HPRN: Holistic Prior-Embedded Relation Network for Spectral Super-Resolution

Chaoxiong Wu*, Student Member, IEEE, Jiaojiao Li*, Member, IEEE, Rui Song*, Member, IEEE, Yunsong Li*, Member, IEEE, and Qian Du*, Fellow, IEEE

Abstract—Spectral super-resolution (SSR) refers to the hyperspectral image (HSI) recovery from an RGB counterpart. Due to the one-to-many nature of the SSR problem, a single RGB image can be reprojected to many HSIs. The key to tackle this ill-posed problem is to plug into multisource prior information such as the natural spatial context prior of RGB images, deep feature prior, or inherent statistical prior of HSIs so as to effectively alleviate the degree of ill-posedness. However, most current approaches only consider the general and limited priors in their customized convolutional neural networks (CNNs), which leads to the inability to guarantee the confidence and fidelity of reconstructed spectra. In this article, we propose a novel holistic prior-embedded relation network (HPRN) to integrate comprehensive priors to regularize and optimize the solution space of SSR. Basically, the core framework is delicately assembled by several multiresidual relation blocks (MRBs) that fully facilitate the transmission and utilization of the low-frequency content prior of RGBs. Innovatively, the semantic prior of RGB inputs is introduced to mark category attributes, and a semantic-driven spatial relation module (SSRM) is invented to perform the feature aggregation of clustered similar ranges for refining recovered characteristics. In addition, we develop a transformer-based channel relation module (TCRM), which breaks the habit of employing scalars as the descriptors of channelwise relations in the previous deep feature prior and replaces them with certain vectors to make the mapping function more robust and smoother. In order to maintain the mathematical correlation and spectral consistency between hyperspectral bands, the second-order prior constraints (SOPCs) are incorporated into the loss function to guide the HSI reconstruction. Finally, extensive experimental results on four benchmarks demonstrate that our HPRN can reach the state-of-the-art performance for SSR quantitatively and qualitatively. Furthermore, the effectiveness and usefulness of the reconstructed spectra are verified by the classification results on the remote sensing dataset. Codes are available at https://github.com/Deep-imageLab/HPRN.

Index Terms—Holistic prior-embedded relation, multiresidual, second-order prior constraints (SOPCs), semantic-driven, spectral super-resolution (SSR), transformer-based channel relation module (TCRM).

I. INTRODUCTION

HYPERSONTICAL imaging can cover a wider range of the electromagnetic spectrum than the ordinary RGB cameras [1]. The captured hyperspectral images (HSIs) usually contain a great quantity of spectral bands. Such abundant spectral information reflects intrinsic properties of objects and promotes the rapid applications of HSIs in remote sensing [2], [3], [4], image super-resolution [5], [6], medical diagnosis [7], [8], image classification [9], [10], [11], [12], image denoising [13], and unmixing [14], to name a few. However, the current hyperspectral imaging technology exists some practical bottlenecks from the hardware perspective, which are sensitive to the acquisition of high-quality HSIs. Typically, the inescapable loss of obtaining an HSI cube with higher spectral resolution is either the worse time resolution, such as whisk- and push-broom scanning [15], [16], or lower spatial resolution, such as integrated devices based on variable-filter design [17]. This tradeoff lies in that these types of equipment must operate in a 1-D or 2-D scanning mode to acquire 3-D data. At the expense of sacrificing moderate spatial or spectral resolution, recent advances in snapshot HSIs imaging can accomplish the hyperspectral video shooting, which can merely meet the coarse-grained application analysis [18], [19], [20], [21]. Also, these relevant imaging facilities inevitably need to bear expensive hardware overhead and burden.

To cope with the above shortcomings of hardware factors, researchers turn their energies to the software perspective to excavate solutions, for example, computed tomography imaging spectrometers [28], compressed sensing techniques [29], and multiplexed illuminations [30], [31], [32], [33]. However, these solutions sometimes rely on special environments and require significant postprocessing calculations to restore

---

*Corresponding author: Jiaojiao Li.

Manuscript received 7 February 2022; revised 7 June 2022, 31 August 2022, 7 December 2022, and 4 January 2023; accepted 15 March 2023. Date of current version 6 August 2024. This work was supported in part by the Fundamental Research Funds for the Central Universities under Grant JBF220101; in part by the State Key Laboratory of Geo-Information Engineering under Grant SKLGE2020-M-5-1; in part by the Science, Technology on Space Intelligent Control Laboratory under Grant ZDSYS-2019-03; in part by the Innovation Fund of Xidian University under Grant 5001-20190215456; in part by the Open Research Fund of CAS Key Laboratory of Spectral Imaging Technology under Grant LSIT201924W; in part by the Wuhu and Xidian University Special Fund for Industry-University-Research Cooperation under Grant XWYCY-012021002; in part by the National Nature Science Foundation of China under Grant 61901343; in part by the 111 Project under Grant B08038; and in part by the Youth Innovation Team of Shaanxi Universities. (Corresponding author: Jiaojiao Li.)

Chaoxiong Wu, Rui Song, and Yunsong Li are with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi’an 710071, China (e-mail: cxwu1120@163.com; rsong@xidian.edu.cn; ysl@mail.xidian.edu.cn).

Jiaojiao Li is with the State Key Laboratory of Integrated Service Networks, Xidian University, Xi’an 710071, China, and also with the CAS Key Laboratory of Spectral Imaging Technology, Chinese Academy of Sciences, Xi’an 710119, China (e-mail: jji@xidian.edu.cn).

Qian Du is with the Department of Electronic and Computer Engineering, Mississippi State University, Starkville, MS 39762 USA (e-mail: du@ece.msstate.edu).

Digital Object Identifier 10.1109/TNNLS.2023.3260828

Authorized license use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
the complete spectral signatures. Later, some sparse coding schemes are proposed to directly utilize a single RGB image to recover the corresponding HSI [34], [35], [36]. Through leveraging statistical sparse priors of HSIs, they first build an overcomplete dictionary and then treat spectral reconstruction as a linear transformation from RGB images to HSIs. Advantageously, these algorithms not only create a convenient way to obtain HSIs but also substantially reduce hardware costs from HSI imaging devices to RGB cameras. Unfortunately, it is an extremely ill-posed conversion to predict high-dimensional spectral stimuli from 3-D RGB signals. Thus, only involving the simple and handcrafted sparse prior in the RGB-to-HSI transition, such sparse coding ways, cannot guarantee the authenticity and feasibility of the generated spectrum. More importantly, the linear transformation is not enough to characterize the hypothesis space of the underdetermined spectral super-resolution (SSR) issue.

In recent years, deep CNNs have been widely exploited in the SSR task [22], [23], [37], [38], [39], [40]. To promote the development of SSR technology, New Trends in Image Restoration and Enhancement (NTIRE) has organized two SSR competitions in 2018 [41] and 2020 [42], where numerous CNN-based methods have been proposed for hyperspectral recovery [24], [25], [26], [27], [43], [44], [45]. Through extracting the abstract deep feature prior from the dataset of RGB–HSI pairs, these approaches learn end-to-end nonlinear mappings between the RGB inputs and the HSI counterparts. Compared with the previous linear transition of sparse dictionary learning, CNN-based algorithms can greatly intensify the precision of estimated spectral signatures.

Nevertheless, several drawbacks still exist in the current CNN-based models. In particular, most of the CNN techniques almost only explore the general contextual prior of RGB inputs and learn the implicit deep feature prior via a customized network. Due to considering the generic and limited priors, these networks can merely alleviate the redundancy of this ill-posed problem to a certain extent and provide an acceptable but relatively low-precision recovered spectral from many alternative results. On this basis, there are also a few models to introduce more advanced deep feature prior into SSR, such as certain attention modules [27], [46] to improve the learning ability of CNNs and acquire a smoother solution domain from the RGB input to HSI output. However, existing attention-based modules usually adopted global-averaging scalars as channelwise squeezers, which probably failed to characterize the entire channel information, since global averaging was easily distracted by background clutter. Also, these approaches rarely further combine potential and superior prior knowledge, such as the semantic prior of RGB inputs and intrinsic statistical prior of HSIs, making SSR performance still limited. Besides, there are some SSR researches to integrate the known camera spectral sensitivity prior [47], [48], but this kind of prior is not always available in reality.

