# Enhanced Pansharpening via Quaternion Spatial-Spectral Interactions

Dong Li, Chunhui Luo, Yuanfei Bao, Gang Yang, Jie Xiao, Xueyang Fu, Zheng-Jun Zha*
University of Science and Technology of China, Hefei, China
dongli6@mail.ustc.edu.cn, zhazj@ustc.edu.cn

## Abstract

*Pansharpening aims to generate high-resolution multispectral (MS) images by fusing panchromatic (PAN) images with corresponding low-resolution MS images. However, many existing methods struggle to fully capture spatial and spectral interactions, limiting their effectiveness. To address this, we propose a novel quaternion-based spatial-spectral interaction network that enhances pansharpening by leveraging the compact representation capabilities of quaternions for high-dimensional data. Our method consists of three key components: the quaternion global spectral interaction branch, the quaternion local spatial structure awareness branch, and the quaternion spatial-spectral interaction branch. The first applies the quaternion Fourier transform to convert multi-channel features into the frequency domain as a whole, enabling global information interaction while preserving inter-channel dependencies, which aids spectral fidelity. The second uses a customized spatial quaternion representation, combined with a window-shifting strategy, to maintain local spatial dependencies while promoting spatial interactions, which helps inject spatial details. The last integrates the two pathways within the quaternion framework to enrich spatial-spectral interactions for richer representations. By utilizing quaternion's multi-dimensional representation and parameter-sharing properties, our method achieves a more compact and efficient cross-resolution, multi-band information integration, significantly improving the quality of the fused image. Extensive experiments validate the proposed method's effectiveness and its superior performance over current SOTA techniques. The code is available at https://github.com/dongli8/QuatPanNet.*

## 1. Introduction

In remote sensing, satellite sensors face physical limitations that prevent the direct acquisition of high-resolution multispectral (MS) images [45]. Pansharpening addresses this by merging low-resolution multispectral (LRMS) images with panchromatic (PAN) images to produce high-resolution multispectral (HRMS) images. As a crucial preprocessing step, pansharpening is widely applied across fields [38, 43]. Conceptually, it is a fusion super-resolution task, focused on incorporating fine spatial details while preserving spectral fidelity, requiring strong inter-channel dependency maintenance and the preservation of subtle local spatial structures.

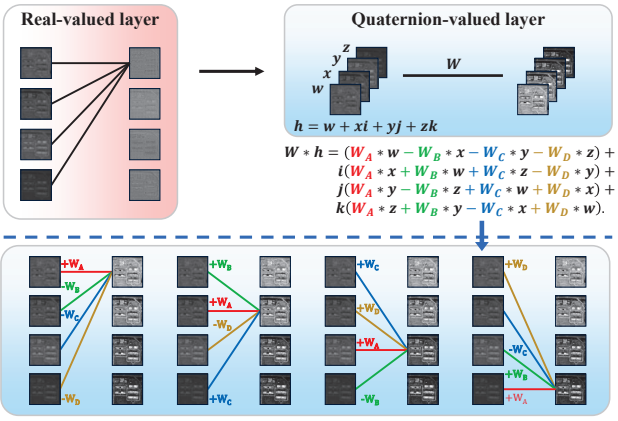


Figure 1. Comparison between standard real-valued layer and quaternion-valued layer. The real-valued operation typically stacks the input multi-channel features, while the quaternion-valued layer, due to quaternion weight sharing, has the ability to learn the latent relationships within the input features.

Traditional methods often rely on mathematical models that integrate spatial and spectral data, typically assuming that the pansharpened image is a linear combination of various spectral bands from a high-resolution multispectral image [40]. However, these approaches are limited by their dependence on prior knowledge, which restricts flexibility and adaptability. With the rise of deep learning, convolutional neural networks (CNNs) have been introduced into pansharpening, leading to substantial breakthroughs in performance with increased model complexity [1, 5, 9, 23].

Despite notable successes, existing methods still struggle to capture the intricate spatial-spectral interactions essential

*Corresponding author.

for high-quality pansharpening. These limitations hinder simultaneous feature interaction and inter-feature dependency preservation, leading to compromises in both spectral fidelity and spatial detail. Unlike conventional real-valued networks, quaternion networks treat multidimensional features as unified entities, allowing them to effectively encode internal dependencies [26, 27]. Additionally, quaternion networks demonstrate strong modeling capabilities for both inter- and intra-correlations within the quaternion latent space [3]. Thus, we aim to harness quaternion operations to further elevate pansharpening performance.

**Our Motivation.** PAN images provide high-resolution spatial details, while MS images offer rich multispectral information [11, 24]. A successful pansharpening model must preserve both spatial details and spectral fidelity during fusion. This requires capturing the unique features of PAN and MS images while leveraging their complementary information to produce accurate and comprehensive fused outputs. Traditional approaches often handle multi-channel inputs by stacking and mapping them to a single output channel, which causes information loss. Consequently, they struggle to capture inter-channel correlations and interactions between spatial and spectral data effectively. To compensate, some methods increase the number of output channels to improve representational capacity [39], but this also increases model complexity. Furthermore, these methods typically extract features in real-number space, limiting their ability to capture directional information such as edges and structures. In contrast, quaternions represent multiple spectral channels as distinct components of a compact quaternion representation, as shown in Figure 1. This approach integrates multi-band information efficiently while reducing information loss through parameter-sharing quaternion operations. These features make quaternions particularly well-suited for pansharpening tasks that involve multi-source data inputs.

Based on the analysis above, we propose a novel quaternion-based pansharpening framework called QuatPanNet. QuatPanNet is designed to balance spectral fidelity and spatial detail richness by leveraging quaternion-based representations and operations to enhance interactions among multispectral, spatial structure, and spatial-spectral features. The framework consists of three core components, i.e., Quaternion Global Spectral Interaction Branch, Quaternion Local Spatial Structure Awareness Branch, and Quaternion Spatial-Spectral Interaction Branch. The first one uses the Quaternion Fourier Transform (QFT) to process multi-channel features holistically in the frequency domain. By enabling global interactions while preserving inter-channel dependencies, it effectively extracts global spectral representations, which are critical for maintaining spectral fidelity. The second one employs a customized spatial quaternion transformation method and a window-shifting strategy, this branch focuses on preserving local structural dependencies and enhancing spatial interactions. It ensures that spatial details are well-integrated while maintaining the structural integrity of the input data. The last one, which built on the outputs of the first two branches, integrates dual-path features into a unified quaternion representation. This integration enhances the complementarity and coherence of information during the spatial-spectral fusion process. By utilizing the compact and efficient representation of PAN and MS image data in quaternion space, QuatPanNet facilitates effective interactions among spatial and spectral features. This approach reduces parameter complexity while achieving high spectral fidelity and rich spatial resolution, leading to superior fusion quality. In summary, the contributions of this work are as follows:

- We propose a novel quaternion-based pansharpening framework, QuatPanNet, designed to comprehensively harness quaternions for this task.
- We design customized quaternion-based spatial and spectral interaction modules for pansharpening. The former facilitates global information interaction while preserving inter-spectral dependencies, and the latter promotes spatial interaction while maintaining local structural dependencies.
- We introduce a quaternion-based spatial-spectral interaction branch, leveraging the high-dimensional representational power of quaternions to improve the fusion of spectral and spatial features.

