applied sciences

MDPI

Article

MCIDN: Deblurring Network for Metal Corrosion Images

Jiaxiang Wang [1,†], Meng Wan [2,†], Pufen Zhang [1], Sijie Chang [1], Hao Du [1], Peng Shi [1,*], Hongying Yu [3], Dongbai Sun [1,4,*], Jue Wang [2] and Yangang Wang [2]

[1] National Center for Materials Service Safety, University of Science and Technology Beijing, Beijing 100083, China; d202110535@xs.ustb.edu.cn (J.W.); d202210559@xs.ustb.edu.cn (P.Z.); d202310583@xs.ustb.edu.cn (S.C.); duhaoai.1101@gmail.com (H.D.)
[2] Computer Network Information Center, Chinese Academy of Sciences, Beijing 100083, China; wanmengdamon@cnic.cn (M.W.); wangjue@sccas.cn (J.W.); wangyg@sccas.cn (Y.W.)
[3] School of Materials, Sun Yat-Sen University, Shenzhen 581007, China; yuhy26@mail.sysu.edu.cn
[4] School of Materials Science and Engineering, Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Sun Yat-Sen University, Guangzhou 581006, China
* Correspondence: pshi@ustb.edu.cn (P.S.); sundongbai@mail.sysu.edu.cn (D.S.)
† These authors contributed equally to this work.

Abstract: The analysis of corrosion images is crucial in materials science, where acquiring clear images is fundamental for subsequent analysis. The goal of deblurring metal corrosion images is to reconstruct clear images from degraded ones. To the best of our knowledge, this study introduces the first paired blurry-sharp image dataset specifically designed for the metal corrosion domain, filling a critical gap in the existing research. This innovative approach effectively addresses the unique challenges associated with deblurring metal corrosion images. We propose a novel metal corrosion images deblurring network (MCIDN) that employs a dual-domain attention mechanism, integrating both spatial and frequency domains to enhance image clarity. This innovative approach effectively addresses the unique challenges associated with deblurring metal corrosion images. While self-attention is widely used in visual tasks, its quadratic complexity often leads to high computational costs. To address this issue, we introduce a new spatial channel attention module (SCAM) that employs dynamic group convolutions to achieve self-attention, effectively integrating information from local regions and enhancing representation learning capabilities. Recognizing the critical role of frequency components in image restoration, we develop a frequency channel attention module (FCAM) that selectively focuses on different frequency components of images, thereby enhancing deblurring performance. These two attention modules are seamlessly integrated into our network. Compared to existing methods, our approach demonstrates superior performance on datasets of blurry metal corrosion images, achieving a peak signal-to-noise ratio (PSNR) of 32.8645 dB and a structural similarity (SSIM) of 0.9768. These metrics indicate that our method provides clearer and more detailed reconstructions, significantly enhancing the image quality.

Keywords: metal corrosion; image deblurring; dual-domain learning; deep learning

check for updates

Citation: Wang, J.; Wan, M.; Zhang, P.; Chang, S.; Du, H.; Shi, P.; Yu, H.; Sun, D.; Wang, J.; Wang, Y. MCIDN: Deblurring Network for Metal Corrosion Images. *Appl. Sci.* **2024**, *14*, 11565. https://doi.org/10.3390/app142411565

Academic Editor: Pedro Couto

Received: 23 October 2024
Revised: 9 December 2024
Accepted: 10 December 2024
Published: 11 December 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 industrial and engineering applications, monitoring and analyzing metal corrosion is paramount due to its impact on structural stability, economic risks, and safety hazards. Leveraging image processing technologies for corrosion detection has become a critical area of research. Recent advancements in image processing have significantly improved corrosion detection [1], identification [2], and segmentation [3]. However, image quality is often compromised by blurring and noise from environmental conditions and capture limitations, affecting detection and recognition accuracy. Therefore, effective image deblurring techniques for metal corrosion images are essential. The improved clarity of reconstructed images has significant implications for materials science research, enabling

detailed analysis of corrosion patterns. It also holds great potential for industrial applications, including bridge degradation assessment, pipeline corrosion detection, and quality control. In these areas, accurate identification of corrosion can enhance safety, optimize maintenance schedules, and reduce associated costs.

Image deblurring is a fundamental low-level vision challenge encountered across various fields, including remote sensing, medical imaging, photography, autonomous driving, and underwater imaging [4–8]. Traditional non-blind deblurring assumes a known blur kernel and employs techniques like Wiener filtering and Richardson–Lucy deconvolution [9,10]. Traditional blind deblurring, where both the blur kernel and the sharp image are unknown, relies on manually crafted priors such as L0 sparsity, low-rank, and dark channel priors [11–14]. These traditional methods often involve complex optimization processes and suboptimal performance.

Recent advances in deep learning have catalyzed the development of Convolutional Neural Networks (CNNs) for image deblurring. These networks effectively translated blurry images into their sharp counterparts [15–17]. Although these methods improved performance, their limited receptive field restricted their ability to model long-range dependencies effectively. Transformer-based approaches addressed the limitation of CNNs by leveraging self-attention mechanisms to enhance image deblurring performance [18,19]. However, this led to a significant increase in computational complexity. Despite some efforts to mitigate the quadratic complexity, the computational overhead remained substantial, and the performance gains over CNNs were relatively modest [20].

In this paper, we propose an efficient spatial channel attention module (SCAM) that combines the benefits of group convolution and self-attention. Our SCAM, normalized by Tanh and sharing weights across the group dimension, emulates self-attention operations through a group convolution mechanism. Unlike traditional self-attention, which aggregates information across extensive regions, SCAM focuses on localized areas to reduce computational complexity. It generates integration weights for every pixel pair based on contextual information, enhancing efficiency through shared weights.

Most deep learning-based deblurring methods focus on the spatial domain, often neglecting frequency domain differences. Recent efforts have aimed to reduce the frequency gap between clear and blurred image pairs. For instance, Zou et al. [21] introduced a wavelet reconstruction module that decouples features into different frequency bands, though it adds computational complexity. Mao et al. [22] process high-frequency signals through normal residual branches and use Fourier transform-based branches for full-frequency signals, but shared inputs can cause interference. Zhang et al. [23] use multiple complex OctConvs for frequency separation, but frequent downsampling and upsampling introduce additional computational burdens.

To leverage frequency features effectively, we propose a frequency channel attention module (FCAM) designed to facilitate interactions among feature channels, enhancing frequency representation learning for image deblurring. FCAM selectively extracts relevant frequency components, aiding image deblurring through channel-wise attention. During reconstruction, FCAM enriches the high-frequency signal spectrum, significantly improving deblurring outcomes. By attenuating the lowest frequency components within features, FCAM accentuates high-frequency signals, refining the deblurring process.

The principal contributions of this paper are as follows:

- We propose a novel metal corrosion image deblurring network (MCIDN) based on deep learning principles. The network employs an encoder–decoder architecture designed to enhance image deblurring by enriching representation learning in both spatial and frequency domains.
- We introduce a SCAM, integrating the strengths of convolution and attention mechanisms to effectively capture local dependencies crucial for precise deblurring.
- We develop an FCAM to aggregate essential frequency components, highlighting our novel approach in utilizing frequency information for image deblurring.

- We create a comprehensive dataset of blurry metal corrosion images, capturing a wide range of corrosion ratings to ensure the robustness and practical applicability of our deblurring model in real-world scenarios. Extensive experimental results demonstrate that our method achieves superior performance on this dataset, effectively handling various ratings of metal corrosion blur.

2. Related Works

The method proposed in this paper enhances deblurring performance in metal corrosion images through dual-domain learning. This section reviews key aspects related to the proposed method, including deblurring datasets, image deblurring techniques, attention-based image restoration, and frequency learning in image restoration.

2.1. Deblurry Dataset

The development of high-quality datasets is crucial for model training in deep learning, as they significantly influence training outcomes and their applicability in real-world scenarios. In the field of high-level computer vision, prominent datasets such as ImageNet [24], COCO [25], and PASCAL VOC [26] have been developed. These datasets focus on a broad spectrum of real-world objects and scenes, encompassing categories like animals, everyday objects, and complex environments. Similarly, in the medical domain, critical datasets include fastMRI [27], which consists of knee and brain MRI scans; the BraTS [28] dataset for brain tumor segmentation; and the MIMIC-CXR [29] dataset for chest X-ray images, aimed at predicting conditions such as atelectasis, edema, and cardiomegaly. In the realm of remote sensing, datasets like DOTA [30] for object detection in aerial images and HRSC2016 [31] for ship detection and segmentation are vital for advancing vision tasks. These datasets are indispensable for high-level vision tasks, including image classification, detection, and segmentation, providing extensive annotations and a broad range of categories that enhance the accuracy and robustness of models across various applications.

