtained Dice Similarity Coefficient (DSC), Intersection over Union (IoU), and Pixel Accuracy of our network are 0.785, 0.659 and 0.964, respectively.

The subsequent sections of this paper are summarized as follows. In Section 2, we review the related work. Then we introduce our method in Section 3, and describe the structure of DPF-net in detail. Then, Section 4 shows our experimental conditions and results, including the comparison with some other existing methods and ablation experiments. Finally, in Section 5, we summarize the work of this paper.

2. Related work

2.1 U-net and its variants for medical image segmentation

U-net and its variants have been applied in many use cases of medical image segmentation [11, 12, 13]. Aiming at the problem of insufficient performance of U-net when the segmentation object is small and has blurred noisy boundaries, KiU-net [10] combined U-net and an over-complete network to form a dual path network, and connected the features of each layer of the two networks by a cross residual fusion block and forwarded them to another network, so that the networks on both sides can obtain complementary features. However, this method required to upsample the image first, which dramatically increased the parameters of the network, and therefore limited the network to using a deeper number of layers. Currently, many studies on image segmentation of placental related diseases mainly focused on magnetic resonance imaging (MRI) [15, 16, 17, 18] and ultrasound images [19, 20, 21, 22], as well as the segmentation of the placenta. However, two problems prevent these successful methods for MRI and ultrasound images from working for fetoscopic images. First of all, the field of view of fetoscopic images is much smaller than that of MRI and ultrasound images. What is worse, there are great differences in image quality and vessel shape between these two types of images.

For fetoscopic images, Sadda et al. [23] proposed to segment blood vessels in the intraoperative video of fetoscopic laser photocoagulation through a fully convolutional network. The network adopted the same structure of U-net, which is composed of three contraction modules and three expansion modules. In particular, each contraction / expansion module included eight convolutional layers, and the network had a total of 25 layers. However, the accuracy of the network was still relatively low, where the Dice score is only 0.55. Furthermore, this result was far from current mainstream deep learning methods. Bano Sophia et al. [14] used U-net to segment the placental blood vessels of 483 sampled frames from fetoscopic video, using different backbones, including vanilla U-net [5], VGG16 [24], Resnet-

50 and Resnet-101 [25]. This was a two-stage method, where the first stage in the method completed the segmentation of blood vessels, while the second stage completed the registration based on the segmented probability map. The sum of the cross-entropy loss and the Jaccard loss was treated as the loss function of the training network. However, this work mainly focused on registration, not segmentation.

2.2 Deep convolutional neural networks for medical image segmentation

Besides U-net and its variant implementations, there are many other deep convolutional neural networks proposed for the same task. For example, Generative Adversarial Network (GAN) [26, 27, 28] was also popular for medical image segmentation. For the ultrasound images, Torrents-Barrena et al. [29] segmented the placenta using a conditional Generative Adversarial Network (cGAN), and then segmented the placental vessels using a modified spatial kernelized fuzzy C-means algorithm and Markov random field. In its follow-up work [30], a stacked generative adversarial network with a stacking structure was also proposed. The functions of the network included the generation of synthetic ultrasound images, the segmentation of various types of placental tissues including placental vessels and the reconstruction of placental shadows. These studies are meaningful for the treatment of Twin-to-Twin Transfusion Syndrome (TTTS). However, all of them did not solve the aforementioned problems of blood vessel segmentation in fetoscopic images.

3. Proposed Method

For the aforementioned considerations, we propose an encoder-decoder network for blood vessel segmentation in fetoscopic images, as shown in **Fig. 2**. Firstly, images are inputted to the encoder network, before a feature map with rich semantic information is extracted through all the encoder blocks. Next, the high-dimensional feature map is sent to the decoder of the network. At the same time, the feature information of different depths in the encoder is transmitted to the decoder through a skip connection. Then, the final segmented image is obtained. In order to improve the extraction of image information, inspired by the work of KiU-net [10], the network path is multiplied to become a dual-path structure. In this way, every path in the network owns a pair of encoder and decoder, which are connected to the corresponding position of the other path through the switching connection, where we implement the information fusion that is crucial to the final segmentation. Besides this, we also have a specific design of the structure for all encoder and decoder blocks. We will explain them in detail in the following subsections.

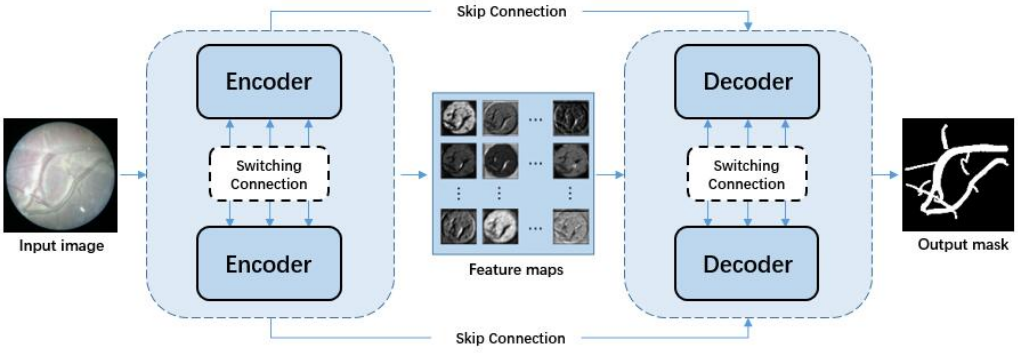

Fig. 2. A overview of the proposed DPF-net method.

