Endmember-Assisted Camera Response Function Learning, Toward Improving Hyperspectral Image Super-Resolution Performance

Jiangsan Zhao, Ying Qu, *Member, IEEE*, Seishi Ninomiya, and Wei Guo

***Abstract*— The camera response function (CRF) that projects hyperspectral radiance to the corresponding RGB images is important for most hyperspectral image super-resolution (HSI-SR) models. In contrast to most studies that focus on improving HSI-SR performance through new architectures, we aim to prevent the model performance drop by learning the CRF of any given HSIs and RGB image from the same scene in an unsupervised manner, independent of the HSI-SR network. Accordingly, we first decompose the given RGB image into endmembers and an abundance map using the Dirichlet autoencoder architecture. Thereafter, a linear CRF learning network is optimized to project the reference HSIs to the RGB image that can be similarly decomposed like the given RGB, assuming that objects in both images share the same endmembers and abundance map. The quality of the RGB images generated from the learned CRFs is compared with that of the corresponding ground-truth images based on the true CRFs of two consumer-level cameras Nikon 700D and Canon 500D. We demonstrate that the effectively learned CRFs can prevent significant performance drop in three popular HSI-SR models on RGB images from different categories of standard datasets of CAVE, ICVL, Chikusei, Cuprite, Salinas, and KSC. The successfully learned CRF using the method proposed in this study would largely promote a wider implementation of HSI-SR models since tremendous performance drop can be prevented practically.**

***Index Terms*— Abundance map, camera response function (CRF), endmember, hyperspectral image (HSI), super-resolution, unsupervised deep learning.**

I. Introduction

HYPERSPECTRAL images (HSIs) capture the radiation of chemicals and biological substances by capturing spectra in the visible to the short-wave infrared region and are therefore advantageous compared with regular RGB [1]. The additional information contained in HSIs enables its widescale employment in research and industry, including remote sensing [2], [3], ecology [4], health care [5], and the wood and food industries [6]–[8].

Despite the broad application potential, further expansion of HSI application is limited by the inadequacies of the devices used to capture HSIs. Two types of hyperspectral imaging devices are commercially available. Scanning-based imaging systems are limited by the complicated setups, low portability, and low temporal resolution [9]. Snapshot-based imaging systems capture the target entirely in a single integration, but multiple filters need to be mounted on the sensor, resulting in larger sensor size and consequently lower spatial resolution of the captured images [10], [11]. Although the loss of temporal resolution is compensated, HSIs that are high in both spatial and spectral resolutions, that is, high-resolution HSI (HR-HSI), are ideal for real-world applications [12]. However, the limitations of the scanning and snapshot-based hyperspectral imaging devices make HR-HSIs acquisition extremely difficult. To overcome the limitations of the hyperspectral imaging systems, that is, the low temporal or spatial resolution, some coding-based approaches in computational photography have been developed based on compressive sensing theory. Complicatedly designed hyperspectral cameras using coded apertures have been proposed as a solution [13]–[15]. Furthermore, a fusion-based snapshot camera generates HR-HSIs by fusing the low-resolution HSIs (LR-HSIs) and a high-resolution RGB image (HR-RGB) captured by the constituent hyperspectral and RGB camera inside [11], [15], [16]. However, the broader application scenarios of these hyperspectral imaging systems are limited owing to the highly complicated design and expensive production.

HSI super-resolution (HSI-SR) approaches have been established to bypass these restrictions of hardware and imaging conditions to acquire high-quality HR-HSIs at a low cost [17]–[22]. Deep learning approaches are increasingly applied in many areas of research and industry [23]–[25] and have been implemented to solve HSI-SR problems as well [26]. The consumer-level RGB imaging sensors can easily capture images with high spatial resolution at the cost of fewer spectral bands and provide the spatial details and color basis for HSI-SR tasks. Currently, there are two major approaches to realize HSI-SR models through deep learning. One is a direct reconstruction model that is developed based on low-resolution RGB images (LR-RGBs) and LR-HSIs [19], [27]. The trained

Manuscript received December 26, 2021; revised February 15, 2022, April 25, 2022, and May 17, 2022; accepted June 8, 2022. Date of publication June 13, 2022; date of current version June 22, 2022. This work was supported in part by the Japan Science and Technology Agency (JST) AIP Acceleration Research "Studies of CPS platform to raise big-data-driven AI agriculture" and in part by the Strategic International Collaborative Research Program (SICORP) under Grant JPMJSC16H2. *(Corresponding author: Jiangsan Zhao.)*

Jiangsan Zhao and Wei Guo are with the Laboratory of Field Phenomics, Graduate School of Agriculture and Life Sciences, The University of Tokyo, Tokyo 188-0002, Japan (e-mail: jiangsan-zhao@g.ecc.u-tokyo.ac.jp; guowei@g.ecc.u-tokyo.ac.jp).

Ying Qu is with the Advanced Imaging and Collaborative Information Processing Group, Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville, TN 37996 USA (e-mail: yqu3@vols.utk.edu).

Seishi Ninomiya is with the Laboratory of Field Phenomics, Graduate School of Agriculture and Life Sciences, The University of Tokyo, Tokyo 188-0002, Japan, and also with the Plant Phenomics Research Center, Nanjing Agricultural University, Nanjing 210095, China (e-mail: snino@g.ecc.u-tokyo.ac.jp).

Digital Object Identifier 10.1109/TGRS.2022.3182425

model can subsequently be used to reconstruct HR-HSIs based on the HR-RGB. The alternative is a fusion-based model that generates HR-HSIs by fusing HR-RGB with LR-HSIs from the same scene directly [28], [29]. Both have their own merits and limitations. Direct reconstruction models in HSI-SR can be easy to train because no registration is needed as the LR-RGB that is used for training is converted directly from LR-HSIs given that the camera response function (CRF) of the HR-RGB is known [30]. However, because the spatial information of HR-RGB images is not considered when building direct reconstruction models, the model performance can suffer severely from the low SR scale and increased spectral distortion [31]. In contrast, fusion-based HSI-SR models generally achieve higher-quality HR-HSIs and a larger super-resolution scale than direct reconstruction methods [18]. Nonetheless, strict registration between HR-RGB and LR-HSIs is usually required additionally to allow fusion-based HSI-SR models to achieve excellent performance.

HR-RGB plays a key role in developing both types of the aforementioned HSI-SR models. Based on the theory of HSI-SR, the recovery of HSIs from the RGB images is ill-posed because it is a regression from a lower-dimensional pixel vector, RGB, to a higher-dimensional hyperspectral counterpart, for example, [31], [32]. The intuitive HSI-SR theory strongly suggests that the output hyperspectral reflectance is highly dependent on the input pixel values, that is, either the magnitude or the vector direction, as confirmed in several studies [18], [19], [30].

The different pixel values or appearances of RGB images of the same object are mainly determined by the CRF that projects the spectral radiance to the RGB values mimicking the human perception of these real-world objects [33]. Given the spectral radiance $X(x, y, \lambda)$ at location (x, y), the intensity recorded by the commercial RGB camera for the pth color filter can be expressed as follows:

$$Y_p(x, y) = \int_{\lambda} C_p(\lambda) X(x, y, \lambda) \mathrm{d}\lambda \tag{1}$$

where $\lambda C_p(\lambda)$ is the corresponding CRF at channel p. The discrete form of (1) can be written as follows:

$$Y_p(x, y) = \sum_{b=1}^{B} C_p(\lambda_b) X(x, y, \lambda_b) \tag{2}$$

where λ_b $(b = 1, 2, \ldots, B)$ is the discrete representation of wavelength λ, and B is the number of spectral bands that is mostly 31.

Most of HSI-SR models are input RGB image appearance-dependent [19], [34]. Several studies have confirmed that the performance of HSI-SR models is highly sensitive to the CRFs used to prepare the input RGB images [17], [30]. The prediction performance of a direct reconstruction HSI-SR model benefited significantly from the selection of the optimal CRF that is close enough to the ground-truth CRF of the existing HR-RGB image [35]. Fusion-based HSI-SR models have also shown improved performance by selecting the optimal CRF from the existing CRF dictionaries [17], [30]. Therefore, the correct CRF of the input RGB image must be used along with HSIs for both the training and the subsequent predictions of HSI-SR models.

However, the main drawback of most of the current HSI-SR models, that is, of both the direction reconstruction-based models [18]–[20], [36] and fusion-based models [11], [37], [38] is that they assume that the CRF of HR-RGB used for training is known beforehand and it will stay the same when testing on new RGB images [39]. The outstanding fusion-based HSI-SR model, unregistered and unsupervised mutual Dirichlet-Net (μ^2MDN), generates HR-HSIs based on unregistered LR-HSIs and HR-RGB images in an unsupervised manner, however, still strongly depends on a solid CRF to manage the task [18]. The CRF selection strategy used in the outstanding direct reconstruction HSI-SR model proposed by Fu *et al.* [17] in 2021 relied entirely on the existence of adequately calibrated CRF dictionaries; the model would not function effectively when the RGB images were sourced from a new camera with unknown CRF. Once the CRF dictionary is not given or input RGB images are produced by cameras with unknown CRFs, current HSI-SR models will suffer from a severe decrease in HSI-SR performance and be much less effective in real-world applications. Therefore, a method that can be easily employed to learn a CRF from any given reference HSIs and RGB image is urgently required. Based on the learned CRF, the performance of the HSI-SR models could be largely maintained at least in future studies.

