IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 21, 2018, accepted August 5, 2018, date of publication August 10, 2018, date of current version September 5, 2018.

*Digital Object Identifier 10.1109/ACCESS.2018.2864592*

# Multichannel Fully Convolutional Network for Coronary Artery Segmentation in X-Ray Angiograms

**JINGFAN FAN[1,2], JIAN YANG[1], YACHEN WANG[1], SIYUAN YANG[1], DANNI AI[1], YONG HUANG[1], HONG SONG[3], AIMIN HAO[2], AND YONGTIAN WANG[1]**

[1]Beijing Engineering Research Center of Mixed Reality and Advanced Display, School of Optics and Electronics, Beijing Institute of Technology, Beijing 100081, China
[2]State Key Laboratory of Virtual Technology and Systems, Beihang University, Beijing 100083, China
[3]School of Software, Beijing Institute of Technology, Beijing 100081, China

Corresponding author: Jian Yang (jyang@bit.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC0107900 and in part by the National Science Foundation Program of China under Grant 61672099, Grant 81627803, Grant 61501030, and Grant 61527827.

**ABSTRACT** Accurate segmentation of coronary arteries in X-ray angiograms is an important step for the quantitative study of coronary artery disease. However, accurate segmentation is a challenging task because coronary arteries are thin tubular structures with relatively low contrast and the presence of artifacts. In this paper, a novel deep-learning-based method is proposed to automatically segment the coronary artery from angiograms by using multichannel fully convolutional networks. Since the artifacts appear in both live images (after the injection of contrast material) and mask images (before the injection of contrast material) and the blood vessels appear only in live images, we take the mask images into consideration to distinguish real blood vessel structures from artifacts. Therefore, both live images and mask images are used as multichannel inputs to provide enhanced vascular structure information. The hierarchical features are then automatically learned to characterize the spatial associations between vessel and background and are further used to achieve the final segmentation. In addition, a dense matching between the live image and mask image is processed for a precise initial alignment. The experimental results demonstrate that our method is effective and robust for coronary artery segmentation, compared with several state-of-the-art methods.

**INDEX TERMS**
Coronary artery, fully convolutional network, dense matching, U-net.

## I. INTRODUCTION

X-ray angiography is the most common image modality applied in the clinical diagnosis of coronary artery disease (CAD) due to its powerful ability to perspectively inspect the vascular structure of the coronary artery [1], [2]. An increasing number of studies have demonstrated that the accurate segmentation of the vascular structure from angiograms can assist doctors in diagnosis and treatment planning. However, X-ray angiography also brings some unavoidable imaging problems, such as low contrast, motion artifacts caused by heart beating and respiration, and integration effects of the catheter and spine in the angiograms [3]. These problems make accurate segmentation of the vascular structure a challenging problem for the objective diagnosis of CAD.

To extract the vascular structure, the traditional segmentation methods usually focus on specific image filters [4]–[9], tracking operators [10]–[12] and optimization models [13]–[16]. The specific image filters, including wavelets [4], Gaussian filter [6], Gabor filter [7], and Hessian matrix [8], are employed to enhance the tubular structure in vascular images while suppressing the background area. Then, the threshold-based classification methods can simply and quickly achieve blood vessel segmentation on the enhanced images. The tracking methods [10]–[12] start with some chosen seed points marked in the blood vessel and then iteratively connect them based on topology to obtain a complete vessel. Additionally, the active contour model [13], region growing model [14], graph cut model [15], and conditional random field model [16] are the most popular

optimization models applied to determine the boundary of the segmented vascular structures by maximizing the feature differences. However, these traditional segmentation methods generally require complicated preprocessing steps, e.g., specifying regions of interest (ROIs), to guide the segmentation procedure. Moreover, their performances are often influenced by the low quality of X-ray angiograms and motion artifacts.

Recently, deep convolutional neural networks (CNN) have shown outstanding performance in various natural and medical image computing tasks, such as image recognition [17] and semantic segmentation [18], [19]. The CNN model has also been applied to vessel segmentation, especially retinal vessel segmentation [20]–[22]. All the methods are implemented via a patch-based learning strategy, i.e., CNN is used as a classifier to classify whether a center point from a selected patch belongs to the vessel region. Rather than taking only a local patch with limited size into consideration, fully convolutional networks (FCN) [18], [19] combine the global semantic information with local details to efficiently produce end-to-end segmentation by using advanced encoder-decoder architectures on full-size images. Existing FCN models in the literature [23] have achieved great success for the segmentation of neuronal structures. However, when these state-of-the-art methods are applied to angiograms, they perform poorly in regions that are visually similar to vessel regions, such as artifacts and catheters. Therefore, segmentation by using only one source of information (e.g., original images), may not be sufficient to effectively distinguish real vascular structures from artifacts in angiograms.

For this reason, a multichannel fully convolutional neural network is proposed in this paper for automatic segmentation of coronary arteries in X-ray angiograms. As an extension of the original FCN, the proposed method applies convolution, pooling, and deconvolution operators to achieve end-to-end segmentation. Specifically, it adopts multichannel inputs to provide both live images (after the injection of contrast material) and mask images (before the injection of contrast material) for precise identification of the highlighted area of a coronary artery based on the differences. In addition, the motion between the live image and mask image caused by heart beating and respiration is corrected by hierarchical matching the correspondence between the images. The experimental results on 148 X-ray angiographic image sequences demonstrate the superior performance of the proposed method compared with state-of-the-art methods.

