Global Information Compensation Network for Image Denoising

Shifei Ding[1,2] , **Qidong Wang**[1,2] , **Lili Guo**[1,2*]

[1]School of Computer Science and Technology, China University of Mining and Technology
[2]Mine Digitization Engineering Research Center of Ministry of Education, China University of Mining and Technology
dingsf@cumt.edu.cn, 1730853234@qq.com, liliguo@cumt.edu.cn

Abstract

In image denoising research, discriminative models have achieved impressive results which mainly owes to the powerful ability of convolutional networks in local feature extraction. However, there is still room for improvement due to insufficient utilization of global information. Although using fully connected layers or increasing network depth can supplement global information, this results in a significant increase in parameters and computational cost. To address these issues, we propose a global information compensation network (GICN) for image denoising in this paper. Firstly, at the shallow network part, we propose a global feature mining block that enhances the network's ability to extract global information by combining non-local blocks and the Fourier transform while improving the interpretability of the model. Secondly, between the encoder and decoder, we propose a cross-scale feature aggregation block to fuse information at different scales. Finally, we employ attention blocks to improve skip connections to better capture long-distance dependencies. Extensive experimental results show that our proposed GICN effectively compensates for global information, achieves a balance between denoising efficiency and effect, and surpasses mainstream methods in multiple benchmark tests.

1 Introduction

As an indispensable information carrier, images play an important role in recording and conveying content. However, due to the influence of multiple complex factors from equipment and the outside world, images will inevitably be mixed with varying degrees of noise. In order to grasp more useful information from images, people are becoming increasingly demanding about image quality. Image denoising, as one of the low-level visual research, indirectly affects the effects of many important visual tasks [Brempong *et al.*, 2022; Cheng *et al.*, 2024] and has practical value in many fields [Li *et al.*, 2023; Cui *et al.*, 2024]. Therefore, research on image denoising technology has always been a significant topic in the field of computer vision and image processing.

In early research, traditional image denoising methods can be broadly categorized into spatial-domain approaches [Hawwar and Reza, 2002; Buades *et al.*, 2005], transform-domain techniques [Starck *et al.*, 2002; Dabov *et al.*, 2007; Maggioni *et al.*, 2012], and sparsity-based methods [Elad and Aharon, 2006; Lu *et al.*, 2015]. While these methods achieved certain levels of noise suppression, they often struggled with increasing image complexity, high computational costs, and limited scalability. Moreover, traditional techniques typically lack robustness against diverse and complex noise patterns, making them less suitable for real-world scenarios. With the advent of deep learning, many of these limitations have been effectively addressed, ushering in a new era of data-driven denoising strategies.

Building on this foundation, convolutional neural network (CNN)-based denoising methods have demonstrated significant advantages. The local receptive field in a convolutional neural network (CNN) is one of the key factors for its effectiveness. The convolutional layer of CNN uses small and local convolution kernels to extract local features of the image. These local receptive fields allow the network to focus on the local structure and texture of the image, helping to capture the changes in noise within local areas, which is very useful for removing local noise. Deep methods such as deep convolutional denoising network (DnCNN) [Zhang *et al.*, 2017a], FFDNet [Zhang *et al.*, 2018], and IRCNN [Zhang *et al.*, 2017b] all reflect this idea when designing.

Although CNNs have a powerful effect in extracting local information, relying solely on local information is not enough to fully remove noise. This limitation arises because the noise in the image is usually not evenly distributed and may be stronger in some areas and weaker in other areas. Moreover, boundary and structural information in the image need to be preserved. These boundaries and structures may span local regions, requiring global information to properly protect them from being mistaken for noise.

To better capture global information, a straightforward approach is to increase the depth of the network. However, this often leads to issues such as vanishing gradients and training instability. In practical applications, excessively deep networks are difficult to train and may suffer from overfitting or convergence problems. A widely adopted and effective alter-

*Corresponding author.

native is to incorporate residual or dense connections, which introduce shortcut paths across layers through skip connections. Recent methods such as RDN [Zhang *et al.*, 2020], and MWDCNN [Tian *et al.*, 2023] have employed this strategy and demonstrated strong performance. Although these connections improve information flow by facilitating gradient propagation and feature reuse, they are not explicitly designed to model long-range dependencies. As a result, their ability to capture global contextual information remains limited, which can hinder performance on denoising tasks requiring holistic image understanding.

The self-attention mechanism in Transformers enables the model to capture global dependencies when processing images, offering a significant advantage in understanding their overall structure and content. Recently, several Transformer-based models—such as Restormer [Zamir *et al.*, 2022], Uformer [Wang *et al.*, 2022], HWformer [Tian *et al.*, 2024], and LIDFormer [Zhou *et al.*, 2024]—have been proposed and have achieved state-of-the-art performance in image restoration tasks. However, a major limitation of these methods lies in their high computational cost and the large number of model parameters, which restricts their practicality in resource-constrained environments.

In this paper, we propose a **G**lobal **I**nformation **C**ompensation **N**etwork (**GICN**) for image denoising, which aims to effectively integrate rich global information from multiple perspectives while remaining computationally efficient. To achieve this, we develop a comprehensive global information compensation strategy. Specifically, non-local mechanisms are employed to capture long-range dependencies between pixels, while Fourier transforms provide complementary global context in the frequency domain. The combination of these two approaches enables a more thorough understanding of the image's structure and content. Furthermore, we adopt an encoder-decoder architecture for its strong feature representation capabilities. Within this framework, the incorporation of dynamic convolutions [Phutke *et al.*, 2023] and residual blocks allows for more flexible and adaptive receptive fields. Additionally, cross-scale feature fusion facilitates the extraction of contextual information across different spatial resolutions. Finally, we enhance global information flow by introducing attention mechanisms into the skip connections between the encoder and decoder. This holistic design enables the model to better leverage global features while maintaining a lightweight structure. In summary, the main contributions of this paper can be summarized as follows:

- We introduce a global feature mining block that combines non-local blocks and Fourier transforms to effectively compensate for global information by capturing long-distance dependencies between pixels and global information in the frequency domain.
- Between the encoder and decoder, we propose a cross-scale feature aggregation block that compensates for contextual global information by fusing features from different scales.
- We dynamically balance local and global information by introducing attention mechanisms in the skip connections between the encoder and decoder, improving the utilization efficiency of global information.
- Extensive experiments demonstrate that our proposed method effectively compensates for global information and achieves state-of-the-art denoising performance on multiple synthetic noise datasets and real-world noise datasets while maintaining high efficiency.

