NonRegSRNet: a Non-rigid Registration Hyperspectral Super-Resolution Network

Ke Zheng, Lianru Gao, *Senior Member, IEEE,* Danfeng Hong, *Senior Member, IEEE,* Bing Zhang, *Fellow, IEEE,* and Jocelyn Chanussot, *Fellow, IEEE*

***Abstract*—Due to the limitations of imaging systems, satellite hyperspectral imagery (HSI), which yields rich spectral information in many channels, often suffers from poor spatial resolution. HSI super-resolution (SR) refers to the fusion of high spatial resolution multispectral imagery (MSI) and low spatial resolution HSI to generate HSI that has both a high spatial and high spectral resolution. However, most existing SR methods assume that the two original images used are perfectly registered: in reality, nonrigid deformation areas can exist locally in the two images even if prior registration of the control points has been carried out. To address this problem, we propose a novel unsupervised spectral unmixing and image deformation correction network – NonRegSRNet – with multi-modal and multi-task learning that can be used for the joint registration of HSI and MSI and to produce SR imagery. More specifically, NonRegSRNet integrates the dense registration and SR tasks into a unified model that includes a triplet convolutional neural network. This allows these two tasks to complement each other so that better registration and SR results can be achieved. Furthermore, because the point spread function (PSF) and spectral response function (SRF) are often unavailable, two special convolutional layers are designed to adaptively learn the parameters of the PSF and SRF, which makes the proposed model more adaptable. Experimental results demonstrate that the proposed method has the ability to produce highly accurate and stable reconstructed images under complex non-rigid deformation conditions. (Code available at https://github.com/saber-zero/NonRegSRNet)**

***Index Terms*—Hyperspectral Image, Super-Resolution, Non-rigid Registration, Convolutional Neural Network, Adaptive Learning.**

I. INTRODUCTION

HYperspectral imagery (HSI) corresponds to a data cube that contains spatial information in hundreds of spectral bands. In contrast to traditional imagery that consists of one or only a few bands, the narrow bandwidths and wide spectral coverage of HSI, mean that HSI can be widely used in many applications, such as forestry, agriculture and environmental monitoring.

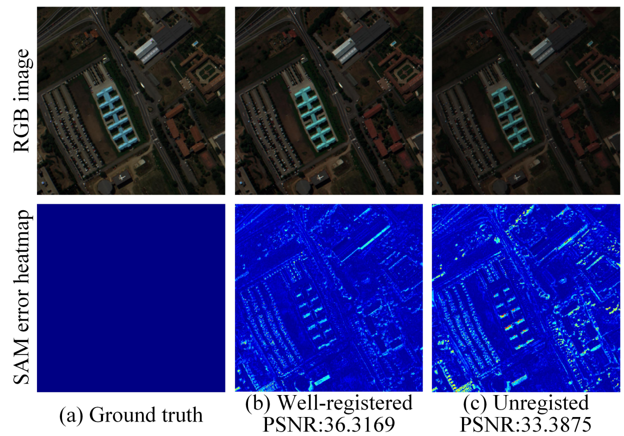

Fig. 1. Comparison of ideal registration and non-rigid deformation showing how the registration accuracy seriously affects the results of the SR reconstruction. Here, CNMF [5] is used as the criterion for the comparison.

However, due to the limitations of imaging system hardware, the penalty for this high spectral resolution is a lower ground sampling distance (GSD), which causes details to be less visible and leads to pixel-mixing effects, thus causing serious problems for the wider application of HSI [1]. By comparison, multispectral imagery (MSI) consists of only a few bands; however, it has a higher spatial resolution and image quality. Naturally, a commonly used method of enhancing the spatial resolution of HSI is to fuse it with high spatial resolution MSI of the same region. This process is known as HSI super-resolution (SR) – HSI SR combines the advantages of the two types of imagery to generate imagery that has both high spatial and spectral resolutions. Using the resulting imagery, further detailed work, such as fine classification or high-precision detection and monitoring can be carried out [2]–[4].

Deformable image registration is of fundamental importance to the fusion of HSI and MSI as the results of the fusion are seriously affected by the accuracy of the registration [6]. Deformable image registration means a dense, non-linear correspondence is established between the HSI and MSI pairs. From Fig. 1, it can be seen that image registration is essential to help mitigate the effects of deformation and facilitate the generation of better fusion results.

This work was supported in part by the National Key R&D Program of China (Grant No. 2021YFB3900502), by the National Natural Science Foundation of China under Grant 42030111 and Grant 41871245, by the China Postdoctoral Science Foundation under Grant 2021M693234, by the MIAI@Grenoble Alpes (ANR-19-P3IA-0003) and the AXA Research Fund.
(*Corresponding author: Lianru Gao*)

K. Zheng, L. Gao, and D. Hong are with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China. (e-mail: zhengkevic@aircas.ac.cn; gaolr@aircas.ac.cn; hongdf@aircas.ac.cn)

B. Zhang is with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China, and also with the College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China. (e-mail: zb@radi.ac.cn)

J. Chanussot is with the Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, 38000 Grenoble, France, and also with the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China. (e-mail: jocelyn@hi.is)

A. Motivation

Recently, many HSI SR methods have been developed and these can be used to produce highly accurate results. However, most of these methods focus only on the fusion process and assume that both the HSI and MSI have already been well registered. Generally, HSI and MSI can be easily registered using rigid registration methods if both types of imagery are captured by the same satellite platform. However, in most cases, the data acquired by different platforms tend to suffer from the effects of complex noises or spectral variabilities in the imaging process [7]. Further, the acquisition time and satellite location for the two types of imagery are different, usually, a non-rigid deformation may be comprised between the two sets of data. This requires the multi-modal data [8] to be transformed into a unified representation so that the non-rigid registration can be achieved. When trying to unify the data representation, the correlation between the HSI and MSI is particularly important. The SR process involves finding this relationship, meaning that a good SR method is of great help in image registration. Good image registration is also an important prerequisite for HSI SR, which means that we believe these two processes can complement each other and help to achieve better registration and fusion results. Nevertheless, there is little published research on the integration of image registration and SR into a single model, especially in the case of non-rigid registration. In this article, we will describe the benefits of the integration of non-rigid registration and HSI SR.

B. Challenges

The main challenges in combining non-rigid registration and image fusion can be summarized as follows:

- **Images.** As they are acquired using different sensors, the spatial and spectral differences between HSI and MSI are large, which results in insufficient information sharing.
- **Consistent space.** As different sensors have very different spatial and spectral response characteristics, it is necessary to transform arbitrary HSI and MSI inputs into a uniform representation space to perform registration.
- **Training strategy.** In the case of remote sensing imagery, the lack of ground-truth deformations and real reconstructed images means that supervised fusion and registration methods serve no purpose. Even if simulated deformations or images are used as the training set, the performance will be restricted due to the limited quality of the training data.
- **Integration.** As the registration and fusion process are based on different principles, it is important to find a way to integrate these two processes into a unified model in which the two processes complement each other.

C. Contributions

The contribution of this paper can be summarized as follows:

- Inspired by the recent success of multi-modal and multi-task deep learning networks, a novel unsupervised non-rigid registration and SR network for HSI and MSI is proposed.
- Registration is integrated into the SR network, and it is shown that this effectively improves the reconstruction accuracy under non-rigid deformation.
- The proposed network is capable of adaptively learning spatial and spectral response functions to improve the stability of the reconstructed imagery.

II. Related Work

A. Related work on HSI SR

Currently used SR methods can be categorized into different types [9]: component substitution-based (CS-based) method [10], multiresolution analysis-based (MRA-based) methods [11], sparse representation methods [12], Bayesian-based methods [13], unmixing-based methods [5] and deep learning-based methods [14].

CS-based and MRA-based methods are derived from pansharpening methods and aim to adapt pansharpening techniques to the HSI SR problem. CS-based methods attempt to replace the intensity component of the HSI with the MSI. The most commonly used of these methods is the adaptive Gram-Schmidt algorithm (GSA) [10] which integrates the effect of the spectral response function (SRF) into the Gram-Schmidt algorithm. In the MRA-based methods, the spatial information extracted from the MSI is injected into the HSI to enhance spatial information [15]. Wavelet coefficient integration and 3-D inverse wavelet transform technology have also been used to fuse HSI and MSI [16]. Selva *et al.* proposed an MRA-based method called hypersharpening in which a high-resolution image was reconstructed for every band of the HSI by using linear regression to produce a linear combination of MSI bands [17]. In general, pansharpening based methods have a high computational efficiency but suffer from unreliable reconstruction quality [18]. Bayesian-based methods maximize the posterior probability density using prior constrains for the HSI SR task [19]. Akhtar *et al.* [20] proposed a Bayesian sparse representation framework to solve the HSI SR problem that inferred the probability distributions for the spectral basis and computed sparse codes for the high-resolution imagery. Kawakami *et al.* [21] applied an unmixing algorithm to estimate the representation of a spectral basis and then used this representation in conjunction with the MSI to produce the reconstructed image. Qi *et al.* [22] integrated a Sylvester equation-based explicit solution into the Bayesian HSI SR task – the resulting method was given the name 'fast fusion based on Sylvester equation' (FUSE). As they are easily interpreted and comprehensible, unmixing-based methods have attracted considerable attention [5], [23]–[26]. Coupled non-negative matrix factorization (CNMF) [5], which uses non-negative matrix factorization to estimate the endmembers and their abundances, is one of the mostly commonly used algorithms for alternately unmixing the HSI and MSI. Similarly, the method described by Lanaaras *et al.* [25], alternately updates the endmembers and their abundances by solving two matrix factorizations. HySure [24] integrates subspace HSI SR with total variation regularization to minimize a convex objective function with respect to subspace coefficients. To extend the matrix factorization, tensor factorization model has been developed to the HSI SR problem [27]–[33]. Chang *et al.* [27] propose a unified low-rank tensor recovery model by treating the singular values differently for HSI restoration tasks. Dian *et al.* [28] proposed a novel non-local sparse tensor factorization based HSI SR by estimating sparse core tensor and dictionaries.

In recent years, deep learning has been outperformed in many computer vision tasks [34]–[36] and also been introduced into HSI SR. Based on the number of input images, deep learning based HSI SR methods can be classified into two types: single HSI SR [37]–[43] and fusion based HSI SR [14], [44]–[53]. Although there have many studies on single HSI SR, compared with fusion-based HSI SR, using single HSI SR, it is more difficult to reconstruct an image with a large scale factor, especially for the low-resolution HSI obtained by satellite sensors. Therefore, in this paper, we focus on fusion-based HSI SR. Dian *et al.* propose an innovative work [44] for the HSI and MSI fusion, which effectively combines the imaging model of the fusion with the powerful learning ability of CNN. Xie *et al.* [46] designed an iterative algorithm to solve the fusion problem by exploiting the proximal gradient method under the low-rankness prior of the observation image. Taking consideration of the large resolution difference in spatial domain of HSI and MSI, Han *et al.* [45] designed a multi-level multi-scale fusion network that gradually changed the feature sizes. Zhang *et al.* [47] integrated residual channel attention and dense blocks to learn spatial-spectral correlation for HSI reconstruction. Wei *et al.* [49] proposed a deep recursive network to implicitly incorporate the deep structure as the regularized prior.

Most of the deep learning based methods described above are supervised learning algorithms that train the known high-resolution HSI or degraded priors, which means that they are difficult to apply in practice when the high-resolution HSI is unavailable [51]. Qu *et al.* [14] proposed an unsupervised deep learning network with sparse Dirichlet distribution for the fusion of HSI and MSI. Zhang *et al.* [48] incorporated spatial and spectral degeneration estimation into a deep blind HSI SR network by utilizing an image-specific generator network to produce the latent HSI. Wang *et al.* [54] also proposed a blind deep SR network that iteratively and alternately optimized estimates the observated data and the fusion process. Zhu *et al.* [51] proposed a progressive zero-centric residual network to learn high-resolution zero-centric residual imagery in a progressive fashion. Zhang *et al.* [55] first implicitly learn a general image prior using deep networks and then adapt it to a special hyperspectral image to improve the generalization of the model. Wei *et al.* [56] proposed an unsupervised recurrence-based HSI SR network that used pixel-aware refinement and which utilized the intermediate reconstruction results to self-supervise unsupervised learning.

