Assessing Full-Resolution Pansharpening Quality: A Comparative Study of Methods and Measurements

Xiaodi Guan, Fan Li, *Senior Member, IEEE*, Xingliang Zhang, Mingyang Ma, and Shaohui Mei, *Senior Member, IEEE*

***Abstract*—Since the introduction of pansharpening, quality assessment has played a pivotal role in related remote sensing research to ensure the overall system's reliability. Full-resolution (FR) quality assessment is a debated research topic for applications. However, FR assessment faces challenges due to the absence of reference compared to reduced-resolution assessment. Moreover, the lack of ground truth makes measuring quality metrics even more challenging. To summarize the current measures for these two challenges, this article presents a comprehensive study of FR methods. We review various FR techniques and analyze how they extract spatial and spectral features from pansharpened images without reference to high-resolution multispectral images. A classification approach is proposed to group these methods based on their shared characteristics, making it easier for researchers and practitioners to compare and select the most appropriate FR method for specific applications. Furthermore, we provide a summary of strategies for measuring FR performance in the absence of ground truth. These strategies are classified into subjective and objective approaches. In addition, we conduct a board analysis on a large-scale public pansharpened database with unified measuring criteria. This unified analysis allows us to present experiments from a statistical perspective, measure FR protocols' performance, and provide a broad qualitative and quantitative analysis. Overall, this study contributes to the development of pansharpening and provides guidance for selecting appropriate FR methods, as well as strategies for measuring FR performance.**

***Index Terms*—Full-resolution quality assessment, image quality assessment (IQA), pansharpening, remote sensing.**

I. Introduction

DUE to limitations in remote sensing satellite sensors and systems, such as restricted incoming radiation energy, onboard storage capacity, and data transmission, remote sensing images obtained must make a tradeoff between spatial and spectral resolutions [1]. Whereas most satellites provide high-resolution (HR) panchromatic (PAN) images alongside low-resolution (LR) multispectral (MS) images, rather than HR-MS imagery. The combination of an MS image with a fine spatial resolution PAN image, called PAN sharpening (pansharpening), can overcome the limitation by incorporating complementary spatial and spectral advantages from HR-PAN and LR-MS images [2], [3]. An effective PAN method can increase the availability of remote sensing images and be highly advantageous for subsequent remote sensing tasks such as change and target detection.

Numerous pansharpening methods have been developed over the years. Vivone et al. [4] categorized these methods into two groups: component substitution (CS) and multiresolution analysis (MRA). CS-based methods use forward spectral transformation to replace a specific spatial part of MS with PAN, followed by an inverse spectral transformation to obtain the HR-MS image [5], [6], [7], [8], [9], [10], [11], [12]. This category includes classical methods such as principal component analysis [13], intensity–hue–saturation [14], and Gram–Schmidt (GS) methods [9], [15]. MRA-based methods use different MRA tools, such as wavelet transform [16] and Laplacian pyramid [17], to split the input images into low- and high-frequency components. A specific method is then employed to regenerate the HR-MS pansharpened image by fusing the low- and high-frequency components. Well-known techniques in this category include additive wavelet luminance proportional (AWLP) and generalized Laplacian pyramid (GLP)-based methods [18], [19], [20], [21], [22], [23]. Variational optimization (VO)-based pansharpening methods have also been proposed, which are based on variational theory and rely on energy function optimization [24], [25], [26], [27], [28], [29], [30], [31]. Deep learning (DL)-based methods establish a complex mapping between the original images and the pansharpening images based on the deep neural network (DNN). With the supervision of large-scale paired images, the networks learn to generate the HR-MS image [32], [33], [34], [35], [36], [37], [38], [39], [40], [41].

Since the topic of pansharpening was put forward, quality assessment has been pivotal in related research. The construction of a rational mathematical model can facilitate the swift and efficient evaluation of fused HR-MS images, thereby contributing to the control and optimization of the remote sensing system. First, due to the inherent absence of HR-MS images, the applicability of the pansharpened images is essential for future processing. Pansharpening often performs as a preliminary

Manuscript received 6 May 2023; revised 29 June 2023; accepted 19 July 2023. Date of publication 24 July 2023; date of current version 1 August 2023. This work was supported in part by the Major Science and Technology Projects in Xinjiang Uygur Autonomous Region under Grant 2022A02012-2, in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2023-JC-JQ-51, and in part by the Xi'an Science and Technology Project under Grant 21RGZN0009. *(Corresponding authors: Fan Li; Shaohui Mei.)*

Xiaodi Guan, Fan Li, and Mingyang Ma are with the School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: gxd1997@stu.xjtu.edu.cn; lifan@mail.xjtu.edu.cn; mamingyang@mail.xjtu.edu.cn).

Xingliang Zhang is with the College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China (e-mail: liangliangxju@163.com).

Shaohui Mei is with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China (e-mail: meish@nwpu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3298104

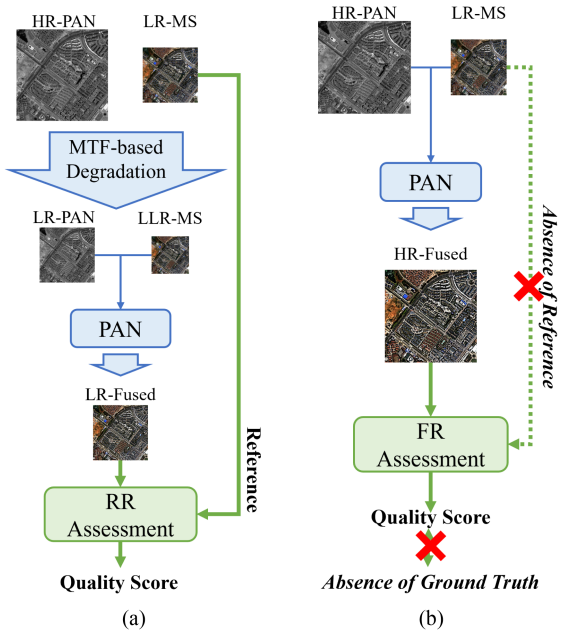

Fig. 1. Comparison of (a) reduced-resolution and (b) full-resolution pansharpening quality assessment.

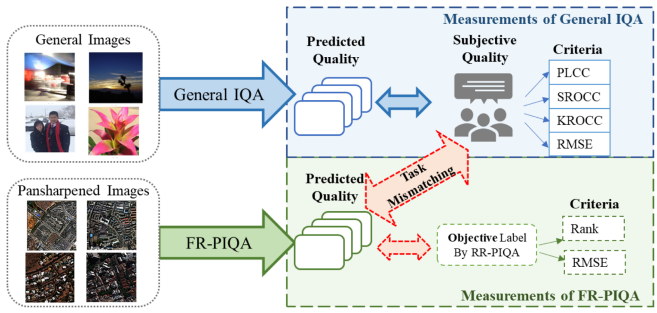

Fig. 2. Comparison between measurements of general IQA and FR-PIQA. The top is the measurements of general IQA, and the bottom is the FR-PIQA.

step for other remote sensing tasks, such as change detection [42], [43], segmentation [44], scene classification [45], and so forth. HR-MS images obtained by effective pansharpening methods are highly desirable. Hence, the quality assessment of pansharpening HR-MS images should be given much attention. Second, from the perspective of a satellite system, an efficient quality assessment method can optimize system efficiency under constraints such as bandwidth. In [42], it was noted that although pansharpened images result in higher quality change detection maps, the presence of pansharpening artifacts adversely affects the detection process. As a result, an appropriate pansharpening method is critical to the efficient functioning of the entire remote sensing system. For this purpose, specialized and targeted quality assessment methods must be developed.

In a word, accurate quality evaluation is crucial to ensure the reliability of remote sensing systems. However, assessing the quality of HR-MS images remains challenging due to the complexity of multimodal fusion and the absence of HR reference images. To address this issue, Wald et al. [46], [47] proposed two characteristics that pansharpened image evaluation should meet: 1) *consistency:* any HR-MS image when degraded to its original resolution should be as identical as possible to the original MS data; and 2) *synthesis:* the fusion product should be as identical as possible to the MS imagery that the corresponding sensor would observe at the highest resolution.

The two characteristics mentioned above give rise to two kinds of pansharpened image quality assessment (PIQA) approach: reduced-resolution (RR) evaluation and full-resolution (FR) evaluation. RR evaluation is based on the synthesis principle, as depicted in Fig. 1(a). The original HR-PAN and LR-MS images are spatially degraded by ideal filters (e.g., modulation transfer function-matched filters) based on their spatial resolution ratio. Then, the degraded MS images are pansharpened by the degraded PAN images, and the original MS image serves as the reference image for quality assessment. Accordingly, several RR quality assessment metrics with reference are developed, including spectral angle mapper (SAM) [48], peak signal-to-noise ratio (PSNR) [49], structural similarity (SSIM) [50], erreur relative globale adimensionnelle de synthése (ERGAS) [51], correlation coefficient (CC) [52], universal image quality index (UIQI) [53], the UIQI of pansharpened MS imagery Q_4 [54], and $Q2^n$ that could be regarded as the multiband extended version of Q_4 [55].

Although LR-MS is available as a reference in RR evaluation, degradation poses constraints on the entire procedure. As Selva et al. [56] pointed out, the scale-invariant hypothesis may be invalid, making it difficult to guarantee the quality of the degraded pansharpened images to be the same as that of pristine ones. Due to the inherent flaws, applying RR evaluation in remote sensing systems can be challenging. To achieve a more precise quality evaluation, several studies focus on FR evaluation. This approach evaluates the quality of the pansharpened image at HR-PAN resolution without the HR-MS reference image or any degradation in the fusion process, as depicted in Fig. 1(b).

However, due to such a direct evaluation process, the FR quality assessment approach faces two major challenges in implementation: 1) how to evaluate the image quality when a reference image is absent and 2) how to measure the effectiveness of the FR methods when ground truth is absent, as depicted by the *red cross* in Fig. 1(b). In the RR approach, researchers evaluated the dissimilarity between pansharpened images and the original MS images to predict image quality. To address reference-absence evaluation, researchers developed hand-crafted and learning features on spatial and spectral domains to abstract image features for FR quality evaluation without comparison.

When it comes to measuring the effectiveness of quality assessment methods, as Fig. 2 revealed, the general image quality assessment (IQA) is done by comparing the subjective (human-rated) data and predicting quality scores to measure the effectiveness of the IQA techniques. This enables an accurate measurement how well IQA techniques perform in comparison to human visual perception. However, for the task of PIQA, it is challenging to obtain a large number of satellite images due to professional issues. In addition, remote sensing data

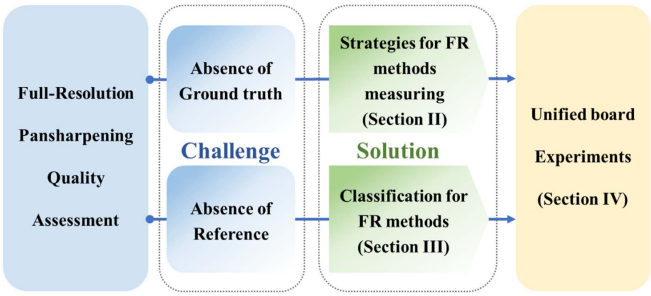

Fig. 3. Framework of the study, containing the solution to the absence of ground truth and reference, corresponding to Sections II and III. And the board experiments are in Section IV.

possess distinct properties essential for professional use, making it difficult to rely on human grading and create a sizable subjective database. To overcome these challenges, researchers have employed two strategies to measure the efficiency of FR methods: 1) objective strategy based on RR and 2) subjective strategy based on the human visual system (HVS). The objective strategy regards the RR evaluation as a benchmark and adopts the RR predicting results as the proxy ground truth to evaluate the FR methods. On the other hand, the subjective strategy adheres to the implementation of general IQA and utilizes human-rated data to complete the HVS-related measurement of the FR method. By utilizing these two kinds of approaches, researchers can achieve a quantitative comprehension of the effectiveness of FR methods. Such a measuring approach can be valuable for uncovering potential limitations or issues with FR algorithms and can ultimately lead to improvements in method design and implementation.