2 Related Work

2.1 CNN Models for Image Denoising

Under the influence of data-driven approaches, CNN-based methods have achieved remarkable success in image denoising and have become mainstream solutions. Early works such as MLP [Burger *et al.*, 2012] and TNRD [Chen and Pock, 2016] demonstrated the potential of learning mappings or image priors for denoising. DnCNN [Zhang *et al.*, 2017a], by integrating residual learning and batch normalization, achieved state-of-the-art performance and established a strong baseline. Building on this, methods like DudeNet [Tian *et al.*, 2021] and BRDNet [Tian *et al.*, 2020b] introduced dual-path designs to enhance learning capacity. A key strength of CNNs lies in their local receptive fields, which enable the extraction of fine-grained details and effective removal of local noise. However, their limited global perception often leads to incomplete noise removal, especially around image boundaries and in complex structures. This highlights the necessity of incorporating global information to enhance denoising quality and contextual understanding.

Increasing the depth of neural networks can indirectly improve their ability to capture global context. Recent advanced methods [Zhang *et al.*, 2020; Liu *et al.*, 2020; Tian *et al.*, 2023] have widely adopted residual and dense connections to enhance feature propagation and facilitate gradient flow. However, these techniques do not explicitly introduce global information. This motivates us to design a dedicated component that can effectively extract and integrate global context into the denoising process.

The encoder-decoder architecture is effective in capturing contextual information within images. The encoder extracts local features and transmits contextual cues to the decoder, facilitating a better understanding of pixel relationships and improving denoising performance. This structure has been widely adopted in existing works such as CBDNet [Guo *et al.*, 2019], SADNet [Chang *et al.*, 2020], DeamNet [Ren *et al.*, 2021], MPRNet [Zamir *et al.*, 2021], and RFFNet [Wang *et al.*, 2024]. Inspired by their success, we also adopt an encoder-decoder framework. However, unlike previous methods, our design places greater emphasis on global information compensation through cross-scale information aggregation.

2.2 Attention Mechanism

Attention mechanisms have become essential components in image restoration due to their ability to selectively emphasize informative features. For instance, ADNet [Tian *et al.*, 2020a] introduces attention modules to help denoising networks focus on noise patterns embedded in complex backgrounds. RIDNet [Anwar and Barnes, 2019] leverages feature attention to capture inter-channel dependencies, while

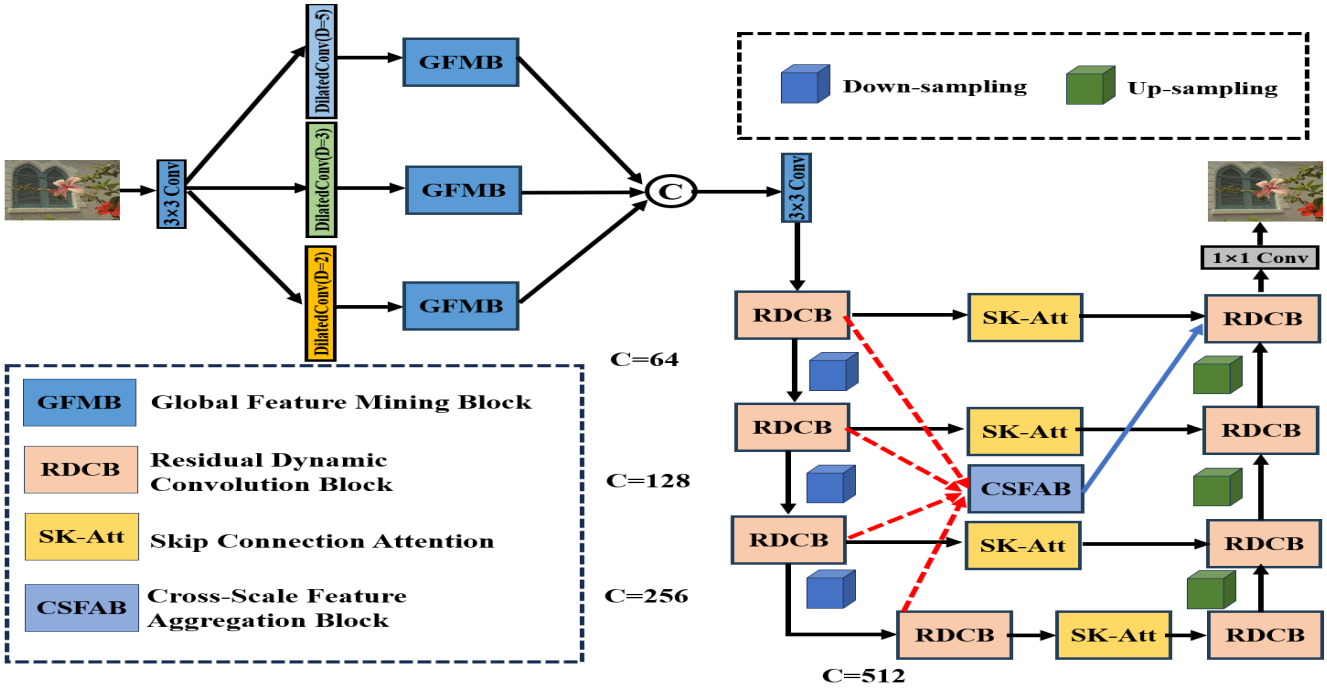

Figure 1: Illustration of our proposed global information compensation network (GICN).