B. Related work on Registration

Many remote sensing registration methods have been developed for over a decade [57]–[60]. These traditional registration algorithms commonly use an iterative approach to progressively optimize the problem under constraint [61]. Traditional registration methods are computationally intensive, which causes them to run slowly [62]. The recent great success of deep learning has led to significant progress in image-pair registration, especially for the registration of medical image [63]. Most of these deep learning approaches can be categorized as learning-based methods including supervised and unsupervised methods. Supervised registration methods [64]–[66] require the preparation of sufficiently large training data sets for a fixed input and moving image pairs with the corresponding ground truth transformation;however, their performance is limited by the amount of training data available [67]. For remote sensing data, the difficulty of acquiring reliable ground truth also remains a significant obstacle. In contrast, unsupervised methods only require a similarity measurement between the fixed image and the warped moving image, which is more similar to how traditional registration algorithms work [68]. The optimization is driven by the similarity loss functions – these functions learn the degree of similarity between the image pairs and are more suitable for application to remote sensing data.

The spatial transformer network (STN) [69] is a key innovation in unsupervised registration methods and can be inserted anywhere in a deep-learning network to learn the parameters of the transformation of the input feature map. This means that the network then has the ability to learn the translational invariance of the affine transformation. Li *et al.* [70] introduced a full convolutional network (FCN) to perform non-rigid registration of 3D brain magnetic resonance (MR) imagery using self-supervision. Normalized cross correlation (NCC) was used as the loss function to evaluate the similarity between the warped and fixed images. Based on this idea, de Vos *et al.* [71] used FCN to register 4D cardiac cine MR data. Shu *et al.* [72] proposed a coarse-to-fine unsupervised deformable registration method where the mean squared error (MSE) was used as the loss function between the fixed and warped moving images. Balakrishnan *et al.* [73], [74] proposed an unsupervised learning-based method for medical image registration – VoxelMorph – that could be used to predict a dense deformation field. Zhao *et al.* [68] designed a recursive cascaded network for performing progressive deformation for the registration of deformable images.

C. Related work on HSI Registration and SR

Although the HSI SR problem has been studied by many pioneering researchers, the simultaneously carrying out of SR and registration tasks has been less well investigated. Yokoya *et al.* [75] designed a cross-calibration and fusion method for EO-1 and Terra data that can be considered an early method for registration and fusion. However, using the correlation coefficient is difficult to handle large-scale differences [6]. Recently, several methods that explicitly employ rigid registration operations to enhance the stability of the fusion process have been developed. Lin *et al.* [76] made use of the spectral sparsity to restore misaligned parts of high-resolution HSI and simultaneously employed spectral and spectral structure correlation to restore the aligned areas. Similarly, Fu *et al.*

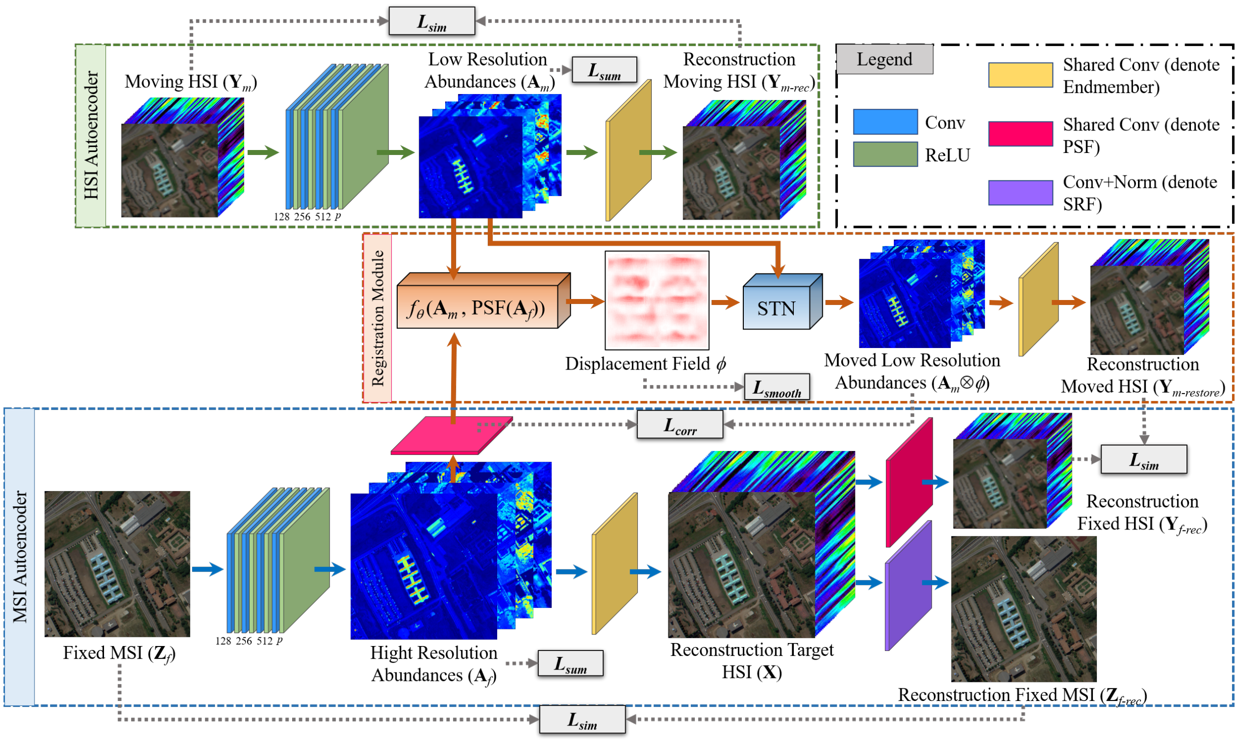

Fig. 2. Flowchart showing the details of our NonRegSRNet for HSI SR and registration. The network includes three sub-modules: the MSI autoencoder, the HSI autoencoder and the registration module.

[77] also presented an approach for simultaneous HSI super-resolution and geometric alignment of a pair of images with drastically contrasting spatial resolutions. By incorporating a spatial transformer network (STN), Nie *et al.* [78] proposed an unsupervised deep learning network to simultaneously achieve HSI SR and registration. However, the methods mentioned above can only be applied to image pairs with affine transformation or rigid deformation.

To handle nonrigid deformation, Zhou *et al.* [6], [79] proposed a registration method that minimized a least-squares objective function by utilizing the point spread function (PSF). After the registration, the fusion method utilized a low-dimensional manifold invariant [80] with local linear transformations to achieve HSI SR. Overall, this approach consists of two-steps: registration followed by image fusion. Qu *et al.* [81] adopted mutual information to capture the non-linear statistical dependencies between the representation from two input data. By maximizing the mutual information, spatial correlations can be characterized to reduce the spectral distortion.

III. PROPOSED APPROACH

A. Overview

The challenges that we faced in this study drive us to design a network that incorporated both registration and fusion, had the ability to carry out these functions successfully in a balanced way and was robust. Fig. 2 shows the framework of the proposed NonRegSRNet: it consists mainly of an HSI autoencoder subnetwork, an MSI autoencoder subnetwork and a registration subnetwork. Specifically, the input images to be registered and fused are fed into the autoencoder networks to solve the linear spectral unmixing with the same endmember components. The registration network acts as a link that re-establishes the spatial relations between the abundances, whereas the spatial deformation correction is performed using displacement field prediction and the spatial transformer network. The outputs of each subnetwork and the corresponding self-similar component parts are fed into the loss function to drive the unsupervised learning.

B. HSI/MSI Autoencoder Networks

In the registration formulation, the moving image (also also called as source image) is warped to register with the fixed image (also called as target image) [74]. We can assume that the high-resolution multispectral image is the fixed image as it contains more detailed information about the spatial features of the ground objects and that the low-resolution hyperspectral image is the moving image. Let $\mathbf{Y}_m$ and $\mathbf{Z}_f$ denote the input low-resolution hyperspectral image and high-resolution multispectral image defined over 3-D domains respectively: $\mathbf{Y}_m \in \mathbb{R}^{m \times n \times L}$ and $\mathbf{Z}_f \in \mathbb{R}^{M \times N \times l}$, where m and n are the width and heigh of $\mathbf{Y}_m$, respectively, and L is the number of spectral bands; and M, N are the width, heigh of $\mathbf{Z}_f$, respectively and l is the number of bands. For convenience, we call the low-resolution hyperspectral image and high-resolution multispectral image as LrHSI and HrMSI, respectively. The HSI SR aims to produce a high-resolution HSI $\mathbf{X} \in \mathbb{R}^{M \times N \times L}$($e.g.$ $l \leq L, n \leq N$ and $m \leq M$). The relation between $\mathbf{X}$ and the observations $\mathbf{Y}_m$ and $\mathbf{Z}_f$ can be formulated as:

$$\mathbf{Y}_m = g(\mathbf{S}\mathbf{X}) + \mathbf{E}_s, \tag{1}$$

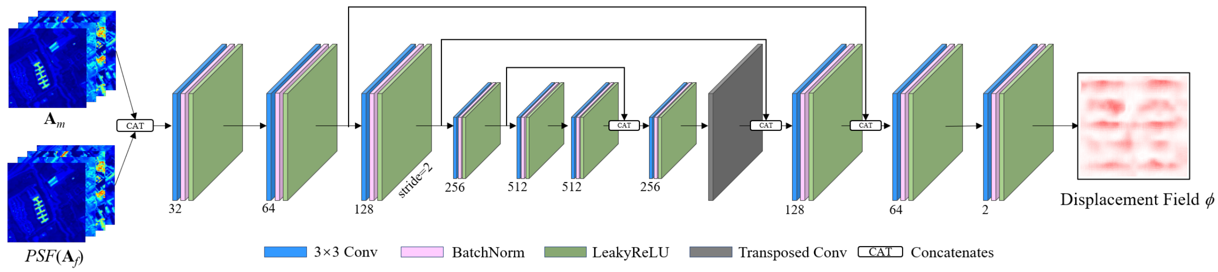

Fig. 3. Part of the registration module used to predict the displacement field θ. This part of the module can be represented by $f_\theta(\mathbf{A}_m, PSF(\mathbf{A}_f))$.

$$\mathbf{Z}_f = \mathbf{XR} + \mathbf{E}_r, \tag{2}$$

where $\mathbf{S} \in \mathbb{R}^{mn \times MN}$ denotes the spatial spread matrix representing the transform of the PSF from high-resolution image to low-resolution image. $\mathbf{R} \in \mathbb{R}^{L \times l}$ denotes the spectral response transform matrix representing the transform of SRF from the hyperspectral sensor to the multispectral sensor. $g()$ denotes the spatial deformation transformation representing image geometric distortion caused by the lens distortion, the perspective of the sensor optics, the motion of the scanning system or the platform altitude, *etc.* $\mathbf{E}_s$ and $\mathbf{E}_r$ are the residuals.

In the HSI autoencoder, a set of 2-D convolutional layers and ReLU layers are stacked so that $\mathbf{Y}_m$ is mapped to its corresponding low-resolution abundances $\mathbf{A}_m \in \mathbb{R}^{m \times n \times p}$, where p is the number of endmembers. In these layers, the kernel sizes of the convolutional layer are set to 1×1 to preserve the spatial structure of the input image. The number of output channels for these layers is shown in Fig. 2, where p is the output channels of the last convolutional layer. The reason we define $\mathbf{A}_m$ as the abundances is that a shared one-layer convolutional layer - shown as the yellow layer in Fig. 2 - is applied over $\mathbf{A}_m$ to implement matrix multiplication, where the kernel size of this convolutional layer is 1×1 and no bias is defined in this convolutional layer. Each convolutional kernel is element-wise product with each pixel of the abundance. Therefore, the parameters of the convolutional layer can be defined as the endmembers $\mathbf{E} \in \mathbb{R}^{p \times 1 \times 1 \times L}$. The output of the shared convolutional layer can be defined as:

$$\mathbf{Y}_{m-rec} = \mathbf{A}_m \times \mathbf{E}, \tag{3}$$

where $\mathbf{Y}_{m-rec} \in \mathbb{R}^{m \times n \times L}$ is the reconstruction of the input $\mathbf{Y}_m$. This means that the linear spectral unmixing approach is embedded into the reconstruction process.

