CAU-net: A Novel Convolutional Neural Network for Coronary Artery Segmentation in Digital Substraction Angiography

Rui-Qi Li[1,2], Gui-Bin Bian[1,2], Xiao-Hu Zhou[1,2], Xiaoliang Xie[1,2], Zhen-Liang Ni[1,2], and Zengguang Hou[1,2,3,4](✉)

[1] State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
zengguang.hou@ia.ac.cn
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] CAS Center for Excellence in Brain Science and Intelligence Technology, Beijing 100190, China
[4] Joint Laboratory of Intelligence Science and Technology, Institute of Systems Engineering, Macau University of Science and Technology, Macau, China

Abstract. Coronary artery analysis plays an important role in the diagnosis and treatment of coronary heart disease. Coronary artery segmentation, as an important part of quantitative researc h on coronary heart disease, has become the main topic in coronary artery analysis. In this paper, a deep convolutional neural network (CNN) based method called Coronary Artery U-net (CAU-net) is proposed for the automatic segmentation of coronary arteries in digital subtraction angiography (DSA) images. CAU-net is a variant of U-net. Based on the observation that coronary arteries are composed of many vessels with the same appearance but different thicknesses, a novel multi-scale feature fusion method is proposed in CAU-net. Besides, a new dataset is proposed to solve the problem of no available public dataset on coronary arteries segmentation, which is also one of our contributions. Our dataset contains 538 image samples, which is relatively large compared with the public datasets of other vessel segmentation tasks. In our dataset, a new labeling method is applied to ensure the purity of the labeling samples. From the experimental results, we prove that CAU-net can make significant improvements compared with the vanilla U-net, and achieve the state-of-the-art performance compared with other traditional segmentation methods and deep learning methods.

Keywords: Machine learning · Coronary artery · Segmentation

1 Introduction

Coronary heart disease (CHD) has been one of the leading causes of death worldwide. As the standard modality to diagnose CHD, Digital Substraction

H. Yang et al. (Eds.): ICONIP 2020, LNCS 12532, pp. 185–196, 2020.
https://doi.org/10.1007/978-3-030-63830-6_16

Angiography (DSA) examination has been widely used and considered as the "gold standard". By observing the coronary arteries in the DSA image, the doctor can determine whether there is stenosis in the coronary arteries and get some useful information about the stenosis. Besides, DSA examination has been implemented at all stages of percutaneous coronary intervention (PCI), a kind of minimally invasive treatment and also the primary treatment for CHD. During PCI, by observing and analyzing the coronary arteries in DSA images, surgeons can: 1. determine the surgical plan and select the appropriate surgical instruments before the intervention; 2. get navigation information for delivering surgical instruments during the intervention, 3. check the surgical results after the intervention. However, all of the above useful information can be obtained from the segmentation results of coronary arteries. In conclusion, as a key component of the quantitative study of CHD, coronary artery segmentation in DSA images plays a crucial role in clinical diagnosis and surgical planning.

Recent years have witnessed the rapid development of deep convolutional neural networks (CNN) for medical segmentation. Especially after the appearance of U-net [1], a large number of its variants were proposed for various segmentation tasks and achieved the state-of-the-art results on different medical datasets. As a supervised method, deep learning algorithms need sufficient training samples with annotations to obtain good performance. For the segmentation task, the training image needs to be manually labeled by category at pixel-level.

However, the pixel-wise annotation of blood vessels is time-consuming, because blood vessels have complex shapes and various edge details. This is the reason why there is still no public dataset for coronary artery segmentation, and also the reason why there are so few deep learning methods for coronary artery segmentation. The same difficulty also occurred in other vessel segmentation tasks: the main public datasets of the retinal vessel segmentation task basically contain fewer than 100 samples. In order to solve the lack of dataset, a dataset containing 538 samples is established in this paper. More details on the dataset will be introduced in the Sect. 4.

