Rank Learning Based Full-Resolution Quality Evaluation Method for Pansharpened Images

Xiaodi Guan, Fan Li, *Senior Member, IEEE*, Haixia Bi, and Lijiao Gong, *Senior Member, IEEE*

***Abstract*—Full-resolution quality evaluation model for pansharpened images is significant for remote sensing applications, yet presents a challenge of the absence of reference compared with the reduced-resolution approach. To predict the image quality accurately, it is necessary to consider the distortion during the pansharpening process. Based on an observation that the quality of pairwise images can more easily be ranked, we propose a rank learning based full-resolution quality evaluation method for pansharpened images. Our approach begins with the synthesizing of ranked distortion images in spatial and spectral domains. Then, we develop a pansharpening distortion-perceiving model. This model employs spatial and spectral Siamese networks to perceive distortions and applies a pair-wise learning strategy for ranked images. Consequently, we establish a distortion-guided full-resolution quality evaluation framework for pansharpening. This framework integrates the spatial and spectral distortion-perceiving network and is enhanced with a dimension alignment module and a discrepancy Rrpresentation module, enabling effective distortion extraction among high-resolution multispectral, panchromatic, and low-resolution multispectral images. We conducted a series of experiments on a large-scale public pansharpened database. The experimental results demonstrate the effectiveness of our proposed approach.**

***Index Terms*—Full-resolution quality evaluation, pansharpened image, pansharpening, rank learning.**

I. Introduction

EFFICIENCY of the satellite systems is limited by the inadequate performance of hardware, for example, the constrained radiation energy, and the underdeveloped data transmission capabilities. Acquiring remote sensing images necessitates a strategic compromise between spatial and spectral resolution [1], [2], [3]. The majority of satellites nowadays provide low-resolution multispectral (LR-MS) imagery along with high-resolution panchromatic imagery (HR-Pan), as an alternative to high-resolution multispectral images (HR-MS). Panchromatic sharpening (pansharpening) emerges as a solution to this challenge [4]. By integrating the finer spatial resolution provided by Pan with the spectral information from MS, pansharpening methods combine the strengths of HR-Pan and LR-MS imagery and generate super-resolved HR-MS images.

Quality evaluation has played a crucial role in pansharpening research since its introduction. Developing a robust mathematical model can expedite the evaluation process of fused HR-MS images, thus enhancing the management and refinement of remote sensing systems. Primarily, pansharpening always serves as an initial phase for the further various applications, including change detection, segmentation, and scene classification [5], [6], [7]. The study by Bovolo et al. [6] highlighted that while pansharpened images lead to superior quality maps for change detection, the related artifacts hamper detecting accuracy. Second, from a system perspective, effective quality evaluation can help pursue optimal performance within bandwidth limitations. Therefore, selecting an effective pansharpening technique becomes crucial for optimizing the overall performance of remote sensing systems. To achieve this goal, tailored quality assessment methodologies must be meticulously designed.

Nevertheless, evaluating the quality of HR-MS imaging poses challenges for the intricacies of cross-modal combination and the lack of high-resolution references. To tackle the challenge, Wald et al. introduced two essential criteria for evaluating pansharpened images: Consistency and Synthesis [8], [9].

These criteria lead to two categories of quality evaluation approaches for pansharpening: Full-Resolution (FR) and Reduced-Resolution (RR). RR involves spatially degrading the original HR-Pan and LR-MS images using ideal filters, such as modulation transfer function-matched filters (MTF). Subsequently, the pansharpening is performed on the filtered MS and the filtered Pan, while the pristine MS is used to evaluate pansharpening quality as the reference imagery. Several RR approaches have been proposed. The full-reference image quality evaluating metrics, like structural similarity (SSIM) [10] and peak signal-to-noise ratio (PSNR) [11], are adopted as the RR for remote sensing quality evaluation. Special-designed methods are developed, such as SAM [12], ERGAS [13], CC [14], UIQI [15], Q_4[16], and $Q2^n$[17], which represents an extended multiband version of Q_4.

While LR-MS serves as a reference in RR evaluation, constraints arise due to degradation throughout the process.

Manuscript received 8 April 2024; revised 23 May 2024 and 10 June 2024; accepted 18 June 2024. Date of publication 24 June 2024; date of current version 8 July 2024. 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 National Natural Science Foundation of China under Grant 62071369 and Grant 42201394. *(Corresponding author: Haixia Bi.)*

Xiaodi Guan, Fan Li, and Haixia Bi 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; haixia.bi@xjtu.edu.cn).

Lijiao Gong is with the College of Mechanical and Electrical Engineering, Shihezi University, Shihezi 832003, China (e-mail: glj_mac@shzu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3418551

Selva et al. [18] highlighted the potential invalidity of the scale-invariant hypothesis, rendering it challenging to ensure that degraded pansharpened images maintain the quality of pristine ones. The implications of these limitations pose challenges in implementing RR evaluation. To improve the accuracy of quality assessment, numerous studies have redirected their focus toward FR evaluation. This method assesses the quality of pansharpened images at a high resolution, with no need for an HR-MS as reference and degrading during fusion.

Currently, quality without reference (QNR) protocol is widely recognized as the predominant FR evaluation protocol for pansharpening. The QNR [19] index assesses spectral distortion by comparing image quality index (QI) pre- and postfusion of multispectral bands, as well as spatial distortion by comparing QI values pre- and postfusion of each multispectral band with Pan. Over the years, various modifications of the QNR protocol, referred to as QNR-like protocols, have been introduced [20], [21], [22], [23]. Furthermore, alternative methods have emerged. Quality estimation by fitting [24] and its further version under Kalman filtering [25] was proposed. In addition, the joint quality measure and its variant version utilized SSIM and designed a novel similarity measurement, instead of the traditional QI [26], [27].

As machine learning and deep learning strategies advance, researchers are increasingly concentrating on extracting deep-level features based the AI methods for quality prediction [28], [29], [30], [31] and other remote sensing tasks [32], [33], [34], [35], [36], [37], [38], [39], [40], [41], [42]. Researchers often regressively predict quality scores by analyzing and comparing the features of Pan, LR-MS, and HR-MS imagery [29], [43], [44], [45], [46], [47]. Such deep learning methods have been proven effective. For general image quality assessment (IQA) tasks, direct training through widespread deep learning frameworks is no longer sufficient to meet quality assessment requirements. Most general IQA methods involve targeted feature mining and training strategies based on the characteristics of distortion information [30], [48], [49], [50]. They mainly focus on mimiking the human visual system (HVS) to achieve better quality perception. However, due to the absence of reference images, distortion information is difficult to filter out directly through a simple comparison. Consequently, current FR quality assessment methods based on deep learning lack targeted feature analyzing and utilization for distortion information in the panchromatic sharpening process.

To overcome the challenge of developing distortion features, we propose a full-resolution quality evaluation method for pansharpened images learning from rankings. We first design a generation strategy for multilevel panchromatic and multispectral images to artificially simulate the distortion during the pansharpening process. The various distorted ranks of images are generated by the strategy. With the generated ranked images, we then establish the spectral and spatial distortion-perceiving Siamese networks to specifically extract distortion features. Finally, based on the Siamese network branches, we construct a full-resolution quality evaluation architecture to predict the pansharpened image quality scores. A series of experiments demonstrate the effectiveness of our proposed method.

The main contributions of this article are as follows:

1) A Ranked Distortion Synthesizing Strategy for Pansharpening Has Been Designed: In the absence of reference images, extracting distortion information from pansharpened images poses a challenge. Through a detailed analysis of the pansharpening process, a set of distortion-generation strategies has been formulated. Artificial distortion synthesis not only facilitates the targeted extraction and utilization of distortion features for subsequent deep learning processes, but also expands the training dataset, thereby enhancing the robustness of deep models.

2) A Pansharpening Distortion-Perceiving Model is Developed Based on Rank Learning: It has been observed that directly assessing distortion is challenging, whereas comparing the quality of paired images is easier. Therefore, we have designed spatial and spectral Siamese network structures, as well as developed a pairwise learning method for ranked images to perceive the distortion. The method significantly reduces the difficulty of extracting pansharpening distortion features and improves the accuracy of feature representation.

3) We Propose a Distortion-Guided Full-Resolution Quality Evaluation Framework for Pansharpening: We construct the distortion-guided quality predicting architecture, which inherited the distortion-perceiving networks in the spatial and spectral domains. We designed a dimension alignment module, which enabled the distortion extraction between HR-MS and the original Pan, LR-MS. Finally, through the integration of discrepancy features, we achieved distortion-guided FR quality evaluation for pansharpening.

This article is organized as follows: Section I serves as the introduction to the entire work. Section II provides a summary of the related works on pansharpening algorithms and FR quality evaluation methods based on deep learning. Our proposed rank learning based FR method is detailed in Section III. Section IV presents the experimental results. Finally, Section V concludes this article.

II. Related Work

A. Pansharpening Algorithms

Over the years, a multitude of pansharpening algorithms have been proposed. Two categories are summarized in [51]: multiresolution analysis (MRA) and component substitution (CS). For MRA-based techniques, they first separate images into low- and high-frequency parts relying on wavelet transform [52], Laplacian pyramid [53] and so on MRA tools [54], [55]. Subsequently, reconstruct the HR-MS by combining the two parts. On the other hand, CS-based algorithms substitute the component of MS, then generate the HR-MS with inverse transformation process [56], [57], [58], [59].

