HYPERSPECTRAL IMAGE SUPER-RESOLUTION WITH DEEP PRIORS AND DEGRADATION MODEL INVERSION

Xiuheng Wang[*], Jie Chen[†], Cédric Richard[*]

[*] Université Côte d'Azur, CNRS, OCA, France

[†] School of Marine Science and Technology, Northwestern Polytechnical University, China

xiuheng.wang@oca.eu, dr.jie.chen@ieee.org, cedric.richard@unice.fr

ABSTRACT

To overcome inherent hardware limitations of hyperspectral imaging systems with respect to their spatial resolution, fusion-based hyperspectral image (HSI) super-resolution is attracting increasing attention. This technique aims to fuse a low-resolution (LR) HSI and a conventional high-resolution (HR) RGB image in order to obtain an HR HSI. Recently, deep learning architectures have been used to address the HSI super-resolution problem and have achieved remarkable performance. However, they ignore the degradation model even though this model has a clear physical interpretation and may contribute to improve the performance. We address this problem by proposing a method that, on the one hand, makes use of the linear degradation model in the data-fidelity term of the objective function and, on the other hand, utilizes the output of a convolutional neural network for designing a deep prior regularizer in spectral and spatial gradient domains. Experiments show the performance improvement achieved with this strategy.

Index Terms— Super-resolution, hyperspectral imaging, optimization, spectral-spatial gradient domain, deep learning

1. INTRODUCTION

Hyperspectral imaging systems collect images of a scene using contiguous spectral bands ranging from ultraviolet to visible and infrared. Hyperspectral imaging is beneficial in applications as diverse as remote surveillance, medicine, and environmental monitoring. Unfortunately, many factors such as noise and low sensing resolution can cause image degradation during the acquisition process. Image restoration methods are often employed before further processing. Recently, it has become popular to restore an HR HSI by fusing an LR HSI and an HR RGB image. This framework, often called *fusion-based HSI super-resolution* [1], raises the challenge of accounting correlation between spectral bands while ensuring spatial consistency [2].

In the scope of this work, we consider the problem of constructing an HR HSI $\mathcal{X} \in \mathbb{R}^{B \times L \times W}$ via the fusion of an LR HSI $\mathcal{Y} \in \mathbb{R}^{B \times l \times w}$ and an HR RGB image $\mathcal{Z} \in \mathbb{R}^{b \times L \times W}$, where B and b are the numbers of spectral channels of the hyperspectral image and the RGB image, respectively, with $B > b$, and $(L \times W)$ and $(l \times w)$ denote the dimensions of the HR image and the LR image, respectively, with $L > l$ and $W > w$. Images $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ can be reshaped in matrix forms $\mathbf{X} \in \mathbb{R}^{B \times N}$, $\mathbf{Y} \in \mathbb{R}^{B \times n}$ and $\mathbf{Z} \in \mathbb{R}^{b \times N}$, respectively, where $N = L \times W$ and $n = l \times w$ are the numbers of pixels in each band of the HR and LR images. Considering a linear degradation model, $\mathbf{Y}$ and $\mathbf{Z}$ can be interpreted as down-sampled versions
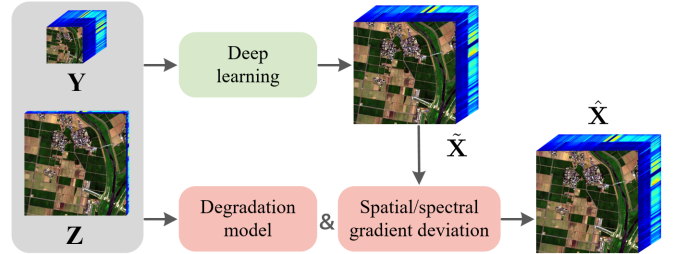

Fig. 1. HSI super-resolution with deep priors and model degradation inversion accounting for spatial/spectral gradient deviation of HSIs.

