

Research on Image Denoising Technology Based on Deep Learning

Shenxiang Zhu^{1*}

¹ School of Information and Intelligence Engineering, University of Sanya, Sanya, China

Email Address

1505408568@qq.com (Shenxiang Zhu)

*Correspondence: 1505408568@qq.com

Received: 21 February 2025; **Accepted:** 28 March 2025; **Published:** 10 April 2025

Abstract:

Image denoising is a technique that utilizes the contextual information of image sequences to remove noise, thereby restoring clear images. It is one of the important research topics in the field of computer vision. With the development of artificial intelligence technology, deep learning has been widely applied in the field of image denoising, becoming an effective solution for processing image noise. This paper analyzes deep learning methods for image denoising, detailing the concepts behind these methods based on network structures, and summarizes their advantages and disadvantages. Finally, it concludes with a summary of key issues in image denoising research and discusses future research trends in this field.

Keywords:

Image Denoising, Real Noise, Synthetic Noise, Deep Learning

1. Introduction

Research on image denoising is an important component of the field of computer vision. In recent years, deep learning-based denoising methods have been successfully applied to synthetic noise, but their generalization performance on real noise is relatively poor [1]. Real noise refers to the noise present in images captured by photographic devices under conditions such as poor lighting, camera shake, object motion, spatial pixel misalignment, and color brightness mismatch. This type of noise is characterized by unknown noise levels, diverse noise types, complex noise distributions, and difficulty in parameterization. In contrast, synthetic noise refers to noise types that conform to a certain probability distribution, with noise levels that can be set independently, such as Gaussian noise, salt-and-pepper noise, and speckle noise [2]. Currently, image denoising methods are widely used in various fields, including remote sensing image processing, medical image analysis, face and fingerprint recognition, among others [3].

Image denoising methods can be primarily divided into two categories: traditional denoising methods based on handcrafted features and denoising methods based on deep learning. Traditional image denoising methods based on handcrafted features modify transform coefficients using techniques such as discrete cosine transform or wavelet transform, and then compute local similarity using the average neighborhood

value. The NLM method and BM3D method achieve outstanding results in terms of image fidelity and visual quality by utilizing self-similar patches [4,5]. Based on these two methods, many variant methods such as SADCT, SAPCA, and NLB have been proposed, which seek self-similar patches in different transform domains to recover more edge information [6,7,8]. However, traditional denoising methods rely on assumptions about the original image for encoding image features, which leads to a lower degree of matching in real images, reducing the performance and flexibility of the methods in practical applications. Additionally, the feature extraction process of these methods is cumbersome, time-consuming, and computationally intensive, making them unsuitable for handling real noise with complex distributions.

This paper summarizes and categorizes current deep learning-based image denoising methods. It first provides an overview of the image denoising process, categorizing it based on foundational deep learning networks. Then, it systematically and comprehensively reviews and introduces recent image denoising methods, comparing and summarizing the advantages, limitations, and applicable scenarios of different methods. Finally, it discusses the main issues currently faced by image denoising methods and the directions for future research.

2. Image Denoising Methods Based on Deep Learning

Due to the limitations of electronic devices themselves and the influence of external environments, images often contain various types of noise, such as Gaussian noise, Poisson noise, salt-and-pepper noise, and speckle noise, each with different frequency and spatial characteristics. Traditional image denoising methods typically rely on prior knowledge and manually designed filters, such as median filtering and Gaussian filtering. Although these methods can reduce noise to some extent, they often have limited effectiveness when dealing with complex scenes and details. In recent years, the rise of deep learning has provided new ideas and technical means for image denoising, and commonly used deep learning-based image denoising techniques can be categorized as follows.

2.1. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are one of the most widely used architectures in deep learning. They significantly enhance the efficiency of feature extraction through local connections, weight sharing, and pooling operations. Many researchers have employed CNNs to build image denoising models to capture local features and spatial relationships in images. For example, DnCNN (Denoising Convolutional Neural Network) is a CNN-based denoising network that utilizes a multi-layer convolutional structure to effectively remove Gaussian noise. DnCNN employs residual learning to compare the original image with the noisy image, ultimately outputting a clear image. This model has demonstrated good denoising performance across multiple datasets. A notable advantage of CNNs is their hierarchical processing capability, which allows the model to adaptively capture multi-scale information, providing greater flexibility when handling noise. As the depth of the model increases, CNNs can learn high-level, abstract features, which is particularly important for denoising in complex scenarios.