Furthermore, researchers have proposed variational optimization (VO) approaches that leverage variational principles and optimization of energy function [60], [61], [62], [63]. Deep learning algorithms build intricate mappings between original and pansharpened images through deep neural networks (DNNs). With extensive image data training, these networks are capable of producing HR-MS imagery [64], [65], [66], [67], [68], [69], [70].

B. Deep Learning-Based FR Assessment for Pansharpened Images

Nowadays, FRQA approaches face two major challenges: 1) evaluate quality when a reference image is absent; 2) measure the effectiveness of the FR methods when ground truth is absent. For the deep learning-based FRQA approaches, the DNNs extract the distortion-relevant features to overcome the absence of reference. Existing methods adopt the dominant deep model, such as multivariate Gaussian (MVG) fitting, Siamese VGGnet, and ResNet architectures. These effective deep architectures also perform well on the FRQA task.

Facing the challenge of lacking ground-truth, researchers proposed two strategies. First, some researchers conducted subjective experiments to collect amounts of opinion scores and built the quality assessment database, similar to the common practices in the field of general image/video quality assessment. But the subjective data have not been public yet. Second, other researchers adopted the RR metrics as the proxy ground truth. The strategy can be regarded as the no-reference quality assessment task, and is easy to conduct. Thus, we also choose the proxy label strategy in our work.

Meng et al. [44] introduced a model that extracts features sensitive to distortions in both spectral and spatial domains. They utilized MVG for training on the extracted features. The model incorporates spatial features of HR-Pan and spectral invariant features of MS. Subsequently, the generated HR-MS is fed to a testing model. The final quality prediction is obtained by the discrepancy between the outputs of the two models.

In addition, researchers developed an opinion-aware methodology for evaluation in [43]. This approach extracts characteristics in the spectral domain and several classical metrics for MS data, which can comprehensively represent distortions. Then, researchers trained an MVG model based on a raw MS database. The method integrates both subjective and objective knowledge in quality assessment, ensuring a comprehensive assessment.

As techniques evolve, DNNs have progressed in autonomously extracting profound features pertinent to quality perception, improving quality evaluation. Researchers constructed a Siamese network to collectively learn the representations of pansharpening-relevant knowledge in [29]. Researchers utilized the Siamese architecture to directly extract the feature from Pan/Fused/MS images. Pan and fused images are fed into the spatial model. Then, the model extracted the feature of Pan and fused images and supervised by the collected spatial DMOS. For the spectral model, MS and fused images are fed into the model. Because of the collected subjective DMOS, the model can directly perceive the degradation caused by pansharpening.

Furthermore, Badal et al. [45] also developed a DNN-based algorithm for evaluating the quality of pansharpened images. This algorithm mimics the RR metrics like $Q2^n$ and SAM under the FR circumstance. Researchers constructed a model with a well-designed Pseudo HRMS and Pansharpenewd Image Frature Extractor, to obtain the pansharpened features. With the utilization of DNN, the algorithm autonomously learns distinctive knowledge for quality perception, providing an accurate quality evaluation of pansharpened images. It is observed that current deep learning-based FR evaluation models lack targeted extraction for distortion information, indicating the potential for further improvement.

In comparison to existing works based on deep learning, our research introduces rank generation and training strategies that enable straightforward implementation of quality rank assessment. These strategies can provide an intermediate stage, namely the comparative assessment of quality levels, which is easily implementable compared with complex quality evaluation problems. Through the establishment of this intermediate stage, achieving more precise quality score predictions becomes easier.

III. Rank Learning Based FR Quality Evaluation Method

A. Overview

The insufficient analysis of distortion information motivates us to propose a distortion-guided full-resolution quality evaluation method. The flowchart is depicted in Fig. 1. Our approach is based on an observation that we can artificially simulate different distortion levels within the pansharpening process. For example, in the spatial domain, we apply multiple levels of Gaussian blur (GB) and Gaussian noise (GN) to the original Pan image. These generated images can be easily ranked because we do know that the added GB does deteriorate quality. Apart from generating simulated distorted images, we designed a ranked distortion-guided training strategy and model structure. Using a Siamese network structure, paired rank learning is performed on the original Pan and MS and generated information with rank labels to enable the deep network to develop distortion perception capabilities. After rank learning, we conduct fine-tuning on the network branch by feeding the pansharpened images with quality scores to address the FR quality evaluation task. The supervised strategy of our model is the same as that of [45], adopting the supervised RR metric as the optimization objective of the model. This approach can achieve similar performance to the RR approach without a direct reference. The pipeline is outlined as follows:

1) Ranked Distortion Synthesizing: We synthetically degrade spatial and spectral domain information. While we struggle to obtain precise degraded quality scores, we know which of a pair of information has better quality. The specific degradation process is detailed in Section III-B.

2) Pansharpening Distortion-Perceiving Model Learning From Ranks: We train Siamese networks in the spatial and spectral domains by utilizing the generated ranked distortion information. For specifics on network structure and training strategies, refer to Section III-C, where the Siamese network can ultimately rank the quality of spatial and spectral domain information based on the degree of distortion.

3) Distortion-Guided Full-Resolution Pansharpening Quality Evaluation: We extract one branch each from the spatial and spectral distortion-perceiving Siamese networks as the foundation of the framework. We integrate spatial dimension reduction modules and spectral size reduction modules to meet the feature comparison requirements of Pan *versus* HR-MS and

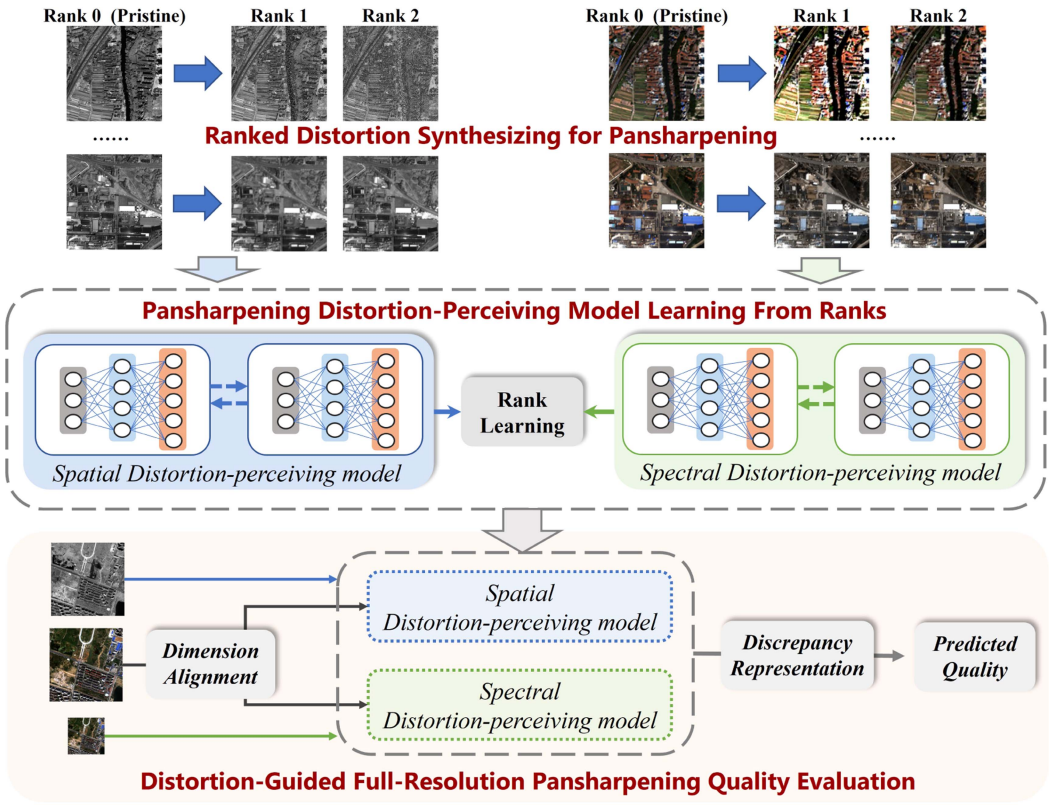


Fig. 1. Flowchart of our proposed FRQA method. The method consists of three main modules: Ranked distortion synthesizing, Pansharpening distortion-perceiving model learning from ranks and the distortion-guided full-resolution Pansharpening quality evaluation. Sample images are collected from IKONOS dataset.

MS *versus* HR-MS, as detailed in Section III-D. Finally, using the distortion-perceiving features and discrepancies between the Pan *versus* HR-MS and MS *versus* HR-MS as intermediate knowledge, regression is conducted to obtain the quality scores of pansharpened images.

B. Ranked Distortion Synthesizing

In previous studies, researchers have repeatedly emphasized the importance of preserving spatial and spectral information in the process of pansharpening [65]. It was pointed out that inevitable spatial artifacts and spectral distortion occur during the pansharpening process, which can impact subsequent tasks [6]. Motivated by this, we consider artificially synthesizing distortions directly on the original Pan and MS images to help the black-box networks learn more distinct distortion information.