Similarly, the input HrMSI $\mathbf{Z}_f$ is fed into a group of convolutional layers and ReLU layers to generate high-resolution abundances given by $\mathbf{A}_f$. After being processed by the shared convolutional layer $\mathbf{E}$, $\mathbf{A}_f$ is transformed into the target HrHSI $\mathbf{X}$, which can be defined as:

$$\mathbf{X}\text{=}\mathbf{A}_f \times \mathbf{E}. \tag{4}$$

In order to form a closed loop for unsupervised learning, a convolutional layer with kernel size 1×1 and a normalization layer are combined at the back of the target image $\mathbf{X}$ to act as the SRF process - this is shown as purple layer in Fig. 2. The output of this process is the reconstructed HrMSI $\mathbf{Z}_{f-rec}$, which can be defined as:

$$\mathbf{Z}_{f-rec} = SRF(\mathbf{X}), \tag{5}$$

A more detailed explanation of $SRF()$ is given below:

$$z_i = SRF(x_\lambda) = \frac{\sum\limits_{\lambda=\lambda_{i,L}}^{\lambda_{i,U}} w_{i,\lambda} x_\lambda}{\sum\limits_{\lambda=\lambda_{i,L}}^{\lambda_{i,U}} w_{i,\lambda}}, \tag{6}$$

where $w_{i,\lambda}$ denotes the weights of the convolutional layer, x_λ is the band in $\mathbf{X}$ with wavelength λ, z_i is the band in $\mathbf{Z}$ with wavelength i, U and L are the spectral coverage upper and lower bound of the i-th band of $\mathbf{Z}$.

The PSF process gives the local spatial correlation between the low-resolution image and the high-resolution image, which means that, under the premise of the two images are well registered, a pixel in the low-resolution image is a weighted combination of the corresponding pixel and its neighboring pixels in the high-resolution image. According to the results of our previous work [53], the parameters of the PSF can be learned by placing a convolutional layer with one input channel and one output channel on the high-resolution image to generate the low-resolution image. The kernel size and the stride of this convolutional layer should be set equal to the scale factor. This convolutional layer can be applied to $\mathbf{A}_f$ and $\mathbf{X}$ to transform the high-resolution abundances and image into the low-resolution uniform representation space in preparation for the registration. The $\mathbf{A}_f$ processed using the PSF is denoted by $PSF(\mathbf{A}_f)$ and the fixed reconstructed HSI can be formulated as:

$$\mathbf{Y}_{f-rec} = PSF(\mathbf{X}). \tag{7}$$

C. Registration Module

In this section, we describe the registration module which was used for dense matching of corresponding pixels. Generally, the full-frame registration of HSI and MSI can be easily achieved using affine alignment methods that are applied to large-scale remote sensing imagery. However, it is more difficult to deal with local nonlinear deformation, such as the deformation of an individual object or a local area. Therefore, in this study, we assumed that $\mathbf{Y}_m$ and $\mathbf{Z}_f$ were affinely aligned during the preprocessing and that only nonlinear deformation of the two images existed.

As shown in Fig. 3, the module function can be denoted by $f_\theta(\mathbf{A}_m, PSF(\mathbf{A}_f))$, where θ are the parameters of the convolutional layers and the module network that resembles UNet [82]. The concatenated $\mathbf{A}_m$ and $PSF(\mathbf{A}_f)$ are fed into $f_\theta()$, where $\mathbf{A}_m$ represents the required registration abundances obtained using LrHSI autoencoder and $PSF(\mathbf{A}_f)$ is the output of PSF process obtained from HrMSI autoencoder. The output of this module is the displacement fields $\phi \in \mathbb{R}^{m\times n\times 2}$, where m and n are the width and height of ϕ, and 2 is the number of feature maps for the x- and y-axes. Displacement fields represent the displacement vectors for all the points of the image that is displaced from $\mathbf{A}_m$ to $PSF(\mathbf{A}_f)$. Displacement fields can be formulated as:

$$\phi{=}f_\theta(\mathbf{A}_m, PSF(\mathbf{A}_f)). \tag{8}$$

In this module, the convolutional layers extract the hierarchical features of the input abundances pairs to generate the displacement field, ϕ. Both the encoder and decoder consist of a stacked convolutional layer, batch normalization layer and LeakReLU layer. In the encoder part, as shown in Fig. 3, we set the stride of convolutional layer to 2 to reduce the size of the feature maps used for forming the multi-scale feature representation. In the decoder part, a transposed convolutional layer was used to increase the size of the feature maps to restore them to their original size. However, to enhance the feature diversity, similar to the image/feature pyramid, concatenated skip connections were used to connect features at different levels.

In order to register the deformed moving abundance $\mathbf{A}_m$, we introduce spatial transformer network (STN) [74] to warp $\mathbf{A}_m$ using ϕ, where ϕ represents the offset of input pixels. Firstly, we need to generate a standard mesh that is the same as the size of the deformation field. Secondly, the standard mesh plus the deformation field to form a sampling grid. Thirdly, the grid values were normalized to [-1, 1] for the next re-sampling. Finally, the normalized grid is used to re-sample $\mathbf{A}_m$ to generate the output moved image $\mathbf{A}_m\otimes\phi$. After processing by the shared endmember convolutional layer, the interpolated abundances can be transformed into the registered LrHSI $\mathbf{Y}_{m-restored}$, which can be written as:

$$\mathbf{Y}_{m-restored} = \mathbf{A}_m \otimes \phi \times \mathbf{E}, \tag{9}$$

where $\otimes$ denotes the grid sample, $\mathbf{A}_m$ is the moving abundance, ϕ is the displacement field, and $\mathbf{E}$ represents the parameters of the shared convolutional layer.

D. Loss Functions

To train our model using an unsupervised method, we defined several loss functions to constrain the registration and SR tasks, which was also illustrated in Fig. 2.

1) Unsupervised Similarity: An unsupervised similarity loss $\mathcal{L}_{sim}$ measures the similarity between the input data and the outputs. We defined the L1 norm as our similarity function:

$$\begin{aligned}\mathcal{L}_{sim} = &\left\|\mathbf{Y}_m - \mathbf{Y}_{m-rec}\right\|_1 + \alpha\left\|\mathbf{Z}_f - \mathbf{Z}_{f-rec}\right\|_1 \\ &+ \beta\left\|\mathbf{Y}_{m-restore} - \mathbf{Y}_{f-rec}\right\|_1,\end{aligned} \tag{10}$$

where α and β are trade-off parameters used to balance the contributions of different losses.

2) Correlation Coefficient: We used the correlation coefficient to minimize the similarity between $PSF(\mathbf{A}_f)$ and $\mathbf{A}_m \otimes \phi$:

$$Corr(PSF(\mathbf{A}_f), \mathbf{A}_m\otimes\phi) = \frac{Cov(PSF(\mathbf{A}_f), \mathbf{A}_m\otimes\phi)}{\sqrt{Var(PSF(\mathbf{A}_f))Var(\mathbf{A}_m\otimes\phi))}}, \tag{11}$$

where Cov is the covariance and Var is the variance. A higher correlation coefficient means a higher similarity. Therefore the loss function used was:

$$\mathcal{L}_{corr}(PSF(\mathbf{A}_f), \mathbf{A}_m\otimes\phi) = 1 - Corr(PSF(\mathbf{A}_f), \mathbf{A}_m\otimes\phi). \tag{12}$$

3) Smooth Constraint: Miniminzing the $\mathcal{L}_{corr}$ may encourage ϕ to be non-smooth; therefore, we defined a diffusion regularizer to smooth the displacement field ϕ:

$$\mathcal{L}_{smooth}(\phi) = \left\|\frac{\partial\phi}{\partial x}\right\|_2 + \left\|\frac{\partial\phi}{\partial y}\right\|_2, \tag{13}$$

where the differences between neighboring pixels can be used to compute the approximate spatial gradients: $\frac{\partial\phi}{\partial x} \approx \phi(x+1,y) - \phi(x,y)$, $\frac{\partial\phi}{\partial y} \approx \phi(x,y+1) - \phi(x,y)$.

4) Summation Constraint: Regarding the spectral unmixing discussed in this paper, we wanted the abundance to satisfy sum-to-one and non-negative constraints. To this end, we introduced a summation constraint on the abundance $\mathbf{A}_m$, $PSF(\mathbf{A}_f)$ and $\mathbf{A}_f$:

$$\begin{aligned}\mathcal{L}_{sum}(\mathbf{A}_m, PSF(\mathbf{A}_f), \mathbf{A}_f) = &\left\|1 - \sum_{i=1}^{P}\mathbf{A}_m^i\right\|_1 + \left\|1 - \sum_{i=1}^{P}\mathbf{A}_f^i\right\|_1 \\ &+ \left\|1 - \sum_{i=1}^{P}PSF(\mathbf{A}_f)^i\right\|_1,\end{aligned} \tag{14}$$

where P denotes the the number of endmembers. To satisfy the non-negative constraint, a clamp function was added at the back of the abundance $\mathbf{A}_m$ and $\mathbf{A}_f$.

The parameters of the PSF layer, SRF and endmember layers should also meet the non-negative constraint. Therefore, we imposed a clamp function on these parameters after the parameters were updated during each back-propagation to force these weights to lie within the range [0, 1].

In summary, the final loss function for the proposed network can be written as:

$$\mathcal{L}_{total} = \mathcal{L}_{sim} + \gamma\mathcal{L}_{corr} + \delta\mathcal{L}_{smooth} + \mu\mathcal{L}_{sum}, \tag{15}$$

where γ, δ and μ are the trade-off parameters.

IV. EXPERIMENTS AND RESULTS

In the section, we describe the application of our proposed registration and SR method under different experimental conditions. First, we describe the datasets used. To get an accurate assessment of the quality of the registration and SR tasks, we carried out simulation experiments. Following this, we performed a sensitivity analysis in which we investigated the performance of the hyperparameters and used an ablation

TABLE I
ABLATION STUDY.

	MSI*	HSI*	registration*	α	β	γ	δ	μ	SAM	ERGAS	mPSNR
well-registerted	✔	✔	✗	✔	✔	✔	✔	✔	36.894	79.243	22.993
	✔	✔	✔	✗	✔	✔	✔	✔	5.475	29.828	20.020
	✔	✔	✔	✔	✗	✔	✔	✔	36.992	828.916	21.281
	✔	✔	✔	✔	✔	✗	✔	✔	4.597	4.070	32.231
	✔	✔	✔	✔	✔	✔	✗	✔	4.438	3.232	38.187
	✔	✔	✔	✔	✔	✔	✔	✗	4.434	3.101	38.235
	✔	✔	✔	✔	✔	✔	✔	✔	**4.420**	**3.037**	**38.475**
2 pixels deformation	✔	✔	✗	✔	✔	✔	✔	✔	39.488	161.913	22.527
	✔	✔	✔	✗	✔	✔	✔	✔	5.465	33.538	33.538
	✔	✔	✔	✔	✗	✔	✔	✔	53.989	2405.3	17.684
	✔	✔	✔	✔	✔	✗	✔	✔	4.989	5.833	29.779
	✔	✔	✔	✔	✔	✔	✗	✔	4.499	3.104	38.167
	✔	✔	✔	✔	✔	✔	✔	✗	4.477	3.056	38.231
	✔	✔	✔	✔	✔	✔	✔	✔	**4.439**	**3.043**	**38.386**

MSI*, HSI* and registration* represents MSI-Autoencoder, HSI-Autoencoder and Registration module, respectively.

