Convolutional neural network-based image enhancement for x-ray percutaneous coronary intervention

Marco Pavoni
Yongjun Chang
Sang-ho Park
Orjan Smedby

Convolutional neural network-based image enhancement for x-ray percutaneous coronary intervention

Marco Pavoni,^a Yongjun Chang,^{b,*} Sang-ho Park,^c and Orjan Smedby^b

^aPolitecnico di Torino, Corso Duca degli Abruzzi, Torino, Italy

^bKTH Royal Institute of Technology MTH, Halsövägen, Huddinge, Sweden

^cSoonchunhyang University, Cheonan Hospital, Division of Cardiology, Department of Internal Medicine, Cheonan, Republic of Korea

Abstract. Percutaneous coronary intervention (PCI) uses x-ray images, which may give high radiation dose and high concentrations of contrast media, leading to the risk of radiation-induced injury and nephropathy. These drawbacks can be reduced by using lower doses of x-rays and contrast media, with the disadvantage of noisier PCI images with less contrast. Vessel-edge-preserving convolutional neural networks (CNN) were designed to denoise simulated low x-ray dose PCI images, created by adding artificial noise to high-dose images. Objective functions of the designed CNNs have been optimized to achieve an edge-preserving effect of vessel walls, and the results of the proposed objective functions were evaluated qualitatively and quantitatively. Finally, the proposed CNN-based method was compared with two state-of-the-art denoising methods: K-SVD and block-matching and 3D filtering. The results showed promising performance of the proposed CNN-based method for PCI image enhancement with interesting capabilities of CNNs for real-time denoising and contrast enhancement tasks. © 2018 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.5.2.024006](https://doi.org/10.1117/1.JMI.5.2.024006)]

Keywords: convolutional neural networks; deep learning; image denoising; edge-preserving image enhancement; loss function.

Paper 18035R received Feb. 26, 2018; accepted for publication Jun. 4, 2018; published online Jun. 28, 2018.

1 Introduction

Worldwide, cardiovascular diseases are the major cause of death. More specifically, atherosclerotic cardiovascular diseases (ACDs) are the most important cause of mortality. Among all the ACDs, coronary heart disease (CHD) is one of the most widespread, with 370,000 deceased annually for CHDs in the United States.¹ The most common medical procedures for CHD treatment are percutaneous coronary intervention (PCI) and coronary artery bypass grafting. Both procedures have the aim of restoring an adequate blood supply in the cardiac tissue. A PCI procedure uses x-ray images showing the coronary vessels to locate the stenosis, guide the catheters, and evaluate the outcome. Thus, an angiographic system and iodinate radiographic contrast agent (IRCA) are required to produce images and make visible the vessels under x-ray irradiation, respectively. During the PCI procedure, the angiographic system can be employed for fluoroscopy or fluorography. The former is a low-dose mode with the aim of locating catheters inside the body. The latter, called cine-mode, is a high-dose mode with the diagnostic aim of detecting atherosclerotic vessels and evaluating the final outcome after the procedure, and it requires images with higher quality than fluoroscopy. To perform PCI correctly, the procedure requires fluoroscopy irradiation for prolonged time and several cine-mode angiographic runs, combined with high IRCA dose. This may have two main side effects. The first is the risk of radiation-induced skin injuries, with also the possible development of radiation-induced cancer of chest organs.^{2,3} The other is contrast-induced nephropathy

due to the high concentration of contrast media injected into the patient.⁴ These side effects can be reduced with lower doses of both x-rays and IRCA, but the resulting images suffer from higher noise level and weaker contrast compared with PCI images of high-dose x-rays and IRCA.

The overall goal of our work is image enhancement of PCI images with lower doses of both x-rays (fluoroscopy mode) and IRCA by applying a deep learning-based method⁵ to obtain images more similar to high-dose PCI images (cine-mode with high contrast media concentration). In this way, patients' quality of life would be improved since lower doses of x-rays and contrast media could be employed during the operation. In addition, the risk of exposure to high-dose x-rays for medical staff would also decrease. We will simply refer to low- and high-dose in both x-ray radiation and IRCA as low- and high-dose, respectively, from now on. It is fundamental that the designed methods are fast in terms of computational time needed for image processing in order for the methods to be applied to a real-time intervention procedure.

Convolutional neural networks (CNNs),⁶ a specific type of artificial neural network widely used with image-based datasets, were employed to achieve the objectives. A preliminary study about simulated low x-ray dose image denoising was introduced⁷ by creating simulated images containing artificial additive noise on high-dose PCI images. Poisson, Gaussian, and the mixture of both noises were investigated at different noise levels to evaluate the capabilities of designed CNNs for denoising simulated low-dose PCI images.

*Address all correspondence to: Yongjun Chang, E-mail: yongjun@kth.se

The specific contributions of this paper may be summarized as follows: first, compared with the results in the previous study,⁷ the obtained results are improved by designing new loss functions for combining denoising with an edge-preserving effect on blood vessels. In addition, our proposed methods are compared with two state-of-the-art denoising methods to demonstrate the effectiveness of the proposed ones in terms of image quality and computation.

