

Research Article

X-Ray Breast Images Denoising Method Based on the Convolutional Autoencoder

Chensheng Yang, Junbo Ye, Yanwei Wang , and Chengling Song

School of Mechanical Engineer, Heilongjiang University of Science & Technology, Harbin, Heilongjiang, China

Correspondence should be addressed to Yanwei Wang; wangyanwei@usth.edu.cn

Received 2 June 2022; Revised 23 August 2022; Accepted 3 October 2022; Published 22 November 2022

Academic Editor: Xiaofeng Li

Copyright © 2022 Chensheng Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Considering the potential risk of X-ray to patients, denoising of low-dose X-ray medical images is imperative. Inspired by deep learning, a convolutional autoencoder method for X-ray breast image denoising is proposed in this paper. First, image symmetry and flip are used to increase the number of images in the public dataset; second, the number of samples is increased further by image cropping segmentation, adding simulated noise, and producing the dataset. Finally, a convolutional autoencoder neural network model is constructed, and clean and noisy images are fed into it to complete the training. The results show that this method effectively removes noise while retaining image details in X-ray breast images, yielding higher peak signal-to-noise ratio and structural similarity index values than classical and novel denoising methods.

1. Introduction

Medical images are frequently utilized in modern clinical diagnosis and therapy to aid in disease diagnosis and treatment evaluation, among other things. They provide a crucial foundation for clinical diagnosis and treatment. Unlike natural photos, medical images generate a lot of signal-related noise during the creation process; therefore, the contrast is lower and the noise is more visible [1]. As a type of medical imaging, X-ray is of great value in early breast cancer detection. Patients should receive mammography with the lowest possible radiation dose [2]. The most common way to reduce the radiation dose is to reduce the X-ray flux by decreasing the operating current and shortening the exposure time of the X-ray tube. However, the weaker the X-ray flux, the more-noisy the reconstructed image will be. Noise can blur the image, obscure important information in the image, and make disease analysis and diagnosis more difficult. Therefore, it is important to study how to remove the noise from X-ray images.

In the past few decades, many scholars have been proposing new image denoising algorithms as image noise has been intensively studied. Traditional image denoising

models can be classified into four categories based on spatial domain, transform domain, sparse representation, and natural statistics. Among them, the representative methods are the median filtering method based on the spatial domain [3], which ignores the characteristics of each pixel, and the image will be more seriously blurred after denoising; the BLS-GSM [4] based on the transform domain lose some useful information while denoising; the NLSC [5] based on the sparse representation has a long computation time and low denoising efficiency. The natural statistics-based BM3D [6] can only filter a specific noise. Although these algorithms are effective in removing noise, they often inevitably result in loss of texture information in the medical image and excessive smoothing of edges, which adversely affects diagnosis. It is still a challenging problem to remove the noise while retaining the detailed information in the medical image.

With the improvement of hardware computing power, the powerful learning and fitting capabilities of neural networks have shown great potential in image processing [7]. For example, convolutional neural networks (CNNs) have been applied to image classification, target detection, image segmentation, image denoising, etc. Currently, many scholars

have used CNN to denoise medical images such as low-dose CT images [8, 9], OCT images [10], and MRI images [11, 12], ultrasonography images [13], and so on. Kim et al. improved the BM3D method by proposing a method to assign different weights to each block according to the degree of denoising [14]. Although the detailed information of the image can be well recovered, the Gibbs effect will be produced after denoising, and the artifacts will be produced which cannot be eliminated. Guo et al. proposed a median filtering method based on adaptive two-level threshold to solve the problems of low contrast and blurred boundary of traditional weighted median filters [15]. This method has a very good denoising effect on CT images of COVID-19, but it is not suitable for denoising mixed noise, which is easy to cause image blur and discontinuity. Jia et al. proposed a pyramid dilated CNN [16], which uses dilated convolution to expand the network's receptive field and obtain more image details. This method has good denoising effect on both gray image and color image. However, the dilation rate needs be adjusted according to the size of the object in the input image to avoid image discontinuity. Huang et al. proposed a denoising GAN based on the U-Net discriminator [17], which can not only give feedback to each pixel in the image but also use U-Net to focus on the global structure at the semantic level. This method has a very good denoising effect on low-dose CT images, but the model is complicated and difficult to train. However, there are few studies on the denoising of X-ray breast images. This inspired us to use a CNN-based denoising model to remove noise from X-ray breast images and improve the quality of X-ray breast images. This paper proposes an X-ray breast image denoising method based on a convolutional autoencoder. First is by expanding the public breast dataset of Mammographic Image Analysis Society MiniMammographic Database (MIAS) and then centrally cropping the key parts to intercept the data containing key medical information while further expanding the number of samples; second is by adding Gaussian noise and salt and pepper noise to the sample data to generate noisy images and then combining the clean and noisy images into a dataset; and the final one is building an autoencoder denoising model based on CNN and completing the training. The results of the experiments show that the method can effectively remove various levels of blending noise in medical images while retaining image detail texture. The following are the main contributions of this paper: (1) We achieved better results than the current denoising methods, with higher peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM); (2) we proposed a mathematical model to simulate blending noise which was implemented by code to add different levels of blending noise in medical images; (3) in the MIAS dataset, the performance of this method is tested from the perspective of multiple index analysis, which fully verifies the efficiency and superiority of this method; and (4) we completed end-to-end modeling for medical image denoising.

2. Proposed Method

2.1. Simulated Medical Image Noise. In imaging, complex noise sources include Gaussian, impulse, pretzel, and scatter

noise [18]. Salt and pepper noise, Gaussian noise, and other types of noise are common in medical images. Because noise-free images are not readily available in clinics, and the network model of proposed method requires high-resolution noise-free images to train the network. As a result, the open dataset is chosen, and the open dataset is a clear clean image free of noise obtained by adding simulated noise to the image to obtain paired training data to train the noise removal network model. In this paper, noise is added to an image by adding noise components to the values of the image's corresponding pixels. The noise model can be defined as follows:

$$N_{(h,w,c)} = O_{(h,w,c)} + Y_{(h,w,c)} + G_{(h,w,c)}, \quad (1)$$

where (h, w, c) represents the pixel point in the image, h represents the image height, w represents the image width, c represents the number of channels, and when $c = 1$ represents the single-channel gray image, $O_{(h,w,c)}$ represents original images, $Y_{(h,w,c)}$ represents salt and pepper noise data, and $G_{(h,w,c)}$ represents Gaussian noise data.