Therefore, in this study, we propose a novel network that can learn the CRF from any given reference HSIs and RGB image of the same scene from random initiation in an unsupervised manner. The major aim is to relieve the CRF dependence that is currently a big limitation for most HSI-SR models through learning the highly faithful CRF between the acquired low-cost LR-HSIs and HR-RGB easily. A specifically designed autoencoder is first trained to decompose the existing HR-RGB into endmembers and the corresponding HR-abundance map—the weights of different endmembers of each pixel. An LR-RGB is first generated from the existing LR-HSIs through a linear network and subsequently decomposed into endmembers and an LR-abundance map using the same autoencoder used in HR-RGB decomposition. Based on the assumption that the same object in both LR-RGB and HR-RGB is composed of the same endmembers and thus the same abundance map, only when the LR-RGB shares the same color space as HR-RGB, can the same endmembers and abundance map be decomposed by the same decomposition autoencoder. In this manner, the CRF that is embedded in the linear network can be learned to project LR-HSIs onto the target LR-RGB that shares the same color space as the HR-RGB. The fidelity of the learned CRFs from the two cameras is further tested on three categories of datasets using three sets of state-of-the-art HSI-SR models, residual convolutional neural network for the hyperspectral reconstruction (HSCNNR), μ^2MDN, and weighted low-rank tensor recovery (WLRTR).

The main contributions of this study are summarized as follows: First, we propose an RGB image decomposition network as an intermediate step for CRF learning. Second, we propose to calculate the loss of the Gram matrix of abundance maps

in different sizes during CRF learning. Third, integrated loss functions, combining $l_{2,1}$-norm loss and spectral information divergence (SID) loss, are used to direct the convergence of the CRF learning model. With the above stand-alone CRF learning network, excellent CRFs can be found to generate RGB images that are in the same color space as target RGB ones, which ultimately preserves the performances of HSI-SR models while without relying on any existing CRFs.

II. Related Work

A. Camera Response Functions

The appearances of produced RGB images differ in different camera brands because of the CRFs used [33], [40]. The CRF of a specific camera brand is determined by the preferences of the manufacturer and is mostly unavailable to consumers [30]. The CRF of a set of RGB images is unavailable if the source device capturing them is missing. Extensive research has been conducted to search for a CRF of a camera. The commonly used experimental approach to determine the CRF is to record the spectral response by exposing the target to monochromatic light [41]. However, this procedure is costly and time-consuming and must be repeated for RGB images from which the camera settings are changed. Although a variety of statistical methods can be used for CRF estimation, an object of known specific radiance spectrum is generally used in the scene to impose additional information [33], [40], [42]–[46]. Therefore, more effective methods that can be easily used to determine a CRF between the scene radiance and RGB values are highly preferred. In addition, most of the studies focus only on CRF searching, and few have used the learned CRFs to solve practical problems pertaining to HSI-SR.

B. Hyperspectral Image Super-Resolution

HSI-SR is an approach that can generate HR-HSIs without complex designed coded-aperture sensors or precisely arranged hybrid systems that overcome the existing hardware limitations and expands the application areas of HSIs by increasing accessibility [19]–[22]. HSI-SR models take advantage of the existing LR-HSIs and HR-RGBs, both of which are available from the existing devices at low cost [17], [18], for the production of HR-HSIs. Fusion-based HSI-SR methods can be grouped into two categories: one category is traditional statistic-based methods including pan-sharpening [47], [48], matrix factorization [49], [50], and tensor representation [51]–[54]; the other category is a deep convolutional neural network (CNN)-based one [17], [55], [56]. However, fusion-based approaches are generally limited by both registration issues, the point spread function (PSF), and the spectral relationship, the CRF, of the existing image pairs [18], [57]. The recently proposed μ^2MDN successfully solved the dependence on image registration and ended with much less performance loss when testing on unregistered image pairs [18]. Another group of HSI-SR methods is to reconstruct HSIs from a single RGB image using CNN-based networks [34], [58], [59]. Because the generated HR-HSIs rely on the spatial resolution of the HR-RGB totally and thus do not have image registration problems like fusion-based ones. However, the CRF has to be known in advance in order to prepare the LR-HSIs and LR-RGB pairs for model training. The direct reconstruction methods are also limited by the smaller SR scale which is generally around 8, while it can be as large as 32 for fusion-based approaches [31]. As the strong dependence on well-registered image pairs during HSI-SR is now largely solved, its broader application potential is still mainly and strongly restricted by the heavy reliance on the existing CRF [57].

We focus on the effective learning of the faithful CRF between any given HSIs and RGB pair of the same scene. The excellent performance of the learned CRF is confirmed not only by their abilities to render high-quality RGB from the corresponding HSIs, but also by their capability to prevent the performance drop of the adequately trained HSI-SR models, HSCNNR, μ^2MDN, and WLRTR, on the RGB images generated from three different benchmark HSIs datasets of indoor, natural scene, and remote sensing, using two CRFs of two consumer-level cameras, Canon 500D and Nikon 700D.

III. Problem Formulation

HR-RGB $X \in R^{HW\times l}$, where H, W, and l are the height, weight, and a number of spectral bands, respectively. LR-HSIs $Y \in R^{hw\times L}$, where h, w, and L represent the corresponding height, weight, and spectral bands, respectively. The aim is to learn a CRF that maps LR-HSIs to their low-spectral-counterpart LR-RGB $R^{hw\times l}$.

It has been shown that RGB images can be decomposed into an abundance map and the corresponding endmembers [60]. This decomposition is formulated as follows:

$$X^{HW\times l} = A^{HW\times\theta}E^{\theta\times l} \tag{3}$$

where $X^{HW\times l}$ is the 2-D matrix unfolded from the RGB image, HW is the total number of pixels, and $l = 3$ is the RGB color channels; $A^{HW\times\theta}$ represents the abundance map matrix with HW rows and θ columns, $E^{\theta\times l}$ is the endmember matrix with θ rows, and $l = 3$ is the RGB channel. θ represents the number of endmembers to be determined based on the complexity of the image that is generally greater than the number of RGB channels. Because both HR-RGB and LR-HSIs belong to the same scene, we assume that the same material of the same scene in both HR and the converted LR-RGB images should share the same endmembers, E, and thus the corresponding coefficients in abundance map A as well. $X^{hw\times l}$ is the unfolded 2-D matrix of the LR-RGB, where hw is the number of pixels and $l = 3$ is the RGB channel. Decomposition can be formulated as follows:

$$X^{hw\times l} = A^{hw\times\theta}E^{\theta\times l} \tag{4}$$

where hw in $A^{hw\times\theta}$ represents the number of pixels of the LR-RGB image, and θ in $A^{hw\times\theta}$ is the same number of endmember shared by the HR-RGB image.

IV. Proposed Approach

A. Network Architecture

The entire network is separated into subnetworks of endmember decomposition and CRF learning.

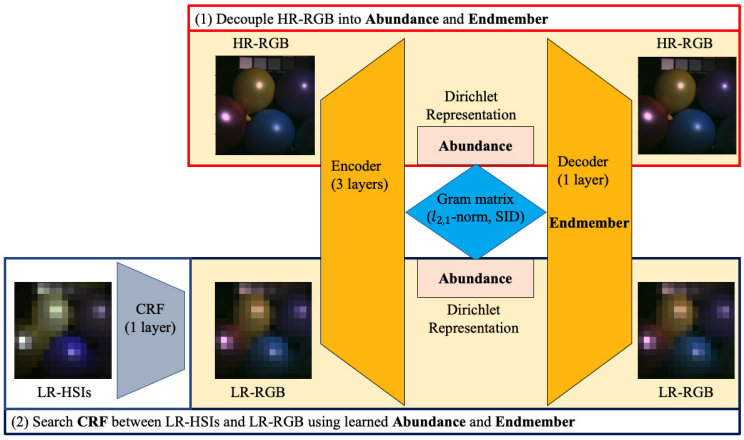

Fig. 1. Architecture of the endmember-assisted CRF learning. Both $l_{2,1}$-norm and SID losses are used to measure the differences of these Gram matrices of the abundance maps of HR-RGB and generated LR-RGB.

1) Endmember Decomposition: A specifically designed autoencoder was used to decompose the HR-RGB image into endmembers and the corresponding abundance map, highlighted in the red rectangle in Fig. 1 (top). We first define the input HR-RGB image as X, the feature representation $\tilde{A}$, and the endmembers as E. The encoder was composed of a fully connected three-layer linear network f_{en}: $X \rightarrow \tilde{A}$, which maps the input data to the latent feature representation. To improve the representation ability of the final output layer of the encoder f_{en}, the previous layers were densely connected with the final layer. In order to transform the feature representation $\tilde{A}$, to the abundance map A, meeting the physical constraints of nonnegative and sum-to-one, a stick-breaking algorithm that has been successfully used by Qu *et al.* [18], [61], [62] in several studies is applied. The derived abundance map, A, follows a Dirichlet distribution that satisfies the physical constraints of nonnegative and sum-to-one naturally.

The decoder, f_{de}: $A \rightarrow X$, is a one-layer-only nine-node linear network that attempts to reconstruct the input HR-RGB image from an extracted abundance map A. Thus, the derived 9 $\times$ 3 weight matrix, E, of the decoder can be treated as the endmembers of the input HR-RGB image X. There are nine endmembers in E and each endmember is composed of three elements. Endmembers should also follow the same value range as the reflectance used in this study during the decomposition process.