Specifically, in the spatial domain, we introduce two types of distortions, GB and GN, to simulate the spatial artifacts generated during pansharpening. GB distortion involves blurring the image using a Gaussian function. Given a Pan image $I(x, y)$ and a Gaussian function $G(x, y)$, the Gaussian blurred image $I_{\text{GB}}(x, y)$ can be expressed as follows:

$$I_{\text{GB}}(x, y) = I(x, y) * G(x, y) \tag{1}$$

where $*$ denotes the convolution operation. The equation of the Gaussian function $G(x, y)$ is

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

where σ represents the standard deviation of the Gaussian function, controlling the degree of blur. In this work, we utilized two different sizes of Gaussian kernels (i.e., different σ values) to achieve varying levels of distortion. GN simulates the distortion caused by random factors in the pansharpening process. Given a Pan image $I(x, y)$, the image $I_{\text{GN}}(x, y)$ after adding GN can be expressed as follows:

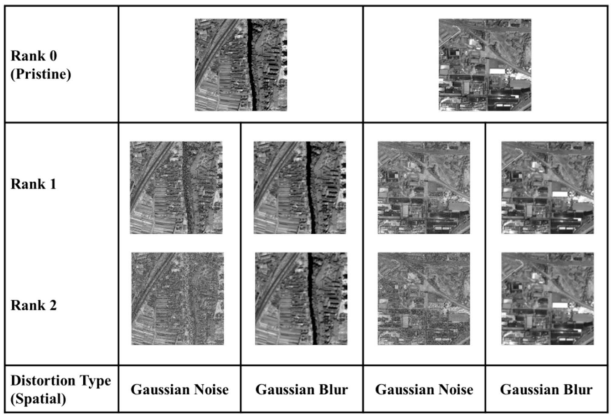
$$I_{\text{GN}}(x, y) = I(x, y) + N(x, y) \tag{3}$$

where $N(x, y)$ represents a random variable following the Gaussian distribution, indicating GN. The mathematical expression of $N(x, y)$ is

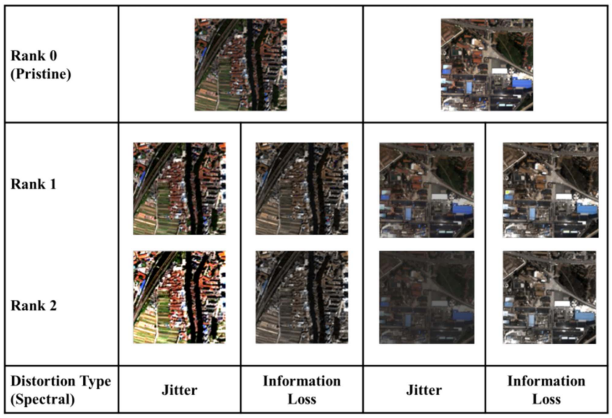
$$N(x, y) = \mathcal{X} \times \sigma, \quad \mathcal{X} \sim \mathcal{N}(0, 1) \tag{4}$$

where, $\mathcal{X}$ is a random value following the standard normal distribution with mean 0 and variance 1, while σ is the parameter controlling the noise intensity. In this work, two values of σ were set to generate two intensity levels of GN. We show two example sets of distorted Pan images in Fig. 2(a), with two levels of spatial distortion.

For the spectral domain, we introduce jitter and smoothing effects to the MS to simulate the spectral distortion during pansharpening. For jitter, we mimic the contrast and brightness adjustment methods to randomly combine spectral information distortion across the full bands. For a given MS $M(x, y, z)$, the $M_{\text{Jitter}}(x, y, z)$ after adding spectral distortion can be expressed



(a)



(b)

Fig. 2. Two sets of image examples from the IKONOS dataset, with synthetically spatial and spectral distortions. Spatial distortions consist of GN and GB, while spectral distortions include jitter and information loss. Each image was manually manipulated with four types of distortions at two levels each. After the degradation, both spatial and spectral distortions are directly represented. (a) Spatial distortion. (b) Spectral distortion.

as follows:

$$M_{\text{Jitter}}(x,y,z) = M(x,y,z) + \alpha \cdot (M(x,y,z) - \mu) + \mu + \beta \tag{5}$$

where μ is the mean value of the original pixel, and α and β are parameters controlling the jittering intensity. To achieve as random jitter distortion as possible, we set α and β as two randomly generated numbers within certain ranges to produce two levels of spectral distortion intensity.

Besides, we perform Savitzky–Golay (SG) smoothing on each band of the spectral information at each position to simulate information loss during the pansharpening process. For a given MS $M(x,y,z)$, the $M_{\text{Loss}}(x,y,z)$ after information loss can be expressed as follows:

$$\hat{M}_{\text{Loss}}(x,y,z) = \sum_{j=-m}^{m} c_j M(x,y,z+j) \tag{6}$$

where $\hat{M}_{\text{Loss}}(x,y,z)$ is the smoothed value of position (x,y) in band z, m is the half-length of the filter, and c_j are the coefficients of the SG filter. The coefficients c_j of the SG smoothing are obtained by least squares fitting. In our work, we set two pairs of c_j and m values to generate two levels of spectral information loss distortion. In Fig. 2(b), two sets of MS images with their two-level distorted variant are depicted.

From Fig. 2, it can be observed that our designed distortion simulation strategy can generate spatial and spectral information with distinct levels of distortion. This generated multilevel distortion information contains distinct distortion features and their inherent distortion level labels, aiding deep networks in more targeted perception training of distortion characteristics.

C. Pansharpening Distortion-Perceiving Model Learning From Ranks

We can utilize the ranked distortion information generated in Section III-B to train distortion-perceiving networks for the spatial and spectral domain, laying a foundational prior knowledge base for full-resolution pansharpening quality predicting. We introduce the Siamese network to learn from the information rankings, which is a network with two identical network branches and a special loss module. Pairwise images and labels serve as inputs to the network, resulting in two outputs which are passed to the loss module. The gradients of the loss function to all model parameters are calculated through backpropagation and updated using the stochastic gradient methods.

As depicted in Fig. 3, for a given image $I(x,y)$ as the network input, the output feature representation of $I(x,y)$ denoted as $f(I(x,y);\theta)$. $f(I(x,y);\theta)$ is obtained from the activation of the last layer of the network, where θ represents the network parameters. As our ultimate goal is to directly predict the quality scores, in the Siamese network, the output of the last layer is a single scalar. Since the purpose of this section is to predict image quality rank, we adopt the pairwise hinge ranking loss to optimize the learning process

$$L(I_1, I_2, \theta) = \max(0, f(I_2,\theta) - f(I_1,\theta) + \epsilon) \tag{7}$$

where ϵ is the margin. The loss function is to calculate the discrepancy between $f(I_2,\theta)$ and $f(I_1,\theta)$, and compare it with ϵ. The gradient calculation of the loss function is as follows:

$$\nabla_\theta L = \begin{cases} 0, \text{if } f(x_2;\theta) - f(x_1;\theta) + \varepsilon \le 0 \\ \nabla_\theta f(x_2;\theta) - \nabla_\theta f(x_1;\theta) \quad \text{otherwise.} \end{cases} \tag{8}$$

In our work, the quality level of input I_1 is always higher than I_2, which allows this loss function to assess whether the predicted quality level aligns with the ground truth. If the predicted quality level of I_1 is higher than I_2 (indicating that the network prediction is correct), the loss function returns 0, and the network does not update the gradient. Otherwise, it returns the discrepancy value, decreases the gradient of the higher scores, and increases the gradient for lower scores. This loss module helps the network learn the differences in multilevel distortions within the error margin.

Our network backbone adopts the popular ResNet50 and adjusts the number of channels in the first layer network, especially for the input Pan/MS. Given the gradient of the loss function to the model parameter θ, we utilize the adaptive moment estimation (Adam) optimizer to train the Siamese network. The

