



Images denoising for COVID-19 chest X-ray based on multi-resolution parallel residual CNN

Xiaoben Jiang¹ · Yu Zhu¹ · Bingbing Zheng¹ · Dawei Yang²

Received: 28 January 2021 / Revised: 19 May 2021 / Accepted: 9 June 2021 / Published online: 28 June 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Chest X-ray (CXR) is a medical imaging technology that is common and economical to use in clinical. Recently, coronavirus (COVID-19) has spread worldwide, and the second wave is rebounding strongly now with the coming winter that has a detrimental effect on the global economy and health. To make pre-diagnosis of COVID-19 as soon as possible, and reduce the work pressure of medical staff, making use of deep learning networks to detect positive CXR images of infected patients is a critical step. However, there are complex edge structures and rich texture details in the CXR images susceptible to noise that can interfere with the diagnosis of the machines and the doctors. Therefore, in this paper, we proposed a novel multi-resolution parallel residual CNN (named MPR-CNN) for CXR images denoising and special application for COVID-19 which can improve the image quality. The core of MPR-CNN consists of several essential modules. (a) Multi-resolution parallel convolution streams are utilized for extracting more reliable spatial and semantic information in multi-scale features. (b) Efficient channel and spatial attention can let the network focus more on texture details in CXR images with fewer parameters. (c) The adaptive multi-resolution feature fusion method based on attention is utilized to improve the expression of the network. On the whole, MPR-CNN can simultaneously retain spatial information in the shallow layers with high resolution and semantic information in the deep layers with low resolution. Comprehensive experiments demonstrate that our MPR-CNN can better retain the texture structure details in CXR images. Additionally, extensive experiments show that our MPR-CNN has a positive impact on CXR images classification and detection of COVID-19 cases from denoised CXR images.

Keywords COVID-19 · CXR images denoising · Multi-resolution · Residual CNN

1 Introduction

Compared with computed tomography (CT), CXR is not only cheap but also has lower radiation, that can reduce the harm to human. COVID-19 is a respiratory disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [1, 2] and has spread rapidly around the world. The epidemic grows even faster with severe new variants of

coronavirus and has greatly damaged the global economy and health. Until May 2021, more than 160 million confirmed cases and 3.32 million deaths have been registered in more than 200 countries and territories. During the diagnosis and assessment of disease progression, the radiologists can perform multiple CXR examinations on the patient to accurately evaluate the curative effect since most COVID-19 infected patients were diagnosed with pneumonia [3]. However, with the increase of confirmed cases, it is not only a huge burden and time-consuming process performed for radiologists to check CXR images, but also difficult to ensure the accuracy of evaluation since annotations of CXR images are often highly influenced by clinical experience [4].

Recently, deep learning [5–11] models have attained significant advancements in the field of medical image analysis by training on enough labeled data and fine-tuning its millions of parameters [12, 13]. Therefore, it is becoming more and more important to use deep learning models to analyze CXR images of COVID-19 infected patients, to relieve the

✉ Yu Zhu
zhuyu@ecust.edu.cn

✉ Dawei Yang
yang_dw@hotmail.com

¹ School of Information Science and Technology, East China University of Science and Technology, Shanghai 200237, People's Republic of China

² Department of Pulmonary and Critical Care Medicine, Zhongshan Hospital, Fudan University, Shanghai 200032, People's Republic of China

shortage of medical resources and the overload of doctors. Here, Ouyang et al. [14] used a dual-sampling attention network to detect of COVID-19 cases. [15] proposed a novel PSSPN model for classification between COVID-19, secondary pulmonary tuberculosis, community-captured pneumonia, and healthy subjects. DenseNet-OTLS method [16] achieved better performances than state-of-the-art approaches in diagnosing COVID-19. [17, 18] both utilized CNN to segment COVID-19 infection in CT images. And Shi et al. [19] make a review of imaging data acquisition, segmentation, and diagnosis for COVID-19 using AI (artificial intelligence). The above works are all typical methods of COVID-19 image analysis.

Nevertheless, there are various types of noise in CXR images, such as ground-glass opacity, bilateral abnormalities, and interstitial abnormalities. Especially, low-dose CXR images susceptible to noise are complicated and fuzzy likely to interfere with the diagnosis of machines and doctors [20]. Therefore, obtaining clearer details in CXR images and improving the images quality by denoising is of great significance [21, 22].

Due to the high practical value, the medical image denoising method [20, 23–27] has been extensively studied for a long time. Mondal et al. [28] and Raj et al. [29] used discrete wavelet technology [30] for medical image denoising. The methods are simple to calculate and run faster, but they both had an unsatisfactory performance in removing Gaussian white noise (GWN) widely existing in medical images. In addition to classic filtering [31–33] and transform domain medical images denoising method [24, 25, 34], non-local mean (NLM) [35, 36] and block-matching and 3D filtering (BM3D) [37, 38] based on the self-similarity show promising denoising performance.

Although traditional medical image denoising algorithms can improve the quality of medical images to a certain extent, they usually need to manually selected parameters and complex optimized algorithms [39], and enable to preserve texture details effectively [20]. Recently, deep learning methods [40–46], given enough data, have significant advances in images denoising than those traditional hand-crafted methods. They are significantly different in several key respects. First of all, deep learning methods do not need to manually adjust the parameters and complicated optimization algorithm. Moreover, deep learning methods can be competent for many varied noise tasks through different training data. However, the proposed methods above still have some obvious weaknesses. (1) Most of these methods ignore the connection between shallow layers and deep layers. (2) Some of these deep networks fail to extract information from feature maps effectively. (3) Lack of efficient multi-resolution feature fusion method. Given these, in this paper, we proposed a novel multi-resolution parallel residual CNN for CXR images denoising. There is spatial

information in the shallow layers with high resolution and semantic information in the deep layers with low resolution. We utilize the multi-resolution parallel convolution streams to connect the spatial and semantic information. The ECSA module is proposed to make the network focus more on texture details in CXR images with fewer parameters. We usually directly add or concatenate multiple resolution feature maps. However, they both provide limited expressive power to the network. Therefore, we design the AMFF method based on attention to improve the expression of the network.

The main contributions of this work are summarized as follows:

- (1) Multi-resolution parallel convolution flows are used to fuse information from high-resolution and low-resolution features. It is also used to enhance the robustness of the model.
- (2) An ECSA model combining effective channel and spatial attention is proposed to make the network pay more attention to the texture details of CXR images while reducing the parameters.
- (3) To improve the representation of the network, an attention-based AMFF method is used, which adaptively fuses multi-resolution features, rather than simply combining and summing features.
- (4) To verify the impact of the MPR-CNN, we design abundant experiments for CXR images classification. The outstanding results demonstrate the ability of our network to detect of COVID-19 cases from denoised CXR images.

The remainder of this paper is organized as follows. Section 2 provides a brief survey of related work. In Sect. 3, our MPR-CNN was first presented and then illustrates the loss function and optimization. In Sect. 4, extensive experiments are conducted to evaluate. Finally, several summaries and future work are given in Sect. 5.

2 Related work

In this paper, we proposed the MPR-CNN model for CXR images denoising. With the rapidly growing CXR images of confirmed cases, there is a pressing necessity to enhance the images quality for improved COVID-19 detection. To better understand the composition and the core of the model, we briefly describe the representative methods for each of the central studied problems.

2.1 Deep learning methods for images denoising

Deep learning has become a dominant machine learning method in image processing, such as image classification [7],

image recognition [47], and image denoising, which have demonstrated great potential and remarkable performance due to flexible and powerful plug-in components in deep learning [39]. Burger et al. [48] first utilized the multilayer perceptron (MLP) for image denoising and the extensive experiments demonstrate that MLP has similar or even better representation power than the hand-crafted BM3D. Besides, GANs [45, 46] that are frameworks to estimate generative models are also fine choices to suppress the noise. Generally, the framework consists of a generative network (G) and a discriminative network (D), ruling the game theory.

In terms of improving the efficiency of denoising, CNNs can be regarded as a modular part, and some classic optimization methods can be inserted to restore potential clean images, which is effective for processing noisy images. DnCNN [43] and IRCNN [49] both use a full convolution network with a signal-scale feature for image denoising. An encoder–decoder method was utilized in [50–53]. First, the input is gradually mapped to the low resolution representation, and then the stepwise reverse mapping is applied to the original resolution. Although these CNNs have achieved progressive results, they still have limitations. Full convolution networks do not use any downsampling operations, so the feature maps have more precise spatial details. However, these networks are less efficient in encoding contextual information due to their limited acceptance field. On the other hand, encoder–decoder methods lost fine spatial details, although gaining more context information.

