

Article

EDUNet++: An Enhanced Denoising Unet++ for Ice-Covered Transmission Line Images

Yu Zhang ^{1,2,3,*}, Yinke Dou ^{1,3,4,*} , Liangliang Zhao ³, Yangyang Jiao ³ and Dongliang Guo ³¹ College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030024, China² Department of Automation, Taiyuan Institute of Technology, Taiyuan 030008, China³ Shanxi Energy Internet Research Institute, Taiyuan 030032, China; zhaoliangliang_97@163.com (L.Z.); jiaoyyru@163.com (Y.J.); 987hans@163.com (D.G.)⁴ Key Laboratory of Cleaner Intelligent Control on Coal & Electricity, Ministry of Education, Taiyuan 030024, China

* Correspondence: zhang.yu.edu@163.com (Y.Z.); douyinke@tyut.edu.cn (Y.D.)

Abstract: New technology has made it possible to monitor and analyze the condition of ice-covered transmission lines based on images. However, the collected images are frequently accompanied by noise, which results in inaccurate monitoring. Therefore, this paper proposes an enhanced denoising Unet++ for ice-covered transmission line images (EDUNet++). This algorithm mainly comprises three modules: a feature encoding and decoding module (FEADM), a shared source feature fusion module (SSFFM), and an error correction module (ECM). In the FEADM, a residual attention module (RAM) and a multilevel feature attention module (MFAM) are proposed. The RAM incorporates the cascaded residual structure and hybrid attention mechanism, that effectively preserve the mapping of feature information. The MFAM uses dilated convolution to obtain features at different levels, and then uses feature attention for weighting. This module effectively combines local and global features, which can better capture the details and texture information in the image. In the SSFFM, the source features are fused to preserve low-frequency information like texture and edges in the image, hence enhancing the realism and clarity of the image. The ECM utilizes the discrepancy between the generated image and the original image to effectively capture all the potential information in the image, hence enhancing the realism of the generated image. We employ a novel piecewise joint loss. On the dataset of ice-covered transmission lines, PSNR (peak signal to noise ratio) and SSIM (structural similarity) achieved values of 29.765 dB and 0.968, respectively. Additionally, the visual effects exhibited more distinct detailed features. The proposed method exhibits superior noise suppression capabilities and robustness compared to alternative approaches.

Keywords: ice-covered transmission lines; image denoising; multilevel feature extraction; source feature fusion; error attention



Citation: Zhang, Y.; Dou, Y.; Zhao, L.; Jiao, Y.; Guo, D. EDUNet++: An Enhanced Denoising Unet++ for Ice-Covered Transmission Line Images. *Electronics* **2024**, *13*, 2085. <https://doi.org/10.3390/electronics13112085>

Academic Editors: Phivos Mylonas, Katia Lida Kermanidis and Manolis Maragoudakis

Received: 25 April 2024

Revised: 25 May 2024

Accepted: 26 May 2024

Published: 27 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent years, with the continuous development of the power system, the overall length of State Grid transmission lines in China had reached 1.142 million kilometers by the end of 2020. During the winter season in northern China, the transmission lines are susceptible to ice and snow coverage due to variations in air humidity, wind speed, temperature, and other factors, that poses safety and stability concerns to the cables [1,2]. Transmission line ice-coverage will not only lead to the decrease of line load capacity, but may also cause equipment failure, power outage and other problems that seriously affect the operation of the power grid [3–5]. Therefore, in order to accurately monitor the condition of power transmission lines, we designed a suspended intelligent vibration de-icing robot as shown in Figure 1. The robot is suspended on the cables between the towers, and the camera carried on the side of the robot takes an upward shot of the cables to detect the snow and ice covering the cables, and to analyze the ice-covered condition

of the cables using artificial intelligence. It then destroys the thin ice on the cable surface through autonomous judgment and vibration. Robots have the ability to independently perceive their environment, make decisions, and take actions without requiring external control or intervention.



Figure 1. Suspended intelligent vibration de-icing robot: (A) static display, (B) deicing process.

However, a crucial aspect for robots to effectively monitor and ensure the safe functioning of transmission lines is to collect high-quality images. In the actual work process, the images collected by robots usually contain noisy distortions caused by swaying in breeze, drifting snow, frost, etc. The presence of these noises hampers the robot's ability to accurately assess the cable icing status and make accurate actions, hence exacerbating the cable icing problem [6,7]. Therefore, denoising the images of ice-covered transmission lines is a key step to improve image quality and accurate monitoring. By employing denoising techniques, we can acquire a precise and unambiguous representation of the cable's actual condition. This information is helpful for the robot to assess the presence of ice coating on the cable and serves as a reliable and timely basis for the autonomous vibration system to make accurate judgments. Consequently, the robot can adjust its vibration amplitude accordingly to effectively remove the ice, thereby ensuring the safe operation of the transmission lines [8,9].

At present, there is a limited amount of research on image denoising of transmission lines, so we rely on the knowledge and expertise gained from the field of image processing to inform our approach. Traditional image denoising algorithms primarily denoise images by considering the statistical properties of individual pixels or local neighborhoods, such as median filtering [10], mean filtering [11], and NLM [12]. These methods are simple to implement and easy to understand, but they are more inclined to process local information of the image, may have certain limitations in retaining the global structure and texture of the image, and are difficult to adapt to complex image structures and changes. With the advancement of deep learning, researchers have increasingly utilized this technology for image denoising, such as FFDNet [13], CTNet [14], and DRSformer [15]. These methods typically employ a neural network model to learn the mapping relationship between noisy images and clean images by training a large number of image pairs. They can acquire knowledge about the intricate arrangement and qualities of the image, resulting in the successful elimination of noisy disturbances and the preservation of a greater amount of detailed data. However, in order to achieve the optimal denoising effect for some forms of noise or complex image structures, it is necessary to employ specific designs and make adjustments [16,17].

Therefore, this paper aims to investigate and refine current image denoising algorithms so as to be most effective for the unique image features of transmission lines. As a result, we propose an enhanced denoising Unet++ for ice-covered transmission line images (EDUNet++). This algorithm uses the Unet++ model to encode and decode images. For this stage, in order to enhance the feature extraction capabilities of the model, to better retain the structure and content of the image, and to adapt to noise at different scales, we propose a residual attention module (RAM). The RAM can effectively suppress invalid information in features and enhances the details of the image. In order to combine different levels of features and capture image contextual information, we propose a multilevel feature

attention module (MFAM). The MFAM effectively combines local and global features and focuses on different features at different levels, which can better capture details and texture information in images. In order to utilize the details, texture and other information in the low-level features, we propose a source feature fusion module (SSFFM). The SSFFM avoids excessive smoothness of the image caused by over-reliance on high-level features and enhances the realism and clarity of the image. In order to minimize the discrepancy between the generated image and the original image, we propose an error correction module (ECM). The ECM acts on generated images by calculating the error between images, thus generating more realistic images. During the model training process, in order to improve the model's recovery of the detailed information within the generated image, we used a segmented fusion loss function. Based on the above viewpoints, the main contributions of this paper are as follows:

- (1) The residual attention module (RAM) and the multilevel feature attention module (MFAM) are proposed in the feature encoding and decoding module (FEADM) to improve the feature extraction capabilities of the model, to effectively combine local and global features, and to suppress the influence of noise information.
- (2) The shared source feature fusion module (SSFFM) is proposed to enhance the model's utilization of source feature information, to understand image information at different levels, and to improve the denoising performance of the model.
- (3) The error correction module (ECM) is proposed to enhance the model's ability to fully capture the rich potential information in the image, and to help generate more realistic images by calculating the error.
- (4) The piecewise joint loss function is innovatively employed in this model. Initially, Mean Squared Error (MSE) loss is utilized to enhance the model's convergence speed during the early stage of training. Subsequently, a joint function consisting of Charbonnier loss and structural similarity index measure (SSIM) loss is employed in the later stage to enhance the model's ability to handle outliers effectively.
- (5) A new denoising model is proposed, called EDUNet++. A large number of experiments show that the EDUNet++ is superior to current denoising algorithms in both quantitative and qualitative aspects.