The rest of the paper is organized as follows: Section 2 overviews the background about image enhancement methods based on deep learning. In Sec. 3, the available data and methods are described. In Sec. 4, the performed experiments and results are presented besides a comparison with two state-of-the-art denoising methods. Conclusions are stated in Sec. 5.

2 Background

Image enhancement is a wide concept and several procedures can be carried out to improve image quality. Super-resolution (SR) and denoising are two tasks in medical applications for which deep learning-based methods have been designed and have outperformed existing state-of-the-art methods for the same tasks.^{8–13}

2.1 Super-Resolution Task with CNNs

The SR approach is an increase of spatial resolution to get a final image with higher quality. Bahrami et al.⁸ proposed a CNN for the reconstruction of 7T-like MRI images starting from 3T brain MRI images. The performance showed higher PSNR values in comparison with other methods, such as histogram matching. In the CNN architecture, zero-padding was not used during convolution and a three-dimensional (3-D) patch was used as input of the network. This led to a patch that shrinks along the net, to have only one value as CNN output. This value substitutes the old center voxel value of the patch.

Lee et al.⁹ proposed a CNN for increasing the spatial resolution of x-ray cardiac images. The designed method was compared with other SR methods such as linear, cubic spline, and multiatlas patch match (MAPM). In terms of PSNR and structural similarity (SSIM), the deep learning-based method outperforms the others. To get an output image with higher spatial resolution with CNN, the authors employed a deconvolution layer before the convolutional layer sequence, where an upscaling operation with zero-padding was performed in the deconvolution layer. This kind of layer is strictly needed for having an output image with a higher number of pixels or voxels, resulting in higher spatial resolution. This is not useful for denoising, because its main goal is noise reduction by preserving the original image size. However, upsampling can be used to combine denoising and SR so that denoised images with higher spatial resolution are obtained. However, this option requires ground truth images with higher resolution, which were not available in our study.

2.2 CNN-Based Image Denoising

Image denoising for x-ray image sequences with low radiation dose is one of the proposed study objectives. Several CNN designs were applied to denoising tasks for both natural images^{10–12} and medical data.⁸

Wang et al.¹⁰ proposed a CNN architecture for natural image denoising, with a linear structure composed of a sequence of four convolutional layers without pooling layers, whereas a

rectified linear unit (ReLU) was used as the activation function. The second to last layer is a fully connected layer defined as a convolution layer with only one kernel, so only one image is obtained. Finally, a sigmoid activation function is applied on the image to obtain the denoised image. Using a fully connected layer following the convolutional layers sequence is reasonable and similar to its common use for classification tasks, where in general a vector containing class information is needed as CNN output. With this architecture, the designed method outperforms K-SVD and field of expert algorithms at high noise variance in terms of PSNR. Moreover, the needed image processing time is extremely low compared with the other two methods.¹⁰

A study on low-dose CT image denoising was proposed by Chen et al.,¹³ simulated by adding Poisson noise to normal-dose CT images. The CNN architecture includes three convolutional layers without pooling and employed ReLU as an activation function. The input of the network was composed by image patches to have a bigger database for CNN training. The proposed method, compared with K-SVD and block-matching and 3D filtering (BM3D) state-of-the-art methods, outperforms them in terms both of SSIM and PSNR.¹³ Moreover, previous studies^{10,13} suggest data augmentation for increasing database size by applying image transformation such as translation, rotation, and warping on the original image, which may be useful when a small database is available.

3 Data and Methods

3.1 Data

The dataset used for the study was composed of nine high-dose and nine low-dose frame sequences from nine different patients treated with PCI. All the data were provided by Soonchunhyang University Cheonan Hospital, South Korea. All the data were selected from patients who already had PCI by the clinician at the hospital (one of the authors). During the procedure, cine-mode with high-dose x-ray or fluoroscopic mode with low-dose x-ray were taken by the decision of the clinician depending on the need of high or low quality of images, with high- and low-dose contrast agent, considering patients' safety including their renal function. All the x-ray dose and the dose of contrast agent were cautiously selected and determined by the clinician within the acceptable range during the procedure, with consent forms for PCI signed by patients. As a pilot study to investigate the possibility of the proposed CNN-based method, this study was approved by Institutional Review Board of Soonchunhyang University Cheonan Hospital.

A sequence of one of these patients studied the left anterior descending artery, a coronary vessel supplying most of the interventricular septum. The sequences of the other eight patients depicted the right coronary artery, a major vessel supplying blood to the myocardium. For two patients, the high-dose sequences were obtained with cine-mode irradiation and IRCA injection with a concentration of 5 mL. For the remaining patients, 6-mL contrast media concentration was used for the high-dose sequences. In the low-dose sequences, fluoroscopy mode combined with 3-mL IRCA injection was used. All the frames have a size of 512×512 pixels coded on 8 bits (256-gray levels).