2.2 Multi-resolution features fusion

Multi-resolution features fusion is an important process to improve the denoising of CXR images. The low-level features with higher resolution, contain more position and detailed information. However, they have less semantic information and more noise due to less convolution. In contrast, high-level features with richer feature information, but the resolution is very low, and the perception of details is unsatisfactory. The purpose of feature fusion is to merge the features extracted from the input into new features that are more expressive than the original one. The classic feature fusion methods are mainly divided into summation [54, 55] and concatenation [56]. Assuming the dimension of the two respective input features are p and q , and the dimension of the output feature Z by concatenation is shown in Eq. (1).

$$\text{Dim}(Z) = p + q \quad (1)$$

The number of channels is increased, but the information in each channel is maintained the same. In contrast, assuming the two respective input features are x and y , and the value of output characteristic Z is shown in Eq. (2). Here, λ represents a constant.

$$Z = x + \lambda y \quad (2)$$

However, they both provide limited expressive power to the network. Inspired by this reason, we design the AMFF method based on attention to improve the expression of the network.

2.3 Attention mechanism

Recently, lots of works [57–60] utilize channel attention or spatial attention to improve the performance of deep learning as an effective module. Hu et al. [57] first proposed a squeeze and excitation network (SENet) to pay attention to the relationship between channels. The weight of each channel is squeezed by global average pooling (GAP) and fully connection layers. Zhang et al. [60] propose a residual non-local attention network to address the issue that the uneven distribution of information in the corrupted images. [59] combines the channel and spatial attention to improve the feature extraction ability of networks.

The attention mechanism enables the network to learn where to concentrate and promotes the network to focus on the target object. The channel attention mechanism enhances or suppresses different channels for different tasks, by modeling the weights of each feature channel. The essence of spatial attention is to locate the target and perform some transformations or obtain weights. These attention mechanisms can improve the expression of the features by establishing dependencies between channels, or weighted spatial attention masks. However, these methods still need a large cost on memory and computation complexity.

3 Proposed method

In this section, we introduce the proposed CXR images denoising network MPR-CNN in detail, containing MNEB, ECSA, and AMFF. The ECSA module is designed to make the network focus more on texture details in CXR images and reduce the parameters by 1D convolution instead of full connection layer. The AMFF module based on attention, rather than simple concatenation or summation for feature fusion, is utilized to improve the expression of the network. The MNEB is utilized for fusing information from high and low resolution features, which is included the ECSA and the AMFF. Also, the whole network uses residual blocks to reduce the difficulty of network learning. Further, the SSIML1 loss and the cosine annealing strategy [61] are set to train our MPR-CNN. We will describe these methods in later subsections.

3.1 Network architecture

The network architecture of the proposed MPR-CNN consisted of ECSA, AMFF, and MNEB is shown in Fig. 1. Here, “DS” and “US” stand for downsampling and upsampling, respectively. First, the MPR-CNN applies a convolutional layer with the filter size of $1 \times 3 \times 3 \times 48$ to extract low-level features from the input X (noisy CXR images). Then, the feature maps pass through several layers of MNEB modules that will describe in Sect. 3.2. The MNEB is the fundamental building block of MPR-CNN. Next, we use a convolutional layer with filter size of $48 \times 3 \times 3 \times 1$ again to obtain the desired residual image $R(X)$. At last, we can subtract $R(X)$ from X to get the output (denoised CXR images).

3.2 Multi-resolution noise extraction block (MNEB)

The architecture of the MNEB is shown in the dotted box above Fig. 1. The full convolution with filter size of $48 \times 3 \times 3 \times 48$ is utilized to keep more precise spatial details and performing with filter size of $48 \times 3 \times 3 \times 96$ and $4 \times$ downsampling with filter size of $48 \times 3 \times 3 \times 192$ on the original features to gain more context information. Then, we use the ECSA module that will describe in Sect. 3.3 to focus more on texture details in CXR images and reduce the parameters as well. Next, $2 \times$ upsampling with filter size of $96 \times 3 \times 3 \times 48$ and $4 \times$ downsampling with filter size of $192 \times 3 \times 3 \times 48$ are applied to restore to

original feature maps size. Further, the AMFF which is utilized to fuse multi-resolution features will be described in Sect. 3.4. Finally, a convolutional layer with filter size of $48 \times 3 \times 3 \times 48$ is applied to extract the residual information from feature maps again. The MNEB module also uses residual learning as same as the whole network to reduce the difficulty of network learning. Multi-resolution parallel convolution streams are utilized for fusing information from high and low resolution features, as well as to enhance the robustness of the model.

3.3 Efficient channel and spatial attention (ECSA)

As shown in Fig. 2, the ECSA module is made up of channel attention and spatial attention, making the network focus more on texture details in CXR images and reduce the parameters as well. The channel attention branch is designed to enhance or suppress different channels for CXR images denoising by modeling the weights of each feature channel. Global average pooling (GAP) is applied to squeeze the input feature maps $M_C \in R^{H \times W \times C}$ and yield a feature descriptor $d \in R^{1 \times 1 \times C}$. The excitation operator usually passed through two fully connected layers to dimension reduction and cross channel interaction. However, dimension reduction has side effects on the prediction of channel attention. Therefore, we utilize the 1D convolution with kernel sizes of 5 and 2 paddings to replace the two fully connected layers. The complexity of this method is tiny, and

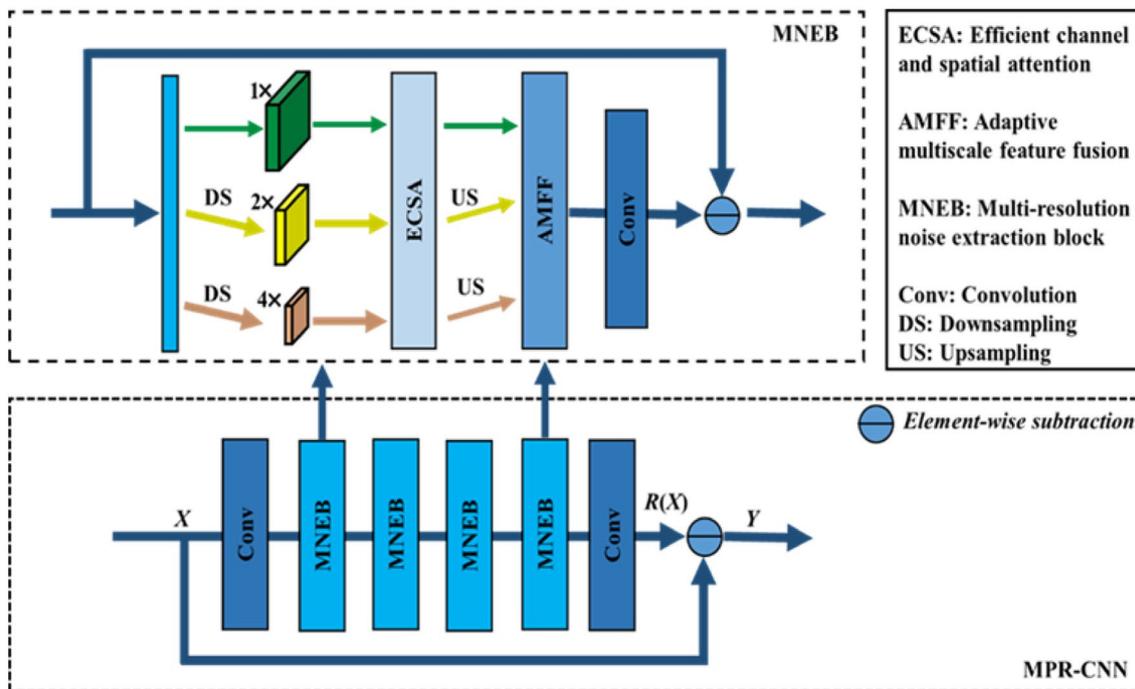


Fig. 1 Network architecture of the proposed MPR-CNN

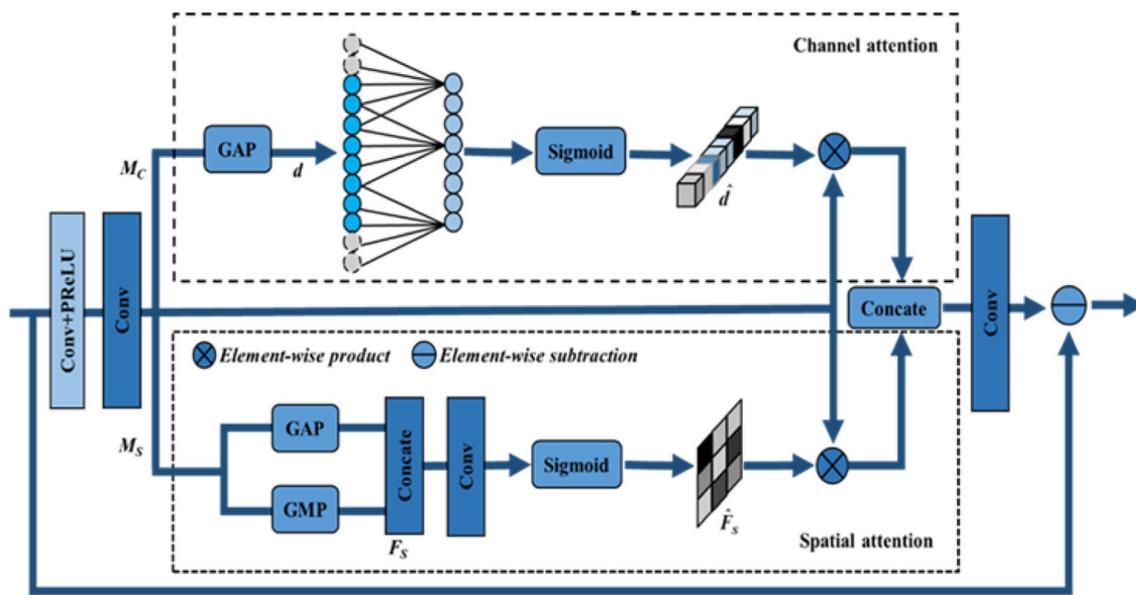


Fig. 2 Efficient channel and spatial attention

the promotion effect is significant. Next, the sigmoid gating is applied to generate activations $\hat{d} \in R^{1 \times 1 \times C}$. Finally, the output of the channel attention branch is obtained by multiplying M_C and \hat{d} .