2. Related Work

Traditional image denoising methods mainly relied on prior knowledge. Dabov [18] proposed the block-matching and 3D filtering method (BM3D), which uses the self-similarity existing in natural images to match adjacent image blocks, then integrates similar blocks through domain transformation to form a denoised image. Li et al. [19] proposed an adaptive matching and tracking algorithm. First, the sparse coefficients are calculated; then the K singular value decomposition algorithm is used to train the dictionary into an adaptive dictionary that can effectively reflect the image structure; finally, the sparse coefficients are combined with an adaptive dictionary for image reconstruction. Qian et al. [20] used sparse coding to optimize the block matching results. This approach utilizes non-local information and introduces graph Laplacian regularization to maintain local information, but image details are severely lost in the presence of strong noise. The feature extraction process of traditional methods is complex, computationally intensive and time-consuming, and has limitations when dealing with complex noise [21].

In recent years, deep learning-based methods have achieved many results in the field of image denoising. Convolutional neural networks have two major characteristics: local perception and parameter sharing, and good results in image feature extraction and recognition [22]. Zhang et al. [23] proposed a denoising convolutional neural network (DnCNN) model that combines batch normalization and residual learning. The effect of denoising uniformly distributed Gaussian noise is good, but the convolution and pooling layers used result in a relatively small receptive field, resulting in limited feature extraction capabilities, and the inability to fully capture contextual information for larger image structures. Guo et al. [24] proposed the convolutional blind denoising network (CBDNet),

which uses a fully convolutional network to estimate the noise level to achieve an adaptive denoising effect. However, the fully convolutional network pays more attention to local information and therefore has a negative impact on global consistency. Ding et al. [25] employed a dilated convolution to construct a two-stage blind denoising network to obtain comprehensive features within sensing ranges of different sizes. However, the network structure is relatively simple, resulting in limited information capture. Huang et al. [26] proposed a channel affine self-attention-based progressively updated network (CasaPuNet), which adopts a multi-stage progressive update method, uses channel affine self-attention to extract channel information from input features, and adaptively fuses multi-stage features through skip connections and residual structures. Mou et al. [27] integrated local and non-local attention mechanisms into deep networks to restore images containing complex textures. Zamir et al. [28] proposed a multi-stage progressive image restoration network (MPRNet), which exchanges information at different stages to reduce the loss of detailed information. Potlapalli et al. [29] proposed a prompt-based learning approach (PromptIR), which generates intermediate representations based on input images through multiple decoding stages to reflect semantic features, and uses these intermediate representations as guidance to restore images in the decoding stage. However, this method requires more computing resources for the model during training and inference.

The introduction of the self-attention mechanism in transformer neural networks solves the problem of the limited receptive field of the convolution operator and the inability of the network to flexibly adapt to the input content [30]. In recent years, the performance of transformer-based modules in visual tasks has also been significantly improved. Wang et al. [31] proposed a general U-shaped transformer network (Uformer) which, to a certain extent, solves the limitations of traditional convolutional neural networks in denoising problems, but lacks clear spatial position information. Yao et al. [32] proposed a dense residual transformer network (DenSformer), that combines a transformer with the ideas of dense connection and residual connection to bridge and fuse features between different layers. Transformer-based methods apply a self-attention mechanism to capture the long-range dependence of image patches. However, they primarily focus on structure-level characteristics and overlook the enhancement of pixel-level features. This oversight can result in the presence of residual textures in the denoised images.

3. Network Structure

The network structure proposed in this article is shown in Figure 2 and has three primary modules: feature encoding and decoding module (FEADM), shared source feature fusion module (SSFFM), and error correction module (ECM). The input of the network is a non-uniform noise image, I_{LR} . First, a convolution layer is utilized to extract the initial feature, F_0 . Then, in order to restore as much image texture and edge information as possible and reasonably remove the noise corresponding to the image content, F_0 is input into the FEADM. After that, the fused features are encoded and decoded by the residual attention module (RAM) and multilevel feature attention module (MFAM). Finally, the reconstructed noise-free image is recorded as P_{HR} . The process is expressed as:

$$F_0 = \text{Conv}(I_{LR}) \quad (1)$$

$$F_1 = \text{MFAM}(\text{Concat}(\text{UP}(\text{RAM}(\text{Down}(\text{RAM}(F_0)) + F_0))), \text{RAM}(F_0)) \quad (2)$$

$$P_{HR} = \{\text{RAM}(\cdot), \text{MFAM}(\cdot), F_0\} \quad (3)$$

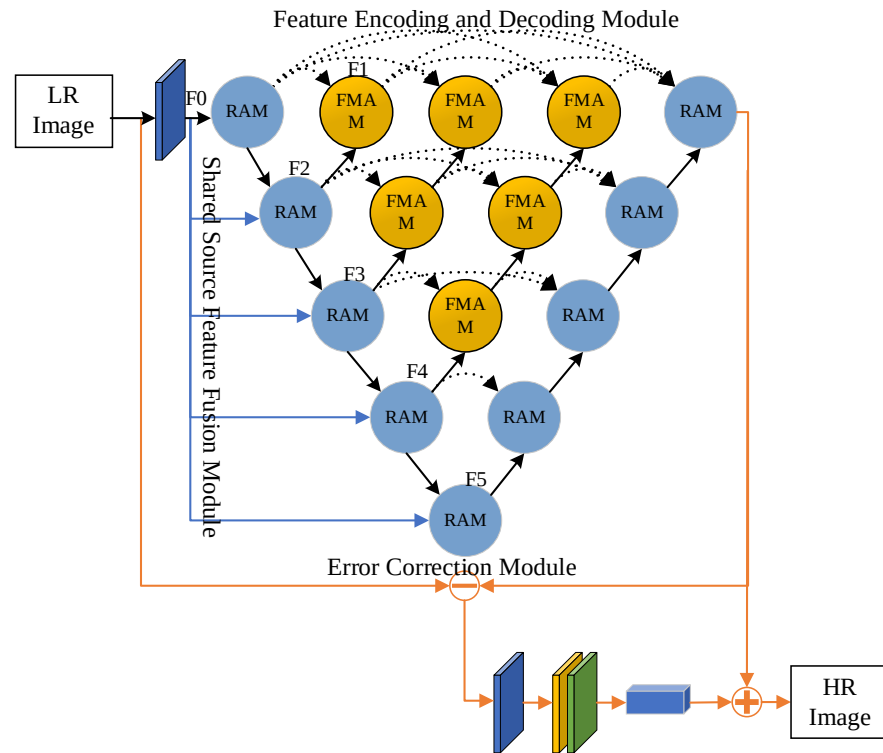


Figure 2. Network structure.

In the above formula, $Conv(\cdot)$ represents the convolution operation; $F1$ represents the intermediate feature in the feature encoding and decoding module; $Down(\cdot)$ represents the feature down-sampling operation; $UP(\cdot)$ represents the feature up-sampling operation; $Concat(\cdot)$ represents the feature splicing operation; $\{RAM(\cdot), MFAM(\cdot), F0\}$ represents the residual attention module (RAM), multilevel feature attention module (MFAM) and source feature $F0$ involved in the feature encoding and decoding module process; and P_{HR} represents the reconstruction of the noise-free image through the encoding and decoding module.

