Recent Research Advances in Image Denoising

Yutong Yao *

Beijing University of Posts and Telecommunications, Beijing, China

* Corresponding Author Email: yaoyt2023@bupt.edu.cn

Abstract. Image denoising is a key area in image processing, especially important in applications such as medical imaging and remote sensing as the demand for high-quality images grows. This paper focuses on the latest research on deep learning-based denoising techniques, including Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), and Swin Transformer, and explores their principles, architectures, and effects of removing complex noise while maintaining image details. Deep learning models are capable of learning complex image features and adapting to different noise patterns and perform well in CT image and synthetic aperture radar (SAR) image denoising. Deep learning models have significant advantages over traditional methods in dealing with complex noise and preserving details. However, these methods still face challenges such as high computational requirements, dependence on large-scale datasets, and difficulty in adapting to realistic noise. This paper also discusses possible solutions to cope with these problems by creating real-world noisy datasets, optimizing model architectures and reducing computational costs, and looks at future research directions.

Keywords: Image denoising; deep learning; convolutional neural network; generative adversarial network; swin transformer.

1. Introduction

Images are widely used in many fields as one of the main information carriers. The clarity and quality of images are crucial for information transmission. Image noise is classified as Gaussian noise, Quantization noise, Speckle noise, Salt-and-pepper noise, and Periodic noise [1]. The quality of the image will be affected by these noises, which will affect follow-up processing tasks. Therefore, image denoising becomes one of the crucial steps in image processing.

To solve this problem, researchers used classical image-denoising methods, including mean filtering, Gaussian filtering, and median filtering. Researchers used these methods to remove the noise by processing the average of the pixel neighbourhood or other statistical methods [2]. However, these methods may blur the edges and details in the image while removing noise, resulting in a lack of clarity. Then, wavelet transform, and other nonlinear methods are introduced into image denoising [3]. Wavelet transform can process images at multiple scales, isolate high-frequency noise, and retain edge information effectively. However, it faces challenges such as significant computational demands and sensitivity to parameter tuning.

Researchers have been applying deep learning models to image-denoising problems recently because of the quick advancement of deep learning techniques. This type of approach can produce a stronger denoising ability by automatically learning the image's complicated features. In addition, it also shows great potential to improve detail reproduction further while generating high-quality noise-free images. Compared with the traditional models, the deep learning models have better noise removal performance and can adapt to different types of noise.

This paper gives an overview of recent advances in image denoising, especially in deep learning algorithms. This paper will discuss the use of convolutional neural networks and other models. At the same time, the application of deep learning denoising algorithm in practice is explained.

2. Deep learning

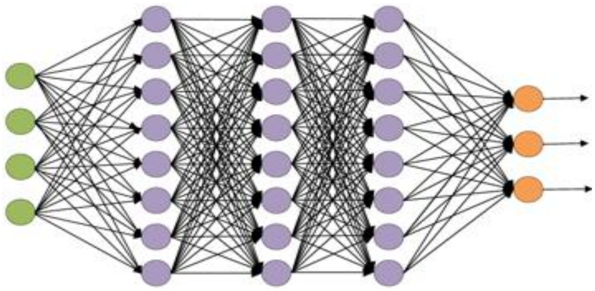

Fig 1. Deep learning model.

The structure of deep neural networks is mainly composed of an input layer, a hidden layer, and an output layer [4]. The input layer receives the raw data, as shown in Fig. 1, and processes it using several hidden layers with weighted connections. The output layer then presents the findings or projections. This design allows deep networks to model complex functions. The deep neural networks work like this:

$$y = f(x) = \sigma(W^l \cdots \sigma(W^2\sigma(W^1x + b^1) + b^2) \cdots + b^l) \tag{1}$$

The input vector is denoted by X, the weight matrix by W, the bias vector by b, the number of layers by l, and the output vector by y.