The spatial attention branch is designed to locate the target and perform some transformations. Given a feature map $M_S \in R^{H \times W \times C}$, GAP, and global max pooling (GMP) are first applied to extract the information along the channel dimensions and then concatenating them to generate a feature map $F_S \in R^{H \times W \times 2}$. Next, the F_S passes through a convolution layer and sigmoid activation to generate a spatial attention feature map $\hat{F}_S \in R^{H \times W \times 1}$. Finally, the output of the spatial attention branch is obtained by multiplying M_S and \hat{F}_S .

The overall pipeline of the ECSA module, a convolution layer with kernel size of 3×3 is first applied to extract the low-level features and PReLU is to improve the nonlinear characteristics of the network. After another convolution layer with kernel size of 3×3 , the feature maps pass through both the channel and spatial attention in parallel. Next, we concatenated the feature maps along the spatial and channel dimensions. Finally, a convolution layer with kernel size of 3×3 is used to extract the residual information from feature maps again. The ECSA module is also a residual block.

3.4 Adaptive multi-resolution feature fusion (AMFF)

As shown in Fig. 3, we design the AMFF method based on attention rather than directly add or concatenate multiple resolution feature maps to improve the expression of the network. We first fuse the multiple resolution feature maps

by element-wise sum as shown in Eq. (3) and get the feature maps M_{in} .

$$M_{in} = M_1 + M_2 + M_3 \tag{3}$$

, where M_1 , M_2 , and M_3 represent $1 \times$, $2 \times$, and $4 \times$ feature maps, respectively. Then, the M passes through the GAP to extract the average information along the channel dimension and gain a feature descriptor $D \in R^{1 \times 1 \times C}$. Further, we use global depthwise convolution (GDC) in that the number of convolution groups is the same as the channel number, and the size of convolution kernel is the same as that of input feature map, to assign each position a learnable weight and get a new descriptor $\hat{D} \in R^{1 \times 1 \times C}$. Next, we still utilize the 1D convolution with kernel sizes of 5 and 2 padding to cross channel interaction and keep the channel dimension unchanged. Afterward, the sigmoid gating is applied to generate three different attention activations $S_1 \in R^{1 \times 1 \times C}$, $S_2 \in R^{1 \times 1 \times C}$, $S_3 \in R^{1 \times 1 \times C}$. Finally, the output M_{out} of the AMFF after recalibration and aggregation is defined in Eq. (4).

$$M_{out} = M_1 \cdot S_1 + M_2 \cdot S_2 + M_3 \cdot S_3 \tag{4}$$

3.5 Loss function and optimization

We propose the MSL1 loss to train our MPR-CNN by adding multi-scale structural similarity (MS_SSIM) [62, 63] and L1 loss. On one hand, MS-SSIM can preserve the contrast in high-frequency regions in CXR images,

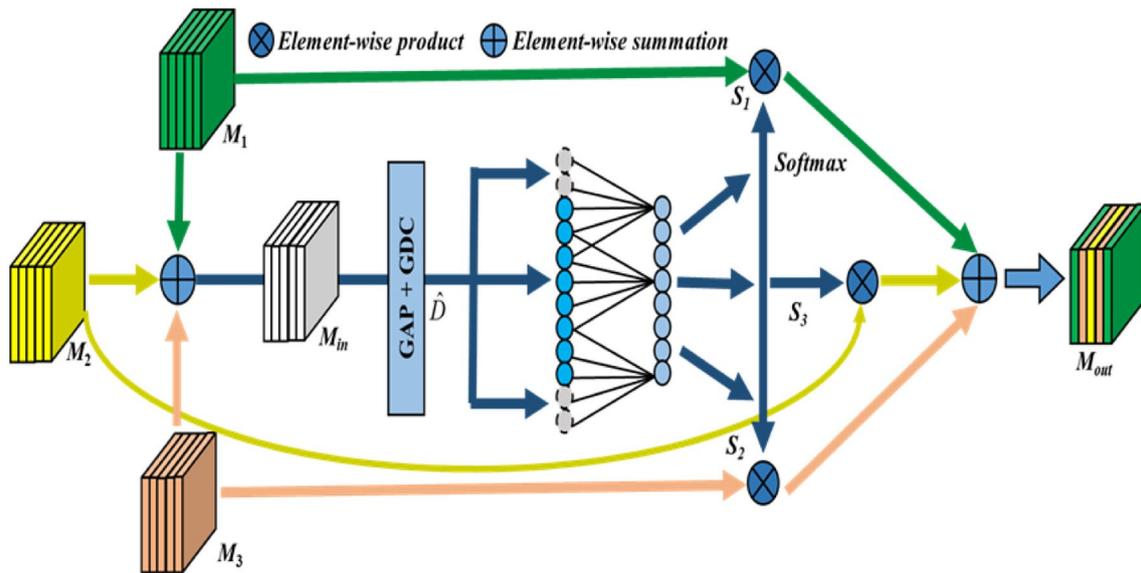


Fig. 3 Schematic for adaptive multi-resolution feature fusion

on the other, L1 loss can keep the color and brightness of CXR images. SSIM is a theoretical method based on image structure similarity and the general form of the SSIM index between signal x and y is defined as Eq. (5), where, l , c , and s represent the light, contrast, and structure. α , β and γ are parameters to allocate the weight of the three attributes in SSIM.

$$SSIM(x, y) = l(x, y)^\alpha \cdot c(x, y)^\beta \cdot s(x, y)^\gamma \tag{5}$$

Here, we set $\alpha = \beta = \gamma = 1$ and the final SSIM index is shown in Eq. (6), where μ_x and μ_y are the mean values of signal x and y , σ_x and σ_y are the variances, and σ_{xy} are the covariances of signal x and y . C1 and C2 are small normal numbers to avoid the situation where the denominator is zero in Eq. (6).

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1) (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1) (\sigma_x^2 + \sigma_y^2 + C_2)} \tag{6}$$

Furthermore, MS-SSIM performs SSIM evaluation on images of different resolutions through downsampling which can merge more structural information. Thus, the MS_SSIM loss is shown in Eq. (7), where X denotes the noisy CXR image and the Y represents the clean CXR image.

$$MS_SSIM((R(X), V(X - Y))) = \frac{1}{N} \sum_{i=1}^N \{1 - SSIM(R(X_i) - V(X_i - Y_i))\} \tag{7}$$

And, Eq. (8) describes the L1 loss.

$$L1(R(X), V(X - Y)) = \frac{1}{N} \sum_{i=1}^N |R(X_i) - V(X_i - Y_i)| \tag{8}$$

The overall MSL1 loss is given by Eq. (9), which expresses the loss error between the desired residual image $V(X-Y)$ and estimated one $R(X)$ from noisy CXR image, and the θ is a constant which is set to 0.2 for all the experiments by ablation study.

$$MSL1 = (1 - \theta) L1 + \theta * MS_SSIM \tag{9}$$

We utilize PSNR as the evaluation for our MSL1 loss, which is shown in Table 6 in Sect. 4.4.

In addition, as shown in Eq. (10), the cosine annealing strategy is set as an optimization method and decreases the learning rate from initial value $5e-4$ to $5e-6$ during training. Here, the η stands for initial value and T is empirically set as 5.

$$\eta_t = \frac{1}{2} \left(1 + \cos \left(\frac{t\pi}{T} \right) \right) \eta \tag{10}$$

4 Experiments

In this section, we first describe the datasets and then give the implementation details. Next, we compare our MPR-CNN with some state-of-the-art denoising methods. Furthermore, ablation studies are designed to explore the impact of each of our architectural components and choices on the final performance. Finally, we innovatively verify the impact of the MPR-CNN for CXR images classification.

4.1 Datasets

We evaluate the denoising performance of our MPR-CNN via COVID-19 radiography database [64], which consists of 1341 normal CXR images, 1345 CXR images of viral pneumonia, and 219 CXR images of COVID-19, collecting by several research organizations. The size of CXR images is 1024×1024 . Here, we randomly select 400 normal CXR images, 400 CXR images of viral pneumonia, and 170 CXR images of COVID-19 as training data. Then, we select 30 images in each category as validation data, and 15 images in each category as test data. To speed up the training process and keep as much detail as possible in each CXR image, we extract the training data to 128×128 and gain $16 \times 10,552$ patches (16 is the mini-batch size and 10,552 represents an iterative number for one training epoch) as training label by scaling and rotation. The main noise in CXR image is the granular noise, which is caused by the receiving device (film), and the granular noise accords with Gaussian distribution. So we decided to add white Gaussian noise (WGN) in the patches with standard deviation to simulate the low-dose noisy CXR images with $\sigma = [0, 55]$ different level noise as the input of MPR-CNN (Fig. 4).