Toward the aforementioned problems, we take full advantage of multisource prior information for SSR and propose a novel holistic prior-embedded relation network (HPRN). To be specific, the backbone is repeatedly stacked via several multiresidual relation blocks (MRBs). Profiting from such multiresidual connection paradigm, the low-frequency context prior of RGB images is adequately utilized through the deep network. Furthermore, a semantic-driven spatial relation module (SSRM) is well-designed innovatively at the tail of the presented HPRN. Taking embedded semantic prior of RGB inputs as the category indexes for resolving HSIs, this module can carry out the feature aggregation across the similar spectral signatures using a semantic-based relation matrix. Such SSRM can selectively model the correlation learning of category-consistent patterns while reducing the attention of discrepant ones, which further effectively refines the quality of reconstructed spectra. In addition, we investigate a transformer-based channel relation module (TCRM), which leverages certain vectors instead of scalars as the squeezers of channelwise relation to extract feature interrelations. Coupled with transformer-style feature interactions, our TCRM can make the mapping function more robust, potentially regularizing and smoothing the one-to-many solution space. Attentively, bandwise correlations of HSIs are intrinsic and nonignorable priors determined by the hyperspectral imaging principle. To reinforce hyperspectral bandwise statistical correlations and spectral continuity, the second-order prior constraint (SOPC) is added to the loss function as an auxiliary term to assist the network training process. Experimental results on four prevalent SSR benchmarks demonstrate that our HPRN can achieve the state-of-the-art performance quantitatively and qualitatively. Meanwhile, the classification results on the remote sensing dataset validate the effectiveness and usefulness of the reconstructed spectrum. As shown in Fig. 1, our HPRN can acquire better visual quality and less reconstruction errors over other superior SSR methods.

The main contributions of this article can be summarized as follows.

1) We present a novel HPRN for SSR. Multisource and abundant priors, including spatial contexts of RGB images, semantic information of RGB signals, deep feature prior, and bandwise correlations of HSIs, are incorporated into the end-to-end mapping function, which can effectively alleviate the ill-posedness of the underconstrained SSR problem, further promoting the accuracy and fidelity of recovered HSIs.

2) Through embedding semantic prior of RGB inputs, a trainable SSRM is well-designed innovatively to perform the feature aggregation among the clustered similar characteristics of the reconstructed HSIs. Such SSRM can accomplish semantic-guided correlation learning of category-consistent patterns and effectively achieve spectral optimization of the estimating HSIs at the end of our HPRN.

3) We develop a TCRM, which leverages certain vectors rather than scalars as the descriptors of channelwise relations to explore feature interdependencies. Together with transformer-style feature interactions, this module can obtain more discriminative learning power to produce a more robust and smoother one-to-many mapping function potentially.
4) The SOPC is additionally incorporated into the loss function to assist the network learning, playing a role in maintaining the hyperspectral bandwise statistical correlations and spectral continuity. Mathematically, the SOPC term can further assist the $L_1$ loss to make the space of possible one-to-many mappings smaller to implement a high-precision spectral recovery.

5) Extensive experimental results demonstrate that our HPRN can surpass the state-of-the-art SSR methods under multiple evaluation metrics on four established benchmarks. Also, the effectiveness and usefulness of the reconstructed spectrum are proved by the classification results on the remote sensing dataset.

II. RELATED WORKS

A. Traditional Methods

In the last few decades, hyperspectral imaging has been successfully proven to be beneficial for many applications in the fields of environment, remote sensing, geography, and so on. With the progress of science and technology, hyperspectral imaging has also been continuously developed and upgraded. Traditional whisk- or push-broom imagers employ the point-by-point or line-by-line scanning manners to collect the scene radiations and reflections [15], [16]. Although these devices can acquire contiguous spectral bands among the larger range electromagnetic spectrums, they are extremely slow and time-consuming. To fulfill fast acquisition of HSIs, coded aperture snapshot hyperspectral cameras are devised to perform spectral imaging of a dynamic scene at a video rate [18], [19], [20], [21]. Despite the faster speed, there is usually a loss of spatial or spectral resolution. Unavoidably, the hardware components of the imaging equipment are still expensive. Thus, conventional RGB cameras are tried to capture the spectral signatures of natural objects [30], [31], [32], [33]. For example, Goel et al. [32] designed a low-cost multispectral system with a digital camera and a software approach that automatically analyzes the scene under controlled lighting. Oh et al. [33] built a framework for reconstructing HSIs by using multiple RGB cameras and introduced an algorithm to combine and convert these different RGB measurements into a single HSI. However, such methods have to face with a huge computational burden.

Afterward, some researchers proposed sparse coding approaches for reconstructing the corresponding HSI from a single RGB image. First, Robles-Kelly [34] extracted a set of prototypes from the training set and utilized sparse coding to predict the irradiances of HSIs. Then Arad and Ben-Shahar [35] published a large database of natural HSIs for SSR and learned an overcomplete dictionary via the K-SVD algorithm [49]. Given an RGB pixel, the dictionary representation of hyperspectral signatures would be computed with the orthogonal match pursuit iterations [50]. Purposefully, Aeschbacher et al. [36] reimplemented and pushed the performance of the work [35] and then introduced a novel shallow network based on the super-resolution method [51]. Eventually, the time-consuming and laborious hyperspectral imaging is transformed into the spectral reconstruction from an RGB signal cost-effectively and quickly. Regrettably, the SSR problem is ill-posed, which means that multiple HSIs solutions can project the same RGB input. Since the abovementioned methods only involve the handcrafted sparse prior of HSIs, they have a hard time choosing a high-confidence and high-quality spectrum from many candidates. To acquire a more accurate spectral estimation, the more priors should be plugged into the SSR process to relieve the underdetermination.

B. CNN-Based Models

Deep CNNs have become the main-stream solution to many tasks, such as super-resolution [52], [53], pansharpening [54], [55], [56], classification [57], and dehazing [58], [59]. In the early work, Galliani et al. [22] proposed a pioneering CNN model for SSR. To boost the SSR accuracy, Yan et al. [23] presented a multiscale CNN (MSCNN) to explicitly map the input RGB image to an HSI. Improved upon HSCNN [37], Shi et al. [25] built a hyperspectral reconstruction network HSCNN+, which achieved the first place on both the “Clean” and “RealWorld” tracks in the NTIRE 2018 Spectral Reconstruction Challenge [41]. In 2020, NTIRE organized the second Spectral Reconstruction Challenge [42], and a new set of CNN-based models were proposed for SSR. Among them, Li et al. [43] developed an adaptive weighted attention network for more powerful feature expression. Zhao et al. [27] adopted the PixelShuffle layer as interlevel interaction and proposed a four-level hierarchical regression network (HRNet) to explore...
context information for the spectral recovery. Moreover, Zhang et al. [26] focused on the size-specific receptive field centered at each pixel and investigated a pixel-aware deep function-mixture network (FMNet) to solve the SSR issue. He et al. [39] proposed an optimization-driven network combining with the data-driven algorithm to improve the model interpretability. Li et al. [47] presented a deep hybrid 2-D–3-D CNN with dual second-order attention, which can excavate sufficient spatial–spectral context information. Recently, Zhu et al. [38] utilized the classic gradient descent algorithm explicitly and designed a lightweight neural network, namely, AGD-Net with a multistage architecture to handle the reconstruction. Dian et al. [60] presented an imaging-model-guided network, which exploited the imaging model of SSR and spatial–spectral characteristics. Benefiting by the ability of CNN to implicitly extract abstract deep feature prior from the dataset of large RGB–HSI pairs, CNN-based methods effectively regularized the one-to-many solution and heightened the accuracy of predicting spectral signatures.