To systematically understand how current researchers solve the challenges of FR (absence of reference and ground truth), we provide a summary and comparison of current FR methods. While there is an existing survey on FR quality assessment [57], the work focuses on contrasting four FR methods and highlighting the methodologies. In this article, we give an overview of how the FR methods address the aforementioned challenges by synthesizing and comparing various FR methods. In summary, the major contributions are as follows, and the framework is depicted in Fig. 3.

1) *A classification approach for FR methods based on feature extraction*: In this study, we reviewed various FR methods and analyzed how they extract spatial and spectral features from pansharpened images without reference to HR-MS images. We also presented a classification approach for FR methods based on the measure of feature extraction. For practitioners, grouping these methods can help to compare and select the most appropriate FR method for specific requirements, which are considered with multiple factors such as spatial and dynamic range, data type, and the level of complexity.
2) *The summary of strategies for FR method measuring*: Facing the absence of ground truth, we concluded the strategy on how to generate the proxy ground truth for performance measurements. The strategies are classified into subjective and objective ones, and the two kinds of strategies are qualitatively analyzed.
3) *Unified board analysis on a large-scale public database*: A board analysis is conducted on a large-scale public pansharpened database with unified measuring criteria. Experiments from a perspective of statistical analysis are presented to support the theoretical analysis of the FR protocols and to evaluate their performance, providing a wide qualitative and quantitative analysis.

The rest of this article is organized as follows. Section I presents the introduction of the whole work. Section II concludes the strategies for FR method measuring, while the summary and classification of FR methods are provided in Section III. Experimental results and statistical discussion are presented in Section IV. Finally, Section V concludes this article.

II. Solution to Absence of Ground Truth: Objective and Subjective

As mentioned in Section I, due to the lack of ground truth, measuring of the FR methods is ambiguous, which may hinder the development of FR methods. Currently, researchers employ two kinds of strategies to face and solve the challenge: 1) objective strategy based on RR and 2) subjective strategy based on the HVS.

A. Objective Strategy

Since the RR methods can refer to the original MS images and predict the pansharpened image quality by comparing the corresponding data in the spatial and spectral domains, some researchers consider that the prediction of RR is more reliable and reasonable. Therefore, most studies regard the well-known RR protocols as a benchmark and adopt the RR predicting results as the proxy ground truth to measure the FR methods. By comparing the prediction of FR methods with the proxy ground truth, the effectiveness of the FR methods is measured. The mainly adopted measuring criteria are global quality rank and root-mean-square error (RMSE). Specifically, QNR [58], FQNR [59], RQNR [60], HQNR [61], quality estimation by fitting (QEF) [62], and joint quality measure (JQM) [63] are compared with SAM [48], ERGAS [51], and Q_4 [54] on the global quality rank. Besides the rank comparison, researchers also compute the RMSE with $Q2^n$ prediction to measure the performance of Kalman QEF (KQEF) [64], CQE [65], and ODA analyzer [66].

The objective strategy relies on proxy ground truth, which can be accessed easily via the public RR protocols. However, this strategy has some flaws. First, as was said in Section I, due to the inherent degradation process, the prediction of the RR protocols may not always be precise and consistent with the pristine pansharpened images. This can affect the accuracy of measuring results. Second, the performance of different RR methods is not always robust. In our work, we selected six images from public datasets and compute the SAM [48], ERGAS [51], and $Q2^n$ [55] metrics of the images with the aid of AWLP [67] in the RR approach, as depicted in Fig. 4. As shown in Table I, the predicting results vary among different RR methods, and the

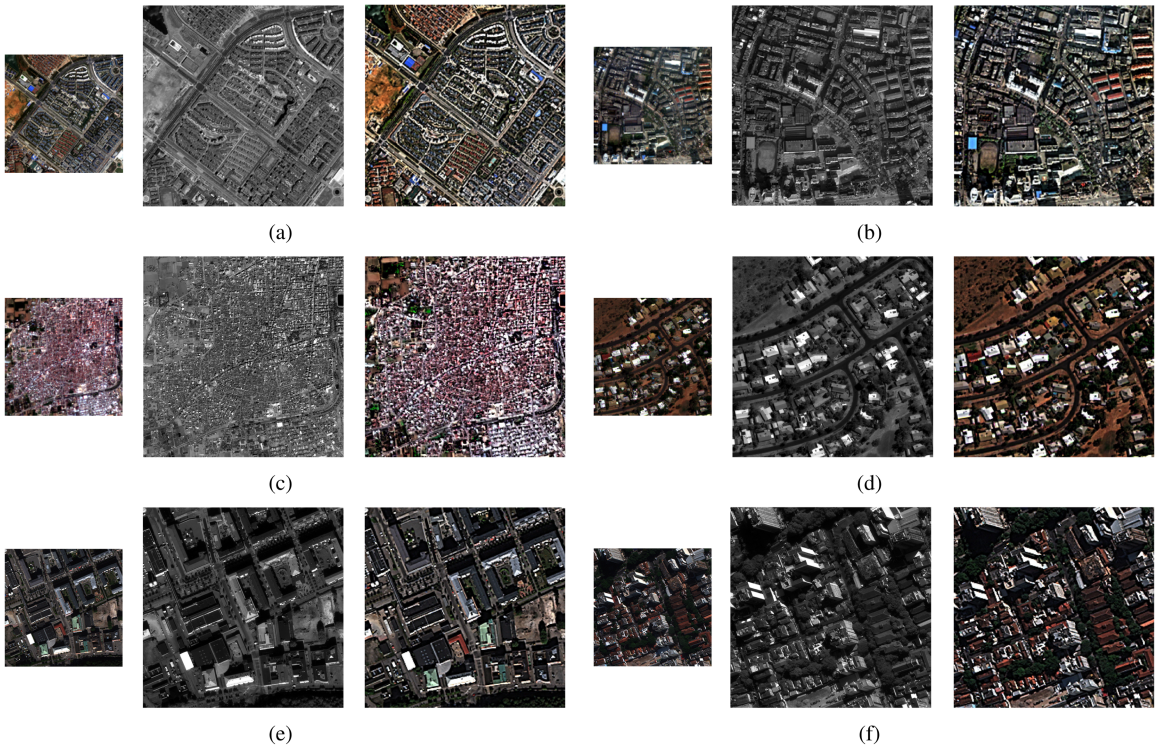

Fig. 4. Six pairs of LR-MS, HR-PAN, and HR-MS images. (a) Captured by IKONOS, with a spatial dimension of 1 and 4 m, and a spectral dimension of one and four bands. (b) Captured by QuickBird, with a spatial dimension of 0.61 and 2.44 m, and a spectral dimension of one and four bands. (c) Captured by GaoFen-1, with a spatial dimension of 2 and 8 m, and a spectral dimension of one and four bands. (d) Captured by WorldView-4, with a spatial dimension of 0.31 and 1.24 m, and a spectral dimension of one and four bands. (e) Captured by WorldView-2, with a spatial dimension of 0.5 and 2 m, and a spectral dimension of one and eight bands. (f) Captured by WorldView-3, with a spatial dimension of 0.31 and 1.24 m, and a spectral dimension of one and eight bands. The HR-MS images are generated by AWLP [67].

TABLE I
DIFFERENT RR PREDICTING RESULTS OF SIX IMAGES FROM PUBLIC DATASETS

Image	SAM (–)	ERGAS (–)	SSIM (+)	$Q2^n$ (+)
(a)	3.3856 (Rank 3)	1.0576 (Rank 2)	0.9994 (Rank 2)	0.7154 (Rank 4)
(b)	1.6792 (Rank 1)	1.0294 (Rank 1)	0.9998 (Rank 1)	0.6317 (Rank 5)
(c)	3.0260 (Rank 2)	2.9532 (Rank 5)	0.9983 (Rank 5)	0.4172 (Rank 6)
(d)	4.4069 (Rank 4)	2.0602 (Rank 3)	0.9970 (Rank 6)	0.8377 (Rank 2)
(e)	7.4164 (Rank 5)	2.6888 (Rank 4)	0.9990 (Rank 3)	0.8410 (Rank 1)
(f)	8.0319 (Rank 6)	2.9914 (Rank 6)	0.9986 (Rank 4)	0.8045 (Rank 3)

predicted rank of the example images is not consistent. The error may be attributable to the various computational approaches and mathematical tools in different RR protocols. However, given the difficulty in obtaining ground truth, the advantage of the objective strategy lies in its efficiency and convenience. With such high efficiency, occasional errors and instability can also be tolerated.

B. Subjective Strategy

Confined with the inconsistency of the RR protocols, some researchers worked to build large-scale datasets that will provide uniform subjective data based on the HVS as the ground truth [66], [68]. The subjective dataset in [68] was established on the human rating of 13 620 pansharpened images acquired from IKONOS, QuickBird, GaoFen-1, WorldView-2, WorldView-3, and WorldView-4 satellite sensors by 28 participants. Before the formal subjective test, all the participants were asked to view 60 testing data and gave corresponding scores to understand the rating criteria. In the formal test, each participant tested 300 data items at each time. Furthermore, original subjective scores were further processed by excluding unreliable subjects and normalized to obtain the difference mean opinion scores. And the subjective dataset in [66] containing 135 images from IKONOS was established based on 33 participants' ratings. Especially, the subjects were asked to rate the distortion present in the images such as artifacts, insufficient color representation, lack of sharpness, and over-/underexposure.

Based on the subjective dataset, some FR methods are measured with the HVS-related ground truth, including [66], [69], [70], [71]. For these methods based on the subjective strategy, the measuring metrics, Spearman rank-order correlation coefficient (SROCC), Pearson linear correlation coefficient (PLCC), Kendall rank-order correlation coefficient (KROCC),

and RMSE, are the same as those used in general IQA. These metrics are intended to calculate the correlation between predicted results and ground truth.

As for the subjective strategy, it takes human perception as the standard and involves perception by numerous volunteers and data preprocessing. This makes the subjective ground truth more robust than the proxy labels generated by the objective strategy based on RR protocols. However, there are some drawbacks to consider. First, the acquisition of remote sensing images is expensive, and building such a large database is time consuming and labor intensive. Second, pansharpening is primarily used for machine analysis rather than human viewing. Therefore, the ground truth generated by human observation and perception may be misleading to machine-related tasks. Third, the subjective ground truth may lag behind the developing pansharpening method. We can only adopt the existing pansharpening methods to process the MS and PAN images and then conduct a subjective study to get the reliable ground truth. The problem is that advanced pansharpening methods in the future might not have corresponding ground truth available, and this issue cannot be solved completely. In summary, the subjective strategy is more robust but has limitations due to the high cost of acquiring images, the subjective nature of human perception, and potential issues with keeping pace with advancements in pansharpening methods.

III. Solution to FR Quality Assessment

From a practical perspective, the use of FR assessment is advantageous as it allows for direct validation on a large scale. However, FR presents a greater scientific challenge compared to RR, primarily due to the need for a suitable strategy to overcome the lack of a reference image without resorting to spatial degradation. The process of FR methods involves extracting both spectral and spatial information to represent the pansharpened image and then assessing its quality. Based on feature acquisition, FR methods can be divided into two main categories: hand-crafted and learning features. The following are descriptions of each category.

A. Methods Based on Hand-Crafted Features

During the early stages of FR methods, the primary approach to assessment was based on hand-crafted features of spatial and spectral distortions. Researchers developed specific parameters to represent the features that differentiate distortions. These methods can be divided into two groups based on whether the features are filtered to obtain a partial proxy reference or not.

1) Quality Assessment With Pristine Features: One of the earliest FR evaluation methods was proposed by Zhou et al. [72]. This method employed the high-pass Laplacian filter to separate the high-spatial-frequency components. The spatial CC was then extracted as the spatial feature by using the CC between the spatial details of the pansharpened image and those of the original PAN image. In addition, the spectral distortion was calculated band by band by utilizing the average absolute difference between the pansharpened band and the interpolated MS band. The spatial and spectral distortion are calculated as follows:

$$D_s = \mathrm{CC}(F_H, P_H) \tag{1}$$

$$D_{\lambda,b} = \frac{1}{I \times J} \sum_i \sum_j \left| F_{i,j,b} - \tilde{M}_{i,j,b} \right| \tag{2}$$

where F is the pansharpened HR-MS image, P is the original HR-PAN image, M is the original LR-MS image, and $\tilde{M}$ is the MS image upsampled to PAN scale. F_H and P_H indicate the high-pass components of F and P, respectively. CC denotes the CC computing. Besides, i and j are the pixel locations, I and J denote the spatial dimension sizes, and b denotes the bth spectral band.