Also, we further make classification experiments to verify the impact of the MPR-CNN for CXR images classification. To balance classification data while classifying, we collect another 605 CXR images of COVID-19 from three

datasets: (1) Fig. 1 COVID-19 chest X-ray Dataset [65], (2) COVID-19 Image Data Collection [66], and (3) ActualMed COVID-19 chest X-ray Dataset [67]. There are three types of cases that are COVID-19, normal, and viral pneumonia. The detailed component distribution of the classified dataset is shown in Table 1.

4.2 Implementation details

The proposed MPR-CNN is end-to-end trainable that does not require any pre-training of sub-modules. The model is trained with the Adam optimizer ($\beta_1 = 0.9$, and $\beta_2 = 0.999$) that is an extension of the stochastic gradient descent algorithm, and the cosine annealing strategy is set as an optimization method and decreases the learning rate from initial value $5e-4$ to $5e-6$ during training. The mini-batch size is set as 16 and 10,552 iterations for one training epoch. We train 30 epochs to fit our model. Specifically, we apply

Table 1 Component distribution of the classified dataset

Types	Training data	Test data	Total
COVID-19	665	166	824
Normal	1073	268	1341
Viral Pneumonia	1076	269	1345
Total	2814	703	3510

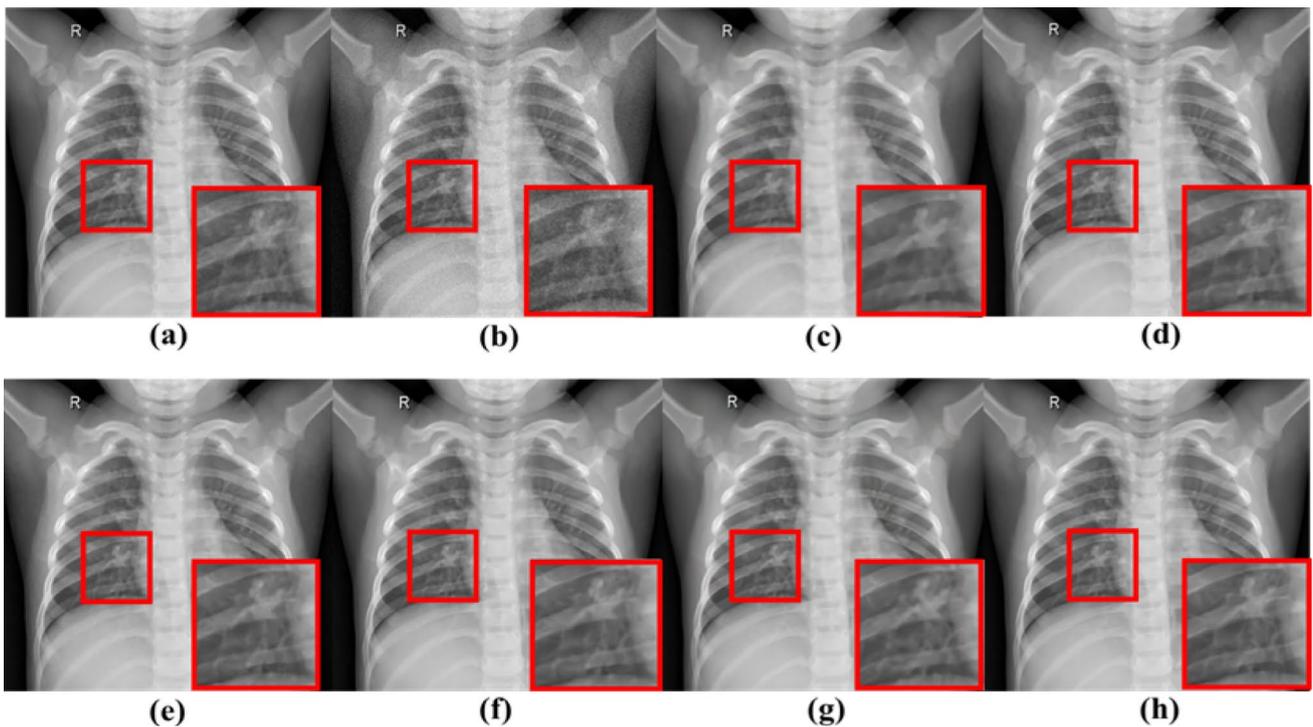


Fig. 4 Denoising results of different methods on normal CXR image with $\sigma=15$. **a** Original CXR image, **b** noisy CXR image/24.731 dB, **c** NLN/38.313 dB, **d** DnCNN/38.402 dB, **e** IRCNN/38.432 dB, **f** FFDNet/38.682 dB, **g** ESRGAN/38.786 dB, **h** MPR-CNN/38.784 dB

Pytorch 1.4.0 and Python 3.5 to train and test the proposed MPR-CNN in CXR image denoising on the Ubuntu 16.04 from a PC, composed of an Intel Core i7-7800X CPU with 3.50 GHz, a RAM 16G, and two Nvidia GeForce RTX 2080 Ti GPU.

4.3 Comparisons with state-of-the-art denoising methods

In this subsection, we test the denoising performance of our MPR-CNN in terms of both subjective and objective evaluation, comparing with 7 state-of-the-art denoising methods, such as NLM [35], BM3D [37], DnCNN [43], IRCNN [49], FFDNet [53], SRGAN [45], and ESRGAN [46].

In terms of subjective evaluation, we compare the denoising ability for different noise levels and different scaling factors with peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) index. PSNR is a measure to evaluate the ability of model to remove noise, while the SSIM is a measure of the similarity of two images. The value is higher, the corresponding denoising method has a better performance. Table 2 describes the average PSNR (dB) and SSIM of different methods on test data with different noise levels of 15, 25, and 40. The proposed MPR-CNN achieves the best performance on noise levels of 15, 25, and 50 in SSIM, although the value of PSNR is a little bit lower than ESPGAN when $\sigma = 15$. Also, when $\sigma = 25$ the value of PSNR is 0.84 dB, 0.627 dB, and 0.034 dB more than BM3D, DnCNN, and ESRGAN, respectively. Especially, the value of SSIM is 0.037, 0.020, and 0.011 more than NLM, IRCNN, and SRGAN. It is noted that our MPR-CNN achieves excellent results on denoised tasks of different noise levels (Fig. 5).

Table 3 describes the average PSNR (dB) and SSIM of different methods on three types of CXR images with noise levels of 30. It is noted that our MPR-CNN is also superior to competing methods on each type of CXR image and has

a better denoising performance on types of viral pneumonia and COVID-19 than the normal type.

Moreover, to evaluate the ability of the proposed model for the blind Gaussian denoising, we also added WGN to Fig. 6a with standard deviation $\sigma = \{10, 15, 20, 25, 30, 35, 40, 45\}$ and the line chart of PSNR and SSIM are shown in Fig. 7. The blue solid lines represent the denoising result of our MPR-CNN, and one can clearly see that the values of PSNR and SSIM of our MPR-CNN are higher than other competing methods at most time, although the value of PSNR is a little bit lower than ESPGAN when $\sigma < 25$. Figure 7 demonstrates that our proposed MPR-CNN is robust for the blind CXR images denoising.

Average PSNR (dB) and SSIM of different methods on test data with different scaling factors are shown in Table 4. Here, we set three scaling factors, $\times 2$, $\times 4$, and $\times 8$, and the corresponding CXR images sizes are 512×512 , 256×256 , and 128×128 . The noise level is still set to 25. According to Table 3, it is noted that our MPR-CNN has a better performance in denoising CXR images with different scaling factors than other methods.

For computation time, we select 4 state-of-the-art denoising methods to perform the test for CXR images denoising. The size of the CXR image is set as 128×128 , 256×256 , and 512×512 as illustrated in Table 5. From that, we can find that the inference time of our MPR-CNN is very competitive in contrast to other popular methods.