Except for the dataset, there are also some difficulties that need to be solved in the task of coronary artery segmentation in DSA images. First, the signal-to-noise ratio of DSA images is low, which is a great challenge for accurate segmentation. Because noise blurs the boundaries of coronary arteries. Second, DSA images belong to grayscale images, so the low diversity of color makes many things in DSA images have similar appearances with vessels and are easily misclassified into coronary arteries, such as the outline of the spine and ribs. Third, there are also many motion artifacts caused by heart beating and respiration. These problems make accurate segmentation of the coronary artery a challenging problem. To solve these problems, the algorithm should have more denoising and recognition capabilities, rather than the ability to extract small vessels.

Our contributions are as follows: (1) We set up a new dataset for the usage of coronary artery segmentation. The dataset will be introduced in Sect. 4. (2) CAU-net, a newly designed CNN modified on U-net, is proposed. In CAU-net, a novel multi-scale feature fusion method is proposed. (3) We apply both traditional segmentation methods and deep learning methods on our dataset for comparison. From the comparison results, it can be seen that our method achieves the state-of-the-art performance.

2 Related Work

Up to now, the existing methods for coronary artery segmentation in DSA images are mainly the traditional segmentation methods. These methods have similar workflows. First vessel enhancement filters [2,3] based on Hessian measures are used to analyze the local second-order profile. These morphological filters can provide vesselness features. Then thresholding or other post-processing methods (level set [4], region growing [5], active contour [6]) are utilized to obtain the final segmentation result. These traditional segmentation approaches have two common problems. First, the generalization performance of these methods is poor. This is because these methods have many hyperparameters that need to be selected, and the selection of these parameters is entirely empirical. Besides, for these methods, performance varies greatly under different hyperparameters. Second, these methods only extract appearance features and cannot extract the semantic features, so any structure with a similar appearance with blood vessels is easy to be classified as coronary arteries. Because of the lack of available public datasets for training, few deep learning methods have been used for coronary artery segmentation. Fan *et al.* [7] and Yang *et al.* [8] both propose a deep learning method to achieve the segmentation of coronary arteries. But different from our task, their inputs not only include the image to be segmented but also include a background image without the contrast agent (they call it 'mask image').

Although there are few supervised learning methods for coronary artery segmentation, many supervised learning and even deep learning methods have emerged in other blood vessel segmentation tasks with public datasets. The most common task is retinal vessel segmentation in the fundus images. For retinal vessel segmentation, researchers have established some public datasets, including DRIVE [9] and STARE [10]. So a number of supervised methods have emerged, including traditional supervised learning methods (k-nearest neighbors [9], support vector machine (SVM) [11], conditional random fields (CRFs) [12]), and many deep learning methods [13,14].

3 Method

U-net is commonly used for medical image segmentation tasks because of their good performance and efficient use of GPU memory. Compared with other segmentation methods, U-net has a small number of parameters (the vanilla U-net only has 31M parameters), which make U-net is very suitable for the training of medical image datasets which only contain a small number of samples. In our network, we further reduce the channel numbers of all convolution layers to a