Extensive experiments on multiple satellite datasets demonstrate that our method achieves superior qualitative and quantitative results with fewer parameters and effectively generalizes to full-resolution scenes.

## 2. Related Works

### 2.1. Pansharpening

Existing pansharpening methods are divided into traditional ones and deep - learning - based ones. Traditional ones mainly consist of Component Substitution (CS) [31, 34], Multi - Resolution Analysis (MRA) [25, 29, 35], and Variational Optimization (VO) [8, 15, 21]. However, their heavy dependence on manually designed features and prior knowledge restricts their applicability [37]. PNN [23] is the first to introduce CNN into the field of pansharpening, achieving better results than traditional methods. Since then, deep learning-based approaches have dominated this field. MS-DCNN [42] combines multi-scale feature extraction and residual learning, enhancing the quality of the fused images. INNformer [47], based on the Transformer architecture, integrates an invertible neural network module for lossless information to model long-range dependencies and efficient feature fusion, successfully introducing the Transformer architecture into this domain. The fusion of frequency do-

main features in images is also a key focus for researchers [36, 48]. For instance, FAME-Net [16] proposes a method that combines MOE with frequency domain information, enabling the network to dynamically learn high-frequency information in remote sensing images. HFIN [33] explores the relationship between PAN and LRMS images using local Fourier information integration. In addition to designing pansharpening models, Zhu *et al*. developed the first upsampling module specifically for pansharpening methods, called PGCU [52].

### 2.2. Quaternion

Quaternions are the natural extension of complex numbers to four dimensions, proposed by Hamilton in 1843 [13]. A quaternion consists of four components, including one real part and three imaginary parts. Quaternions, with their four components, are particularly well-suited for three-dimensional and four-dimensional feature vectors, as seen in applications such as image processing and robotic kinematics [2, 30]. In 2001, Pei *et al*. [28] introduced the quaternion model for color images, marking the beginning of quaternion applications in the field of image processing. The advantages of quaternion wavelet transforms [18], quaternion principal component analysis [44], and other quaternion-based color image processing techniques [41] have been demonstrated to extract more representative features for color images. These techniques have achieved promising results in advanced visual tasks such as color image classification. With the wave of deep learning, many researchers have begun to explore the application of quaternions in the field of deep learning [10, 26]. Xu *et al*. [51] proposes QCNN, which combines quaternions with CNNs, achieving faster convergence of the loss function while preserving color information more effectively than real-valued CNNs. Kusamichi *et al*. [19] studies color night vision and finds that using quaternion networks yields better results in extracting color information from dim images.

## 3. Method

In this section, we first introduce the relevant properties of quaternions and their operators, and then present the proposed quaternion pan-sharpening network, as shown in Figures 2 and 3. We then provide a detailed explanation of the fundamental components of our method, including three quaternion-based branches: a) the global interaction branch utilizes the Quaternion Fourier Transform and quaternion convolution to maintain channel interdepenendies while representing global information; b) the local structure-aware branch models local spatial dependencies through a specially designed spatial quaternion transformation method combined with a window-shifting strategy; and c) the dual-domain interaction branch leverages the orthogonal relationships inherent in quaternions to model dual-domain correlations and promote complementary feature learning. Finally, we present the optimization strategy employed in our approach.

### 3.1. Quaternion Operators

Quaternion, proposed by Hamilton in 1843, extends complex numbers to a four-dimensional space. A quaternion $Q \in \mathbb{H}$ can be expressed as follows [32]:

$$Q = w1 + xi + yj + zk, \tag{1}$$

where $w$, $x$, $y$, $z$ are all real numbers, and 1, $i$, $j$, $k$ are the quaternion unit basis. $w1$ refers to the real component of the number, and $xi + yj + zk$ to the imaginary component. In the definition of quaternions, they must satisfy the following identities, known as the Hamilton rules:

$$i^2 = j^2 = k^2 = ijk = -1. \tag{2}$$

We propose the use of quaternions and their associated operators to uncover both intra- and inter-relations between the features of high spatial resolution PAN and MS images. This approach is inspired by the unique feature-handling capabilities inherent in quaternion representation [3]. Consequently, we introduce two quaternion operators: quaternion convolution and the quaternion Fourier transform.

**Quaternion Convolution**

In the real-valued space, convolution involves convolving a filter matrix with a vector. Extending this operation to quaternions, quaternion convolution is achieved by convolving a quaternion filter matrix with a quaternion vector. To maintain consistency with the defined quaternions, we represent the computation of quaternion convolution using real-valued matrices. This computation is performed by convolving the quaternion filter matrix $W = A+Bi+Cj+Dk$ with the quaternion vector $Q = w+xi+yj+zk$, where $A$, $B$, $C$, and $D$ are real-valued matrices. The convolution operation between $W$ and $Q$ is denoted as:

$$\begin{aligned} W * Q =& (A*w - B*x - C*y - D*z) + \\ & i(A*x + B*w + C*z - D*y) + \\ & j(A*y - B*z + C*w + D*x) + \\ & K(A*z + B*y - C*x + D*w). \end{aligned} \tag{3}$$

It can be seen that the result of quaternion convolution produces a unique linear combination of each coordinate axis for every single-axis result. This stems from the structure of quaternion multiplication and enforces interactions between each axis of the kernel and each axis of the image. This contrasts with real-valued convolution, which simply multiplies each channel of the kernel with the corresponding channel of the image. The advantage of this approach is that quaternion convolution promotes cross-axis interactions and information extraction through the reuse of filters across each axis and combination [27].