2.1. Convolutional neural networks

Convolutional neural networks typically consist of an input layer, a convolutional layer, a pooling layer, a nonlinear layer, and a fully connected layer [5].

The input layer receives the input data, typically an image of $h \times w \times c$ size. Convolutional layers extract image features. Each convolutional layer contains multiple filters, which are used to create feature images. The pooling layer is used to reduce the spatial size of the feature image and the computational complexity of the model. Nonlinear layers introduce nonlinearity through activation functions, enabling convolutional neural networks to learn and represent more intricate functional relationships. The fully connected layer combines the features extracted by the earlier convolutional and pooling layers to produce the final output.

In convolutional neural network denoising, DNCNN is one of the most classic and commonly used denoising methods [6].

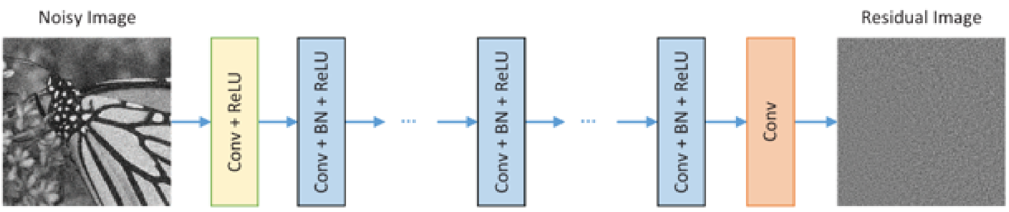

Fig 2. DNCNN model.

As shown in Fig.2, the DNCNN model is based on CNN and uses repetitive convolutional blocks. The output of the model is the noise image. The difference between the noise image and the expected noise is then used to determine the image-denoising result.

By learning the differences between noisy and clean images, the DNCNN model can separate noise from the original noisy image. During training, the model uses mean square error to measure the discrepancies between the noisy and clean images.

Compared with traditional networks, DNCNN can be used in more complex noise patterns with a significant improvement in denoising performance [7]. The residual learning method is used to learn the noise components in the image. In this manner, the model can more rapidly focus on the noisy areas, thereby preserving the image's detailed structure and enhancing the efficiency of the denoising process.

2.2. Generative Adversarial Networks

Goodfellow proposed a GAN generative adversarial network model [8]. In GAN, two neural networks learn by playing games with each other and reach the Nash equilibrium state through constant iteration.

The discriminator D and generator G are the two components that make up the generative adversarial network. Generator G produces data by simulating a machine, learning the pattern of real sample distributions, and creating pseudo-samples resembling real ones [9]. Discriminator D needs to determine whether the input data is from the data generated by generator G or the real samples. Train D to determine the source of the data as accurately as possible, and train G to make the data generated conform to the distribution of the real sample as much as possible. Through this way, it will reach the Nash equilibrium state. The structure and workflow of the GAN model are depicted in Fig. 3.

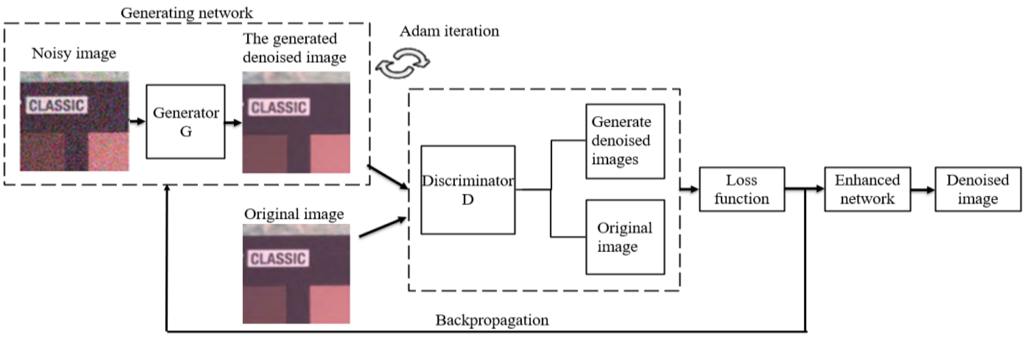

Fig 3. GAN model.