The most widespread FR assessment method to date is the QNR protocol [58]. It determines the quality of the pansharpened image by evaluating the preservation of the original relationships of the input pansharpening data. Changes in intrarelationships between MS bands are regarded as spectral distortions, while changes in interrelationships between MS and PAN are regarded as spatial distortions. And the difference operator UIQI is used to assess the similarity of the relationship. The spatial and spectral distortion of QNR is calculated as follows:

$$D_s = \frac{1}{B} \sum_{b=1}^{B} |Q(F_b, P) - Q(M_b, P_L)| \tag{3}$$

$$D_\lambda = \frac{1}{B \times (B-1)} \sum_{b=1}^{B} \sum_{c=1, c \neq b}^{B} \left| Q(F_b, F_c) - Q(\tilde{M}_b, \tilde{M}_c) \right| \tag{4}$$

$$\mathrm{QNR} = (1 - D_\lambda)^\alpha (1 - D_s)^\beta \tag{5}$$

where P_L is the downsampled PAN image.

The QNR protocol has been a widely used FR assessment method for evaluating the quality of pansharpened images. Several variations of QNR have been developed, including the general QNR (GQNR) [73], which is a QNR-like protocol without filtering. The GQNR approach addresses the issue of inaccurate spatial distortion prediction caused by spectral overlap between the PAN and short-wave infrared bands in WorldView-3 images. To overcome this issue, the GQNR protocol applies the natural image quality evaluator (NIQE) [74] model to characterize spatial distortion and then integrates the spectral distortion index in the QNR to obtain the final quality prediction, calculated as follows:

$$D_s^G = \mathrm{NIQE}(F) \tag{6}$$

$$D_\lambda^G = D_\lambda \tag{7}$$

$$\mathrm{GQNR} = D_s^G \cdot D_\lambda^G \tag{8}$$

where NIQE denotes the natural image quality evaluator model proposed in [74].

The methods that do not involve filtering are the most straightforward means of evaluation and offer direct calculation. However, they may not provide an accurate assessment of spectral quality. This is because when such methods measure spectral distortion, the image with the least spectral distortion will be

one that has no spatial features combined, such as a simple interpolated MS image. To address this paradox, some methods incorporate spectral degradation to assess quality.

2) Quality Assessment With Filtered Features: The FQNR protocol [59] was the first to significantly modify the original QNR from the perspective of spectral distortion. It is widely accepted that any color difference between the pansharpened MS and the original MS data is an indicator of spectral distortions. However, direct evaluation is impossible as there is no-reference (NR) MS image. The FQNR protocol has introduced an innovative approach to assess the spectral preservation of the pansharpened images by applying Wald's protocol's consistency property. MTF-matched filters are utilized to execute the degradation of the MS bands, and a nearly ideal filter is used for the PAN images' consistency property evaluation. For spectral distortion, each fused MS band is spatially filtered and decimated with a specific MTF-matched filter, and the $Q2^n$ index between the original MS data and the degraded pansharpened MS images is calculated. After that, a unit complementary value is employed to compute the spectral distortion index, calculated as follows:

$$D_\lambda^F = 1 - Q2^n(F_{L\downarrow}, M) \tag{9}$$

where $F_{L\downarrow}$ indicates the decimated fused images by MTF filtering.

For spatial distortion, the QNR measures spatial distortions by computing changes in the relationships between the PAN and MS bands. The FQNR differs from the QNR in that it compares the relationships between the high-frequency components of the PAN and those of the MS bands across scales. PAN and MS are computed at each scale and compared in terms of the UIQI before their absolute difference is measured and averaged for the number of spectral bands. The spatial distortion and FQNR are calculated as follows:

$$D_s^F = \frac{1}{B}\sum_{b=1}^{B} |Q(M_{b,H}, P_{L\downarrow H}) - Q(F_{b,H}, P_H)| \tag{10}$$

$$\text{FQNR} = (1 - D_\lambda^F)^\alpha (1 - D_s^F)^\beta \tag{11}$$

where $M_{b,H}$, $F_{b,H}$, and P_H indicate the high-pass LR-MS, fused HR-MS, and HR-PAN image by MTF filtering, respectively, and $P_{L\downarrow H}$ is the high-pass decimated LR-PAN image.

The HQNR, proposed in [61], combines the advantages of both the FQNR and the QNR. It utilizes the successful consistency approach of the FQNR for assessing spectral quality. In addition, the HQNR uses the spatial distortion technique of the QNR to evaluate spatial quality. The HQNR is calculated as follows:

$$\text{HQNR} = (1 - D_\lambda^F)^\alpha (1 - D_s)^\beta. \tag{12}$$

The RQNR provides a further QNR variation [60]. The FQNR's spectral consistency approach is adopted. The innovation of the RQNR is in the spatial quality assessment. According to the reasoning, the MS bands after being pansharpened should match the PAN band. Therefore, the PAN can be described as a linear combination of the MS bands as a linear combination. Therefore, the RQNR leverages the coefficient of determination of the multivariate linear regression (MLR) between the fused MS and the FR PAN image to assess spatial fidelity.

According to [60], the RQNR is a variation of QNR that adopts the spectral consistency approach used by the FQNR but also includes the spatial quality assessment. The method involves calculating a MLR between the fused MS and the FR PAN image, where the PAN can be described as a linear combination of the MS bands. The coefficient of determination from this MLR is then used to assess spatial fidelity. This innovation in the RQNR allows for a more comprehensive evaluation of image fusion quality.

Besides the QNR-like protocols, other methods are also proposed. In [62], a multiscale approach-based quality assessment framework called QEF was proposed. QEF recasts the FR assessment as a regression problem, where the metrics computed at several degraded resolution scales are independent variables and the metrics computed at the full scale are dependent variables. QEF iteratively performs the spatial degradation of Wald's protocol to generate quality/distortion metric functions of the ground sample distance (GSD). The quality at the full scale is then inferred through proper interpolation techniques. However, the fundamental drawback of this method is the absence of FR measurements in the quality estimation procedure.

The KQEF was developed based on the QEF method to address the lack of an FR index in the framework [64]. To overcome this flaw, a sequential Bayesian framework methodology using Kalman filters was proposed. The KQEF integrates the QEF multiresolution framework without the use of RR indices. At each scale of degraded resolution, the HQNR index and the $Q2^n$ index are generated, and two sequences of quality measurements, which are a function of the GSD, are used for the Kalman filter to infer quality prediction at the full scale. A less computationally demanding version of KQEF is proposed in [65], where the degradation iteration of reproducing fusion at multiple coarser resolutions and evaluation of an index without reference is required at each scale.

The JQM was proposed as another procedure for evaluating quality in [63] and [75]. Unlike QNR, the JQM calculates a weighted sum of two terms. The first term is the spatial quality, which is determined by comparing a low-pass filtered pansharpened image and the original MS image at an LR/RR scale. The second term is the spectral quality, which is determined by measuring the intensity estimated from the spectrally weighted pansharpened MS image and the original PAN image at an HR scale. The JQM differs from other QNR-like protocols by using SSIM metrics and a composite similarity measure based on means, standard deviations, and correlation coefficient (CMSC), rather than the conventional UIQI. The author contends that the CMSC's offering of translation invariance is more relevant for distant sensing applications like clustering and change detection than the UIQI. And the SSIM version of JQM can be calculated as

$$D_s^{\text{JQM}} = \text{SSIM}\left(P, \frac{1}{B}\sum_{b=1}^{B} F_b\right) \tag{13}$$

$$D_\lambda^{\text{JQM}} = \frac{1}{B}\sum_{b=1}^{B} \text{SSIM}(M_b, F_{L\downarrow}). \tag{14}$$

This kind of methods utilizing filtering rely on Wald's protocols' consistency property to assess quality. However, it is important to note that filtering/degradation cannot be ignored as it may have significant impacts, similar to the RR methods. Furthermore, these methods often use simple hand-crafted features, which may not be able to adequately represent complex structures and both spectral and spatial distortions, as opposed to just their interrelationships.

B. Methods Based on Learning Features

With the development of machine learning (ML) and DL strategies, researchers are also focusing on extracting learning-based features for FR evaluation.

Unlike the stepwise evaluation of the spectral, spatial, and final overall quality score using a weighted combination, which can introduce inconsistent results, Meng et al. [70] directly analyze the quality of fused images. To increase the robustness of the evaluation model, the proposed method is patch based rather than relying on a single pixel. The evaluation model considers all the spatial and spectral features that are sensitive to spatial and spectral distortions of the fused image. These features are simultaneously trained based on online benchmark multivariate Gaussian (MVG) fitting. Assuming that the spectral relations between the ideal pansharpened and original MS images should not be changed, a pristine benchmark assessment model is first created using the spatial characteristics extracted from the original HR-PAN image and the spectral features. The spatial information from the fused picture and the spectral features from the spectral variation between the fused and MS images serve as the foundation for a testing assessment model. The distance between the benchmark and testing models is the final metric used to evaluate the quality of the fused image.

Similarly, Zhou et al. [71] proposed an opinion-aware method to evaluate quality. The method extracts features from the spectral bands and typical information indices of the MS images that comprehensively reflect spatial distortion, spectral distortion, and the effects of pansharpening on applications. Using the features extracted from the pristine MS image training dataset, an MVG model is learned to evaluate image quality. This approach enables the model to capture both the objective and subjective aspects of image fusion evaluation.

With the continuous development of techniques, DNNs have become capable of automatically extracting more deep features relevant to quality assessment and optimizing these features using the backpropagation method to improve prediction performance. As a result, the DNN has been applied to various IQA and video quality assessment methods and provides a promising option for addressing the challenging FR task. Bao et al. [69] designed a Siamese framework to collaboratively learn the spatial, spectral, and overall quality of the fused image. The feature extraction layer's parameters in the spatial and spectral evaluation models are frozen for the overall evaluation model, which improves accuracy and convergence speed. This approach effectively captures spatial–spectral features from different sources and results in more robust and representative metrics for evaluating FR performance.

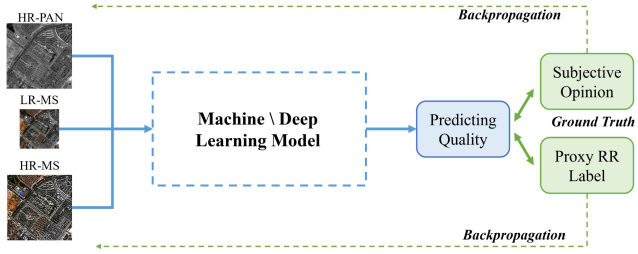

Fig. 5. Framework of the kind of methods that are based on learning features.

Badal et al. [76] proposed a learning-based NR approach to assess the quality of pansharpened images. This approach can predict state-of-the-art reference-based measures such as $Q2^n$ and SAM without requiring a reference image. They then designed an end-to-end deep pansharpening IQA network that computes the similarity of deep features fused from the PAN and input LR MS images with similar features extracted from the given pansharpened image. By leveraging the power of DL, this method can automatically learn discriminative spatial–spectral features and accurately evaluate the quality of a pansharpened image without relying on a reference image. Furthermore, the proposed NR evaluation approach is beneficial for applications where a ground truth reference image may not be available.

This kind of learning-based method in IQA can extract more effective and targeted features for quality evaluation and can be depicted in Fig. 5. However, they may face challenges when ground truth is not available, which is a common issue in various research areas. Breaking through the limit of ground truth can enable the DL method to make more significant progress in PIQA. By synthesizing large-scale training datasets with well-designed subjective experiments, DNNs can learn to assess image quality using visual cues that are similar to human perception. In addition, transfer learning and self-supervised learning techniques can also be leveraged to reduce the dependence on large amounts of labeled data and customize the network for specific tasks. Therefore, there is great potential for DL to make strides in PIQA.

IV. Experiments

A. Experimental Setup

1) Databases: We chose a large-scale benchmark dataset for pansharpening performance evaluation for a comprehensive comparison [77]. The dataset consists of 2270 pairs of HR-PAN and LR-MS images, and the details are shown in Table II. The dataset was grouped by remote sensing satellite types, consisting of 200 pairs of IKONOS PAN/MS images, 500 pairs of QuickBird PAN/MS images, 410 pairs of GaoFen-1 PAN/MS images, 500 pairs of WorldView-2 PAN/MS images, 500 pairs of WorldView-4 PAN/MS images, and 160 pairs of WorldView-3 PAN/MS images. The PAN images have a spatial dimension of 1024×024, and the MS images have a spatial dimension of 256×256. In the spectral dimension, the MS images from IKONOS, QuickBird, GaoFen-1, and WorldView-4 have four bands, and the WorldView-2 and WorldView-3 MS images have eight.