The remainder of this paper is organized as follows. Section II.A proposes the multichannel inputs in the proposed FCN framework; Section II.B presents the network architecture; Section II.C introduces the matching method for the live and mask images; Section II.D introduces the preparation of the training data. Section III presents the experimental results, followed by a discussion of future directions and applications in Section IV.

## II. METHOD

### A. MULTICHANNEL INPUTS

Many CNN-based vessel segmentation methods adopt a patch-based classification strategy [20]–[22] that regards blood vessel detection as a pixel-wise binary classification problem. The trained CNN classifier estimates whether the center position of a patch is in a coronary artery or not. As a result, the limited range of the local patch transforms the segmentation into several independent problems for each pixel, and the segmentation results may be discontinuous and noisy. To take the global semantic information into consideration, the FCN-based [18], [19] segmentation model [23] introduces an encoder-decoder architecture for end-to-end segmentation. However, artifact regions and catheter regions are very similar to the vessel regions in angiographic images, and this similarity may lead to incorrect classification in these regions. A set of examples is shown in Fig. 1, where Fig. 1 (a) presents a live image with clear vascular structure. Fig. 1 (a1)-(a4) show four zoomed regions on the live image that represent an artifact region, diaphragm region, catheter region, and real vessel region, respectively. Tubular structures are present in all four zoomed images, so they are all likely to be considered to be potential vascular structures if the only reference is the live image.

Hence, we propose multichannel-input-based FCN to address this limitation, inspired by digital subtraction technology [24], [25]. The mask image shown in Fig. 1 (b) is selected from the same angiographic image sequence of Fig. 1 (a), where Fig. 1 (b1)-(b4) correspond to Fig. 1 (a1)-(a4). The three fake vessel regions Fig. 1 (b1)-(b3) are highly similar to the live image Fig. 1 (b1)-(b3), whereas the real vessel regions in Fig. 1 (b4) and Fig. 1 (a4) are completely different because the live and mask images were taken before and after injection of the contrast material, respectively. Significant differences occur in the regions of vessel, while the background remains consistent, except for heartbeat and respiratory motion. Therefore, we introduce an image registration method for the live and mask images in Section II.C to obtain the aligned mask image shown in Fig. 1 (c). Clearly, Fig. 1 (c1)-(c4) are aligned to the position of Fig. 1 (a1)-(a4), especially Fig. 1 (c2). Then, Fig. 1 (a4) is easily identified as the real vessel region, while Fig. 1 (a1)-(a3) are fake vessel regions. Therefore, in the proposed method, the aligned mask image is also fed into the FCN model with the live image as multichannel inputs to obtain a better segmentation result.

### B. NETWORK ARCHITECTURE

Fig. 2 shows the architecture of our proposed network, which is a variant FCN model [18], [19], including multiple convolutional layers, max pooling layers, deconvolutional layers, and fully convolutional layers, just like U-net [23]. In the encoder path, the multichannel inputs first go through the 64 channels of the $3 \times 3$ convolutional layer with zero

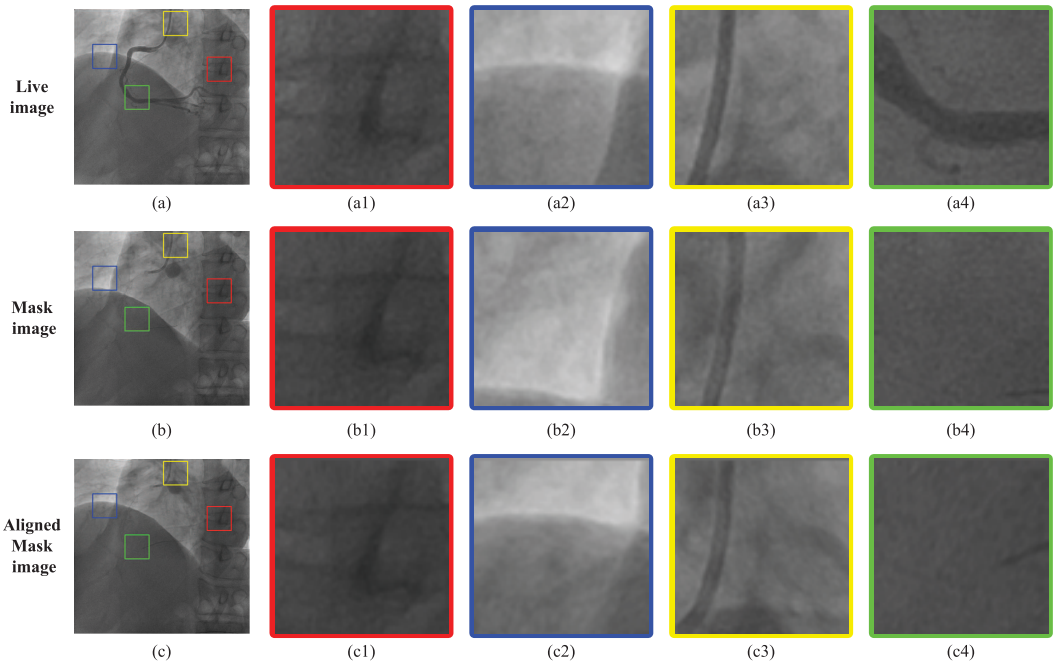



**FIGURE 1.** Example patches from the live image, mask image, and aligned mask image. (a) Live image. (b) Mask image. (c) Aligned mask image. (a1)-(a4) Zoomed patches from the artifact region, diaphragm region, catheter region, and real vessel region. (b1)-(b4) Corresponding zoomed patches of (a1)-(a4) in the mask image. (c1)-(c4) Corresponding zoomed patches of (a1)-(a4) in the aligned mask image.