However, P_{HR} still has some distorted texture information in local areas, so after correction by the ECM, the noise-free image I_{HR} is generated. The process is expressed as:

$$I_{HR} = Error(I_{LR} - P_{HR}) \quad (4)$$

In this formula, $Error(\cdot)$ represents the ECM. The inputs are P_{HR} and I_{LR} . The error feature representation is generated by subtracting the two images, and then the features are represented by attention; finally, P_{HR} is used to correct the blurred texture information in the image.

3.1. Feature Encoding and Decoding Module

3.1.1. Residual Attention Module (RAM)

In order to enhance the model's ability to accurately capture crucial information in the image and preserve valuable details while removing noise, we propose the residual attention module (RAM) as shown in Figure 3.

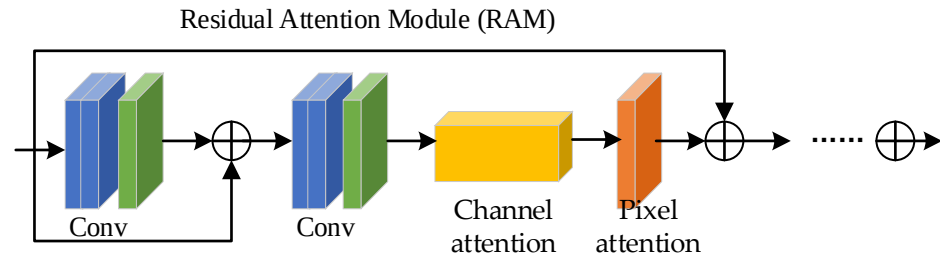


Figure 3. Residual attention module.

The RAM learns a mapping that preserves the original image information by utilizing skip connections and uses previously learned features to preserve the details of the image structure. This module enhances the network's training process and enables more efficient learning of the image features. This module consists of M residual blocks, and each residual block includes two convolution operations, two residual connections and a combined attention fusion module consisting of channel attention and pixel attention. Channel attention can make the network pay more attention to channels that are particularly important for image denoising tasks, helping to improve the model's perception of different features. Pixel attention allows the network to focus more intensively on pixels that are important to the image denoising task, helping the network to better capture details and structural information in the image. The combined attention fusion module utilizes channel-level and pixel-level information to enable the model to more comprehensively understand the structure and content of the image and better adapt to noise at different scales and frequencies. By integrating these two levels of information, the model can be more accurate and selectively amplify or diminish specific elements of an image in a more inclusive manner, hence improving overall denoising performance. The output F_N of the N residual block is:

$$F_N = PA(CA(Conv(Conv(F_{N-1}) + F_{N-1}))) + F_{N-1} \quad (5)$$

In Formula (5), F_{N-1} represents the input of the N residual block, $CA(\cdot)$ represents the channel attention operation, and $PA(\cdot)$ represents the pixel attention operation. First, F_{N-1} undergoes the convolution operation, then performs the residual addition operation with F_{N-1} to obtain the added features. Next, these added features pass through the convolution operation and the channel attention and pixel attention operations. Finally, the output feature F_N is obtained by performing the residual addition operation with F_{N-1} .

3.1.2. Multilevel Feature Attention Module (MFAM)

In order to improve the model's understanding of image content and structure, we propose the multilevel feature attention module (MFAM) for different scales and different feature extractions as shown in Figure 4.

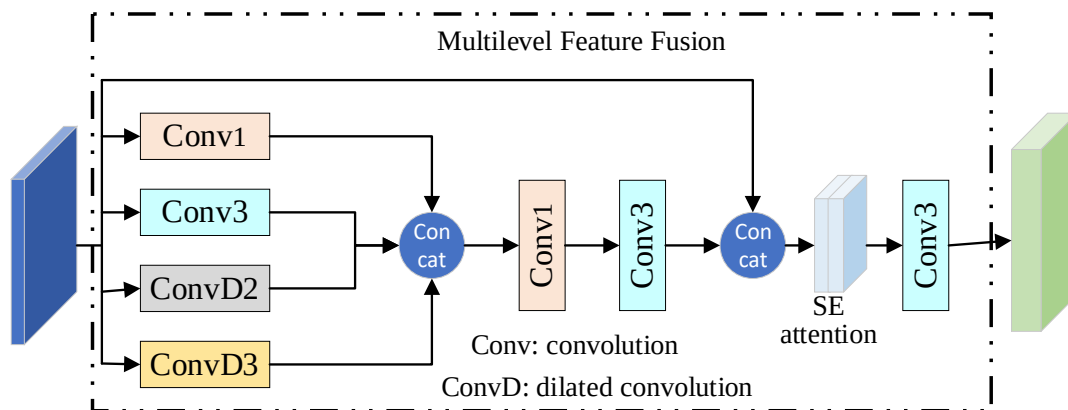


Figure 4. Multilevel feature attention module.

The MFAM uses different convolution kernels to extract and capture features of different scales and abstraction levels. The smaller convolution kernels are more suitable for capturing local features of the image, while larger convolution kernels can be used to obtain broader global features, thus enhancing the model's understanding of the image by fusing multi-level and multi-scale feature information. In order to reduce the number of parameters, we use dilated convolutions instead of traditional convolutions. Therefore, we use 3×3 convolutions with dilation rates of 2 and 3 in place of 5×5 and 7×7 , respectively. After getting diverse features, we combine them using a concatenate operation. After module feature fusion, we use the SE attention mechanism to calculate the weight of each channel, dynamically adjust the importance of features at each scale, and adaptively focus on the most important feature information for the denoising task, thereby improving the performance of the model. The MFAM helps enhance the generalization ability of the model, can better adapt to image features of different scales and structures, and is more robust for the processing of different samples and noise distributions. The output F_M of MFAM is:

$$F_a = C_1(\text{Concat}(C_1(F_i), C_3(F_i), C_{D2}(F_i), C_{D3}(F_i))) \quad (6)$$

$$F_M = C_3(\text{SE}(\text{Concat}(C_3(F_a), F_i))) \quad (7)$$

In the formula, F_i represents the input feature, C_1 represents the 1×1 convolution, C_3 represents the 3×3 convolution, C_{D2} represents the 3×3 dilated convolution with the dilation rate of 2, and C_{D3} represents the 3×3 dilated convolution with the dilation rate of 3. $\text{Concat}(\cdot)$ represents the feature splice operation, F_a represents the intermediate feature, $\text{SE}(\cdot)$ represents the SE attention mechanism, and F_M represents the input feature. First, F_i performs convolution operations through convolution kernels of different sizes to obtain features of different scales, and the different features are then spliced together. Next, the spliced features pass through 1×1 and 3×3 convolution operations to obtain F_a . Finally, F_a and F_i perform concat operation and SE attention to obtain the output feature, F_M .

3.2. Shared Source Feature Fusion Module (SSFFM)

Low-level features typically encompass the fundamental structures and intricate elements within images, and they play an important part in the process of image denoising. Therefore, the shared source feature fusion module (SSFFM) is proposed, which corresponds to the down-sampling part on the left side of UNet++ as depicted in Figure 2. The F_0 is combined prior to each down-sampling process primarily because it contains low-frequency information such as smaller details, textures, and edges in the image. Subtle texture plays a vital role in enhancing the authenticity and sharpness of the image, while also preventing distortion produced by excessive smoothness. Edge information is the transition between different areas in an image, so enhancing these edges helps improve the clarity of the image and makes objects and structures easier to identify. High-level features usually contain more abstract information, and over-reliance on it may cause the image to be over-smoothed and lose some of the details of the original image during denoising. Therefore, fusing source features can help retain more details, improve the overall visual quality of the image, and make the image look more natural and realistic. The module process is expressed as:

$$F_2 = \text{RAM}(\text{RAM}(F_0) + F_0) \quad (8)$$

$$F_3 = \text{RAM}(\text{RAM}(F_2) + F_0) \quad (9)$$

$$F_4 = \text{RAM}(\text{RAM}(F_3) + F_0) \quad (10)$$

$$F_5 = \text{RAM}(\text{RAM}(F_4) + F_0) \quad (11)$$

In the formula, F_0 represents the source feature, and F_2 , F_3 , F_4 , and F_5 represent intermediate features. The combination of source features and high-level features makes full use of various information in the image, that enables the model to understand and

process image information at different levels and improves the denoising performance of the model.