3.1 The Encoder and Decoder Blocks

The structure of our encoder and decoder blocks is shown in **Fig. 3**. After inputted to the network, the image first enters the encoder blocks and a feature map of this image is obtained. In each encoder block, two convolutions are performed, using a kernel with a size of 3 × 3, and then a ReLU layer follows. Between the convolution and ReLu layers, we introduce batch normalization. This operation plays an important role in accelerating the convergence of network training and therefore stabilizes the model training.

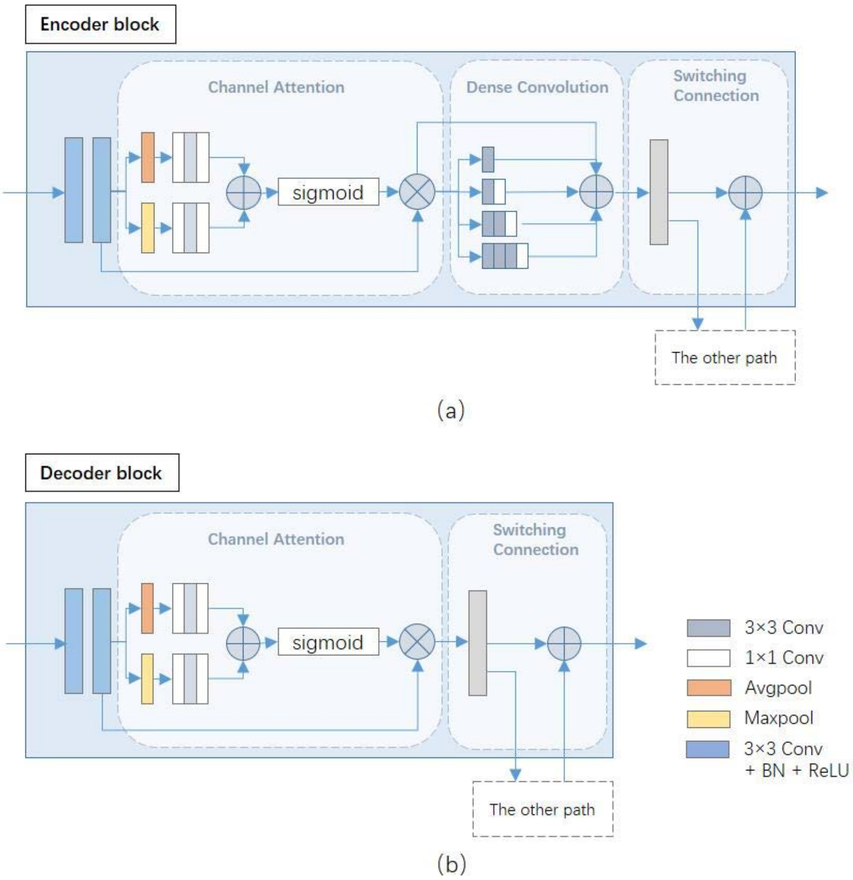

Fig. 3. The structure of (a) the encoder block and (b) the decoder block of DPF-net.

Due to the narrow field of view of the fetoscope, only a local part of the vascular tree can be observed, resulting in significant changes in the size of the placental vessels in the fetoscopic images. To solve this problem, inspired by Inception-ResNet [31] and CE-Net [32], we introduce a dense continuous convolution structure into the encoder block to obtain multi-scale features. Through a continuous 3 × 3 convolution of different numbers, this structure enables the encoder block to have multiple size receptive fields. By adding this structure to each encoder block, the network can deeply mine the multi-scale information in the image. At the same time, the shortcut connection of the residual network is also introduced to avoid the gradient vanishing problem. It is noted that, different from CE-Net, we do not use the dilated convolution. Instead, a continuous ordinary 3 × 3 convolution is used. According to our experimental observations, this ordinary convolution is more suitable than the dilated convolution in blood vessel segmentation of fetoscopic images. We believe that it is the aforementioned issue that many small structures are contained in the blood vessels produces a negative influence. Besides this, too large receptive fields are unnecessary for this segmentation task.

Each channel of the feature map represents a detector for different information of the image. In order to emphasize the most relevant channels, we introduce a channel attention mechanism, inspired by convolutional block attention module [33], between the double convolution and the dense convolution. For an input feature map $F \in \mathbb{R}^{C \times H \times W}$, two context descriptors with a size of $C \times 1 \times 1$ are obtained through a global average pooling and a global max pooling, respectively. Then, the channel attention map $M_C \in \mathbb{R}^{C \times 1 \times 1}$ is obtained through the shared two-layer neural network and the sigmoid layer. Finally, M_C and F are merged by an element-wise summation to obtain the output:

$$F' = F * M_C. \tag{1}$$

The reason why we choose this channel attention mechanism, instead of the spatial attention mechanism, is because the position of the blood vessel as the detection target in the fetoscopic image is not fixed and is likely to appear in any position in the visual area. The spatial attention mechanism mainly helps focus on specific areas. However, this is not helpful for our segmentation task. The experiment results also show that adding a spatial attention mechanism generates no improvement in the segmentation.

