

Vessel Layer Separation in X-ray Angiograms with Fully Convolutional Network

Haidong Hao^a, Hua Ma^{b,*}, Theo van Walsum^b

^aFaculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, the Netherlands;

^bBiomedical Imaging Group Rotterdam, Erasmus MC, Rotterdam, the Netherlands;

*corresponding author: h.ma@erasmusmc.nl

ABSTRACT

Percutaneous coronary intervention is a minimally-invasive procedure to treat coronary artery disease. In such procedures, X-ray angiography, a real-time imaging technique, is commonly used for image guidance to identify lesion sites and navigate catheters and guide-wires within coronary arteries. Due to the physical nature of X-ray imaging, photon energy undergoes absorption when penetrating tissues, rendering a 2D projection image of a 3D scene, in which semi-transparent structures overlap with each other. The overlapping structures make robust information processing of X-ray images challenging. To tackle this issue, layer separation techniques for X-ray images were proposed to separate those structures into different image layers based on structure appearance or motion pattern. These techniques have been proven effective for vessel enhancement in X-ray angiograms. However, layer separation approaches still suffer either from spurious structures or non-real-time processing, which prevent their application in clinics. Purpose of this work is to investigate whether vessel layer separation from X-ray angiography images is possible via a data-driven strategy. To this end, we develop and evaluate a deep learning based method to extract the vessel layer. More specifically, U-Net, a fully convolutional network architecture, was trained to separate the vessel layer from the background. The results of our experiments show good vessel layer separation on 42 clinical sequences. Compared to the previous state-of-the-art, our proposed method has similar performance but runs much faster, which makes it a potential real-time clinical application.

Keywords: X-ray Angiogram, Vessel Layer Separation, Fully Convolutional Network, U-Net, Deep Learning

1. INTRODUCTION

1.1 Motivation

Percutaneous coronary intervention (PCI) is a minimally-invasive procedure to treat coronary arteries disease. In such procedures, a catheter with a pre-mounted stent is introduced to the lesion site through the femoral or radial artery; during such procedures, X-ray angiography is used to visualize the blood vessels, and enables clinicians to navigate catheters and guidewires within the coronary artery. Since X-ray image formation is based on photon energy absorption of various tissues along the rays, the image can be seen as a superposition of 2D projections of 3D anatomical structures, such as spine, lung and heart, which become opaque or semi-transparent structures in X-ray images. These structures normally overlap with each other, which makes robust information processing in X-ray angiograms difficult.

1.2 Related works

Layer separation techniques were introduced to separate structures in X-ray images into different layers so that each layer contains structures of similar appearance or motion pattern. Existing layer separation methods for X-ray fluoroscopic sequences can be generally grouped into two classes:¹ motion-based methods²³ which rely on estimation of the layered motion, and motion-free approaches¹⁴⁵⁶ that do not require to estimate the layered motion.

Further author information: (Send correspondence to Hua Ma)

Hua Ma: E-mail:h.ma@erasmusmc.nl

Among motion-based methods, Zhang et al.² proposed a method to separate each fluoroscopic image into a static layer, a slow movement layer and a fast movement layer based on the observation that different anatomical structures have various motion patterns. Similarly, Zhu et al.³ developed a method separating each fluoroscopic image into a background layer and a coronary layer based on a Bayesian framework which combined dense motion estimation, uncertainty propagation and statistical fusion together.

In motion-free methods, the background layer and/or foreground (vessel) layer of each fluoroscopic image are modeled under certain hypotheses. Tang et al.⁵ proposed an approach which was based on an assumption that the vessel layer and background layer are reconstructed from independent signals and then utilized independent component analysis (ICA) to solve the problem. A method based on the robust principal component analysis (RPCA), which was used to detect and track the stent graft delivery device automatically in 2D fluoroscopic sequences, was proposed by D. Volpi et al.⁶ Similarly, Ma et al.¹ proposed a background modelling based approach which separates an X-ray angiogram sequence into a breathing layer, a quasi-static layer and a vessel layer using morphological closing and online robust PCA (OR-PCA) proposed by Feng et al.⁷ This method is one of the few that run online, which takes one frame as input each time and updates the background model based on the frame. Compared to its parental offline approach proposed by Ma et al.,⁴ this method showed similar performance and achieved a processing rate up to 6 frames per second (fps), which is still slower than common image acquisition rates (7.5-15 fps).

The purpose of this work is separating vessel layer from X-ray angiograms. We intend to model the task as an image-to-image mapping problem, particularly, from the domain of the original X-ray angiogram to the domain of the layer that contains mainly the vessel structures.

Recently, deep learning approaches such as fully convolutional networks (FCNs)⁸⁹ have been developed to find such mappings. A FCN is a deep learning network architecture that has no fully connected layers. Unlike traditional convolutional neural network (CNN) for classification tasks, FCN outputs an image that has the same size as the input image(s). Typical FCNs contain an encoding path, which is similar to traditional CNNs to encode image features, and decoding path which map the learned features to pixel-wise information, e.g. a semantic segmentation map. The deep network structure consists of many convolutional layers that allows learning more powerful image representation which serves as a key to success on many object classification and segmentation tasks. Among the various FCN architectures, Noh, H. et al.⁹ proposed a FCN consisting of a convolution part and a deconvolution part to segment RGB images. The convolution part extracts features from input and transforms to feature representations, whereas the deconvolution part reconstructs the object segmentation from the feature representations. Similarly, U-Net¹⁰ connects features that are learned from the encoding path to the decoding path to facilitate feature decoding and pixel-wise information reconstruction on different scales with skip connections. This network architecture has shown exceptional performances of segmentation tasks of biomedical images,¹⁰ but it has not been explored yet to what extent it can be applied to other tasks such as vessel layer separation.

1.3 Overview and contributions

The purpose of this work is to develop a robust layer separation method that can run in real-time, so as to be clinically applicable. In particular, we focus on vessel layer separation directly, as this layer contains most structures of interest, such as coronary arteries, guiding catheters and guidewires. The basic idea of this work makes use of the recent advances in deep learning which has shown good performance on many computer vision tasks and medical imaging applications. Particularly, U-Net,¹⁰ a fully convolutional network was used for separating the vessel layer from X-ray angiograms, which runs online and real time. We also proposed a weight mask for each training sample by morphologically dilating the inverted gray level of the reference vessel layer to calculate the training loss and let the network focus on learning features from the vessel area.