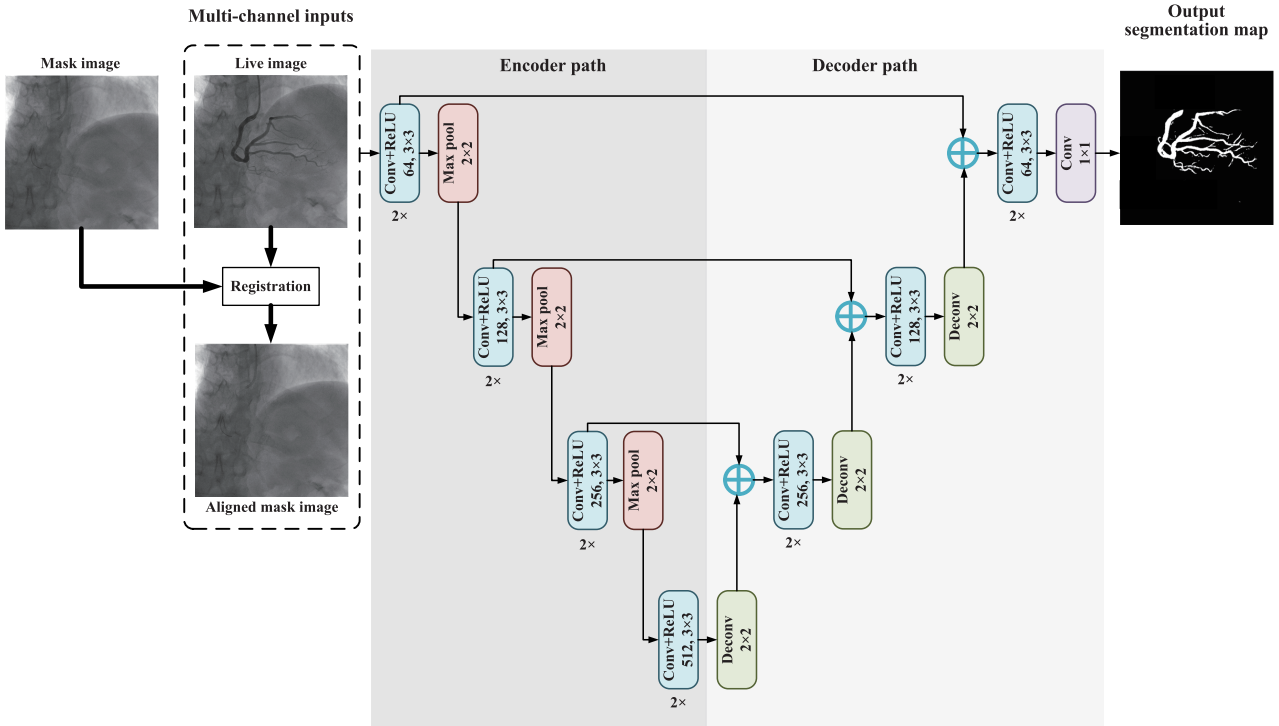



**FIGURE 2.** Architecture of the proposed deep learning network.

padding (stride of 1), followed by rectified linear unit (ReLU) activation [26]. After these two convolutions, a 2 × 2 max pooling layer is connected to downsample the feature maps. The same process is repeated three times to reach the lowest resolution level, and the channel number is doubled during each downsampling.

Symmetric to the encoder path, the decoder path increases the resolution level three times by using a 2 × 2 deconvolutional layer. Similar convolutional layers with ReLU activation are performed twice in each resolution level, and the channel number is halved during each upsampling. Since the up sampling deconvolutional layer loses the sharpness of the feature maps, at each resolution level, a concatenation operator is included to fuse the upsampled feature maps with the same level feature maps obtained from the previous encoder path. Then, both the global contextual information from the low-level features and the spatial details from the previous convolution guide the precise segmentation. The final segmentation map is achieved by a 1 × 1 fully convolutional layer and sigmoid activation. The coefficients of the network can be learned using the training images with ground-truth segmentations (achieved by the data preparation method described in Section II.D) of the coronary arteries.

### C. MATCHING OF LIVE AND MASK IMAGES

Mask images are widely used in angiography subtraction, as they usually have similar backgrounds to those of the live images [24], [25]. However, as live and mask images are acquired at different times, considerable displacements exist between them, as shown in Fig. 1 (a2) and (b2). To address the displacements, the hierarchical deformable dense matching method [27] is used to generate the matching correspondences between the live and mask images; then, the aligned mask images are registered to the live images, as shown in Fig. 1 (a2) and (c2).

Hierarchical deformable dense matching [27] is a general matching framework that allows robust determination of the dense correspondences between two images. The algorithm consists of two major steps. First, the multilevel correlation maps, which indicate the similarity scores between the multisize local patches of a feature point from the live image and each position in the mask image, are calculated by a bottom-up strategy. Second, a top-down strategy is used to estimate the movement of the feature points by iteratively maximizing the similarity scores in multilevel correlation maps. This process is also guided by the motion information obtained from a level higher than the current one (except the top level).

### D. TRAINING DATA PREPARATION

148 angiography image sequences are included in this study. We choose one pair of live and mask images from each sequence to form 148 image pairs. The resolution of the images is 512 × 512. All the live images were segmented manually by experts from the collaborating hospital to form the ground truth. In total, 130 pairs of angiograms are taken as the training set and the other 18 pairs are chosen as the test set.