The extracted feature maps are then inputted to the decoder blocks after four encoder blocks and three corresponding downsampling layers. We adopt a similar structure in both the encoder and decoder blocks, except that the decoder block does not contain the dense continuous convolution structure. It has been proven that skip connections in U-net can improve segmentation performance by fusing multi-level features. Therefore, we introduce the same skip connections as U-net in our DPF-net. Starting from the second decoder block, the input of each block becomes the result of concatenation of the output of the previous block and the corresponding feature map brought by the skip connection. Through skip connections, the low-level features that are learned by the encoder blocks are combined with the high-level feature information from the decoder blocks. The concatenating operation can better preserve the information in the feature maps, allowing the network to make local predictions that respect the global structure.

3.2 The Switching Connection

In order to better link the two paths to fuse their respective information, we introduce a switching connection between the corresponding encoder block and decoder block in each path. For the kth encoder module $(k > 1)$, the input is:

$$I_{ek} = O_{e(k-1)} + O'_{e(k-1)}, \tag{2}$$

where $O_{e(k-1)}$ is the output of the previous encoder block, $O'_{e(k-1)}$ is the output of the corresponding encoder block of another path. For the kth decoder module $(k > 1)$, the input is:

$$I_{dk} = (O_{d(k-1)} + O'_{d(k-1)}) \oplus O_k^p, \tag{3}$$

where O_k^p is the corresponding feature map from the skip connection. $\oplus$ represents the concatenating operation. The switching connection can strengthen the relation between the two paths, so that each module of the network can obtain more information and improve the segmentation effect. We performed ablation experiments to confirm the role of this structure.

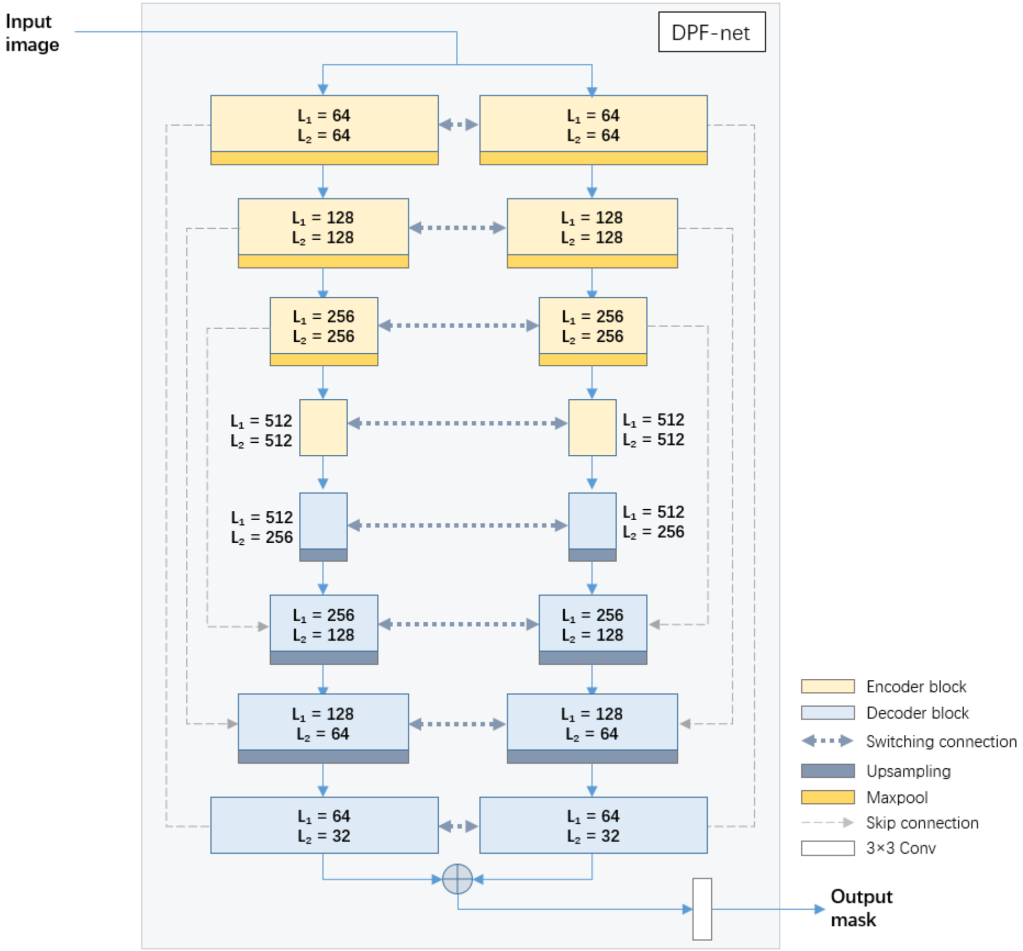

Fig. 4. The structure of DPF-net. L_1 and L_2 at the encoder blocks and decoder blocks in the figure represent the number of channels of the first and second convolution layers in the block respectively.