To introduce more advanced deep feature prior, there are also a few CNN-based models to design certain attention modules for learning interdependencies among intermediate features. The attention mechanism is generally regarded as an explanatory high-level prior, which imitates the human visual system that automatically captures salient parts of objects [61]. In the deep CNNs, the attention block can explore the correlation between deep features and adaptively focus on more informative ones. This can boost the learning power of the network and make the end-to-end function smoother than plain CNNs. Nathan et al. [46] presented a lightweight residual dense attention net based on attention mechanisms. Li et al. [62] put forward the 2-D channel and 3-D bandwise attention modules in their work. Peng et al. [44] devised a pixel attention module to rescale pixelwise features in all feature maps. These attention-based blocks were accustomed to adopt the scalars computed by the global average pooling as channelwise squeezers. Nevertheless, this global-averaging scalar probably failed to characterize the entire channel information since it was easily distracted by background clutter and anomaly. Thus, to acquire the more robust feature representations, we replace scalars with certain vectors to learn feature interdependencies and develop a TCRM. Meanwhile, based on that the transformer model has recently proved to be an effective tool in the computer vision [63], [64], we introduce transformer-style feature interactions into the proposed TCRM to enhance the expression ability. Furthermore, the one-to-many mapping space is regularized indirectly, and thus, we can obtain a more robust RGB-to-HSI solution.

Moreover, several related works exploited prior category information of specific objects into the spectral reconstruction. Han et al. [65] employed the unsupervised clustering to divide each RGB image into multiple classes and established the nonlinear spectral mapping for each class. Yan et al. [66] labeled a C2H-Data laboriously and directly loaded the semantic prior information into the intermediate layers of their C2H-Net. However, the former inefficiently performs the classification and reconstruction tasks step by step, while the latter relies on labor-intensive manual labeling. Innovatively, this semantic prior from RGB signals is subtly plugged into our network, and an SSRM is meticulously designed at the end of the HPRN. Such SSRM can carry out correlation learning and feature aggregation across category-consistent ranges, where the category indexes are assigned by the built-in semantic prior knowledge of RGB signals. Without extensive manual labeling and inefficiently performing spectral mapping for each class, the SSRM module can fulfill the spectral optimization of the estimating HSIs, which can jointly work with the HSIs recovery process. With additional statistical bandwise correlations prior of HSIs, our HPRN can availably mitigate uncertainty of the underdetermined SSR problem, thus increasing the precision of estimating spectra.

III. PROPOSED METHOD

A. Preliminaries and Motivation

Let \( \mathbf{I}_{\text{RGB}} \in \mathbb{R}^{B \times H \times W} \) and \( \mathbf{I}_{\text{HSI}} \in \mathbb{R}^{B \times H \times W} \) denote the given RGB image and corresponding ground HSI, respectively, where \( B, H, \) and \( W \) are the band, height, and width of an HSI cube, respectively. When the camera spectral sensitivity prior \( \mathbf{P} \in \mathbb{R}^{B \times B} \) is known, the RGB signal can be converted from the HSI measurement as follows [67]:

\[
\mathbf{I}_{\text{RGB}} = \mathbf{P} \mathbf{I}_{\text{HSI}}. \tag{1}
\]

The conversion from \( \mathbf{I}_{\text{HSI}} \) to \( \mathbf{I}_{\text{RGB}} \) is a projection from higher dimension to lower one. Conversely, recovering HSIs from RGB inputs is obviously an ill-posed problem, indicating that one RGB image can be reprojected to multiple HSIs. Thus, integrating a large amount of prior information into SSR is beneficial to regularize the solution space, which can further advance the fidelity of reconstructed HSIs. Since the camera spectral sensitivity prior is sometimes unknown in reality, mathematically, the super-resolved HSI \( \mathbf{I}_{\text{SSR}} \in \mathbb{R}^{B \times H \times W} \) can be obtained by the following formula:

\[
\mathbf{I}_{\text{SSR}} = \mathbf{P}^{-1} \mathbf{I}_{\text{RGB}} \tag{2}
\]

where \( \mathbf{P}(\cdot) \) is the prior term containing various available prior information from the SSR task. In our proposed HPRN, we merge abundant priors, including semantic categories of RGB signals, more advanced deep feature prior, and bandwise correlations of HSIs besides the general contextual prior of RGB inputs, which can constrain and optimize the one-to-many solution domain, further improving the quality of reprojected spectra.

B. Network Architecture

Fig. 2 gives an illustration of the presented HPRN. Initially, a convolutional layer is placed for shallow feature-prior learning from the input RGB context prior

\[
\mathbf{F}_0 = H_{\text{SPF}}(\mathbf{I}_{\text{RGB}}) \tag{3}
\]

where \( H_{\text{SPF}}(\cdot) \) stands for the convolution function. Then, the obtained shallow feature prior \( \mathbf{F}_0 \) is fed into the backbone network for the deep feature-prior extraction. Concretely, the
basic framework is composed of several MRBs. The whole process is expressed as

$$F_{DFP} = H_{MRBs}(F_0)$$  \hspace{1cm} (4)

where $H_{MRBs}()$ denotes the deep feature-prior extraction. Compared with the classical residual module, our MRB adopts the multiresidual connections, which can make the best of the low-frequency context-prior of RGB images. Next, the deep feature prior $F_{DFP}$ is injected into another convolutional layer and interacts with the shallow feature prior $F_0$ to form a global residual summation (GRS)

$$F_{GRS} = H_{GRS}(F_{DFP}) + F_0$$  \hspace{1cm} (5)

where $H_{GRS}()$ is the convolutional weights. Such operation can avoid causing gradients to disappear or explode and strengthen the stability of network training. At the end of our HPRN, we set up the reconstruction part consisting of a single convolution and the developed SSRM. The output feature of the convolution is denoted as $F_{SSR}$, which is regarded as one of the inputs of the latter SSRM module. The whole reconstruction part is represented as a function as follows:

$$I_{SSR} = H_{Rec}(F_{GRS}) = H_{HPRN}(I_{RGB})$$  \hspace{1cm} (6)

where $I_{SSR}$ and $H_{HPRN}()$ denote the final reconstructed HSI and the function of our presented HPRN, respectively. The detailed implementation of SSRM is described in Section III-D.

C. Transformer-Based Channel Relation Module

In this section, we would introduce the developed TCRM module in detail. Ideally, given an input feature map $F \in \mathbb{R}^{C \times H \times W}$, $F = [f_1, f_2, \ldots, f_C]$, where $C$ is the number of channels., as shown in Fig. 3, the feature $F$ is first divided into $4 \times 4$ grids evenly along the spatial dimension. Then, the local-range-averaging pooling is employed to generate the squeezed vectors $F_{SV} \in \mathbb{R}^{C \times 4 \times 4}$ for all channels

$$F_{SV} = H_{LAP}(F_{in})$$  \hspace{1cm} (7)

where $H_{LAP}()$ denotes the local average pooling function. Compared with global-averaging single-number scalar to characterize the entire channel information, one local-range-averaging multiple-number vector can provide a more robust representation. The reason is that the prominent targets generally exist in a certain local position, and the global-average downsampling may be biased by the background noise and anomaly in other positions. Moreover, a vector of multiple
numbers contains richer and finer information than a scalar of a single number. To capture nonlocal (also local) relations among channels, a transformer-based block is applied to perform the feature interaction. Concretely, we adopt its core component i.e., the multihead self-attention layer, which can perform the feature interaction. Among channels, a transformer-based block is applied to numbers containing richer and finer information than a scalar.