2.1.1. Generation of Gaussian Noise. Gaussian noise is a kind of noise that obeys Gaussian distribution, where μ represents the expectation of Gaussian noise data and σ represents the variance of Gaussian noise data. In this paper, the method of adding Gaussian noise to medical images is as follows:

$$Z \sim N(\mu, \sigma), \quad (2)$$

$$G_{(h,w,c)} = kZ_{(h,w,c)},$$

where k represents noise intensity, $Z_{(h,w,c)}$ represents denoising images, considering the pixel of images which may over 255 before adding noise, and the pixel of the noisy image is limited to avoid data overflow in the computer, and the restriction is as follows:

$$N_{(h,w,c)} \begin{cases} 0, N_{(h,w,c)} < 0, \\ 255, 255 < N_{(h,w,c)}, \\ N_{(h,w,c)}, 0 \leq N_{(h,w,c)} \leq 255. \end{cases} \quad (3)$$

2.1.2. Salt and Pepper Noise Generation. Salt and pepper noise generation is also a common noise in the image, as the name suggests, and pepper represents black, and salt represents white. Salt and pepper noise represents white or black points. The generation way is as follows:

$$N_{(h,w,c)} \begin{cases} 255, 0 \leq W \leq 1 - \text{SNR} \text{ and } Z = 1, \\ 0, 0 \leq W \leq 1 - \text{SNR} \text{ and } Z = 0, \\ O_{(h,w,c)}, \text{ other,} \end{cases} \quad (4)$$

$$W \sim U(0, 1),$$

$$Z \sim B(1, 0.5),$$

where SNR represents signal-to-noise ratio, which is expressed in the range of $[0, 1]$, W obeys the uniform distribution of parameter $[0, 1]$, Z obeys the 0.1 distribution when $P(1) = 0.5$ and $W \in [0, (1 - \text{SNR})]$, and the value of

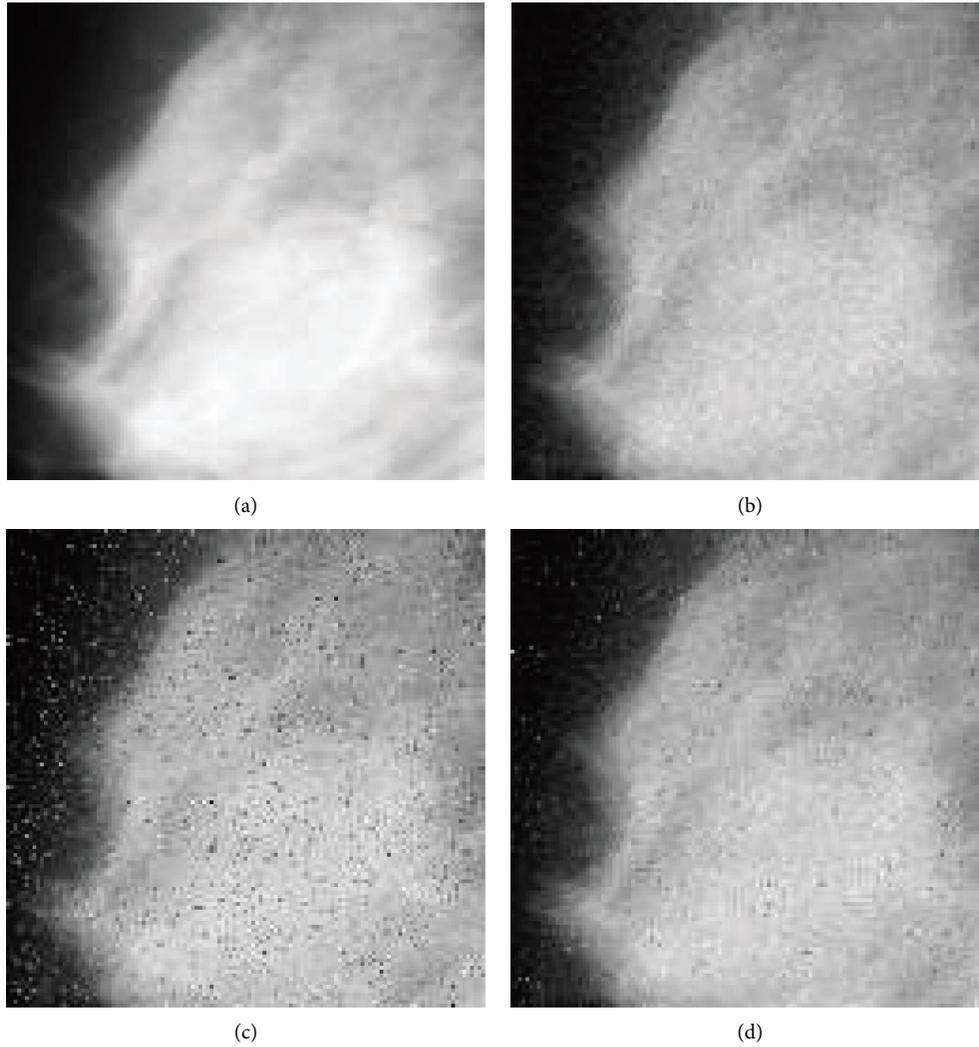


FIGURE 1: Noise images. (a) Origin image. (b) Gaussian noise. (c) Salt and pepper noise. (d) Mixed noise.

pixels in the original image in each channel has a 50% chance of being 255 or 0. When the pixel value is 255, it shows black salt and pepper noise points, and when it is 0, it shows white salt and pepper noise points.

In Figure 1, (a) represents the original image, (b) represents the image after adding Gaussian noise, (c) represents the image after adding salt and pepper noise, and (d) represents the mixed image after adding Gaussian noise and salt and pepper noise.

