

RESEARCH

Open Access



DMF-Net: a deep multi-level semantic fusion network for high-resolution chest CT and X-ray image de-noising

Tapan Kumar Nayak^{1†}, Chandra Sekhara Rao Annavarappu^{1†}, Soumya Ranjan Nayak^{2†} and Berihun Molla Gedefaw^{3**†}

Abstract

Medical images such as CT and X-ray have been widely used for the detection of several chest infections and lung diseases. However, these images are susceptible to different types of noise, and it is hard to remove these noises due to their complex distribution. The presence of such noise significantly deteriorates the quality of the images and significantly affects the diagnosis performance. Hence, the design of an effective de-noising technique is highly essential to remove the noise from chest CT and X-ray images prior to further processing. Deep learning methods, mainly, CNN have shown tremendous progress on de-noising tasks. However, existing CNN based models estimate the noise from the final layers, which may not carry adequate details of the image. To tackle this issue, in this paper a deep multi-level semantic fusion network is proposed, called DMF-Net for the removal of noise from chest CT and X-ray images. The DMF-Net mainly comprises of a dilated convolutional feature extraction block, a cascaded feature learning block (CFLB) and a noise fusion block (NFB) followed by a prominent feature extraction block. The CFLB cascades the features from different levels (convolutional layers) which are later fed to NFB to attain correct noise prediction. Finally, the Prominent Feature Extraction Block (PFEB) produces the clean image. To validate the proposed de-noising technique, a separate and a mixed dataset containing high-resolution CT and X-ray images with specific and blind noise are used. Experimental results indicate the effectiveness of the DMF-Net compared to other state-of-the-art methods in the context of peak signal-to-noise ratio (PSNR) and structural similarity measurement (SSIM) while drastically cutting down on the processing power needed.

Keywords Cascaded Feature, DMF-Net, CT, X-Ray

Introduction

X-ray and Computed Tomography (CT) are two distinct imaging techniques used frequently in the medical field for corporal inspection of human lungs. An X-ray is a commonly used and widely available imaging technique, while a CT scan is similar to MRI in that it produces high-quality body organ images. Both types of scans produce images in different ways. The CT scan takes pictures of body organs from all angles, enhancing its accuracy, while X-rays use electromagnetic waves to flow through the patient's body, producing black-and-white photos of the inside structure. In

[†]Tapan Kumar Nayak, Chandra Sekhara Rao Annavarappu, Soumya Ranjan Nayak and Berihun Molla Gedefaw. All authors contributed equally to this work.

*Correspondence:

Berihun Molla Gedefaw
berihunmolla44@gmail.com

¹ Department of CSE, IIT(ISM) Dhanbad, Sardar Patel Nagar, Dhanbad 826004, Jharkhand, India

² School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar 751024, Odisha, India

³ Department of Health Informatics, Arba Minch University College of Medicine and Health Science, Arba Minch, Ethiopia



radiology, CT scans and X-rays are frequently used for diagnostic purposes.

The quality of CT and X-ray pictures has deteriorated for a variety of causes, including the following:

Blurred image: These medical pictures are blurred due to incorrect protocol parameters and patient movement [1]. Patient movement is caused by various variables, including an increase in heart rate, respiration, fluctuations in the number of pixels while scanning in a uniform material, and the patient's unwillingness to cooperate. Blurring rises in direct proportion to the degree of movement.

Field of View (FOV): When the resolution is changed to take a picture of a smaller or larger area, the image quality also gets worse.

Artefact: Artefacts are incidental, supplementary graphics that appear alongside the main image, like metal artifacts, beam hardening, and faulty equipment, accidental damage, or inappropriate input can result in artefacts.

It has a large impact on reducing the model's accuracy during classification, detection, segmentation, and registration when the picture quality is influenced by various amounts of noise [2–4]. As a result, it is important to make a model for removing noise from medical images like CT and X-rays as part of the pre-processing [5, 6].

To denoise the images various models are used can be categorized in two types (1) traditional model and (2) deep models.

(1) Traditional models: Different traditional methods used are linear smoothing, median filtering, wiener filtering, anisotropic diffusion, and wavelet-based methods.

- i. Linear smoothing: Here a noisy image y convolved with a Gaussian filter k , to clean up an image.

$$\hat{x} = y \times k \quad (1)$$

It can also be done in a fourier domain as follows:

$$\hat{X} = Y \odot K \quad (2)$$

Here, capital letters stand for the Fourier transform of their counterparts (for example, $Y = F(y)$, where F is the Fourier transform), and \odot stands for the element-wise product. In the Fourier domain, K is also gaussian.

- ii. Median Filtering: Median filtering can be used instead of linear smoothing. The idea behind median filtering is to take an image and work on it one pixel at a time. Also, every pixel is supplemented by the value that is in the middle of a group of pixels that are close to it. So, the method can also be seen as a way to filter, although the filter is not linear.
- iii. Wiener Filter: Linear prediction, signal restoration, and channel equalization are just a few examples

of the many uses for Wiener filters. This technique works for both additive noise and multiplicative noise.

- iv. Anisotropic diffusion: Anisotropic diffusion, an iterative method that uses smoothing, can be used to remove noise from images. This strategy makes an effort to meet the following conditions: (a) Object borders must be kept intact, and (b) noise must be effectively filtered out in areas of high similarity. The approach is so-called because its mathematical underpinnings are similar to those of heat diffusion equations and because its smoothing or diffusion is applied in discrete regions rather than globally.
- v. Wavelet-based method: Here, an image is converted into a wavelet domain and wavelet coefficients. Then, the inverse wavelet transform is used to get the denoised image.

(2) Deep models: Deep learning was first introduced for image de-noising in 1989 by Chiang and Sullivan. Here the proposed neural network uses blur function and additive noise to get a clean image. The network then used weight values to get rid of complicated noise [7]. To cut down on the high cost of computation, a feedforward network had been suggested to achieve a tradeoff among both de-noising performance and efficiency [8]. After that, further optimization techniques were employed to speed up the convergence of the network and improve the performance of de-noising [9]. Also by raising the depth or modifying the activation function, novel network designs shown competitive in removing noise [10]. But these models require parameters to be set manually and it got resolved with gradient descent [11, 12]. Because of the aforementioned reasons, convolutional neural networks (CNNs) were proposed with vanishing gradients and different activation functions such as sigmoid [13] and tanh [14] but it needs a computationally effective platform to implement. As a result the ImageNet challenge started in 2012 brings different pre-trained models like AlexNet, VGGNet, MobileNet [15–17] etc. to deal with. These models were used for de-noising image starting from 2015 [18, 19]. Image de-noising methods mostly use the mathematical model $y = x + \eta$ to get all the clean images, x , where η and y are just additive noise with standard deviation (σ) and a noisy image, respectively. Many publications used this formulation as the bedrock for their models [20] in the past. Zhang et al. [21] devised a deep convolutional neural network entitled De-noising CNN with Batch Normalization (BN) and Residual Learning (RL) to reconstruct the clean image. Autoencoders and stacked sparse autoencoders have been applied to enhance image de-noising performance

with higher efficiency for spatial correlations [22]. The information from the final layer got incorporated in most CNN-based models, while low-level information remained overlooked. Although the previous methods are more appealing toward image restoration, they have distinguished drawbacks: (1) The dense network [23] does not use the shallow or hidden layers effectively. (2) Most of the methods trade-off with complex background which hides required features.

Therefore the proposed DMF-Net is composed of Dilated Feature Extraction Block (DFEB), Cascaded Feature Block(CFB), and Prominent Feature Refinement Block(PFRB) trained with decaying the learning rate. The contributions of this work are as follows,

- The proposed DMF-Net contains dilated convolution and batch normalization is used with a unique idea of combined feature matrix from different layers.
- The proposed model is trained and evaluated with CT and X-ray images separately and combined.
- The proposed model addresses the high-level abstraction of radiographs, hence eliminating the requirement of a handcrafted feature extraction process.
- The recent research trends often give importance to other layers for feature extraction not only the last layer, as it diversifies the image features. Getting inspired by this literature, an effort has been made with other layers.

Related work

Different CNN models have been designed for object detection and retrieval of the clear image with modified preexisting networks or designing networks with different plug-ins to improve the model results [24, 25]. Most of the CNNs has been designed with the aim of improving efficiency and accuracy w.r.t. de-noising. Dabov et al. [20] proposed a model with enhanced sparsity to convert 2D array to 3d array. They got a significant improvement with specially designed weiner filter. Zhang et al. [26] proposed weighted nuclear norm minimization(wnnm) technique to take the advantage of nonlocal self similarity. Buades et al. [27] put emphasis on nonlocal means(nlm) to focus on structural preservation of image. Portilla et al. [28] used statistical model with bayesian estimator to eliminate gaussian noise. Chen et al. [29] proposed residual encoder with decoder convolutional network(RED-CNN) by patch based training for CT-images. Gondara et al. [30] developed an autoencoder with convolutional layers to deal with heterogenous data to address less computational complexity. Kang et al. [22] designed an algorithm with directional wavelet to address the photon starvation in CT-images. Zheng et al. [31] enhanced the spatial adaptivity of nlm with element-wise fractal dimension. Duan et al. [32]

proposed a new second-order total generalised variation (TGV) decay model to get rid of speckle noise. Yang et al. [33] developed tensor based adaptive control for principal component analysis with a searching window for image de-noising. Zohair et al. [34] proposed a phase preserving approach got better peak signal noise ratio(psnr). Trinh et al. [35] formulated quadratic programming on weighted image block resulting better performance. Chen et al. [36] introduced deep boosting framework integrating with various cnn to generate required features for noise removal. Also the increased training samples has shown better performance in suppressing the noise. As experimented with a generative adversarial network (GAN) [37] a discriminative network has been used to increase the samples for the training purposes.