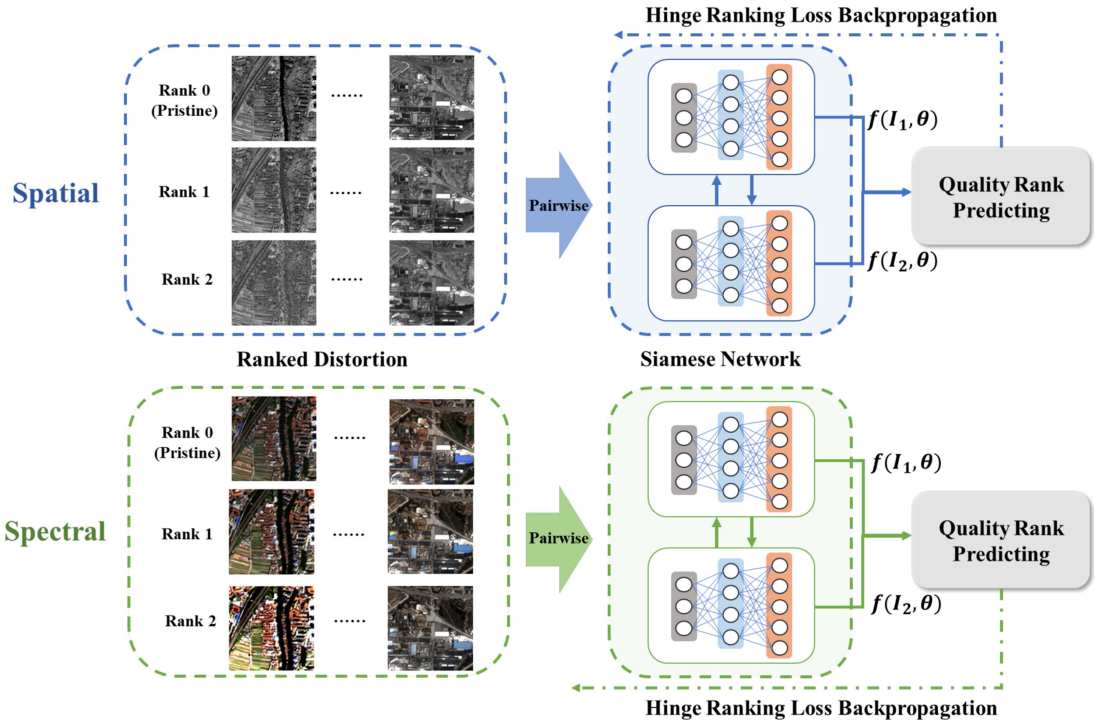


Fig. 3. Flowchart of distortion-perceiving model learning from ranks. After generating the ranked images, we pairwise feed them into the spatial and spectral Siamese networks. With the hinge ranking loss backpropagation, the models can perceive the distortion and predict the quality rank.

Adam optimizer is with an initial learning rate of 1e-4. The learning rate is scaled by 0.8 every 10 epochs and 100 epochs are required for training.

Through the well-designed ranking learning strategy, our trained spatial and spectral distortion-perceiving network learns various distortion characteristics and their quantified representations. The trained networks can serve as a pretraining foundation for FR quality evaluation networks, aiding targeted learning and representation of the pansharpening distortion features.

D. Distortion-Guided Full-Resolution Pansharpening Quality Evaluation

After the rank learning in Sections III-B and III-C, we developed two deep networks which capable of perceiving spatial and spectral distortion levels, but this cannot directly predict full-resolution pansharpening quality scores. There are three obstacles: 1) Due to differences in information channels and dimensional sizes, HRMS cannot be directly input into the pre-trained distortion-aware network; 2) The network can only perceive the distortion levels of HRMS and the original Pan/MS without a specific discrepancy representation strategy; 3) Lack of optimization goal with ground truth. To overcome the obstacles, we design the dimension alignment and discrepancy representation modules and adopt a label-generating strategy in [45]. The framework is depicted in Fig. 4.

1) Dimension Alignment: Since the depth of HRMS is different from Pan and the size is different from MS, it cannot be directly input into the pre-trained network. Therefore, a dimension alignment module was added, including a spatial channel squeezing and a spectral size reduction. Traditional averaging or downsampling was not used, instead convolutional processing. Integrating the dimension alignment into the complete optimization process of quality evaluation can establish a nonlinear mapping between the downsizing and the final quality prediction.

The spatial channel squeezing consists of three stacked convolution layers, a convolution layer with a kernel size of $1 \times 1 \times C/2$, a convolution layer with a size of $3 \times 3 \times 1$, and a convolution layer with a size of $1 \times 1 \times 1$. Each layer is followed by BatchNorm and Relu to enhance the network robustness. The specific structure is as shown in Fig. 5(a). After spatial channel squeezing, the input HRMS dimension of $H \times W \times C$ is aligned with Pan as $H \times W \times 1$, meeting the requirements of the spatial distortion-perceiving network.

The spectral size reduction consists of two convolution layers with a kernel size of $3 \times 3 \times C$, a stride of 2, and padding of 1. Similar to the spatial domain, each convolution layer is followed by BatchNorm and Relu activation. The specific structure is as shown in Fig. 5(b).

After spectral channel reduction, the input HRMS dimension of $H \times W \times C$ is aligned with MS as $H/4 \times W/4 \times C$, meeting the requirements of the spectral distortion-perceiving network.

2) Discrepancy Representation: We freeze the parameters of the pretrained spatial and spectral distortion-perceiving networks, then feed the original Pan/MS and the dimension-aligned HRMS to them. As a result, we obtain the spatial and spectral distortion features of the same dimension from the last fully connected layer of the backbone ResNet50 network.

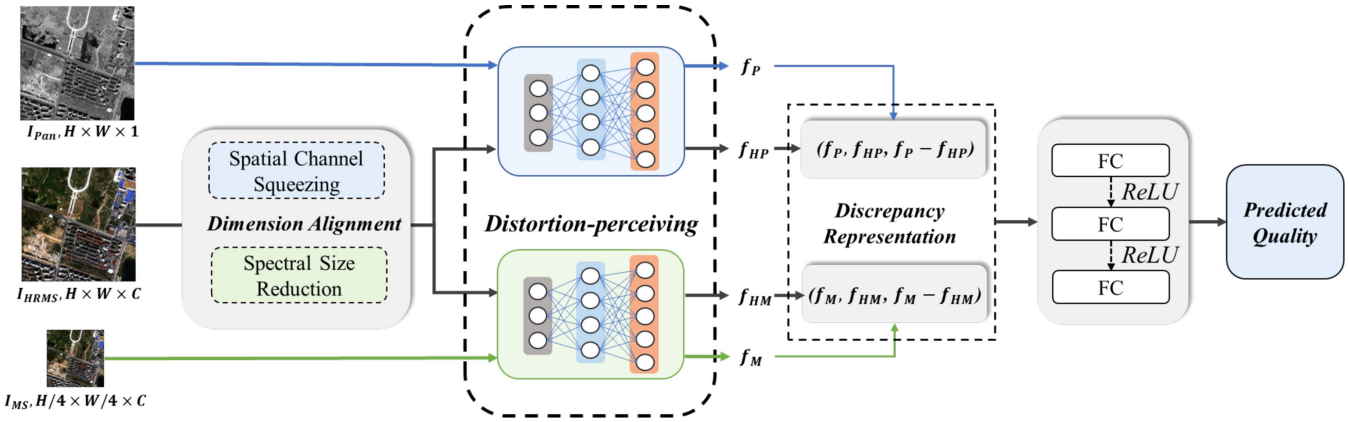


Fig. 4. Flowchart of distortion-guided full-resolution pansharpening quality evaluation. For the different dimension of Pan/MS and HRMS images, we first design the dimension alignment module to reduce the spatial and spectral size of HRMS. With the alignment, the HRMS can be fed into the pre-trained distortion-perceiving networks. The difference between the distortion-perceiving features is then used to represent the quality degradation. Finally, the regression group outputs the predicted quality.

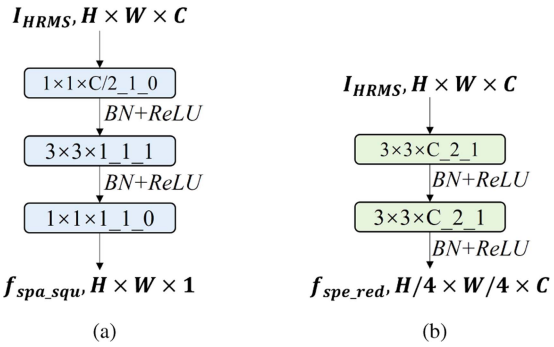


Fig. 5. Detailed structures of dimension alignment module. (a) Spatial channel squeezing; (b) Spectral size reduction. The convolutional layers are defined in terms of kernel size_stride_padding. BN denotes batch-normalization, and ReLU denotes rectified linear activation function.

Specifically, we denote f_P and f_{HP} as the output features of the spatial network with inputs of Pan and HR-MS, and f_M and f_{HM} as the output features of the spectral network with inputs of MS and HR-MS. To assess the quality score of HR-MS, we take the difference between f_P and f_{HP}, $f_P - f_{\mathrm{HP}}$, to characterize the spatial distortion during pansharpening. Similarly, we take the difference between f_M and f_{HM}, $f_M - f_{\mathrm{HM}}$, to characterize the spectral distortion between HR-MS and MS. Ultimately, we concat the spatial and spectral features to represent the distortion of pansharpening, mapping them to the final predicted score.

Specifically, we combine the features of $(f_P, f_{\mathrm{HP}}, f_P - f_{\mathrm{HP}})$ and $(f_M, f_{\mathrm{HM}}, f_M - f_{\mathrm{HM}})$, and add a regression mapping group. The regression mapping group consists of three sets of fully connected layers and Relu activation layers, ultimately outputting a single-dimensional scalar as the quality prediction result of pansharpening. The specific implementation of the regression group is as Fig. 4.