Since only a few manually annotated images are available for training, data augmentation is necessary to teach the network the desired invariance and robustness properties. In the case of different imaging positions, we randomly choose 9 sets of rotation and translation invariance for each pair of images to obtain 1300 pairs of images, including the original images. In addition, random elastic deformations of the training samples appear to be the key concept to train a segmentation network with very few annotated images [23]. Hence, we generate 3 random smooth deformations on each image pair to obtain 5200 pairs of images for training by using random displacement vectors on a coarse 3 × 3 grid.

## III. EXPERIMENTAL RESULTS

In this section, we first present the experimental settings and the competing methods, then evaluate the effectiveness of the improved network proposed in this paper, and finally compare the segmentation results achieved by different methods.

### A. EXPERIMENTAL SETTINGS

As mentioned in Section II.D, there are 130 pairs of manually annotated angiograms taken as the training set and 18 pairs of angiograms chosen as the test set. Each angiographic image is 512 × 512 pixels with 256 gray levels. The network is implemented using Caffe [28] and optimized using SGD [29] using a computer with a single GPU (i.e., NVIDIA GTX 1080 8GB). We set the initial learning rate to 1e-2, and multiply it by 0.5 after every 10000 steps.

To quantitatively compare the segmentation performance of our method with that of state-of-the-art methods, we use various evaluation metrics. First, for the segmentation result, true positive (TP) denotes the number of predicted vascular pixels inside the ground-truth vascular segmentation; false positive (FP) denotes the number of predicted vascular pixels outside the ground-truth vascular segmentation; true negative (TN) denotes the number of predicted background pixels outside the ground-truth vascular segmentation; false negative (FN) denotes the number of predicted background pixels inside the ground-truth vascular segmentation. The accuracy (Acc), Sensitivity (Sen), Specificity (Spe), Precision (Pre), and F1-score are defined as:

$$\begin{cases} Acc = \dfrac{TP + TN}{TP + TN + FP + FN}, \quad Sen = \dfrac{TP}{TP + FN} \\ Spe = \dfrac{TN}{TN + FP}, \quad Pre = \dfrac{TP}{TP + FP}, \quad F1 = \dfrac{2 \cdot Pre \cdot Sen}{Pre + Sen} \end{cases} \tag{1}$$

The above five indicators range in [0, 1], and the larger the score, the better the segmentation.

### B. COMPETING METHODS

In this paper, we propose a multichannel input segmentation network for X-ray angiograms based on the U-net structure. Two models are trained by the prepared training data, one with the live and mask images, named multichannel segmentation network (MSN), and the other with the live and the aligned mask images, named multichannel segmentation network with aligned inputs (MSN-A), which is the final version of the proposed method. To evaluate the segmentation

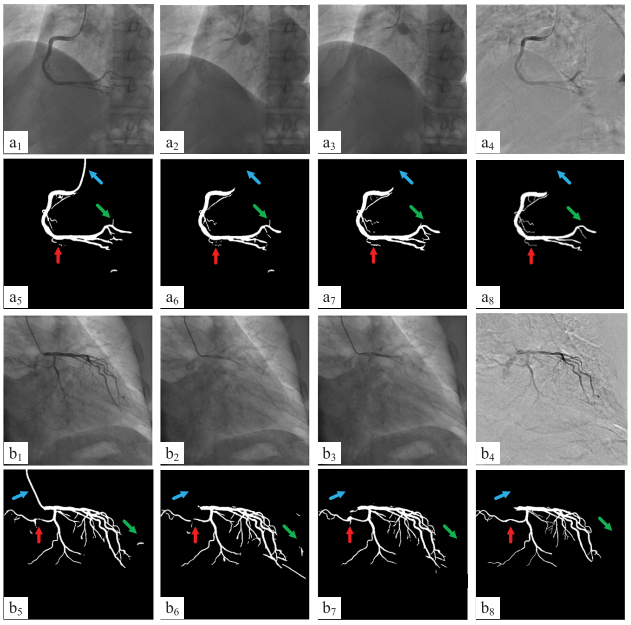



**FIGURE 3.** Illustration of the segmentation results achieved by different networks. (a1) Live image. (a2) Mask image. (a3) Aligned mask image. (a4) Subtraction image of the live image and the aligned mask image. (a5) Segmentation result by U-net. (a6) Segmentation result by MSN. (a7) Segmentation result by MSN-A. (a8) Ground truth of segmentation. (b1)-(b8) The images from LCA, corresponding to (a1)-(a8) from RCA.

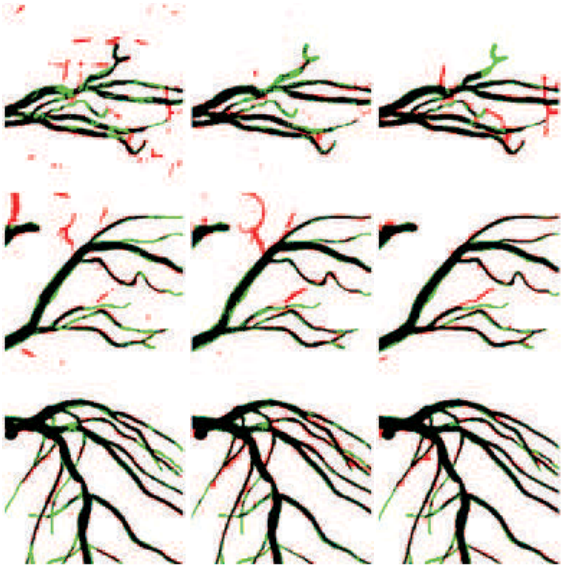


**FIGURE 4.** Zoomed illustration of the segmentation results achieved by different networks. First column: Segmentation result by U-net. Second column: Segmentation result by MSN. Third column: Segmentation result by MSN-A. Different rows of images show zoomed vessels in different regions, in which black, white, red, and green regions indicate TP, TN, FP, and FN, respectively.

performance, four baseline methods are chosen as competing methods, namely, the conditional random filed (CRF) method [16], LevelSet method [9], CNN method [20], and original U-net method [23]. These four competing methods are briefly introduced as follows.

