IPOL IMAGE PROCESSING ON LINE

Published in Image Processing On Line on YYYY–MM–DD.
Submitted on YYYY–MM–DD, accepted on YYYY–MM–DD.
ISSN 2105–1232 © YYYY IPOL & the authors CC–BY–NC–SA
This article is available online with supplementary materials,
software, datasets and online demo at
https://doi.org/10.5201/ipol

QMSANet: A Quaternion Multi-Scale Attention Network for Color Image Denoising

Yi Liu[1], Qi Xie[2], Yu Guo[1], Guoqing Chen[1], Boying Wu[3], Deyu Meng[2], Jean-Michel Morel[4], Qiyu Jin[1], Michael Kwok-Po Ng[5],

[1] Inner Mongolia University, Hohhot, China (liuyi_123456789@126.com; yuguomath@aliyun.com; cgq@imu.edu.cn; qyjin2015@aliyun.com)
[2] Xi'an Jiaotong University, Xi'an, China (xie.qi@mail.xjtu.edu.cn; dymeng@mail.xjtu.edu.cn)
[3] Harbin Institute of Technolog, Harbin, China (mathwby@hit.edu.cn)
[4] City University of Hong Kong, Kowloon Tong, Hong Kong (moreljeanmichel@gmail.com)
[5] Hong Kong Baptist University, Kowloon Tong, Hong Kong (michael-ng@hkbu.edu.hk)

PREPRINT April 17, 2025

Abstract

Color image denoising is a critical task in computer vision, often hindered by the underutilization of inter-channel correlations, resulting in color distortion and loss of fine details. We propose QMSANet, a Quaternion Multi-Scale Attention Network, to address these challenges by leveraging quaternion operations for color image denoising. Operating in the quaternion domain, QMSANet preserves channel dependencies across all processing stages, enhancing noise suppression and detail retention. The network comprises three innovative modules: the Quaternion Multi-Scale Sparse Block (QMSB) for extracting multi-scale features with sparsity enforcement, the Quaternion Stacked Enhancement Block (QSEB) for refining deep features through inter-channel interactions, and the Lightweight Quaternion Attention Block (LQAB) for adaptively focusing on salient features with minimal computational overhead. These modules collectively mitigate color deviation, detail loss, and edge artifacts. Extensive experiments on benchmark datasets demonstrate that QMSANet outperforms state-of-the-art denoising models in both synthetic and real-world noisy conditions. Typically, a blind denoiser exhibits diminished performance in comparison to a non-blind denoiser. However, QMASNet-B, a blind denoiser constructed based on our model, also surpasses most of the comparison models. At $\sigma = 15$, QMASNet and QMASNet-B achieve PSNR improvements of 0.53 dB and 0.50 dB, respectively, compared to the state-of-the-art method on CBSD68. Visual comparisons further highlight its ability to preserve structural details. QMSANet offers a balanced, efficient solution for high-quality color image denoising, with significant potential for real-world applications.

Keywords: Quaternion, Multi-Scale, Attention, Color image denoising

1 Introduction

Color image denoising is crucial in computer vision, as noise from electronic interference, sensor limitations, and environmental factors degrades image quality and impacts subsequent vision tasks [1]. This degradation particularly affects critical applications including object detection [2], autonomous navigation [3], and remote sensing [4]. Image denoising serves as a critical preprocessing step, aiming to remove noise while preserving structural integrity and information content.

Image denoising methods are divided into two main categories: traditional methods and learning-based methods. Traditional methods focus on mathematical models that leverage image properties to differentiate between noise and true image features. Early advancements introduced algorithms like Nonlocal Means (NLM) [5], which utilizes nonlocal self-similarity (NSS) to generate pointwise estimates that reduce noise while preserving structure [6]. The BM3D algorithm [7,8] further employed block matching in the wavelet domain to effectively separate noise from image details, establishing a benchmark for traditional methods [9]. However, these methods exhibit limitations in processing high-dimensional color images and adapting to complex real-world noise patterns, motivating subsequent methodological developments.

The emergence of sparse representation theory revolutionized denoising by focusing on precise modeling of local image structures and efficient handling of high-dimensional data through sparse representations. The K-SVD algorithm [10] and learning-based methods [11] effectively preserved fine image details. Subsequent improvements in sparse coding models, through enhanced data fidelity and regularization terms, further boosted denoising performance [12]. Some specific quaternion-based models, such as pQSTV [13], QWNNM [14], and QNMF [15], leverage their unique mathematical structures to process color images. These methods are effective in reducing the impact of color artifacts and distortions in image recovery. However, traditional models, often reliant on handcrafted mathematical frameworks, struggle in dynamic, real-world scenarios due to high computational costs and limited adaptability to diverse noise patterns.

The rise of deep learning has revolutionized the field of image denoising [16–18], addressing many limitations of traditional approaches by harnessing the powerful learning capabilities of Convolutional Neural Networks (CNNs) and other architectures. DnCNN [19], for instance, reframes image denoising as a noise prediction problem, with a focus on predicting residual images to separate noise from the underlying image. ADNet [20] employs the attention mechanism to circumvent the constraints of network depth, exerting a guiding influence on feature extraction. More advanced approaches like MWDCNN [21] integrate signal processing techniques with dynamic convolutions to optimize noise suppression across diverse image types and depths. Hierarchical networks such as HNN [22] further enhance performance through multi-scale feature extraction that preserves both global structures and local details. Transformers, a recent breakthrough in deep learning, offer compelling solutions for complex noise conditions by effectively modeling long-range dependencies in image data [23, 24]. Models such as CTNet [25], HWformer [26], and SwinIR [27] exemplify the synergy between the local feature extraction strengths of CNNs and the global dependency modeling capabilities of Transformers, leveraging both short- and long-range dependencies to enhance image quality. While deep learning-based approaches have advanced denoising performance, most models process RGB channels independently, failing to exploit inter-channel correlations inherent in color images. This oversight leads to suboptimal noise suppression, color distortion, and loss of fine details—issues exacerbated in complex real-world scenarios.

Building on real-valued CNN-based color image denoising models, we recognize the strong capabilities of end-to-end denoising networks. Recently, Quaternion Convolutional Neural Networks (QCNNs) [28] have leveraged quaternions to enhance CNNs' ability to capture correlations among the three color channels. This advancement highlights the potential of quaternions beyond traditional image processing techniques [29, 30]. However, existing QCNN-based denoising methods [31]

still face two key challenges: (1) Many incorporate traditional optimization frameworks, complicating end-to-end learning and limiting adaptability; (2) Their architectures lack mechanisms to effectively prioritize salient features or adapt to multi-scale noise patterns, leading to computational inefficiency and incomplete detail preservation. Moreover, while attention mechanisms have demonstrated effectiveness in real-valued networks, their integration into quaternion-based models remains underexplored, particularly in balancing computational cost and performance.

To address these limitations, we propose QMSANet, a novel framework comprising three key modules: Quaternion Multi-Scale Sparse Block (QMSB), Quaternion Stacked Enhancement Block (QSEB), and Lightweight Quaternion Attention Block (LQAB). First, QMSANet preserves strong inter-channel correlations—reflected in consistent intensity variations—throughout the denoising pipeline, mitigating distortion. Second, it achieves accurate detail recovery through efficient multi-scale information extraction and deep mining of complex features. By integrating quaternion operations with a lightweight attention mechanism, QMSANet improves noise suppression while maintaining structural fidelity. Finally, our approach balances performance and computational efficiency, providing a practical alternative to state-of-the-art methods that rely on iterative refinement or heavy transformer architectures, which are unsuitable for real-time applications.

In summary, the primary contributions of our paper are as follows:

(i) Innovative QMSANet Framework: QMSANet leverages quaternion operations to enhance color image denoising, preserving inter-channel correlations, reducing color distortion, and improving detail retention across all processing stages.

(ii) Advanced Module Design: QMSB, QSEB, and LQAB modules enable multi-scale feature extraction, deep feature enhancement, and efficient attention, significantly boosting noise suppression and detail preservation in complex scenarios.

(iii) Superior denoising performance: We conducted extensive color image denoising experiments on several widely used datasets, demonstrating that our algorithm surpasses state-of-the-art methods. Specifically, our method achieves a 0.53 dB PSNR improvement on CBSD68 at σ=15, while maintaining competitive computational efficiency.

The remainder of the paper is organized as follows: Section 2 reviews related denoising models and provides background on the proposed methods. The proposed network architecture is then elaborated in Section 3, while Section 4 provides theoretical and experimental analyses of QMSANet. We give the experimental results in Section 5 and conclude the whole paper in Section 6.

2 RELATED WORK

2.1 Deep CNNs for Color Image

Despite significant advancements in CNNs for color images, further research and innovation remain necessary. DnCNN [19] proposes an end-to-end image denoising network based on real-valued convolutions, with its residual learning paradigm widely adopted in subsequent research. As CNN depth increases, the guidance from shallower layers to deeper layers diminishes, often leading to inadequate feature extraction. To address this, Tian et al. [20] proposed an attention mechanism to guide CNNs in feature extraction, effectively mitigating depth-related limitations.

While most CNN models are generally effective, they often overlook differences between color and grayscale images. Grayscale denoising typically targets single-channel features, whereas color images consist of three correlated channels that share structural information. To utilize these inter-channel correlations, Zhu et al. [28] introduced quaternion convolution in CNNs, forming the QCNN

definition. This approach enables interactive computation across the three color channels, enhancing inter-channel feature extraction and improving image restoration. However, to our knowledge, a general and powerful quaternion deep learning network for color image denoising that consistently preserves inter-channel correlations throughout the denoising process remains unexplored.

2.2 Quaternion Algebra and Convolution

The quaternion $\dot{\mathbf{q}} \in \mathbb{H}$ is a kind of hypercomplex number proposed by Hamilton [32], which extends the concept of complex numbers and can be represented as:

$$\dot{\mathbf{q}} = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}, \tag{1}$$

where a, b, c, d are real numbers, $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$ are imaginary units, a is usually considered to be the real part, and b, c, d are the imaginary parts and satisfy the following relation:

$$\begin{aligned} &\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1, \mathbf{ij} = \mathbf{k}, \mathbf{ji} = -\mathbf{k}, \\ &\mathbf{jk} = \mathbf{i}, \mathbf{kj} = -\mathbf{i}, \mathbf{ki} = \mathbf{j}, \mathbf{ik} = -\mathbf{j}. \end{aligned} \tag{2}$$

Quaternion matrices represent color images, enabling interaction among the three channels through quaternion operations. This effectively leverages the strong correlations among the channel information, thereby enhancing image restoration results. Let the input color image be represented by the quaternion matrix $\dot{\mathbf{X}} \in \mathbb{H}^{M\times N}$

$$\dot{\mathbf{X}} = \mathbf{X_0} + \mathbf{X_r}\mathbf{i} + \mathbf{X_g}\mathbf{j} + \mathbf{X_b}\mathbf{k}, \tag{3}$$

where $\mathbf{X_a} \in \mathbb{R}^{M\times N}(\mathbf{a} = \mathbf{0}, \mathbf{r}, \mathbf{g}, \mathbf{b})$. When a quaternion matrix represents a color image, the real part $\mathbf{X_0}$ of the quaternion matrix is set to zero.