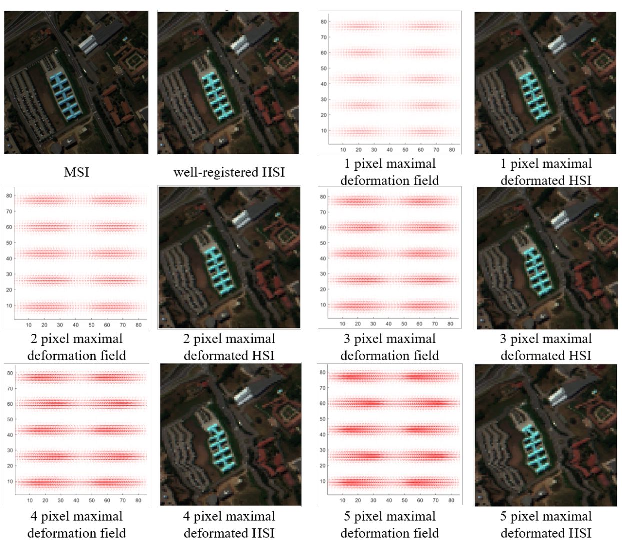

Fig. 4. Simulated Pavia University data with different maximal deformations applied to simulate non-rigid transformations.

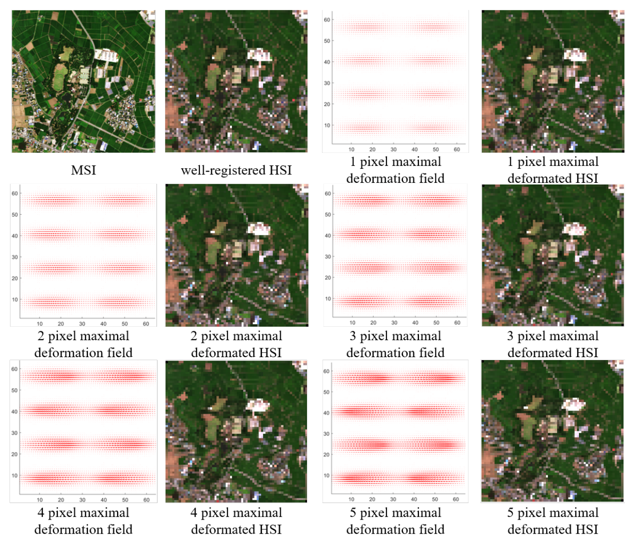

Fig. 5. Simulated Chikusei data with different maximal deformations applied to simulate non-rigid transformations.

TABLE II
REGISTRATION QUANTITATIVE COMPARISON UNDER DIFFERENT DEFORMATION MAGNITUDE.

		Maximal Deformation Magnitude				
		1 pixel	2 pixels	3 pixels	4 pixels	5 pixels
Pav*	LSQ	**0.3193**	0.6162	0.8556	1.1036	1.2236
	Ours	0.3343	**0.6016**	**0.8535**	**0.9943**	**1.1241**
Chi*	LSQ	0.6137	0.7679	0.8795	1.0177	1.1950
	Ours	**0.4468**	**0.6670**	**0.8228**	**1.0630**	**1.1765**
Wa*	LSQ	0.3100	0.5924	0.8155	1.0283	1.2007
	Ours	**0.3080**	**0.5316**	**0.7679**	**1.0102**	**1.1468**

Pav*, Chi* and Wa* represent Pavia University, Chikusei and Washington, D.C. data, respectively.

study to verify the effectiveness of proposed method. Finally in this section, comparisons with different registration and SR methods are made.

A. Experimental Data and Implementation Details

Simulation data. Three widely used HSI datasets were used as simulation data in our experiment. These were the Pavia University dataset, Chikusei dataset and Washington, D.C. dataset. The Pavia University dataset was acquired by the ROSIS-3 sensor in 2003. The original Pavia data consist of 610×340 pixels in 115 bands covering the range 430 nm to 840 nm with a GSD of 1.3m. We selected the top-left part of the image consisting of 340×340 pixels and 103 bands for use in our experiment. The Chikusei dataset was captured by the Headwall airborne hyperspectral sensor over Chikusei, Ibaraki, Japan. The original data comprise 2517×2335 pixels and 128 bands in the spectral range from 363 nm to 1018 nm with a GSD of 2.5m. We used the top-right corner of the HSI covering 512×512 pixels. The Washington, D.C. data were acquired by the HYDICE sensor in 1995. This dataset covers 1280×307 pixels in 210 bands in the spectral range 400 nm to 2500 nm and has a GSD of 2.5m. After removing

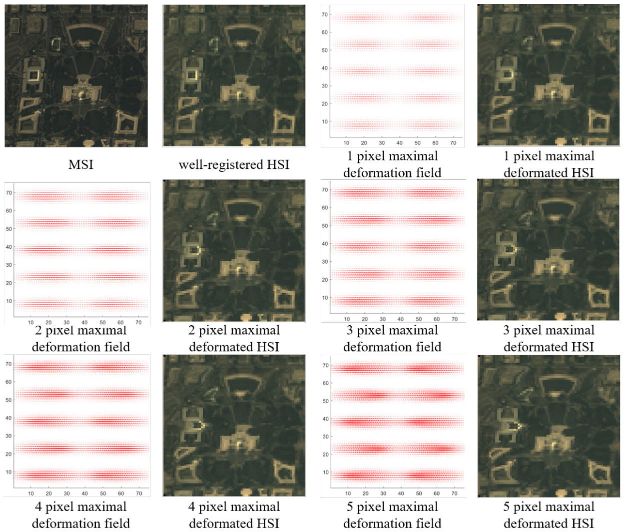

Fig. 6. Simulated Washington, D.C. data with different maximal deformations applied to simulate non-rigid transformations.

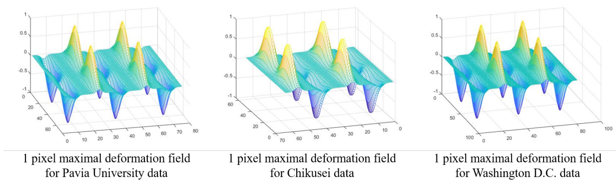

Fig. 7. 3D visualization diagram for deformation fields.

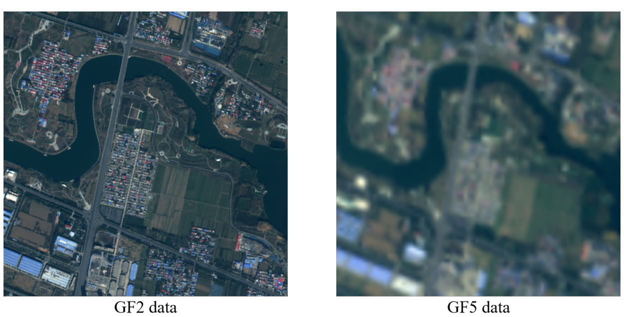

Fig. 8. Color composite of the GF2 and GF5 data.

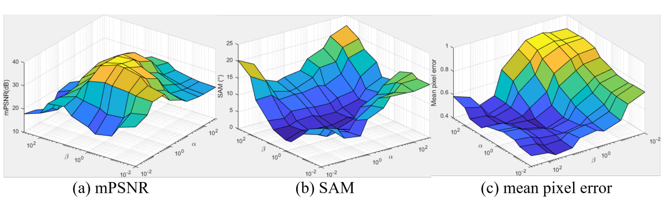

Fig. 9. Results of the hyperparameter analysis for the trade-off weights α and βbased on the Pavia University dataset.

the noisy bands, a 300×300 -pixel section with 191 bands from the lower part of the image was selected for use in our experiment.

Simulation details. The original HrHSI was used as a reference for conducting the quality assessment. The LrHSI was generated using an isotropic Gaussian PSF. The SR scale ratios were set as 4, 8 and 4 for the Pavia University, Chikusei, and Washington, D.C. datasets, respectively. We used

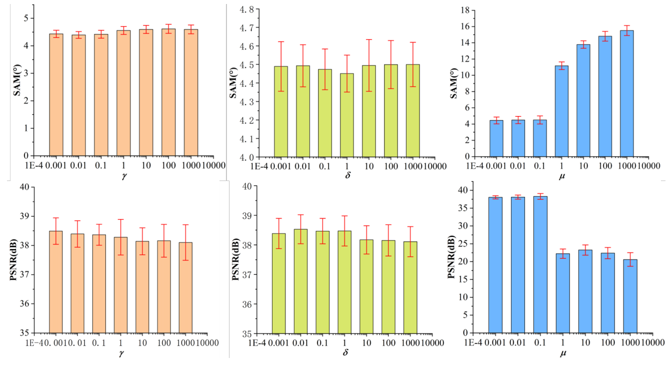

Fig. 10. Results of the sensitivity analysis for the parameters γ, δ and μ.

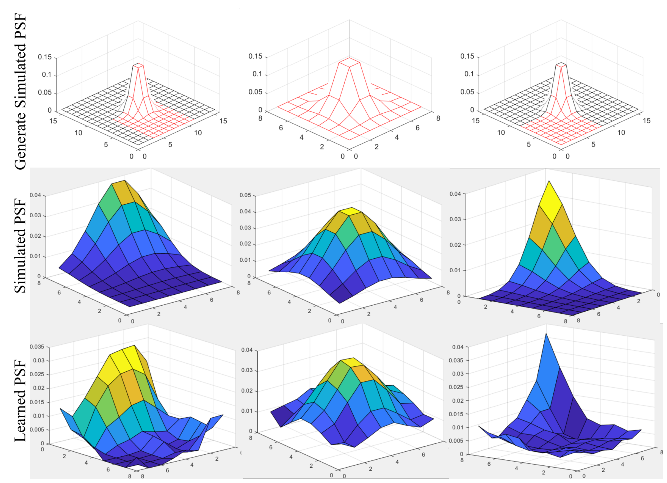

Fig. 11. Visualization of different learned PSF kernels.

the blue–green–red Landsat-8 SRF, the blue–green–red–near-infrared SRF and the blue to SWIR2 SRF to generate the HrMSI data for the Pavia University, Chikusei and HrMSI Washington, D.C. datasets, respectively.

To simulate the non-rigid deformed LrHSI, we linearly combine multiple Gaussian distributions to generate the deformation field. The non-linear transformation can be formed as $\mathcal{T}(x)=x+v(x)$ [6]. $v(x)$ can be formulated as Gaussian mixture distribution $v(x)=\sum_{k} \mathbf{c}_k \mathcal{N}\left(x \mid \mu_k, \sigma^2\right)$, where k is the kth Gaussian component, $\mathbf{c}_k$ is the kth mixture coefficient, u_k is the mean of the kth component and σ is the standard deviation. The non-linear transformation was applied to the spatially downsampled LrHSI (see Fig. 4, 5 and 6), where the 3D visualization diagrams of the deformation fields are shown in Fig.7.