3.3 The Dual-Path Fetoscopic network

In this study, we have the full network, i.e., the Dual-Path Fetoscopic network. The detailed structure of our network is shown in **Fig. 4**. We utilize 4 pairs of encoder and decoder blocks on each path, and used Maxpool layers for downsampling in the first three encoder blocks. Three corresponding upsampling layers are used in the decoding part to restore the image size. Unlike the way in [10], in which one path used upsampling layers in the decoder, we employ downsampling in the decoder on both paths and form a symmetrical structure, which can avoid the difficulty of training due to too many parameters, thus allowing us to increase the number of network layers. In the figure, L_1 is the number of channels of the first convolution layer of the encoder / decoder blocks, and L_2 is that of the second convolution layer. At the last encoder block, the network obtains a high-dimensional feature map with channels of 512. At the same time, the feature maps with different dimensions obtained by each encoder block are spliced and fused in the decoder through the skip connection. At the end of the network, the outputs of the two paths are added to get the final segmentation image.

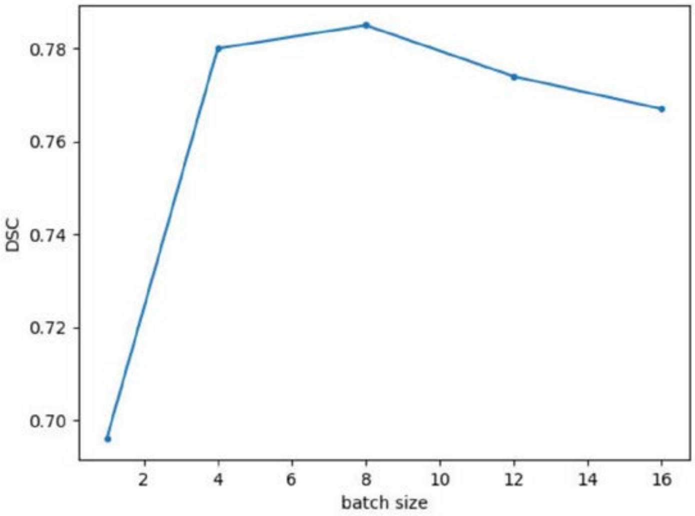

Fig. 5. DSC of DPF-net using different values of batch size.

4. Experimental results and analysis

4.1 Dataset

We conducted experiments on the Fetoscopy Placenta dataset [14]. The images were captured from a set of vivo videos of 6 TTTS laser ablation cases, including a total of 483 images. The size of each image is 448 × 448, and there is a corresponding binary image of blood vessel annotation. The numbers of images from all cases are 121, 101, 39, 88, 37, and 97, respectively.

In order to shorten the training time, the size of images and the corresponding mask images are adjusted to 128 × 128 by using bilinear interpolation, thereby reducing the computational burden of the model training. Although the image size is reduced, the experiment shows that the network trained with this size of images can still show good performance. At the same time, as all networks are trained under the same conditions, resizing the images will not significantly affect the comparison of different models, and can greatly reduce the training time. Therefore, we uniformly use the reduced images for subsequent experiments. We expand the number of images three times through random horizontal or vertical flip, random rotation, and random distort. 1/6 of the dataset is randomly divided to test the performance of networks.

4.2 Implementation Details

The implementation is based on PyTorch. The versions of all implementation techniques are PyTorch 1.11.0, CUDA 11.6, CuDNN 8.4.0 and Python 3.9.1. The experiments were run on RTX A6000 GPU. Our network is trained using a combined loss function of cross-entropy and Jaccard[14]. The loss function is as follows:

$$L = -\frac{\sum[ylog\hat{y}+(1-y)log1-\hat{y}]}{N} + [1 - \frac{\sum(y\cdot\hat{y})+\epsilon}{\sum(y+\hat{y})-\sum(y\cdot\hat{y})+\epsilon}], \tag{4}$$

where y and $\hat{y}$ are prediction tensor and the ground-truth tensor, respectively. Adam is set as the optimizer, and the learning rate is set to 0.0001. According to our preliminary experiments, most of the benchmark networks get better training results when setting the batch size to 8. Therefore, we set the batch size to 8 in the experiments. As shown in **Fig. 5**, we have the best DSC of DPF-net when training the network with batch size set to 8. Each network was trained for 100 epochs and ensured it converged. In order to evaluate the quality of network segmentation, we use three indicators, namely DSC, IoU and Pixel Accuracy, which are commonly used in segmentation tasks [14].

4.3 Experimental Results

The benchmark results of these metrics obtained on the Fetoscopy Placenta dataset are listed in **Table 1**. We compare our method with U-net [5] and its variant implementations, including Attention U-net [6], KiU-net [10] and CE-net [32]. We can see that our DPF-net produces a DSC of 0.785, which is higher than the current best result of U-Net (0.742). The improvement rate is around 5.80%. Comparing to the result of KiU-net (0.632), the improvement is up to 24.20%. This confirms our observation that U-net and its variants were far from being suitable in the task of fetoscopic placental vessel segmentation, despite U-net was working well in other tasks of medical image segmentation. What is more, our network gains the highest scores in IoU (0.659) and Pixel accuracy (0.964) as well. The improvement rates are 8.29% of IoU and 0.62% of Pixel Accuracy. All of these benchmarking results demonstrate the impressing performance of our DPF-net.

Table 1. Comparison of different deep learning methods.