1) **CRF** [16]: CRF is an extensively used image segmentation method. In this method, image segmentations are mapped to graphs, where each pixel represents a node, and every node is connected with an edge to its neighbors according to a certain connectivity rule [30]. Then, energy minimization helps to achieve the segmentation based on the graphs.
2) **LevelSet** [9]: The LevelSet model combines both line and edge detection using quadrature filters across multiple scales. The filter result gives well-defined vessels as linear structures, while distinct edges facilitate a robust segmentation.
3) **CNN** [20]: CNN is used as a classifier to classify whether a center point from a selected patch belongs to the vessel region on the live images. To train the CNN model, the same training data (Section II.D) are included to extract the training patches following the method in [20].
4) **U-net** [23]: The original U-Net is also an encoder-decoder-architecture-based FCN model that uses only live image, instead of multichannel inputs. For fair comparison, all the training settings (Section III.A) are consistent with those of the proposed method.

### C. EFFECTIVENESS OF THE IMPROVED NETWORK

In this subsection, we compare the proposed MSN and MSN-A to the original U-net model to evaluate the effectiveness of the improved network, especially with the multichannel inputs and matching method for the live and mask images. Fig. 3 presents two sets of segmentation results on the left coronary artery (LCA) and right coronary artery (RCA). First, the aligned mask images (Fig. 3 (a3) and (b3)) effectively reduce the interference of the motion caused by heart beating and respiration between the live images (Fig. 3 (a1) and (b1)) and the mask images (Fig. 3 (a2) and (b2)). Fig. 3 (a4) and (b4) show the subtraction images of the live images and the aligned mask image, and they give a rough vascular enhancement but cannot achieve accurate binary segmentation. Fig. 3 (a5)-(a8) and (b5)-(b8) illustrate the segmentation results achieved by U-net, MSN, MSN-A and the manually annotated ground truth, respectively. In these figures, three salient regions, namely the catheter, vessel-like artifacts, and difficult to detect thin vessels, are marked with blue, green, and red arrows, respectively. Compared to U-net, the proposed MSN and MSN-A effectively identify the catheter (blue arrow) as a non-vascular region. Additionally, MSN-A achieves the most accurate vascular segmentation in the thin vessel region (red arrow) and the strongest noise removal ability in the artifact region (green arrow).

We also quantitatively evaluate the accuracy of the segmentation results based on the ground truth manually annotated by experts, as shown in Table 1. By comparing the results obtained by MSN with those obtained by the original