Image restoration tasks, considered as low-level vision challenges, encompass image dehazing [32], deblurring [20], denoising [33], deraining [34], and desnowing [35]. Datasets like SOTS [36] and Dense-Haze [37] for dehazing, Snow100K [35] for desnowing, GoPro [38] and HIDE [39] for deblurring, and Rain100L and Rain100H [40] for deraining are predominantly derived from everyday scenarios featuring common subjects like people, vehicles, and buildings.

In the field of material science, particularly regarding metal corrosion, there is a notable lack of specific datasets. This gap highlights the necessity for tailored datasets in this domain, as existing ones are inadequate for tasks such as metal corrosion image deblurring. Therefore, in this paper, we have created a blurry metal corrosion dataset. This dataset can be used to train models for removing blurriness from corrosion images, facilitating subsequent tasks such as corrosion level prediction, corrosion type detection, and corrosion area segmentation.

2.2. End-to-End Image Deblurring

Recent advancements in CNNs have significantly improved image deblurring, outperforming traditional methods. DeepDeblur [38] introduced direct regression between blurred and sharp images using a multi-scale CNN, setting a new standard. Subsequent methods like the scale-recurrent network by Tao et al. [41] and selective parameter sharing by Gao et al. [42] further enhanced performance.

As large models yield diminishing returns, Zhang et al. [15] proposed a pyramid-like hierarchical network for coarse-to-fine processing. Zamir et al. [43] improved feature exploration across network stages, while Cho et al. [44] reduced computational complexity with a multi-input/output structure. However, the spatial invariance of convolutional operations still limits comprehensive contextual modeling.

To address these limitations, Transformer-based models have been introduced, offering significant advancements in image processing. Zamir et al. developed Restormer [18], which

utilizes MDTA and GDFN modules to facilitate long-range pixel interactions. SwinIR [45] enhances efficiency through localized self-attention mechanisms. Kong et al. proposed FSAS [46], incorporating element-wise products to optimize self-attention. Tsai et al. introduced Stripformer [20], leveraging strip tokens to effectively capture blur across various orientations. Additionally, Li et al. [47] amalgamated multiple attention mechanisms to achieve comprehensive deblurring. These innovations collectively pushed the boundaries of deblurring techniques, showcasing the potential of Transformer-based approaches in overcoming traditional challenges. These methods primarily focused on spatial dimensions, often neglecting channel interaction with frequency. Our approach integrates spatial and frequency channels to enhance deblurring performance in blurry metal corrosion images. To provide a clear overview of the recent advancements in image deblurring, Table 1 summarizes key methods, their methodologies, datasets, and evaluation metrics.

Table 1. Summary of recent advancements in image deblurring methods.

Methods	Dataset	Network	Metrics	Application
DeepDeblur [38]	GoPro	CNN	PSNR, SSIM	General images
DMPHN [15]	GoPro	CNN	PSNR, SSIM	General images
MPRNet [16]	GoPro	CNN	PSNR, SSIM	General images
HINet [48]	GoPro	CNN	PSNR, SSIM	General images
Restormer [18]	GoPro	Transformer	PSNR, SSIM	General images
NAFNet [33]	GoPro	CNN	PSNR, SSIM	General images
SRN-DeblurNet [41]	GoPro	RNN	PSNR, SSIM	General images
DeblurGan-v2 [49]	GoPro	GAN	PSNR, SSIM	General images
DBGAN [50]	GoPro	GAN	PSNR, SSIM	General images
SwinIR [45]	GoPro	Transformer	PSNR, SSIM	General images
Stripformer [20]	GoPro	Transformer	PSNR, SSIM	General images
GRL [47]	GoPro	Transformer	PSNR, SSIM	General images

These methods primarily focused on spatial dimensions, often neglecting channel interaction with frequency. Our approach integrates spatial and frequency channels to enhance deblurring performance in blurry metal corrosion images.

2.3. Attention Mechanisms in Image Restoration

Attention mechanisms have significantly improved high-level vision tasks, leading to their adoption in low-level vision tasks to focus on crucial information and ignore less relevant details [51,52]. For instance, Qin et al. [53] introduced a feature fusion attention network for image dehazing that flexibly processes different channels and pixels. Zamir et al. [43] developed a multi-stage architecture that uses attention mechanisms to reweight local features at each stage, while Chen et al. [33] proposed a network with simplified channel attention and gating mechanisms. However, these methods mainly operate in the spatial attention without addressing the frequency attention.

In this paper, we introduce a spatial channel attention module that focuses on blurry areas and a frequency channel attention module that enhances high-frequency details, collectively improving deblurring performance.

2.4. Frequency Learning in Image Restoration

From the perspective of image restoration, image deblurring requires addressing various frequency signals to achieve clearer images. Low-frequency signals correspond to smooth transitions, while high-frequency signals are associated with abrupt changes such as edges and textures. The Fast Fourier Transform (FFT) facilitates efficient global feature modeling by leveraging the convolution theorem. Huang et al. [54] propose a Fourier-based exposure correction network, employing amplitude and phase sub-networks to reconstruct luminance and structural information. Guo et al. [55] introduce a spatial-

frequency attention network that uses FFT for global information modeling. Li et al. [56] embed FFT into a cascaded network to enhance brightness and reduce noise in low-light images. Mao et al. [57] convert spatial features into the frequency domain for learning across frequencies.

Additionally, wavelet transform has been incorporated into image restoration networks. Huang et al. [58] propose a wavelet-based diffusion model, learning high-quality image distribution from low-quality wavelet spectra. Zou et al. [21] suggest using distinct convolutional layers for wavelet components, while Chen et al. [59] design a wavelet loss using dual-tree wavelet transforms to capture complex snow shapes. Techniques like filtering [60,61], OctConv [23], and Discrete Cosine Transform [62] have also been explored.

In this paper, we introduce a comprehensive approach combining spatial and frequency domain mechanisms. Our SCAM enables the model to focus on blurry areas within the image, identifying regions requiring deblurring. Concurrently, our FCAM enhances high-frequency details in these regions, significantly improving overall deblurring performance. This dual-domain approach allows the model to flexibly handle diverse frequency information with lower complexity, effectively aggregating essential frequency components for enhanced image restoration, particularly for metal corrosion images.

3. Method

The schematic diagram of the model architecture is shown in Figure 1. First, we introduce the overall architecture of our MCIDN model. Next, we delve into the details of the proposed SCAM and FCAM modules. Finally, we describe the custom loss function designed specifically for our model.

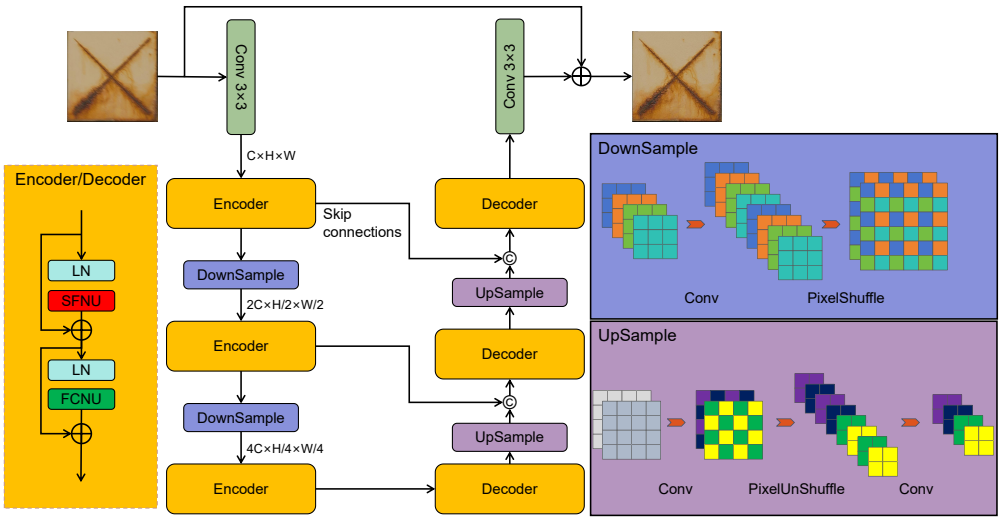


Figure 1. Overall architecture of the proposed MCIDN.