The whole frames were not employed as input to the CNN, but each of them was divided into overlapping patches, used later as input to the network to get the enhanced patch as CNN output. Finally, the whole enhanced frame was

reconstructed from its overlapping patches by averaging. There were two main reasons for this choice. First, due to the meager available database, a patch-based approach allows having a larger dataset as input, which is crucial for the training phase. Second, an image can be described with local structures, detectable with a small patch if its size is correctly chosen. Among the nine-patient image sequences, 676 frames were available, containing 152,100 patches with a size of 64×64 pixels and 33,124 patches with a size of 128×128 pixels.

3.2 Methods for Simulated Low-Dose PCI Image Denoising with Edge-Preserving Effect

Regarding the CNN architecture, the same neural network structure as that employed in previous work⁷ was used in this study. It is composed of a sequence of four convolutional layers, where each one includes a convolution layer followed by a detector layer with ReLU as activation function. The last CNN layer is the loss function layer. Kernel size and number of kernels for each convolution layer are 3×3 and 64, respectively. Input patch size is 64×64 pixels when mean square error (MSE) is used as loss function, whereas 128×128 pixels patch size is employed when SSIM is the chosen loss function for the CNN training. All these hyperparameters were chosen based on an extensive set of trial and error experiments. More information about the CNN structure designing and hyperparameters selection is described in the previous study.⁷ The CNN architecture employed in previous work should be suitable for this work because input data for the network are the same for both MSE- and SSIM-based CNN in terms of the patch size and the number of patches. A schematic representation of the CNN architecture employed is shown in Fig. 1. In Fig. 1, the output of the CNN is compared with the desired patch (the green one), with which the loss function is computed. Starting from the loss function computation, the networks weights and biases are updated relying on the gradient descent algorithm as backpropagation optimizer.

The preliminary results⁷ showed that the designed CNNs are able to reduce the noise level of patches and reconstructed frames, with an increase of SSIM index. However, a weak blurring effect occurs on major vessel walls resulting from denoising, which reduces the contrast of the structures of interest. For this reason, further improvement was performed to preserve the vessel edges and their contrast as much as possible, coupled with a general noise intensity reduction. To preserve edges while reducing image noise levels, both MSE- and SSIM-

Fig. 1 Schematic description of the CNN architecture considered in this study.

based loss functions in the previous work⁷ were modified. This was needed because the standard definition of these two loss functions may not be able to preserve efficiently the structures of interest in PCI images, such as vessels edges, with the risk of blurring those details. Their new definitions are described in the following paragraphs.

3.2.1 Edge-preserving MSE loss function

The original MSE-based loss function that computes the mean squared error between the output batch and the desired batch is defined as

$$L(X, Y) = \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^M (x_{i,k} - y_{i,k})^2, \quad (1)$$

where N is the number of patches contained in the training batch, M is the number of pixels contained in one patch, X is the denoised patch with $x_{i,k}$ as k 'th pixel of the i 'th output patch, and Y is the clean patch with $y_{i,k}$ as k 'th pixel of the i 'th desired patch. Since the available database is not large, N was chosen equal to 1 to have a big enough number of CNN weights updates within the same epoch.

This loss function considers only a squared error between an output image and the desired one, without taking into account the structures contained in the image, such as vessels. The MSE is reduced uniformly on the image, whereas it would be preferable if the error decreased more where there are vessels and their edges, so the output structures could preserve their details better. This aim could be achieved by coupling the general MSE with a weighted MSE, where the weights depend upon the edges and discontinuities inside the images are, such as the vessel walls. The weighted MSE represents the regularization term, and it has an edge-preserving function. The general MSE is called data term and it has to be balanced with the regularization term. The new loss function structure is the following:

$$L(X, Y) = \text{MSE}(X, Y) + \lambda \times \text{MSE}(X, Y) \circ \Gamma(X), \quad (2)$$

where the first addend is the data term whereas the second addend is the regularization term; $\Gamma(X)$ is a function containing the weights for the MSE with which the Hadamard product is computed; λ parameter determines the importance of the edge-preserving effect in the whole loss function computation.

The gradient magnitude of the output image was used as $\Gamma(X)$ function since this operator is able to detect image discontinuities, such as edges. It is defined as follows:

$$\Gamma(X) = \sqrt{X_x^2 + X_y^2}, \quad (3)$$

where X_x and X_y are the x -wise and y -wise first-order derivative of the output image X , respectively, computed by convolving X with the corresponding 3×3 Sobel operator matrices. For performing the kernel operation, we chose the Sobel operator considering complexity and noise sensitivity. Because processing time is important in a real-time PCI procedure, we only considered low-complexity filters including Roberts and Laplacian operators.¹⁴ Among those operators, Sobel operator produced edge thickness values suitable for MSE optimization while the other operators produced too narrow edges, making the MSE optimization difficult as well as increasing sensitivity to noise.

Fig. 2 (a) Original frame and (b) its gradient magnitude image.

The gradient magnitude is able to show the edges of an image, such as vessel walls. By multiplying, element-by-element, the gradient magnitude with the MSE, the error increases where the gradient magnitude is higher, i.e., where vessel edges are located. In this way, the loss function optimization, which is encouraged where its value is higher, is more focused where there are edges of image structures. The gradient magnitude of a sample frame appears in Fig. 2.