Network	DSC	IoU	pixel accuracy
U-net	0.742	0.608	0.958
KiU-net	0.632	0.489	0.943
Attention U-net	0.741	0.605	0.958
CE-net	0.709	0.569	0.955
DPF-net	**0.785**	**0.659**	**0.964**

Besides these quantitative results, we also visualize the segmentation results of different methods, as shown in **Fig. 6**. In particular, the first two columns are the inputted fetoscopic images and their corresponding placental vessel annotation. Columns from third to seventh are the segmentation results by our DPF-net and all other benchmark methods, including U-net, Attention U-net, KiU-net and CE-net. It can be seen that although placental blood vessels can be recognized under various lighting conditions, there are still some challenges to achieving more accurate segmentation. For example, in the fifth row of **Fig. 6**, the boundary of the blood vessel in the original image is not obvious, which increases the difficulty of segmentation. Although our network performs better than other methods, there are still some obvious deficiencies. The edges of the vessels in the first and second rows are also missing. Although the segmentation results retain the basic shape of the vessels, some information is still lost. Obviously, our DPF-net produces much more promising segmentation results than current U-net implementations.

Table 2. Experimental results of our backbone network with different components.

Components	DSC	IoU	pixel accuracy
Backbone	0.754	0.619	0.959
Backbone + SC	0.756	0.623	0.960
Backbone + CA	0.755	0.621	0.960
Backbone + DC	0.766	0.636	0.962
Backbone + CA + SC	0.762	0.631	0.961
Backbone + CA + DC	0.774	0.646	0.963
Backbone + DC + SC	0.770	0.637	0.961
Backbone + DC + SC + CA	**0.785**	**0.659**	**0.964**

4.4 Ablation Experiment

In order to verify the effectiveness of the key components in our network, we also conducted a set of ablation experiments under the same conditions. We study and compare the usage and combination of all components in the backbone, including the switching connection (SC), the channel attention (CA) and the dense convolution (DC). The results are listed in **Table 2**.

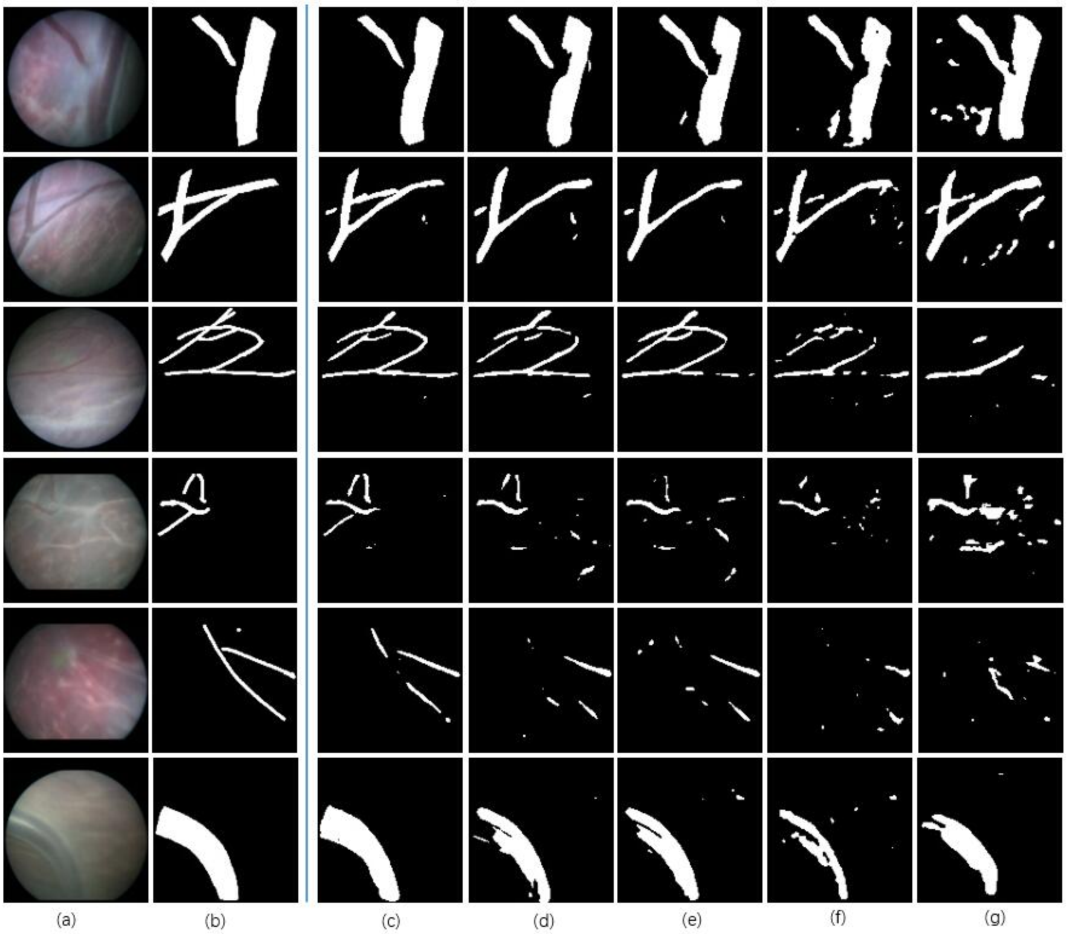

Fig. 6. Segmentation results of placental vessels using different methods, where columns from left to right are (a) the original fetoscopic images, (b) the vessel annotations, and the results by (c) DPF-net (ours), (d) U-net, (e) Attention U-net, (f) KiU-net, and (g) CE-net.