2) CRF Learning: Because the LR-HSIs and HR-RGB images belong to the same scene, the assumption is that objects of the same material appearing in both HR-RGB and LR-RGB that are projected from LR-HSIs by a target CRF share the same endmembers E, and the corresponding abundance map indicating the weights of endmembers composing each pixel of the RGB image. Moreover, the output abundance map of the encoder is dependent on the input RGB vectors, as per the characteristic of a linear network. Therefore, if both the HR-RGB and LR-RGB can be decomposed into the same endmembers using the same encoder f_{en} and decoder f_{de}, setup, then they will also share the same color space which is determined by the CRF. Based on the theory mentioned above, a one-layer three-node linear CRF learning network f_{crf} is developed to map the LR-HSIs to an LR-RGB that is to share the same endmembers E and abundance map A, with the HR-RGB. Therefore, the decomposed abundance maps, A_{HR} and A_{LR}, are forced to be as similar as possible during the optimization of f_{crf} for CRF learning.

3) Gram Matrix of Abundance Map: Almost all loss functions require the input matrices to be of the same dimension. However, the abundance maps, $A^{HW\times 9}$ and $A^{hw\times 9}$, decomposed from HR-RGB and LR-RGB differ in size. The challenge for CRF learning is the transformation of differently sized abundance maps to the same size, such that regular loss functions can be used to quantify their differences. The Gram matrices of these abundance maps are one of the few options that can represent abundance maps of different sizes excellently. The product of the transposed abundance map and the abundance map itself is treated as its Gram matrix that is a square matrix with the same number of rows and columns as the number of endmembers $\text{GM}^{9\times 9}$. For example, the abundance maps of an HR-RGB and LR-RGB can be formulated in 9 $\times$ 9 matrices as $\text{GM}_{\text{HR}}^{9\times 9}$ and $\text{GM}_{\text{LR}}^{9\times 9}$, respectively. Commonly used effective loss functions can be used as the Gram matrices of both HR and LR abundance maps, $\text{GM}_{\text{HR}}^{9\times 9}$ and $\text{GM}_{\text{LR}}^{9\times 9}$, are of the same size, 9 $\times$ 9. Another reason why the Gram matrix is used is that the coupled features of spatial and abundance information suggested by intuitive calculation are effective in representing the abundance map [63]. Regular loss functions can thereafter be applied to help the CRF learning through regulating the Gram matrices indirectly.

B. Optimization and Implementation Details

1) Loss Function: Loss functions play a key role in regulating model convergence in deep learning. To best facilitate the convergence of the CRF learning model, a composite loss function—a combination of collaborative $l_{2,1}$-norm and SID loss—was incorporated into the CRF learning network, instead of the conventional counterparts. The $l_{2,1}$-norm loss is the sequential application of the l_2-norm and l_1-norm loss on each pixel vector to better regulate the model convergence during spectral recovery [18]. SID [64] measures the vector shape dissimilarity that is complementary to $l_{2,1}$-norm that is the

magnitude difference. The combination of collaborative $l_{2,1}$-norm loss (5) and SID loss (6) between the Gram matrices is supposed to facilitate the model convergence, as both functions provide complementary measures of both magnitude and similarity of a vector. The final loss of a model is the sum of the individual loss that can be formulated as (7)

$$l_{2,1\text{loss}} = \times \times \sum_{i=1}^{m}\sum_{j=1}^{n}\left(\left(\text{GM}_{\text{HR}}^{9\times 9} - \text{GM}_{\text{LR}}^{9\times 9}\right)^2\right)^{1/2} \tag{5}$$

$$\text{SID}_{\text{loss}} = \left(\frac{1}{m}\right)\sum_{i=1}^{m}\left(\frac{\text{GM}_{\text{HR}}^{9\times 9}}{\left|\text{GM}_{\text{HR}}^{9\times 9}\right|} - \frac{\text{GM}_{LR}^{9\times 9}}{\left|\text{GM}_{LR}^{9\times 9}\right|}\right) \times \left(\log\left(\frac{\text{GM}_{\text{HR}}^{9\times 9}}{\left|\text{GM}_{\text{HR}}^{9\times 9}\right|}\right) - \log\left(\frac{\text{GM}_{LR}^{9\times 9}}{\left|\text{GM}_{\text{LR}}^{9\times 9}\right|}\right)\right). \tag{6}$$

$\text{GM}_{\text{HR}}^{9\times 9}$ and $\text{GM}_{\text{LR}}^{9\times 9}$ are the Gram matrices of the abundance maps corresponding to HR-RGB and LR-HSIs, respectively; m and n represent the number of rows and columns of the matrices, respectively,

$$\text{Loss}_{\text{total}} = \lambda_1 l_{2,1\text{loss}} + \lambda_2 \text{SID}_{\text{loss}}. \tag{7}$$

λ_1 and λ_2 represent the different weights given to the subcomponents of the total loss.

2) Training: Before feeding into the network, the spectral vectors in the LR-HSIs and HR-RGB images were transformed to zero-mean vectors. The number of output nodes of the encoder was determined by the scene complexity of the HR-RGB image, and nine were selected after preliminary screening. The weights of both the decoder and CRF learning networks were randomly initiated as default in the Pytorch framework [65]. The detailed structure of the RGB image decomposition network is illustrated in Fig. 1. Both the HR-RGB image decomposition and CRF learning processes stop until the losses no longer decrease any further for over 1000 epochs. The weights of the CRF network are extracted as the CRF that is used to map the LR-HSIs to the LR-RGB image in preparation for the HSI-SR models.

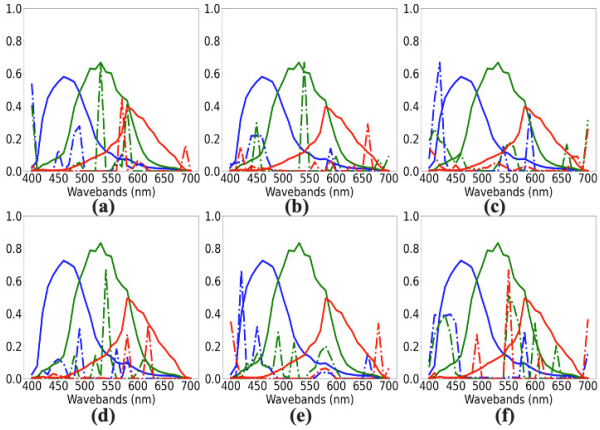

Fig. 2. Comparisons of the learned CRFs and true CRFs of Canon 500D camera on three datasets CAVE, ICVL, and Chikusei. (a)–(c) Learned and true Canon CRFs are indicated by dashed–dotted and solid lines, respectively. (d)–(f) Learned and true Nikon CRFs are indicated by dashed–dotted and solid lines, respectively.

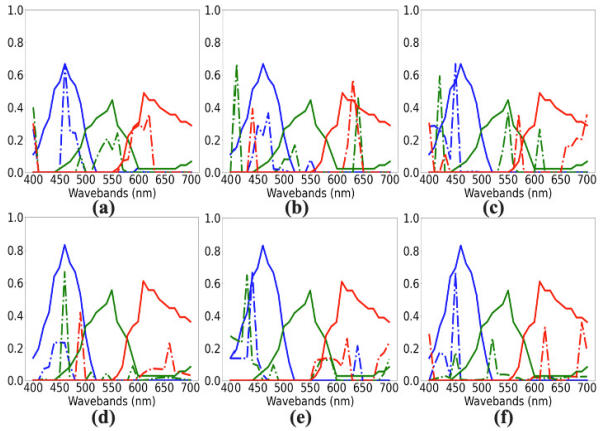

Fig. 3. Comparisons of the learned CRFs and true CRFs of Nikon 700D camera on three datasets Cuprite, Salinas, and KSC. (a)–(c) Learned and true Canon CRFs are indicated by dashed–dotted and solid lines, respectively. (d)–(f) Learned and true Nikon CRFs are indicated by dashed–dotted and solid lines, respectively.

V. Experiments and Results

In this section, we first review the HR-HSIs datasets and the setup of our experiments. Thereafter, we provide the results of HR-RGB image decomposition, learned CRFs, and the quality of the RGB generated with learned CRFs to demonstrate the effectiveness of the proposed method in CRF learning. With the RGB images rendered by the learned CRF (learned RGB images for short), HSI-SR models from two categories, HSCNNR, μ^2MDN, and WLRTR, were trained and externally tested on RGB images with CRFs of two consumer-level cameras, Canon 500D [40] and Nikon 700D [18]. Extensive experiments were conducted on benchmark datasets of indoor and natural scenes and remote sensing.