Methodology

This section contains an extensive description of the DMF-Net. The noisy X-ray and CT scans are developed by adding a particular noise value with the clean image, which has been described as $Y = X + \eta$, used as a source of data to DICN, where X : clean image, Y : input noisy image with η : external noise. The basic goal is to get the DICN to learn with a noisy image and then replicate a clean or noise-free image. The main aim of the model is to make it learn the noisy data to predict the clean image.

Feature dialation

The architecture with dilated convolutions [38] does the expansion of the receptive field, which increases accuracy with no resolution loss. Integrating multi-scale information with sub-sampling in image classification models reduces resolution. It helps in the area expansion of image data without pooling. The main aim is to cover wider pixel information to convolve for the output feature with the same cost of computation. Here the dilation factor(d) determines what will be the result by convolving with different values of d. The same kernel parameter can be used to retrieve more information. The value of d=1 means the kernel gets mapped with the same size of input but from d=2 onwards one pixel gets skipped while mapping with input.

For a dilated convolution or a d-dilated convolution named *d is used. The familiar discrete convolution is simply the 1-dilated convolution. Let $f_0, f_1, \dots, f_n : Z \rightarrow R$ be the functions and $k_0, k_1, \dots, k_n : \Omega \rightarrow R$ be the 3×3 filters. Kernel filters has been applied with increasing dilation:

The convolution operation is performed as defined $(F * k)(p) = X$

$$(F * k)(p) = \sum_{m+t=p} F(m)k(t) \quad (3)$$

For the dilated convolution operation, let d is the dilation factor then $(F *_d k)(p) = X$

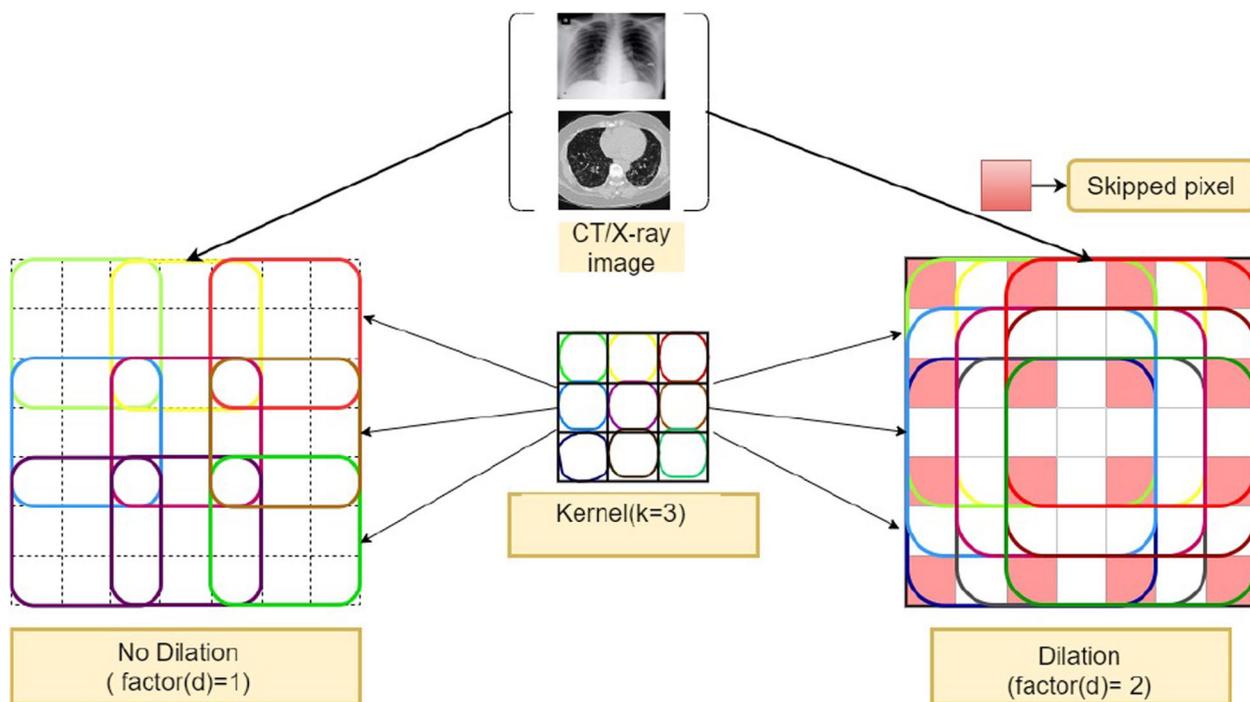


Fig. 1 Convolution with(d=2) and without(d=1) dilation

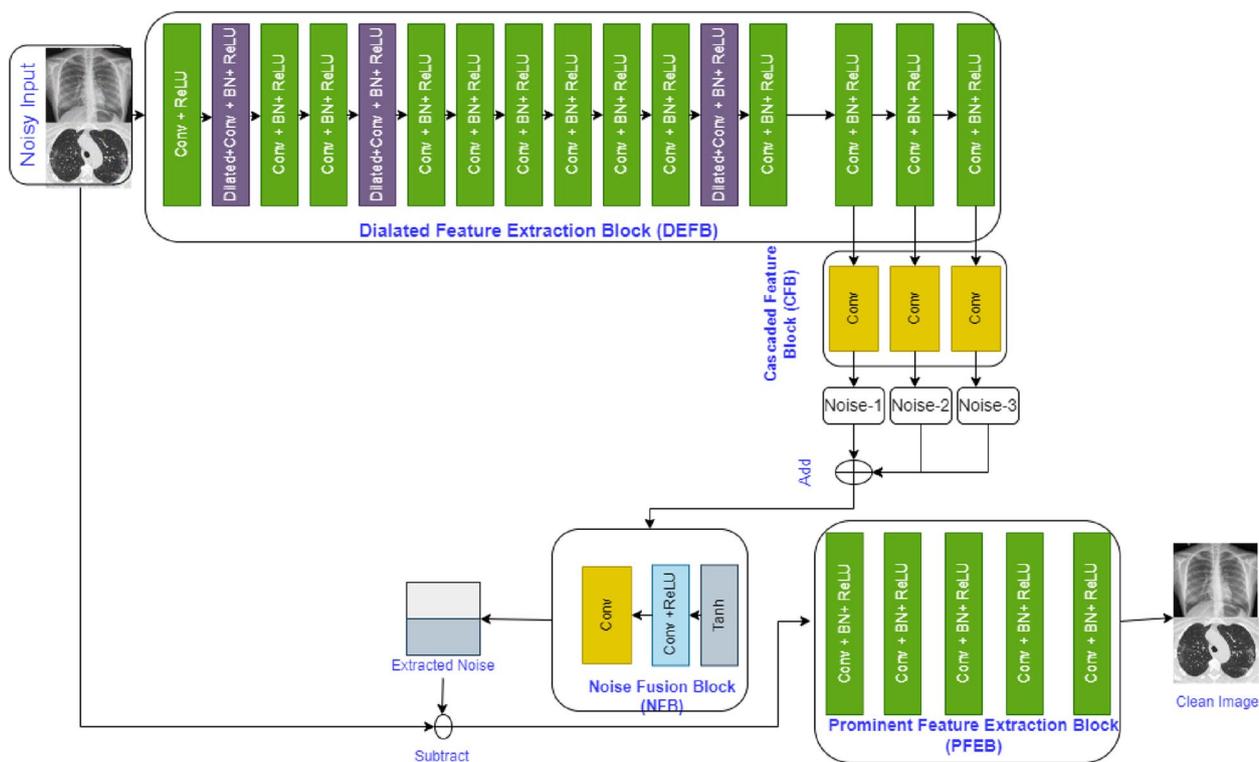


Fig. 2 Proposed model DMF-Net with different blocks

$$(F *_{d} k)(p) = \sum_{m+dt=p} F(m)k(t) \tag{4}$$

Here $*_{d}$ is assumed as the dilation factor. During convolution operation, the dilation factor determines how many pixels will be skipped with d dilation, d-1 no. of pixels skipped from input receptive(p) during convolution performed with the kernel. The Fig. 1 shows how the convolution performed with the skipping of pixels with different dilation factor.

Proposed DMF-Net architecture

The proposed architecture DMF-Net in Fig. 2 with the layer description in Table 1, contains three parts, DFEB, CFB and PFRB. The noisy images are the input to the DFEB box which has sub-modules like convolution, convolution with batch normalization and dilated convolution with batch normalization. CFB block does the inner layer cascading by adding the output features from low, mid and high levels. Its aim is to give all the levels equal chances for

the extraction of the noisy feature. These get concatenated to enhance the feature responsible for the noise. DMF-Net is designed with total of 26 convolution layers. The Dilated Feature Extraction Block consist of dilated 3×3 with ReLU [39] activation operation and normalising of several batches (BN) [40] and operation of ReLU activation (blue box).