FFA-Net [Qin *et al.*, 2020] combines channel and pixel attention for more effective feature fusion in image dehazing. Recent studies [Wu *et al.*, 2024; Jiang *et al.*, 2025] demonstrate that integrating various forms of attention can significantly enhance the representational capacity of networks and promote multi-perspective feature modeling, which is particularly beneficial for image restoration tasks.

In parallel, Transformer-based models [Kim *et al.*, 2024; Zhou *et al.*, 2024; Tian *et al.*, 2024]—built upon self-attention mechanisms—have shown remarkable capacity for modeling long-range dependencies and global context, which is also crucial in image restoration tasks. Inspired by both CNN-based and Transformer-based attention designs, we incorporate attention mechanisms into the skip connections of our encoder-decoder framework. This enhancement enables better global information flow and more effective integration of low-level and high-level features.

2.3 Leveraging Frequency Domain Representations in Image Restoration

The use of frequency domain representations has emerged as a promising direction in image restoration, offering complementary advantages to spatial-domain learning by capturing global structures and patterns. WINNet [Huang and Dragotti, 2022] and MWDCNN [Tian *et al.*, 2023] further integrate wavelet-based representations into deep networks to improve interpretability and multi-scale feature extraction. In the context of image super-resolution, Fourier-based methods [Fuoli *et al.*, 2021] have demonstrated their effectiveness in modeling global frequency components that are often difficult to capture using standard convolutional operations. Motivated by these findings, we introduce the Fourier transformation as a complementary module to enhance global information modeling. It helps compensate for the limitations of spatial features, particularly in capturing long-range dependencies and structural consistency.

3 Proposed Method

3.1 Overall Network Architecture

The architecture schematic of the Global Information Compensation Network (GICN) proposed by us is shown in Fig. 1. Its primary design objective is to compensate for global information for deep networks from multiple aspects. GICN consists of two main processing sub-networks: global information mining in the shallow part and the primary denoising network. First, we assume the input noisy image is denoted as I_{N}, the processing in the shallow stage as F_s, the processing of the primary denoising network as F_m, and the clean image as I_{C}. The denoising process of GICN can be represented by the following equation.

$$F_{\mathrm{GICN}}\left(I_{\mathrm{N}}\right) = F_m\left(F_s\left(I_{\mathrm{N}}\right)\right). \tag{1}$$

For the loss function part, similar to mainstream methods, we adopt the MSE loss as the model's loss function. The cost function of GICN can be represented by the following equation:

$$J_{\mathrm{GICN}} = \frac{1}{2}\left\|I_{\mathrm{C}} - F_{\mathrm{GICN}}\left(I_{\mathrm{N}}\right)\right\|_2^2, \tag{2}$$

where J_{GICN} represents the cost function of the GICN, I_{N} is the input noisy image, F_{GICN} is the overall processing by GICN, I_{C} is the clean image, and $\|\cdot\|_2$ represents the L2 norm, which measures the difference between the clean image and the output of GICN.

In the shallow part of the network, we first apply 3x3 convolutions to transform the channel dimension to 64. Then, it is passed in parallel into three network branches. In each branch, we perform dilated convolutions with different dilation rates of 2, 3, and 5. This allows us to capture features at different scales, considering both local and global information simultaneously. Subsequently, all three branches go through a global feature mining block for extracting global information, aiding the network in better understanding the global context and semantic information of the input image. Finally, the outputs from the three paths are concatenated using a Concat operation, enabling the network to simultaneously utilize features of different scales and semantics extracted from different branches, providing abundant global features for subsequent subnetworks. The above process can be described by the following equations:

$$X_2^{d2} = \text{DilatedConv}(X_{\text{conv}}, 64, 3, \text{dilation} = 2), \quad (3)$$

$$X_3^{d3} = \text{DilatedConv}(X_{\text{conv}}, 64, 3, \text{dilation} = 3), \quad (4)$$

$$X_5^{d5} = \text{DilatedConv}(X_{\text{conv}}, 64, 3, \text{dilation} = 5), \quad (5)$$

where X_2^{d2}, X_3^{d3}, and X_5^{d5} represent three parallel branches, each performing dilated convolutions with different dilation rates.

$$X_{\text{concat}} = \text{Concat}(F_{\text{global}}(X_2^{d2}), F_{\text{global}}(X_3^{d3}), F_{\text{global}}(X_5^{d5})), \quad (6)$$

where $F_{\text{global}}(X_i^d)$ represents the output of the global feature mining block executed on branch i with dilation rate d.

After processing in the shallow stage, the result is then fed into the main denoising network, which follows an encoder-decoder structure with skip connections. In total, there are 4 encoding and 4 decoding stages. In the encoding stages, the number of channels doubles, and the feature maps are downscaled, aiding in feature representation learning. In the decoding stages, the number of channels halves, and the feature maps are upscaled, achieved through downsampling and upsampling operations. It is noteworthy that, unlike conventional methods, we utilize Residual Dynamic Convolution Blocks (RDCB) instead of regular residual blocks, as shown in Fig. 2. We introduce dynamic convolutions combined with residual blocks to provide a flexible receptive field, which helps in capturing both local and global features effectively. The working process of the RDCB block can be described by the following equations:

$$X_{\text{dynamic}} = \text{PReLU}(\text{DynamicConv}(X, 3 \times 3)), \quad (7)$$

$$X_{\text{output}} = X + \text{Conv}(X_{\text{dynamic}}, 3 \times 3), \quad (8)$$

where X_{dynamic} represents the output of the PReLU activation applied to the 3x3 dynamic convolution. The final output feature map X_{output} is obtained by adding the result of the regular convolution with X_{dynamic} to the input feature map X.

In the main denoising network, two other crucial improvements are cross-scale feature aggregation and the incorporation of channel attention mechanism into skip connections. The former assists the model in capturing pixel dependencies at different scales while compensating for global information. The latter helps the network enhance its utilization of global information.