A. Dataset and Experimental Setting

Six widely used benchmark HSI datasets are investigated in this section, including the CAVE [66], ICVL [19], Chikusei [21], Cuprite [67], Salinas [68], and KSC datasets [69]. The CAVE dataset includes images of 32 indoor scenes. Each image is 512 × 512 pixels, with 31 spectral bands ranging from 400 to 700 nm. The ICVL dataset contains natural scenes captured by the Specim NIR sensor in Israel in 2017. The original HSIs are 1000 × 1000 pixels with 31 spectral bands. The Chikusei dataset was acquired by a hyperspectral imaging sensor over Chikusei, Japan, in 2014. The original HSIs consist of 2517 × 2335 pixels and 128 bands with a spectral range of 363–1018 nm. In our study, we cropped a 1000 × 1000 patch from the center. "f970619t01p02_r02_sc03.a.rfl" from online Cuprite data was included in this study. Cuprite was captured by the airborne visible/infrared imaging spectrometer (AVIRIS) sensor over the Cuprite mine district in Nevada, USA; it contains 224 spectral bands in the range of 400–2500 nm. Salinas was collected in the 224 band using an AVIRIS sensor over Salinas Valley, California. The KSC dataset was collected by NASA AVIRIS, with spectral bands from 0.4 to 2.5 μm, a spatial resolution of 18 m, spatial size of (512, 614), and spectral bands of 176 totally. Only 31 spectral

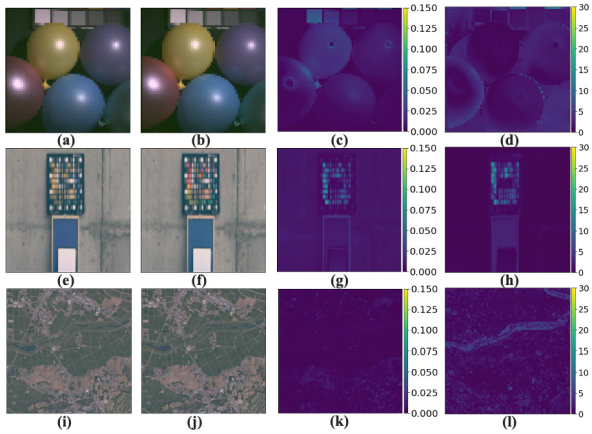

Fig. 4. SAM and average absolute difference of the learned Nikon RGB images of CAVE, ICVL, and Chikusei datasets. (a), (e), and (i) Learned Nikon RGB images. (b), (f), and (j) True Nikon RGB images. (c), (g), and (k) Average absolute differences. (d), (h), and (l) SAMs.

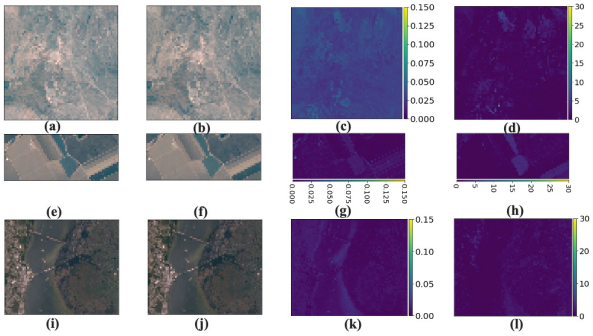

Fig. 5. SAM and average absolute difference of the learned Nikon RGB images of Cuprite, Salinas, and KSC datasets. (a), (e), and (i) Learned Nikon RGB images. (b), (f), and (j) True Nikon RGB images. (c), (g), and (k) Average absolute differences. (d), (h), and (l) SAMs.

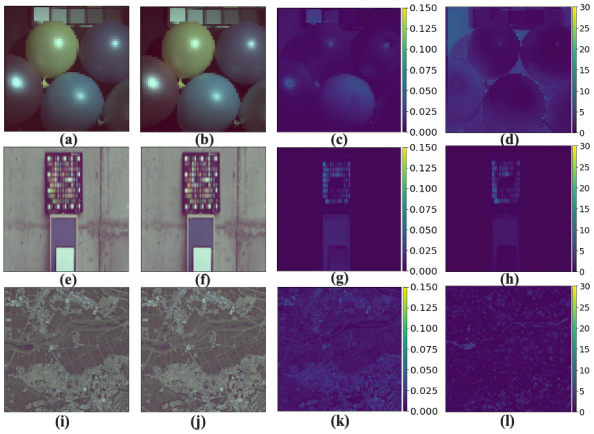

Fig. 6. SAM and average absolute difference of the learned Canon RGB images and CAVE, ICVL, and Chikusei datasets. (a), (e), and (i) Learned Canon RGB images. (b), (f), and (j) True Canon RGB images. (c), (g), and (k) Average absolute differences. (d), (h), and (l) SAMs.

bands considered within the wavelength range of 400–700 nm from the Chikusei, Cuprite, Salinas, and KSC datasets were used in this study.

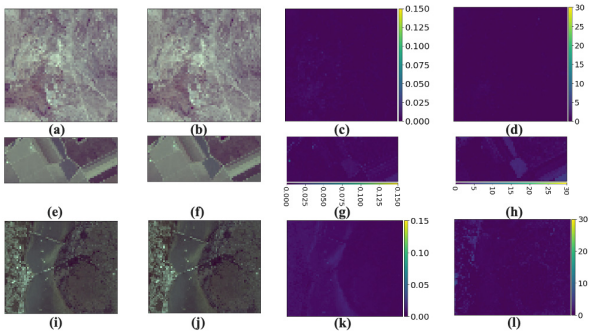

Fig. 7. SAM and average absolute difference of the learned Canon RGB images of Cuprite, Salinas, and KSC datasets. (a), (e), and (i) Learned Canon RGB images. (b), (f), and (j) True Canon RGB images. (c), (g), and (k) Average absolute differences. (d), (h), and (l) SAMs.

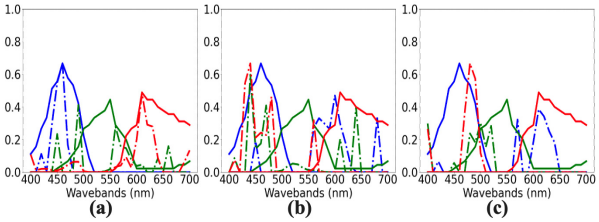

Fig. 8. Comparisons between the learned CRFs with different loss functions and true CRFs of Nikon 700D cameras. (a) Learned CRF using combined loss functions of $l_{2,1}$-norm and SID versus true Canon CRF are indicated by dashed–dotted and solid lines, respectively. (b) Learned CRF using $l_{2,1}$-norm loss only versus true Nikon CRF are indicated by dashed–dotted and solid lines, respectively. (c) Learned CRF using SID loss only versus true Nikon CRF are indicated by dashed–dotted and solid lines, respectively.

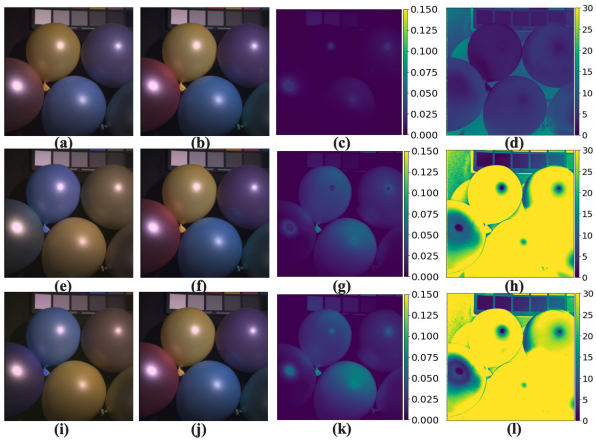

Fig. 9. SAM and average absolute differences of the learned Nikon RGB images using different loss functions. (a), (e), and (i) Nikon RGB images converted with combined, $l_{2,1}$-norm and SID loss, respectively. (b), (f), and (j) True Nikon RGB images. (c), (g), and (k) Average absolute differences. (d), (h), and (l) SAMs.

Two commonly used CRFs of both Nikon 700D (Nikon) [18] and Canon 500D (Canon) [33] cameras were employed in the generation of ground-truth HR-RGB images. We adopted a Gaussian filter to obtain the LR-HSIs from the HR-HSIs by constructing a filter with the same width as the super-resolution scale and 0.5 valued deviations. A super-resolution scale of 8 was set for all six datasets.

TABLE I

EVALUATIONS ON LEARNED RGB IMAGES OF BOTH CANON500D AND NIKON 700D CAMERAS ON CAVE, ICVL, CHIKUSEI, CUPRITE, SALINAS, AND KSC DATASETS

Evaluation method	CAVE		ICVL		Chikusei		Cuprite		Salinas		KSC	
	Canon	Nikon	Canon	Nikon	Canon	Nikon	Canon	Nikon	Canon	Nikon	Canon	Nikon
SAM ↓	2.50	2.05	0.50	0.81	1.15	1.37	0.27	0.53	1.01	0.98	0.78	0.84
PSNR ↑	43.03	41.84	55.0	39.85	39.03	50.91	53.30	36.93	48.12	44.40	45.74	43.12
ERGAS ↓	8.62	10.39	2.65	4.64	16.22	6.02	0.96	4.12	2.73	3.09	5.41	6.83

*Arrows indicate the direction towards the best performance

TABLE II

PERFORMANCES OF THE LEARNED NIKON MODELS OF μ^2MDN ON TRUE NIKON AND CANON HR-RGB IMAGES OF CAVE, ICVL, CHIKUSEI, CUPRITE, SALINAS, AND KSC DATASETS

Dataset	Training			Testing					
	Learned Nikon RGB			True Nikon RGB			True Canon RGB		
	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓
CAVE	**4.79**	**42.33**	**0.44**	7.53	35.38	0.47	28.17	20.49	2.14
ICVL	9.53	41.06	1.58	**8.55**	**43.19**	**1.32**	19.61	26.57	2.38
Chikusei	**6.68**	**50.94**	**1.64**	8.59	47.23	2.67	16.10	34.87	6.43
Cuprite	16.22	**46.08**	**0.12**	9.94	40.00	0.15	39.42	26.63	0.61
Salinas	7.92	**45.11**	1.08	**7.88**	44.87	**1.04**	26.26	28.92	4.11
KSC	**10.95**	**51.10**	8.26	12.11	49.40	**8.16**	31.76	39.33	18.80

*Arrows indicate the direction towards the best performance

TABLE III

PERFORMANCES OF LEARNED NIKON MODELS OF HSCNNR ON TRUE NIKON AND CANON HR-RGB IMAGES OF CAVE, ICVL, CHIKUSEI, CUPRITE, SALINAS, AND KSC DATASETS

Dataset	Training			Testing					
	Learned Nikon RGB			True Nikon RGB			True Canon RGB		
	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓
CAVE	**5.18**	**39.66**	1.82	5.53	38.58	**1.55**	21.55	26.72	4.92
ICVL	9.11	40.14	**1.44**	**8.70**	**40.51**	1.68	17.55	32.49	2.55
Chikusei	**7.57**	**49.11**	**5.03**	7.90	48.08	11.62	33.75	34.34	6.17
Cuprite	13.60	37.90	0.74	**12.9**	**38.16**	**0.67**	24.27	30.63	1.06
Salinas	8.28	**44.46**	**1.49**	**7.57**	44.43	3.05	42.24	28.03	5.95
KSC	**11.03**	**50.17**	2.16	12.55	48.43	**2.06**	32.55	38.20	10.74

*Arrows indicate the direction towards the best performance

B. HSI-SR Models

To test the fidelity of the learned CRF for HSI-SR tasks, three state-of-the-art architectures, HSCNNR [20], μ^2MDN [18], and WLRTR [52] were tested in this study. The models were trained using the suggested hyperparameters and tuned to maximize their performance. With two reference CRFs from Nikon 700D and Canon 500D, three categories of HSI-SR models, six models in total were built for each dataset. Further tests were done on HR-RGB images that were rotated by 90° in an anticlockwise direction to check the abilities of μ^2MDN and WLRTR in handling unregistered image pairs during HSI-SR.

