Electronics and Electrical Engineering

AIMS Electronics and Electrical Engineering, 9(4): 500–540.
DOI: 10.3934/electreng.2025023
Received: 28 May 2025
Revised: 04 August 2025
Accepted: 25 August 2025
Published: 04 September 2025

https://www.aimspress.com/journal/ElectrEng

Research article

Robust color multi-focus image fusion using quaternion sparse representation and spatial information

Wei Liu[1,2], Wanqing Li[1], Xia Zhao[3] and Fang Zhu[4,*]

[1] College of Mathematics and Computer Science, Tongling University, Tongling, Anhui 244000, China
[2] Anhui Engineering Research Center Of Intelligent Manufacturing Of Copper-based Materials, Tongling, Anhui 244000, China
[3] Academic Affairs Division, Tongling University, Tongling, Anhui 244000, China
[4] Department of Mathematics, Ministry of General Education, Anhui Xinhua University, Hefei 230088, China

* **Correspondence:** Email: zhufang198710@163.com; Tel: +86-136-4551-3684.

Abstract: Most sparse representation (SR)-based fusion methods handle color channels separately, which easily causes hue distortion and color saturation reduction in fused images. To address these issues, we propose a novel color multi-focus image fusion method based on quaternion sparse representation (QSR). QSR employs the quaternion matrix to model color images in a holistic way that can fully exploit high correlations among color channels and preserve the inherent color structures of source images in reconstruction results. We first learn a clear quaternion dictionary on a high-quality image set and a blurry quaternion dictionary on a Gaussian-blurred version of the same set. To accurately estimate the focus information of each image patch, the salience and sparsity features are computed by collaboratively employing the sparse coefficients of the current patch and its neighboring patches deduced from the QSR model under the two learned dictionaries. Then, an activity measurement of each patch is defined by exploiting the two features. Finally, a maximum fusion rule is adopted to get the composited sparse coefficients. The experimental results show that our method successfully avoids color distortion in the fused images and performs qualitatively and quantitatively better than some recent SR-based fusion methods.

Keywords: color image; multi-focus image fusion; quaternion sparse representation; spatial information; activity measurement

1. Introduction

In recent years, there has been a growing interest in multi-focus image fusion (MFIF). This stems from the inherent limitation of optical lenses, which can only capture partially focused images. However, in numerous practical scenarios, fully focused images are necessary, leading to the pursuit of advanced techniques. MFIF serves as a potent method for extracting focus details from several partially focused images of a single scene, each captured with distinct focus parameters, thereby generating a completely sharp image [1]. At present, MFIF technology has been successfully applied in many fields, such as micro-image fusion [2], visual sensor networks [3], visual power patrol inspection [4], and optical microscopy [5].

In recent years, many efficient MFIF methods have been presented, which are divided into three groups: spatial domain–based methods, transform domain–based methods, and deep learning–based methods [6]. In spatial domain–based methods, the earliest method used the mean pixel values as the results, causing severe detail information loss and blurring effects. As a result, researchers presented several focus measures [multi-scale morphological gradient (MSMG) [7], dense SIFT [8], and LBP [9], for instance] as the criteria to distinguish the focused and defocused pixels of source images. These methods make full use of the relationship between surrounding pixels and, in most cases, perform well for fusing multi-focus images, but may suffer from discontinuity in the results. Recently, many effective MFIF methods based on Markov random field optimization [10], multi-matting [11], conditional random field optimization [12], lazy random walks [13], multiscale fuzzy quality assessment [14], Gaussian curvature filter [15], and parameter adaptive dual channel dynamic threshold neural P systems [16] have been proposed to mitigate artifacts and provide spatial consistency. These can get satisfactory results; however, most of them cannot deal with small-sized focused regions and the transitional zone between focused and defocused regions. The post-processing technology is used to fill the “holes” and remove noise in the decision maps, which may produce incorrect results and unclear boundaries.

Transform domain–based methods typically encompass three key stages: image decomposition, coefficient integration, and inverse transformation. The multi-scale transform (MST) is a common transform tool used in transform domain–based methods. Widely used MST-based methods include pyramid-based [17,18], wavelet-based [19,20], and geometric analysis–based methods [21,22]. Recently, a variety of MSTs that leverage image filtering have been introduced to MFIF, yielding promising outcomes [23–25]. While MST effectively captures the intrinsic features of images, determining the most suitable transform for source images with diverse content remains a complex challenge.

In recent years, deep learning–based methods have become a very active direction in the field of image fusion. Usually, deep learning–based methods can be divided into supervised methods and unsupervised methods. The models used in supervised methods mainly include convolutional neural networks (CNN) and generative adversarial networks (GAN). In 2017, Liu et al. [26] proposed the first CNN-based method. This method utilizes a CNN model to learn a direct mapping from the source image to the fused image, which integrates the clarity information of all source images. Subsequently, many CNN-based methods have been proposed and achieved good results [27–31]. In addition, GAN has also been used for image fusion [32–35]. The FuseGAN proposed by Guo et al. [32] was the first GAN-based method. FuseGAN expresses the fusion problem as an image-to-image conversion problem and utilizes the least squares GAN objective function to

enhance its training stability. In recent years, Transformer models, along with newly emerging generative models, have brought new vitality to MFIF. Notably, Ma [36] harnessed the capabilities of a Swin Transformer model, successfully addressing the challenge of maintaining information across extended distances. Additionally, FusionDiff [37] pioneered the integration of the denoising diffusion probabilistic model into fusion tasks, significantly elevating generation quality through a meticulous process of iterative optimization. Zhu et al. introduced TC-MOA [38], a cohesive and generalized image fusion model that leverages the Vision Transformer architecture. Meanwhile, Hao et al. [39] put forth a comprehensive, large selective kernel network, designed to grasp long-term dependencies spanning diverse scales. This method adeptly assigns weights to the features refined by an array of extensive deep kernels. Furthermore, several researchers have advocated for the utilization of hybrid technology in generating fusion decision maps. Liu et al. [40] innovatively crafted a residual architecture, incorporating a multi-scale feature extraction module and a dual attention module, to derive the decision map. Wang et al. [41] introduced edge-preserving techniques within neural networks, presenting Y-Net to produce the decision map with precision. Duan et al. [42] merged CNNs with Transformers, creating a synergy that yields the fusion decision map. Shao et al. [43] developed a pioneering model for MFIF, seamlessly integrating a U-Net architecture with both a Swin Transformer and CNN. This combination of CNN and Transformer was strategically employed to extract both local and global information, enhancing the overall efficacy of the fusion process. Mamba, representing an innovative advancement in state space models (SSMs), transcends the static parameter constraints of conventional SSMs by incorporating an input-driven selective state mechanism. This groundbreaking model has also found its way into the realm of MFIF [44,45], further expanding its applications and demonstrating its versatility.

Supervised methods require a large amount of labeled training data. In fact, most supervised methods use synthetic data during the training process. Recently, scholars have proposed some unsupervised methods. In 2019, Mustafa et al. [46] proposed an unsupervised method based on CNN, which combines pixel difference and structural similarity (SSIM) as the loss function. In 2020, Xu et al. [47] proposed the FusionDN, which formulated different image fusion tasks into a unified dense connected network. In 2021, Ma et al. [48] proposed a method based on encoder–decoder networks, which also uses structural similarity as part of the loss function. In 2023, Hu et al. [49] introduced the ZMFF method, a novel approach that harnesses the power of deep image prior networks and deep mask prior networks. This method not only produces crisp and clear fused images but also concurrently generates focused maps for each source image, enhancing the overall quality and detail of the fusion process. In addition, researchers use techniques such as cross-scale transformation and pyramid fusion to ensure the semantic and visual coherence of the fused images. Fang et al. [50] utilized the dilated residual dense network to extract comprehensive global features from images. Liu et al. [51] introduced an adaptive feature concatenation attention network, which facilitates seamless feature fusion by amalgamating deep and shallow features through an adaptive cross-layer coordinate attention module. This module derives insights from semantic and edge information, adopting a holistic perspective. Li et al. [52] presented a groundbreaking method that incorporates a gradient-intensity joint proportional constraint, based on the generative adversarial network, for MFIF. Meanwhile, Zhai et al. [53] developed multilayer semantic interaction leveraging dynamic transformers. This innovation achieves parallel multi-scale attention feature computation by integrating multi-head self-attention with deep separable convolution, further advancing the fusion capabilities. Jiang et al. [54] proposed a novel technique rooted in a multi-scale neural network,

complemented by a SpatialSwin autoencoder-based matting for MFIF. This approach adeptly tackles challenges such as boundary precision and texture preservation, underscoring its robustness and practicality. In theory, deep learning–based methods can achieve better fusion performance than traditional methods in image fusion tasks. However, existing deep learning–based methods have not shown significant advantages over traditional methods. The main reason for this phenomenon is that the existing training data is poor, and large-scale, real benchmark data is lacking.

In addition, SR has also been applied to image fusion in the past ten years due to its powerful capability of image representation. In SR-based methods, the way of constructing a dictionary and the activity measurement of each image patch play important roles in fusion performance. Generally, a dictionary can be obtained by pre-constructing-based and learning-based methods. The first uses fixed bases to construct a dictionary. For example, Yang et al. used a discrete cosine transform basis as the dictionary for image fusion and super-resolution [55]. Dong et al. presented a fusion method based on a curvelet transform dictionary [56]. The second method uses a pre-collected image set or the source images themselves to learn an overcomplete dictionary using some training algorithms. The learned overcomplete dictionary can be further classified as a globally trained dictionary or an adaptively trained dictionary. In [57], Liu et al. presented an MST and SR-based fusion method, in which numerous high-quality nature images were used to learn a global dictionary. Zhu et al. constructed a discriminative dictionary by combining several compact sub-dictionaries [58]. In [59], Zhang et al. proposed a multi-task robust sparse representation model to fuse images with misregistration, and the source images were used to build the adaptive dictionary. In [60], Yin et al. learned several adaptive sub-dictionaries from the source image itself and then combined them to construct a joint dictionary for MFIF. The pre-constructing dictionary-based SR model is fast to implement but is restricted to the contents of the image. Compared with pre-constructing dictionaries, the basis atoms of learned dictionaries have richer information and stronger representation capability for source images. Thus, the learned dictionary–based fusion methods usually produce better fusion results than fixed dictionary–based ones.

The activity measurement of each image patch is also important for the fusion performance. The widely used activity measurement is defined as the l_1-norm of sparse coefficients of the local patch [55,57,58]. In addition, some novel activity measurements are also designed to select the focused image patches from source images. For example, Nejati et al. defined an activity level of each patch by using the correlation between sparse coefficients and the training pooled features [10]. Yin et al. used a weighted multi-norm of sparse coefficients as the activity measure for each patch [60]. Zhang et al. distinguished the focused and defocused patches from source images by employing the sparse reconstruction errors of the SR model [59,61]. Recently, they designed a focus measure by collaboratively using the sparse coefficients and sparse errors of each super-pixel [62].

Although these SR-based methods obtained good performance, there are still several limitations in MFIF. First, most methods focus on fusing grayscale source images, but little attention has been paid to designing specific SR models for color MFIF. For color source images, there are three main methods of processing used in the sparse coding process. The first way is to separate the color image into multi-channel ones and then sparsely encode each color channel independently. The second way is to concatenate all color channels to form a large monochromatic image and then perform sparse coding on it. Both ways ignore the strong correlation among color channels, which may cause color bias in the results. The third way is to transform the RGB color space into some independent space (such as YCbCr and YUV space) and then perform sparse coding on the luminance channel (Y

channel). However, how to effectively fuse the chrominance components of source images is another problem. In recent years, quaternion theory has been introduced to solve classical color image processing problems. The commonly used way of quaternion representation (QR) of color images is to represent each color pixel as a pure quaternion. Then, we can apply fruitful quaternion algebra theories for image processing tasks. This representation has some attractive merits, such as processing all channels simultaneously, conveniently performing transformations in 3D space, and retaining the high-level correlation of all channels. Owing to these superiorities of the QR, numerous QR-based processing methods have been developed [63–69]. For example, Xu et al. combined the merits of SR and QR to develop a QSR model and achieved exciting results for color image denoising [65]. In [68], Zou et al. proposed a quaternion collaborative representation model with successful application to face recognition. To explore the underlying low-rank property of the quaternion matrix, Chen et al. proposed a low-rank quaternion approximation model for color image denoising and inpainting [69]. These methods achieved exciting results for many classic color image processing tasks, but few QR-based fusion methods have been reported.

The second limitation is that the design of the activity measurement of sparse coefficients of each patch did not adequately consider spatial information. Most of them are only designed on the local patch, while the information of neighboring patches is ignored. Thus, these activity measurements may lack adequate ability to differentiate between the focused and defocused patches. The fusion methods in [59,61,62] designed activity measurements by employing spatial contextual information, but they suffered from "jagged" artifacts and too smooth transitional regions in the results, which are inherent problems of spatial domain–based methods. In fact, the principle of symmetry can serve as a guideline for identifying and highlighting significant features of the source image that exhibit certain structural or pattern characteristics. For instance, when processing images containing object edges or textures, if these elements display some form of symmetry, they are likely to carry crucial visual information. Leveraging the principle of symmetry can assist algorithms in more effectively detecting these features and ensuring their prominent representation in the fused image.

Concerning the limitations mentioned above, we propose a novel fusion framework based on the QSR model to generate a fully focused image, which consists of the following four steps:

(1) Two different quaternion dictionaries, namely a clear quaternion dictionary and a blurry quaternion dictionary, are learned from a pre-collected high-quality image set and its Gaussian blurred version by employing the K-quaternion singular value decomposition (K-QSVD) method. The reason for constructing two different dictionaries is as follows: The dictionaries used in existing SR-based methods are learned from the clear image patch set or the mixed image patch set (a mixture of clear image patches and blurry image patches). These image patch sets may have smooth data; thus, the learned dictionaries usually contain both sharp components and smooth components. In the sparse coding process, the clear patch and blurry patch will be decomposed into a weighted summation of atoms with similar numbers. This weakens the ability of the sparsity feature (l_0-norm of sparse coefficients) to differentiate between the focused and defocused patches. In the following experiments, we found that when coding each image patch and the corresponding blurred patch under the blurry quaternion dictionary, the respective required number of atoms shows quantitatively different results. This diverged effect indicates that the required number of atoms in the sparse coding process can characterize the focus information of each patch. Therefore, we will use this clue to design a simple but expressive feature for focused patch detection.

(2) Each source image is divided into a set of image patches by the sliding-window operator, and then each patch is represented by the quaternion vector form. After that, the clear quaternion sparse coefficients and the blurry ones of each image patch are estimated by a simultaneous quaternion orthogonal matching pursuit (SQOMP) algorithm and a quaternion orthogonal matching pursuit (QOMP) algorithm, respectively.
(3) The salience feature and sparsity feature of each image patch are computed from the clear quaternion sparse coefficients and blurry quaternion sparse coefficients, respectively. Then, a new activity measurement of each patch is computed by exploiting both salience and sparsity features. In the process of feature extraction, we evaluate the focus information of each patch by jointly using the local information of the current patch and its neighboring patches. Thus, the extracted features can properly differentiate between focused and defocused patches.
(4) A maximum fusion rule is adopted to combine the clear quaternion sparse coefficients of source images, and then we can precisely reconstruct the fused image using the composited sparse coefficients and the clear quaternion dictionary. A schematic diagram of our method is shown in Figure 1. The contributions are elaborated as follows:

(i) We present a novel color MFIF via a QSR model. To the best of the author's knowledge, this is the first time that the QSR is used to solve the problem of color MFIF. Unlike most SR-based methods handling color channels separately, the QSR combines the merits of QR and SR, which can transform all color channels into a sparse space uniformly and can preserve the inherent color structures of source images completely in reconstruction results. Therefore, our method can successfully prevent color distortion in the fused images.

(ii) Two discriminating features for each image patch are computed from the quaternion sparse coefficients deduced from the QSR model under the specifically learned quaternion dictionaries. The feature calculation process fully considers the spatial neighboring information to enhance its discriminant power. Moreover, the salience feature and sparsity feature are unitized to calculate the activity measurement, which can comprehensively evaluate the clarity attributes of each patch.

(iii) Experimental results show that our method can preserve well the important features of source images and can successfully avoid color distortion in the fused images. In addition, our method performs favorably against some recent SR-based methods in subjective perception and objective assessment.

The remaining contents are arranged as follows: Section 2 briefly reviews the basic concepts of quaternion algebra and the related theory of QSR. Section 3 describes the color MFIF via the QSR model and spatial information in detail. Sections 4 and 5 provide experimental results and concluding remarks, respectively.

2. Preliminaries

This section introduces some background knowledge, including quaternion algebra, QR of color images, and the related theory of QSR. Within this paper, we refer to scalar variables, vector variables, and matrix variables in the real space ($\mathbb{R}$) using normal letters (e.g., x), boldface lowercase letters (e.g., $\mathbf{x}$), and bold capital letters (e.g., $\mathbf{X}$), respectively. In the quaternion space ($\mathbb{H}$), we use a dot above variables to represent the quaternion variables, e.g., $\dot{x}, \dot{\mathbf{x}}$ and $\dot{\mathbf{X}}$.

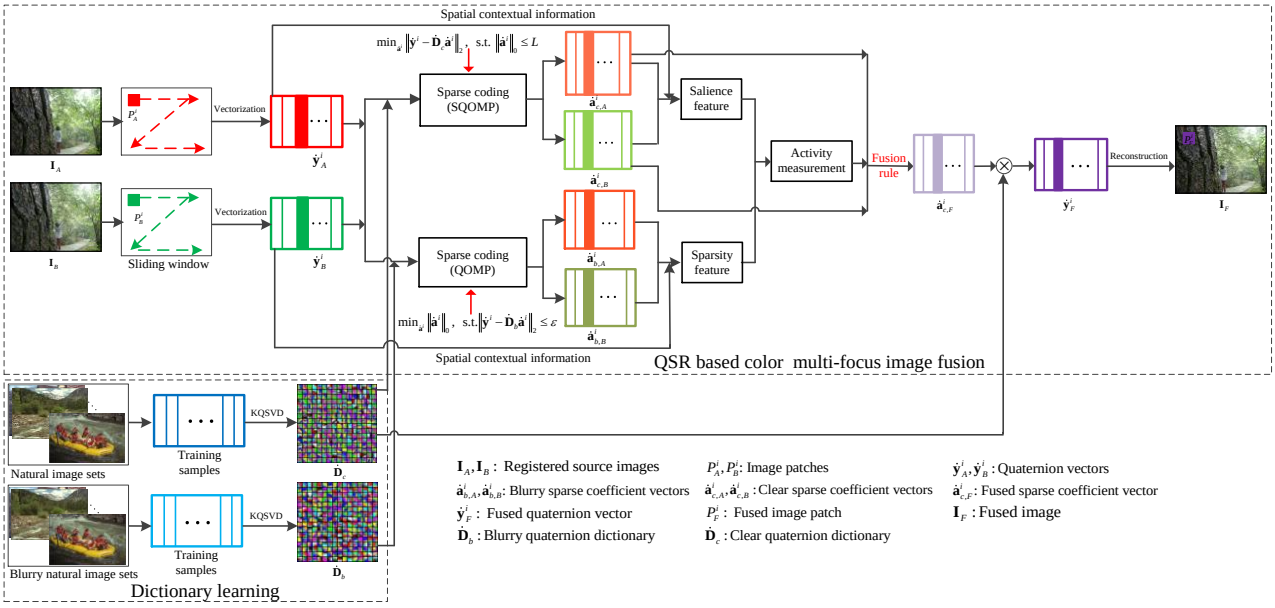


Figure 1. The framework of our color MFIF, using the QSR model and spatial information. Input all the source images, with each one initially being divided into a series of image patches and subsequently expressed in the form of quaternion vectors. Subsequently, the clear quaternion sparse coefficients, along with their corresponding blurry counterparts for each image patch, are estimated utilizing the SQOMP and QOMP algorithms, respectively, based on the two pre-trained quaternion dictionaries. Next, an activity measurement for each patch is computed by leveraging both the salience feature and the sparsity feature, which are derived from the quaternion sparse coefficients. Finally, the fused image is constructed by using the composited sparse coefficients and the clear quaternion dictionary.