3.3. Error Correction Module (ECM)

The high-frequency information within the image generated by the encoding and decoding module will be partially lost. In order to fully capture the rich potential information contained in noisy images, we propose an error correction module (ECM) as shown in Figure 5, where I_{LR} is the noisy image, P_{HR} is the encoder-generated image and I_{HR} is the denoised image. This module transfers P_{HR} to I_{LR} and subsequently mitigates the various levels of mistakes by computing the errors of both pictures. It allocates greater importance to regions with significant errors and rectifies errors dispersed throughout different pixel areas. ECM can strike a balance between reducing noise and preserving image quality while denoising, preventing excessive destruction of image content and aiding in the creation of more lifelike images. The module process is expressed as:

$$F_e = \text{Conv}(P_{HR} - I_{LR}) \quad (12)$$

$$E_{\text{attention}} = \sigma(\text{Conv}(\delta(\text{Conv}(F_e)))) \quad (13)$$

$$F_{EA} = E_{\text{attention}} \otimes F_e \quad (14)$$

$$I_{HR} = F_{EA} + P_{HR} \quad (15)$$

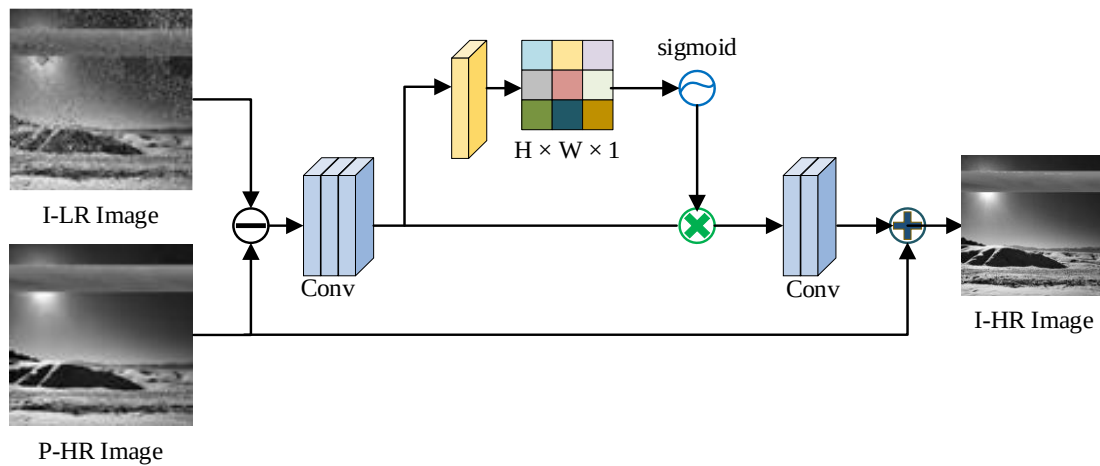


Figure 5. Error correction module.

In the formula, F_e represents the error feature of the P_{HR} and I_{LR} , δ represents the Relu activation function, σ represents the Sigmoid function, $E_{\text{attention}}$ represents the error attention, F_{EA} represents the error feature after the attention operation, and I_{HR} represents the final generated HD images. This module inputs P_{HR} and I_{LR} : first, it obtains the image error features through a subtraction operation, then transforms the dimensions from $C \times H \times W$ to $1 \times H \times W$ using a convolution operation. Next, the attention of the error feature is obtained through the Sigmoid operation, and is weighted and combined with the error feature maps of P_{HR} and I_{LR} to obtain the module's output, F_{EA} . Finally, the F_{EA} is added to the P_{HR} to obtain the final denoised image, I_{HR} .

3.4. Loss Function

When training a model for image denoising, the choice of loss function can have an impact on the quality and characteristics of the generated images. The MSE Loss is often used for image denoising tasks, due to its simplicity in calculation and ease of optimization. However, the model-generated denoised image may be over-smoothed and lose some detailed information, performing poorly on the high-frequency part of the noise.

Charbonnier Loss [33] is less sensitive to outliers than MSE Loss and pays more attention to detailed information, thus helping to retain the subtle structure and texture in the image. SSIM Loss [34] considers brightness, contrast and structural similarity. It is more in line with human visual perception, allowing it to help retain the structure and details of the image and to make the generated image closer to human eye perception. Therefore, we use MSE Loss to calculate the model's loss during the initial phase of training, enabling the model to be quickly optimized. Then, we use the joint loss operation that combines Charbonnier Loss and SSIM Loss to retain more detailed information of the image and improve the denoising performance of the model.

$$MSE\ Loss(I_{pred}, I_{gt}) = \frac{1}{N} \sum_{i=1}^N (I_{pred}(i) - I_{gt}(i))^2 \quad (16)$$

$$Charbonnier\ Loss(I_{pred}, I_{gt}) = \sqrt{\epsilon^2 + (I_{pred} - I_{gt})^2} \quad (17)$$

$$\begin{aligned} l(I_{pred}, I_{gt}) &= \frac{2 \cdot \mu_{pred} \cdot \mu_{gt} + c_1}{\mu_{pred}^2 + \mu_{gt}^2 + c_1} \\ c(I_{pred}, I_{gt}) &= \frac{2 \cdot \sigma_{pred} \cdot \sigma_{gt} + c_2}{\sigma_{pred}^2 + \sigma_{gt}^2 + c_2} \\ s(I_{pred}, I_{gt}) &= \frac{\sigma_{pred,gt} + c_3}{\sigma_{pred} \cdot \sigma_{gt} + c_3} \\ SSIM\ Loss(I_{pred}, I_{gt}) &= 1 - SSIM(I_{pred}, I_{gt}) = 1 - l \cdot c \cdot s \end{aligned} \quad (18)$$

In the formula, I_{pred} represents the image generated by the model, I_{gt} represents the target image, N represents the number of pixels, ϵ is a small positive number which is commonly used to avoid the denominator being zero, μ represents the mean, σ represents the variance, and c_1, c_2, c_3 are constants used for stable calculations.

4. Experiment

4.1. Datasets and Evaluation Metrics

Since there is no complete dataset of ice-covered transmission lines at present, we simulated the snow-covered environment of transmission lines in winter in a low-temperature laboratory, and then used the camera carried by the deicing robot to take pictures, and finally collected 280 training sets and 40 test sets. At the same time, in the winter snow environment, the deicing robot was used to take 50 images of the actual cable covered with snow and ice. This experiment employed Gaussian noise to replicate the real noise environment. We partitioned each image into 64 distinct regions and introduced Gaussian noise with varying standard deviations in each region to create a diverse noise distribution. This approach prevents the noise distribution from being too uniform, which could lead to reduced generalization capabilities.

We used SSIM and PSNR as experimental evaluation indicators. In addition, we used the Adam optimizer to train the proposed network. The initial learning rate was 0.003 and we then dynamically updated the learning rate according to the epoch. The max-epoch was 300. The proposed model was implemented with the Pytorch framework on RTX-4090 and the running environment was Ubuntu 20.04. 6.

4.2. Ablation Experiment

In order to verify the effectiveness and necessity of the design components in the proposed method, we used different components to perform experiments on the dataset, including RAM, FMAM, SSFFM, and ECM. The experiment was carried out with Gaussian noise having a standard deviation ranging from 20 to 50. The results are shown in Table 1.