The experimental results show that as a single component, the dense continuous convolution structure can greatly improve the network performance. When the switching connection and/or the channel attention block are set alone, they do not evidently improve the network performance. However, when they are combined and used together, the network greatly benefits. When adding all the components at the same time, the network achieves the best segmentation. All of these demonstrate the effectiveness of our method.

4.5 Discussion

As can be seen from **Table 1**, our network achieves the highest dice score. Although Attention U-net introduces the attention mechanism on the basis of U-net, it does not significantly improve the segmentation effect. KiU-net performs poorly in this task because it has fewer network layers and thus lacks the ability to learn the features in fetoscopic images. At the same time, the method with upsampling first limits the number of layers, because increasing the number of layers while upsampling will increase the difficulty of the network training. In our experiments, CE-net also fails to achieve good results. We believe that it is due to the dilated convolution, which is likely to be not suitable for this task. The continuous dilated convolution increases the receptive field. However, it also leads to some information loss from adjacent pixels. Considering that there is some tiny structure in placental vessels to be identified, the dilated convolution is therefore not helpful for this segmentation task. We replace the continuous convolution with the structure of DAC [32] in our implementation, the dilated convolutions replace the convolutions in this part and the relevant parameters of the dilated convolutions are the same as those of DAC. The rest of the network remains unchanged. The results are shown in **Table 3**. Besides this, we also compare the results using spatial attention and channel attention by replacing the attention module and leaving the rest of the network unchanged. The results are listed together in **Table 3**. Among them, the spatial and channel attention scheme adopts the same structure as CBAM [33]. It turns out that the channel attention (0.785 of DSC) is superior to the spatial attention (0.771) in this task.

Table 3. Accuracy comparison of structure of encoder and decoder blocks.

Components	Structure	DSC
Attention mechanism	No attention mechanism	0.770
	Spatial attention	0.771
	Channel attention	**0.785**
	Spatial and channel attention	0.772
Dense convolution block	Dense convolution block	0.774
	Ordinary convolution	**0.785**

The existence of highlights on blood vessels is also a negative influencing factor for effective segmentation. **Fig. 7** shows two sets of examples where the highlights negatively affect the segmentation. Sometimes there are tubular highlights on the blood vessels, and the color of these highlights is close to the background, which leads to a wrong recognition of the vessel highlights from the background, and therefore results in a wrong segmentation of one vessel into two vessels (when the highlight is in the middle of the vessel), or losing the edge of the segmented vessel. The red arrow and circles in **Fig. 7** indicate the missing parts due to highlights. The same issues can be seen in the first and last rows of **Fig. 6**. In the last row, although our network performs well, other networks have been affected to some extent. In the first line, all networks are missing to varying degrees. Subsequent research can aim at this problem, reducing the impact of highlight on segmentation through data enhancement or post-processing methods, so as to further improve the segmentation effect.

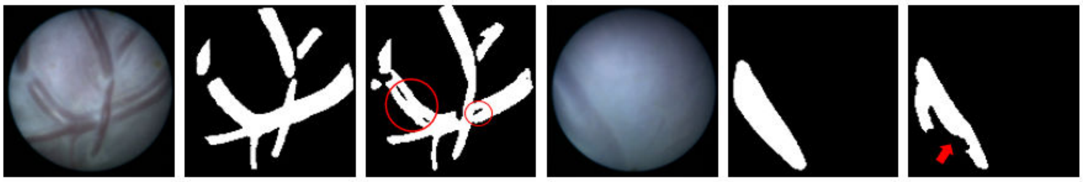

Fig. 7. Examples of vascular highlights negatively affecting the segmentation. Red circles and arrow indicate the partial loss of the segmented blood vessels caused by highlights.

Table 4. Comparison of U-net and DPF-net in terms of parameters, floating point operations, and memory usage.

Network	Parameters	GFlops	memory usage
U-net	**31.04M**	**11.53**	**101.00MB**
DPF-net	39.95M	32.68	217.61MB

In our work, the proposed DPF-net achieves the highest segmentation accuracy, the complex structure of the encoder-decoder blocks also brings computational overhead. As shown in Table 4, the parameters, floating point operations, and memory usage of DPF-net have increased.

5. Conclusion

In this paper, we propose a placental vessel segmentation network for fetoscopic images. The network has two paths and an encoder-decoder structure. We designed its encoder blocks and decoder blocks and introduced channel attention mechanism, dense continuous convolution structure and switching connection. Ablation experiments show that these structures can improve the segmentation performance of the network. After experimental verification on the public dataset, our method can be used in the task of fetoscopic image vascular segmentation. Compared with the existing mainstream methods, our method has achieved the highest scores. The framework can also be extended to other related medical image segmentation tasks.

Currently, there are only a few available fetoscopic image datasets labeled with blood vessels. Therefore, we only use one dataset provided by [14] in our experiment. At the same time, there may be a small number of annotation errors in the dataset, which will also affect the training of the network. In the future, we will continue to evaluate our method using more data.