It adds non-linearity to the network. The Batch Normalization performed in different layers enhances the training with an improved learning rate. Here from layer numbers 14th, 15th and 16th, the feature has been taken with convolution. This has been done keeping in mind that the features from other layers may enhance the noisy information. The three features from these layers are concatenated and passed through a noise fusion block(NFB). The NFB contains Tan hyperbolic function with convolution and ReLU. Finally, it gets convolved to give the final net extracted noise. The NFB module refines the fused noise for the final considered net noise feature (NNF). This NNF is subtracted from the noisy image to give a clear image after passing through a prominent feature extraction block(PFRB). It

Table 1 Detailed description of layers in DMF-Net

Operation Layer		No of Filters	Size of Each Filters	Dilation	Stride Value	Padding Value	Size of Output Image
Input Image		-	-	-	-	-	$128 \times 128 \times 1$
Convolution Layer (one)	Convolution	64	$3 \times 3 \times 1$	1	1×1	1×1	$128 \times 128 \times 64$
	ReLU	-	-	-	-	-	$128 \times 128 \times 64$
Convolution Layer (one)	Convolution+BN	64	$3 \times 3 \times 1$	2	1×1	1×1	$128 \times 128 \times 64$
	ReLU	-	-	-	-	-	$128 \times 128 \times 64$
Convolution Layer (two)	Convolution+BN	64	$3 \times 3 \times 1$	1	1×1	1×1	$128 \times 128 \times 64$
	ReLU	-	-	-	-	-	$128 \times 128 \times 64$
Convolution Layer (one)	Convolution+BN	64	$3 \times 3 \times 1$	2	1×1	1×1	$128 \times 128 \times 64$
	ReLU	-	-	-	-	-	$128 \times 128 \times 64$
Convolution Layer (six)	Convolution+BN	64	$3 \times 3 \times 1$	1	1×1	1×1	$128 \times 128 \times 64$
	ReLU	-	-	-	-	-	$128 \times 128 \times 64$
Convolution Layer (one)	Convolution+BN	64	$3 \times 3 \times 1$	2	1×1	1×1	$128 \times 128 \times 64$
	ReLU	-	-	-	-	-	$128 \times 128 \times 64$
Convolution Layer (four)	Convolution+BN	64	$3 \times 3 \times 1$	1	1×1	1×1	$128 \times 128 \times 64$
	ReLU	-	-	-	-	-	$128 \times 128 \times 64$
Convolution Layer (three)	Convolution	1	$3 \times 3 \times 1$	1	1×1	1×1	$128 \times 128 \times 1$
	-	-	-	-	-	-	$128 \times 128 \times 1$
	Tanh	-	-	-	-	-	$128 \times 128 \times 1$
Convolution Layer (one)	Convolution	1	$3 \times 3 \times 1$	1	1×1	1×1	$128 \times 128 \times 1$
	ReLU	-	-	-	-	-	$128 \times 128 \times 1$
Convolution Layer (one)	Convolution+BN	1	$3 \times 3 \times 1$	1	1×1	1×1	$128 \times 128 \times 1$
	ReLU	-	-	-	-	-	$128 \times 128 \times 1$
Convolution Layer (one)	Convolution+BN	64	$3 \times 3 \times 1$	1	1×1	1×1	$128 \times 128 \times 64$
	ReLU	-	-	-	-	-	$128 \times 128 \times 64$
Convolution Layer (three)	Convolution+BN	64	$3 \times 3 \times 1$	1	1×1	1×1	$128 \times 128 \times 64$
	ReLU	-	-	-	-	-	$128 \times 128 \times 64$
Convolution Layer (one)	Convolution+BN	1	$3 \times 3 \times 1$	1	1×1	1×1	$128 \times 128 \times 1$
	ReLU	-	-	-	-	-	$128 \times 128 \times 1$

Table 2 Using CT and X-Ray images individually and combined, PSNR and SSIM calculated at different noise levels, i.e. specific noise (N_{sp}) and blind noise(N_b)

Trained	Image	Noise=15		Noise=20		Noise=25	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Specific Noise	CT(No Aug)	29.33	0.8585	27.99	0.8230	26.95	0.7981
	CT(Aug)	29.06	0.8592	27.95	0.8316	26.62	0.7902
	X-Ray(No Aug)	30.94	0.8848	29.72	0.8560	28.72	0.8282
	X-Ray(Aug)	31.03	0.8896	29.95	0.8637	29.19	0.8442
	CT+X-Ray	30.24	0.8794	28.91	0.8445	28.32	0.8314
Blind Noise	CT(No Aug)	26.66	0.7973	27.10	0.8082	24.96	0.7444
	CT(Aug)	28.23	0.8294	27.11	0.7934	26.57	0.7787
	X-Ray(No Aug)	28.59	0.8445	27.72	0.8185	26.94	0.7926
	X-Ray(Aug)	28.92	0.8525	28.01	0.8240	27.19	0.7954
	CT+X-Ray	29.30	0.8582	28.37	0.8270	27.53	0.8020

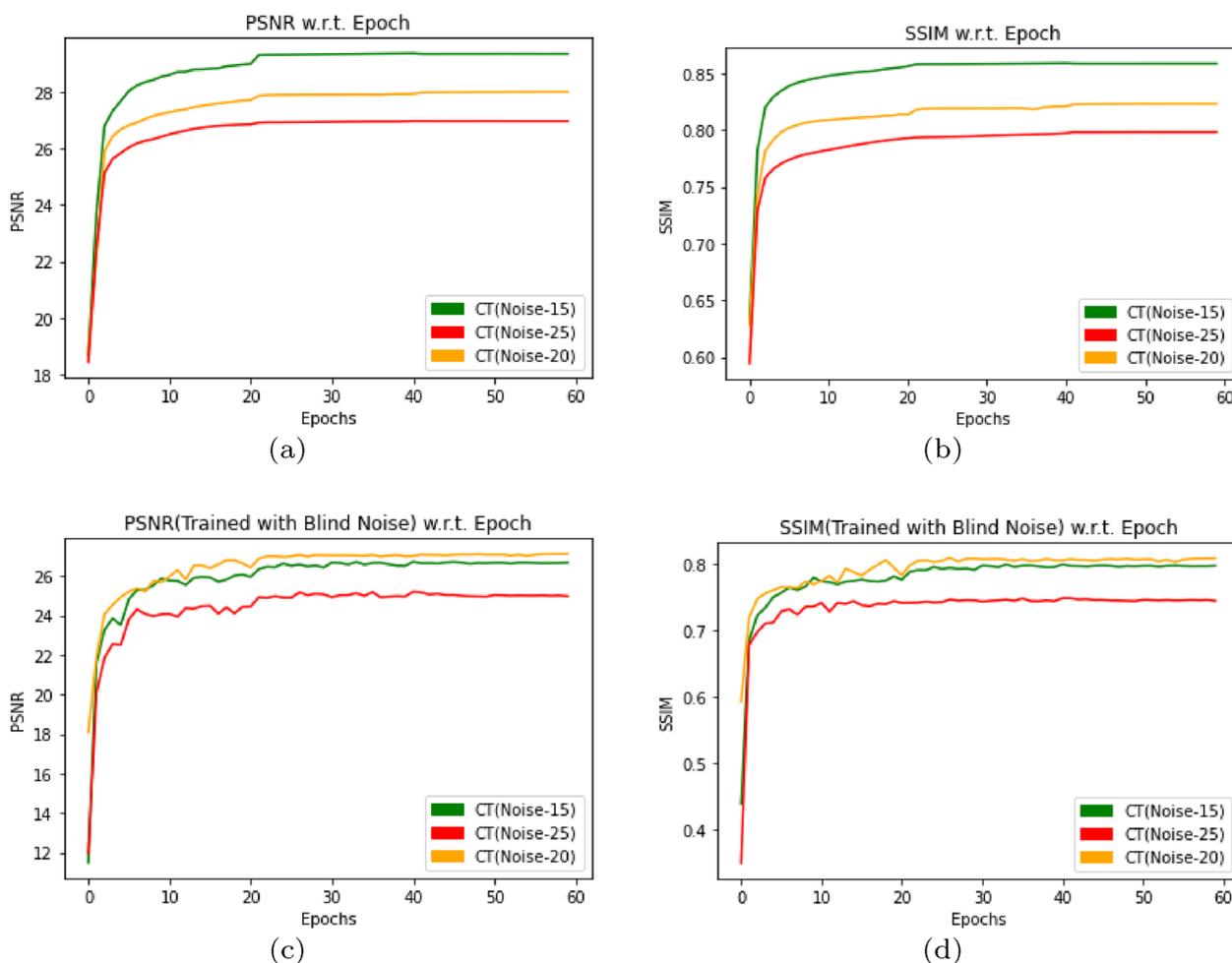


Fig. 3 PSNR and SSIM w.r.t. Epoch by using DMF-Net without augmentation and with specific noise in (a, b) and blind noise in (c, d) for noise level(σ) 15,20 and 25

contains four modules with one convolution with ReLU and three convolutions with Batch Normalization and ReLU. The PFRB block enhances the image quality by considering the prominent features. Here throughout the model same kernel size has been used i.e. 3×3 kernel. The final image size and the input image size is same. The 128×128 image size has been considered throughout the experiment by keeping the stride and padding one in every convolution layer. In order to extract more complete features and thus improve the model potential.

