Article

A No-Reference Quality Assessment Method for Hyperspectral Sharpened Images via Benford's Law

Xiankun Hao [1], Xu Li [1,*], Jingying Wu [1], Baoguo Wei [1], Yujuan Song [2] and Bo Li [1]

[1] School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China; haoxiankun@mail.nwpu.edu.cn (X.H.)

[2] Beijing Institute of Control and Electronics Technology, Beijing 100045, China

[*] Correspondence: lixu@nwpu.edu.cn

Abstract: In recent years, hyperspectral (HS) sharpening technology has received high attention and HS sharpened images have been widely applied. However, the quality assessment of HS sharpened images has not been well addressed and is still limited to the use of full-reference quality evaluation. In this paper, a novel no-reference quality assessment method based on Benford's law for HS sharpened images is proposed. Without a reference image, the proposed method detects fusion distortion by performing first digit distribution on three quality perception features in HS sharpened images, using the standard Benford's law as a benchmark. The experiment evaluates 10 HS fusion methods on three HS datasets and selects four full-reference metrics and four no-reference metrics to compare with the proposed method. The experimental results demonstrate the superior performance of the proposed method.

Keywords: image quality assessment; no-reference; hyperspectral sharpened image; Benford's law; first digit distribution

1. Introduction

With the rapid development of imaging sensor technology in recent years, more and more airborne and satellite hyperspectral sensors have been able to capture hyperspectral (HS) images. Due to the rich spectral information, HS images have a wide range of applications in remote sensing, such as geological monitoring [1–3], object tracking [4], atmospheric monitoring [5–7], etc. Because of the trade-off issue in HS images acquisition, where high spectral resolution comes at the cost of low spatial resolution, the practical application of HS data has been greatly limited. In the past decade, many HS sharpening methods have emerged, which produce fused HS images with high spatial resolution (also known as HS sharpened images). In practice, high-quality fused HS images are very desirable in some specific tasks, such as anomaly detection [8], image classification [9], change detection [10], etc. The quality assessment of HS sharpened images is crucial for evaluating the quality of the fusion products from different HS sharpening methods, providing useful feedback to fusion design and valuable guidance for downstream processing tasks.

Nowadays, there are well-known satellites and airborne platforms capturing and providing HS images. The Hyperion sensor equipped on NASA's Earth Observing-1 satellite (EO-1) can simultaneously capture HS or multispectral (MS) images with a spatial resolution of 30 m and panchromatic (PAN) images with a spatial resolution of 10 m, covering the spectral range of 400–2500 nm [11]. Another famous airborne sensor is the Next-Generation Airborne Visible Infrared Imaging Spectrometer (AVIRIS-NG), which collects HS images from 2012 to the present and supports major activities in North America, India, and Europe [12]. AVIRIS-NG covers a spectral range of 380–2510 nm and has a spatial resolution of 0.3–20 m. The Hyperspectral Imager Suite (HISUI) recently developed by Japan has been recording HS and MS data since 2020, with a spatial resolution of 30 m

for HS images and 5 m for MS ones [13]. The Prisma satellite is managed and operated by the Italian Space Agency and was launched in 2019. It collects HS images with a spatial resolution of 30 m and MS images with a spatial resolution of 5 m, covering a spectral range of 400–2500 nm [14]. In addition, the new Environmental Mapping and Analysis Program (EnMAP) HS satellite of Germany was launched in 2022, which can provide 224 bands and has a spectral range of 420–2450 nm, and a spatial resolution of 30 m.

The emergence of new HS sensors and the availability of HS images have stimulated the further development of HS image enhancement technology including HS super resolution and HS sharpening. HS super resolution processes a single input HS image without relying on any auxiliary image/information, while HS sharpening is the fusion of HS images with MS/PAN images to obtain the high spatial resolution HS images. The existing HS sharpening methods can be roughly divided into four categories: classical pansharpening-based methods, matrix factorization (MF)-based methods, tensor representation (TR)-based methods, and deep convolution neural network (CNN)-based methods. Pansharpening (also known as multispectral sharpening) is a traditional fusion technique that utilizes PAN images to improve the spatial resolution of MS images. For the fusion of HS and MS images by using band selection or band synthesis strategy to assign each MS band to several corresponding HS bands, HS sharpening can be seen as an extension of multiple pansharpening [15]. In this way, pansharpening methods can be applied to fuse HS images. Two representative pansharpening methods are component substitution (CS)-based methods and multi-resolution analysis (MRA)-based methods. The former includes intensity–hue–saturation [16], principal component analysis (PCA) [17], Gram-Schmidt adaptive (GSA) [18], etc., while the latter includes the generalized Laplacian pyramid (GLP) [19], smoothing filtered-based intensity modulation (SFIM) [20], etc. MF-based methods is used to estimate the spectral basis and coefficients and regard HS sharpening as an optimization problem. The typical MF-based fusion methods include sparse MF [21], fast fusion based on Sylvester equation [22], non-negative structured sparse representation [23], maximum a posteriori estimation with a stochastic mixing model (MAP-SMM) [24], HS super-resolution (HySure) [25], and coupled non-negative matrix factorization [26]. TR-based methods treat HS and MS data as tensors and decompose them in different ways. Tucker decomposition and Canonical polyadic (CP) decomposition are widely used. The representative TR-based fusion methods include coupled sparse tensor factorization [27], nonlocal coupled tensor CP [28], low tensor multi-rank regularization (LTMR) [29], low tensor-train rank representation [30], unidirectional total variation with tucker decomposition (UTV) [31], tensor subspace representation-based regularization model (IR-TenSR) [32]. The deep CNN-based methods often train a network between ground-truth images (i.e., original HS images) and input images (i.e., low spatial resolution HS and high spatial resolution MS images). In recent years, many deep CNN-based methods have been proposed. The representative deep CNN-based methods include CNN-based fusion (CNN-Fus) [33], two-stream fusion network [34], HS image super-resolution network [35], spatial-spectral residual network [36] and so on. Although there are many different HS sharpening methods, how to accurately evaluate the quality of the HS sharpened images has not been well resolved in the fusion field. Due to the lack of ideal high spatial resolution HS images that can be referenced, no-reference quality assessment of HS sharpened images is still a challenging issue.

In general, quality assessment of fused remote sensing images consists of subjective evaluation and objective evaluation. Subjective evaluation, also known as qualitative evaluation, is the subjective visual evaluation of the fusion result by human observers, and the implementation of the evaluation often requires a lot of manpower and material resources. Objective evaluation, namely quantitative evaluation, is the use of various algorithms or metrics to calculate specific numerical values representing image quality. Figure 1 shows the process of HS sharpening with its objective quality assessment. At present, objective measurement can be classified into full-reference (FR) evaluation and no-reference (NR) evaluation. FR evaluation is more popular, since it is convenient and

reliable [37]. FR evaluation needs the reference image as the ground truth and uses it to measure the distorted image. However, the reality is that high spatial resolution (HR) HS images, i.e., the references, are extremely difficult to obtain or do not exist at all. Therefore, FR evaluation of HS sharpened images is usually conducted at a degraded scale. The specific procedure is to degrade both the original HS image and the original HR-MS image to respectively obtain the low spatial resolution (LR) HS image and the degraded LR-MS image. Thus, the original HS image is used as the reference. Then the LR-HS and the LR-MS images are fused together to obtain the fused images (i.e., the HS sharpened images). Finally, the FR quality assessment employs the reference image to measure the distortions of the HS sharpened images. Commonly used FR metrics include peak signal-to-noise ratio (PSNR) [38], spectral angle mapper (SAM) [39], erreur relative globale adimensionnelle de synthèse (ERGAS) [40], structural similarity (SSIM) [41], Q^{2n} [42], etc. Due to the inability of FR assessment to directly provide evaluation results for full-resolution fused images, some scholars shifted the research towards NR evaluation. In recent years, many metrics designed for multispectral sharpening have been proposed for NR evaluation. The typical NR metrics include Quality with No Reference (QNR) [43], filtered QNR (FQNR) [44], regression-based QNR (RQNR) [45], hybrid QNR (HQNR) [46], generalized QNR (GQNR) [47], etc. Moreover, some researchers proposed NR metrics based on a multivariate Gaussian distribution (MVG) model from the perspective of statistical characteristics [48]. With the development of deep learning, some novel evaluation methods based on a deep learning network have also been proposed [49]. However, most of the current NR metrics cannot be directly used for the quality assessment of HS sharpened images. For example, the well-known QNR metric requires measuring the difference between two bands, while HS sharpened images typically have hundreds of bands, which undoubtedly leads to a huge computational burden. To the best of our knowledge, there are few evaluation metrics specifically designed for NR quality assessment of HS sharpened images. Due to the large number of bands in HS images, HS image data are characterized by rich mathematical and statistical characteristics. Therefore, it is possible to use statistical distribution models to analyze the quality of HS sharpened images. Wu et al. extracted the first digit distribution (FDD) feature of the angle component of the hyperspherical color domain in pansharpened MS images and used it to evaluate the spectral quality of the fused images [50]. Inspired by this, a novel NR quality assessment method for HS sharpened images based on Benford's law is designed in the paper, and extensive experiments are conducted to verify the effectiveness of the proposed method. The contributions of this paper are as follows:

- Three statistical features following the standard Benford's law are discovered from HS images. The features are the FDDs of (1) the singular values of low-frequency coefficients from discrete wavelet transform (DWT), (2) the complementary values of the high-frequency correlation between HS bands and MS bands, and (3) the information preservation coefficients between HS bands and MS bands at different resolution scales.

- An effective NR evaluation method for HS sharpened images via Benford's law is proposed. The proposed method extracts the three FDD features from the HS sharpened images, then computes the distance between the extracted FDD features and the standard FDD feature, and finally obtains the evaluation results.

- Comprehensive experiments are conducted to verify the proposed method. The band assignment algorithm is applied to the four NR evaluation metrics of multispectral sharpening, and these adapted metrics are used to evaluate the HS sharpened images. Furthermore, the experiments adopt four commonly used FR evaluation metrics, three HS datasets, and 10 fusion methods to verify the accuracy and robustness of the proposed method.
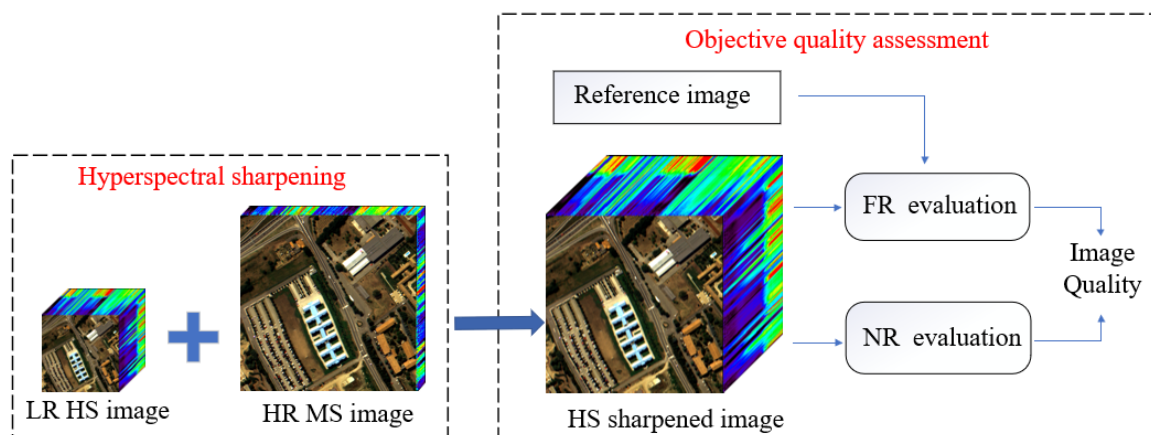
Figure 1. The process of HS sharpening and objective quality assessment.

The organizational structure of this paper is as follows: Section 2 briefly reviews related works. Section 3 provides a detailed introduction to the proposed method. Section 4 reports various experiments to validate the effectiveness of the proposed method. Section 5 discusses the performance of all related NR quality assessment metrics based on the experimental results and analysis. Finally, Section 6 concludes the paper.

2. Related Works

2.1. Wald's Protocol

Lucien Wald et al. proposed the first protocol to quantitatively evaluate the quality of multispectral sharpening (often referred to as the Wald's protocol) [51]. Wald's protocol is a widely used remote sensing image fusion quality evaluation strategy. It was further discussed and established in [52,53]. The protocol mentions that the pansharpened images should have two properties: *consistency* and *synthesis*.

The *consistency* property is conducted at the observation resolution of imaging sensors. If an HR fused image $\hat{\mathbf{F}}$ is spatially degraded to the LR fused image $\hat{\mathbf{F}}_L$ by using the same modulation transfer function (MTF) as the original HS imaging sensor, then the image $\hat{\mathbf{F}}_L$ should be as consistent as possible with the original LR-HS image $\mathbf{H}$. The *consistency* can be represented as

$$D_L(\hat{\mathbf{F}}_{\mathbf{L}i}, \mathbf{H}_i) \leq \varepsilon_i, \; i = 1, 2, \ldots, N \tag{1}$$

where $D_L(\mathbf{A}, \mathbf{B})$ represents the difference between two images $\mathbf{A}$ and $\mathbf{B}$ at a low spatial resolution, ε_i describes the error which should be small enough, i is the band index, and N is the number of bands in the HS image.

The *synthesis* property consists of two sub-properties. One states that the pansharpened image should be as identical as possible to the image acquired via the MS sensor if it had the high spatial resolution of the PAN sensor. In other words, the HR fused images $\hat{\mathbf{F}}$ should be as consistent as possible with the reference images $\mathbf{G}$. It can be represented as

$$D_H(\hat{\mathbf{F}}_i, \mathbf{G}_i) \leq \varepsilon_i, \; i = 1, 2, \ldots, N \tag{2}$$

where $D_H(\mathbf{A}, \mathbf{B})$ represents the difference between two images $\mathbf{A}$ and $\mathbf{B}$ at the resolution level of the fused images, i.e., the full resolution scale.

The other sub-property states that the spectral vectors of the HR fused images $\hat{\mathbf{F}}$ and reference images $\mathbf{G}$ should be as consistent as possible, which can be represented as

$$D_{SP}(\hat{\mathbf{F}}, \mathbf{G}) \leq \varepsilon \tag{3}$$

where $D_{SP}(\mathbf{A}, \mathbf{B})$ represents the distance between the spectral vectors of two images $\mathbf{A}$ and $\mathbf{B}$ at the resolution level of the fused image. The smaller the distance, the higher the spectral similarity.

2.2. FR Quality Assessment

FR quality assessment is a commonly used approach for evaluating the quality of fusion images. Currently, the mainstream of FR quality assessment follows the synthesis of the Wald's protocol which requires higher resolution images as the references, although [54] mentioned that *consistency* can also be used as an evaluation strategy for FR. However, in practical applications, it is hard to obtain the images using higher spatial resolution imaging sensors. The reference image is not available. In FR quality assessment of image fusion, it is feasible to perform spatial degradation on all input images (PAN, MS, and HS images) in order to use the original HS or MS images as the references. FR assessment is therefore called reduced resolution assessment. Since the degraded fusion images can be directly compared with the reference images, FR evaluation results are usually considered more accurate and reliable. Most of the current quality assessments of HS sharpened images are almost always based on FR quality assessment [37].

The FR quality assessment of HS sharpening has been an open problem. Firstly, the reliability of FR assessment will decrease as the resolution scale differences increase [38], which often occurs in HS sharpening. Secondly, the difficulty in obtaining MTF-filter parameters during the degradation process is a potential influencing factor. Alternatively, some researchers use uniform filtering [31,32] and Gaussian filtering [29,30,36] to implement the spatial degradation. The size and the variance of Gaussian kernels used are different, making the accuracy of spatial degradation unreliable. Thirdly, due to the lack of benchmark HS datasets, the current way adopted by researchers is to simulate LR-HS and HR-MS images from the original HS images as the fusion inputs, and then use the original HS images as the reference for quality evaluation [38,55], which is inconsistent with the actual implementation of FR quality assessment.

2.3. NR Quality Assessment

NR quality assessment, also known as full resolution quality assessment, does not have a reference, but instead uses the input images, namely HR-PAN or HR-MS images, as the spatial or spectral references. NR assessment is more in line with practical evaluation processes. Although many NR evaluation methods designed for multispectral sharpening have been proposed, there is little dedicated NR quality assessment method for HS sharpened images. The most widespread NR evaluation metric is QNR [43]. It concerns two distortions: spatial distortion and spectral distortion. The spatial distortion is measured using the difference in the Q values between PAN and MS images at different scales, while the spectral distortion is measured using the change in the Q values among MS bands. There have been some variants and improvements originating from the QNR metric so far. These variants include FQNR [44], RQNR [45], HQNR [46], and GQNR [47], to name a few. Also, some other NR methods have been designed. The quality-estimation-by-fitting (QEF) method predicts the evaluation results via linear regression from the FR evaluation results at different degraded resolution scales [56]. Kalman QEF (KQEF) is subsequently designed to improve the QEF [57]. Moreover, the MVG-based QNR method (MQNR) extracts multiple spatial and spectral features from the fused images to build the distorted MVG model. The distance between the benchmark MVG model and the distorted MVG model is used to indicate the evaluation results [48]. With the help of large-scale datasets, deep-learning-based NR methods have also emerged and been used for PAN-MS fusion quality assessment [49].