The convolution kernel $\dot{\mathbf{W}} \in \mathbb{H}^{M\times N}$ can be expressed in the form of a quaternion matrix:

$$\dot{\mathbf{W}} = \mathbf{W_0} + \mathbf{W_r}\mathbf{i} + \mathbf{W_g}\mathbf{j} + \mathbf{W_b}\mathbf{k}, \tag{4}$$

where $\mathbf{W_t} \in \mathbb{R}^{M\times N}(\mathbf{t} = \mathbf{0}, \mathbf{r}, \mathbf{g}, \mathbf{b})$.

Quaternion convolution operator $\dot{\circledast}$ is defined by

$$\begin{aligned} \dot{\mathbf{F}} &= \dot{\mathbf{W}} \dot{\circledast} \dot{\mathbf{X}}, \\ &= (\mathbf{W_0} \circledast \mathbf{X_0} - \mathbf{W_r} \circledast \mathbf{X_r} - \mathbf{W_g} \circledast \mathbf{X_g} - \mathbf{W_b} \circledast \mathbf{X_b}) \\ &+ (\mathbf{W_0} \circledast \mathbf{X_r} + \mathbf{W_r} \circledast \mathbf{X_0} + \mathbf{W_g} \circledast \mathbf{X_b} - \mathbf{W_b} \circledast \mathbf{X_g})\mathbf{i} \\ &+ (\mathbf{W_0} \circledast \mathbf{X_g} - \mathbf{W_r} \circledast \mathbf{X_b} + \mathbf{W_g} \circledast \mathbf{X_0} + \mathbf{W_b} \circledast \mathbf{X_r})\mathbf{j} \\ &+ (\mathbf{W_0} \circledast \mathbf{X_b} + \mathbf{W_r} \circledast \mathbf{X_g} - \mathbf{W_g} \circledast \mathbf{X_r} + \mathbf{W_b} \circledast \mathbf{X_0})\mathbf{k}, \end{aligned} \tag{5}$$

where $\circledast$ is the real-valued convolution operator, $\dot{\mathbf{F}}$ denotes the output of quaternion convolution, i.e., quaternion feature map [28, 29, 33]. Different from the real-valued convolution, the quaternion convolution is the convolution operation $\dot{\circledast}$ between each component of $\dot{\mathbf{X}}$ and $\dot{\mathbf{W}}$ with each other. Quaternion convolution is a transformative process that involves rotating and scaling to identify the optimal representation of a pixel within a limited color space of a color image [34]. It perfectly describes the relationship between different channels of the color image.

Most existing QCNN-based deep learning models for color image denoising still demonstrate limited performance. In contrast, we propose a powerful quaternion-domain denoising network architecture that shows potential to surpass conventional real-valued approaches.

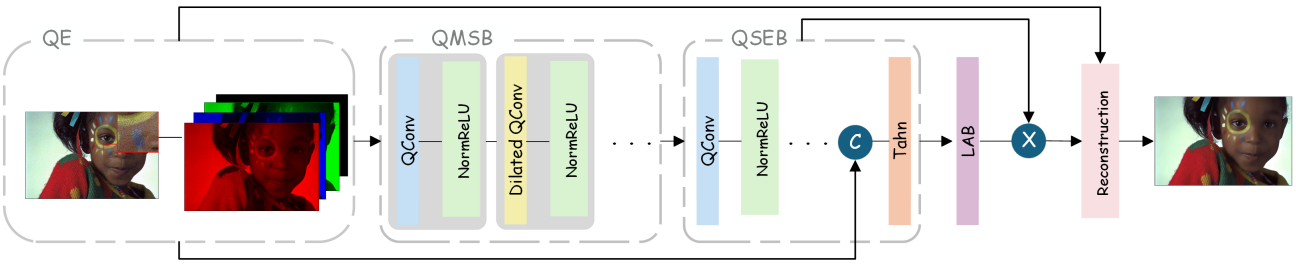


Figure 1: Network architecture of the proposed QMSANet, is composed of five parts: a Quaternion Encoding Block (QE), a Quaternion-based Stacked Enhanced Block (QMSB), a Quaternion-based Stacked Enhanced Block (QSEB), a Lightweight Quaternion Attention Block (LQAB), and a Reconstruction Block (RB).

2.3 SimAM Attention

Identifying key image features is crucial for training an effective image denoising model. Traditional CNNs assign equal weight to all image regions, but attention mechanisms allow the model to focus on important areas, reducing the computational load from irrelevant information and enhancing key feature extraction [35]. A Simple, Parameter-Free Attention Module (SimAM) [36] is an attention mechanism that can be utilized in the domain of computer vision, with a foundation in neuroscientific theories [37]. Unlike many existing attentions [38–40], which are complex and typically focus on either channel or spatial dimensions, SimAM attention considers both spatial and channel dimensions, providing a simpler architecture.

Based on neuroscience theory, the most informative neurons are those that exhibit firing patterns different from their neighbors, and active neurons can inhibit the activity of surrounding neurons. SimAM employs this point by defining an energy function for each neuron in the feature map, with a particular emphasis on those exhibiting significant spatial inhibition effects [41]. The approach effectively acquires the three-dimensional weights of each neuron in the feature map without introducing additional network parameters. Furthermore, it considers both the spatial and channel dimensions, thereby reducing the necessity for extensive network tuning.

In light of this, we propose a computationally efficient lightweight attention block tailored for quaternion-valued feature representations.

3 QMSANet

This section presents the Quaternion Multi-Scale Attention Network (QMSANet), a novel architecture designed for color image denoising. We first outline the overall framework and the role of each component, followed by the loss function definition. Subsequent subsections provide detailed descriptions of the key modules.

3.1 Network Architecture

The QMSANet architecture, depicted in Fig. 1, comprises five core components: the Quaternion Encoding (QE) module, the Quaternion Multi-Scale Sparse Block (QMSB), the Quaternion Stacked Enhancement Block (QSEB), the Lightweight Quaternion Attention Block (LQAB), and the Reconstruction Block (RB). These modules collectively leverage quaternion algebra to preserve inter-channel correlations, extract multi-scale features, and focus on important features.

The QE module transforms the input RGB image into a quaternion representation, enabling subsequent quaternion-based processing. The QMSB employs quaternion convolutions and dilated

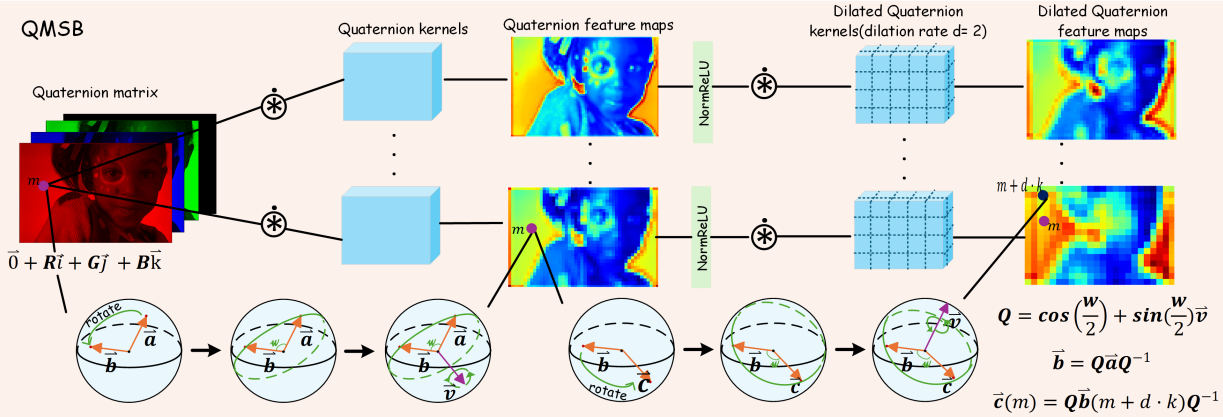


Figure 2: Internal structure diagram of Quaternion Multi-Scale Sparse Block (QMSB).

quaternion convolutions to capture multi-scale features, balancing local detail preservation with global context awareness. The QSEB, a deep feature enhancement module, mitigates information loss in deeper layers through stacked quaternion convolutions and long-path connections. The LQAB adaptively focuses on salient features using a lightweight attention mechanism tailored for quaternion data. Finally, the RB reconstructs the denoised image via residual learning. Detailed explanations of each module follow.

3.2 Loss Function

QMSANet is trained to minimize the mean squared error (MSE) between the predicted residual and the true noise component. The loss function is defined as:

$$L(\theta) = \frac{1}{2N} \sum_{i=1}^{N} ||f_{\text{QMSANet}}(I_N^i) - \widetilde{I}_C^i||^2, \tag{6}$$

where θ represents the trainable parameters, f_{QMSANet} denotes the network's estimated denoised image, N is the number of training samples, $\widetilde{I}_C^i$ and I_N^i are the given i-th clean image with zero-channel-filled real part, and the noisy image, respectively. After obtaining the noise prediction results, perform a residual operation by subtracting them from the input noisy image encoded as a quaternion to obtain an estimated denoised image f_{QMSANet}. Discard the real part padding of this estimated denoised image and retain only the last three R, G, B channels to recover a color denoised image. This formulation guides the network to effectively separate noise from the underlying image content.

3.3 Quaternion Encoding

The Quaternion Encoding (QE) module transforms an input RGB color image $I_{\text{input}} \in \mathbb{R}^{3\times H\times W}$ into a quaternion representation, enabling subsequent processing in the quaternion domain. This step is foundational to QMSANet, as it prepares the conditions for exploiting inter-channel correlations inherent in color images.