Training strategy

The mixed data training approach for deep network construction has been proposed here. The data set contains both CT and X-ray image data for the training. The basic goal of mixed training is for the model to learn a number of desirable properties. The features change drastically when a clinical image is added with a certain type of noise. Because

of this diversity in noisy images, it has been experimented with here to train the model with mixed clinical images. The results found here are also promising. Also, the DMF-Net is trained separately with CT images, augmented CT images, X-ray images, and augmented X-ray images with normally distributed specific noise levels (15, 20, 25, respectively) and blind noise. The model is trained upto 60 epochs with dynamic learning rate ($\alpha=0.001, 0.0001, 0.00001, 0.000001$) has been considered for the smooth convergence. Learning rate ($\alpha = 0.001$ (upto 20 epoch), 0.0001 ($20 < \text{epoch} \leq 40$), 0.00001 ($40 < \text{epoch} \leq 50$) and 0.000001 ($50 < \text{epoch} \leq 60$)). The contribution of low, mid, and high-level aspects was properly considered by integrating the features from the last three layers to determine noise.

Features from the last three-layer have been added to find the net feature $f(L)$.

$$f_{DMF-Net}(L) = f_{DMF-Net}(L_{17}) + f_{DMF-Net}(L_{18}) + f_{DMF-Net}(L_{19}) \tag{5}$$

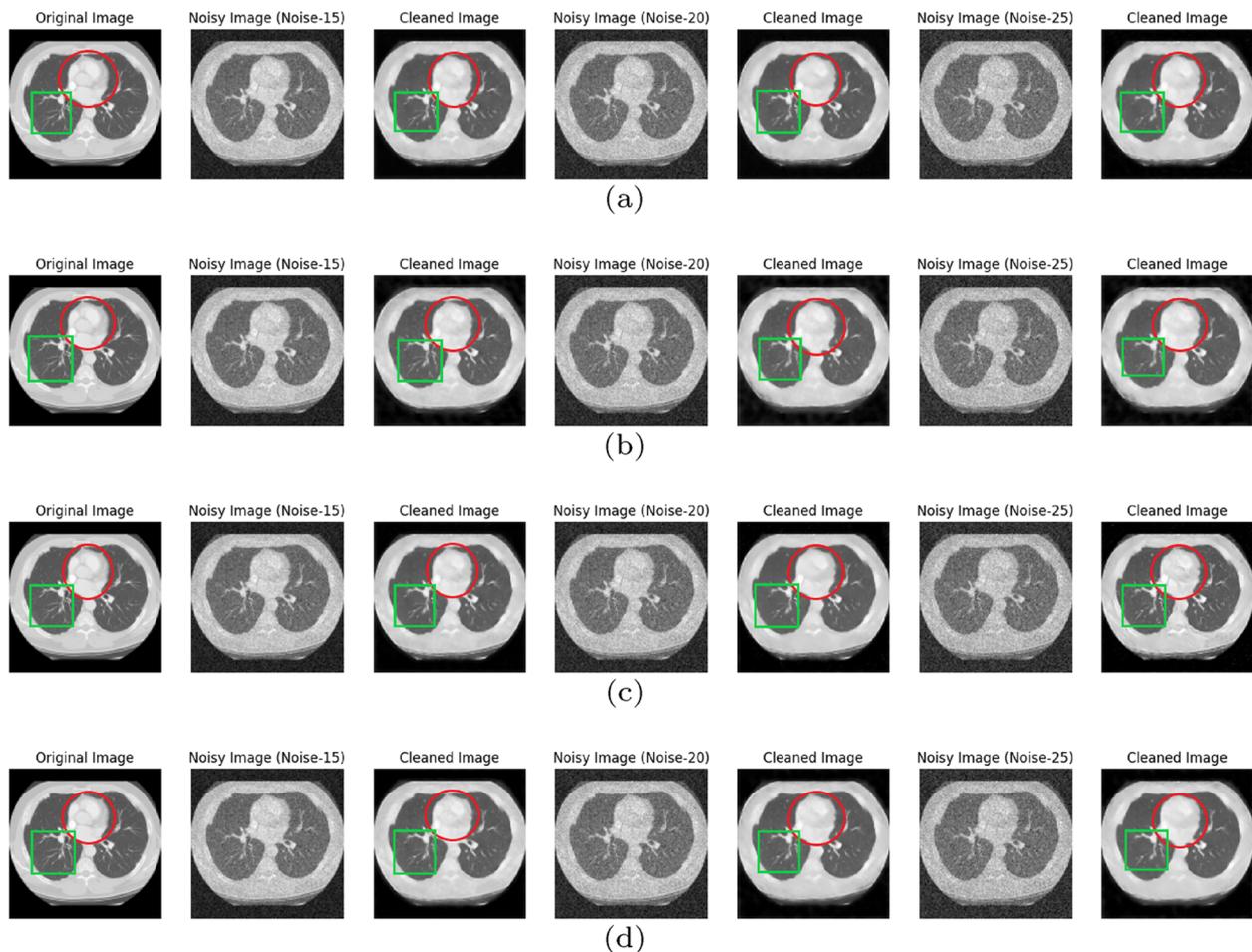


Fig. 4 Original CT image with noisy($N_{sp}=15,20,25$ and $N_{bl}=0$ to 55) and denoised or cleaned image by DMF-Net where (a and c) shows for specific noise without and with augmentation in (b and d) shows for blind noise without and with augmentation; red circle and green box to observe inter lungs area and left lungs respectively with original image

where L_{17}, L_{18}, L_{19} are layer 17, layer 18 and layer 19. The final feature after convolution is considered as the final noise feature $f_{DMF-Net}(N)$. The predicted denoised image (I_p) has been determined between the input noisy image (I_{noisy}) and $f(N)$.

$$I_p = I_{noisy} - f_{DMF-Net}(N) \tag{6}$$

$$Loss(L) = \|I - I_p\|^2 \tag{7}$$

Finally, the difference between the original noise-free picture (I) and the anticipated noise-free image is computed (I_p). After the loss is calculated, the optimization is done using adam optimizer [41] which is considered during training. The number of parameters get optimized due to the noise features extracted during training. However, the dynamic learning rate with respect to different epochs helps in faster convergence. Because the loss calculated in gradient gets transmitted due to the learning

function, reflecting the variation between the original noise and predicted noise with the parameter.

Experimental results

Details about the data as well as the implementation

Here our data folder contains CT image files as well as X-Ray image files. CT files of counts 1647 are collected from kaggle.com [42, 43] and GitHub.com [44], each one of grayscale lung CT slices of size 128×128 . The data folder is again divided in two folders as CT-training and CT-testing. Randomly the files are chosen from the training folder containing 1352 files with a ratio of 90:10 as training : validation and 295 files for testing. 1550 lungs X-Ray files are collected from kaggle.com [45] and distributed in two folders as X-Ray-training with 1292 files and X-Ray-testing with 258 files. Files from X-Ray-training folder are again chosen randomly as training : validation with ratio 90:10.