3) Optimization Strategy: Currently, there is no publicly available database with quality-labeled ground truth for the FR pansharpening task. Inspired by [45], we take the quality evaluation results of the RR index as ground truth, enabling the optimization goal of the FR pansharpening quality evaluation network to achieve predictive performance as close as possible to the reduced-resolution pansharpening quality evaluation. For the quality evaluation task, we choose the mean squared error (MSE) loss function to minimize the error between the predicted quality and the ground truth quality. The loss function is defined as follows:

$$L^q = \|\hat{Q} - Q\|_2 \tag{9}$$

where $\hat{Q}$ denotes the predicted quality score, and Q refers to the actual value.

Adam optimizer with an initial learning rate of 1e-5, is adopted. The learning rate is scaled by 0.5 every 10 epochs and 100 epochs are required for training.

Eventually, a distortion-guided full-resolution pansharpening quality evaluation framework is established, capable of perceiving and jointly processing spatial and spectral distortions based on the rank learning network, and ultimately mapping them to quality scores to achieve the goal of quality prediction.

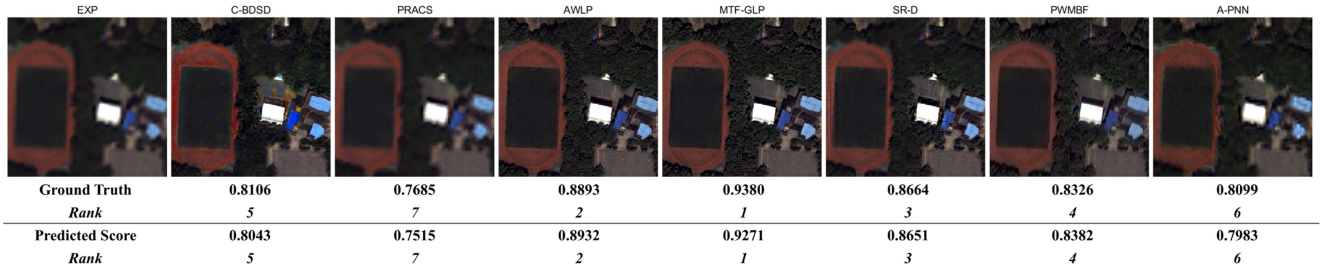
IV. Experiments

A. Experimental Setup

Datasets: We used a public benchmark database for pansharpening for the evaluations [71]. 2270 sets of MS, Pan imagery are collected in the database, where 200 sets are captured by the IKONOS sensor, 500 by QuickBird, 410 by GaoFen-1, 500 by WorldView-2, 160 by WorldView-3, and 500 by WorldView-4. All the Pan images have a size of 1024 × 1024. MS imagery, captured by IKONOS, QuickBird, GaoFen-1, and WorldView-4, are in a size of 256 × 256 × 4, while the others in 256 × 256 × 8.

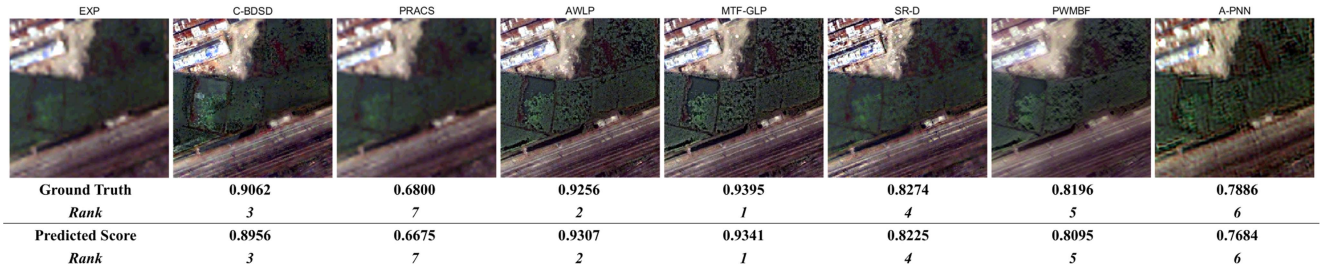
For efficient deep learning, we extract the image patches to match the model training. We extract Pan patches in a size of 256 × 256, and MS in 64 × 64 × 4/8. Thus, the pansharpened patches have a size of 256 × 256 × 4/8.

To evaluate the quality of pansharpened images, we apply different pansharpening algorithms to these patches. We



	EXP	C-BDSD	PRACS	AWLP	MTF-GLP	SR-D	PWMBF	A-PNN
Ground Truth		0.8106	0.7685	0.8893	0.9380	0.8664	0.8326	0.8099
Rank		*5*	*7*	*2*	*1*	*3*	*4*	*6*
Predicted Score		0.8043	0.7515	0.8932	0.9271	0.8651	0.8382	0.7983
Rank		*5*	*7*	*2*	*1*	*3*	*4*	*6*

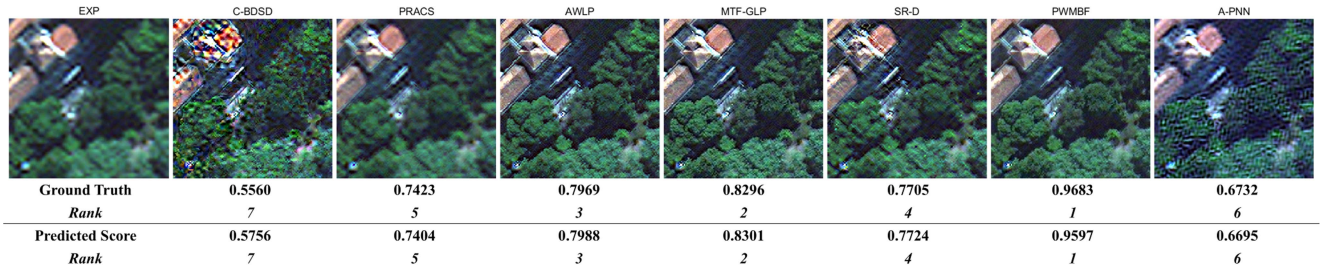
(a)



	EXP	C-BDSD	PRACS	AWLP	MTF-GLP	SR-D	PWMBF	A-PNN
Ground Truth		0.9062	0.6800	0.9256	0.9395	0.8274	0.8196	0.7886
Rank		*3*	*7*	*2*	*1*	*4*	*5*	*6*
Predicted Score		0.8956	0.6675	0.9307	0.9341	0.8225	0.8095	0.7684
Rank		*3*	*7*	*2*	*1*	*4*	*5*	*6*

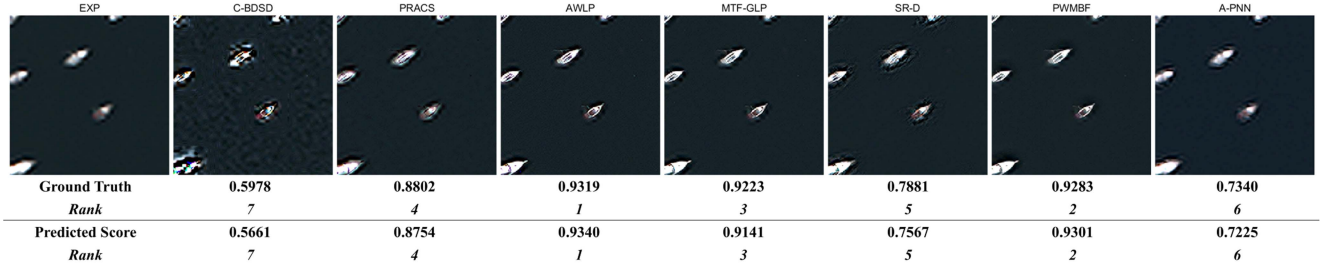
(b)

Fig. 6. Two sets of image examples captured by IKONOS sensor, pansharpened with seven different methods. (a) Urban scene; (b) Green vegetation. Each image is accompanied by the ground-truth $Q2^n$ values and their corresponding predicted quality scores with their respective quality rankings.



	EXP	C-BDSD	PRACS	AWLP	MTF-GLP	SR-D	PWMBF	A-PNN
Ground Truth		0.5560	0.7423	0.7969	0.8296	0.7705	0.9683	0.6732
Rank		*7*	*5*	*3*	*2*	*4*	*1*	*6*
Predicted Score		0.5756	0.7404	0.7988	0.8301	0.7724	0.9597	0.6695
Rank		*7*	*5*	*3*	*2*	*4*	*1*	*6*

(a)



	EXP	C-BDSD	PRACS	AWLP	MTF-GLP	SR-D	PWMBF	A-PNN
Ground Truth		0.5978	0.8802	0.9319	0.9223	0.7881	0.9283	0.7340
Rank		*7*	*4*	*1*	*3*	*5*	*2*	*6*
Predicted Score		0.5661	0.8754	0.9340	0.9141	0.7567	0.9301	0.7225
Rank		*7*	*4*	*1*	*3*	*5*	*2*	*6*

(b)

Fig. 7. Two sets of image examples captured by WorldView-3 sensor, pansharpened with seven different methods. (a) Green vegetation; (b) Water scenario. Each image is accompanied by the ground-truth $Q2^n$ values and their corresponding predicted quality scores with their respective quality rankings.

adopted seven kinds of pansharpening algorithms to generate the pansharpened patches, including the CS-based Methods (C-BDSD [58] and PRACS [59]), MRA-based Methods (AWLP [72] and MTF-GLP [54]), VO-based Methods(SR-D [61] and PWMBF [62]) and DL-based Methods(A-PNN [70]). Figs. 6 and 7 highlight image examples captured by IKONOS and WorldView-3 sensors, with their ground truth $Q2^n$ scores for the seven chosen pansharpening algorithms.

