

Chest X-ray image denoising method based on deep convolution neural network

ISSN 1751-9659

Received on 28th February 2019

Revised 28th April 2019

Accepted on 26th June 2019

E-First on 9th August 2019

doi: 10.1049/iet-ipr.2019.0241

www.ietdl.org

Yan Jin¹ ✉, Xiao-Ben Jiang¹, Zhen-kun Wei¹, Yuan Li¹¹College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, People's Republic of China

✉ E-mail: jy@zjut.edu.cn

Abstract: To improve the visual effect of chest X-ray images and reduce the noise interference in disease diagnosis based on the chest X-ray images, the authors proposed an image denoising model based on deep convolution neural network. They utilise batch normalisation to solve the problem of performance degradation due to the increase of neural network layers, and use residual learning of the distribution of noise in noisy X-ray images. Specifically, the depthwise separable convolution is used to accelerate the convergence speed of network model, shorten the training time, and improve accuracy of the model. Compared to the several popular or the state-of-the-art denoising algorithms, their extensive experiments demonstrate that their method can not only achieve better denoising effects, but also significantly reduce the complexity of the network and shorten the computation time.

1 Introduction

Chest X-ray is a medical imaging technology that is economical and easy to use. Compared to computed tomography (CT), traditional X-rays are not only cheap, but also have lower radiation. Modern medical research shows that 1.5–2% of tumours may be due to high radiation dose of CT [1]. Especially, X-rays are widely used for diagnosing diseases such as pneumonia, pneumothorax, interstitial lung disease, heart failure, bone fracture, hiatal hernia, and so on. Moreover, chest X-ray is a standard screening method in physical examinations that over 300,000,000 people took in 2014 all over China. This number is still increasing, resulting in hundreds of millions of chest X-ray images per year [2].

Although chest X-rays are playing a more and more important role in medical examinations, chest X-rays are not only subject to various types and levels of noise during image transmission and acquisition, but also have streak artefact [3] which has direct or indirect influence that complicates the diagnosis, analysis, and treatment process [4]. Therefore, denoising and restoring chest X-ray images are of great significance. The denoising process perhaps disturbs the quality of the original image, which may lead to false decisions either by humans or machines. Therefore, the accuracy of noise removal should be considered as much as possible [5]. Image denoising is one of the most fundamental research topics in the field of computer vision. The medical image denoising problem has been extensively studied by scholars because of its high practical value. In most recent studies, there are mainly two types of image denoising methods: traditional denoising algorithms and deep learning methods based on neural network. Discrete wavelet is one of the traditional image denoising methods [6–8]. The major advantages of wavelet decomposition and reconstruction method are its simple algorithm and fast calculation [9]. However, the performance is relatively poor when denoising white noise. Arif *et al.* [10] proposed an isotropic diffusion filter and Bhonsle *et al.* [11] proposed a bilateral filter. Both improved the effect of removing the white noise, but the effect of denoising on the edges of the image is not satisfactory. The shearlet transform [12], the curvelet transform [13], and discrete cosine transform (DCT) non-local image denoising method [14] are utilised to improve the effect of denoising on the edges of the medical images, but the noise artefacts on the medical images cannot be effectively removed. Zhang *et al.* [15] have utilised nonlocal means-based regularization to measure noise artefacts and significantly improve the quality of low-dose CT. Dabov *et al.* [16] proposed block-

matching and 3-D transform-domain collaborative filtering (BM3D) strategy by grouping similar 2D image fragments (e.g. blocks) into 3D data arrays, which reduces calculation time while effectively suppressing noise artefacts.

However, the above algorithms also cause the loss of edge details while suppresses noise and artefacts. With the recent development of deep learning [17–19], deep convolutional neural networks [20] have become more and more widely used in image denoising fields. Denoising autoencoder (DAE) [21] and convolutional denoising autoencoders (CNN DAE) [22] are an extension of classical autoencoder and are extensively studied for medical image denoising. The models attempted to learn the denoised image from its noisy version; however, the robustness of the models to different noise types is not good enough, and the actual test effect is not as good as the training performance. To address this problem, Zhang *et al.* [23] proposed a model (DnCNN) for Gaussian noise removal. The DnCNN constructed a deep convolution networks to learn the distribution of Gaussian noise using residual learning [24] and batch normalisation (BN) [25]. Compared to the above traditional denoising algorithms and CNN DAE, better denoising results were achieved. The methods proposed in [5, 26], respectively, achieved low-dose CT and chest X-ray image restorations by constructing different deep residual networks. In this paper, we propose our deep convolutional neural network (denoted as X-ReCNN; here, ‘X’ stands for X-ray images, the ‘Re’ means restoration, and the ‘CNN’ means convolution neural network) used for the chest X-ray denoising, which combines the residual learning, BN, and depthwise separable convolution [27]. This method preserves as much detailed information of the original image as possible while denoising. Our X-ReCNN model can not only obtain high effectiveness, but also greatly reduce parameters of the network and shorten the training time. Compared to some popular denoising algorithms such as nonlocal means (NLM), BM3D, and the advanced model DnCNN, our extensive experiments demonstrate that our X-ReCNN can achieve the optimal effect not only in subjective visual effect, but also on objective evaluation index [(peak-signal-to-noise ratio (PSNR), structural similarity index (SSIM), and root-mean-squared error (RMSE))].

The contributions of our work can be summarised as follows:

- We design a deep feed-forward convolutional neural network model which directly approximates the noise from a noisy chest X-ray image.

- ii. We improve the model performance by combining the residual learning, depthwise separable convolution, and BN. The residual learning is used as learning algorithm, the depthwise separable convolution is used to reduce network parameters and improve accuracy as well as shorten training time, also BN is incorporated to boost model performance.
- iii. We get the X-ReCNN-S model and the X-ReCNN-B model separately by using different training data for X-ReCNN model. The X-ReCNN-S model is for removing known noise levels and the X-ReCNN-B model for blind Gaussian denoising.

The remainder of this paper is organised as follows. Section 2 provides a brief survey of related work. Section 3 first presents our X-ReCNN model, and then illustrates the training details. In Section 4, extensive experiments are conducted to evaluate X-ReCNN. Finally, several concluding remarks are given in Section 5.

2 Related works

2.1 Deep convolutional neural networks

Deep convolutional neural networks are the feed-forward neural networks with deep structure and using convolution or correlation calculation, which are a kind of the representative deep learning algorithms of [18, 28]. Compared to the traditional multi-layer perceptron, convolutional neural networks significantly reduce the parameters and decrease the computational complexity by sharing weights. The traditional convolutional neural network consists of input layer, convolution layer, pooling layer, fully connected layer, and output layer. Convolutional layer has utilised convolution kernel to pick up multiple angles and levels of features in the image, and uses the result as input to the pooling layer. The benefit of the pooling layer is to reduce the dimension of features and enhance the robustness of features. Recently, simpler and more advanced methods of convolutional neural networks have been developed such as rectified linear unit (ReLU) [$\max(0, \cdot)$], BN, residual learning, and depthwise separable convolution which improve the learning ability of convolutional neural networks.