The encoding process begins by creating a zero-filled channel [29] with dimensions identical to a single RGB channel ($H \times W$). This zero tensor is concatenated with the three RGB channels along the channel dimension, yielding a four-channel quaternion tensor $O_{QE} \in \mathbb{H}^{H\times W}$, where the dimensions transition from $(3, H, W)$ to $(4, H, W)$. Mathematically, the quaternion representation is

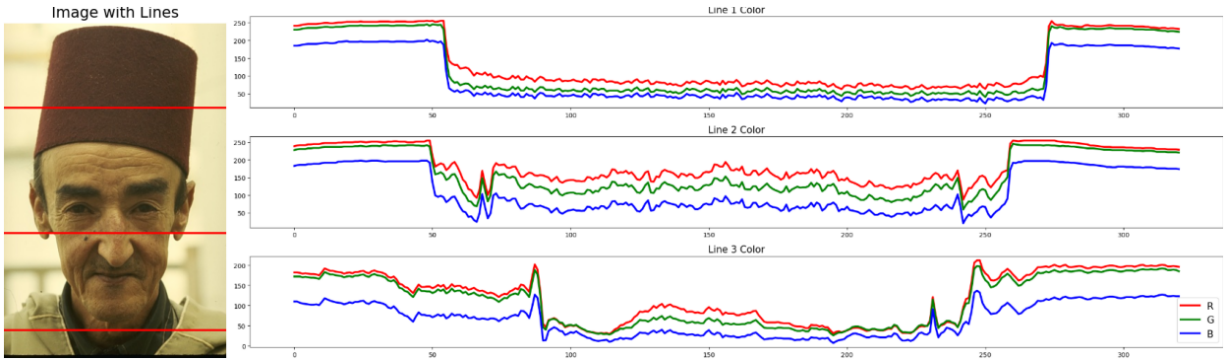
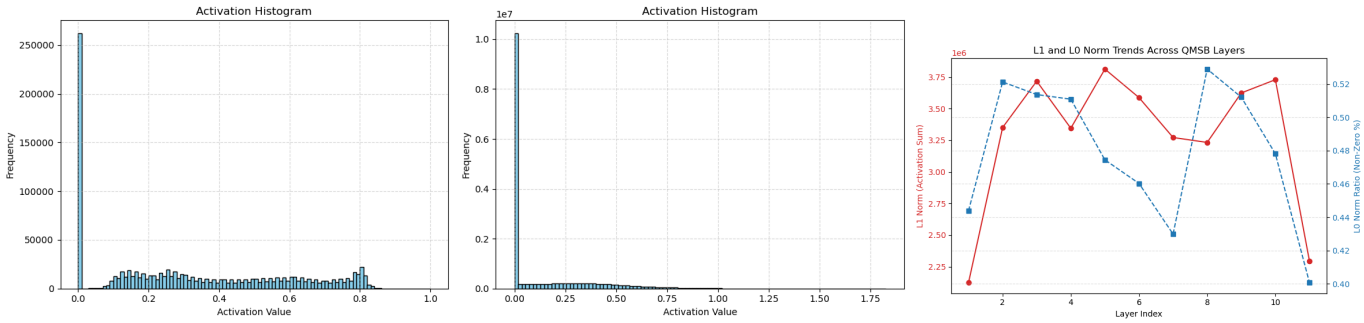


Figure 3: The intensity variation curves of the red, green, and blue channels traverse different regions of the image (cap, eyes, and neck).



(a) (b) (c)

Figure 4: Charting the "sparsity" of QMSB modules using a noisy image from the CC dataset. Subfigures (a) and (b) display the histograms of activation values in feature maps before and after QMSB module, respectively. Subfigure (c) presents the L_1 and L_0 norm curves across different layers of the QMSB module.

expressed as:

$$O_{QE} = \mathbf{0} + R\mathbf{i} + G\mathbf{j} + B\mathbf{k},$$

where R, G and B denote the red, green, and blue channels, respectively, and the real part is set to zero. The QE operation is formalized as:

$$O_{QE} = f_{QE}(I_{input}), \tag{7}$$

where f_{QE} is the encoding function. This transformation preserves the original image content without loss or alteration, standardizing the data for quaternion convolution operations in subsequent modules. By embedding the RGB channels into the imaginary components of a quaternion, QE ensures that inter-channel dependencies are maintained throughout the denoising pipeline, a critical advantage over real-valued representations.

3.4 Quaternion Multi-Scale Sparse Block

The Quaternion Multi-Scale Sparse Block (QMSB) is a pivotal component of QMSANet, designed to exploit the inherent properties of color images—namely, strong inter-channel correlations and information sparsity—to achieve robust denoising. As illustrated in Fig. 2, QMSB operates in the quaternion domain, enabling it to jointly model the dependencies among RGB channels. This addresses the limitations of traditional real-valued convolutions, which treat each channel independently. The

design of QMSB centers around three key functions: modeling inter-channel correlations, extracting multi-scale features, and enforcing sparsity for robust noise suppression.

In color images, intensity variations across RGB channels exhibit high consistency, as illustrated in Fig. 3. QMSB leverages this property by representing the input image as a quaternion matrix. Quaternion convolution, defined in (5), processes these channels holistically, enabling comprehensive inter-channel interactions that enhance feature representation over traditional methods.

To address the multi-scale nature of noise and image structures, QMSB integrates standard quaternion convolution with dilated quaternion convolution. Quaternion convolution, with a 3×3 kernel, captures fine-scale features such as edges and textures, while dilated quaternion convolution, with a dilation rate of 2, expands the receptive field to model large-scale structures and complex noise patterns in textured or flat regions. This alternating pattern—implemented across a 12-layer block with dilated convolutions at layers 2, 5, 9, and 12—ensures complementary extraction of local and global features, improving denoising efficacy. Each layer is followed by BatchNorm2d [42] and ReLU [43], accelerating convergence and enhancing noise-content separation.

QMSB also enforces sparsity through a structured design of convolution operations. The quaternion dilated convolution reduces dependence on locally continuous pixels, improving global feature representation while lowering computational costs. Simultaneously, quaternion convolution preserves fine-scale details, suppressing local noise. Together, these operations enforce spatial sparsity, allowing efficient utilization of multi-scale information. As illustrated in Fig. 4, QMSB modules significantly increase spatial sparsity in feature maps. The activation histograms (a-b) before and after passing through the QMSB module show a significant increase in near-zero activation, while the cross-layer L_1 and L_0 norm trend in sub-figure (c) shows a decrease in activation density but preserves information size. These results empirically validate that QMSB structurally enhances spatial sparsity by suppressing redundant responses and focusing on salient patterns. The consistent intensity variations across channels (Fig. 3) further indicate that holistic channel processing reduces redundancy and enhances sparsity. Channel sparsity is implicitly enforced via quaternion multiplication, which adaptively retains critical inter-channel interactions.

The QMSB process is formalized as:

$$O_{QMSB} = f_{QMSB}(O_{QE}), \tag{8}$$

where f_{QMSB} denotes the QMSB function, transforming the quaternion-encoded input O_{QE} into a multi-scale, sparse feature map O_{QMSB}. This output serves as input to the subsequent Quaternion Stacked Enhancement Block (QSEB).

Compared to prior multi-scale approaches like MWDCNN [21], QMSB uniquely integrates quaternion operations with sparsity constraints, offering superior inter-channel modeling and computational efficiency. Experimental results (Section 5) validate its contribution to QMSANet's state-of-the-art performance, particularly in preserving fine details and suppressing complex noise patterns.

3.5 Quaternion-based Stacked Enhanced Block

As network depth increases, deeper layers often face challenges in effectively leveraging features extracted from shallower layers, leading to information degradation and diminished expressive capacity—a phenomenon termed the "fatigue phase". To mitigate this, we introduce the Quaternion-based Stacked Enhanced Block (QSEB), a novel module designed to enhance feature propagation, stabilize training, and boost denoising performance by integrating multi-scale features and preserving critical low-level information within the quaternion domain.

The QSEB module operates as a feature enhancement bridge, ensuring robust information flow from shallow to deep layers. It comprises three key components: a four-layer quaternion convolutional

stack, a long-path connection, and a Tanh activation layer. The convolutional stack, built with quaternion convolutions, progressively refines features by modeling inter-channel correlations among RGB components. Each of the first three layers is equipped with BatchNorm2d [42] and ReLU [43] to stabilize training and introduce nonlinearity, enabling the capture of complex textures and noise patterns. The fourth layer consolidates these features, preparing them for subsequent processing.

To counter information loss in deeper layers, QSEB incorporates a long-path connection that directly links the noisy input image (via O_{QE}) and the output of the preceding Quaternion Multi-Scale Sparse Block (O_{QMSB}) to the deeper layers. This design preserves both the original noise distribution information and the extracted primary feature information. The incorporation of these information both prevents the deep network from losing the noise distribution prior and corrects the coarse-grained output of the QMSB, which is crucial for the network to more accurately separate signal from noise. By facilitating gradient flow across the network, the long-path connection mitigates vanishing gradient issues, enhancing training stability and feature utilization compared to standard feedforward architectures.

The Tanh activation layer concludes the QSEB. By confining the output range to $[-1, 1]$, it directly ensures that each component of the quaternion remains within the unit range, effectively averting amplitude explosion. This normalization approach furnishes gradients for both positive and negative inputs, preserving the positive-negative relationships among the quaternion components. As a result, it facilitates the QSEB module to further refine the features. Moreover, the bounded nature of the Tanh output is compatible with the attention mechanism of the subsequently connected LQAB module. This compatibility prevents the attention weights from being dominated by extreme values.

Together, these components form a cohesive module that amplifies QMSANet's capacity to recover intricate details while suppressing noise effectively.

The QSEB operation is formalized as:

$$O_{QSEB} = f_{QSEB}(O_{QE}, O_{QMSB}), \tag{9}$$

where f_{QSEB} denotes the QSEB function, integrating the quaternion-encoded input O_{QE} and the multi-scale features O_{QMSB} into an enhanced feature map O_{QSEB}, which is then fed into the Lightweight Quaternion Attention Block (LQAB).

Compared to prior enhancement strategies, such as those in TSP-RDANet [44] or hierarchical networks like HNN [22], QSEB distinguishes itself by leveraging quaternion operations to maintain inter-channel consistency and employing long-path connections to preserve shallow-layer features. Ablation studies (Section 5) confirm that QSEB significantly improves detail retention and noise suppression, contributing to QMSANet's superior performance on benchmark datasets. This module exemplifies a balanced approach to deep feature enhancement, making it particularly effective for complex color image denoising tasks.

3.6 Lightweight Quaternion Attention Block

Attention mechanisms are critical for directing a network's focus toward salient features, yet their integration into quaternion-based models remains underexplored, often incurring high computational costs in traditional designs. To address this, we propose the Lightweight Quaternion Attention Block (LQAB), a novel module tailored for quaternion-valued feature maps, which enhances feature discriminability while minimizing computational overhead in QMSANet.