Table 1. Ablation experiment results of different components.

Unet++	RAM	FMAM	SSFFM	ECM	PSNR (dB)	SSIM
✓					28.457	0.926
✓	✓				28.514	0.942
✓	✓	✓			28.846	0.945
✓	✓	✓	✓		29.295	0.953
✓	✓	✓	✓	✓	29.765	0.968

As shown in Table 1, the model solely utilizes the Unet++ basic network for image denoising, and the achieved values for PSNR and SSIM were 28.457 dB and 0.926, respectively. With the addition of the RAM, PSNR and SSIM were improved by 0.057 dB and 0.016. It was proved that RAM enabled the model to better retain image structure details, to improve the model's feature extraction capabilities, and to provide more efficient training and learning of image features by the network. With the addition of the FMAM, PSNR and SSIM were improved by 0.332 dB and 0.003. It was proved that the FMAM could effectively combine local and global features and focus on different features at different levels to better capture details and texture information in images. With the addition of the SSFFM, PSNR and SSIM were improved by 0.449 dB and 0.008. It was verified that SSFFM could effectively fuse features to help preserve the details of the image and avoid distortion caused by over-smoothing. With the addition of the ECM, PSNR and SSIM were improved by 0.47 dB and 0.015. This reflects the importance of error in correcting pixels in different areas. The ECM ensured that the content of the image was not destroyed during the model denoising process, thus helping to generate more realistic images. The ablation experiments illustrate the effectiveness and necessity of the proposed components for image denoising.

4.3. Comparison with State-of-the-Art Methods

In order to verify the effectiveness and advancement of the proposed method for creating a ice-covered transmission line dataset, we compared it with state-of-the-art denoising methods, including BM3D, CBDNet, MPRNet, Uformer and PromptIR. Experiments were performed using Gaussian noise standard deviations ranging from 10 to 40, 20 to 50, and 30 to 60. The results are shown in Table 2.

Table 2. Quantitative comparison of different methods on resulting dataset.

	10–40		20–50		30–60	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM	PSNR (dB)	SSIM
BM3D	20.892	0.805	19.393	0.727	17.198	0.615
CBDNet	29.128	0.920	27.595	0.890	25.622	0.841
MPRNet	30.582	0.956	28.067	0.933	26.068	0.892
Uformer	31.080	0.970	28.914	0.957	27.557	0.938
PromptIR	31.043	0.972	29.150	0.959	27.864	0.945
EDUNet++	31.055	0.978	29.765	0.968	28.007	0.957

As shown in Table 2, the PSNR and SSIM values of the proposed method achieve optimal results in different noise environments. When the standard deviation of Gaussian noise falls within the range of 10 to 40, the PSNR value of Uformer achieves its peak at 31.080 dB, that is 0.025 dB higher than the PSNR value of the proposed method. However, the SSIM value of Uformer is 0.008 lower than that of the proposed method. The denoising effect of BM3D is poor in various noise conditions, as seen by the lowest achieved values of PSNR and SSIM. As the noise environment grows increasingly intricate, the PSNR and SSIM values decline faster and faster, indicating that the image denoising effect is getting worse and worse. For Gaussian noise with a standard deviation ranging from 20 to 50, the PSNR and SSIM values of this method are 1.698 dB and 0.035 higher than those of MPRNet. Similarly, for a standard deviation ranging from 30 to 60, the PSNR

and SSIM values are 0.45 dB and 0.019 higher than from Uformer. It can be seen that our method has significant advantages over current deep learning methods. Furthermore, the data in the table clearly demonstrate that when the noise environment becomes more and more complex, the PSNR and SSIM value decay of the proposed method is minimal, indicating that the proposed method possesses robust denoising capabilities for complex noise environments. We conducted a comparison with available state-of-the-art denoising methods and confirmed the usefulness and superiority of our proposed method in image denoising of ice-covered transmission lines.

4.4. Visual Effects Comparison

Quantitative indicators in image denoising can quantitatively evaluate the performance of the denoising algorithm, but they do not fully reflect the impact of the denoising algorithm on the visual perception of the image. In order to consider the denoising of visual perception of the proposed method, we compare both the original and denoised renderings in different noise environments. Figure 6 shows the denoising renderings of different components when the Gaussian noise standard deviation ranged from 20 to 50. Figure 7 shows the denoising of Gaussian noise of different methods in three noise environments.

As shown in Figure 6D exhibits significant noise in most regions, lacks clarity in depicting cable structure and detailed information, and appears relatively blurry. Figure 6E looks relatively clear, with only a small amount of noise on the surface. However, the cable's outline exhibits an irregular form and the structural details appear indistinct, indicating that RAM can effectively suppress the noise on the image surface. In Figure 6F, the surface noise shadow is basically completely removed and the structural details of the cable are clearly presented, but there is still some blurriness, indicating that FMAM can effectively combine local and global features to better capture details and texture information in images and effectively suppress the impact of noise. In Figure 6G, the structural details of the cable are completely clear, and there are no noise shadows on the image surface. However, there is a certain gap in the surface details of the cable, indicating that SSFFM can achieve effective fusion of features and enhance the effect of image denoising. In Figure 6H, the detailed information on the surface of the cable is clearly reproduced, such as brightness, darkness, texture, etc., indicating that ECM can effectively correct errors in the generated image.

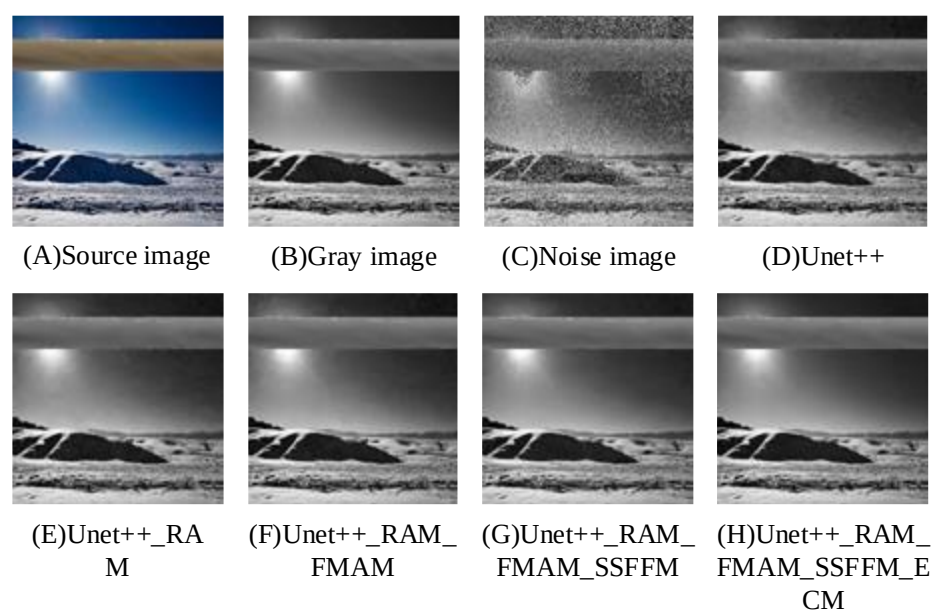


Figure 6. Visual effects of different components: (A) source image, (B) gray image, (C) noise image, (D) Unet++, (E) Unet++ and RAM, (F) Unet++, RAM, and FMAM, (G) Unet++, RAM, FMAM, and SSFFM, (H) Unet++, RAM, FMAM, SSFFM, and ECM.