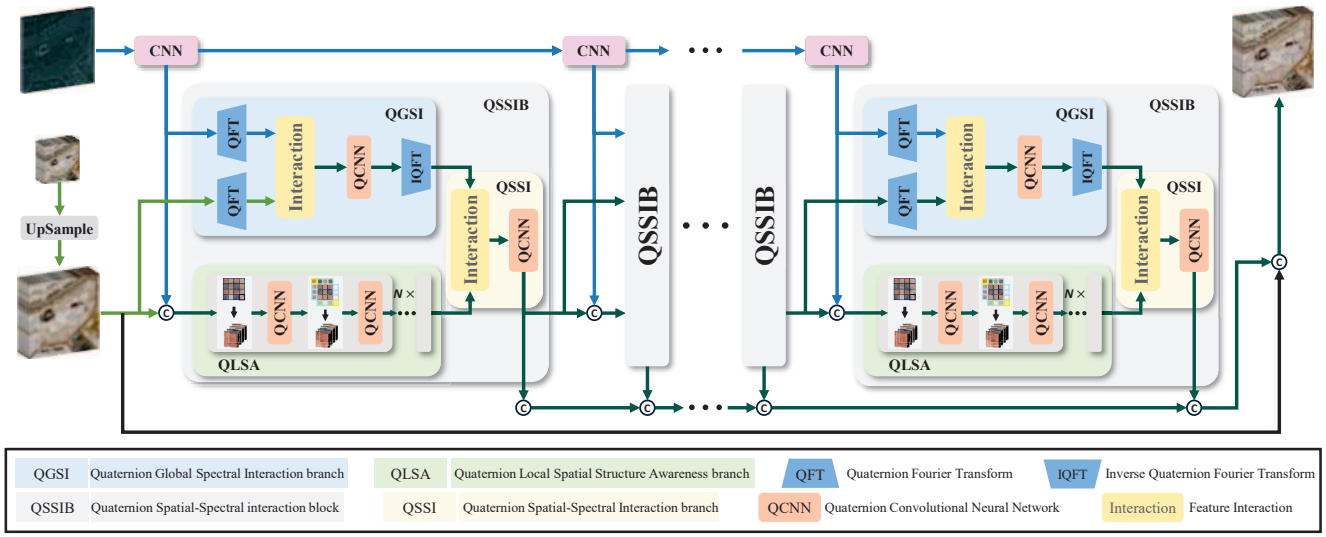


Figure 2. The framework of our proposed pan-sharpening network QuatPanNet. It consists of several Quaternion Spatial-Spectral Interaction Blocks (QSSIB). The Quaternion Spatial-Spectral Interaction Block is shown in Figure 3.

**Quaternion Fourier Transform**

The Fourier Transform plays a crucial role in various image processing tasks, including pansharpening [50]. After decomposing the input image into multiple channels, the Discrete Fourier Transform (DFT) can be applied to each channel individually. However, this approach neglects the mutual information among different channels, which can hinder the performance of pansharpening. The Discrete Quaternion Fourier Transform (DQFT) offers the possibility of simultaneously transforming multi-channel images, as it can transform quaternions holistically [4].

Due to the non-commutative nature of quaternion multiplication, there are various types of Quaternion Fourier Transforms (QFT), such as left-side, right-side and two-side QFT [6]. In this paper, we utilize a typical QFT, specifically the left-sided QFT. For an image $\dot{f}$ of size $H \times W$, the Discrete Quaternion Fourier Transform is defined as follows.

$$\mathcal{F}(\dot{f})(u,v) = \frac{1}{\sqrt{HW}} \sum_{x=0}^{H-1} \sum_{y=0}^{W-1} e^{-\dot{\mu} 2\pi (\frac{x}{H}u + \frac{y}{H}v)} \dot{f}(x,y), \tag{4}$$

where $\dot{f}(x,y)$ is pixel of $\dot{f}$ and $\dot{\mu}$ is unit pure quaternion. $(u,v)$ is the coordinate in the quaternion frequency domain. $\mathcal{F}^{-1}(\dot{f})$ defines the inverse quaternion Fourier transform accordingly. Inspired by the traditional Fourier transform, the modulus of $\mathcal{F}(\dot{f})(u,v)$ is taken as the amplitude component, and the angles with the three imaginary axes are taken as the phase components.

It is worth noting that the QFT is not a simple substitution of the imaginary unit in the two-dimensional Fourier Transform with quaternions; rather, its basis functions possess orthogonal imaginary components, which preserve the information and mutual relationships among different components in the frequency domain. Consequently, some studies leverage QFT to treat multiple features as a unified whole to enhance performance [4, 46]. Furthermore, the efficiency of QFT is also noteworthy. For example, when comparing the use of QFT with three separate conventional Fourier Transforms for processing color images, the number of real multiplications and additions required by QFT is approximately 75% of that needed for the three separate transforms [30]. For pansharpening, we propose utilizing the QFT to treat multispectral data as a unified whole, facilitating global information interactions in the frequency domain while preserving inter-channel dependencies.

### 3.2. Framework

Based on the above analysis, we propose a novel pansharpening approach based on quaternion information interaction, as illustrated in Figure 2. Given a panchromatic (PAN) image $P \in \mathbb{R}^{H \times W \times 1}$ and a low-resolution multispectral (LRMS) image $L \in \mathbb{R}^{H/r \times W/r \times C}$, the network first uses convolution to project the bicubic upsampled r-times $L$ into a shallow feature representation. Simultaneously, $P$ is processed through multiple cascaded convolution layers to extract effective information for high-resolution multispectral (HRMS) reconstruction. Then, the features of the PAN and LRMS images are jointly processed through the core building module, the quaternion spatial-spectral interaction block (QSSIB), for continuous global information interaction and local structure-aware integration. Next, a convolution layer transforms all features gathered by the N QSSIBs back into the image space, and finally, these features are combined with the residual to obtain the final output.

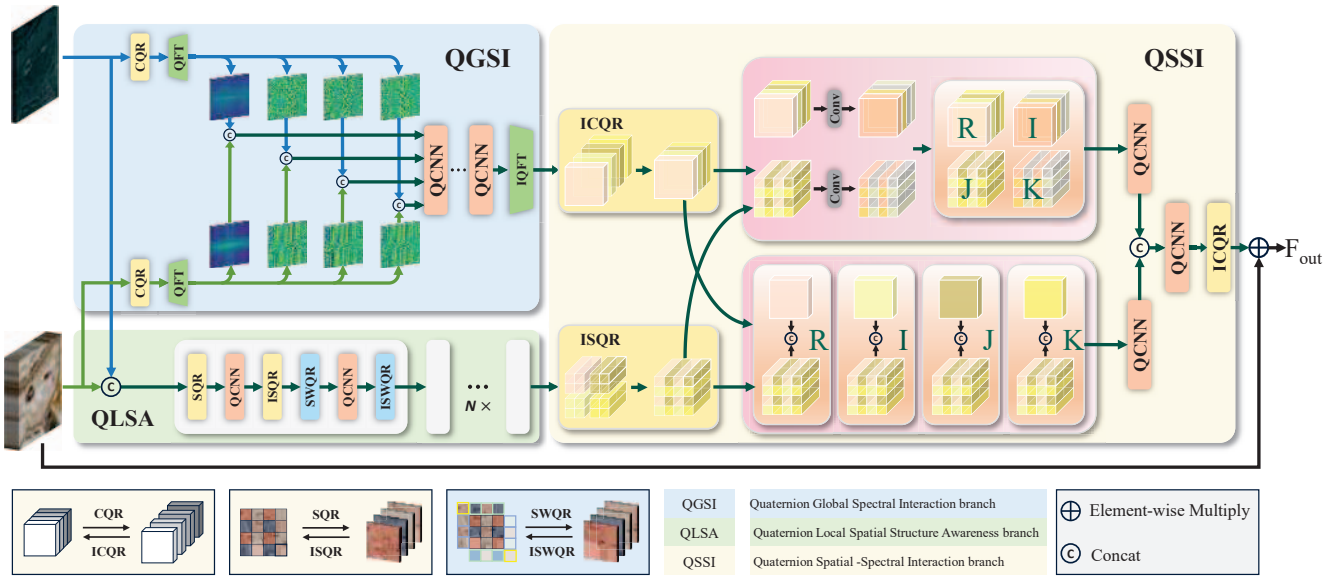


Figure 3. The detailed flowchart of the proposed core building module quaternion spatial-spectral interaction block (QSSIB), consisting of three components: quaternion global spectral interaction branch (QGSI), quaternion local spatial structure awareness branch (QLSA) and quaternion spatial-spectral interaction branch (QSSI).