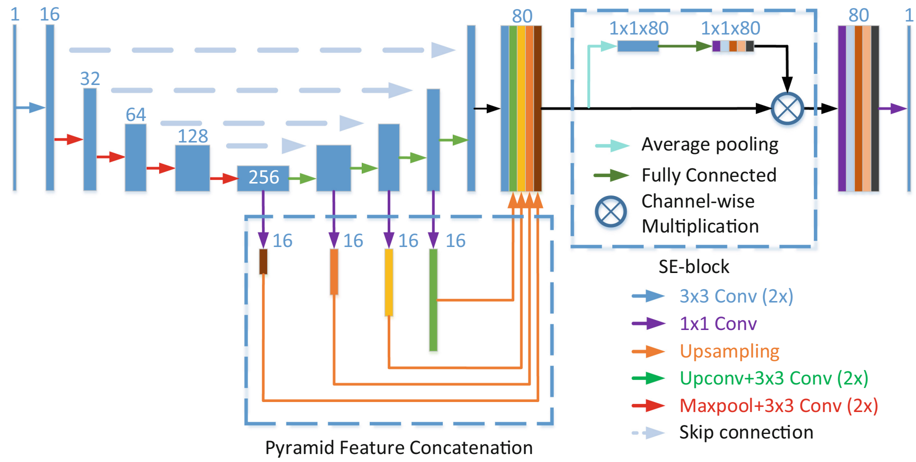

Fig. 1. Architecture of CAU-net

quarter of the former number to reduce the inference time and the risk of over-fitting, and the parameter number of U-net is reduced to about 2M. In order to make U-net more suitable for the task of coronary artery segmentation, we propose a novel feature fusion module on it and named it Coronary Artery U-net (CAU-net), as shown in Fig. 1.

3.1 Feature Fusion Module

Coronary arteries are composed of many vessels with the same appearance but different thicknesses. These vessels with different thickness will have different activations in the feature maps at different scales. Based on this observation, we believe that the idea of multi-scale feature fusion is helpful to improve the coronary artery segmentation results.

The most famous multi-scale feature fusion method is FPN (Feature Pyramid Networks) [15], as shown in Fig. 2(a), which has been widely used in many detection models. In fact, from the perspective of structure, U-net's decoder structure itself is very similar to the structure of the FPN, as shown in Fig. 2(b). The main difference between them is that the feature maps of different scales in FPN are used to generate the detection results, while U-net only uses the last feature maps to predict the final segmentation results. As shown in Fig. 2(c), we also hope to propose a network that can use the feature maps of each scale to predict a segmentation result, and then fuse all segmentation results into one, just like what FPN did. But the segmentation results at different scales can not be fused like the way of detection task. Therefore, in implementation, we choose to fuse all feature maps at different scales first, and then generate the final segmentation results, as shown in Fig. 2(d).

As shown in Fig. 1, we first unify all the feature maps of different scales to the same size and concatenate them, which we called PFC (Pyramid Feature

Concatenation). And then an SE-block is applied to assign the fusion weights to the feature maps of different scales. This fusion weights act as the weights for the fusion of the final segmentation results of different scales. The final segmentation result will be predicted by the fusion feature maps. We treat this result as the fusion of multiple segmentation results generated by feature maps of different scales. In summary, our feature fusion module consists of two parts: PFC and SE-block.

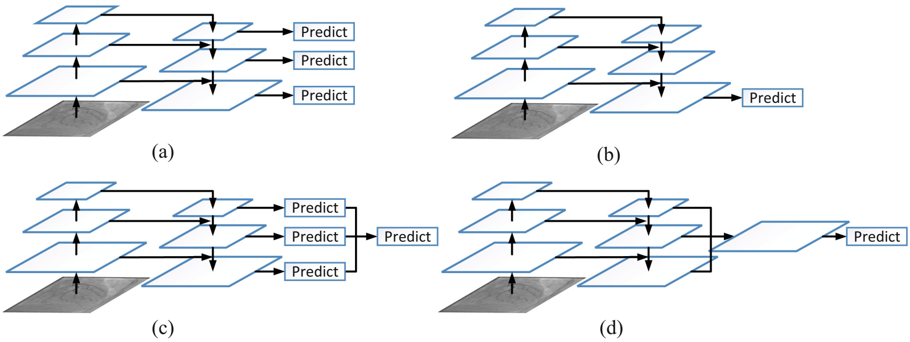

Fig. 2. Some schematic structures of networks: (a) Schematic diagram of FPN. (b) Schematic diagram of U-net. (c) The structure we want to propose. (d) The structure we actually propose

3.2 Pyramid Feature Concatenation

PFC is used to fuse all feature maps of each scale. The feature maps are first reduced to the same channels as the output feature maps by a 1×1 convolution and then are upsampled to the same size as the segmentation result, as shown in Fig. 1. The reason for using concatenation instead of addition in PFC is to ensure that the features of each scale can be completely retained to generate the final segmentation results.