LQAB processes the quaternion feature map $\dot{\mathbf{X}} \in \mathbb{H}^{C\times M\times N}$ output by the Quaternion Stacked Enhancement Block (QSEB), where C, M, and N denote the number of channels, height, and width, respectively. As shown in Fig. 5, the module comprises two stages: channel compression and attention

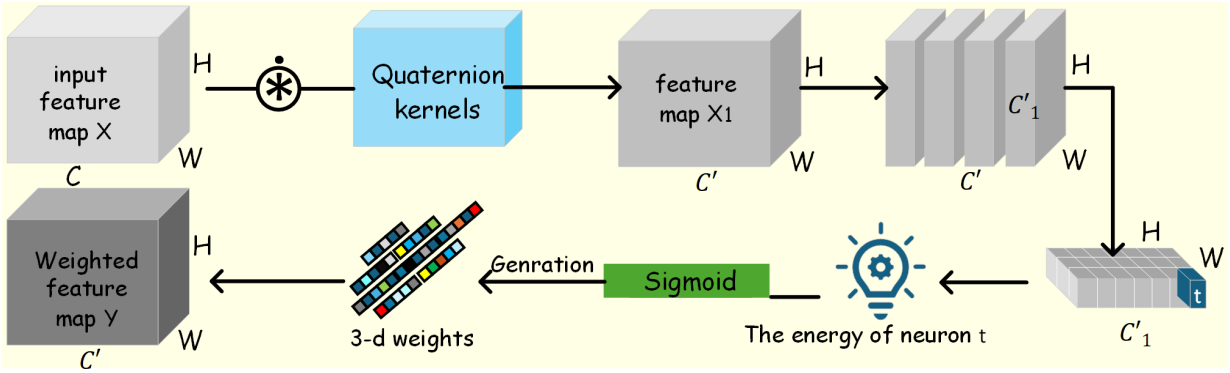


Figure 5: Internal structure diagram of Lightweight Quaternion Attention Block (LQAB).

weighting. In the first stage, a single-layer quaternion convolution reduces the channel dimension from 8 to 4, producing an intermediate feature map $\dot{\mathbf{X}}_1 \in \mathbb{H}^{4\times M\times N}$. This compression, leveraging the quaternion convolution defined in (5), eliminates redundant information while preserving inter-channel correlations, producing a compact yet discriminative representation for subsequent attention computation.

Next, inspired by the SimAM attention mechanism [36], we model the intermediate feature map locally to assess the importance of each pixel in the entire map. For each neuron t in a channel of ${C_1}'$, an energy function quantifies its significance based on spatial suppression effects:

$$e_t(w_t, b_t, y, x_i) = (y_t - \hat{t})^2 + \frac{1}{N-1}\sum_{i=1}^{N-1}(y_0 - \hat{x}_i)^2, \tag{10}$$

where $\hat{t}$ is the target neuron, x_i represents the other neurons in a single channel of the feature map, $N = H \times W$ is the number of spatial positions, and w_t and b_t denote the weight and bias transformations, respectively. Neurons with higher energy values—indicating greater deviation from their surroundings—are assigned higher attention scores, reflecting their importance to the feature map.

The intermediate feature map is then weighted by these attention scores, producing the final output $\dot{\mathbf{Y}} \in \mathbb{H}^{{C_1}'\times H\times W}$. This process amplifies responses in high-energy regions while suppressing those in low-energy areas, enabling QMSANet to focus on salient details and filter noise effectively. The LQAB operation is expressed as:

$$I_R = f_{LQAB}(O_{QSEB}), \tag{11}$$

where f_{LQAB} denotes the LQAB function, transforming the QSEB output O_{QSEB} into a refined residual map I_R for image reconstruction.

LQAB's lightweight nature stems from two design choices. First, by reducing the number of channels, LQAB effectively eliminates redundant or noisy information, which not only reduces computational overhead but also enhances the quality of the extracted features. Second, the attention mechanism relies solely on statistical computations (energy function and weighting), avoiding the parameter-heavy layers typical of methods. This efficiency, combined with quaternion-based processing, distinguishes LQAB from existing attention modules, offering a practical balance of performance and cost.

Compared to real-valued attention mechanisms, LQAB leverages quaternion representations to maintain inter-channel consistency, enhancing detail preservation in color images. As shown in Fig. 6, thermodynamic images generated by the LQAB module highlight key detail regions with prominent

(a) (b)

Figure 6: Thermodynamic images of the LQAB module are shown using some CC dataset images. (The left is the real noise image and the right is the LQAB module's corresponding thermodynamic image.)

red markings, indicating higher attentional weights assigned to these areas. Experimental results (Section 5) further confirm its effectiveness in QMSANet, particularly in complex noise scenarios, demonstrating its role as a lightweight, quaternion-based attention solution for improved denoising performance.

3.7 Reconstruction block

The Reconstruction Block (RB) serves as the final stage of QMSANet, tasked with synthesizing the denoised image from the refined features produced by preceding modules. By leveraging residual learning principles [45], RB integrates the noisy input image with the learned residual map to achieve a balance between noise suppression and detail preservation, a critical requirement for high-fidelity color image denoising.

RB takes as input the residual map $I_R \in \mathbb{R}^{3\times H\times W}$, generated by the Lightweight Quaternion Attention Block (LQAB), which encapsulates the estimated noise component across RGB channels. The block reconstructs the clean image estimate I_{output} through a subtraction operation:

$$I_{\text{output}} = I_{\text{input}} - I_R, \tag{12}$$

where $I_{\text{input}} \in \mathbb{R}^{3\times H\times W}$ is the original noisy RGB image. The residual map I_R is derived from the sequential processing of the input through QMSANet's core modules:

$$I_R = f_{LQAB}\left(f_{QSEB}\left(O_{QE}, f_{QMSB}\left(f_{QE}(I_{\text{input}})\right)\right)\right), \tag{13}$$

where f_{QE}, f_{QMSB}, f_{QSEB}, and f_{LQAB} represent the functions of the Quaternion Encoding (QE), Quaternion Multi-Scale Sparse Block (QMSB), Quaternion Stacked Enhancement Block (QSEB), and LQAB, respectively. This formulation ensures that I_R captures noise patterns while preserving inter-channel correlations encoded in the quaternion domain.

4 Network Analysis

To provide a rigorous theoretical foundation for QMSANet, this section analyzes the core properties of quaternion convolutions underpinning its architecture. We focus on two critical aspects: the translation equivariance of quaternion convolutions, which ensures spatial consistency, and their parameter efficiency, which enhances computational practicality. These properties are substantiated through mathematical derivations and empirical comparisons with real-valued convolutional networks.

4.1 Translation Equivariance of Quaternion Convolution

Translation equivariance is a fundamental property of convolutional operations, ensuring that spatial shifts in the input are preserved in the output—a cornerstone of traditional CNNs [46]. To verify

this for quaternion convolutions, we represent the input feature map as a quaternion-valued function $\dot{f} : \mathbb{Z}^2 \to \mathbb{H}$ over a bounded domain, where for a coordinate $z = (z_1, z_2)$, $\dot{f}(z) = f_a(z) + f_b(z)\mathbf{i} + f_c(z)\mathbf{j} + f_d(z)\mathbf{k}$, with $f_a, f_b, f_c, f_d \in \mathbb{R}$. The quaternion filter is similarly defined as $\dot{\psi}(z) = \psi_a(z) + \psi_b(z)\mathbf{i} + \psi_c(z)\mathbf{j} + \psi_d(z)\mathbf{k}$, where $\psi_i \in \mathbb{R}$ for $i \in \{a, b, c, d\}$.

The quaternion convolution operation is given by:

$$[\dot{f}\dot{\circledast}\dot{\psi}](x) = \sum_{y\in\mathbb{Z}^2} \dot{f}(y) \cdot \dot{\psi}(x - y), \tag{14}$$

where "$\dot{\circledast}$" denotes the quaternion convolution operator, and "$\cdot$" adheres to quaternion multiplication rules (see Section 2.2). Expanding this yields:

$$\begin{aligned}
&\dot{f}(y) \cdot \dot{\psi}(x - y) \\
&= [f_a(y)\psi_a(x - y) - f_b(y)\psi_b(x - y) \\
&\qquad - f_c(y)\psi_c(x - y) - f_d(y)\psi_d(x - y)] \\
&\quad + [f_a(y)\psi_b(x - y) + f_b(y)\psi_a(x - y) \\
&\qquad + f_c(y)\psi_d(x - y) - f_d(y)\psi_c(x - y)]\mathbf{i} \\
&\quad + [f_a(y)\psi_c(x - y) - f_b(y)\psi_d(x - y) \\
&\qquad + f_c(y)\psi_a(x - y) + f_d(y)\psi_b(x - y)]\mathbf{j} \\
&\quad + [f_a(y)\psi_d(x - y) + f_b(y)\psi_c(x - y) \\
&\qquad - f_c(y)\psi_b(x - y) + f_d(y)\psi_a(x - y)]\mathbf{k}.
\end{aligned} \tag{15}$$

For a translation vector $t \in \mathbb{Z}^2$, the translated input is $L_t\dot{f}(y) = \dot{f}(y - t)$. We test equivariance by examining the convolution of the translated input:

$$\begin{aligned}
[[L_t\dot{f}]\dot{\circledast}\dot{\psi}](x) &= \sum_{y\in\mathbb{Z}^2} \dot{f}(y - t) \cdot \dot{\psi}(x - y) \\
&= \sum_{y'\in\mathbb{Z}^2} \dot{f}(y') \cdot \dot{\psi}((x - t) - y') \\
&= [\dot{f}\dot{\circledast}\dot{\psi}](x - t) \\
&= [L_t[\dot{f}\dot{\circledast}\dot{\psi}]](x)
\end{aligned} \tag{16}$$

where substitution $y' = y - t$. This derivation confirms that quaternion convolution is translation-equivariant, mirroring the behavior of real-valued convolution while leveraging quaternion algebra to model inter-channel dependencies more effectively.

4.2 Parameter Efficiency and Comparative Analysis

Quaternion convolution also offers significant parameter efficiency, a key advantage for resource-constrained applications. Consider a real-valued convolution $\text{Conv2d}(C_{\text{in}}, C_{\text{out}}, 3 \times 3)$ with $C_{\text{in}} = C_{\text{out}} = 64$; the parameter count is $64 \times 64 \times 3 \times 3 = 36,864$. In contrast, quaternion convolution distributes channels across four components (real, $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$), reducing the effective channel dimensions to $C_{\text{in}}/4$ and $C_{\text{out}}/4$. For the same $C_{\text{in}} = C_{\text{out}} = 64$, the parameter count is $4 \times (16 \times 16 \times 3 \times 3) = 9,216$, a fourfold reduction due to the shared weight structure across quaternion components.