Implementation Details: The database is randomly partitioned into 70% for training, 10% for validation, and 20% for testing. The training, validation, and testing sets are nonoverlapping. This process is repeated 10 times, and the final experimental results are obtained by averaging the results from the 10 testing sets.

Performance Criteria: In the absence of ground truth, we adopted the RR protocol $Q2^n$ as the reference, the same as [45].

TABLE I
PERFORMANCE COMPARISON OF OUR PROPOSED METHOD AND THE FOUR PUBLIC FR METRICS ON IKONOS, QUICKBIRD, GAOFEN-1 AND WORLDVIEW-4

Sensors	**IKONOS**				**QuickBird**			
Methods	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
QNR	0.5639	0.3863	0.2749	0.1229	0.4444	0.4423	0.3180	0.1482
FQNR	0.7632	0.7889	0.6244	0.1052	0.4621	0.4579	0.3300	0.1311
HQNR	0.7969	0.7727	0.5991	0.1054	0.5451	0.5235	0.3749	0.1317
RQNR	0.7734	0.8052	0.6496	0.0954	0.4272	0.4243	0.3049	0.1325
Proposed	**0.8911**	**0.8865**	**0.7194**	**0.0685**	**0.9022**	**0.9101**	**0.7323**	**0.0608**
Sensors	**GaoFen-1**				**WorldView-4**			
Methods	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
QNR	0.2486	0.2694	0.2002	0.1759	0.5183	0.5571	0.4034	0.1450
FQNR	0.7307	0.7369	0.5887	0.1113	0.6738	0.7235	0.5898	0.1273
HQNR	0.6291	0.6409	0.4891	0.1361	0.6845	0.7439	0.5895	0.1042
RQNR	0.4710	0.4474	0.3425	0.1496	0.5872	0.6737	0.5278	0.1292
Proposed	**0.8986**	**0.8972**	**0.7081**	**0.0685**	**0.8790**	**0.8768**	**0.6624**	**0.0894**

The IKONOS, QuickBird, GaoFen-1 and WorldView-4 sensors capture MS imagery with 4 bands. The best performance are in bold.

TABLE II
PERFORMANCE COMPARISON OF OUR PROPOSED METHOD AND THE FOUR PUBLIC FR METRICS ON WORLDVIEW-2 AND WORLDVIEW-3

Sensors	**WorldView-2**				**WorldView-3**			
Methods	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
QNR	0.9182	0.8677	0.7078	0.0694	0.8820	0.8699	0.6558	0.0855
FQNR	0.9096	0.9045	0.7531	0.0703	0.8636	0.8595	0.7030	0.0939
HQNR	0.9230	0.9111	0.7672	0.0655	0.8937	0.8753	0.7236	0.0756
RQNR	0.9081	0.9087	0.7617	0.0692	0.8593	0.8639	0.7042	0.0884
Proposed	**0.9491**	**0.9512**	**0.7949**	**0.0590**	**0.9022**	**0.9056**	**0.7467**	**0.0681**

The two sensors capture MS imagery with 8 bands. The best performance is in bold.

Model performance is assessed using four public metrics in the quality evaluation tasks: Pearson linear correlation coefficient (PLCC), Spearman rank-order correlation coefficient (SRCC), Kendall rank-order correlation coefficient (KRCC), and root mean squared error (RMSE). Higher SRCC, PLCC, and KRCC, along with lower RMSE, indicate a better prediction.

B. Performance Comparisons

For performance evaluation on FR pansharpening, we compare the proposed model with four public FR approaches: QNR [19], FQNR [23], HQNR [22], and RQNR [20] for comparison.

In Table I, we summarize the experimental results on imagery collected by IKONOS, QuickBird, GaoFen-1, and Worldview-4. The four sensors capture MS imagery with 4 bands. Table II presents results on the imagery collected by Worldview-2 and Worldview-3, where MS is with 8 bands. We computed the weighted average results in Table III, to represent the overall performance of quality evaluation methods. The weight of each sensor is proportional to the number of captured images.

As presented in Tables I and II, the proposed method achieves the best performance in PLCC, SRCC, KRCC, and RMSE for all the sensors. Also, it shows the best overall performance as depicted in Table III.

TABLE III
COMPARISON OF THE WEIGHTED-AVERAGE EXPERIMENTAL RESULTS

Weighted-Average Results				
Methods	PLCC	SROCC	KROCC	RMSE
QNR	0.5710	0.5552	0.4214	0.1284
FQNR	0.7106	0.7226	0.5793	0.1083
HQNR	0.7209	0.7253	0.5735	0.1055
RQNR	0.6372	0.6546	0.5199	0.1145
Proposed	**0.9057**	**0.9070**	**0.7262**	**0.0692**

The best performances are in bold.

Our method has made significant progress on 4-band data compared to 8-band data. Compared with the 8-band MS, the raw data of the 4-band is insufficient, resulting in the ineffectiveness of the methods based on handcrafted features. However, through our method of distortion synthesizing and rank learning, we overcome the issue of data shortage.

Furthermore, driven by deep learning, the distorted features are effectively perceived, forming a more accurate mapping with quality scores.

In Table II, our method still achieved improvement on 8-band data for the well-designed training strategy. It can also be observed that HQNR performs very well. Compared with

TABLE IV
ABLATION EXPERIMENTS FOR THE PROPOSED MODEL

Sensors	**IKONOS**				**QuickBird**			
Methods	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
Backbone	0.7783	0.7657	0.6098	0.0975	0.7767	0.7810	0.6183	0.0981
Spectral RL	0.8425	0.8412	0.6776	0.0916	0.8580	0.8678	0.6875	0.0765
Spatial RL	0.8503	0.8394	0.6732	0.0873	0.8616	0.8633	0.6880	0.0760
RL	0.8846	0.8791	0.7017	0.0751	0.8895	0.8932	0.7154	0.0721
Proposed	**0.8911**	**0.8865**	**0.7194**	**0.0685**	**0.9022**	**0.9101**	**0.7323**	**0.0608**
Sensors	**GaoFen-1**				**WorldView-4**			
Methods	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
Backbone	0.7847	0.7764	0.5932	0.1099	0.7110	0.7200	0.4942	0.1068
Spectral RL	0.8489	0.8478	0.6602	0.0882	0.8310	0.8310	0.6189	0.0905
Spatial RL	0.8584	0.8562	0.6634	0.0865	0.8353	0.8304	0.6189	0.0874
RL	0.8827	0.8887	0.6910	0.0772	0.8630	0.8606	0.6442	0.0728
Proposed	**0.8986**	**0.8972**	**0.7081**	**0.0685**	**0.8790**	**0.8768**	**0.6624**	**0.0694**
Sensors	**WorldView-2**				**WorldView-3**			
Methods	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
Backbone	0.8300	0.7875	0.6253	0.0873	0.7313	0.7499	0.5836	0.1082
Spectral RL	0.9006	0.9047	0.7517	0.0705	0.8590	0.8584	0.6990	0.0820
Spatial RL	0.9017	0.9066	0.7480	0.0711	0.8546	0.8609	0.7047	0.0762
RL	0.9332	0.9330	0.7788	0.0603	0.8851	0.8877	0.7294	0.0705
Proposed	**0.9491**	**0.9512**	**0.7949**	**0.0590**	**0.9022**	**0.9056**	**0.7467**	**0.0681**

Backbone denotes the model without rank learning. Spectral RL denotes backbone adding the spectral rank learning. Spatial RL denotes backbone adding the spatial rank learning. RL denotes the proposed model only without discrepancy representation. The best performances are in bold.

the other three methods based on handcrafted features, HQNR adopts indicators excelling in both spectral and spatial distortion measuring.

In addition to excellent performance, our method also demonstrates greater stability. QNR, FQNR, and RQNR perform excellently in Table II but are extremely unstable in Table I. Through rank learning and quality regression, we extract distortion-perceiving features from various data, exhibiting stable perception capability across all sensors. However, manually crafted features may fail to adequately grasp the complexity of the data, leading to unstable results.

As results in Table III, our proposed model achieves the best weighted-average performance, and HQNR and FQNR show the similar performance, placing them in second and third place. The overall performance comparison also demonstrates the good generalization and robustness of our method. Besides, the HQNR and FQNR share the same spectral distortion measurement. The performance also validates the effectiveness of their spectral feature extractor.

Figs. 6 and 7 illustrate a comparison of predicted $Q2^n$ scores with their corresponding ground truth and quality ranks. Four example patches from the IKONOS and WorldView-3 sensors using various pansharpening techniques are shown, including C-BDSD, PRACS, AWLP, MTF-GLP, SR-D, PWMBF, A-PNN, and the plain LR-MS upsampling (EXP). These techniques are evaluated at different quality levels of ground truth $Q2^n$. The observed performance rankings for sample image patches captured by IKONOS and WordView-3 using the respective pansharpening algorithms align with the predicted $Q2^n$ score trends. These results collectively underscore the robustness of our proposed method across data captured by diverse satellite sensors.