Acknowledgement

This research was supported by the Science and Technology Project of Sichuan (Grant Nos. 2022ZHCG0033, 2023ZHCG0005, 2023ZHCG0008), States Key Laboratory of Air Traffic Management System and Technology (Grant No: SKLATM202202), the Chengdu Science and Technology Project (Grant No: 2022-YF05-00068-SN), and Innovation Ability Construction Project for Sichuan Provincial Engineering Research Center of Communication Technology for Intelligent IoT (2303-510109-04-03-318020).

References

[1] J. Sommer, A. M. Nuyt, F. Audibert, et al., “Outcomes in extremely premature infants with twin–twin transfusion syndrome treated by laser therapy,” *Paediatrics & Child Health*, vol. 23, no. suppl_1, pp. e20–e21, 2018. Article (CrossRef Link)

[2] K. Hecher, H. M. Gardiner, A. Diemert, et al., “Long-term outcomes for monochorionic twins after laser therapy in twin-to-twin transfusion syndrome,” *The Lancet Child & Adolescent Health*, vol. 2, no.7, pp. 525-535, 2018. Article (CrossRef Link)

[3] L. Peter, M. Tella-Amo, D. I. Shakir, et al., “Retrieval and registration of long-range overlapping frames for scalable mosaicking of in vivo fetoscopy,” *International journal of computer assisted radiology and surgery*, vol. 13, no.5, pp.713-720, 2018. Article (CrossRef Link)

[4] J. Long, E. Shelhamer, T. Darrell, “Fully convolutional networks for semantic segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640-651, 2017. Article (CrossRef Link)

[5] O. Ronneberger, P. Fischer, T. Brox, “U-net: Convolutional networks for biomedical image segmentation," in *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234-241, 2015. Article (CrossRef Link)

[6] O. Oktay, J. Schlemper, L. L. Folgoc, et al., “Attention u-net: Learning where to look for the pancreas,” *Eprint arXiv:1804.03999*, 2018. Article (CrossRef Link)

[7] M. Z. Alom, M. Hasan, C. Yakopcic, et al., “Recurrent residual convolutional neural network based on u-net (r2u-net) for medical image segmentation,” *Eprint arXiv:1802.06955*, 2018. Article (CrossRef Link)

[8] W. Chen, Y. Zhang, J. He, et al., “Prostate Segmentation using 2D Bridged U-net,” *Eprint arXiv:1807.04459*, 2018. Article (CrossRef Link)

[9] Z. Zhou, S. M. M. Rahman, N. Tajbakhsh, et al., “Unet++: A nested u-net architecture for medical image segmentation,” in *Proc. of International Workshop on Multimodal Learning for Clinical Decision Support*, pp.3-11, 2018. Article (CrossRef Link)

[10] J. M. J. Valanarasu, V. A. Sindagi, I. Hacihaliloglu, et al., “Kiu-net: Towards accurate segmentation of biomedical images using over-complete representations,” in *Proc. of International conference on medical image computing and computer-assisted intervention*, pp. 363-373, 2020. Article (CrossRef Link)

[11] X. Xu, C. Lian, S. Wang, et al., “Asymmetrical multi-task attention U-Net for the segmentation of prostate bed in CT image,” in *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp.470-479, 2020. Article (CrossRef Link)

[12] T. Xiang, C. Zhang, D. Liu, et al., “BiO-Net: learning recurrent bi-directional connections for encoder-decoder architecture,” in *Proc. of International conference on medical image computing and computer-assisted intervention*, pp.74-84, 2020. Article (CrossRef Link)

[13] Z. Liu, H. Wang, S. Zhang, et al., “Nas-scam: Neural architecture search-based spatial and channel joint attention module for nuclei semantic segmentation and classification,” in *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp.263-272, 2020. Article (CrossRef Link)

[14] S. Bano, F. Vasconcelos, L. M. Shepherd, et al., “Deep placental vessel segmentation for fetoscopic mosaicking,” in *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp.763-773, 2020. Article (CrossRef Link)

[15] M. Han, Y. Bao, Z. Sun, et al., “Automatic segmentation of human placenta images with U-Net,” *IEEE Access*, vol. 7, pp.180083-180092, 2019. Article (CrossRef Link)

[16] T. Wu, X. Sun, J. Liu, “Segmentation of uterine area in patients with preclinical placenta previa based on deep learning,” in *Proc. of 2019 6th International conference on information science and control engineering (ICISCE)*, pp.541-544, 2019. Article (CrossRef Link)

[17] G. Wang, M. A. Zuluaga, W. Li, et al., “DeepIGeoS: a deep interactive geodesic framework for medical image segmentation,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no.7, pp.1559-1572, 2019. Article (CrossRef Link)

[18] M. Shahedi, J. D. Dormer, A. D. TT, et al., "Segmentation of uterus and placenta in MR images using a fully convolutional neural network," *Medical Imaging 2020: Computer-Aided Diagnosis*, vol.11314, pp.411-418, 2020. Article (CrossRef Link)

[19] P. Looney, G. N. Stevenson, K. H. Nicolaides, et al., "Fully automated, real-time 3D ultrasound segmentation to estimate first trimester placental volume using deep learning," *JCI insight*, vol. 3, no.11, 2018. Article (CrossRef Link)

[20] X. Yang, L. Yu, S. Li, et al., "Towards automated semantic segmentation in prenatal volumetric ultrasound," *IEEE transactions on medical imaging*, vol. 38, no.1, pp.180-193, 2019. Article (CrossRef Link)