Some researchers have attempted to apply NR quality assessment metrics of multispectral sharpening to evaluate HS sharpened images. In the practice of using PAN images to increase the spatial resolution of HS images, Javier et al. used drones, orthophotos, and satellite high spatial resolution data to enhance the resolution of airborne and satellite HS images. When conducting NR quality assessment, they calculated the spatial correlation coefficients (sCC) between each band of the fused HS image and the PAN image and averaged all sCC results as an evaluation of spatial quality. In addition, they computed the SAM between the degraded fused images and the original HS images to evaluate

the spectral quality [58]. Meng et al. initially designed MQNR metric for the NR quality assessment of multispectral sharpening and applied the MQNR to HS sharpened images to verify its effectiveness [48]. In the cases of HS-MS image fusion, NR quality assessment is feasible in a band assignment way. The band assignment algorithm originated from HS-MS image fusion. According to the spectral overlap of MS and HS images, it divides the HS bands into multiple groups corresponding to each of MS bands, and then uses the pansharpening-based method to fuse the HS bands and MS images in each group [59]. Rocco et al. first applied the band assignment algorithm to NR quality assessment of HS sharpened images, when they conducted a study on HS sharpening using multiplatform data [15]. In NR quality evaluation, they proposed to assign each MS band to the related fused HS bands and obtained multiple MS-HS groups. Then, they adopted the spatial index of RQNR to evaluate the spatial quality of each group and the spectral index of FQNR to calculate the SAM between the degraded fused image and the original HS image. Another very interesting evaluation method is the task-based approach. Kawulok et al. retrieved the air pollution index, i.e., NO_2 column from HS sharpened images and compared it with the ground-truth NO_2 column to get the evaluation results [60]. Overall, although NR quality assessment faces more challenges than FR quality assessment, it is more practical and has less limitations.

2.4. Image Quality Assessment Based on Benford's Law

Benford's law is a numerical law. In 1881, Simon Newcomb discovered that the first few pages of the logarithmic table were more broken than the other pages, indicating a higher probability of the first digit being 1 in the values used for calculation [61]. Subsequently, Frank Benford rediscovered this phenomenon in 1935. After extensive testing and research, he found that as long as there were enough samples of data, the probability distribution of the first digit of the data followed a logarithmic criterion [62]. Benford's law has received attention from people since then. At present, data from many fields in reality always indicate compliance with Benford's law, such as mathematics [63], physics [64], social sciences [65], and environmental sciences [66]. Until today, Benford's law still guides people to explore the principles behind it [67].

Ou et al. found that FDD features extracted from the discrete cosine transform (DCT) coefficients of natural images are highly sensitive to white noise, Gaussian blur, and fast fading, and applied Benford's law to the quality assessment of natural images [68]. It has been verified in [69–71] that the high-frequency coefficients of discrete wavelet transform (DWT), shearlet coefficients, and singular values follow the standard Benford's law. Therefore, the standard Benford's law has been applied to distortion classification [69] and natural images quality assessment [70,71]. Additionally, the generalized Benford's law [72] was used to evaluate the quality of natural images with external noise such as Gaussian white noise, JPEG compression distortion, blur distortion, etc. For the first time, Wu et al. introduced Benford's law into the quality assessment of multispectral sharpening, who employed the extracted FDD feature of the angle components of the pansharpened images in the hyperspherical color domain to evaluate the spectral quality [50]. However, the method exhibits poor performance in evaluating HS sharpened images. Inspired by this, this paper proposes an NR method to evaluate the quality of HS sharpened images with the help of Benford's law.

3. Proposed Method

Figure 2 shows the flowchart of the proposed method. The method consists of two parts. One part is the FDD feature extraction. The FDD features include a low-frequency FDD feature, a high-frequency FDD feature, and a Q-based FDD feature. The other part is the distance measurement, which calculates the difference between the distributions: the extracted FDD features and the FDD feature of the standard Benford's law. A detailed introduction to each part of the proposed method will be provided in this section.
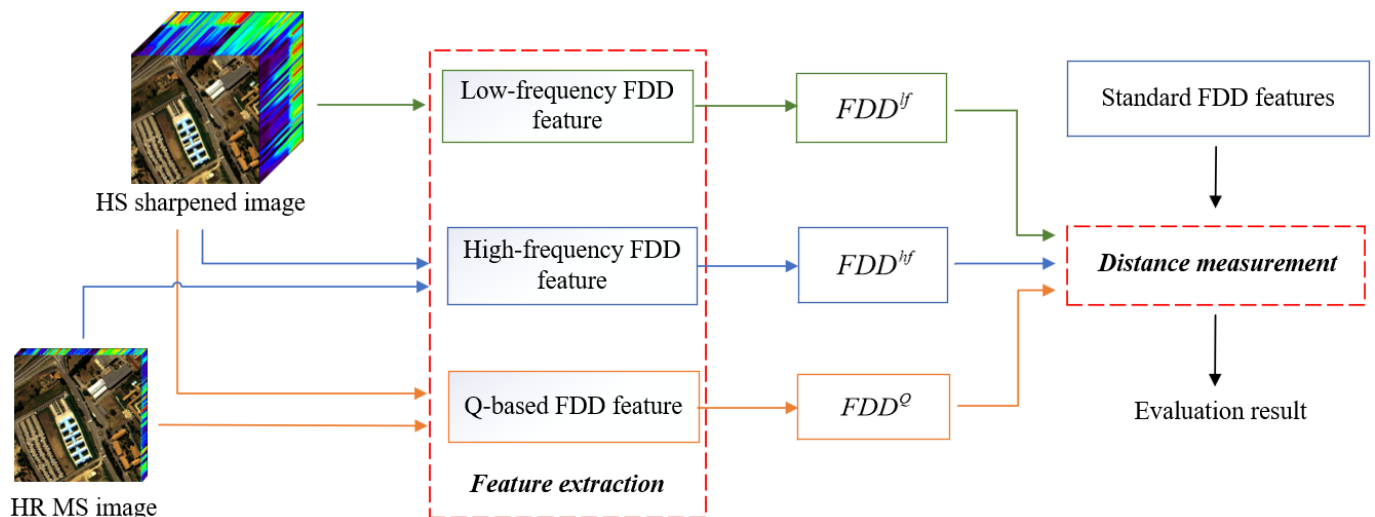
Figure 2. Flowchart of the proposed method.

3.1. Standard Benford's Law

The standard Benford's law records the distribution probability of the first digits about a certain type of data. The first digit refers to the first non-zero digit of a numerical value from left to right. For example, the first digit of 2.45 is 2 and the first digit of -0.32 is 3. The standard Benford's law is based on this numerical law, which holds that the distribution of the first digits of a large amount of data about a certain thing follows a logarithmic distribution. It can be described using the following:

$$P(n) = \log_{10} \left(1 + \frac{1}{n}\right), \ n = 1, 2, \ldots, 9, \tag{4}$$

where $P(n)$ records the FDD feature of the first digit being n. The standard Benford's law can be represented by a nine dimensional FDD feature, i.e., $[P(1), P(2), \ldots, P(9)]$. The standard FDD feature is denoted as FDD^{std} in (5) and the first digit distribution is plotted in Figure 3.

$$FDD^{std} = [0.3010, 0.1761, 0.1249, 0.0969, 0.0792, 0.0669, 0.0580, 0.0512, 0.0459] \tag{5}$$
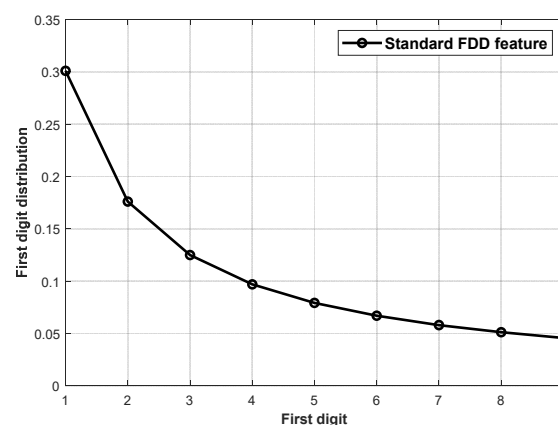
Figure 3. The distribution of the FDD^{std}.

It is assumed that the FDD features extracted from the ideal fusion products should conform to the standard Benford's law as much as possible, so that the standard FDD feature FDD^{std} can be used as the reference model. Fortunately, three quality perception features in high-quality HS image data are found. Through extensive experiments, it is

verified that the features comply with the standard Benford's law and are sensitive to distortions in HS sharpened images.

3.2. Quality Perception Features

HS images often have dozens or hundreds of spectral bands, which contain a large amount of information and have complex potential statistical characteristics. With the help of Benford's law, it is possible to identify some stable features extracted from the HS images for quality assessment. After extensive experiments, three types of FDD features are found out, including a low-frequency FDD feature, a high-frequency FDD feature, and a Q-based FDD feature, which can effectively perceive the distortions of fusion results. A detailed description of the three types of FDD features is provided below.

3.2.1. Low-Frequency FDD Feature

Singular value decomposition has been widely used in image denoising in which the degree of image distortion is determined by detecting singular values. Moreover, the compact support property of Meyer wavelets makes them important and valuable in signal processing, image reconstruction, and analysis of distribution functions [73]. Motivated by this, this work tries to find the representative statistical features to characterize the quality of the HS images. Through lots of experiments on HS images, it is found that the FDD features of the singular values of the low-frequency coefficients within Meyer wavelet domain follow the standard Benford's law. Firstly, Meyer wavelet transform is performed on each band of the HS image to obtain the low-frequency coefficients of the approximation images. Then, singular value decomposition is used so the coefficients and the singular values of all HS bands are obtained. Finally, the low-frequency FDD (denoted as FDD^{lf}) feature is calculated based on the singular values.

To verify the effectiveness of the low-frequency FDD feature, experiments of the feature extraction are conducted on three different HS datasets, i.e., Pavia University dataset, Salinas dataset, and Cuprite dataset (see the detailed information of the datasets in Section 4.1). Figure 4 displays the distribution of the three FDD^{lf} features from the datasets and the distribution of the FDD^{std}. Table 1 reports the specific values of the FDD^{std} and FDD^{lf} features.

Table 1. The specific values of the FDD^{std} and FDD^{lf} features from three HS datasets. The first column represents the sources of the extracted FDD^{lf} features and the FDD^{std}. P(n) records the FDD^{std} and FDD^{lf} features of the first digit being *n.* The last column provides the symmetric Kullback–Leibler divergence (*sKL*) between the FDD^{std} and FDD^{lf} features.

	P(1)	*P*(2)	*P*(3)	*P*(4)	*P*(5)	*P*(6)	*P*(7)	*P*(8)	*P*(9)	*sKL*
Pavia University	0.2969	0.1784	0.1262	0.0968	0.0791	0.0708	0.0569	0.0490	0.0460	3.1×10^{-4}
Salinas	0.3021	0.1699	0.1218	0.0975	0.0829	0.0681	0.0600	0.0517	0.0460	4.2×10^{-4}
Cuprite	0.3004	0.1687	0.1294	0.0980	0.0800	0.0686	0.0578	0.0525	0.0446	4.4×10^{-4}
FDD^{std}	0.3010	0.1761	0.1249	0.0969	0.0792	0.0669	0.0580	0.0512	0.0459	0

To compare the difference between the FDD features, symmetric Kullback–Leibler divergence (*sKL*) is used to measure the distance between the FDD^{std} and FDD^{lf}. The calculation formula for *sKL* is as follows:

$$sKL(A, B) = \frac{1}{2} KL(A, B) + \frac{1}{2} KL(B, A) \tag{6}$$

where *A* and *B* are the first digit distributions to be measured and Kullback–Leibler (*KL*) divergence is given by

$$KL(A, B) = \sum_{i=1}^{9} A(i) \log_2 \frac{A(i)}{B(i)} \tag{7}$$
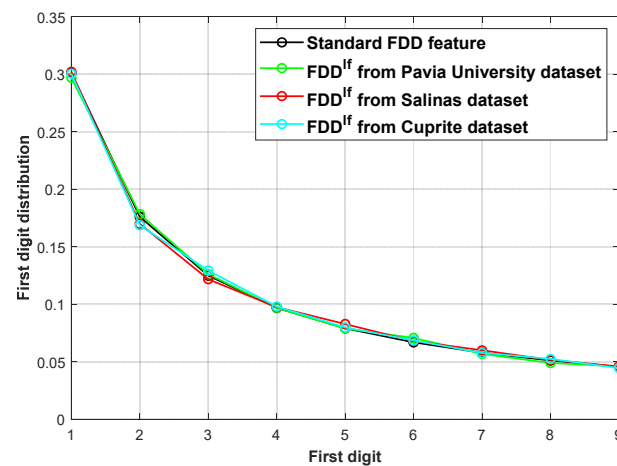
Figure 4. The distributions of the FDD^{std} and FDD^{lf} features from three HS datasets.

From the values in Table 1, it can be seen that the sKL divergence between the FDD^{lf} features and the FDD^{std} are close to 10^{-4}. The distance between two probability distributions is very small. Based on the observations of Figure 4 and Table 1, it can be concluded that the FDD^{lf} feature of the original HS images conform to the standard Benford's law. This makes it possible to evaluate the quality of HS sharpened images based on the low-frequency FDD feature.

In order to test the sensitivity of the FDD^{lf} feature to the quality of distorted HS images, two experiments are conducted on the HS datasets. In the first experiment, different types of noise are added to the datasets separately. The noises include Gaussian white noise with a mean of 0 and a gradually increasing standard deviation, Gaussian blur noise with a gradually increasing standard deviation, and multiplicative noise with a gradually increasing standard deviation. In the second experiment, four HS sharpened images, i.e., IR-TenSR [32], UTV [31], SFIM [20], MAP-SMM [24], and EXP (i.e., the upsampled image by cubic interpolation) with increasing fusion distortions are selected. Then, the FDD^{lf} features are extracted from these distorted images and the FDD^{std} is used as the benchmark for distance measurement. Taking the experiment on the Pavia University dataset as an example, Figure 5 shows the distribution of the FDD^{lf} features from the different distorted images and Table 2 gives the values of the FDD^{lf} features from the HS sharpened images.

Table 2. The specific values of the FDD^{std} and FDD^{lf} features from different fused images of the Pavia University dataset.

	P(1)	*P*(2)	*P*(3)	*P*(4)	*P*(5)	*P*(6)	*P*(7)	*P*(8)	*P*(9)	*sKL*
Reference	0.2969	0.1784	0.1262	0.0968	0.0791	0.0708	0.0569	0.0490	0.0460	3.1×10^{-4}
IR-TenSR [32]	0.3025	0.1809	0.1270	0.0932	0.0788	0.0650	0.0568	0.0504	0.0453	3.0×10^{-4}
UTV [31]	0.2911	0.1841	0.1318	0.0965	0.0818	0.0675	0.0522	0.0507	0.0444	1.3×10^{-3}
SFIM [20]	0.2893	0.1661	0.1284	0.1005	0.0822	0.0730	0.0632	0.0508	0.0466	1.7×10^{-3}
MAP-SMM [24]	0.2817	0.1641	0.1286	0.1066	0.0809	0.0744	0.0631	0.0534	0.0472	3.3×10^{-3}
EXP	0.3336	0.2179	0.1234	0.0822	0.0624	0.0469	0.0460	0.0447	0.0428	2.1×10^{-2}
FDD^{std}	0.3010	0.1761	0.1249	0.0969	0.0792	0.0669	0.0580	0.0512	0.0459	0