2.1. Quaternion algebra

Hamilton extended the 2D complex number into a 4D hypercomplex number, thereby introducing the notion of quaternions [70]. A quaternion $\dot{q} \in \mathbb{H}$ is defined as follows:

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

where a is called the real part of $\dot{q}$, $b\mathrm{i}+c\,\mathrm{j}+d\,\mathrm{k}$ is the imaginary part, and $a,b,c,d \in \mathbb{R}$, i,j and k are operation units of the quaternion, which follow the rules that $\mathrm{i}^2 = \mathrm{j}^2 = \mathrm{k}^2 = \mathrm{ijk} = -1$. If the real part $a = 0$, $\dot{q}$ is called a pure quaternion.

Let $\dot{q}_1, \dot{q}_2 \in \mathbb{H}$, $\lambda \in \mathbb{R}$, $\dot{\mathbf{p}} = [\dot{p}_1, \dot{p}_2, \ldots, \dot{p}_N]^T \in \mathbb{H}^N$ and $\dot{\mathbf{q}} = [\dot{q}_1, \dot{q}_2, \ldots, \dot{q}_N]^T \in \mathbb{H}^N$. Then, we present some fundamental algebraic operations of quaternion systems as follows.

(1) Addition: $\dot{q}_1 + \dot{q}_2 = (a_1 + a_2) + (b_1 + b_2)\mathrm{i} + (c_1 + c_2)\,\mathrm{j} + (d_1 + d_2)\,\mathrm{k}$.

(2) Multiplication: $\lambda\dot{q}_1 = \lambda a_1 + \lambda b_1\mathrm{i} + \lambda c_1\,\mathrm{j} + \lambda d_1\,\mathrm{k}$,

$$\begin{aligned}\dot{q}_1\dot{q}_2 = {} & (a_1a_2 - b_1b_2 - c_1c_2 - d_1d_2) + (a_1b_2 + b_1a_2 + c_1d_2 - d_1c_2)\mathrm{i} \\ & + (a_1c_2 - b_1d_2 + c_1a_2 + d_1b_2)\,\mathrm{j} + (a_1d_2 + b_1c_2 - c_1b_2 + d_1a_2)\,\mathrm{k}\end{aligned}.$$

In general, the multiplication between two quaternions $\dot{q}_1$ and $\dot{q}_2$ is non-commutative, i.e., $\dot{q}_1\dot{q}_2 \neq \dot{q}_2\dot{q}_1$. This is a key difference between the quaternion system and the complex system.

(3) The conjugate, modulus, and inverse of the quaternion $\dot{q}_1$ are defined as:

$\bar{\dot{q}}_1 = a_1 - (b_1\,\mathrm{i} + c_1\,\mathrm{j} + d_1\,\mathrm{k})$, $|\dot{q}_1| = \sqrt{\dot{q}_1\bar{\dot{q}}_1} = \sqrt{\bar{\dot{q}}_1\dot{q}_1} = \sqrt{a_1^2 + b_1^2 + c_1^2 + d_1^2}$ and $\dot{q}_1^{-1} = \frac{\bar{\dot{q}}_1}{|\dot{q}_1|^2}$.

(4) The inner product of two quaternion vectors $\dot{\mathbf{p}}, \dot{\mathbf{q}}$ is:

$<\dot{\mathbf{p}}, \dot{\mathbf{q}}> = \dot{\mathbf{p}}^H \dot{\mathbf{q}} = \sum_{n=1}^{N} \dot{p}_n \dot{q}_n$.

where $\dot{\mathbf{p}}^H = [\bar{\dot{p}}_1, \bar{\dot{p}}_2, \ldots, \bar{\dot{p}}_N]$ is the conjugate transpose of $\dot{\mathbf{p}}$.

(5) The l_0-, l_1-, and l_2- norms of the quaternion vector $\dot{\mathbf{p}}$ are:

$\|\dot{\mathbf{p}}\|_0 = \#(n \,|\, |\dot{p}_n| \neq 0)$, $\|\dot{\mathbf{p}}\|_1 = \sum_{n=1}^{N} |\dot{p}_n|$, and $\|\dot{\mathbf{p}}\|_2 = (\sum_{n=1}^{N} (|\dot{p}_n|)^2)^{\frac{1}{2}}$.

More theories of quaternion algebra can be found in [49].

2.2. Quaternion representation of color image

To capture the high correlation between color channels, the pure quaternion matrix is widely used to represent color image $\mathbf{I}$ [63–69], namely,

$$\dot{\mathbf{Q}} = \mathbf{R}\,\mathrm{i} + \mathbf{G}\,\mathrm{j} + \mathbf{B}\,\mathrm{k} \tag{2}$$

where $\dot{\mathbf{Q}}$ is the QR of $\mathbf{I}$, and $\mathbf{R}, \mathbf{G}$, and $\mathbf{B}$ are the red, green, and blue channel matrices, respectively. It can be seen from Eq. (2) that using the QR can transform a given color image into a quaternion matrix uniquely. Thus, the algebraic operations of the quaternion are capable of handling all color channels simultaneously, which can preserve the inherent color structures of source images well in image processing tasks.

2.3. Quaternion representation of color image

The fundamental idea of SR is that a specified signal can be constructed as a weighted summation of a *few* atoms from an overcomplete dictionary. Due to the powerful capability of image representation, many SR-based image processing tasks have achieved thrilling results [55–62]. The existing SR-based methods perform well for grayscale images, but the performance degrades significantly when processing color images. The reason is that the traditional SR model treats color channels separately or stacks them as a vector. These two ways are inconsistent with the mechanism of the human visual system, which essentially handles all color channels in parallel. To better capture the inter-relationship among color channels, Xu proposed a QSR model and achieved exciting results for color image denoising [65]. The QSR employs the quaternion matrix to model color images in a holistic way, which can perfectly preserve the inherent color structures of original images in the reconstruction results. Thus, in this paper, we employ it for color MFIF. The QSR model will be briefly introduced.

To boost the computational efficiency, the QSR model usually focuses on patch-based processing. For a color image $\mathbf{I}$, let $\mathbf{p}$ be a color image patch of size $\sqrt{n} \times \sqrt{n}$ in image $\mathbf{I}$; then, the quaternion vector form of $\mathbf{p}$ is denoted as $\dot{\mathbf{v}} = \mathbf{0} + \mathbf{v}_r\,\mathrm{i} + \mathbf{v}_g\,\mathrm{j} + \mathbf{v}_b\,\mathrm{k}$, $\dot{\mathbf{v}} \in \mathbb{H}^n$, where $\mathbf{v}_c \in \mathbb{R}^n, c = r, g, b$ is the real vector form of patch $\mathbf{p}$ for each channel. More formally, the QSR model has a linear form as:

$$\dot{\mathbf{v}} = \dot{\mathbf{D}}\dot{\mathbf{s}} \tag{3}$$

where $\dot{\mathbf{D}} = \mathbf{D}_o + \mathbf{D}_r\,\mathrm{i} + \mathbf{D}_g\,\mathrm{j} + \mathbf{D}_b\,\mathrm{k}$, $\dot{\mathbf{D}} \in \mathbb{H}^{n \times K} (n < K)$ is an overcomplete quaternion dictionary containing K quaternion atoms, and $\dot{\mathbf{s}} = \mathbf{s}_0 + \mathbf{s}_1\,\mathrm{i} + \mathbf{s}_2\,\mathrm{j} + \mathbf{s}_3\,\mathrm{k}$, $\dot{\mathbf{s}} \in \mathbb{H}^K$ is the unknown quaternion sparse coefficient vector. Generally, the solution of Equation (3) is not unique because the dictionary is overcomplete. The goal

of QSR is to obtain a solution that contains the smallest number of non-zero entries. Mathematically, the sparest solution $\dot{\mathbf{s}}$ of Equation (3) can be solved by:

$$\dot{\mathbf{s}} = \arg\min_{\dot{\mathbf{s}}} \left\| \dot{\mathbf{v}} - \dot{\mathbf{D}}\dot{\mathbf{s}} \right\|_2, \ \text{s.t.} \left\| \dot{\mathbf{s}} \right\|_0 \le T \tag{4}$$

or

$$\dot{\mathbf{s}} = \arg\min_{\dot{\mathbf{s}}} \left\| \dot{\mathbf{s}} \right\|_0, \ \text{s.t.} \left\| \dot{\mathbf{v}} - \dot{\mathbf{D}}\dot{\mathbf{s}} \right\|_2 \le \varepsilon \tag{5}$$

where T and ε are two types of terminal criteria. The QOMP algorithm usually solves the above NP-hard problem because of its high efficiency [65].

Rewriting the QSR model in Equation (3), we can obtain that the quaternion coefficient vector obeys the following constraints:

$$\begin{cases} \mathbf{0} = \mathbf{D}_o\mathbf{s}_0 - \mathbf{D}_r\mathbf{s}_1 - \mathbf{D}_g\mathbf{s}_2 - \mathbf{D}_b\mathbf{s}_3 \\ \mathbf{v}_r = \mathbf{D}_o\mathbf{s}_1 + \mathbf{D}_r\mathbf{s}_0 + \mathbf{D}_g\mathbf{s}_3 - \mathbf{D}_b\mathbf{s}_2 \\ \mathbf{v}_g = \mathbf{D}_o\mathbf{s}_2 - \mathbf{D}_r\mathbf{s}_3 + \mathbf{D}_g\mathbf{s}_0 + \mathbf{D}_b\mathbf{s}_1 \\ \mathbf{v}_b = \mathbf{D}_o\mathbf{s}_3 + \mathbf{D}_r\mathbf{s}_2 - \mathbf{D}_g\mathbf{s}_1 + \mathbf{D}_b\mathbf{s}_0 \end{cases} \tag{6}$$

Equation (6) enforces explicit constraints on the correlations between color channels. Different from the traditional SR models that select atoms from the respective channel dictionary, each $\mathbf{v}_c$ is linearly dependent on four channel dictionaries in the QSR model. Moreover, the coefficient vector should be in the null space of $[\mathbf{D}_o, \mathbf{D}_r, \mathbf{D}_g, \mathbf{D}_b]$. By learning the quaternion dictionary $\dot{\mathbf{D}}$ in an appropriate way, the inter-relationship among all color channels for each patch $\mathbf{p}$ can be well preserved. Thus, the inherent color structures of source images can be well preserved during the reconstruction process, which is very useful to color MFIF.

3. Color multi-focus image fusion based on quaternion sparse representation and spatial information

In this section, we will use the QSR model to sparse code each color source image, which is attributed to the clear advantage over the traditional SR model. Apart from the selection of a sparse model, the constructed dictionary, the activity measurement of each image patch, and the coefficient integration rule are the other key issues in representation-based fusion methods, which will affect the final fusion performance. In the following subsections, we will discuss in detail how to solve these issues.

3.1. Quaternion dictionary constructing

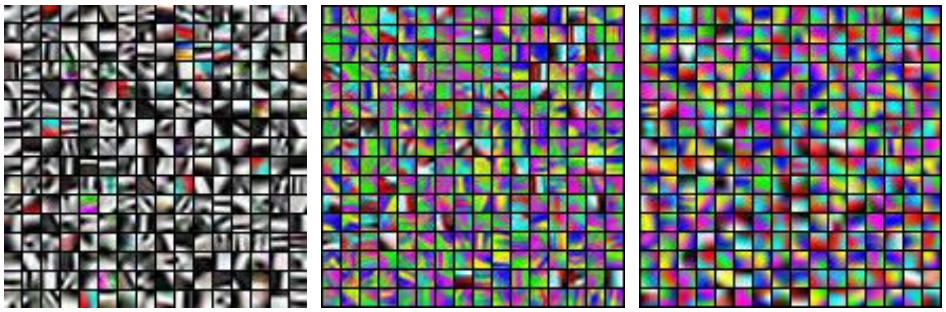
The representation capability of the SR model is determined by the over-dictionary. Generally, the dictionary can be obtained by the pre-constructing-based method and the learning-based method. Compared with the pre-constructing-based dictionary, the learned dictionary has richer information on basis atoms and has better adaptability for images with various contents. Thus, in this paper, we use the learning technique to construct the quaternion dictionary.

For that, we first sample N training image patches $\mathbf{p}_i, i = 1, \ldots, N$ of size $\sqrt{n} \times \sqrt{n}$, which are randomly cropped from a series of high-quality color images. Then, each patch $\mathbf{p}_i$ is normalized to zero mean value and is transformed into vector form $\mathbf{v}_i$ via lexicographic ordering. Now, we get a training matrix $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N] \in \mathbb{R}^{n \times N \times 3}$. Let the QR of the training matrix be $\dot{\mathbf{V}} = [\dot{\mathbf{v}}_1, \dot{\mathbf{v}}_2, \ldots, \dot{\mathbf{v}}_N] \in \mathbb{H}^{n \times N}$. The training process of the quaternion dictionary is formulated as a below-optimization problem by extending the model in Equation (3):

$$\left\{\hat{\dot{\mathbf{D}}},\hat{\dot{\mathbf{S}}}\right\}=\arg\min_{\dot{\mathbf{D}},\dot{\mathbf{S}}}\left\|\dot{\mathbf{V}}-\dot{\mathbf{D}}\dot{\mathbf{S}}\right\|_2+\lambda\left\|\dot{\mathbf{S}}\right\|_0 \tag{7}$$

where $\dot{\mathbf{D}}=[\dot{\mathbf{d}}_1,\dot{\mathbf{d}}_2,\ldots,\dot{\mathbf{d}}_K]\in\mathbb{H}^{n\times K}$ denotes the quaternion dictionary matrix containing K quaternion atoms, $\dot{\mathbf{S}}=[\dot{\mathbf{s}}_1,\dot{\mathbf{s}}_2,\ldots,\dot{\mathbf{s}}_N]\in\mathbb{H}^{K\times N}$ is the quaternion sparse coefficient matrix in which each $\dot{\mathbf{s}}_i$ is the representation coefficient to reconstruct $\dot{\mathbf{v}}_i$, and $\lambda>0$ is the regularization parameter. In Eq (7), the dictionary matrix and coefficient matrix are both unknown variables. There are several algorithms to solve the above optimization problem, such as ML [71], MOD [72], and K-SVD [73]. Among them, the most frequently used is the K-SVD algorithm because of its high computational efficiency. However, the original K-SVD algorithm is designed for a real number space. In [65], Xu et al. proposed a K-QSVD algorithm that extends the standard K-SVD algorithm into the quaternion space. Here, we also employ the K-QSVD algorithm to train the optimal quaternion dictionary. The detailed computation process can be found in [65].

Figure 2 gives a comparison of color dictionaries trained by the K-SVD and K-QSVD algorithms with different SR models. In this experiment, 100,000 color image patches with a size of 8×8 are first randomly cropped from 20 natural color images. Then, we train dictionaries with 256 atoms by using different SR models. The maximum number of atoms used in patch decomposition is set to 16. To visualize the learned quaternion dictionary, we enforce the real part $\dot{\mathbf{D}}_o=\mathbf{0}$ by using the linear correlation of four-channel dictionaries. The dictionaries learned by the K-SVD algorithm using the concatenation model [74] and by the K-QSVD algorithm using the quaternion sparse model are illustrated in Figure 2(a) and (b), respectively. It can be seen from Figure 2 that these two dictionaries contain many edge-like components, reasonably representing the spatial structure. However, the atoms of the dictionary learned by the K-SVD algorithm are colorless, which indicates that they are not plentiful enough to represent various colors. Thus, bias and color washing issues will be introduced in the reconstruction results. Compared with the dictionary learned by the K-SVD algorithm, the atoms of the quaternion dictionary contain rich color information, which encodes spatial morphologies and diversified colors well. Meanwhile, we train a specific dictionary on the Gaussian blurred version ($\sigma=2$) of the same set of color image patches and call it a blurry quaternion dictionary. The blurry quaternion dictionary is shown in Figure 2(c), which presents distinctly different structures from the quaternion dictionary trained on clear image patches and contains almost no sharp patterns. This difference between the two quaternion dictionaries reveals that the blur effect of the image will influence the fundamental atoms in the training results. Moreover, the clear and blurry quaternion dictionaries are not interchangeable when reconstructing image patches.



(a) (b) (c)

Figure 2. Visualizing different dictionaries for color image sparse representation. (a) Using the K-SVD algorithm learned dictionary on the clear image dataset. (b) Using the K-QSVD algorithm learned clear quaternion dictionary on the high-quality image dataset. (c) Using the K-QSVD algorithm learned blurry quaternion dictionary on the Gaussian blurred image dataset.

3.2. Quaternion sparse representation for color multi-focus image

Considering that the MFIF method usually relies on local information, a patch-wise processing strategy with a QSR model is implemented, similar to existing SR-based methods. To make the SR shift invariant, a sliding-window technique is utilized to extract the image patches, which is very important to the fusion performance. In existing SR-based fusion methods, sparse coding for each patch is commonly performed by using only one learned dictionary that contains both sharp components and smooth components. The clear patch and blurry patch will be decomposed with a similar number of atoms, which will be weakened by the discrimination ability of the sparsity feature (l_0-norm of sparse coefficients). Moreover, as mentioned above, the clear and blurry quaternion dictionaries contain distinctly different structures of atoms and are not interchangeable when performing patch reconstruction. Therefore, it was unwise to use only one dictionary for representing image patches in the MFIF. In the following QSR-based color MFIF method, we first learn two quaternion dictionaries, namely the clear and blurry quaternion dictionaries, to sparsely encode each image patch and then try to construct two features from different aspects to comprehensively reflect the significant information of each patch, which is one main difference from existing SR-based fusion methods.