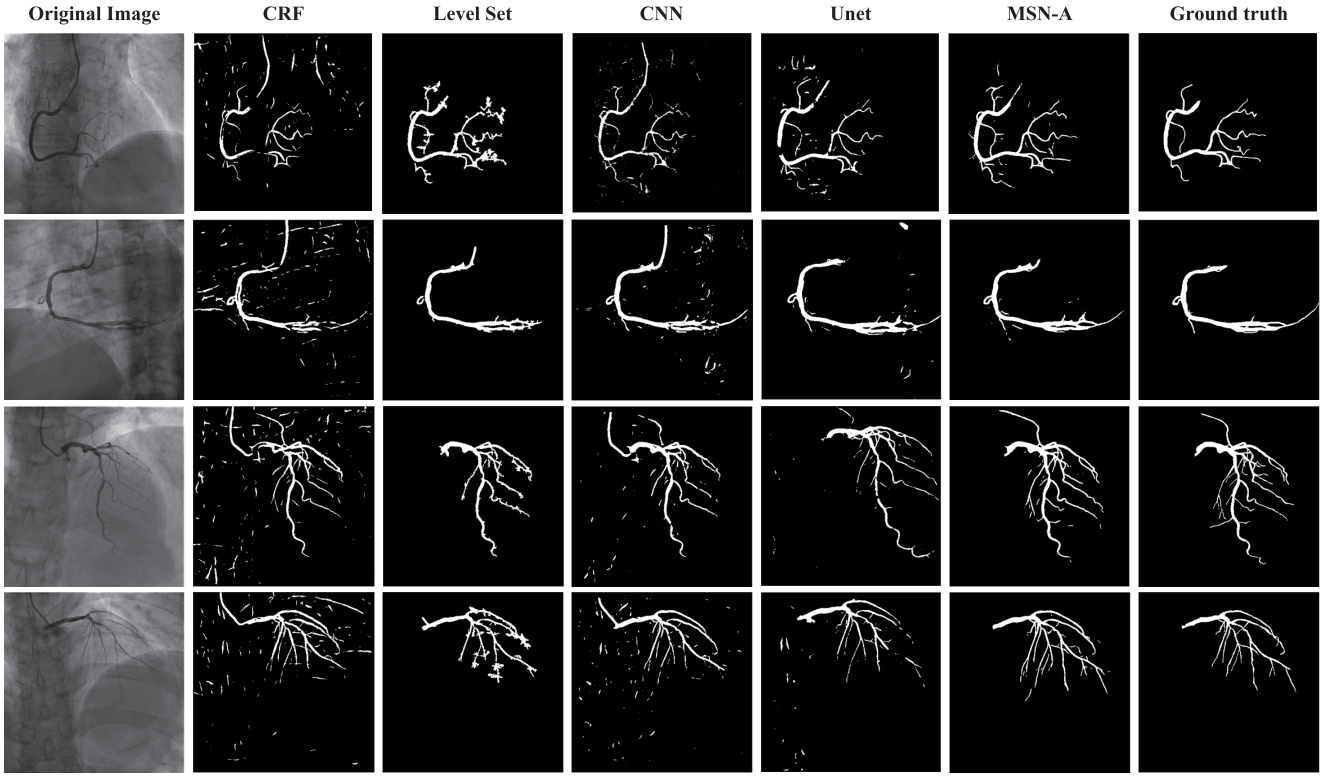



**FIGURE 5.** Illustration of the segmentation results achieved by CRF, LevelSet, CNN, U-net, and the proposed MSN-A; the four rows indicate four sets of results.

**TABLE 1.** The average segmentation performance, in terms of Pre, Sen, Spe, Acc and F1-score, obtained by U-net, MSN, and MSN-A for the ten testing images.

| Methods | Pre | Sen | Spe | Acc | F1-score |
|---|---|---|---|---|---|
| U-net | 0.8192 | 0.8215 | 0.9947 | 0.9876 | 0.8203 |
| MSN | 0.8206 | 0.8649 | 0.9923 | 0.9875 | 0.8402 |
| MSN-A | **0.8678** | **0.8773** | **0.9954** | **0.9881** | **0.8725** |

**TABLE 2.** The average Sen value of the segmentation performance on vascular regions with different thickness obtained by U-net, MSN, and MSN-A.

| Methods | Large | Medium | Small |
|---|---|---|---|
| U-net | 0.9423 | 0.8347 | 0.7231 |
| MSN | 0.9500 | 0.9134 | 0.7468 |
| MSN-A | **0.9505** | **0.9211** | **0.7678** |

U-Net, we observe that multichannel inputs are beneficial for improving the performance of U-Net (i.e., average F1-score is improved from 0.8203 to 0.8402). Furthermore, by incorporating the image registration method, MSN-A achieves the best performance (i.e., average F1-score is improved from 0.8402 to 0.8725). This result demonstrates that it is possible to effectively improve the segmentation performance by using live images and aligned mask images as multichannel inputs.

A more detailed view of the segmentation results is shown in Fig. 4. The FP region (red) and the FN region (green) are the smallest in the segmentation result of the MSN-A method. Compared to the rough edges and clutter noise in the segmentation results of U-net and MSN, the results obtained by our MSN-A method have smooth edges and fine vascular segmentation detail.

To evaluate the sensitivity of the segmentation of tiny vascular structures, we divided the vascular regions into three types according to the size of the radius (i.e., large: the radius is larger than 5 pixels; medium: the radius is in 3 to 5 pixels; small: the radius is smaller than 3 pixels). The average Sen value is calculated for the segmentation results, as shown in Table 2. The results show that the Sen value of the three methods was relatively close in the area of the large vessel regions, while the advantages of the proposed MSN-A method were obvious in the small and medium vessel regions, i.e., the Sen value increases from 0.8347 to 0.9211 for medium vessels and from 0.7231 to 0.7678 for small vessels.

### D. EVALUATION ON SEGMENTATION RESULTS

After verifying the effectiveness of the improved network, the comparison of its performance with the state-of-the-art methods is given in Fig. 5. The first column of Fig. 5 shows four live images from the testing data; the second to sixth columns show the segmentation results obtained by CRF,

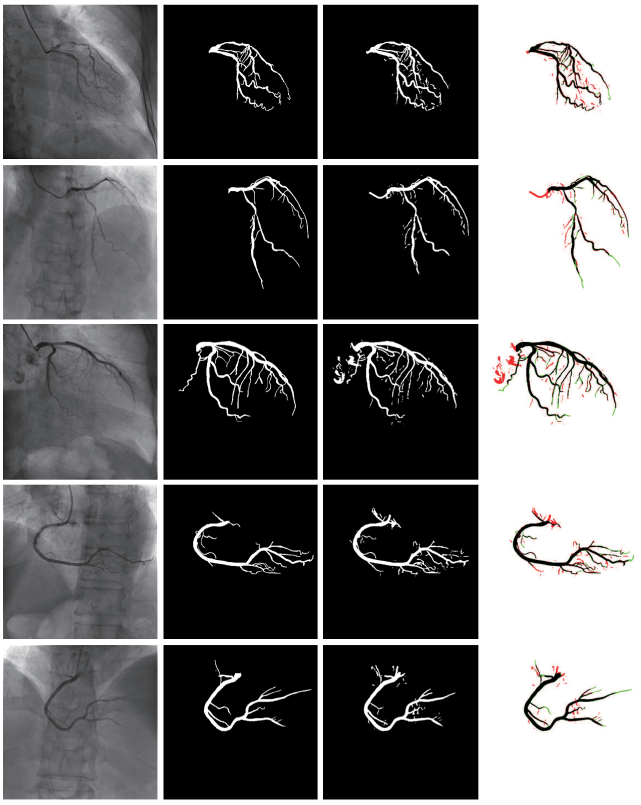


**FIGURE 6.** More results from our proposed method. First column: original live images. Second column: ground-truth segmentations. Third column: the segmentation results obtained by the proposed MSN-A method. Fourth column: the color maps of the segmentation errors, where black, white, red, and green regions indicate TP, TN, FP, and FN.