4.4 Ablation studies

We design ablation studies to explore the impact of each of our architectural components and choices on the final performance. All the ablation experiments use the same test data, adding WGN with standard deviation $\sigma = 25$ to simulate the low-dose noisy CXR images. First, we analyze the impact of different loss function for denoising CXR images in Table 6. It shows that the proposed MSL1 loss with $\theta = 0.2$ has the

Table 2 Average PSNR (dB) and SSIM of different methods on test data with different noise levels of 15, 25, and 40

Methods	NLM	BM3D	DnCNN	IRCNN	FFDNet	SRGAN	ESRGAN	MPR-CNN
$\sigma = 15$								
PSNR	37.751	37.964	38.114	38.203	38.463	38.456	38.653	38.482
SSIM	0.901	0.919	0.900	0.913	0.921	0.913	0.916	0.922
$\sigma = 25$								
PSNR	35.872	36.040	36.253	36.452	36.732	36.661	36.846	36.880
SSIM	0.878	0.892	0.873	0.896	0.903	0.894	0.902	0.905
$\sigma = 40$								
PSNR	34.067	34.249	34.434	34.563	35.023	34.986	35.136	35.286
SSIM	0.854	0.874	0.850	0.871	0.885	0.880	0.887	0.891

Bold vaue indicates that the best denoising results

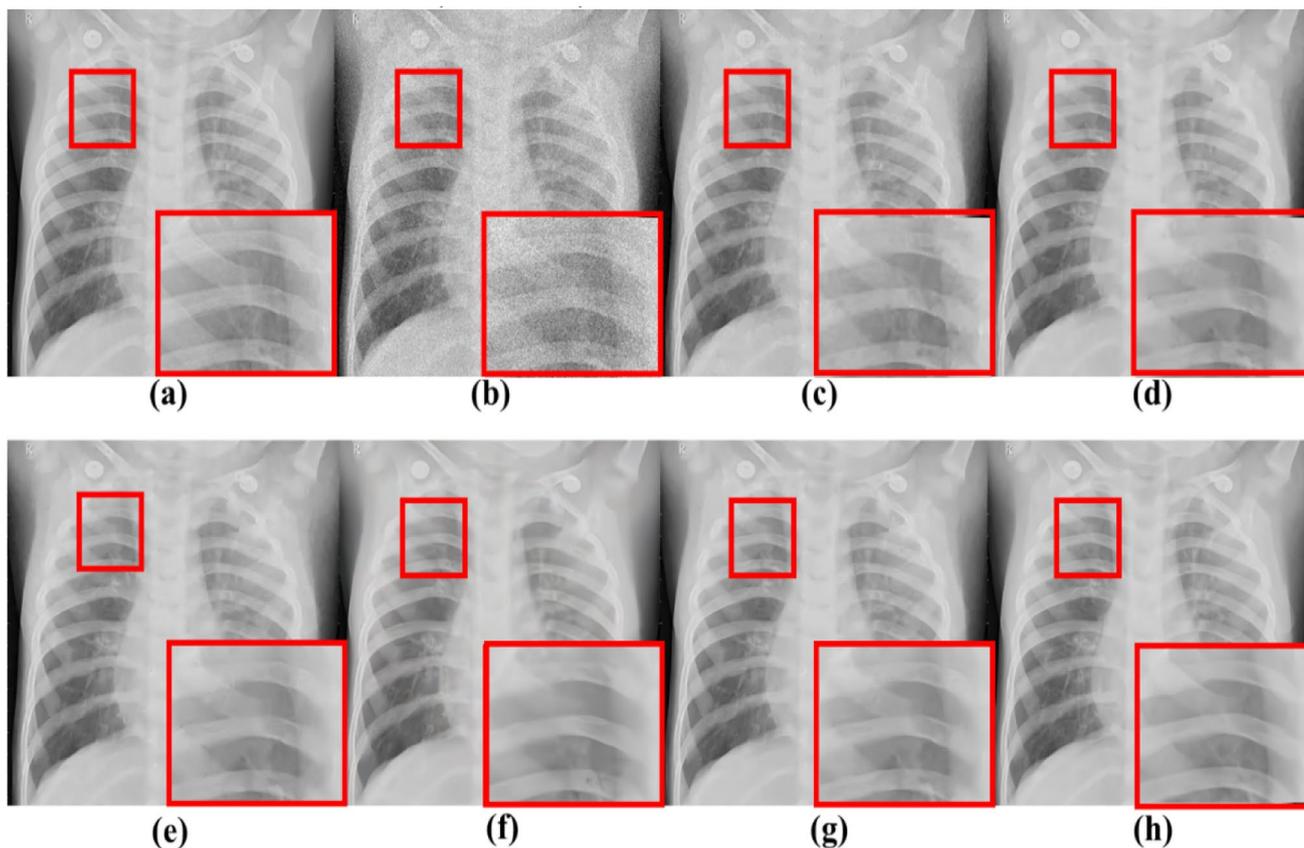


Fig. 5 Denoising results of different methods on CXR image of viral pneumonia with $\sigma=25$. **a** Original CXR image, **b** noisy CXR image/20.306 dB, **c** BM3D/36.482 dB, **d** DnCNN/36.931 dB, **e** IRCNN/36.973 dB, **f** SRGAN/37.154 dB, **g** FFDNet/37.263 dB, **h** MPR-CNN/37.577 dB

Table 3 Average PSNR (dB) and SSIM of different methods on three types of CXR images with noise levels of 30

Methods	DnCNN	IRCNN	FFDNet	ESRGAN	MPR-CNN
<i>Normal</i>					
PSNR	34.702	34.683	35.263	35.373	35.484
SSIM	0.830	0.858	0.878	0.880	0.883
<i>Viral pneumonia</i>					
PSNR	36.192	36.053	36.771	36.971	37.011
SSIM	0.884	0.898	0.917	0.919	0.922
<i>COVID-19</i>					
PSNR	35.761	35.544	36.231	36.296	36.368
SSIM	0.876	0.877	0.892	0.894	0.895

Bold vaue indicates that the best denoising results

most outstanding denoised performance than other loss function, which increases 0.243 dB more than L1 loss and 0.46 dB more than MS-SSIM. Furthermore, the SSIM also has a certain promotion. It could be concluded that MSL1 loss can preserve the contrast in high-frequency regions in CXR images and keep the color and brightness as well.

Then, we study the influence of the number of multi-resolution streams in the MNEB for the CXR images denoising quality in Table 7. According to Table 7, we can note that the MNEB with two different resolution streams is better than one of single one and three different resolution streams have the best performance. Therefore, it could be concluded that increasing the number of streams can provide a significant improvement for CXR images denoised and the MNEB is important to improve the CXR images quality.

Finally, in Table 8, we make ablation studies on the impact of proposed ECSA and AMFF for CXR images denoised. From the first three columns, we can note that the AMFF based on attention, rather than simple concatenation or summation for feature fusion, can improve the expression of the network, which increases by 0.318 dB more than summation and 0.108 dB more than concatenation. Moreover, it is also evident from Table 8 that the ECSA module has a positive effect on our MPR-CNN, which, respectively, increases by 0.019, 0.020, 0.020 in SSIM.

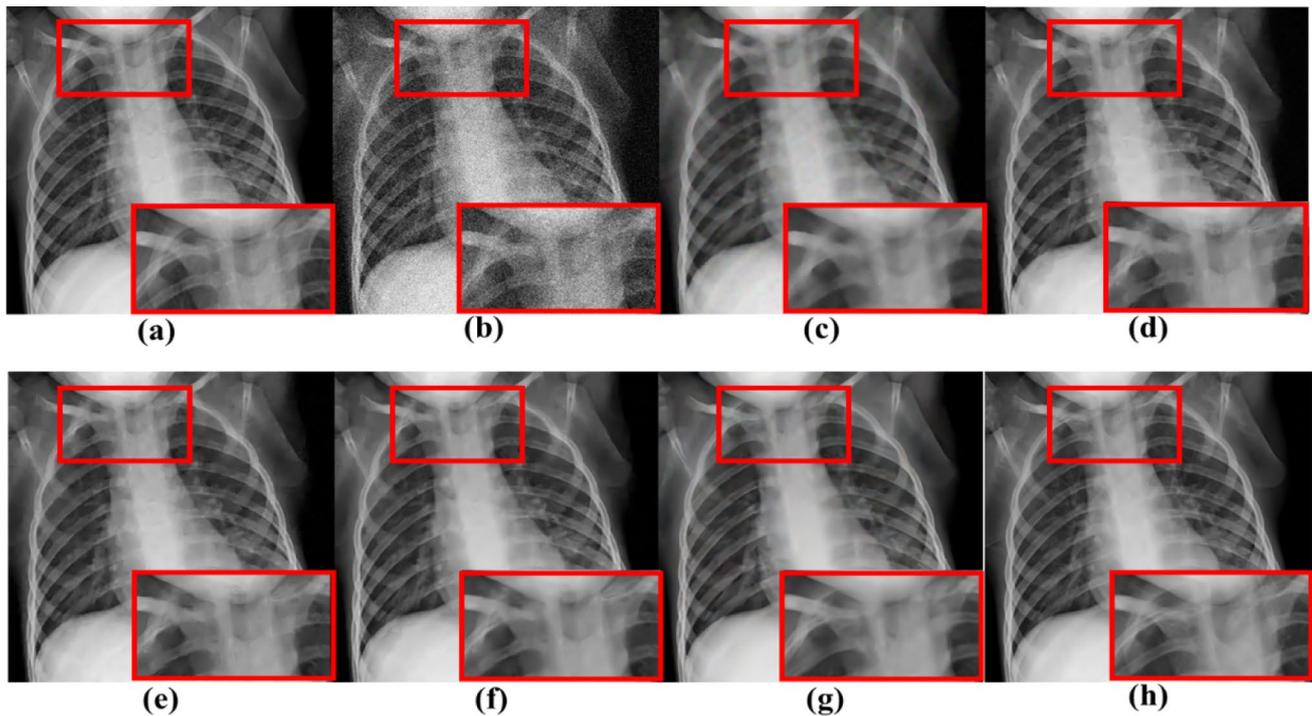


Fig. 6 Denoising results of different methods on CXR image of COVID-19 with $\sigma=40$. **a** Original CXR image, **b** noisy CXR image/16.732 dB, **c** NLM/33.566 dB, **d** DnCNN/34.403 dB, **e** IRCNN/34.982 dB, **f** FFDNet/35.192 dB, **g** ESRGAN/35.626 dB, **h** MPR-CNN/35.632 dB

Extensive ablation experiments prove that the proposed MS-SSIM loss can preserve more detailed information in CXR images as well as the MNEB, and the ECSA and AMFF both have positive influences on the final CXR images quality.