Table 1 empirically validates this efficiency. With 64 channels, QMSANet using quaternion convolution achieves a PSNR of 32.28 dB on CBSD68 at $\sigma = 25$ with only 0.13M parameters and 0.21G FLOPs, outperforming its real-valued counterpart (31.02 dB) with the same parameter and

computational budget. Increasing the real-valued model's channels to 64 to match the quaternion network leads to a higher cost (0.52M parameters, 0.84G FLOPs) while still yielding a lower PSNR.

These results demonstrate that quaternion convolution not only reduces parameter overhead but also enhances denoising efficacy by leveraging inter-channel correlations. Moreover, it preserves essential geometric properties, such as translation equivariance, while improving both efficiency and performance over real-valued convolutions. This dual advantage positions QMSANet as a compelling advancement in color image denoising, as further evidenced by the experimental results in Section 5.

Table 1: Ablation study of key components. Best results are in bold and second best results are underlined. The same settings are applied to Table 2, 3, 5 and 6.

Model	Channels	CBSD68	Kodak	McMaster	FlOPs	Params
QMSANet	64	**32.28**	**33.27**	<u>33.25</u>	0.21	0.13
QMSANet	128	**32.28**	<u>33.25</u>	**33.39**	0.84	0.52
QMSANet w.o. QCNN	64	31.33	32.39	32.57	0.84	0.52
QMSANet w.o. QCNN	32	31.02	32.01	32.06	0.21	0.13

5 Experiments

This section evaluates our proposed method against 18 baseline methods on four widely used datasets, demonstrating highly satisfactory results. Additionally, we conduct ablation studies to further validate the effectiveness of our approach.

5.1 Experimental Settings

5.1.1 Training datasets

We conduct two primary denoising experiments: one for synthetic colored noise removal and another for real-world colored noise removal. Through repeated experiments, we determine an optimal training set of 4,259 images, comprising 400 images from the Color Berkeley Segmentation Dataset (CBSD) [55] and 3,859 images from WED [56]. Since synthetic noise may not fully capture the diversity of real-world noise patterns, we develop a real-noise denoising model trained on the PolyU dataset [57], which consists of 80 noisy images captured in 40 distinct scenes. To enhance scale-invariant feature learning, we apply double-triple interpolation with reduction factors ranging from 0.7 to 1.0 for dataset expansion. Additionally, eight geometric transformations are used to improve viewpoint invariance and model generalization.

5.1.2 Test datasets

To comprehensively assess model performance, we evaluate it on four benchmark datasets: CBSD68 [55], Kodak24 [58], McMaster [59], and CC [60].

5.1.3 Implementation details

All experiments are conducted on the same computing system, equipped with a 13th Gen Intel Core i9-13900HX CPU and an NVIDIA GeForce RTX 4060 GPU.

Table 2: PSNR (dB) results of different methods on CBSD68, Kodak24, and McMaster.

Methods	CBSD68				Kodak24				McMaster			
	$\sigma=15$	$\sigma=25$	$\sigma=35$	$\sigma=50$	$\sigma=15$	$\sigma=25$	$\sigma=35$	$\sigma=50$	$\sigma=15$	$\sigma=25$	$\sigma=35$	$\sigma=50$
CBM3D [47]	33.44	30.69	29.01	27.35	34.39	31.83	30.23	28.63	34.03	31.67	30.12	28.50
DnCNN [19]	33.90	31.10	29.58	27.79	34.69	32.03	30.46	28.83	34.21	31.73	30.14	28.55
IRCNN [48]	33.87	31.18	29.50	27.88	34.69	32.15	30.55	28.94	34.58	32.18	30.60	28.91
FFDNET [49]	33.88	31.22	29.59	27.97	34.75	32.25	30.69	29.11	34.66	32.36	30.83	29.19
ADNet [20]	33.99	31.31	29.66	28.04	34.76	32.26	30.68	29.10	34.93	32.56	31.00	29.36
AirNet [50]	34.14	31.48	-	28.23	35.09	32.63	-	29.53	35.05	32.79	-	29.66
VIRNet [51]	34.27	31.65	30.04	28.45	35.27	32.86	31.34	29.81	35.32	33.09	31.59	30.02
DRANet [52]	34.18	31.56	29.96	28.37	35.02	32.59	31.06	29.50	35.09	32.84	31.36	29.77
QCNN* [28]	33.85	30.93	-	27.53	34.49	31.96	-	28.73	33.91	31.99	-	28.58
DQRNDL* [31]	34.19	31.79	-	28.80	35.25	33.01	-	30.08	35.38	33.31	-	<u>30.43</u>
TSP-RDANet [44]	34.07	31.45	29.84	28.25	34.87	32.59	31.06	29.50	35.09	32.84	31.36	29.77
DMID-d [53]	34.45	31.86	30.26	28.72	35.51	33.12	31.61	<u>30.14</u>	**35.72**	**33.49**	**32.00**	**30.50**
CFMNet* [54]	34.26	31.64	-	28.46	35.09	32.69	-	29.64	35.23	32.98	-	29.99
CTNet [25]	34.36	31.70	30.06	28.43	35.28	32.82	31.26	29.67	35.54	33.21	31.67	30.02
Ours	**34.98**	**32.33**	**30.66**	**29.06**	**35.81**	**33.33**	**31.71**	**30.16**	**35.72**	<u>33.39</u>	<u>31.77</u>	30.17
Ours-B	<u>34.95</u>	<u>32.28</u>	<u>30.64</u>	<u>28.97</u>	<u>35.78</u>	<u>33.27</u>	<u>31.68</u>	30.03	<u>35.55</u>	33.25	31.70	30.01

The method marked with $*$ indicates that the PSNR values are sourced from the original paper.

Table 3: SSIM results of different methods on CBSD68, Kodak24, and McMaster.

Methods	CBSD68				Kodak24				McMaster			
	$\sigma=15$	$\sigma=25$	$\sigma=35$	$\sigma=50$	$\sigma=15$	$\sigma=25$	$\sigma=35$	$\sigma=50$	$\sigma=15$	$\sigma=25$	$\sigma=35$	$\sigma=50$
CBM3D [47]	0.9269	0.8740	0.8278	0.7701	0.9196	0.8709	0.8289	0.7775	0.9130	0.8707	0.8329	0.7831
DnCNN [19]	0.9312	0.8832	0.8452	0.7833	0.9224	0.8756	0.8390	0.7830	0.9171	0.8757	0.8412	0.7964
IRCNN [48]	0.9285	0.8824	0.8403	0.7998	0.9210	0.8779	0.8398	0.7943	0.9195	0.8818	0.8486	0.8070
FFDNET [49]	0.9290	0.8821	0.8408	0.7887	0.9224	0.8780	0.7952	0.7952	0.9216	0.8862	0.8550	0.8150
ADNet [20]	0.9335	0.8888	0.8487	0.7974	0.9247	0.8826	0.8445	0.7993	0.9286	0.8942	0.8640	0.8245
AirNet [50]	0.9356	0.8928	-	0.8057	0.9288	0.8895	-	0.8121	0.9293	0.8981	-	0.8343
VIRNet [51]	0.9340	0.8918	0.8549	0.8082	0.9289	0.8912	0.8589	0.8186	0.9312	0.9017	0.8759	0.8433
DRANet [52]	0.9326	0.8895	0.8520	0.8048	0.9266	0.8876	0.8551	0.8131	0.9274	0.8964	0.8698	0.8354
QCNN* [28]	0.9264	0.8714	-	0.7671	0.9210	0.8690	-	0.7892	0.9100	0.8796	-	0.7973
DQRNDL* [31]	0.9333	0.8950	-	0.8215	0.9288	0.8931	-	0.8261	0.9331	0.9077	-	**0.8588**
TSP-RDANet [44]	0.9316	0.8878	0.8498	0.8016	0.9247	0.8847	0.8507	0.8080	0.9252	0.8936	0.8662	0.8304
DMID-d [53]	0.9359	0.8952	0.8587	0.8155	0.9309	0.8950	0.8627	<u>0.8270</u>	0.9360	0.9085	<u>0.8839</u>	<u>0.8553</u>
CFMNet* [54]	0.9376	0.8969	-	0.8149	0.9301	0.8927	-	0.8195	0.9336	0.9043	-	0.8453
CTNet [25]	0.9378	0.8963	0.8590	0.8107	0.9309	0.8929	0.8596	0.8170	0.9360	0.9062	0.8795	0.8437
Ours	<u>0.9442</u>	**0.9125**	**0.8763**	**0.8377**	<u>0.9384</u>	**0.9077**	**0.8759**	**0.8416**	**0.9373**	**0.9132**	**0.8849**	0.8545
Ours-B	**0.9446**	<u>0.9085</u>	<u>0.8759</u>	<u>0.8337</u>	**0.9390**	<u>0.9056</u>	<u>0.8759</u>	<u>0.8371</u>	<u>0.9365</u>	<u>0.9086</u>	0.8836	0.8494

The method marked with $*$ indicates that the SSIM values are sourced from the original paper.

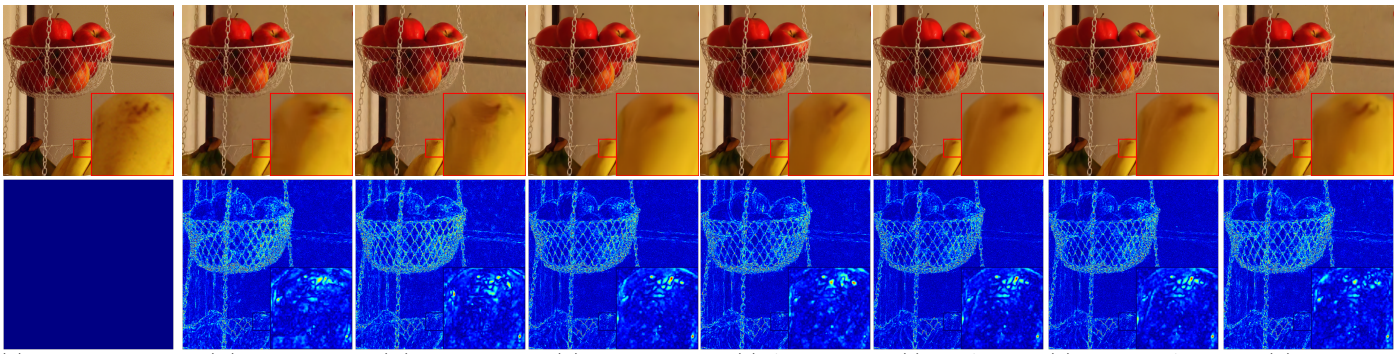
The QMSANet denoising model is trained for 100 iterations with a batch size of 128. The initial learning rate is set to 1×10^{-3} and remains constant for the first 20 epochs. It is then reduced to 1×10^{-4} from epochs 21 to 50 and further decays to 1×10^{-5} in the final 50 epochs.