LevelSet, CNN, U-net, and the proposed MSN-A method, respectively; and the last column shows the ground-truth segmentations of the angiograms. Overall, all methods can segment most of the vascular regions. However, they perform very differently in the artifact regions. The CRF, CNN and U-net methods perform poorly in removing catheter and sparse artifacts, as these regions are very similar to the vascular regions. The LevelSet method and our method perform well in removing both catheters and artifacts, but the LevelSet method does not perform well in extracting vascular details. Moreover, the LevelSet method requires many seed points that are carefully selected in the vascular regions by manually labeling. By contrast, our method is fully automatic and free of parameter tuning. Additionally, the automatic segmentations obtained by the proposed MSN-A are more consistent with the manual ground truth in these examples, especially for the small vascular structures.

The quantitative segmentation results of the testing images obtained by our MSN-A method and the four competing methods are reported in Table 3. First, compared with the conventional segmentation methods (i.e., CRF and LevelSet) and deep-learning-based segmentation methods (i.e., CNN and U-net), the proposed MSN-A method achieves good results in coronary artery segmentation in terms of all five evaluation criteria (i.e., Pre, Sen, Spe, Acc and F1-score).

**TABLE 3.** Segmentation results for different methods.

| Methods | Pre | Sen | Spe | Acc | F1-score | Runtime (s) |
|---|---|---|---|---|---|---|
| CRF | 0.7353 | 0.5975 | 0.9906 | 0.9744 | 0.6578 | 1.14 |
| LevelSet | 0.7725 | 0.7394 | 0.9910 | 0.9810 | 0.7534 | 7.65 |
| CNN | 0.7311 | 0.5473 | 0.9919 | 0.9730 | 0.6219 | 70.33 |
| U-net | 0.8192 | 0.7087 | 0.9955 | 0.9851 | 0.7600 | **0.11** |
| MSN-A | **0.8547** | **0.8773** | **0.9974** | **0.9881** | **0.8659** | 0.12 |

Second, in terms of the computational time required to segment one image, U-net (0.11 s) and MSN-A (0.12 s) are the least time consuming, and the conventional methods take longer (1.14 s for CRF and 7.65 s for LevelSet). Although CNN is a deep-learning-based segmentation method, it is the slowest (70.33 s) method due to the patch-based learning strategy. Five additional segmentation results for angiographic images are shown in Fig.6.

## IV. CONCLUSION

In this study, we have proposed a multichannel FCN model (MSN-A) to achieve efficient end-to-end segmentation of vascular structures in X-ray angiograms. MSN-A adopts a multichannel input strategy to comprehensively characterize the vascular structural information from live images and mask images complementarily. To correct the motion caused by heart beating and respiration, a hierarchical dense matching method is employed to align the mask images to the live images. Compared to the typical single-channel CNN or U-net method, taking the aligned mask image into consideration is suitable for segmentation in low-contrast angiograms, especially for removing catheters and artifacts. Experiments demonstrate that the proposed end-to-end segmentation method is very effective and efficient and can achieve better results than those of three state-of-the-art methods. Moreover, the proposed MSN-A method can be further improved in the future by taking all the angiographic image sequences into consideration instead of choosing one live image and one mask image.

## REFERENCES

[1] J. Yang, Y. Wang, Y. Liu, S. Tang, and W. Chen, "Novel approach for 3-D reconstruction of coronary arteries from two uncalibrated angiographic images," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1563–1572, Jul. 2009.

[2] W. Cong, J. Yang, D. Ai, Y. Chen, Y. Liu, and Y. Wang, "Quantitative analysis of deformable model-based 3-D reconstruction of coronary artery from multiple angiograms," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 7, pp. 1389–1400, Jul. 2016.

[3] Y. Chen *et al.*, "Artifact suppressed dictionary learning for low-dose CT image processing," *IEEE Trans. Med. Imag.*, vol. 33, no. 12, pp. 2271–2292, Dec. 2014.

[4] P. Bankhead, C. N. Scholfield, J. G. McGeown, and T. M. Curtis, "Fast retinal vessel detection and measurement using wavelets and edge location refinement," *PLoS ONE*, vol. 7, no. 3, p. e32435, 2012.

[5] F. M'hiri, N. L. T. Hoang, L. Duong, and M. Cheriet, "A new adaptive framework for tubular structures segmentation in X-ray angiography," in *Proc. 11th Int. Conf. Inf. Sci., Signal Process. Appl. (ISSPA)*, 2012, pp. 496–500.

[6] Y. Li, S. Zhou, J. Wu, X. Ma, and K. Peng, "A novel method of vessel segmentation for X-ray coronary angiography images," in *Proc. 4th Int. Conf. Comput. Inf. Sci. (ICCIS)*, 2012, pp. 468–471.