C. Evaluation Metrics

Three complementary and widely used evaluation metrics, including a peak signal-to-noise ratio (PSNR), a spectral angle mapper (SAM) [63], and erreur relative globale adimensionnellede synthese (ERGAS) [71], are used to quantitatively measure the quality of reconstructed HR-HSIs. SAM measures the spectral angle difference between the predicted and ground-truth reflectance of each pixel. PSNR and ERGAS are mean square error (mse)-based bandwise indices indicating spatial fidelity and global quality, respectively. The best model performance results from the prediction with the maximized PSNR, while minimizing SAM and ERGAS.

TABLE IV

PERFORMANCES OF LEARNED CANON AND NIKON CRFS OF WLRTR ON TRUE CANON AND NIKON HR-RGB IMAGES OF CAVE, ICVL, CHIKUSEI, CUPRITE, SALINAS, AND KSC DATASETS

Dataset	Learned Nikon CRF						Learned Canon CRF					
	True Nikon RGB			True Canon RGB			True Nikon RGB			True Canon RGB		
	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓
CAVE	**1.62**	**44.40**	**2.23**	3.85	26.79	17.97	4.85	33.17	10.40	**2.19**	**39.45**	**2.94**
ICVL	**5.54**	**45.01**	**4.85**	9.62	29.81	6.33	9.46	33.34	3.98	**6.68**	**43.05**	**3.89**
Chikusei	**8.29**	**42.13**	**7.05**	16.70	32.67	14.31	16.04	35.89	**9.20**	**11.40**	**42.95**	9.77
Cuprite	**10.62**	**39.05**	**0.60**	24.58	29.43	1.87	21.18	32.45	**0.86**	**17.59**	**34.63**	1.38
Salinas	**5.48**	**43.75**	**2.14**	12.48	29.23	2.89	11.71	31.53	2.29	**7.56**	**40.81**	**1.19**
KSC	**10.93**	**49.08**	**3.71**	20.53	37.92	14.32	18.29	40.69	5.21	**15.43**	**46.60**	**3.77**

*Arrows indicate the direction towards the best performance

TABLE V

PERFORMANCES OF THE LEARNED NIKON MODELS OF μ^2MDN ON ROTATED TRUE NIKON AND CANON HR-RGB IMAGES OF CAVE AND SALINAS DATASETS

Dataset	Training			Testing					
	Learned Nikon RGB			True Nikon RGB			True Canon RGB		
	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓
CAVE	**4.30**	**42.20**	**0.38**	4.51	41.35	0.40	21.94	26.65	0.94
Salinas	**7.02**	**46.11**	0.80	**7.02**	46.45	**0.61**	37.04	28.34	2.04

*Arrows indicate the direction towards the best performance

TABLE VI

PERFORMANCES OF LEARNED CANON AND NIKON MODELS OF WLRTR ON ROTATED TRUE CANON AND NIKON HR-RGB IMAGES OF CAVE AND SALINAS DATASETS

Dataset	Learned Nikon CRF						Learned Canon CRF					
	True Nikon RGB			True Canon RGB			True Canon RGB			True Nikon RGB		
	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓
CAVE	**15.97**	**27.81**	**12.84**	16.37	23.84	22.34	15.03	**28.18**	**14.66**	**15.02**	25.53	14.98
Salinas	**42.74**	**28.85**	**2.97**	48.74	26.12	3.00	**37.80**	**29.53**	2.87	41.32	27.61	**2.61**

*Arrows indicate the direction towards the best performance

1) Evaluations on Learned CRFs: Because pixel values of RGB images are calculated from the integrated radiance through CRF, and pixel values are important for HSI-SR models, different CRFs may result in the same RGB image. Therefore, the final quality of the CRFs is evaluated based on the differences between the regenerated RGB images and the corresponding ground-truth images in this study.

2) Evaluations on HR-HSIs Recovery: We focus on testing the performance of the HSI-SR models trained on the learned RGB images of Nikon and Canon cameras (learned Nikon and Canon RGB images), within which the trained models are called learned Nikon or Canon μ^2MDN, learned Nikon or Canon HSCNNR, and learned Nikon or Canon WLRTR, respectively. The performances of the learned μ^2MDN, HSCNNR, and WLRTR are based on the quality of the reconstructed HR-HSIs in comparison to ground-truth images through testing on RGB images converted by the true CRFs of both Nikon and Canon cameras (true Nikon and Canon RGB images).

3) Results on Learned CRFs: The learned CRFs are shown in Figs. 2 and 3, all of which are evidently different from the true CRFs of both Nikon and Canon cameras, regardless of the dataset used. Interestingly, the peaks of the corresponding channels in these learned and true CRFs are located extremely close to each other in most cases (see Figs. 2 and 3).

4) Quality of RGB Images With Learned CRFs: The SAM and average absolute differences between the learned RGB images (leftmost column) and true RGB images (second

TABLE VII

PERFORMANCES OF LEARNED CANON MODELS OF μ^2MDN ON TRUE CANON AND NIKON HR-RGB IMAGES OF CAVE, ICVL, CHIKUSEI, CUPRITE, SALINAS, AND KSC DATASETS

Dataset	Training			Testing					
	Learned Canon RGB			True Canon RGB			True Nikon RGB		
	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓
CAVE	**4.79**	**42.33**	**0.44**	7.53	35.38	0.47	28.17	20.49	2.14
ICVL	9.53	41.06	1.58	**8.55**	**43.19**	**1.32**	19.61	26.57	2.38
Chikusei	**6.68**	**50.94**	**1.64**	8.59	47.23	2.67	16.10	34.87	6.43
Cuprite	**11.21**	39.02	0.19	11.26	**39.58**	**0.16**	26.63	21.71	0.97
Salinas	**6.57**	**46.35**	**0.80**	8.62	41.88	1.45	21.22	26.78	7.23
KSC	**11.37**	**50.61**	3.69	12.05	48.98	**3.68**	35.29	31.30	29.73

*Arrows indicate the direction towards the best performance

TABLE VIII

PERFORMANCES OF LEARNED CANON MODELS OF HSCNNR ON TRUE CANON AND NIKON HR-RGB IMAGES OF CAVE, ICVL, CHIKUSEI, CUPRITE, SALINAS, AND KSC DATASETS

Dataset	Training			Testing					
	Learned Canon RGB			True Canon RGB			True Nikon RGB		
	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓
CAVE	7.65	**37.18**	6.08	**7.45**	36.67	**4.27**	22.12	21.42	17.03
ICVL	**8.34**	40.67	1.03	8.92	**41.15**	**0.96**	28.55	22.89	3.32
Chikusei	**13.97**	**44.28**	**2.41**	14.36	43.46	6.94	20.05	33.25	13.88
Cuprite	**16.28**	**35.28**	1.23	**16.28**	35.36	**1.34**	33.05	22.63	2.95
Salinas	**7.54**	**45.33**	**0.39**	9.19	41.02	0.72	25.98	24.67	4.09
KSC	16.30	**47.67**	11.13	**15.98**	47.14	**10.16**	27.34	34.75	16.94

*Arrows indicate the direction towards the best performance

TABLE IX

PERFORMANCES OF LEARNED CANON MODELS OF μ^2MDN ON ROTATED TRUE CANON AND NIKON HR-RGB IMAGES OF CAVE, ICVL, CHIKUSEI, CUPRITE, SALINAS, AND KSC DATASETS

Dataset	Training			Testing					
	Learned Canon RGB			True Canon RGB			True Nikon RGB		
	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓	SAM ↓	PSNR ↑	ERGAS ↓
CAVE	**4.78**	**42.39**	**0.24**	5.51	39.83	0.47	30.33	20.08	1.87
Salinas	**6.99**	**45.51**	**1.57**	8.72	42.14	1.90	19.70	27.90	7.11

*Arrows indicate the direction towards the best performance

leftmost column) of the Nikon and Canon cameras are shown in Figs. 4–7, respectively. Generally, the differences between learned and ground-truth RGB images are hardly noticeable to the human eye. More quality parameters between learned and true RGB images are summarized in Table I. Slightly more obvious errors can be found on the learned Nikon RGB