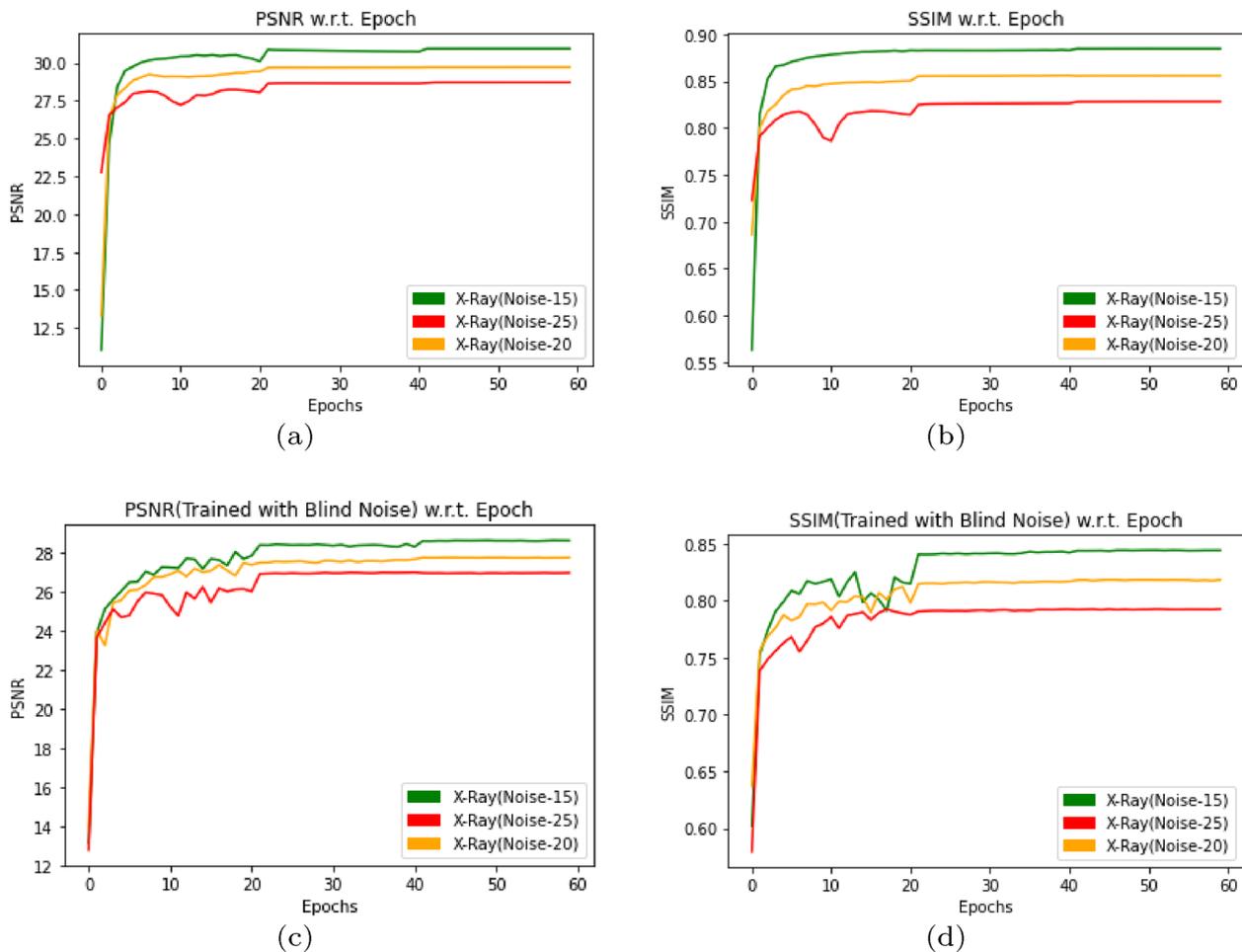


Fig. 5 PSNR and SSIM w.r.t. Epoch by using DMF-Net without augmentation and with specific noise in (a, b) and blind noise in (c, d) for noise level 15,20 and 25

Our deep model DMF-Net is developed in PyTorch. Here Adam optimizer is set for training the network with starting learning rate of 0.001 and it is scaled down by 10 after 20, 40 and 50 epochs. Training is done with 60 epochs with a batch size of 32. After training the network, the states of the model for 60 epochs have been saved with 60 .pth files. The system set up for the training and testing is NVIDIA-SMI 460.32.03 and CUDA version 11.2 with Tesla T4 and 32 GB RAM.

Evaluation metrics

The peak signal to noise ratio (PSNR) and the structural similarity index (SSIM) are employed as assessment criteria in this study. The mathematical formulation of PSNR is

$$PSNR = 20 \log_{10} \frac{MAX_f}{(MSE)^{\frac{1}{2}}} \tag{8}$$

where, peak signal to noise ratio (PSNR) is the ratio of the maximum pixel value of a noise-free image to the maximum pixel value of the image. and $(MSE)^{\frac{1}{2}}$ is the root mean square error.

$$MSE = \frac{1}{pq} \sum_0^{p-1} \sum_0^{q-1} \|I(i,j) - I'(i,j)\|^2 \tag{9}$$

where $I(i,j)$ is the pixel data of our noise free image, $I'(i,j)$ represents the pixel data of denoised image predicted by model. p: total rows of image data pixels and q: total columns of image data pixels. SSIM measures the incasant difference between similar images. It never conclude about the original or denoised image.

$$SSIM(I, I') = \frac{(2\mu_I \mu_{I'} + c_1)(2\sigma_{II'} + c_2)}{(\mu_I^2 + \mu_{I'}^2 + c_1)(\sigma_I^2 + \sigma_{I'}^2 + c_2)} \tag{10}$$

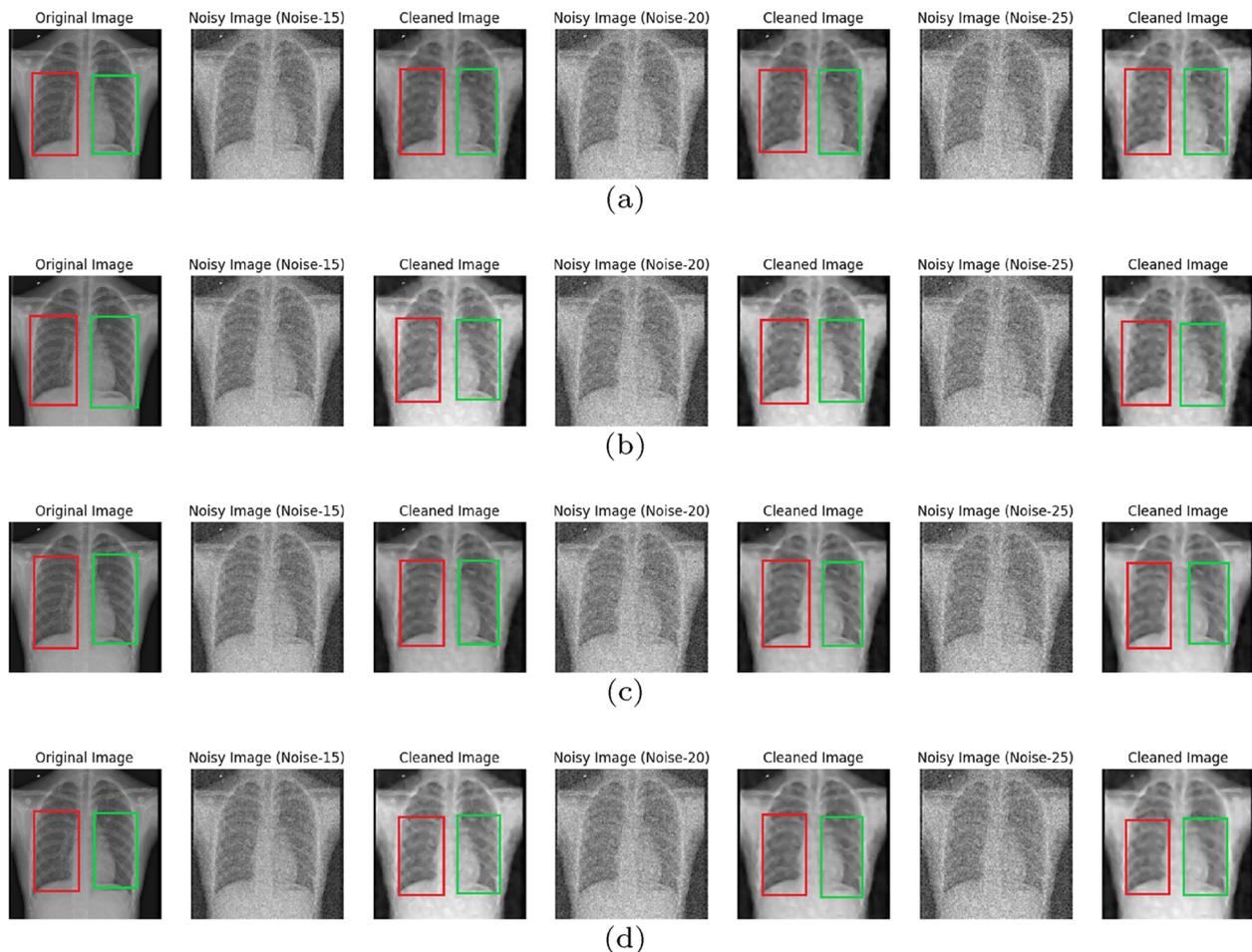


Fig. 6 Original X-Ray image with noisy($N_{sp}=15,20,25$ and $N_{bl}=0$ to 55) and denoised or cleaned image by DMF-Net where (a and c) shows for specific noise without and with augmentation in (b and d) shows for blind noise without and with augmentation; red box and green box to observe left lungs and right lungs area respectively with original image

where μ_I and $\mu_{I'}$ are mean of image I and I' , σ_I and $\sigma_{I'}$ are standard deviation(s.d.) of image I and I' and $\sigma_{II'}$ is the covariance of I and I' with c_1c_2 as constants.

To observe the performance, DMF-Net is trained with specific noise and blind noise. Specific noise and blind noise has been normalized with noise level N_{sp} (15, 20, 25) and N_{bl} (0 to 55) for creating the noisy X-Ray as well as CT image. The table 2 shows the PSNR and SSIM resulted by training the model with CT image, X-Ray image and mixed images separately.

De-noising CT image where DMF-Net trained with CT image

In this section, DMF-Net trained with 128×128 CT image files in the CT-training folder containing 1352 files with specific noise(N_{sp}) and blind noise(N_{bl}). The states of the model is saved for 60 epochs for evaluating

the model with PSNR and SSIM. Again the model gets trained with augmented CT images created by flipping the image right-left and up-down with angle $90^\circ, 180^\circ$ and 270° randomly, generating 2458 image files. Also the states of the model for 60 epochs has been saved for evaluating the model with PSNR and SSIM. Without augmentation the trained model for a test data of 295 images achieved average PSNR and SSIM for $N_{sp}=15,20$ and 25 with [(26.66,0.7973),(27.09,0.8082) and (24.96, 0.7444)] respectively and for $N_{bl}=0$ to 55 with [(26.66,0.7973),(27.10, 0.8082),(24.96, 0.7444)] respectively as shown in Fig. 3(a) and (b). With augmentation the model after 60 epochs results average PSNR and SSIM for $N_{sp}=15,20$ and 25 with [(28.23,0.8294),(27.11, 0.7934) and (26.57,0.7787)] respectively and for $N_{bl}=0$ to 55 with [(28.23,0.8294),(27.11, 0.7934), (26.57,0.7787)] respectively as shown in Fig. 3(c) and (d). The plot for PSNR and SSIM w.r.t. epochs are described in Fig. 3 with