2. METHOD

2.1 Network architectures

In this work, the architecture of U-Net introduced by Ronneberger, O. et al.¹⁰ was utilized to map from original X-ray angiograms to vessel layer images. As shown in [Figure 1](#), firstly, the resolution of input images is 512×512 ; secondly, each convolution block has two 3×3 convolution layers with a ReLU (Rectified Linear Unit)

activation; thirdly, max pooling with stride 2 was used between convolution blocks in the encoding path and an upconvolution layer was used to connect two successive convolution blocks in the decoding path, which consists of an upsampling operation followed by a 2×2 convolution without ReLU; lastly, the final read-out layer has one 1×1 convolution filter with a Sigmoid activation, outputting the vessel layer image with size 512×512 .

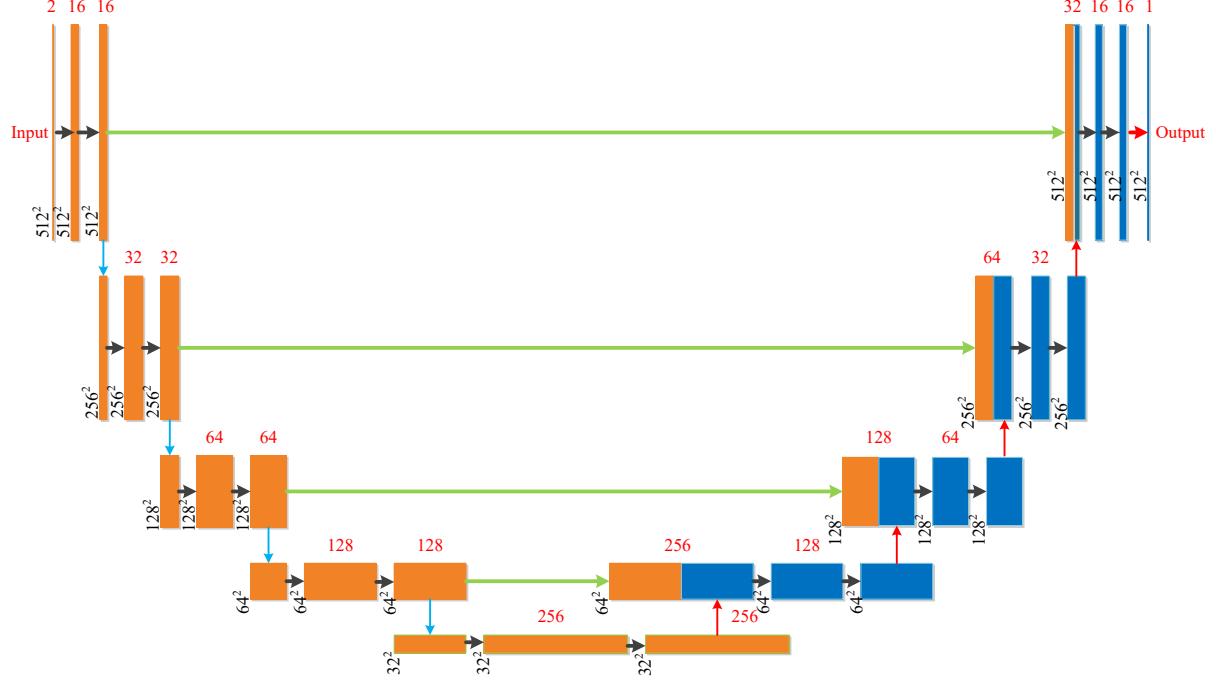


Figure 1. U-Net architecture. Left part and right part indicate the encoding path and decoding path, respectively. Each box indicates a feature map; the number on top of each box denote the number of feature maps; the number at the lower left edge of each box is the size of corresponding feature maps; different color arrows denote different operations (black arrow: 3×3 convolution with a ReLU activation, blue arrow: max pooling with stride 2, green arrow: skip connection, red up arrow: upsampling operation followed by a 2×2 convolution without ReLU, red right arrow: 1×1 convolution filter with a Sigmoid activation); the orange boxes in the decoding path represent corresponding copied feature maps from the encoding path.

2.2 Learning target

In this work, we did not follow the way in many previous works that let a network to fit to human-annotated labels. As the exact pixel values are usually not clearly defined in saliency maps, such as the vessel layer of X-ray angiogram images, which makes it difficult to obtain the “ground truth” vessel layer with human annotation, we employed the method proposed by Ma et al.⁴ to generate the vessel layer as the target of our learning task. The parameter setting described by the authors was also used in our study.

2.3 Loss function

A data set (\mathbf{x}, \mathbf{y}) is employed to train the network, in which \mathbf{x} is the input that is the original X-ray angiograms in this work, and \mathbf{y} denotes the corresponding reference vessel layers (learning target). Let $f(\mathbf{x}, \theta)$ denote U-Net as a mapping function, in which θ is the learnable parameters of U-Net. The problem can be formulated as $\hat{\mathbf{y}} = f(\mathbf{x}, \theta)$, in which $\hat{\mathbf{y}}$ denotes the separated vessel layer output by U-Net. Then, the vessel layer separation problem requires to find the optimal parameter θ which minimizes a loss function representing the difference between $\hat{\mathbf{y}}$ and the learning target \mathbf{y} .

Let \mathbf{y} and $\hat{\mathbf{y}}$ denote samples drawn from probability distribution $p(\mathbf{y})$ and $p(\hat{\mathbf{y}})$, respectively. The normalized pixel value of each pixel in both original angiogram and the corresponding reference image can be regarded as

the probability of that pixel belonging to the background and the probability of that pixel belonging to the blood vessel is one minus the normalized pixel value.¹¹

To quantify the difference between the prediction and the learning target, we use binary cross entropy (BCE) shown in [Equation \(1\)](#):

$$BCE = -\frac{1}{w \times h} \sum_{i=1}^w \sum_{j=1}^h [(1 - y_{i,j}) \log(1 - \hat{y}_{i,j}) + y_{i,j} \log \hat{y}_{i,j}] \quad (1)$$

in which, (w, h) is the size of the image; $y_{i,j}$ and $\hat{y}_{i,j}$ denote the probability of the pixel belonging to the background for the labeled and predicted pixel, respectively.