[7] B. Felfelian *et al.*, "Vessel segmentation in low contrast X-ray angiogram images," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 375–379.
[8] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Berlin, Germany: Springer, 1998, pp. 130–137.
[9] G. Läthén, J. Jonasson, and M. Borga, "Blood vessel segmentation using multi-scale quadrature filtering," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 762–767, 2010.
[10] Y. Chen *et al.*, "Curve-like structure extraction using minimal path propagation with backtracking," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 988–1003, Feb. 2016.
[11] L. Grady, "Random walks for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 11, pp. 1768–1783, Nov. 2006.
[12] F. M'hiri, L. Duong, C. Desrosiers, and M. Cheriet, "Vesselwalker: Coronary arteries segmentation using random walks and hessian-based vesselness filter," in *Proc. IEEE 10th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2013, pp. 918–921.
[13] M. T. Dehkordi, A. M. D. Hoseini, S. Sadri, and H. Soltanianzadeh, "Local feature fitting active contour for segmenting vessels in angiograms," *IET Comput. Vis.*, vol. 8, no. 3, pp. 161–170, 2013.
[14] M. Jin, R. Li, J. Jiang, and B. Qin, "Extracting contrast-filled vessels in X-ray angiography by graduated rpca with motion coherency constraint," *Pattern Recognit.*, vol. 63, pp. 653–666, Mar. 2017.
[15] M. W. K. Law and A. C. S. Chung, "Efficient implementation for spherical flux computation and its application to vascular segmentation," *IEEE Trans. Image Process.*, vol. 18, no. 3, pp. 596–612, Mar. 2009.
[16] J. I. Orlando, E. Prokofyeva, and M. B. Blaschko, "A discriminatively trained fully connected conditional random field model for blood vessel segmentation in fundus images," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 1, pp. 16–27, Jan. 2017.
[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[18] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.
[19] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
[20] P. Liskowski and K. Krawiec, "Segmenting retinal blood vessels with deep neural networks," *IEEE Trans. Med. Imag.*, vol. 35, no. 11, pp. 2369–2380, Nov. 2016.
[21] A. Dasgupta and S. Singh, "A fully convolutional neural network based structured prediction approach towards the retinal vessel segmentation," in *Proc. IEEE 14th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2017, pp. 248–251.
[22] H. Fu, Y. Xu, D. W. K. Wong, and J. Liu, "Retinal vessel segmentation via deep learning network and fully-connected conditional random fields," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 698–701.
[23] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.
[24] J. Yang, Y. Wang, S. Tang, S. Zhou, Y. Liu, and W. Chen, "Multiresolution elastic registration of X-ray angiography images using thin-plate spline," *IEEE Trans. Nucl. Sci.*, vol. 54, no. 1, pp. 152–166, Feb. 2007.
[25] S. Tang, Y. Wang, and Y.-W. Chen, "Application of ICA to X-ray coronary digital subtraction angiography," *Neurocomputing*, vol. 79, pp. 168–172, Mar. 2012.
[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[27] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "DeepMatching: Hierarchical deformable dense matching," *Int. J. Comput. Vis.*, vol. 120, no. 3, pp. 300–323, 2016.
[28] Y. Jia *et al.*, "Caffe," in *Proc. ACM Int. Conf. Multimedia (MM)*, 2014, pp. 675–678.
[29] C. Zhang *et al.*, "Theory of deep learning III: Generalization properties of SGD," Center Brains, Minds Mach., Cambridge, U.K., 2017.
[30] S. Kumar and M. Hebert, "Discriminative random fields: A discriminative framework for contextual interaction in classification," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 1150–1157.

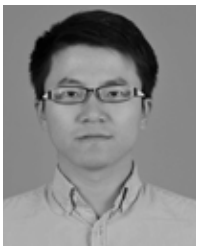
**JINGFAN FAN** received the Ph.D. degree in optical engineering from the Beijing Institute of Technology in 2016. He is currently a Post-Doctoral Researcher with the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University. He focuses his research interests on medical image processing, computer vision, and augmented reality.

**JIAN YANG** received the Ph.D. degree in optical engineering from the Beijing Institute of Technology, China, in 2007. He is currently a Professor with the School of Optoelectronics, Beijing Institute of Technology. He focuses his research interests on medical image processing, augmented reality, and computer vision.

**YACHEN WANG** received the B.S. degree in optical engineering from the Beijing Institute of Technology, China, in 2017. He is currently pursuing the master's degree with the School of Optoelectronics, Beijing Institute of Technology. His research interests include medical image processing and computer vision.

**SIYUAN YANG** received the B.S. and M.S. degrees in optical engineering from the Beijing Institute of Technology in 2014 and 2017, respectively. He is currently an Engineer at Meitu Inc. His research interests include medical image processing and computer vision.

**DANNI AI** received the B.S. and M.S. degrees from Xi'an Jiaotong University in 2005 and 2008, respectively, and the Ph.D. degree from Ritsumeikan University, Japan, in 2011. She is currently an Associate Professor with the School of Optics and Photonics, Beijing Institute of Technology. Her research interests include medical image analysis, surgical navigation, virtual reality, and augmented reality.



**YONG HUANG** received the Ph.D. degree from the Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MA, USA, in 2013. He was a Post-Doctoral Research Fellow with the Department of Electrical and Computer Engineering, Johns Hopkins University, from 2013 to 2014. He is currently an Associate Professor at the School of Optics and Photonics, Beijing Institute of Technology, China. His research interests include biomedical imaging, image processing, and high-performance computation.

**HONG SONG** received the Ph.D. degree in computer science from the Beijing Institute of Technology, China, in 2004. She is currently an Associate Professor with the School of Software, Beijing Institute of Technology. She focuses her research interests on medical image processing, augmented reality, and computer vision.

**AIMIN HAO** received the B.S., M.S., and Ph.D. degrees from Beihang University, all in computer science. He is currently a Professor with the Computer Science School and an Associate Director of the State Key Laboratory of Virtual Reality Technology and Systems, Beihang University. His research interests include virtual reality, computer simulation, computer graphics, geometric modeling, image processing, and computer vision.

**YONGTIAN WANG** received the Ph.D. degree in optics from the University of Reading, U.K., in 1986. He is currently a Yangtze River Scholar, and the Director of the Optoelectronics and Information Technology Center, Beijing Institute of Technology. His research interests include optical design and CAD, virtual reality, and augmented reality technologies and applications.