of $\mathbf{X} \in \mathbb{R}^{B \times N}$ in the spatial and spectral domains, respectively:

$$\mathbf{Y} = \mathbf{XBS}, \quad \mathbf{Z} = \mathbf{RX}, \qquad (1)$$

where $\mathbf{B} \in \mathbb{R}^{N \times N}$ is the spatial blurring matrix, $\mathbf{S} \in \mathbb{R}^{N \times n}$ is a downsampling operator with factor $s = L/l$, and $\mathbf{R} \in \mathbb{R}^{b \times B}$ denotes the spectral response function (SRF) of the RGB camera. The HSI super-resolution problem can be formulated as the estimation of $\mathbf{X}$ given the observed data $\mathbf{Y}$ and $\mathbf{Z}$, with known system responses $\mathbf{B}$, $\mathbf{S}$ and $\mathbf{R}$. Based on (1), this problem can be written as the minimization of an unconstrained objective function of the form:

$$J(\mathbf{X}) = \|\mathbf{Y} - \mathbf{XBS}\|_F^2 + \|\mathbf{Z} - \mathbf{RX}\|_F^2 + \varphi(\mathbf{X}) \qquad (2)$$

where $\varphi(\cdot)$ is some regularization function. Reconstructing $\mathbf{X}$ from $\mathbf{Y}$ and $\mathbf{Z}$ by minimizing (2) without $\varphi(\cdot)$ is a highly ill-posed problem. This justifies the use of $\varphi(\cdot)$ to constrain the solution space by promoting prior information on $\mathbf{X}$.

Under the classic optimization framework, handcrafted forms of $\varphi(\cdot)$, which promote the sparsity, spatial continuity and edge preserving, have been intensively studied for this problem [1, 3–6], and also for other highly related problems [7, 8]. However, designing a powerful $\varphi(\cdot)$ is not trivial and may also cause difficulty in finding optimal solutions. Inspired by the success of deep learning in computer vision, convolutional neural networks (CNN) have been used to get $\mathbf{X}$ from the fusion of $\mathbf{Y}$ and $\mathbf{Z}$ [9–12]. Deep learning methods require less handcrafted prior information on $\mathbf{X}$ and have been shown to achieve significant performance enhancement compared to model-based methods such as (2). However, they need massive data for training and may not be consistent with the physical degradation model of form (1) involving $\mathbf{Y}$ and $\mathbf{Z}$. A brief review of these methods is given in Section 2.

To leverage the merits of both model-based and deep learning methods, recent approaches have started to plug the output of a CNN, denoted as $\tilde{\mathbf{X}}$, into the objective function (2) as a *deep prior regularizer* [13–15]. Specially, the Frobenius norm $\varphi(\mathbf{X}) = \|\mathbf{X} - \tilde{\mathbf{X}}\|_F^2$ is

considered in [13, 14]. It allows the use of a fast off-the-shelf solver. The 2D Total Variation (TV) norm $\varphi(\mathbf{X}) = \|\mathbf{X} - \tilde{\mathbf{X}}\|_{TV}^2 + \|\mathbf{X}\|_{TV}$ is used in [15]. It is applied to each band independently from the others to enforce smoothness in the spatial domain. Nevertheless, none of these methods simultaneously exploits the spectral-spatial gradient information for enhancing fusion process.

The aim of this paper is to introduce a novel strategy for HSI super-resolution that, on the one hand, makes use of the physical linear degradation model in the data-fidelity term of the objective function and, on the other hand, exploits the spectral-spatial gradient difference of HSIs using a deep prior regularizer from the output of a convolutional neural network. Experimental results show the performance improvement achieved with this strategy.

Notation: $\mathcal{X}$, $\mathbf{X}$ and $\mathbf{x}$ refer to the same 3D image $\mathcal{X}$: matrix $\mathbf{X}$ is obtained by arranging the pixel column vectors in $\mathcal{X}$ next to each other; vector $\mathbf{x}$ is obtained by stacking the columns of $\mathbf{X}$ on top of each other. This notation system also works for other images.

2. RELATED WORKS

In this section, we provide a brief overview on some existing methods to facilitate the presentation of our method.