3.1. Architecture

Our network employs an encoder–decoder architecture, enabling the generation of feature maps at different scales, as illustrated in Figure 1. The encoder–decoder block in the network is comprised of large residual dual-attention group networks, with downsampling modules utilized between encoders and upsampling modules between decoders. The encoders adjust the channel size through convolution operation, allowing cascading with the corresponding decoder for output.

The process begins with the input of a blurred image, from which shallow features are extracted using a 3×3 convolutional layer. These features are further processed through successive encoder blocks. Upon reaching the deepest features via the final encoder, the decoders are employed to progressively restore high-resolution features. During this restoration process, feature fusion between corresponding encoders and decoders facilitates enhanced image reconstruction.

Given a blurry image of dimensions $H \times W \times 3$, where $H \times W$ represents the height and width, and 3 denotes the RGB channels, an initial 3×3 convolutional layer extracts shallow features of size $H_{\downarrow} \times W_{\downarrow} \times C_{\uparrow}$. These features are progressively enriched with deep features through multiple encoder subnetworks. Subsequent downsampling modules following each encoder expand the channel dimensions while reducing the spatial resolution. The deep features produced by the final encoder are input into the decoders, where subsequent upsampling modules compress the channels and increase the resolution, gradually restoring the features to the original image size and producing sharp features. The encoder and corresponding decoder are combined through skip connections to aid in the reconstruction task, where 3×3 convolutions halve the feature channel dimensions. Following the final decoder, a 3×3 convolution and a skip connection with the input image features are employed to generate the sharp image.

The internal structure of both the encoder and decoder is identical, each comprising two residual attention networks. In the first residual network, the input first passes through a Layer Normalization (LN) layer, followed by the spatial frequency network unit (SFNU), producing the output, which is then connected residually to the input. The SFNU includes the SCAM and FCAM, which will be elaborated upon later. The output of the first residual network is the input of the second residual network. In the second residual network, the input similarly passes first through the LN layer, then through the feedforward channel network unit (FCNU), producing the output, which is residually connected to the input.

3.2. Spatial Frequency Network Unit (SFNU) and Feedforward Channel Network Unit (FCNU)

The structures of the SFNU and FCNU are shown in Figure 2.

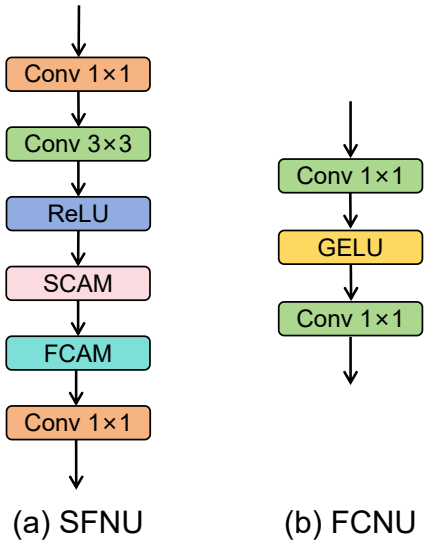


Figure 2. The structures of (**a**) SFNU and (**b**) FCNU.

By observing Figure 2a, the input to the network is first normalized using LN. The input feature map is then transformed through a 1×1 convolution for channel transformation, followed by a 3×3 convolution to alter the feature map's dimensions. Nonlinearity is introduced via the ReLU activation function, followed by the SCAM and FCAM. The process ultimately outputs through a 1×1 convolution to adjust the number of channels. As illustrated by the encoder in Figure 1, the output of the SFNU is residually connected to the input before the LN layer.

As illustrated in Figure 2, although the FCNU only employs 1×1 convolutions, it remains both effective and efficient. The network first undergoes a 1×1 expansion of the number of channels, increasing the feature dimension and enabling the network to capture more information and details. Subsequently, nonlinearity is introduced through the GELU activation function, allowing the model to handle complex patterns and features, thereby

facilitating the learning of more intricate mapping relationships. Finally, the number of channels is compressed through another 1×1 convolution. These operations ensure the model's performance while maintaining high computational efficiency. This structure can adaptively adjust the number of channels while preserving low computational complexity, thereby enhancing the network's expressive power.

3.3. Spatial Channel Attention Module (SCAM)

The architecture of the SCAM is illustrated in Figure 3. This module employs a convolutional network to generate weights between different channels, facilitating enhanced feature representation.

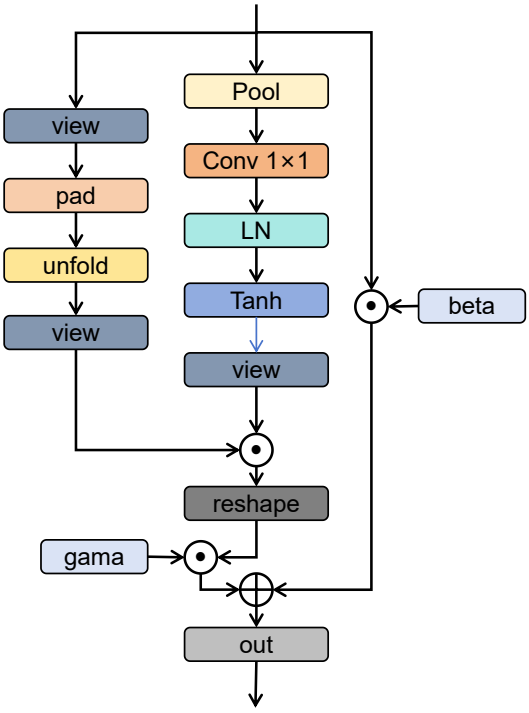


Figure 3. The architecture of proposed SCAM.

First, adaptive average pooling (AAP) is applied to the input feature map I to obtain global channel information. Then, a 1×1 convolution ($W_{1\times1}$) is used to compress or expand the number of channels in the globally pooled features. Next, the convolution output is normalized using LN. Following normalization, the features undergo a nonlinear transformation via the Tanh activation function. Finally, the activated features are processed by the function V to generate attention weights W^1. The process of generating attention weights can be formally expressed as

$$W^1 = V_1(Tanh(LN(W_{1\times1}(AAP(I))))) \tag{1}$$

where AAP, $W_{1\times1}$, LN, $Tanh$, and V correspond to adaptive average pooling, 1×1 convolutional layer, layer normalization, Tanh activation function, and dimension transformation, respectively. The use of the V function for dimension transformation here matches the operation of multiplying it with the feature tensor of another path. The other path first performs a dimensional transformation on the input tensor, followed by a sliding window operation to extract local features. Subsequently, the shape of these features is transformed. The specific process is detailed below:

$$W^2 = V_3(Unfold(Pad(V_2(I)))) \tag{2}$$

where Pad and $Unfold$ represent the reflection padding operation and the sliding window operation. The final output of the SCAM can be formulated as follows:

$$\widehat{I}_c = \beta I + \gamma \mathrm{R}\left(W^1 W^2\right) \tag{3}$$

where R denotes the reshape operation, γ and β are two learnable parameters.

3.4. Frequency Channel Attention Module (FCAM)

In addition to the SCAM, we introduce the FCAM for image deblurring, as illustrated in Figure 4. Recognizing the differences in high-frequency information between blurred and sharp images, we initially employ convolutional operators to enhance the high-frequency features, similar to high-pass filters. Based on the learned attention weights, we then amplify the useful frequency components and merge these enhanced features to facilitate image restoration.

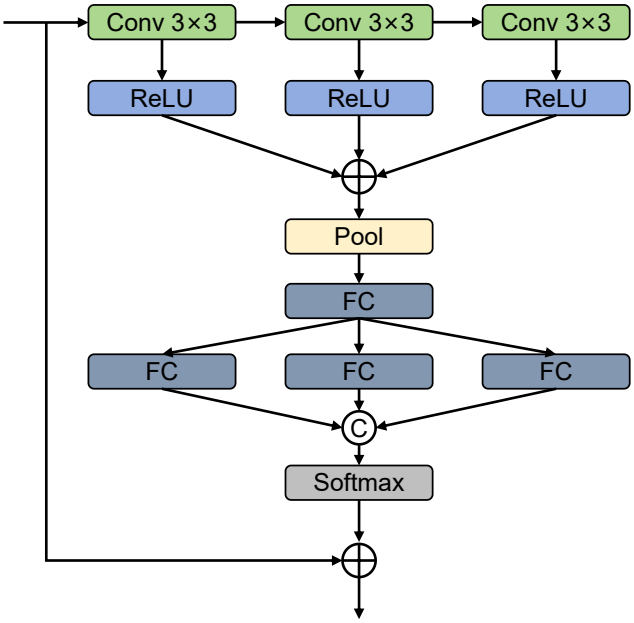


Figure 4. The architecture of proposed FCAM.