### 3.3. Key Building Components

As shown in Figure 3, the fundamental building block of our method comprises three key elements: the quaternion global spectral interaction branch (QGSI), the quaternion local spatial structure awareness branch (QLSA), and the quaternion spatial-spectral interaction branch (QSSI). To convert images into quaternion representations, we design customized methods, including channel quaternion representation (CQR), spatial quaternion representation (SQR), and shifted window quaternion representation (SWQR). When transforming quaternion representations back into image features, we denote the corresponding inverse operations as ICQR, ISQR, and ISWQR, respectively. (For more details on quaternion representation, refer to the supplementary material.) These three key components, in conjunction with the quaternion representations of the images, jointly facilitate channel interactions during multispectral fusion and local interactions for spatial resolution enhancement.

**Quaternion Global Spectral Interaction**

The global interaction branch utilizes the Quaternion Fourier Transform (QFT) to transform multi-channel features as a whole into the frequency domain, facilitating global information interactions. Since the input to QFT must be in quaternion form and we aim to maintain inter-channel interdependencies, we design a channel quaternion representation (CQR). This approach divides the image features equally along the channel dimension into four parts, which are then used as the four components of a quaternion. Let the features of the MS and PAN images be denoted as $F_{ms}$ and $F_p$, respectively; their channel-dimensional quaternion representations are denoted as $Q^c_{ms}$ and $Q^c_p$. We then use QFT to transform the MS and PAN features into the frequency domain. Some studies on image restoration indicate that decomposing amplitude and phase components facilitates the decoupling of degradation [17, 50]. Therefore, we decompose the frequency domain features into amplitude and phase components, with the corresponding process represented as

$$\mathcal{A}(Q^c_p), \mathcal{P}_1(Q^c_p), \mathcal{P}_2(Q^c_p), \mathcal{P}_3(Q^c_p) = \mathcal{F}(Q^c_p), \quad (5)$$

$$\mathcal{A}(Q^c_{ms}), \mathcal{P}_1(Q^c_{ms}), \mathcal{P}_2(Q^c_{ms}), \mathcal{P}_3(Q^c_{ms}) = \mathcal{F}(Q^c_{ms}), \quad (6)$$

where $\mathcal{A}(\cdot)$ represents the amplitude, and $\mathcal{P}_1(\cdot)$ , $\mathcal{P}_2(\cdot)$, and $\mathcal{P}_3(\cdot)$ represent the phases corresponding to the three imaginary axes. We then concatenate the amplitude and phase spectra of the PAN and MS images accordingly to form new quaternions. Next, we process the newly formed quaternions using quaternion convolution, thereby enhancing global representations in the frequency domain while promoting interactions between channels.

$$\begin{aligned} Q^c_f = \mathcal{OQ}(\mathcal{C}\left[\mathcal{A}(Q^c_p), \mathcal{A}(Q^c_{ms})\right], \mathcal{C}\left[\mathcal{P}_1(Q^c_p), \mathcal{P}_1(Q^c_{ms})\right] \\ , \mathcal{C}\left[\mathcal{P}_2(Q^c_p), \mathcal{P}_2(Q^c_{ms})\right], \mathcal{C}\left[\mathcal{P}_3(Q^c_p), \mathcal{P}_3(Q^c_{ms})\right]) \end{aligned} \quad (7)$$

where $\mathcal{C}$ denotes concatenation along the channel dimension, $\mathcal{OQ}$ is quaternion convolution. The fused amplitude and phase components are the transformed back to the spatial domain using the inverse discrete quaternion Fourier transform.

$$Q^c_{fre} = \mathcal{F}^{-1}(Q^c_f). \quad (8)$$

By representing the image in a channel-dimensional quaternion form, the quaternion operator treats the multi-channel data as a unified whole for information interaction.

Additionally, according to spectral convolution theorem [7], processing information in the Fourier space enables capturing global frequency representations. In summary, this branch generates the global information representation with strengthened channel interactions $Q^c_{fre}$.

**Quaternion Local Spatial Structure Awareness**

Pansharpening is essentially a fusion-based super-resolution task, making the injection of spatial resolution critically important. Enhancing spatial resolution requires not only global information from the frequency domain but also local information interactions in the spatial domain. For pansharpening, preserving local structures during information interactions is particularly significant, as it enhances the network's ability to recover fine details.

We aim to leverage the balance of enhanced information interaction and specific information preservation offered by quaternions to improve the network's spatial super-resolution capabilities. To this end, we design the spatial quaternion representation (SQR). It divides the image into several $2 \times 2$ windows in the spatial dimension, with each local window's four pixels serving as the four components of a quaternion, as illustrated in Figure 3. This approach explicitly ensures dependencies among local pixels, preserving the local structure. Based on this representation, we use quaternion convolutions (QCNN) to process the features.

However, this representation method lacks cross-window connections, which limits spatial interactions. Inspired by [22], we design the shifted window quaternion representation (SWQR) based on SQR. This allows the network to alternately employ two partitioning configurations, thereby enhancing the effectiveness of spatial quaternion representation. Through the combined effect of this method and quaternion convolution, this branch preserves local spatial dependencies while promoting spatial interactions, thereby obtaining the local representation $Q^s_{spa}$.

The information representations of the two branches are highly complementary, with one branch responsible for global, frequency-domain, and channel interactions, and the other focused on local, spatial-domain, and spatial interactions. Therefore, further interaction and integration between them facilitate mutual compensation.

**Quaternion Spatial-Spectral Interaction**

The two branches described above respectively obtain quaternion representations in the channel dimension, $Q^c_{fre} \in \mathbb{H}^{H\times W\times C/4}$, and in the spatial dimension, $Q^s_{spa} \in \mathbb{H}^{H/2\times W/2\times C}$. The quaternion dual-domain information interaction branch mainly consists of two parts: one is a quaternion-based global fusion, and the other is the channel-aware quaternion spatial feature injection.

Quaternion-based global fusion merges the features obtained from the two branches into a single quaternion for unified processing, thereby enhancing the correlation between the two through the structure of the quaternion. First, both $Q^c_{fre}$ and $Q^s_{spa}$ are converted back into image feature representations. $Q^c_{fre}$ is combined along the channel dimension to obtain $F_{fre} \in \mathbb{R}^{H\times W\times C}$, and $Q^s_{spa}$ is rearranged spatially to obtain $F_{spa} \in \mathbb{R}^{H\times W\times C}$. To achieve a richer feature representation, we apply convolutions to $F_{fre}$ and $F_{spa}$ to obtain $F'_{fre}$ and $F'_{spa}$, respectively. We then combine these four features into a quaternion $Q_f = F_{fre} + F_{spa}i + F'_{fre}j + F'_{spa}k$ . We use quaternion convolution to process $Q_f \in \mathbb{H}^{H\times W\times C}$ as a whole, producing $Q_a \in \mathbb{H}^{H\times W\times C/4}$.