The diagram of SSRM is shown in Fig. 4. Intuitively, the input of SSRM module is constituted of the semantic prior \( S_{\text{RGB}} \) of the RGB image and the feature \( F_S \). \( S_{\text{RGB}} \) is obtained through the widely used simple linear iterative clustering (SLIC) superpixel algorithm [69]. \( F_S \) comes from the reconstruction part in Section III-B. Formally, we seed \( F_S \) to two 1 x 1 convolutional layers \( \phi(\cdot) \) and \( \psi(\cdot) \) to generate two new features \( D \) and \( E \), respectively, where \( \{D, E\} \subseteq \mathbb{R}^{B \times H \times W} \). Then, \( S_{\text{RGB}} \) is embedded as the indexes to unfold and order the features \( \{D, E\} \) along the \( H \times W \) direction

\[
\{D', E'\} = H_{\text{UO}}(\{D, E\})
\]
where \( \{D', E'\} \in \mathbb{R}^{B \times (H \times W)} \). Each row of \( D' \) and \( E' \) is sorted according to the category indexes. To aggregate features of semantically consistent contents in parallel, we divide the \( H \times W \) 1-D \( B \)-length features from \( \{D', E'\} \) into \( G \) groups spatially, and the results are expressed as \( \{D''_g, E''_g\} \in \mathbb{R}^{G \times B \times N} \). Ideally, \( H \times W \) should be exactly divisible by \( G \). When it cannot be satisfied in fact, the mirror filling of boundaries is necessary. For \( \{D'', E''\} \), the \( N \)-1-D \( B \)-length items of most groups belong to the same label. Inevitably, since the feature number of each category is actually unbalanced, the \( N \) elements in partial groups contain a few features of the neighboring category. Next, we reshape and transpose these two features to \( \{D'' \in \mathbb{R}^{G \times N \times B}, E'' \in \mathbb{R}^{G \times B \times N}\} \). A semantic-embedded relation matrix \( Z \) can be obtained by

\[
Z = \text{softmax}(D'' \otimes E'') \quad (12)
\]

where \( \otimes \) denotes the batch matrix multiplication and \( Z \in \mathbb{R}^{G \times N \times N} \). Briefly, this computation broadcasts along the ground dimension. For the \( g \)-th group \( Z_g \in \mathbb{R}^{N \times N} \), \( N \) weights of its \( i \)-th row encode the dependence between the \( i \)-th 1-D \( B \)-length feature of \( D''_g \) and all the 1-D \( B \)-length features of \( E''_g \). Similarly, another \( 1 \times 1 \) convolution \( \chi(\cdot) \) is adopted to acquire the new feature \( Y \in \mathbb{R}^{B \times H \times W} \). Also, the same dimension conversion as \( D \) is implemented to become \( Y'' \in \mathbb{R}^{G \times N \times B} \). Finally, the feature aggregation across the clustered similar characteristics is performed by

\[
I_{SSR} = H_{FR}(Z \otimes Y'') \quad (13)
\]

where \( H_{FR}(\cdot) \) is the folding and reordering operations, which are the inverse process of \( H_{UO}(\cdot) \). At this point, we illustrate the process of the SSRM with one scale (i.e., the number of categories and Fig. 4 shows eight example categories). To reduce the error of once segmentation result, we employ multiscale SLIC to produce several clustering mappings with different numbers of categories parallelly. Finally, multiscale SSRM is combined by each single scale result via a \( 1 \times 1 \) convolution.

E. Second-Order Prior Constraint

Second-order statistics play a critical role to explore the correlations of different samples [43], [70]. From the imaging principle of HSIs, there are mathematical correlations among different bands and spectral consistency along the band direction. Taking the reconstructed spectral \( I_{SSR} \in \mathbb{R}^{B \times H \times W} \) as an example, we reshape it as \( I_{SSR} \in \mathbb{R}^{B \times (H \times W)} \), which means that there are \( B \) samples of \( n = H \times W \) length. Then, the sample covariance matrix can be calculated by

\[
X_{SSR} = I_{SSR} \otimes \Sigma \otimes I_{SSR}^T \quad (14)
\]

where \( \Sigma = (1/n)(\Sigma - (1/n)1) \). \( \Sigma \) and \( 1 \) are \( n \times n \) identity matrix and matrix of all ones. \( \otimes \) denotes the matrix multiplication. The element in the \( i \)-th row and \( j \)-th column of matrix \( X_{SSR} \in \mathbb{R}^{B \times B} \) represents the correlation between the \( i \)-th band and the \( j \)-th one for \( I_{SSR} \). In the same way, we can get the normalized covariance matrix \( X_{HSI} \) of ground HSI \( I_{HSI} \). In order to maintain the mathematical correlation and spectral consistency between hyperspectral bands, the SOPCs are incorporated into the loss function

\[
L(\Theta) = \|I_{HSI} - I_{SSR}\|_1 + \tau \|X_{HSI} - X_{SSR}\|_1 \quad (15)
\]

where \( \Theta \) denotes the parameter set of our HPRN and \( \tau \) is the tradeoff weight. In our experimental settings, the tradeoff value was experimentally tried with \{0.2, 0.5, 1, 2, 5, 10\} and finally determined to be 2 using the NTIRE2018 validation set. From the mathematical perspective, the SOPC term can help \( L1 \) loss to compress the space of the possible one-to-many mapping function to achieve a high-precision spectral recovery.

IV. EXPERIMENTS

A. Experimental Setting

1) SSR Benchmark Datasets: The presented HPRN is evaluated on four public datasets, i.e., both “Clean” and “Real World” tracks of two SSR challenges NTIRE2018\(^1\) [41] and NTIRE2020\(^2\) [42]. For “Clean” tracks, the HSIs are estimated from noise-free RGB images, which are calculated numerically using the ground-truth HSI and given spectral sensitivity functions. “Real World” tracks simulate capturing by an uncalibrated and unknown camera, where the HSIs are recovered from noisy JPEG-compressed RGB images. Certainly, the camera response functions for the same tracks of the NTIRE2020 are changed over the NTIRE2018. A Specim PS Kappa DX4 hyperspectral camera is used for collecting NTIRE2018 datasets with a rotary stage for spatial scanning, while NTIRE2020 datasets are acquired by a stand-alone, battery-powered, push-broom spectral imaging system equipped with a Specim IQ mobile hyperspectral camera. The NTIRE2018 dataset consists of 256 training RGB–HSI pairs, five validating RGB–HSI pairs, and ten testing RGB inputs. The HSIs contain 31 spectral bands (400–700 nm at roughly 10-nm increments) with 1392 × 1300 size. Also, there are 450 training RGB–HSI pairs, ten validating RGB–HSI pairs, and 20 testing RGB images. All HSIs possess the size of 512 × 482 with 31 bands from 400 to 700 nm at 10-nm steps. Due to the unavailable official testing spectra, we decide to take the official verification set as our test set and randomly choose several images from the official training as our verification set in this article. The rest of the official training is adopted as our training set. The specific description refers to Tables I and II.

2) Remote Sensing Dataset: The University of Pavia (UP)\(^3\) dataset is chosen to verify the effectiveness of the reconstructed spectrum based on classification results [71]. The

---

1https://data.vision.ee.ethz.ch/cvl/ntire18/
2https://data.vision.ee.ethz.ch/cvl/ntire20/
3http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes#Pavia_University_scene
TABLE II
DETAILED DATASET DIVISION OF NTIRE2020 “CLEAN” AND “REAL WORLD” TRACKS

| Our testing set | Our verification set | Our training set |
|-----------------|----------------------|------------------|
| 451,453,455,456,457, 079,089,255,304,363, 372,387,422,434,446. | The remain of the 459,462,463,464,465. | The remain of the official training set. |

Fig. 5. UP dataset. (a) Simulated RGB input. (b) False-color image (25th, 45th, and 70th bands). (c) Label. (d) Standard training samples.