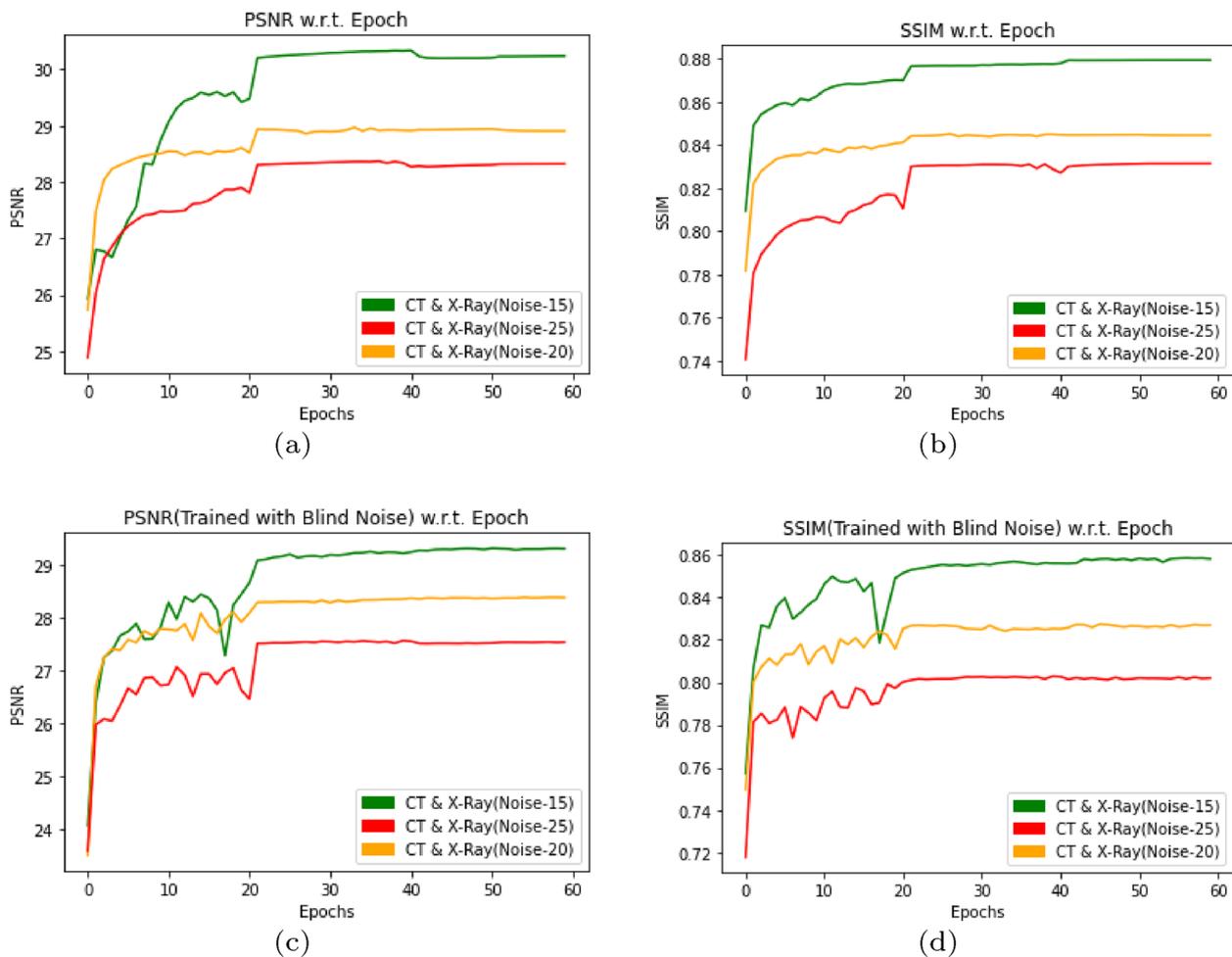


Fig. 7 PSNR and SSIM w.r.t. Epoch by using DMF-Net without augmentation and with specific noise in (a, b) and blind noise in (c, d) for noise level 15,20 and 25

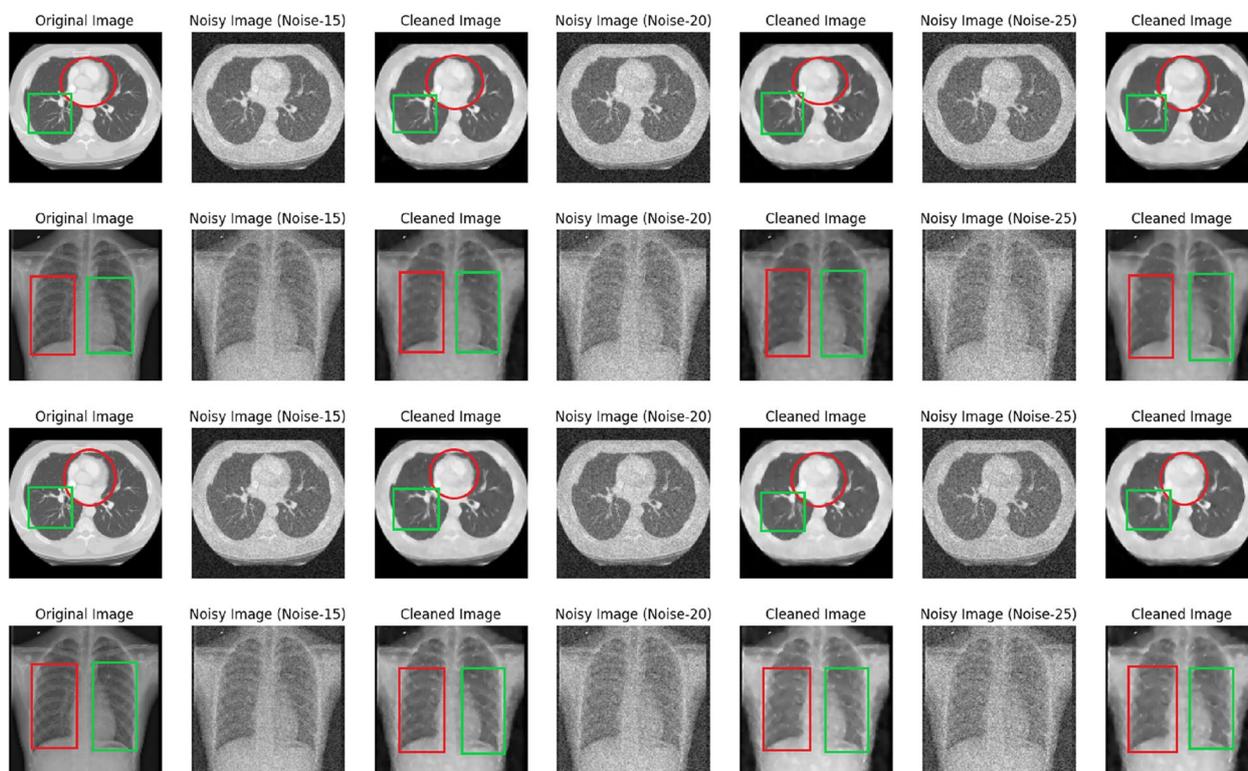


Fig. 8 Original CT and X-Ray image with noisy($N_{sp}=15,20,25$ and $N_{bl}=0$ to 55) and denoised or cleaned image by DMF-Net where same images of Figs. 4 and 6 are shown with specific noise and blind noise

Table 3 Multiple methods were used to acquire PSNR and SSIM measurements at varying CT and X-Ray noise levels (blind and specified noise levels); the best results are highlighted in bold

Trained	Methods	Noise(σ)=15		Noise(σ)=20		Noise(σ)=25	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Specific Noise	BM3D [20]	21.99	0.3778	18.63	0.2671	16.08	0.1381
	WNNM [26]	21.92	0.3669	18.23	0.2356	16.09	0.1345
	NLM [27]	24.61	0.7290	22.32	0.6431	20.29	0.5430
	BLS-GSM [28]	30.60	0.8581	29.09	0.8146	27.99	0.7766
	RED-CNN [29]	30.46	0.8878	23.98	0.7109	18.19	0.6691
	Autoencoder(ANN) [30]	30.30	0.8091	22.11	0.7209	21.34	0.6788
	CNN and Wavelets [22]	30.21	0.8912	23.12	0.7217	21.87	0.6823
Blind Noise	Non-local means [31]	28.71	0.8371	23.28	0.7154	18.76	0.6544
	Freq. Domain FFT [32]	29.85	0.8822	22.91	0.6992	17.21	0.6141
	Adapt. Tensor, PCA [33]	29.46	0.8731	22.74	0.6963	18.31	0.6129
	Coeff. Driven Variation [46]	28.75	0.8563	22.41	0.6871	17.78	0.5821
	Phase-preserving [34]	28.13	0.8429	22.09	0.6776	17.13	0.6011
	Optimal Weight [35]	30.17	0.8879	23.08	0.7050	-	-
	Wavelet and Sparse [47]	30.23	0.8078	22.52	0.7053	-	-
	DMF-Net(proposed)	31.03	0.8896	29.95	0.8637	29.19	0.8442
	DMF-Net(proposed)	29.30	0.8582	28.37	0.8270	27.53	0.8020