Real data. In this study, we have also verified the proposed method on real data, where GF2 data and GF5 data were used. The GF2 data was acquired by GF2-Panchromatic Multispectral (PMS) sensor on 13-Nov-2019 over Xuchang city, Henan province, China. This data covers 7304×7304 pixels with a GSD of 4m in 4 bands, including Blue-Green-Red-NearIR bands. The GF5 data was acquired by GF5-Advanced Hyperspectral Imager (AHSI) on 10-Nov-2019. It consists of 2083×2008 pixels in 330 bands covering the range 390 nm to 2500 nm with a GSD 30m. We used ENVI ROI tools to select a part of two images for the experiment, where a part of

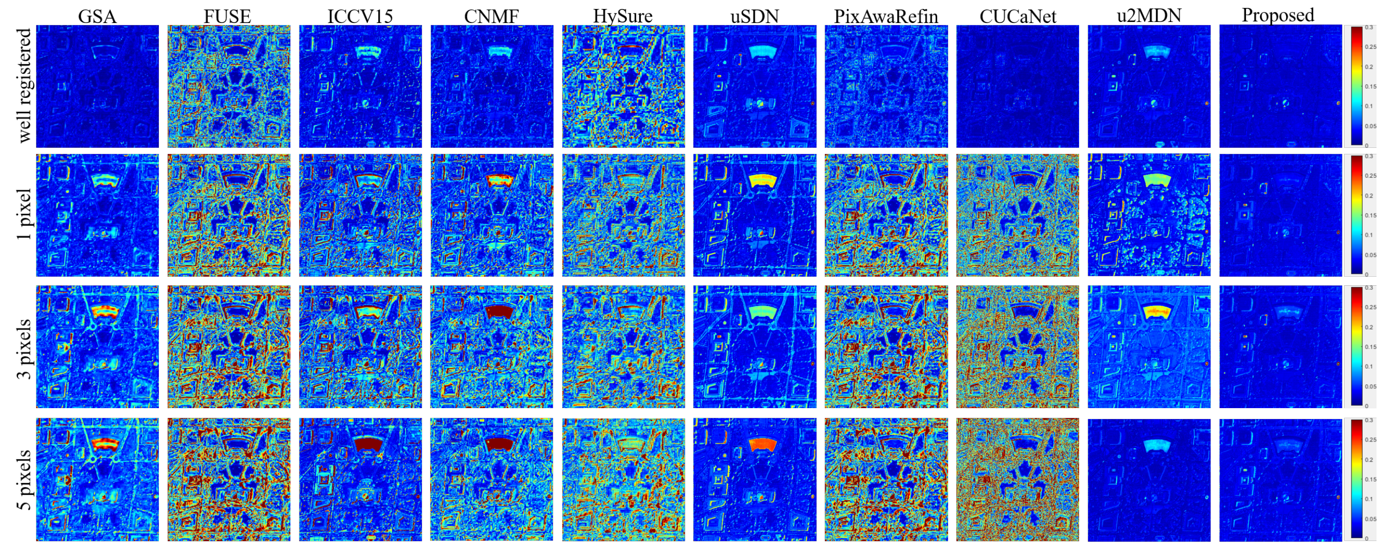

Fig. 12. Visual comparison of the quality of the results obtained under a range of deformation conditions when applying different SR methods to the Washington, D.C. data. The values in the error images are those of the MRAE. 'well-registered' means that the input data consist of simulated perfectly registered data. "1 pixel" means that, first, the LSQ free-form method [6] is used to register the input data whose maximum deformation magnitude is 1 pixel; different SR methods are then used to fuse the registered images. Similarly, "3 pixels" and "5 pixels" indicate that the maximum deformation magnitudes are 3 pixels and 5 pixels, respectively.

TABLE III
THE QUALITY ASSESSMENT COMPARISON ON WASHINGTONG, D.C. DATA. THE BEST RESULTS ARE SHOWN IN BOLD.

	Maximal Deformation Magnitude											
	well-registration		1 pixel		2 pixel		3 pixel		4 pixel		5 pixel	
Quality	SAM	PSNR	SAM	PSNR	SAM	PSNR	SAM	PSNR	SAM	PSNR	SAM	PSNR
GSA	1.7659	40.6485	3.2517	35.2123	3.9126	33.6086	4.1861	33.0094	4.3056	32.4682	4.9615	31.4813
FUSE	5.5815	28.8595	7.4453	26.4161	8.0423	25.8442	8.3183	25.5777	8.4248	25.3308	8.8301	24.9587
ICCV15	2.2644	36.9433	5.3499	31.0462	6.2442	29.9275	6.5231	29.6324	6.5053	29.4433	4.2015	30.1084
CNMF	2.0190	37.4056	3.9408	30.5595	4.7932	28.5659	5.4317	28.4536	5.3757	27.1936	6.4829	25.3495
HySure	5.9193	30.5364	7.1883	27.9395	6.9179	27.5599	6.9594	27.6011	7.2599	25.9982	7.4751	25.6595
uSDN	2.2519	36.5113	2.8438	35.3096	2.3976	35.0996	3.0334	35.0956	2.3594	35.5588	3.0230	34.4248
PixAwaRefin	2.7825	34.2016	7.6682	26.5661	8.4228	26.0796	8.9295	25.4634	8.9284	25.6152	9.4940	25.2611
CUCaNet	1.4873	**40.7987**	7.8772	26.1870	9.5363	24.3924	10.0328	24.0147	11.1339	23.5353	11.4123	23.1614
u2MDN	2.0566	35.7389	2.9347	32.2905	4.4656	31.2905	4.3619	31.3248	4.7136	31.3804	2.0352	37.1595
Proposed	**1.4103**	39.3093	**1.4674**	**38.2439**	**1.6677**	**38.8564**	**1.7376**	**38.6041**	**1.7642**	**38.1867**	**1.6952**	**39.0536**

600×600 pixels section from the GF2 data was selected and 78×78 pixels section from GF5 was selected. In order to make the scale factor be an integer, we upsampled the GF5 data to a size of 100×100 pixels. In addition, we use relative radiation normalization to linear normalize the spectrum of GF2 data to the corresponding spectrum of GF5 data. The RGB images are shown in Fig8

Experimental environment. The proposed network was implemented on the PyTorch framework [83], and was trained using an Adam optimizer set to its default parameters [84]. Kaiming parameter initialization was used for all the network layers. The inital learning rate was set to 0.005 with linear step decay schedules set from 2000 to 10000 epochs.

Quality assessment. For the experiment, we used several widely used quality indices to compare the registration and reconstruction quality: these included the mean pixel error, spectral angle mapper (SAM), erreur relative globale adimensionnelle de synthese (ERGAS), mean peak signal-to-noise ratio (mPSNR), structural similarity index measure (SSIM), mean square error (MSE) and mean relative absolute error (MRAE) [6], [75].

B. Sensitivity Analysis

Ablation study. Our proposed method consists of three basic modules: the HSI autoencoder, MSI autoencoder and registration module. To investigate the performance of different combinations of components under different deformation conditions, we performed a simple experiment using the Pavia University data. Table I shows the experimental results. It can be seen that better performance is available only when the three modules are combined. And we also test the ablation study for the loss functions. The experiment shows that the trade-off parameters α, β and γ play an important role in improving the performance. In comparison, θ and μ also useful to stabilize the result.

Analysis of hyperparameters. In the proposed method, several trade-off parameters are used to balance the weights of the different loss functions. As the autoencoder is driven by

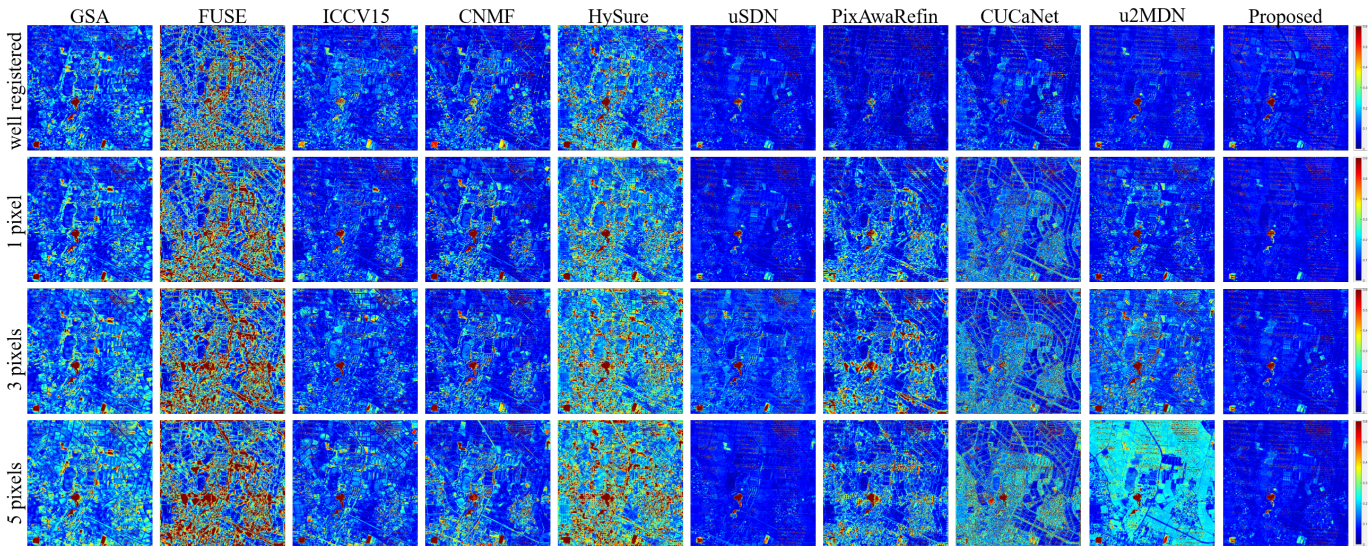

Fig. 13. Visual comparison of the quality of the results obtained under a range of deformation conditions when applying different SR methods to the Chikusei data. "well-registered" means that the input data consist of simulated perfectly registered data. "1 pixel" means that, first, the LSQ free-form method [6] is used to register the input data whose maximum deformation magnitude is 1 pixel; different SR methods are then used to fuse the registered images. Similarly, "3 pixels" and "5 pixels" indicate that the maximum deformation magnitudes are 3 pixels and 5 pixels, respectively. to register the input data whose maximum deformation magnitude is 1 pixel, and then different SR methods are used to fuse the registered images. Similarly, "3 pixel" and "5 pixel" indicate the maximum deformation magnitude are 3 pixels and 5 pixels, respectively.

TABLE IV
THE QUALITY ASSESSMENT COMPARISON ON CHIKUSEI DATA. THE BEST RESULTS ARE SHOWN IN BOLD.

	Maximal Deformation Magnitude											
	well-registration		1 pixel		2 pixel		3 pixel		4 pixel		5 pixel	
Quality	SAM	PSNR	SAM	PSNR	SAM	PSNR	SAM	PSNR	SAM	PSNR	SAM	PSNR
GSA	3.8250	34.8901	4.2855	34.2813	4.5532	33.8065	5.0290	33.0524	5.3537	32.4267	5.7613	31.9719
FUSE	4.9040	27.3041	5.7819	26.2111	5.9671	26.0157	6.2546	25.7282	6.3768	25.5476	6.5955	25.3155
ICCV15	2.6857	35.4171	2.6720	35.4157	2.6929	35.1356	3.1156	34.0847	3.5472	32.5006	3.5619	31.9751
CNMF	3.1302	33.1852	3.3271	36.0164	3.245	35.9315	2.9832	34.8000	3.2396	32.8205	4.3615	32.5540
HySure	4.8299	32.2683	5.3104	31.7977	5.6831	31.2274	6.5563	30.1148	6.8996	29.5827	7.9536	28.4353
uSDN	2.5324	41.0687	2.6281	40.4927	2.7395	40.2941	4.9915	39.7581	3.3215	40.0190	4.9425	40.2180
PixAwaRefin	**2.0827**	40.2701	4.4631	32.2967	4.6216	31.4546	5.2675	31.1170	5.4171	30.4846	5.4102	30.0807
CUCaNet	2.6742	40.4798	4.8033	27.9320	5.1783	25.1761	5.3624	24.9317	6.1006	24.2003	6.4182	23.8122
u2MDN	2.5907	41.0254	2.6910	40.8319	2.7169	40.0414	3.3759	37.3539	2.8885	38.8059	4.0506	35.9707
Proposed	2.5688	**41.1742**	**2.5704**	**41.2253**	**2.6054**	**40.8901**	**2.6158**	**40.5483**	**2.6124**	**40.5996**	**2.5925**	**40.3664**

the reconstruction errors, we first compared the performance of the model using different values of the parameters α and β and with the remaining parameters set to 0.01 by default. A grid search was used to find the optimal parameter values and the results are shown in Fig.9. It can be found that good results were obtained for the target image for small values of α and β , with the best results being achieved when $\alpha = \beta = 1$.

Fig. 10 shows the results of the sensitivity analysis for γ, δ and μ. The reconstruction accuracy is relatively little affected by γ and δ but is much more sensitive to μ. We set $\gamma = 0.001, \delta = 1, \mu = 0.001$ for the subsequent experiments.

Registration comparison. To compare the performance of different non-rigid registration methods, we chose least-squares non-rigid registration [6] as the benchmark method for comparison. The registration errors obtained for different deformation magnitudes using the three datasets are shown in Table.II. Here, 'LSQ' denotes the LSQ free-form non-rigid registration method. It can be seen that, in most cases, a higher registration accuracy can be achieved using the proposed method. This is because the proposed method is combined with iterative estimation of the response function and registration, which ensures more consistent registration relative to other methods.