2.2. Convolutional Autoencoder Model. The concept of autoencoder was first proposed by Rumelhart et al. [19], which was originally applied to dimension reduction of complex data. The autoencoder includes encoding and decoding and trains the network through reverse propagation, so that the output is equal to the input [20]. The encoder can reduce the dimension of data compression, reduce the amount of data, and retain the most critical feature information. The function of the decoder is opposite to the encoder. The decoder restores the compressed data and restores the input data through decoding. The full

connection layer of the traditional autoencoder will stretch the data into one dimension, thus losing the spatial information of the two-dimensional image data. The CNN has strong performance in extracting spatial feature information of images, which can compensate for the loss of spatial information when an autoencoder extracts features. Based on the traditional autoencoder, convolutional autoencoder combines the advantages of CNN and realizes the deep neural network through the superposition of convolution layer, activation layer, and pooling layer to complete the extraction of image detail features [21]. This paper uses the convolution layer, pooling layer, activation layer, and deconvolution layer in the CNN to construct an encoder and decoder. Encoders and decoders can be defined as

$$\begin{aligned} M &= f(W_i * X + b_i), \\ N &= f(W_i * X + b_i), \end{aligned} \quad (5)$$

where W_i and b_i represent the weight matrix and bias of each convolution layer, respectively; i represents the i convolution layers; f and h represent the encoder and decoder, respectively; X represents the input medical image data; M

represents the feature information of input image data after feature extraction by the encoder; and N represents the medical image data generated after the feature information is decoded.

2.3. Loss Function and Optimization Algorithm. The mean squared difference loss function is used in this paper to compute the difference between the denoised noisy image and the clear image. The mean squared difference of image reconstruction in this paper is as follows:

$$\text{MSE} = \frac{1}{M} \sum_{i=1}^M (f_p(y_i, \theta) - f_i)^2, \quad (6)$$

$$\theta = \{w, b\},$$

where M represents the number of training samples; θ represents the parameters in the network model; w represents weight; and b represents a bias. $f_p(y_i, \theta)$ represents the output image of the denoising model, and f_i denotes the noise-free image corresponding to the output image of the model. When the mean square difference is smaller, the reconstruction effect is better. To make the model converge faster and better [22], the learning process of the model is optimized using the Adam algorithm, which is based on the gradient descent method but differs from the traditional stochastic gradient descent algorithm that can make the model converge faster and better. To update all the weights, stochastic gradient descent uses a single learning rate that does not change during the training process. In contrast, the Adam algorithm calculates the first-order moment estimates and second-order moment estimates of the gradient to design independent adaptive learning rates for different parameters, which has significant advantages for model optimization of large-scale datasets. The Adam algorithm's optimization procedure is as follows:

$$\left\{ \begin{array}{l} g_t = \nabla_{\theta} F_t(\theta_{t-1}), \\ m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \\ v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \\ \hat{m}_t = \frac{m_t}{(1 - \beta_1^t)}, \\ \hat{v}_t = \frac{v_t}{(1 - \beta_2^t)}, \\ \theta_i^t = \theta_{i-1}^t - \alpha \frac{\hat{m}_i^t}{\sqrt{\hat{v}_i^t + \varepsilon}}, \end{array} \right. \quad (7)$$

where t is the current time step; F is the optimization function of the model; g_t is the gradient of the optimization objective function F at the time step t ; θ is the model parameter vector; and θ_i^t is the value of the first element θ_i in the parameter

vector θ when t time step. α is the step length; m is the first-order moment estimation of gradient; β_1 is the exponential decay rate of m ; v is the second-order moment estimation of gradient; β_2 is the exponential decay rate of v ; and $\beta_1, \beta_2 \in [0, 1)$; $\varepsilon = 10^{-8}$ equation (6) represents each update of the parameter vector θ . However, as the neural network is trained, the distribution of the parameters of the next layer will change from the parameters of the previous layer. This causes a constant change in the distribution of parameters, which slows down training. This issue can be addressed by incorporating a normalization layer into the network model and performing normalization for each small batch of samples [23]. Suppose the input of a layer in the model is $x = (x^1, x^2, \dots, x^n)$ and the set of samples is $B = (x_1, x_2, \dots, x_m)$. The batch normalization method is as follows:

$$\begin{aligned} \mu_B &= \frac{1}{m} \sum_{i=1}^m x_i, \\ \sigma_B^2 &= \frac{1}{m} \sum_{i=1}^m (x_i - \mu_B)^2, \\ x^{(\hat{n})} &= \frac{x^{(n)} - \mu_B}{\sigma_B}, \\ y^{(n)} &= \gamma^{(n)} * x^{(\hat{n})} + \beta^{(n)}, \end{aligned} \quad (8)$$

where $x^{(n)}$ is the n th dimension of the input x ; μ_B is the sample set B 's expectation; σ_B^2 is the variance of the sample set B ; $x^{(k)}$ is the input's regularization result; $y^{(n)}$ is the batch regularization result of $x^{(n)}$; and $\gamma^{(n)}$ and $\beta^{(n)}$ is the parameters to be learned. The network model species' parameters are then continuously updated using back-propagation and optimization algorithms to produce the best denoising model.

3. Design of the Denoising Method

3.1. Denoising Process. Figure 2 depicts the overall flow of the denoising method. To begin with, a public dataset is chosen for data preprocessing. The MIAS dataset with 322 high-definition 1024 * 1024 mammographic images are used in this paper. The MIAS dataset is then cropped, rotated, and flipped to increase the volume of training data to 3000 images with 336 * 336 resolution. Second, simulated noise is added to the cropped and expanded clear image data, and the noise images and clear images are combined into a dataset that can be passed into the model, with the samples in the dataset divided into an 8:1:1 training, test, and validation set. The convolutional autoencoder neural network model is then built, and the training set is fed into the model to complete the model's training. The validation set is then fed into the trained model, which checks the model's denoising effect and outputs the model parameters. Finally, the model is loaded and the test set is fed into it to test the model's denoising effect.