TABLE II
DETAILS OF THE LARGE-SCALE BENCHMARK PANSHARPENING DATASET

Satellite sensors	Number of images		Spatial dimension	Spectral dimension	Dimension size
IKONOS	200	PAN	1 m	1 band	1024×1024
		MS	4 m	4 bands	256×256×4
QuickBird	500	PAN	0.61 m	1 band	1024×1024
		MS	2.44 m	4 bands	256×256×4
GaoFen-1	410	PAN	2 m	1 band	1024×1024
		MS	8 m	4 bands	256×256×4
WorldView-2	500	PAN	0.5 m	1 band	1024×1024
		MS	2 m	8 bands	256×256×8
WorldView-3	160	PAN	0.31 m	1 band	1024×1024
		MS	1.24 m	8 bands	256×256×8
WorldView-4	500	PAN	0.31 m	1 band	1024×1024
		MS	1.24 m	4 bands	256×256×4

2) Implementation Details: The experimental validation relies on 25 pansharpening methods belonging to the four main categories recalled in Section I: CS-based [5], [6], [7], [8], [9], [10], [11], [12], MRA-based [11], [18], [19], [20], [21], [22], [23], [67], VO-based [26], [27], [28], [29], [30], [31], and DL-based [40], [41], with an ideal interpolator (EXP). Among the FR quality assessment described in Section III, we adopt four pubic metrics: QNR [58], FQNR [59], HQNR [61], and RQNR [60] with their spatial and spectral distortion indices for quantitative and qualitative comparison.

We employed the above 25 pansharpening methods to generate the HR-MS images, where a total of 56 750 images based on 2270 pairs of PAN/MS images are obtained. The codes we used are from the public MATLAB toolbox proposed in [78]. After that, we assess the quality of 56 750 images based on four FR methods. All the experiments are conducted on MATLAB2022R.

3) Performance Criteria: To face the absence of ground truth, we adopted the objective strategy: RR protocol $Q2^n$ as the reference. Besides the global numerical comparison, we also measure the FR methods' performance using the SROCC, PLCC, KROCC, and RMSE between the predicted quality scores and ground truth generated by $Q2^n$ in Table III.

Given N distorted images, the ground truth and predicted quality is represented by y_i and $\hat{y}_i$, respectively. The SROCC is calculated as follows:

$$\text{SROCC} = 1 - \frac{6\sum_{i-1}^{N}(v_i - p_i)^2}{N(N^2-1)} \tag{15}$$

where v_i is the rank of the ground truth y_i, and p_i is the rank of the predicted quality u_i of all N images.

The PLCC is calculated as follows:

$$\text{PLCC} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2}} \tag{16}$$

where $\bar{y}$ and $\bar{\hat{y}}$ are the means of y and y_i, respectively.

The KROCC is calculated as follows:

$$\text{KROCC} = \frac{N_c - N_d}{\frac{1}{2}N(N-1)} \tag{17}$$

where N_c is the number of concordant pairs in the observations, and N_d is the number of discordant pairs.

The RMSE is calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}(y_i - \hat{y}_i)^2}. \tag{18}$$

Higher SROCC, PLCC, and KROCC values and lower RMSE values represent a quality assessment method closer to $Q2^n$.

B. Performance of the FR Methods

To evaluate the effectiveness of the four FR quality evaluation metrics, we conducted a large-scale macroscopic quality evaluation with results in Table III. We assessed the FR quality of all 56 750 HR-MS images generated in the six individual databases. The FR quality prediction results were compared with the $Q2^n$ proxy ground truth for statistical evaluation.

As indicated in Table III, the evaluation of different algorithms for spatial distortion has shown relatively promising results. Among the algorithms considered, FQNR [59], HQNR [61], and RQNR [60] exhibit excellent performance and robustness across all the datasets. These metrics outperform the spatial distortion index in QNR and demonstrate more consistent performance across multiple datasets. Specifically, while the D_λ of QNR performs well on the WorldView-2 database, it shows poor performance on other databases. In contrast, D_λ^F, D_λ^H, and D_λ^R show superior performance on each respective database. This improved performance can be attributed to their tailored feature extraction strategy based on MTF filtering. The MTF filtering approach provides more informative and reference-based measurements compared to QNR's simple distance computation, resulting in better performance.

In terms of spectral distortion evaluation, D_s and D_s^H show better performance compared to other spectral distortion indices. However, these metrics still have room for improvement. One of the challenges in measuring spectral distortion is that it is not as apparent as spatial distortion in PAN sharpening. Exploring spectral distortion requires considering information from multiple frequency bands, making it more complex and challenging. In addition, the use of Q_{2n} as a proxy ground truth may not provide a sufficient measurement of spectral distortion information, leading to generally lower performance of the spectral indicator. In summary, the feature extraction and measurement of spectral distortion are still key and challenging points of pansharpening.

Overall, among the four algorithms considered, the HQNR algorithm demonstrates the best performance. This is because it combines indicators that exhibit superior performance in both spatial and spectral domains. The FQNR and RQNR algorithms also show competitive performance and relative stability. In contrast, the QNR algorithm performs poorly and is unstable, particularly showing high performance on WorldView-2 and WorldView-3 datasets but extremely poor performance on the GaoFen-1 dataset. This discrepancy can be attributed to the relatively simple measurement method used in QNR, which may not effectively capture the complexity of the data.

However, there is still room for improvement in the measurement of spectral distortion in all the methods. Enhancing

TABLE III
STATISTICAL RESULTS OF THE FOUR PUBLIC FR METHODS WITH THEIR SUBCOMPONENTS

Dataset	IKONOS				QuickBird				GaoFen-1			
Metrics	PLCC	SROCC	KROCC	RMSE	PLCC	SROCC	KROCC	RMSE	PLCC	SROCC	KROCC	RMSE
D_λ	0.5363	0.3275	0.2327	0.3737	0.4967	0.5073	0.3625	0.4228	0.1996	0.2535	0.1800	0.5347
D_s	0.6350	0.4576	0.3246	0.2889	0.3995	0.4132	0.2976	0.3537	0.2574	0.2861	0.2106	0.4531
QNR	0.5639	0.3863	0.2749	0.4977	0.4444	0.4423	0.3180	0.3828	0.2486	0.2694	0.2002	0.2059
D_λ^F	0.7721	**0.8103**	**0.6533**	**0.1675**	0.5161	0.5166	0.3747	0.2788	**0.7714**	**0.7747**	**0.6227**	0.4336
D_s^F	0.2068	0.2646	0.1862	0.4190	0.1249	0.1158	0.0818	0.4621	0.4908	0.4980	0.3655	0.5070
FQNR	0.7632	0.7889	0.6244	0.5245	0.4621	0.4579	0.3300	0.3113	0.7307	0.7369	0.5887	**0.1313**
D_λ^H	0.7721	**0.8103**	**0.6533**	**0.1675**	0.5161	0.5166	0.3747	**0.2786**	**0.7714**	**0.7747**	**0.6227**	0.4336
D_s^H	0.6350	0.4576	0.3246	0.2889	0.3995	0.4132	0.2976	0.3537	0.2572	0.2861	0.2106	0.4531
HQNR	**0.7969**	0.7727	0.5991	0.5004	**0.5451**	**0.5235**	**0.3749**	0.3171	0.6291	0.6409	0.4891	0.1641
D_λ^R	0.7721	**0.8103**	**0.6533**	**0.1675**	0.5161	0.5166	0.3747	0.2788	**0.7714**	**0.7747**	**0.6227**	0.4336
D_s^R	0.1632	0.1993	0.1384	0.4590	0.1117	0.1293	0.0826	0.5056	0.1970	0.2272	0.1582	0.4939
RQNR	0.7734	0.8052	0.6496	0.5476	0.4272	0.4243	0.3049	0.3251	0.4710	0.4474	0.3425	0.1896
Dataset	WorldView-2				WorldView-3				WorldView-4			
Metrics	PLCC	SROCC	KROCC	RMSE	PLCC	SROCC	KROCC	RMSE	PLCC	SROCC	KROCC	RMSE
D_λ	0.9020	0.8447	0.6887	0.3283	0.7930	0.7099	0.5675	0.3601	0.5916	0.6596	0.4945	0.3744
D_s	0.8943	0.8514	0.6893	0.2900	0.8666	0.7881	0.6293	0.2794	0.4161	0.4283	0.3049	0.3323
QNR	0.9182	0.8677	0.7078	0.5433	0.8820	0.8099	0.6558	0.5542	0.5183	0.5571	0.4034	0.5086
D_λ^F	0.9070	0.9013	0.7512	**0.1429**	0.8685	**0.8781**	0.7224	**0.1482**	0.6722	0.7406	**0.6105**	**0.2336**
D_s^F	0.6299	0.6224	0.4414	0.4100	0.6307	0.6075	0.4537	0.3900	0.2950	0.3570	0.2469	0.4160
FQNR	0.9096	0.9045	0.7531	0.5322	0.8636	0.8595	0.7030	0.5539	0.6738	0.7235	0.5898	0.4713
D_λ^H	0.9070	0.9013	0.7512	**0.1429**	0.8685	**0.8781**	0.7224	**0.1482**	0.6722	0.7406	**0.6105**	**0.2336**
D_s^H	0.8943	0.8514	0.6893	0.2900	0.8666	0.7881	0.6293	0.2794	0.4161	0.4283	0.3049	0.3323
HQNR	**0.9230**	**0.9111**	**0.7672**	0.5506	**0.8937**	0.8753	**0.7236**	0.5658	**0.6845**	**0.7439**	0.5895	0.4423
D_λ^R	0.9070	0.9013	0.7512	**0.1429**	0.8685	**0.8781**	0.7224	**0.1482**	0.6722	0.7406	**0.6105**	**0.2336**
D_s^R	0.1476	0.1748	0.1109	0.4547	0.2371	0.2406	0.1628	0.4135	0.2381	0.3741	0.2541	0.4321
RQNR	0.9081	0.9087	0.7617	0.5444	0.8593	0.8639	0.7042	0.5694	0.5872	0.6737	0.5278	0.4692

The bold indicates the best performance.

the measurement of spectral distortion is a crucial and challenging aspect of PIQA. Future developments in FR algorithms can greatly benefit from innovations that substantially improve the performance of spectral distortion measurement. Further advancements in spectral distortion extraction and evaluation methods are needed to enhance the overall performance of FR algorithms.

C. Numerical Comparison of Pansharpening Methods on Individual Datasets

This section is dedicated to presenting the experimental results obtained at FR. We compute the average value of the HR-MS image quality score for the specific pansharpening method. Tables IV–IX report the numerical scores in the six individual datasets for all of the QNR, FQNR, HQNR, RQNR, and their subindices. And Figs. 6 and 7 show the outcomes of 26 pansharpening methods generated from two pairs of example images in IKONOS and WorldView-3, while the quality scores predicted by QNR are attached. The comparison of the results will be conducted by dividing the section into four subsections, each representing a class of pansharpening methods.

1) CS-Based Methods: In the CS-based class, BDSD-PC, C-GSA, and PRACS have been shown to provide high numerical scores. Moreover, C-BDSD and C-GSA outperform the original BDSD and GSA methods. The segmentation algorithm helps to constrain the range of spatial distortion generation, leading to a considerable reduction in distortion. This reduction in distortion is reflected in the high numerical scores achieved by BDSD-PC, C-GSA, and PRACS, while C-BDSD and C-GSA outperform their original versions. Among the methods in this class, the PRACS method stands out as the most effective pansharpening method, as it obtains the best results in terms of the chosen quality indexes without reference. This method rewards the balance between detail enhancement and consistency with respect to the original images.