[21] V. A. Zimmer, A. Gomez, E. Skelton, et al., "Towards whole placenta segmentation at late gestation using multi-view ultrasound images," in *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp.628-636, 2019. Article (CrossRef Link)

[22] P. Looney, Y. Yin, S. L. Collins, et al., "Fully Automated 3-D Ultrasound Segmentation of the Placenta, Amniotic Fluid, and Fetus for Early Pregnancy Assessment," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol.68, no.6, pp.2038-2047, 2021. Article (CrossRef Link)

[23] P. Sadda, M. Imamoglu, M. Dombrowski, et al., "Deep-learned placental vessel segmentation for intraoperative video enhancement in fetoscopic surgery," *International journal of computer assisted radiology and surgery*, vol.14, no.2, pp.227-235, 2019. Article (CrossRef Link)

[24] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Eprint arXiv:1409.1556*, 2014. Article (CrossRef Link)

[25] K. He, X. Zhang, S. Ren, et al., "Deep residual learning for image recognition," in *Proc. of the IEEE conference on computer vision and pattern recognition*, pp.770-778, 2016. Article (CrossRef Link)

[26] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., "Generative adversarial networks," *Communications of the ACM*, vol.63, no.11, pp.139-144, 2020. Article (CrossRef Link)

[27] H. Gong, J. Liu, B. Chen, et al., "ResAttenGAN: Simultaneous segmentation of multiple spinal structures on axial lumbar MRI image using residual attention and adversarial learning," *Artificial Intelligence in Medicine*, vol.124, p.102243, 2022. Article (CrossRef Link)

[28] K. S. Kumar, N. Suganthi, S. Muppidi, et al., "FSPBO-DQN: SeGAN based segmentation and Fractional Student Psychology Optimization enabled Deep Q Network for skin cancer detection in IoT applications," *Artificial Intelligence in Medicine*, vol.129, p.102299, 2022. Article (CrossRef Link)

[29] J. Torrents-Barrena, G. Piella, N. Masoller, et al., "Automatic segmentation of the placenta and its peripheral vasculature in volumetric ultrasound for TTTS fetal surgery," in *Proc. of 2019 IEEE 16th International Symposium on Biomedical Imaging*, pp.772-775, 2019. Article (CrossRef Link)

[30] J. Torrents-Barrena, G. Piella, B. Valenzuela-Alcaraz, et al., "TTTS-STgan: stacked generative adversarial networks for TTTS fetal surgery planning based on 3D ultrasound," *IEEE Transactions on Medical Imaging*, vol.39, no.11, pp.3595-3606, 2020. Article (CrossRef Link)

[31] C. Szegedy, S. Ioffe, V. Vanhoucke, et al., "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. of Thirty-first AAAI conference on artificial intelligence*, 2016. Article (CrossRef Link)

[32] Z. Gu, J. Cheng, H. Fu, et al., "Ce-net: Context encoder network for 2d medical image segmentation," *IEEE transactions on medical imaging*, vol.38, no.10, pp.2281-2292, 2019. Article (CrossRef Link)

[33] S. Woo, J. Park, J. Y. Lee, et al., "Cbam: Convolutional block attention module," in *Proc. of the European conference on computer vision*, pp.3-19, 2018. Article (CrossRef Link)

Yunbo Rao (M'16) received the B.S. degree from Sichuan Normal University, in 2003, and the M.E. and Ph.D. degrees from the School of Computer Science and Engineering, University of Electronic Science and Technology of China, in 2006 and 2012, respectively. From 2009 to 2011, he was a Visiting Scholar of Electrical Engineering with the University of Washington, Seattle, USA. Since 2012, he has been with the School of Information and Software Engineering, University of Electronic Science and Technology of China, where he is currently an Associate Professor. His research interests include image segmentation, 3-D reconstruction, video enhancement, and medical image processing.

Tian Tan studied at the University of Electronic Science and Technology of China and received a bachelor's degree from Northeastern University. His research interests include image segmentation, image enhancement and medical image processing.

Shaoning Zeng (M'18) received his B.S. and M.S. from Beihang University, Beijing, China, in 2004 and 2007, respectively. He received the Ph.D. degree in Computer Science from the University of Macau in 2020. Currently, he is an Associate Professor in the Yangtze Delta Region Institute (Huzhou) of University of Electronic Science and Technology of China. His research interests include computer vision, multimedia analysis and deep learning. Most of his research work has been open-sourced in https://github.com/zengsn/research.

Zhanglin Cheng received the Ph.D degree from the Institude of Automation, Chinese Academy of Sciences, in 2008. He is currently an Associate Professor with the Shenzhen Key Laboratory for Visual Computing and Analytics, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences. His research interests and experirnce include wide range of topics including computer graphics, computer vision, and visualization.

Jihong Sun received the M.D. degree in clinical medicine from Zhejiang University School of Medicine, Hangzhou, China, in 2009. He is currently an Attending Radiologist and Chief Physician in Department of Radiology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, China. He has authored or co-authored numerous medical and nanomaterial articles in well-known international journals such as Radiology, Nature communications and Advanced Materials. His current research interests include molecular imaging, artificial intelligence and radiomics.