Channel-aware quaternion spatial feature injection involves injecting spatial features $F_{spa}$ from the local structure-aware branch into the four components of the quaternion representation $Q^c_{fre}$ in the channel dimension, thereby facilitating information fusion in both spatial and channel domains. Specifically, we replicate the image feature $F_{spa}$ four times and concatenate each copy with one of the four components of the quaternion feature $Q^c_{fre}$ to obtain $Q_i \in \mathbb{H}^{H\times W\times 5C/4}$:

$$Q_i = \mathcal{C}\left[Q^c_{fre}, (F_{spa} + F_{spa}i + F_{spa}j + F_{spa}k)\right]. \quad (9)$$

We then feed $Q_i$ into quaternion convolution to obtain $Q_b \in \mathbb{H}^{H\times W\times C/4}$. Afterward, we concatenate $Q_a$ and $Q_b$, and apply quaternion convolution for further fusion to obtain $Q_{out} \in \mathbb{H}^{H\times W\times C/4}$.

$$Q_{out} = \mathrm{QConv}(\mathcal{C}\left[Q_a, Q_b\right]), \quad (10)$$

where QConv denotes quaternion convolution. We then convert $Q_{out}$ into the image feature representation $F_{out} \in \mathbb{R}^{H\times W\times C}$. Finally, a residual learning mechanism is applied, adding the input feature $F_{ms}$ to the fused feature. The proposed network achieves effective fusion through the seamless integration of the two quaternion representations.

### 3.4. Optimization

Let $Y$ and $GT$ denote the network's output and the corresponding ground truth, respectively. We use a joint spatial-frequency domain loss to supervise the network training. In the frequency domain, we first apply the quaternion Fourier transform to convert $Y$ and $GT$ into the quaternion Fourier space. Then, we compute the L1 norm of the real and imaginary differences between $Y$ and $GT$, and sum them to obtain the quaternion Fourier domain loss.

$$\begin{aligned}\mathcal{L}_{fre} = &\left\|R(Y) - R(GT)\right\|_1 + \left\|I(Y) - I(GT)\right\|_1 \\ &+ \left\|J(Y) - J(GT)\right\|_1 + \left\|K(Y) - K(GT)\right\|_1,\end{aligned} \quad (11)$$

where $R(\cdot)$, $I(\cdot)$, $J(\cdot)$, and $K(\cdot)$ represent the real part and three imaginary parts of the quaternion respectively. In the spatial domain, we employ the L1 loss. Finally, the overall loss function is as follows

$$\mathcal{L} = \left\|Y - GT\right\|_1 + \lambda\mathcal{L}_{fre}, \quad (12)$$

where $\lambda$ is the weighting factor and set to 0.1 empirically.

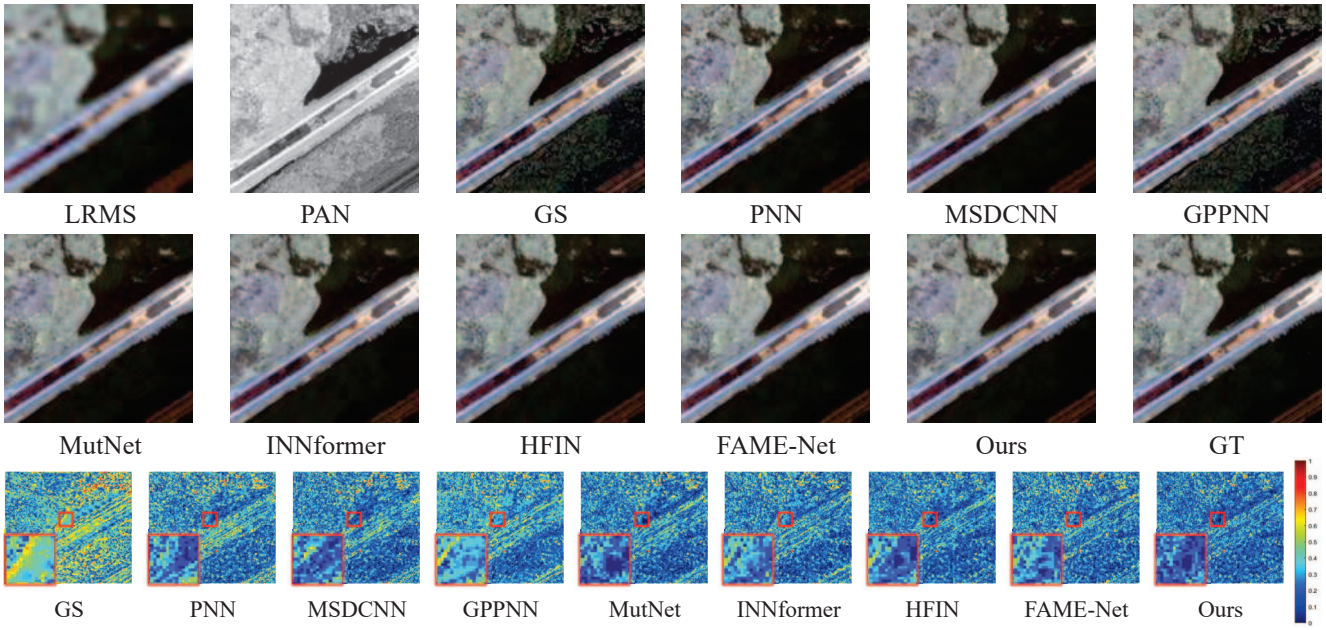


Figure 4. The visual comparisons between other pan-sharpening methods and our method on WorldView-II satellite

Table 1. Experimental results of all the competing methods on the three benchmark datasets. The best and the second best values are highlighted in **bold** and underline, respectively.