Specifically, given an input I, the FCAM utilizes a sequence of convolutional layers to enrich the high-frequency information. The process can be formalized as follows:

$$F_1 = f^1_{3\times3}(I) \tag{4}$$

$$F_2 = f^2_{3\times3}(F_1) \tag{5}$$

$$F_3 = f^2_{3\times3}(F_2) \tag{6}$$

where $f^n_{3\times3}$ denotes a 3×3 convolutional layer. This explanation is initially presented within the context of a dual-frequency scenario for simplicity, yet the approach is readily extendable to accommodate multifrequency scenarios. Following this, the network is tailored to modulate individual frequency components and subsequently amalgamate them. To this end, channel attention weights are computed as follows:

$$W_i = L_i(L_0(AAP(\sum_i F_i))), i \in \{1,2,3\} \tag{7}$$

where F_i represents the input feature shape, and L denotes a fully connected layer. We use concat operation to add up the attention weights of the three channels, and then generate the final attention weights.

$$\left[W_1', W_2', W_3'\right]_c = \frac{e^{[W_1, W_2, W_3]_c}}{\sum_{j=1}^{3C} e^{[W_1, W_2, W_3]_j}} \tag{8}$$

In Equation (8), $[\cdot, \cdot, \cdot]$ denotes the concatenation operation. The final output of FCAM can be generated using the following equation:

$$\hat{I} = W_{1\times1}\left(F_1 W_1' + F_2 W_2' + F_3 W_3'\right) + I \tag{9}$$

3.5. Loss Function

We adopt multiple loss functions to train our network. All loss functions include spatial domain loss L_s, frequency loss L_f, and edge loss L_e. The total loss function is expressed as follows:

$$L = L_s + \lambda_1 L_f + \lambda_2 L_e \tag{10}$$

where λ_1 and λ_2 are set to 0.1 and 0.01, respectively, to maintain a balanced learning process between the loss functions.

3.5.1. Spatial Domains Loss

The spatial domain loss function L_s measures the discrepancy between the predicted and ground truth images in the pixel domain. The L_s loss is expressed as

$$L_s = \sum_{r=1}^{R} \frac{1}{S_r} ||\hat{y}_r - y_r||_1 \tag{11}$$

where $\hat{y}_r$ denotes the predicted image at the r-th scale, y_r is the corresponding ground truth image, R represents the index of the multi-scale output, and S_r is the normalization factor indicating the total number of elements in $\hat{y}_r$. The L_s loss directly quantifies the pixel-wise differences between the predicted and ground truth images, ensuring visual similarity.

3.5.2. Frequency Domain Loss

The frequency domain loss function L_f assesses the discrepancy between the predicted and ground truth images in the frequency domain. Using the Fourier transform F, the images are converted from the spatial domain to the frequency domain, highlighting their frequency characteristics. It is defined as

$$L_f = \sum_{r=1}^{R} \frac{1}{S_r} ||F(\hat{y}_r) - F(y_r)||_1 \tag{12}$$

where $F(\cdot)$ represents the Fourier transform. The frequency domain loss captures the global structures and repetitive patterns of the images, which is crucial for preserving the details and overall structure of the images.

3.5.3. Edge Loss

The edge loss function L_e enhances the edge features of the images, ensuring that the generated images have sharper and more defined edges. We use the Sobel operator to

extract the edge information of the images and then calculate the differences in the edge maps between the predicted and ground truth images. Edge loss L_e is expressed as

$$L_e = \sum_{r=1}^{R} \frac{1}{S_r} ||S(\hat{y}_r) - S(y_r)||_1 \tag{13}$$

where $S(\cdot)$ denotes the edge detection operation using the Sobel operator. The edge loss significantly improves the detail representation of the images, particularly in the edge and texture regions.

By integrating these three loss functions, our model optimizes the images at the pixel level, frequency characteristics, and edge details, thereby producing more realistic and detail-rich images.

4. Experiment

In this section, we first introduce the blurry metal corrosion images dataset and the experimental setup details. Next, we present the experimental results. The best performance in the given tables is highlighted in bold, followed by the results of the ablation studies.

4.1. Dataset Construction

In this paper, we create a dataset containing images of blurry metal corrosion, primarily referencing the corrosion images collected by Yin et al. [63]. According to the ASTM D1654 standard, corrosion ratings range from 0 to 10, with a rating below 5 indicating severe corrosion and a rating of 10 representing no corrosion, corresponding to the initial condition of the metal corrosion test. For this study, we select corrosion ratings between 5 and 9, as they provide a more representative and challenging range for evaluating corrosion effects. These images are meticulously selected by material scientists and categorized into 5 to 9 distinct corrosion ratings, resulting in a total of 600 images with 120 images per rating. The corrosion ratings are annotated based on over a decade of laboratory corrosion testing by material scientists, encompassing common corrosion scenarios encountered in daily life and offering substantial practical application insights.

Our dataset, depicted in Figure 5, consists of images that contain one or two scratches. The images are labeled with varying corrosion ratings, where a higher rating signifies less severe corrosion. To increase the challenge of deblurring corrosion images, our dataset includes metal samples with various background colors to prevent the background from being overly monotonous. We then apply an algorithm to blur these images, creating pairs of blurred and sharp images for training our metal corrosion images deblurring model. As outlined in Table 2, the dataset is randomly divided into training, validation, and test sets with a ratio of 8:1:1, ensuring no overlap of images across these sets to promote model generalization to new images.

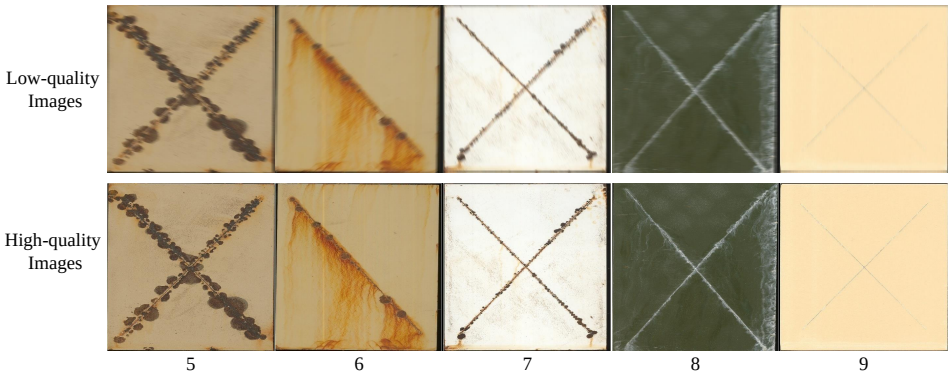


Figure 5. Some low-quality and high-quality images of corrosion rating class from 5 to 9 in our dataset.

Table 2. Detailed divison of the metal corrosion images.

	Total	Training Set	Validation Set	Test Set
Original images	600	480	60	60
Blurry images	600	480	60	60
Total	1200	960	120	120

As shown in Table 3, to balance the impact of each category on the model, our dataset for each corrosion rating comprises 120 images. This balanced approach aims to enhance the generalizability of the trained model.

Table 3. Detailed distribution of different classes of corrosion ratings.