For instance, denoising techniques N2N explore the mapping relationship between target pixels and input pixels through self-supervised learning by leveraging the independence of pixels to achieve denoising objectives. Figure 1 shows a comparison of the network structure of traditional self-supervised training models (left) and the

N2V method (right). However, self-supervised learning methods often overlook the dependencies between spatial information, leading to insufficient capability in expressing noise with the extracted features. Additionally, the adjustment of network training parameters lacks flexibility, making it difficult to effectively capture the complex mapping relationship between noisy images and clear images.

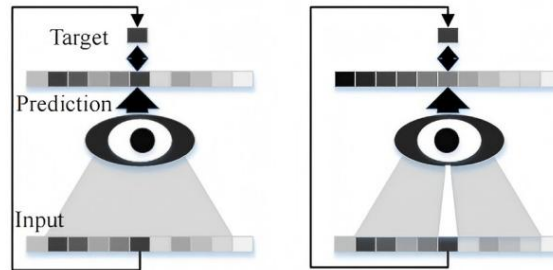


Figure 1. Self-supervised network structure comparison diagram.

Supervised learning-based image denoising techniques, including DnCNN-B, FFDNet, and CBDNet, employ Gaussian mixture models to train on sample images exhibiting various levels of noise, thereby validating the effectiveness of these methods on real noisy images. A prevalent challenge associated with this approach is its limited adaptability to noisy images that fall outside the training domain. The Mixture of Experts (MoE) framework offers a promising resolution to this issue by training a specialized expert for each representative domain within the dataset, evaluating and classifying the usefulness of the data, and minimizing the MoE loss through unsupervised clustering techniques [9].

CNN-based image denoising methods primarily focus on the extraction of feature information through convolutional layers and the optimization of network structures. However, they also exhibit certain limitations, mainly reflected in two aspects: on one hand, a single convolutional neural network lacks memory functionality, resulting in a substantial loss of shallow pixel-level information during the pooling process, which causes residual noise to remain in the denoised image. On the other hand, increasing the number of convolutional layers to enhance feature representation capabilities leads to an increase in the number of parameters, greater computational demand, and longer time consumption, complicating the network and increasing the difficulty of training.

2.2. Residual Networks

The core idea of Residual Networks (ResNet) is to optimize the network training process by introducing residual learning. This approach allows the network to effectively learn the residuals between the input and the target, rather than learning the target directly, significantly reducing the learning difficulty. In denoising tasks, the input is a noisy image, and the target is a clear image. Residual networks facilitate more efficient information flow within the network through shortcut connections between layers, promoting feature extraction and reconstruction. Denoising methods based on residual networks typically include three parts: preprocessing, network structure design, and post-processing. First, preprocessing the input image can reduce the initial impact of noise. Next, designing an appropriate residual network structure is crucial. Researchers continuously explore deeper network layers and more effective activation functions to enhance denoising performance. Finally, the post-processing step can further improve the visual quality of the image, such as applying image

enhancement techniques to capture finer details. The denoising concept can be described by the following formula:

$$f_0 = M_e(x) \quad (1)$$

$$f_r = M_{fl}(f_0) \quad (2)$$

$$\hat{y} = M_r(f_r) \quad (3)$$

Although denoising methods designed based on the residual network concept perform well, frequent use of shallow information can lead to overfitting in the network. To address this issue, Cha et al. utilized data augmentation and L2-SP (starting point) regularization to adaptively fine-tune network parameters, thereby minimizing the risk of overfitting [10].

Contemporary image denoising methodologies frequently utilize standard RGB images as training samples, neglecting the influence of the sensor device on image quality during the image signal processing (ISP) phase. This oversight results in a persistent discrepancy between RGB images and the original RAW images captured by imaging devices, leading to flawed and imprecise noise modeling techniques. To mitigate this issue, Zamir et al. introduced a cyclic mirror transformation approach termed CycleISP, which comprises two branches: RGB2RAW and RAW2RGB, as depicted in Figure 2 [11].