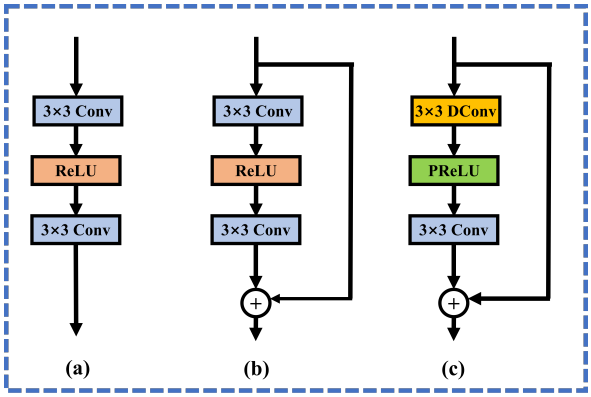

Figure 2: Structure illustration of Residual Dynamic Convolution Block, (a) Represents a Regular Convolution Block, (b) Represents a Standard Residual Convolution Block, (c) Represents the Residual Dynamic Convolution Block we designed.

3.2 Global Feature Mining Block

The construction of the Global Feature Mining Block consists of two parallel branches, as illustrated in Fig. 3. One branch handles frequency domain information, while the other utilizes a non-local block to process spatial information. The frequency domain branch first transforms the features into the frequency domain using Fast Fourier Transform (FFT), followed by a 1×1 convolution. Subsequently, the features are converted back to channel features through inverse transformation. The spatial branch employs a non-local block and is processed through three 1×1 convolutions labeled as α, β, and γ. The α and β branches are combined using element-wise multiplication ($\alpha \odot \beta$), and the resulting combination is further fused with the γ branch using the same operation. Finally, the result of the combination of α, β, and γ branches, i.e., $(\alpha \odot \beta \odot \gamma)$ is added to the input features via a residual connection. Ultimately, the information from both the frequency domain and spatial branches is combined through addition. These operations enable the organic fusion of spatial and frequency domain information, enhancing the network's perception of global information. The above process, using a dilated convolution branch with a dilation rate of 2, can be described by the following equation:

$$F_{\text{r}} = \text{IFFT}(\text{Conv}_{1\times1}(\text{FFT}(X_2^{d2}))), \quad (9)$$

$$S_{\text{r}} = ((\text{Conv}^{\alpha}(X_2^{d2}) \odot \text{Conv}^{\beta}(X_2^{d2})) \odot \text{Conv}^{\gamma}(X_2^{d2})) + X_2^{d2}, \quad (10)$$

$$Z = F_{\text{r}} + S_{\text{r}}, \quad (11)$$

where FFT represents Fast Fourier Transform, IFFT represents Inverse Fast Fourier Transform, $\text{Conv}^{\alpha}, \text{Conv}^{\beta}$, and Conv^{γ} represent convolution operations with different weights (represented by α, β, γ).

3.3 Cross-Scale Feature Aggregation Block

To better combine features across different scales, we propose a cross-scale feature aggregation block, as shown in Fig. 4. We resize multiple features generated from different encoding layers (H×W×C, H/2×W/2×2C, H/4×W/4×4C,

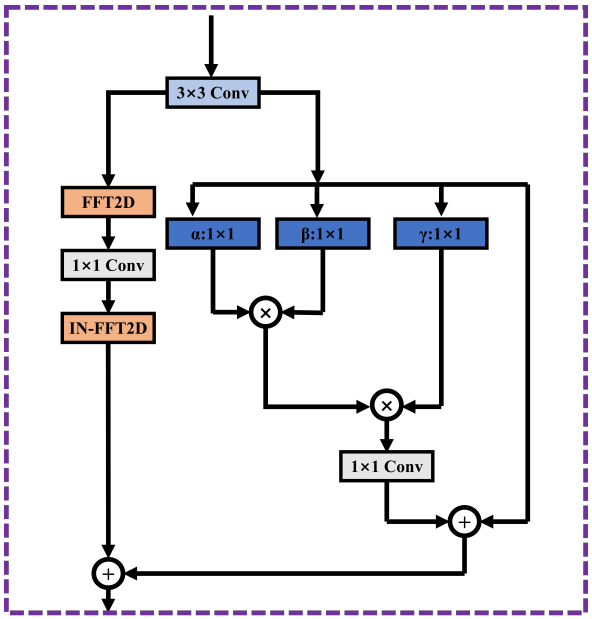

Figure 3: Illustration of Global Feature Mining Block (GFMB).

H/8×W/8×8C) to a unified scale (H×W×C). This process involves reshaping and using convolution operations to adjust the resolution and channel count of the feature maps. Then, by employing additive merging of these features, we can maintain stable feature dimensions while capturing image details and contextual information from local to global scales, thereby enhancing the model's generalization ability. Finally, feeding these cross-scale merged features into the last layer of the decoder enables the decoder to better reconstruct detailed information in the image. These operations can be represented by the following equations:

$$F_s' = \text{Conv}(\text{Reshape}(F_s, H \times W), C), \tag{12}$$

$$F_{\text{merged}} = \sum_s F_s', \tag{13}$$

$$F_{\text{output}} = \text{Decoder}(F_{\text{merged}}), \tag{14}$$

where F_s' represents the process of adjusting feature maps at each scale s to a unified scale $H \times W \times C$. F_{merged} describes the merging of all adjusted feature maps

3.4 Skip Connection Attention Block

In the Skip Connection stage, to enhance the network's utilization efficiency of global information, we introduced a channel attention mechanism, as shown in Fig. 5. Firstly, we perform average pooling on the concatenated results of the corresponding layers from the encoder and decoder. Then, we flatten the result through a 1x1 convolution, and finally, we generate weights through the sigmoid function. This can be represented by the following equation:

$$W = \text{sigmoid}(\text{Conv}(\text{AvgPool}(\text{ConcatRes}))), \tag{15}$$

where W represents the Weight, $Conv$ denotes the 1×1 convolution operation, $AvgPool$ stands for average pooling, and $ConcatRes$ signifies the concatenated results from the encoder and decoder layers.