| Methods | Params (M) | WorldView II | | | | WorldView III | | | | GaoFen2 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | PSNR ↑ | SSIM ↑ | SAM ↓ | ERGAS ↓ | PSNR ↑ | SSIM ↑ | SAM ↓ | ERGAS ↓ | PSNR ↑ | SSIM ↑ | SAM ↓ | ERGAS ↓ |
| Brovey | - | 35.8646 | 0.9216 | 0.0403 | 1.8238 | 22.5060 | 0.5466 | 0.1159 | 8.2331 | 37.7974 | 0.9026 | 0.0218 | 1.3720 |
| IHS | - | 35.2962 | 0.9027 | 0.0461 | 2.0278 | 22.5579 | 0.5354 | 0.1266 | 8.3616 | 38.1754 | 0.9100 | 0.0243 | 1.5336 |
| GS | - | 35.6376 | 0.9176 | 0.0423 | 1.8774 | 22.5608 | 0.5470 | 0.1217 | 8.2433 | 37.2260 | 0.9034 | 0.0309 | 1.6736 |
| PNN | 0.0689 | 40.7550 | 0.9624 | 0.0259 | 1.0646 | 29.9418 | 0.9121 | 0.0824 | 3.3206 | 43.1208 | 0.9704 | 0.0172 | 0.8528 |
| MSDCNN | 0.2390 | 41.3355 | 0.9664 | 0.0242 | 0.994 | 30.3038 | 0.9184 | 0.0782 | 3.1884 | 45.6874 | 0.9827 | 0.0135 | 0.6389 |
| GPPNN | 0.1198 | 41.1622 | 0.9684 | 0.0244 | 1.0315 | 30.1785 | 0.9175 | 0.0776 | 3.2593 | 44.2145 | 0.9815 | 0.0137 | 0.7361 |
| MutNet | 0.0714 | 41.6773 | 0.9705 | 0.0224 | 0.9519 | 30.4907 | 0.9223 | 0.0749 | 3.1125 | 47.3042 | 0.9892 | 0.0102 | 0.5481 |
| INNformer | 0.0706 | 41.6903 | 0.9704 | 0.0227 | 0.9514 | 30.5365 | 0.9225 | 0.0747 | 3.0997 | 47.3528 | 0.9893 | 0.0102 | 0.5479 |
| HFIN | 0.0772 | 42.2319 | 0.9714 | 0.0215 | 0.8807 | 30.6147 | 0.9203 | 0.0742 | 3.0786 | 48.8783 | 0.9898 | 0.0093 | 0.4591 |
| FAME-Net | 0.1408 | 42.0262 | 0.9723 | 0.0215 | 0.9172 | 30.9903 | 0.9287 | 0.0697 | 2.9531 | 47.6721 | 0.9898 | 0.0098 | 0.5242 |
| Ours | 0.0387 | **42.4846** | **0.9738** | **0.0208** | **0.8554** | **31.1766** | **0.9332** | **0.0521** | **2.5761** | **49.1423** | **0.9901** | **0.0090** | **0.4474** |

# 4. Experiments

## 4.1. Dataset and benchmarks

We conduct our experiments on three classic datasets. WorldView-II (WV2), Gaofen2 (GF2) and WorldView-III (WV3). The WorldView-II dataset consists of 760 image pairs for training and 80 image pairs for testing. The WorldView-III dataset includes 2150 image pairs for training and 200 image pairs for testing. The GaoFen-2 dataset comprises 2712 image pairs for training and 200 image pairs for testing. To validate the effectiveness of our approach, we compare it with several state-of-the-art pansharpening methods, including PNN [23], MSDCNN [42], GPPNN [39], MutNet [49], INNformer [47], HFIN [33] and FAME-Net [16], as well as traditional pansharpening methods such as GS [20], Brovey [12], and IHS [14].

In our experiments, all deep learning models are implemented using PyTorch and trained on an NVIDIA GeForce GTX 3090 GPU. For each set, the multispectral (MS) images are cropped into patches with the size of 32×32, and the corresponding panchromatic (PAN) images are of size 128×128. We employ common evaluation metrics, including PSNR, SSIM, SAM, and ERGAS. Additionally, we utilize three widely-used no-reference IQA metrics for real-world full-resolution scenes: $D_\lambda$,$D_S$ and QNR. More descriptions of the implementation details and dataset are provided in the supplementary material.

Table 2. Non-reference metrics on full-resolution dataset.

| | PNN | MSDCNN | GPPNN | INNformer | MutNet | HFIN | FAME-Net | Ours |
|---|---|---|---|---|---|---|---|---|
| $D_\lambda \downarrow$ | 0.0746 | 0.0734 | 0.0782 | 0.0697 | 0.0694 | 0.0710 | 0.0674 | **0.0566** |
| $D_s \downarrow$ | 0.1164 | 0.1151 | 0.1253 | 0.1128 | 0.1118 | 0.1098 | 0.1121 | **0.1055** |
| QNR ↑ | 0.8191 | 0.8215 | 0.8073 | 0.8253 | 0.8247 | 0.8261 | 0.8291 | **0.8447** |

## 4.2. Comparison with state-of-the-art methods

**Evaluation on reduced-resolution scene.** To evaluate the effectiveness of our method, we conducted experiments on three datasets and compared it with ten other state-of-the-art pansharpening methods. Table 1 summarizes the evaluation metrics for the three datasets, where the best result for each metric is highlighted in red and the second-best result in blue. The results demonstrate that our method significantly outperforms the others across all evaluation metrics, achieving the best performance on all three datasets. Compared to the second-best method in terms of PSNR for each dataset, our method shows an improvement of 0.25 dB on the WV2 dataset, 0.18 dB on the GF2 dataset, and 0.27 dB on the WV3 dataset, indicating the consistency of our method with the true high-resolution images. We also conducted qualitative experiments, Figures 4 present samples from the WV2 datasets. The top two rows show the comparison of experimental results, while the last row displays the mean squared error maps of each network's results compared to the ground truth. It is evident that the results obtained by our network are closest to the ground truth, with the smallest discrepancies.

**Evaluation on full-resolution scene.** We also conduct full-resolution analysis in real-world scenarios to validate the generalization capability of our method. We perform experiments on an additional 200 sets of the GF2 dataset. Due to the absence of high-resolution multispectral images for actual real-world scenarios, we employ three commonly used no-reference metrics, $D_\lambda$, $D_S$, and QNR for evaluation. The experimental results, as shown in Table 2, demonstrate that our method achieves the best performance across all three metrics.

Table 3. The results of different configurations on WorldViewII. The best and the second best values are highlighted in **bold** and underline, respectively.

| Configuration | QGSI | QLSA | QSSI | PSNR ↑ | SSIM ↑ | SAM ↓ | ERGAS ↓ |
|---|---|---|---|---|---|---|---|
| I | ✗ | ✓ | ✓ | 41.7356 | 0.9673 | 0.0247 | 0.9337 |
| II | ✓ | ✗ | ✓ | 42.0217 | 0.9706 | 0.0219 | 0.9134 |
| III | ✓ | ✓ | ✗ | 42.1734 | 0.9713 | 0.0221 | 0.9012 |
| Ours | ✓ | ✓ | ✓ | **42.4846** | **0.9738** | **0.0208** | **0.8554** |

## 4.3. Ablation Experiments

We perform an ablation study using the WorldView-II dataset to further demonstrate the effectiveness of our method. The quaternion global spectral interaction branch (QGSI), quaternion local spatial structure awareness branch (QLSA), and quaternion spatial-spectral interaction branch (QSSI) are the core components of the proposed method, and we conduct separate ablations for each. Additionally, we also explore the impact of the number of basic building blocks used in the network through ablation studies.