The generator G takes a corrupted image as input and generates a cleaner version of it. The discriminator D receives both the clean reference image and the output from G, producing a value between [0, 1] that indicates how closely the generated image resembles the clean one [10].

The created adversarial network image may successfully restore the image's edge and detail information by effectively removing the actual noise. The processed image has good visual quality and can make the image details more visible.

2.3. Transformer

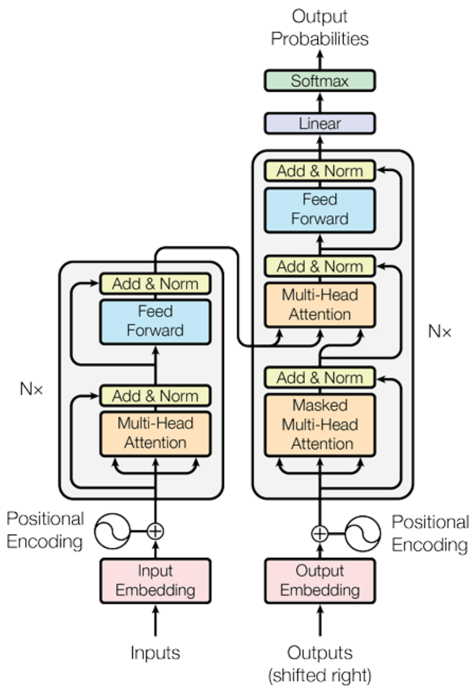

Fig 4. Transformer model.

Fig. 4 illustrates the structure of the transformer model, a network architecture built on an attention mechanism. It primarily consists of two components: the Encoder and the Decoder [11, 12].

Each of the encoder's several layers is mostly made up of a feedforward neural network and an inattention mechanism. The self-attention mechanism captures the global relationship of the input sequence, ensuring that the information at each position can fully reflect the entire sequence content. Then, the information is processed using a feedforward neural network.

The decoder is responsible for converting the encoder's information into a target sequence consisting of several layers, each with three core mechanisms: self-attention, encoder-decoder attention mechanism, and feedforward neural network. The self-attention mechanism ensures that the decoding depends on the output of the previous step to prevent information leakage. The encoder-decoder attention mechanism captures the relationship between the encoder output and the current decoder output to ensure that the correct symbols are generated according to the context. The feedforward neural network processes the information.

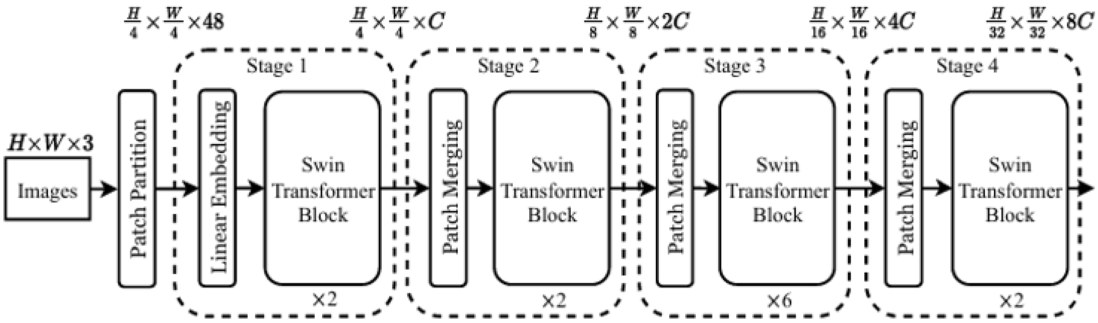

Fig 5. Swin Transformer model.