C. Ablation Study

1) Rank Learning: We conducted ablation experiments to determine the effect of the rank learning strategy on the FRQA task. Experimental results are shown in Table IV.

Initially, we directly utilized the original backbone network ResNet for feature extraction, only adding a dimension alignment module without conducting discrepancy representation. The network's training labels and loss function remained consistent with the proposed model. While the features directly output by ResNet can represent certain information, they are not specifically designed for pansharpening distortion perception, resulting in unsatisfactory test results, corresponding to the model named *Backbone* in the Table.

Furthermore, we conducted separate experiments to determine the significance of spatial and spectral rank learning in the final model. Specifically, the model termed *Spectral RL* indicates that spatial features are directly extracted by ResNet, while spectral features are derived from a spectral distortion perception network trained with rank learning. Conversely, the model named *Spatial RL* states that spectral features are directly extracted by ResNet, with spatial features extracted by a spatial distortion perception network trained with rank learning. Performance improvements were observed in integrating rank learning, as highlighted in Table IV, indicating that the rank

TABLE V
PERFORMANCE COMPARISON BASED ON THE TWO KINDS OF TRAINING STRATEGY

Sensors	**IKONOS**				**QuickBird**			
Strategy	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
I	0.8665	0.8631	0.6863	0.0761	0.8794	0.8759	0.7111	0.0701
II	0.8911	0.8865	0.7194	0.0685	0.9022	0.9101	0.7323	0.0608
Improvement	*+0.0246*	*+0.0234*	*+0.0331*	*-0.0076*	*+0.0228*	*+0.0342*	*+0.0212*	*-0.0093*
Sensors	**GaoFen-1**				**WorldView-4**			
Strategy	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
I	0.8799	0.8804	0.7012	0.0704	0.8405	0.8410	0.6181	0.0819
II	0.8986	0.8972	0.7081	0.0685	0.8790	0.8768	0.6624	0.0694
Improvement	*+0.0187*	*+0.0168*	*+0.0069*	*-0.0019*	*+0.0385*	*+0.0358*	*+0.0443*	*-0.0125*
Sensors	**WorldView-2**				**WorldView-3**			
Strategy	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
I	0.9315	0.9324	0.7641	0.0632	0.8997	0.9001	0.7182	0.0690
II	0.9491	0.9512	0.7949	0.0590	0.9022	0.9056	0.7467	0.0681
Improvement	*+0.0176*	*+0.0188*	*+0.0278*	*-0.0042*	*+0.0025*	*+0.0055*	*+0.0285*	*-0.0009*

Strategy I indicates the model trained with the images pansharpened by one specific method. Strategy II indicates the model trained with the images pansharpened by eight kinds of Methods.

learning strategy enhances the network's capability to perceive distortion features in both spatial and spectral domains.

The model designated as *RL* in the table utilizes the spectral and spatial distortion-perceiving networks to extract separate distortion features. Results demonstrate that the combination of spatial and spectral distortion features significantly improves performance. Also, the progress indicates that joint perception and prediction of various types of distortions can establish a more reasonable and comprehensive mapping relationship, greatly benefiting the effectiveness of quality prediction.

2) Discrepancy Representation: We conducted an ablation experiment on the discrepancy representation module to express the necessity of the differential module. Experiment results are shown in Table IV, and the differences between RL and proposed demonstrate that, compared to direct regression of features, adding a discrepancy representation module can more directly express the differences between distorted and original imagery. Such an operation aligns with quality assessment tasks, aiding in the learning of deep networks. This also suggests that designing more precise differential representation methods may still improve network performance.

D. Discussion

1) Training Strategy: In the experiment, we employed two training strategies to develop the quality evaluation model. The *Strategy I* involved specific model training and testing for each pansharpening method, and the final performance was averaged over the seven models. The *Strategy II* unified the images generated by multiple pansharpening methods as inputs for training and testing a general model, suggested as [45].

The comparative results of the two strategies and the improvements are shown in Table V. The results indicate that the second training strategy possesses superior and more stable predictive performance.

We attribute this to two factors. First, imagery generated by various pansharpening methods increases the training data, enhancing the adequacy of network training, which is crucial for deep learning. Particularly for IKONOS, compared to other sensors, it collected a limited 200 sets of data. Its performance under training with a single pansharpening method is relatively unsatisfactory. After mixed large-scale training, its performance significantly improved.

Second, the distortions generated by a single pansharpening method may be limited, restricting the image quality within a certain range and causing data limitations for network training. In contrast, distortions from various pansharpening methods are more diverse and random, covering a more comprehensive range of image quality. In addition, the artificially simulated pansharpening distortions generated by our method can better match these comprehensive distortion characteristics.

Therefore, the general training strategy enhances the rationality of network training data, and the comprehensiveness of network knowledge learning, and thus helps enhance the predictive performance of the network. Therefore, we ultimately adopted *Strategy II* for model training and testing.

2) Distortion Ranks: We synthesize the one-level and three-level distortion to compare with the adopted two-level distortion version. For one-level, the synthetic distortion is the Rank 2 distortion in adopted version. For three-level, we add more severe distortion in spatial and spectral domain. Two sets of image examples from the IKONOS dataset, with synthetically spatial and spectral distortions, are shown in Fig. 8.

We conduct experiments to show the performance of different distortion levels, with results shown in Table VI. As the results show, the model trained with two-level distortion performs the best. For the one-level version, the synthetic distortion is too limited to represent the quality degradation that occurs during the pansharpening process, resulting in poor performance. For the three-level version, its performance is satisfactory, but worse

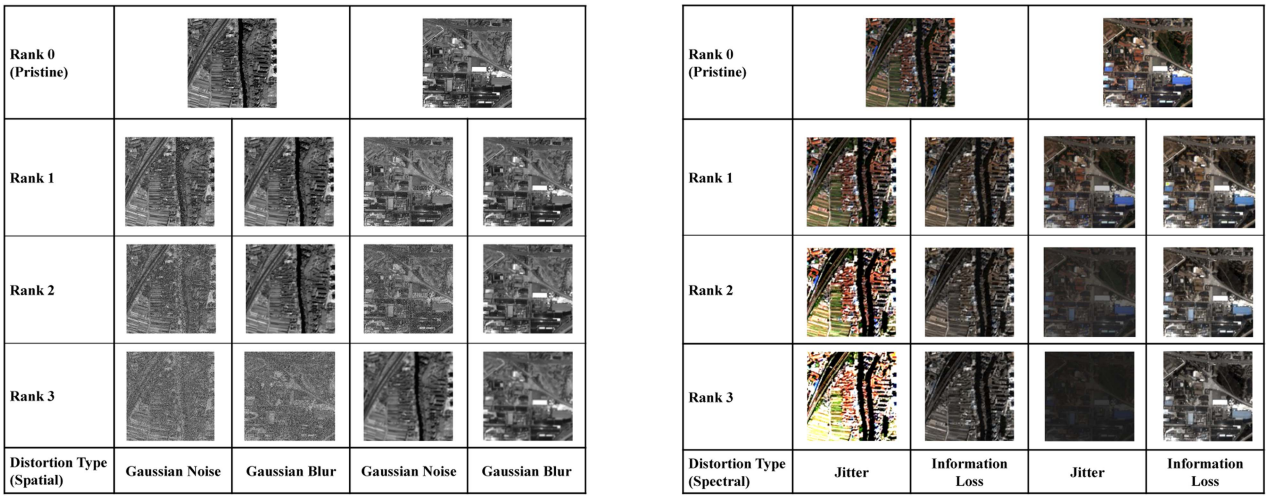


Fig. 8. Two sets of image examples from the IKONOS dataset, with synthetically spatial and spectral distortions. Three ranks of distortion images are generated to discuss the impact of the generation level on the final performance. (a) Spatial distortion. (b) Spectral distortion.

TABLE VI
PERFORMANCE COMPARISON OF THE DIFFERENT GENERATING DISTORTION RANKS

Sensors	**IKONOS**				**QuickBird**			
Distortion Rank	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
One	0.8124	0.8092	0.6159	0.1091	0.8291	0.814	0.6211	0.0992
Two (Adopted)	**0.8911**	**0.8865**	**0.7194**	**0.0685**	**0.9022**	**0.9101**	**0.7323**	**0.0608**
Three	0.8788	0.8792	0.7011	0.0912	0.8866	0.8819	0.7209	0.0719
Sensors	**GaoFen-1**				**WorldView-4**			
Distortion Rank	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
One	0.8011	0.7992	0.6061	0.1082	0.789	0.7822	0.6212	0.1101
Two (Adopted)	**0.8986**	**0.8972**	**0.7081**	**0.0685**	**0.8790**	**0.8768**	**0.6624**	**0.0894**
Three	0.8722	0.8696	0.6877	0.0721	0.8554	0.8415	0.6355	0.091
Sensors	**WorldView-2**				**WorldView-3**			
Distortion Rank	PLCC	SRCC	KRCC	RMSE	PLCC	SRCC	KRCC	RMSE
One	0.9021	0.9044	0.7589	0.0658	0.8741	0.8678	0.7245	0.0757
Two (Adopted)	**0.9491**	**0.9512**	**0.7949**	**0.0590**	**0.9022**	**0.9056**	**0.7467**	**0.0681**
Three	0.9351	0.9301	0.7681	0.0603	0.8879	0.8699	0.7312	0.0729

The best performances are in bold.

than the two-level ones. The generated Rank 3 images are impaired severely with extremely low quality, which is rare in the process of pansharpening. So, the additional three-level distortion generated cannot better represent the distortion characteristics generated during the pansharpening process. Therefore, we train our model with two-level synthetic distortion to achieve the best predicting of pansharpening quality.