2.1. Model-based methods

Solving HSI super-resolution and unmixing in a joint framework has been demonstrated to significantly improve the super-resolution performance. In [1], the authors perform a joint unmixing of $\mathbf{Y}$ and $\mathbf{Z}$ that allows them to split the initial optimization problem into two constrained quadratic sub-problems, which can be solved efficiently. In [3], the authors reconstruct the latent $\mathbf{X}$ by using a variable splitting technique w.r.t. its endmembers and their corresponding abundances. Considering the similarity between neighboring pixels, the authors in [4] enforce group-sparsity and non-negativity properties in small image cubes. In [5], the authors introduce a non-negative sparse coding method to exploit the sparsity of pixels and non-local spatial similarity of the latent $\mathbf{X}$. Another strategy in HSI super-resolution consists of tensor-based factorization. For instance, a graph Laplacian-guided coupled tensor decomposition model is proposed by the authors in [6] to exploit the spatial-spectral information of HSI and RGB images.

2.2. Deep learning methods

Recently, deep learning has been proved to be an effective data-driven technique for solving HSI super-resolution problem. In [9], the authors design a CNN with 3D convolution to fuse $\mathbf{Y}$ and $\mathbf{Z}$ and obtain $\mathbf{X}$ in an end-to-end manner. The PanNet architecture is proposed in [10] to preserve spectral and spatial information in HSI super-resolution problem. In [11], the authors unfold an iterative algorithm into a deep network called MHF-net. This algorithm is based on a novel LR/HR fusion model which takes the degradation models of $\mathbf{Y}$ and $\mathbf{Z}$ as well as the low-rankness of $\mathbf{X}$ into consideration. In [12], a two-stage network based on unsupervised adaption learning is proposed to learn priors of $\mathbf{X}$ while estimating the unknown spatial degradation.

3. THE PROPOSED METHOD

Considering the HSI super-resolution problem defined by the unconstrained objective function (2), we propose to jointly consider the super-resolution from the physical model, i.e., the first two terms in (2), and the data-driven prior information denoted by $\tilde{\mathbf{X}}$. In this work, a prior image from the output of an CNN is used to enhance the

physical model result via $\varphi(\mathbf{X})$ by reducing the difference between the optimization variable $\mathbf{X}$ and $\tilde{\mathbf{X}}$ in spectral and spatial gradient domains respectively. Under this design, the objective function (2) becomes:

$$J(\mathbf{X}) = \|\mathbf{Y} - \mathbf{XBS}\|_F^2 + \|\mathbf{Z} - \mathbf{RX}\|_F^2 + \varphi(\mathbf{X})$$

$$\text{with } \varphi(\mathbf{X}) = \mu\|\mathbf{D}(\mathbf{x} - \tilde{\mathbf{x}})\|^2 + \nu\|\mathbf{E}(\mathbf{x} - \tilde{\mathbf{x}})\|^2 \tag{3}$$

$$\text{and} \qquad \tilde{\mathbf{X}} = \text{CNN}(\mathbf{Y}, \mathbf{Z})$$

where $\mathbf{x} \in \mathbb{R}^{BN}$ and $\tilde{\mathbf{x}} \in \mathbb{R}^{BN}$ denote the vectors obtained by stacking the columns of the matrices $\mathbf{X}$ and $\tilde{\mathbf{X}}$, respectively and CNN denotes a trained CNN with the inputs $\mathbf{Y}$ and $\mathbf{Z}$ to produce $\tilde{\mathbf{X}}$. The first two terms of $J(\mathbf{X})$ guarantee that the candidate solution is consistent with the degradation model (1). The last term of $J(\mathbf{X})$ are regularization terms used to promote prescribed properties, with positive hyper-parameters μ and ν. These two properties consist of smooth error maps along the spatial and spectral dimensions, obtained with matrices $\mathbf{D}$ and $\mathbf{E}$ defined as follows.