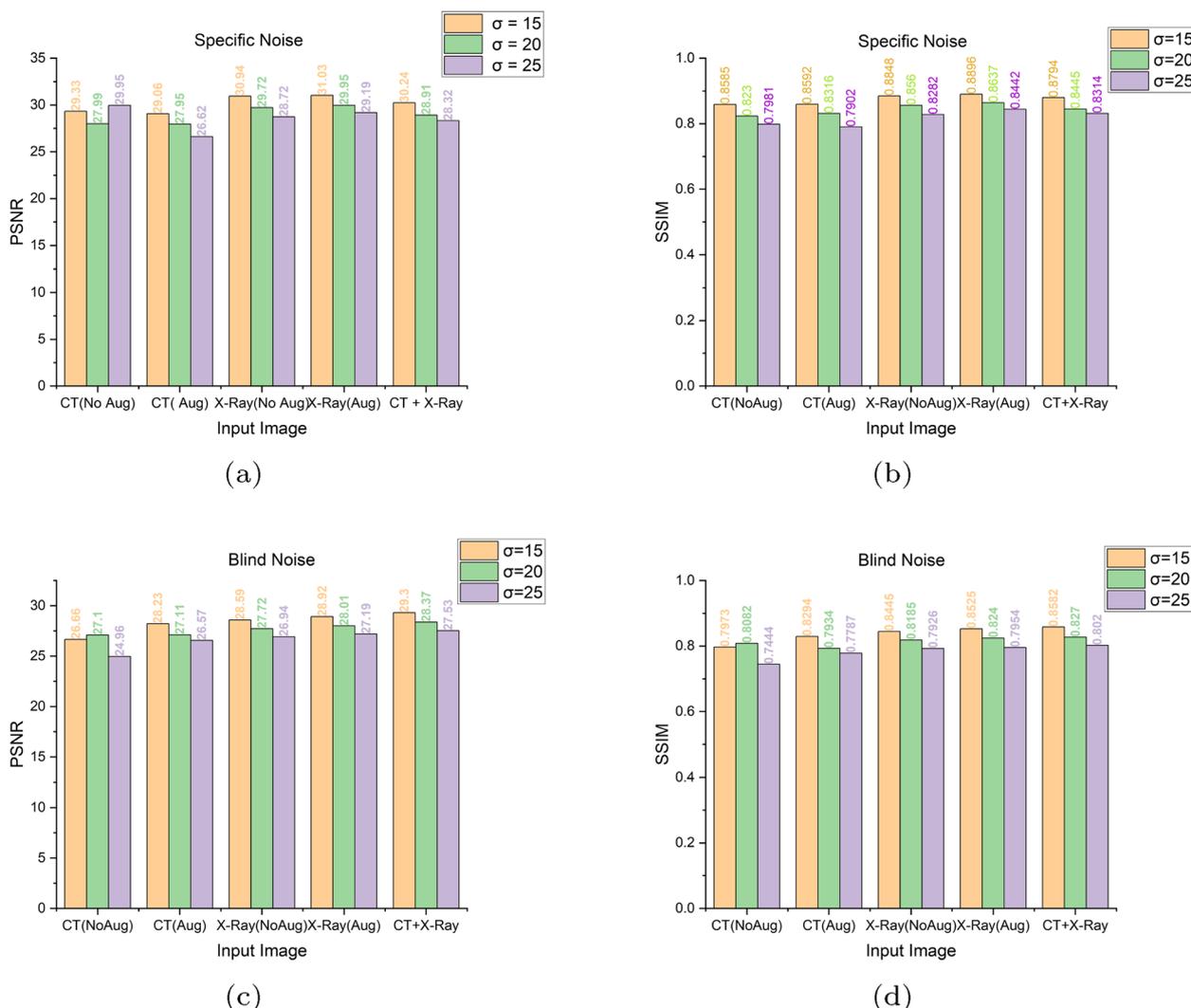


Fig. 9 PSNR and SSIM w.r.t. input CT and X-Ray, where (a, b) shows result when the model trained and validated with specific noise($\sigma=15,20$ and 25) and (c, d) shows result when the model trained with blind noise and validated with noise level($\sigma=15,20$ and 25)

different noise values. The resulted denoised image has been displayed in Fig. 4 for different noise type and the peculiar areas has been selected to compare with original image.

De-noising X-Ray image where DMF-Net trained with X-Ray image

In this section, DMF-Net trained with 128×128 X-Ray image files in X-Ray-training folder containing 1292 files with specific(N_{sp}) and blind(N_{bl}) noise observing PSNR and SSIM. Also the model performance is observed for augmented images generated by flipping right-left, up-down with angle $90^\circ, 180^\circ$ and 270° randomly, generating

2348 image files. The model is trained for 60 epochs. The trained model for a test data of 258 images achieved average PSNR and SSIM for $N_{sp}=15,20$ and 25 with [(30.94, 0.8848), (29.72,0.8560) and (28.72,0.8282)] respectively and for $N_{bl}=0$ to 55 with [(28.59,0.8445), (27.72, 0.8185),(26.94, 0.7926)] respectively as shown in Fig. 5(a) and (b). But with augmentation the model after 60 epochs results average PSNR and SSIM for $N_{sp}=15,20$ and 25 with [(31.03,0.8896), (29.95, 0.8637) and (29.19,0.8442)] respectively and for $N_{bl}=0$ to 55 with [(28.92,0.8525),(28.01, 0.8240),(27.19, 7954)] respectively as shown in Fig. 5(c) and (d) and some resulted images are shown in Fig. 6 for the visual comparison of some peculiar areas.

De-noising both CT and X-Ray image where DMF-Net trained with mixed image

In this section, DMF-Net trained with mixed data set of 128×128 X-Ray image and CT images. The data set contains 1292 X-Ray image files and 1365 CT image files. The PSNR and SSIM has been observed for the model after training with specific noise (N_{sp}) and blind noise (N_{bl}). The model trained only without augmentation for 60 epochs and it is tested with 553 X-Ray and CT-images with $N_{sp}=15, 20$ and 25 resulting average PSNR and SSIM with [(30.24, 0.8794), (28.91, 0.8445) and (28.32, 0.8314)] respectively as shown in Fig. 7(a) and (b). Also trained with blind noise $N_{bl}=0$ to 55 resulting average PSNR and SSIM with [(29.30, 0.8582), (28.37, 0.8270) and (27.53, 0.8020)] respectively as shown in Fig. 7(c) and (d). The input CT and X-Ray images with output denoised images has been shown in Fig. 8.

State of art comparison

The proposed model DMF-Net compared with other de-noising models like BM3D [20], WNNM [26], NLM [27], BLS-GSM [28], RED-CNN [29], autoencoder [30], CNN-Wavelets [22], non-local means [31], frequency domain FFT [32], adaptive tensor with PCA [33], coefficient driven variation [46], phase preserving [34], optimal weight [35] and wavelet with sparse [30]. The experimental results on specified and blind noise levels are shown in Table 3 to verify the suggested model's performance. The above methods are evaluated with three different noise levels of 15, 20 and 25. At higher noise levels, our model outperformed the BM3D, WNNM, NLM, and BLS-GSM techniques in terms of PSNR and SSIM. Also our model performance has been evaluated with different medical images like X-Ray and CT as in Table 2, it is found that the performance is better in different two different types of images taken separately or mixed with blind noise and in specific noise. All the result has been compared with different existing models and our proposed model has shown the improved performance over all these.

Conclusion

In this work, we developed DMF-Net with a strategy of training the network with for de-noising CT and X-Ray images. Here the network is designed with dilated feature extraction block (DEFB), cascaded feature block (CFB) and prominent feature extraction block (PFEB) block. The model consist of total 26 convolutional block where 18 convolution+BN+ReLU (green box), 3 dilated convolution+ReLU (violet box), 4 convolution (orange box), 1 convolution+ ReLU (navy blue box) and Tanh

function. The features of the last layers in the cascaded feature block are added with the intention to consider the low, mid and high-level features for evaluating the net noise. The clean image is extracted from the prominent feature extraction block after several convolutions. Here the model is trained and tested separately for CT images, X-Ray images and mixed images for specific and blind noise. The Fig. 9 shows result when the model trained and validated with specific noise ($\sigma=15, 20$ and 25) (Fig. 9(a), (b)) and Fig. 9(c), (d) shows result when the model trained with blind noise and validated with noise level ($\sigma=15, 20$ and 25). The network was trained for 60 epochs with just a dynamic decaying learning rate, and it was noticed that in the case of blind noise, mixed images perform better than individual images with and without augmentation and that a good result can also be found in specific noise when using peak signal-noise ratio (PSNR) and structural similarity index measurement (SSIM) as evaluation metrics. As a consequence, the proposed model can learn features taken from mixed images and satisfactorily denoise both CT and X-ray images. Situations where sufficient data is not readily available in advance for training, or when existing data must be modified to account for novel patterns, are ideal applications for real-time machine learning. The preserved state of the trained DMF-Net can be deployed there as an event-driven model to provide real-time generation of clean images with levels of precision.

Acknowledgements

Thanks to IIT(ISM) Dhanbad for providing me with the platform to work and special thanks to Dr.A.C.S. Rao who has given direction to accomplish the work.

Authors' contributions

All the authors have equally participated as follows: Conceptualization and design by Tapan Kumar Nayak, Investigation, and methodology by Chandra Sekhara Rao Drafting the article and revising it critically for important intellectual content by Soumya Ranjan Nayak Final Proofreading and interpretation of the data by Berihun Molla Gedefaw. All authors have agreed and read the published version of the manuscript.