However, this regularization term has to be balanced with the data term that considers the general MSE to denoise the input image. This balance is achieved with the parameter λ , which determines the importance of the regularization term of the cost function in Eq. (2). High λ values result in images of high edge-preserving effects while leading to low PSNR and SSIM values over nonedge pixels in images. Thus, appropriate λ values are investigated to balance the data term and the regularization term, resulting in an overall patch denoising as well as edge-preserving effect. Table 1 shows the average PSNR and SSIM values of the whole patches depending on different values of λ ranging from 0.5 to 4, where input noisy patches were based on Poisson noise as in Sec. 3.2.3. Based on Table 1, the CNN performed the best with λ equal to 1 in terms of both PSNR and SSIM. From visual inspection of an image patch within a sampled frame, the overall denoising effect and the edge-preserving effect were noticeable with λ equal to 1,

which agrees with the quantitative results while no other value of λ gives an obviously visually superior result in Fig. 3. Thus, the new edge-preserving MSE-based loss function is defined as follows:

$$L(X, Y) = \frac{1}{N} \sum_{i=1}^N \sum_{k=1}^M [(x_{i,k} - y_{i,k})^2 + \lambda \cdot (x_{i,k} - y_{i,k})^2 \cdot \Gamma(x_{i,k})], \quad (4)$$

where $\Gamma(x_{i,k})$ is the gradient magnitude value at k 'th pixel position of the i 'th patch and the value of λ is set to 1.

As an output image Y generated from the CNN with an input image X becomes close to the desired output image in our optimization process using stochastic gradient descent (SGD) with a learning rate small enough, the loss function in Eq. (2) also converges because λ is a constant and $\Gamma(X)$ computed from the input image X is finite. In terms of convergence to the global minimum, SGD is known to converge to the band of local minima close to the global minimum as the network size increases.¹⁵ In our study, we assessed convergence and overfitting by watching the trends of training and validation errors during the network training. In training the edge-preserving MSE CNN as an example, convergence was attained

Table 1 PSNR and SSIM values for edge-preserving MSE CNN, depending on λ ranging from 0.5 to 4 (mean values and standard deviations are computed over the whole dataset).

λ	PSNR (dB): mean $\pm$ std			SSIM: mean $\pm$ std		
	Noisy	Denoised	Gain	Noisy	Denoised	Gain
0.5	31.61 $\pm$ 0.38	35.87 $\pm$ 0.38	4.26	0.703 $\pm$ 0.027	0.870 $\pm$ 0.011	0.167
1	31.61 $\pm$ 0.38	36.42 $\pm$ 0.40	4.81	0.703 $\pm$ 0.027	0.875 $\pm$ 0.011	0.172
2	31.61 $\pm$ 0.38	35.09 $\pm$ 0.40	3.48	0.703 $\pm$ 0.027	0.834 $\pm$ 0.013	0.131
3	31.61 $\pm$ 0.38	35.31 $\pm$ 0.41	3.70	0.703 $\pm$ 0.027	0.845 $\pm$ 0.013	0.142
4	31.61 $\pm$ 0.38	35.05 $\pm$ 0.41	3.44	0.703 $\pm$ 0.027	0.839 $\pm$ 0.013	0.136

Fig. 3 Visual inspection of vessel details in a sampled frame denoised by edge-preserving MSE CNN depending on different λ values ranging from 0.5 to 4.

Fig. 4 Training (blue) and validation (red) errors over the number of weights updates (training steps) for edge-preserving MSE CNN based on Eq. (4).

within 600 steps without overfitting, for both training and validation in Fig. 4.

3.2.2 Edge- and contrast-preserving SSIM loss function

SSIM-based loss function was chosen also because of the consistency of SSIM index with human vision perception. Indeed, this parameter considers three fundamental visual elements: the luminance, contrast, and structure of the image. The original SSIM-based loss function definition is defined as follows:¹⁶

$$L(X, Y) = 1 - \text{SSIM}(X, Y), \quad (5)$$

where $\text{SSIM}(X, Y)$ is the SSIM between the output image X and the desired one Y , defined as

$$\text{SSIM}(X, Y) = \frac{(2\mu_X\mu_Y + c_1)(2\sigma_{XY} + c_2)}{(\mu_X^2 + \mu_Y^2 + c_1)(\sigma_X^2 + \sigma_Y^2 + c_2)}, \quad (6)$$

where μ_X and μ_Y are the average of X and Y , respectively; σ_X and σ_Y are the variance of X and Y , respectively; σ_{XY} is the covariance of X and Y ; $c_1 = L \times k_1$ and $c_2 = L \times k_2$ constants where L is the dynamic range of gray levels (255) whereas k_1 and k_2 are equal to 0.01 and 0.03, respectively.