Matrix $\mathbf{D}$ can be designed by choosing first a convolution kernel $\mathcal{D}$. We consider the Laplacian filter $\mathcal{D}_\ell$ for each channel ℓ:

$$\begin{pmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{pmatrix}. \tag{4}$$

We construct an $N \times N$ block-Toeplitz matrix $\mathbf{D}_\ell$ with N Toeplitz blocks; see [2] for details. Imposing periodic boundary conditions on $\mathbf{D}_\ell$, it can be reformulated as a block circulant matrix with circulant blocks, a structure denoted as *circulant-block-circulant* (CBC). This property allows to diagonalize $\mathbf{D}_\ell$ with 2D Fourier transforms. This leads to matrix $\mathbf{D}$ in (3) with block-diagonal structure:

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{D}_B \end{pmatrix}. \tag{5}$$

Matrix $\mathbf{E}$ cannot be block-diagonal since it operates across channels. A typical choice is to penalize the variation between two adjacent channels with a first-order derivative filter $\mathcal{E}_0 = [1 \ -1]$ along the spectral dimension. The convolution matrix, of size $(B-1) \times B$, is then given by:

$$\mathbf{E}_0 = \begin{pmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -1 & 1 \end{pmatrix}. \tag{6}$$

This yields:

$$\mathbf{E} = \mathbf{E}_0 \otimes \mathbf{I}_N \tag{7}$$

where $\otimes$ denotes the Kronecker product of two matrices and $\mathbf{I}_N$ is the $N \times N$ identity matrix. Although $\mathbf{D}$ and $\mathbf{E}$ are large matrices of size $BN \times BN$, they are not explicitly stored or used in practice. As can be seen in Subsection 4.3, $\|\mathbf{D}(\mathbf{x} - \tilde{\mathbf{x}})\|^2$ and $\|\mathbf{E}(\mathbf{x} - \tilde{\mathbf{x}})\|^2$ are actually computed in an efficient manner with Fourier transforms.

4. NUMERICAL OPTIMIZATION

We shall now introduce the algorithm to estimate $\mathbf{X}$ by minimizing the objective function (3). First, we use a variable splitting technique, i.e., the Half Quadratic Splitting (HQS) [16] algorithm, to decompose the optimization problem into iterative sub-problems. One

Algorithm 1 Solving the Sylvester equation w.r.t. $\mathbf{X}_{k+1}$

Input: $\mathbf{Y}, \mathbf{Z}, \mathbf{B}, \mathbf{S}, \mathbf{R}, \mathbf{V}_k, \rho$.
Output: $\mathbf{X}_{k+1}$.
  Initialize $\mathbf{C_1}, \mathbf{C_2}, \mathbf{C_3}$ via (13)
  Compute the eigen-decomposition of $\mathbf{B}$ as $\mathbf{FDF}^H$
  $\overline{\mathbf{D}} = \mathbf{D}(\mathbf{1}_s \otimes \mathbf{I}_n)$
  Compute the eigen-decomposition of $\mathbf{C_1}$ as $\mathbf{C_1} = \mathbf{Q\Lambda Q}^{-1}$
  $\overline{\mathbf{C_3}} = \mathbf{Q}^{-1}\mathbf{C_3F}$
  Compute auxiliary matrix $\overline{\mathbf{X}}$ channel by channel
  for $\ell = 1$ to B **do**
    $\overline{\mathbf{X}}_\ell = \lambda_\ell^{-1}(\overline{\mathbf{C_3}})_i - \lambda_\ell^{-1}(\overline{\mathbf{C_3}})_k \overline{\mathbf{D}}(\lambda_\ell s\mathbf{I}_n + \sum_{t=1}^{s} \mathbf{D}_t^2)\overline{\mathbf{D}}^H$
  end for
  $\mathbf{X}_{k+1} = \mathbf{Q}\overline{\mathbf{X}}\mathbf{F}^H$

Algorithm 2 HSI super-resolution accounting for spectral-spatial gradient deviation

Input: $\mathbf{Y}, \mathbf{Z}, \mathbf{B}, \mathbf{S}, \mathbf{R}, \tilde{\mathbf{X}}, \mu, \nu, \rho$
Output: $\hat{\mathbf{X}}$
  Initialize $\mathbf{V}_0 = \tilde{\mathbf{X}}, k = 0, \mu' = \mu/\rho, \nu' = \nu/\rho$
  while stopping criterion is not met and $k < K$ **do**
    Update $\mathbf{X}_{k+1}$ via Algorithm 1
    For $\ell = 1, \ldots, B$, compute 2D DFTs of data $\mathcal{X}_{k+1,\ell}$ and $\tilde{\mathcal{X}}_\ell$
    For each $\mathbf{f}$, compute $\underline{\mathcal{V}}_\ell(\mathbf{f})$ using (18) and (19)
    For $\ell = 1, \ldots, B$, compute $\mathbf{v}_{k+1,\ell}$ from $\underline{\mathcal{V}}_\ell$ via inverse 2D DFTs
    $k = k + 1$
  end while
  $\hat{\mathbf{X}} = \mathbf{X}_K$