Corrosion Rating	Total	Training Set	Validation Set	Test Set
5	120	96	12	12
6	120	96	12	12
7	120	96	12	12
8	120	96	12	12
9	120	96	12	12
Total	600	480	60	60

4.2. Implementation Details

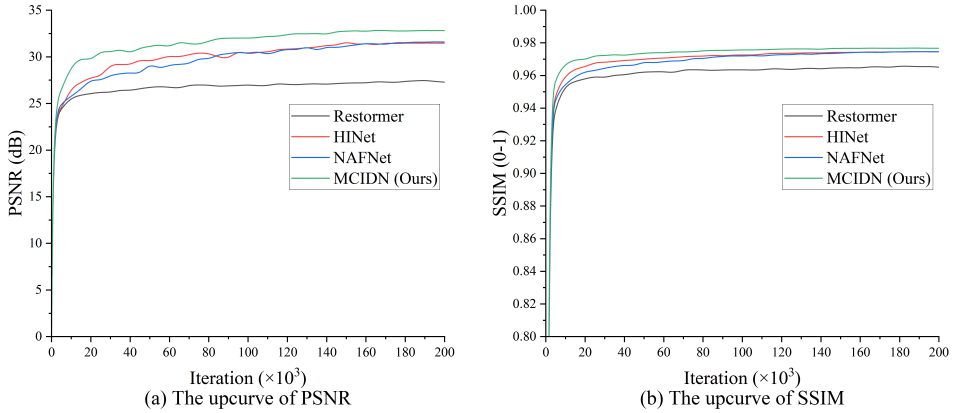
The model slices the blurry corrosion images into 256 × 256 patches for training, while the original image size is used for testing. Both training and testing are conducted on an NVIDIA Tesla V100 GPU. During training, random horizontal flipping and rotation are used for data augmentation. We employ the AdamW optimizer, with the beta parameters set to 0.9. The learning rate decreases from $1e^{-3}$ to $1e^{-7}$ using cosine annealing, and the weight decay is set to $1e^{-3}$. MACs are tested on images with a patch size of 256 × 256.

4.3. Experiment Result

4.3.1. Training Process Comparison

Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) are widely used metrics for evaluating the performance of image deblurring algorithms. PSNR quantifies the difference between the deblurred image and the original image by measuring the pixel-wise mean squared error (MSE). Higher PSNR values indicate that the deblurred image is closer to the original image, thus signifying better image quality. SSIM, on the other hand, evaluates the perceptual similarity between the deblurred image and the original image by comparing luminance, contrast, and structural information. Higher SSIM values denote greater similarity and, therefore, better visual quality. In the context of image deblurring, achieving high PSNR and SSIM values signifies that the algorithm effectively restores image details and reduces blurriness, resulting in clearer and more accurate images.

We compare the training process curves of our method with several state-of-the-art (SOTA) methods on the blurry metal corrosion dataset, as illustrated in Figure 6. The horizontal axis represents 20,000 iterations of the model. Figure 6a shows the PSNR curves for different methods as a function of iteration count, while Figure 6b presents the SSIM curves, with SSIM values ranging from 0 to 1. For enhanced visibility, the Y-axis for SSIM starts from 0.8.



(a) The upcurve of PSNR (b) The upcurve of SSIM

Figure 6. The PSNR and SSIM curves comparing SOTA methods with our proposed method. The curves illustrate the performance improvements achieved by our method over the iterations.

The PSNR curve of Restormer demonstrates the slowest change, underscoring the challenging nature of training on the blurry metal corrosion dataset. HINet and NAFNet exhibit very similar trends in their PSNR and SSIM curves, with their final results being very close. In contrast, our method shows the fastest improvement in both PSNR and SSIM curves, achieving superior final results compared to the other methods. Specifically, our method reaches a final PSNR of 32.8645 dB and an SSIM of 0.9768.

This superior performance can be attributed to the effective integration of spatial and frequency attention mechanisms, which allow our model to better capture and enhance relevant features in the blurry metal corrosion images. The rapid convergence of our method indicates a more efficient learning process, which is critical for practical applications where training time is a significant factor. Additionally, the high final values of PSNR and SSIM suggest that our model not only learns faster but also generalizes better to the task of deblurring metal corrosion images, making it a robust and reliable solution for industrial applications.

4.3.2. Quantitative Comparison

We evaluate our method against current SOTA methods on the metal corrosion dataset, where each SOTA method is retrained and fine-tuned specifically for this dataset. The results of these comparisons are summarized in Table 4. Our MCIDN model achieves superior performance in both PSNR and SSIM metrics. Specifically, MCIDN attains a PSNR of 32.8645 dB on the blurry metal corrosion dataset, which is 1.2607 dB higher than NAFNet. Although the SSIM improvement over NAFNet is relatively smaller at 0.0019, it is still noteworthy.

Table 4. Quantitative comparison of our proposed network with previous methods.

Methods	PSNR	SSIM	Params(M)	FLOPs(G)
Blurry Image	23.4609	0.9229	-	-
Stripformer [20]	26.2827	0.9543	19.7	170.46
MPRNet [16]	27.1775	0.9649	20.1	588
Restormer [18]	27.4252	0.9660	26.1	140
DMPHN [15]	27.5943	0.9607	21.7	-
HINet [48]	31.4780	0.9748	88.7	170.7
NAFNet-64 [33]	31.6038	0.9749	63.36	67.8
MCIDN (Ours)	32.8645	0.9768	28.59	35.3

Our method also surpasses HINet by 1.3865 dB in PSNR and 0.002 in SSIM. When compared to DMPHN, our model exhibits a significant improvement of 5.2702 dB in PSNR and 0.0061 in SSIM. Additionally, our approach outperforms other competitive methods including Stripformer, MPRNet, and Restormer.

The outstanding performance of our MCIDN model can be attributed to the innovative integration of spatial and frequency attention mechanisms, which enhance the model's ability to capture and refine crucial features in blurry metal corrosion images. This integration allows for a more precise restoration of image details, leading to higher PSNR and SSIM values.

Moreover, our model maintains a competitive range in terms of parameters and FLOPs, ensuring that the enhanced performance does not come at the cost of significantly increased computational resources. This balance between performance and efficiency makes our MCIDN model not only highly effective but also practical for real-world applications where computational resources may be limited.

The significant improvements in both PSNR and SSIM metrics across various comparisons underscore the robustness and generalizability of our approach. These results demonstrate that our method sets a new benchmark for deblurring tasks in the context of metal corrosion images, making it a valuable tool for industrial applications where image clarity is critical for accurate analysis and decision-making.

4.3.3. Qualitative Comparison

Figure 7 presents the visual deblurring effects of various methods on blurry metal corrosion images. The "Blurry Image" denotes the original low-quality image, while the "Reference" signifies the sharp and high-quality image. Each image also displays the PSNR values of the deblurring results from different methods.

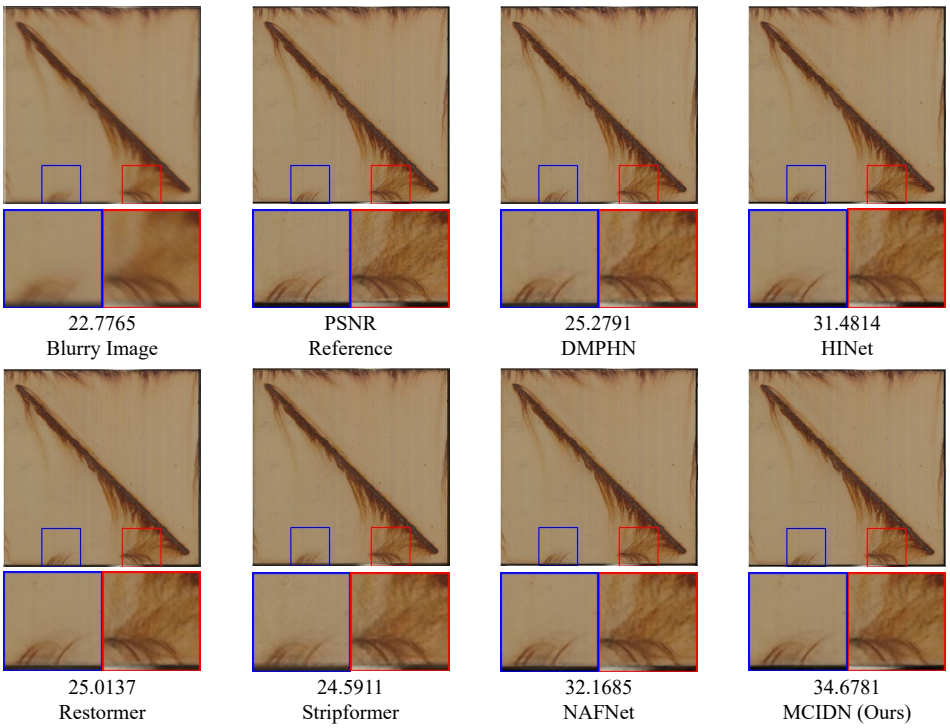


Figure 7. Qualitative comparison of image deblurring methods on blurry metal corrosion dataset.

Observing the results from the Stripformer method, we notice that the corrosion details remain relatively blurred, indicating insufficient improvement. The results from Restormer and DMPHN also exhibit noticeable blurriness with discontinuous boundaries, highlighting their limited restoration capabilities. HINet's results, although improved, still lack sharpness in edge details, which is critical for accurate corrosion detection. NAFNet performs relatively well but still presents unclear edges and corrosion details that require further refinement.