Funding

No external funding was associated with this research study.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the Kaggle repository, [<https://www.kaggle.com/datasets/plameneduardo/sarscov2-ctscan-dataset>][<https://www.kaggle.com/datasets/andrewmvd/pediatric-pneumonia-chest-xray>].

Declarations

Ethics approval and consent to participate

This manuscript has not been submitted to, nor is under review at, another journal or other publishing venue. We declare that this paper is original and has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that all have approved the order of authors listed in the paper of us. All the patients' shared data have been approved by the Ethical Committee of the Public Hospital of the Government Employees of Sao Paulo (HSPM),

Sao Paulo/Brazil and the same has been cited in the reference section of this paper. However, due to ethical concerns about patients' privacy information, the Kaggle database presented the dataset in TIFF format files. The same tiff file has been used in this current research experimentation. The study reported in this manuscript doesn't require any involving human participants, human data, or human tissue.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 24 July 2023 Accepted: 24 September 2023

Published online: 09 October 2023

References

- Diwakar M, Kumar M. A review on CT image noise and its denoising. *Biomed Signal Process Control*. 2018;42:73–88.
- Zhang L, Li J, Yin K, Jiang Z, Li T, Hu R, et al. Computed tomography angiography-based analysis of high-risk intracerebral haemorrhage patients by employing a mathematical model. *BMC Bioinformatics*. 2019;20(7):109–16.
- Hu H, Pan N, Wang J, Yin T, Ye R. Automatic segmentation of left ventricle from cardiac MRI via deep learning and region constrained dynamic programming. *Neurocomputing*. 2019;347:139–48.
- Wang C, Song H, Chen L, Li Q, Yang J, Hu XT, et al. Automatic liver segmentation using multi-plane integrated fully convolutional neural networks. In: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2018. p. 1–6.
- Xu J, Li H, Liang Z, Zhang D, Zhang L. Real-world noisy image denoising: A new benchmark. *arXiv preprint arXiv:1804.02603*. 2018.
- Nayak SR, Nayak J, Sinha U, Arora V, Ghosh U, Satapathy SC. An automated lightweight deep neural network for diagnosis of COVID-19 from chest X-ray images. *Arab J Sci Eng*. 2023;48:11085–102.
- Chiang YW, Sullivan B. Multi-frame image restoration using a neural network. In: Proceedings of the 32nd Midwest Symposium on Circuits and Systems. IEEE; 1989. p. 744–747.
- Nakamura K, Okamura H, Wada M, Nagata K, Tamura T. Endotoxin-induced serum factor that stimulates gamma interferon production. *Infect Immun*. 1989;57(2):590–5.
- Bedini L, Tonazzini A. Image restoration preserving discontinuities: the Bayesian approach and neural networks. *Image Vis Comput*. 1992;10(2):108–18.
- Sivakumar K, Desai UB. Image restoration using a multilayer perceptron with a multilevel sigmoidal function. *IEEE Trans Signal Proc*. 1993;41(5):2018–22.
- Lee CC, de Gyvez JP. Color image processing in a cellular neural-network environment. *IEEE Trans Neural Netw*. 1996;7(5):1086–98.
- Labianca R, Pessi MA, Zamparelli G. Treatment of colorectal cancer. *Drugs*. 1997;53(4):593–607.
- Marreiros AC, Daunizeau J, Kiebel SJ, Friston KJ. Population dynamics: variance and the sigmoid activation function. *Neuroimage*. 2008;42(1):147–57.
- Jarrett K, Kavukcuoglu K, Ranzato M, LeCun Y. What is the best multi-stage architecture for object recognition? In: 2009 IEEE 12th international conference on computer vision. IEEE; 2009. p. 2146–53.
- Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*. 2012.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*. 2014.
- Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*. 2017.
- Liang J, Liu R. Stacked denoising autoencoder and dropout together to prevent overfitting in deep neural network. In: 2015 8th international congress on image and signal processing (CISP). IEEE; 2015. p. 697–701.
- Xu J, Zhang L, Zuo W, Zhang D, Feng X. Patch group based nonlocal self-similarity prior learning for image denoising. In: Proceedings of the IEEE international conference on computer vision. IEEE Conference; 2015. p. 244–252. <https://ieeexplore.ieee.org/document/7410393>.
- Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans Image Process*. 2007;16(8):2080–95.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE Conference; 2016. p. 770–778. <https://ieeexplore.ieee.org/document/7780459>.
- Kang E, Min J, Ye JC. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Med Phys*. 2017;44(10):360–75.
- Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y. Residual dense network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE Conference; 2018. p. 2472–2481. <https://ieeexplore.ieee.org/document/8578360>.
- Wang H, Wang Q, Gao M, Li P, Zuo W. Multi-scale location-aware kernel representation for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE Conference; 2018. p. 1248–1257. <https://ieeexplore.ieee.org/document/8578234>.
- Ren D, Zuo W, Hu Q, Zhu P, Meng D. Progressive image deraining networks: a better and simpler baseline. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE Computer Society; 2019. p. 3937–3946. <https://www.computer.org/csdl/proceedings-article/cvpr/2019/329300d932/1gyrWFIIOfe>.
- Gu S, Zhang L, Zuo W, Feng X. Weighted nuclear norm minimization with application to image denoising. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE Conference; 2014. p. 2862–2869. <https://ieeexplore.ieee.org/document/6909762>.
- Buades A, Coll B, Morel JM. A review of image denoising algorithms, with a new one. *Multiscale Model Simul*. 2005;4(2):490–530.
- Portilla J, Strela V, Wainwright MJ, Simoncelli EP. Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Trans Image Process*. 2003;12(11):1338–51.
- Chen H, Zhang Y, Kalra MK, Lin F, Chen Y, Liao P, et al. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Trans Med Imaging*. 2017;36(12):2524–35.
- Gondara L. Medical image denoising using convolutional denoising autoencoders. In: 2016 IEEE 16th international conference on data mining workshops (ICDMW). IEEE; 2016. p. 241–6.
- Zheng X, Liao Z, Hu S, Li M, Zhou J. Improving spatial adaptivity of nonlocal means in low-dosed CT imaging using pointwise fractal dimension. *Comput Math Methods Med*. 2013;2013:1–8. https://downloads.hindawi.com/journals/cm/2013/902143.pdf?_gl=1*1ixtmat*_ga*MTk4NzQxMjMzNS4xNjg3ODU4ODcw*_ga_NF5QFMJTV5*MTY5NjI1NzA4MS44MCAxLjE2OTYyNTcxMDYuMzUuMC4w&_ga=2.89033048.2122471803.1696242501-1987412335.1687858870.
- Duan J, Lu W, Tench C, Gottlob I, Proudlock F, Samani NN, et al. Denoising optical coherence tomography using second order total generalized variation decomposition. *Biomed Signal Process Control*. 2016;24:120–7.
- Ai D, Yang J, Fan J, Cong W, Wang Y. Adaptive tensor-based principal component analysis for low-dose CT image denoising. *PLoS ONE*. 2015;10(5):0126914.
- Al-Ameen Z, Sulong G, Rehman A, Al-Rodhaan M, Saba T, Al-Dhelaan A. Phase-preserving approach in denoising computed tomography medical images. *Comput Methods Biomech Biomed Eng Imaging Vis*. 2017;5(1):16–26.
- Trinh DH, Luong M, Rocchisani JM, Pham CD, Pham HD, Dibos F. An optimal weight method for CT image denoising. *J Electron Sci Technol*. 2012;10(2):124–9.
- Chen C, Xiong Z, Tian X, Wu F. Deep boosting for image denoising. In: Proceedings of the European Conference on Computer Vision (ECCV). Springer Conference; 2018. p. 3–18. https://link.springer.com/chapter/10.1007/978-3-030-01252-6_1.
- Tripathi S, Lipton ZC, Nguyen TQ. Correction by projection: Denoising images with generative adversarial networks. *arXiv preprint arXiv:1803.04477*. 2018.
- Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*. 2015.
- Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. In: *Icml*. ACM Digital Library; 2010. <https://dl.acm.org/doi/10.5555/3104322.3104425>.

40. Singh KK, Siddhartha M, Singh A. Diagnosis of coronavirus disease (covid-19) from chest x-ray images using modified xceptionnet. *Rom J Inf Sci Tech.* 2020;23(657):91–115.
41. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980.* 2014.
42. Angelov P, Almeida Soares E. SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification. *MedRxiv.* 2020. <https://www.medrxiv.org/content/10.1101/2020.04.24.20078584v3#~:text=In%20this%20paper%2C%20we%20build,hospitals%20from%20Sao%20Paulo%2C%20Brazil.>
43. Soares E, Angelov P, Biaso S, Froes MH, Abe DK. SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification. *MedRxiv.* 2020:2020–04.
44. Yang X, He X, Zhao J, Zhang Y, Zhang S, Xie P. COVID-CT-dataset: a CT scan dataset about COVID-19. *arXiv preprint arXiv:2003.13865.* 2020.
45. Kermany DS, Goldbaum M, Cai W, Valentim CC, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell.* 2018;172(5):1122–31.
46. Al-Ameen Z, Sulong G. Attenuating noise from computed tomography medical images using a coefficients-driven total variation denoising algorithm. *Int J Imaging Syst Technol.* 2014;24(4):350–8.
47. Ghadrdan S, Alirezaie J, Dillenseger JL, Babyn P. Low-dose computed tomography image denoising based on joint wavelet and sparse representation. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE; 2014. p. 3325–8.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