2) MRA-Based Methods: The MRA class comprises high-performance methods, including AWLP, MTF-GLP-HPM-H, and C-MTF-GLP-CBD. The MTF-GLP-HPM-H method is particularly effective in excluding haze from the pansharpening

TABLE IV
FR NUMERICAL SCORES FOR THE IKONOS DATASET

Algorithm	QNR	D_λ	D_s	FQNR	D_λ^F	D_s^F	HQNR	D_λ^H	D_s^H	RQNR	D_λ^R	D_s^R
EXP	0.8304	**0.0000**	0.1563	0.6984	0.0405	0.3032	0.7792	0.0405	0.1563	0.6510	0.0405	0.2901
BT-H	0.8463	0.0749	0.0822	0.9103	0.0668	0.0157	0.8564	0.0668	0.0822	0.8995	0.0668	0.0363
BDSD	0.8988	0.0387	0.0617	0.8839	0.0902	0.0223	0.8580	0.0902	0.0617	0.8821	0.0902	0.0202
C-BDSD	0.9385	0.0167	0.0392	0.8901	0.0901	0.0205	0.8740	0.0901	0.0392	0.8783	0.0901	0.0359
BDSD-PC	0.8992	0.0416	0.0675	0.8918	0.0894	0.0264	0.8458	0.0894	0.0675	0.8843	0.0894	0.0187
GS	0.9115	0.0278	0.0578	0.8019	0.1491	0.0566	0.8063	0.1491	0.0578	0.8451	0.1491	0.0073
GSA	0.8344	0.0659	0.1036	0.9262	0.0652	**0.0045**	0.8356	0.0652	0.1036	0.9318	0.0652	**0.0008**
C-GSA	0.8930	0.0734	0.0370	0.8364	0.0635	0.1018	0.9004	0.0635	0.0370	0.9342	0.0635	0.0014
PRACS	0.9004	0.0381	0.0622	0.9002	0.0576	0.0423	0.8834	0.0576	0.0622	0.9265	0.0576	0.0164
AWLP	0.8142	0.0844	0.0933	0.9631	0.0286	0.0116	0.8732	0.0286	0.0933	0.9437	0.0286	0.0319
MTF-GLP	0.8235	0.0868	0.0978	0.9518	0.0301	0.0198	0.8768	0.0301	0.0978	0.9397	0.0301	0.0319
MTF-GLP-FS	0.8441	0.0746	0.0916	**0.9653**	0.0289	0.0094	0.8821	0.0289	0.0916	0.9431	0.0289	0.0295
MTF-GLP-HPM	0.8318	0.0856	0.0905	0.9507	0.0268	0.0217	0.8911	0.0268	0.0905	0.9315	0.0268	0.0403
MTF-GLP-HPM-H	0.8458	0.0801	0.0751	0.9482	0.0249	0.0256	0.8967	0.0249	0.0751	0.9358	0.0249	0.0376
MTF-GLP-HPM-R	0.8557	0.0654	0.0848	0.9598	0.0291	0.0134	0.8864	0.0291	0.0848	0.9332	0.0291	0.0381
MTF-GLP-CBD	0.8479	0.0768	0.0918	0.9619	0.0295	0.0104	0.8835	0.0295	0.0918	0.9443	0.0295	0.0277
C-MTF-GLP-CBD	0.9490	0.0171	0.0342	0.8959	0.0317	0.0761	0.9386	0.0317	0.0342	0.9268	0.0317	0.0508
MF	0.8690	0.0622	0.0730	0.9258	0.0425	0.0334	0.8890	0.0425	0.0730	0.9035	0.0425	0.0601
FE-HPM	0.8647	0.0622	0.0801	0.9473	0.0220	0.0282	0.8879	0.0220	0.0801	0.9323	0.0220	0.0399
SR-D	0.9038	0.0487	0.0529	0.8928	**0.0186**	0.0869	0.9328	**0.0186**	0.0529	0.8382	**0.0186**	0.1613
PWMBF	0.8165	0.0884	0.1058	0.9006	0.0830	0.0169	0.8128	0.0830	0.1058	0.9035	0.0830	0.0148
TV	0.9507	0.0172	0.0310	0.9104	0.0483	0.0425	0.9268	0.0483	0.0310	0.9177	0.0483	0.0386
RR	0.4684	0.3186	0.3069	0.7124	0.0794	0.2340	0.5314	0.0794	0.3069	0.7829	0.0794	0.1503
PNN	0.9258	0.0265	0.0510	0.9386	0.0375	0.0254	0.9143	0.0375	0.0510	**0.9489**	0.0375	0.0140
PNN-IDX	0.9289	0.0494	**0.0245**	0.9057	0.0611	0.0360	0.9182	0.0611	**0.0245**	0.9278	0.0611	0.0145
A-PNN	0.9591	0.0170	0.0521	0.9227	0.0304	0.0506	**0.9399**	0.0304	0.0521	0.9469	0.0304	0.0310
A-PNN-FT	**0.9616**	0.0158	0.0337	0.9174	0.0319	0.0627	0.9387	0.0319	0.0337	0.9376	0.0319	0.0315

Bold indicates the best overall results, and blue color indicates the best results for each category.

TABLE V
FR NUMERICAL SCORES FOR THE QUICKBIRD DATASET

Algorithm	QNR	D_λ	D_s	FQNR	D_λ^F	D_s^F	HQNR	D_λ^H	D_s^H	RQNR	D_λ^R	D_s^R
EXP	0.8497	**0.0000**	0.0850	0.6685	0.0520	0.1617	0.7308	0.0520	0.0850	0.8082	0.0520	0.0471
BT-H	0.8679	0.1283	0.2275	0.8064	0.1472	0.0742	0.8513	0.1472	0.2275	0.8528	0.1472	**0.0003**
BDSD	0.8771	0.0672	0.1780	0.8148	0.1312	0.0586	0.8570	0.1312	0.1780	0.8847	0.1312	0.0009
C-BDSD	0.8845	0.0386	0.1234	0.8261	0.1298	0.0575	0.8618	0.1298	0.1234	0.8924	0.0898	0.0004
BDSD-PC	0.8668	0.0699	0.1802	0.8500	0.0897	0.0587	0.8571	0.0897	0.1802	0.8873	0.0897	0.0009
GS	0.8269	0.0933	0.2045	0.8133	0.0983	0.0700	0.8547	0.0983	0.2045	0.8743	0.0983	0.0005
GSA	0.8658	0.0701	0.1800	0.8358	0.0813	0.0662	0.8514	0.0813	0.1800	0.8123	0.0813	0.0021
C-GSA	0.8099	0.1049	0.2170	0.8356	0.0823	0.0661	0.8541	0.0823	0.2170	0.8786	0.0823	0.0007
PRACS	0.8192	0.0514	0.1387	0.8369	0.0826	0.0656	0.8641	0.0826	0.1387	0.8803	0.0826	0.0006
AWLP	0.8197	0.1232	0.2130	**0.9073**	0.0211	0.0639	0.8602	0.0211	0.2130	0.9339	0.0211	0.0069
MTF-GLP	0.7792	0.1354	0.2256	0.8804	0.0264	0.0660	0.8578	0.0264	0.2256	0.9129	0.0264	0.0068
MTF-GLP-FS	0.7729	0.0813	0.1631	0.8734	0.0215	0.0882	0.8660	0.0215	0.1631	**0.9589**	0.0215	0.0069
MTF-GLP-HPM	0.7714	0.0831	0.1633	0.8736	0.0213	0.0661	0.8662	0.0213	0.1633	0.9570	0.0213	0.0080
MTF-GLP-HPM-H	0.8035	0.1273	0.1999	0.8945	0.0204	0.0647	0.8605	0.0204	0.1999	0.9156	0.0204	0.0094
MTF-GLP-HPM-R	0.8007	0.1359	0.2226	0.8854	0.0260	0.0685	0.8583	0.0260	0.2226	0.9093	0.0260	0.0091
MTF-GLP-CBD	0.7986	0.0826	0.1655	0.8730	0.0273	0.0674	0.8657	0.0273	0.1655	0.9574	0.0273	0.0066
C-MTF-GLP-CBD	0.7939	0.0804	0.1433	0.8752	0.0220	0.0695	0.8690	0.0220	0.1433	0.9541	0.0220	0.0119
MF	0.7889	0.1311	0.2141	0.8473	0.3068	0.0688	0.8400	0.3068	0.2141	0.8582	0.3068	0.0104
FE-HPM	0.6854	0.1321	0.2176	0.8669	0.0282	0.0698	0.8566	0.0282	0.2176	0.8911	0.0282	0.0078
SR-D	**0.9179**	0.0346	**0.0492**	0.8764	**0.0152**	0.0982	**0.9065**	**0.0152**	**0.0492**	0.9150	**0.0152**	0.0103
PWMBF	0.8778	0.0928	0.1444	0.8644	0.0317	**0.0564**	0.8466	0.0317	0.1444	0.8367	0.0317	0.0135

Bold indicates the best overall results, and blue color indicates the best results for each category.

TABLE VI
FR NUMERICAL SCORES FOR THE GAOFEN-1 DATASET

Algorithm	QNR	D_λ	D_s	FQNR	D_λ^F	D_s^F	HQNR	D_λ^H	D_s^H	RQNR	D_λ^R	D_s^R
EXP	0.8466	**0.0000**	0.1331	0.6047	0.3409	0.1711	0.6369	0.3409	0.1331	0.6078	0.3409	0.1569
BT-H	0.8809	0.1196	0.1965	0.7929	0.1264	0.1322	0.8651	0.1264	0.1965	0.9044	0.1264	**0.0344**
BDSD	0.9594	0.1198	0.0987	0.6504	0.1497	0.1429	0.7378	0.1497	0.0987	0.8732	0.1497	0.1053
C-BDSD	0.9286	0.1282	0.1048	0.7635	0.1489	0.0833	0.8199	0.1489	0.1048	0.7597	0.1489	0.1964
BDSD-PC	0.9125	0.1562	0.1692	0.6977	0.1220	0.1208	0.8323	0.1220	0.1692	0.8239	0.1220	0.0731
GS	0.9095	0.1378	0.1467	0.7925	0.1145	0.1000	0.8469	0.1145	0.1467	0.8590	0.1145	0.0380
GSA	0.9524	0.0933	0.1341	0.7649	0.1309	0.1355	0.8594	0.1309	0.1341	0.8612	0.1309	0.0367
C-GSA	0.8929	0.1518	0.1498	0.7653	0.1121	0.1483	0.8295	0.1121	0.1498	0.8981	0.1121	0.0524
PRACS	0.9524	0.0580	0.1344	0.8261	0.1058	0.1176	0.9425	0.1058	0.1344	0.9013	0.1058	0.0861
AWLP	0.9182	0.1138	0.0966	0.8694	0.0831	0.0752	0.9481	**0.0531**	0.0966	0.9165	0.0831	0.0436
MTF-GLP	0.8868	0.1612	0.1365	0.8204	0.0831	0.0915	0.8981	0.0831	0.1365	0.8962	0.0831	0.0894
MTF-GLP-FS	0.8749	0.1735	0.1031	0.8682	0.1120	0.1011	0.9094	0.0920	0.1031	0.9119	0.1120	0.0657
MTF-GLP-HPM	0.9117	0.1568	0.1053	0.8288	0.0845	0.1056	0.9195	0.0845	0.1053	0.9269	0.0845	0.0463
MTF-GLP-HPM-H	0.8155	0.1755	0.0796	0.8060	0.1068	0.1101	0.9080	0.0968	0.0796	0.9253	0.1068	0.0476
MTF-GLP-HPM-R	0.8742	0.1331	0.1890	0.8463	0.0836	0.1330	0.8692	0.0836	0.1890	0.8774	0.0836	0.0635
MTF-GLP-CBD	0.8974	0.1206	0.0993	0.8453	0.1138	0.1010	0.9308	0.0638	0.0993	0.9286	0.1138	0.0364
C-MTF-GLP-CBD	**0.9623**	0.1099	**0.0655**	0.8808	0.1106	0.1245	**0.9538**	0.0611	**0.0655**	**0.9679**	0.0706	0.0407
MF	0.8710	0.1171	0.1612	0.8647	0.1165	0.1055	0.8913	0.1065	0.1612	0.8805	0.1165	0.0834
FE-HPM	0.8819	0.1686	0.1766	0.8304	0.1046	0.1437	0.8903	0.1046	0.1766	0.9310	0.1046	0.1029
SR-D	0.9479	0.0609	0.1160	**0.9090**	**0.0690**	**0.0539**	0.8901	0.0890	0.1160	0.9563	**0.0690**	0.0361
PWMBF	0.8992	0.1125	0.1250	0.8773	0.0703	0.1010	0.8369	0.1603	0.1250	0.8710	0.0703	0.0727

Bold indicates the best overall results, and blue color indicates the best results for each category.