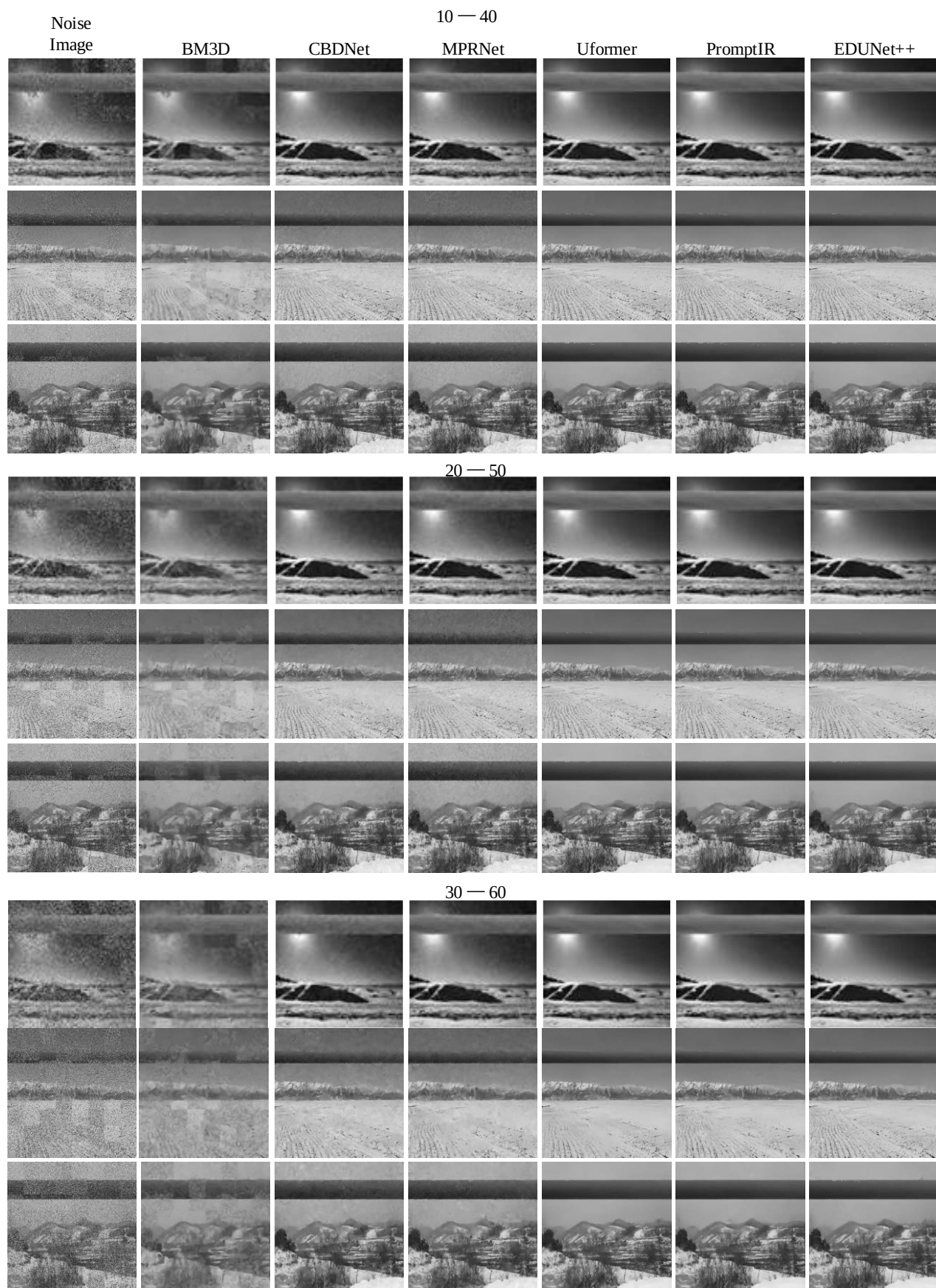


Figure 7. Visual effects of different methods.

As shown in Figure 7, the BM3D method contains a large amount of noise information, and as the complexity of the noise environment increases, more noise remains on the image surface. Although CBDNet and MPRNet have greatly reduced the impact of noise on the image surface, there are still certain blur artifacts and the structural details of the image are not clearly presented. Uformer and PromptIR can effectively reduce blur artifacts on the image surface, and the structure of the image is relatively complete without serious distortion, but the surface details of the image are missing. By contrast, our method has a more obvious denoising effect on non-uniform noise that can effectively remove the noise influence of the image, reduce blur artifacts, and restore smoother cable contours and edges, along with high-frequency detailed information on the cable surface. By conducting the visual perception comparison, we have confirmed that the proposed method can effectively remove complex non-uniform noise, reconstruct the clarity and definition of cable textures, and generate more realistic images. Our method has greater advantages in denoising images of an ice-covered cable and lays a good foundation for a robot to analyze ice thickness on ice-covered lines.

4.5. Image Ice Edge Detection

In order to verify the accuracy and reliability of the proposed method for robot analysis of ice thickness on an ice-covered cable, we conducted cable edge detection experiments. With Gaussian noise standard deviations ranging from 10 to 40, 20–50 and 30–60, the detection results are shown in Figure 8. The actual cable test results are shown in Figure 9.

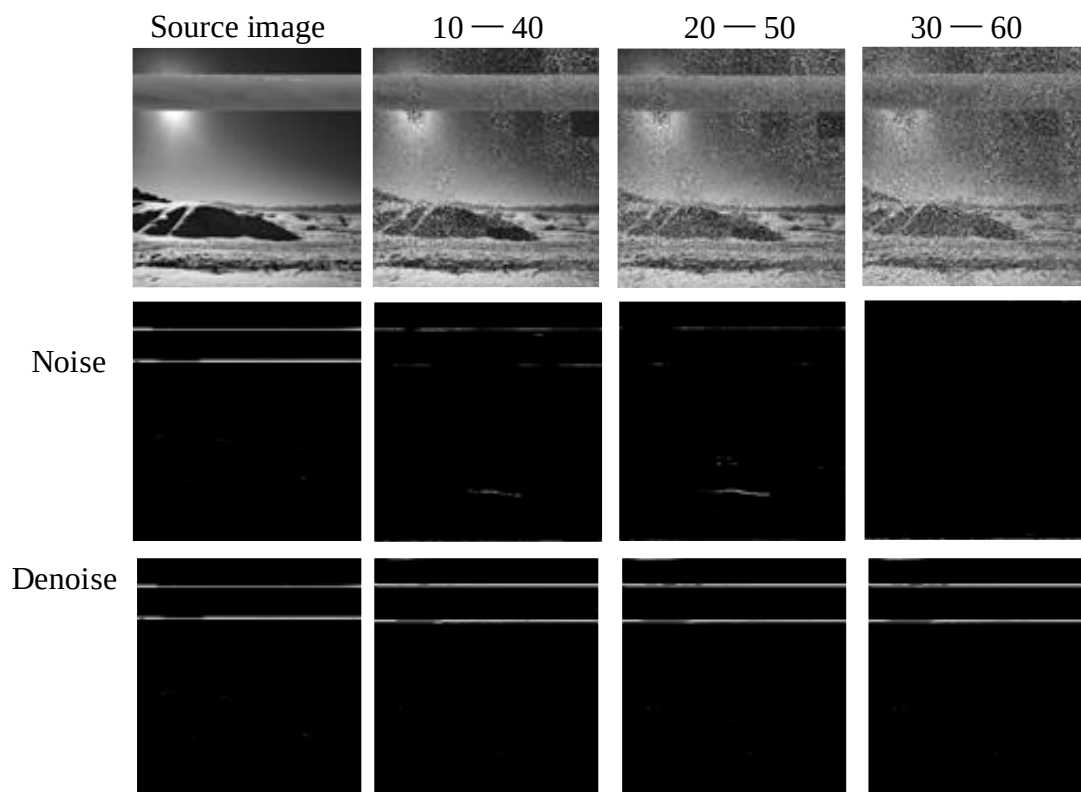


Figure 8. Cable edge detection results.