Figure 2 depicts the process of the denoising method used in this paper: first, the public dataset is converted into a

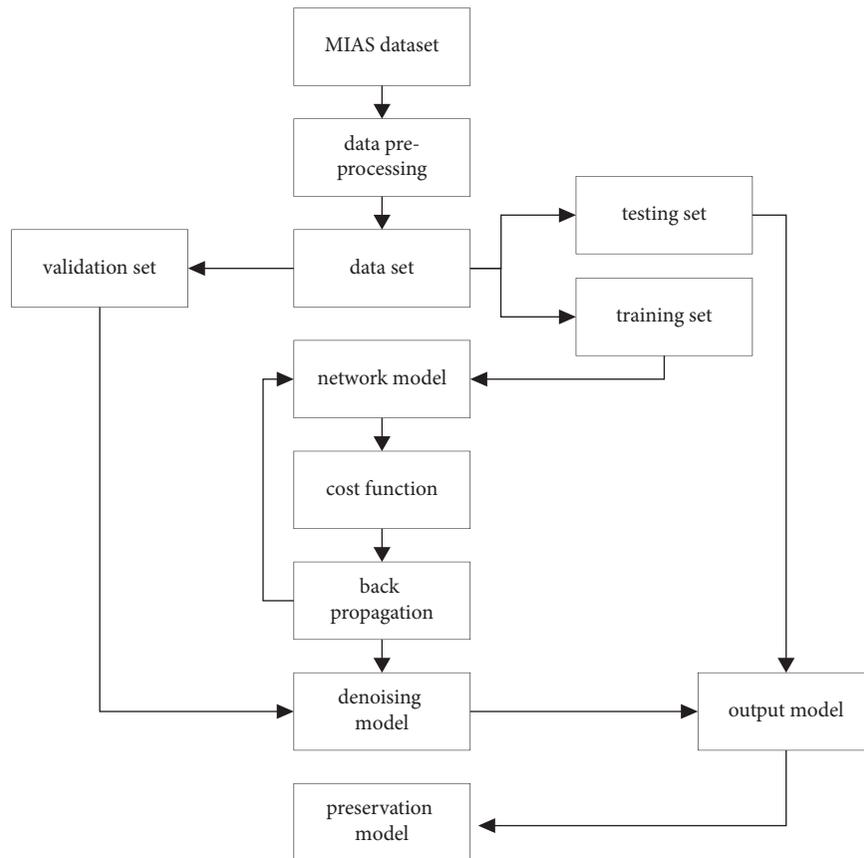


FIGURE 2: Flow chart of the denoising method.

dataset, and then the dataset is divided into a training set, a validation set, and a test set. The model training is completed with the help of the training set. The validation set is used to determine whether the trained model is overfitting. The test set is used to validate the model's denoising effect. Finally, the model is then saved.

3.2. Model Structure. The model's goal is to create an end-to-end mapping of noisy image noise to a clear image. The process of converting a noisy image to a clear image is known as image denoising. This paper's convolutional autoencoder denoising network is divided into two parts: encoder and decoder. The encoder compresses and downscales the input noisy image through feature extraction via convolution before transforming it into an abstract mathematical description. By deconvolution, the decoder converts this abstract description into an image. As the cost function is reduced, the decoder generates images that are increasingly similar to the target image. The structural model of convolutional autoencoder proposed in this paper is shown in Figure 3, where the input is the noisy image and the output is the image with completed noise reduction processing. The model contains a total of 12 convolutional layers, the first 6 layers belong to the encoder structure, and the last 6 layers belong to the decoder structure. The first four layers of the encoder contain 32 convolutional kernels, the fifth layer contains

64 convolutional kernels, and the sixth layer contains 128 convolutional kernels. The first two layers of the decoder contain 64 convolutional kernels, layers 3 and 4 contain 32 convolutional kernels, layer 5 contains 16 convolutional kernels, and the last layer contains 1 convolutional kernel. The size of all convolutional kernels in the model is 3×3 . A 3×3 convolution kernel has the same perceptual field after three convolutions as a 7×7 convolution kernel after 1 convolution; that is, a small-sized convolution kernel can obtain the same perceptual field as a large-sized convolution kernel through multiple convolution operations, and the multiple kernel can increase the network depth and improve the network's nonlinear fitting ability [24]. Furthermore, the small convolution kernel reduces the number of parameters and improves the model's convergence speed. As a result, small-sized convolutional kernels of 3×3 are used in the models developed in this paper. A ReLU layer is added as the activation function to ensure the gradient descent speed and model convergence. After the encoder's third layer, the model adds a maximum pooling layer. The pooling layer can reduce redundant information while also expanding the receptive field.

Figure 3 depicts the end-to-end medical image denoising model developed in this paper, with the noisy image as the input, and the regenerated image after noise reduction as the output.

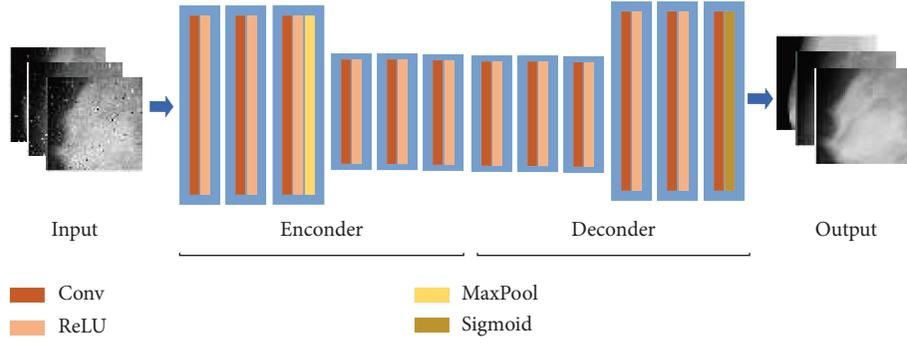


FIGURE 3: Structure model of the convolutional autoencoder.