4.5 Verify the impact of the MPR-CNN for CXR images classification

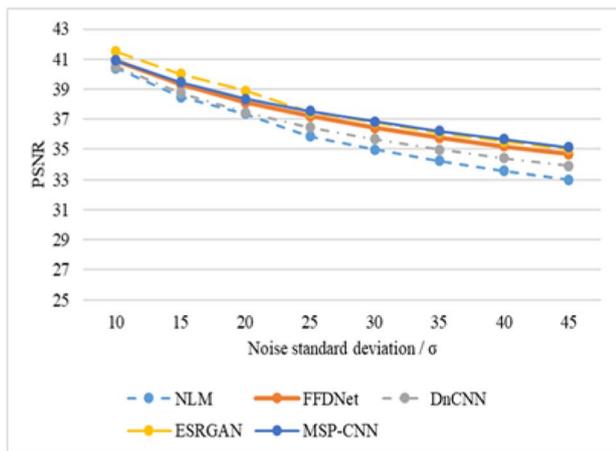
In this subsection, we not only make the denoising experiments but also use the denoised CXR images by MPR-CNN to classify CXR images. To evaluate the effectiveness of the MPR-CNN, we use three classic classified networks: ResNet18 [47], VGG19 [68], DenseNet121 [69] (Fig. 8).

We use a classified dataset that has been introduced in Sect. 4.1 for CXR images classification. The images size are set to 512×512 , which adding WGN with standard deviation $\sigma = 20$ to simulate the low-dose noisy CXR images. Here, the vertical data represents the true value, while horizontal data stands for the predicted one. Especially, the number of diagonals represents the correct classifications. Moreover, the normal cases correctly classified, respectively, increase by 5 and 6 after denoising using VGG19 and DenseNet121. Then, the viral Pneumonia cases correctly classified, respectively, increase by 63 and 38 after denoising using ReseNet18 and VGG19. Especially, the correctly classified COVID-19 cases, respectively, increase by 11, 17, and 4

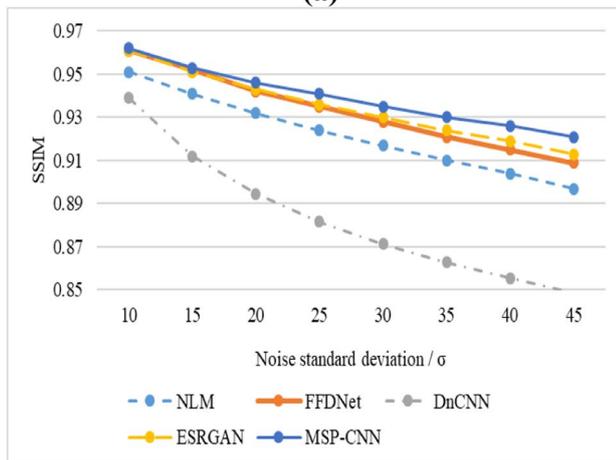
using ResNet18, VGG19, and DenseNet121. Hence, it can clearly note that the MPR-CNN has a positive impact on the CXR images classification.

The classification effects between different models denoised by DnCNN and MPR-CNN are shown in Table 9. To quantify the classified networks, we calculated the test accuracy (ACC), sensitivity (SEN), and precision (PRE) of each infection type on the above classified dataset. Here, the higher value the SEN corresponds to the lower the probability of missing positive cases. Moreover, the higher value the PRE results in the lower the probability of misdiagnosing negative cases. After denoising by MPR-CNN, the ACCs of the ResNet18, VGG19, and DenseNet121 are, respectively, improved by 8.96%, 8.53%, and 8.52%, while the PREs, respectively, improved by 7.41%, 7.26%, and 7.37%. Comparing to DnCNN, the SENs have improved by 0.56%, 1.13%, and 1.56%, respectively, using ResNet18, VGG19, and DenseNet121. Meanwhile, the classification performance of denoised CXR image by MPR-CNN is very close to original one. The ACCs have just decreased by 0.57%, 0.30%, and 0.32% using ResNet18, VGG19, and DenseNet121.

Furthermore, it could be concluded that classification models fed into CXR images by MPR-CNN have a lower probability of missing COVID-19 cases, as well as a lower probability of misdiagnosing negative cases.



(a)



(b)

Fig.7 Value of PSNR and SSIM of denoised results of Fig. 6a using different method

Table 4 Average PSNR (dB) and SSIM of different methods on test data with different scaling factors

Methods	NLM	BM3D	DnCNN	IRCNN	FFDNet	SRGAN	ESRGAN	MPR-CNN
×2								
PSNR	34.083	34.235	34.672	34.832	35.032	35.021	35.613	35.606
SSIM	0.860	0.872	0.861	0.883	0.888	0.886	0.900	0.901
×4								
PSNR	31.741	31.885	32.283	32.373	32.432	32.316	32.513	32.950
SSIM	0.831	0.843	0.833	0.847	0.851	0.848	0.855	0.871
×8								
PSNR	29.181	29.319	29.729	29.852	29.783	29.701	29.713	29.730
SSIM	0.814	0.823	0.814	0.820	0.822	0.809	0.812	0.831

Bold vau indicates that the best denoising results

Table 5 Computation time of 4 popular denoising methods for the noisy images of sizes 256×256, 512×512, and 1024×1024

Method/s	128×128	256×256	512×512
BM3D	0.284	0.562	2.486
FFDNet	0.032	0.062	0.134
ESRGAN	0.078	0.163	0.378
MPR-CNN	0.029	0.061	0.127

5 Conclusion

In this paper, we propose a novel MPR-CNN for CXR images denoising and special application for COVID-19 that can improve the images quality. Multi-resolution parallel convolution streams are utilized for fusing information from both high and low resolution features. The ECSA module is proposed to make the network focus more on texture details in CXR images as well as to reduce the parameters. The AMFF method based on attention is utilized to improve the expression of the network rather than simple concatenation or summation for feature fusion. The MSL1 loss is utilized to preserve the contrast in high-frequency regions in CXR images and keep the color and brightness as well. The extensive experiments demonstrate that all the proposed methods have significant impacts on CXR images denoising. Comparing to competing methods, our MPR-CNN has the best performance in both subjective visual evaluation and objective indicators. It is noted that our proposed MPR-CNN is very robust for blind CXR images denoising. Moreover, extensive experiments show that the proposed MPR-CNN has a positive impact on CXR images classification and detection of COVID-19 cases from denoised CXR images.

Table 6 Average PSNR and SSIM of denoised CXR images by different loss functions

Loss function	L1 loss	MS-SSIM	MSL1 ($\theta = 0.1$)	MSL1 ($\theta = 0.2$)	MSL1 ($\theta = 0.3$)
PSNR/dB	36.727	36.420	36.849	36.880	36.773
SSIM	0.900	0.892	0.903	0.905	0.901

Bold vaue indicates that the best denoising results

Table 7 Result of different components of multi-resolution streams in MNEB

1×	✓			✓	✓		✓
2×		✓		✓		✓	✓
4×			✓		✓	✓	✓
PSNR/dB	35.732	36.171	35.900	36.405	36.226	36.607	36.880
SSIM	0.875	0.886	0.880	0.891	0.888	0.897	0.905

Bold vaue indicates that the best denoising results

Table 8 Influence of individual components of MNEB

ECSA				✓	✓	✓
Summation	✓			✓		
Concatenation		✓			✓	
AMFF			✓			✓
PSNR/dB	35.808	36.018	36.126	36.567	36.796	36.880
SSIM	0.877	0.882	0.885	0.896	0.902	0.905