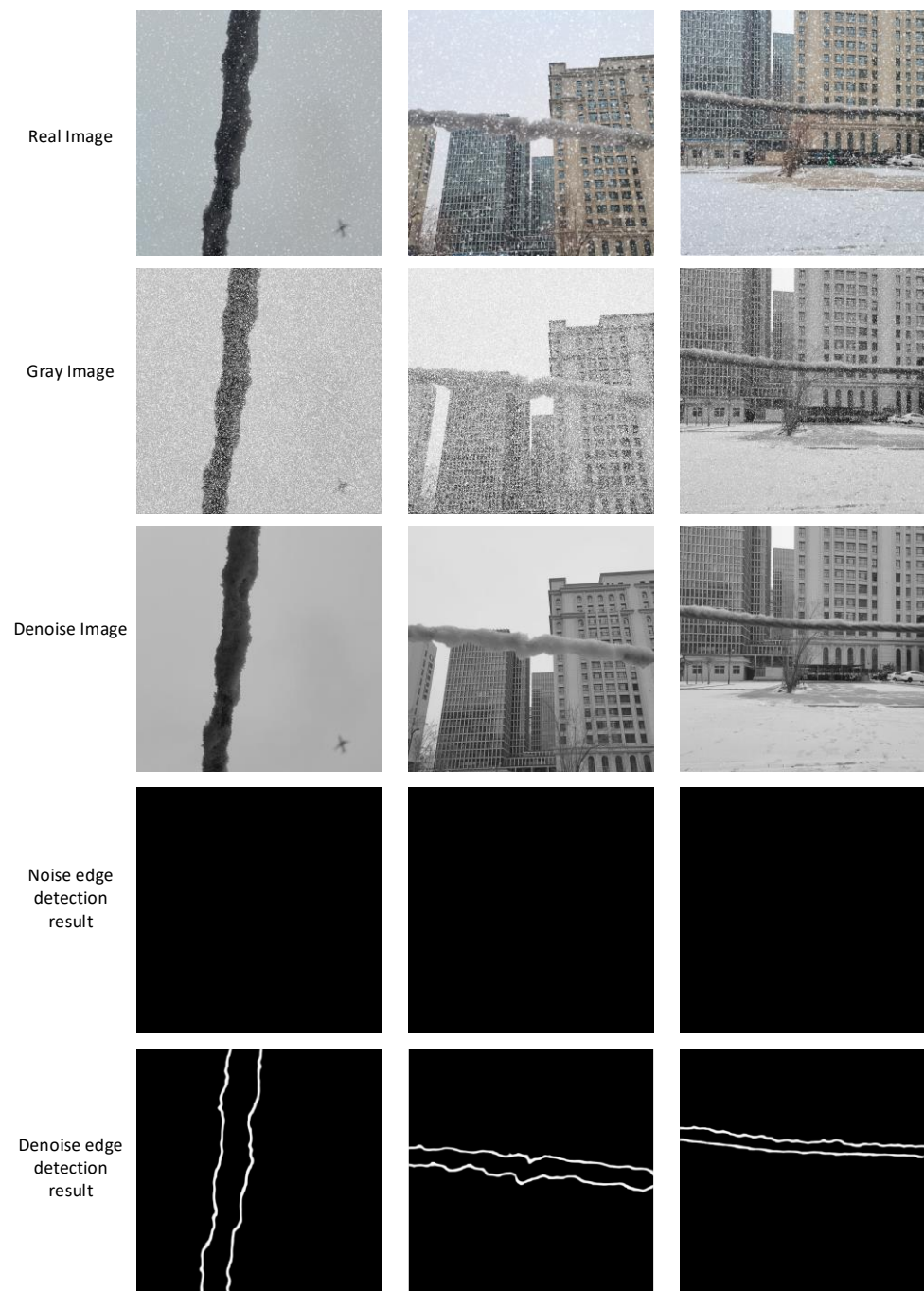


Figure 9. Actual cable edge detection results.

As shown in Figure 8, the second row shows the results of cable edge detection under different noise environments. When the Gaussian noise standard deviation ranges from 10 to 40, the upper edge detection of the cable is intermittent, and only a small part of the lower edge is detected. As the standard deviation of Gaussian noise falls within the range of 20 to 50, the upper edge discontinuity gets more and more serious, and the lower edge is almost not detected. When the Gaussian noise standard deviation ranges from 30 to 60, the edge of the cable is not detected. The results show that the presence of noise has a serious impact on the accuracy of the robot's detection of the ice thickness on an ice-covered cable. As the level of noise interference rises, the robot's detection effect becomes worse. The third row shows the cable edge detection results after denoising by our proposed method.

It demonstrates that, regardless of the type of noise present, the edge of the cable can be accurately detected and the detection results are also reliable.

As shown in Figure 9, the third row shows the denoising result of the image under a snowy environment in winter. This shows that our proposed method can achieve good denoising effects, successfully remove the snowflake interference on the image surface and restore clear cable information. The fourth row shows the cable edge detection results in a snowy environment. The edge detection model cannot detect the cable edge information due to the interference of snowflakes. The fifth row shows the cable edge detection results after denoising by our method. The edge of the cable is accurately detected, indicating that the proposed method can eliminate the noise in the image, enhance the clarity of the image, improve the edge detection accuracy, make the subsequent analysis of the snow and ice thickness of the cable more accurate and reliable, and eventually enable efficacious deicing by the robot.

5. Conclusions

In this paper, we propose an enhanced denoising Unet++ in order to improve the quality of images of ice-covered transmission lines (EDUNet++). The algorithm consists of three crucial modules: a feature encoding and decoding module (FEADM), a shared source feature fusion module (SSFFM) and an error correction module (ECM). Specifically, a residual attention module (RAM) and a multilevel feature attention module (MFAM) are proposed in the FEADM. The RAM incorporates the cascaded residual structure and hybrid attention mechanism, that effectively preserve the mapping of feature information. The MFAM uses dilated convolution to obtain features at different levels, and then uses SE attention for weighting, that effectively combines local and global features and focuses on different features at different levels. The SSFFM realizes the effective transmission line of source features, that enhances the fusion of features and realizes mutual mapping between features. The ECM implements the correction of errors between images. On the ice-covered transmission lines dataset, which we verified through quantitative and qualitative experiments, the experimental results show that the SSIM and PSNR reached 29.765 dB and 0.968, and the visual effects also exhibit enhanced clarity in capturing finer details. This confirms the robustness and reliability of our algorithm in complex noise environments. This method surpasses traditional methods and other deep learning methods; it can effectively suppress complex noise in images, learn the loss of image details and signal distortion caused by noise, and restores smoother cable outlines and clear surface detail information. This research provides a fresh outlook on the clarity and information recovery of ice-covered transmission line images, and provides strong support for detecting the status of ice-covered transmission lines. In future research, it will be deployed on de-icing robots and mounted in actual ice-covered transmission line environments to ensure its scalability in a wider range of application scenarios. We believe that the results of this research will provide valuable reference and inspiration for advanced detection of transmission line conditions.

Author Contributions: Conceptualization, Y.Z.; methodology, Y.Z.; software, Y.Z. and Y.D.; validation, Y.Z. and L.Z.; formal analysis, Y.Z. and Y.D.; investigation, L.Z. and Y.J.; resources, Y.D.; data curation, L.Z., Y.J. and D.G.; writing—original draft preparation, L.Z.; writing—review and editing, Y.Z. and Y.D.; visualization, D.G.; supervision, Y.Z. and L.Z.; project administration, Y.D.; funding acquisition, Y.Z. and Y.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Shanxi Provincial Higher Education Science and Technology Innovation Project (Grant number 2022L524) and Shanxi Provincial Key Research and Development Project (Grant number 202102060301020).