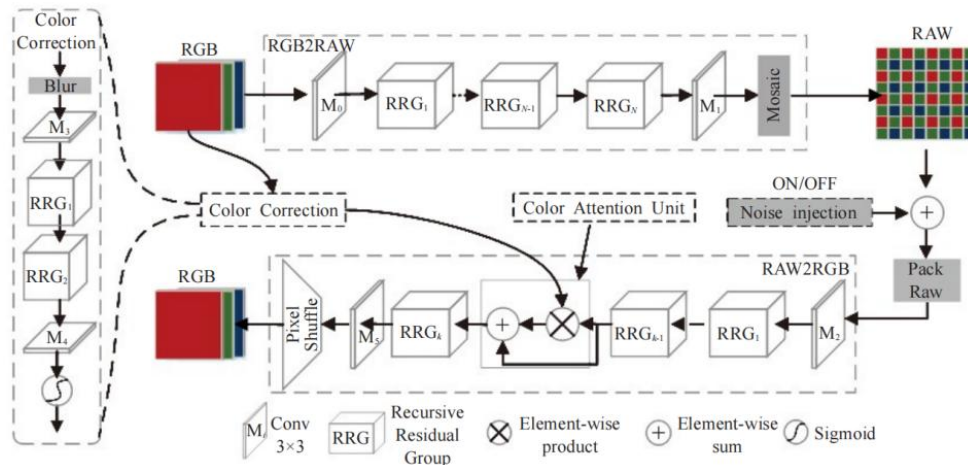


Figure 2. CycleISP network structure diagram.

2.3. Generative Adversarial Networks

Many denoising methods that utilize Convolutional Neural Networks (CNN) and Residual Networks (ResNet) often depend on having paired training samples, consisting of both noise-free, clear images and the corresponding noisy images, to perform supervised learning effectively. This requirement for paired samples creates a significant challenge in real-world applications where such perfectly matched pairs might not be readily available. Consequently, researchers and practitioners face difficulties in obtaining the necessary data for training these models. On the other hand, Generative Adversarial Networks (GANs) are known for their powerful learning capabilities. They can effectively generate realistic noisy images using sophisticated adversarial learning strategies. This unique characteristic of GANs helps to mitigate part of the problem concerning the shortage of paired training samples. In this way, GANs provide a promising alternative by enabling the generation of diverse

training data without the need for explicitly paired noise-free and noisy image samples.

Chen et al. developed a model that extracts noise information from real noisy images utilizing Generative Adversarial Networks (GANs), subsequently synthesizing new noisy images by integrating randomly generated noise blocks from the generator with the original clear images [12]. The SRGAN and ADGAN methodologies adopt a similar approach for constructing paired training data. In contrast to GCBD, the ADGAN approach incorporates a feature pyramid attention network to enhance the network's capacity for feature extraction during noise modeling. Addressing challenges akin to those examined by the CycleISP methodology, Isola et al. proposed a framework to mitigate the effects of sensor damage on signals. They extracted feature information at five distinct frequencies from the raw sensor data to serve as input for the Denoiser Network. This strategy lessens the dependency on dense filters, thereby reducing computational burden and enhancing efficiency [13]. Furthermore, this approach improves the robustness and stability of the network by implementing a blend of supervised and unsupervised training techniques.

GAN-based image denoising approaches leverage adversarial learning strategies to model data distributions through the interaction between generators and discriminators. However, these methods exhibit notable limitations in two primary areas: first, the generative model's distribution is not explicitly articulated, leading to challenges in interpretability. Second, the necessity for synchronization in parameter updates between the generator and discriminator complicates the generation of discrete data, and thus far, no methodology has successfully attained Nash equilibrium, resulting in training instabilities within the network.

2.4. Graph Neural Networks

Graph Neural Networks (GNN) demonstrate a significant aptitude for processing unstructured and complex datasets. Valsesia et al. developed an innovative denoising method known as GCDN, grounded in Graph Convolutional Networks (GCN), as shown in Figure 3. This approach mitigates the loss of edge information by implementing an Edge Attention mechanism and utilizes filters of sizes 3×3 , 5×5 , and 7×7 to facilitate multi-scale hierarchical feature extraction, thereby enhancing the effective use of feature information [14].