In contrast, our MCIDN method demonstrates the best visual results, with significantly clearer corrosion contours and richer detail. This superior performance underscores the

effectiveness of our approach to enhancing image clarity and preserving critical details, which are essential for accurate analysis in practical applications. The MCIDN method's ability to produce sharp, detailed images from blurry inputs highlights its robustness and generalizability across different types of corrosion images.

The clear delineation of corrosion features in images processed by MCIDN not only facilitates better visual interpretation but also enhances the potential for automated analysis and decision-making in industrial applications. The comprehensive evaluation demonstrates that our MCIDN method sets a new benchmark for deblurring tasks, effectively bridging the gap between blurred and clear images. This advancement is particularly valuable in industrial contexts where the accurate depiction of corrosion can lead to more precise diagnostics and maintenance decisions.

More examples of original images and their deblurred versions using our method are shown in Figure 8. Figure 8a presents the original low-quality corrosion image, Figure 8b displays the high-quality reference image, and Figure 8c illustrates the deblurred image generated by our method. We can observe that the deblurred metal corrosion images produced by our method exhibit clear edges and corrosion details, closely resembling the sharp, original images. This demonstrates the effectiveness of our approach to restoring the clarity and detail of metal corrosion images, making them highly similar to their high-quality counterparts.



(a) Original Image (b) Reference (c) Ours

Figure 8. Examples of images after deblurring using the proposed method.

To clearly demonstrate the deblurring effectiveness of our method, Figure 9 presents the edge maps and high-frequency maps of the metal corrosion images. It is evident that there are significant differences in edge information and high-frequency details between the blurry corrosion images and the sharp corrosion images. By examining the edge maps and high-frequency maps of the deblurred images produced by our method, we can observe more distinct edges and detailed texture information. This improvement is attributed to the SCAM and FCAM proposed in this paper.

Upon examining these images, our model demonstrates excellent deblurring performance, effectively restoring complex structures and intricate corrosion details in blurry metal corrosion images. The results indicate that MCIDN not only enhances the overall image clarity but also preserves essential features that are critical for accurate corrosion analysis.

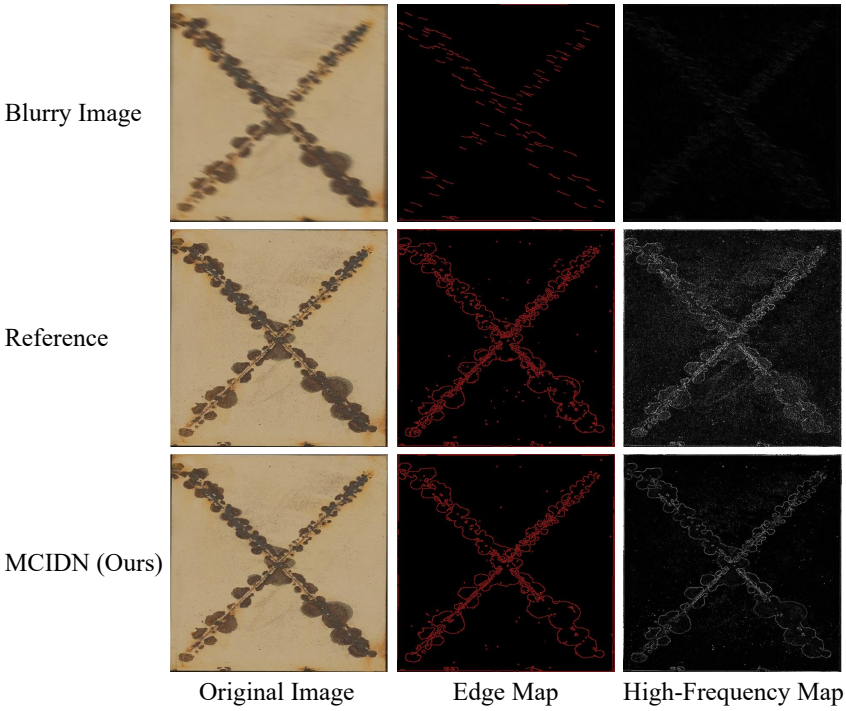


Figure 9. A visual comparison of edge and high-frequency maps of the blurry image, sharp image, and deblurred image using the proposed method. It is evident that the blurry images of metal corrosion exhibit significant loss of edge and high-frequency information. The proposed method effectively restores the edge details and high-frequency components of metal corrosion images.

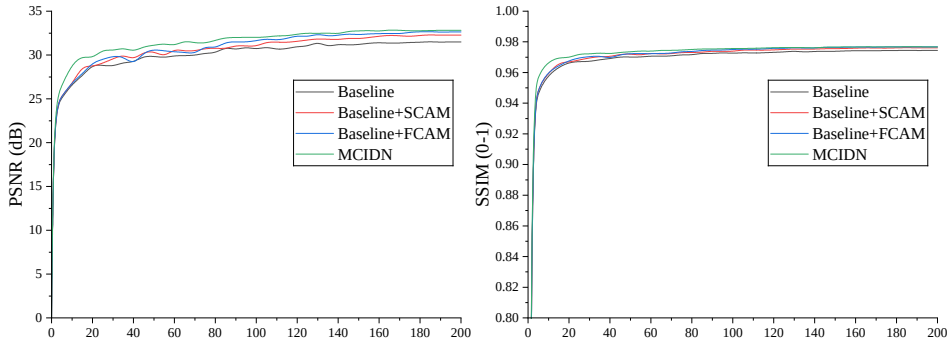
4.4. Ablation Studies

We conduct extensive ablation studies to verify the effectiveness of SCAM and FCAM in our model on blurry metal corrosion images. In this section, our baseline network does not include SCAM and FCAM. The experimental results show that adding these two components effectively enhances the performance of the baseline model. We also validated the effectiveness of the multiple loss functions employed in this paper.

4.4.1. Influence of Each Module

Figure 10 shows the PSNR and SSIM curves for the baseline model and the models with each added component, where the curve labeled MCIDN represents the baseline network with both SCAM and FCAM added. The results indicate that the baseline model has limited effectiveness. When SCAM and FCAM are added separately, the PSNR curve shows improvements, and the SSIM curve also improves. This is due to the effectiveness of the spatial and frequency-domain attention mechanisms in SCAM and FCAM, respectively. When both components are added to the baseline network, the curve achieves the best results, indicating the significant improvement provided by these components.

Table 5 shows that our baseline network achieves a PSNR of 31.5345 dB on the blurry metal corrosion dataset. Adding SCAM to the baseline increases the PSNR by 0.7761 dB. Adding FCAM, which integrates frequency signals, improves the PSNR by 1.1415 dB. The complete model, incorporating both SCAM and FCAM, achieves the highest PSNR and SSIM scores. The experimental results demonstrate the effectiveness of our component design. Compared to the baseline model, our method produces smoother edges and retains finer details in the corrosion images.



(a) The upcurve of PSNR (b) The upcurve of SSIM

Figure 10. PSNR and SSIM curves from ablation experiments on different components.

Table 5. Ablation studies for the proposed modules on blurry metal corrosion dataset.

Baseline	SCAM	FCAM	PSNR	SSIM	Params(M)	FLOPs(G)
✓	✗	✗	31.5345	0.9745	17.0	18.53
✓	✓	✗	32.3106	0.9762	22.3	18.54
✓	✗	✓	32.6760	0.9767	28.2	21.02
✓	✓	✓	32.8645	0.9768	32.7	21.02

4.4.2. Study of Loss Functions

To validate the effectiveness of our multiple loss functions, we conduct a series of ablation experiments. Table 6 presents the contributions of each component in the loss function, demonstrating that every part enhances the model's performance. When utilizing only the spatial domain loss function, the PSNR is 32.0136 dB. The inclusion of either the frequency domain loss or edge loss individually results in varying degrees of improvement in both PSNR and SSIM. This is because the frequency domain loss improves the network's capacity to recover high-frequency information, resulting in clearer texture details. Simultaneously, the edge loss preserves sharp image contours and reduces blurring. When employing the total loss function, the PSNR and SSIM reach their highest values. These results indicate that our loss function plays a crucial role in deblurring metal corrosion images, enabling the model to more accurately reconstruct the corroded structures and edge contours.

Table 6. Verify the effectiveness of each loss in MCIDN.