V. CONCLUSION

This article presents a novel approach to address the challenges in FR quality evaluation for pansharpening. First, we introduce a ranked distortion synthesizing strategy to extract distortion information from generated images without reference images. Subsequently, a pansharpening distortion-perceiving model is developed based on rank learning. We develop spatial and spectral Siamese network structures to perceive the distortion and utilize a pair-wise learning method for ranked images. Finally, we construct a distortion-guided full-resolution quality evaluation framework for pansharpening. The framework incorporates the rank-learning Siamese network and is complemented with a dimension alignment module and discrepancy representation module to facilitate distortion extraction among HR-MS, Pan, and MS images. Furthermore, we conducted comprehensive experiments on a public remote sensing database. The experimental results highlight the superior performance of the proposed method. While the work has improved evaluation accuracy through rank generation and training strategies, we recognize the importance of incorporating specialized network design to further enhance prediction accuracy. This modification can potentially lead to more precise results and improved overall performance of our model.

REFERENCES

[1] H. Wang, S. Cheng, Y. Li, and A. Du, "Lightweight remote-sensing image super-resolution via attention-based multilevel feature fusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Nov. 2023, Art. no. 2005715.

[2] X. Chao and Y. Li, "Semisupervised few-shot remote sensing image classification based on KNN distance entropy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 8798–8805, Oct. 2022.

[3] M. Zhou, J. Huang, D. Hong, F. Zhao, C. Li, and J. Chanussot, "Rethinking pan-sharpening in closed-loop regularization," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2023.3279931.

[4] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "25 years of pansharpening: A critical review and new developments," *Signal Image Process. Remote Sens.*, pp. 533–548, 2012.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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, no. 1/2, pp. 4–18, 2003.

[10] 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.

[11] 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.

[12] 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. JPL, Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, pp. 147–149.

[13] 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" SEE/URISCA*, 2000, pp. 99–103.

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

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

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

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

[27] G. Palubinskas, "Quality assessment of pan-sharpening methods," in *Proc. IEEE Geosci. Remote Sens. Symp.*, 2014, pp. 2526–2529.

[28] Y. Li, J. Yang, Z. Zhang, J. Wen, and P. Kumar, "Healthcare data quality assessment for cybersecurity intelligence," *IEEE Trans. Ind. Inform.*, vol. 19, no. 1, pp. 841–848, Jan. 2023.

[29] 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, May 2022, Art. no. 5410311.

[30] X. Yang, F. Li, and H. Liu, "TTL-IQA: Transitive transfer learning based no-reference image quality assessment," *IEEE Trans. Multimedia*, vol. 23, pp. 4326–4340, 2021.

[31] Y. Li and S. Ercisli, "Explainable human-in-the-loop healthcare image information quality assessment and selection," *CAAI Trans. Intell. Technol.*, 2023.

[32] J. Wang, F. Li, Y. An, X. Zhang, and H. Sun, "Toward robust LiDAR-Camera fusion in BEV space via mutual deformable attention and temporal aggregation," *IEEE Trans. Circuits Syst. Video Technol.*, to be published, doi: 10.1109/TCSVT.2024.3366664.

[33] D. Hong, J. Yao, C. Li, D. Meng, N. Yokoya, and J. Chanussot, "Decoupled-and-coupled networks: Self-supervised hyperspectral image super-resolution with subpixel fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Oct. 2023, Art. no. 5410311.

[34] S. Cheng, R. Chan, and A. Du, "CACFTNet: A hybrid Cov-attention and cross-layer fusion transformer network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–17, Mar. 2024.

[35] Z. Kuang, H. Bi, F. Li, C. Xu, and J. Sun, "Polarimetry-inspired contrastive learning for class-imbalanced PolSAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, May 2024, Art. no. 5410311.

[36] C. Li et al., "CasFormer: Cascaded transformers for fusion-aware computational hyperspectral imaging," *Inf. Fusion*, vol. 108, pp. 102408–102419, 2024.

[37] X. Zhang, S. Cheng, L. Wang, and H. Li, "Asymmetric cross-attention hierarchical network based on CNN and transformer for bitemporal remote sensing images change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Feb. 2023, Art. no. 2000415.

[38] Y. Li, J. Yang, and J. Wen, "Entropy-based redundancy analysis and information screening," *Digit. Commun. Netw.*, vol. 9, no. 5, pp. 1061–1069, 2023.

[39] C. Li, B. Zhang, D. Hong, J. Yao, and J. Chanussot, "LRR-Net: An interpretable deep unfolding network for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, May 2023, Art. no. 5513412.

[40] N. Wang, H. Bi, F. Li, C. Xu, and J. Gao, "Self-distillation-based polarimetric image classification with noisy and sparse labels," *Remote Sens.*, vol. 15, no. 24, 2023, Art. no. 5751.

[41] S. Cheng, L. Wang, A. Du, and Y. Li, "Bidirectional focused semantic alignment attention network for cross-modal retrieval," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2021, pp. 4340–4344.

[42] D. Hong et al., "SpectralGPT: Spectral remote sensing foundation model," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published, doi: 10.1109/TPAMI.2024.3362475.

[43] 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.

[44] X. Meng et al., "A blind full-resolution quality evaluation method for pansharpening," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jun. 2022, Art. no. 5513412.

[45] 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, Aug. 2022.

[46] G. Vivone, M. D. Mura, A. Garzelli, and F. Pacifici, "A benchmarking protocol for pansharpening: Dataset, preprocessing, and quality assessment," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6102–6118, Jun. 2021.

[47] X. Guan, F. Li, X. Zhang, M. Ma, and S. Mei, "Assessing full-resolution pansharpening quality: A comparative study of methods and measurements," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 6860–6875, Jul. 2023.
[48] X. Liu, J. Van de Weijer, and A. D. Bagdanov, "RankIQA: Learning from rankings for no-reference image quality assessment," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1040–1049.
[49] F. Li, Y. Zhang, and P. C. Cosman, "MMMNet: An end-to-end multi-task deep convolution neural network with multi-scale and multi-hierarchy fusion for blind image quality assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4798–4811, Dec. 2021.
[50] B. Hu, G. Zhu, L. Li, J. Gan, W. Li, and X. Gao, "Blind image quality index with cross-domain interaction and cross-scale integration," *IEEE Trans. Multimedia*, vol. 26, pp. 2729–2739, 2024.
[51] G. Vivone et al., "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, May 2015.
[52] H. Li, B. Manjunath, and S. Mitra, "Multisensor image fusion using the wavelet transform," *Graphical Models Image Process.*, vol. 57, no. 3, pp. 235–245, 1995.
[53] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. 31, no. 4, pp. 532–540, Apr. 1983.
[54] 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.
[55] 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.
[56] 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.
[57] 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.
[58] 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.
[59] 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.
[60] G. Vivone et al., "Pansharpening based on semiblind deconvolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1997–2010, Apr. 2015.
[61] 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.
[62] 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.
[63] 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.
[64] 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.
[65] 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.
[66] 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.
[67] 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, Jul. 2020.
[68] 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.
[69] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, 2016, Art. no. 594.
[70] 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.
[71] 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.
[72] 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.

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 signal and information processing 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 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, 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.

Haixia Bi received the B.S. and M.S. degrees in computer science and technology from the Ocean University of China, Qingdao, China, in 2003 and 2006, respectively, and the Ph.D. degree in computer science and technology from Xi'an Jiaotong University, Xi'an, China, in 2018.

She was a Post-Doctoral Research Fellow with the University of Derby, Derby, U.K., from 2018 to 2019, and the University of Bristol, Bristol, U.K., from 2019 to 2021. She is currently an Associate Professor with the School of Information and Communication Engineering, Xi'an Jiaotong University. Her research interests include machine learning and remote sensing image processing.

Dr. Bi was the recipient of the Best Reviewer Award of the IEEE Geoscience and Remote Sensing Letters in 2019 and 2021. She serves as a Guest Editor of Remote Sensing special issue.

Lijiao Gong (Senior Member, IEEE) received the M.S. degree in measuring technology and instruments from Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2008, and the Ph.D. degree in instrumentation science and technology from University of Science and Technology of China (USTC), Hefei, China, in 2015.

She is currently a Professor with Shihezi University, Shihezi, China. She is the author/coauthor of more than 100 technical papers and three books, and holds more than 15 issued/pending patents. She is serving as a member of the International Standards Committee IEC SC 8 A. Her research interests are in the areas of multiagent cooperative control and its application, the source-grid coordinated control technologies for renewables.