2.2 Batch normalisation

BN was originally proposed to solve the problem that the distribution of input data in each layer will often be significantly different from that before the parameter is updated, which result in the network constantly adapting to the new data distribution and being hard to train. The core of this method is not only to standardise the input layer, but also to standardise the input of each inter layer (before the activation function) so that the output data obey a normal distribution with 0 mean and 1 variance. Let $B = \{x_1, x_2, x_3, \dots, x_m\}$ be a sample set, then the formula for batch regularisation is as follows, where x from i to m is the element in B :

$$\mu_B = \frac{1}{m} \sum_{i=1}^m x_i \quad (1)$$

$$\sigma_B^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2 \quad (2)$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} \quad (3)$$

$$y_i = \gamma \hat{x}_i + \beta, \quad (4)$$

where μ_B and σ_B^2 are, respectively, the mean and variance of the sample set B , \hat{x}_i represents the regularised output value of x_i , and ε is a small positive number to avoid the denominator of, for example (3) is zero. Here, γ and β , respectively, stand for translation and scaling of the normalised parameters, which need to

be obtained through training. Many researchers have utilised BN to accelerate the convergence of the network and improve the generalisation ability of the network [25].

2.3 Residual learning

Residual learning [24] was originally proposed to solve the performance degradation problem, i.e. even the training accuracy begins to degrade along with the increasing of network depth which is not because of overfitting. Fig. 1 shows the framework of residual learning framework.

We utilise the residual learning to change the optimisation target from $H(X) = F(X) + X$ to $H(X) - X$. The benefit of this jump structure is that the optimisation target can be turned to approach zero, which is much easier to be learnt than the original unreferenced mapping. In this paper, X denotes undegraded chest X-ray, $H(X)$ stands for a degraded noisy map, and $F(X)$ represents a noise map. Therefore, our learning objective is to learn the noise distribution in the noisy image instead of directly learning the noisy map, which makes it easier for the network to find local optimal solutions and shorten training time.

2.4 Depthwise separable convolution

Fig. 2 shows a traditional convolution process for a three-channel colour picture with the size of $5 \times 5 \times 3$. In Fig. 2, we suppose there are four output channels. After the convolution operation with convolution kernel of 3×3 , the total required parameters are $3 \times 3 \times 4 = 108$.

Compared to the traditional convolution process, depthwise separable convolution decomposes a convolution operation into two steps. First, as shown in Fig. 3, we perform depthwise convolution (DW). Unlike the traditional convolution shown in

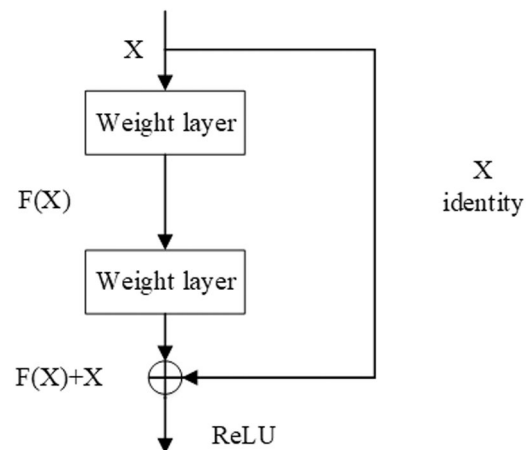


Fig. 1 Residual learning framework

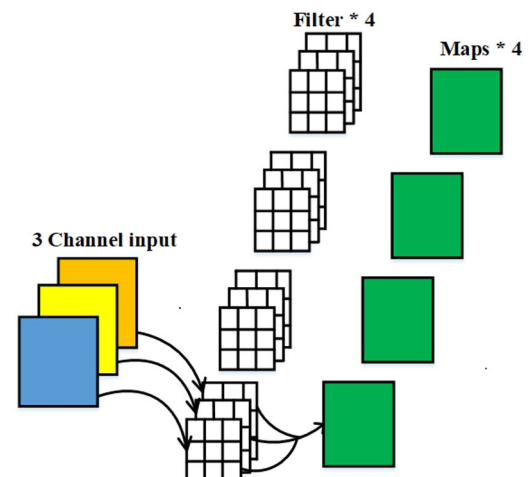


Fig. 2 Traditional convolution

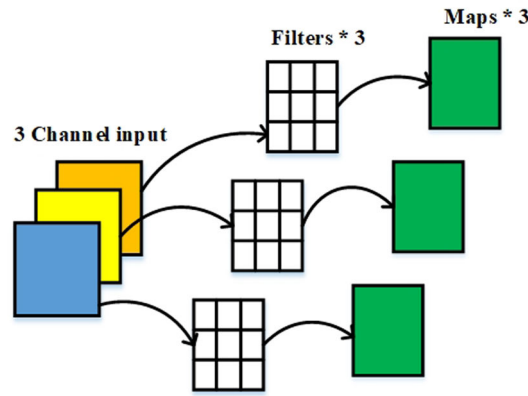


Fig. 3 DW

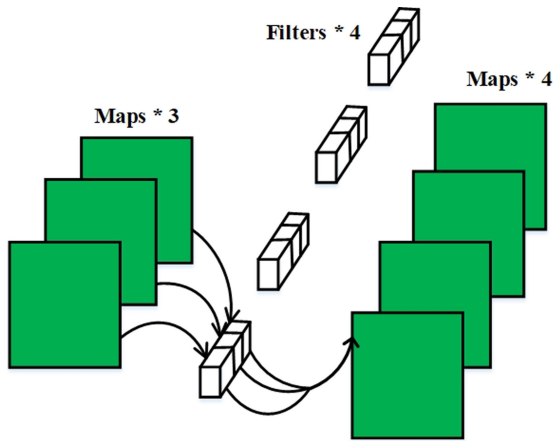


Fig. 4 PW

Fig. 2, DW, an independent operation on each channel of the same picture, is operated entirely in a 2D plane, number of convolution kernels is the same as the previous layer.

The number of output feature maps after DW is the same as the channel number of the input layer, resulting in the inability to expand the feature map. Moreover, DW does not effectively extract the feature information of different channels in the same spatial position because it is performed on each channel of the input layer independently. Therefore, in the next step, we utilise pointwise convolution (PW) to combine these feature maps to generate a new feature map.

As shown in Fig. 4, the convolution kernel size is $1 \times 1 \times M$, where M is the number of the channels on the previous layer. So the total parameter number of the depthwise separable convolution is $3 \times 3 \times 3 + 3 \times 1 \times 1 \times 4 = 39$.

Compared to the traditional convolution, the depthwise separable convolution can greatly reduce the amount of the parameters. Especially, the effect is more obvious as the number of output feature maps increases. The advantages of reducing the parameters are that it can speed up network convergence, shorten training time, and improve accuracy [27].