From Figure 5a–c, it can be seen that for different types of external noise, as the noise increases, the FDD^{lf} features of the distorted images further deviate from the FDD^{std}, indicating that the FDD^{lf} features exhibit good performance in measuring external noise. In Figure 5d and Table 2, the FDD^{lf} features of the different fused images and EXP image with increasing fusion distortion are presented. Table 2 shows that as the degree of fusion distortion increases, the values of sKL divergence also monotonically increase. This phenomenon can also be observed in Figure 5d. The two experiments demonstrate that the FDD^{lf} feature

can be selected as a quality perception feature to evaluate the quality of HS sharpened images.
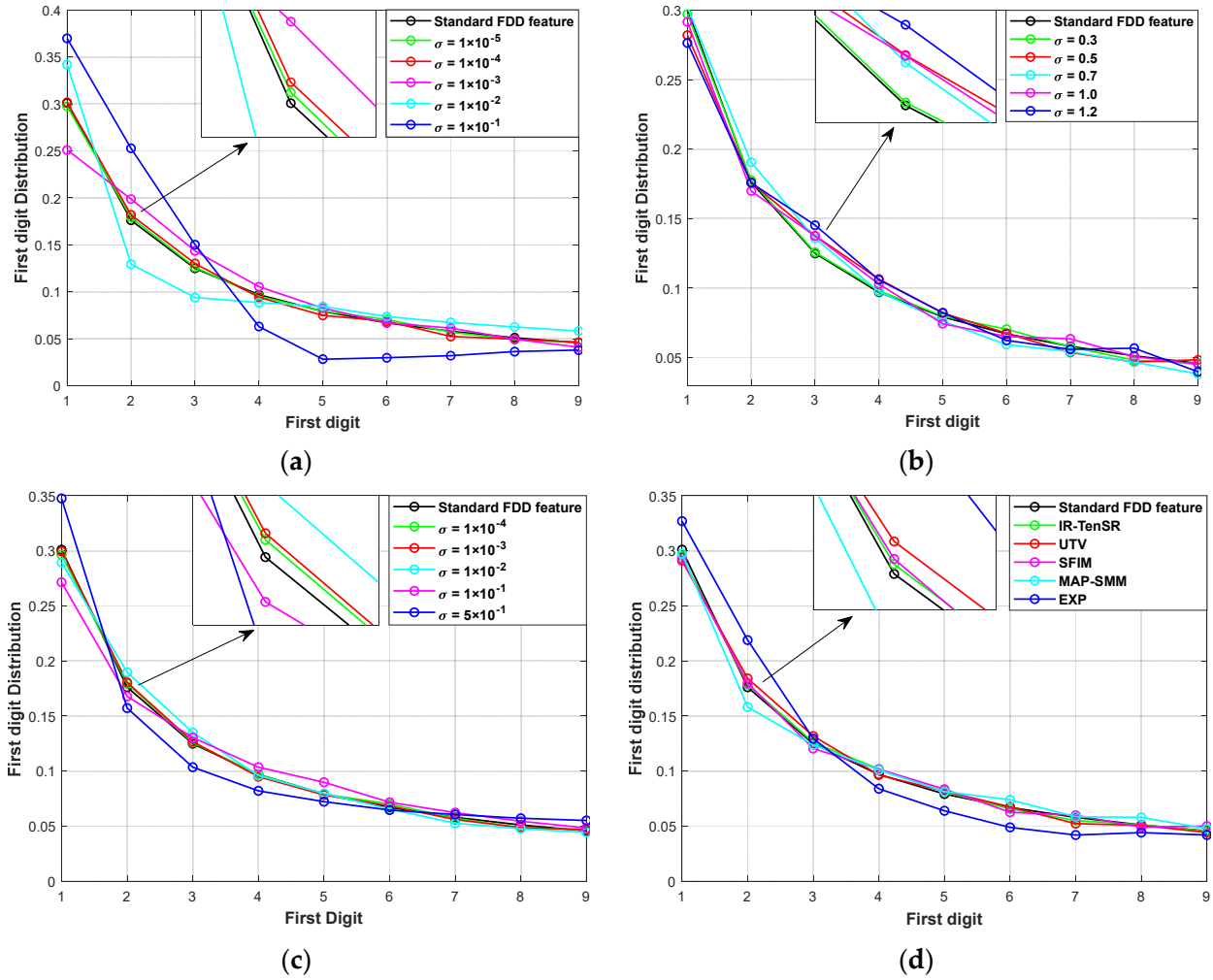
Figure 5. The distributions of the FDD^{lf} features from different distorted images of the Pavia University dataset: (**a**) images distorted by Gaussian white noise with gradually increasing standard deviation; (**b**) images distorted by Gaussian blur noise with gradually increasing standard deviation; (**c**) images distorted by multiplicative noise with gradually increasing standard deviation; (**d**) HS sharpened images with increasing fusion distortion.

3.2.2. High-Frequency FDD Feature

For the quality evaluation of natural images, the FDD features extracted from high-frequency coefficients of DWT, DCT, and shearlet have been proven to be effective [69–71]. Unfortunately, these typical FDD features are ineffective for hyperspectral images and cannot be used for quality evaluation of HS sharpened images. This is because the distortion of natural images often comes from external noise, such as image transmission, image compression, etc. The distortion of fused images, as an internal distortion, is generated in the fusion process. Therefore, the FDD features of natural images may view fusion distortion as the structure of the image itself, rather than distortion. To tackle with this problem, a new way is designed to obtain an effective high-frequency FDD feature as a quality perception feature.

(1)   The band assignment on the fused HS image and the HR-MS image is performed according to [59]. Suppose the original HS image of N bands and the HR-MS image of M bands are fused, the HS sharpened image will have N bands. After band assignment, M groups can be obtained so that each group contains one MS band and

several corresponding HS bands. Then, the quality assessment of HS sharpened image can be treated as a combination of multiple pansharpening quality assessment cases.

(2) For each group, the high-frequency components are extracted from the HS bands and the MS band, respectively. The extraction process of the high-frequency component of HS/MS band is as follows:

$$\mathbf{X}^{hf} = \mathbf{X} - LPF_{Gauss}(\mathbf{X}) \tag{8}$$

where $\mathbf{X}$ is the MS/HS band, $LPF_{Gauss}(\cdot)$ is Gaussian low-pass filtering, and $\mathbf{X}^{hf}$ is the high-frequency component of $\mathbf{X}$.

Next, the high-frequency components of HS/MS bands are divided into non-overlapping blocks and the correlation coefficients between the blocks of each HS band and the MS band are calculated. The complementary values of the correlation coefficients are calculated as

$$\mathbf{HF}^{\mathrm{D}}_{i,k} = 1 - CC(\mathbf{F}^{hf}_{i,k}, \mathbf{M}^{hf}_k), \; i = 1, 2, \ldots, T; \; k = 1, 2, \ldots, M \tag{9}$$

where $\mathbf{HF}^{\mathrm{D}}_{i,k}$ is the high-frequency difference matrix between the ith fused HS band and the MS band in the kth group, $\mathbf{F}^{hf}_{i,k}$ is the high-frequency component of the ith fused HS band in the kth group, and $\mathbf{M}^{hf}_k$ is the high-frequency component of the MS band in the same group. $CC(\cdot)$ is the calculation of correlation coefficients in blocks, T is the number of HS bands in kth group, and M is the number of the groups. The set of all high-frequency difference matrices is denoted as $\mathbf{HF}^{\mathrm{D}}$. The lower the high-frequency similarity between two bands, the higher the values of $\mathbf{HF}^{\mathrm{D}}$, which facilitates the detection of the spatial distortion in the HS sharpened images.

(3) The high-frequency FDD feature, denoted as FDD^{hf} is obtained by counting the number of the first digit of $\mathbf{HF}^{\mathrm{D}}$.

Similar to Section 3.2.1, the high-frequency FDD feature is experimentally verified to comply with the standard Benford's law (The detailed information of the fused HS image and HR-MS image in the experiment are provided in Section 4.1). Then, FDD^{hf} features from three HS datasets are extracted separately and compared with the FDD^{std}. The FDD^{hf} features of three HS datasets are presented in Figure 6 and Table 3. It can be seen that the values of sKL are very small reaching the order of 10^{-3} or 10^{-4}, and the curves of the FDD^{hf} features match well with the FDD^{std}. The experimental results demonstrate that the FDD^{hf} can serve as a stable high-frequency feature representing the quality of HS images.
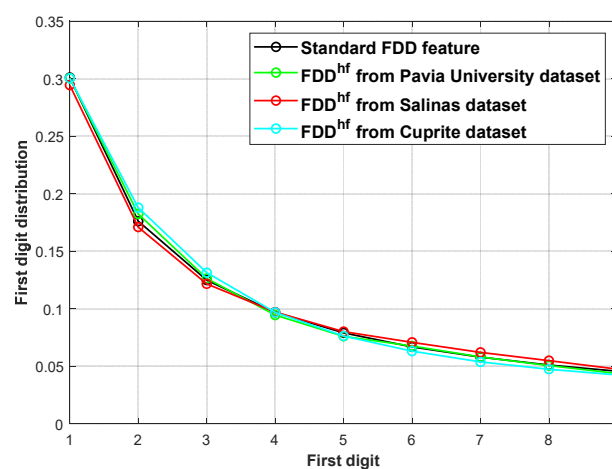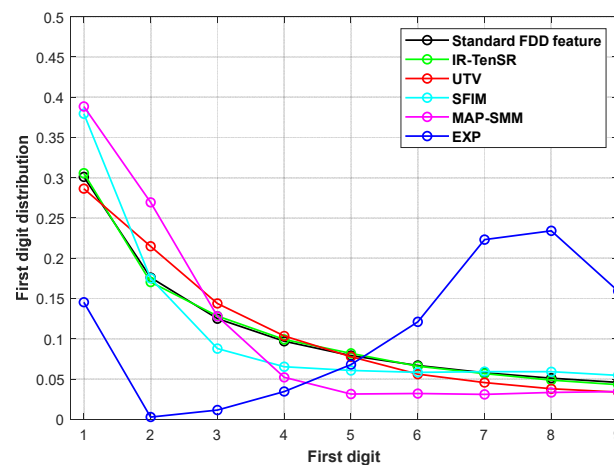

Figure 6. The distributions of the FDD^{std} and FDD^{hf} features from three HS datasets.

Table 3. The specified values of the FDD^{std} and the FDD^{hf} features from three HS datasets.

	P(1)	P(2)	P(3)	P(4)	P(5)	P(6)	P(7)	P(8)	P(9)	sKL
Pavia University	0.3004	0.1824	0.1264	0.0945	0.0761	0.0677	0.0582	0.0506	0.0438	3.7×10^{-4}
Salinas	0.2944	0.1710	0.1217	0.0971	0.0802	0.0709	0.0621	0.0550	0.0476	9.0×10^{-4}
Cuprite	0.3004	0.1879	0.1314	0.0969	0.0764	0.0632	0.0538	0.0475	0.0426	1.6×10^{-3}
FDD^{std}	0.3010	0.1761	0.1249	0.0969	0.0792	0.0669	0.0580	0.0512	0.0459	0

Furthermore, FDD^{hf} features are extracted from four different fused images and EXP on three HS datasets. Figure 7 and Table 4 show the experimental results on the Pavia University dataset. In Figure 7, it can be seen that the FDD^{hf} features of different fused images significantly deviate from the FDD^{std} and the FDD^{hf} feature of EXP appears the maximum deviation. From the values of sKL in Table 4, it can be seen that as the distortion of the fused image increases, the values of the FDD^{hf} features also increase and the trend of the change is consistent. The experiment demonstrates that the high-frequency FDD feature is an appropriate quality perception feature for HS images and has the ability to detect fusion distortion.

Figure 7. The distributions of the FDD^{std} and FDD^{hf} features from the fused images of the Pavia University dataset.

Table 4. The specific values of the FDD^{std} and FDD^{hf} features from the fused images of the Pavia University dataset.

	P(1)	P(2)	P(3)	P(4)	P(5)	P(6)	P(7)	P(8)	P(9)	sKL
Reference	0.3004	0.1824	0.1264	0.0945	0.0761	0.0677	0.0582	0.0506	0.0438	3.7×10^{-4}
IR-TenSR [32]	0.3062	0.1705	0.1274	0.0993	0.0817	0.0659	0.0572	0.0488	0.0430	5.6×10^{-4}
UTV [31]	0.2864	0.2149	0.1439	0.1035	0.0778	0.0561	0.0457	0.0379	0.0339	1.7×10^{-2}
SFIM [20]	0.3806	0.1754	0.0885	0.0649	0.0606	0.0575	0.0591	0.0588	0.0547	3.8×10^{-2}
MAP-SMM [24]	0.3905	0.2671	0.1279	0.0517	0.0315	0.0313	0.0314	0.0335	0.0351	1.3×10^{-1}
EXP	0.1457	0.0041	0.0099	0.0357	0.0660	0.1236	0.2201	0.2334	0.1614	1.28
FDD^{std}	0.3010	0.1761	0.1249	0.0969	0.0792	0.0669	0.0580	0.0512	0.0459	0

3.2.3. Q-Based FDD Feature

Since the Q index was proposed by Wang et al. in 2002 [74], it has demonstrated a powerful ability to measure the similarity between two images. Several variations of the Q-based QNRs, such as FQNR, RQNR, and HQNR, are the most commonly used in NR-quality assessment of multispectral sharpening. Motivated by this, it is meaningful to find a Q-based FDD feature satisfying the assumption of the preservation of the statistical Q through the hyperspectral sharpening process.

(1) Similar to Section 3.2.2, the band assignment is performed to the HS sharpened image and the HR-MS image and *M* groups are obtained. In each group, there are one MS band and several HS bands.

(2) For each group, the Q values between the MS and HS bands at different resolution scales are calculated. For a certain resolution scale, the Q values are calculated as [74]

$$Q(\mathbf{X}, \mathbf{Y}) \triangleq \frac{4\sigma_{\mathbf{XY}} \cdot \overline{\mathbf{X}} \cdot \overline{\mathbf{Y}}}{(\sigma_{\mathbf{X}}^2 + \sigma_{\mathbf{Y}}^2)(\overline{\mathbf{X}}^2 + \overline{\mathbf{Y}}^2)} \tag{10}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are the reference MS band and the fused HS band, respectively. $\sigma_{\mathbf{XY}}$ is the covariance between $\mathbf{X}$ and $\mathbf{Y}$, $\overline{\mathbf{X}}$ and $\overline{\mathbf{Y}}$ are the means, and $\sigma_{\mathbf{X}}^2$ and $\sigma_{\mathbf{Y}}^2$ are the variances of $\mathbf{X}$ and $\mathbf{Y}$, respectively. The series of Q-value at different resolution scales are given by

$$\mathbf{Q}_{i,k}^{D(r)} \triangleq Q(\mathbf{M}_k, \hat{\mathbf{F}}_{i,k}) - Q(\widetilde{\mathbf{M}}_k^{(r)}, \mathbf{H}_{i,k}^{(r)}), \quad i = 1, 2, \ldots, T; \ k = 1, 2, \ldots, M \tag{11}$$

where $\hat{\mathbf{F}}_{i,k}$ is the *i*th fused HS band and $\mathbf{M}_k$ is the HR-MS band in the *k*th group. $\mathbf{H}_{i,k}^{(r)}$ is the *i*th original LR-HS band in the *k*th group and $\widetilde{\mathbf{M}}_k^{(r)}$ is the spatially degraded version of $\mathbf{M}_k$, in which the superscript (*r*) indicates the spatial resolution ratio between the fused image and the original image. *T* is the number of HS bands in the *k*th group and *M* is the number of groups. Similarly, the set of all $\mathbf{Q}_{i,k}^{D(r)}$ is denoted as $\mathbf{Q}^{D(r)}$.

(3) The Q-based FDD feature, denoted as *FDD*^{Q}, is obtained by counting the number of the first digit on the $\mathbf{Q}^{D(r)}$ matrix.

Also, the experiment on the Pavia University dataset is taken as an example to show the experimental validation. Considering the different spatial resolutions of LR-HS images and HR-MS images in practice, the original HS image is degraded at five different resolution scales and the degradation ratios are 2, 4, 6, 8, and 10. At each resolution scale, the corresponding Q-based FDD features are extracted, i.e., $\mathbf{Q}^{D(2)}$, $\mathbf{Q}^{D(4)}$, $\mathbf{Q}^{D(6)}$, $\mathbf{Q}^{D(8)}$, and $\mathbf{Q}^{D(10)}$. The experiment results are shown in Figure 8 and Table 5. It can be seen that the Q-based FDD features at different scales have stable values around 10^{-3} and basically match with the standard FDD feature. Therefore, for HR-MS and LR-HS images with different spatial resolution ratios in practice, the *FDD*^{Q} can be used as a feature to represent the quality of HS images.
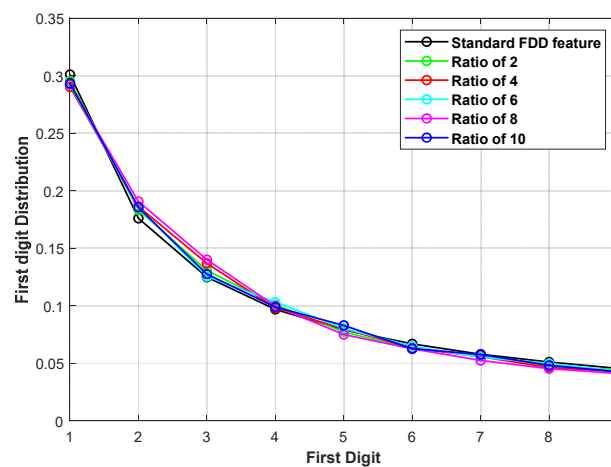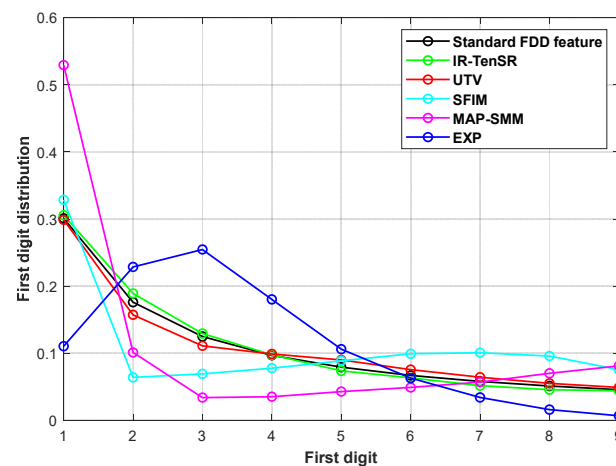

Figure 8. The distributions of the *FDD*^{std} and *FDD*^{Q} features from the Pavia University dataset at different resolution ratios.