A new transformer, the Swin transformer depicted in Fig. 5, is proposed and applied in the field of image denoising [13]. In this transformer model, the input image, with size $H \times M \times 3$ is initially divided into 4 × 4 patches during the patch partitioning step. Each patch has a feature dimension of 48, transforming the image from $H \times M \times 3$ to $\frac{H}{4} \times \frac{M}{4} \times 48$. The process then proceeds through four stages. In the first stage, the image undergoes linear embedding, where its feature dimensions are projected onto a new dimension, followed by a Swin Transformer block. Stage 2-4 is composed of a patch merging and San swing transformer block, patch merging results in different feature dimensions. In stage2-4, the feature dimension is$2C, 4C, 8C$ respectively, and the output resolution is $\frac{H}{8} \times \frac{M}{8}$, $\frac{H}{16} \times \frac{M}{16}$, $\frac{H}{32} \times \frac{M}{32}$. In this process, the resolution of the feature dimension will decrease in different stages, so that the window self-attention learning content will be different, and the learning features will be various.

Swin Transformer improves some problems existing in traditional transformer models [14]. Swin Transformer handles large-scale variations between images and sequences using a shift window mechanism, which helps decrease both computational cost and parameter count. Such an algorithm enables the model to efficiently handle pixel-level tasks. Swin Transformer's layered design and local windowing-based attention mechanism enable it to capture fine-grained spatial details, effectively solving intensive prediction tasks.

3. Application

3.1. CT image

Deep learning-based image-denoising technology is widely used in the medical field, especially in CT imaging. They use CT imaging to observe the structure of organs in the human body, allowing them to evaluate any lesions present. The quality of these images has a considerable impact on the accuracy of clinical diagnoses. One of the main challenges in CT imaging is to get the original image from a noisy environment and obtain high-quality medical images [15]. There is a "photon starvation" effect, which causes the projected data to be polluted with noise [16]. The CT image reconstructed from the projected data not only has obvious noise but also produces fringe artifacts. The radiation dose affects the noise intensity in the CT image, and when the radiation dose is reduced, the CT image will inevitably be polluted by noise. Therefore, CT has a high demand for high-quality image-denoising while preserving more detail.

In the traditional denoising methods, people usually use the spatial domain denoising and transformation domain denoising methods. In the spatial domain denoising, the spatial filter is used to rely on the low-pass filter for denoising. Noise usually occupies the high-frequency spectrum area, and the spatial filter can remove the noise within a reasonable range. However, when the noise is large, the noise removal may be incomplete and faces a series of issues such as blurred image edges and loss of image details. This cannot meet the needs of CT image clarity in medical diagnosis.

On this basis, the transformation domain method is used to denoise the image. Compared with the spatial domain, the transformation domain has better performance, especially the BM3D [17]. BM3D has undergone improvements and extensions by researchers and has been successfully applied to medical image denoising tasks, leading to a significant enhancement in image quality. However, while the image quality is improved, the denoised image is too smooth and there is a loss of detail.

The convolutional neural network can feature learning and mapping, which can effectively process the denoising work of CT images. The denoising algorithm using convolutional neural networks can preserve the image structure information, especially the texture structure, ensuring that the local details of the image remain more distinct. A new and state-of-the-art single-level CT image denoising framework is proposed, which combines the CNN denoising model with other denoising methods [18]. A CNN model was used to denoise the noisy CT picture that resulted from applying additive Gaussian white noise. After using this framework for denoising, the edge information and other details of the CT image can be effectively retained, and no artefacts will be generated during the whole noise reduction process, which greatly enhances the quality of the image.

3.2. Synthetic Aperture Radar Image

Deep learning-based image denoising is also widely used in radar imaging. Synthetic aperture radar can provide images that are unaffected by sunlight, cloud cover, and weather conditions [19]. It can monitor the Earth's surface in a reliable, continuous, and global manner. This is a high-resolution imaging radar in which target objects in each small area are reflected by receiving reflections generated by many scattering centres. Each scattering centre forms an independent sub-echo signal with varying phases and amplitudes. When these signals converge, they make up the total reflected signal in that area. Random biases arise because of the radar signal's attenuation in the SAR imaging system. A typical particle interference in SAR images is speckle noise, which can blur features and lower image quality [20]. This will have a significant impact on later tasks like image categorization and target recognition.