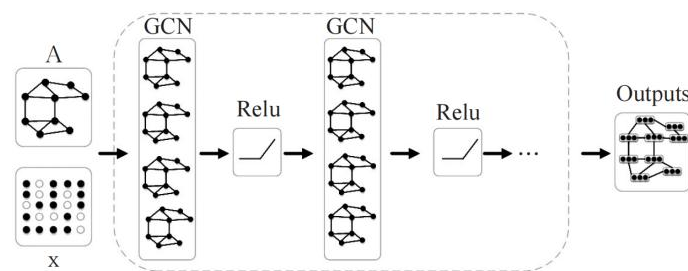


Figure 3. GCN network structure diagram.

The network topology of Graph Convolutional Networks (GCNs) is particularly well-suited to handle dense noise distributions that may arise in various applications. However, the complexities inherent in the architecture of GCNs, coupled with the dynamic fluctuations in how nodes are ordered and processed, can lead to inconsistent and unstable training outcomes. This variability in training results can be particularly challenging for practitioners seeking reliable performance from these models. Additionally, GCNs exhibit a tendency to fall into the trap of overfitting, especially

when they are applied to datasets comprised of homogeneous data types. This overfitting phenomenon can severely limit the generalizability of the model to new, unseen data instances. In an effort to tackle the persistent issue of overfitting within graph networks, researchers such as Behjati et al. implemented a novel approach in which they dynamically constructed neighborhood graphs. These graphs incorporated intricate feature hierarchies that leverage the principle of similarity invariance among high-dimensional features. This strategy aims to effectively attenuate the adverse effects of noise distribution, thereby enhancing the robustness of GCNs. The approach demonstrated significant resilience against the structured noise that is often produced by processes such as LiDAR scanning, showing promise for improving the reliability and accuracy of models operating in noisy environments [15].

Table 1 provides a detailed summary and overview of deep learning-based image denoising methods from multiple perspectives, including advantages, limitations, applicable noise types, and suitable scenarios.

Table 1. Comparison summary of image denoising methods based on deep learning.

Networks	Advantages	Disadvantages
CNN	Self-supervised Learning Denoising: (1) Does not require paired training samples, overcoming the issue of insufficient real image paired training samples; (2) The model structure is relatively simple, and the training process is relatively stable; (3) Fewer parameters, resulting in lower memory usage. Supervised Learning Denoising: (1) Can autonomously and quickly extract features, and the improved CNN denoising methods have relatively good feature representation capabilities; (2) The network structure is easy to expand and optimize.	Self-supervised Learning Denoising: (1) Feature extraction is easily influenced by surrounding pixels, leading to a significant deviation between predicted and actual values; (2) Insufficient feature representation capability for complex data distributions. Supervised Learning Denoising: (1) Low utilization of shallow pixel-level information, making it easy to lose texture details; (2) Insufficient feature representation capability, making it difficult to remove dense noise; (3) The balance between denoising and detail preservation is difficult to control, leading to issues such as noise residue or excessive smoothing.
ResNet	(1) Effectively addresses the problems of vanishing and exploding gradients, accelerating convergence speed; (2) Maximizes information flow, enabling the computation of long-range correlations and improving the utilization of low-frequency and high-frequency information.	(1) Designing network structures and optimizing objective functions is becoming increasingly difficult; (2) The use of dense connections can lead to overfitting, affecting the consistency between objective evaluation metrics and subjective visual effects.
GAN	It can generate realistic noise images, augmenting the real image dataset and addressing the issue of insufficient training samples.	(1) Issues such as unstable network training, slow convergence speed, and uncontrollable models are existed; (2) The distribution of the generative model is not explicitly expressed, resulting in poor interpretability.
GNN	(1) Suitable for processing unstructured data; (2) The topological structure of graph networks can effectively fit complex noise distributions.	(1) The size of the graph is uncontrollable, and the topological structure is complex, lacking a fixed node order as a reference; (2) Unstable dynamic topological structures can reduce the expressive power of features.

3. Conclusion and Prospect

With the ongoing advancement of image denoising technologies, an increasing number of methods are dedicated to effectively eliminating realistic noise. Successfully executing the preprocessing phase of image denoising establishes a robust foundation for subsequent research in related image processing domains. This paper provides a comprehensive summary and classification of image denoising techniques developed in recent years. Traditional image denoising approaches, while established, often fail to adequately account for the diversity of noise types and typically rely on prior knowledge, which confines their efficacy to local feature extraction and overlooks the significance of global information. Furthermore, these models tend to exhibit limited generalizability, resulting in suboptimal performance when applied to real-world noisy images. In contrast, deep learning-based image denoising methods possess powerful representation learning capabilities; models trained on a large number of samples can capture more accurate noise characteristics. Recent developments in image denoising methods, particularly GAN-based and GNN-based approaches, have shown rapid progress, demonstrating significant advantages in constructing training samples and handling unstructured noise distributions, respectively.