The general definition of SSIM is the product of three expressions: luminance, contrast, and structure. This expression can be defined as follows:

$$\text{SSIM}(X, Y) = \text{lum}(X, Y)^\alpha \cdot \text{cont}(X, Y)^\beta \cdot \text{struct}(X, Y)^\gamma, \quad (7)$$

where lum , cont , and struct are the three functions depending on mean, variance, covariance, and standard deviation of X and Y ,¹⁶ representing the luminance, contrast, and structure comparison between X and Y , respectively; α , β , and γ are the three exponents determining the importance of luminance, contrast, and structure for SSIM computation, respectively.

If α , β , and γ are all equal to one, the SSIM formulation becomes Eq. (6). To preserve better both contrast and edges, it is interesting to use different weights for the three functions composing the SSIM index. More specifically, an increase of β and γ parameters may help to preserve more the image contrast and the

structure, both useful for an edge-preserving effect. Several experiments were performed with different values of these three parameters, and the resulting frames have been compared with those obtained with standard SSIM-based loss function ($\alpha = \beta = \gamma = 1$). The difference between two frames obtained with these two different methods helps to visualize if there is an increase of contrast with edge-preserving effect. Different combinations of these three parameters were tested to obtain an edge-preserving effect coupled with a denoising effect.

3.2.3 Fluoroscopy noise level simulation

Starting from the noise model,⁷ a more precise simulation of fluoroscopy photon noise was performed and then used in the experiments. The evaluation was done with a visual analysis of actual fluoroscopy PCI images of low dose, by focusing on the image background noise pattern. Then, this pattern was simulated by adding Poisson noise only one time on cine-mode images and by comparing the resulting image with the low-dose one. By analyzing the two noise patterns, the original one seemed weakly blurred when compared with the artificial one. For this reason, the original fluoroscopy noise level and pattern was simulated by applying a convolution operation on the artificial noisy image with a 3×3 pixel Gaussian kernel with variance equal to 0.25. In this way, a very little blurring effect acting on the added Poisson noise results in a noise pattern slightly more similar to the fluoroscopy one, even if the simulated result looks still noisier. To avoid a blurring effect that would affect the image structures of interest, the variance of the Gaussian window and its size cannot be increased too much. A comparison between an original low-dose image, a high-dose image, and the simulated low x-ray dose image shown in Fig. 5.

4 Experimental Results

4.1 Results for Simulated Low-Dose PCI Image Denoising

4.1.1 Visual inspection

The main experiment with simulated low-dose PCI images included training of CNNs with the optimized loss functions,

Fig. 5 Comparison of (a) the actual low-dose image, (b) the simulated low-dose image by adding Poisson noise on the high-dose image followed by convolution with a 3×3 pixel Gaussian kernel, and (c) the actual high-dose image.

Fig. 6 Denoising results of original SSIM CNN ($\alpha = \beta = \gamma = 1$). (a) Noisy, (b) denoised, and (c) desired frame.

followed by visual evaluation of the edge-preserving effect. For both edge-preserving and original loss functions, visual results appear in Figs. 6–8, showing a sample frame including a part of coronary arteries from the sequence for validation. The input images are generated by adding Poisson noise on high-dose cine-mode images to simulate low-dose fluoroscopy mode images described in Sec. 3.2.3. It can be noticed that compared with the result carried out by the CNN with the standard SSIM-based loss function (original SSIM CNN) with Eqs. (5) and (6), the CNN with the edge-preserving SSIM in Eq. (7) with increasing values of β and γ set to 30 (EP SSIM CNN) and the CNN with the edge-preserving MSE-based loss functions in Eq. (4) (EP MSE CNN) led to edge-preserving effects on vessel edges.

The edge-preserving effect is more visible in Fig. 9. The images in Fig. 9 represent the differences between the resulting image with the original loss function and the image outcome with the optimization of EP SSIM [Fig. 9(a)] and EP MSE [Fig. 9(b)]. In these images, a positive area means that the region

is darker with the optimized loss function, whereas a negative area is brighter with the optimized loss function. So, increased vessel edge enhancement occurs when the inside of the vessel boundary is positive, whereas the outside of the vessel boundary is negative, in the difference image. For the optimized EP SSIM CNN result, the contrast and the edge-preserving effect act on the major part of vessels and the effect is stronger and more homogeneous compared with the optimized EP MSE CNN result. With the optimized EP MSE-based loss function, the edge-preserving effect acts only on some vessel structure, as can be noticed when comparing the two images. For these reasons, the EP SSIM-based loss function leads to a stronger edge-preserving and contrast-enhancing effect compared with the EP MSE-based loss function.

4.1.2 Quantitative evaluation

Quantitative results were computed to evaluate and compare the edge-preserving effect and the overall image enhancement.

Fig. 7 Denoising result of EP SSIM CNN ($\alpha = 1$, $\beta = 30$, and $\gamma = 30$). (a) Noisy, (b) denoised, and (c) desired frame.

Fig. 8 Denoising result of EP MSE. (a) Noisy, (b) denoised, and (c) desired frame.

Fig. 9 Difference image between the denoised frame with original SSIM CNN and the denoised frame with EP CNNs. (a) EP SSIM CNN and (b) EP MSE CNN.