Reflective Optics System Imaging Spectrometer (ROSIS) sensor is employed to collect these data over the area of Pavia, Northern Italy. The scene has 103 spectral bands with a range of 430–860 nm, and the spatial resolution is 1.3 m with the size of 610 × 340. There are about 43,923 samples labeled with nine classes. According to [71], the 53rd, 31st, and 7th bands of the original HSI are selected to form the simulated RGB data. The simulated RGB input, the false-color image of the HSI, the ground-truth label, and the standard training samples are drawn in Fig. 5(a)–(d), respectively. The corresponding testing set is the remaining part of the ground-truth label except the training samples.

3) Quantitative Metrics: To evaluate the performance of HPRN quantitatively, five metrics, including the mean relative absolute error (MRAE), root-mean-square error (RMSE), spectral angle mapper (SAM), peak signal-to-noise ratio (PSNR), and average structural similarity (ASSIM), are utilized.

\[
\text{MRAE} = \frac{1}{N} \sum_{p=1}^{N} \left| \frac{I_{\text{HSI}}^{(p)} - I_{\text{SSR}}^{(p)}}{I_{\text{HSI}}^{(p)}} \right| 
\]

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{p=1}^{N} \left( I_{\text{HSI}}^{(p)} - I_{\text{SSR}}^{(p)} \right)^2} 
\]

\[
\text{SAM} = \frac{1}{M} \sum_{v=1}^{M} \frac{180}{\pi} \arccos \left( \frac{\langle I_{\text{HSI}}^{(v)}, I_{\text{SSR}}^{(v)} \rangle}{\| I_{\text{HSI}}^{(v)} \|_2 \| I_{\text{SSR}}^{(v)} \|_2} \right) 
\]

\[
\text{PSNR} = -10 \log_{10} \left( \frac{\text{MSE}(I_{\text{HSI}}^{(b)}, I_{\text{SSR}}^{(b)})}{B} \right) 
\]

\[
\text{ASSIM} = \frac{1}{B} \sum_{b=1}^{B} \text{SSIM}(I_{\text{HSI}}^{(b)}, I_{\text{SSR}}^{(b)}) 
\]

where \(I_{\text{HSI}}^{(p)}\) and \(I_{\text{SSR}}^{(p)}\) represent the \(p\)th pixel of the super-resolved and ground-truth HSIs, respectively.

\(\langle I_{\text{HSI}}^{(v)}, I_{\text{SSR}}^{(v)} \rangle\) are the dot product of the two \(v\)th spectral vectors between the ground truth and the estimated spectrum. \(\| \cdot \|_2\) denotes the 2-norm function. \(N, M, \), and \(B\) are the number of pixels, spectral vectors, and bands of the HSI cube, respectively. \(I_{\text{HSI}}^{(b)}\) and \(I_{\text{SSR}}^{(b)}\) are the \(b\)th band of the ground truth and the recovered HSI. MSE(·) computes the mean squared error of the inputs. SSIM(·) calculates the SSIM value of a typical band. Mathematically, the smaller the MRAE, RMSE, and SAM, the better predicted HSIs. Moreover, the larger the PSNR and SSIM, the fewer errors between the reconstructed result and ground truth.

4) Implementation Details: For the NTIRE2020 and NTIRE2018 datasets, 64 × 64 patches are cropped from the original RGB–HSI counterparts in the training procedure. The ADAM [72] is adopted as the optimizer of our HPRN. The exponential decay rates are set as \(\beta_1 = 0.9\) and \(\beta_2 = 0.99\) for the first and second moment estimates, respectively. The learning rate is set to 0.00012, and the polynomial function is applied as the decay means with power = 1.5. As for the backbone framework, there are \(M = 10\) MRBs, where all intermediate features have 200 channels. The parameter \(r\) in TCRM is 16 and the group size \(g = 64\) in SRRM. Besides, the best results are reported within 100 epochs, and the proposed HPRN network has been implemented on the Pytorch framework via an NVIDIA 2080Ti GPU. The package of automatic mixed precision is also employed to speed up the training of the network. For the remote sensing UP dataset, we arrange the 11 × 11 neighbor patch to incorporate the spectral–spatial statistics according to the positions of training and testing samples, following the general setting of the CNN-based HSI classification method [71].

B. Results on SSR Benchmark Datasets

To evaluate the robustness and generalization of the proposed network, we compare with seven state-of-the-art methods, including Galliani [22], MSCNN [23], UNet [24], HSCNN+ [25], FMNet [26], HRNet [27], and AGDNet [38]. For a fair comparison, we run their released models on the training set of this article. Also, the best model is chosen via the validation set and further evaluated on the final testing set on four established benchmarks.

1) Quantitative Results: The quantitative results of final test set of the NTIRE2018 and NTIRE2020 tracks are summarized in Tables III and IV, respectively. As we can see, our method consistently outperforms other approaches across the whole benchmarks in terms of five numerical metrics. From MRAE and RMSE assessments, our HPRN obtains the smallest values on all the datasets, which suggests that our estimating spectrum is the most accurate. It is worth noting that the proposed network improves SAM by 18.5%, 2.9%, 4.9%, and 1.2% than the second best results on the NTIRE2018 “Clean,” NTIRE2018 “Real World,” NTIRE2020 “Clean,” and NTIRE2020 “Real World” tracks. This indicates that our reconstructed HSIs contain better spectral continuity and authenticity. For structural measurement PSNR and SSIM, our method acquires larger improvement compared with these existing models. The reason may be that HPRN can adequately
mine and utilize spatial contexts and semantic category priors of RGB images.

2) **Qualitative Evaluation:** To further illustrate the superior performance of our network, Fig. 6 shows certain examples of MRAE heat maps generated by our HPRN and other state-of-the-art methods. The “BGU_HS_00263” HSI (530 nm) of NTIRE2018 “Clean,” “BGU_HS_00265” HSI (550 nm) of NTIRE2018 “Real World,” “ARAD_HS_00453” HSI (570 nm) of NTIRE2020 “Clean,” and “ARAD_HS_00457” HSI (590 nm) of NTIRE2020 “Real World” datasets from top to bottom. The best view on the screen.

Fig. 6. Visual comparison on the “BGU_HS_00263” HSI (530 nm) of NTIRE2018 “Clean,” “BGU_HS_00265” HSI (550 nm) of NTIRE2018 “Real World,” “ARAD_HS_00453” HSI (570 nm) of NTIRE2020 “Clean,” and “ARAD_HS_00457” HSI (590 nm) of NTIRE2020 “Real World” datasets from top to bottom. The best view on the screen.
TABLE III
QUANTITATIVE RESULTS OF FINAL TEST SET OF THE NTIRE2018 “CLEAN” AND “REAL WORLD” TRACKS. THE BEST AND SECOND BEST RESULTS ARE HIGHLIGHTED AND UNDERLINED, RESPECTIVELY

| Model         | Track 1: Clean | Track 2: Real World |
|---------------|----------------|---------------------|
|               | MRAE($) | RMSE($) | SAM($) | PSNR($) | SSIM(↑) | MRAE($) | RMSE($) | SAM($) | PSNR($) | SSIM(↑) |
| Galliani[22]  | 0.05130 | 37.68  | 1.775  | 43.08   | 0.99606 | 0.06613 | 57.98  | 2.180  | 37.66   | 0.99155 |
| MSCNN[23]     | 0.03304 | 25.99  | 1.725  | 45.45   | 0.99748 | 0.04542 | 32.90  | 2.190  | 43.33   | 0.99405 |
| UNet[24]      | 0.01552 | 15.39  | 1.144  | 53.06   | 0.99897 | 0.03151 | 24.15  | 1.730  | 46.55   | 0.99564 |
| FMNet[26]     | 0.01404 | 13.22  | 1.010  | 53.87   | 0.99923 | 0.03053 | 23.43  | 1.660  | 46.83   | 0.99588 |
| HSCNN+[25]    | 0.00373 | 13.02  | 0.988  | 54.02   | 0.99922 | 0.03042 | 23.60  | 1.667  | 46.84   | 0.99577 |
| HRNet[27]     | 0.00395 | 13.38  | 1.015  | 54.00   | 0.99923 | 0.02915 | 22.84  | 1.580  | 47.17   | 0.99601 |
| AGD-Net[38]   | 0.01463 | 14.13  | 1.077  | 49.86   | 0.99906 | 0.03131 | 23.87  | 1.709  | 44.99   | 0.99560 |
| Ours          | 0.01155 | 10.44  | 0.805  | 55.14   | 0.99943 | 0.02896 | 22.18  | 1.534  | 47.12   | 0.99603 |