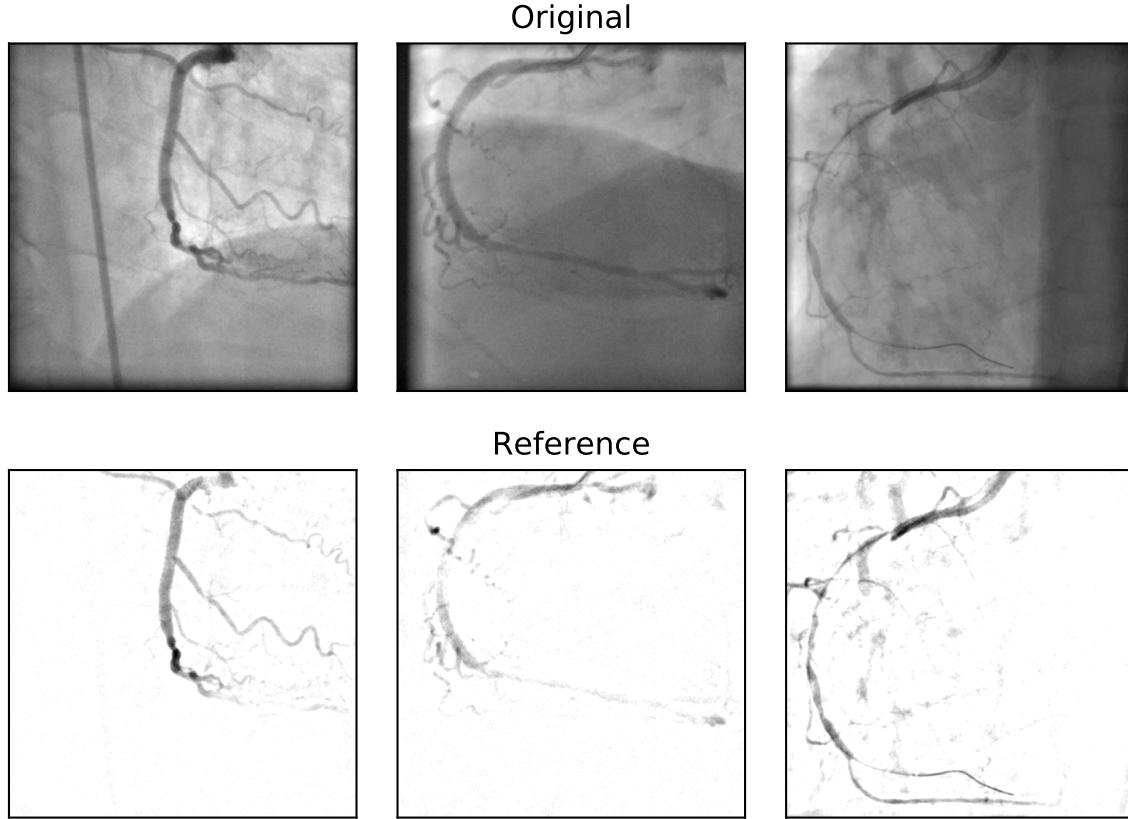


Figure 2. Examples of original angiogram and the corresponding reference vessel layer: (First Row) the original X-ray angiogram; (Second Row) the reference vessel layer generated by the method of Ma et al.⁴

From the reference images in [Figure 2](#), it can be seen that the vessel structures possess a small area in the complete image, while the background area takes up the majority; in other words, information from the vessel and background are imbalanced. To offset the imbalance of prevalence of vessel pixels and background pixels¹² and let the network focus on learning features from the vessel area, we created a weight mask for each training sample by morphologically dilating the inverted gray level of the reference vessel layer, so as to weight the vessel pixels more. An example of weight mask is shown in [Figure 3](#). Using the weight mask, another loss function, weighted binary cross entropy (ωBCE) is shown in [Equation \(2\)](#):

$$\omega BCE = -\frac{1}{w \times h} \sum_{i=1}^w \sum_{j=1}^h \omega_{i,j} [(1 - y_{i,j}) \log(1 - \hat{y}_{i,j}) + y_{i,j} \log \hat{y}_{i,j}] \quad (2)$$

in which, $\omega_{i,j}$ is the pixel value at location (i,j) of the weight mask.

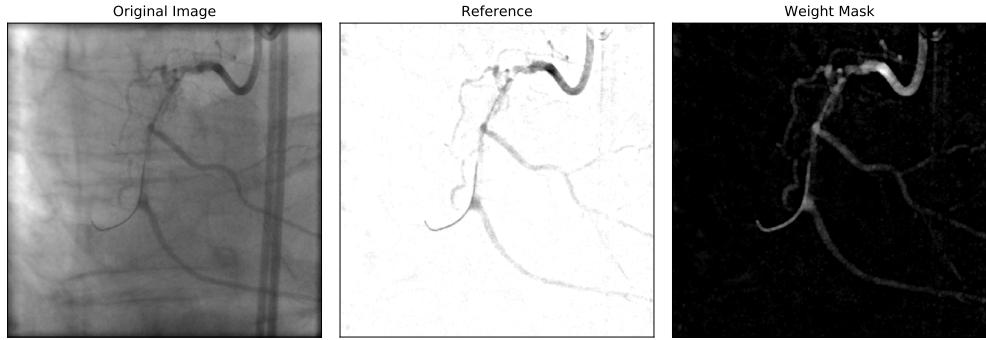


Figure 3. An example of original angiogram, reference image, weight mask: (left) the original X-ray image; (middle) the reference vessel layer; (right) the weight mask generated from the reference image with morphological dilation; The weight mask was used to calculate the training loss.

2.4 Utilization of temporal information

Because X-ray angiograms are time-series data, to take the advantage of the temporal information contained in the data set, apart from the original X-ray images, we also use difference images as additional input channels. This is expected to ignore the static structures and let the network focus on moving objects. Firstly, the difference image between the current frame and its previous frame as Channel 1 (Ch1); secondly, the other difference image which is the current frame minus the first frame as Channel 2 (Ch2); With these images, we constructed two different types of inputs for the network: the two-channel input (2Ch), which uses the original X-ray images (Channel 0) and Ch1, and the three-channel input (3Ch) with all three channels. An example of each channel input from clinical angiogram is shown in [Figure 4](#).

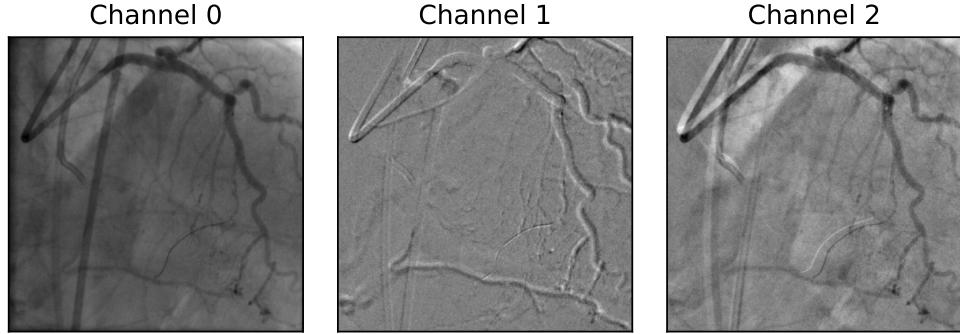


Figure 4. An example of each channel input from clinical angiogram. (left) the current frame; (middle) the current frame minus its previous frame; (right) the current frame minus the first frame.