Table 4. PSNR values of Our model with different number of blocks on WorldViewII. The best and the second best values are highlighted in **bold** and underline, respectively.

| Stages (K) | PSNR ↑ | SSIM ↑ | SAM ↓ | ERGAS ↓ |
|---|---|---|---|---|
| 1 | 41.8378 | 0.9603 | 0.0249 | 0.9879 |
| 2 | 41.9002 | 0.9647 | 0.0238 | 0.9725 |
| 3 | 42.0913 | 0.9698 | 0.0234 | 0.9578 |
| 4 | 42.2369 | 0.9724 | 0.0215 | 0.9007 |
| 5 | **42.4846** | **0.9738** | **0.0208** | **0.8554** |

To validate the effectiveness of the three key quaternion components, QGSI, QLSA, and QSSI, we replace their quaternion operations with real-valued operations, as shown in Table 3. The results demonstrate that removing each quaternion component leads to a decline in performance, thus confirming the effectiveness of the quaternion's multi-dimensional representation and parameter-sharing mechanism for pansharpening.

We conduct experiments on networks with varying numbers of core building blocks to evaluate the impact of the number of core modules, as shown in Table 4. As observed, the model performance improves with the increase in the number of Quaternion Spatial-Spectral Interaction Blocks (QSSIB), though this comes at the cost of increased resource consumption. We balance performance and resource consumption by using the default setting of 5. More ablation experiments are shown in the supplementary materials.

## 5. Conclusion

In this paper, we propose QuatPanNet, a novel quaternion-based spatial-spectral interaction network that enhances pansharpening by leveraging quaternions' compact representation for high-dimensional data. Our work demonstrates the advantages of a unified quaternion framework for effectively addressing the pansharpening problem. We custom-design quaternion-based spectral and spatial interaction modules: the former promotes global information interaction while preserving spectral dependencies, and the latter fosters spatial interaction while maintaining local structural dependencies. Building on these, we integrate the dual-path features into a unified quaternion representation. By utilizing the compact and efficient representation of PAN and MS image data in the quaternion space, QuatPanNet facilitates effective interactions between spatial and spectral features. Our method achieves exceptional fusion quality with a reduced number of parameters.

## Acknowledgments

This work is supported by the National Natural Science Foundation of China (NSFC) under Grants 62225207, 62436008, 62422609 and 62276243.

## References

[1] Paolo Addesso, Gemine Vivone, Rocco Restaino, and Jocelyn Chanussot. A data-driven model-based regression applied to panchromatic sharpening. *IEEE Transactions on Image Processing*, 29:7779–7794, 2020. 1

[2] Nicholas A Aspragathos and John K Dimitros. A comparative study of three methods for robot kinematics. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 28(2):135–145, 1998. 3

[3] Qinglong Cao, Zhengqin Xu, Yuntian Chen, Chao Ma, and Xiaokang Yang. Domain prompt learning with quaternion networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26637–26646, 2024. 2, 3

[4] Ping Cui, Cui-Ming Zou, Bang-Ju Wang, and Feng Shi. Color image fusion in quaternion spectral domain. In *2023 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)*, pages 56–61. IEEE, 2023. 4

[5] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 1

[6] Todd A Ell, Nicolas Le Bihan, and Stephen J Sangwine. *Quaternion Fourier transforms for signal and image processing*. John Wiley & Sons, 2014. 4

[7] Matteo Frigo and Steven G Johnson. Fftw: An adaptive software architecture for the fft. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181)*, pages 1381–1384. IEEE, 1998. 6

[8] Xueyang Fu, Zihuang Lin, Yue Huang, and Xinghao Ding. A variational pan-sharpening with local gradient constraints. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10265–10274, 2019. 2

[9] Xueyang Fu, Wu Wang, Yue Huang, Xinghao Ding, and John Paisley. Deep multiscale detail networks for multiband spectral image sharpening. *IEEE Transactions on Neural Networks and Learning Systems*, 32(5):2090–2104, 2020. 1

[10] Chase J Gaudet and Anthony S Maida. Deep quaternion networks. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2018. 3

[11] Pedram Ghamisi, Behnood Rasti, Naoto Yokoya, Qunming Wang, Bernhard Hofle, Lorenzo Bruzzone, Francesca Bovolo, Mingmin Chi, Katharina Anders, Richard Gloaguen, et al. Multisource and multitemporal data fusion in remote sensing: A comprehensive review of the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 7(1):6–39, 2019. 2

[12] Alan R Gillespie, Anne B Kahle, and Richard E Walker. Color enhancement of highly correlated images. ii. channel ratio and "chromaticity" transformation techniques. *Remote Sensing of Environment*, 22(3):343–365, 1987. 7

[13] William Rowan Hamilton. *Lectures on Quaternions: Containing a Systematic Statement of a New Mathematical Method; of which the Principles Were Communicated in 1843 to the Royal Irish Academy; and which Has Since Formed the Subject of Successive Courses of Lectures, Delivered in 1848 and Subsequent Years, in the Halls of Trinity College, Dublin: with Numerous Illustrative Diagrams, and with Some Geometrical and Physical Applications*. Hodges and Smith, 1853. 3

[14] R Haydn. Application of the ihs color transform to the processing of multisensor data and image enhancement. In *Proc. of the International Symposium on Remote Sensing of Arid and Semi-Arid Lands, Cairo, Egypt, 1982*, 1982. 7

[15] Xiyan He, Laurent Condat, José M Bioucas-Dias, Jocelyn Chanussot, and Junshi Xia. A new pansharpening method based on spatial and spectral sparsity priors. *IEEE Transactions on Image Processing*, 23(9):4160–4174, 2014. 2

[16] Xuanhua He, Keyu Yan, Rui Li, Chengjun Xie, Jie Zhang, and Man Zhou. Frequency-adaptive pan-sharpening with mixture of experts. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2121–2129, 2024. 3, 7

[17] Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. Deep fourier-based exposure correction network with spatial-frequency interaction. In *European Conference on Computer Vision*, pages 163–180. Springer, 2022. 5

[18] Creed F Jones III. *Color face recognition using quaternionic Gabor filters*. PhD thesis, Virginia Polytechnic Institute and State University, 2005. 3

[19] Hiromi Kusamichi, Teijiro Isokawa, Nobuyuki Matsui, Yuzo Ogawa, and Kazuaki Maeda. A new scheme for color night vision by quaternion neural network. In *Proceedings of the 2nd international conference on autonomous robots and agents*. Citeseer, 2004. 3

[20] Craig A Laben and Bernard V Brower. Process for enhancing the spatial resolution of multispectral imagery using pansharpening, 2000. US Patent 6,011,875. 7

[21] Pengfei Liu, Liang Xiao, and Tao Li. A variational pansharpening method based on spatial fractional-order geometry and spectral–spatial low-rank priors. *IEEE Transactions on Geoscience and Remote Sensing*, 56(3):1788–1802, 2017. 2

[22] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 6

[23] Giuseppe Masi, Davide Cozzolino, Luisa Verdoliva, and Giuseppe Scarpa. Pansharpening by convolutional neural networks. *Remote Sensing*, 8(7):594, 2016. 1, 2, 7