TABLE IV
QUANTITATIVE RESULTS OF FINAL TEST SET OF THE NTIRE2020 “CLEAN” AND “REAL WORLD” TRACKS. THE BEST AND SECOND BEST RESULTS ARE HIGHLIGHTED AND UNDERLINED, RESPECTIVELY

| Model         | Track 1: Clean | Track 2: Real World |
|---------------|----------------|---------------------|
|               | MRAE($) | RMSE($) | SAM($) | PSNR($) | SSIM(↑) | MRAE($) | RMSE($) | SAM($) | PSNR($) | SSIM(↑) |
| Galliani[22]  | 0.10720 | 0.0275 | 4.351  | 34.03   | 0.98237 | 0.11959 | 0.0352 | 4.761  | 31.24   | 0.96686 |
| MSCNN[23]     | 0.06835 | 0.0198 | 3.514  | 37.31   | 0.99085 | 0.09023 | 0.0214 | 4.271  | 35.28   | 0.97567 |
| UNet[24]      | 0.04389 | 0.0161 | 2.923  | 42.72   | 0.99444 | 0.07274 | 0.0193 | 3.725  | 36.75   | 0.98091 |
| HSCNN+[25]    | 0.04097 | 0.0154 | 2.776  | 42.83   | 0.99527 | 0.07148 | 0.0191 | 3.668  | 36.80   | 0.98180 |
| FMNet[26]     | 0.04030 | 0.0148 | 2.688  | 43.08   | 0.99595 | 0.07217 | 0.0190 | 3.744  | 36.89   | 0.98215 |
| HRNet[27]     | 0.04045 | 0.0154 | 2.729  | 43.26   | 0.99601 | 0.07156 | 0.0187 | 3.687  | 36.96   | 0.98284 |
| AGD-Net[38]   | 0.04131 | 0.0146 | 2.838  | 43.44   | 0.99536 | 0.07517 | 0.0199 | 3.922  | 36.61   | 0.98081 |
| Ours          | 0.03784 | 0.0134 | 2.557  | 44.19   | 0.99654 | 0.07048 | 0.0185 | 3.623  | 37.12   | 0.98727 |

Fig. 7. Spectral response curves of selected several spatial points from the reconstructed HSIs. (a) and (b) NTIRE2020 “Clean” and “Real World” tracks, respectively. (c) and (d) NTIRE2018 “Clean” and “Real World” track, respectively.

World” are from top to bottom. Visually, the bluer the displayed color, the lower the error of predicting HSIs. Overall, we can find that the presented HPRN produces the optimal visualizations, which are much close to the ground truth in various challenging scenarios. In particular, compared with the two earlier approaches Galliani and MSCNN, the superiority of the proposed HPRN is quite obvious. The visual performance is also in line with the numerical comparisons. In terms of spectral dimensions, Fig. 7 shows the spectral response curves of selected several spatial points from the reconstructed HSIs. The change of the spectral curve is more representative of the essential attributes of the ground objects. From the qualitative evaluation, our recovered HSIs yield a spectral curve that is more consistent with the ground truth, which can also be reflected through the SAM criterion in Tables III and IV. To be specific, this observation can infer that the SOPC plays an important role in maintaining the hyperspectral bandwise statistical correlations and spectral continuity.

3) Model Efficiency Analyses: Table V states the comparisons of computational parameters and running times of recent deep CNN-based SSR methods. The actual memory footprint of the model is represented as “size” in the second column of Table V, which refers to the memory size occupied by the weights of trained-finished CNN-based models. The running time refers to the average inference time of multiple 512 × 512 images on an NVIDIA 2080ti GPU and an Intel Core i9-9900 CPU. Specifically, the segmentation preprocessing is performed simultaneously, while the original images are cropped as numerous patches on the CPU, and then, they are all prestored on the disk. The input of our model includes the RGB image and the corresponding segmentation preprocessing...
Fig. 8. Classification results on the UP dataset. (a) False-color image. (b) Label. (c) Simulated RGB. (d) Reconstructed HSI. (e) Ground-truth HSI. (f) Categories.

TABLE V

| Method    | Params(M) | Sizes(MB) | Times(ms) | MRAE(\textdegree) |
|-----------|-----------|-----------|-----------|-------------------|
| Galliani  | 2.124     | 8.30      | 9.04      | 0.05130           |
| MSCNN     | 26.69     | 101       | 3.12      | 0.03304           |
| UNet      | 1.044     | 3.94      | 1.28      | 0.01552           |
| FMNet     | 6.948     | 26.5      | 5.67      | 0.01404           |
| HSCNN+    | 19.55     | 74.6      | 3.47      | 0.01373           |
| HRNet     | 31.71     | 120       | 13.0      | 0.01395           |
| AGD-Net   | 0.772     | 1.68      | 12.3      | 0.01463           |
| Ours      | 14.89     | 56.8      | 12.7      | 0.01155           |

result from the CPU to the GPU, while the input of other comparison methods is the RGB signal. Furthermore, the runtime represents the inference efficiency of the CNN-based model on the GPU, without considering the preprocessing operations on the CPU. Before the measurement, the machine will be warmed in advance. Due to the frequent index operations of SSRM, the testing time of the proposed HPRN reaches 12.72 ms. Among these networks, the parameters of Galliani, UNet, and AGD-Net are small, but the performance is not high. Compared with MSCNN, HSCNN+, and HRNet, our approach contains less parameters and better precision, which indicates that our method can balance the algorithm performance and model complexity.

C. Results on Remote Sensing Dataset

To investigate the applicability of the reconstructed HSIs, we introduce the hyperspectral classification task to evaluate the SSR performance. Here, we adopt the popular UP dataset to compare different classificational results, and the pixels are classified into nine classes through employing a classical CNN-based hyperspectral classification network [73]. This model consists of three convolutional units and two fully connected layers. Also, Table VI shows the corresponding scores, including overall accuracy (OA), average accuracy (AA), and Kappa coefficient (Kappa). OA is the proportion of correctly classified pixels to all pixels. AA represents the average of classification accuracy for each class. Kappa indicates the degree of consistency between the classification result and the ground truth. These three metrics follow the same rule: the larger the value, the better the classification effect. We can observe that our reconstructed HSI achieves better OA, AA, and Kappa than the simulated RGB image and is closer to the classification performance of using the ground-truth label, which is also reflected in Fig. 8. This measurement demonstrates that the presented HPRN can effectively predict spectral information on the remote sensing dataset. Meanwhile, since the UP dataset contains 103 bands, which is more than the 31 bands of NTIRE2018 and NTIRE2020, this result also verifies that our algorithm has the ability to reconstruct more bands. To fulfill this, only the number of output channels of the last convolution layer needs to be modified according to the number of bands in the dataset.