We optimize the training process using the RMSProp [61] algorithm and employ mean squared error (MSE) [62] as the loss function.

5.2 Color Image Denoising

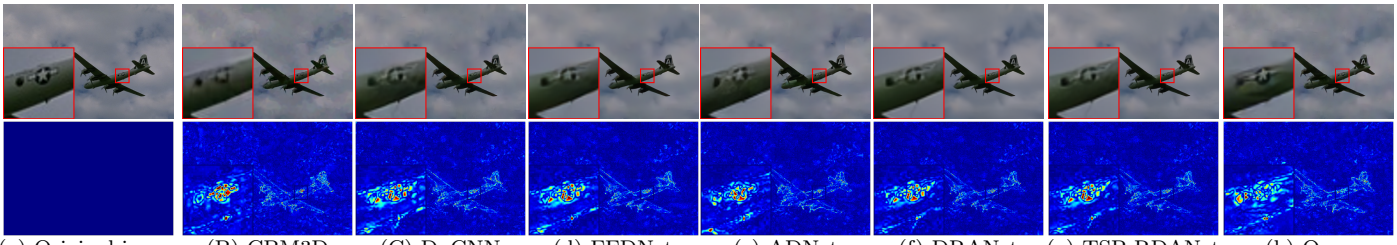
To evaluate the denoising performance of our model on synthetic noisy color images, we compare it with 14 state-of-the-art algorithms across three widely used benchmark datasets. These methods include the traditional CBM3D model [47], classical deep-learning-based denoising models such as DnCNN [19], IRCNN [48], and FFDNet [49], as well as more recent approaches, including ADNet [20], AirNet [50], MWDCNN [21], VIRNet [51], DRANet [52], QCNN [28], DQRNDL [31], TSP-RDANet [44], DMID-d [53], CFMNet [54], and CTNet [25].

We assessed our model comprehensively using peak signal-to-noise ratio (PSNR) [63], structural



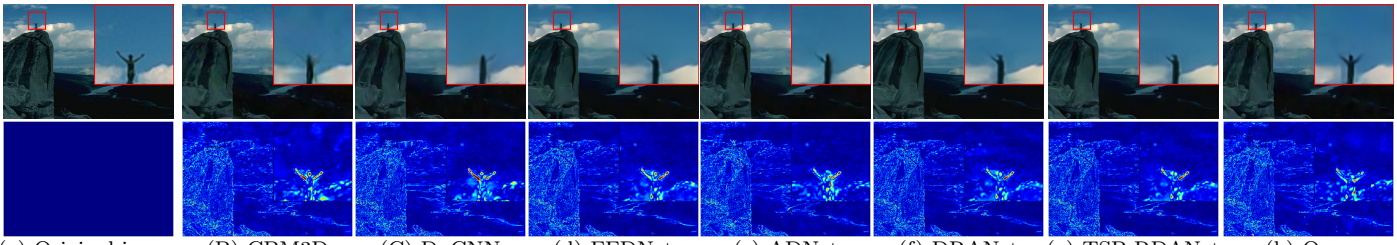
(a) Original image (B) CBM3D 33.72 dB (C) DnCNN 33.60 dB (d) FFDNet 34.44 dB (e) ADNet 34.63 dB (f) DRANet 35.02 dB (g) TSP-RDANet 34.80 dB (h) Ours 35.35 dB

Figure 7: Denoising visualization results of different methods for the image "12" from the McMaster dataset with a noise level of 25.



(a) Original image (B) CBM3D 36.96 dB (C) DnCNN 36.08 dB (d) FFDNet 38.29 dB (e) ADNet 38.39 dB (f) DRANet 39.11 dB (g) TSP-RDANet 38.89 dB (h) Ours 39.22 dB

Figure 8: Denoising visualization results of different methods for the image "3096" from the CBSD68 dataset with a noise level of 35.



(a) Original image (B) CBM3D 31.45 dB (C) DnCNN 31.53 dB (d) FFDNet 31.77 dB (e) ADNet 31.96 dB (f) DRANet 32.23 dB (g) TSP-RDANet 32.13 dB (h) Ours 32.8 dB

Figure 9: Denoising visualization results of different methods for the image "14037" from the CBSD68 dataset with a noise level of 50.

Table 4: Running time of 6 popular denoising methods for the noisy images of sizes 256 × 256, 512 × 512 and 1024 × 1024.

Device	Methods	256*256	512*512	1024*1024
GPU	DnCNN [19]	<u>0.012</u>	<u>0.016</u>	<u>0.034</u>
	IRCNN [48]	0.013	**0.013**	**0.020**
	FFDNET [49]	**0.011**	**0.013**	**0.020**
	ADNet [20]	0.014	0.018	0.039
	AirNet [50]	0.120	0.481	2.182
	VIRNet [51]	0.107	0.477	1.919
	DRANet [52]	0.781	1.652	7.664
	TSP-RDANet [44]	0.459	1.340	4.298
	CTNet [25]	2.408	10.444	49.413
	Ours	0.020	0.022	0.043

similarity (SSIM) [64], and runtime metrics. Notably, we also evaluated a blind denoising variant, QMSANet-B. QMSANet-B model is implemented by assigning each training sample a randomly sampled Gaussian noise level, promoting noise-agnostic adaptation.

Tables 2 and 3 present the PSNR and SSIM results across CBSD68, Kodak24, and McMaster datasets for noise levels $\sigma = 15, 25, 35, 50$. The results indicate that our method outperforms most baseline models in both PSNR and SSIM. For instance, at $\sigma = 15$ on the CBSD68 dataset, QMSANet achieves a 0.53 dB higher PSNR than the leading diffusion-based method [53], while QMSANet-B improves by 0.50 dB. In comparison to the non-blind denoising method QMSANet, which employs the a priori information of the noise, the blind denoiser QMSANet-B often results in a certain loss of denoising performance due to the absence of precise knowledge regarding the noise properties. However, QMSANet-B also surpasses the majority of the comparison models on both PSNR and SSIM metrics, thereby substantiating that our model attains superior denoising capabilities under both blind and non-blind conditions. For $\sigma = 25$ and 35, our model consistently surpasses all baselines in both PSNR and SSIM on CBSD68 and Kodak datasets. Although our method does not achieve the highest overall performance on the McMaster dataset, it excels in specific scenarios. It attains the highest PSNR at $\sigma = 15$ and the best SSIM for $\sigma = 15, 25, 35$, remaining competitive across other cases. DMID-d [53] performs slightly better on some metrics, but requires computationally expensive multi-step iterative sampling. In contrast, our model achieves efficient denoising in a single forward pass, balancing performance and computational efficiency.

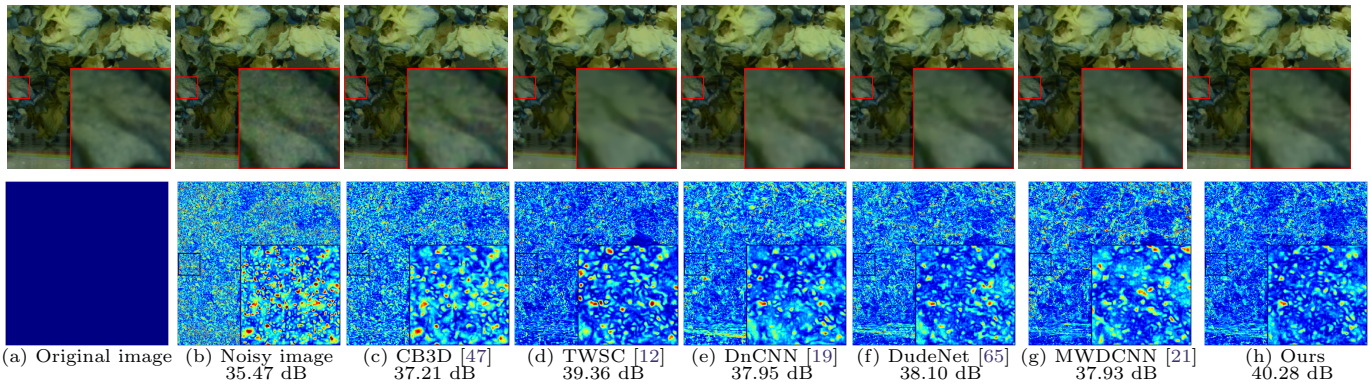
Figs. 7–9 present a visual comparison of denoising performance across methods. The first row of each figure shows denoised color images from competing algorithms, while the second row displays pseudo-color residual images generated by computing absolute differences between noisy inputs and denoised outputs, followed by normalization and pseudo-coloring. In Fig. 7, our method effectively preserves the distinct black spots on bananas within the zoomed region. Fig. 8 demonstrates superior reconstruction of aircraft logo details, with our approach accurately recovering the pentagram's sharp edges compared to blurred results from other methods. As shown in Fig. 9, our algorithm preserves both arms of the subject, whereas alternatives only partially recover one arm. These qualitative comparisons underscore our method's enhanced structural detail preservation during noise reduction. The results demonstrate improved visual fidelity over existing approaches, positioning our framework as a promising solution for high-quality image denoising.

Table 4 reports the runtime of ten deep-learning-based denoising methods across different image sizes. Each model was evaluated using pre-trained weights at $\sigma = 25$, with the reported values representing the average execution time over ten runs. As shown, DnCNN [19], IRCNN [48], and FFDNet [49] achieve the shortest runtimes, making them ideal for real-time applications. Although our method requires approximately twice the runtime of these models, it remains highly competitive, achieving a substantial PSNR improvement of over 1 dB. In contrast, while AirNet [50], VIRNet [51], DRANet [52], TSP-RDANet [44], and CTNet [25] outperform DnCNN, IRCNN, and FFDNet in

Table 5: PSNR (dB) / SSIM results of different methods on real noisy images.