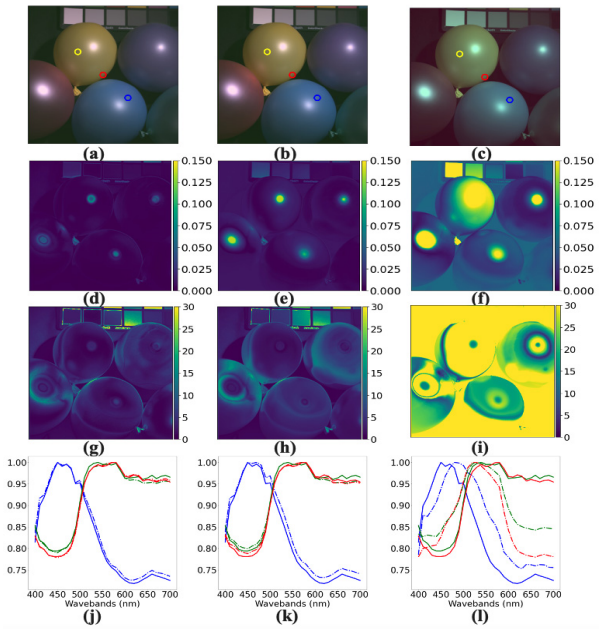

Fig. 10. Performances of learned Nikon μ^2MDN model in HR-HSI generation based on the true Nikon and Canon RGB images of CAVE. (a)–(c) Learned and true Nikon and true Canon RGB images. (d)–(f) Average absolute difference. (g)–(i) SAMs. (j)–(l) Visualization of reflectance differences of three selected pixels circled in the RGB images in (a)–(c).

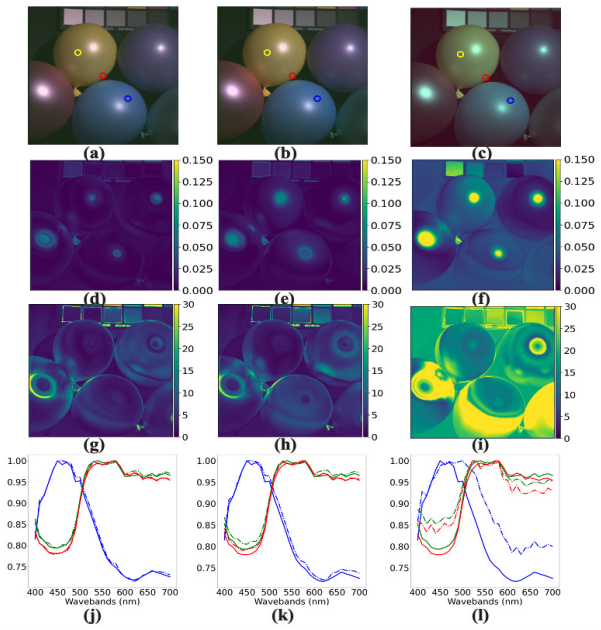

Fig. 12. Performances of the learned Nikon HSCNNR model in HR-HSI generation based on the true Nikon and Canon RGB images of CAVE. (a)–(c) Learned and true Nikon and true Canon RGB images. (d)–(f) Average absolute difference. (g)–(i) SAMs. (j)–(l) Visualization of reflectance differences of three selected pixels circled in the RGB images in (a)–(c).

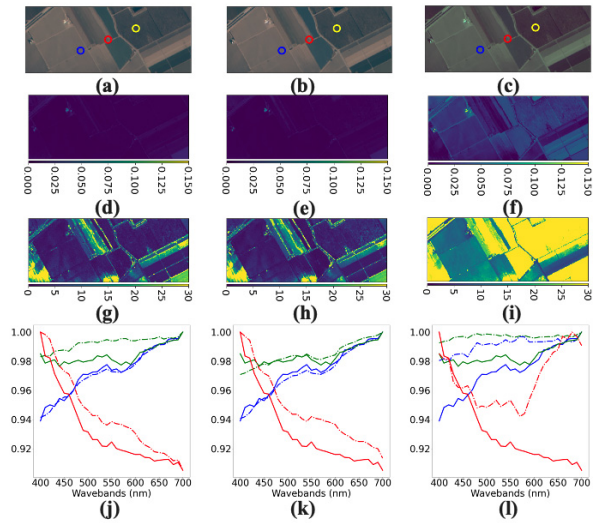

Fig. 11. Performances of the learned Nikon μ^2MDN model in HR-HSI generation based on the true Nikon and Canon RGB images of Salinas. (a)–(c) Learned and true Nikon and true Canon RGB images. (d)–(f) Average absolute difference. (g)–(i) SAMs. (j)–(l) Visualization of reflectance differences of three selected pixels circled in the RGB images in (a)–(c).

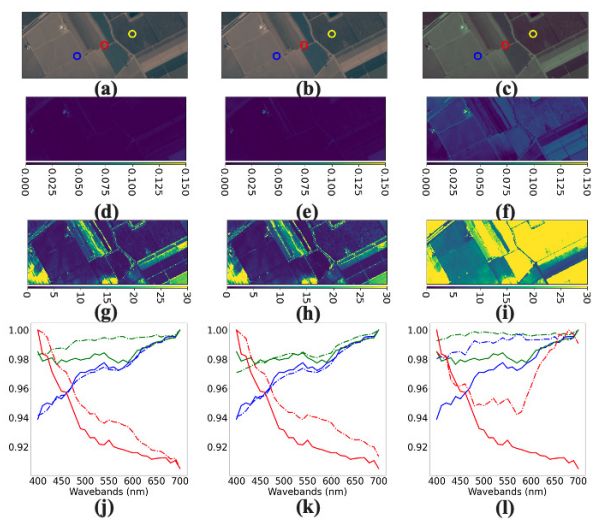

Fig. 13. Performances of the learned Nikon HSCNNR model in HR-HSI generation based on the true Nikon and Canon RGB images of Salinas. (a)–(c) Learned and true Nikon and true Canon RGB images. (d)–(f) Average absolute difference. (g)–(i) SAMs. (j)–(l) Visualization of reflectance differences of three selected pixels circled in the RGB images in (a)–(c).

images of ICVL that can also be seen more clearly in the error maps of SAM and average absolute differences shown in subplots *g* and *h* in Fig. 4. The variances in the quality of the learned RGB images probably resulted from the interaction of the selected CRFs and the benchmark datasets used because they were trained similarly with the same model architectures.

5) Ablation Study on Loss Functions: The CRF learning using individual loss functions of either $l_{2,1}$-norm loss or SID loss are compared with CRF learned using combined losses of $l_{2,1}$-norm loss and SID loss (see Figs. 8 and 9). Individual loss cannot help, while only when they are combined can high-quality CRFs be learned.

6) Performance of Models Developed Based on Learned Nikon RGB Images: The performances of μ^2MDN, HSCNNR, and WLRTR models that were trained on the registered pairs of learned Nikon RGB images of different datasets, as shown in Tables II–IV, respectively, with the best results highlighted in bold. All three models involving learned Nikon

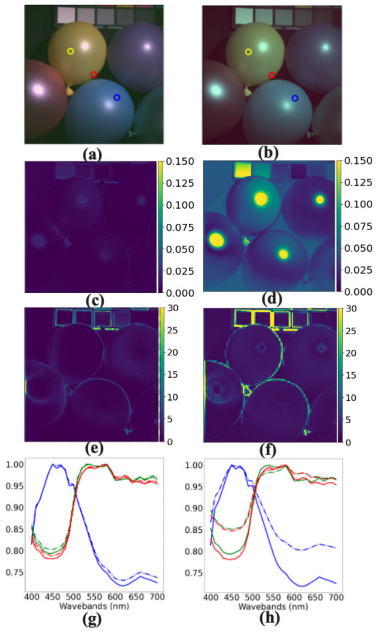

Fig. 14. Performances of the learned Nikon WLRTR model in HR-HSI generation based on the true Nikon and Canon RGB images of CAVE. (a) and (b) True Nikon and true Canon RGB images. (c) and (d) Average absolute difference. (e) and (f) SAMs. (g) and (h) Visualization of reflectance differences of three selected pixels circled in RGB images in (a) and (b).

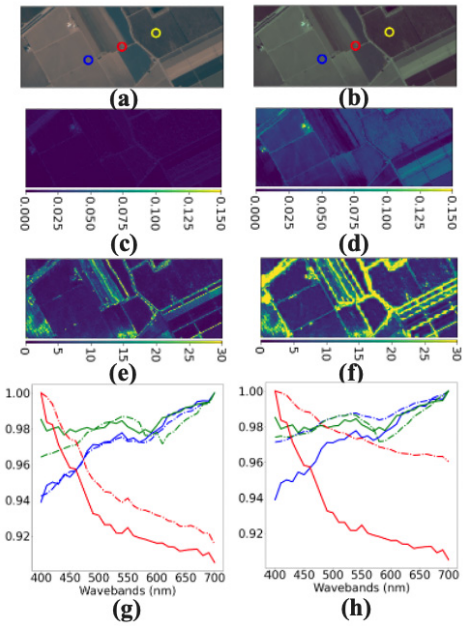

Fig. 15. Performances of the learned Nikon WLRTR model in HR-HSI generation based on the true Nikon and Canon RGB images of Salinas. (a) and (b) True Nikon and true Canon RGB images. (c) and (d) Average absolute difference. (e) and (f) SAMs. (g) and (h) Visualization of reflectance differences of three selected pixels circled in RGB images in (a) and (b).