Table 5. The specific values of the FDD^{std} and FDD^Q features from the Pavia University dataset at different resolution ratios.

	P(1)	P(2)	P(3)	P(4)	P(5)	P(6)	P(7)	P(8)	P(9)	sKL
Ratio of 2	0.2954	0.1834	0.1307	0.1010	0.0774	0.0626	0.0573	0.0477	0.0445	1.0×10^{-3}
Ratio of 4	0.2902	0.1863	0.1369	0.0980	0.0800	0.0634	0.0558	0.0465	0.0428	2.2×10^{-3}
Ratio of 6	0.2921	0.1855	0.1255	0.1033	0.0804	0.0649	0.0553	0.0499	0.0431	1.1×10^{-3}
Ratio of 8	0.2924	0.1907	0.1401	0.0998	0.0751	0.0627	0.0526	0.0453	0.0413	3.9×10^{-3}
Ratio of 10	0.2933	0.1856	0.1276	0.0989	0.0830	0.0628	0.0577	0.0483	0.0428	1.2×10^{-3}
FDD^{std}	0.3010	0.1761	0.1249	0.0969	0.0792	0.0669	0.0580	0.0512	0.0459	0

Taking the $\mathbf{Q}^{D(4)}$ as an example, the Q-based FDD features from four fused images are extracted. Figure 9 and Table 6 show their distributions and specific values. It can be seen that the Q-based FDD feature of IR-TenSR with lower distortion are closer to the standard FDD feature. Due to the more severe fusion distortion of SFIM, MAP-SMM, and EXP, their Q-based FDD features deviate further from the standard FDD feature. Therefore, the Q-based FDD feature can effectively perceive fusion distortion.

Figure 9. The distributions of the FDD^{std} and FDD^Q features from the fused images of the Pavia University dataset.

Table 6. The specific values of the FDD^{std} and FDD^Q features from the fused images of the Pavia University dataset.

	P(1)	P(2)	P(3)	P(4)	P(5)	P(6)	P(7)	P(8)	P(9)	sKL
Reference	0.2902	0.1863	0.1369	0.0980	0.0800	0.0634	0.0558	0.0465	0.0428	2.2×10^{-3}
IR-TenSR [32]	0.3056	0.1891	0.1294	0.0975	0.0737	0.0633	0.0516	0.0456	0.0442	2.3×10^{-3}
UTV [31]	0.2987	0.1573	0.1111	0.0989	0.0900	0.0756	0.0643	0.0552	0.0490	5.4×10^{-3}
SFIM [20]	0.3286	0.0640	0.0693	0.0778	0.0888	0.0992	0.1009	0.0959	0.0755	1.7×10^{-1}
MAP-SMM [24]	0.5293	0.1012	0.0340	0.0353	0.0429	0.0490	0.0570	0.0701	0.0813	2.9×10^{-1}
EXP	0.1105	0.2286	0.2546	0.1802	0.1061	0.0629	0.0342	0.0160	0.0069	3.5×10^{-1}
FDD^{std}	0.3010	0.1761	0.1249	0.0969	0.0792	0.0669	0.0580	0.0512	0.0459	0

3.3. Distance Measurement

In the above experiments, three quality perception features for HS sharpening have been identified and validated on different fused images of different HS datasets. Three quality perception features are the low-frequency FDD feature FDD^{lf}, the high-frequency FDD feature FDD^{hf}, and the Q-based FDD feature FDD^Q. The features measure the fusion distortion from different perspectives. Therefore, it is necessary to integrate them in order

to comprehensively evaluate the quality of HS sharpened images. Here, the standard FDD feature FDD^{std} is taken as the benchmark for distance measurement. Specifically, Manhattan distance is used to measure the distance between the distorted FDD features and the standard FDD feature. The distance measurement is calculated as

$$Q_{FDD} = 1 - \sum_{i=1}^{9} \left| FDD_i^{std} - FDD_i^{lf} \right| + \left| FDD_i^{std} - FDD_i^{hf} \right| + \left| FDD_i^{std} - FDD_i^{Q} \right| \quad (12)$$

The ideal value of Q_{FDD} is 1, and the larger the value, the higher the quality of the HS sharpened image.

4. Experiments

4.1. Datasets

In the experiments, three typical HS datasets, i.e., Pavia University dataset [55], Salinas dataset [9], and Cuprite dataset [55] are selected.

- The Pavia University dataset was collected using German Reflective Optics Spectrographic Imaging System (ROSIS-03) sensors and recorded partial scenes of Pavia city, Italy. The dataset contains 103 available bands, covering a spectral range of 0.43–0.86 μm. The dimension is 610 × 340 pixels with a spatial resolution of 1.3 m.
- The Salinas dataset was captured using the AVIRIS sensor of NASA and recorded parts of the Salinas Valley in California. The dataset has 204 available bands and covers a spectral range of 0.4–2.5 μm. The spatial resolution of the HS images is 3.7 m with the dimension of 512 × 217 pixels.
- The Cuprite dataset was also collected using the AVIRIS sensor and recorded parts of the Cuprite area in Nevada. The original dataset contains 224 bands and the spectral range covers from 0.37–2.48 μm. The bad bands with low signal-to-noise are removed and 185 bands are retained for the experiment. The dimension of the HS images is 512 × 612 pixels with a spatial resolution of 20 m.

To generate LR-HS and HR-MS images for HS-MS fusion, spatial and spectral simulations are conducted on these HS datasets. The simulation process is the same as that in [55]. Specifically, spatial filtering is performed on the original HS image to obtain the LR-HS image. The convolutional kernel of the spatial filtering is a 7 × 7 Gaussian kernel with zero-mean and a standard deviation of 3. The spatial resolution of the obtained LR-HS image is 1/4 of the original HS image. To obtain HR-MS image, the spectral response functions of the IKONOS sensor and the WorldView3 sensor are used as the spectral filters. The spectral coverage range of the IKONOS sensor is 0.45–0.9 μm [55], while the spectral coverage range of the Worldview3 sensor is 0.4–2.365 μm [15]. Therefore, the Pavia University dataset is filtered using the spectral response function of IKONOS and HR-MS image with four bands is obtained. Similarly, the Salinas dataset and Cuprite dataset are filtered using the spectral response function of WorldView3 and HR-MS images with 16 bands are obtained.

4.2. HS Sharpening Methods

Among the four categories of fusion methods mentioned in Section 1, ten HS sharpening methods are selected.

- Classical pansharpening-based methods. Four methods, including PCA [17], GSA [18], GLP [19], and SFIM [20] have been selected. PCA and GSA are typical CS-based fusion methods, which are built on principal component transformation and Gram-Schmidt orthogonal transformation, respectively. They have been widely used in many practical applications. GLP and SFIM are representative MRA-based fusion methods, which improve the spatial resolution of LR images by injecting the spatial details obtained through spatial filtering. After the band assignment, these methods can be applied to fuse HS images, since HS sharpening can be seen as a multi pansharpening cases.

- MF-based methods. Two methods, MAP-SMM [24] and HySure [25], have been selected. MAP-SMM utilizes a stochastic mixing model to estimate the underlying spectral scene content and develops a cost function to optimize the estimated HS data relative to the input HS/MS images. HySure formulates the HS sharpening as the minimization of a convex objective function with respect to subspace coefficients.
- TR-based methods. Three methods, including LTMR [29], UTV [31], and IR-TenSR [32] have been selected. LTMR learns the spectral subspace from the LR-HS image via singular value decomposition, and then estimates the coefficients via the low tensor multi-rank prior. UTV utilizes the classical Tucker decomposition to decompose the target HR-HS image as a sparse core tensor multiplied by the dictionary matrices along with the three modes. Then it conducts proximal alternating optimization scheme and the alternating direction method of multipliers to iteratively solve the proposed model. IR-TenSR integrates the global spectral–spatial low-rank and the nonlocal self-similarity priors of HR-HS image. Then it uses an iterative regularization procedure and develops an algorithm based on the proximal alternating minimization method to solve the proposed model.
- Deep CNN-based method. CNN-Fus method [33] is selected for experiment. CNN-Fus learns the subspace from LR-HS image via singular value decomposition. Then it approximates the desired HR-HS image with the low-dimensional subspace multiplied by the coefficients. It uses the well-trained CNN designed for gray image denoising to regularize the estimation of coefficients.

4.3. Quality Assessment Metrics

To evaluate the quality of HS sharpened images, four FR metrics are selected: PSNR [38], SAM [39], ERGAS [40], and Q^{2n} [42]. PSNR evaluates the spatial reconstruction quality, SAM is the metric for evaluating angular distortion of spectra, and ERGAS and Q^{2n} measure the overall distortion. Furthermore, the proposed method is compared with four NR evaluation metrics including QNR [43], FQNR [44], RQNR [45], and MQNR [48]. As mentioned in Section 3.2.2, after applying the band assignment to the HS and MS bands, the evaluation of the fused images can be seen as a combination of multiple pansharpening evaluation cases. To obtain the final evaluation results, these NR metrics are performed to HS sharpened images in each group and the evaluations of all groups are averaged. In addition, the spatial distortion indexes and spectral distortion indexes of the NR metrics are analyzed in Section 4.5.4. The spatial distortion index and spectral distortion index of QNR are respectively denoted as D_{Qs} and $D_{Q\lambda}$. Similarly, D_{Fs} and $D_{F\lambda}$ are the spatial and spectral indexes of FQNR, D_{Ms} and $D_{M\lambda}$ are the spatial and spectral indexes of MQNR. D_{Rs} is the spatial index of RQNR.

To analyze the reliability and accuracy of all NR metrics, the evaluation results of FR metrics are used as the testing benchmark. In other words, if the result of an NR metric is consistent with the results of FR metrics, then the NR metric is considered accurate. Pearson linear correlation coefficient (PLCC) [71], Spearman rank-order correlation coefficient (SROCC) [71], and Kendall rank-order correlation coefficient (KROCC) [71] are used to calculate the correlation between the results of NR metrics and the benchmark. PLCC reflects the evaluation accuracy of the NR metrics. SROCC and KROCC are used to reflect the evaluation monotonicity of the NR metrics. Better performance should provide higher PLCC, SROCC, and KROCC values.

4.4. Experiment Environment

To ensure a fair comparison, all experiments in this paper are coded on MATLAB 2021a, using a laptop equipped with memory of 16.0 GB RAM, AMD Ryzen 7 6800H CPU, Radeon Graphics, and NVIDIA GeForce RTX 3060 GPU.

4.5. Experimental Results and Analysis

4.5.1. Subjective Evaluation

Figures 10–12 respectively show the sub scenes in the fusion results of the Pavia University, Salinas, and Cuprite HS datasets. Although it is difficult to visually analyze the results of HS sharpening due to hundreds of HS bands, spectral filtering is performed on the fused HS images to obtain MS images, and then select the red, green, and blue bands for true color composite display. The zoomed view of a local area is displayed in the upper right. The corresponding RMSE heat maps are also provided below each fused image.
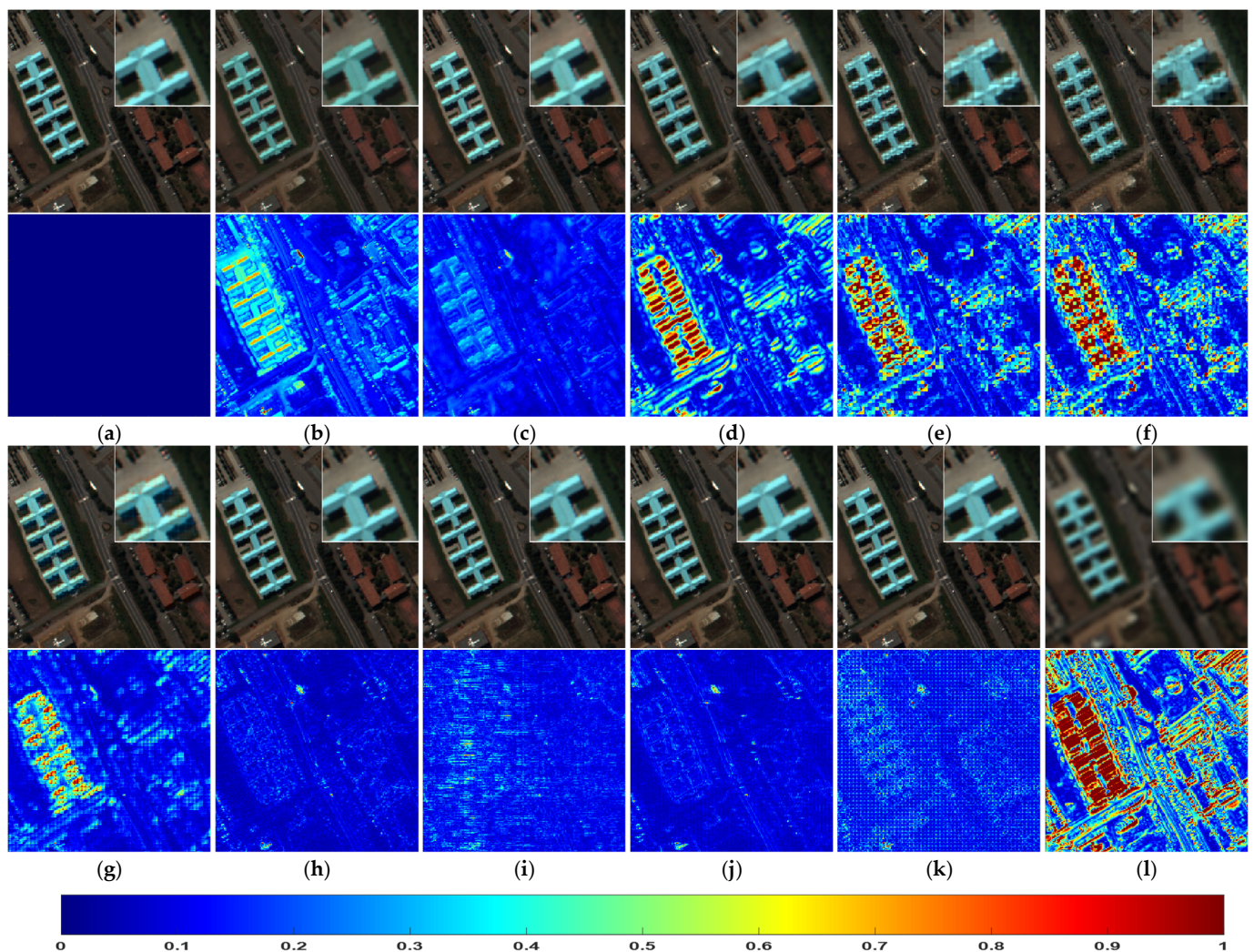
Figure 10. Fused images of different HS sharpening methods on Pavia University dataset: (**a**) REF; (**b**) PCA; (**c**) GSA; (**d**) GLP; (**e**) SFIM; (**f**) MAP-SMM; (**g**) HySure; (**h**) LTMR; (**i**) UTV; (**j**) IR-TenSR; (**k**) CNN-Fus; (**l**) EXP.