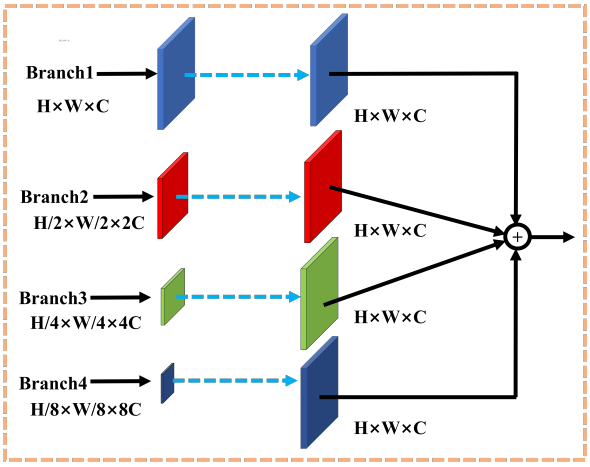

Figure 4: Illustration of Cross-Scale Feature Aggregation Block (CSFAB).

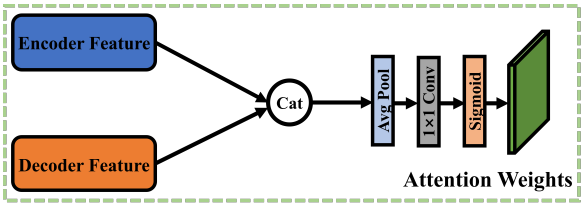

Figure 5: Illustration of Skip Connection Attention.

4 Experiments and Analysis

4.1 Experimental Setup

Training and test datasets. In this paper, we used a composite dataset for Gaussian denoising, consisting of 400 images from [Martin *et al.*, 2001a], 800 from [Timofte *et al.*, 2017], and 3000 from [Ma *et al.*, 2016]. Grayscale and color versions were used separately to train corresponding denoising models. All images were cropped into 96×96 patches with random rotations for data augmentation. For real-world denoising, we trained on a combination of SIDD [Abdelhamed *et al.*, 2018] and RENOIR [Anaya and Barbu, 2018], using 256×256 image patches.

In the experiment, we employed four datasets for testing synthetic noise: BSD68 [Martin *et al.*, 2001b], Kodak24 [Russakovsky *et al.*, 2015], McMaster [Zhang *et al.*, 2011], and CBSD68 [Martin *et al.*, 2001b]. For real image denoising, we conducted testing using three datasets: SIDD [Abdelhamed *et al.*, 2018], Nam [Nam *et al.*, 2016], and PolyU [Xu *et al.*, 2018].

Implementation Details. We implemented GICN using PyTorch 1.10 on an NVIDIA RTX 3090 GPU. The model was trained for 100 epochs with a batch size of 128 using the Adam optimizer. The initial learning rate was set to 1e-4 and decayed to 1e-5 after 30 epochs.

4.2 Gaussian Image Denoising

To validate the effectiveness of our proposed GICN, we conducted comparisons with ten state-of-the-art denoising methods previously introduced. We evaluated performance on four

Dataset	σ	DnCNN	IRCNN	FFDNet	RIDNet	SADNet	DudeNet	DeamNet	WINNet	MWDCNN	DRANet	GICN
BSD68	15	31.73	31.63	31.64	31.81	31.68	31.78	31.91	31.72	31.77	31.79	**31.92**
	25	29.23	29.15	29.19	29.34	29.31	29.29	29.44	29.27	29.28	29.36	**29.46**
	50	26.23	26.19	26.29	26.40	26.51	26.31	**26.54**	29.27	26.29	26.47	26.51
Kodak24	35	30.64	30.43	30.56	30.70	30.76	30.69	30.89	30.73	30.88	30.90	**30.93**
	50	29.02	28.81	28.99	29.15	29.17	29.10	N/A	29.18	29.27	29.50	**29.61**
	75	27.58	27.39	27.25	27.35	27.48	27.39	N/A	27.44	27.57	27.59	**27.69**
McMaster	35	30.91	30.59	30.76	30.81	30.94	30.99	31.09	30.83	31.05	31.08	**31.21**
	50	29.21	28.91	29.14	29.37	29.43	29.12	29.68	29.26	N/A	29.77	**29.81**
	75	27.35	27.32	27.29	27.39	27.48	N/A	N/A	27.07	27.37	27.39	**27.54**
CBSD68	35	29.65	29.50	29.57	29.72	29.78	29.71	29.84	N/A	29.81	29.83	**29.87**
	50	28.01	27.86	27.96	28.03	28.07	28.09	28.21	27.99	28.13	28.37	**28.48**
	75	26.42	26.30	26.24	26.36	26.47	26.40	N/A	26.35	26.40	26.45	**26.51**

Table 1: Denoising performance (PSNR) for different noise levels (σ) on BSD68, Kodak24, McMaster, and CBSD68 datasets.

Method	**SIDD**		**Nam**		**PolyU**	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DnCNN	38.56	0.910	34.95	0.885	35.74	0.878
CBDNet	38.68	0.909	40.02	0.969	37.85	0.956
RIDNet	38.71	0.913	39.20	0.973	38.07	0.957
DeamNet	39.47	0.955	40.05	0.974	39.20	0.958
SADNet	39.46	0.957	39.89	0.979	39.35	0.958
MPRNet	39.71	0.958	39.97	0.987	39.98	0.972
Uformer-S	**39.77**	**0.970**	N/A	N/A	39.95	0.976
DRANet	39.50	0.957	39.79	0.974	39.15	0.955
LIDFormer	39.62	0.958	N/A	N/A	N/A	N/A
HWFormer	39.72	0.956	N/A	N/A	39.92	0.965
LAN (Restormer)	N/A	N/A	38.86	0.965	39.30	0.969
GICN	39.73	0.962	**40.25**	**0.987**	**40.38**	**0.981**

Table 2: Average PSNR (dB) and SSIM results of different methods for real-world image denoising on SIDD, Nam, and PolyU datasets.