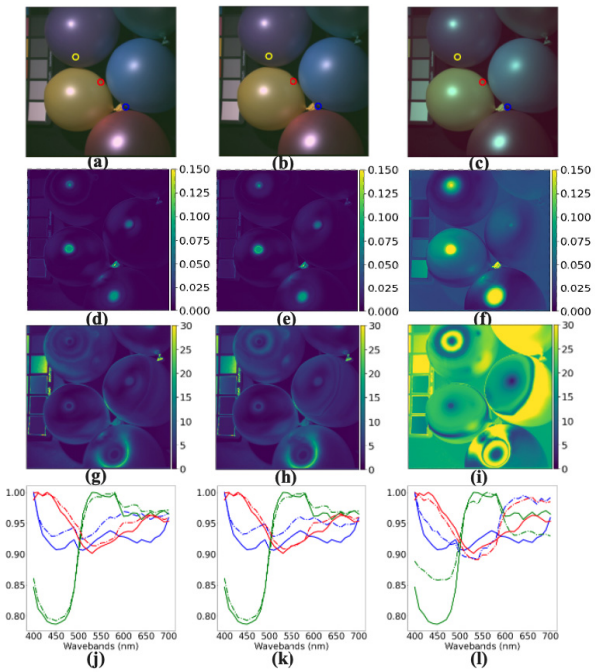

Fig. 16. Performances of the learned Nikon μ^2MDN model in HR-HSI generation based on the rotated true Nikon and Canon RGB images of CAVE. (a)–(c) Learned and true Nikon and true Canon RGB images. (d)–(f) Average absolute difference. (g)–(i) SAMs. (j)–(l) Visualization of reflectance differences of three selected pixels circled in RGB images in (a)–(c).

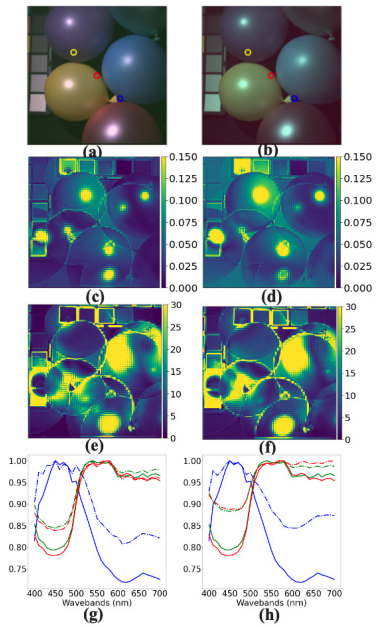

Fig. 17. Performances of the learned Nikon WLRTR model in HR-HSI generation based on the rotated true Nikon and Canon RGB images of CAVE. (a) and (b) Learned and true Nikon and true Canon RGB images. (c) and (d) Average absolute difference. (e) and (f) SAMs. (g) and (h) Visualization of reflectance differences of three selected pixels circled in RGB images in (a) and (b).

CRF as a component, μ^2MDN, HSCNNR, and WLRTR, suffer from slight to no performance drop when testing on true Nikon RGB images, while a significant performance drop is observed when testing on true Canon RGB images. The visualized error maps of two example datasets, CAVE, Salinas for μ^2MDN, HSCNNR, and WLRTR models are shown in Figs. 10–15, respectively. The performances of μ^2MDN and WLRTR models with rotated HR-RGB images as input are shown in Tables V and VI, where models of μ^2MDN are invariant to the image rotation, while WLRTR models suffer

from a big performance drop. Examples of error maps of μ^2MDN and WLRTR models on rotated image pairs of the CAVE dataset are shown in Figs. 16 and 17, respectively.

7) Performance of Models Built Based on Learned Canon RGB Images: Both types of HSI-SR models, learned Canon μ^2MDN (see Table VII), learned Canon HSCNNR (see Table VIII), and learned from WLRTR (see Table IV), performed equally on the true Canon RGB images, while exhibiting a tremendous performance decrease on true Nikon RGB images of all six datasets CAVE, ICVL, Chikusei, Cuprite, Salinas, and KSC. The rotation of input HR-RGB images has a marginal effect on the model performance of μ^2MDN (see Table IX) while is detrimental to WLRTR models.

VI. Conclusion

In this study, we demonstrate the stand-alone network to learn the CRF between any given HSIs and RGB image of the same scene in an unsupervised manner, which significantly preserved the performance of HSI-SR models. Additionally, the simple yet effective CRF learning process can largely replace the conventional extensive searching or supervised learning methods based on the selected corresponding pixels in HSIs and RGB images. Similar strategies can be implemented to learn mapping functions that map images covering different wavebands of the same scene, for example, multispectral image (MSI) to RGB or HSIs to MSI when we would like to apply SR models under these situations; RGB images of different appearances can also be mapped from one to another similarly. Through learning a CRF that is specific to the target set of RGB images with the proposed network, the heavy dependence of all HSI-SR models, including HSCNNR, μ^2MDN, and WLRTR models in this study, on RGB images with known CRFs can be largely relieved. Higher-fidelity RGB images might be generated through embedding CRF in nonlinear networks and thus the performances of HSI-SR models be better preserved with future efforts.

References

[1] X. Dong, M. Jakobi, S. Wang, M. H. Köhler, X. Zhang, and A. W. Koch, "A review of hyperspectral imaging for nanoscale materials research," *Appl. Spectrosc. Rev.*, vol. 54, no. 4, pp. 285–305, Apr. 2019.

[2] A. K. Mahlein *et al.*, "Plant disease detection by hyperspectral imaging: From the lab to the field," *Adv. Animal Biosci.*, vol. 8, no. 2, pp. 238–243, 2017.

[3] H. Qi, T. Paz-Kagan, A. Karnieli, X. Jin, and S. Li, "Evaluating calibration methods for predicting soil available nutrients using hyperspectral VNIR data," *Soil Tillage Res.*, vol. 175, pp. 267–275, Jan. 2018.

[4] C. K. Vance, D. R. Tolleson, K. Kinoshita, J. Rodriguez, and W. J. Foley, "Near infrared spectroscopy in wildlife and biodiversity," *J. Near Infr. Spectrosc.*, vol. 24, no. 1, pp. 1–25, Feb. 2016.

[5] I. O. Afara, I. Prasadam, Z. Arabshahi, Y. Xiao, and A. Oloyede, "Monitoring osteoarthritis progression using near infrared (NIR) spectroscopy," *Sci. Rep.*, vol. 7, no. 1, p. 11463, Dec. 2017.

[6] S. Tsuchikawa and H. Kobori, "A review of recent application of near infrared spectroscopy to wood science and technology," *J. Wood Sci.*, vol. 61, no. 3, pp. 213–220, Jun. 2015.

[7] D. F. Barbin, G. ElMasry, D.-W. Sun, P. Allen, and N. Morsy, "Non-destructive assessment of microbial contamination in porcine meat using NIR hyperspectral imaging," *Innov. Food Sci. Emerg. Technol.*, vol. 17, pp. 180–191, Jan. 2013.

[8] P. Menesatti *et al.*, "Laboratory vs. in-field spectral proximal sensing for early detection of fusarium head blight infection in durum wheat," *Biosystems Eng.*, vol. 114, no. 3, pp. 289–293, Mar. 2013.

[9] A. Signoroni, M. Savardi, A. Baronio, and S. Benini, "Deep learning meets hyperspectral image analysis: A multidisciplinary review," *J. Imag.*, vol. 5, no. 5, p. 52, 2019.

[10] X. Cao, H. Du, X. Tong, Q. Dai, and S. Lin, "A prism-mask system for multispectral video acquisition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2423–2435, Dec. 2011.

[11] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3586–3594.

[12] S. Thomas *et al.*, "Benefits of hyperspectral imaging for plant disease detection and plant protection: A technical perspective," *J. Plant Diseases Protection*, vol. 125, no. 1, pp. 5–20, Feb. 2018.

[13] L. Wang, Z. Xiong, D. Gao, G. Shi, and F. Wu, "Dual-camera design for coded aperture snapshot spectral imaging," *Appl. Opt.*, vol. 54, no. 4, pp. 848–858, 2015.

[14] X. Lin, Y. Liu, J. Wu, and Q. Dai, "Spatial–spectral encoded compressive hyperspectral imaging," *ACM Trans. Graph.*, vol. 33, no. 6, pp. 1–11, Nov. 2014.

[15] R. Kawakami, Y. Matsushita, J. Wright, M. Ben-Ezra, Y.-W. Tai, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *Proc. CVPR*, Jun. 2011, pp. 2329–2336.

[16] W. Dong *et al.*, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, May 2016.

[17] Y. Fu, T. Zhang, Y. Zheng, D. Zhang, and H. Huang, "Joint camera spectral response selection and hyperspectral image recovery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 256–272, Jan. 2020.

[18] Y. Qu, H. Qi, C. Kwan, N. Yokoya, and J. Chanussot, "Unsupervised and unregistered hyperspectral image super-resolution with mutual Dirichlet-net," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–18, 2021.

[19] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 19–34.

[20] Z. Shi, C. Chen, Z. Xiong, D. Liu, and F. Wu, "HSCNN+: Advanced CNN-based hyperspectral recovery from RGB images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 939–947.

[21] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, Jun. 2017.

[22] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, Jun. 2014.

[23] E. Rahimy, "Deep learning applications in ophthalmology," *Current Opinion Ophthalmol.*, vol. 29, no. 3, pp. 254–260, 2018.

[24] Q. Rao and J. Frtunikj, "Deep learning for self-driving cars: Chances and challenges," in *Proc. 1st Int. Workshop Softw. Eng. AI Auto. Syst.*, May 2018, pp. 35–38.

[25] S. Ghosal *et al.*, "A weakly supervised deep learning framework for sorghum head detection and counting," *Plant Phenomics*, vol. 2019, p. 1525874, 2019.