3. Our X-ray image denoising model

The purpose of image denoising is to remove the noise from the noisy image and to retain the original image information to the greatest extent. In, for example (5), set chest X-ray image as $f(h, v)$ which can be seen as the result of adding the original image $u(h, v)$ and the noise $n(h, v)$, where h and v represent the horizontal and vertical coordinates of the pixels, respectively

$$f(h, v) = u(h, v) + n(h, v) \quad (5)$$

Different from the traditional neural network directly taking the mapping of $f(h, v)$ to $u(h, v)$, the proposed model in this paper constructs the residual learning which takes the mapping from

$f(h, v)$ to $n(h, v)$, and then minimise the loss function, for example the equation below:

$$\arg \min \| f(h, v) - n(h, v) \|_2^2 \quad (6)$$

It is much easier to let the network indirectly learn the noise $n(h, v)$ in the image compared with directly learning to get a clean image $u(h, v)$. Then, we can get the denoised image by subtracting the noise image from the low-dose chest X-ray image.

3.1 Network architecture

In traditional convolution operations, many researchers often deepen the network to improve the model accuracy. However, the model accuracy begins to degrade along with the increasing of network depth. In addition, the model is prone to overfitting [29] and falls into the local optimal solution during training.

To solve the problems, we proposed a new network architecture X-ReCNN. We utilised BN and residual learning to ensure the accuracy of the model will not decrease with the increasing of network depth. Moreover, we use the depthwise separable convolution operation to reduce the parameter of the network which can further prevent overfitting and accelerate the convergence speed of the network.

Given the X-ReCNN with 15 layers, there are three types of layers, shown in Fig. 5. Here, the 'Conv' means convolution operation, the '(DW+PW) Conv' stands for the depthwise separable convolution, and ReLU is the activation function.

- Conv + ReLU:** For the first layer, the noise characteristics of the chest X-ray film are extracted and ReLU [ReLU, max(0, ·)] are then utilised for non-linearity.
- (DW + PW) Conv BN + ReLU:** For layers 2–14, unlike the previous layer, DW and PW are utilised for reducing the parameters and preventing overfitting.
- Conv:** For the last layer, convolution operations are used to reconstruct the output.

Finally, the whole network forms a large residual unit. The residual learning strategy is used to continuously learn $n(h, v)$. BN and DW separable convolution operations are used to accelerate network convergence as well as improve network performance.

The parameter settings of the convolution kernel and the channel number in each layer are shown in Table 1.

3.2 Training data

To train the model X-ReCNN, first we considered the images of chest radiographs with nodules and non-nodules from Japan Society of Radiological Technology (JSRT) database. This is a publicly available database with 247 chest radiographs collected from 13 institutions in Japan and one in the USA. All images used in the database (lung nodules and non-nodules) have been examined by CT images to confirm the presence (or the absence)

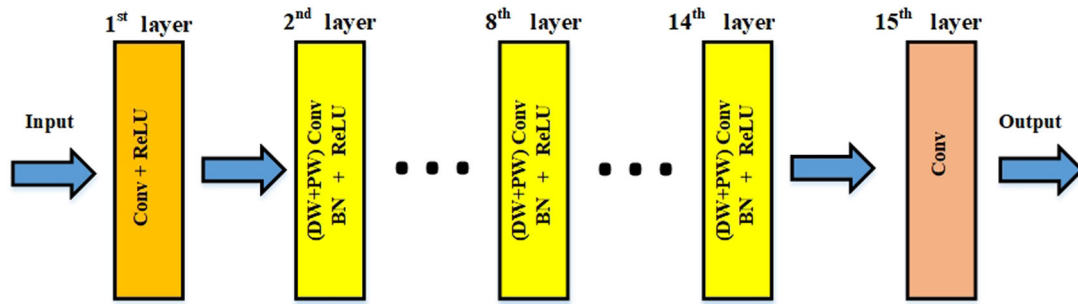


Fig. 5 Architecture of the X-ReCNN network

Table 1 Related parameters of X-ReCNN

Layer	Kernel size	Channel number
1st	3×3	48
2nd–14th	depthwise: 3×3 PW: 1×1	48
15th	3×3	1

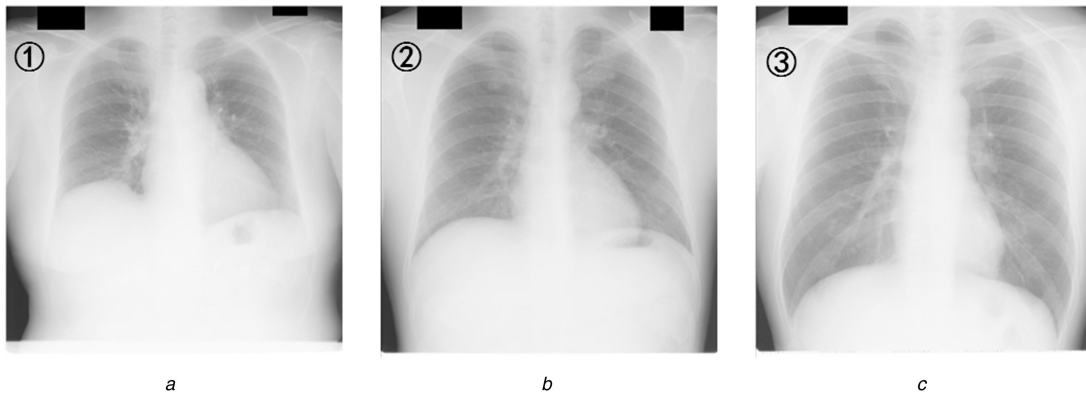


Fig. 6 Testing images

(a) JPCNN092, (b) JPCNN093, (c) JPCNN089

of nodules. The image size is 2048×2048 , the pixel size is 0.175 mm, and the optical density ranges from 0 to 3.5 (high density corresponds to a high pixel value and low density to a low pixel value). We randomly selected 225 images as the training set and the other 15 images are used for test. To enhance the data and speed up the training of the network, we first scale the size of original image to 512×512 . Then, we set the patch size as 120×120 and crop 64×2096 patches by scaling and rotation.

3.3 Training parameters

On the one hand, we first added white Gaussian noise (WGN) with standard deviations $\sigma_1 = 15$ and $\sigma_2 = 25$, to the sampled image blocks to simulate the noisy images. Then, we trained our model X-ReCNN by using the paired original images and the noisy images with fixed level noise to get the model X-ReCNN-S for removing known noise levels.

On the other hand, we also added WGN with standard deviation $\sigma = [0, 45]$ to the sampled image blocks to simulate the noisy images with different level noises. Then, we trained the model by using the paired original images and the noisy images with different level noises to get the blind Gaussian denoising model X-ReCNN-B.