3. EXPERIMENTS

3.1 Data set

Two types of data set were used in our experiments: clinical X-ray angiograms (XA) and synthetic low-contrast XA. Each dataset was divided into a training set, a validation set and a test set. The training set contains 2892 images from 26 sequences, the validating set contains 924 images from 6 sequences, and the test set contains 1068 images from 10 sequences.

3.1.1 Clinical X-ray angiograms

The same clinical X-ray angiograms used by Ma et al.¹ were used in this work. All images were resized to 512×512 as the network input, and the pixel values were normalized to the range between 0 and 1.

3.1.2 Synthetic low-contrast X-ray angiogram

Contrast agent for vessel visualization may cause kidney diseases or allergic reactions,¹³ so the dose of the contrast agent used in clinical application should be under control to ensure clear visualization while not causing harm to patients. To assess if our proposed method may be used to decrease contrast agent concentrations, we evaluate the performance of our methods on low-contrast data, and the method proposed by Ma et al.¹ was used to synthesize a 80% amount contrast data set and a 50% amount contrast data set. Both the two data set were preprocessed following the same procedure in section 3.1.1.

3.2 Evaluation metrics

To quantify the performance of vessel layer separation, the contrast-to-noise ratio (CNR) defined in Equation (3) was employed as one evaluation metric, which indicates the normalized difference between the average pixel value of the foreground and background. To evaluate the CNR for global and local scale, we adopted the foreground and background masks defined in Ma et al.¹ which are shown in Figure 5.

$$CNR = \frac{|\mu_F - \mu_B|}{\sigma_B} \quad (3)$$

in which, the mean of foreground and background pixel values are denoted by μ_F and μ_B ; the standard deviation of the background pixel values is denoted by σ_B . The global CNR measures the relation of the contrast between foreground and the whole background pixel intensities to the standard deviation of the whole background pixel intensities. On the other hand, the local CNR demonstrates the relation of the contrast between foreground and partial background surrounding the foreground pixel intensities to the standard deviation of the partial background pixel intensities.

In addition to CNR, which evaluates the contrast in a single image, we also adopt Structural SIMilarity

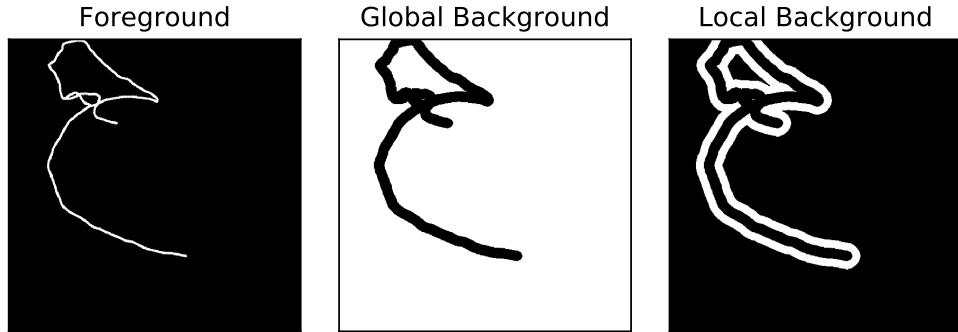


Figure 5. An example of foreground and background proposed by Ma et al.¹ (left) foreground (white area); (middle) global background (white area); (right) local background (white area).

(SSIM) proposed by Wang et al.¹⁴ to quantify the similarity between the predicted image and the reference image, in which the luminance, contrast and structure similarities between a reference image and a predicted image were measured independently and then all the luminance, contrast and structure similarities were combined for calculating the total similarity. The definition of SSIM follows Equation (4).

$$SSIM = \frac{(2\mu_t\mu_p + C_1)(2\sigma_{tp} + C_2)}{(\mu_t^2 + \mu_p^2 + C_1)(\sigma_t^2 + \sigma_p^2 + C_2)} \quad (4)$$

in which, μ_t and μ_p are the means of the reference image and the corresponding prediction image, respectively; σ_t^2 and σ_p^2 are the corresponding variances; σ_{tp} is the covariance between the reference image and the corresponding prediction image; $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$, where $K_1 = 0.01$, $K_2 = 0.03$ proposed by Wang et al.¹⁴ were adopted here, and L is the range of the pixel value, i.e. 1 in our work. Mean SSIM (MSSIM) is the average SSIM of Q pairs of images, which is defined by [Equation \(5\)](#).

$$MSSIM = \frac{1}{Q} \sum_{k=1}^Q SSIM_k \quad (5)$$

To compare the similarity between the reference image and the prediction of the network in vessel area and the whole image, we also adopted the foreground and background masks ([Figure 5](#)) defined in Ma et al¹ to calculate the global MSSIM (the whole image) and local MSSIM (the vessel area).

3.3 Experiment 1: hyper parameters tuning

The optimal hyper-parameter setting needs to be found, such as learning rate, epoch number for training, network architecture, number of filters. We searched the optimal hyper-parameters by comparing the average CNR, MSSIM and prediction examples between different hyper-parameter settings in the following way: first, we tuned the hyper-parameters with ωBCE loss function using two-channel input. After doing several pilot experiments with arbitrarily chosen hyper-parameter settings, we selected a combination of these hyper parameters as a reference, which is shown in the second row of [Table 1](#), and then arranged four sub-experiments (Ex1-Ex4) to find the optimal learning rate, epoch number, network architecture, and number of filters of the first convolution layer. In all the sub-experiments of [Table 1](#), the weight mask of ωBCE loss function were generated by the corresponding reference image of the training data with the steps described as below:

Step1: Invert the pixel value of the reference image to make the vessel area with larger pixel value than the background;

Step 2: Dilate the resulted image from step 1 using a 3×3 square kernel;

Step 3: Normalize the pixel value of the weight mask in the range from 0 to 1.

Table 1. Sub-experiments of Hyper parameters selection with ωBCE loss function. UNet7 is an alternative architecture of U-Net as shown in [Figure 6](#) and UNet9 is as shown in [Figure 1](#).

Experiment	Learning rate	Epoch No.	Architecture	Filter No. (f)
Reference	1e-5	50	UNet9	16
Ex1	1e-3, 5e-5, 5e-4	50	UNet9	16
Ex2	1e-5	30, 70	UNet9	16
Ex3	1e-5	50	UNet9	8, 32
Ex4	1e-5	50	UNet7	8, 16, 32

To compare the performances between two-channel and three-channel input, after finding the best hyper parameter combination for the ωBCE loss function based on two-channel input, another sub-experiments was conducted by replacing two-channel input with three-channel input.