For the Pavia University dataset, it can be seen from the true color images in Figure 10 that SFIM and MAP-SMM images have obvious blocking distortion, EXP appears too blurry, and other fused images have better visual effects. The RMSE heat maps can help us easily distinguish the subtle visual difference among the fused results. The images in Figure 10h,j show that LTMR and IR-TenSR have the minimal spatial distortion. GSA, a traditional pansharpening method, has also achieved good fusion result. GLP, SFIM, MAP-SMM show the poor results as the significant distortions spread uniformly in the whole scene.
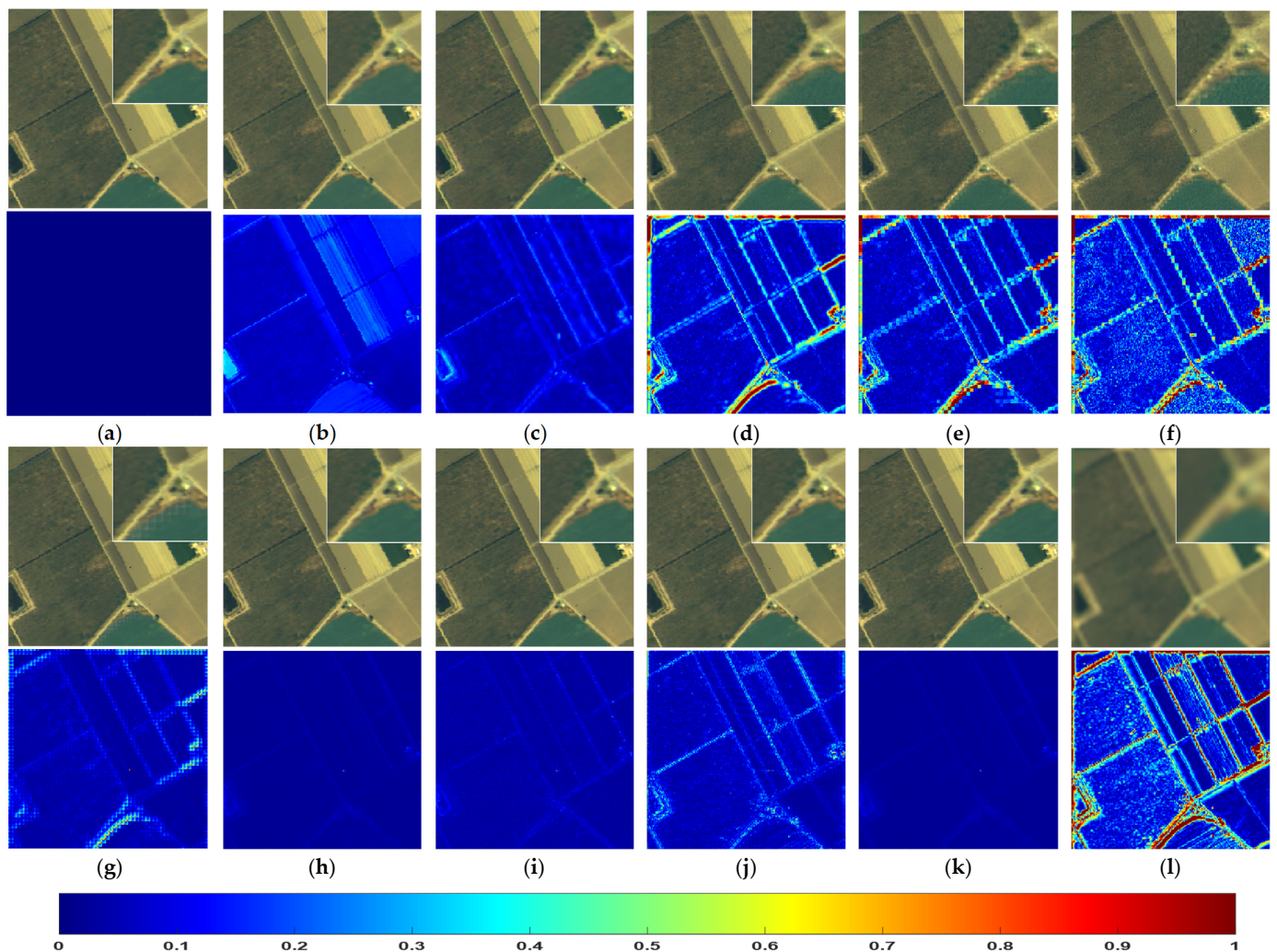
Figure 11. Fused images of different HS sharpening fusion methods on Salinas dataset: (**a**) REF; (**b**) PCA; (**c**) GSA; (**d**) GLP; (**e**) SFIM; (**f**) MAP-SMM; (**g**) HySure; (**h**) LTMR; (**i**) UTV; (**j**) IR-TenSR; (**k**) CNN-Fus; (**l**) EXP.

For the Salinas dataset, the fused images and their RMSE heat maps are presented in Figure 11. Figure 11 shows a farmland scene, mostly in flat areas. From the zoomed windows, it can be seen that the TR-based methods and deep CNN-based method, i.e., LTMR, UTV, IR-TenSR, and CNN-Fus are very similar to Figure 11a, the reference image. Moreover, Figure 11d–g exhibit apparent block distortions, especially in SFIM and MAP-SMM images. From the heat maps, it can be observed that the CNN-Fus, LTMR, and UTV methods have achieved excellent fusion results. GLP, SFIM and MAP-SMM still show poor results, especially around the roads.

Figure 12 shows the fusion results and RMSE heat maps on the Cuprite dataset. There are almost no man-made buildings or vegetations in the scene and the types of ground objects are relatively single. Despite of the low spatial resolution of the dataset, most fusion methods still achieve good fusion results, making it difficult to distinguish the differences in true color images. With the help of the RMSE heat maps, it can be seen that GSA and UTV have the smallest error, LTMR and CNN-Fus achieve suboptimal results, and GLP, SFIM, and MAPSMM methods are still the poorest. It is worth mentioning that, due to the hundreds of bands in HS sharpened images, it is difficult to distinguish spectral distortion accurately from the true color images only.
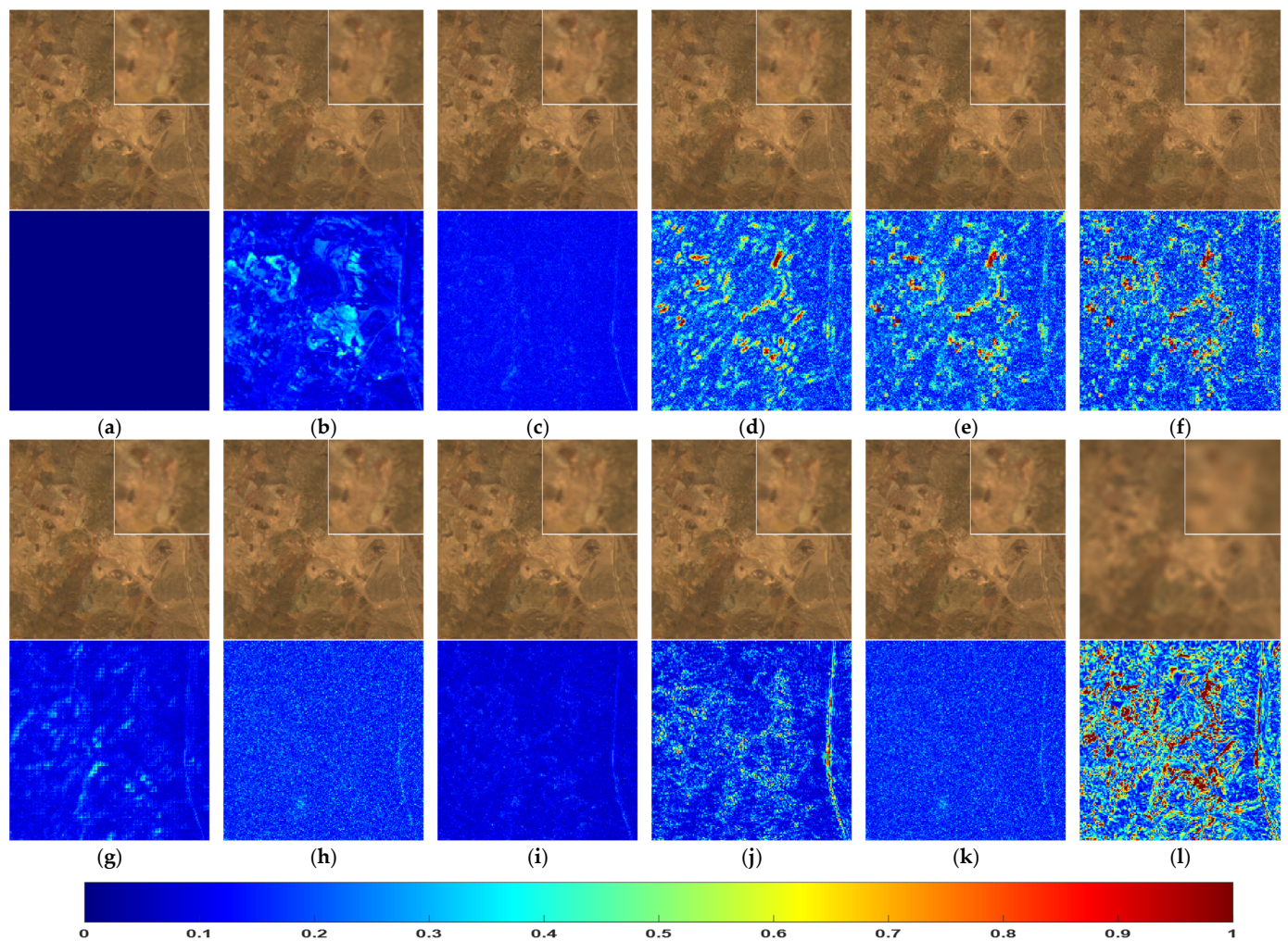
Figure 12. Fused images of different HS sharpening methods on Cuprite dataset: (**a**) REF; (**b**) PCA; (**c**) GSA; (**d**) GLP; (**e**) SFIM; (**f**) MAP-SMM; (**g**) HySure; (**h**) LTMR; (**i**) UTV; (**j**) IR-TenSR; (**k**) CNN-Fus; (**l**) EXP.

4.5.2. FR and NR Evaluations

The LR-HS and HR-MS images (mentioned in Section 4.1) are fused to obtain the fused images. Then all fusion products are evaluated using four FR metrics and five NR metrics and the experimental results are presented in Table 7, Table 8, Table 9 and Figures 13–15. For FR evaluation, the original HS images are taken as the reference and four FR metrics are used to evaluate. For NR evaluation, band assignment is conducted on the fused images and the HR-MS images to get multiple groups. Each group contains one MS band and several HS bands.

Each group is evaluated using NR metrics and the final NR evaluation results are obtained by averaging the NR evaluation results of each group. Table 7, Table 8, Table 9 report the FR and NR evaluation results of the fused images on three HS datasets. For each metric, the best result is marked in bold, and the second-best result is underlined. For a more intuitive comparison, the evaluation results of each metric are displayed in a line graph. Figures 13–15 are the line graphs for different datasets, in which the best five results are marked with yellow blocks.

Table 7. The FR and NR evaluation results of the fused images on Pavia University dataset (best results are in bold and the second-best results are underlined).

	FR					NR			
	PSNR	SAM	ERGAS	Q^{2n}	Proposed	QNR	FQNR	RQNR	MQNR
PCA [17]	32.7516	4.8990	3.6755	0.9384	0.9052	0.9539	0.8868	0.9402	11.2499
GSA [18]	36.6391	3.8547	2.5563	0.9644	**0.9761**	0.9514	0.8852	0.9654	12.4637
GLP [19]	30.1203	4.8156	5.0499	0.9122	0.8026	0.9152	0.6900	0.8384	10.7750
SFIM [20]	29.6753	4.5224	5.3376	0.9057	0.7962	0.9094	0.5469	0.8228	<u>9.6054</u>
MAP-SMM [24]	28.4848	5.4159	6.0371	0.8685	0.6682	0.8739	0.4478	0.7805	11.5903
HySure [25]	34.5272	4.9741	3.1551	0.9609	0.9609	<u>0.9543</u>	0.8813	0.9652	13.6281
LTMR [29]	**41.5977**	**3.3518**	**1.9020**	**0.9794**	<u>0.9756</u>	**0.9590**	**0.9527**	<u>0.9980</u>	14.6542
UTV [31]	37.1334	4.7330	2.8079	0.9553	0.9213	0.9007	0.8468	0.9908	16.5310
IR-TenSR [32]	<u>40.4947</u>	<u>3.4947</u>	<u>2.0962</u>	<u>0.9764</u>	0.9708	0.9335	<u>0.9293</u>	**0.9981**	17.2970
CNN-Fus [33]	38.2684	5.3699	2.9606	0.9392	0.9346	0.8615	0.9014	0.9980	29.7144
EXP	24.5021	7.4930	9.5317	0.6217	0.4117	0.6290	0.1470	0.4851	**8.9306**

Table 8. The FR and NR evaluation results of the fused images on Salinas dataset (best results are in bold and the second-best results are underlined).

	FR					NR			
	PSNR	SAM	ERGAS	Q^{2n}	Proposed	QNR	FQNR	RQNR	MQNR
PCA [17]	41.0053	1.1651	1.4423	0.9561	0.9704	0.7855	0.8986	0.9359	5.6958
GSA [18]	42.0334	0.9958	1.4939	0.9513	0.9724	0.7284	0.8831	0.9457	6.0773
GLP [19]	33.8706	1.6669	2.2878	0.8407	0.8921	0.7577	0.6338	0.8855	4.4391
SFIM [20]	33.5175	1.5912	2.3925	0.8478	0.8725	0.7456	0.5518	0.8856	**4.0835**
MAP-SMM [24]	32.2169	1.9270	2.7237	0.7988	0.8030	0.7662	0.4740	0.8559	5.0511
HySure [25]	39.8855	0.9986	1.6597	0.9301	0.9726	0.7885	0.8687	0.9310	5.1651
LTMR [29]	<u>50.3966</u>	<u>0.4058</u>	**1.3581**	<u>0.9738</u>	<u>0.9817</u>	<u>0.7889</u>	**0.9406**	<u>0.9685</u>	5.7486
UTV [31]	45.7645	0.5914	1.3272	0.9570	0.9487	**0.8037**	0.9279	0.9660	4.9139
IR-TenSR [32]	42.5068	1.1308	1.7279	0.9030	0.9160	0.4030	0.8712	**0.9745**	5.4985
CNN-Fus [33]	**50.4034**	**0.4055**	<u>1.3569</u>	**0.9739**	**0.9820**	0.7888	<u>0.9402</u>	0.9683	5.7423
EXP	29.5208	2.2622	3.5258	0.6552	0.6636	0.7062	0.2314	0.7413	<u>4.3056</u>

Table 9. The FR and NR evaluation results of the fused images on Cuprite dataset (best results are in bold and the second-best results are underlined).

	FR					NR			
	PSNR	SAM	ERGAS	Q^{2n}	Proposed	QNR	FQNR	RQNR	MQNR
PCA [17]	39.5539	0.6362	0.4708	0.9752	0.8605	0.8484	**0.9409**	0.9643	<u>3.5417</u>
GSA [18]	45.2870	0.6361	0.3104	0.9867	0.8759	0.6106	0.8305	0.9714	4.0883
GLP [19]	35.0414	0.6187	0.7746	0.9255	0.6276	0.7136	0.5303	0.8599	5.1500
SFIM [20]	34.8900	0.6224	0.7880	0.9244	0.6071	0.7172	0.4070	0.8604	5.0715
MAP-SMM [24]	33.6162	0.6049	0.9124	0.8910	0.5451	0.8188	0.3665	0.8395	4.8478
HySure [25]	42.2074	0.6558	0.3708	0.9833	0.8428	**0.9299**	0.8761	0.9745	4.4016
LTMR [29]	**48.5105**	**0.5219**	**0.2667**	**0.9898**	**0.9401**	0.3162	0.8196	0.9794	7.4048
UTV [31]	42.6072	0.6022	0.3548	0.9794	0.807	<u>0.9167</u>	<u>0.9207</u>	<u>0.9876</u>	**3.3382**
IR-TenSR [32]	41.3697	0.9291	0.5971	0.9496	0.8385	0.2819	0.8205	**0.9985**	7.4925
CNN-Fus [33]	<u>48.5053</u>	<u>0.5220</u>	<u>0.2668</u>	<u>0.9897</u>	<u>0.9400</u>	0.3150	0.8199	0.9796	7.4009
EXP	29.2470	1.0039	1.5053	0.6427	0.3065	0.5774	0.0893	0.6508	5.0503
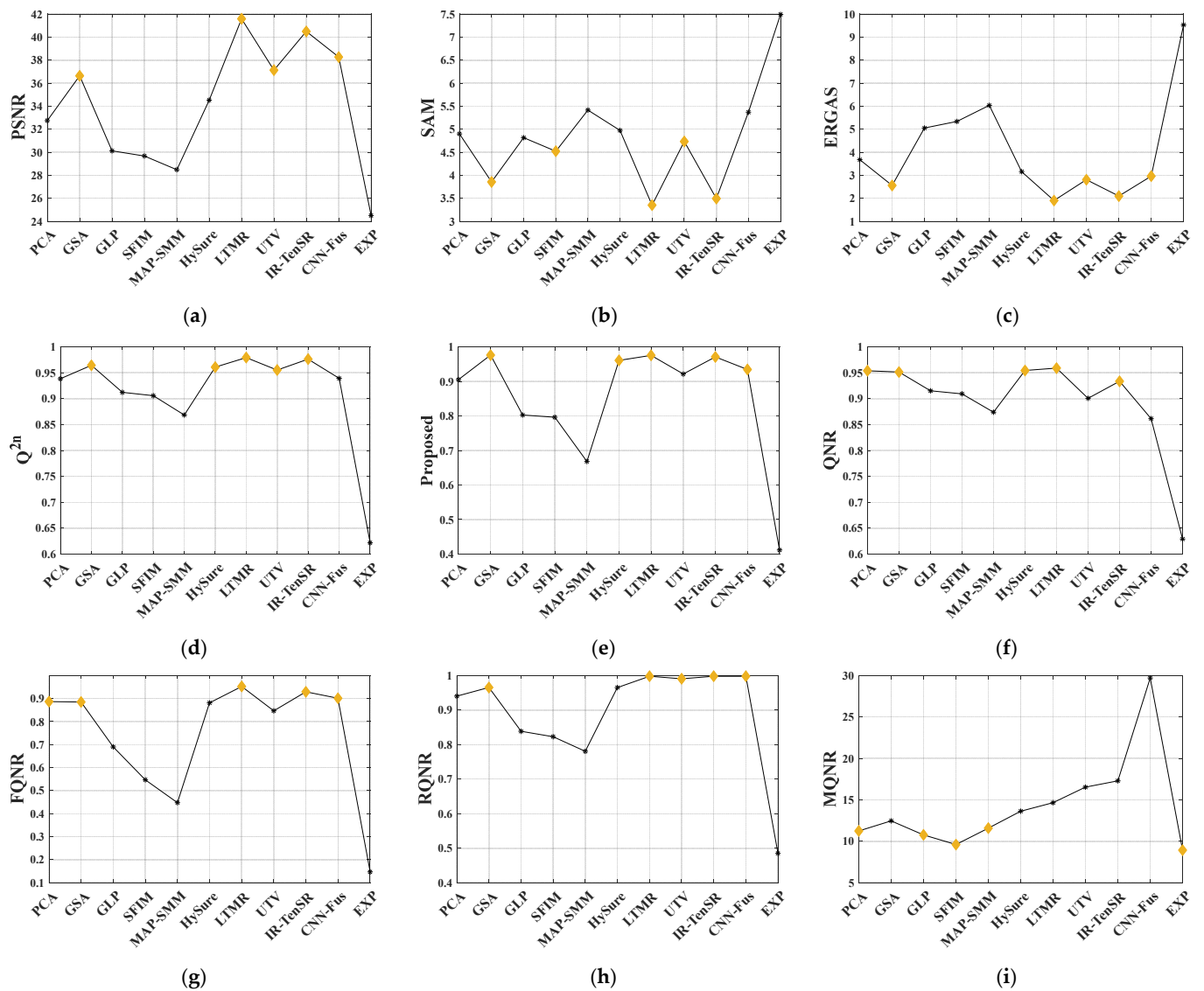
Figure 13. The evaluation values of FR and NR metrics on Pavia University dataset: (**a**) PSNR; (**b**) SAM; (**c**) ERGAS; (**d**) Q^{2n}; (**e**) Proposed; (**f**) QNR; (**g**) FQNR; (**h**) RQNR; (**i**) MQNR.