The network weights are initialised with a Gaussian distribution, and the convolution of each layer uses a Gaussian distribution with mean zero and variance 0.1. Adam optimisation method [30] was adopted in the training. The initial learning rate is set to 0.001, and gradually decays to 10^5 as the training progresses. The loss function of the network is shown in, for example (6). The mini-batch size is set as 64. L2 regularisation [31] was used, and weighting coefficients were set as 0.001 in our experiments. We train 50 epochs for our X-ReCNN-B and X-ReCNN-S models.

Unless otherwise stated, all the experiments were carried out in the Pycharm environment running on a PC with Inter Core i7-7800K CPU and Nvidia GeForce GTX 1080Ti.

4 Experimental results

We applied our models to synthetic images corrupted by additive WGN with different standard deviation and a real X-ray image to evaluate the performance of our models. Comparisons of denoising performance are made among NLM [32], BM3D [16], DnCNN [23], and our models. We also test denoising time of the different algorithms.

4.1 On synthetic images

In the first experiment, in order to verify the validity and accuracy of the X-ReCNN-B and X-ReCNN-S, we selected 15 chest X-ray images (JPCNN147-154, JPCNN087-093) from the JSRT dataset as test data, which are not included during training dataset. Fig. 6 shows three chest X-ray images of the test dataset. The test images were set to the size of 512×512 , corrupted by WGN with standard deviations $\sigma_1 = 15$ and $\sigma_2 = 25$, and divided into two groups for test. Compared to the NLM, BM3D, and DnCNN, the advantages of our model are further illustrated. We mainly utilised subjective and objective evaluation criteria to evaluate the experimental results. In terms of subjective evaluation, the denoising algorithms are evaluated by observing the overall visual effect of the denoised images and the retention of edge details. In terms of objective evaluation, we use PSNR, SSIM, and RMSE to evaluate our algorithms.

For Fig. 6a, we added WGN with the standard deviation $\sigma_1 = 15$, as shown in Fig. 7a. The denoising effects of different

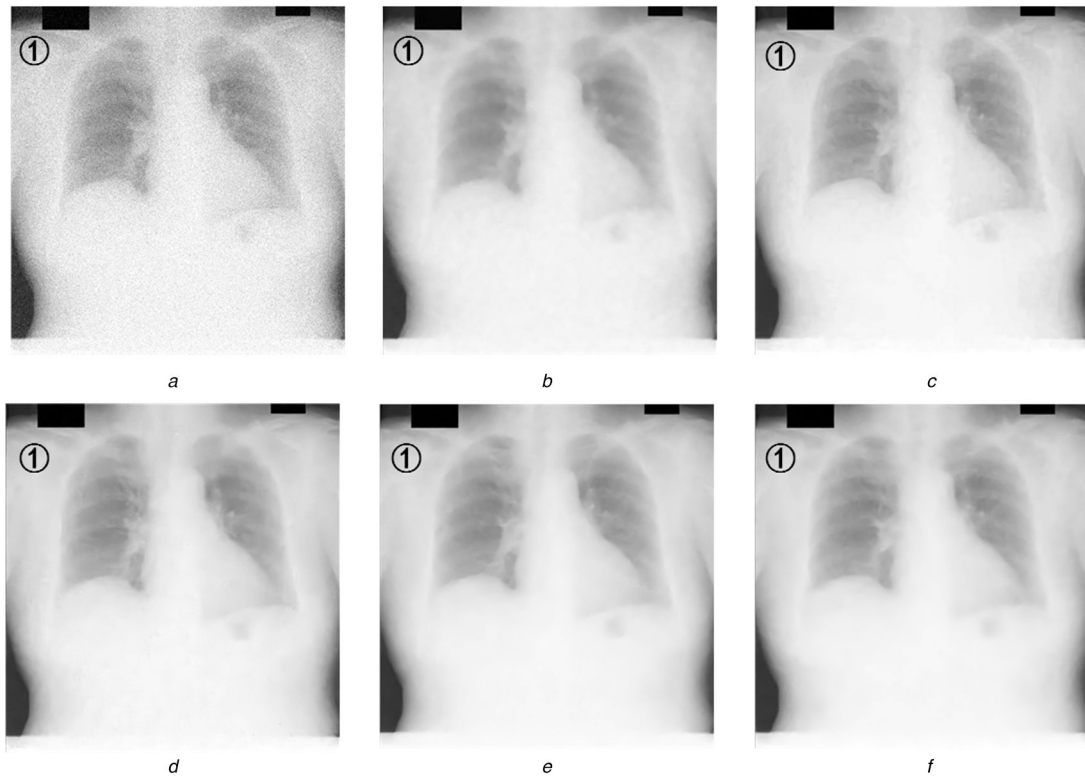


Fig. 7 Denoising results and their PSNRs of Fig. 6a with noise standard deviation 15 by different algorithms

(a) Noisy image/25.061 dB, (b) NLM/38.668 dB, (c) BM3D/41.539 dB, (d) DnCNN/41.827 dB, (e) X-ReCNN-B/42.270 dB, (f) X-ReCNN-S/42.316 dB

algorithms are shown in Figs. 7b–f. One can clearly see that NLM is not satisfactory, and there is still much obvious noise. Although BM3D and DnCNN are significantly improved, the edges and textures between the bilateral lungs are not maintained well. Figs. 7e and f demonstrate the denoising effect of X-ReCNN-B and X-ReCNN-S. It can be seen that the details of the inside of the chest are retained to the utmost extent while denoising and enhancing the edge effect. For Fig. 6b, we added WGN with the standard deviation $\sigma_2 = 25$, as shown in Fig. 8a. The denoised images and their PSNRs by different algorithms are shown in Figs. 8b–f. Just like Fig. 7, one can still see that the denoised images (Figs. 8e and f) by the X-ReCNN have better performance. Figs. 9b–f show the denoising results of the square box region of Fig. 9a using different algorithms. From these images, one can clearly see that though NLM removes most of the noise, it also blurs the image. The denoising effect of BM3D and DnCNN is improved, but the edge details of the ribs are not maintained well. The denoised images by X-ReCNN-S and X-ReCNN-B are closer to the original image than the denoised images by other models. X-ReCNN models can preserve the edge details between the ribs much better while denoising.

In Fig. 7, the PSNR value of X-ReCNN-S is the highest, and has 3.648, 0.777, and 0.489 dB more than NLM, BM3D, and DnCNN, respectively. The PSNR of X-ReCNN-B is closed to X-ReCNN-S very much. In Fig. 8, the PSNR value of X-ReCNN-B is the highest, and has 4.729, 1.039, and 0.282 dB more than NLM, BM3D, and DnCNN, respectively. X-ReCNN-S is slightly less than X-ReCNN-B.