4. Experimental Analysis and Results

4.1. Experimental Set-Up

- (1) *Data for Training and Testing.* To train the proposed model's denoising ability under different levels of noise, we used the MIAS dataset to generate the training dataset. In the dataset used for model training, four different intensities of Gaussian noise and four different intensities of salt and pepper noise were set. The standard deviations of Gaussian noise are set to 40, 50, 60, and 70, respectively; the signal-to-noise ratios of salt and pepper noise were set to 0.99, 0.97, 0.95, and 0.93, respectively. The change of mixed noise intensity in the proposed method is completed by setting parameters in the program. An image training set consists of 2400 $336 * 336$ MIAS breast X-ray images. The test set and validation set were 300 MIAS breast X-ray images of $336 * 336$.
- (2) *Super Parameter Setting and Model Training.* Based on a large number of experimental results, the depth of the model is determined to be 12 layers by weighing the receptive field and the model depth. When the model is trained, the number of mini-batch is 4, the initial learning rate is 0.01, and after three rounds of iteration, the learning rate is 0.0001, so that the model continues to learn. If the loss function of the model does not change after five consecutive iterations, the model stops iterative training.
- (3) *Experimental Environment.* To train the model with a better denoising effect, a large number of experiments are performed in this paper. The experimental environment mainly includes hardware and software. The hardware configuration is mainly Intel Core i7 4810M CPU and NVIDIA GeForce960M GPU, in which the graphics card memory is 2G, and the running memory is 12G. The software environment includes operating system Windows 10 64-bit, programming language Python 3.8, deep learning framework PyTorch 1.10, and Python data integration package Anaconda.
- (4) *Evaluation Indicators.* Because it is a medical image, this paper employs both objective and subjective evaluation methods. The subjective

evaluation is the human's subjective reaction to the denoised image. PSNR is the most commonly used image quality evaluation criterion among image researchers, but the visual effect of an image is not identical to the PSNR. Because the PSNR index is not always the highest for the best visual effect in a set of images, it cannot be used as an absolute criterion to assess image quality [25]. SSIM structural similarity, on the other hand, comprehensively takes into account nonstructural distortions such as brightness and contrast as well as structural distortions such as junction noise intensity and blurring degree [26]. This means that in the context of visual perception, SSIM can better reflect the image quality. Therefore, in this experiment, we comprehensively considered PSNR and SSIM to evaluate the denoising effect of the model. The calculation formula is as follows:

$$\text{PSNR} = 10 * \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right),$$

$$\left\{ \begin{array}{l} l(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}, \\ c(x, y) = \frac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + c_2}, \\ s(x, y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}, \end{array} \right. \quad (9)$$

$$\text{SSIM}(x, y) = [l(x, y)^\alpha * c(x, y)^\beta * s(x, y)^\gamma],$$

where MAX is the maximum value of pixels in the image, and all the images involved in this paper are 8-bit color depth, so $\text{MAX} = 255$; MSE is the mean square value; x and y are input images; μ is the mean value of the input image; σ is the variance of the input image; and α, β, γ are power exponents used to adjust the importance of input images. Since the input image is equally important, here we take the power

TABLE 1: The PSNR (dB) and SSIM value of different methods at the MIAS dataset (SNR = 0.99, $k = 0.1$).

Different methods	$\sigma = 40$		$\sigma = 50$		$\sigma = 60$		$\sigma = 70$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Noise image	23.37	0.33	22.76	0.27	21.97	0.21	21.21	0.17
Literature [14]	37.39	0.90	36.19	0.87	34.83	0.82	34.22	0.80
Literature [15]	37.51	0.90	37.24	0.89	35.26	0.83	34.68	0.81
Literature [16]	38.37	0.91	38.03	0.90	37.37	0.88	35.29	0.84
Literature [17]	38.91	0.92	38.25	0.91	37.28	0.88	35.87	0.86
Proposed	38.95	0.93	38.31	0.92	37.48	0.91	36.36	0.90

TABLE 2: The PSNR (dB) and SSIM of different methods at the MIAS dataset ($k = 0.1$, $\sigma = 60$).

Different methods	SNR = 0.99		SNR = 0.97		SNR = 0.95		SNR = 0.93	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Noise image	21.97	0.21	19.22	0.14	17.55	0.10	16.32	0.07
Literature [14]	37.15	0.90	36.15	0.88	34.72	0.85	32.41	0.82
Literature [15]	36.82	0.89	35.64	0.87	33.48	0.84	31.52	0.81
Literature [16]	37.24	0.91	36.23	0.88	36.04	0.88	34.91	0.85
Literature [17]	37.29	0.91	36.59	0.89	36.21	0.88	35.73	0.87
Proposed	37.48	0.91	37.07	0.90	36.47	0.89	35.93	0.88