Metric	DnCNN	DeamNet	SADNet	MPRNet	Uformer	GICN
Params (M)	0.56	2.23	4.77	15.74	20.63	2.98
Time (s)	0.01	0.05	0.13	0.28	0.87	0.03

Table 3: Runtime and parameter comparisons of selected state-of-the-art image denoising methods on 256×256 resolution inputs using an RTX 3090 GPU.

publicly available datasets. For grayscale image denoising, we used the BSD68 dataset with noise levels $\sigma = 15, 25, 50$. For color image denoising, we tested on three datasets under noise levels $\sigma = 35, 50, 75$. The experimental results are summarized in Table 1. Quantitative analysis reveals that GICN consistently outperforms all compared methods across different scenarios, notably surpassing the advanced MWDCNN, SADNet [Chang *et al.*, 2020], and DRANet [Wu *et al.*, 2024] by up to 0.35 dB. Moreover, our method maintains robust denoising performance even under high noise levels.

In addition to quantitative comparisons, we visualized the denoised images obtained from our method and other approaches. The results are shown in Fig. 6. Upon observation, it is evident that our method preserves more textures and edge details compared to the other methods.

4.3 Real-World Image Denoising

We compare our proposed GICN with advanced methods to validate its real noise reduction capability. Specifically, we compare against DnCNN [Zhang *et al.*, 2017a], CBDNet [Anwar and Barnes, 2019], RIDNet [Anwar and Barnes, 2019], DeamNet [Ren *et al.*, 2021], SADNet [Chang *et al.*, 2020], MPRNet [Zamir *et al.*, 2021], Uformer-S [Wang *et al.*, 2022], DRANet [Wu *et al.*, 2024], LIDFormer [Zhou *et al.*, 2024], HWFormer[Tian *et al.*, 2024]and LAN [Kim *et al.*, 2024]. The experimental results are shown in Table 2 and 3. We can conclude that GICN outperforms other state-of-the-art methods on both datasets, notably excelling on the Nam and PolyU datasets. On the SIDD dataset, achieving performance similar to Uformer-S, while significantly reducing the number of parameters compared to Uformer-S, balances both denoising performance and computational resources. In summary, it can be demonstrated that GICN exhibits excellent denoising performance.

4.4 Ablation Studies

We conducted ablation studies to evaluate the individual and combined contributions of the proposed GFMB, CSFAB, and SK-Att modules, as well as to assess the impact of DRCB. To ensure fairness and clarity, experiments were performed under two settings: Gaussian noise with $\sigma = 15$ on the BSD68 dataset, and real-world noise on the Nam dataset. The average PSNR, SSIM, and model parameter counts are reported in Tables 4 and 5.

Effectiveness of GFMB. The impact of GFMB is assessed by conducting experiments involving the removal of this module, as shown in the experimental results in Tables 4 and 5. It can be observed that when GFMB is removed from GICN, the performance decreases from 31.92dB to 31.81dB on the Gaussian synthetic noise test dataset. The change is even more pronounced on the real noise dataset, indicating that this module contributes to performance improvement by providing the model with more comprehensive global information. Furthermore, when GFMB is removed in conjunction with other components, the performance experiences further varying degrees of decline, suggesting its combined interaction with other components.

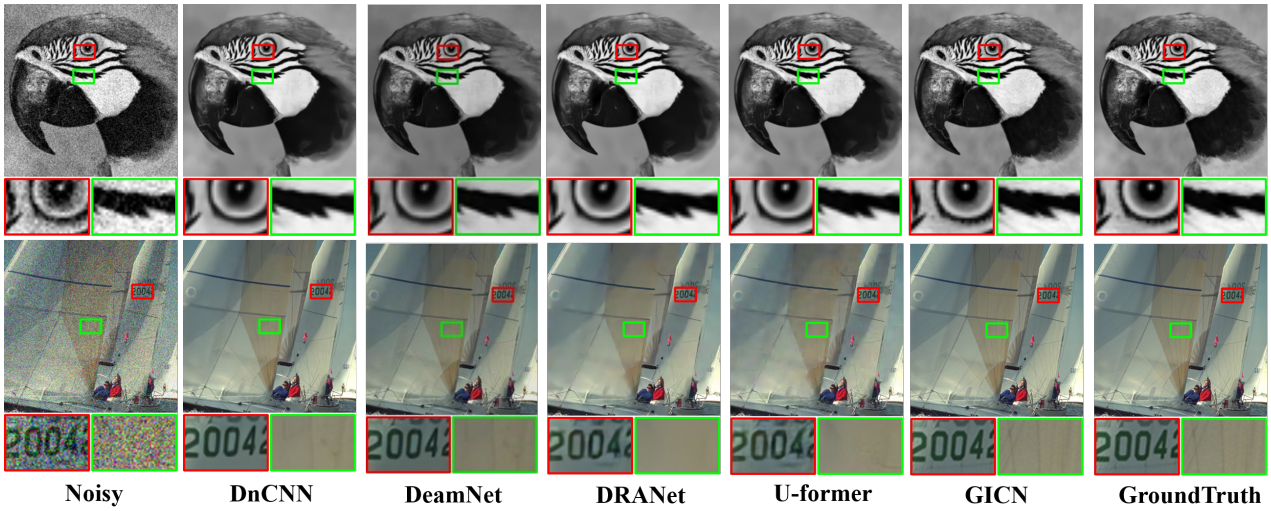

Figure 6: Visual comparisons of various denoising methods. The noise level is $\sigma = 25$.