[26] Z. Xiong, Z. Shi, H. Li, L. Wang, D. Liu, and F. Wu, "HSCNN: CNN-based hyperspectral image recovery from spectrally undersampled projections," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2017, pp. 518–525.

[27] B. Arad, R. Timofte, O. Ben-Shahar, Y.-T. Lin, and G. D. Finlayson, "NTIRE 2020 challenge on spectral reconstruction from an RGB image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2020, pp. 446–447.

[28] J. Yao, D. Hong, J. Chanussot, D. Meng, X. Zhu, and Z. Xu, "Cross-attention in coupled unmixing nets for unsupervised hyperspectral super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 208–224.

[29] J. Jiang, H. Sun, X. Liu, and J. Ma, "Learning spatial–spectral prior for super-resolution of hyperspectral imagery," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1082–1096, 2020.

[30] B. Kaya, Y. B. Can, and R. Timofte, "Towards spectral estimation from a single RGB image in the wild," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3546–3555.

[31] W. Wei, J. Nie, Y. Li, L. Zhang, and Y. Zhang, "Deep recursive network for hyperspectral image super-resolution," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1233–1244, 2020.

[32] T. Stiebel and D. Merhof, "Brightness invariant deep spectral super-resolution," *Sensors*, vol. 20, no. 20, p. 5789, Oct. 2020.
[33] R. Kawakami, H. Zhao, R. T. Tan, and K. Ikeuchi, "Camera spectral sensitivity and white balance estimation from sky images," *Int. J. Comput. Vis.*, vol. 105, no. 3, pp. 187–204, 2013.
[34] S. Nie, L. Gu, Y. Zheng, A. Lam, N. Ono, and I. Sato, "Deeply learned filter response functions for hyperspectral reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4767–4776.
[35] B. Arad and O. Ben-Shahar, "Filter selection for hyperspectral estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3153–3161.
[36] R. M. H. Nguyen, D. K. Prasad, and M. S. Brown, "Training-based spectral reconstruction from a single RGB image," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 186–201.
[37] E. Wycoff, T.-H. Chan, K. Jia, W.-K. Ma, and Y. Ma, "A non-negative sparse promoting algorithm for high resolution hyperspectral imaging," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 1409–1413.
[38] W. Wei, J. Nie, L. Zhang, and Y. Zhang, "Unsupervised recurrent hyperspectral imagery super-resolution using pixel-aware refinement," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2020.
[39] B. Sovdat, M. Kadunc, M. Batič, and G. Milčinski, "Natural color representation of Sentinel-2 data," *Remote Sens. Environ.*, vol. 225, pp. 392–402, May 2019.
[40] J. Jiang, D. Liu, J. Gu, and S. Susstrunk, "What is the space of spectral sensitivity functions for digital color cameras?" in *Proc. IEEE Workshop Appl. Comput. Vis. (WACV)*, Jan. 2013, pp. 168–179.
[41] G. Finlayson, M. M. Darrodi, and M. Mackiewicz, "Rank-based camera spectral sensitivity estimation," *J. Opt. Soc. Amer.*, vol. 33, no. 4, pp. 589–599, 2016.
[42] Y. Li, C. Wang, J. Zhao, and Q. Yuan, "Efficient spectral reconstruction using a trichromatic camera via sample optimization," *Vis. Comput.*, vol. 34, no. 12, pp. 1773–1783, Dec. 2018.
[43] C. P. Huynh and A. Robles-Kelly, "Recovery of spectral sensitivity functions from a colour chart image under unknown spectrally smooth illumination," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 708–713.
[44] S. Han, Y. Matsushita, I. Sato, T. Okabe, and Y. Sato, "Camera spectral sensitivity estimation from a single image under unknown illumination by using fluorescence," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 805–812.
[45] G. D. Finlayson, S. D. Hordley, and M. S. Drew, "Removing shadows from images," in *Proc. Eur. Conf. Comput. Vis.*, 2002, pp. 823–836.
[46] M. M. Darrodi, G. Finlayson, T. Goodman, and M. Mackiewicz, "Reference data set for camera spectral sensitivity estimation," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 32, no. 3, pp. 381–391, 2015.
[47] X. Meng, H. Shen, Q. Yuan, H. Li, L. Zhang, and W. Sun, "Pansharpening for cloud-contaminated very high-resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2840–2854, May 2018.
[48] J. Ma, W. Yu, C. Chen, P. Liang, X. Guo, and J. Jiang, "Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion," *Inf. Fusion*, vol. 62, pp. 110–120, Oct. 2020.
[49] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2011.
[50] J. Liu, Z. Wu, L. Xiao, J. Sun, and H. Yan, "A truncated matrix decomposition for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 29, pp. 8028–8042, 2020.
[51] 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.
[52] Y. Chang, L. Yan, X. Zhao, H. Fang, Z. Zhang, and S. Zhong, "Weighted low-rank tensor recovery for hyperspectral image restoration," *IEEE Trans. Cybern.*, vol. 50, no. 11, pp. 4558–4572, Nov. 2020.
[53] S. Yang, M. Wang, P. Li, L. Jin, B. Wu, and L. Jiao, "Compressive hyperspectral imaging via sparse tensor and nonlinear compressed sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 5943–5957, Nov. 2015.
[54] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, 2013.
[55] Y. Gao, H. Li, J. Dong, and G. Feng, "A deep convolutional network for medical image super-resolution," in *Proc. Chin. Autom. Congr. (CAC)*, Oct. 2014, pp. 184–199.
[56] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, p. 594, Jul. 2016.
[57] R. Dian, S. Li, B. Sun, and A. Guo, "Recent advances and new guidelines on hyperspectral and multispectral image fusion," *Inf. Fusion*, vol. 69, pp. 40–51, May 2021.
[58] Y. Yan, L. Zhang, J. Li, W. Wei, and Y. Zhang, "Accurate spectral super-resolution from single RGB image using multi-scale CNN," in *Proc. Chin. Conf. Pattern Recognit. Comput. Vis. (PRCV)*, 2018, pp. 206–217.
[59] T. Stiebei, S. Koppers, P. Seltsam, and D. Merhof, "Reconstructing spectral images from RGB-images using a convolutional neural network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2018, pp. 948–953.
[60] I. Omer and M. Werman, "Color lines: Image specific color representation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2004, p. 2.
[61] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse Dirichlet-net for hyperspectral image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2511–2520.
[62] Y. Qu, R. K. Baghbaderani, and H. Qi, "Few-shot hyperspectral image classification through multitask transfer learning," in *Proc. 10th Workshop Hyperspectral Imag. Signal Process., Evol. Remote Sens. (WHISPERS)*, Sep. 2019, pp. 1–5.
[63] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep photo style transfer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4990–4998.
[64] C. Chang, "Spectral information divergence for hyperspectral image analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 1, Jun. 1999, pp. 509–511.
[65] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 8026–8037.
[66] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.
[67] *Airbone Visible/Infrared Imaging Spectrometer*. [Online]. Available: https://aviris.jpl.nasa.gov/data/free_data.html
[68] S. Shahriyar, M. Paul, M. Murshed, and M. Ali, "Lossless hyperspectral image compression using binary tree based decomposition," in *Proc. Int. Conf. Digit. Image Computing: Techn. Appl. (DICTA)*, Nov. 2016, pp. 1–8.
[69] A. Alcolea, M. E. Paoletti, J. M. Haut, J. Resano, and A. Plaza, "Inference in supervised spectral classifiers for on-board hyperspectral imaging: An overview," *Remote Sens.*, vol. 12, no. 3, p. 534, Feb. 2020.
[70] F. A. Kruse *et al.*, "The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data," *Remote Sens. Environ.*, vol. 44, nos. 2–3, pp. 145–163, 1993.
[71] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?" in *Proc. 3rd Conf. Fusion Earth Data, Merging Point Meas., Raster Maps Remotely Sensed Images*, 2000, pp. 99–103.

Jiangsan Zhao received the Ph.D. degree in crop science from the University of Natural Resources and Life Sciences, Vienna, Austria, in 2017.

He currently holds a postdoctoral position at the Graduate School of Agricultural and Life Science, The University of Tokyo, Tokyo, Japan. His research interests include hyperspectral imaging, artificial intelligence, and computer vision in plant phenotyping.

Ying Qu (Member, IEEE) received the B.S. degree in automatics and the M.S. degree in pattern recognition and artificial intelligence from Northeastern University, Shenyang, China, in 2008 and 2010, respectively, and the Ph.D. degree in computer engineering from The University of Tennessee, Knoxville, TN, USA, in 2017.

She is currently a Research Associate with the Department of Electrical Engineering and Computer Science, The University of Tennessee. Her research interests include remote sensing, artificial intelligence, and computer vision.

Dr. Qu was a recipient of the Best Student Paper Awards at the International Geoscience and Remote Sensing Symposium (IGARSS) in 2016.

Seishi Ninomiya received the Ph.D. degree in agriculture from The University of Tokyo, Tokyo, Japan, in 1982.

He is currently a Project Professor with the Graduate School of Agricultural and Life Science, The University of Tokyo, and a Visiting Professor with the Plant Phenomics Research Center, Nanjing Agricultural University, Nanjing, China. His research interests include agro-informatics, statistics, breeding science, and plant phenomics.

Wei Guo received the Ph.D. degree in agriculture from The University of Tokyo, Tokyo, Japan, in 2014.

He is currently a Project Associate Professor with the Graduate School of Agricultural and Life Science, The University of Tokyo. His research interests include machine learning, computer vision, and plant phenomics.