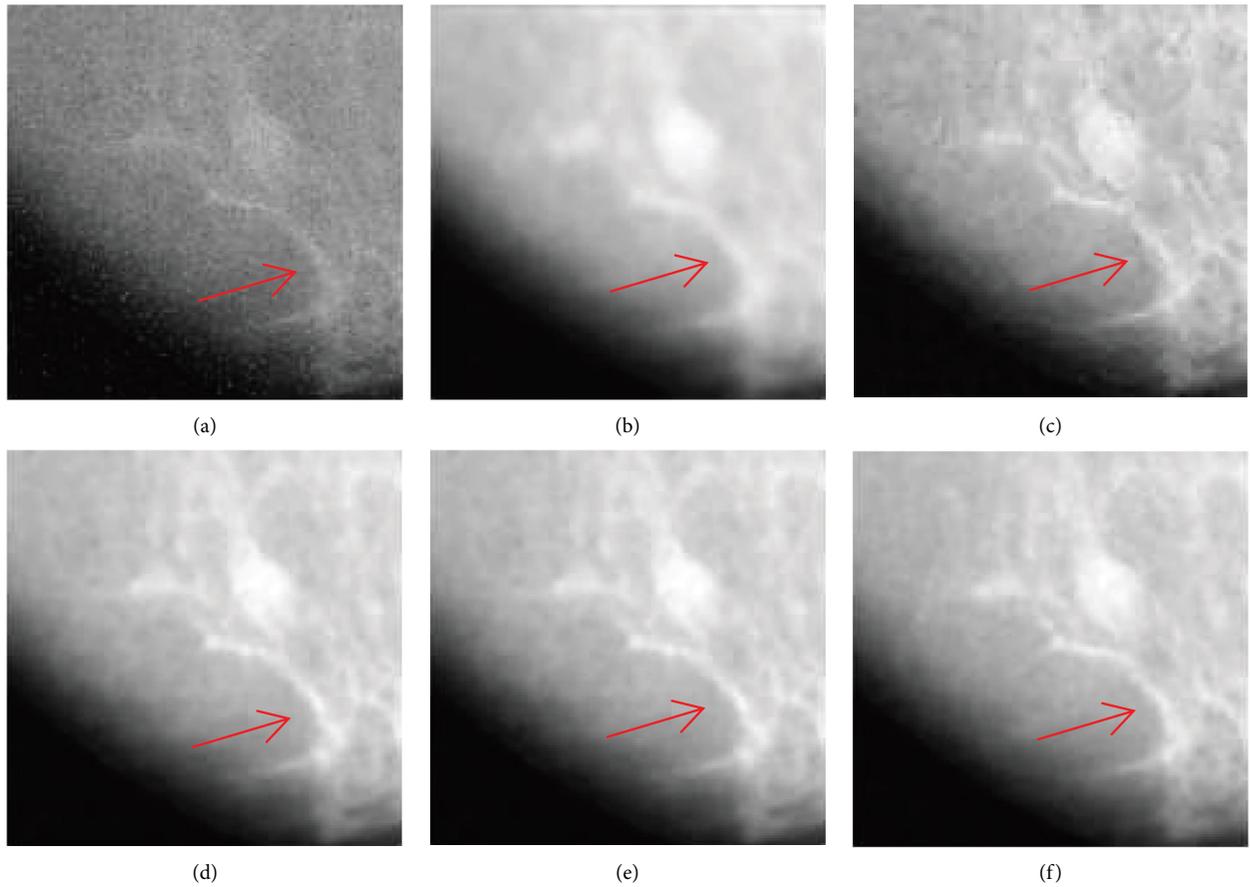


FIGURE 4: Denoising results of different methods. (a) Noisy image (PSNR = 21.53 dB) (SSIM = 0.26). (b) Literature [14] (PSNR = 33.47 dB) (SSIM = 0.87). (c) Literature [15] (PSNR = 33.58) (SSIM = 0.83). (d) Literature [16] (PSNR = 36.18 dB) (SSIM = 0.88). (e) Literature [17] (PSNR = 37.41 dB) (SSIM = 0.90). (f) Proposed method (PSNR = 37.83 dB) (SSIM = 0.91).

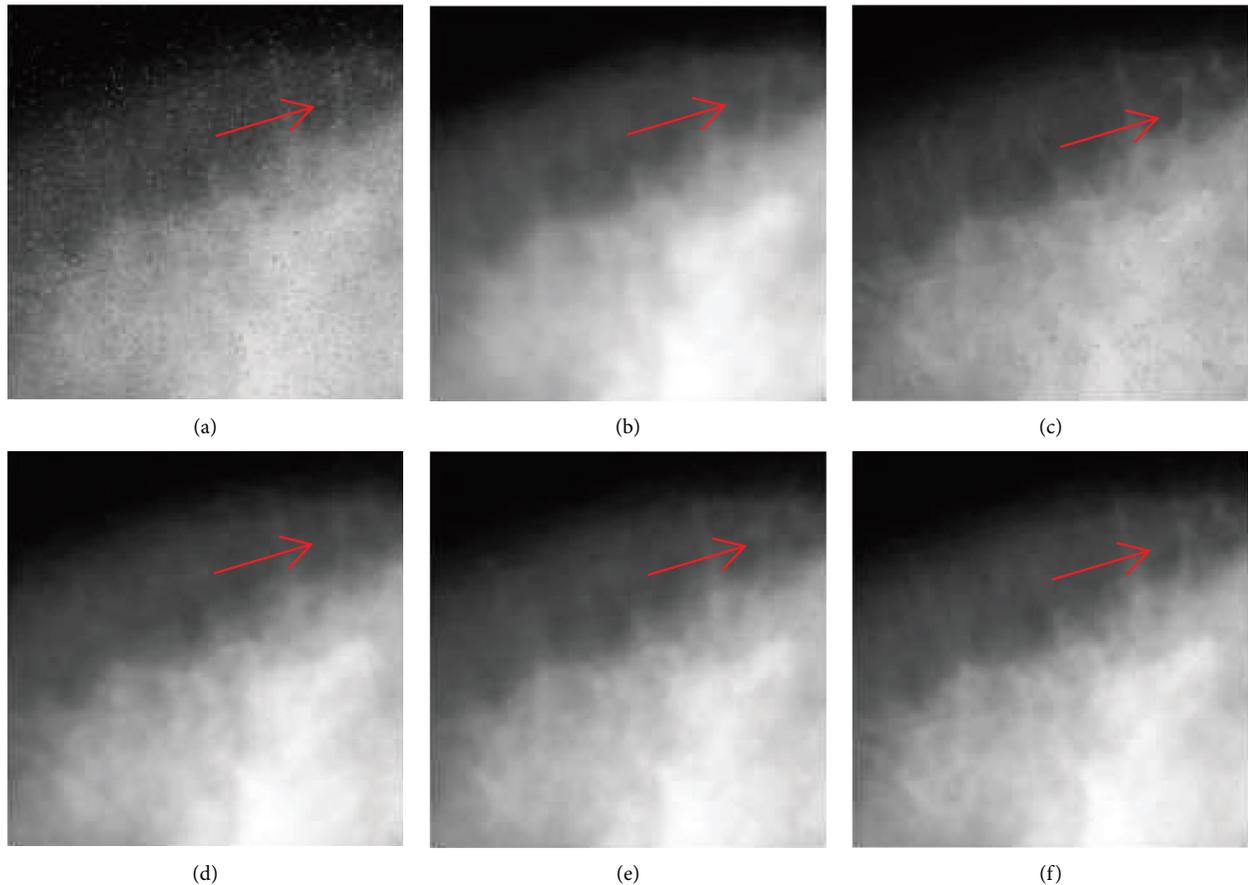


FIGURE 5: Denoising results of different methods. (a) Noisy image (PSNR = 22.58 dB) (SSIM = 0.27). (b) Literature [14] (PSNR = 32.17 dB) (SSIM = 0.51). (c) Literature [15] (PSNR = 33.78 dB) (SSIM = 0.75). (d) Literature [16] (PSNR = 33.49 dB) (SSIM = 0.82). (e) Literature [17] (PSNR = 34.91 dB) (SSIM = 0.86). (f) Proposed method (PSNR = 35.17 dB) (SSIM = 0.87).

index as 1; l represents brightness, c represents contrast, and s represents structure.