Point spread function. In order to verify the ability of the proposed model to estimate the PSF, we used some regions of Gaussian distribution as the simulated PSF, as shown in the first row of Fig. 11. The region of the red grid is the selected simulated PSF. The second row shows three different simulated PSF kernels to generate LrHSI. Here, the Chikusei data was used as the example and the PSF kernel size was 8. The unsupervised estimated PSFs are visualized in Fig.11 as 'learned PSF'. All the overall shapes of the estimated PSFs appear to be very similar to those of the original PSFs. Locally, the estimated values differ from the real ones, which causes the learned PSFs to be slightly less smooth than the simulated ones. This is because a slight error may induce irregularity

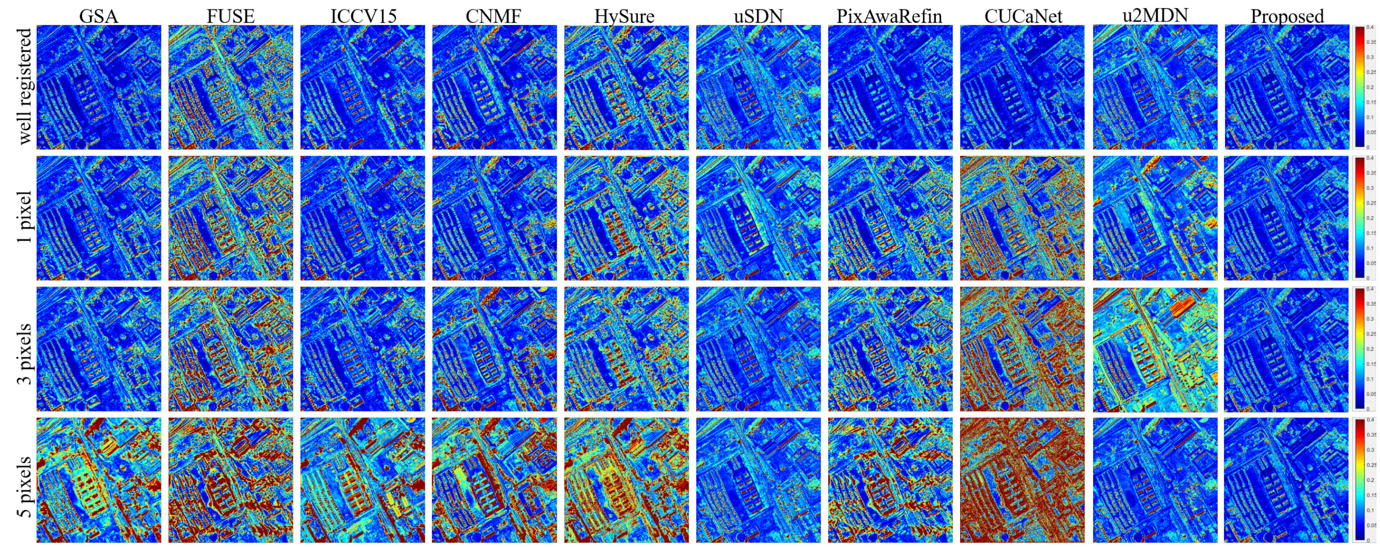

Fig. 14. Visual comparison of the quality of the results obtained under a range of deformation conditions when applying different SR methods to the Pavia University data. "well-registered" means that the input data consist of simulated perfectly registered data. "1 pixel" means that, first, the LSQ free-form method [6] is used to register the input data whose maximum deformation magnitude is 1 pixel; different SR methods are then used to fuse the registered images. Similarly, "3 pixels" and "5 pixels" indicate that the maximum deformation magnitudes are 3 pixels and 5 pixels, respectively.

TABLE V
THE QUALITY ASSESSMENT COMPARISON ON PAVIA UNIVERSITY DATA. THE BEST RESULTS ARE SHOWN IN BOLD.

	Maximal Deformation Magnitude											
	well-registration		1 pixel		2 pixel		3 pixel		4 pixel		5 pixel	
Quality	SAM	PSNR	SAM	PSNR	SAM	PSNR	SAM	PSNR	SAM	PSNR	SAM	PSNR
GSA	3.8896	38.9235	4.4522	36.4441	4.5748	35.8442	4.7615	35.1957	7.3171	27.4971	7.1827	27.7433
FUSE	5.1997	28.8678	6.2619	27.1366	6.5061	26.8179	6.9164	26.4390	9.2621	23.5087	9.2382	23.6174
ICCV15	4.1506	36.7114	4.5927	34.8555	4.5268	34.2199	4.7780	33.7327	7.3236	25.9954	7.2962	26.0470
CNMF	4.3170	36.2256	4.4631	34.7291	4.6751	34.2358	5.1176	33.0868	10.2142	28.1299	12.2259	26.9543
HySure	6.7719	32.7940	8.3750	31.4689	8.7945	31.1361	8.7580	31.0591	10.6249	25.6221	11.1790	25.7066
uSDN	5.6939	36.5644	5.7492	35.4093	5.9451	35.6661	5.5216	36.0068	6.1332	34.8830	5.8769	35.5245
PixAwaRefin	**3.4739**	36.6302	5.4893	32.0982	5.7758	31.9425	6.3066	31.2817	9.0162	28.3988	8.7958	28.5164
CUCaNet	3.8272	**39.7208**	6.9915	26.5037	7.5346	25.9277	8.9666	24.1868	12.0766	20.9322	13.1459	20.4993
u2MDN	5.8040	36.1886	6.6152	32.5035	6.3940	33.9035	8.6036	29.8932	8.0481	31.2687	5.3330	36.4715
Proposed	3.6917	38.6478	**4.3475**	**38.3214**	**4.2960**	**38.5405**	**4.3450**	**38.5491**	**4.3918**	**38.4520**	**4.3676**	**38.2425**

in the shape due to the small original numerical values. Even so, in general the estimated PSFs match the real PSFs very closely.

C. Comparison with other methods

In this part, we first evaluate the performance on nine SR methods by using well-registered input date. Secondly, LSQ free-form non-rigid registration method [6] was used to register the deformed images and the fusion results obtained using different SR methods based on these registered images were compared.

To fully compare the different SR methods, a set of baseline methods were used for comparison: these included traditional methods, e.g., GSA [10], FUSE [22], ICCV^{15} [25], CNMF [5] and HySure [24]; and deep learning methods, e.g., uSDN [14], PixAwaRefin [50], CUCaNet [52] and u2MDN [81]. The parameters of all the compared methods were tuned to their optimal values.

Washington, D.C. data. We first evaluated the visual quality of the fused images obtained using the Washington, D.C. data. Fig.12 shows the errors in the fused imagery obtained using different methods under different deformation conditions. The sizes of the errors are given using the MRAE. The first row of Fig.12 shows the errors obtained under well-registered conditions. It can be seen that most of the compared algorithms produce good fusion results under such conditions – the exceptions are FUSE and HySure, which produce serious ground object boundary errors. Slight material-dependent errors can be seen in the results for ICCV^{15}, CNMF and uSDN.

The second to fourth rows in Fig.12 represent the SR results after registration and fusion processing successively. The deformed images were first registered using the LSQ free-form method [6] and then processed using different SR methods. The proposed method was not included in this as it takes the deformed image as the input and outputs the SR results directly. Due to the limits of space, here, we only show the performance results obtained under three sets of deformation conditions: those with maximum deformations of 1 pixel, 3 pixels and 5 pixels. Compared to the well-registered input, even though the input to the SR method was processed using

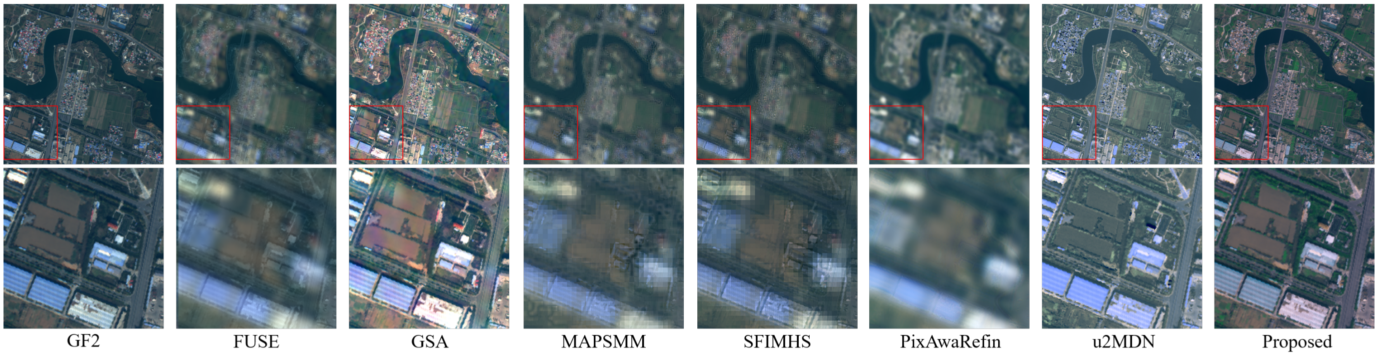

Fig. 15. Reconstruction results of different comparison methods on the GF2 data and GF5 data. The first row shows the reconstruction results of the entire image. The second row shows the local magnified images.

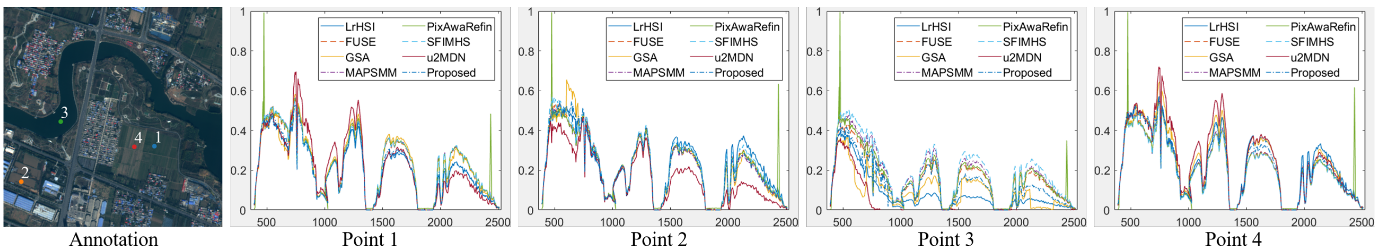

Fig. 16. Reconstructed spectral curve at different points for the comparison method on the GF2 and GF5 data.

the registration method, the fusion results obtained using the different methods are very different. The results for CUCaNet are markedly different because this method uses cross attention to transfer cross-domain information between two stream networks. Since there is still a certain degree of deformation after image registration, the use of cross-domain information will lead to a poor performance even if the deformation is slight. The results obtained using uSDN exhibit material-dependent errors that result in the reconstruction of some of the ground objects being poor. The GSA results gradually become poorer as the deformation magnitude increases. The results obtained using the remaining methods are relatively poor. The u2MDN shows excellent performance under the condition of 5 pixels maximum displacement benefited from adopting mutual information to capture the non-linear statistical dependencies between the input images. But the stability of this method is weak when dealing with different deformations.

Table.III shows the quality measure results obtained using the Washington, D.C. data under different conditions. The results for GSA, CUCaNet, uSDN, u2MDN and the proposed method show a better reconstruction accuracy than the other methods under well-registered conditions; however, an increase in deformation leads to a sudden worsening of the performance of GSA and CUCaNet. Although the results for uSDN are inferior to those obtained using GSA and CUCaNet, the stability of uSDN is better, indicating that uSDN is less sensitive to the deformation. However, from a comparison of all of the methods, it can be seen that the proposed method is the most stable and produces the best performance under a range of deformation conditions.

Chikusei data. The visual quality comparison for the Chikusei data is shown in Fig.13. Compared with the Washington, D.C. data, the reconstruction results obtained using the GSA suffer degraded performance. This indicates that the performance of the GSA is affected by the imaging device used and the area that is imaged. In constrast, the visual reconstruction results obtained using PixAwaRefin and CUCaNet are better than for the Washington, D.C. data. Compared with the other methods, both uSDN and the proposed method produce good, stable reconstruction results under varying deformation conditions. Compared with Washington, D.C. data, the visual results of u2MDN can not achieve stable results in this set of data.