Local standard deviation (LSTD) was considered to assess the edge-preserving effect quantitatively. Standard deviation of image luminance is suitable for contrast measurements, as showed by Moulden et al.¹⁷ If this metric is computed locally, local edge-preserving and contrast enhancement can be assessed. This metric was computed in this study by taking into account the pixels contained in a 3×3 window that slides within the whole output denoised image, resulting in an LSTD value at the center of the window. Figure 10 shows the maps representing LSTD values of a sample frame for original SSIM CNN and EP CNNs. From Fig. 10, it can be noticed that the LSTD of EP CNNs have higher values than original SSIM CNN at the edges of coronary arteries. This means that there is a stronger abrupt variation of image luminance,

resulting in a stronger vessel edge contrast. Moreover, the high LSTD regions corresponding to the vessel edges are narrower with the optimized loss functions than with the indicating a weaker blurring effect on that area. This evaluation is also supported by the maximum value of the LSTD in a frame, which is located on a vessel edge. The mean and standard deviation of the maximum LSTD values for each CNN scheme were computed over all the available frames. Its value for the original SSIM-based CNN is 25.46, whereas for the edge-preserving MSE and SSIM-based CNNs are 31.19 and 29.76, respectively, with a positive gain of 5.73 and 4.29, respectively.

Finally, a frame enhancement metric was defined as follows to combine the denoising effect with the overall frame edge-preserving effect:

Fig. 10 Comparison of LSTD maps among the CNN with original SSIM and EP CNNs. (a) LSTD of original SSIM CNN, (b) LSTD of EP SSIM CNN, and (c) LSTD of EP MSE CNN. The scale is the same for all the images.

Table 2 Frame enhancement comparison among original SSIM CNN and two edge-preserving CNNs, which are EP SSIM CNN and EP MSE CNN.

	Original SSIM	EP SSIM CNN	EP MSE CNN
Frame enhancement in Ref. 11	944 ± 235	1071 ± 297	1212 ± 369

$$\text{frame enhancement} = \text{PSNR}_{\text{frame}} \times \max(\text{LSTD}_{\text{frame}}), \quad (8)$$

where $\text{PSNR}_{\text{frame}}$ is the PSNR between the i 'th output frame and the corresponding desired frame and $\max(\text{LSTD}_{\text{frame}})$ is the highest value of the LSTD for the current frame. The mean and standard deviation of the frame enhancement metric computed over all the available frames for each CNN scheme are shown in Table 2 and in terms of this metric, the edge-preserving CNNs outperform the original SSIM CNN in accordance with the visual inspection.

4.2 Comparison with State-of-the-Art Denoising Methods

To compare the proposed CNN performances with state-of-the-art denoising methods in terms of denoising effect and image enhancement, a dictionary denoising method based on K-

SVD^{18,19} and the BM3D method²⁰ was used. This comparison was addressed qualitatively and quantitatively. A visual inspection of a representative frame before and after the image processing was included for a subjective discussion. Moreover, PSNR and SSIM were used as quantitative metrics for evaluating the denoising and image-enhancement effects.

Since these two state-of-the-art denoising methods are mostly designed for removing Gaussian noise from a target image, the quantitative analysis was performed by adding Gaussian noise on high-dose frames. Then, these noisy frames were used as input for each denoising method and the resulting denoised images were evaluated. Then, the mean and standard deviation for each method were computed among all the frames of the available patients for performance comparison in denoising.

The K-SVD-based method relies on a dictionary of 64 atoms, with a size of 8×8 pixels. These atoms were trained with K-SVD algorithms for 10 iterations using as training sets 110,592 noisy patches of 8×8 size extracted from representative frames of each available patient at three different Gaussian noise levels. All the other settings for this denoising method were the same proposed by previous work.¹⁸ BM3D method is based on a collaborative filtering of 3-D groups containing similar 2-D image fragments called blocks. The denoising settings were the same employed in previous work.²⁰ Both for K-SVD and BM3D, the available MATLAB toolboxes were used for implementing the denoising processing.^{18–20} These two state-of-the-art denoising methods were compared with

the different CNN-based methods (original MSE CNN, EP MSE CNN, original SSIM CNN, and EP SSIM CNN), at low, medium, and high Gaussian noise level, with standard deviation σ equal to 8, 12, and 16, respectively.

4.2.1 Visual inspection

Visual inspection of denoising results with three methods of BM3D, K-SVD, and CNNs-based methods was performed. The results of visual inspection for an example frame magnified details with low Gaussian noise are shown in Fig. 11. It can be noticed that BM3D, K-SVD, and CNN-based methods are able to

reduce the noise level, resulting in the output images cleaner than the input frame. Visually, the BM3D method produced the cleanest image if compared with the others. The K-SVD method produced denoised images with vessel edges that look more blurred than the proposed method. The EP SSIM CNN-based method is also comparable with BM3D in terms of edge-preserving effect. From this visual inspection, it can be assessed that EP SSIM CNN produces images that are not cleaner than BM3D but comparable with it. With the proposed EP SSIM CNN method, the vessel edges are better preserved compared with K-SVD denoising results. The other CNNs show no better results than EP SSIM CNN from a visual inspection point of view.