Dataset	Methods							
	CBM3D [47]	DnCNN [19]	TWSC [12]	ADNet* [20]	DudeNet [65]	MWDCNN [21]	WACAFRN* [66]	Ours
Canon 5D ISO=3200 1	38.99/0.9719	37.26/0.9623	**40.55**/0.9793	35.96/0.9630	36.66/0.9702	36.97/0.9725	40.12/**0.9819**	38.60/0.9790
Canon 5D ISO=3200 2	35.84/0.9479	34.13/0.9467	35.92/0.9368	36.11/0.9398	36.70/0.9477	36.01/0.9491	37.18/0.9602	**37.88**/**0.9644**
Canon 5D ISO=3200 3	35.76/0.9603	34.09/0.9451	33.83/0.9467	34.87/0.9475	35.03/0.9509	34.80/0.9478	37.03/0.9695	**37.35**/**0.9714**
Nikon D600 ISO=3200 1	34.24/0.9327	33.62/0.9370	35.36/0.9584	33.94/0.9452	33.72/0.9381	33.91/0.9497	35.47/0.9565	**36.61**/**0.9655**
Nikon D600 ISO=3200 2	34.99/0.9124	34.48/0.9182	37.09/**0.9569**	34.33/0.9110	34.70/0.9245	34.88/0.9302	37.12/0.9557	**37.50**/0.9553
Nikon D600 ISO=3200 3	36.65/0.9118	35.41/0.9362	41.13/**0.9863**	38.87/0.9351	37.98/0.9364	37.02/0.9370	40.35/0.9840	**41.22**/0.9684
Nikon D800 ISO=1600 1	37.21/0.9411	37.95/0.9575	39.36/0.9678	37.61/0.9458	38.10/0.9591	37.93/0.9572	39.33/0.9681	**40.28**/**0.9732**
Nikon D800 ISO=1600 2	38.00/0.9376	36.08/0.9320	**41.91**/0.9811	38.24/0.9601	39.15/0.9684	37.49/0.9359	41.28/**0.9812**	41.06/0.9767
Nikon D800 ISO=1600 3	36.73/0.9099	35.48/0.9230	38.81/0.9468	36.89/0.9341	36.14/0.9311	38.44/0.9416	39.23/0.9537	**40.44**/**0.9581**
Average	36.49/0.9362	35.39/0.9398	38.22/0.9622	36.31/0.9424	36.46/0.9394	36.38/0.9468	38.57/0.9679	**38.99**/**0.9680**

The method marked with ∗ indicates that the PSNR and SSIM values are sourced from the original paper.



(a) Original image (b) Noisy image 35.47 dB (c) CB3D [47] 37.21 dB (d) TWSC [12] 39.36 dB (e) DnCNN [19] 37.95 dB (f) DudeNet [65] 38.10 dB (g) MWDCNN [21] 37.93 dB (h) Ours 40.28 dB

Figure 10: Ablation visualization results of real image denoising of "d800_iso1600_1" image from CC dataset by different methods.

denoising quality, they incur a 10 to 200-fold increase in computational cost, yet still fall short of our method's performance. This underscores the efficiency of our approach, which not only achieves a strong balance between runtime and performance but also delivers superior image restoration quality.

The experimental results demonstrate that our proposed color image denoising model excels across various noise intensities and image datasets. The superior performance metrics validate the model's robustness, computational efficiency and broad applicability.

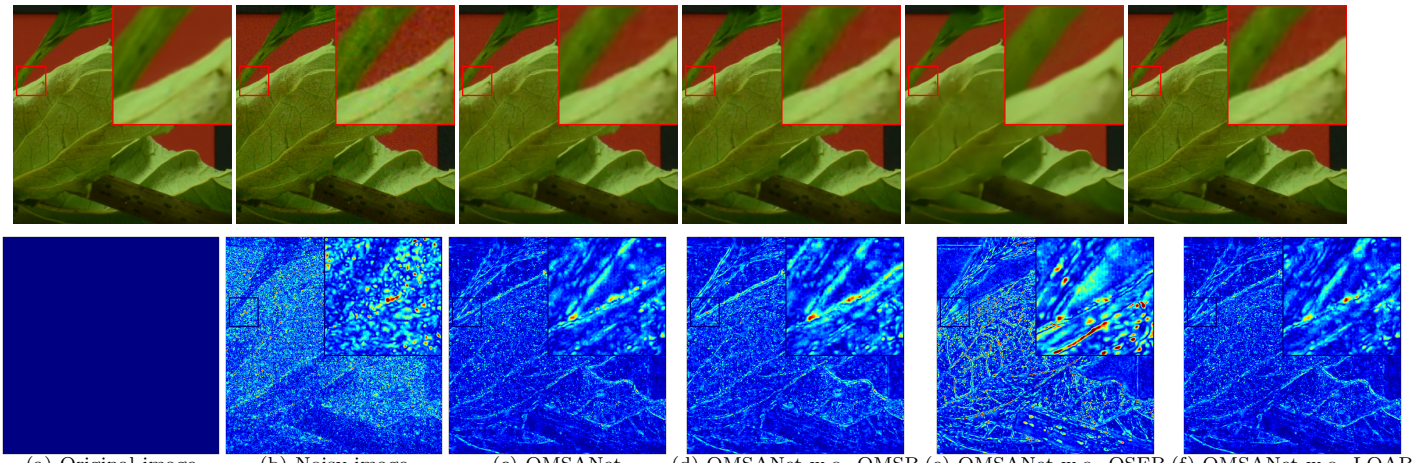
5.3 Experimental Analysis on Real-World Denoising

On the real-world noisy image CC dataset [60], we compare with traditional denoising models, including CBM3D [47] and TWSC [12], as well as state-of-the-art deep learning-based models such as DnCNN [19], ADNet [20], DudeNet [65], MWDCNN [21], and WACAFRN [66]. Notably, we rigorously evaluate all quantitative results for the baseline models to ensure reliability and provide meaningful comparisons.

Table 5 presents the PSNR and SSIM results for denoised images from a subset of the CC dataset. Our method achieves outstanding PSNR performance on specific images. For instance, the "Nikon D600 ISO=3200 3" image attains a PSNR of 41.22 dB, indicating minimal distortion and high-quality denoising. Similarly, our approach excels in SSIM, demonstrating effective structural preservation. Specifically, the "Nikon D800 ISO=1600 1" image achieves an SSIM of 0.9732, closely resembling the original structure. On average, our method surpasses all state-of-the-art models by 0.42 dB in PSNR, highlighting its superior denoising capability while maintaining image integrity. Compared to other approaches, our method offers distinct advantages for practical applications.

Fig. 10 illustrates QMSANet's effectiveness in real-world denoising. The zoom in region reveals that our method removes noise while preserving fine petal details without introducing artifacts. Al-

though TWSC [12] achieves a comparable PSNR, its residual images exhibit frequent red-highlighted areas, indicating significant detail loss during reconstruction. In contrast, QMSANet's minimal light-blue residuals confirm superior structure preservation.



(a) Original image (b) Noisy image 36.06 dB (c) QMSANet 40.44 dB (d) QMSANet w.o. QMSB 39.32 dB (e) QMSANet w.o. QSEB 36.89 dB (f) QMSANet w.o. LQAB 40.28 dB

Figure 11: Visualization results of real image denoising of "d800_iso1600_3" image from CC dataset by different methods.

5.4 Ablation Study

To dissect the contribution of each QMSANet component, we conducted a systematic ablation study using the CC dataset as a baseline. We evaluated the impact of progressively removing the Quaternion Multi-Scale Sparse Block (QMSB), the Quaternion Stacked Enhancement Block (QSEB), and the Lightweight Quaternion Attention Block (LQAB). Table 6 quantifies the contribution of each module, demonstrating their complementary benefits to denoising performance.

Table 6: Ablation study of key components.

QMSB	QSEB	LQAB	PSNR(dB)	FlOPs(G)	Params(M)
✗	✓	✓	37.71	4.16973	0.02018
✓	✗	✓	36.62	19.68865	0.09544
✓	✓	✗	<u>38.00</u>	27.44648	0.13306
✓	✓	✓	**38.15**	27.44812	0.13307

In the complete QMSANet model, the parameter count increases by only 0.0075% compared to the model without the LQAB. This illustrates that our LQAB effectively improves denoising performance by 0.15 dB while maintaining a lightweight architecture, achieving its design objective.

To assess the role of QMSB, we compare the first and last rows of Table 6. The PSNR values indicate that QMSB is essential for capturing both local details and global structures, leading to more effective noise removal. This is further supported by the pseudo-colored residual images in Fig. 10, which illustrate QMSB's ability to preserve fine details and reduce artifacts. Similarly, to evaluate QSEB, we analyze the second and last rows of Table 6 and examine Fig. 11. Results show that QSEB enhances the network's ability to distinguish signal from noise while mitigating learning fatigue in deeper layers.

These ablation experiments provide strong empirical evidence of the scientific validity and effectiveness of our proposed innovations.

6 Conclusion

We introduce QMSANet, a Quaternion Multi-Scale Attention Network designed for color image denoising, addressing the critical challenge of preserving inter-channel correlations in RGB images. By operating in the quaternion domain, QMSANet ensures consistent channel dependency modeling, effectively minimizing color distortion and detail loss commonly observed in traditional methods. The network integrates three innovative modules: the Quaternion Multi-Scale Sparse Block (QMSB) for multi-scale feature extraction, the Quaternion Stacked Enhancement Block (QSEB) for deep feature refinement, and the Lightweight Quaternion Attention Block (LQAB) for efficient attention allocation. These components collectively enhance noise suppression while retaining fine structural details. Extensive experimental evaluations demonstrate that QMSANet significantly surpasses existing state-of-the-art denoising methods across multiple benchmark datasets in both synthetic and real-noise settings. This work underscores the potential of quaternion-based deep learning for high-quality image restoration, offering a robust and efficient solution for color image denoising challenges.

References

[1] L. Fan, F. Zhang, H. Fan, and C. ming Zhang, "Brief review of image denoising techniques," *Vis. Comput. Ind. Biomed. Art*, vol. 2, 2019.

[2] M. Shi and H. Wang, "Infrared dim and small target detection based on denoising autoencoder network," *Mobile Netw. Appl.*, vol. 25, no. 4, pp. 1469–1483, 2020.

[3] Y. Lee, J. Lee, H. Ahn, and M. Jeon, "Snider: Single noisy image denoising and rectification for improving license plate recognition," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2019, pp. 0–0.

[4] Z. Wang, M. K. Ng, L. Zhuang, L. Gao, and B. Zhang, "Nonlocal self-similarity-based hyperspectral remote sensing image denoising with 3-d convolutional neural network," *IEEE Trans. Geosci. Remote Sensing*, vol. 60, pp. 1–17, 2022.