Suppose $\mathbf{I}_j \in \mathbb{R}^{W\times H\times 3}, j=1,\ldots,J$ are J geometrically registered color source images. By applying the sliding-window technique with a step size of stp pixels, we can extract all possible patches of size $\sqrt{n}\times\sqrt{n}$ for each source image. Let $\{\mathbf{p}_j^i\}_{i=1}^{L}$ denote all the extracted patches of source image $\mathbf{I}_j$, where $L = nw\times nh$, $nw = \begin{cases} sw+1, & \text{if } ((W-\sqrt{n}+1)-((sw-1)\times stp+1))>0 \\ sw, & \text{otherwise} \end{cases}$, $nh = \begin{cases} sh+1, & \text{if } ((H-\sqrt{n}+1)-((sh-1)\times stp+1))>0 \\ sh, & \text{otherwise} \end{cases}$, $sw = \left\lceil \frac{W-\sqrt{n}+1}{stp} \right\rceil$ and $sh = \left\lceil \frac{H-\sqrt{n}+1}{stp} \right\rceil$, and $\lceil\cdot\rceil$ denotes a ceiling function. To facilitate the analysis, each patch $\mathbf{p}_j^i$ of the source image $\mathbf{I}_j$ is first subtracted from its mean $\mathbf{m}_j^i$ and then transformed into vector form $\mathbf{v}_j^i$ via lexicographic ordering. After that, we obtain the QR of $\mathbf{v}_j^i$ as $\dot{\mathbf{v}}_j^i$ and calculate the clear and blurry quaternion sparse coefficient vectors by:

$$\hat{\dot{\mathbf{s}}}_{c,j}^i = \arg\min_{\dot{\mathbf{s}}_{c,j}^i} \left\| \dot{\mathbf{v}}_j^i - \dot{\mathbf{D}}_c \dot{\mathbf{s}}_{c,j}^i \right\|_2, \ \text{s.t.} \left\| \dot{\mathbf{s}}_{c,j}^i \right\|_0 \le T \tag{8}$$

and

$$\hat{\dot{\mathbf{s}}}_{b,j}^i = \arg\min_{\dot{\mathbf{s}}_{b,j}^i} \left\| \dot{\mathbf{s}}_{b,j}^i \right\|_0, \ \text{s.t.} \left\| \dot{\mathbf{v}}_j^i - \dot{\mathbf{D}}_b \dot{\mathbf{s}}_{b,j}^i \right\|_2 \le \varepsilon \tag{9}$$

where $\dot{\mathbf{D}}_c$ and $\dot{\mathbf{D}}_b$ are the clear and blurry quaternion dictionaries learned on the high-quality image set and the Gaussian blurred version of the high-quality image set, respectively, and $\dot{\mathbf{s}}_{c,j}^i$ and $\dot{\mathbf{s}}_{b,j}^i$ are the clear and blurry quaternion sparse coefficient vectors for the i-th patch $\mathbf{p}_j^i$. The QOMP algorithm [65] can efficiently solve the above problem of signal decomposition. However, for the QOMP algorithm, different source images may be decomposed into the weighted summation of different subsets of a given dictionary atom. This is analogous to decomposing the corresponding image patches of source images by different transform tools, which might invalidate the frequently used salience feature (l_1-norm of sparse coefficients). For the MFIF, we expect that the corresponding patches of source images are represented by the same subset of dictionary atoms when solving Equation (8). The simultaneous orthogonal matching pursuit (SOMP) algorithm [75] can meet this requirement, which uses the same set of dictionary atoms to represent image patches from different source images. In this paper, we first extend the SOMP algorithm into quaternion space to devise an

SQOMP algorithm and then utilize the SQOMP to estimate the clear quaternion sparse coefficients $\{\dot{\mathbf{s}}_{c,j}^{i}\}_{j=1}^{J}$ for the J quaternion vectors of image patches $\{\dot{\mathbf{v}}_{j}^{i}\}_{j=1}^{J}$. The main steps of SQOMP are given in Algorithm 1. The SQOMP terminates precisely after T iterations. Nevertheless, it is noteworthy that the stopping criterion typically encompasses two components: one based on the iteration count and another contingent upon the norm of the residual. Specifically, if the norm of the residual falls below a predefined threshold, the iteration ceases. Various norms can be employed for the latter criterion. For further reading on this topic, the interested reader is referred to the related literature [75]. From Algorithm 1, it can be found that the QOMP algorithm is a special case of the SQOMP algorithm when $J=1$. The solution of Equation (9) is more inclined to obtain the minimum error between the original and reconstructed signals; thus, the QOMP algorithm is used to estimate the blurry quaternion sparse coefficients $\{\dot{\mathbf{s}}_{b,j}^{i}\}_{j=1}^{J}$.

Algorithm 1. Simultaneous quaternion orthogonal matching pursuit algorithm

Input: Quaternion dictionary $\dot{\mathbf{D}} \in \mathbb{H}^{n\times K}$, quaternion signals $\{\dot{\mathbf{v}}_j\}_{j=1}^{J}$, $\dot{\mathbf{v}}_j \in \mathbb{H}^n$, the maximum number of non-zero sparse coefficients T, a stopping threshold δ.

Output: Quaternion sparse coefficients $\{\dot{\mathbf{s}}_j\}_{j=1}^{J}$, $\dot{\mathbf{s}}_j \in \mathbb{H}^K$.

1. Initialization: Residual signal $\dot{\mathbf{r}}_j^{(0)} = \dot{\mathbf{v}}_j, j=1,\ldots,J$, atom set $\dot{\mathbf{A}}^{(0)} = \varnothing$, and the iteration counter $itr = 1$.

2. Iteration procedure:

(1) For every atom $\dot{\mathbf{d}}_t \in \dot{\mathbf{D}} \setminus \dot{\mathbf{A}}^{(itr-1)}$, compute the correlation value between $\dot{\mathbf{d}}_t$ and the residual signal of last iteration $\dot{\mathbf{r}}_j^{(itr-1)}$:

$$C_t^{(itr)} = \sum_{j=1}^{J} \left\| \left\langle \dot{\mathbf{d}}_t, \dot{\mathbf{r}}_j^{(itr-1)} \right\rangle \right\| = \sum_{j=1}^{J} \left\| \dot{\mathbf{d}}_t^{H} \dot{\mathbf{r}}_j^{(itr-1)} \right\|.$$

(2) Find an atom $\dot{\mathbf{d}}^{(itr)}$ that solves the easy optimization problem:

$\hat{t}^{(itr)} = \arg\max_t C_t^{(itr)}$, $\dot{\mathbf{d}}^{(itr)} = [\dot{\mathbf{D}} \setminus \dot{\mathbf{A}}^{(itr-1)}]_{\hat{t}_{itr}}$.

This optimization process ensures that all signals can be perfectly reconstructed simultaneously.

(3) Update the atom set: $\dot{\mathbf{A}}^{(itr)} = \left[\dot{\mathbf{A}}^{(itr-1)}, \dot{\mathbf{d}}^{(itr)}\right]$.

(4) Calculate new sparse coefficients by projecting each input quaternion signal onto the atom set $\dot{\mathbf{A}}^{(itr)}$:

$\dot{\mathbf{s}}_j^{(itr)} = \arg\min_{\dot{\mathbf{s}}_j} \left\| \dot{\mathbf{v}}_j - \dot{\mathbf{A}}^{(itr)} \dot{\mathbf{s}}_j \right\| = ((\dot{\mathbf{A}}^{(itr)})^H \dot{\mathbf{A}}^{(itr)})^{-1} (\dot{\mathbf{A}}^{(itr)})^H \dot{\mathbf{v}}_j = (\dot{\mathbf{A}}^{(itr)})^{\triangle} \dot{\mathbf{v}}_j$, for $j=1,\ldots,J$.

where the superscript $\triangle$ represents the operator of quaternion pseudo-inverse.

(5) Update the residual signal: $\dot{\mathbf{r}}_j^{(itr)} = \dot{\mathbf{v}}_j - \dot{\mathbf{A}}^{(itr)} \dot{\mathbf{s}}_j^{(itr)}$, for $j=1,\ldots,J$.

(6) Increment the iteration counter $itr = itr+1$, and go back to step (1) unless the stopping criterion is met, i.e., $\sum_{j=1}^{J} \left\| \dot{\mathbf{r}}_j^{(itr)} \right\|_2 \le \delta$ or $itr <= T$.

3.3. Fusion rule

In this subsection, we will solve the remaining two key issues of the SR-based methods, namely the activity measurement of each image patch and the coefficient combination rule. The activity measurement identifies the saliency of sparse coefficients, and the combination rule decides how much useful information of source images is transferred to the composited image. In the SR domain, the l_1 -norm of sparse coefficients of each patch reflects how much detail-level structural information it

contains. A larger value indicates that the image patch brings more saliency information. Thus, the l_1-norm of sparse coefficients is commonly used as the activity measurement of each patch [55,57,58]. Meanwhile, the l_0-norm of sparse coefficients means how many atoms contribute to represent each image patch. It seems more reasonable that different features are extracted from the sparse coefficients to assess the significance of the corresponding image patch. In this paper, we build two features, namely salience feature and sparsity feature, to assess the significance of image patches. The salience feature is computed by the l_1-norm of sparse coefficients obtained by Equation (8). The sparsity feature is described by the l_0-norm of sparse coefficients over a specific dictionary trained on blurry data, which is different from traditional methods. Additionally, the calculation process of features fully considers spatial information to improve its robustness. To comprehensively assess the focus information of each patch, we design a new activity measurement by exploiting both salience and sparsity features. The detailed computation process of activity measurement is discussed as follows.

For source image $\mathbf{I}_j$, we first extract all possible image patches $\left\{\mathbf{p}_j^i\right\}_{i=1}^L$ by the sliding window operator and then transform them into vector form $\left\{\mathbf{v}_j^i\right\}_{i=1}^L$ in the same way as in the previous subsection, where L is the number of extracted patches. For the i-th image patch $\mathbf{p}_j^i$, we calculate the clear and blurry quaternion sparse coefficient vectors $\dot{\mathbf{s}}_{c,j}^i$ and $\dot{\mathbf{s}}_{b,j}^i$ by using Equation (8) and Equation (9), respectively. Considering that the l_1-norm of clear quaternion sparse coefficient vectors $\dot{\mathbf{s}}_{c,j}^i$ can represent the sharpness degree of the image patch and that the neighboring patches generally possess similar focus information in multi-focus images, the salience feature $\mathbf{SA}_j^i$ of patch $\mathbf{p}_j^i$ is built as:

$$\mathbf{SA}_j^i = \left\|\dot{\mathbf{s}}_{c,j}^i\right\|_1 + \sum_{\mathbf{p}_j^k \in \Gamma(\mathbf{p}_j^i)} \omega_{i,k} \left\|\dot{\mathbf{s}}_{c,j}^k\right\|_1 \tag{10}$$

where $\Gamma(\mathbf{p}_j^i)$ is the set of 8-connected neighboring patches of $\mathbf{p}_j^i$, and the weight $\omega_{i,k}$ is calculated by:

$$\omega_{i,k} = \begin{cases} \exp\left(-\dfrac{\left\|\mathbf{v}_j^i - \mathbf{v}_j^k\right\|_2}{2\sigma^2}\right), & \text{if } \mathbf{p}_j^k \in \Gamma(\mathbf{p}_j^i) \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

where σ is a kernel parameter and is usually set to 0.5 [76]. The salience feature defined in Equation (10) fully takes into account the spatial contents, which can improve its discriminant power.

Moreover, the l_0-norm of sparse coefficients reflects the concentration ratio of elementary information of the image patch, and its value indicates how many atoms are used to represent the image patch. The l_0-norm-based feature may produce ambiguity when distinguishing the focused and defocused patches. This will make the frequently used method invalid in which sparse coefficients with a larger number of non-zero elements are selected as fused coefficients. The rationale is that a dictionary learned on clear training data possesses both sharp and smooth components, and the required number of atoms for reconstructing clear image patches will be similar to that for reconstructing the corresponding blurry patch. In [77], Shi et al. proposed an effective sparsity feature for detecting just noticeable blur. Inspired by the idea of [77], we learn a specific dictionary on blurry data and build a discriminate feature for measuring the significance of image patches. In this dictionary, there are numerous blurry bases and barely sharp components. Thus, a large number of atoms may be needed when reconstructing a clear patch with sharp structures by blurry bases. In contrast, a blurry patch requires fewer atoms for optimal reconstruction. To prove this hypothesis, we conduct two experiments by counting the number of non-zero elements for representing image patches.

The first experiment is conducted on a clear image patch with a size of 8×8 and the Gaussian blurred version with different blur degrees. The blurry quaternion dictionary used for sparse coding

is shown in Figure 2(c). The decomposition results are shown in Figure 3. We can see from Figure 3 that the number of atoms used to represent the original clear patch is almost as many as its dimension. As the blur strength becomes severe, the variation in patch content decreases. The required number of atoms for representing these patches is reduced accordingly.

The generality of the phenomenon shown in Figure 3 is validated in the second experiment. We first randomly crop 100,000 color patches with a size of 8×8 from 50 natural color images. To avoid ambiguous results, patches with small variance are removed. Then, the sampled patches are smoothed with Gaussian standard deviations ranging from 0.5 to 5. In total, we collect approximately 750,000 image patches with different blur degrees. Figure 4 lists the numbers of non-zero sparse coefficients under different blurriness levels. In Figure 4, the height of each bar represents the mean value of non-zero sparse coefficients corresponding to a particular blurriness, and the gray line above each bar denotes the standard deviation. We can see from Figure 4(a) that the value dramatically decreases when $\sigma \le 2$, and the values for each blur degree are fairly consistent under a small standard variation. In addition, we also use a clear quaternion dictionary, which is shown in Figure 2(b), to represent these sampled patches, and the results are shown in Figure 4(b). It is easily found that the rate of descent of values in Figure 4(b) is slower than that in Figure 4(a). This manifests in that the extracted feature using the blurry quaternion dictionary is more sensitive to blur strength than the one using the clear quaternion dictionary. We believe this is because the blurry quaternion dictionary captures more elementary information to represent blurred image patches. The second experiment statistically proves the robust and stable relationship between the blur degree and the value of non-zero sparse coefficients for image patch representation using a blurry quaternion dictionary. Thus, we utilize this cue as the blur indicator and build the sparsity feature $\mathbf{SP}_j^i$ of patch $\mathbf{p}_j^i$ by:

$$\mathbf{SP}_j^i = \left\| \dot{\mathbf{s}}_{b,j}^i \right\|_0 + \sum_{\mathbf{p}_j^k \in \Gamma(\mathbf{p}_j^i)} \omega_{i,k} \left\| \dot{\mathbf{s}}_{b,j}^k \right\|_0 \tag{12}$$

Now, the new activity measurement $\mathbf{AM}_j^i$ of patch $\mathbf{p}_j^i$ is obtained by linearly combining the salience feature and sparsity feature:

$$\mathbf{AM}_j^i = n\mathbf{SA}_j^i \times n\mathbf{SP}_j^i \tag{13}$$

where $n\mathbf{SA}_j^i$ and $n\mathbf{SP}_j^i$ are the normalized values of $\mathbf{SA}_j^i$ and $\mathbf{SP}_j^i$, respectively. After the constructed activity measurement of the image patch, the following issue is how to merge sparse coefficients of source images into the counterparts of the composited image. The typical combination rules are the averaging-based and absolute maximum-based rules. The former rule selects the fused coefficients by weighted averaging all the sparse coefficients of source images. It can resist the noise in the fusion process, but the detailed structures of fused images will be smoothed. For the color MFIF, we expect that all the valuable information from source images is transferred into the fused image. In reality, establishing such an ideal rule is almost impossible. A more practical approach is preserving the most important information among source images into the fused image. Therefore, the maximum activity measurement rule is employed to choose fused coefficients in this paper. The fused clear quaternion coefficient vector $\dot{\mathbf{s}}_{c,F}^i$ of the i-th image patch is obtained by choosing the clear quaternion coefficient vector of the source image with the maximum activity measurement as follows:

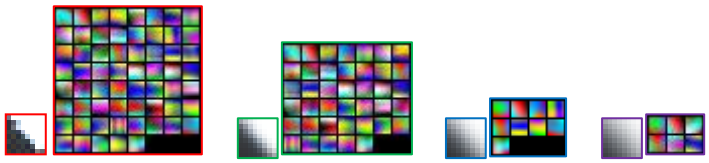
$$\dot{\mathbf{s}}_{c,F}^i = \dot{\mathbf{s}}_{c,j^*}^i, \quad j^* = \arg\max_j \{\mathbf{AM}_j^i, j = 1, \ldots, J\} \tag{14}$$

The corresponding QR of the fused mean value is $\dot{\mathbf{m}}_F^i = [\dot{m}_{j^*}^i, \dot{m}_{j^*}^i, \ldots, \dot{m}_{j^*}^i]^T \in \mathbb{H}^n$. Now, the quaternion vector form of the i-th fused image patch is:

$$\dot{\mathbf{v}}_F^i = \dot{\mathbf{D}}_c \dot{\mathbf{s}}_{c,F}^i + \dot{\mathbf{m}}_F^i \tag{15}$$

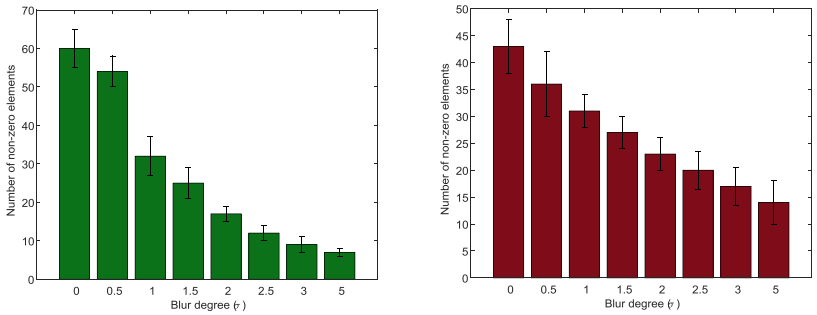
After obtaining all the quaternion vector forms of fused patches $\left\{\dot{\mathbf{v}}_F^i\right\}_{i=1}^L$, we transform them into a quaternion matrix $\dot{\mathbf{I}}_F$ by using the inverse process of patch extraction. Since the fused patches

overlap, each value of $\dot{\mathbf{I}}_F$ is divided by its accumulation times. Finally, the color fused image $\mathbf{I}_F$ is obtained by extracting the imaginary part of $\dot{\mathbf{I}}_F$.



(a) (b) (c) (d)

Figure 3. The atoms used for representing image patches with different blur degrees. The left part in each sub-image is the image patch to be decomposed, and the right part is the used atoms. (a) Original patch, 61 atoms. (b) Standard deviation $\sigma = 1$, 40 atoms. (c) Standard deviation $\sigma = 2$, 9 atoms. (d) Standard deviation $\sigma = 3$, 6 atoms.



(a) (b)

Figure 4. The average numbers of non-zero sparse coefficients under different blurriness levels. (a) The decomposition of image patches using a blurry quaternion dictionary. (b) The decomposition of image patches using a clear quaternion dictionary.

4. Experiments

This section provides extensive experiments to validate the feasibility of our color MFIF. First, we present the corresponding experimental setting. Second, the influences of some key parameters on the performance of our method are discussed. Finally, the experimental results are analyzed subjectively and objectively.

4.1. Experimental setting

To validate the effectiveness and robustness of our method, experiments have been conducted on four publicly accessible datasets. The test datasets comprise the Lytro dataset [10], consisting of 24 pairs of authentic multi-focus images derived from light field data, the MFFW dataset [78], which contains 13 pairs of multi-focus images of varying resolutions, the MFI-WHU dataset [79], produced from blurred all-focus images and comprising 120 pairs of synthetic multi-focus images, and the Real-MFF dataset [80], generated from light field images and containing an extensive 710 pairs of multi-focus images. Additionally, four sets were collected from the Internet [81]. Source images with different contents stand for various cases encountered in practice. A portion of source image pairs are displayed in Figure 5.