---

sub-problem related to data fidelity terms is solved based on a fast Sylvester analytical solver. Another sub-problem involving regularizers is solved efficiently using 2D discrete Fourier transform.

4.1. Variable splitting based on HQS

HQS is employed to decouple the data fidelity terms and the regularization terms in (3). By introducing an auxiliary variable $\mathbf{V} \in \mathbb{R}^{B \times N}$, minimization of (3) can be reformulated as:

$$\hat{\mathbf{X}} = \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{XBS}\|_F^2 + \|\mathbf{Z} - \mathbf{RX}\|_F^2 + \mu\|\mathbf{D}(\mathbf{v} - \tilde{\mathbf{x}})\|^2$$
$$+ \nu\|\mathbf{E}(\mathbf{v} - \tilde{\mathbf{x}})\|^2 \qquad \text{s.t.} \qquad \mathbf{V} = \mathbf{X}. \tag{8}$$

where $\mathbf{v} \in \mathbb{R}^{BN}$ denote the vector obtained by stacking the columns of the matrice $\mathbf{V}$. The augmented Lagrangian function is given by:

$$\mathcal{L}_\rho(\mathbf{X}, \mathbf{V}) = \|\mathbf{Y} - \mathbf{XBS}\|_F^2 + \|\mathbf{Z} - \mathbf{RX}\|_F^2 + \rho\|\mathbf{X} - \mathbf{V}\|_F^2$$
$$+ \nu\|\mathbf{E}(\mathbf{v} - \tilde{\mathbf{x}})\|^2 + \mu\|\mathbf{D}(\mathbf{v} - \tilde{\mathbf{x}})\|^2 \tag{9}$$

where ρ is a positive penalty parameter. HQS method then minimizes (9) via the following steps:

$$\mathbf{X}_{k+1} = \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{XBS}\|_F^2 + \|\mathbf{Z} - \mathbf{RX}\|_F^2 + \rho\|\mathbf{X} - \mathbf{V}_k\|_F^2 \tag{10}$$

$$\mathbf{v}_{k+1} = \min_{\mathbf{v}} \rho\|\mathbf{x}_{k+1} - \mathbf{v}\|^2 + \mu\|\mathbf{D}(\mathbf{v} - \tilde{\mathbf{x}})\|^2 + \nu\|\mathbf{E}(\mathbf{v} - \tilde{\mathbf{x}})\|^2 \tag{11}$$

We can observe that the fidelity terms and the regularization terms are decoupled in the two sub-problems (10) and (11). Now we can address the above minimization problems in an efficient manner by iteratively minimizing with respect to $\mathbf{X}$ and $\mathbf{v}$, independently. These two steps are as follows.

4.2. Optimization w.r.t. X

Let us assume that the blurring matrix $\mathbf{B}$ has a CBC structure. Under this widely accepted assumption, matrix $\mathbf{B}$ can be decomposed as follows: $\mathbf{B} = \mathbf{FDF}^H$ with $\mathbf{F} \in \mathbb{R}^{N \times N}$ the DFT matrix, $\mathbf{D} \in \mathbb{R}^{N \times N}$ a diagonal matrix, and ^{H} the conjugate transpose.

To solve sub-problem (10), we set the gradient of the objective function (10) w.r.t. $\mathbf{X}$ to zero. Thus, $\mathbf{X}_{k+1}$ is the solution of the Sylvester equation:

$$\mathbf{C_1}\mathbf{X}_{k+1} + \mathbf{X}_{k+1}\mathbf{C_2} = \mathbf{C_3} \tag{12}$$

where

$$\mathbf{C_1} = \mathbf{R}^T\mathbf{R} + \mu\mathbf{I}_B$$
$$\mathbf{C_2} = (\mathbf{BS})(\mathbf{BS})^T \tag{13}$$
$$\mathbf{C_3} = \mathbf{R}^T\mathbf{Z} + \mathbf{Y}(\mathbf{BS})^T + \rho\mathbf{V}_k$$

and $\mathbf{I}_B$ is the identity matrix of size $B \times B$.