Tables 2 and 3 show the averages PSNR, RMSE, and SSIM values of denoised images of 15 test images with noise standard deviations 15 and 25, respectively. Objective evaluation indicators PSNR, RMSE, and SSIM further demonstrate the effectiveness of X-ReCNN. PSNR and RMSE are indicators for evaluating the denoising ability of the algorithm. The larger the value of PSNR is, the stronger the denoising ability of the corresponding algorithm has, while the smaller the value of RMSE has, the better the performance. SSIM is an indicator to measure the similarity of the structure of two images. The larger the value of SSIM, the better the algorithm retains image details. Best results are designated in bold.

From Table 2, where $\sigma = 15$, one can see that our X-ReCNN-S model is the best one, its PSNR values are 2.529, 0.474, and 0.445 dB more than NLM, BM3D, and DnCNN, respectively, whereas its SSIM values are 0.0202, 0.0077, and 0.0052 more than NLM, BM3D, and DnCNN.

From Table 3, where $\sigma = 25$, the PSNR value of our X-ReCNN-B are 4.327, 0.982, and 0.446 dB more than other three algorithms, whereas its SSIM values are 0.0378, 0.015, and 0.0087 more than other three algorithms, respectively.

To test the ability of our X-ReCNN-B for the blind Gaussian denoising, we added WGN with standard deviation $\sigma = \{15, 20, 25, 30, 35, 40, 45\}$ to Fig. 6c, then recovered the noisy images using NLM, BM3D, DnCNN, and X-ReCNN-B. From the denoised results shown in Figs. 10 and 11, one can clearly see that PSNR and SSIM values of our X-ReCNN-B are higher than NLM, BM3D, and DnCNN.

To discuss the influence of BN, residual learning, and DW separable convolution on chest X-ray image denoising, we made the following changes in our X-ReCNN network (called as default network):

- Remove all BN layers from our X-ReCNN and get a network which is without BN layers.
- Let the network directly learn noise images without using residual learning.
- Replace depthwise separable convolution with traditional convolution and get a network which is without DW and PW.

We used the same training data such as X-ReCNN-B to train the three changed networks. Then, we utilised these trained networks and X-ReCNN-B to denoise the mentioned 15 test images (512×512) corrupted by WGN with $\sigma = 35$. We still took the average of PSNR, RMSE, and SSIM values of denoised images as the objective evaluation criteria. From Table 4, one can see that the X-ReCNN-B is better than all other models which indicate BN, residual learning, and DW separable convolution all play a positive role in improving the denoising performance of the network. Especially, DW separable convolution increases the PSNR of the model by 0.906 and residual learning improves the PSNR by 0.8752.

In Fig. 12, the X -axis represents training steps and the Y -axis stands for training error. Fig. 12 appears that the networks with DW and PW operation lead to lower convergence loss in the training steps, indicating that the depthwise separable convolution boosts the generalisation ability.

4.2 On real X-ray images

In the second experiment, for demonstrating the denoising effectiveness of our X-ReCNN-B on real X-ray images, we test the model on real noisy chest X-ray image of size 202×250 . Since there is not a clean image, the quality of recovered image cannot be evaluated by any of PSNR, RMSE, and SSIM. Here, method noise

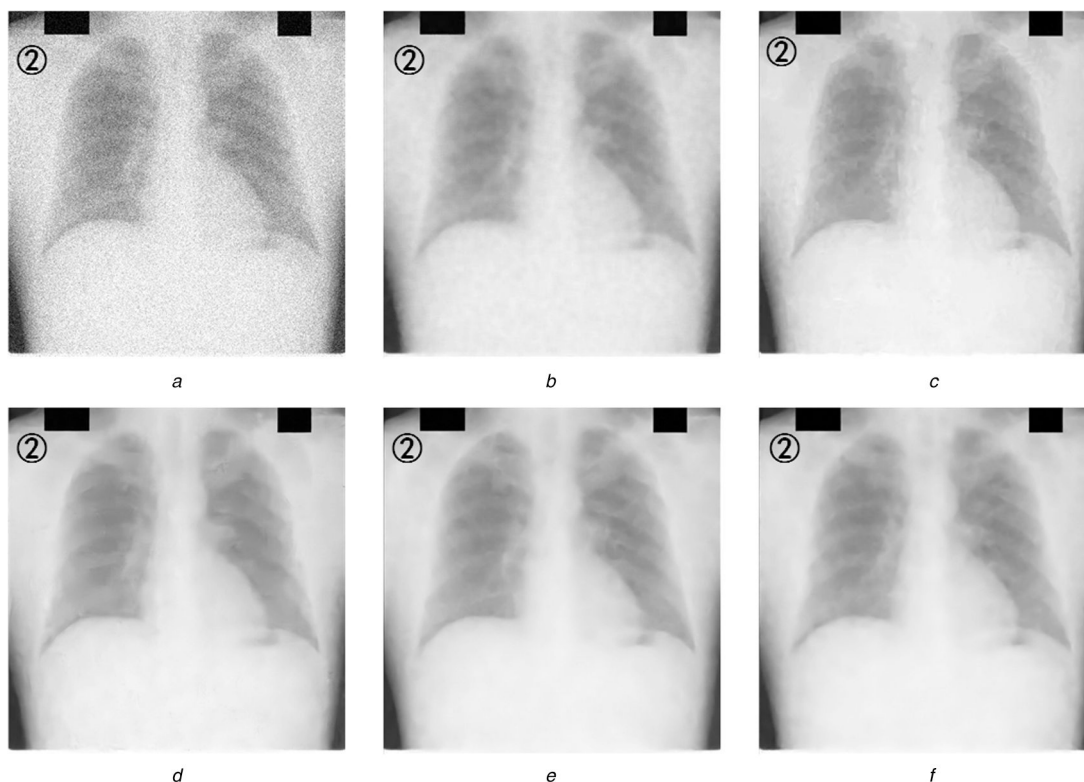


Fig. 8 Denoising results and their PSNRs of Fig. 6b with noise standard deviation 25 by different algorithms

(a) Noisy image/20.818 dB, (b) NLM/34.902 dB, (c) BM3D/38.592 dB, (d) DnCNN/39.349 dB, (e) X-ReCNN-B/39.631 dB, (f) X-ReCNN-S/39.567 dB