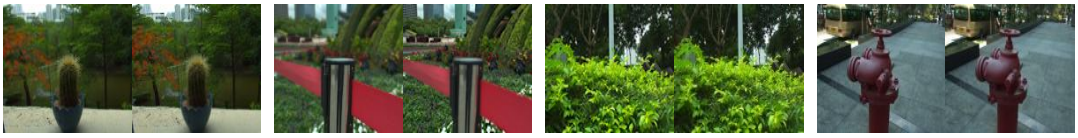
We select 13 popular MFIF methods in the comparison experiments to demonstrate the superiority of our method (QSR method, for short). Specifically, we use one classical MST-based method, the nonsubsampled contourlet transform-based method (NSCT) [21], and six recently representation-based methods, which are the sparse representation-based method (SR) [82], the nonsubsampled contourlet transform and sparse representation-based method (NSCT-SR) [57], the sparse representation-based method with an adaptive dictionary learned on pre-collection training data (ASR-1) [83], the sparse representation-based method with an adaptive dictionary learned on source images themselves (ASR-2) [60], the jointly clustered patch dictionary-based method (JCPD) [84], and the dictionary learning and low-rank representation-based method (DL-LRR) [85]. Additionally, we use three effective spatial domain-based methods, including the multiscale fuzzy quality-based method (MFQ) [14], the multi-dictionary linear sparse representation-based method (MDLSR) [86], and the parameter adaptive dual channel dynamic threshold neural P systems-based method (PADCDTNPS) [16], as well as three effective deep learning–based methods, namely the CNN-based method (CNN) [26], the swin transformer network-based method (SWIN) [87], and the latent feature-guided diffusion Transformer model-based method (LFDT) [88]. The parameters of the comparison methods are the same as those in related papers. All the selected representation-based methods fuse color channels separately, except for the JCPD method. The JCPD method first converts the RGB color space into YCbCr color space and then performs a fusion rule on the Y channel. Finally, the inverse color space transform is performed over the merged luminance component and chrominance components to obtain the fused image.

To quantitatively assess the quality of fused images obtained by different methods, six mainstream assessment metrics are used in the experiments: the information theory–based metric (Q_{NMI}) [89], the feature-based metric (Q_G) [90], the structural similarity–based metric (Q_Y) [91], the human perception–based metric (Q_{CB}) [92], the visual information fidelity–based metric (Q_{VIFF}) [93], and the colorfulness index of color image (Q_{CF}) [94]. The first four metrics can estimate the amount of information, sharpness information, structural information, and significant features of source images preserved in the fused results, respectively. The metric Q_{VIFF}, which is grounded in the human visual system and natural scene statistics theory, serves to quantify the visual information fidelity between the fused image and each of the source images. The metric Q_{CF} uses a combination of image statistics to represent colorfulness. For each metric, a higher value implies superior performance.

(a) The Lytro dataset

(b) The MFFW dataset

(c) The MFI-WHU dataset



(d) The Real-MFF dataset

Figure 5. Examples of some source image sets from the selective datasets.

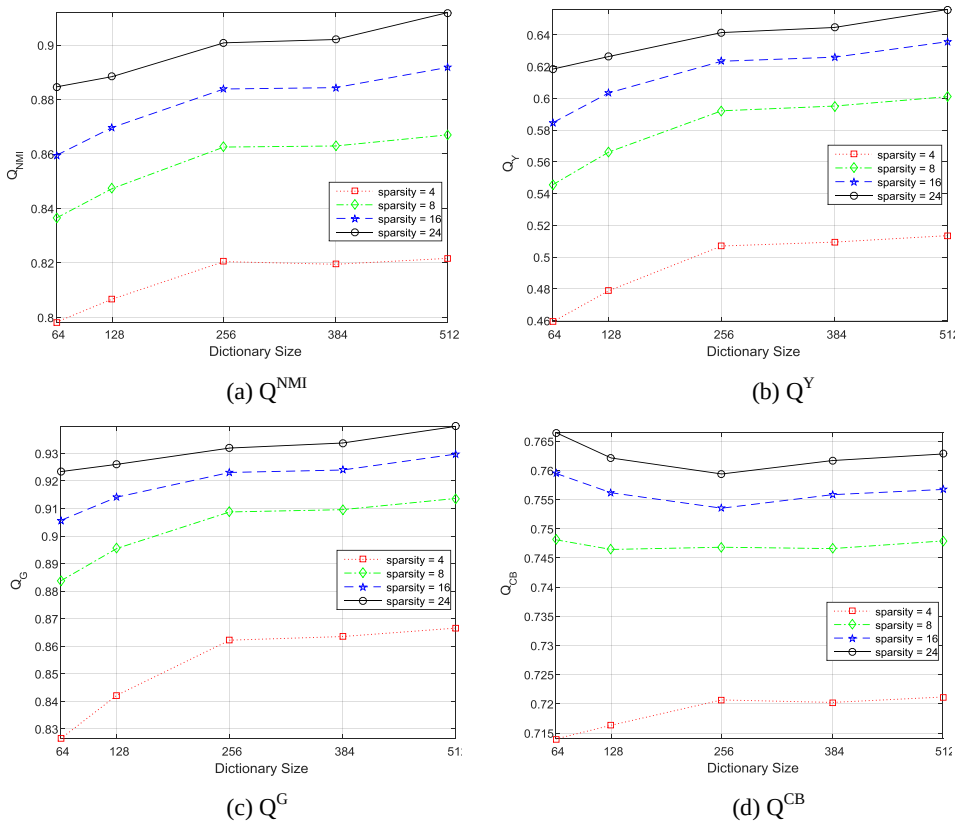
4.2. Effects of parameters

In this subsection, we will determine two key parameters in our method, including the number of dictionary atoms K when constructing dictionary and the sparsity level T in Equation (8). To accomplish this, the experiments are conducted on the Lytro dataset, and the impacts of parameters on the fusion results are assessed by using the averaged values of four metrics.

When training a quaternion dictionary, we first require collecting training data that consist of many image patches. As the size of the patch increases, the size of the dictionary increases accordingly. Thus, the efficiency of sparse representation for image patches becomes lower. If the size of the patch is too small, the containing information will not be sufficient for distinguishing whether it is a focused patch or not. As suggested in the literature [55,58,60,73], the size of the patch is also set to 8×8 in the constructing dictionary. In the experiment, the clear training data consist of 100,000 color image patches, which are randomly cropped from 20 high-resolution color images. These color images are obtained from [95]. The blurred training data are obtained by blurring the clear training data with a Gaussian kernel of $\sigma = 2$. We try to produce different blurred data by smoothing the clear data with different Gaussian variances. It is found that this configuration is sufficiently feasible when extracting the sparsity feature. Then, we train a set of clear and blurry quaternion dictionaries, where the numbers of atoms K are set to 64, 128, 256, 384, and 512, respectively. We also test four different sparsity levels T in the experiment, including 4, 8, 16, and 24. The assessment results of the QSR method based on these dictionaries for all testing images are shown in Figure 6.

It can be seen from Figure 6 that almost all of the metrics increase with the number of dictionary atoms, while the improvements in the values are not obvious when the number of atoms $K > 256$ increases. We note that learning a dictionary with a large number of atoms may have some side-effects. The first is that the fusion methods using this dictionary will demand higher computational costs. The second is that the dictionary may over-fit the training data. Therefore, the number of atoms is set to $K = 256$. Moreover, all metrics always increase with sparsity levels. However, a large sparsity level means that more atoms are used to represent image patches, which also increases the computational cost. As shown in Figure 6, all metrics achieve nice results when the sparsity level T is set to 16. Apart from the above two key parameters, there are another two parameters in the QSR method, namely the step size of the sliding-window operator stp and the global error ε in Equation (9). To reduce spatial artifacts and blocking effects, the step size is usually set to one pixel in many SR-based fusion methods. These methods will require large memory storage and high computational cost. To make full use of spatial information in extracting features, we find that a proportion of overlapped region reaching 50% of each image patch can obtain satisfactory fusion results. This setting can effectively reduce computational complexity. For the global error ε, we select a small value, expecting that the reconstruction results are close to the original image patches. The reason is that the detail-level structure information is critical to perceive image blur. We set the global error ε to 0.07, as suggested in the literature [77].

In summary, the parameters of the QSR method are set as $K = 256$, $stp = 4$, $T = 16$, and $\varepsilon = 0.07$ in the following experiments.



(a) Q^{NMI} (b) Q^{Y}

(c) Q^{G} (d) Q^{CB}

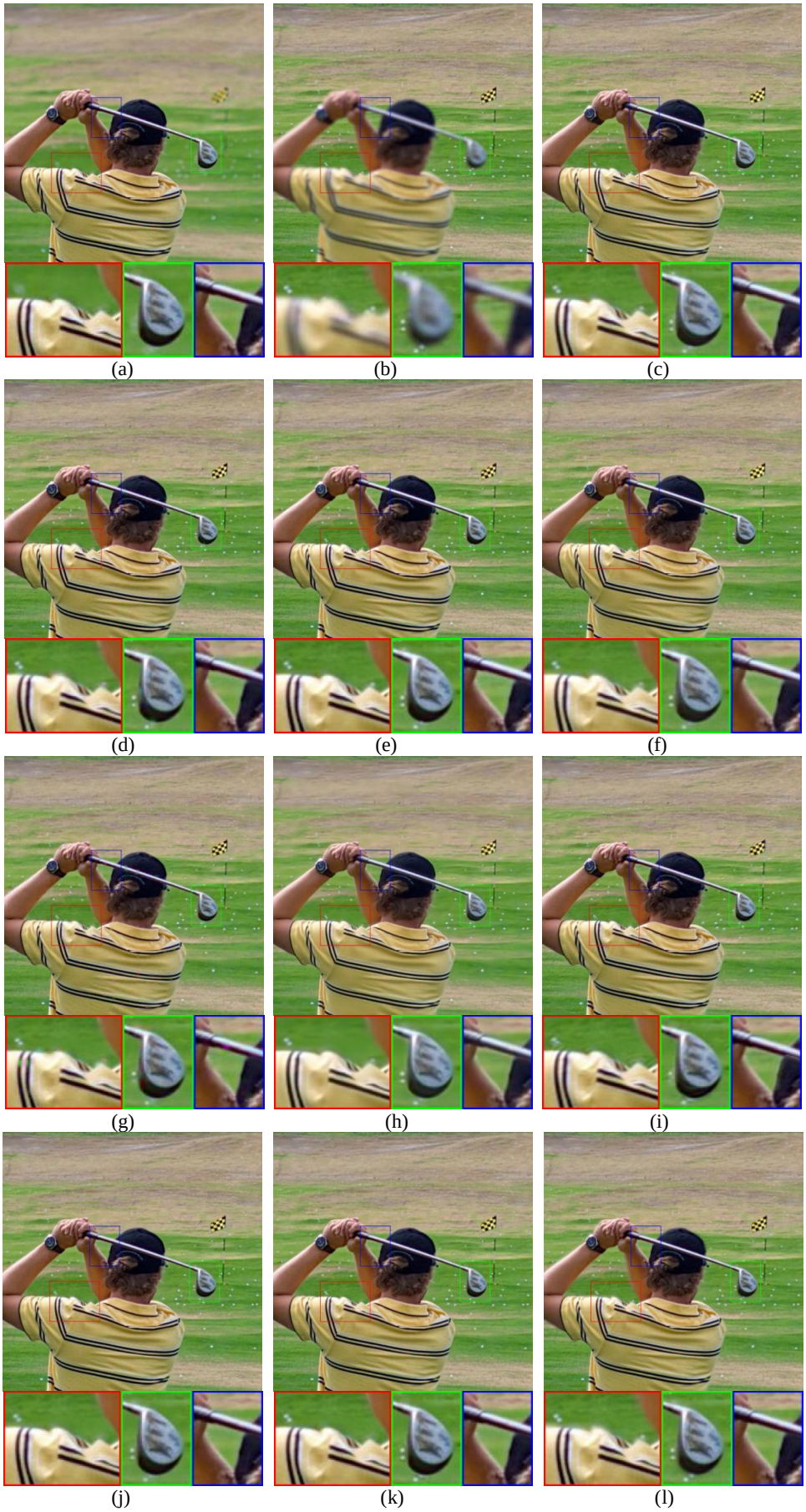
Figure 6. Mean assessment results vs. the size of the dictionary and sparsity level.

4.3. Experimental results and discussion

4.3.1. Qualitative comparisons

In this subsection, we first give the qualitative comparison of different fusion methods by comparing the visual quality of fused images. To validate the effectiveness and universality of the QSR method, eight representative experiments are chosen for comparative analysis. The selected source images contain various types, such as objects of different scales, different brightness variations, different number of source images, and good or poor registration. The fused results achieved through various methods are depicted in Figures 7–19. To clearly show the details of fused images, the local enlarged patches are also presented below each image.

Figure 7 displays the fused results for the Golfer source images from the Lytro dataset, a challenging scenario due to a small focused region (a triangular meadow) surrounded by a large defocused area and small-scale golf balls located near the boundaries of the foreground object. While most methods appear to perform well overall, several exhibit specific drawbacks: the SR, ASR-2, and DL-LRR methods introduce artifacts at object boundaries, with SR and ASR-2 also showing a hue bias; the JCPD method produces an overly smooth result due to its average fusion rule. A closer examination reveals that the PADCDTNPS, SWIN, LFDT, and QSR methods are more effective at preserving the detail of the golf balls near the foreground, whereas this information is eroded to varying degrees by the other techniques. Furthermore, when it comes to reproducing the focus information of the triangular meadow from the second source image, only the SWIN and QSR methods succeed, as all other methods fail to capture this detail correctly.

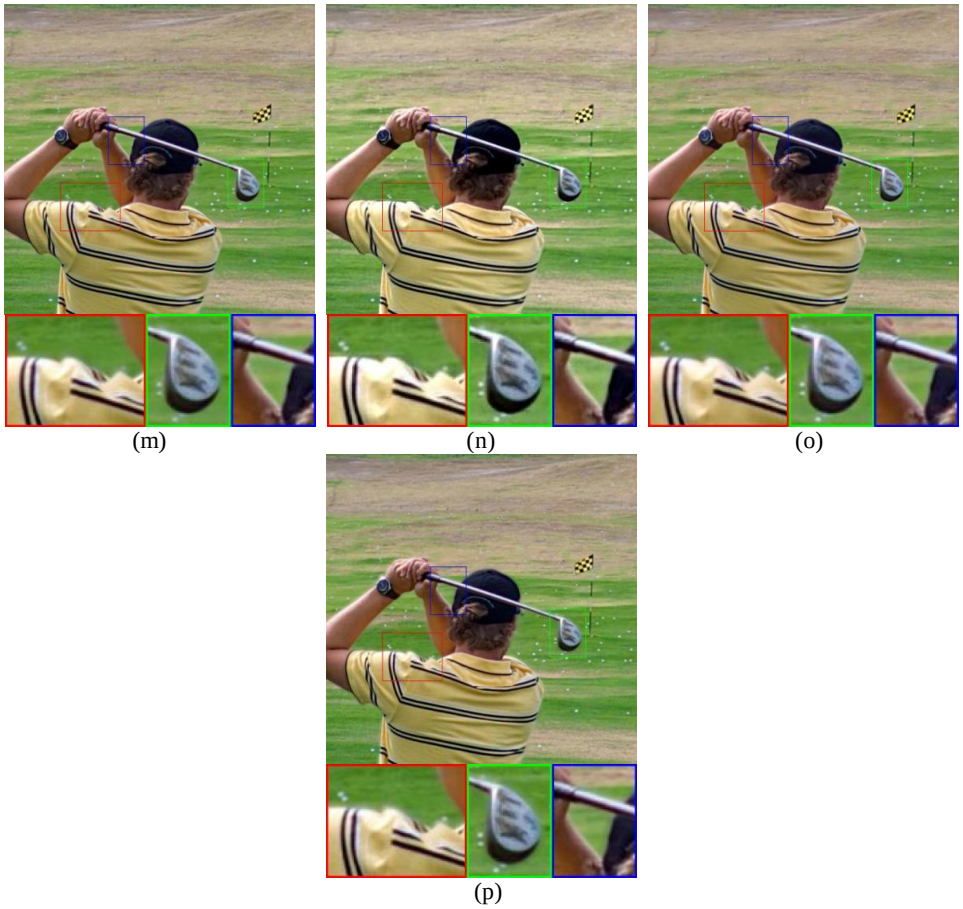


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

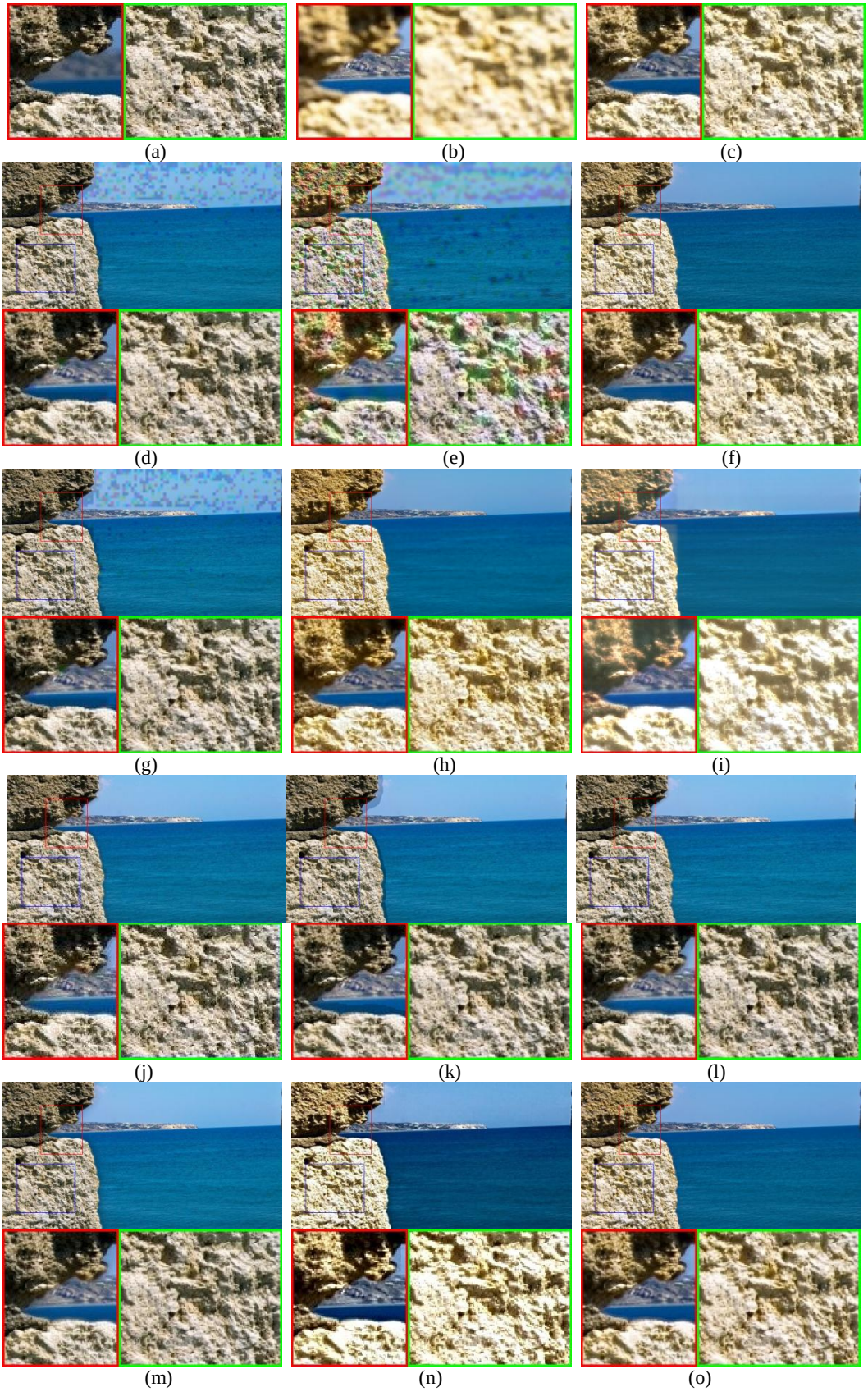
(j) (k) (l)



(m) (n) (o) (p)

Figure 7. Illustration of the fused images on the Golfer set from the Lytro dataset. (a)–(b) Source images. (c)–(p) Fusion results obtained by the NSCT, SR, NSCT-SR, ASR-1, ASR-2, JCPD, DL-LRR, MFQ, MDLSR, PADCDTNPS, CNN, SWIN, LFDT, and QSR, respectively.