4.2. Results and Discussion. To show the proposed method's effectiveness in removing noise from X-ray breast pictures and its benefits in maintaining texture. In this study, we conduct an experimental comparison of the proposed method's denoising performance with that of novel denoising methods. These methods include those mentioned in literatures [14], [15], [16], and [17]. In our experiments, we denoise the dataset constructed based on MIAS, and the experimental results are shown in Tables 1 and 2. All methods present visually well denoised results to some degree. When $\sigma = 40$, the PSNR of literature [16] and literature [15] are 0.98 dB and 0.12 dB greater than literature [14], respectively. The PSNR of literature [17] is 0.54 dB higher than literature [16], while the PSNR of the proposed approach in this study is 0.04 dB higher than literature [17]. The average SSIM values of literature [16] are improved by 0.01 compared to literature [14] and literature [15], and the SSIM of literature [17] is improved by 0.01 compared to literature [16]. The SSIM of the proposed method in this paper is improved by 0.01 compared to literature [17]. In addition, with the increase in noise intensity, $\sigma = 50$, $\sigma = 60$,

and $\sigma = 70$, and it can be seen from the experimental data in Table 1 that the proposed method in this paper performs well on PSNR and SSIM. When SNR = 0.99, the PSNR of literature [16] and literature [14] are 0.42 dB and 0.33 dB greater than literature [15], respectively. Literature [17] is 0.05 dB greater than literature [16], while the PSNR of the proposed approach in this study is 0.19 dB higher than literature [17]. The average SSIM values of literature [16], literature [17] and literature [14] are improved by 0.02, 0.02, and 0.01 compared to literature [15], respectively, and the SSIM of the proposed method has the same value with literature [17] and literature [16]. In addition, with the decrease in signal-to-noise ratio, SNR = 0.97, SNR = 0.95, and SNR = 0.93, and it can be seen from the experimental data in Table 2 that the proposed method performs well on PSNR and SSIM.

We chose two images from the test set to display in Figures 4 and 5 to further confirm the visualization effect of the technique for X-ray image denoising described in this research. In Figure 4, the noise intensity factor is 0.1, the variance of the noisy image is = 0.6, and the SNR is 0.99. In Figure 5, the noise intensity factor is 0.1, the variance of the noisy image is 0.6, and the SNR is 0.93. It is observed that the literature [14] produces a significant blurring effect while removing noise and fails to recover the detailed texture information of the image. A weakness can be seen in

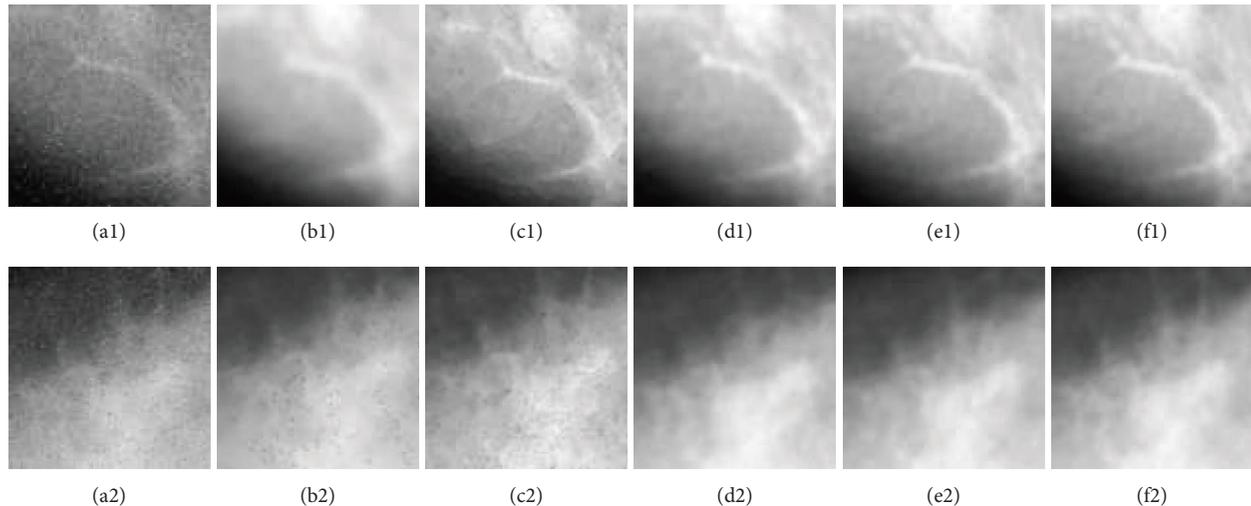


FIGURE 6: Local detail enlarged image.

the harmonious interplay between noise removal and preservation of edge features and texture information. Literature [15] is a highly competitive technique that has shown higher performance in both noise removal and texture information preservation. The Gibbs effect, which cannot be removed, will be produced by BM3D since it executes the filtering operation in the transform domain, which is comparable to windowing the image signal. After denoising, the Gibbs effect will produce pseudo-textures that resemble scratches. The clinical diagnosis might be significantly impacted by pseudo-texture. Literature [16] is a relatively novel method at present, and the pyramid convolution method expands the perceptual field without increasing the number of parameters, and thus more image details can be obtained. Figure d shows the denoising results of the method in the literature [16] for mixed noise, which not only removes the noise in the image but also retains the image detail information. The proposed method obtains a better denoising effect than literature [16], and the image detail texture is clearer. Literature [17] is a U-net based denoising method that has a simple structure and uses a completely different feature fusion approach to stitch the features in channels to form thicker features. Figure e shows the denoising results of the method in the literature [17] for mixed noise, and the noise in the image is removed more cleanly. The proposed denoising method in this paper obtains comparable denoising results to those of the literature [17].

To better show the detailed information about image denoising, we zoom in on the areas marked by red arrows in Figures 4 and 5, as shown in Figure 6. Figures 6(a1)–6(f1) correspond to the part pointed by the red arrow in Figure 4, respectively.