From the results in Table 7, it can be seen that LTMR and IR-TenSR achieve the best or second-best results among FR metrics, and EXP has the worst results in terms of FR metrics. For Tables 8 and 9, CNN-Fus and LTMR are always the best or second-best among FR metrics while EXP is the worst. This phenomenon is basically consistent with the subjective analysis results in Figures 10–12. From Figures 13–15, it can be seen that there are some correlations among the four curves of FR metrics, such as the optimal evaluation results among them basically including GSA, LTMR, UTV, and CNN-Fus. In the performance of NR metrics, the trend of the proposed method is very similar to Q^{2n} and ERGAS, indicating that the proposed method has shown excellent performance in evaluating overall quality. FQNR and RQNR also have a similar trend to the Q^{2n} curves, but their evaluations are not accurate in methods with good fusion results. For example, in the evaluations of LTMR, UTV, IR-TenSR, and CNN-Fus, the yellow blocks in RQNR are very close. For QNR, it has good evaluation ability on the Pavia University dataset, but its evaluation performance significantly decreases on the other two datasets, as its curves are clearly unrelated. The curves of MQNR are inconsistent with those of FR metric, which means it cannot be used for accurate evaluation. In addition, the evaluation results of the NR metrics have a certain

correlation with PSNR and SAM, but the correlations are not stable enough, especially on the Cuprite dataset.
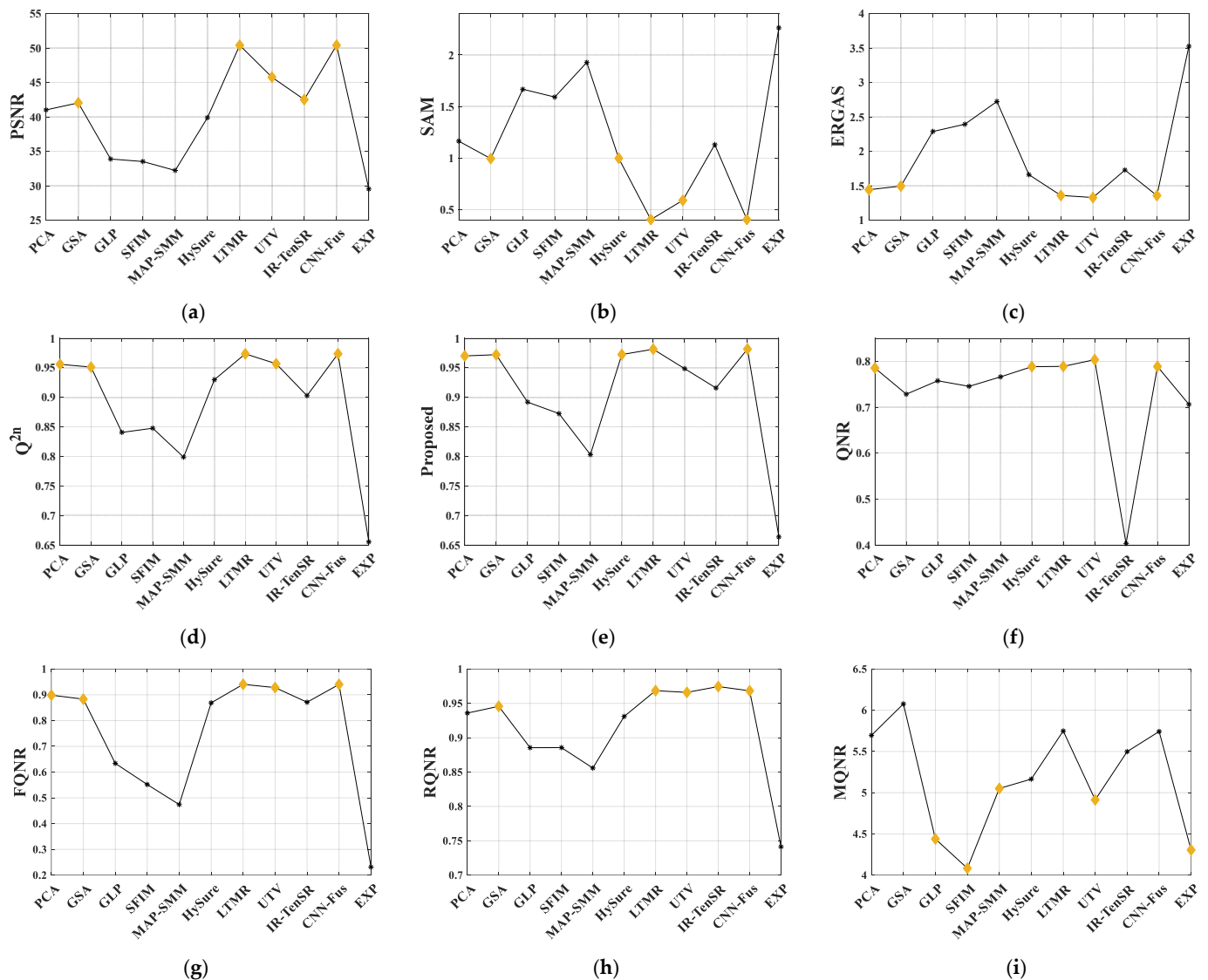
Figure 14. The evaluation values of FR and NR metrics on Salinas dataset: (**a**) PSNR; (**b**) SAM; (**c**) ERGAS; (**d**) Q^{2n}; (**e**) Proposed; (**f**) QNR; (**g**) FQNR; (**h**) RQNR; (**i**) MQNR.

4.5.3. The Consistency between FR and NR Metrics

FR quality assessment is commonly considered more reliable than NR evaluation, since the reference is available and used in FR. Therefore, it is reasonable to use the evaluation conclusion of FR as a reliability and accuracy benchmark for validating NR metrics. In order to verify the consistency between FR and NR metrics and analyze the differences, consistency experiments are conducted in which the evaluation results of Q^{2n}, SAM, and PSNR are considered as representatives of overall quality, spectral quality, and spatial quality, respectively. Then the PLCC, SROCC, and KROCC are calculated between all NR metrics and Q^{2n}, SAM, and PSNR separately. Tables 10–12 report the correlations on three datasets. The higher the values, the more consistent with the FR evaluation. The negative values indicate that the evaluation results of the two metrics have opposite trend. For a more intuitive analysis, the correlations are plotted in the form of a bar graph in Figures 16–18, in which the negative values are set to 0.
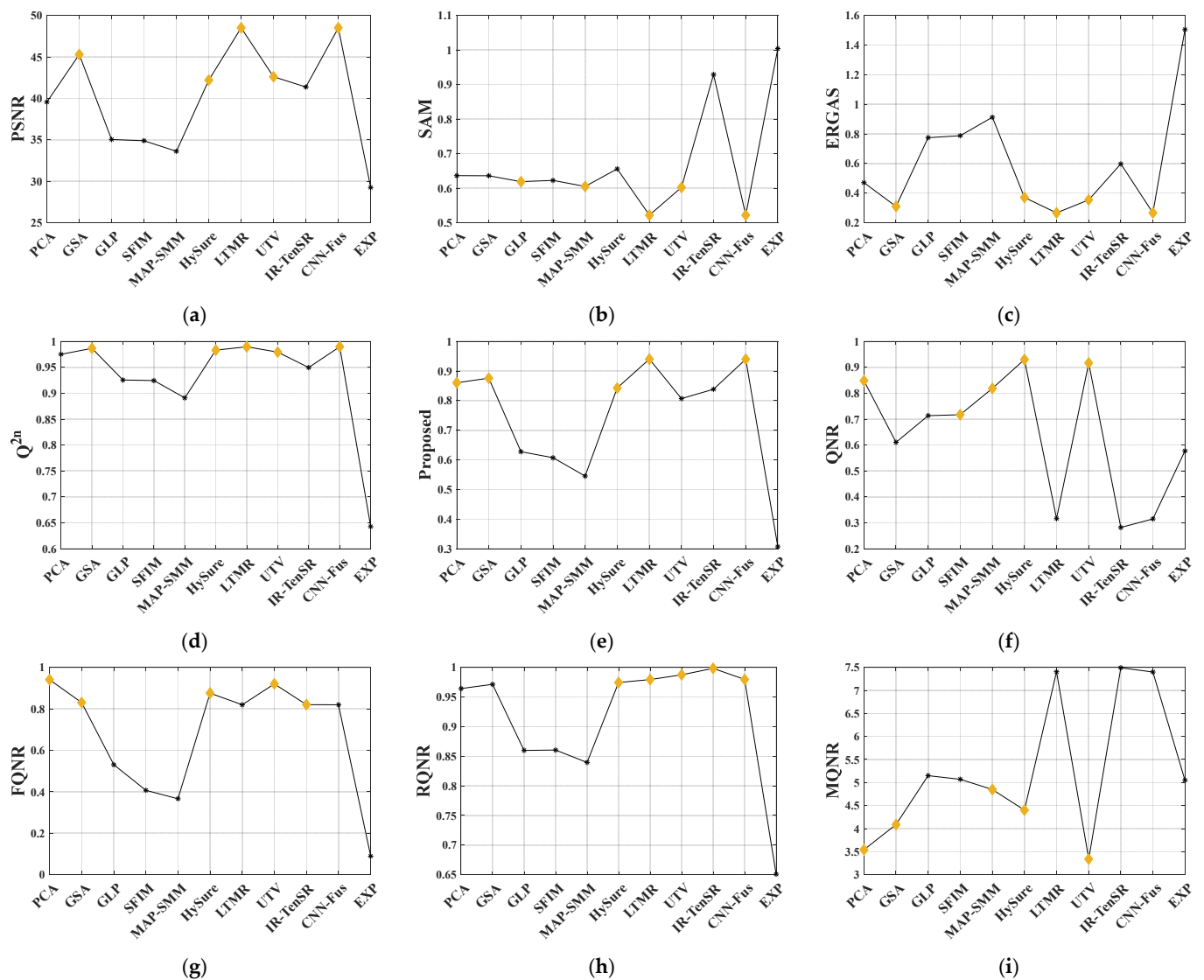
Figure 15. The evaluation values of FR and NR metrics on Cuprite dataset: (**a**) PSNR; (**b**) SAM; (**c**) ERGAS; (**d**) Q^{2n}; (**e**) Proposed; (**f**) QNR; (**g**) FQNR; (**h**) RQNR; (**i**) MQNR.

Table 10. PLCC, SROCC, and KROCC between Q^{2n} and all NR metrics on three HS datasets (best results are in bold and the second-best results are underlined).
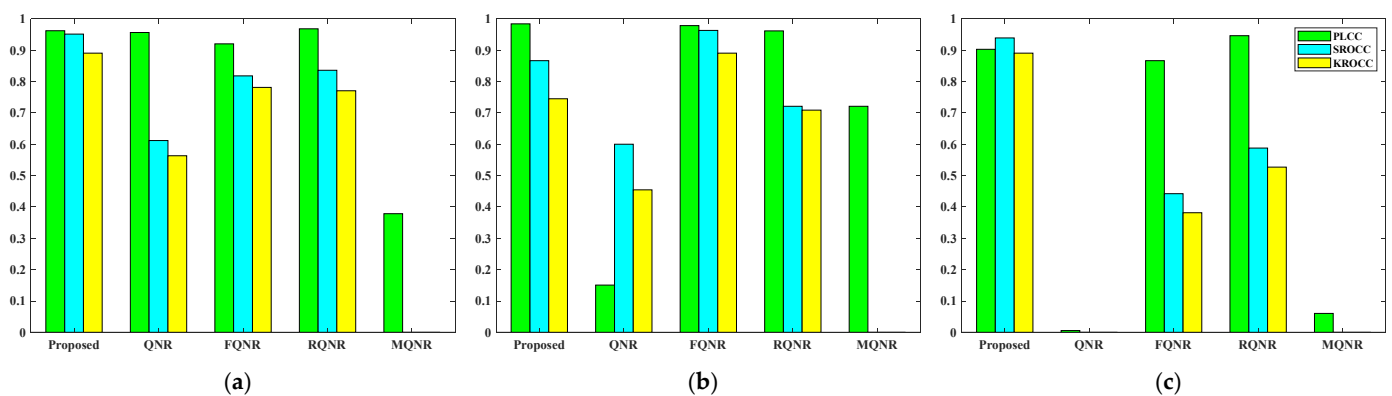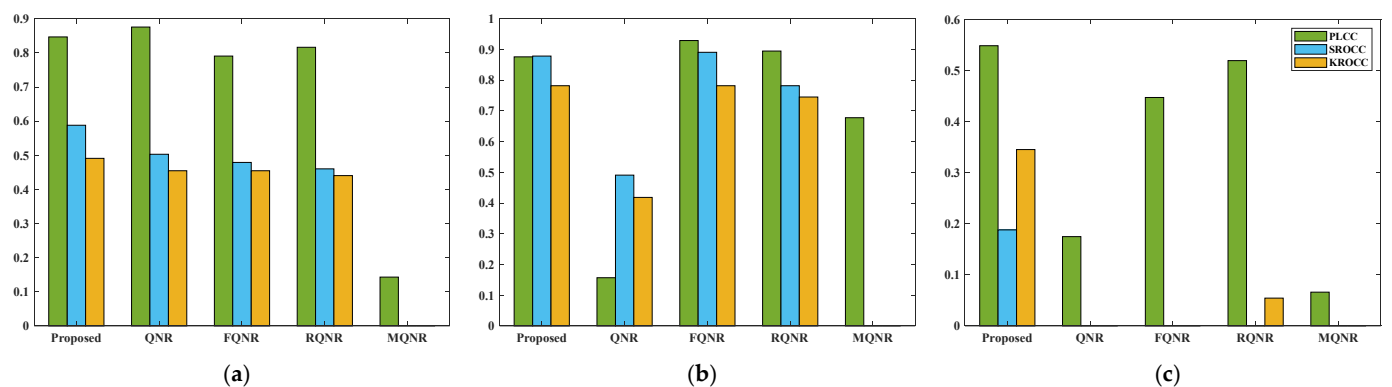
	Pavia University			Salinas			Cuprite		
	PLCC	**SROCC**	**KROCC**	**PLCC**	**SROCC**	**KROCC**	**PLCC**	**SROCC**	**KROCC**
Proposed	<u>0.9622</u>	**0.9515**	**0.8909**	**0.9846**	<u>0.8667</u>	<u>0.7455</u>	<u>0.9031</u>	**0.9394**	**0.8909**
QNR	0.9569	0.6121	0.5636	0.1506	0.6000	0.4545	0.0054	−0.5273	−0.0545
FQNR	0.9204	0.8182	<u>0.7818</u>	<u>0.9787</u>	**0.9636**	**0.8909**	0.8671	0.4424	0.3818
RQNR	**0.9685**	<u>0.8363</u>	0.7707	0.9618	0.7212	0.7091	**0.9467**	<u>0.5879</u>	<u>0.5273</u>
MQNR	0.3784	−0.1303	0.5273	0.7211	−1.2121	0.4909	0.0604	−0.4061	0.0909