Effectiveness of CSFAB. Observing Table 5, it can be noted that when we remove CSFAB from GICN, there is a performance decrease of 0.10dB. This indicates that cross-scale feature aggregation plays a positive role in enhancing denoising performance. Similarly, CSFAB, in combination with other components, also plays an important role in enhancing performance.

Effectiveness of SK-Att. SK-Att is a lightweight component designed to improve the utilization of global information in the skip connection part through channel attention mechanisms. According to Tables 4 and 5, SK-Att, at the cost of adding an additional 0.04 million parameters, leads to a performance improvement of 0.06dB on the BSD68 dataset and 0.14dB on the Nam dataset. This demonstrates the efficacy of this module. Furthermore, when we simultaneously remove GFMB, CSFAB, and SK-Att, the performance is the poorest on both test sets, indicating that the combination of these three components is the most effective.

Effectiveness of RDCB. To assess the contribution of RDCB, we replaced it with a standard RCB and evaluated performance under the same conditions. We found that the parameters used by RDCB and RCB were similar, but when using the conventional RCB, GICN's performance decreased by 0.04dB in synthetic noise tasks and 0.08dB in real noise tasks. This indirectly demonstrates that RDCB is superior to RCB in our architecture. Therefore, we use RDCB to build the backbone denoising network.

Model	PSNR	Params
GICN without GFMB	31.81	2.84M
Replace RDCB with RCB	31.88	2.97M
GICN without CSFAB	31.82	2.75M
GICN without SK-Att	31.86	2.94M
GICN without GFMB and CSFAB	31.72	2.61M
GICN without GFMB and SK-Att	31.75	2.80M
GICN without CSFAB and SK-Att	31.77	2.71M
GICN without GFMB, CSFAB, and SK-Att	31.67	2.57M
GICN	**31.92**	**2.98M**

Table 4: Ablation study on the individual and combined impact of key modules in GICN under a noise level of 15 on the BSD68 test dataset.

Model	PSNR	Params
GICN without GFMB	40.09	2.84M
Replace RDCB with RCB	40.17	2.97M
GICN without CSFAB	40.15	2.75M
GICN without SK-Att	40.11	2.94M
GICN without GFMB and CSFAB	39.96	2.61M
GICN without GFMB and SK-Att	39.92	2.80M
GICN without CSFAB and SK-Att	40.02	2.71M
GICN without GFMB, CSFAB, and SK-Att	39.85	2.57M
GICN	**40.25**	**2.98M**

Table 5: Ablation study on the individual and combined impact of key modules in GICN on the real-world denoising dataset Nam.

5 Conclusions

In this paper, we propose a Global Information Compensation Network (GICN) for image denoising, which integrates frequency-domain modeling, multi-scale feature aggregation, and attention mechanisms to better handle complex noise. Specifically, we design a global feature mining module that combines non-local operations with Fourier transforms to capture long-range dependencies; a cross-scale aggregation module to enhance semantic context representation; and attention-based skip connections to dynamically fuse local and global features. Experiments on both synthetic and real-world datasets demonstrate that GICN outperforms existing methods, with ablation studies validating the contribution of each module. Compared to Transformer-based models, GICN offers competitive global modeling with significantly lower computational cost, achieving a better trade-off between performance and efficiency.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant No.62276265 and No.62406326.

References

[Abdelhamed *et al.*, 2018] Abdelrahman Abdelhamed, Stephen Lin, and M.S Brown. A high-quality denoising dataset for smartphone cameras. In *CVPR*, pages 1692–1700, 2018.

[Anaya and Barbu, 2018] Josue Anaya and Adrian Barbu. Renoir–a dataset for real low-light image noise reduction. *Journal of Visual Communication and Image Representation*, 51:144–154, 2018.

[Anwar and Barnes, 2019] Saeed Anwar and Nick Barnes. Real image denoising with feature attention. In *ICCV*, pages 3155–3164, 2019.

[Brempong *et al.*, 2022] Emmanuel Asiedu Brempong, Simon Kornblith, Ting Chen, Niki Parmar, Matthias Minderer, and Mohammad Norouzi. Denoising pretraining for semantic segmentation. In *ICCV*, pages 4175–4186, 2022.

[Buades *et al.*, 2005] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *CVPR*, volume 2, pages 60–65, 2005.

[Burger *et al.*, 2012] Harold C Burger, Christian J Schuler, and Stefan Harmeling. Image denoising: Can plain neural networks compete with bm3d? In *CVPR*, pages 2392–2399, 2012.

[Chang *et al.*, 2020] Meng Chang, Qi Li, Huajun Feng, and Zhihai Xu. Spatial-adaptive network for single image denoising. In *ECCV*, pages 171–187, 2020.

[Chen and Pock, 2016] Yunjin Chen and Thomas Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE transactions on pattern analysis and machine intelligence*, 39(6):1256–1272, 2016.

[Cheng *et al.*, 2024] Jun Cheng, Dong Liang, and Shan Tan. Transfer clip for generalizable image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25974–25984, 2024.

[Cui *et al.*, 2024] Jiaqi Cui, Yan Wang, Luping Zhou, Yuchen Fei, Jiliu Zhou, and Dinggang Shen. 3d point-based multi-modal context clusters gan for low-dose pet image denoising. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[Dabov *et al.*, 2007] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007.

[Elad and Aharon, 2006] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736–3745, 2006.

[Fuoli *et al.*, 2021] Dario Fuoli, Luc Van Gool, and Radu Timofte. Fourier space losses for efficient perceptual image super-resolution. In *ICCV*, pages 2360–2369, 2021.

[Guo *et al.*, 2019] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *CVPR*, pages 1712–1722, 2019.

[Hawwar and Reza, 2002] Yousef Hawwar and Ali Reza. Spatially adaptive multiplicative noise image denoising technique. *IEEE Transactions on Image Processing*, 11(12):1397–1404, 2002.

[Huang and Dragotti, 2022] Jun-Jie Huang and Pier Luigi Dragotti. Winnet: Wavelet-inspired invertible network for image denoising. *IEEE Transactions on Image Processing*, 31:4377–4392, 2022.