[24] Xiangchao Meng, Yiming Xiong, Feng Shao, Huanfeng Shen, Weiwei Sun, Gang Yang, Qiangqiang Yuan, Randi Fu, and Hongyan Zhang. A large-scale benchmark data set for evaluating pansharpening performance: Overview and implementation. *IEEE Geoscience and Remote Sensing Magazine*, 9(1):18–52, 2020. 2

[25] Jorge Nunez, Xavier Otazu, Octavi Fors, Albert Prades, Vicenc Pala, and Roman Arbiol. Multiresolution-based image fusion with additive wavelet decomposition. *IEEE Transactions on Geoscience and Remote sensing*, 37(3):1204–1211, 1999. 2

[26] Titouan Parcollet, Mirco Ravanelli, Mohamed Morchid, Georges Linarès, Chiheb Trabelsi, Renato De Mori, and Yoshua Bengio. Quaternion recurrent neural networks. *arXiv preprint arXiv:1806.04418*, 2018. 2, 3

[27] Titouan Parcollet, Ying Zhang, Mohamed Morchid, Chiheb Trabelsi, Georges Linarès, Renato De Mori, and Yoshua Bengio. Quaternion convolutional neural networks for end-to-end automatic speech recognition. *arXiv preprint arXiv:1806.07789*, 2018. 2, 3

[28] Soo-Chang Pei, Jian-Jiun Ding, and Ja-Han Chang. Efficient implementation of quaternion fourier transform, convolution, and correlation by 2-d complex fft. *IEEE transactions on signal processing*, 49(11):2783–2797, 2001. 3

[29] Thierry Ranchin and Lucien Wald. Fusion of high spatial and spectral resolution images: The arsis concept and its implementation. *Photogrammetric engineering and remote sensing*, 66(1):49–61, 2000. 2

[30] Stephen John Sangwine. Fourier transforms of colour images using quaternion or hypercomplex, numbers. *Electronics letters*, 32(21):1979–1980, 1996. 3, 4

[31] Vijay P Shah, Nicolas H Younan, and Roger L King. An efficient pan-sharpening method via a combined adaptive pca approach and contourlets. *IEEE transactions on geoscience and remote sensing*, 46(5):1323–1335, 2008. 2

[32] Lilong Shi and Brian Funt. Quaternion color texture segmentation. *Computer Vision and image understanding*, 107(1-2): 88–96, 2007. 3

[33] Jiangtong Tan, Jie Huang, Naishan Zheng, Man Zhou, Keyu Yan, Danfeng Hong, and Feng Zhao. Revisiting spatial-frequency information integration from a hierarchical perspective for panchromatic and multi-spectral image fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 25922–25931, 2024. 3, 7

[34] Te-Ming Tu, Shun-Chi Su, Hsuen-Chyun Shyu, and Ping S Huang. A new look at ihs-like image fusion methods. *Information fusion*, 2(3):177–186, 2001. 2

[35] Gemine Vivone, Rocco Restaino, Mauro Dalla Mura, Giorgio Licciardi, and Jocelyn Chanussot. Contrast and error-based fusion schemes for multispectral image pansharpening. *IEEE Geoscience and Remote Sensing Letters*, 11(5): 930–934, 2013. 2

[36] Dong Wang, Yunpeng Bai, and Ying Li. Multispectral pan-sharpening via dual-channel convolutional network with convolutional lstm based hierarchical spatial-spectral feature fusion, 2020. 3

[37] Yingying Wang, Xuanhua He, Yuhang Dong, Yunlong Lin, Yue Huang, and Xinghao Ding. Cross-modality interaction network for pan-sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 2024. 2

[38] Chen Wu, Bo Du, Xiaohui Cui, and Liangpei Zhang. A post-classification change detection method based on iterative slow feature analysis and bayesian soft fusion. *Remote Sensing of Environment*, 199:241–255, 2017. 1

[39] Shuang Xu, Jiangshe Zhang, Zixiang Zhao, Kai Sun, Junmin Liu, and Chunxia Zhang. Deep gradient projection networks for pan-sharpening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1366–1375, 2021. 2, 7

[40] Gang Yang, Xiangyong Cao, Wenzhe Xiao, Man Zhou, Aiping Liu, Xun Chen, and Deyu Meng. Panflownet: A flow-based deep network for pan-sharpening. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 16857–16867, 2023. 1

[41] Licheng Yu, Yi Xu, Hongteng Xu, and Hao Zhang. Quaternion-based sparse representation of color image. In *2013 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–7. IEEE, 2013. 3

[42] Qiangqiang Yuan, Yancong Wei, Xiangchao Meng, Huanfeng Shen, and Liangpei Zhang. A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(3):978–989, 2018. 2, 7

[43] Xiaohui Yuan, Jianfang Shi, and Lichuan Gu. A review of deep learning methods for semantic segmentation of remote sensing imagery. *Expert Systems with Applications*, 169: 114417, 2021. 1

[44] Rui Zeng, Jiasong Wu, Zhuhong Shao, Yang Chen, Beijing Chen, Lotfi Senhadji, and Huazhong Shu. Color image classification via quaternion principal component analysis network. *Neurocomputing*, 216:416–428, 2016. 3

[45] Hao Zhang and Jiayi Ma. Gtp-pnet: A residual learning network based on gradient transformation prior for pansharpening. *ISPRS Journal of Photogrammetry and Remote Sensing*, 172:223–239, 2021. 1

[46] Yunzuo Zhang, Jiayu Zhang, Ruixue Liu, Pengfei Zhu, and Yameng Liu. Key frame extraction based on quaternion fourier transform with multiple features fusion. *Expert Systems with Applications*, 216:119467, 2023. 4

[47] Man Zhou, Jie Huang, Yanchi Fang, Xueyang Fu, and Aiping Liu. Pan-sharpening with customized transformer and invertible neural network. In *Proceedings of the AAAI conference on artificial intelligence*, pages 3553–3561, 2022. 2, 7

[48] Man Zhou, Jie Huang, Keyu Yan, Hu Yu, Xueyang Fu, Aiping Liu, Xian Wei, and Feng Zhao. Spatial-frequency domain information integration for pan-sharpening. In *European conference on computer vision*, pages 274–291. Springer, 2022. 3

[49] Man Zhou, Keyu Yan, Jie Huang, Zihe Yang, Xueyang Fu, and Feng Zhao. Mutual information-driven pan-sharpening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1798–1808, 2022. 7

[50] Man Zhou, Jie Huang, Chun-Le Guo, and Chongyi Li. Fourmer: An efficient global modeling paradigm for image restoration. In *International conference on machine learning*, pages 42589–42601. PMLR, 2023. 4, 5

[51] Xuanyu Zhu, Yi Xu, Hongteng Xu, and Changjian Chen. Quaternion convolutional neural networks. In *Proceedings of*

*the European conference on computer vision (ECCV)*, pages 631–647, 2018. 3

[52] Zeyu Zhu, Xiangyong Cao, Man Zhou, Junhao Huang, and Deyu Meng. Probability-based global cross-modal upsampling for pansharpening. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14039–14048, 2023. 3