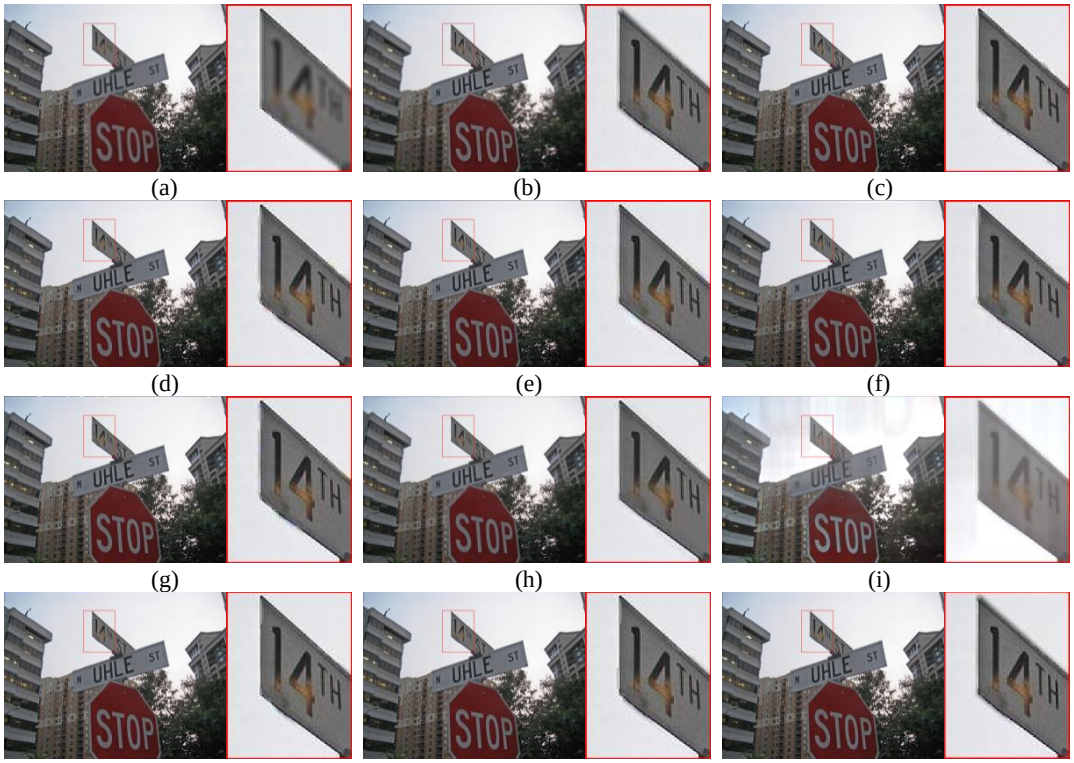
Figure 8 displays the fused results for the Sea source images from the MFFW dataset, a challenging case involving images captured under different brightness conditions and a defocused region with rich gradient information, particularly noticeable in the stone part. The JCPD and DL-LRR methods suffer from severe color bias on these stones, a flaw attributed to their reliance on absolute coefficients as focus measures, which leads to ambiguous recognition in areas with complex gradients in both the foreground and background; JCPD's visual quality is further degraded by its use of only one chrominance component. While the SR, NSCTSR, and ASR-2 methods introduce few artifacts in the stone area, they produce obvious artifacts in the sky regions. In contrast, the SWIN, LFDT, and QSR methods demonstrate superior performance by preserving more focus information from the source images, resulting in a fused output with clear and well-defined transitional regions.


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)
(m)
(n)
(o)

(p)

Figure 8. Illustration of the fused images on the Sea set from the MFFW dataset. (a)–(b) Source images. (c)–(p) Fusion results obtained by the NSCT, SR, NSCT-SR, ASR-1, ASR-2, JCPD, DL-LRR, MFQ, MDLSR, PADCDTNPS, CNN, SWIN, LFDT, and QSR, respectively.

Figure 9 displays the fused results for the Traffic sign source images from the MFI-WHU dataset, which are impeccably aligned with minimal defocus diffusion. The JCPD method produces a poor visual effect, as shown in Figure 9(i). Meanwhile, the ASR-1, ASR-2, and PADCDTNPS methods introduce artificial segmentation boundaries at the transitions between focused and defocused areas. Although the MFQ, MDLSR, and CNN methods avoid significant misjudgments in large background areas, their segmentation boundaries lack the necessary fineness. In contrast, the SWIN, LFDT, and QSR methods stand out by effectively preserving all critical information from the source images while ensuring sharp boundary contours, ultimately achieving the highest quality in the fused results.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

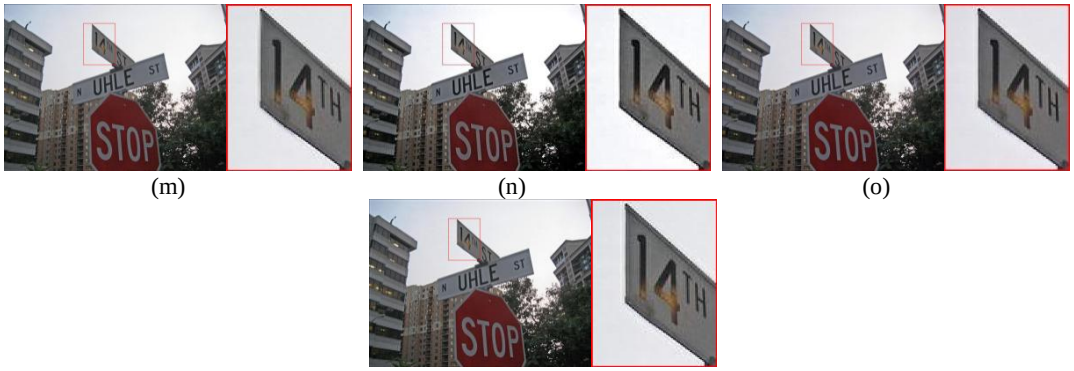


Figure 9. Illustration of the fused images on the Traffic sign set from the MFI-WHU dataset. (a)–(b) Source images. (c)–(p) Fusion results obtained by the NSCT, SR, NSCT-SR, ASR-1, ASR-2, JCPD, DL-LRR, MFQ, MDLSR, PADCDTNPS, CNN, SWIN, LFDT, and QSR, respectively.

Figure 10 displays the fused images for the Flower source images from the Real-MFF dataset, with locally enlarged versions presented in Figure 11. The source images are perfectly aligned but feature irregular edges on the flower and branches, posing a significant challenge. Several methods, including ASR-2, MFQ, MDLSR, PADCDTNPS, and CNN, struggle to accurately differentiate between focused and defocused areas, resulting in prominent blurred edges, particularly on the branches. The MDLSR method, as shown in Figure 10(k), exhibits especially severe edge blurring. While the SR-based methods show some improvement in reducing this blurring, their results still fall short of ideal. In contrast, the SWIN, LFDT, and QSR methods produce the most vivid and visually pleasing fusion images, successfully preserving the intricate details and sharp contours of the source images.

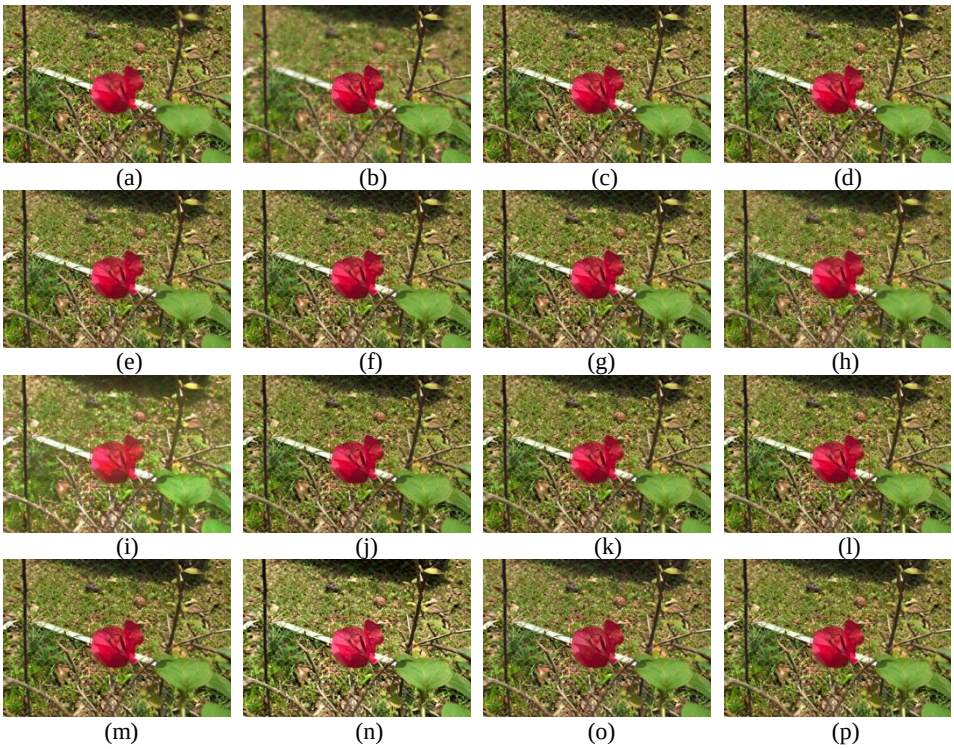


Figure 10. Illustration of the fused images on the Flower set from the Real-MFF dataset. (a)–(b) Source images. (c)–(p) Fusion results obtained by the NSCT, SR, NSCT-SR, ASR-1, ASR-2, JCPD, DL-LRR, MFQ, MDLSR, PADCDTNPS, CNN, SWIN, LFDT, and QSR, respectively.

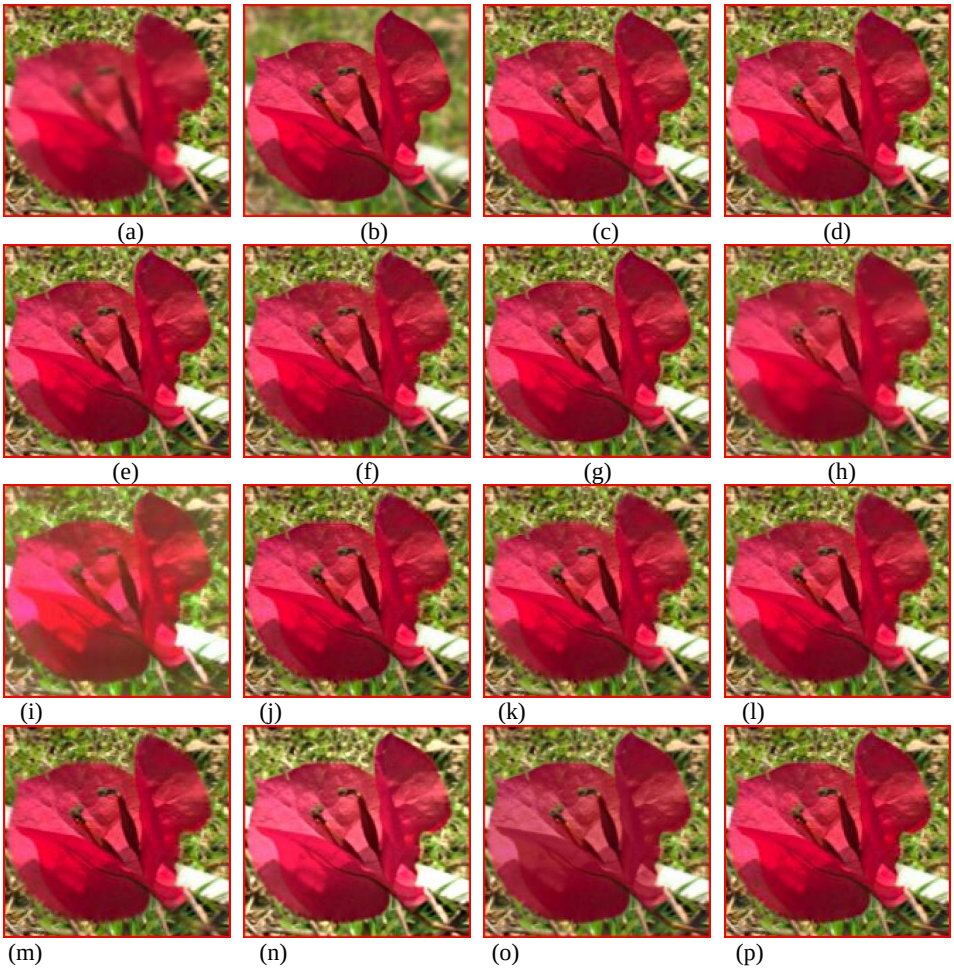


Figure 11. Local enlarged regions of the Flower fused images in Figure 10. (a)–(b) Source images. (c)–(p) Fusion results obtained by the NSCT, SR, NSCT-SR, ASR-1, ASR-2, JCPD, DL-LRR, MFQ, MDLSR , PADCDTNPS, CNN, SWIN, LFDT, and QSR, respectively.

The problem of image misregistration is a main challenge for the MFIF task, especially for the transform domain-based methods. Two fusion experiments are presented in the following, conducted on source images with misregistration caused by different reasons.

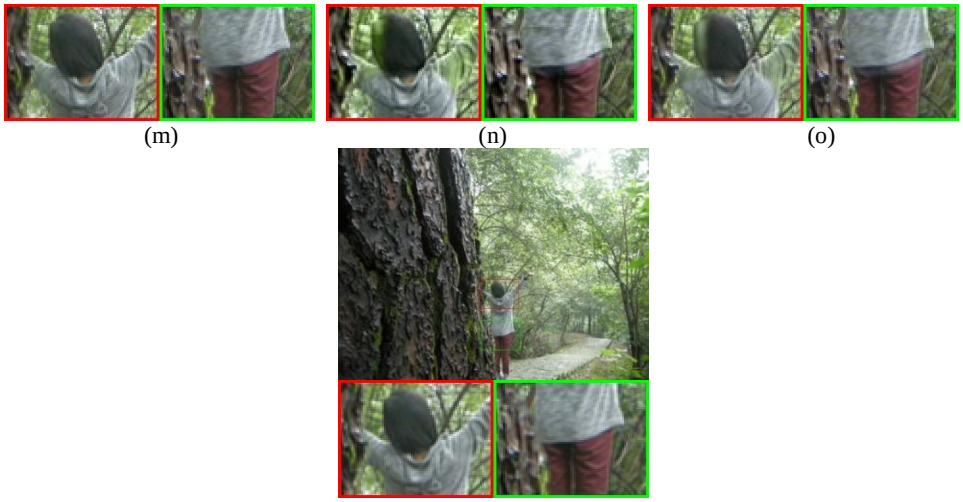
The first experiment, conducted on the Girl source images shown in Figure 12(a) and (b), presents a significant challenge due to misregistration caused by movement in the arm and head. Generally, spatial domain–based methods outperform transform domain–based ones, with the notable exception of the QSR method. As seen in the enlarged regions, transform domain–based methods like NSCT, NSCT-SR, and ASR-1 produce obvious artifacts on the head and body. While the sliding-window technique in the SR and ASR-2 methods somewhat alleviates these issues, artifacts persist. Furthermore, the JCPD and DL-LRR methods exhibit noticeable color bias. In contrast, spatial domain methods such as MFQ, MDLSR, and PADCDTNPS perform well, yielding clear transitional regions. Ultimately, the CNN, SWIN, LFDT, and QSR methods achieve the best visual effects. Notably, the QSR method is particularly effective, as its designed features capture the intrinsic focus characteristics of the source images, allowing it to robustly resist the performance degradation caused by moving objects.

(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)



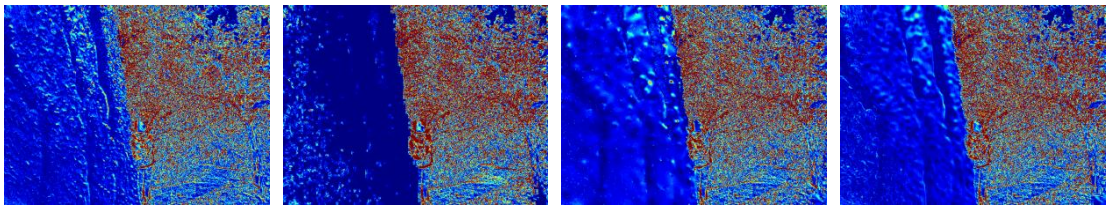
(m) (n) (o) (p)

Figure 12. Illustration of the fused images on the Girl set. (a)–(b) Source images. (c)–(p) Fusion results obtained by the NSCT, SR, NSCT-SR, ASR-1, ASR-2, JCPD, DL-LRR, MFQ, MDLSR, PADCDTNPS, CNN, SWIN, LFDT, and QSR, respectively.

To further assess the quality of the fused images, we also provide difference images, serving as a means to evaluate the effectiveness of fusion methods in retaining the information from the source images. The difference image $\mathbf{I}_D$ is given by the following formula:

$$\mathbf{I}_D = \gamma \left| \mathbf{I}_F - \mathbf{I}_n \right| \tag{16}$$

where $\gamma > 1$ represents a parameter designed to present a more straightforward result; in our experiments, we have fixed $\gamma = 10$. Additionally, $\mathbf{I}_F$ and $\mathbf{I}_n$ are the fused image and nth source image. In the difference image, the pixels are anticipated to appear red in the defocused areas and blue in the focused areas. The difference images in Figure 13, using source image 1 [Figure 12(a)] as the reference, highlight the ability of fusion methods to transfer focus information. Source image 1 features a clear foreground and a blurry background. The MFQ and CNN methods yield satisfactory results, demonstrating their capability to precisely discern focused areas by using a binary decision map that minimizes residual information in the difference images. In contrast, other methods lose some degree of focus information. However, because the focused region in the lower-right corner of source image 1 is small and dispersed, the CNN method inevitably retains some residual focusing information, leading to artifacts or blurring in the fused image. Although our method's difference image also retains some information from the focused area, it excels at recognizing even these small, scattered regions, which consequently results in a fused image with a superior visual effect.