Data Availability Statement: The data presented in this study are available in this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Dong, B.B.; Jiang, X.L.; Yin, F.H. Development and prospect of monitoring and prevention methods of icing disaster in China power grid. *IET Generation. Transm. Distrib.* **2022**, *16*, 4480–4493. [CrossRef]
- Chen, H.; Zhu, X.D.; Xin, Y.C. Review of research on deicing technology for ice-covered transmission lines. *Jilin Electr. Power* **2022**, *50*, 30–34+48. [CrossRef]
- Zhang, Z.J. A Review of Icing and Anti-Icing Technology for Transmission Lines. *Energies* **2023**, *16*, 601. [CrossRef]
- Daniel, D.; Chowdhury, J.; Hangan, H. Joint wind and ice hazard for transmission lines in mountainous terrain. *J. Wind. Eng. Ind. Aerodyn.* **2023**, *232*, 105276. [CrossRef]
- Deng, H.G. A brief analysis of factors affecting transmission line operation safety and preventive measures. *China Equip. Eng.* **2021**, 157–159. [CrossRef]
- Tan, Z.C.; Liu, L.C.; Lei, Y.F.; Xia, J.; Liu, J.Z.; Wang, H.N. Structural design of distribution network line deicing robot. *Integr. Circuit Appl.* **2023**, *40*, 220–221. [CrossRef]
- Li, H.Y.; Li, X.; Xie, J.W. Research on visual positioning method of transmission line deicing robot. *Mech. Des. Manuf.* **2023**, 198–202. [CrossRef]
- Fan, W.; Zhang, S.H.; Zhu, W.D.; Zhu, H. An efficient dynamic formulation for the vibration analysis of a multi-span power transmission line excited by a moving deicing robot. *Appl. Math. Model.* **2022**, *103*, 619–635. [CrossRef]
- Qiao, H.; Chen, J.H.; Huang, X. A survey of brain-inspired intelligent robots: Integration of vision, decision, motion control, and musculoskeletal systems. *IEEE Trans. Cybern.* **2021**, *52*, 11267–11280. [CrossRef]
- George, G.; Oommen, R.M.; Shelly, S.; Philipose, S.S.; Varghese, A.M. A survey on various median filtering techniques for removal of impulse noise from digital image. In Proceedings of the 2018 Conference on Emerging Devices and Smart Systems (ICEDSS), Tiruchengode, India, 2–3 March 2018; pp. 235–238. [CrossRef]
- Xu, W.Z.; Xiao, C.Q.; Jia, Z.R.; Han, Y.P. Digital image denoising method based on mean filter. In Proceedings of the 2020 International Conference on Computer Engineering and Application (ICCEA), Guangzhou, China, 18–20 March 2020; pp. 857–859. [CrossRef]
- Li, F.; Wang, J. Research on image denoising algorithm based on improved wavelet threshold and non-local mean filtering. In Proceedings of the 2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP), Nanjing, China, 22–24 October 2021; pp. 493–497. [CrossRef]
- Zhang, K.; Zuo, W.M.; Zhang, L. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans. Image Process.* **2018**, *27*, 4608–4622. [CrossRef]
- Tian, C.W.; Zhen, M.H.; Zuo, W.M.; Zhang, S.C.; Zhang, Y.N.; Lin, J.W. A cross Transformer for image denoising. *Inf. Fusion* **2024**, *102*, 102043. [CrossRef]
- Chen, X.; Li, H.; Li, M.Q.; Pan, J.S. Learning A Sparse Transformer Network for Effective Image Deraining. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 5896–5905. [CrossRef]
- Zhai, L.J.; Wang, Y.H.; Cui, S.X.; Zhou, Y. A Comprehensive Review of Deep Learning-Based Real-World Image Restoration. *IEEE Access* **2023**, *11*, 21049–21067. [CrossRef]
- Zhu, M.-L.; Zhao, L.-L.; Xiao, L. Image Denoising Based on GAN with Optimization Algorithm. *Electronics* **2022**, *11*, 2445. [CrossRef]
- Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **2007**, *16*, 2080–2095. [CrossRef] [PubMed]
- Li, G.H.; Li, J.J.; Fan, H. Image Denoising Algorithm Based on Adaptive Matching Pursuit. *Comput. Sci.* **2020**, *47*, 176–185. Available online: <https://link.cnki.net/urlid/50.1075.tp.20190816.1713.061> (accessed on 24 April 2024).
- Qian, C.; Chang, D.X. Image denoising algorithm based on graph laplacian regularized sparse transform learning. *J. Comput. Eng. Appl.* **2022**, *58*, 232–239. [CrossRef]
- Srujana, P.; Kumar, D.S.; Ramadevi, B. Comparison of Image Denoising using Convolutional Neural Network (CNN) with Traditional Method. In Proceedings of the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 8–10 April 2021; pp. 826–831. [CrossRef]
- Kumar, V.S.; Samalla, K. Image denoising analysis by Deep Learning Algorithms. In Proceedings of the 2022 6th International Conference on Electronics, Communication and Aerospace Technology, Coimbatore, India, 1–3 December 2022; pp. 1307–1311. [CrossRef]
- Zhang, K.; Zuo, W.M.; Chen, Y.J.; Meng, D.Y.; Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [CrossRef] [PubMed]
- Guo, S.; Yan, Z.F.; Zhang, K.; Zuo, W.M.; Zhang, L. Toward convolutional blind denoising of real photographs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1712–1722. [CrossRef]
- Ding, Y.H.; Wu, H.; Kong, F.L. Two-stage blind denoising for real image noise. *Chin. J. Image Graph.* **2023**, *28*, 2026–2036. [CrossRef]
- Huang, J.; Liu, X.; Pan, Y.Z.; He, X.H.; Ren, C. CasaPuNet: Channel affine self-attention based progressively updated network for real image denoising. *IEEE Trans. Ind. Inform.* **2022**, *19*, 9145–9156. [CrossRef]

27. Mou, C.; Zhang, J.; Fan, X.P.; Liu, H.F.; Wang, R.G. COLA-Net: Collaborative attention network for image restoration. *IEEE Trans. Multimed.* **2021**, *24*, 1366–1377. [[CrossRef](#)]
28. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H.; Shao, L. Multi-stage progressive image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14821–14831. [[CrossRef](#)]
29. Potlapalli, V.; Zamir, S.W.; Khan, S.; Khan, F.S. PromptIR: Prompting for All-in-One Blind Image Restoration. *arXiv* **2023**, arXiv:2306.13090. [[CrossRef](#)]
30. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *Advances in neural information processing systems. arXiv* **2017**, arXiv:1706.03762. [[CrossRef](#)]
31. Wang, Z.D.; Cun, X.D.; Bao, J.M.; Zhou, W.G.; Liu, J.Z.; Li, H.Q. Uformer: A general u-shaped transformer for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 17683–17693. [[CrossRef](#)]
32. Yao, C.; Jin, S.; Liu, M.; Ban, X. Dense residual Transformer for image denoising. *Electronics* **2022**, *11*, 418. [[CrossRef](#)]
33. Yang, Z.Y.; Pan, D.; Shi, P. Joint image dehazing and super-resolution: Closed shared source residual attention fusion network. *IEEE Access* **2021**, *9*, 105477–105492. [[CrossRef](#)]
34. Hwang, J.; Yu, C.S.; Shin, Y. SAR-to-optical image translation using SSIM and perceptual loss based cycle-consistent GAN. In Proceedings of the 2020 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Republic of Korea, 21–23 October 2020; pp. 191–194. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.