TABLE VII
FR NUMERICAL SCORES FOR THE WORLDVIEW-2 DATASET

Algorithm	QNR	D_λ	D_s	FQNR	D_λ^F	D_s^F	HQNR	D_λ^H	D_s^H	RQNR	D_λ^R	D_s^R
EXP	0.8053	**0.0000**	0.0946	0.5802	0.3061	0.1395	0.6319	0.3061	0.0946	0.6029	0.3061	0.1274
BT-H	0.8199	0.1115	0.1394	0.7402	0.1018	0.1065	0.8154	0.1018	0.1394	0.8579	0.1018	0.0030
BDSD	0.8947	0.1042	0.0923	0.6121	0.1534	0.0875	0.7137	0.1534	0.0923	0.6708	0.1534	0.0651
C-BDSD	**0.9530**	0.0630	0.0895	0.7069	0.1248	0.0547	0.7784	0.1248	0.0895	0.7060	0.1248	0.1560
BDSD-PC	0.8921	0.1009	0.1067	0.6829	0.1203	0.1185	0.7982	0.1203	0.1067	0.7708	0.1203	0.0526
GS	0.9089	0.0852	0.1294	0.7376	0.1052	0.0996	0.8145	0.1052	0.1294	0.8472	0.1052	0.0028
GSA	0.8914	0.0804	0.1214	0.7093	0.1150	0.0989	0.7944	0.1150	0.1214	0.8178	0.1150	0.0046
C-GSA	0.8608	0.0930	0.1263	0.7432	0.1039	0.0927	0.8177	0.1039	0.1263	0.8509	0.1039	0.0068
PRACS	0.9245	0.0683	0.0943	0.8078	0.0861	0.0645	0.9073	0.0861	0.0943	0.8883	0.0861	**0.0021**
AWLP	0.8850	0.0873	0.1140	0.8312	0.0802	0.0588	0.8888	0.0802	0.1140	0.9003	0.0802	0.0309
MTF-GLP	0.8237	0.1160	0.1310	0.8061	0.0825	0.0905	0.8680	0.0825	0.1310	0.8937	0.0825	0.0308
MTF-GLP-FS	0.8490	0.1057	0.1244	0.8206	0.0786	0.0854	0.8845	0.0786	0.1244	0.9073	0.0786	0.0278
MTF-GLP-HPM	0.8634	0.1001	0.1204	0.8222	0.0798	0.0793	0.8858	0.0798	0.1204	0.9010	0.0798	0.0322
MTF-GLP-HPM-H	0.8012	0.1249	0.1386	0.7996	0.0844	0.0911	0.8564	0.0844	0.1386	0.8875	0.0844	0.0322
MTF-GLP-HPM-R	0.8382	0.1111	0.1265	0.7961	0.0882	0.0832	0.8581	0.0882	0.1265	0.8718	0.0882	0.0390
MTF-GLP-CBD	0.8529	0.1041	0.1232	0.8214	0.0786	0.0845	0.8856	0.0786	0.1232	**0.9076**	0.0786	0.0208
C-MTF-GLP-CBD	0.9491	0.0674	0.0887	0.8259	0.0800	0.0647	0.9049	0.0800	0.0887	0.9015	0.0800	0.0313
MF	0.8389	0.1101	0.1255	0.7994	0.0890	0.0766	0.8563	0.0890	0.1255	0.8660	0.0890	0.0440
FE-HPM	0.8463	0.1071	0.1240	0.8087	0.0846	0.0797	0.8695	0.0846	0.1240	0.8832	0.0846	0.0381
SR-D	0.9253	0.0309	**0.0390**	**0.8812**	**0.0697**	**0.0276**	**0.9694**	**0.0697**	**0.0390**	0.8977	**0.0697**	0.0372
PWMBF	0.8750	0.0735	0.0825	0.8759	0.0922	0.0517	0.8213	0.0922	0.1825	0.8566	0.0922	0.0529
TV	0.9580	0.0282	0.0227	0.8593	0.0964	0.0645	0.9491	0.0964	0.0227	0.9021	0.0868	0.0352

Bold indicates the best overall results, and blue color indicates the best results for each category.

TABLE VIII
FR NUMERICAL SCORES FOR THE WORLDVIEW-4 DATASET

Algorithm	QNR	D_λ	D_s	FQNR	D_λ^F	D_s^F	HQNR	D_λ^H	D_s^H	RQNR	D_λ^R	D_s^R
EXP	0.8441	**0.0000**	0.0656	0.6746	0.2197	0.1321	0.7315	0.2197	0.0656	0.6991	0.2197	0.1018
BT-H	0.8734	0.0765	0.0553	0.8249	0.1184	0.0385	0.8336	0.1184	0.0553	0.8722	0.1065	0.0088
BDSD	0.9017	0.0606	0.0503	0.8539	0.1076	0.0374	0.8670	0.1076	0.0503	0.9046	0.0969	0.0038
C-BDSD	**0.9674**	0.0310	0.0351	0.8792	0.0970	0.0398	0.9214	0.0970	0.0351	0.9229	0.0873	0.0181
BDSD-PC	0.8969	0.0623	0.0515	0.8541	0.1075	0.0377	0.8648	0.1075	0.0515	0.9051	0.0967	0.0037
GS	0.8560	0.0858	0.0588	0.8225	0.1192	0.0397	0.8220	0.1192	0.0588	0.8758	0.1072	0.0007
GSA	0.8997	0.0687	0.0483	0.7342	0.1523	0.0443	0.7663	0.1523	0.0483	0.7845	0.1370	0.0035
C-GSA	0.8420	0.0923	0.0618	0.8268	0.1184	0.0360	0.8186	0.1184	0.0618	0.8778	0.1066	**0.0006**
PRACS	0.9538	0.0379	0.0379	0.8793	0.0943	0.0496	0.9224	0.0943	0.0379	0.9247	0.0849	0.0256
AWLP	0.9136	0.0767	0.0411	**0.9432**	0.0735	0.0357	0.9680	0.0735	0.0411	**0.9635**	0.0735	0.0429
MTF-GLP	0.8720	0.0998	0.0486	0.9270	0.0800	0.0356	0.9345	0.0800	0.0486	0.9486	0.0800	0.0408
MTF-GLP-FS	0.9008	0.0835	0.0432	0.9251	0.0794	0.0400	0.9479	0.0794	0.0432	0.9501	0.0794	0.0401
MTF-GLP-HPM	0.9030	0.0819	0.0429	0.9245	0.0797	0.0399	0.9483	0.0797	0.0429	0.9330	0.0797	0.0676
MTF-GLP-HPM-H	0.8992	0.0833	0.0439	0.9263	0.0791	0.0389	0.9491	0.0791	0.0439	0.9279	0.0791	0.0691
MTF-GLP-HPM-R	0.8745	0.0982	0.0482	0.9236	0.0814	**0.0352**	0.9323	0.0814	0.0482	0.9208	0.0814	0.0784
MTF-GLP-CBD	0.9008	0.0833	0.0433	0.9246	0.0796	0.0401	0.9475	0.0796	0.0433	0.9499	0.0796	0.0400
C-MTF-GLP-CBD	0.9298	0.0715	0.0365	0.9277	0.0764	0.0454	0.9677	0.0764	0.0365	0.9508	0.0764	0.0486
MF	0.9226	0.0990	0.0502	0.9116	0.0856	0.0366	0.9183	0.0856	0.0502	0.9190	0.0856	0.0645
FE-HPM	0.9475	0.1001	0.0489	0.9098	0.0865	0.0358	0.9193	0.0865	0.0489	0.9146	0.0778	0.0687
SR-D	0.9468	0.0336	**0.0201**	0.9276	**0.0627**	0.0862	**0.9683**	**0.0627**	**0.0201**	0.9459	**0.0565**	0.0960
PWMBF	0.9179	0.0918	0.0492	0.8612	0.1034	0.0427	0.8790	0.1034	0.0492	0.8980	0.0931	0.0268
TV	0.9580	0.0282	0.0227	0.8593	0.0964	0.0645	0.9491	0.0964	0.0227	0.9021	0.0868	0.0452

Bold indicates the best overall results, and blue color indicates the best results for each category.

TABLE IX
FR NUMERICAL SCORES FOR THE WORLDVIEW-3 DATASET

Algorithm	QNR	D_λ	D_s	FQNR	D_λ^F	D_s^F	HQNR	D_λ^H	D_s^H	RQNR	D_λ^R	D_s^R
EXP	0.8818	**0.0000**	0.1182	0.6739	0.0662	0.2747	0.8261	0.0662	0.1182	0.8105	0.0662	0.1392
BT-H	0.8802	0.0569	0.0669	0.8493	0.1040	0.0598	0.8403	0.1040	0.0669	0.9003	0.1040	0.0006
BDSD	0.9348	0.0235	0.0480	0.7749	0.1604	0.0747	0.8043	0.1604	0.0480	0.7794	0.1604	0.0649
C-BDSD	0.9413	0.0266	0.0392	0.7959	0.1438	0.0754	0.8057	0.1438	0.0392	0.7524	0.1438	0.1048
BDSD-PC	**0.9421**	0.0261	**0.0282**	0.7658	0.1562	0.0847	0.8179	0.1562	**0.0282**	0.7886	0.1562	0.0639
GS	0.9301	0.0121	0.0564	0.8104	0.1139	0.0758	0.8415	0.1139	0.0564	0.8854	0.1139	**0.0005**
GSA	0.8959	0.0474	0.0628	0.8385	0.1014	0.0656	0.8476	0.1014	0.0628	0.8958	0.1014	0.0018
C-GSA	0.9314	0.0296	0.0450	0.8524	0.0947	0.0592	0.8657	0.0947	0.0450	0.9029	0.0947	0.0019
PRACS	0.9317	0.0231	0.0459	0.8473	0.0831	0.0763	0.8769	0.0831	0.0459	0.9128	0.0831	0.0060
AWLP	0.8128	0.0461	0.0430	0.9266	0.0237	0.0478	0.9288	0.0237	0.0430	**0.9318**	0.0237	0.0319
MTF-GLP	0.8783	0.0679	0.0973	0.9306	0.0305	0.0418	0.9088	0.0305	0.0973	0.9304	0.0305	0.0410
MTF-GLP-FS	0.8843	0.0606	0.0907	0.9309	0.0308	0.0392	0.9136	0.0308	0.0907	0.9314	0.0308	0.0392
MTF-GLP-HPM	0.8803	0.0614	0.0913	0.9513	0.0328	0.0354	0.9088	0.0328	0.0913	0.9232	0.0328	0.0471
MTF-GLP-HPM-H	0.8370	0.0915	0.0746	0.9472	0.0309	0.0412	0.8890	0.0309	0.0746	0.9208	0.0309	0.0518
MTF-GLP-HPM-R	0.8910	0.0559	0.0839	**0.9598**	0.0328	0.0354	0.8892	0.0328	0.0839	0.9254	0.0328	0.0438
MTF-GLP-CBD	0.8896	0.0579	0.0945	0.9593	0.0310	0.0403	0.8827	0.0310	0.0945	0.9316	0.0310	0.0384
C-MTF-GLP-CBD	0.9329	0.0378	0.0339	0.8987	0.0362	0.0541	0.8837	0.0362	0.0339	0.9316	0.0362	0.0397
MF	0.8906	0.0655	0.0491	0.9412	0.0336	0.0278	0.9178	0.0336	0.0491	0.9038	0.0336	0.0649
FE-HPM	0.8854	0.0607	0.0583	0.9360	0.0328	0.0294	0.9127	0.0328	0.0583	0.9189	0.0328	0.0508
SR-D	0.9173	0.0433	0.0451	0.9516	**0.0188**	0.0318	**0.9422**	**0.0188**	0.0451	0.8691	**0.0188**	0.1145
PWMBF	0.8076	0.1107	0.0972	0.8654	0.0652	0.0759	0.8493	0.0652	0.0972	0.8947	0.0652	0.0461
TV	0.9407	0.0231	0.0379	0.9528	0.0329	**0.0138**	0.9315	0.0329	0.0379	0.9170	0.0329	0.0509
PNN	0.8967	0.0568	0.0516	0.8410	0.0597	0.1047	0.8918	0.0597	0.0516	0.8651	0.0597	0.0791
PNN-IDX	0.7946	0.0904	0.1268	0.7732	0.0894	0.1564	0.7992	0.0894	0.1268	0.7487	0.0894	0.1870
A-PNN	0.8687	0.0637	0.0731	0.8348	0.0398	0.1305	0.8910	0.0398	0.0731	0.8518	0.0398	0.1125
A-PNN-FT	0.9004	0.0580	0.0435	0.8431	0.1068	0.0549	0.8533	0.1068	0.0435	0.8419	0.1068	0.0562

Bold indicates the best overall results, and blue color indicates the best results for each category.