(a) (b) (c) (d)

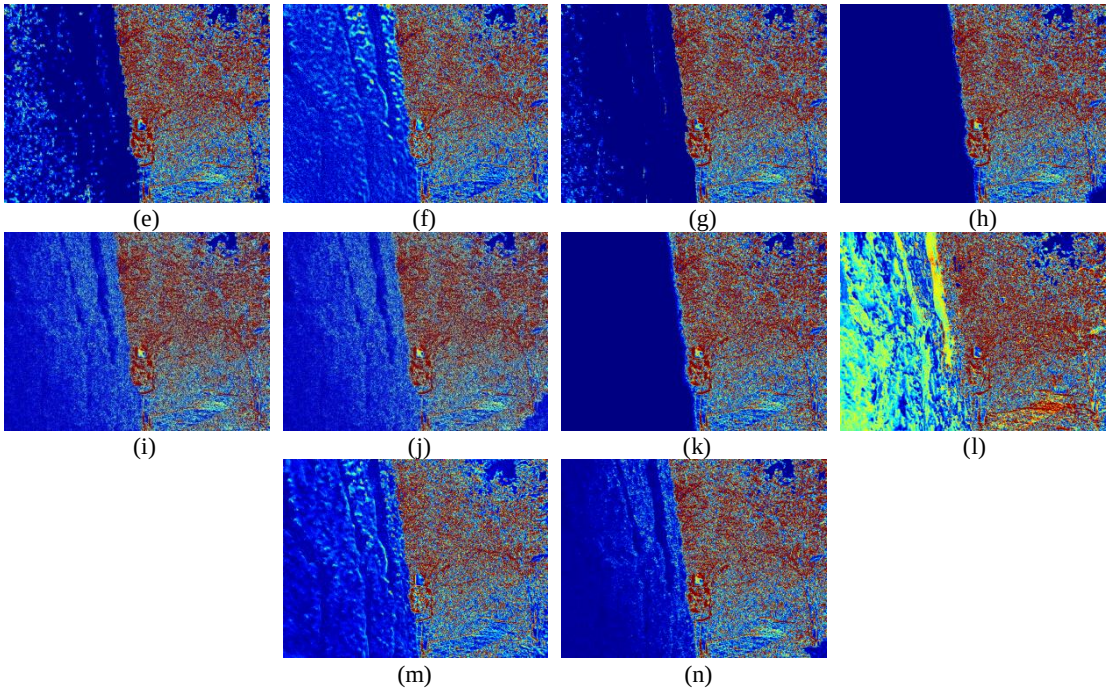
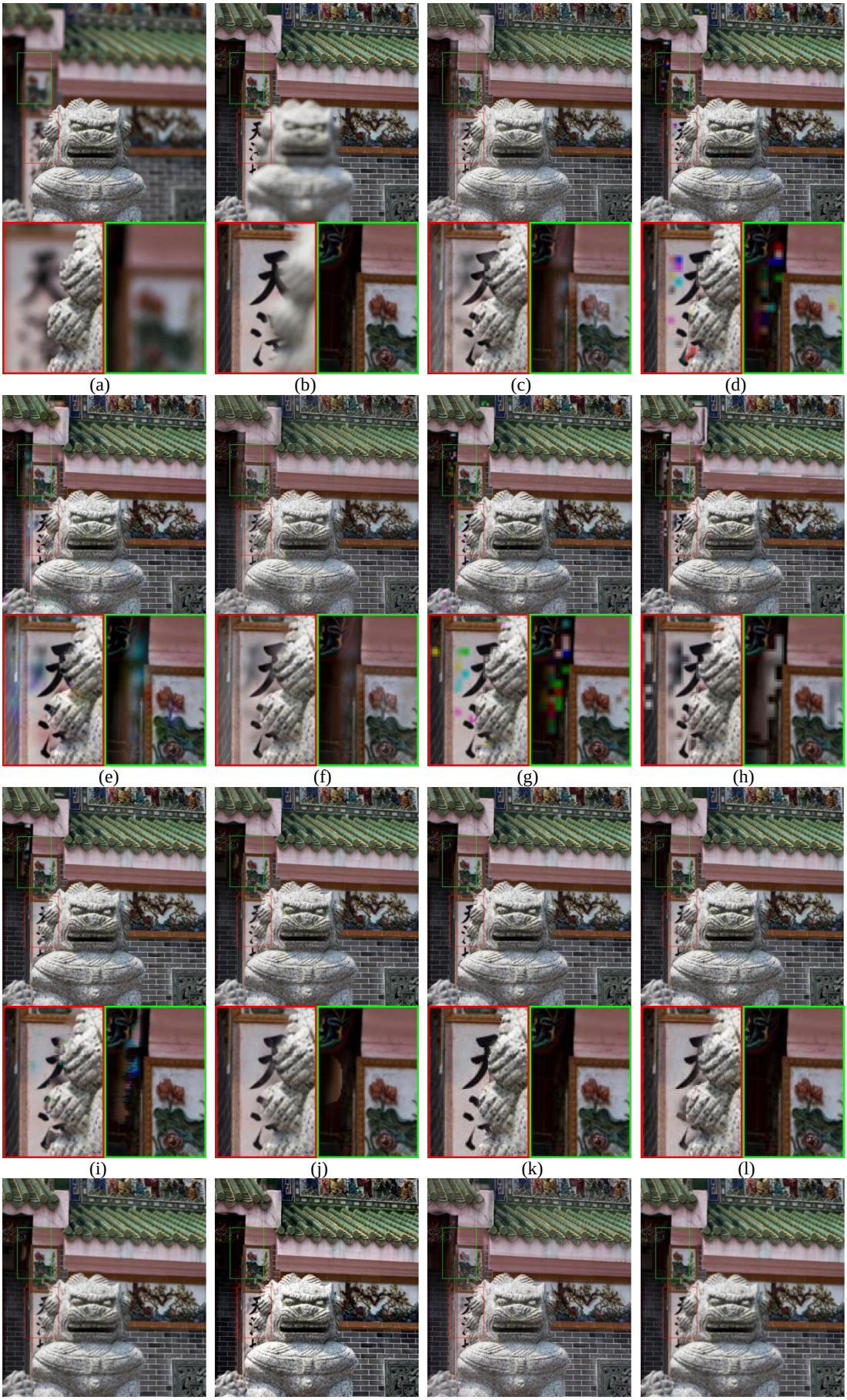


Figure 13. Difference images between the source image 1 [Figure 12(a)] and each of the fused images in Figure 12. (a) NSCT; (b) SR; (c) NSCT-SR; (d) ASR-1; (e) ASR-2; (f) JCPD; (g) DL-LRR; (h) MFQ; (i) MDLSR; (j) PADCDTNPS; (k) CNN; (l) SWIN; (m) LFDT; (n) QSR.

The second experiment is conducted on Lion source images. Because the viewpoint is changed when capturing source images, they are not registered well with each other. This is another reason causing image misregistration. Many methods, including SR, NSCT-SR, ASR-2, and DL-LRR, produce severe color bias because their channel-by-channel processing fails to preserve inherent color structures. Others, like NSCT, ASR-1, and JCPD, introduce numerous artifacts, with JCPD also exhibiting significant blurring. These failures are largely due to the zoom effect creating texture-less focused regions and weak edges in defocused areas, making correct focus detection difficult. In contrast, spatial domain and deep learning methods [Figure 14(j–o)] are superior, showing fewer artifacts and no color distortion. However, MFQ and PADCDTNPS suffer from unclear transitions due to overly smoothed decision maps, while SWIN and LFDT, though satisfactory, still show slight blurring in these regions. The QSR method excels by leveraging its strong color image representation and spatial feature extraction, accurately distinguishing focused and defocused regions to achieve the best visual effect. Figure 15 shows the difference images from subtracting the first source image from fused images. Source image 1 has a clear foreground (lion) but blurry background. Transform domain–based methods (NSCT, NSCT-SR, ASR-1, and JCPD) retain excessive defocused information in focused areas due to poor misregistration handling. The MFQ, MDLSR, and PADCDTNPS methods show boundary blurriness, while CNN, SWIN, and LFDT methods suffer from edge diffusion. In contrast, the QSR method accurately identifies focused regions with clear boundaries between focused and defocused areas, and our method additionally reduces chromatic distortion in the fusion results.

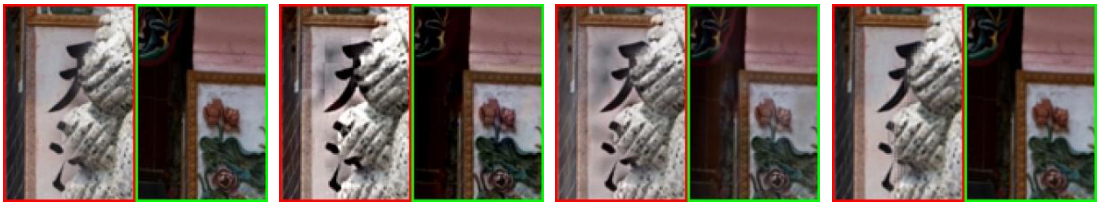
We can infer that, based on the above two experiments, the QSR method is able to deal with the issue of image misregistration well.



(a) (b) (c) (d)

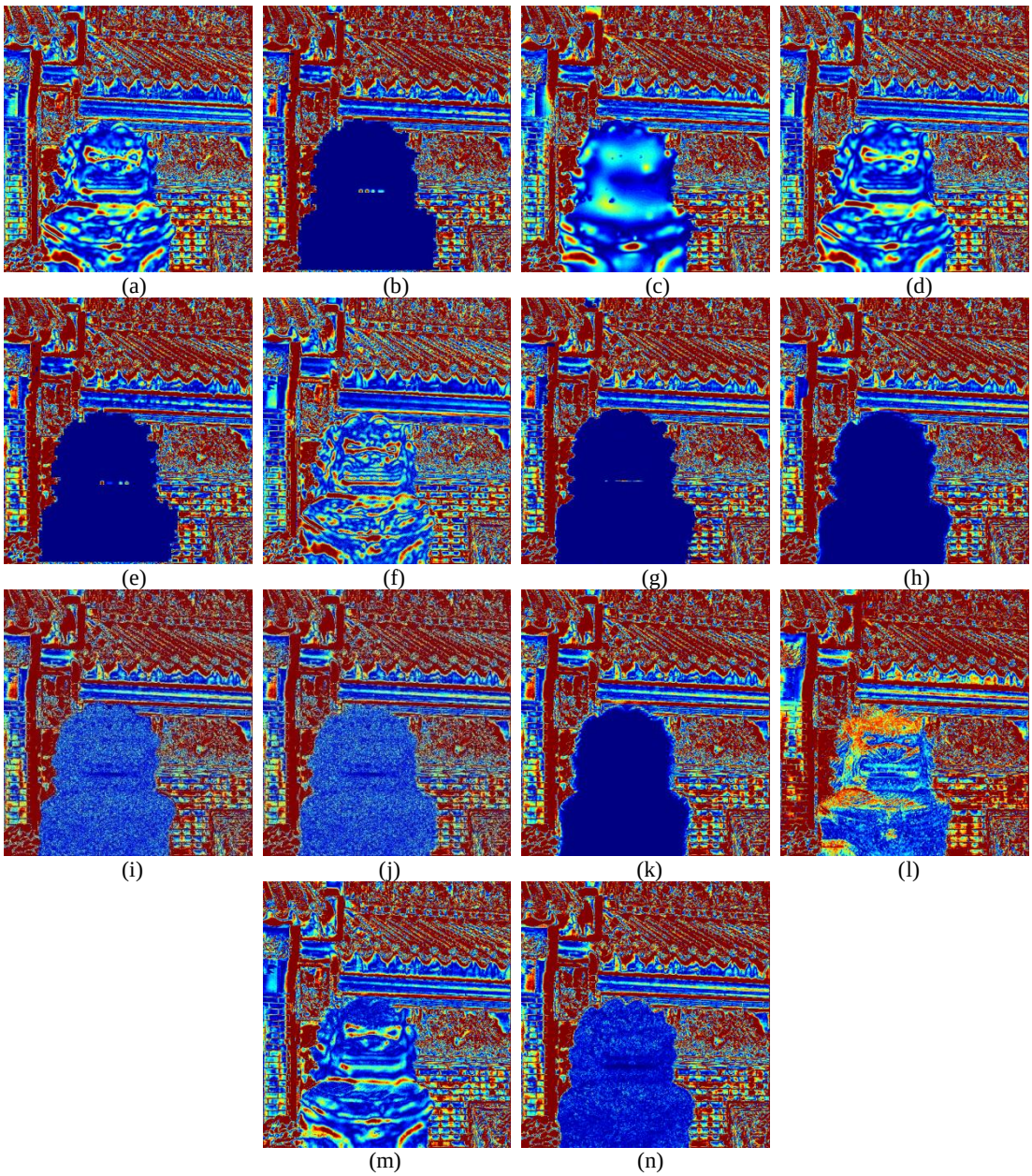
(e) (f) (g) (h)

(i) (j) (k) (l)



(m) (n) (o) (p)

Figure 14. Illustration of the fused images on the Lion set. (a)–(b) Source images. (c)–(p) Fusion results obtained by the NSCT, SR, NSCT-SR, ASR-1, ASR-2, JCPD, DL-LRR, MFQ, MDLSR, PADCDTNPS, CNN, SWIN, LFDT, and QSR, respectively.



(a) (b) (c) (d)

(e) (f) (g) (h)

(i) (j) (k) (l)

(m) (n)

Figure 15. Difference images between the source image 1 [Figure 14(a)] and each of the fused images in Figure 14. (a) NSCT; (b) SR; (c) NSCT-SR; (d) ASR-1; (e) ASR-2; (f) JCPD; (g) DL-LRR; (h) MFQ; (i) MDLSR; (j) PADCDTNPS; (k) CNN; (l) SWIN; (m) LFDT; (n) QSR.

In the following, we assess the efficacy of the proposed method in the task of micro multi-focus image fusion. In these experiments, we have selected several methods that performed well in previous experiments for comparative analysis.

Figure 16 displays a sequence of thirteen Bug multi-focus microscopic images, downloaded from [96]. The fused images are presented in Figure 17. Each source image has a shallow depth of field, creating irregularly shaped focused regions. In addition, the white fluffs on the feeler of Bug occupy a small area in the clear source image, but they occupy a larger area in the blurred images, making it very difficult to accurately and completely extract the focused regions. As shown in Figure 17, the PADCDTNPS method struggles with weak contrast, leaving blurred regions, while SWIN and LFDT underperform due to the subtle features and indistinct boundaries in these microscopic images. Though NSCT and SR methods improve visual quality, they lose some fluff details (such as the white fluffs on the feeler). The MFQ and CNN retain the most focused information but introduce artifacts. In contrast, the QSR method produces a clearer result with fewer artifacts, outperforming the other approaches.

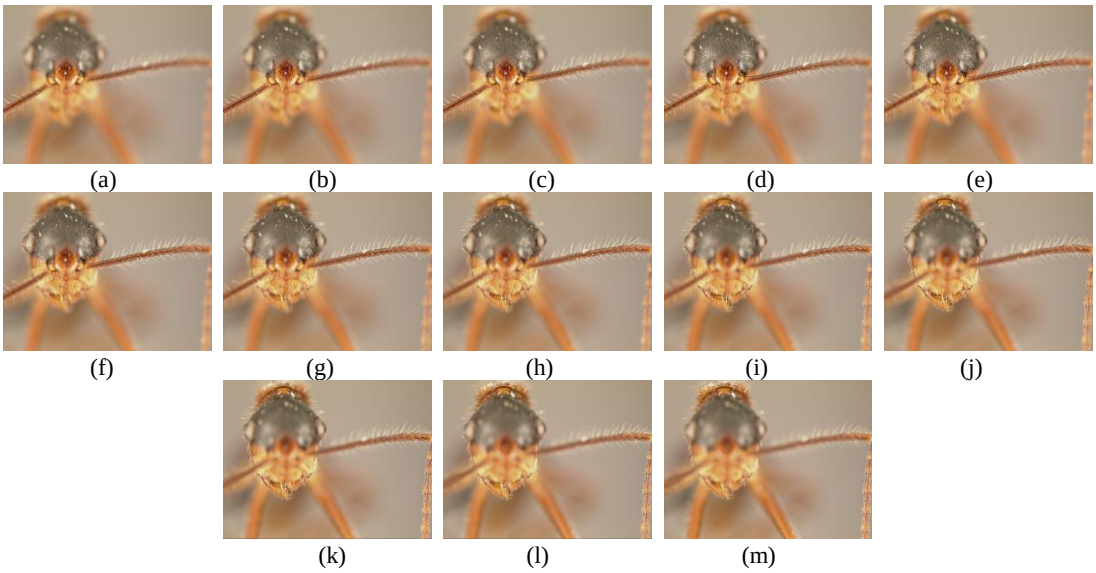
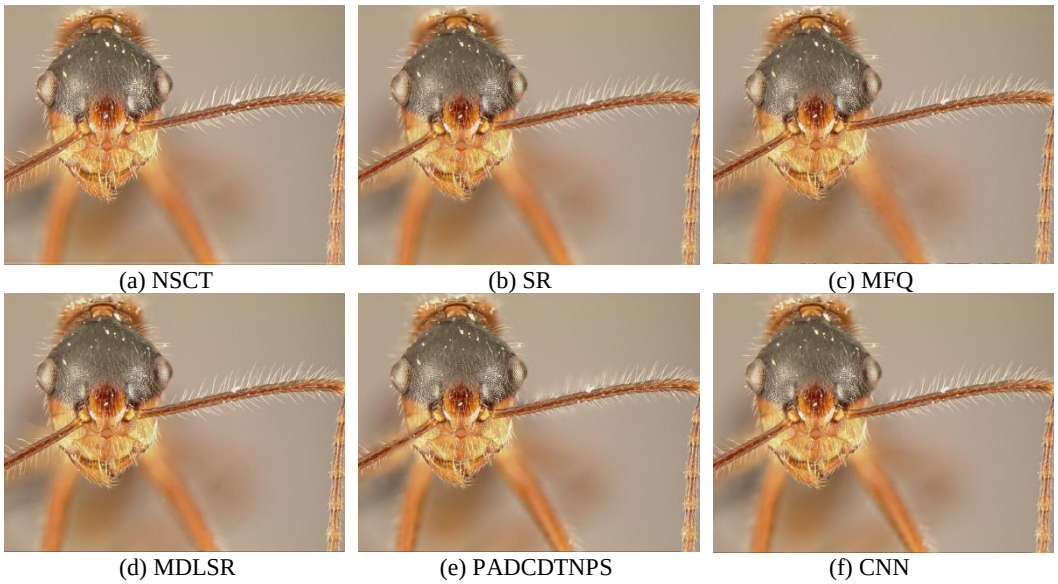


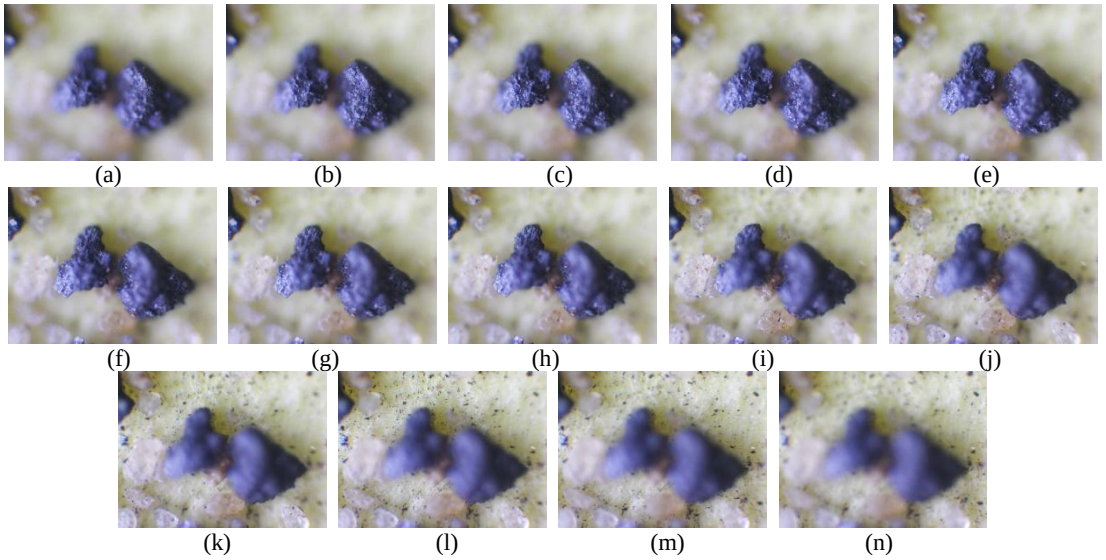
Figure 16. The Bug multi-focus microscopic image sequence.



(g) SWIN (h) LFDT (i) QSR

Figure 17. Illustration of the fused images on the Bug set.