D. Ablation Studies

In this section, we conduct extensive ablation studies to thoroughly analyze the proposed HPRN. All experimental results are trained using the NTIRE2018 “Clean” training set and reported on the NTIRE2018 “Clean” validating data. The baseline model is the backbone network consisting of ten MRBs, which only contain ordinary convolutional layers and trained by a single $L_1$ loss constraint. For the TCRM, we explore its critical embedded location and the squeezed vector size. Correspondingly, we evaluate the effects of sharing embedding functions, the group size, and the scale settings for the SSRM. Subsequently, the combined performances of SSRM, TCRM, and SOPC are studied in detail.

1) Explore the Location of the TCRM: We investigate the influence of different positions for the developed TCRM and the results are listed in Table VII. The 1-pos, 2-pos, and 3-pos denote before the first, second, and third residual connection of the MRB, respectively. When our TCRM is embedded before the third residual summation of the MRB, it can produce better MRAE and SAM metrics than other two
TABLE VII
EXPLORE THE LOCATION OF THE TCRM WITH THE SQUEEZED VECTOR SIZE 4 × 4

| Description  | MRAE(%) | SAM(%) |
|--------------|---------|--------|
| baseline     | 0.01538 | 1.089  |
| 1-pos        | 0.01493 | 1.036  |
| 2-pos        | 0.01501 | 1.049  |
| 3-pos        | 0.01487 | 1.020  |
| multi-pos    | 0.01481 | 1.011  |

TABLE VIII
EXPLORE THE SQUEEZED VECTOR SIZE OF THE TCRM

| Description | MRAE(%) | SAM(%) | Params(M) | MACs(K) |
|-------------|---------|--------|-----------|---------|
| baseline    | 0.01538 | 1.089  | 8.25      | 48.04   |
| 1 × 1       | 0.01516 | 1.051  | 10.3      | 58.24   |
| 2 × 2       | 0.01504 | 1.030  | 8.38      | 48.64   |
| 4 × 4       | 0.01487 | 1.020  | 41.1      | 218.1   |
| 8 × 8       | 0.01504 | 1.042  |           |         |

TABLE IX
EFFECTS OF SHARING WEIGHTS OF EMBEDDING FUNCTIONS FOR THE SSRM

| Description | MRAE(%) | SAM(%) |
|-------------|---------|--------|
| baseline    | 0.01538 | 1.089  |
| ϕ(·) = ψ(·) | 0.01494 | 1.055  |
| ϕ(·) ≠ ψ(·) | 0.01514 | 1.067  |

2) Explore the Squeezed Vector Size of the TCRM: As discussed above, the simple scalars are replaced with certain vectors as the squeezers of channelwise relation to extract feature interdependences. Table VIII summarizes the impact of different squeezed vector sizes. “Params” and “MACs” denote the additional number of parameters and multiply–accumulate operations relative to the baseline model. Results imply that 4 × 4 = 16 length vector can provide the more robust feature representations. Once the size of this vector is further enlarged, for example, 8 × 8, the model complexity also increases accordingly, which may lead to low-efficient transformer-style feature interactions and reduce the accuracy of SSR.

3) Effects of Sharing Embedding Functions for the SSRM: Table IX explores the effects of sharing weights of embedding functions for the SSRM. We can see that ϕ(·) = ψ(·) not only achieves lower MRAE and SAM measurements but also saves the number of parameters. The main reason lies in that sharing embedding functions has a greater possibility to generate a more discriminative semantic-based relation matrix. Once the two embedding functions are different, the clustered-similar feature aggregation can be biased by their transformed projections. Therefore, we adopt the sharing embedding functions in the SSRM.

4) Effects of the Group Size of the SSRM: The feature aggregation of SSRM is performed based on semantic-driven groups; hence, the group size closely affects the effectiveness of SSRM. An appropriate G value can assign category-consistent elements to each group as much as possible, so as to fulfill highly effective correlation learning. To be specific, we set G = {16, 36, 64, 100, 144, 256} and the experimental result is shown in Fig. 9. From the beginning, as the pixels involved in the correlation calculation increase, the error naturally continues to decrease. However, when G is further enlarged, some newly introduced parts with little correlation would have a counterproductive effect, leading to a growing error. This shows that it is more crucial to choose informative features than to consider more.

5) Effects of the Scales Settings of the SSRM: Some errors on the clustering boundary may exist through running once SLIC algorithm with one scale (i.e., the number of categories). Moreover, different RGB signals may also be identified to inconsistent number of category attributes. Due to the two reasons, the multiscale SLIC is employed parallelly, and the final multiscale SSRM is weighted by each single scale one. Table X reveals the scale settings of the SSRM. We try the four scales with various combinations to compare their effects and adopt the scale settings of 8, 12, 16, and 20 based on the experimental results.

6) Effectiveness of Different Modules: The proposed method consists of the three key modules, including the developed TCRM, well-designed SSRM, and incorporated SOPC.
To determine the necessity of these components, we evaluate the performance of HPRN with its variants that remove a part of the network structure in Table XI. With TCRM, with SSRM, and with SOPC represent the variant containing the single key module on the baseline network. Their results successively prove the effectiveness of each individual section, compared with the baseline MRAE, SAM, PSNR, and SSIM scores. Immediately after, the above three components are combined in pairs, and the corresponding results produce less error than using any single module. Finally, the presented HPRN model containing all parts achieves the best effect, which demonstrates that all modules are necessary for the proposed method to effectively alleviate the ill-posedness of this underconstrained SSR problem and promote the accuracy of recovered HSIs.

V. CONCLUSION
In this article, a novel HPRN is proposed for SSR. Multisource and abundant priors, including spatial contexts of RGB images, semantic categories of RGB signals, deep feature prior, and bandwise correlations of HSIs, are incorporated into an end-to-end mapping function, which can effectively alleviate the ill-posedness of the SSR problem. To be specific, the backbone network is repeatedly stacked by several MRBs, where the low-frequency context prior of RGB images is adequately utilized by the deep network. Through embedding semantic prior of RGB inputs, a trainable SSRM is well-designed innovatively to perform the category-guided feature aggregation and effectively achieve spectral optimization of the estimating HSIs. Furthermore, a TCRM is investigated to leverage one local-range-averaging multiple-number vector to represent the channel of a deep feature. Coupled with transformer-style feature interactions, the TCRM can obtain more discriminative learning power and make the RGB-to-HSI solution more robust. Finally, the SOPC is incorporated into the loss function to guide the HSI reconstruction to maintain the mathematical correlation and spectral consistency among hyperspectral bands. Extensive experimental results demonstrate that the presented HPRN can achieve superior performance on four SSR benchmarks under quantitative and perceptual comparisons. Also, the effectiveness of the estimating spectrum is verified by the classification results on the remote sensing dataset. In the future, the work of this research will mainly focus on the lightweight CNN-based methods and unsupervised deep learning.