3.4 Experiment 2: Compare with other methods

The best ωBCE method was compared with one of the best method proposed by Ma et al.¹ using CNR as the evaluation metric, which is the OR-PCA method with closed-form solution (CF) as the subspace basis updating strategy and sliding window (SW) as the past information downweighting technique ($SW + CF$). The optimal parameters of $SW + CF$ method are as below: the intrinsic rank of the subspace basis $r = 5$, the regularization parameters $\lambda_1 = \lambda_2 = 2.1$ and the window size $t_0 = 3$.

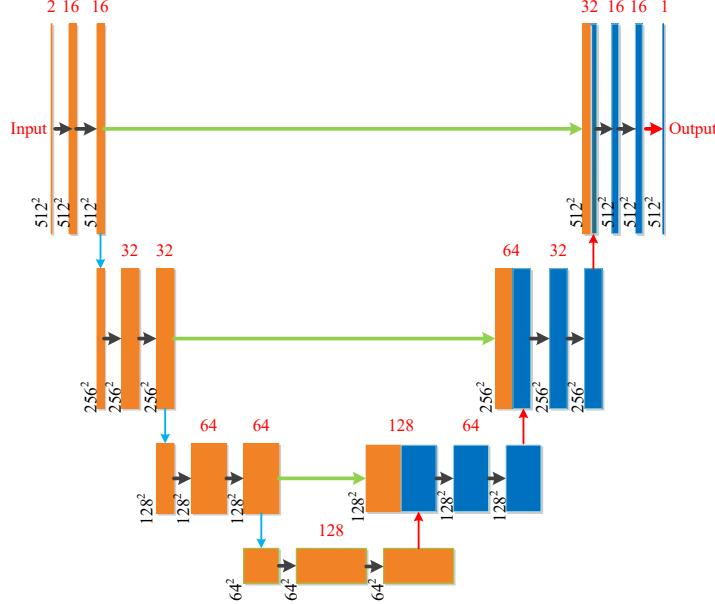


Figure 6. An alternative architecture of U-Net.

3.5 Experiment 3: performance of low contrast data

As shown in [Table 2](#), to evaluate the performances of our methods on low contrast data, we replaced the clinical X-ray angiograms with a 80% amount contrast data set (80%) and a 50% amount contrast data set (50%), respectively. The reference are the same as those in Experiment 1. Because the vessels in the low contrast data are subtle, especially in the 50% amount contrast data set, we employed three-channel input to train, evaluate and test the network in addition to two-channel input.

Table 2. Sub-experiments of Hyper parameters selection with low contrast data (50% and 80%, respectively).

Experiment	Loss Function	Epoch No.	Data Set
Ex5	ωBCE	50, 70	2Ch, 3Ch

3.6 Implementation

The network was trained and evaluated on the Dutch national supercomputer with an NVIDIA Tesla K40m GPU using Keras with Tensorflow as the backend. The network parameter θ were trained using an ADAM optimizer.¹⁵

4. RESULTS AND DISCUSSION

4.1 Optimal hyper parameters

Three of the best hyper parameter combinations based on the ωBCE loss function are shown in [Table 3](#) and the corresponding average CNR and MSSIM are shown in [Figure 7](#) and [Table 4](#). For the architecture, UNet9 is similar to UNet7 in terms of CNR and MSSIM, but there are much less parameters in UNet7, so there is a trade-off between performance and speed. If the compute capability of GPUs is limited, we can choose UNet7 and get acceptable results. For the input, both 2Ch and 3Ch got similar performance. The reason may be that the vessel structure is very clear in the clinical angiograms as can be seen in [Figure 2](#), so the two-channel input

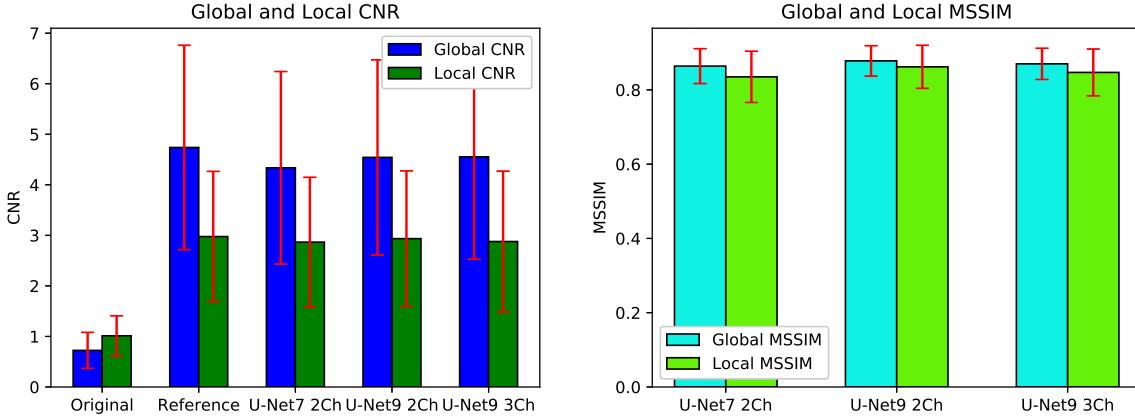


Figure 7. Average CNR and MSSIM of various ω BCE methods

is adequate to achieve results similar to the three-channel input. Figure 8 illustrates two prediction examples of the three hyper parameter combinations listed in Table 3; it also shows that the predictions of our ω BCE methods are worse than the reference image, especially in the background area. The catheter in the predictions in row 1 and the spine structures in row 2 are visible, which may be because both catheter and spine in the original angiograms are similar in both structure and colour to the vessel.

Table 3. Three of the best hyper parameter combinations of ω BCE loss function.

Learning rate	Epoch No.	Architecture	Filter No. (f)	Dataset
1e-5	50	UNet7	16	2Ch
1e-5	50	UNet9	16	2Ch
1e-5	50	UNet9	16	3Ch

Table 4. The average CNR and MSSIM of various ω BCE methods. (mean \pm standard deviation)

Method	Local CNR	Global CNR	Local MSSIM	Global MSSIM
Reference	2.976 ± 1.289	4.739 ± 2.021	1	1
UNet7-2Ch	2.866 ± 1.282	4.336 ± 1.905	0.835 ± 0.069	0.864 ± 0.047
UNet9-2Ch	2.935 ± 1.342	4.543 ± 1.930	0.862 ± 0.058	0.878 ± 0.041
UNet9-3Ch	2.878 ± 1.392	4.551 ± 2.023	0.847 ± 0.063	0.870 ± 0.042

We also assessed whether the hyper parameter combinations are statistically significantly different in terms of average CNR and MSSIM, for which we employed a two-sided Wilcoxon signed-rank test.¹⁶ The results are shown in Table 5, in which, UNet7 is statistically significantly different from both the two UNet9 hyper parameter combinations except local CNR; the two UNet9 hyper parameter combinations are not statistically significantly different except MSSIMs.