In addition to drawing on the idea of FPN, we explain the role of PFC from another perspective. Because high-level semantic information is the key to removing noise in DSA images, PFC can help the segmentation results obtain more semantic features to reduce the noise, which is the main reason why PFC works.

3.3 SE-Block

SE-block was first introduced in [16], the features are first squeezed by global average pooling to generate channel-wise statistics, then two fully connected (FC) layers are applied to fully capture channel-wise dependencies. At last, multiply the channel-wise dependencies by the original feature maps to get the excitation feature maps, as shown in Fig. 1.

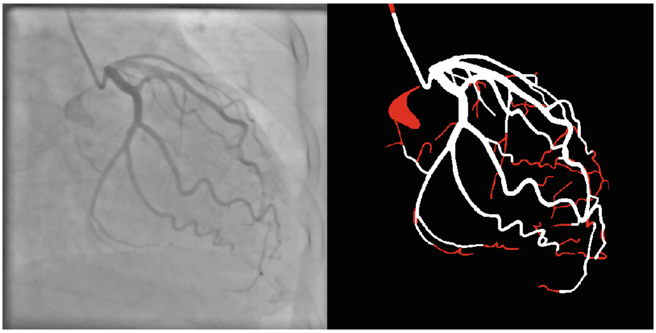

Fig. 3. An example of the sample in our dataset. (left) The original image. (right) Corresponding label (Color figure online)

By using an SE-block, we can obtain the importance of the feature maps at different levels to the final segmentation result. At the same time, it can balance the importance of semantic information and appearance information, so that the final result can get a good performance in removing the noise and preserving the details of blood vessels.

3.4 Loss Function

Following the SE-block, the final feature maps are generated. A 1×1 convolution layer is applied to the final feature maps to calculate the segmentation result. We use sigmoid cross-entropy as the loss function of the segmentation result:

$$Loss = -(y log y^{*} + (1-y) log(1-y^{*})) \tag{1}$$

In this equation, y is the segmentation result. y^{*} is the ground truths of segmentation.

4 Dataset

We established a dataset to verify the superiority of our model in coronary artery segmentation. Our dataset contains 538 samples from 36 DSA sequences in total, including all 412 samples from 7 DSA sequences and 126 samples randomly selected from 29 DSA sequences. All samples were labeled at the pixel-level.

Because of the special image-forming principle of DSA, we used a new way to label our dataset. Instead of labeling all pixels into background and foreground, all pixels are labeled into three categories: coronary artery (White), background (Black), and uncertainty (Red), as you can see from Fig. 3. Uncertainty means that we cannot determine whether some pixels should be marked as background or foreground when labeling. There are several reasons why a pixel may be labeled as uncertainty:

1. The contrast agent, which makes the coronary artery visible, is liquid. The liquid doesn't have a definite boundary. It's just because of the constraints of the vessel wall that make the boundary of the vessel appears in DSA images. But the axial direction of the vessel, also the direction of blood flow, is unconstrained, so along this direction, there are always some pixels where it's hard to say whether it should be labeled as the vessel or not.
2. Again, since the contrast agent is liquid, there will be a phenomenon that the contrast agent will partially backflow at the catheter orifice and produce smoke, makes it difficult to label the category of some pixels covered by smoke.
3. There are often shadows on the border of DSA images. When the blood vessels go out of the image range, the pixels at the boundaries will be covered by shadows. We also think that these pixels are difficult to mark.
4. Because not enough contrast agent flowing in, some small vessels can barely show, which makes some pixels hard to label.

These pixels are annotated as a new category because they are regarded as bad samples. In other deep learning tasks bad samples should be removed from datasets. Due to the special image-forming principle of DSA, these pixels are inevitable in DSA images, so we annotate them as a new category. Loss generated by these pixels is not calculated in the training, and the performance of these pixels is not calculated in the test.