Figure 18 depicts a sequence of 14 Diamond multi-focus microscopic source images, sourced from [97]. Figure 19 shows the fused images produced by various fusion methods. The Diamond source images feature small, irregularly shaped, focused objects with indistinct boundaries and subtle background characteristics, presenting a considerable challenge for spatial domain–based methods to accurately extract the focused regions. As shown in Figure 19, the PADCDTNPS and CNN methods misclassify pixels and blur the background due to subtle gradient variations, while SR, MFQ, and MDLSR retain focused regions but still produce artifacts. The performance of the two deep learning methods (SWIN and LFDT) on this dataset remains subpar, rendering them largely incapable of accomplishing this type of image fusion task. In contrast, the QSR and NSCT methods achieve the best visual results, successfully fusing the perfectly registered source images.

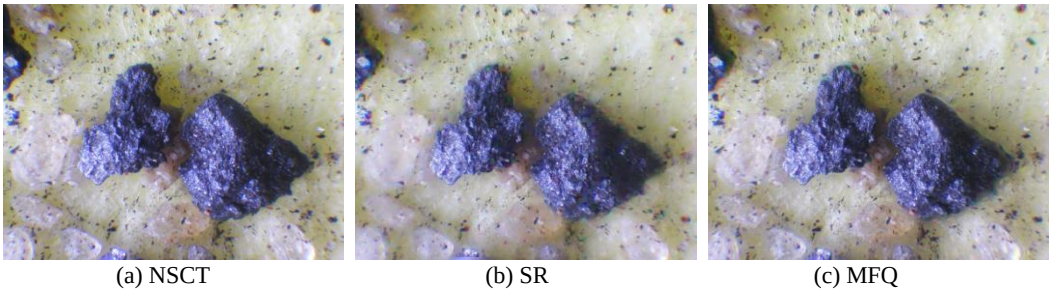


(a) (b) (c) (d) (e)

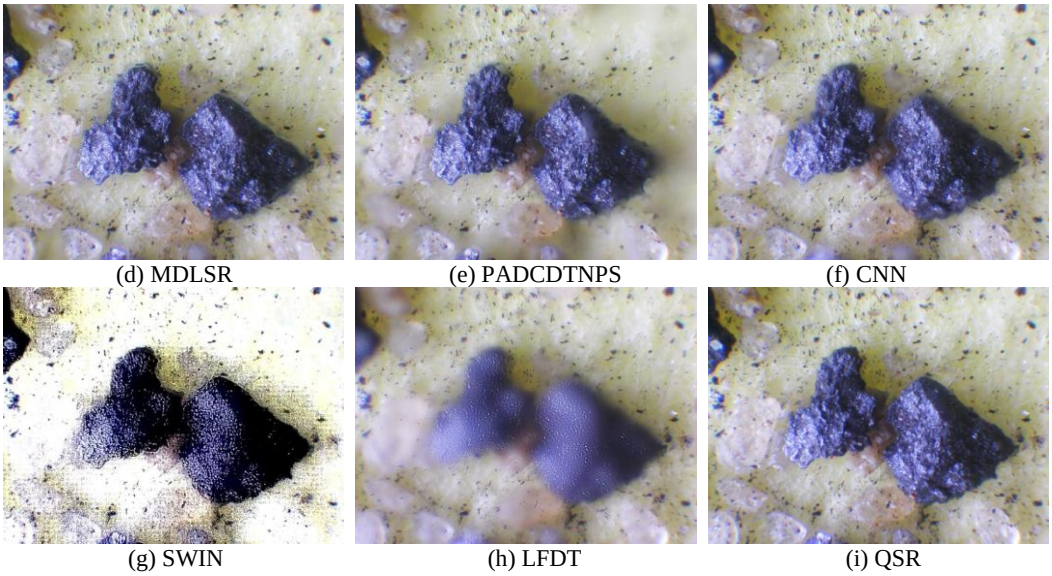
(f) (g) (h) (i) (j)

(k) (l) (m) (n)

Figure 18. The Diamond multi-focus microscopic image sequence.



(a) NSCT (b) SR (c) MFQ



(d) MDLSR (e) PADCDTNPS (f) CNN

(g) SWIN (h) LFDT (i) QSR

Figure 19. Illustration of the fused images on the Diamond set.

4.3.2. Quantitative comparisons

Beyond subjective evaluation, quantitative assessments are crucial for validating fusion methods. To test the QSR method's robustness, we compared it with other methods using six metrics, with the mean results for each dataset presented in Tables 1–4 (best results in bold). To observe the assessment results intuitively, the mean assessment values of all fused images are plotted in Figure 20. The horizontal coordinate axis represents different fusion methods, and the vertical coordinate axis represents the corresponding mean assessment values.

Globally, the QSR method outperforms transform domain–based methods and competes well with advanced spatial domain–based methods. As shown in Tables 1–4, the JCPD method performs poorly due to its simplistic averaging rule, which loses significant image information. Compared to the classical NSCT method, QSR shows noticeable improvement across all metrics. While QSR slightly underperforms the SR method on the Q_{NMI} metric, it leads in others, as representation-based methods using absolute sparse coefficients as focus measures fail to properly capture image focus characteristics. Spatial domain–based methods generally outperform transform domain–based methods on metrics like Q_{NMI} because they directly select source pixels, whereas transform methods average pixel values during fusion. For sharpness and structural detail metrics (Q_G, Q_Y), QSR achieves the highest scores, indicating superior preservation of source image details. For color fidelity (Q_{CF}), the SWIN method leads due to its specialized training, yet QSR still outperforms channel-by-channel fusion methods, demonstrating richer color and better alignment with human perception. Although deep learning–based methods may excel in some objective metrics, they fail in subjective evaluations for micro multi-focus fusion, limiting their applicability. Figure 20 confirms that no single method is best across all metrics, but QSR consistently ranks among the top, proving its robustness.

In summary, we can conclude that, based on the above analysis and discussion, the QSR exhibits competitive performance for color MFIF.

Table 1. The average objective assessment values of the fusion results on the Lytro dataset.

Methods	NSCT	SR	NSCT-SR	ASR-1	ASR-2	JCPD	DL-LRR	MFQ	MDLSR	PADCDTNPS	CNN	SWIN	LFDT	QSR
Q_{NMI}	0.9187	1.1058	0.9978	0.9442	1.0623	0.8461	0.8953	1.1649	**1.1876**	1.1324	1.1512	0.7624	0.9177	1.0806
Q_G	0.6713	0.6933	0.6989	0.6926	0.6713	0.5065	0.6045	0.7259	**0.7268**	0.7241	0.7250	0.6101	0.6594	0.7215
Q_Y	0.9529	0.9646	0.9666	0.9642	0.9509	0.8182	0.9170	0.9884	**0.9893**	0.9735	0.9875	0.8815	0.9383	0.9788
Q_{CB}	0.7169	0.7677	0.7705	0.7207	0.7424	0.6242	0.7379	0.8108	0.8112	0.8075	0.8084	0.6415	0.7172	**0.8136**
Q_{VIFF}	0.9070	0.9353	0.9559	0.9020	0.9187	0.8494	0.9140	0.9462	0.9454	0.8976	0.9438	**1.0985**	0.9302	0.9753
Q_{CF}	56.7926	58.6251	60.0364	55.5944	58.2348	51.7600	48.4340	60.7473	61.3433	51.3987	60.2280	**102.7381**	44.0223	62.9275

Table 2. The average objective assessment values of the fusion results on the MFFW dataset.

Methods	NSCT	SR	NSCT-SR	ASR-1	ASR-2	JCPD	DL-LRR	MFQ	MDLSR	PADCDTNPS	CNN	SWIN	LFDT	QSR
Q_{NMI}	0.8194	1.0005	0.8115	0.8307	0.9690	0.7774	0.6299	0.7453	**1.1682**	1.1441	1.0881	0.7206	0.8377	1.0003
Q_G	0.6267	0.6138	0.5984	0.6246	0.5982	0.4277	0.4191	0.4933	0.7053	0.7019	0.6802	0.5297	0.6106	**0.7066**
Q_Y	0.9014	0.8419	0.8601	0.9024	0.8297	0.7185	0.7048	0.7846	0.9861	0.9838	0.9737	0.8064	0.8788	**0.9884**
Q_{CB}	0.6465	0.6574	0.6474	0.6338	0.6423	0.5502	0.5322	0.6813	**0.7634**	0.7574	0.7438	0.6047	0.6436	0.7583
Q_{VIFF}	0.8229	0.8331	0.8428	0.7995	0.8251	0.7327	0.7159	0.8258	0.8424	0.8410	0.8424	**0.9958**	0.8298	0.8568
Q_{CF}	57.2515	53.0816	50.8122	58.0223	52.5061	60.2992	14.0557	60.4846	54.3381	53.5005	51.6109	**84.6801**	49.5275	68.4776

Table 3. The average objective assessment values of the fusion results on the MFI-WHU dataset.

Methods	NSCT	SR	NSCT-SR	ASR-1	ASR-2	JCPD	DL-LRR	MFQ	MDLSR	PADCDTNPS	CNN	SWIN	LFDT	QSR
Q_{NMI}	0.9773	1.1728	1.0914	1.0030	1.1509	0.8382	0.6312	**1.2203**	1.2180	1.1975	1.1837	0.7594	0.9202	1.1209
Q_G	0.6996	0.7276	0.7235	0.7142	0.7212	0.5157	0.4407	0.7401	0.7410	0.7340	0.7384	0.6269	0.6877	**0.7408**
Q_Y	0.9650	0.9784	0.9780	0.9707	0.9743	0.8356	0.7689	0.9880	0.9898	0.9893	0.9886	0.8893	0.9483	**0.9904**
Q_{CB}	0.7823	0.8100	0.8047	0.7754	0.8041	0.7103	0.5752	0.8255	**0.8289**	0.8232	0.8270	0.6525	0.7431	0.8235
Q_{VIFF}	0.9762	0.9804	0.9807	0.9660	0.9765	0.9073	0.8280	0.9832	0.9813	0.9765	0.9804	**1.1623**	0.9869	0.9867
Q_{CF}	24.2006	29.6607	26.2794	25.3644	29.1587	18.4981	11.9365	30.6521	30.4822	29.7861	29.3495	**51.3819**	13.9346	30.5525

Table 4. The average objective assessment values of the fusion results on the Real-MFF dataset.

Methods	NSCT	SR	NSCT-SR	ASR-1	ASR-2	JCPD	DL-LRR	MFQ	MDLSR	PADCDTNPS	CNN	SWIN	LFDT	QSR
Q_{NMI}	0.9253	0.9305	0.9253	0.9229	0.9245	0.8884	0.7000	0.9292	0.9294	0.9298	0.9306	0.8290	**1.1001**	0.9302
Q_G	0.5379	0.5221	0.5358	0.5232	0.5356	0.4370	0.4457	0.5308	0.5343	0.5346	0.5348	0.5450	**0.7136**	0.5534
Q_Y	0.8217	0.8089	0.8210	0.8086	0.8205	0.7161	0.7271	0.8169	0.8218	0.8216	0.8215	0.7932	**0.9304**	0.8227
Q_{CB}	0.7619	0.7533	0.7674	0.7486	0.7675	0.7044	0.5994	0.7656	0.7721	0.7700	0.7712	0.6021	**0.7849**	0.7716
Q_{VIFF}	0.9494	0.9215	0.9600	0.9241	0.9617	0.9378	0.9560	0.9424	0.9523	0.9448	0.9487	**1.1801**	0.9875	0.9663
Q_{CF}	28.5522	29.1079	30.3819	27.1160	30.3339	23.9809	9.1969	31.0155	32.2488	31.5310	31.6324	**82.5738**	21.5427	31.6494

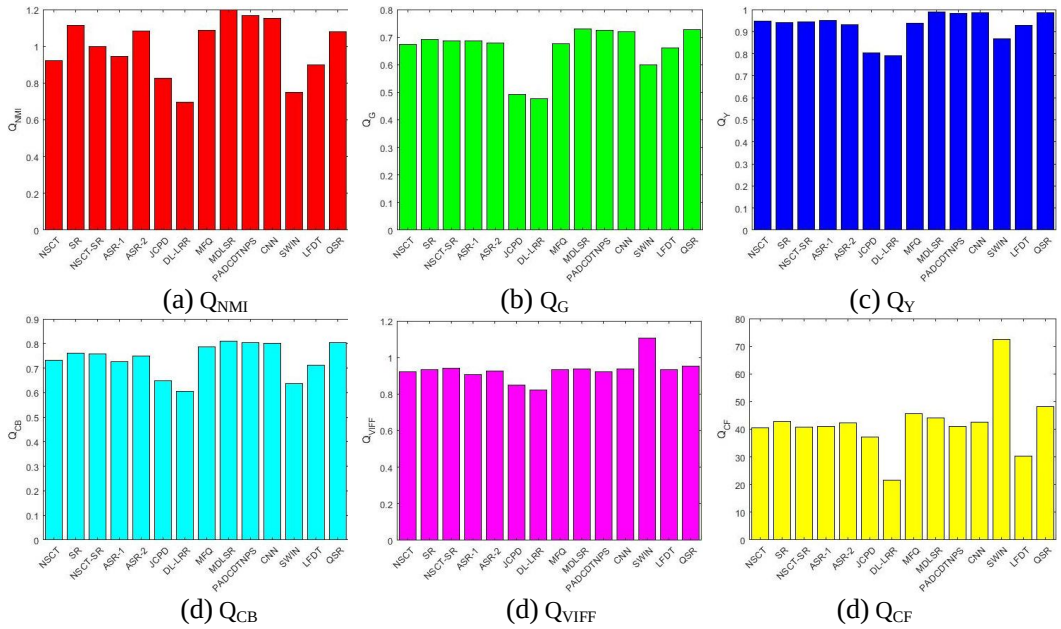


Figure 20. Mean objective assessment values of fused images.

4.3.3. Efficiency comparisons

Finally, we compare the computational efficiency of different methods. The mean running time for fusing the Lytro dataset is listed in Table 3. All traditional methods are conducted in the MATLAB R2014b environment on a computer with an Intel i5 CPU and 8 GB of RAM. For the experiments of the two deep learning methods (SWIN and LFDT), they were conducted on an NVIDIA GeForce RTX 4080 GPU and a 13th Gen Intel(R) Core(TM) i7-13700F 2.10GHz processor. The proposed network is implemented on the PyTorch platform using the PyCharm tool.

Table 5 shows that the LFDT is the most efficient method, while DL-LRR is the slowest due to its computationally intensive patch classification and low-rank representation. Among representation-based methods, JCPD is faster than QSR as it only processes luminance components, while other methods like SR, NSCT-SR, and ASR variants are slower. QSR is approximately twice as fast as NSCT-SR and ASR-1, and three times faster than ASR-2, because its sliding window has a 50% overlap compared to the 90% overlap in other methods. Although QSR is not the most efficient, its advantages in micro-image processing make it suitable for applications where speed is not critical. Future work will focus on improving its computational efficiency.

Table 5. Computational cost comparison.

Methods	NSCT	SR	NSCT-SR	ASR-1	ASR-2	JCPD	DL-LRR	MFQ	MDLSR	PADCDTNPS	CNN	SWIN	LFDT	QSR
Time (second)	75.78	98.56	166.21	159.51	264.86	35.04	896.64	11.61	1.60	2.44	36.28	1.03	0.64	89.06

5. Conclusions

In this paper, a novel color MFIF method using the QSR model is presented. First, all the color channels of source images are transformed into the sparse space uniformly by the QSR model, rather than coding color channels as in traditional SR-based methods. Then, two discriminating features,

namely the salience and sparsity features, are designed to capture the intrinsic focus characteristics of each image patch. Particularly, we use the blurry sparse coefficients deduced from the QSR model under a blurry quaternion dictionary learned on blurred training data instead of employing the clear sparse coefficients to compute the sparsity feature. Additionally, the computation process of the two features fully considers the spatial information to enhance their discriminant power. Owing to the excellent ability of QSR for color image representation and making full use of spatial information in feature extraction, our method can differentiate between focused and defocused patches accurately and reduce or even eliminate the spatial artifacts and color distortion in the fused images. The QSR method is experimentally demonstrated to be superior to some recent representation-based methods and competitive with some advanced spatial domain–based methods and deep learning–based methods under various situations.

Author contributions

Wei Liu: Writing–review & editing, Writing–original draft, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. Wanqing Li: Writing–original draft, Visualization, Software, Methodology, Investigation. Xia Zhao: Writing–review & editing, Investigation, Funding acquisition. Fang Zhu: Writing‑review & editing, Software, Investigation, Funding acquisition, Conceptualization.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors would like to thank the editor and anonymous reviewers for their detailed review, valuable comments and constructive suggestions. This work has been supported by the Key Project of Natural Science Research in Anhui Province (Grant No.2022AH051750, No.2024AH050611), in part by the Excellent Innovative Research Team of universities in Anhui Province (Grant No.2023AH010056), in part by the Quality Engineering Project of the Department of Education Anhui Province (Grant No.2024dzxkc096), in part by the Anhui Province Youth Teacher Training Action Project (Grant No.YQYB2024071) and in part by the Fundamental Research Funds for the Tongling University (Grant No. 2022tlxyrc11).

Conflict of interest

The authors declare no conflicts of interest in this paper.