Figures 6(a2)–6(f2) correspond to the part pointed by the red arrow in Figure 6, respectively. From Figures 6(b1), 6(b2), 6(c1), and 6(c2), it can be found that the literature [14, 15] cannot produce a good denoising effect on the mixed noise in X-ray breast images, and the denoised images are blurred, and there are artifacts. Figures 6(d1), 6(d2), 6(e1), and 6(e2) show the enlarged details after denoising the noisy images in literature [16, 17]. Besides,

these two methods can remove the noise in X-ray breast images and retain the details better. Figures 6(f1) and 6(f2) show the detailed information of the X-ray breast image denoised by the proposed method. It can be seen that the denoising effect of the proposed method on X-ray breast images is slightly better than that of the literature [16, 17]. By showing the enlarged details, it can be seen that the proposed method has a better denoising effect on X-ray breast images, compared with the current novel denoising methods.

5. Conclusions

In this paper, the image denoising method is described. In view of the shortcomings of the current traditional denoising methods, a convolution autoencoder neural network denoising model for medical images is proposed by using the powerful extraction ability of the convolution neural network for image information and the learning ability of the autoencoder network structure. A solution to mixed noise in medical images is presented. The proposed method has the following advantages: (1) Denoising effect is obvious, in the removal of mixed noise in medical images, while retaining the details of the image texture information; (2) model denoising is fast and can complete a large number of medical image denoising tasks in a short time, which has a clinical application value; (3) end-to-end modeling is realized, and the denoised image can be obtained by inputting the noise image without manually adjusting the parameters. In addition, the experiment in this paper is limited by the memory of graphic processing unit. In the future, we can consider increasing the memory of the graphic processing unit and further optimize the model from the total amount of sample data and mini-batch.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest with any financial organizations regarding the material reported in this manuscript.

Acknowledgments

This work was supported by Natural Science Foundation of Heilongjiang Province of China under grant number LH2021F039.

References

- [1] S. V. Mohd Sagheer and S. N. George, "A review on medical image denoising algorithms," *Biomedical Signal Processing and Control*, vol. 61, p. 102036, 2020.
- [2] K. A. S. H. Kulathilake, N. A. Abdullah, A. Q. M. Sabri, and K. W. Lai, "A review on Deep Learning approaches for low-dose Computed Tomography restoration," *Complex & Intelligent Systems*, pp. 1–33, 2021.
- [3] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: interscale orthonormal wavelet thresholding," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 593–606, Mar 2007.
- [4] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, Nov 2003.
- [5] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Proceedings of the 2009 IEEE 12th International Conference on Computer Vision (ICCV)*, Kyoto, Japan, September 2010.
- [6] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug 2007.
- [7] G. Wang, "A perspective on deep imaging," *IEEE Access*, vol. 4, pp. 8914–8924, 2016.
- [8] H. M. Shan, Y. Zhang, Q. Yang et al., "Correction for "3D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2D trained network" [jun 18 1522-1534]," *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, p. 2750, Dec 2018.
- [9] Y. Lei, J. Zhang, and H. Shan, "Strided self-supervised low-dose ct denoising for lung nodule classification," *Phenomics*, vol. 1, 2021.
- [10] F. Shi, N. Cai, Y. Gu et al., "DeSpecNet: a CNN-based method for speckle reduction in retinal optical coherence tomography images," *Physics in Medicine and Biology*, vol. 64, no. 17, p. 175010, 2019.
- [11] B. Srinivas and G. Sasibhushana Rao, "A novel DeepCNN model for denoising analysis of MRI brain tumour images," *International Journal of Intelligent Information and Database Systems*, vol. 13, no. 2/3/4, pp. 393–410, 2020.
- [12] P. C. Tripathi and S. Bag, "CNN-DMRI: a convolutional neural network for denoising of magnetic resonance images," *Pattern Recognition Letters*, vol. 135, pp. 57–63, Jul 2020.
- [13] L. Bo, K. Xu, D. Feng, H. Mi, and Z. Jian, "Denoising convolutional auto-encoder based B-mode ultrasound tongue image feature extraction," in *Proceedings of the ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Brighton, UK, May 2019.
- [14] G. I. Ri, S. J. Kim, and M. S. Kim, "Improved BM3D method with modified block-matching and multi-scaled images," *Multimedia Tools and Applications*, vol. 81, no. 9, pp. 12661–12679, Apr 2022.
- [15] S. L. Guo, G. W. Wang, L. N. Han, X. W. Song, and W. T. Yang, "COVID-19 CT image denoising algorithm based on adaptive threshold and optimized weighted median filter," *Biomedical Signal Processing and Control*, vol. 75, p. 103552, 2022.
- [16] X. L. Jia, Y. L. Peng, J. Li, Y. H. Xin, B. Ge, and S. G. Liu, "Pyramid dilated convolutional neural network for image denoising," *Journal of Electronic Imaging*, vol. 31, no. 02, Article ID 023024, 2022.
- [17] Z. Z. Huang, J. P. Zhang, Y. Zhang, and H. M. Shan, "DUGAN: generative adversarial networks with dual-domain U-Net-Based discriminators for low-dose CT denoising," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.
- [18] A. E. Ilesanmi and T. O. Ilesanmi, "Methods for image denoising using convolutional neural network: a review," *Complex & Intelligent Systems*, vol. 7, no. 5, pp. 2179–2198, 2021.
- [19] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, Oct 1986.
- [20] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: a review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, Aug 2013.
- [21] J. Masci, U. Meier, D. Ciresan, and J. Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *International Conference on Artificial Neural Networks*, pp. 52–59, Springer, Berlin, Heidelberg, 2011.
- [22] D. Kingma and J. J. C. S. Ba, *Adam: A Method for Stochastic Optimization*, Computer Science, San Diego, CA, USA, 2014.
- [23] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning JMLR.org*, pmlr, New York, NY, USA, June 2015.
- [24] K. Simonyan and A. J. C. S. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <https://arxiv.org/abs/1409.1556>.
- [25] W. Zhou and A. C. J. I. S. P. M. Bovik, "Mean squared error: love it or leave it? A new look at Signal Fidelity Measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.
- [26] A. Horé and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proceedings of the 20th International Conference on Pattern Recognition, ICPR 2010*, pp. 23–26, Istanbul, Turkey, August 2010.