While deep learning-based denoising techniques have achieved remarkable advancements and the effectiveness of these methods continues to improve steadily, there are still several key areas that warrant further exploration and development. First and foremost, one of the most pressing challenges is the application of denoising methods to real-time video processing. Recent studies in the field of image denoising have predominantly concentrated on static images, leaving a noticeable gap in research that specifically addresses the complexities of video denoising. A critical research direction for the future lies in figuring out how to effectively leverage the temporal relationships that exist between frames in a video sequence. This capability is essential for achieving real-time denoising, which holds significant practical value across various applications, such as live streaming, video conferencing, and gaming. The need for efficient and effective real-time solutions is becoming increasingly important as the demand for high-quality video content continues to rise.

Additionally, it is important to recognize that denoising a single image poses a greater challenge compared to the denoising of video sequences. This is primarily due to the fact that a single frame image is treated as an independent input, lacking the context provided by adjacent time sequences. As a result, the method is unable to learn from the differences in neighboring chrominance, which complicates the denoising process. This unique challenge associated with single image denoising makes it a particularly difficult problem to solve and highlights it as one of the significant research directions that should be prioritized in future studies. Addressing these challenges will be crucial for the continued advancement of denoising techniques and their applicability in real-world scenarios.

Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

References

- [1] Chen, R.H.; Lai, Z.Y.; Qian, Y.R. Improved image denoising method for generative adversarial networks. *Computer Engineering and Applications*, 2021, 57(5), 168-172.
- [2] Tan, T.; Qian, Y.; Hu, H. et al. Adaptive very deep convolutional residual network for noise robust speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2018, 26(8), 1393-1405.
- [3] Xu, J.; Li, H.; Liang, Z, et al. Real-world noisy image denoising: A new benchmark. arXiv preprint arXiv:1804.02603, 2018.
- [4] Izadi, S.; Sutton, D.; Hamarneh, G. Image denoising in the deep learning era. *Artificial Intelligence Review*, 2023, 56(7), 5929-5974.
- [5] Elad, M.; Kawar, B.; Vaksman G. Image denoising: The deep learning revolution and beyond—a survey paper. *SIAM Journal on Imaging Sciences*, 2023, 16(3), 1594-1654.
- [6] Liu, D.; Wen, B.; Jiao, J, et al. Connecting image denoising and high-level vision tasks via deep learning. *IEEE Transactions on Image Processing*, 2020, 29, 3695-3706.
- [7] Jebur, R.S.; Zabil, M.H.B.M.; Hammood, D.A. et al. A comprehensive review of image denoising in deep learning. *Multimedia Tools and Applications*, 2024, 83(20), 58181-58199.
- [8] Hashimoto, F.; Onishi, Y.; Ote, K, et al. Deep learning-based PET image denoising and reconstruction: a review. *Radiological physics and technology*, 2024, 17(1), 24-46.
- [9] Mei, Y.; Fan, Y.; Zhang, Y. et al. Pyramid attention network for image restoration. *International Journal of Computer Vision*, 2023, 131(12), 3207-3225.
- [10] Kim, D.W.; Ryun, Chung.J.; Jung, S.W. Grdn: Grouped residual dense network for real image denoising and gan-based real-world noise modeling[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2019; pp. 0-0.
- [11] Zamir, S.W.; Arora, A.; Khan, S. et al. Cycleisp: Real image restoration via improved data synthesis[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020; pp. 2696-2705.
- [12] Chen, J.; Chen, J.; Chao, H. et al. Image blind denoising with generative adversarial network based noise modeling[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018; pp. 3155-3164.
- [13] Isola, P.; Zhu, J.Y.; Zhou, T. et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017; pp. 1125-1134.

- [14] Valsesia, D.; Fracastoro, G.; Magli, E. Deep graph-convolutional image denoising. *IEEE Transactions on Image Processing*, 2020, 29, 8226-8237.
- [15] Behjati, P.; Rodriguez P.; Mehri A. et al. Overnet: Lightweight multi-scale super-resolution with overscaling network[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021; pp. 2694-2703.



© 2025 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)