Fig. 11 Denoising results of magnified coronary details with the proposed CNN-based methods, BM3D and K-SVD.

Table 3 Quantitative results for low, medium, and high Gaussian noise experiment in terms of two image enhancement metrics: PSNR and SSIM. BM3D and K-SVD are compared with the designed CNNs. The gain is defined as the reconstructed output frame metric value minus the input frame metric value.

Denoising method	PSNR gain (mean $\pm$ standard deviation)			SSIM gain (mean $\pm$ standard deviation)		
	Noise level			Noise level		
	Low	Medium	High	Low	Medium	High
Original MSE CNN	7.17 ± 0.41	9.13 ± 0.45	10.55 ± 0.48	0.269 ± 0.017	0.424 ± 0.020	0.524 ± 0.019
EP MSE CNN	7.04 ± 0.37	8.95 ± 0.38	10.30 ± 0.42	0.266 ± 0.016	0.419 ± 0.019	0.513 ± 0.019
Original SSIM CNN	6.64 ± 0.32	5.10 ± 0.37	8.78 ± 0.28	0.273 ± 0.013	0.428 ± 0.011	0.530 ± 0.019
EP SSIM CNN	6.87 ± 0.26	8.16 ± 0.24	9.83 ± 0.26	0.275 ± 0.016	0.428 ± 0.019	0.528 ± 0.018
K-SVD	6.33 ± 0.37	8.65 ± 0.44	10.46 ± 0.51	0.238 ± 0.018	0.399 ± 0.020	0.507 ± 0.021
BM3D	7.27 ± 0.39	9.40 ± 0.38	11.08 ± 0.38	0.262 ± 0.018	0.420 ± 0.021	0.528 ± 0.020

4.2.2 Quantitative results

The results of comparison at low ($\sigma = 8$), medium ($\sigma = 12$), and high Gaussian noise level ($\sigma = 16$) were evaluated in terms of PSNR and SSIM as summarized in Table 3. In terms of PSNR, original MSE CNN among the proposed CNN-based methods outperformed K-SVD for all the noise levels. However, it was not true for comparison with the BM3D method. The difference ranged between 0.10 and 0.53 dB. In terms of SSIM, all the CNN-based methods outperformed K-SVD in all the experimental conditions. Moreover, the SSIM-based CNN methods (original SSIM CNN and EP SSIM CNN) outperformed

BM3D in all the experimental conditions. It is interesting that the proposed CNNs always showed a lower standard deviation compared with the other methods in SSIM gain, suggesting a higher stability of performance in terms of SSIM than K-SVD and BM3D.

Another important consideration in this study is the computational time needed for processing an image. Since this study would require an image enhancement method able to work in an actual PCI procedure, the most suitable method has to guarantee low-computational time for image reconstruction as well as good image-enhancement effects including denoising and contrast increase. For this reason, we measured the processing time for reconstructing frames in seconds. We ran 100 denoising test experiments at different noise levels. As a result, the proposed CNN-methods are clearly faster than BM3D and K-SVD, with the average frame processing time equal to 2.08 s, whereas the other methods perform a frame denoising within 5.02 and 14.26 s on average, respectively, as shown in Table 4.

Table 4 Computational time per frame (mean $\pm$ standard deviation).

	BM3D	K-SVD	CNN
Frame processing time (s)	5.02 ± 0.13	14.26 ± 8.25	2.08 ± 0.12

Fig. 12 Preliminary result of image enhancement with actual low-dose PCI images. (a) The network input frame, (b) the output frame, and (c) the high-dose reference frame. The SSIM indexes with respect to the reference high-dose frame are placed on the top of the corresponding frames.

5 Discussion

In this study, image enhancement was investigated with CNN-based deep learning for simulated low-dose PCI images to restore the images as closely as possible to the corresponding high-dose images. For simulated fluoroscopy images, vessel contrast enhancement was attempted by designing edge-preserving loss functions, as well as denoising at different Poisson noise levels. The experimental results showed that CNNs using edge-preserving optimized loss functions can reduce the noise level and increase the contrast-enhancing effect on vessel edges. In PSNR, original MSE CNN has performance comparable with BM3D and better than K-SVD methods. In terms of SSIM, however, original SSIM CNN and EP SSIM CNN outperform K-SVD and BM3D. In addition, the proposed methods reduce the computation time for processing a frame by 85% and 59% compared with the K-SVD and BM3D methods, respectively. This processing time is important for future applications dealing with actual PCI procedures in real-time intervention procedures.

There is one main limitation in the study: the number of available datasets. Nine patient sequences may be considered a meager dataset for neural networks training, which generally improves the performance using a large amount of data as input. This motivated an overlapping-patch-based approach to increase the amount of data for the network, composed of thousands of patches instead of hundreds of frames.