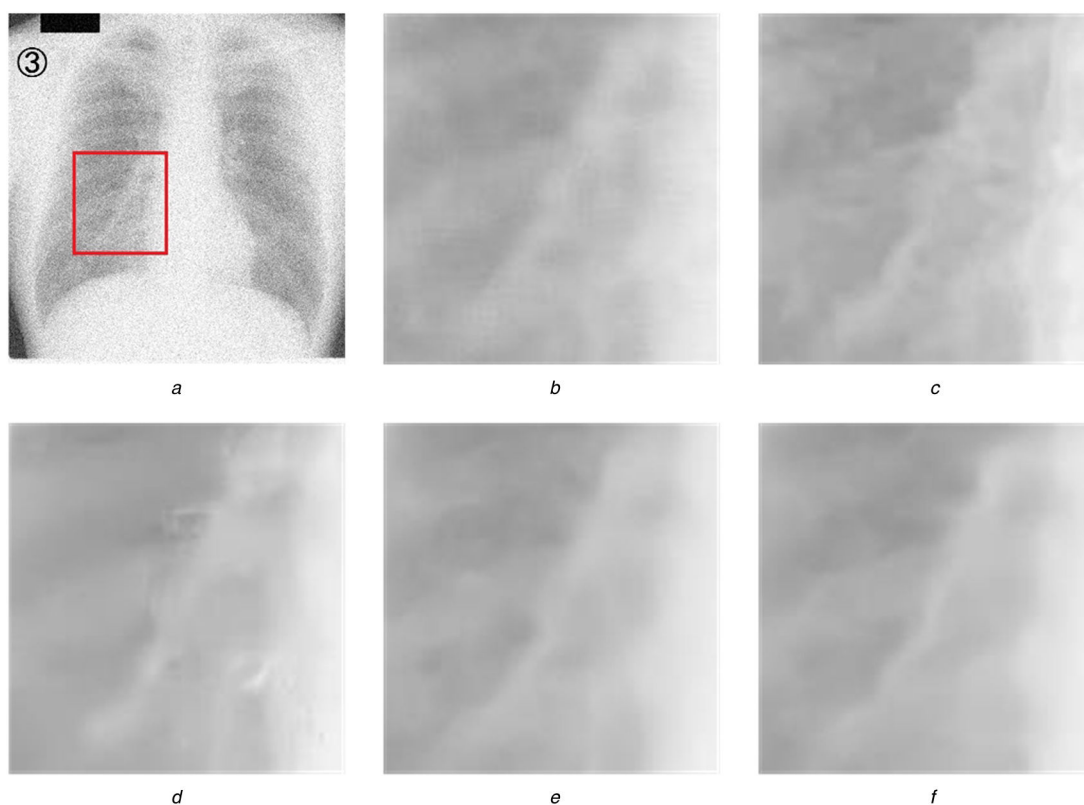


Fig. 9 Local zoomed-in zone of denoised images of Fig. 6c with noise standard deviation 25

(a) Noisy image, (b) NLM, (c) BM3D, (d) DnCNN, (e) X-ReCNN-B, (f) X-ReCNN-S

Table 2 Average PSNR, RMSE, and SSIM of denoised images by different algorithms of test images with noise standard deviation 15

	PSNR, dB	RMSE	SSIM
noise image/ $\sigma = 15$	24.935	14.4482	0.2751
NLM	38.893	2.8984	0.9545
BM3D	40.948	2.2892	0.9670
DnCNN	40.977	2.2839	0.9693
X-ReCNN-B	41.371	2.1851	0.9742
X-ReCNN-S	41.422	2.1689	0.9745

The bold values indicates the best results.

Table 3 Average PSNR, RMSE, and SSIM of denoised images by different algorithms of test images with noise standard deviation 25

	PSNR, dB	RMSE	SSIM
noise image/ $\sigma = 25$	20.816	23.2179	0.1350
NLM	34.998	4.5427	0.9293
BM3D	38.343	3.0892	0.9521
DnCNN	38.879	2.9135	0.9584
X-ReCNN-B	39.325	2.7641	0.9671
X-ReCNN-S	39.157	2.8215	0.9674

The bold values indicates the best results.

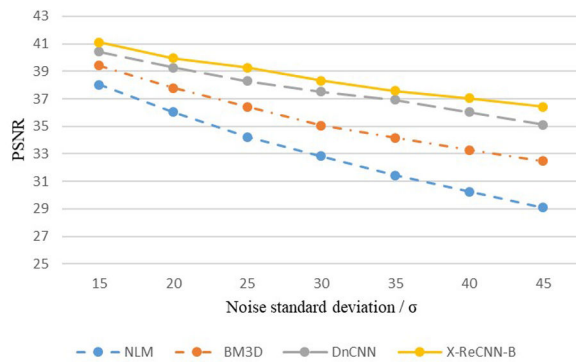


Fig. 10 PSNR values of denoised results of noisy JPCNN089 image using different methods

Table 4 Denoised results of 15 noisy test images with noise standard deviation 35 using Model 1, Model 2, Model 3, and X-ReCNN-B with different network structures

	PSNR, dB	RMSE	SSIM
default	37.8629	3.4418	0.9646
without BN layers	37.3902	3.4546	0.9627
without residual learning	36.9877	3.6742	0.9524
without DW and PW	36.9569	3.6684	0.9519

The bold values indicates the best results.

proposed in [32] is chosen as evaluation criterion, which is independent of the clean image. The method noise of a denoising approach is defined by $\hat{n} = f - \hat{u}$; here, f represents a noisy image such as the real X-ray image, \hat{u} is a denoised image, and \hat{n} shows the noise that is removed from the image f . The less edge and texture on \hat{n} , the better the ability of keeping edges and textures is [32]. If a denoising method performs well, the method noise looks like a noise and should contain as little texture details as possible [33]. Combining the visual judgement with the method noise, the experiments were stopped when the textures of the image \hat{n} were the least and the denoised image had a good visual effect.

Fig. 13 shows the denoising results by using NLM, BM3D, DnCNN, and our X-ReCNN-B. The method noises of these methods in Fig. 14 show that the lost textures by our X-ReCNN-B are less than NLM, BM3D, and DnCNN, which means X-ReCNN-

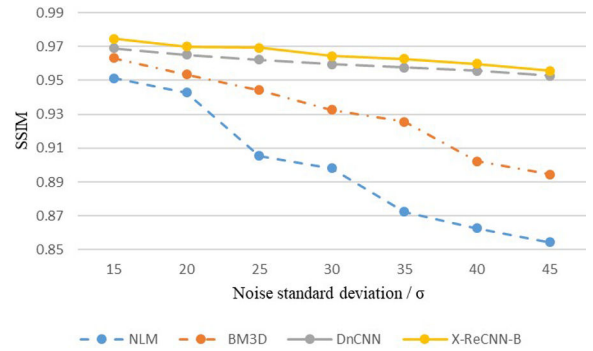


Fig. 11 SSIM values of denoised results of noisy JPCNN089 using different methods

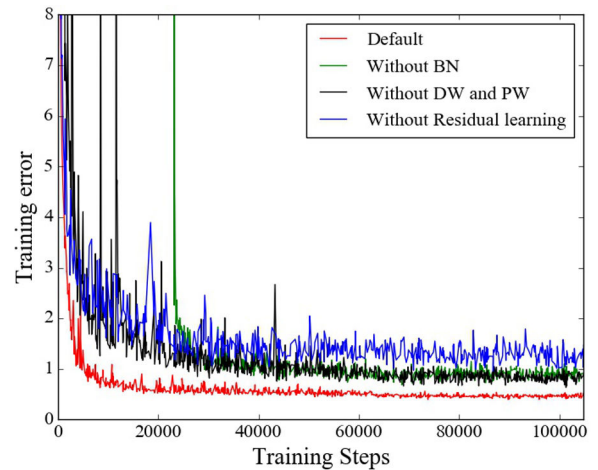


Fig. 12 Training error of four specific models

B has a better performance in preserving textures than NLM, BM3D, and DnCNN while denoising well.