In our experiments, we use 337 samples from 24 DSA sequences as training set, and 201 samples from 12 DSA sequences as testing set.

5 Experiments

5.1 Implementation Details

The network is implemented using Tensorflow, and for optimization, SGD optimizer with a momentum of 0.9 is applied and batch size is 8. We use a initial learning rate of 0.1, and the initial learning rate is multiplied by 0.8 every 4000 steps to avoid overfitting. For data augmentation, random flip, random crop, random grayscale adjustment $[-20, 20]$ and random contrast ratio $[0.8, 1.2]$ are adopted. Training takes about 9 h on an NVIDIA Titan XP for 2000 epochs.

5.2 Evaluation Metrics

Similar to the retinal vessel segmentation task, we also use the metrics including Specificity (Sp), Sensitivity (Se) and Accuracy (Acc) to evaluate the segmentation result. These metrics are defined below:

$$Sp = \frac{TN}{TN + FP}, Se = \frac{TP}{TP + FN}, Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

TP, FN, TN, FP denote true positive, false negative, true negative and false positive, respectively.

Table 1. Results of ablation experiment

Methods	Acc	Se	Sp	Dice
U-net(baseline)	0.9910	**0.8905**	0.9956	0.8817
U-net+PFC	0.9913	0.8836	0.9963	0.8968
U-net+SE	0.9907	0.8873	0.9955	0.8670
U-net+PFC+SE(Ours)	**0.9919**	0.8863	**0.9969**	**0.9035**

*PFC stands for Pyramid Feature Concatenation, SE stands for SE-Block

Since the coronary artery segmentation task contains most of the easily segmented pixels, the improvements of Sp, Se and Acc are not significant. In order to more intuitively reflect the performance differences between different methods, Dice score is used to measure the segmentation performance.

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{3}$$

5.3 Ablation Experiments

We run a number of ablations to analyze the performance gains of our feature fusion module. Results are shown in Table 1. Because DSA images contain many pixels that can be easily segmented correctly, the performance of U-net on Acc, Se and Sp is already good. But both of PFC and SE-block can significantly improve the segmentation performance, indicating that our method performs better on the pixels that are hard to classify correctly. Good detail performance is important for coronary artery segmentation.

It should be pointed out that we also try to add only SE-block on U-net, but we find that the result gets even worse. That is to say, the SE-block can only work if it is used together with the PFC. This result also indicates that the function of the SE-block is to assign importance to the feature maps at different scales, so it is meaningless to use the SE-block on feature maps at a single scale.

5.4 Comparing with Other Methods

We do not find any specific CNN design for coronary vessel segmentation, so we compared our method with some classic CNN methods for medical image segmentation to prove that our network is superior to other methods for coronary vessel segmentation. We also compare our network with some traditional methods, including an unsupervised method and a supervised method. These competing methods are briefly introduced as follows:

1. **Frangi** [2]: Frangi filter is one of the most popular Hessian-based multiscale filter used for vessel enhancement. The output processed by Frangi filter gives

Table 2. Performance of all segmentation methods on our dataset

Methods	Acc	Se	Sp	Dice
Frangi [2]	0.9591	0.7141	0.9706	0.5532
CRF [12]	0.9676	0.7236	0.9790	0.6213
Attention U-net [18]	0.9908	0.8809	0.9959	0.8811
U-net [1]	0.9910	**0.8905**	0.9956	0.8817
Resnet50 U-net	0.9912	0.8846	0.9962	0.8850
Ours	**0.9919**	0.8863	**0.9969**	**0.9035**

a possibility that each pixel belongs to a blood vessel. To get the segmentation results, a threshold is used to form an unsupervised method. The threshold and other hyperparameters of Frangi filter were optimized using random search.