References

1. Li S, Kang X, Fang LY, Hu J, Yin H (2017) Pixel-level image fusion: A survey of the state of the art. *Inform Fusion* 33: 100–112. https://doi.org/10.1016/j.inffus.2016.05.004

2. Jiang ZG, Han ZG, Chen J, Zhou XK (2004) A wavelet based algorithm for multi-focus micro-image fusion. In *Third International Conference on Image and Graphics (ICIG'04*), 176–179. https://doi.org/10.1109/ICIG.2004.29
3. Sujatha K, Punithavathani DS (2018) Optimized ensemble decision-based multi-focus image fusion using binary genetic grey-wolf optimizer in camera sensor networks. *Multimed Tools Appl* 77: 1735–1759. https://doi.org/10.1007/s11042-016-4312-3
4. Chen Z, Wang D, Gong S, Zhao F (2017) Application of multifocus image fusion in visual power patrol inspection. In *2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference* 1688–1692. https://doi.org/10.1109/IAEAC.2017.8054302
5. Song Y, Li M, Li Q, Sun L (2006) A new wavelet based multi - focus image fusion scheme and its application on optical microscopy. In *2006 IEEE International Conference on Robotics and Biomimetics* 401–405. https://doi.org/10.1109/ROBIO.2006.340210
6. Liu Y, Wang L, Cheng J, Li C, Chen X (2020) Multi-focus image fusion: A Survey of the state of the art. *Inform Fusion* 64: 71–91. https://doi.org/10.1016/j.inffus.2020.06.013
7. Zhou Z, Li S, Wang B (2014) Multi-scale weighted gradient-based fusion for multi-focus images. *Inform Fusion* 20: 60–72. https://doi.org/10.1016/j.inffus.2013.11.005
8. Liu Y, Liu S, Wang Z (2015) Multi-focus image fusion with dense SIFT. *Inform Fusion* 23: 139–155. https://doi.org/10.1016/j.inffus.2014.05.004
9. Yin W, Zhao W, You D, Wang D (2019) Local binary pattern metric-based multi-focus image fusion. *Opt Laser Technol* 110: 62–68. https://doi.org/10.1016/j.optlastec.2018.07.045
10. Nejati M, Samavi S, Shirani S (2015) Multi-focus image fusion using dictionary-based sparse representation. *Inform Fusion* 25: 72–84. https://doi.org/10.1016/j.inffus.2014.10.004
11. Chen Y, Guan J, Cham WK (2018) Robust multi-focus image fusion using edge model and multi-matting. *IEEE T Image Process* 27: 1526–1541. https://doi.org/10.1109/TIP.2017.2779274
12. Bouzos O, Andreadis I, Mitianoudis N (2019) Conditional random field model for robust multi-focus image fusion. *IEEE T Image Process* 28: 5636–5648. https://doi.org/10.1109/TIP.2019.2922097
13. Liu W, Zheng Z, Wang Z F (2021) Robust multi-focus image fusion using lazy random walks with multiscale focus measures. *Signal Process* 179: 107850. https://doi.org/10.1016/j.sigpro.2020.107850
14. Li J, Li X, Li X, Han D, Tan H, Hou Z, et al. (2024) Multi-focus image fusion based on multiscale fuzzy quality assessment. *Digital Signal Process* 153: 104592. https://doi.org/10.1016/j.dsp.2024.104592
15. Adeel H, Riaz MM, Bashir T, Ali SS, Latif S (2024) Shahzad Latif4 Multi-focus image fusion using curvature minimization and morphological filtering. *Multimed Tools Appl* 83: 78625–78639. https://doi.org/10.1007/s11042-024-18654-6
16. Li B, Zhang L, Liu J, Peng H, Wang Q, Liu J (2024) Multi-focus image fusion with parameter adaptive dual channel dynamic threshold neural P systems. *Neural Networks* 179: 106603. https://doi.org/10.1016/j.neunet.2024.106603
17. Toet A (1989) Image fusion by a ratio of lowpass pyramid. *Pattern Recogn Lett* 9: 245–253. https://doi.org/10.1016/0167-8655(89)90003-2
18. Petrovic VS, Xydeas CS (2004) Gradient-based multiresolution image fusion. *IEEE T Image Process* 13: 228–237. https://doi.org/10.1109/TIP.2004.823821

19. Li H, Manjunath B, Mitra S (1995) Multisensor image fusion using the wavelet transform. *Graphical Models Image Process* 57: 235–245. https://doi.org/10.1006/gmip.1995.1022
20. Lewis JJ, O'Callaghan RJ, Nikolov SG, Bull DR, Canagarajah N (2007) Pixel- and region-based image fusion with complex wavelets. *Inform Fusion* 8: 119–130. https://doi.org/10.1016/j.inffus.2005.09.006
21. Zhang Q, Guo B (2009) Multifocus image fusion using the nonsubsampled contourlet transform. *Signal Process* 89: 1334–1346. https://doi.org/10.1016/j.sigpro.2009.01.012
22. Gao G, Xu L, Feng D (2013) Multi-focus image fusion based on nonsubsampled shearlet transform. *IET Image Process* 7: 633–639. https://doi.org/10.1049/iet-ipr.2012.0524
23. Hu J, Li S (2012) The multiscale directional bilateral filter and its application to multisensor image fusion. *Inform Fusion* 13: 196–206. https://doi.org/10.1016/j.inffus.2011.02.003
24. Jian L, Yang X, Zhou Z, Zhou K, Liu K (2018) Multi-scale image fusion through rolling guidance filter. *Future Gener Comput Syst* 83: 310–325. https://doi.org/10.1016/j.future.2018.01.039
25. Liu W, Wang Z (2020) A novel multi-focus image fusion method using multiscale shearing non-local guided averaging filter. *Signal Process* 166: 1–24. https://doi.org/10.1016/j.sigpro.2019.107252
26. Liu Y, Chen X, Peng H, Wang Z (2017) Multi-focus image fusion with a deep convolutional neural network. *Inform Fusion* 36: 191–207. https://doi.org/10.1016/j.inffus.2016.12.001
27. Tang H, Xiao B, Li W, Wang G (2018) Pixel convolutional neural network for multi-focus image fusion. *Inf Sci* 433-434: 125–141. https://doi.org/10.1016/j.ins.2017.12.043
28. Yang Y, Nie Z, Huang S, Lin P, Wu J (2019) Multilevel features convolutional neural network for multifocus image fusion. *IEEE T Comput Imag* 5: 262–273. https://doi.org/10.1109/TCI.2018.2889959
29. Wang Z, Li X, Duan H, Zhang X, Wang H (2019) Multifocus image fusion using convolutional neural networks in the discrete wavelet transform domain. *Multimed Tools Appl* 78: 34483–34512. https://doi.org/10.1007/s11042-019-08070-6
30. Zhai H, Zheng W, Ouyang Y, Pan X, Zhang W (2024) Multi-focus image fusion via interactive transformer and asymmetric soft sharing. *Eng Appl Artif Intell* 133: 107967. https://doi.org/10.1016/j.engappai.2024.107967
31. Ouyang Y, Zhai H, Hu H, Li X, Zeng Z (2025) FusionGCN: Multi-focus image fusion using superpixel features generation GCN and pixel-level feature reconstruction CNN. *Expert Syst Appl* 262: 125665. https://doi.org/10.1016/j.eswa.2024.125665
32. Guo X, Nie R, Cao J, Zhou D, Mei L, He K (2019) FuseGAN: learning to fuse multi-focus image via conditional generative adversarial network. *IEEE Trans Multimedia* 21: 1982–1996. https://doi.org/10.1109/TMM.2019.2895292
33. Huang J, Le Z, Ma Y, Mei X, Fan F (2020) A generative adversarial network with adaptive constraints for multi-focus image fusion. *Neural Comput Appl* 32: 15119–15129. https://doi.org/10.1007/s00521-020-04863-1
34. Li H, Qian W, Nie R, Cao J, Xu D (2023) Siamese conditional generative adversarial network for multi-focus image fusion. *Appl Intell* 53: 17492–17507. https://doi.org/10.1007/s10489-022-04406-2

35. Li J, Li B, Jiang Y (2023) GIPC-GAN: an end-to-end gradient and intensity joint proportional constraint generative adversarial network for multi-focus image fusion. *Complex Intell Syst* 9: 7395–7422. https://doi.org/10.1007/s40747-023-01151-y
36. Ma J, Tang L, Fan F, Huang J, Mei X, Ma Y (2022) SwinFusion: Cross-domain long-range learning for general image fusion via swin transformer. *IEEE/CAA J Autom Sin* 9: 1200–1217. https://doi.org/10.1109/JAS.2022.105686
37. Li M, Pei R, Zheng T, Zhang Y, Fu W (2024) FusionDiff: Multi-focus image fusion using denoising diffusion probabilistic models. *Expert Syst Appl* 238: 121664. https://doi.org/10.1016/j.eswa.2023.121664
38. Zhu P, Sun Y, Cao B, Hu Q (2024) Task-customized mixture of adapters for general image fusion. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7099–7108. https://doi.org/10.1109/CVPR52733.2024.00678
39. Zhai H, Zhang G, Zeng Z, Xu Z, Fang A (2025) LSKN-MFIF: Large selective kernel network for multi-focus image fusion. *Neurocomputing* 635: 129984. https://doi.org/10.1016/j.neucom.2025.129984
40. Liu Y, Wang L, Li H, Chen X (2022) Multi-focus image fusion with deep residual learning and focus property detection. *Inform Fusion* 86: 1–16. https://doi.org/10.1016/j.inffus.2022.06.001
41. Wang Z, Li X, Zhao L, Duan H, Wang S, Liu H, et al. (2023) When multi-focus image fusion networks meet traditional edge-preservation technology. *Int J Comput Vis* 131: 2529–2552. https://doi.org/10.1007/s11263-023-01806-w
42. Duan Z, Luo X, Zhang T (2024) Combining transformers with CNN for multi-focus image fusion. *Expert Syst Appl* 235: 121156. https://doi.org/10.1016/j.eswa.2023.121156
43. Shao X, Jin X, Jiang Q, Miao S, Wang P, Chu X (2024) Multi-focus image fusion basedon transformer and depth information learning. *Comput Electr Eng* 119: 109629. https://doi.org/10.1016/j.compeleceng.2024.109629
44. Xie X, Cui Y, Tan T, Zheng X, Yu Z (2024) Fusionmamba: Dynamic feature enhancement for multimodal image fusion with mamba. *Vis Intell* 2: 37. https://doi.org/10.1007/s44267-024-00072-9
45. Jin X, Zhu P, Yu D, Wozniak M, Jiang Q, Wang P, et al. (2025) Combining depth and frequency features with Mamba for multi-focus image fusion. *Inf Fusion* 124: 103355. https://doi.org/10.1016/j.inffus.2025.103355
46. Mustafa H, Liu F, Yang J, Khan Z, Huang Q (2019) Dense multi-focus fusion net: A deep unsupervised convolutional network for multi-focus image fusion. In: *Proceedings of the International Conference on Artificial Intelligence and Soft Computing*, 153–163. https://doi.org/10.1007/978-3-030-20912-4_15
47. Xu H, Ma J, Le Z, Jiang J, Guo X (2020) FusionDN: A unified densely connected network for image fusion. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, 34: 12484–12491. http://dx.doi.org/10.1609/aaai.v34i07.6936
48. Ma B, Zhu Y, Yin X, Ban X, Huang H, Mukeshimana M (2021) SESF-Fuse: An unsupervised deep model for multi-focus image fusion. *Neural Comput Appl* 33: 5793–5804. https://doi.org/10.1007/s00521-020-05358-9
49. Hu X, Jiang J, Liu X, Ma J (2023) ZMFF: Zero-shot multi-focus image fusion. *Inf Fusion* 92: 127–138. https://doi.org/10.1016/j.inffus.2022.11.014

50. Fang J, Ning X, Mao T, Zhang M, Zhao Y, Hu S, et al. (2024) A multi-focus image fusion network combining dilated convolution with learnable spacings and residual dense network. *Comput Electr Eng* 117: 109299. https://doi.org/10.1016/j.compeleceng.2024.109299
51. Liu S, Peng W, Liu Y, Zhao J, Su Y, Zhang Y (2023) AFCANet: An adaptive feature concatenate attention network for multi-focus image fusion. *J King Saud Univ-Comput Inf Sci* 35: 101751. https://doi.org/10.1016/j.jksuci.2023.101751
52. Li J, Li B, Jiang Y (2023) GIPC-GAN: An end-to-end gradient and intensity joint proportional constraint generative adversarial network for multi-focus image fusion. *Complex Intell Syst* 9: 7395–7422. https://doi.org/10.1007/s40747-023-01151-y
53. Zhai H, Ouyang Y, Luo N, Chen L, Zeng Z (2024) MSI-DTrans: A multi-focus image fusion using multilayer semantic interaction and dynamic transformer. *Displays* 85: 102837. https://doi.org/10.1016/j.displa.2024.102837
54. Jiang S, Yu S (2025) Refined multi-focus image fusion using multi-scale neural network with SpSwin autoencoder-based matting. *Expert Syst Appl* 276: 126980. https://doi.org/10.1016/j.eswa.2025.126980
55. Yin HT, Li ST, Fang L (2013) Simultaneous image fusion and super-resolution using sparse representation. *Inf Fusion* 14: 229–240. https://doi.org/10.1016/j.inffus.2012.01.008
56. Dong L, Yang Q, Wu H, Xiao H, Xu M (2015) High quality multi-spectral and panchromatic image fusion technologies based on curvelet transform. *Neurocomputing* 159: 268–274. https://doi.org/10.1016/j.neucom.2015.01.050
57. Liu Y, Liu S, Wang Z (2015) A general framework for image fusion based on multi-scale transform and sparse representation. *Inf Fusion* 24: 147–164. https://doi.org/10.1016/j.inffus.2014.09.004
58. Zhu Z, Chai Y, Yin H, Li Y, Liu Z (2016) A novel dictionary learning approach for multi-modality medical image fusion. *Neurocomputing* 214: 471–482. https://doi.org/10.1016/j.neucom.2016.06.036
59. Zhang Q, Levine M (2016) Robust multi-focus image fusion using multi-task sparse representation and spatial context. *IEEE T Image Process* 25: 2045–2058. https://doi.org/10.1109/TIP.2016.2524212
60. Tang D, Xiong Q, Yin H, Zhu Z, Li Y (2016) A novel sparse-representation-based multi-focus image fusion approach. *Neurocomputing* 216: 216–229. https://doi.org/10.1016/j.eswa.2022.116737
61. Zhang Q, Shi T, Wang F, Rick S, Han J (2018) Robust sparse representation based multi-focus image fusion with dictionary construction and local spatial consistency. *Pattern Recognit* 83: 299–313. https://doi.org/10.1016/j.patcog.2018.06.003
62. Zhang Q, Wang F, Luo Y, Han J (2021) Exploring a unified low rank representation for multi-focus image fusion. *Pattern Recognit* 113: 107752. https://doi.org/10.1016/j.patcog.2020.107752
63. Subakan Ö, Vemuri B (2011) A Quaternion Framework for Color Image Smoothing and Segmentation. *Int J Comput Vis* 91: 233–250. https://doi.org/10.1007/s11263-010-0388-9
64. Kolaman A, Yadid-Pecht O (2012) Quaternion structural similarity: A new quality index for color images. *IEEE T Image Process* 21: 1526–1536. https://doi.org/10.1109/TIP.2011.2181522

65. Xu Y, Yu L, Yu H, Zhang H, Nguyen T (2015) Vector sparse representation of color image using quaternion matrix analysis. *IEEE T Image Process* 24: 1315–1329. https://doi.org/10.1109/TIP.2015.2397314
66. Lan R, Zhou Y (2016) Quaternion-michelson descriptor for color image classification. *IEEE T Image Process* 25: 5281–5292. https://doi.org/10.1109/TIP.2016.2605922
67. Zou C, Kou K, Wang Y, Tang Y (2021) Quaternion block sparse representation for signal recovery and classification. *Signal Process* 179: 107849. https://doi.org/10.1016/j.sigpro.2020.107849
68. Zou C, Kou K, Wang Y (2016) Quaternion collaborative and sparse representation with application to color face recognition. *IEEE T Image Process* 25: 3287–3302. https://doi.org/10.1109/TIP.2016.2567077
69. Chen Y, Xiao X, Zhou Y (2020) Low-rank quaternion approximation for color image processing. *IEEE T Image Process* 29: 1426–1439. https://doi.org/10.1109/TIP.2019.2941319
70. Hamilton WR (1866) Elements of Quaternions. (Cambridge Library Collection - Mathematics) (W. Hamilton, Ed.). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511707162
71. Lewicki M, Sejnowski T (1998) Learning overcomplete representations. *Neural Comput* 12: 337–365. https://doi.org/10.1162/089976600300015826
72. Skretting K, Husøy J, Aase S (2006) General design algorithm for sparse frame expansions. *Signal Process* 86: 117–126. https://doi.org/10.1016/j.sigpro.2005.04.013
73. Aharon M, Elad M, Bruckstein A (2006) K-SVD: an algorithm for designing over-complete dictionaries for sparse representation. *IEEE T Signal Process* 54: 4311–4322. https://doi.org/10.1109/TSP.2006.881199
74. Mairal J, Elad M, Sapiro G (2008) Sparse representation for color image restoration. *IEEE T Image Process* 17: 53–69. https://doi.org/10.1109/TIP.2007.911828
75. Tropp JA, Gilbert AC, Strauss MJ (2006) Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit. *Signal Process* 86: 572–588. https://doi.org/10.1016/j.sigpro.2005.05.030
76. Li J, Levine M, An X, Xu X, He H (2013) Visual saliency based on scale-space analysis in the frequency domain. *IEEE Trans Pattern Anal Mach Intell* 35: 996–1010. https://doi.org/10.1109/TPAMI.2012.147
77. Shi J, Xu L, Jia J (2015) Just noticeable defocus blur detection and estimation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 657–665. https://doi.org/10.1109/CVPR.2015.7298665
78. Xu S, Wei X, Zhang C, Liu J, Zhang J (2020) MFFW: A new dataset for multi-focus image fusion. arXiv:2002.04780 [cs.CV]. https://doi.org/10.48550/arXiv.2002.04780
79. Zhang H, Le Z, Shao Z, Xu H, Ma J (2021) MFF-GAN: An unsupervised generative adversarial network with adaptive and gradient joint constraints for multi-focus image fusion. *Inf Fusion* 66: 40–53. https://doi.org/10.1016/j.inffus.2020.08.022
80. Zhang J, Liao Q, Liu S, Ma H, Yang W, Xue JH (2020) Real-MFF: A large realistic multi-focus image dataset with ground truth. *Pattern Recognit Lett* 138: 370–377. https://doi.org/10.1016/j.patrec.2020.08.002
81. Multi-focus Image Dataset. Available from: https://github.com/yuliu316316/MSFIN-Fusion.
82. Yang B, Li S (2010) Multifocus image fusion and restoration with sparse representation. *IEEE Trans Instrum Meas* 59: 884–892. https://doi.org/10.1109/TIM.2009.2026612

83. Liu Y, Wang Z (2015) Simultaneous image fusion and denoising with adaptive sparse representation. *IET Image Process* 9(5): 347–357. https://doi.org/10.1049%2Fiet-ipr.2014.0311
84. Kim M, Han D, Ko H (2016) Joint patch clustering-based dictionary learning for multimodal image fusion. *Inf Fusion* 27: 198–214. https://doi.org/10.1016/j.inffus.2015.03.003
85. Li H, Wu X (2017) Multi-focus image fusion using dictionary learning and low-rank representation. In: *Lecture Notes in Computer Science*, 675–686. https://doi.org/10.1007/978-3-319-71607-7_59
86. Wang JW, Qu HJ, Zhang ZS, Xie M (2024) New insights into multi-focus image fusion:a fusion method based on multi-dictionary linear sparse representation and region fusion model. *Inf Fusion* 105: 102230. https://doi.org/10.1016/j.inffus.2024.102230
87. Xie XZ, Guo BY, Li PL, He SY, Zhou SJ (2025) SwinMFF: toward high-fidelity end-to-end multi-focus image fusion via swin transformer-based network. *Vis Comput* 41: 3883–3906. https://doi.org/10.1007/s00371-024-03637-3
88. Yang B, Jiang ZH, Pan D, Yu HY, Gui G, Gui WH (2025) LFDT-Fusion: A latent feature-guided diffusion Transformer model for general image fusion. *Inf Fusion* 113: 102639. https://doi.org/10.1016/j.inffus.2024.102639
89. Hossny M, Nahavandi S, Vreighton D (2008) Comments on 'Information measure for performance of image fusion'. *Electron Lett* 44: 1066–1067. http://dx.doi.org/10.1049/el:20081754
90. Zhao J, Laganiere R, Liu Z (2007) Performance assessment of combinative pixel-level image fusion based on an absolute feature measurement. *Int J Innov Comput Inf Control* 3: 1433–1447. https://api.semanticscholar.org/CorpusID:2510782
91. Wang Z, Bovik A, Sheikh H, Simoncelli E (2004) Image quality assessment: From error visibility to structural similarity. *IEEE T Image Process* 13: 600–612. https://doi.org/10.1109/TIP.2003.819861
92. Chen Y, Blum R (2009) A new automated quality assessment algorithm for image fusion. *Image Vis Comput* 27: 1421–1432. https://doi.org/10.1016/j.imavis.2007.12.002
93. Han Y, Cai Y, Cao Y, Xu X (2013) A new image fusion performance metric based on visual information fidelity. *Inf Fusion* 14: 127–135. https://doi.org/10.1016/j.inffus.2011.08.002
94. Fu YY (2006) Color image quality measures and retrieval. New Jersey, USA: New Jersey Institute of Technology, Dissertations. 745. https://digitalcommons.njit.edu/dissertations/745
95. True Color Kodak Images. Available from: http://r0k.us/graphics/kodak/.
96. Interactive Digital Photomontage. Available from: https://grail.cs.washington.edu/projects/photomontage/.
97. Helicon Focus. Available from: https://www.heliconsoft.com/helicon-focus-gallery/.