To tackle noise in images and enhance their quality, the denoising algorithm in the spatial domain is first applied. The algorithm that uses spatial domain denoising has certain limitations, primarily due to the size of the selection window. As a result, the denoised image can be adversely affected by block effects and artificial textures, leading to blurred edges and loss of detail. This means that the final image may not accurately represent the real environment [21]. Although transformation-domain algorithms can retain image details and avoid over-smoothing, they require image decomposition and reconstruction, leading to high computational demands and potential artefact introduction.

Since then, the transformation domain and the spatial domain have been combined, and a remote sensing image denoising method based on a dual-domain combination has been proposed [22]. The combination of the denoising algorithm of the spatial domain and the transformation domain can preserve the image details while removing most of the noise. Although the dual-domain combination algorithm can achieve effective noise removal, the parameter count and computational demands during the denoising process increase significantly, leading to a considerable drop in efficiency when processing large-scale remote sensing images.

Remote sensing images contain a lot of detailed information, including geological structure and texture features. The introduction of a deep learning denoising algorithm into SAR image denoising allows the model to adapt to various data distributions and noise types through learning, providing strong generalization capabilities and achieving superior denoising results [22]. The denoising model using deep learning can effectively retain the detailed information with a frequency band close to the noise and improve the image quality. At the same time, deep learning-based image denoising methods can reduce the complexity of model calculation, improve the efficiency of the algorithm, and solve the processing problem of large-scale remote sensing images.

4. Conclusion

As the application of images in daily life increases, the requirement for denoising quality is also increasing. This paper provides an overview of recent advancements in image-denoising research. It reviews various deep learning-based denoising techniques. The paper also discusses and analyzes the advantages of each approach and the denoising model of deep learning applied in practice.

Deep learning has greatly contributed to the progress of image-denoising techniques. Convolutional neural networks are extensively utilized due to their remarkable efficiency. By effectively extracting local features, they offer both a theoretical foundation and technical support for advancing image-denoising methods. Among them, the DNCNN model can handle various noise scenes. The generative adversarial network has a good ability to remove complex noise by using the characteristics of the generator and discriminator adversarial training. Swin transformer, which is based on the transformer model, can reduce the computational complexity so that the model can handle pixel-level tasks. These models can promote the development of image-denoising algorithms and improve the performance of denoising.

This paper explores how deep learning algorithms are applied to image denoising, particularly in medical CT imaging and SAR imaging. In the processing of CT images, deep learning algorithms can improve the quality of CT image denoising, while retaining the details in the image, helping doctors to make better diagnoses. In SAR image processing, deep learning algorithms effectively address speckle noise, significantly enhancing image quality. The denoised images can then be used effectively for subsequent tasks, such as target detection and classification.

The introduction of the deep learning model promotes the progress of image denoising algorithm and shows excellent performance in training and application. However, with the increase of noise complexity in practical application scenarios and the demand for efficient processing models, deep learning denoising algorithms still face many challenges.

At present, most of the models are trained with simulated noise, but there is still a gap between the simulated noise and the actual real noise. The real noise is affected by many external factors and is more complex and diverse, so the existing model may have great defects in practical application even if it has a good performance in training. At the same time, deep learning training usually requires clean images and images containing noise to be trained at the same time, but in some scenes, it is difficult to obtain many high-quality paired data for training. The lack of training data also leads to the degradation of the performance of deep learning denoising algorithms. Even if the existing model can reduce the computational complexity, it still faces the problem of high computational cost when processing high-resolution images or real-time image processing, which makes it difficult to meet the actual needs.