2. **CRF** [12]: Conditional random field (CRF) is a widely used image segmentation method. In this method, each pixel represents a node, and each node is connected with an edge to its neighbors. Then, energy minimization helps to achieve segmentation based on the graphs. Although the CRF-based method proposed in [12] is for the retinal vessel segmentation, the method is essentially used for the segmentation of blood vessels, so we still regard it as a comparison method.
3. **Resnet50 U-net**: In many computer vision fields, the use of pre-trained models to speed up training and improve training results has become a common method, especially when training data is insufficient. To verify if this method works on our task, we also tried to replace the encoder part of U-net with a pre-trained model and fine-tune with our data. The pre-trained model we use here is Resnet50 [17], so we name the model Resnet50 U-net.
4. **Attention U-net** [18]: Attention U-net is one of the state-of-the-art algorithms for medical segmentation. A novel attention gate (AG) was proposed in Attention U-net. AGs can implicitly learn to suppress irrelevant regions in an input image while highlighting salient features useful for the segmentation task. This network has been shown to be more effective than original U-net in many organ segmentation tasks. As a variant of U-net, we also apply this model for comparison with our method.

The final results are shown in Table 2. We visualized some results in Fig. 4. According to the results, we can see that whether supervised or unsupervised, the traditional segmentation method does not work well, and many background pixels are segmented into coronary arteries. This is because the features used in the traditional methods are handcraft features, which can only extract the appearance features but not the semantic features, so that all the structures that look like vessels are classified into coronary arteries, and such similar structures happen to be very common in DSA images.

Similarly, U-net also misclassifies a small number of background pixels into coronary arteries, especially when there are only a few positive pixels in the image. We believe this is because U-net only use the feature maps of high resolution to predict segmentation result, and the high-level semantic information are not fully utilized.

From the results, Attention U-net and Resnet50 U-net, as two commonly used methods to improve the accuracy of feature extraction, do not significantly improve the segmentation results compared with U-net. We believe that this is mainly due to the characteristics of DSA images and the characteristics of the coronary arteries. DSA images are grayscale images and relatively simple, so there is little knowledge that can be transferred from the model pre-trained on ImageNet. At the same time, as a single-class segmentation task, it is not significant to calculate the attention of coronary arteries.

According to the results in Table 2, our CAU-net is more suitable for coronary vessel segmentation than other methods because of our novel feature fusion module. Compared to the best performing network, CAU-net improved by 1.85% in the Dice score. As shown in Fig. 4, we could see that our method performed better in the details of coronary vessel segmentation and removal of impurities.

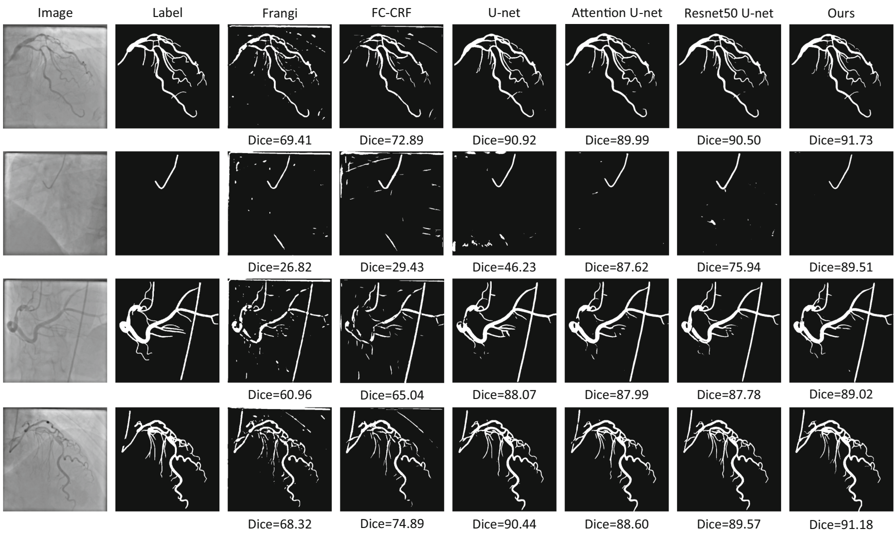

Fig. 4. Illustration of the segmentation results achieved by Frangi, CRF, U-net, Attention U-net, Resnet50 U-net and our method; the four rows indicate four sets of results