4.3 Denoising time

In addition to visual quality, another important aspect for image restoration methods is the running speed. A good denoising algorithm, while ensuring the denoising effect, should also shorten the denoising time as much as possible. We use Fig. 6a with noise standard deviation 25 to test denoising time of the different algorithms. The size of the image is set as 128×128 , 256×256 , and 512×512 . The experimental results in Table 5 show X-ReCNN-S have a relatively high speed though it is little slower than DnCNN when the image is of size 128×128 .

5 Conclusion

In this paper, a deep convolutional neural network is proposed for chest X-ray image denoising, where depthwise separable convolution is adopted to reduce network parameters and improve accuracy as well as shorten training time. The residual learning and BN are integrated to separate noise from noisy observation and speed up the training process. Extensive experimental results demonstrated that our method not only produces favourable image denoising performance for both synthetic images and real images, but also has less denoising time. Our X-ReCNN outperforms NLM, BM3D, and DnCNN in terms of both denoising ability and retention of detailed information. Especially, our X-ReCNN-B model can handle the blind Gaussian denoising with unknown noise level.

6 Acknowledgment

This research was supported by the Zhejiang Provincial Natural Science Foundation of China under Grant no. LY17F010015.

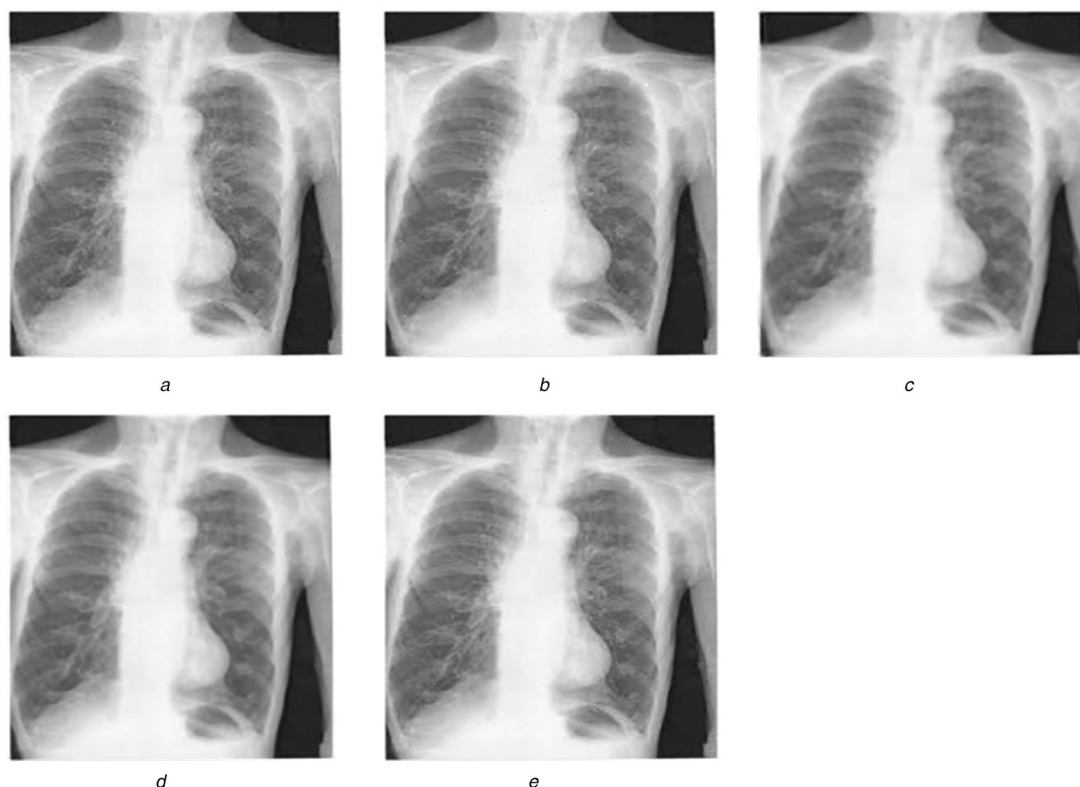


Fig. 13 Test results on real X-ray image

(a) Real X-ray image, (b) Recovered image of NLM (c) Recovered image of BM3D, (d) Recovered image of DnCNN, (e) Recovered image of X-ReCNN-B

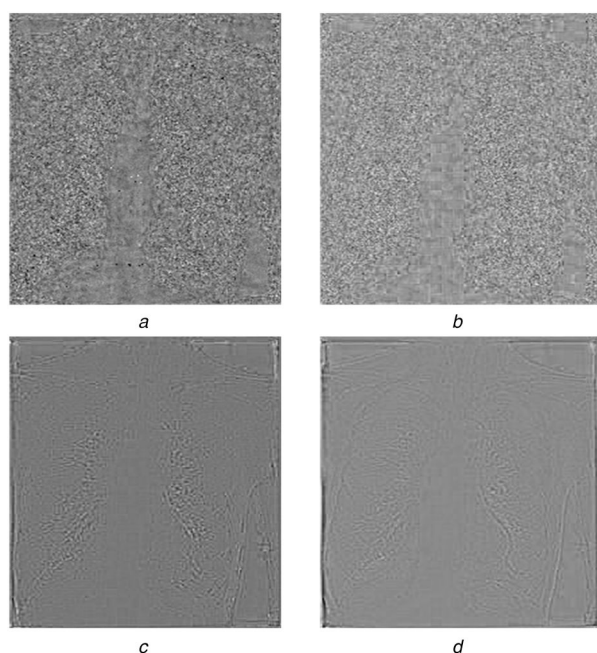


Fig. 14 Residual images of different methods on real X-ray image

(a) NLM, (b) BM3D, (c) DnCNN, (d) X-ReCNN-B

Table 5 Denoising time of different methods of Fig. 6a of size 128×128 , 256×256 , and 512×512 with noise standard deviation 25

Image size	BM3D, s	DnCNN, s	X-ReCNN-S, s
128×128	1.10	0.93	1.08
256×256	1.71	1.24	1.23
512×512	3.41	2.47	2.33

The bold values indicates the best results.