Bold vaue indicates that the best denoising results

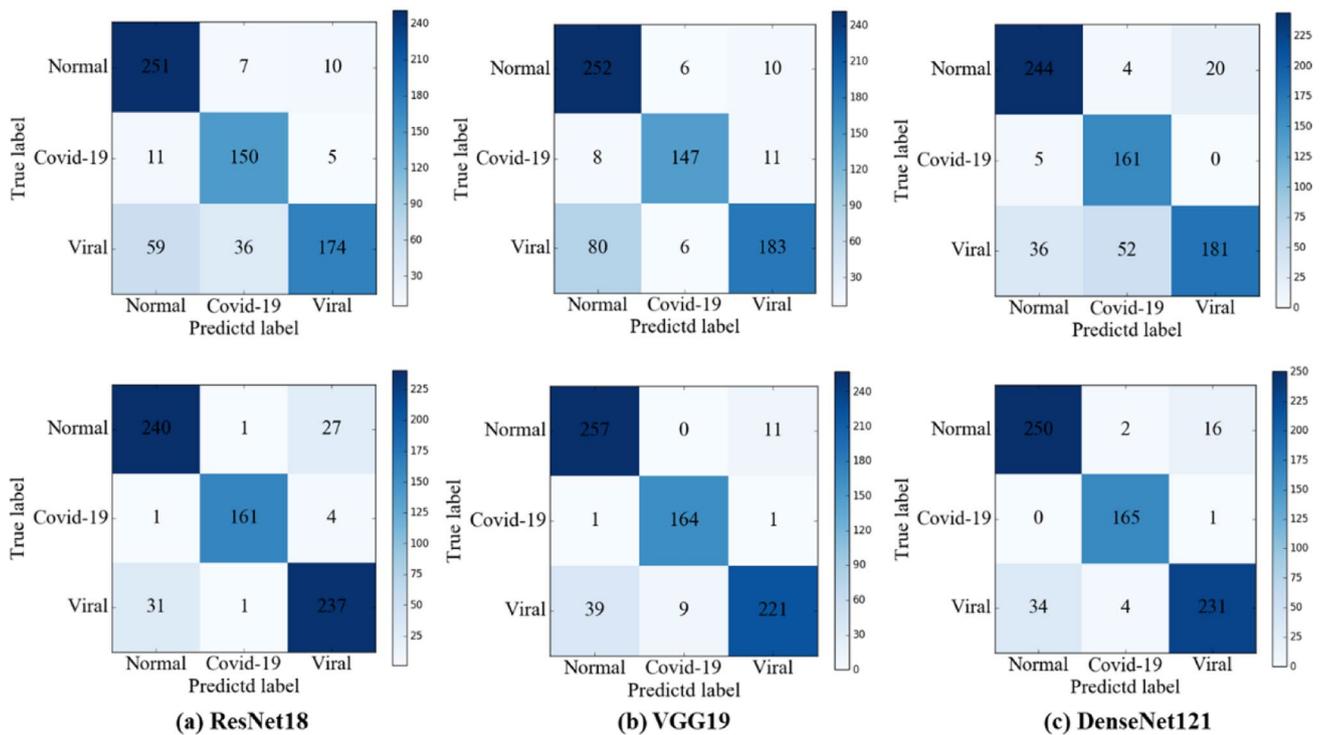


Fig. 8 Comparison of the confusion matrices of different models denoised by MPR-CNN. First line: classification using noisy CXR images. Second line: denoised CXR images by MPR-CNN, then feed into classification models

Table 9 Comparison of classification effects between different models denoised by DnCNN and MPR-CNN

Methods	Noisy CXR images			Denoised by DnCNN			Denoised by MPR-CNN			Original CXR images		
	ACC	SEN	PRE	ACC	SEN	PRE	ACC	SEN	PRE	ACC	SEN	PRE
ResNet18	81.79	81.79	83.39	90.18	90.19	90.19	90.75	90.75	90.80	91.32	91.32	91.45
VGG19	82.79	82.79	84.41	90.19	90.19	90.35	91.32	91.32	91.67	91.62	91.61	91.91
DenseNet 121	83.37	83.36	84.62	90.33	90.33	90.56	91.89	91.89	91.99	92.21	92.30	92.18

On the whole, the proposed MPR-CNN can provide a more clear and rigorous diagnostic basis both for radiologists and machines. We will continue to focus on the development of COVID-19, and our future work will concentrate on effectively reducing the noise artifacts in COVID-19 CXR images with the current powerful method. Improving the quality of COVID-19 CXR images, to classify and detect of COVID-19 cases more accurately from denoised CXR images.

Acknowledgements The authors greatly appreciate the financial supports of Shanghai Pujiang Program (20PJ1402400), Zhongshan Hospital Clinical Research Foundation (2016ZSLC05), Science and Technology Commission of Shanghai Municipality (20DZ2261200) and Shanghai Engineer & Technology Research Center of Internet of Things for Respiratory Medicine (20DZ2254400).

Declarations

Conflict of interest The author(s) declared no conflicts of interest with respect to the research, authorship, and publication of this paper.

References

1. Waheed, A., Goyal, M., Gupta, D., et al.: CovidGAN: data augmentation using auxiliary classifier GAN for improved Covid-19 detection. *IEEE Access* **8**(8), 91916–91923 (2020)
2. Sohrabi, C., Alsafi, Z., O’Neill, N., et al.: World Health Organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). *Int. J. Surg.* **76**, 71–76 (2020)
3. Oh, Y., Park, S., Ye, J.C.: Deep learning COVID-19 features on CXR using limited training data sets. *IEEE Trans. Med. Imaging* **39**(8), 2688–2700 (2020)
4. Ma, J., Wang, Y., An, X., et al.: Towards efficient COVID-19 CT annotation: a benchmark for lung and infection segmentation. [arXiv:2004.12537](https://arxiv.org/abs/2004.12537) (2020)
5. Lecun, Y., Bengio, Y., Hinton, G.: Deep learning. *Nature* **521**(7553), 436–444 (2015)
6. Schmidhuber, J.: Deep learning in neural networks: an overview. *Neural Netw.* **61**, 85–117 (2015)
7. Krizhevsky, A., Sutskever, I., Hinton, G.: ImageNet classification with deep convolutional neural networks. In: *Proceedings of 26th Conference on Neural Information Processing Systems (NIPS)*, pp. 1097–1105 (2012)
8. Yan, C., Gong, B., Wei, Y., et al.: Deep multi-view enhancement hashing for image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **43**(4), 1445–1451 (2020)
9. Yan, C., Shao, B., Zhao, H., et al.: 3D room layout estimation from a single RGB image. *IEEE Trans. Multimed.* **22**(11), 3014–3024 (2020)

10. Yan, C., Li, Z., Zhang, Y., et al.: Depth image denoising using nuclear norm and learning graph model. *ACM Trans. Multimed. Comput. Commun. Appl.* **16**(4), 1–17 (2020)
11. Yan, C., Hao, Y., Li, L., et al.: Task-adaptive attention for image captioning. *IEEE Trans. Circuits Syst. Video Technol.* (2021). <https://doi.org/10.1109/TCSVT.2021.3067449>
12. Litjens, G., Kooi, T., Bejnordi, B.E., et al.: A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017)
13. Wang, S., Tang, C., Sun, J., et al.: Cerebral micro-bleeding detection based on densely connected neural network. *Front. Neurosci.* **13**, 422–432 (2019)
14. Ouyang, X., Huo, J., Xia, L., et al.: Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia. *IEEE Trans. Med. Imaging* **39**(8), 2595–2605 (2020)
15. Wang, S.H., Zhang, Y., Cheng, X., et al.: PSSPNN: PatchShuffle Stochastic Pooling Neural Network for an explainable diagnosis of COVID-19 with multiple-way data augmentation. *Comput. Math. Methods Med.* **2021**, 1–18 (2021)
16. Zhang, Y.-D., Satapathy, S.C., Zhang, X., et al.: Covid-19 diagnosis via DenseNet and optimization of transfer learning setting. *Cognit. Comput.* (2021). <https://doi.org/10.1007/s12559-020-09776-8>
17. Jin, Q., Cui, H., Sun, C., et al.: Domain adaptation based self-correction model for COVID-19 infection segmentation in CT images. *Expert Syst. Appl.* **176**, 114848 (2021)
18. Fan, D.P., Zhou, T., Ji, G.P., et al.: Inf-net: automatic covid-19 lung infection segmentation from CT images. *IEEE Trans. Med. Imaging* **39**(8), 2626–2637 (2020)
19. Shi, F., Wang, J., Shi, J., et al.: Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. *IEEE Rev. Biomed. Eng.* **14**, 4–15 (2020)
20. Jin, Y., Jiang, X.B., Wei, Z.K., et al.: Chest X-ray image denoising method based on deep convolution neural network. *IET Image Proc.* **13**(11), 1970–1978 (2019)
21. Wang, C., Elazab, A., Jia, F., et al.: Automated chest screening based on a hybrid model of transfer learning and convolutional sparse denoising autoencoder. *Biomed. Eng. Online* **17**(1), 63 (2018)
22. Lee, D., Choi, S., Kim, H.J.: Performance evaluation of image denoising developed using convolutional denoising autoencoders in chest radiography. *Nucl. Instrum. Methods B* **884**(11), 97–104 (2018)
23. Wang, Y., Zhou, H.: Total variation wavelet-based medical image denoising. *Int. J. Biomed. Imaging.* **2006**, 89095–89107 (2006)
24. Rabbani, H., Nezafat, R., Gazor, S.: Wavelet-domain medical image denoising using bivariate laplacian mixture model. *IEEE. Trans. Biomed. Eng.* **56**(12), 2826–2837 (2009)
25. Satheesh, S., Prasad, K.: Medical image denoising using adaptive threshold based on contourlet transform. [arXiv:1103.4907](https://arxiv.org/abs/1103.4907) (2011)
26. Li, S., Yin, H., Fang, L.: Group-sparse representation with dictionary learning for medical image denoising and fusion. *IEEE. Trans. Biomed. Eng.* **59**(12), 3450–3459 (2012)