Fig. 6. FR fusion outcomes of a close-up view of the IKONOS exampled image, using a true-color representation. The LR-MS has a dimension size of 256×256×4, the HR-PAN is 1024 × 1024, while the outcomes are 1024×1024×4. The images and the quality predicted by QNR are as follows: BT-H: 0.7292; BDSD: 0.8217; C-BDSD: 0.9019; BDSD-PC: 0.8156; GS: 0.7788; GSA: 0.6703; C-GSA: 0.7110; PRACS: 0.8574; AWLP: 0.6543; MTF-GLP: 0.6417; MTF-GLP-FS: 0.6895; MTF-GLP-HPM: 0.6456; MTF-GLP-HPM-H: 0.7196; MTF-GLP-HPM-R: 0.6936; MTF-GLP-CBD: 0.6893; C-MTF-GLP-CBD: 0.7441; MF: 0.8441; FE-HPM: 0.7022; SR-D: 0.7973; PWMBF: 0.6509; TV: 0.9385; RR: 0.7186; PNN: 0.8335; PNN-IDX: 0.8374; A-PNN: 0.8668; and A-PNN-FT: 0.8649.

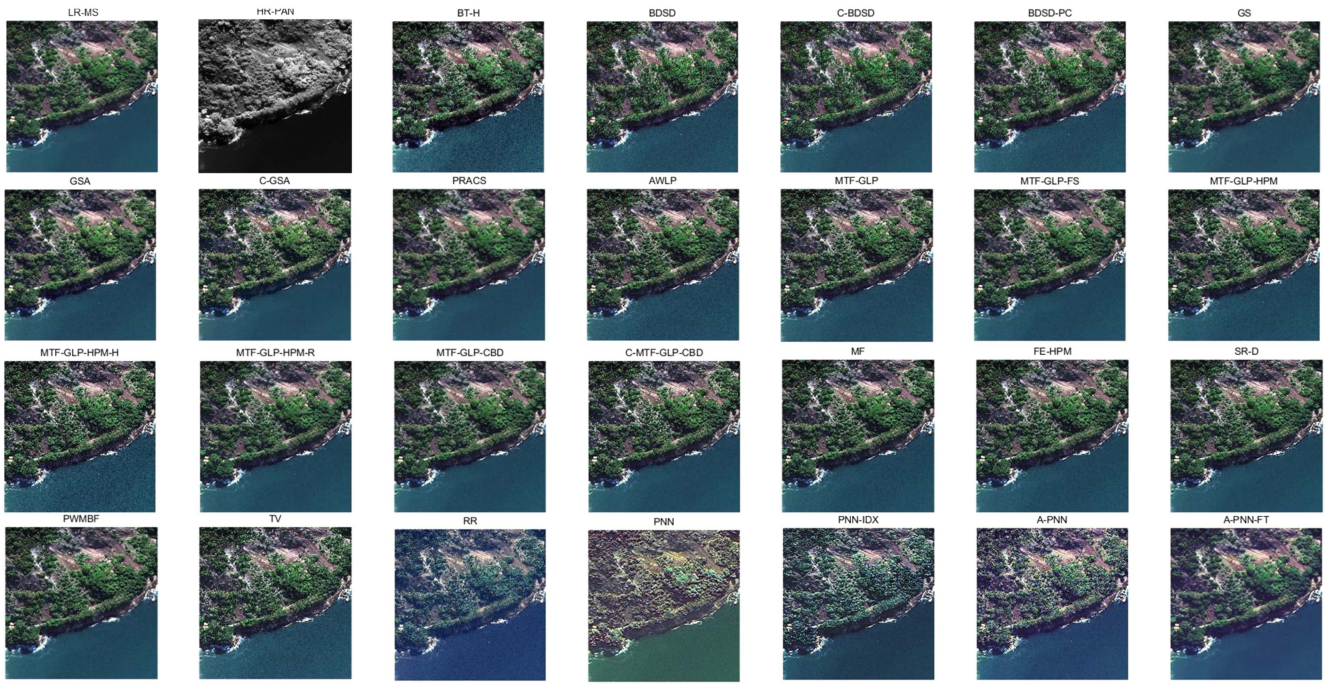

Fig. 7. FR fusion outcomes of a close-up view of the WorldView-3 exampled image, using a true-color representation. The LR-MS has a dimension size of 256×256×8, the HR-PAN is 1024 × 1024, while the outcomes are 1024×1024×8. The images and the quality predicted by QNR are as follows: BT-H: 0.5797; BDSD: 0.7973; C-BDSD: 0.9033; BDSD-PC: 0.7410; GS: 0.8109; GSA: 0.8081; C-GSA: 0.7880; PRACS: 0.8769; AWLP: 0.6990; MTF-GLP: 0.6972; MTF-GLP-FS: 0.8033; MTF-GLP-HPM: 0.7084; MTF-GLP-HPM-H: 0.6772; MTF-GLP-HPM-R: 0.8019; MTF-GLP-CBD: 0.8078; C-MTF-GLP-CBD: 0.7661; MF: 0.7637; FE-HPM: 0.7693; SR-D: 0.9086; PWMBF: 0.7141; TV: 0.7384; RR: 0.6220; PNN: 0.6704; PNN-IDX: 0.7742; A-PNN: 0.7890; and A-PNN-FT: 0.7938.

process, resulting in significant noise reduction. On the other hand, the C-MTF-GLP-CBD method leverages a context-based decision with a regression-based injection model that utilizes local parameter estimation and clustering. The effective learning-based techniques employed in C-MTF-GLP-CBD contribute to reducing distortion in the pansharpening process.

3) VO-Based Methods: The VO-based class includes high-performance methods, with SR-D demonstrating superior performance over all other VO techniques for six datasets. The TV method is particularly effective for the WorldView-3 dataset, thanks to its spatial regularization strategy. However, since GaoFen-1 and QuickBird lack TV and RR model parameters, no experiments were conducted for these datasets. The FE-HPM and PWMBF techniques generally fall behind these algorithms in terms of performance. In comparison to the other class, VO-based methods exhibit less robustness.

4) ML-Based Methods: The performance of ML-based methods at FR is also satisfactory, especially when considering the HQNR metric, which summarizes optimal spatial and spectral quality. However, lower performance is observed on the WorldView-3 dataset. When analyzing the IKONOS data, a progressive performance gain is observed when moving from the basic model PNN to the most advanced model, A-PNN-FT. PNN, PNN-IDX, and A-PNN show a performance drop in WorldView-3, which may be due to a larger statistical shift between the training and test datasets in this case. The significant performance gain of A-PNN-FT over A-PNN on WorldView-3 confirms this intuition.

However, the good performance of A-PNN-FT comes at the cost of additional computational efforts required for a supplemental training phase. Due to the lack of model parameters, experiments were only conducted on IKONOS and WorldView-3.

Based on our analysis of each database, the VO-based methods, particularly the SR-D method, demonstrate the best and most stable performance, with outstanding results on WorldView-2, WorldView-3, WorldView-4, QuickBird, and GaoFen-1. Although we only conducted experiments on IKONOS due to the lack of model parameters, we predict that the ML method can also perform well using effective DL methods. In addition, the MRA-based and CS-based methods exhibit strong performance on these databases. Moreover, the overall performance for the eight-band WorldView-2 and WorldView-3 satellite images is relatively poorer than that of the four-band IKONOS, QuickBird, GaoFen-1, and WorldView-4 satellite images. The lower performance of the eight-band WorldView-2 and WorldView-3 satellite images may indicate that the increased spectral diversity poses a challenge to PAN sharpening and increases the likelihood of distortion.

V. Conclusion

This article presented a comprehensive study of FR-PIQA methods. We proposed a classification approach for FR methods based on feature extraction and analyzed how different techniques extract spatial and spectral features from pansharpened images without reference to HR-MS images. In addition, we provided a summary of strategies to measure FR performance in the absence of ground truth data, including both subjective and objective strategies. Furthermore, we conducted a board analysis on a large-scale public pansharpened database with unified measuring criteria. This analysis demonstrated the strengths and limitations of different FR protocols through qualitative and quantitative analysis from a statistical perspective. Looking to the future, we anticipate the development of ground truth generation methods that should neither be limited to RR indicators nor dependent on human perception. We believe that techniques such as transfer learning and unsupervised learning can facilitate the creation of fair ground truth data. By leveraging DL technology and more reliable ground truth, it is possible to develop better FR-PIQA methods that are more effective at feature extraction and generalization.

References

[1] H. Ghassemian, "A review of remote sensing image fusion methods," *Inf. Fusion*, vol. 32, pp. 75–89, 2016.

[2] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "25 years of pansharpening: A critical review and new developments," in *Signal and Image Processing for Remote Sensing*. Boca Raton, FL, USA: CRC Press, 2012, pp. 533–548.

[3] L. Alparone, S. Baronti, B. Aiazzi, and A. Garzelli, *Remote Sensing Image Fusion*. Boca Raton, FL, USA: CRC Press, 2015.

[4] G. Vivone et al., "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015.

[5] S. Lolli, L. Alparone, A. Garzelli, and G. Vivone, "Haze correction for contrast-based multispectral pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2255–2259, Dec. 2017.

[6] A. Garzelli, F. Nencini, and L. Capobianco, "Optimal MMSE pan sharpening of very high resolution multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 228–236, Jan. 2008.

[7] A. Garzelli, "Pansharpening of multispectral images based on nonlocal parameter optimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2096–2107, Apr. 2015.

[8] G. Vivone, "Robust band-dependent spatial-detail approaches for panchromatic sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6421–6433, Sep. 2019.

[9] C. A. Laben and B. V. Brower, "Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening," US Patent 6 011 875, Jan. 2000.

[10] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS +pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Oct. 2007.

[11] R. Restaino, M. Dalla Mura, G. Vivone, and J. Chanussot, "Context-adaptive pansharpening based on image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 753–766, Feb. 2017.

[12] J. Choi, K. Yu, and Y. Kim, "A new adaptive component-substitution-based satellite image fusion by using partial replacement," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 295–309, Jan. 2011.

[13] P. Chavez, S. C. Sides, and J. A. Anderson, "Comparison of three different methods to merge multiresolution and multispectral data—LANDSAT TM and SPOT panchromatic," *Photogrammetric Eng. Remote Sens.*, vol. 57, no. 3, pp. 295–303, 1991.

[14] W. Carper, T. Lillesand, and R. W. Kiefer, "The use of intensity-hue-saturation transformations for merging SPOT panchromatic and multispectral image data," *Photogrammetric Eng. Remote Sens.*, vol. 56, no. 4, pp. 459–467, 1990.

[15] T.-M. Tu, S.-C. Su, H.-C. Shyu, and P. S. Huang, "A new look at IHS-like image fusion methods," *Inf. Fusion*, vol. 2, no. 3, pp. 177–186, 2001.

[16] H. Li, B. Manjunath, and S. Mitra, "Multisensor image fusion using the wavelet transform," *Graph. Models Image Process.*, vol. 57, no. 3, pp. 235–245, 1995.

[17] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. AC-31, no. 4, pp. 532–540, Apr. 1983.

[18] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "MTF-tailored multiscale fusion of high-resolution MS and pan imagery," *Photogrammetric Eng. Remote Sens.*, vol. 72, no. 5, pp. 591–596, 2006.

[19] G. Vivone, R. Restaino, and J. Chanussot, "Full scale regression-based injection coefficients for panchromatic sharpening," *IEEE Trans. Image Process.*, vol. 27, no. 7, pp. 3418–3431, Jul. 2018.