In summary, the second row of [Table 3](#) (UNet9-2Ch) is the optimal hyper parameter combination among all the combinations listed in [Table 1](#) for our project based on ωBCE loss function.

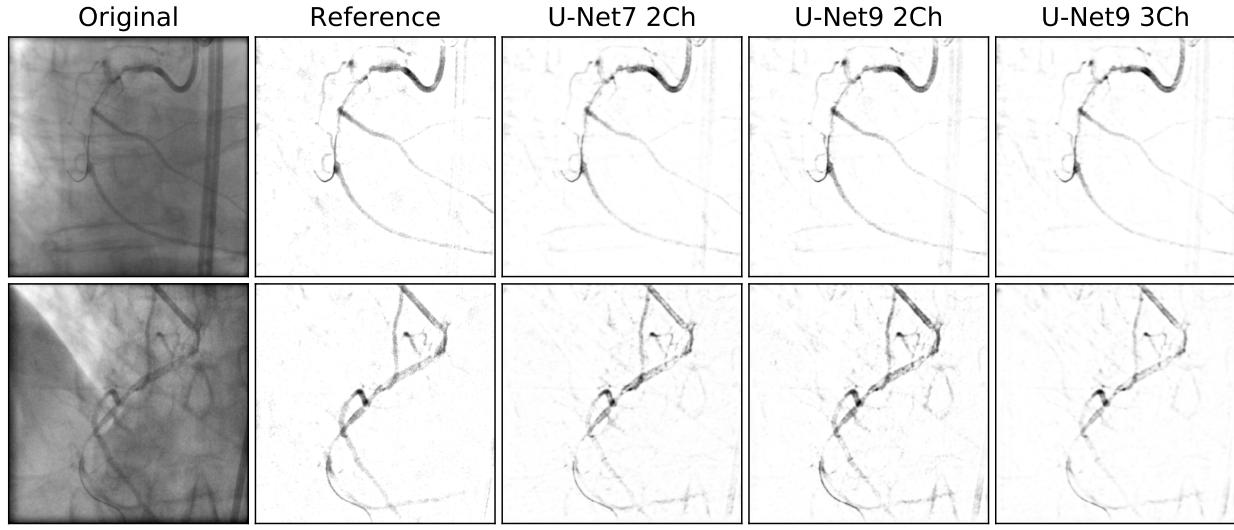


Figure 8. Two Prediction Examples of various ωBCE methods

Table 5. p-values among various ωBCE loss function methods in terms of CNR and MSSIM.

Method 1	Method 2	Local CNR	Global CNR	Local MSSIM	Global MSSIM
UNet7-2Ch	UNet9-2Ch	0.114	<0.001	<0.001	<0.001
UNet7-2Ch	UNet9-3Ch	0.632	0.001	0.003	0.044
UNet9-2Ch	UNet9-3Ch	0.084	0.694	0.001	0.040

4.2 Comparing with other methods

[Figure 10](#) shows examples of vessel layer separation using two different methods. Two frames (the first and second Row) from two different sequences qualitatively exhibit the performances of our methods (ωBCE) and the method $SW + CF$ presented by Ma et al.¹ These results show that the performance of our method are close to the reference image (the second column) and the method of Ma et al¹ (the last column). Compared to the work of Ma et al,¹ the background obtained with our method contains fewer structures, although the vessel area seems slightly worse.

The ωBCE method has similar CNR measures to the reference image. Compared to the method of Ma et al,¹ our method has superior performance on global CNR, but slightly worse on local CNR, as shown in [Figure 9](#) and [Table 6](#). This may be because there are less dark structures in the predictions of our method as shown in [Figure 10](#), which decrease σ_B and increase μ_B , resulting in the increase of global CNR. For local CNR, μ_F of our method are larger than that of the method of Ma et al,¹ which decreases $|\mu_F - \mu_B|$, leading to the decrease of local CNR. In terms of the processing speed, the proposed method achieves a rate of about 18 fps thanks to the use of a GPU. This is faster than the common image acquisition rate in clinics (15 fps). This result demonstrates potential for a real-time clinical application.

The Wilcoxon signed-rank test¹⁶ was done to compare the performance between our ωBCE method and the $SW + CF$ method proposed by Ma et al¹ according to 10 pairs samples. The p-values of local and global CNR

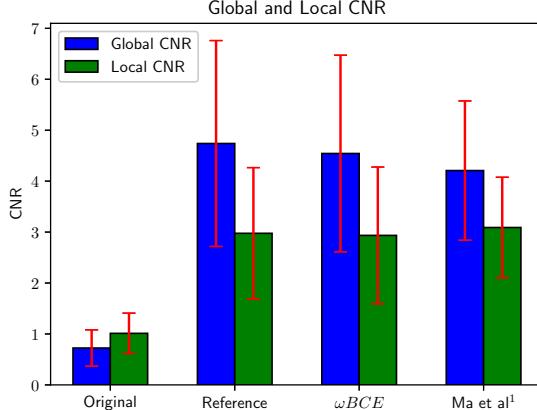


Figure 9. Average CNR of various methods

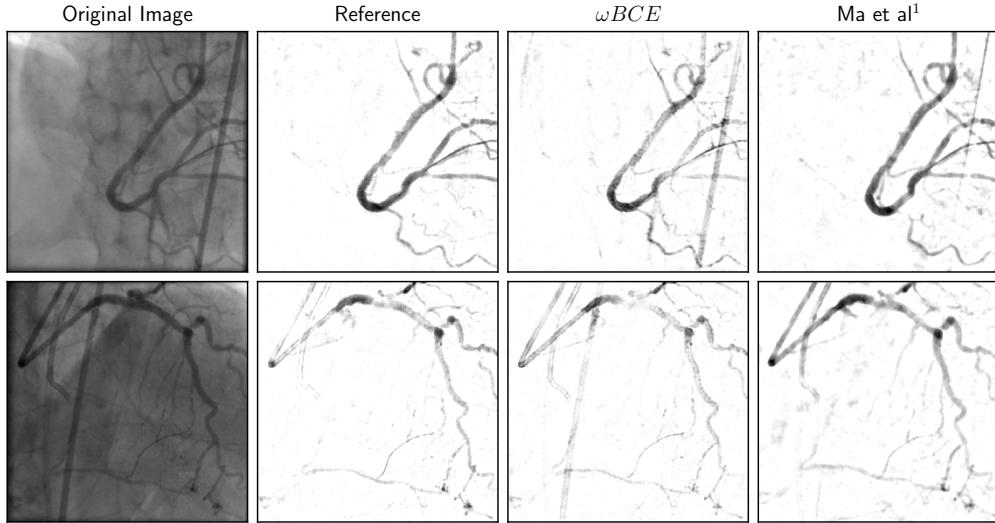


Figure 10. Comparison of our proposed method to the method of Ma et al¹ on two examples

are 0.185 and 0.262, respectively, so there is no statistically significant difference between the two methods in terms of local and global CNR.