REFERENCES
[1] K. Wei, Y. Fu, and H. Huang, “3-D quasi-recurrent neural network for hyperspectral image denoising,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 363–375, Jan. 2021.
[2] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, “Hyperspectral remote sensing data analysis and future challenges,” *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
[3] D. Hong et al., “More diverse means better: Multimodal deep learning meets remote-sensing imagery classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, Aug. 2021.
[4] X. Wu, D. Hong, and J. Chanussot, “Convolutional neural networks for multimodal remote sensing data classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5517010.
[5] R. Dian, S. Li, and L. Fang, “Learning a low tensor-train-rank representation for hyperspectral image super-resolution,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2672–2683, Sep. 2019.
[6] J.-F. Hu, T.-Z. Huang, L.-J. Deng, T.-X. Jiang, G. Vivone, and J. Chanussot, “Hyperspectral image super-resolution via deep spatio-spectral attention convolutional neural networks,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 7251–7265, Dec. 2022.
[7] G. Lu and B. Fei, “Medical hyperspectral imaging: A review,” *J. Biomed. Opt.*, vol. 19, no. 1, Jan. 2014, Art. no. 010901.
[8] X. Wei, W. Li, M. Zhang, and Q. Li, “Medical hyperspectral image classification based on end-to-end fusion deep neural network,” *IEEE Trans. Instrum. Meas.*, vol. 68, no. 11, pp. 4481–4492, Nov. 2019.
[9] B. Xi et al., “Semi-supervised cross-scale graph prototypical network for hyperspectral image classification,” *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 23, 2022, doi: 10.1109/TNNLS.2022.3215751.
[10] J. Li, X. Zhao, Y. Li, Q. Du, B. Xi, and J. Hu, “Classification of hyperspectral imagery using a new fully convolutional neural network,” *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 292–296, Feb. 2018.
[11] D. Hong et al., “SpectralFormer: Rethinking hyperspectral image classification with transformers,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518615.
[12] L. Mou, X. Lu, X. Li, and X. Zhu, “Nonlocal graph convolutional networks for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8246–8257, Dec. 2020.
[13] H. Chen, G. Yang, and H. Zhang, “Hider: A hyperspectral image denoising transformer with spatial–spectral constraints for hybrid noise removal,” *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Oct. 31, 2022, doi: 10.1109/TNNLS.2022.3215751.
[14] D. Hong et al., “Endmember-guided unmixing network (EGU-Net): A general deep learning framework for self-supervised hyperspectral unmixing,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6518–6531, Nov. 2021.
[15] R. O. Green et al., “Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS),” *Remote Sens. Environ.*, vol. 65, no. 3, pp. 227–248, Sep. 1998.
[16] J. James, *Spectrograph Design Fundamentals*. Cambridge, U.K.: Cambridge Univ. Press, 2007.
[17] Y. Y. Schechner and S. K. Nayar, “Generalized mosaicing: Wide field of view multispectral imaging,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 10, pp. 1334–1348, Oct. 2002.
[18] A. Wagadarikar, R. John, R. Willett, and D. Brady, “Single disperser design for coded aperture snapshot spectral imaging,” *Appl. Opt.*, vol. 47, no. 10, p. B44, 2008.

### TABLE XI

| Description | SSRM | TCRM | SOPC | MRAE(\(\downarrow\)) | SAM(\(\downarrow\)) | PSNR(\(\uparrow\)) | SSIM(\(\uparrow\)) |
|-------------|------|------|------|----------------------|------------------|-----------------|------------------|
| baseline    | ×    | ×    | ×    | 0.01538              | 1.089            | 51.35           | 0.99915          |
| with SSRM   | ✓    | ✗    | ✗    | 0.01494              | 1.055            | 51.45           | 0.99919          |
| with TCRM   | ✓    | ✓    | ✗    | 0.01487              | 1.020            | 51.41           | 0.99919          |
| with SOPC   | ✗    | ✓    | ✓    | 0.01520              | 1.075            | 51.39           | 0.99915          |
| w/o TCRM    | ✓    | ✗    | ✓    | 0.01484              | 1.046            | 51.62           | 0.99919          |
| w/o SSRM    | ✓    | ✗    | ✓    | 0.01478              | 1.017            | 51.49           | 0.99922          |
| w/o SOPC    | ✓    | ✓    | ✗    | 0.01455              | 1.018            | 51.65           | 0.99921          |
| HPRN        | ✓    | ✓    | ✓    | 0.01441              | 1.009            | 51.77           | 0.99922          |
H. Wu et al., “CVT: Introducing convolutions to vision transformers,” 2021, arXiv:2103.15808.

X. Han, J. Yu, J.-H. Xue, and W. Sun, “Spectral super-resolution for RGB images using class-based BP neural networks,” in Proc. Digit. Image Comput., Techn. Appl. (DICTA), Dec. 2018, pp. 1–7.

L. Yan, X. Wang, M. Zhao, M. Kaloorazi, J. Chen, and S. Rahardja, “Reconstruction of hyperspectral data from RGB images with prior category information,” IEEE Trans. Comput. Imag., vol. 6, pp. 1070–1081, 2020.

Y. Fu, T. Zhang, Y. Zheng, D. Zhang, and H. Huang, “Joint camera spectral sensitivity selection and hyperspectral image recovery,” in Proc. Eur. Conf. Comput. Vis. (ECCV), Oct. 2018, pp. 788–804.

R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, “SLIC superpixels compared to state-of-the-art superpixel methods,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 11, pp. 2274–2282, Nov. 2011.

D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in ICLR, 2015, vol. 5, no. 2, pp. 848–857.

X. Xu, W. Li, Q. Ran, Q. Du, L. Gao, and B. Zhang, “Multisource remote sensing data classification based on convolutional neural network,” IEEE Trans. Geosci. Remote Sens., vol. 56, no. 2, pp. 937–949, Feb. 2018.

D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” 2014, arXiv:1412.6980.

Yunsong Li (Member, IEEE) received the M.S. degree in telecommunications and information systems and the Ph.D. degree in signal and information processing from Xidian University, Xi’an, China, in 1999 and 2002, respectively.

In 1999, he joined the School of Telecommunications Engineering, Xidian University, where he is currently a Professor. He is also the Director of the Image Coding and Processing Center, State Key Laboratory of Integrated Service Networks, Xidian University.

His research interests focus on image and video processing, hyperspectral image processing, and high-performance computing.

Jiaojiao Li (Member, IEEE) received the B.E. degree in computer science and technology and the M.S. degree in software engineering and the Ph.D. degree in communication and information systems from Xidian University, Xi’an, China, in 2009, 2012, and 2016, respectively. She was an exchange Ph.D. student with Mississippi State University, Starkville, MS, USA, supervised by Dr. Qian Du. She is currently an Associate Professor and a Doctoral Supervisor with the School of Telecommunications Engineering, Xidian University. Her research interests include hyperspectral remote sensing image analysis and processing, and pattern recognition.

Chaoxiong Wu (Student Member, IEEE) received the B.E. degree in telecommunications engineering from Xidian University, Xi’an, China, in 2018, where he is currently pursuing the Ph.D. degree with the State Key Laboratory of Integrated Service Networks.

His research interests include spectral super-resolution, semantic segmentation of remote sensing images, and deep learning.

Y. Han, J. Yu, J.-H. Xue, and W. Sun, “Spectral super-resolution for RGB images using class-based BP neural networks,” in Proc. Digit. Image Comput., Techn. Appl. (DICTA), Dec. 2018, pp. 1–7.

Rui Song (Member, IEEE) received the B.S. degree in telecommunications and the M.S. and Ph.D. degrees in signal and information processing from Xidian University, Xi’an, China, in 2003, 2006, and 2009, respectively.

Since 2019, he has been a Professor with the School of Telecommunications Engineering, Xidian University. He is currently a member of the Image and Video Processing Laboratory, State Key Laboratory of Integrated Service Networks, Xidian University.

His research interests include video coding, computer vision, 3-D computer vision, and remote sensing image processing.

Dr. Song was a fellow of the Shanxi Society of Image and Graphics. He got the Prominent Teacher and Huashan Scholar Engineering Elite Talent of Xidian University in 2015 and the New-Star of Science and Technology supported by Shanxi Province, China.

Dr. Du is a fellow of the SPIE—International Society for Optics and Photonics. She was a recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society (GRSS). She was the Co-Chair of the Data Fusion Technical Committee of the IEEE GRSS from 2009 to 2013. She was the Chair of the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She was the General Chair of the fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing held at Shanghai, China, in 2012. She served as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS), THE JOURNAL OF APPLIED REMOTE SENSING, and the IEEE SIGNAL PROCESSING LETTERS. Since 2016, she has been the Editor-in-Chief of the IEEE JSTARS.