As shown in Table.IV, the results of the quality assessment are similar to those of the visual quality comparison. Under well-registered conditions, compared with Washington D.C. data, the deep learning methods (uSDN, PixAwaRefin, CUCaNet and the proposed method) produce significantly improved results for the Chikusei data. However, it can clearly be seen that the results of PixAwaRefin and CUCaNet have a clear tendency to get worse as the deformation increases. The main reason for this is that the use of uncorrected mutual information results in a poorer performance. In the case of the Washington, D.C. and Chikusei data, the reconstruction accuracy of FUSE and HySure is limited. Overall, the performance of uSDN and the proposed method is good and stable.

Pavia University data. To make a further evaluation of the different methods, we also carried out experiments using the Pavia University data. The results of this are shown in Fig.14 and Table.V. The quantitative results are similar to those for the Washington, D.C. data. Generally, it can be observed that, under well-registered conditions, CUCaNet reconstructs spatial information well and that the overall performance drops as soon as the amount of deformation starts to increase. Of all the methods compared, the results for uSDN and the proposed method are the most stable and accurate. The subspace-based methods (ICCV^{15} and CNMF) produce reasonably good performances but the results are not as good as those obtained using the GSA. These results indicate that the proposed method produces highly accurate and stable results under a variety of deformation conditions.

GF2 data and GF5 data. We further evaluate the proposed method on the GF2 data and GF5 data. It can be seen from Fig.8 that the two input images are not completely aligned due to registration errors caused by the spatial resolution differences. Since some comparison methods cannot generate stable reconstruction results for this data, we only present the results of comparison methods (FUSE [22], GSA [10], MAPSMM [85], SFIMHS [86], PixAwaRefin [50] and u2MDN [81]) as shown in Fig.15 and Fig. 16. Since there is no ground truth, the result of the color-composite image and spectral curve is present for a visual comparison in this experiment. It can be found that GSA, u2MDN and the proposed method have better spatial reconstruction capabilities. But GSA shows edge block error distortion in Fig. 15. This is because GSA sharpens the low-resolution image by directly adding spatial information by calculating the difference between HrMSI and upsampled LrHSI. In comparison, the proposed method shows a better performance for preserving spatial and spectral information.

V. CONCLUSION

In this work, a novel end-to-end unsupervised HSI SR network for reconstructing high-resolution HSI from unregistered low-resolution HSI and high-resolution MSI using multi-modal, multi-task learning was proposed. Specifically, the proposed method integrates non-rigid registration and SR into a unified model that includes a triple convolutional neural network. The use of this network allows the SR and registration to complement each other, which means that better results are obtained for both processes. Furthermore, the proposed network is capable of adaptively learning the spatial and spectral response functions, which enables the model to adapt to a variety of imaging conditions and to produce stable reconstructed images. Extensive experiments conducted on three simulated benchmark datasets and a pair of real data demonstrated the effectiveness of the proposed method and its ability to produce highly accurate and stable reconstructed images under complex non-rigid deformation conditions.

ACKNOWLEDGMENTS

The authors would like to thank Prof. David Landgrebe for providing the Washington DC Mall data, Prof. Paolo Gamba for providing the Pavia University data, and Space Application Laboratory, Department of Advanced Interdisciplinary Studies, the University of Tokyo for providing the Chikusei hyperspectral data. The authors would like to express their appreciation to Dr. Naoto Yokoya for providing MATLAB codes for hyperspectral and multispectral data fusion toolbox.

REFERENCES

[1] D. Hong, W. He, N. Yokoya, J. Yao, L. Gao, L. Zhang, J. Chanussot, and X. Zhu, "Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 2, pp. 52–87, 2021.

[2] X. Wu, D. Hong, J. Chanussot, Y. Xu, R. Tao, and Y. Wang, "Fourier-based rotation-invariant feature boosting: An efficient framework for geospatial object detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 302–306, 2020.

[3] H. Yu, L. Gao, W. Liao, B. Zhang, L. Zhuang, M. Song, and J. Chanussot, "Global spatial and local spectral similarity-based manifold learning group sparse representation for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3043–3056, 2019.

[4] D. Hong, L. Gao, N. Yokoya, J. Yao, J. Chanussot, D. Qian, and B. Zhang, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, 2021.

[5] 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, 2011.

[6] Y. Zhou, A. Rangarajan, and P. D. Gader, "An integrated approach to registration and fusion of hyperspectral and multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, pp. 3020–3033, May 2020.

[7] D. Hong, N. Yokoya, J. Chanussot, and X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. on Image Process.*, vol. 28, no. 4, pp. 1923–1938, 2019.

[8] D. Hong, J. Yao, D. Meng, Z. Xu, and J. Chanussot, "Multimodal gans: Toward crossmodal hyperspectral-multispectral image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5103–5113, 2021.

[9] 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, 2017.

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

[11] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "Mtf-tailored multiscale fusion of high-resolution ms and pan imagery," *PHOTOGRAMM. ENG. REM. S.*, vol. 72, no. 5, pp. 591–596, 2006.

[12] L. Gao, D. Hong, J. Yao, B. Zhang, P. Gamba, and J. Chanussot, "Spectral superresolution of multispectral imagery with joint sparse and low-rank learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2269–2280, 2021.

[13] K. Kotwal and S. Chaudhuri, "A bayesian approach to visualization-oriented hyperspectral image fusion," *Inform. Fusion*, vol. 14, no. 4, pp. 349–360, 2013.

[14] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse dirichlet-net for hyperspectral image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 2511–2520, IEEE, 2018.

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

[16] Y. Zhang and M. He, "Multi-spectral and hyperspectral image fusion using 3-d wavelet transform," *Journal of Electronics (China)*, vol. 24, no. 2, pp. 218–224, 2007.

[17] M. Selva, B. Aiazzi, F. Butera, L. Chiarantini, and S. Baronti, "Hyper-sharpening: A first approach on sim-ga data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3008–3024, 2015.

[18] G. Vivone, L. Alparone, J. Chanussot, M. Dalla Mura, A. Garzelli, G. A. Licciardi, R. Restaino, and L. Wald, "A critical comparison among pansharpening algorithms," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2565–2586, 2014.

[19] Y. Zhang, S. De Backer, and P. Scheunders, "Noise-resistant wavelet-based bayesian fusion of multispectral and hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3834–3843, 2009.

[20] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 3631–3640, 2015.

[21] R. Kawakami, Y. Matsushita, J. Wright, M. Ben-Ezra, Y.-W. Tai, and K. Ikeuchi, "High-resolution hyperspectral imaging via matrix factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 2329–2336, IEEE, 2011.

[22] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Fast fusion of multi-band images based on solving a sylvester equation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4109–4121, 2015.

[23] N. Yokoya, T. Yairi, and A. Iwasaki, "Hyperspectral, multispectral, and panchromatic data fusion based on coupled non-negative matrix factorization," in *Proc. WHISPERS*, pp. 1–4, IEEE, 2011.

[24] 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, 2014.

[25] C. Lanaras, E. Baltsavias, and K. Schindler, "Hyperspectral super-resolution by coupled spectral unmixing," in *Proc. IEEE Int. Conf. Computer Vision*, pp. 3586–3594, IEEE, 2015.

[26] O. Berné, A. Helens, P. Pilleri, and C. Joblin, "Non-negative matrix factorization pansharpening of hyperspectral data: An application to mid-infrared astronomy," in *Proc. WHISPERS*, pp. 1–4, IEEE, 2010.

[27] Y. Chang, L. Yan, X.-L. 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, 2020.

[28] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 5344–5353, 2017.

[29] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4118–4130, 2018.

[30] 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, 2019.

[31] W. He, Y. Chen, N. Yokoya, C. Li, and Q. Zhao, "Hyperspectral super-resolution via coupled tensor ring factorization," *Pattern Recognit.*, vol. 122, p. 108280, 2022.

[32] R. Dian and S. Li, "Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5135–5146, 2019.

[33] R. A. Borsoi, C. Prévost, K. Usevich, D. Brie, J. C. Bermudez, and C. Richard, "Coupled tensor decomposition for hyperspectral and multispectral image fusion with inter-image variability," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 3, pp. 702–717, 2021.

[34] D. Hong, L. Gao, J. Yao, B. Zhang, P. Antonio, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, 2021.

[35] B. Rasti, D. Hong, R. Hang, P. Ghamisi, X. Kang, J. Chanussot, and J. Benediktsson, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, 2020.

[36] D. Hong, Z. Han, J. Yao, L. Gao, B. Zhang, A. Plaza, and J. Chanussot, "Spectralformer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, 2021. DOI: 10.1109/TGRS.2021.3130716.

[37] W. Xie, X. Jia, Y. Li, and J. Lei, "Hyperspectral image super-resolution using deep feature matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6055–6067, 2019.

[38] K. Zheng, L. Gao, Q. Ran, X. Cui, B. Zhang, W. Liao, and S. Jia, "Separable-spectral convolution and inception network for hyperspectral image super-resolution," *Int. J. Mach. Learn. Cyb.*, vol. 10, no. 10, pp. 2593–2607, 2019.

[39] P. Arun, K. M. Buddhiraju, A. Porwal, and J. Chanussot, "Cnn-based super-resolution of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 9, pp. 6106–6121, 2020.

[40] J. Hu, X. Jia, Y. Li, G. He, and M. Zhao, "Hyperspectral image super-resolution via intrafusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7459–7471, 2020.

[41] 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.

[42] Q. Wang, Q. Li, and X. Li, "Hyperspectral image super-resolution using spectrum and feature context," *IEEE Trans. Comput. Imag.*, 2020.

[43] D. Liu, J. Li, and Q. Yuan, "A spectral grouping and attention-driven residual dense network for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, 2021.

[44] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5345–5355, 2018.

[45] X.-H. Han, Y. Zheng, and Y.-W. Chen, "Multi-level and multi-scale spatial and spectral fusion cnn for hyperspectral image super-resolution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pp. 0–0, 2019.

[46] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu, "Multispectral and hyperspectral image fusion by ms/hs fusion net," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 1585–1594, 2019.

[47] T. Zhang, Y. Fu, L. Wang, and H. Huang, "Hyperspectral image reconstruction using deep external and internal learning," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pp. 8559–8568, 2019.

[48] L. Zhang, J. Nie, W. Wei, Y. Li, and Y. Zhang, "Deep blind hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, 2020.

[49] 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.

[50] W. Wei, J. Nie, L. Zhang, and Y. Zhang, "Unsupervised recurrent hyperspectral imagery super-resolution using pixel-aware refinement," *IEEE Trans. Geosci. Remote Sens.*, p. 1–15, 2020.

[51] Z. Zhu, J. Hou, J. Chen, H. Zeng, and J. Zhou, "Hyperspectral image super-resolution via deep progressive zero-centric residual learning," *IEEE Trans. Image Process.*, 2020.

[52] 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. European Conf. Computer Vision (ECCV)*, pp. 208–224, 2020.

[53] K. Zheng, L. Gao, W. Liao, D. Hong, B. Zhang, X. Cui, and J. Chanussot, "Coupled convolutional neural network with adaptive response function learning for unsupervised hyperspectral super resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, pp. 2487—-2502, Mar 2021.

[54] W. Wang, W. Zeng, Y. Huang, X. Ding, and J. Paisley, "Deep blind hyperspectral image fusion," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pp. 4150–4159, 2019.

[55] L. Zhang, J. Nie, W. Wei, Y. Zhang, S. Liao, and L. Shao, "Unsupervised adaptation learning for hyperspectral imagery super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 3073–3082, 2020.

[56] W. Wei, J. Nie, L. Zhang, and Y. Zhang, "Unsupervised recurrent hyperspectral imagery super-resolution using pixel-aware refinement," *IEEE Trans. Geosci. Remote Sens.*, 2020.

[57] Y. Bentoutou, N. Taleb, K. Kpalma, and J. Ronsin, "An automatic image registration for applications in remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 9, pp. 2127–2137, 2005.