4.3 Low contrast data

Channel 1 and channel 2 shown in Figure 4 enhance the vessel structure. We therefore utilized three-channel input to train, evaluate and test low contrast data (50% and 80%) in addition to two-channel input. The performance of three-channel input is better than that of two-channel input. The optimal hyper parameters combination for both 50% and 80% dataset based on ω BCE loss functions is shown in Table 7.

Figure 11 illustrates an prediction example using different contrast data based on the ω BCE method. The first, second and third row are from clinical angiogram, synthesized 80% contrast data and synthesized 50% contrast data, respectively. Columns from left to right show the original image, reference image, predictions of the optimal ω BCE method, respectively. The catheter in all the predictions was not totally removed compared to the reference image, which may be because the vessel and the catheter are similar in colour and structure, and it is difficult for the network to distinguish them. The static structures in the prediction increased as the

Table 6. Average CNR of various methods. (mean \pm standard deviation)

Method	Local CNR	Global CNR
ω BCE	2.935 ± 1.342	4.543 ± 1.930
Ma et al ¹	3.090 ± 0.987	4.208 ± 1.367

Table 7. The optimal hyper parameters combination for both 50% and 80% dataset of ω BCE loss function

Learning rate	Epoch No.	Architecture	Filter No. (f)	Dataset
1e-5	70	UNet9	16	3Ch

concentration of the contrast agent decreased, which may be also because the similarities in colour and structure between vessel and static structures increased. The predictions of 80% data are nearly the same as the predictions of clinical data, although the vessel in the synthesized 80% contrast data is more subtle than that in clinical data. The reason may be that the vessel color is still different from the colour of the most static structures in 80% data. For 50% contrast data, the vessel was enhanced.

As shown in [Figure 12](#) (left part) and [Table 8](#), both clinical angiogram and low contrast data (50% and 80%)

Table 8. Average CNR and MSSIM of different dataset, the numbers 1, 0.8 and 0.5 indicate clinical angiogram, synthesized 80% contrast and 50% contrast data, respectively. (mean \pm standard deviation)

Method	Local CNR	Global CNR	Local MSSIM	Global MSSIM
Reference	2.976 \pm 1.289	4.739 \pm 2.021	1	1
ω BCE1	2.935 \pm 1.342	4.543 \pm 1.930	0.862 \pm 0.058	0.878 \pm 0.041
ω BCE0.8	2.947 \pm 1.371	4.581 \pm 2.077	0.880 \pm 0.055	0.866 \pm 0.037
ω BCE0.5	3.072 \pm 1.472	4.620 \pm 2.112	0.828 \pm 0.060	0.797 \pm 0.041

achieved nearly the same global and local CNR as the reference image, while the prediction examples in [Figure 11](#) are different from the reference. The reason may be that the existence of the catheter and static structures in the predictions decreases μ_B and increases σ_B simultaneously, which makes CNR nearly unchanged. [Figure 12](#) (right part) and [Table 8](#) shows the MSSIM, the local and global MSSIM of both clinical data and 80% contrast data are nearly the same, but 50% data achieved slightly low MSSIM than the other two dataset.

Table 9. p-values between clinical angiogram, synthesized 80% contrast data and synthesized 50% contrast data using ω BCE loss function method, respectively, in terms of CNR and MSSIM.

Method 1	Method 2	Local CNR	Global CNR	Local MSSIM	Global MSSIM
ω BCE1	ω BCE0.8	0.182	0.211	0.0760	0.434
ω BCE1	ω BCE0.5	0.050	0.478	<0.001	<0.001

The Wilcoxon signed-rank test¹⁶ was also utilized to compare the performance between clinical and low contrast dataset, the results of which are shown in [Table 9](#). There is no statistically significant differences between clinical angiogram and synthesized 80% contrast data, although the CNR of synthesized 80% contrast data is lower than that of clinical angiogram.

[Table 9](#) also shows that there are statistically significant differences between clinical angiogram and synthesized 50% contrast data in terms of both local and global MSSIM. Clinical angiogram and synthesized 50% contrast data have no statistically significant difference in terms of CNR, although the CNR of synthesized 50% contrast data is much lower than that of clinical angiogram, which indicates that ω BCE method can enhance the vessel layer.

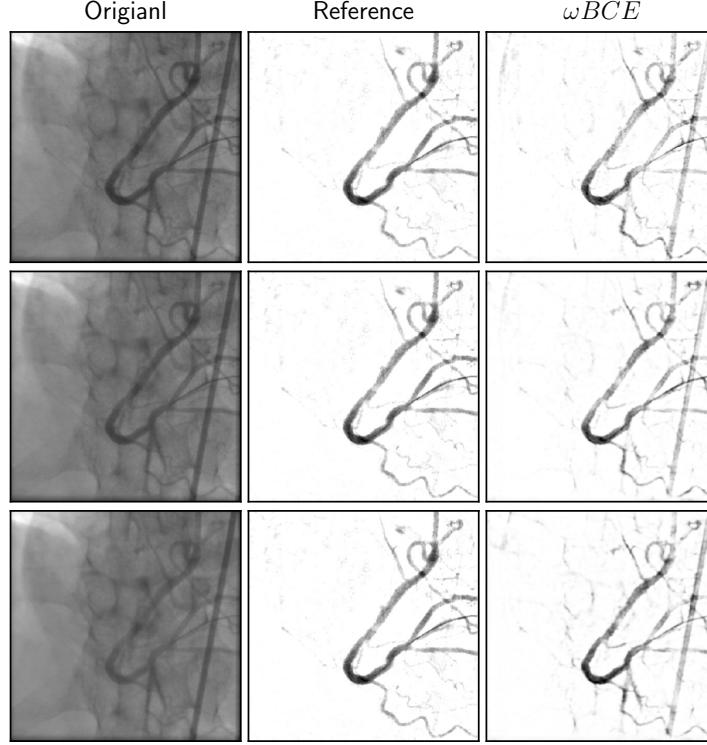


Figure 11. A prediction example of different data with respective optimal methods, 1st Row: Clinical angiogram; 2nd Row: synthesized 80% contrast data; 3rd Row: synthesized 50% contrast data; 3rd Column: ω BCE loss function.