[5] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, 2005, pp. 60–65 vol. 2.

[6] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Modeling & Simulation*, vol. 7, no. 3, pp. 1005–1028, 2009.

[7] K. Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian, "Image denoising with block-matching and 3d filtering," in *J. Electron. Imaging*, 2006.

[8] J. Guo, Y. Guo, Q. Jin, M. Kwok-Po Ng, and S. Wang, "Gaussian patch mixture model guided low-rank covariance matrix minimization for image denoising," *SIAM J. Imag. Sci.*, vol. 15, no. 4, pp. 1601–1622, 2022.

[9] Y. Hou, J. Xu, M. Liu, G. Liu, L. Liu, F. Zhu, and L. Shao, "Nlh: A blind pixel-level non-local method for real-world image denoising," *IEEE Trans. Image Process.*, vol. 29, pp. 5121–5135, 2019.

[10] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.

[11] J. S. Turek, I. Yavneh, and M. Elad, "On mmse and map denoising under sparse representation modeling over a unitary dictionary," *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3526–3535, 2011.

[12] J. Xu, L. Zhang, and D. Zhang, "A trilateral weighted sparse coding scheme for real-world image denoising," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 20–36.

[13] T. Wu, ChaoyanJin, ZhengmengJia, Zhigang, and M. K. Ng, "Total variation based pure quaternion dictionary learning method for color image denoising," *Int. J. Numer. Anal. Model.*, vol. 19, no. 5, pp. 709–737, 2022.

[14] C. Huang, Z. Li, Y. Liu, T. Wu, and T. Zeng, "Quaternion-based weighted nuclear norm minimization for color image restoration," *Pattern Recognit.*, vol. 128, p. 108665, 2022.

[15] Y. Guo, G. Chen, T. Zeng, Q. Jin, and M. K.-P. Ng, "Quaternion nuclear norm minus frobenius norm minimization for color image reconstruction," *Pattern Recognit.*, vol. 158, p. 110986, 2025.

[16] I. Hong, Y. Hwang, and D. Kim, "Efficient deep learning of image denoising using patch complexity local divide and deep conquer," *Pattern Recognit.*, vol. 96, p. 106945, 2019.

[17] W. Shi, F. Jiang, S. Zhang, R. Wang, D. Zhao, and H. Zhou, "Hierarchical residual learning for image denoising," *Signal Process. Image Commun.*, vol. 76, pp. 243–251, 2019.

[18] J. Li, Z. Zhang, X. Liu, C. Feng, X. Wang, L. Lei, and W. Zuo, "Spatially adaptive self-supervised learning for real-world image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 9914–9924.

[19] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, 2017.

[20] C. Tian, Y. Xu, Z. Li, W. Zuo, L. Fei, and H. Liu, "Attention-guided cnn for image denoising," *Neural Netw.*, vol. 124, pp. 117–129, 2020.

[21] C. Tian, M. Zheng, W. Zuo, B. Zhang, Y. Zhang, and D. Zhang, "Multi-stage image denoising with the wavelet transform," *Pattern Recognit.*, vol. 134, p. 109050, 2022.

[22] A. Joshi, N. Akalwadi, C. Mandi, C. Desai, R. A. Tabib, U. Patil, and U. Mudenagudi, "Hnn: Hierarchical noise-deinterlace net towards image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, June 2024, pp. 3007–3016.

[23] H. Yin and S. Ma, "Csformer: Cross-scale features fusion based transformer for image denoising," *IEEE Signal Process. Lett.*, vol. 29, pp. 1809–1813, 2022.

[24] J. Xiao, X. Fu, F. Wu, and Z.-J. Zha, "Stochastic window transformer for image restoration," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, pp. 9315–9329, 2022.

[25] C. Tian, M. Zheng, W. Zuo, S. Zhang, Y. Zhang, and C.-W. Lin, "A cross transformer for image denoising," *Inf. Fusion*, vol. 102, p. 102043, 2023.

[26] C. Tian, M. Zheng, C.-W. Lin, Z. Li, and D. Zhang, "Heterogeneous window transformer for image denoising," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–12, 2024.

[27] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "Swinir: Image restoration using swin transformer," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2021, pp. 1833–1844.

[28] X. Zhu, Y. Xu, H. Xu, and C. Chen, "Quaternion convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham: Springer International Publishing, 2018, pp. 645–661.

[29] J. Miao, K. I. Kou, Y. Yang, L. Yang, and J. Han, "Quaternion matrix completion using untrained quaternion convolutional neural network for color image inpainting," *Signal Process.*, vol. 221, p. 109504, 2023.

[30] J. Miao and K. I. Kou, "Color image recovery using low-rank quaternion matrix completion algorithm," *IEEE Trans. Image Process.*, vol. 31, pp. 190–201, 2022.

[31] Z. Zhou, Y. Chen, and Y. Zhou, "Simultaneously learning deep quaternion reconstruction and noise convolutional dictionary for color image denoising," *IEEE Trans. Emerg. Top. Comput. Intell.*, pp. 1–14, 2024.

[32] Y. Xu, L. Yu, H. Xu, H. Zhang, and T. Nguyen, "Vector sparse representation of color image using quaternion matrix analysis," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1315–1329, 2015.

[33] Y. Cao, Y. Fu, Z. Zhu, and Z. Rao, "Color random valued impulse noise removal based on quaternion convolutional attention denoising network," *IEEE Signal Process. Lett.*, vol. 29, pp. 369–373, 2022.

[34] G. E. Altamirano-Gómez and C. Gershenson, "Quaternion convolutional neural networks: Current advances and future directions," *ArXiv*, vol. abs/2307.08663, 2023.

[35] A. Zafar, D. Aftab, R. Qureshi, X. Fan, P. Chen, J. Wu, H. Ali, S. Nawaz, S. Khan, and M. Shah, "Single stage adaptive multi-attention network for image restoration," *IEEE Trans. Image Process.*, vol. 33, pp. 2924–2935, 2024.

[36] L. Yang, R.-Y. Zhang, L. Li, and X. Xie, "Simam: A simple, parameter-free attention module for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 11 863–11 874.

[37] M. Carrasco, "Visual attention: The past 25 years," *Vision Res.*, vol. 51, pp. 1484–1525, 2011.

[38] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

[39] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "Eca-net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 11 531–11 539.

[40] S. Woo, J. Park, J.-Y. Lee, and I.-S. Kweon, "Cbam: Convolutional block attention module," *ArXiv*, vol. abs/1807.06521, 2018.

[41] B. S. Webb, N. T. Dhruv, S. G. Solomon, C. Tailby, and P. Lennie, "Early and late mechanisms of surround suppression in striate cortex of macaque," *J Neurosci.*, vol. 25, pp. 11 666 – 11 675, 2005.

[42] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 448–456.

[43] M. Makitalo and A. Foi, "Optimal inversion of the generalized anscombe transformation for poisson-gaussian noise," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 91–103, 2012.

[44] W. Wu, A. Ge, G. Lv, Y. Xia, Y. Zhang, and W. Xiong, "Two-stage progressive residual dense attention network for image denoising," *CoRR*, vol. abs/2401.02831, 2024.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[46] T. Cohen and M. Welling, "Group equivariant convolutional networks," *ArXiv*, vol. abs/1602.07576, 2016.

[47] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[48] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep cnn denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3929–3938.

[49] K. Zhang, W. Zuo, and L. Zhang, "Ffdnet: Toward a fast and flexible solution for cnn-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, 2018.

[50] B. Li, X. Liu, P. Hu, Z. Wu, J. Lv, and X. Peng, "All-in-one image restoration for unknown corruption," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 17 452–17 462.

[51] Z. Yue, H. Yong, Q. Zhao, L. Zhang, D. Meng, and K.-Y. K. Wong, "Deep variational network toward blind image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2024.

[52] W. Wu, S. Liu, Y. Xia, and Y. Zhang, "Dual residual attention network for image denoising," *Pattern Recognit.*, vol. 149, p. 110291, 2024.

[53] T. Li, H. Feng, L. Wang, L. Zhu, Z. Xiong, and H. Huang, "Stimulating diffusion model for image denoising via adaptive embedding and ensembling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 12, pp. 8240–8257, 2024.

[54] J. Du, X. Qiao, Z. Yan, H. Zhang, and W. Zuo, "Flexible image denoising model with multi-layer conditional feature modulation," *Pattern Recognit.*, vol. 152, p. 110372, 2024.

[55] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, 2001, pp. 416–423.

[56] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, H. Yong, H. Li, and L. Zhang, "Waterloo exploration database: New challenges for image quality assessment models," *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 1004–1016, 2016.

[57] J. Xu, H. Li, Z. Liang, D. C. Zhang, and L. Zhang, "Real-world noisy image denoising: A new benchmark," *ArXiv*, vol. abs/1804.02603, 2018.

[58] R. Franzen, "Kodak lossless true color image suite," *source: http://r0k. us/graphics/kodak*, vol. 4, no. 2, p. 9, 1999.

[59] L. Zhang, X. Wu, A. Buades, and X. Li, "Color demosaicking by local directional interpolation and nonlocal adaptive thresholding," *J. Electron. Imaging*, vol. 20, no. 2, pp. 023 016–023 016, 2011.

[60] S. Nam, Y. Hwang, Y. Matsushita, and S. J. Kim, "A holistic approach to cross-channel image noise modeling and its application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1683–1691.

[61] M. C. Mukkamala and M. Hein, "Variants of rmsprop and adagrad with logarithmic regret bounds," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2545–2553.

[62] U. Sara, M. Akter, and M. S. Uddin, "Image quality assessment through fsim, ssim, mse and psnr—a comparative study," *Journal of Computer and Communications*, vol. 7, no. 3, pp. 8–18, 2019.

[63] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of psnr in image/video quality assessment," *Electron. Lett*, vol. 44, no. 13, pp. 800–801, 2008.

[64] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *Proc. Int. Conf. Pattern Recognit.*, 2010, pp. 2366–2369.

[65] C. Tian, Y. Xu, W. Zuo, B. Du, C.-W. Lin, and D. C. Zhang, "Designing and training of a dual cnn for image denoising," *Knowl. Based Syst.*, vol. 226, p. 106949, 2020.

[66] S. Ding, Q. Wang, L. Guo, X. Li, L. Ding, and X. Wu, "Wavelet and adaptive coordinate attention guided fine-grained residual network for image denoising," *IEEE Trans. Circuits Syst. Video Technol.*, 2024.

Image Credits