[58] H.-M. Chen, M. K. Arora, and P. K. Varshney, "Mutual information-based image registration for remote sensing data," *Int. J. Remote Sens.*, vol. 24, no. 18, pp. 3701–3706, 2003.

[59] Y. Ye, J. Shan, L. Bruzzone, and L. Shen, "Robust registration of multimodal remote sensing images based on structural similarity," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2941–2958, 2017.

[60] J. Ma, J. Jiang, H. Zhou, J. Zhao, and X. Guo, "Guided locality preserving feature matching for remote sensing image registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4435–4447, 2018.

[61] S. Dawn, V. Saxena, and B. Sharma, "Remote sensing image registration techniques: A survey," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, pp. 103–112, Springer, 2010.

[62] G. Haskins, U. Kruger, and P. Yan, "Deep learning in medical image registration: a survey," *Mach. Vision Appl.*, vol. 31, no. 1, pp. 1–18, 2020.

[63] S. Zhao, T. Lau, J. Luo, I. Eric, C. Chang, and Y. Xu, "Unsupervised 3d end-to-end medical image registration with volume tweening network," *IEEE J. Biomed. Health*, vol. 24, no. 5, pp. 1394–1404, 2019.

[64] X. Cao, J. Yang, J. Zhang, D. Nie, M. Kim, Q. Wang, and D. Shen, "Deformable image registration based on similarity-steered cnn regression," in *Proc. MICCAI*, pp. 300–308, Springer, 2017.

[65] J. Krebs, T. Mansi, H. Delingette, L. Zhang, F. C. Ghesu, S. Miao, A. K. Maier, N. Ayache, R. Liao, and A. Kamen, "Robust non-rigid registration through agent-based action learning," in *Proc. MICCAI*, pp. 344–352, Springer, 2017.

[66] H. Sokooti, B. De Vos, F. Berendsen, B. P. Lelieveldt, I. Išgum, and M. Staring, "Nonrigid image registration using multi-scale 3d convolutional neural networks," in *Proc. MICCAI*, pp. 232–239, Springer, 2017.

[67] N. J. Tustison, B. B. Avants, and J. C. Gee, "Learning image-based spatial transformations via convolutional neural networks: A review," *Magn. Reson. Imaging*, vol. 64, pp. 142–153, 2019.

[68] S. Zhao, Y. Dong, E. I. Chang, Y. Xu, *et al.*, "Recursive cascaded networks for unsupervised medical image registration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pp. 10600–10610, 2019.

[69] M. Jaderberg, K. Simonyan, A. Zisserman, *et al.*, "Spatial transformer networks," in *Proc. Conf. Neural Information Processing Systems (NeurIPS)*, pp. 2017–2025, 2015.

[70] H. Li and Y. Fan, "Non-rigid image registration using self-supervised fully convolutional networks without training data," in *Proc. ISBI*, pp. 1075–1078, IEEE, 2018.

[71] B. D. de Vos, F. F. Berendsen, M. A. Viergever, M. Staring, and I. Išgum, "End-to-end unsupervised deformable image registration with a convolutional neural network," in *Proc. MICCAI*, pp. 204–212, Springer, 2017.

[72] C. Shu, X. Chen, Q. Xie, and H. Han, "An unsupervised network for fast microscopic image registration," in *Progr. Biomed. Opt. Imaging Proc. SPIE*, vol. 10581, p. 105811D, International Society for Optics and Photonics, 2018.

[73] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "An unsupervised learning model for deformable medical image registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 9252–9260, 2018.

[74] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "Voxelmorph: a learning framework for deformable medical image registration," *IEEE Trans. Med. Imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[75] N. Yokoya, N. Mayumi, and A. Iwasaki, "Cross-calibration for data fusion of eo-1/hyperion and terra/aster," *IEEE J. Sel.Topics Appl. Earth Observ. Remote Sens.*, vol. 6, pp. 419–426, Apr 2013.

[76] Y. Lin, Y. Zheng, Y. Fu, and H. Huang, "Hyperspectral image super-resolution under misaligned hybrid camera system," *IET Image Process.*, vol. 12, no. 10, pp. 1824–1831, 2018.

[77] Y. Fu, Y. Zheng, L. Zhang, Y. Zheng, and H. Huang, "Simultaneous hyperspectral image super-resolution and geometric alignment with a hybrid camera system," *Neurocomputing*, Dec 2019.

[78] J. Nie, L. Zhang, W. Wei, C. Ding, and Y. Zhang, "Unsupervised deep hyperspectral super-resolution with unregistered images," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, pp. 1–6, Jul 2020.

[79] Y. Zhou, A. Rangarajan, and P. D. Gader, "Nonrigid registration of hyperspectral and color images with vastly different spatial and spectral resolutions for spectral unmixing and pansharpening," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 86–94, 2017.

[80] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. X. Zhu, "Joint and progressive subspace analysis (jpsa) with spatial-spectral manifold alignment for semisupervised hyperspectral dimensionality reduction," *IEEE Trans. Cybern.*, vol. 51, no. 7, pp. 3602–3615, 2021.

[81] 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.*, pp. 1–18, 2021.

[82] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, pp. 234–241, Springer, 2015.

[83] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Conf. Neural Information Processing Systems (NeurIPS)*, pp. 8024–8035, 2019.

[84] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2014.

[85] M. T. Eismann and R. C. Hardie, "Hyperspectral resolution enhancement using high-resolution multispectral imagery with arbitrary response functions," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 455–465, 2005.

[86] J. Liu, "Smoothing filter-based intensity modulation: A spectral preserve image fusion technique for improving spatial details," *Int. J. Remote Sens.*, vol. 21, no. 18, pp. 3461–3472, 2000.

Ke Zheng received B.S. degree in geographic information system from Shandong Agricultural University, Taian, China, in 2012, and the M.S. and Ph.D. degrees in remote sensing from College of Geosciences and Surveying Engineering, China University of Mining and Technology (Beijing), Beijing, China, in 2016 and 2020, respectively.

Since 2020, he has been a Post-Doctoral Associate with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, Chinese Academy of Science, Beijing, China. His research interests include image processing, machine learning, deep learning and their application in Earth Vision.

Lianru Gao (M'12–SM'18) received the B.S. degree in civil engineering from Tsinghua University, Beijing, China, in 2002, the Ph.D. degree in cartography and geographic information system from Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, China, in 2007.

He is currently a Professor with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute, CAS. He also has been a visiting scholar at the University of Extremadura, Cáceres, Spain, in 2014, and at the Mississippi State University (MSU), Starkville, USA, in 2016. His research focuses on hyperspectral image processing and information extraction. In last ten years, he was the PI of 10 scientific research projects at national and ministerial levels, including projects by the National Natural Science Foundation of China (2016-2019, 2018-2020, 2022-2025), and by the Key Research Program of the CAS (2013-2015) et al. He has published more than 180 peer-reviewed papers, and there are more than 100 journal papers included by Science Citation Index (SCI). He was coauthor of 3 academic books including "Hyperspectral Image Information Extraction" et al. He obtained 29 National Invention Patents in China. He was awarded the Outstanding Science and Technology Achievement Prize of the CAS in 2016, and was supported by the China National Science Fund for Excellent Young Scholars in 2017, and won the Second Prize of The State Scientific and Technological Progress Award in 2018. He received the recognition of the Best Reviewers of the IEEE JSTARS in 2015, and the Best Reviewers of the IEEE TGRS in 2017.

Danfeng Hong (S'16–M'19–SM'21) received the M.Sc. degree (summa cum laude) in computer vision from the College of Information Engineering, Qingdao University, Qingdao, China, in 2015, the Dr. -Ing degree (summa cum laude) from the Signal Processing in Earth Observation (SiPEO), Technical University of Munich (TUM), Munich, Germany, in 2019.

From 2015 to 2019, he was a Research Associate at the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germany. Since 2019, He has been a Research Scientist and led a Spectral Vision Group at IMF, DLR. He is also an Adjunct Scientist at GIPSA-lab, Grenoble INP, CNRS, Univ. Grenoble Alpes, Grenoble, France, from 2020. He is currently with the Key Laboratory of Digital Earth Science, Aerospace Information Research Institute (AIR), Chinese Academy of Sciences (CAS). His research interests include signal / image processing and analysis, hyperspectral remote sensing, machine / deep learning, artificial intelligence, and their applications in Earth Vision.

Dr. Hong is an Editorial Board Member of Remote Sensing and a Topical Associate Editor of the IEEE Transactions on Geoscience and Remote Sensing (TGRS). He was a recipient of the Best Reviewer Award of the IEEE TGRS in 2021 and the Jose Bioucas Dias award for recognizing the outstanding paper at the Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS) in 2021. He is also a Leading Guest Editor of the International Journal of Applied Earth Observation and Geoinformation, the IEEE Journal of Selected Topics in Applied Earth Observations, and Remote Sensing.

Bing Zhang (M'11–SM'12–F'19) received the B.S. degree in geography from Peking University, Beijing, China, in 1991, and the M.S. and Ph.D. degrees in remote sensing from the Institute of Remote Sensing Applications, Chinese Academy of Sciences (CAS), Beijing, China, in 1994 and 2003, respectively.

Currently, he is a Full Professor and the Deputy Director of the Aerospace Information Research Institute, CAS, where he has been leading lots of key scientific projects in the area of hyperspectral remote sensing for more than 25 years. His research interests include the development of Mathematical and Physical models and image processing software for the analysis of hyperspectral remote sensing data in many different areas. He has developed 5 software systems in the image processing and applications. His creative achievements were rewarded 10 important prizes from Chinese government, and special government allowances of the Chinese State Council. He was awarded the National Science Foundation for Distinguished Young Scholars of China in 2013, and was awarded the 2016 Outstanding Science and Technology Achievement Prize of the Chinese Academy of Sciences, the highest level of Awards for the CAS scholars.

Dr. Zhang has authored more than 300 publications, including more than 170 journal papers. He has edited 6 books/contributed book chapters on hyperspectral image processing and subsequent applications. He is the IEEE fellow and currently serving as the Associate Editor for IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. He has been serving as Technical Committee Member of IEEE Workshop on Hyperspectral Image and Signal Processing since 2011, and as the president of hyperspectral remote sensing committee of China National Committee of International Society for Digital Earth since 2012, and as the Standing Director of Chinese Society of Space Research (CSSR) since 2016. He is the Student Paper Competition Committee member in IGARSS from 2015-2019.

Jocelyn Chanussot (M'04–SM'04–F'12) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree from the Université de Savoie, Annecy, France, in 1998. Since 1999, he has been with Grenoble INP, where he is currently a Professor of signal and image processing. His research interests include image analysis, hyperspectral remote sensing, data fusion, machine learning and artificial intelligence. He has been a visiting scholar at Stanford University (USA), KTH (Sweden) and NUS (Singapore). Since 2013, he is an Adjunct Professor of the University of Iceland. In 2015-2017, he was a visiting professor at the University of California, Los Angeles (UCLA). He holds the AXA chair in remote sensing and is an Adjunct professor at the Chinese Academy of Sciences, Aerospace Information research Institute, Beijing.

Dr. Chanussot is the founding President of IEEE Geoscience and Remote Sensing French chapter (2007-2010) which received the 2010 IEEE GRS-S Chapter Excellence Award. He has received multiple outstanding paper awards. He was the Vice-President of the IEEE Geoscience and Remote Sensing Society, in charge of meetings and symposia (2017-2019). He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote sensing (WHISPERS). He was the Chair (2009-2011) and Cochair of the GRS Data Fusion Technical Committee (2005-2008). He was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society (2006-2008) and the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing (2009). He is an Associate Editor for the IEEE Transactions on Geoscience and Remote Sensing, the IEEE Transactions on Image Processing and the Proceedings of the IEEE. He was the Editor-in-Chief of the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2011-2015). In 2014 he served as a Guest Editor for the IEEE Signal Processing Magazine. He is a Fellow of the IEEE, a member of the Institut Universitaire de France (2012-2017) and a Highly Cited Researcher (Clarivate Analytics/Thomson Reuters).