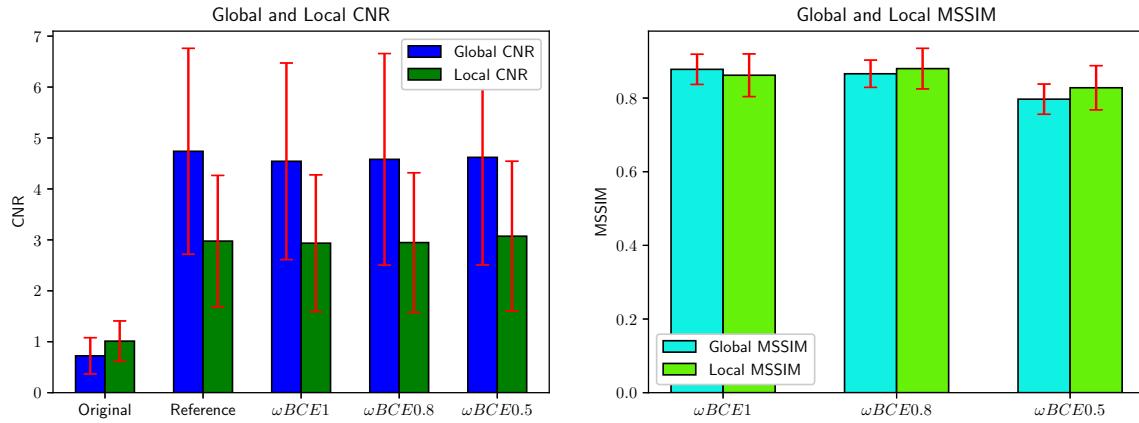


Figure 12. Average CNR and MSSIM of different dataset, the numbers 1, 0.8 and 0.5 indicate clinical angiogram, synthesized 80% contrast and 50% contrast data, respectively.

5. CONCLUSION

We have presented a data-driven method to separate vessel layer from cardiac interventional X-ray angiograms for vessel enhancement. The method uses a fully convolutional network to map the original X-ray image to a vessel layer image in which vessel structures have better visibility. We trained the network with automatically generated images of the vessel structure, the experimental results show that our proposed method is able to compute the

vessel layer and enhance vessel structures. The proposed method shows an improved CNR compared to the original X-ray images, and has a performance that is similar to the state-of-the-art method. As the proposed method has a processing rate of about 18 frames per second, it has potential for real-time clinical application. We also investigated the performance of our method on low contrast dataset and the performance on the 80% contrast dataset is nearly the same as the clinical angiograms, which indicates a potential to reduce the dose of contrast agent in coronary interventions.

ACKNOWLEDGMENTS

This work was supported by NWO (The Netherlands Organisation for Scientific Research), IMAGIC project under the iMIT program (grant number 12703). The authors also acknowledge SURF (Collaborative organisation for ICT in Dutch education and research) for providing computing resources on the Dutch national supercomputer.

REFERENCES

- [1] Ma, H., Hoogendoorn, A., Regar, E., Niessen, W. J., and van Walsum, T., “Automatic online layer separation for vessel enhancement in X-ray angiograms for percutaneous coronary interventions,” *Medical Image Analysis* **39**, 145–161 (2017).
- [2] Zhang, W., Ling, H., Prummer, S., Zhou, K. S., Ostermeier, M., and Comaniciu, D., “Coronary tree extraction using motion layer separation,” in [*International Conference on Medical Image Computing and Computer-Assisted Intervention*], 116–123, Springer (2009).
- [3] Zhu, Y., Prummer, S., Wang, P., Chen, T., Comaniciu, D., and Ostermeier, M., “Dynamic layer separation for coronary DSA and enhancement in fluoroscopic sequences,” in [*International Conference on Medical Image Computing and Computer-Assisted Intervention*], 877–884, Springer (2009).
- [4] Ma, H., Dibildox, G., Banerjee, J., Niessen, W., Schultz, C., Regar, E., and van Walsum, T., “Layer separation for vessel enhancement in interventional X-ray angiograms using morphological filtering and robust PCA,” in [*Workshop on Augmented Environments for Computer-Assisted Interventions*], 104–113, Springer (2015).
- [5] Tang, S., Wang, Y., and Chen, Y.-W., “Application of ICA to X-ray coronary digital subtraction angiography,” *Neurocomputing* **79**, 168–172 (2012).
- [6] Volpi, D., Sarhan, M. H., Ghotbi, R., Navab, N., Mateus, D., and Demirci, S., “Online tracking of interventional devices for endovascular aortic repair,” *International journal of computer assisted radiology and surgery* **10**(6), 773–781 (2015).
- [7] Feng, J., Xu, H., and Yan, S., “Online robust pca via stochastic optimization,” in [*Advances in Neural Information Processing Systems*], 404–412 (2013).
- [8] Long, J., Shelhamer, E., and Darrell, T., “Fully convolutional networks for semantic segmentation,” in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 3431–3440 (2015).
- [9] Noh, H., Hong, S., and Han, B., “Learning deconvolution network for semantic segmentation,” in [*Proceedings of the IEEE International Conference on Computer Vision*], 1520–1528 (2015).
- [10] Ronneberger, O., Fischer, P., and Brox, T., “U-net: Convolutional networks for biomedical image segmentation,” in [*International Conference on Medical Image Computing and Computer-Assisted Intervention*], 234–241, Springer (2015).
- [11] Wei, Y., Liang, X., Chen, Y., Shen, X., Cheng, M.-M., Feng, J., Zhao, Y., and Yan, S., “STC: A simple to complex framework for weakly-supervised semantic segmentation,” *IEEE transactions on pattern analysis and machine intelligence* (2016).
- [12] Patravali, J., Jain, S., and Chilamkurthy, S., “2D-3D fully convolutional neural networks for cardiac MR segmentation,” *arXiv preprint arXiv:1707.09813* (2017).
- [13] Wysowski, D. K. and Nourjah, P., “Deaths attributed to X-ray contrast media on US death certificates,” *American journal of roentgenology* **186**(3), 613–615 (2006).
- [14] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., “Image quality assessment: from error visibility to structural similarity,” *IEEE transactions on image processing* **13**(4), 600–612 (2004).

- [15] Kingma, D. and Ba, J., “ADAM: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980* (2014).
- [16] Siegal, S., [*Nonparametric statistics for the behavioral sciences*], McGraw-hill (1956).