Table 11. PLCC, SROCC, and KROCC between SAM and all NR metrics on three HS datasets (best results are in bold and the second-best results are underlined).
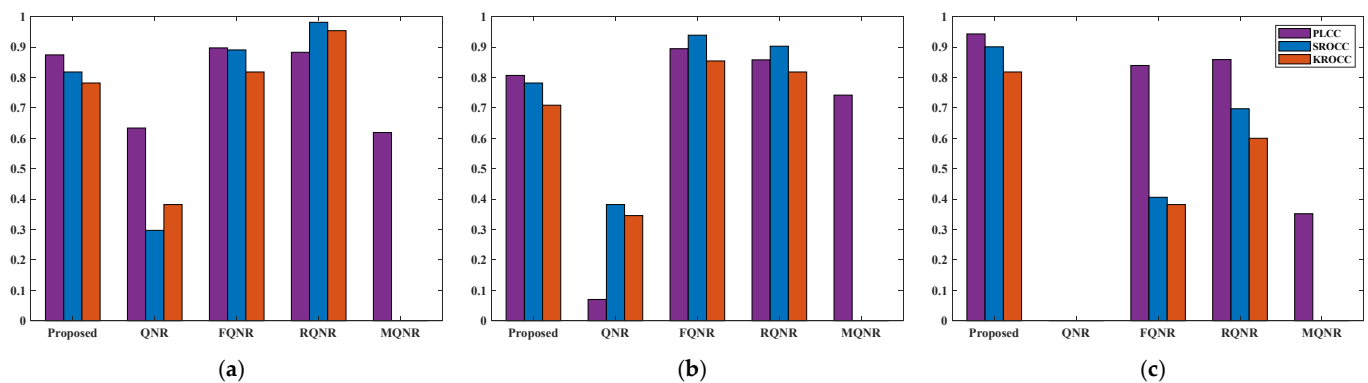
	Pavia University			Salinas			Cuprite		
	PLCC	**SROCC**	**KROCC**	**PLCC**	**SROCC**	**KROCC**	**PLCC**	**SROCC**	**KROCC**
Proposed	<u>0.8466</u>	**0.5879**	**0.4909**	0.8760	<u>0.8788</u>	**0.7818**	**0.5489**	**0.1879**	**0.3455**
QNR	**0.8761**	<u>0.5030</u>	<u>0.4545</u>	0.1570	0.4910	0.4180	0.1748	−0.3939	−0.0182
FQNR	0.7908	0.4788	<u>0.4545</u>	**0.9292**	**0.8909**	**0.7818**	0.4474	−0.3697	−0.0182
RQNR	0.8165	0.4606	0.4404	<u>0.8948</u>	0.7818	<u>0.7455</u>	<u>0.5197</u>	<u>−0.0060</u>	<u>0.0545</u>
MQNR	0.1430	−0.7576	−0.2000	0.6778	−1.2485	−0.4771	0.0662	−0.5030	−0.1273

Table 12. PLCC, SROCC and KROCC between PSNR and all NR metrics on three HS datasets (best results are in bold and the second-best results are underlined).

	Pavia University			Salinas			Cuprite		
	PLCC	**SROCC**	**KROCC**	**PLCC**	**SROCC**	**KROCC**	**PLCC**	**SROCC**	**KROCC**
Proposed	0.8745	0.8182	0.7818	0.8071	0.7818	0.7091	**0.9435**	**0.8909**	**0.8182**
QNR	0.6336	0.2970	0.3818	0.0701	0.0382	0.0346	−0.4017	−0.6364	−0.1273
FQNR	**0.8972**	<u>0.8909</u>	<u>0.8182</u>	**0.8949**	**0.9394**	**0.8545**	0.8399	0.4061	0.3818
RQNR	<u>0.8833</u>	**0.9818**	**0.9542**	<u>0.8581</u>	<u>0.9030</u>	<u>0.8182</u>	<u>0.8591</u>	<u>0.6970</u>	<u>0.6000</u>
MQNR	0.6192	−1.4970	−0.7091	0.7419	−1.2485	−0.5505	0.3517	−0.4788	−0.0909

Figure 16. PLCC, SROCC, and KROCC between all NR metrics and Q^{2n} on three HS datasets: (**a**) Pavia University dataset; (**b**) Salinas dataset; (**c**) Cuprite dataset.

Figure 17. PLCC, SROCC, and KROCC between all NR metrics and SAM on three HS datasets: (**a**) Pavia University dataset; (**b**) Salinas dataset; (**c**) Cuprite dataset.

Figure 18. PLCC, SROCC, and KROCC between all NR metrics and PSNR on three HS datasets: (**a**) Pavia University dataset; (**b**) Salinas dataset; (**c**) Cuprite dataset.

Table 10 and Figure 16 show the correlations between NR metrics and Q^{2n} for evaluating the overall quality. It can be seen that the proposed method has the highest or second-highest correlations with the evaluation results of Q^{2n} on all three datasets. On the Cuprite dataset, there is a significant decrease in the accuracy of the other NR metrics, while the proposed method still maintains excellent evaluation performance. For the FQNR and RQNR metrics, they can only perform the good evaluation performance on a single dataset and show unstable evaluation performance and weak robustness. For the QNR and MQNR metrics, their evaluation results show a weaker correlation or even negative values compared to Q^{2n}, which indicates the two NR metrics are not reliable to evaluate the overall quality.

For the spectral quality evaluation, Table 11 and Figure 17 show the correlations between NR metrics and SAM metric. In Table 11, it can be seen that the proposed method stays in a stable first tier on all datasets and has good robustness, especially on the Cuprite dataset. In contrast, the negative values appear in the other NR metrics, which indicates the metrics cannot accurately evaluate the spectral quality of the Cuprite dataset. Although RQNR achieves the second-best in terms of KROCC, its numerical value (0.0545) is much smaller than the proposed method (0.345). On the Salinas dataset, FQNR is the most relevant to SAM in terms of PLCC and SRCC. However, FQNR performs poorly on the other two datasets. This is similar to its situation in Table 10. QNR shows good correlation on the Pavia University dataset but is unstable on the other datasets.

For the spatial distortion evaluation, Table 12 and Figure 18 show the correlations between NR metrics and PSNR. FQNR and RQNR achieve the highest correlation with PSNR on datasets Pavia University and Salinas. The proposed method is closer to PSNR than other metrics on the Cuprite dataset. The reason is that Benford's law is based on the statistical characteristics of data, making it more sensitive to large-scale spatial features (i.e., lower spatial resolution). The Cuprite dataset has a spatial resolution of 20 m, which is much lower than the other datasets. Therefore, FQNR and RQNR metrics designed specifically for high spatial resolution (i.e., the multispectral sharpening) are more effective in perceiving small-scale spatial features. Additionally, it can be observed in Table 12 that the values of the proposed method are close to FQNR and RQNR on the Pavia University and Salinas datasets and exhibit stability.

From the consistency experiment results, the following conclusions can be drawn: (1) Compared with the other NR metrics, the evaluation results of the proposed method are generally closer to those of FR on all datasets. (2) The proposed method achieves the best results on the Cuprite dataset. (3) Compared with the other NR metrics, the proposed method is more stable and robust for different datasets. (4) QNR and MQNR metrics differ significantly from FR, making the quality evaluation of HS sharpened images unreliable.

Overall, the proposed method exhibits optimal performance in evaluating global and spectral quality and has high robustness. The FQNR and RQNR can only exhibit optimal

performance on one to two datasets and have poor robustness. QNR and MQNR are not suitable for NR quality assessment of HS sharpened images.

4.5.4. Analysis of the Sub-Index in NR Metrics

In order to investigate the performance of FDD^{lf}, FDD^{hf}, and FDD^Q, the sub-indexes of all NR metrics are analyzed and compared by calculating the SROCC between the spectral/spatial distortion index and all FR metrics. SROCC is often used to reflect the effectiveness of the spatial/spectral indexes. The higher the SROCC, the more accurate the ranking of the evaluation results of the sub-index, and the more effective the sub-index is. Tables 13–15 report the SROCC results between all spatial/spectral indexes and all FR metrics. The rows on top represent the spatial indexes of the NR metrics, the rows in the middle represent the spectral indexes of the NR metrics, and the bottom seven rows belong to the sub-indexes proposed in this paper. PSNR reflects the ability to measure the spatial distortion, SAM presents the ability to measure the angular distortion of the spectrum, and ERGAS and Q^{2n} reflect the performance of measuring the overall quality.

Table 13. SROCC between sub-indexes and all FR metrics on Pavia University dataset (best results are in bold and the second-best results are underlined).

	Sub-Index	PSNR	SAM	ERGAS	Q^{2n}
spatial	D_{Qs}	0.8909	**0.6727**	**0.9636**	**1.0000**
	D_{Rs}	**0.9515**	0.4545	0.9030	0.8788
	D_{Fs}	0.7212	0.5273	0.7697	0.8061
	D_{Ms}	0.4909	0.1273	0.5152	0.6364
spectral	$D_{Q\lambda}$	−1.5758	−0.8970	−1.4788	−1.3697
	$D_{F\lambda}$	**0.9576**	**0.4242**	**0.8606**	**0.8121**
	$D_{M\lambda}$	−1.4424	−0.8364	−1.3212	−1.1394
	FDD^{lf}	0.7333	0.0909	0.6121	0.6606
	FDD^{hf}	0.5394	**0.6970**	0.6606	0.7939
	FDD^Q	0.8061	0.6364	0.9273	**0.9636**
	$FDD^{lf} + FDD^{hf}$	0.7758	0.5333	0.8242	0.9091
	$FDD^{lf} + FDD^Q$	0.8061	0.6364	0.9273	**0.9636**
	$FDD^{hf} + FDD^Q$	0.8061	0.6364	0.9273	**0.9636**
	Proposed	**0.8182**	0.5879	**0.9515**	0.9515

Table 14. SROCC between sub-indexes and all FR metrics on Salinas dataset (best results are in bold and the second-best results are underlined).

	Sub-Index	PSNR	SAM	ERGAS	Q^{2n}
spatial	D_{Qs}	0.6788	0.6061	0.7758	0.7152
	D_{Rs}	**0.8727**	0.7273	0.6061	0.6909
	D_{Fs}	**0.8727**	**0.8606**	**0.9455**	**0.9697**
	D_{Ms}	0.6485	0.6485	0.7697	0.7939
spectral	$D_{Q\lambda}$	−1.0424	−1.0546	−0.8606	−0.9455
	$D_{F\lambda}$	**0.9212**	**0.8000**	**0.6788**	**0.7394**
	$D_{M\lambda}$	−1.2121	−1.1152	−1.1394	−1.0909
	FDD^{lf}	0.4545	0.6121	0.5152	0.6485
	FDD^{hf}	0.7455	0.7818	**0.7939**	**0.8788**
	FDD^Q	0.6485	0.7939	0.6849	0.7576
	$FDD^{lf} + FDD^{hf}$	0.7455	0.7818	**0.7939**	**0.8788**
	$FDD^{lf} + FDD^Q$	0.7212	0.8424	0.7333	0.8182
	$FDD^{hf} + FDD^Q$	**0.7818**	**0.8788**	0.7697	0.8667
	Proposed	**0.7818**	**0.8788**	0.7818	0.8667

Table 15. SROCC between sub-indexes and all FR metrics on Cuprite dataset (best results are in bold and the second-best results are underlined).

	Sub-Index	PSNR	SAM	ERGAS	Q^{2n}
spatial	D_{Qs}	<u>0.4364</u>	<u>−0.3394</u>	<u>0.4727</u>	<u>0.4606</u>
	D_{Rs}	<u>0.4364</u>	<u>−0.3394</u>	0.4242	0.4121
	D_{Fs}	**0.5273**	**−0.1515**	**0.6000**	**0.5879**
	D_{Ms}	0.3455	−0.4909	0.3818	0.3930
spectral	$D_{Q\lambda}$	−0.9818	−0.6182	−0.8970	−0.8727
	$D_{F\lambda}$	**0.7273**	**0.0606**	**0.6546**	**0.6061**
	$D_{M\lambda}$	<u>−0.8909</u>	<u>−0.6121</u>	<u>−0.7818</u>	<u>−0.8546</u>
	FDD^{lf}	0.4304	−0.5880	0.4182	0.4545
	FDD^{hf}	0.7333	0.1030	0.8061	0.8182
	FDD^{Q}	<u>0.9152</u>	**0.2242**	0.8667	0.8546
	$FDD^{lf} + FDD^{hf}$	0.7333	0.1030	0.8061	0.8182
	$FDD^{lf} + FDD^{Q}$	0.9030	<u>0.2121</u>	0.8546	0.8424
	$FDD^{hf} + FDD^{Q}$	**0.9515**	0.2000	**0.9273**	**0.9515**
	Proposed	0.8909	0.1879	<u>0.9152</u>	<u>0.9394</u>

For the spatial indexes, it can be seen that D_{Rs} achieves the highest SROCC with PSNR on the first and second dataset, while D_{Fs} obtains optimal performance on the second and third dataset. The SROCC of D_{Ms} are the lowest among all spatial indexes. Moreover, the NR spatial indexes also demonstrate high accuracy in measuring overall distortion, such as the performance of D_{Qs} on the first dataset and D_{Fs} on the second dataset, although their performances are not stable on all datasets. In addition, these NR indexes are not accurate in measuring the spectral distortion, since they show negative SROCC values on the Cuprite dataset.

For the spectral indexes, $D_{F\lambda}$ achieves the best performance among the three NR spectral indexes, and it also demonstrates excellent performance in measuring the spatial and overall distortion on all three datasets. All SROCCs of $D_{Q\lambda}$ and $D_{M\lambda}$ are negative, indicating that they fail to measure fusion distortion. The reason may be that the assumption of invariance in the relationship between HS bands is unreasonable. For example, the spectral distortion of EXP is considered smallest in pansharpening and the results of both SAM and $D_{Q\lambda}$ should be close to 0. In HS sharpening, however, the SAM values of EXP are always the worst and the results of $D_{Q\lambda}$ are still close to 0. The completely opposite evaluation results of SAM and $D_{Q\lambda}$ lead to poor performance of $D_{Q\lambda}$. For $D_{M\lambda}$, as it inherits the assumption of QNR's spectral distortion index, the results of $D_{M\lambda}$ are also unreliable.

For the sub-indexes proposed, each of the FDD features and their pairwise combinations are compared separately with FR metrics. It can be seen that the best evaluation performance is not achieved by using all three FDD features. The performance of different types of FDD features varies on different datasets. FDD^{hf} and FDD^{Q} both have stable evaluation performance, and combining them can obtain more accurate and stable evaluation results. In contrast, the ability of FDD^{lf} to measure various distortions is weak since FDD^{lf} only characterize the statistics of the fused image itself, without considering the relationship between the fused image and the input images (HR-MS and LR-HS). Due to the excellent performance of FDD^{lf} in measuring external distortion, it is retained in Q_{FDD} to enhance the robustness of the proposed method. Overall, from the data in the tables, it can be seen that the sub-indexes of the proposed method exhibit high accuracy and excellent robustness in measuring various distortions, especially global distortion. The proposed method is effective to measure the overall distortion of HS sharpened image.

5. Discussion

5.1. The Distortions in HS Sharpening

For HS sharpening (or MS sharpening), the input images have already been denoised and registered. In other words, the input images have no noises like stripe noise, Gaussian white noise, salt and pepper noise, etc. Similar to the authentic distortions of natural images, the distortions in HS sharpening are inherent in the fused image itself and some external distortions are not considered. Therefore, the quality assessment should pay more attention to the distortions yielded in HS sharpening process. To date, the fusion community consensus is that there exist two types of distortions, i.e., the spatial distortion and the spectral distortion. However, due to the large number of spectral bands in HS sharpened images, it is difficult to separate the spectral distortion and spatial distortion completely. In the experimental results, it can be seen that there exists, to some extent, a correlation between the evaluation results of spectral and spatial indexes. There is coupling between the two distortions. It is more reasonable to describe the overall quality of fusion. To define what the term (spatial/spectral) quality means and quantify the quality in a measurable way make the quality assessment of HS sharpening a complex task.

5.2. The Accuracy of NR Quality Assessment

For the quality assessment of HS sharpened images, the qualitative evaluation is infeasible. The subjective evaluation of HS sharpened bands within the visible spectrum has limitation. Therefore, it is very difficult to verify the accuracy of NR metrics using DMOS (difference mean opinion score) values obtained from subjective evaluation. Due to the rarity of the real reference images, NR quality assessment of the fused images is much debated, especially for NR evaluation of HS sharpened images at full resolution scale. If the ranking of different fusion results in an NR evaluation is consistent with the ranking based on full-reference evaluation, then the NR method is considered effective and reliable. This assumption is a feasible way in current situation. In this paper, experiments are conducted on the consistency between NR and FR metrics to test the accuracy of NR metrics. The experimental results show that the proposed method is generally closer to the evaluation results of FR compared to other NR metrics. However, for the datasets with high spatial resolution (i.e., Pavia University and Salinas), the proposed method does not have significant advantages. FQNR and RQNR metrics have good accuracy in measuring overall and spatial quality on the Pavia University and Salinas datasets but have poor performance on the Cuprite dataset with low spatial resolution, which indicates the metrics are sensitive to the spatial resolution of HS sharpened images and have weak robustness. Compared to FR, QNR and MQNR show opposite evaluation results, so they are not suitable for evaluating HS sharpening.