6 Conclusion

We propose a new CAU-net for automatic segmentation of coronary arteries in DSA images. In this network a new feature fusion module is proposed to fuse multi-scale features, which can significantly improve the performance in coronary artery segmentation. We also establish a coronary artery segmentation dataset with a special labeling method and compared many methods with our CAU-net using this dataset. The experimental results show that our method achieves the state-of-the-art segmentation results in coronary artery segmentation.

Acknowledgments. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFB1311700, the National Natural Science Foundation of China under Grants 61533016, U1913601, and 61421004, the Youth Innovation Promotion Association of CAS under Grant 2020140 and the Strategic Priority Research Program of CAS under Grant XDBS01040100.

References

1. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
2. Frangi, A.F., Niessen, W.J., Vincken, K.L., Viergever, M.A.: Multiscale vessel enhancement filtering. In: Wells, W.M., Colchester, A., Delp, S. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 130–137. Springer, Heidelberg (1998). https://doi.org/10.1007/BFb0056195
3. Manniesing, R., Viergever, M.A., Niessen, W.J.: Vessel enhancing diffusion: a scale space representation of vessel structures. Med. Image Anal. **10**(6), 815–825 (2006)
4. Brieva, J., Gonzalez, E., Gonzalez, F., Bousse, A., Bellanger, J.J.: A level set method for vessel segmentation in coronary angiography. In: IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, pp. 6348–6351 (2005)
5. Kerkeni, A., Benabdallah, A., Manzanera, A., Bedoui, M.H.: A coronary artery segmentation method based on multiscale analysis and region growing. Comput. Med. Imag. Graph. **48**, 49–61 (2016)
6. Dehkordi, M.T., Mohamad, A., Hoseini, D., Sadri, S., Soltanianzadeh, H.: Local feature fitting active contour for segmenting vessels in angiograms. IET Comput. Vis. **8**(3), 161–170 (2014)
7. Fan, J., et al.: Multichannel fully convolutional network for coronary artery segmentation in x-ray angiograms. IEEE Access **6**, 44635–44643 (2018)
8. Yang, S., et al.: Automatic coronary artery segmentation in X-ray angiograms by multiple convolutional neural networks. In: Proceedings of the 3rd International Conference on Multimedia and Image Processing, pp. 31–35 (2018)
9. Staal, J., Abramoff, M.D., Niemeijer, M., Viergever, M.A., van Ginneken, B.: Ridge-based vessel segmentation in color images of the retina. IEEE Trans. Med. Imaging **23**(4), 501–509 (2004)
10. Hoover, A., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. IEEE Trans. Med. Imaging **19**(3), 203–210 (2000)
11. You, X., Peng, Q., Yuan, Y., Cheung, Y., Lei, J.: Segmentation of retinal blood vessels using the radial projection and semi-supervised approach. Pattern Recogn. **44**(10–11), 2314–2324 (2011)
12. Orlando, I., Prokofyeva, E., Blaschko, M.B.: A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images. IEEE Trans. Biomed. Eng. **64**(1), 16–27 (2016)

13. Li, Q., et al.: A cross-modality learning approach for vessel segmentation in retinal images. IEEE Trans. Med. Imaging **35**(1), 109–118 (2016)
14. Wu, Y., Xia, Y., Song, Y., Zhang, Y., Cai, W.: Multiscale network followed network model for retinal vessel segmentation. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11071, pp. 119–126. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00934-2_14
15. Lin, T., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 936–944. IEEE (2017)
16. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7132–7141. IEEE (2018)
17. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9908, pp. 630–645. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46493-0_38
18. Oktay, O., et al.: Attention U-net: learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)