To continue promoting the development of deep learning in the field of image denoising, it can be developed in the following ways in the future.

Data sets of real noise can be built to cover practical application scenarios of various noise types. This will facilitate model training, allowing the model to be more effective in denoising real-world noise. At the same time, to meet the needs of real-time processing and high-resolution image denoising, a more efficient deep learning algorithm model can be developed to further reduce the computational complexity and storage cost.

References

[1] Verma R, Ali J. A comparative study of various types of image noise and efficient noise removal techniques. International Journal of Advanced Research in Computer Science and Software Engineering, 2013, 3(10).

[2] Fan L, Zhang F, Fan H, et al. A brief review of image denoising techniques. Visual Computing for Industry, Biomedicine, and Art, 2019, 2(1): 7.

[3] Zhang Nana, Zhang Yuanyuan, Ding Weiqi. Review of Classical Image Denoising Methods. Control and Instruments in Chemical Industry, 2021, 48(5): 409-412+423.

[4] Xin Jiang. Research and hardware implementation of image denoising algorithm based on convolutional neural network. Xi'an University of Technology, 2024.

[5] Zhang Xiaoqiao. Research on pansharpening method for remote sensing images based on graph convolutional neural network. Xi'an University of Technology, 2024.

[6] Zhang K, Zuo W, Chen Y, et al. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Transactions on Image Processing, 2017, 26(7): 3142-3155.

[7] Ren Xingyu. Improvement of DnCNN image denoising network based on deep learning. Nanjing University, 2020.

[8] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. Advances in Neural Information Processing Systems, 2014, 27.

[9] Wang Meiqin, Yuan Weiwei, Zhang Jiye. Overview of research on generative adversarial network GAN. Computer Engineering and Design, 2021, 42(12): 3389-3395.

[10] Jiang Julang, Yan Huafeng, Liu Guoming. Image real noise removal algorithm based on generative adversarial network. Journal of Anqing Normal University (Natural Science Edition), 2023, 29(4): 74-78.

[11] Vaswani A. Attention is all you need. Advances in Neural Information Processing Systems, 2017.

[12] Chen Jiajun, Liu Bo, Lin Weiwei, Zheng Jianwen, Xie Jiachen. Survey of Transformer-based time series forecasting methods. Computer Science, 1-17 [2024-11-22].

[13] Liu Z, Lin Y, and Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.

[14] Fan CM, Liu TJ, Liu KH. SUNet: Swin transformer UNet for image denoising. Proceedings of the 2022 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2022: 2333-2337.

[15] Cheng Wen. Medical image denoising algorithm via deep learning. Linyi University, 2023.

[16] Zhang Yungang. Research on denoising and super-resolution algorithms for medical CT images based on convolutional neural networks. Wuhan University, 2019.

[17] Zhao T, Hoffman J, McNitt-Gray M, et al. Ultra-low-dose CT image denoising using modified BM3D scheme tailored to data statistics. Medical Physics, 2019, 46(1): 190-198.

[18] Singh P, Diwakar M, Gupta R, et al. A method noise-based convolutional neural network technique for CT image denoising. Electronics, 2022, 11(21): 3535.

[19] Moreira A, Prats-Iraola P, Younis M, et al. A tutorial on synthetic aperture radar. IEEE Geoscience and Remote Sensing Magazine, 2013, 1(1): 6-43.

[20] Zhao X, Ren F, Sun H, et al. Synthetic aperture radar image despeckling based on a deep learning network employing frequency domain decomposition. Electronics, 2024, 13(3): 490.

[21] Tian Shikang. Research on SAR image denoising algorithm based on transformer and convolutional neural network. Hebei University, 2024.

[22] Wang Haoyu, Yang Haitao, Wang Jinyu, Zhou Xixuan, Zhang Honggang, Xu Yifan. Review of image denoising methods for remote sensing. Computer Engineering and Applications, 2024, 60(15): 55-65.