27. Gondara, L.: Medical image denoising using convolutional denoising autoencoders. In: Proceedings of 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW), pp. 241–246 (2016)
28. Mondal, T., Maitra, M.: Denoising and compression of medical image in wavelet 2D. *Int. J. Recent Innov. Trends Comput. Commun.* **2**(2), 1–4 (2014)
29. Raj, V.N.P., Venkateswarlu, T.: Denoising of medical images using undecimated wavelet transform. In: Proceedings of 2011 IEEE Recent Advances in Intelligent Computational Systems (RAICS), pp. 483–488 (2011)
30. Chao, R., Zhang, K., Li, Y.-J.: An image fusion algorithm using wavelet transform. *Acta Electr. Sin.* **32**(5), 750–753 (2004)
31. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Proceedings of 16th International Conference on Computer Vision (ICCV), pp. 839–846 (1998)
32. Bhonsle, D., Chandra, V., Sinha, G.: Medical image denoising using bilateral filter. *Int. J. Image Graph.* **4**(6), 36–43 (2012)
33. Chang, C.C., Hsiao, J.Y., Hsieh, C.P.: An adaptive median filter for image denoising. In: Proceedings of 2nd International Symposium on Intelligent Information Technology Application, pp. 346–350 (2008)
34. Gupta, S., Chauhan, R., Saxena, S.: Locally adaptive wavelet domain Bayesian processor for denoising medical ultrasound images using speckle modelling based on Rayleigh distribution. *IEEE Proc. Vis. Image Signal Proc.* **152**(1), 129–135 (2005)
35. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: Proceedings of 5th IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 60–65 (2005)
36. Mingliang, X., Pei, L., Mingyuan, L., et al.: Medical image denoising by parallel non-local means. *Neurocomputing* **195**, 117–122 (2016)
37. Dabov, K., Foi, A., Katkovnik, V., et al.: Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **16**(8), 2080–2095 (2007)
38. Zhao, T., Hoffman, J., McNitt-Gray, M., et al.: Ultra-low-dose CT image denoising using modified BM3D scheme tailored to data statistics. *Med. Phys.* **46**(1), 190–198 (2019)
39. Tian, C., Xu, Y., Li, Z., et al.: Attention-guided CNN for image denoising. *Neural Netw.* **124**, 117–129 (2020)
40. Anwar, S., Barnes, N.: Real image denoising with feature attention. In: Proceedings of the IEEE International Conference on Computer Vision (ECCV), pp. 3155–3164 (2019)
41. Chen, L.C., Zhu, Y., Papandreou, G., et al.: Encoder–decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818 (2018)
42. Nishio, M., Nagashima, C., Hirabayashi, S., et al.: Convolutional auto-encoder for image denoising of ultra-low-dose CT. *Heliyon* **3**(8), e00393 (2017)
43. Zhang, K., Zuo, W., Chen, Y., et al.: Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans. Image Process.* **26**(7), 3142–3155 (2017)
44. Isogawa, K., Ida, T., Shiodera, T., et al.: Deep shrinkage convolutional neural network for adaptive noise reduction. *IEEE Signal Process. Lett.* **25**(2), 224–228 (2018)
45. Ledig, C., Theis, L., Huszár, F., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4681–4690 (2017)
46. Wang, X., Yu, K., Wu, S., et al.: ESRGAN: enhanced super-resolution generative adversarial networks. In: Proceedings of the European Conference on Computer Vision (ECCV) (2018)
47. He, K., Zhang, X., Ren, S., et al.: Deep residual learning for image recognition. In: Proceedings of IEEE Conference on Computer Vision Pattern Recognition (CVPR), pp. 770–778 (2016)
48. Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising: can plain neural networks compete with BM3D? In: Proceedings of IEEE Conference on Computer Vision Pattern Recognition (CVPR), pp. 2392–2399 (2012)
49. Zhang, K., Zuo, W., Gu, S., et al.: Learning deep CNN denoiser prior for image restoration. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3929–3938 (2017)
50. Kupyń, O., Martyniuk, T., Wu, J., et al.: DeblurGAN-v2: deblurring (orders-of-magnitude) faster and better. In: Proceedings of IEEE International Conference on Computer Vision (ECCV), pp. 8878–8887 (2019)
51. Zhang, Y., Zhang, J., Guo, X.: Kindling the darkness: a practical low-light image enhancer. In: Proceedings of ACM International Conference on Multimedia (ACM MM), pp. 1632–1640 (2019)
52. Mao, X.J., Shen, C., Yang, Y.-B.: Image restoration using very deep convolutional encoder–decoder networks with symmetric skip connections. In: Proceedings of Conference on Neural Information Processing Systems (NIPS), pp. 2810–2818 (2016)
53. Zhang, K., Zuo, W., Zhang, L.: FFDNet: toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans. Image Process.* **27**(9), 4608–4622 (2018)
54. Yang, J., Yang, J.Y.: Generalized K–L transform based combined feature extraction. *Pattern Recognit.* **35**(1), 295–297 (2002)
55. Yang, J., Yang, J.Y., Zhang, D., et al.: Feature fusion: parallel strategy vs. serial strategy. *Pattern Recognit.* **36**(6), 1369–1381 (2003)
56. Liu, C.J., Wechsler, H.: A shape and texture-based enhanced Fisher classifier for face recognition. *IEEE Trans. Image Process.* **10**(4), 598–608 (2001)
57. Hu, J., Shen, L., Albanie, S., et al.: Squeeze and excitation networks. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7132–7141 (2018)
58. Wang, X., Girshick, R., Gupta, A., et al.: Non-local neural networks. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7794–7803 (2018)
59. Woo, S., Park, J., Lee, J.Y., et al.: CBAM: convolutional block attention module. In: Proceedings of IEEE International Conference on Computer Vision (ECCV), pp. 3–19 (2018)
60. Zhang, Y., Li, K., Li, K., et al.: Residual non-local attention networks for image restoration. In: Proceedings of International Conference on Learning Representations (ICLR) (2019)
61. Loshchilov, I., Hutter, F.: SGDR: stochastic gradient descent with warm restarts. [arXiv:1608.03983](https://arxiv.org/abs/1608.03983) (2016)
62. Wang, Z., Bovik, A.C., Sheikh, H.R., et al.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
63. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: Proceedings of 37th Asilomar Conference on Signals, Systems & Computers (ACSSC), pp. 1398–1402 (2003)
64. Radiological Society of North America, COVID-19 radiography database. <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database> (2020)
65. Chung, A.: Figure 1 COVID-19 chest X-ray data initiative. <https://github.com/agchung/Figure1-COVID-chestxray-dataset> (2020)
66. Cohen, J.P., Morrison, P., & Dao, L.: COVID-19 image data collection. <https://github.com/ieee8023/covid-chestxray-dataset> (2020)

67. Chung, A.: Actualmed COVID-19 chest x-ray data initiative. <https://github.com/agchung/Actualmed-COVID-chestxray-dataset> (2020)
68. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
69. Lin Min, C.Q., Yan, S.: Network in network. [arXiv:1312.4400](https://arxiv.org/abs/1312.4400) (2013)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Xiaoben Jiang is pursuing the PhD degree in East China University of Science and Technology. His current research interests include digital image processing and computer vision. His experience includes the denoising method on chest X-ray images and CT images, and detection of COVID-19 cases from denoised CXR images. He has published in journals in the crossing field of medical science and computer vision and has been involved in publicly and privately funded projects.



Yu Zhu received the PhD degree in optical engineering from Nanjing University of Science and Technology, China, in 1999. She is currently a professor in the Department of Electronics and Communication Engineering of East China University of Science and Technology. Her research interests include image processing, computer vision, multimedia communication and deep learning. She has published more than 100 papers in journals and conferences.



Bingbing Zheng obtained the BS degree in Information Science and Engineering from East China University of Science and Technology in 2015 and is pursuing the PhD degree in East China University of Science and Technology. His main research interests include deep learning for medical image processing and computer vision. His experience includes the identification and detection of pulmonary nodules on CT images, the classification and segmentation of prostate on MRI, and the classification and segmentation of COVID-19. He has published in journals and conferences in the crossing field of medical science and computer vision and has been involved in publicly and privately funded projects.



Dr. Dawei Yang is dedicated to early diagnosis of lung cancer and relevant studies, with special interests on the management of pulmonary nodule and validation of diagnostic biomarker panels based on MIOT, CORE, and radiomics artificial intelligence (AI) platform. He is a member of IASLC Prevention, Screening & Early Detection Committee. Since 2011, he has published 17 SCI research articles and 9 as the first author, including those on *Am J Resp Crit Care* (2013), *Can Lett* (2015, 2020), *Cancer* (2015, 2018, 2020), and *Dev Cell* (2020), etc. He acted as a presenter for oral or poster presentation in ATS, WCLC, APSR, ISRD for a couple of times. He is one of the peer reviewers for international journals, such as *J Cell Mol Med*, *J Transl Med*, etc., and managing editor of *Clinical eHealth*.