Even though the current study focuses on denoising and edge-preserving image enhancement on simulated data, the work will be extended to image enhancement for actual low-dose (in radiation and contrast agent) images. In this case, the lack of spatial and temporal correspondence between low-dose and high-dose sequences from the same patient will be a challenge for which spatial and temporal frame registration will be attempted to get high-dose ground truth data corresponding to actual low-dose data for training. A preliminary result of such an experiment is shown in Fig. 12.

In conclusion, CNN-based denoising and edge preservation is promising for near-real-time enhancement of images during PCI procedures.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

References

1. S. Barquera et al., "Global overview of the epidemiology of atherosclerotic cardiovascular disease," *Arch. Med. Res.* **46**(5), 328–338 (2015).
2. M.-O. Bernier et al., "Patient cumulative radiation exposure in interventional cardiology," *Radioprotection* **47**(1), 93–103 (2012).
3. S. Jacob et al., "Risk assessment for potential radiation-induced cancer after lung and bone marrow exposure during interventional cardiology procedures," *Radioprotection* **48**(2), 203–213 (2013).
4. M. Andreucci, T. Faga, and A. Michael, "The renal problems in x-ray based imaging techniques using iodinated radiographic contrast agents," *Open Access J. Sci. Technol.* **3**, 101145 (2015).
5. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**(7553), 436–444 (2015).
6. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, Cambridge, Massachusetts (2016). <http://www.deeplearningbook.org>.
7. M. Pavoni, Y. Chang, and O. Smedby, "Image denoising with convolutional neural networks for percutaneous transluminal coronary angioplasty," in *VipIMAGE 2017: Proc. of the VI ECCOMAS Thematic*

- Conf. on Computational Vision and Medicinal Image Processing*, Porto, Portugal, October 18–20, 2017, pp. 255–265 (2018).
8. K. Bahrami et al., *Convolutional Neural Network for Reconstruction of 7T-like Images from 3t MRI using Appearance and Anatomical Features*, pp. 39–47, Springer International Publishing, Cham, Switzerland (2016).
9. M. Lee et al., "Multi-input cardiac image super-resolution using convolutional neural networks," in *Proc. of 19th Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, Athens, Greece (2016).
10. X. Wang et al., "Deep convolutional architecture for natural image denoising," in *2015 Int. Conf. on Wireless Communications Signal Process (WCSP)*, pp. 1–4 (2015).
11. V. Jain and S. Seung, "Natural image denoising with convolutional networks," in D. Koller et al. Eds., *Advances in Neural Information Processing Systems 21*, pp. 769–776, Curran Associates, Inc. (2009).
12. D. Bodington, "Learned convolutional denoising. CS231n convolutional neural networks for visual recognition," *Course Project Reports*, Stanford University (2015).
13. H. Chen et al., "Low-dose CT denoising with convolutional neural network," in *2017 IEEE 14th Int. Symp. on Biomedical Imaging (ISBI 2017)*, Melbourne, VIC, Australia (2017).
14. R. Maini and H. Aggarwal, "Study and comparison of various image edge detection techniques," *Int. J. Image Process.* **3**(1), 1–9 (2009).
15. A. Choromanska et al., "The loss surfaces of multilayer networks," in *Artificial Intelligence and Statistics*, pp. 192–204 (2015).
16. Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
17. B. Moulden, F. Kingdom, and L. F. Gatley, "The standard deviation of luminance as a metric for contrast in random-dot images," *Perception* **19**(1), 79–101 (1990).
18. M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.* **15**(12), 3736–3745 (2006).
19. M. Aharon, M. Elad, and A. Bruckstein, "k-svd: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.* **54**(11), 4311–4322 (2006).
20. K. Dabov et al., "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.* **16**(8), 2080–2095 (2007).

Marco Pavoni received his bachelor's degree from the Polytechnic University of Turin in biomedical engineering in 2015. In 2017, he received his master's degree from the Polytechnic University of Turin in biomedical engineering, specializing in biomedical instrumentation.

Yongjun Chang received his BS and MS degrees in industrial engineering from Seoul National University and University of Wisconsin–Madison, respectively, and his doctorate degree from KAIST in video signal processing in 2012. He was with Asan Medical Center and KAIST as a research fellow and a research assistant professor, respectively. He has been with KTH Royal Institute of Technology as a visiting researcher since 2016 toward medical image analysis and computer-aided diagnosis systems.

Sang-ho Park graduated from the Medical School at Soochunhyang University in 1999 and received his master's and PhD degrees in 2003 and 2013, respectively. He is a professor in cardiology division of Soonchunhyang Hospital Cheonan Hospital. He is an interventional cardiologist with about 10 years of experience in the field. His main research interest is peripheral and complex coronary intervention in which he has performed a series of studies and many clinical trials.

Orjan Smedby graduated from the Medical School at Uppsala University in 1983 and defended his doctoral thesis there in 1992. From 1999 to 2014, he was a professor of medical radiology at Linköping University. He was one of the founders of the Center for Medical Image Science and Visualization in Linköping. Since 2015, he is a professor in medical image processing and visualization at KTH Royal Institute of Technology. He is also a director of Swedish Bioimaging.