5.3. The Limitations

Although the proposed method is tested by using three different HS datasets, there still may be potential limitations: (1) The input images (i.e., HR-MS and LR-HS images) used when generating fusion results are simulated from the HS dataset itself. The effectiveness of the proposed method is uncertain about heterogeneous or multi-platform HS data. (2) The experimental results show that the proposed method has limitations in detecting the spatial distortion of the HS sharpened images with high spatial resolution. We believe that there is a potential connection between the spatial information of image and Benford's law. The implicit relationship is worth exploring and will help design a more accurate metric.

6. Conclusions

This paper proposes a no-reference quality assessment method of HS sharpened images without the need for a reference image or any other simulations based on Wald's protocol. The method designs three quality perception features, namely FDD^{lf}, FDD^{hf}, and FDD^{Q}, to perceive the quality and employs the standard Benford's law as a benchmark to evaluate the fusion distortion of HS sharpened images. The extensive experiments

are conducted to verify the effectiveness of the proposed method. Compared with the other four commonly used NR metrics, the proposed method is more stable and robust for different HS datasets, although the performance of the FDD features varies on different datasets. In addition, the evaluation results of the proposed method are generally closer to those of FR on all HS datasets. The proposed method can be flexibly used for evaluating spectral, spatial, and overall distortion without relying on the type of sensors and scale of the dataset. To obtain a more accurate Q_{FDD} by applying the weights to the three FDD features should be the subject of future work.

References

1. Bai, L.; Dai, J.; Song, Y.; Liu, Z.; Chen, W.; Wang, N.; Wu, C. Predictive prospecting using remote sensing in a mountainous terrestrial volcanic area, in Western Bangongco–Nujiang Mineralization Belt, Tibet. *Remote Sens.* **2023**, *15*, 4851. [CrossRef]
2. Zhou, Q.; Wang, S.; Guan, K. Advancing airborne hyperspectral data processing and applications for sustainable agriculture using RTM-based machine learning. In Proceedings of the IGARSS 2023–2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 16–21 July 2023.
3. Li, H.; Zhou, B.; Xu, F. Variation analysis of spectral characteristics of reclamation vegetation in a rare earth mining area under environmental stress. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 4408412. [CrossRef]
4. Zhang, Y.; Li, X.; Wei, B.; Li, L.; Yue, S. A Fast Hyperspectral Tracking Method via Channel Selection. *Remote Sens.* **2023**, *15*, 1557. [CrossRef]
5. Xie, M.; Gu, M.; Zhang, C.; Hu, Y.; Yang, T.; Huang, P.; Li, H. Comparative study of the atmospheric gas composition detection capabilities of FY-3D/HIRAS-I and FY-3E/HIRAS-II based on information capacity. *Remote Sens.* **2023**, *15*, 4096. [CrossRef]
6. Acito, N. Improved learning-based approach for atmospheric compensation of VNIR-SWIR hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5512715. [CrossRef]
7. Wright, L.A.; Kindel, B.C.; Pilewskie, P.; Leisso, N.P.; Kampe, T.U.; Schmidt, K.S. Below-cloud atmospheric correction of airborne hyperspectral imagery using simultaneous solar spectral irradiance observations. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 1392–1409. [CrossRef]
8. Li, Z.; Zhang, Y. Hyperspectral anomaly detection via image super-resolution processing and spatial correlation. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 2307–2320. [CrossRef]
9. Liu, J.; Guan, R.; Li, Z.; Zhang, J.; Hu, Y.; Wang, X. Adaptive multi-feature fusion graph convolutional network for hyperspectral image classification. *Remote Sens.* **2023**, *15*, 5438. [CrossRef]
10. Huang, Y.; Zhang, L.; Qi, W.; Huang, C.; Song, R. Contrastive self-supervised two-domain residual attention network with random augmentation pool for hyperspectral change detection. *Remote Sens.* **2023**, *15*, 3739. [CrossRef]
11. Middleton, E.M.; Ungar, S.G.; Mandl, D.J.; Ong, L.; Frye, S.W.; Campbell, P.E.; Landis, D.R.; Young, J.P.; Pollack, N.H. The earth observing one (EO-1) satellite mission: Over a decade in space. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 243–256. [CrossRef]
12. Green, R.O.; Schaepman, M.E.; Mouroulis, P.; Geier, S.; Shaw, L.; Hueini, A.; Bernas, M.; McKinley, I.; Smith, C.; Wehbe, R.; et al. Airborne visible/infrared imaging spectrometer 3 (AVIRIS-3). In Proceedings of the 2022 IEEE Aerospace Conference (AERO), Big Sky, MT, USA, 5–12 March 2022.
13. Yamamoto, S.; Tsuchida, S.; Urai, M.; Mizuochi, H.; Iwao, K.; Iwasaki, A. Initial analysis of spectral smile calibration of hyperspectral imager suite (HISUI) using atmospheric absorption bands. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5534215. [CrossRef]
14. Loizzo, R.; Guarini, R.; Longo, F.; Scopa, T.; Formaro, R.; Facchinetti, C.; Varacalli, G. Prisma: The italian hyperspectral mission. In Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018.

15. Restaino, R.; Vivone, G.; Addesso, P.; Chanussot, J. Hyperspectral sharpening approaches using satellite multiplatform data. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 578–596. [CrossRef]

16. Haydn, R.; Dalke, G.W.; Henkel, J.; Bare, J.E. Application of the IHS color transform to the processing of multisensor data and image enhancement. In Proceedings of the International Symposium on Remote Sensing of Environment, First Thematic Conference: Remote Sensing of Arid and Semi-Arid Lands, Cairo, Egypt, 19–25 January 1982.

17. Chavez, P.S.; Sides, S.C., Jr.; Anderson, J.A. Comparison of three different methods to merge multiresolution and multispectral data: Landsat TM and SPOT panchromatic. *Photogramm. Eng. Remote Sens.* **1991**, *57*, 295–303.

18. Aiazzi, B.; Baronti, S.; Selva, M. Improving component substitution pansharpening through multivariate regression of MS + Pan data. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3230–3239. [CrossRef]

19. Aiazzi, B.; Alparone, L.; Baronti, S.; Garzelli, A.; Selva, M. MTF-tailored multiscale fusion of high-resolution MS and Pan imagery. *Photogramm. Eng. Remote Sens.* **2006**, *72*, 591–596. [CrossRef]

20. Liu, J.G. Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details. *Int. J. Remote Sens.* **2000**, *21*, 3461–3472. [CrossRef]

21. Kawakami, R.; Matsushita, Y.; Wright, J.; Ben-Ezra, M.; Tai, Y.-W.; Ikeuchi, K. High-resolution hyperspectral imaging via matrix factorization. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011.

22. Wei, Q.; Dobigeon, N.; Tourneret, J.-Y. Fast fusion of multiband images based on solving a sylvester equation. *IEEE Trans. Image Process.* **2015**, *24*, 4109–4121. [CrossRef]

23. Dong, W.; Fu, F.; Shi, G.; Cao, X.; Wu, J.; Li, G.; Li, X. Hyperspectral image superresolution via non-negative structured sparse representation. *IEEE Trans. Image Process.* **2016**, *25*, 2337–2352. [CrossRef]

24. Eismann, M.T. Resolution Enhancement of Hyperspectral Imagery Using Maximum a Posteriori Estimation with a Stochastic Mixing Model. Ph.D. Dissertation, Department of Electrical and Computer Engineering, University of Dayton, Dayton, OH, USA, May 2004.

25. Simões, M.; Dias, J.B.; Almeida, L.; Chanussot, J. A Convex formulation for hyperspectral image superresolution via subspace-based regularization. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 3373–3388. [CrossRef]

26. Yokoya, N.; Yairi, T.; Iwasaki, A. Coupled non-negative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 528–537. [CrossRef]

27. Li, S.; Dian, R.; Fang, L.; Bioucas-Dias, J.M. Fusing hyperspectral and multispectral images via coupled sparse tensor factorization. *IEEE Trans. Image Process.* **2018**, *27*, 4118–4130. [CrossRef]

28. Xu, Y.; Wu, Z.; Chanussot, J.; Comon, P.; Wei, Z. Nonlocal coupled tensor CP decomposition for hyperspectral and multispectral image fusion. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 348–362. [CrossRef]

29. Dian, R.; Li, S. Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization. *IEEE Trans. Image Process.* **2019**, *28*, 5135–5146. [CrossRef]

30. Dian, R.; Li, S.; Fang, L. Learning a low tensor-train rank representation for hyperspectral image super-resolution. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 2672–2683. [CrossRef]

31. Xu, T.; Huang, T.; Deng, L.; Zhao, X.; Huang, J. Hyperspectral image superresolution using unidirectional total variation with Tucker decomposition. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 4381–4398. [CrossRef]

32. Xu, T.; Huang, T.; Deng, L.; Yokoya, N. An iterative regularization method based on tensor subspace representation for hyperspectral image super-resolution. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5529316. [CrossRef]

33. Dian, R.; Li, S.; Kang, X. Regularizing hyperspectral and multispectral image fusion by CNN denoiser. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *32*, 1124–1135. [CrossRef]

34. Liu, X.; Liu, Q.; Wang, Y. Remote sensing image fusion based on two-stream fusion network. *Inf. Fusion* **2020**, *55*, 1–15. [CrossRef]

35. Hu, J.-F.; Huang, T.-Z.; Deng, L.-J.; Jiang, T.-X.; Vivone, G.; Chanussot, J. Hyperspectral image super-resolution via deep spatiospectral attention convolutional neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2021**, *33*, 7251–7265. [CrossRef]

36. Zhang, X.; Huang, W.; Wang, Q.; Li, X. SSR-NET: Spatial–spectral reconstruction network for hyperspectral and multispectral image fusion. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 5953–5965. [CrossRef]

37. Vivone, G. Multispectral and hyperspectral image fusion in remote sensing: A survey. *Inf. Fusion* **2023**, *89*, 405–417. [CrossRef]

38. Yokoya, N.; Grohnfeldt, C.; Chanussot, J. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 29–56. [CrossRef]

39. Kruse, F.A.; Lefkoff, A.B.; Boardman, J.W.; Heidebrecht, K.B.; Shapiro, A.T.; Barloon, J.P.; Goetz, A.F.H. The spectral image processing system (SIPS): Interactive visualization and analysis of imaging spectrometer data. *Remote Sens. Environ.* **1993**, *44*, 145–163. [CrossRef]

40. Wald, L. Quality of high resolution synthesised images: Is there a simple criterion? In Proceedings of the Third Conference "Fusion Earth Data: Merging Point Measurements, Raster Maps and Remotely Sensed Images", Sophia Antipolis, France, 26–28 January 2000.

41. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]

42. Garzelli, A.; Nencini, F. Hypercomplex quality assessment of multi/hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2009**, *6*, 662–665. [CrossRef]

43. Alparone, L.; Aiazzi, B.; Baronti, S.; Garzella, A.; Nencini, F.; Selva, M. Multispectral and panchromatic data fusion assessment without reference. *Photogramm. Eng. Remote Sens.* **2008**, *74*, 193–200. [CrossRef]
44. Khan, M.M.; Alparone, L.; Chanussot, J. Pansharpening quality assessment using the modulation transfer functions of instruments. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 3880–3891. [CrossRef]
45. Alparone, L.; Garzelli, A.; Vivone, G. Spatial consistency for full-scale assessment of pansharpening. In Proceedings of the IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018.
46. Aiazzi, B.; Alparone, L.; Baronti, S.; Carlà, R.; Garzelli, A.; Santurri, L. Full scale assessment of pansharpening methods and data products. In Proceedings of the Image and Signal Processing for Remote Sensing XX, Amsterdam, The Netherlands, 22–25 September 2014.
47. Kwan, C.; Budavari, B.; Bovik, A.C.; Marchisio, G. Blind quality assessment of pansharpened Worldview-3 images by using the combinations of pansharpening and hypersharpening paradigms. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1835–1839. [CrossRef]
48. Meng, X.; Bao, K.; Shu, J.; Zhou, B.; Shao, F.; Sun, W.; Li, S. A blind full-resolution quality evaluation method for pansharpening. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5401916. [CrossRef]
49. Bao, K.; Meng, X.; Chai, X.; Shao, F. A blind full resolution assessment method for pansharpened images based on multistream collaborative learning. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5410311. [CrossRef]
50. Wu, J.; Li, X.; Wei, B.; Li, L. A no-reference spectral quality assessment method for multispectral pansharpening. In Proceedings of the IGARSS 2023–2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 16–21 July 2023.
51. Wald, L.; Ranchin, T.; Mangolini, M. Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images. *Photogramm. Eng. Remote Sens.* **1997**, *63*, 691–699.
52. Ranchin, T.; Aiazzi, B.; Alparone, L.; Baronti, S.; Wald, L. Image fusion—The ARSIS concept and some successful implementation schemes. *Isprs J. Photogramm. Remote Sens.* **2003**, *58*, 4–18. [CrossRef]
53. Thomas, C.; Ranchin, T.; Wald, L.; Chanussot, J. Synthesis of multispectral images to high spatial resolution: A critical review of fusion methods based on remote sensing physics. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1301–1312. [CrossRef]
54. Palsson, F.; Sveinsson, J.R.; Ulfarsson, M.O.; Benediktsson, J.A. Quantitative quality evaluation of pansharpened imagery: Consistency versus synthesis. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1247–1259. [CrossRef]
55. Dian, R.; Li, S.; Sun, B.; Guo, A. Recent advances and new guidelines on hyperspectral and multispectral image fusion. *Inf. Fusion* **2021**, *69*, 40–51. [CrossRef]
56. Carlà, R.; Santurri, L.; Aiazzi, B.; Baronti, S. Full-scale assessment of pansharpening through polynomial fitting of multiscale measurements. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6344–6355. [CrossRef]
57. Vivone, G.; Restaino, R.; Chanussot, J. A bayesian procedure for full resolution quality assessment of pansharpened products. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4820–4834. [CrossRef]
58. Marcello, J.; Ibarrola-Ulzurrun, E.; Gonzalo-Martín, C.; Chanussot, J.; Vivone, G. Assessment of hyperspectral sharpening methods for the monitoring of natural areas using multiplatform remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8208–8222. [CrossRef]
59. Picone, D.; Restaino, R.; Vivone, G.; Addesso, P.; Dalla Mura, M.; Chanussot, J. Band assignment approaches for hyperspectral sharpening. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 739–743. [CrossRef]
60. Kawulok, M.; Kowaleczko, P.; Ziaja, M.; Nalepa, J.; Kostrzewa, D.; Latini, D.; Santis, D.; Salvucci, G.; Petracca, I.; Pegna, V.; et al. Understanding the Value of Hyperspectral Image Super-Resolution from Prisma Data. In Proceedings of the IGARSS 2023–2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 16–21 July 2023.
61. Newcomb, S. Note on the frequency of use of the different digits in natural numbers. *Am. J. Math.* **1881**, *4*, 39–40. [CrossRef]
62. Benford, F. The law of anomalous numbers. *Proc. Am. Philos. Soc.* **1938**, *78*, 551–572.
63. Luque, B.; Lacasa, L. The first-digit frequencies of prime numbers and Riemann zeta zeros. *Proc. R. Soc.* **2009**, *465*, 2197–2216. [CrossRef]
64. Burke, J.; Kincanon, E. Benford's law and physical constants: The distribution of initial digits. *Am. J. Phys.* **1991**, *59*, 952. [CrossRef]
65. Leemis, L.M.; Schmeiser, B.W.; Evans, D.L. Survival distributions satisfying Benford's law. *Am. Stat.* **2000**, *54*, 236–241. [CrossRef]
66. Brown, R.J.C. Benford's law and the screening of analytical data: The case of pollutant concentrations in ambient air. *Analyst* **2005**, *130*, 1280–1285. [CrossRef] [PubMed]
67. Wang, L.; Ma, B.-Q. A concise proof of Benford's law. *Fundam. Res.* 2023; *in press*. [CrossRef]
68. Ou, F.-Z.; Wang, Y.-G.; Zhu, G. A novel blind image quality assessment method based on refined natural scene statistics. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019.
69. Al-Bandawi, H.; Deng, G. Classification of image distortion based on the generalized Benford's law. *Multimed. Tools Appl.* **2019**, *78*, 25611–25628. [CrossRef]
70. Varga, D. No-Reference Image quality assessment based on the fusion of statistical and perceptual features. *J. Imaging* **2020**, *6*, 75. [CrossRef] [PubMed]
71. Varga, D. Analysis of Benford's law for no-reference quality assessment of natural, screen-content, and synthetic images. *Electronics* **2021**, *10*, 2378. [CrossRef]
72. Fu, D.; Su, W. A generalized Benford's law for JPEG coefficients and its applications in image forensics. In Proceedings of the Security, Steganography, and Watermarking of Multimedia Contents IX, San Jose, CA, USA, 28 January–1 February 2007; Volume 6505, pp. 1–11.

73. Zhang, X.; Deng, C.; Han, Y. The image space of Meyer wavelet transform. In Proceedings of the 2013 2nd International Conference on Measurement, Information and Control, Harbin, China, 16–18 August 2013.

74. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [CrossRef]