[20] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "An MTF-based spectral distortion minimizing model for pan-sharpening of very high resolution multispectral images of urban areas," in *Proc. 2nd GRSS/ISPRS Joint Workshop Remote Sens. Data Fusion Over Urban Areas*, 2003, pp. 90–94.

[21] G. Vivone, R. Restaino, and J. Chanussot, "A regression-based high-pass modulation pansharpening approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 984–996, Feb. 2018.

[22] L. Alparone, L. Wald, J. Chanussot, C. Thomas, P. Gamba, and L. M. Bruce, "Comparison of pansharpening algorithms: Outcome of the 2006 GRS-S data-fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3012–3021, Oct. 2007.

[23] R. Restaino, G. Vivone, M. Dalla Mura, and J. Chanussot, "Fusion of multispectral and panchromatic images based on morphological operators," *IEEE Trans. Image Process.*, vol. 25, no. 6, pp. 2882–2895, Jun. 2016.

[24] X. Meng, H. Shen, Q. Yuan, H. Li, L. Zhang, and W. Sun, "Pansharpening for cloud-contaminated very high-resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2840–2854, May 2019.

[25] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé, "A variational model for P+ XS image fusion," *Int. J. Comput. Vis.*, vol. 69, no. 1, pp. 43–58, 2006.

[26] G. Vivone et al., "Pansharpening based on semiblind deconvolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1997–2010, Apr. 2015.

[27] M. R. Vicinanza, R. Restaino, G. Vivone, M. D. Mura, and J. Chanussot, "A pansharpening method based on the sparse representation of injected details," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 180–184, Jan. 2015.

[28] F. Palsson, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Model-based fusion of multi- and hyperspectral images using PCA and wavelets," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2652–2663, May 2015.

[29] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "A new pansharpening algorithm based on total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 318–322, Jan. 2014.

[30] F. Palsson, M. O. Ulfarsson, and J. R. Sveinsson, "Model-based reduced-rank pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 4, pp. 656–660, Apr. 2020.

[31] M. O. Ulfarsson, F. Palsson, M. D. Mura, and J. R. Sveinsson, "Sentinel-2 sharpening using a reduced-rank method," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6408–6420, Sep. 2019.

[32] L.-J. Deng et al., "Machine learning in pansharpening: A benchmark, from shallow to deep networks," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 3, pp. 279–315, Sep. 2022.

[33] W. Huang, L. Xiao, Z. Wei, H. Liu, and S. Tang, "A new pan-sharpening method with deep neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1037–1041, May 2015.

[34] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "PanNet: A deep network architecture for pan-sharpening," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1753–1761.

[35] Q. Liu, H. Zhou, Q. Xu, X. Liu, and Y. Wang, "PSGAN: A generative adversarial network for remote sensing image pan-sharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10227–10242, Dec. 2021.

[36] Y. Qu, R. K. Baghbaderani, H. Qi, and C. Kwan, "Unsupervised pansharpening based on self-attention mechanism," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3192–3208, Apr. 2021.

[37] S. Luo, S. Zhou, Y. Feng, and J. Xie, "Pansharpening via unsupervised convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4295–4310, 2020.

[38] Q. Yuan, Y. Wei, X. Meng, H. Shen, and L. Zhang, "A multiscale and multidepth convolutional neural network for remote sensing imagery pan-sharpening," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 978–989, Mar. 2018.

[39] T.-J. Zhang, L.-J. Deng, T.-Z. Huang, J. Chanussot, and G. Vivone, "A triple-double convolutional neural network for panchromatic sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, doi: 10.1109/TNNLS.2022.3155655.

[40] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, 2016, Art. no. 594.

[41] G. Scarpa, S. Vitale, and D. Cozzolino, "Target-adaptive CNN-based pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5443–5457, Sep. 2018.

[42] F. Bovolo, L. Bruzzone, L. Capobianco, A. Garzelli, S. Marchesi, and F. Nencini, "Analysis of the effects of pansharpening in change detection on VHR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 1, pp. 53–57, Jan. 2010.

[43] X. Tang et al., "An unsupervised remote sensing change detection method based on multiscale graph convolutional network and metric learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5609715.

[44] C. Henry, S. M. Azimi, and N. Merkle, "Road segmentation in SAR satellite images with deep fully convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 12, pp. 1867–1871, Dec. 2018.

[45] C. Peng, Y. Li, L. Jiao, and R. Shang, "Efficient convolutional neural architecture search for remote sensing image scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 6092–6105, Jul. 2021.

[46] L. Wald, T. Ranchin, and M. Mangolini, "Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images," *Photogrammetric Eng. Remote Sens.*, vol. 63, no. 6, pp. 691–699, 1997.

[47] T. Ranchin, B. Aiazzi, L. Alparone, S. Baronti, and L. Wald, "Image fusion—The ARSIS concept and some successful implementation schemes," *ISPRS J. Photogrammetry Remote Sens.*, vol. 58, nos. 1/2, pp. 4–18, 2003.

[48] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, pp. 147–149.

[49] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[50] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[51] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?," in *Proc. 3rd Conf. Fusion Earth Data: Merging Point Meas. Raster Maps Remotely Sensed Images*, 2000, pp. 99–103.

[52] L. Zhang, H. Shen, W. Gong, and H. Zhang, "Adjustable model-based fusion method for multispectral and panchromatic images," *IEEE Trans. Systems, Man, Cybern. B, Cybern.*, vol. 42, no. 6, pp. 1693–1704, Dec. 2012.

[53] Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.*, vol. 9, no. 3, pp. 81–84, Mar. 2002.

[54] L. Alparone, S. Baronti, A. Garzelli, and F. Nencini, "A global quality measurement of pan-sharpened multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 4, pp. 313–317, Oct. 2004.

[55] A. Garzelli and F. Nencini, "Hypercomplex quality assessment of multi/hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 662–665, Oct. 2009.

[56] M. Selva, L. Santurri, and S. Baronti, "On the use of the expanded image in quality assessment of pansharpened images," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 320–324, Mar. 2018.

[57] A. Arienzo, G. Vivone, A. Garzelli, L. Alparone, and J. Chanussot, "Full-resolution quality assessment of pansharpening: Theoretical and hands-on approaches," *IEEE Geosci. Remote Sens. Mag.*, vol. 10, no. 3, pp. 168–201, Sep. 2022.

[58] L. Alparone, B. Aiazzi, S. Baronti, A. Garzelli, F. Nencini, and M. Selva, "Multispectral and panchromatic data fusion assessment without reference," *Photogrammetric Eng. Remote Sens.*, vol. 74, no. 2, pp. 193–200, 2008.

[59] M. M. Khan, L. Alparone, and J. Chanussot, "Pansharpening quality assessment using the modulation transfer functions of instruments," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3880–3891, Nov. 2009.

[60] L. Alparone, A. Garzelli, and G. Vivone, "Spatial consistency for full-scale assessment of pansharpening," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 5132–5134.

[61] B. Aiazzi, L. Alparone, S. Baronti, R. Carlà, A. Garzelli, and L. Santurri, "Full-scale assessment of pansharpening methods and data products," *Proc. SPIE*, vol. 9244, 2014, Art. no. 924402.

[62] R. Carlà, L. Santurri, B. Aiazzi, and S. Baronti, "Full-scale assessment of pansharpening through polynomial fitting of multiscale measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6344–6355, Dec. 2015.

[63] G. Palubinskas, "Joint quality measure for evaluation of pansharpening accuracy," *Remote Sens.*, vol. 7, no. 7, pp. 9292–9310, 2015.

[64] G. Vivone, R. Restaino, and J. Chanussot, "A Bayesian procedure for full-resolution quality assessment of pansharpened products," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4820–4834, Aug. 2018.
[65] G. Vivone, P. Addesso, and J. Chanussot, "A combiner-based full resolution quality assessment index for pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 437–441, Mar. 2019.
[66] O. A. Agudelo-Medina, H. D. Benitez-Restrepo, G. Vivone, and A. Bovik, "Perceptual quality assessment of pan-sharpened images," *Remote Sens.*, vol. 11, no. 7, 2019, Art. no. 877.
[67] X. Otazu, M. Gonzalez-Audicana, O. Fors, and J. Nunez, "Introduction of sensor spectral response into image fusion methods. application to wavelet-based methods," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 10, pp. 2376–2385, Oct. 2005.
[68] "A large-scale remote sensing database for subjective and objective quality assessment of pansharpened images," *J. Vis. Commun. Image Represent.*, vol. 73, 2020, Art. no. 102947.
[69] K. Bao, X. Meng, X. Chai, and F. Shao, "A blind full resolution assessment method for pansharpened images based on multistream collaborative learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 541031.
[70] X. Meng et al., "A blind full-resolution quality evaluation method for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5401916.
[71] B. Zhou, F. Shao, X. Meng, R. Fu, and Y.-S. Ho, "No-reference quality assessment for pansharpened images via opinion-unaware learning," *IEEE Access*, vol. 7, pp. 40388–40401, 2019.
[72] J. Zhou, D. L. Civco, and J. A. Silander, "A wavelet transform method to merge landsat TM and spot panchromatic data," *Int. J. Remote Sens.*, vol. 19, no. 4, pp. 743–757, 1998.
[73] C. Kwan, B. Budavari, A. C. Bovik, and G. Marchisio, "Blind quality assessment of fused WorldView-3 images by using the combinations of pansharpening and hypersharpening paradigms," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1835–1839, Oct. 2017.
[74] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.
[75] G. Palubinskas, "Quality assessment of pan-sharpening methods," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2014, pp. 2526–2529.
[76] N. Badal, R. Soundararajan, A. Garg, and A. Patil, "No reference pansharpened image quality assessment through deep feature similarity," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7235–7247, 2022.
[77] X. Meng et al., "A large-scale benchmark data set for evaluating pansharpening performance: Overview and implementation," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 1, pp. 18–52, Mar. 2021.
[78] G. Vivone et al., "A new benchmark based on recent advances in multispectral pansharpening: Revisiting pansharpening with classical and emerging pansharpening methods," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 1, pp. 53–81, Mar. 2021.

Xiaodi Guan received the B.S. degree in information engineering in 2019 from Xi'an Jiaotong University, Xi'an, China, where she is currently working toward the Ph.D. degree in information and communication engineering with the School of Information and Communications Engineering.

Her current research interests include visual quality assessment and enhancement.

Fan Li (Senior Member, IEEE) received the B.S. degree in information engineering and the Ph.D. degree in information and communication engineering from the School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an, China, in 2003 and 2010, respectively.

From 2017 to 2018, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA, USA. He is currently a Professor with the School of Information and Communications Engineering, Xi'an Jiaotong University. He has authored or coauthored more than 80 technical papers. His research interests include multimedia communication, image/video coding, and image/video quality assessment.

Xingliang Zhang received the B.S. degree in network engineering from Kashi University, Kashi, China, in 2022. He is currently working toward the master's degree with Xinjiang University, Urumqi, China.

His research interests include remote sensing quality assessment.

Mingyang Ma received the B.S. degree in communication engineering and the Ph.D. degree in information and communication engineering from Northwestern Polytechnical University, Xi'an, China, in 2015 and 2021, respectively.

He is currently an Assistant Professor and a Postdoctoral Researcher with the School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an. His main research interests include hyperspectral image processing and video content analysis.

Shaohui Mei (Senior Member, IEEE) received the B.S. degree in electronics and information engineering and the Ph.D. degree in signal and information processing from Northwestern Polytechnical University, Xi'an, China, in 2005 and 2011, respectively.

From 2007 to 2008, he was a Visiting Student with the University of Sydney, Camperdown, NSW, Australia. He is currently a Professor with the School of Electronics and Information, Northwestern Polytechnical University. His research interests include hyperspectral remote sensing image processing and applications, intelligent signal and information acquisition and processing, video processing, and pattern recognition.

Dr. Mei was a recipient of the Excellent Doctoral Dissertation Award of Shaanxi Province in 2014 and the Best Paper Award of 2017 IEEE International Symposium on Intelligent Signal Processing and Communication Systems. He was also awarded the Best Reviewer of IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS) in 2019 and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) in 2022. He is the Associate Editor for IEEE TGRS and IEEE JSTARS. He is a Reviewer for more than 20 international famous academic journals.