[Jiang *et al.*, 2025] Bo Jiang, Jinxing Li, Yao Lu, Qing Cai, Huaibo Song, and Guangming Lu. Eficient image denoising using deep learning: A brief survey. *Information Fusion*, page 103013, 2025.

[Kim *et al.*, 2024] Changjin Kim, Tae Hyun Kim, and Sungyong Baik. Lan: Learning to adapt noise for image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25193–25202, 2024.

[Li *et al.*, 2023] Miaoyu Li, Ying Fu, and Yulun Zhang. Spatial-spectral transformer for hyperspectral image denoising. In *AAAI*, volume 37, pages 1368–1376, 2023.

[Liu *et al.*, 2020] Wei Liu, Qiong Yan, and Yuzhi Zhao. Densely self-guided wavelet network for image denoising. In *CVPRW*, pages 432–433, 2020.

[Lu *et al.*, 2015] Ting Lu, Shutao Li, Leyuan Fang, Yi Ma, and Jón Atli Benediktsson. Spectral–spatial adaptive sparse representation for hyperspectral image denoising. *IEEE Transactions on Geoscience and Remote Sensing*, 54(1):373–385, 2015.

[Ma *et al.*, 2016] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 26(2):1004–1016, 2016.

[Maggioni *et al.*, 2012] Matteo Maggioni, Vladimir Katkovnik, Karen Egiazarian, and Alessandro Foi. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE transactions on image processing*, 22(1):119–133, 2012.

[Martin *et al.*, 2001a] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, volume 2, pages 416–423, 2001.

[Martin *et al.*, 2001b] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, pages 416–423, 2001.

[Nam *et al.*, 2016] Seonghyeon Nam, Youngbae Hwang, Yasuyuki Matsushita, and Seon Joo Kim. A holistic approach to cross-channel image noise modeling and its application to image denoising. In *CVPR*, pages 1683–1691, 2016.

[Phutke *et al.*, 2023] Shruti S Phutke, Ashutosh Kulkarni, Santosh Kumar Vipparthi, and Subrahmanyam Murala. Blind image inpainting via omni-dimensional gated attention and wavelet queries. In *CVPR*, pages 1251–1260, 2023.

[Qin *et al.*, 2020] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *AAAI*, volume 34, pages 11908–11915, 2020.

[Ren *et al.*, 2021] Chao Ren, Xiaohai He, Chuncheng Wang, and Zhibo Zhao. Adaptive consistency prior based deep network for image denoising. In *CVPR*, pages 8596–8606, 2021.

[Russakovsky *et al.*, 2015] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *IJCV*, 115(3):211–252, 2015.

[Starck *et al.*, 2002] Jean-Luc Starck, Emmanuel J Candès, and David L Donoho. The curvelet transform for image denoising. *IEEE Transactions on image processing*, 11(6):670–684, 2002.

[Tian *et al.*, 2020a] Chunwei Tian, Yong Xu, Zuoyong Li, Wangmeng Zuo, Lunke Fei, and Hong Liu. Attention-guided cnn for image denoising. *Neural Networks*, 124:117–129, 2020.

[Tian *et al.*, 2020b] Chunwei Tian, Yong Xu, and Wangmeng Zuo. Image denoising using deep cnn with batch renormalization. *Neural Networks*, 121:461–473, 2020.

[Tian *et al.*, 2021] Chunwei Tian, Yong Xu, Wangmeng Zuo, Bo Du, Chia-Wen Lin, and David Zhang. Designing and training of a dual cnn for image denoising. *Knowledge-Based Systems*, 226:106949, 2021.

[Tian *et al.*, 2023] Chunwei Tian, Menghua Zheng, Wangmeng Zuo, Bob Zhang, Yanning Zhang, and David Zhang. Multi-stage image denoising with the wavelet transform. *Pattern Recognition*, 134:109050, 2023.

[Tian *et al.*, 2024] Chunwei Tian, Menghua Zheng, Chia-Wen Lin, Zhiwu Li, and David Zhang. Heterogeneous window transformer for image denoising. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2024.

[Timofte *et al.*, 2017] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *CVPRW*, pages 114–125, 2017.

[Wang *et al.*, 2022] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *CVPR*, pages 17683–17693, 2022.

[Wang *et al.*, 2024] Qiang Wang, Yuning Cui, Yawen Li, Yaping Ruan, Ben Zhu, and Wenqi Ren. Rffnet: Towards robust and flexible fusion for low-light image denoising. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 836–845, 2024.

[Wu *et al.*, 2024] Wencong Wu, Shijie Liu, Yuelong Xia, and Yungang Zhang. Dual residual attention network for image denoising. *Pattern Recognition*, 149:110291, 2024.

[Xu *et al.*, 2018] Jun Xu, Hui Li, Zhetong Liang, David Zhang, and Lei Zhang. Real-world noisy image denoising: A new benchmark. *arXiv preprint arXiv:1804.02603*, 2018.

[Zamir *et al.*, 2021] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, pages 14821–14831, 2021.

[Zamir *et al.*, 2022] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, pages 5728–5739, 2022.

[Zhang *et al.*, 2011] Lei Zhang, Xiaolin Wu, Antoni Buades, and Xin Li. Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. *Journal of Electronic imaging*, 20(2):023016, 2011.

[Zhang *et al.*, 2017a] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE transactions on image processing*, 26(7):3142–3155, 2017.

[Zhang *et al.*, 2017b] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *CVPR*, pages 3929–3938, 2017.

[Zhang *et al.*, 2018] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018.

[Zhang *et al.*, 2020] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image restoration. *IEEE transactions on pattern analysis and machine intelligence*, 43(7):2480–2495, 2020.

[Zhou *et al.*, 2024] Yubo Zhou, Jin Lin, Fangchen Ye, Yanyun Qu, and Yuan Xie. Efficient lightweight image denoising with triple attention transformer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 7704–7712, 2024.