L_s	L_f	L_e	PSNR	SSIM
✓	✗	✗	32.0136	0.9755
✓	✗	✓	32.2078	0.9758
✓	✓	✗	32.4522	0.9763
✓	✓	✓	32.8645	0.9768

5. Limitations and Future Work

Although our method achieves high performance on blurry metal corrosion datasets, it is still constrained by the dataset quality, and the restoration effect on certain corrosion images remains suboptimal. Our future work will focus on developing domain-adaptive algorithms to fully exploit the potential of synthetic blurry metal corrosion datasets for real-world applications. These algorithms aim to bridge the gap between synthetic and real-world data, enhancing the generalization capability and robustness of deblurring models when applied to actual corrosion scenarios. This approach has the potential to significantly improve the practicality and effectiveness of image deblurring techniques in industrial environments.

6. Conclusions

In this paper, we propose a deblurring network for metal corrosion images, called MCIDN. Specifically, we explore the potential of channel interaction in deblurring corrosion images. SCAM generates spatial attention weights and enhances different channels in an efficient convolutional manner, identifying useful aggregated features. Additionally, FCAM refines the spectrum using frequency decomposition and modulation with simpler frequency techniques, integrating useful frequencies with frequency attention weights. To achieve better deblurring performance, we train the model with multiple loss function. The final experimental results demonstrate that our method achieves advanced deblurring performance on blurry metal corrosion images. However, the proposed SCAM and FCAM still have limitations that require further improvement. SCAM demonstrates insufficient robustness in complex scenarios, particularly when dealing with intricate corrosion patterns or significant background noise. To address this, incorporating more context-aware mechanisms and multi-scale feature extraction methods could enhance its adaptability to complex corrosion models. FCAM exhibits limited capability in processing certain high-frequency information. This limitation could be mitigated by integrating advanced techniques such as wavelet transforms or high-order spectral decomposition to capture more comprehensive frequency details and improve performance.

The image-deblurring technology for metal corrosion has broad applications across industrial inspection, equipment maintenance, fault prediction, energy infrastructure management, and the safety assessment of buildings and bridges. By enhancing the accuracy and reliability of corrosion detection, this technology provides critical support for extending equipment lifespan, ensuring structural integrity, optimizing material development, and advancing environmental evaluation. Moreover, it serves as a cornerstone for promoting automation and intelligent analysis in corrosion monitoring and assessment.

Author Contributions: Conceptualization, J.W. (Jiaxiang Wang) and P.S.; methodology, J.W. (Jiaxiang Wang) and M.W.; software, P.Z. and S.C.; validation, H.D. and Y.W.; formal analysis, P.S. and J.W. (Jue Wang); investigation, H.Y.; resources, P.S. and D.S.; data curation, J.W. (Jue Wang); writing—original draft preparation, J.W. (Jiaxiang Wang) and P.Z.; writing—review and editing, P.S. and M.W.; visualization, S.C.; supervision, P.S. and D.S.; project administration, J.W. (Jue Wang) and P.S.; funding acquisition, H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key R&D Program of China (Grant No. 2021YFA1601104), National Key R&D Program of China (Grant No. 2022YFA16038004), National Key R&D Program of China (Grant No. 2022YFA16038002), National Key R&D Program of China (Grant No. 2023YFC3805703) and National Science and Technology Major Project of China (Grant No. J2019-VI-0004-0117).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data will be made available on request.

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

Abbreviations

The following abbreviations are used in this manuscript:

MCIDN	metal corrosion images deblurring network
SCAM	spatial channel attention module
FCAM	frequency channel attention module
PSNR	peak signal-to-noise ratio
SSIM	structural similarity
CNNs	Convolutional Neural Networks
FFT	Fast Fourier Transform
LN	Layer Normalization

FCNU	feedforward channel network unit
SFNU	spatial frequency network unit
MSE	mean squared error

References

1. Forkan, A.R.M.; Kang, Y.B.; Jayaraman, P.P.; Liao, K.; Kaul, R.; Morgan, G.; Ranjan, R.; Sinha, S. CorrDetector: A framework for structural corrosion detection from drone images using ensemble deep learning. *Expert Syst. Appl.* **2022**, *193*, 116461. [CrossRef]
2. Zhu, Y.; Wang, L.; Behnamian, Y.; Song, S.; Wang, R.; Gao, Z.; Hu, W.; Xia, D.H. Metal pitting corrosion characterized by scanning acoustic microscopy and binary image processing. *Corros. Sci.* **2020**, *170*, 108685. [CrossRef]
3. Rahman, A.; Wu, Z.Y.; Kalfarisi, R. Semantic deep learning integrated with RGB feature-based rule optimization for facility surface corrosion detection and evaluation. *J. Comput. Civ. Eng.* **2021**, *35*, 04021018. [CrossRef]
4. Zhang, Z.; Zheng, L.; Piao, Y.; Tao, S.; Xu, W.; Gao, T.; Wu, X. Blind remote sensing image deblurring using local binary pattern prior. *Remote Sens.* **2022**, *14*, 1276. [CrossRef]
5. Cao, S.; Fang, H.; Chen, L.; Zhang, W.; Chang, Y.; Yan, L. Robust blind deblurring under stripe noise for remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 1–17. [CrossRef]
6. Zeng, X.; Dong, Q.; Li, Y. MG-CNFNet: A multiple grained channel normalized fusion networks for medical image deblurring. *Biomed. Signal Process. Control* **2023**, *82*, 104572. [CrossRef]
7. Li, G.; Yang, Y.; Qu, X.; Cao, D.; Li, K. A deep learning based image enhancement approach for autonomous driving at night. *Knowl.-Based Syst.* **2021**, *213*, 106617. [CrossRef]
8. Han, G.; Wang, M.; Zhu, H.; Lin, C. MPDNet: An underwater image deblurring framework with stepwise feature refinement module. *Eng. Appl. Artif. Intell.* **2023**, *126*, 106822. [CrossRef]
9. Dong, J.; Roth, S.; Schiele, B. DWDN: Deep wiener deconvolution network for non-blind image deblurring. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 9960–9976. [CrossRef]
10. Tai, Y.W.; Tan, P.; Brown, M.S. Richardson-lucy deblurring for scenes under a projective motion path. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 1603–1618.
11. Levin, A.; Weiss, Y.; Durand, F.; Freeman, W.T. Understanding and evaluating blind deconvolution algorithms. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, PMiami, FL, USA, 20–25 June 2009 ; IEEE: New York, NY, USA, 2009; pp. 1964–1971.
12. Mairal, J.; Elad, M.; Sapiro, G. Sparse representation for color image restoration. *IEEE Trans. Image Process.* **2007**, *17*, 53–69. [CrossRef] [PubMed]
13. Ren, W.; Cao, X.; Pan, J.; Guo, X.; Zuo, W.; Yang, M.H. Image deblurring via enhanced low-rank prior. *IEEE Trans. Image Process.* **2016**, *25*, 3426–3437. [CrossRef] [PubMed]
14. He, K.; Sun, J.; Tang, X. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 2341–2353. [PubMed]
15. Zhang, H.; Dai, Y.; Li, H.; Koniusz, P. Deep stacked hierarchical multi-patch network for image deblurring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5978–5986.
16. Chu, X.; Chen, L.; Chen, C.; Lu, X. Improving image restoration by revisiting global information aggregation. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 53–71.
17. Kim, Y.; Kwon, H.; Ko, H. Facial image deblurring network for robust illuminance adaptation and key structure restoration. *Eng. Appl. Artif. Intell.* **2024**, *133*, 107959. [CrossRef]
18. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H. Restormer: Efficient transformer for high-resolution image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5728–5739.
19. Ye, J.; Yang, L.; Qiu, C.; Zhang, Z. Joint low-light enhancement and deblurring with structural priors guidance. *Expert Syst. Appl.* **2024**, *249*, 123722. [CrossRef]
20. Tsai, F.J.; Peng, Y.T.; Lin, Y.Y.; Tsai, C.C.; Lin, C.W. Stripformer: Strip transformer for fast image deblurring. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 146–162.
21. Zou, W.; Jiang, M.; Zhang, Y.; Chen, L.; Lu, Z.; Wu, Y. Sdwnet: A straight dilated network with wavelet transformation for image deblurring. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1895–1904.
22. Mao, X.; Liu, Y.; Liu, F.; Li, Q.; Shen, W.; Wang, Y. Intriguing findings of frequency selection for image deblurring. In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 1905–1913.
23. Zhang, Y.; Li, Q.; Qi, M.; Liu, D.; Kong, J.; Wang, J. Multi-scale frequency separation network for image deblurring. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 5525–5537. [CrossRef]
24. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE conference on computer vision and pattern recognition, Miami, FL, USA, 20–25 June 2009; IEEE: New York, NY, USA, 2009; pp. 248–255.

25. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In *Computer Vision—ECCV 2014: Proceedings of the 13th European Conference, Zurich, Switzerland, 6–12 September 2014*; Proceedings, Part V 13; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.
26. Everingham, M.; Van Gool, L.; Williams, C.K.; Winn, J.; Zisserman, A. The pascal visual object classes (voc) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]
27. Zbontar, J.; Knoll, F.; Sriram, A.; Murrell, T.; Huang, Z.; Muckley, M.J.; Defazio, A.; Stern, R.; Johnson, P.; Bruno, M.; et al. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv* **2018**, arXiv:1811.08839.
28. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **2014**, *34*, 1993–2024. [CrossRef]
29. Johnson, A.E.; Pollard, T.J.; Berkowitz, S.J.; Greenbaum, N.R.; Lungren, M.P.; Deng, C.y.; Mark, R.G.; Horng, S. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Sci. Data* **2019**, *6*, 317. [CrossRef]
30. Xia, G.S.; Bai, X.; Ding, J.; Zhu, Z.; Belongie, S.; Luo, J.; Datcu, M.; Pelillo, M.; Zhang, L. DOTA: A large-scale dataset for object detection in aerial images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3974–3983.
31. Liu, Z.; Yuan, L.; Weng, L.; Yang, Y. A high resolution optical satellite image dataset for ship recognition and some new baselines. In Proceedings of the International Conference on Pattern Recognition Applications and Methods, Porto, Portugal, 24–26 February 2017; SciTePress: Setúbal, Portugal, 2017; Volume 2, pp. 324–331.
32. Song, Y.; He, Z.; Qian, H.; Du, X. Vision transformers for single image dehazing. *IEEE Trans. Image Process.* **2023**, *32*, 1927–1941. [CrossRef] [PubMed]
33. Chen, L.; Chu, X.; Zhang, X.; Sun, J. Simple baselines for image restoration. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 17–33.
34. Chen, X.; Li, H.; Li, M.; Pan, J. 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.
35. Liu, Y.F.; Jaw, D.W.; Huang, S.C.; Hwang, J.N. Desnownet: Context-aware deep network for snow removal. *IEEE Trans. Image Process.* **2018**, *27*, 3064–3073. [CrossRef] [PubMed]
36. Li, B.; Ren, W.; Fu, D.; Tao, D.; Feng, D.; Zeng, W.; Wang, Z. Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.* **2018**, *28*, 492–505. [CrossRef] [PubMed]
37. Ancuti, C.O.; Ancuti, C.; Sbert, M.; Timofte, R. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; IEEE: New York, NY, USA, 2019; pp. 1014–1018.
38. Nah, S.; Hyun Kim, T.; Mu Lee, K. Deep multi-scale convolutional neural network for dynamic scene deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3883–3891.
39. Shen, Z.; Wang, W.; Lu, X.; Shen, J.; Ling, H.; Xu, T.; Shao, L. Human-aware motion deblurring. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, South Korea, 27 October–2 November 2019; pp. 5572–5581.
40. Yang, W.; Tan, R.T.; Feng, J.; Liu, J.; Guo, Z.; Yan, S. Deep joint rain detection and removal from a single image. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1357–1366.
41. Tao, X.; Gao, H.; Shen, X.; Wang, J.; Jia, J. Scale-recurrent network for deep image deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA., 18–23 June 2018; pp. 8174–8182.
42. Gao, H.; Tao, X.; Shen, X.; Jia, J. Dynamic scene deblurring with parameter selective sharing and nested skip connections. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3848–3856.
43. 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.
44. Cho, S.J.; Ji, S.W.; Hong, J.P.; Jung, S.W.; Ko, S.J. Rethinking coarse-to-fine approach in single image deblurring. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 4641–4650.
45. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 1833–1844.
46. Kong, L.; Dong, J.; Ge, J.; Li, M.; Pan, J. Efficient frequency domain-based transformers for high-quality image deblurring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Montreal, BC, Canada, 17–24 June 2023; pp. 5886–5895.
47. Li, Y.; Fan, Y.; Xiang, X.; Demandolx, D.; Ranjan, R.; Timofte, R.; Van Gool, L. Efficient and explicit modelling of image hierarchies for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Montreal, BC, Canada, 17–24 June 2023; pp. 18278–18289.
48. Chen, L.; Lu, X.; Zhang, J.; Chu, X.; Chen, C. Hinet: Half instance normalization network for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 182–192.
49. Kupyn, O.; Martyniuk, T.; Wu, J.; Wang, Z. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, South Korea, 27 October–2 November 2019; pp. 8878–8887.

50. Zheng, S.; Zhu, Z.; Zhang, X.; Liu, Z.; Cheng, J.; Zhao, Y. Distribution-induced bidirectional generative adversarial network for graph representation learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 7224–7233.
51. Cui, Y.; Knoll, A. Exploring the potential of channel interactions for image restoration. *Knowl.-Based Syst.* **2023**, *282*, 111156. [CrossRef]
52. Wang, J.; Li, Z.; Shi, P.; Yu, H.; Sun, D. MIDNet: Deblurring Network for Material Microstructure Images. *Comput. Mater. Contin.* **2024**, *79* . [CrossRef]
53. Qin, X.; Wang, Z.; Bai, Y.; Xie, X.; Jia, H. FFA-Net: Feature fusion attention network for single image dehazing. In Proceedings of the AAAI Conference on Artificial Intelligence, New York City, NY, USA, 7–12 February 2020; Volume 34, pp. 11908–11915.
54. Huang, J.; Liu, Y.; Zhao, F.; Yan, K.; Zhang, J.; Huang, Y.; Zhou, M.; Xiong, Z. Deep fourier-based exposure correction network with spatial-frequency interaction. In Proceedings of the European Conference on Computer Vision, Tel Aviv, Israel, 23–27 October 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 163–180.
55. Guo, S.; Yong, H.; Zhang, X.; Ma, J.; Zhang, L. Spatial-frequency attention for image denoising. *arXiv* **2023**, arXiv:2302.13598.
56. Li, C.; Guo, C.L.; Zhou, M.; Liang, Z.; Zhou, S.; Feng, R.; Loy, C.C. Embedding fourier for ultra-high-definition low-light image enhancement. *arXiv* **2023**, arXiv:2302.11831.
57. Mao, X.; Liu, Y.; Shen, W.; Li, Q.; Wang, Y. Deep residual fourier transformation for single image deblurring. *arXiv preprint arXiv:2111.11745* **2021**, *2*, 5.
58. Huang, Y.; Huang, J.; Liu, J.; Yan, M.; Dong, Y.; Lyu, J.; Chen, C.; Chen, S. Wavedm: Wavelet-based diffusion models for image restoration. *IEEE Trans. Multimed.* **2024**, *26*, 7058–7073. [CrossRef]
59. Chen, W.T.; Fang, H.Y.; Hsieh, C.L.; Tsai, C.C.; Chen, I.; Ding, J.J.; Kuo, S.Y. All snow removed: Single image desnowing algorithm using hierarchical dual-tree complex wavelet representation and contradict channel loss. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Nashville, TN, USA, 20–25 June 2021; pp. 4196–4205.
60. Fu, X.; Huang, J.; Ding, X.; Liao, Y.; Paisley, J. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Trans. Image Process.* **2017**, *26*, 2944–2956. [CrossRef]
61. Liu, K.H.; Yeh, C.H.; Chung, J.W.; Chang, C.Y. A motion deblur method based on multi-scale high frequency residual image learning. *IEEE Access* **2020**, *8*, 66025–66036. [CrossRef]
62. Zhu, Z.; Zhang, D.; Wang, Z.; Feng, S.; Duan, P. Spectral Dual-Channel Encoding for Image Dehazing. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *33*, 6236-6248. [CrossRef]
63. Yin, B.; Josselyn, N.; Considine, T.A.; Kelley, J.V.; Rinderspacher, B.; Jensen, R.E.; Snyder, J.F.; Zhang, Z.; Rundensteiner, E.A. Corrosion Image Data Set for Automating Scientific Assessment of Materials. In Proceedings of the BMVC, Online, 22–25 November 2021; p. 102.

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.