7 References

- [1] Smith-Bindman, R., Lipson, J., Marcus, R., *et al.*: 'Radiation dose associated with common computed tomography examinations and the associated lifetime attributable risk of cancer', *Arch. Intern. Med.*, 2009, **169**, (22), pp. 2078–2086
- [2] Dong, Y., Pan, Y., Zhang, J., *et al.*: 'Learning to read chest X-ray images from 16000+ examples using CNN'. Proc. Int. Conf. Connected Health: Applications, Systems and Engineering Technologies, PA, USA, July 2017, pp. 51–57
- [3] Mori, I., Machida, Y., Osanai, M., *et al.*: 'Photon starvation artifacts of X-ray CT: their true cause and a solution', *Radiol. Phys. Technol.*, 2013, **6**, (1), pp. 130–141
- [4] Gao, Q., Olgac, N.: 'Stability analysis for LTI systems with multiple time delays using the bounds of its imaginary spectra', *Syst. Control Lett.*, 2017, **102**, pp. 112–118
- [5] Jifara, W., Jiang, F., Rho, S., *et al.*: 'Medical image denoising using convolutional neural network: a residual learning approach', *J. Supercomput.*, 2019, **75**, (2), pp. 704–718
- [6] Mondal, T., Maitra, M.: 'Denoising and compression of medical image in wavelet 2D', *Int. J. Recent Innov. Trends Comput. Commun.*, 2014, **2**, (2), pp. 1–4
- [7] Mustafa, N., Khan, S.A., Li, J.-P., *et al.*: 'Medical image de-noising schemes using wavelet transform with fixed form thresholding'. Int. Conf. Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China, December 2014, pp. 397–402
- [8] Raj, V.N.P., Venkateswarlu, T.: 'Denoising of medical images using undecimated wavelet transform'. Int. Conf. Recent Intelligent Computational Systems (RAICS), Trivandrum, India, November 2011, pp. 483–488
- [9] Guan, L., Jin-zhao, W., Ming-yu, F.: 'Signal de-noising analysis based on wavelet', *Commun. Technol.*, 2010, **43**, (9), pp. 79–84
- [10] Arif, A.S., Mansor, S., Logeswaran, R.: 'Combined bilateral and anisotropic-diffusion filters for medical image de-noising'. Int. Conf. Research and Development (SCORed), Cyberjaya, Malaysia, February 2012, pp. 420–424
- [11] Bhonsle, D., Chandra, V., Sinha, G.: 'Medical image denoising using bilateral filter', *Int. J. Image Graph. Signal Process.*, 2012, **4**, (6), p. 36
- [12] Zhao, J., Lü, L., Sun, H.: 'Multi-threshold image denoising based on shearlet transform'. Proc. Int. Conf. Applied Mechanics and Materials, Switzerland, 29 August 2010, pp. 2251–2255
- [13] Starck, J.-L., Candès, E.J., Donoho, D.L.: 'The curvelet transform for image denoising', *IEEE Trans. Image Process.*, 2002, **11**, (6), pp. 670–684
- [14] Hu, J., Pu, Y., Wu, X., *et al.*: 'Improved DCT-based nonlocal means filter for MR images denoising', *Comput. Math. Methods Med.*, 2012, **2012**, pp. 14–28
- [15] Zhang, H., Ma, J., Wang, J., *et al.*: 'Statistical image reconstruction for low-dose CT using nonlocal means-based regularization, part II: an adaptive approach', *Comput. Med. Imaging Graph.*, 2015, **43**, (3), pp. 26–35
- [16] Dabov, K., Foi, A., Katkovnik, V., *et al.*: 'Image denoising by sparse 3-D transform-domain collaborative filtering', *IEEE Trans. Image Process.*, 2007, **16**, (8), pp. 2080–2095

- [17] LeCun, Y., Bengio, Y., Hinton, G.: 'Deep learning', *Nature*, 2015, **521**, (7553), p. 436
- [18] Bengio, Y., Lamblin, P., Popovici, D., *et al.*: 'Greedy layer-wise training of deep networks'. Neural Information Proc. Systems (NIPS), Canada, December 2007, pp. 153–160
- [19] Schmidhuber, J.: 'Deep learning in neural networks: an overview', *Neural Netw.*, 2015, **61**, pp. 85–117
- [20] Krizhevsky, A., Sutskever, I., Hinton, G.E.: 'ImageNet classification with deep convolutional neural networks'. Neural Information Proc. Systems (NIPS), NV, USA, December 2012, pp. 1097–1105
- [21] Vincent, P., Larochelle, H., Bengio, Y., *et al.*: 'Extracting and composing robust features with denoising autoencoders'. Proc. 25th Int. Conf. Machine learning, Helsinki, Finland, July 2008, pp. 1096–1103
- [22] Gondara, L.: 'Medical image denoising using convolutional denoising autoencoders'. 2016 IEEE 16th Int. Conf. Data Mining Workshops, Barcelona, Spain, December 2016, pp. 241–246
- [23] Zhang, K., Zuo, W., Chen, Y., *et al.*: 'Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising', *IEEE Trans. Image Process.*, 2017, **26**, (7), pp. 3142–3155
- [24] He, K., Zhang, X., Ren, S., *et al.*: 'Deep residual learning for image recognition'. Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, June 2016, pp. 770–778
- [25] Salimans, T., Kingma, D.P.: 'Weight normalization: a simple reparameterization to accelerate training of deep neural networks'. Neural Information Processing Systems (NIPS), Barcelona, Spain, December 2016, pp. 901–909
- [26] Burger, H.C., Schuler, C.J., Harmeling, S.: 'Image denoising: Can plain neural networks compete with BM3D?'. Computer Vision and Pattern Recognition (CVPR), Providence, USA, June 2012, pp. 2392–2399
- [27] Chollet, F.: 'Xception: deep learning with depthwise separable convolutions'. Computer Vision and Pattern Recognition (CVPR), Hawaii, USA, July 2017, pp. 1251–1258
- [28] Gu, J., Wang, Z., Kuen, J., *et al.*: 'Recent advances in convolutional neural networks', *Pattern Recognit.*, 2018, **77**, pp. 354–377
- [29] Rasmussen, C.E.: '*Gaussian processes in machine learning*' (Summer School on Machine Learning, Berlin, German, 2003), pp. 63–71
- [30] Kingma, D.P., Salimans, T., Welling, M.: 'Variational dropout and the local reparameterization trick'. Neural Information Proc. Systems (NIPS), Montreal, Canada, December 2015, pp. 2575–2583
- [31] Krizhevsky, A., Sutskever, I., Hinton, G.E.: 'ImageNet classification with deep convolutional neural networks'. Neural Information Proc. Systems (NIPS), NV, USA, December 2012, pp. 1097–1105
- [32] Buades, A., Coll, B., Morel, J. M.: 'A review of image denoising algorithms, with a new one'. *SIAM J. Multiscale Model. Simul.*, 2005, **4**, (2), pp. 490–530
- [33] Lu, J., Yang, H.M., Shen, L.X., *et al.*: 'Ultrasound image restoration based on a learned dictionary and a higher-order MRF', *Comput. Math. Appl.*, 2019, **77**, pp. 991–1009