According to the conclusion in [17], the Sylvester equation in (12) has a unique solution when an arbitrary sum of the eigenvalues of $\mathbf{C_1}$ and $\mathbf{C_2}$ is not equal to zero. Matrix $\mathbf{C_1}$ is positive-definite since $\mathbf{R}^T\mathbf{R}$ and $\mathbf{I}_B$ are both positive-definite matrices, and $\mathbf{C_2}$ is positive semi-definite. Thus, any sum of the eigenvalues of $\mathbf{C_1}$ and $\mathbf{C_2}$ is larger than zero, ensuring the uniqueness of the solution of (12). For a fast algorithm solving (12), the interested reader can refer to [18]. The steps are summarized in Algorithm 1.

4.3. Optimization w.r.t. v

Following [2] and recalling the notation system $(\mathcal{X}, \mathbf{X}, \mathbf{x})$ for other images as described at the end of Section 1, we can rewrite the objective function in (11) in 3D image domain with a sum running over spectral channels:

$$\min_{\mathcal{V}} \sum_{\ell=1}^{B} \left(\rho\|\mathcal{X}_{k+1,\ell} - \mathcal{V}_\ell\|_F^2 + \mu\|\mathcal{D}_\ell *_{2D} (\mathcal{V}_\ell - \tilde{\mathcal{X}}_\ell)\|_F^2 \right.$$
$$\left. + \nu\|[\mathcal{E}_0 *_{1D}(\mathcal{V} - \tilde{\mathcal{X}})]_\ell\|_F^2 \right) \tag{14}$$

The operator $*_{2D}$ denotes 2D convolution in the spatial domain while $*_{1D}$ represents 1D convolution across spectral channels. Using $L \times W$ 2D DFT in the spatial domain, and denoting the Fourier transformed quantities with underlined symbols of the spatial frequency variable $\mathbf{f}$, Plancherel theorem allows to rewrite (14) as:

$$\min_{\underline{\mathcal{V}}} \sum_{\ell=1}^{B} \left(\rho\|\underline{\mathcal{X}}_{k+1,\ell} - \underline{\mathcal{V}}_\ell\|_F^2 + \mu\|\underline{\mathcal{D}}_\ell \odot (\underline{\mathcal{V}}_\ell - \underline{\tilde{\mathcal{X}}}_\ell)\|_F^2 \right.$$
$$\left. + \nu\|[\mathcal{E}_0 *_{1D}(\underline{\mathcal{V}} - \underline{\tilde{\mathcal{X}}})]_\ell\|_F^2 \right) \tag{15}$$

with $\odot$ the Hadamard product. Given a channel ℓ, $[\mathcal{E}_0 *_{1D}(\mathcal{V} - \tilde{\mathcal{X}})]_\ell$ is a sum of images $(\mathcal{V} - \tilde{\mathcal{X}})_{\ell'}$ weighted by the coefficients in $\mathcal{E}_0$. As the DFT is not calculated across channels, $[\mathcal{E}_0 *_{1D}(\underline{\mathcal{V}} - \underline{\tilde{\mathcal{X}}})]_\ell$ is also a sum of images $(\underline{\mathcal{V}} - \underline{\tilde{\mathcal{X}}})_{\ell'}$ with the same coefficients.

Since the convolution in $\mathcal{E}_0 *_{1D}(\underline{\mathcal{V}} - \underline{\tilde{\mathcal{X}}})$ is performed in the spectral channel domain, the minimization problem (15) can be separated into independent minimisation sub-problems w.r.t. each spatial frequency $\mathbf{f}$. The optimization problem in (11) can be decomposed into a set of independent least square problems, each corresponding to a point $\mathbf{f}$ in the spatial frequency domain:

$$\min_{\mathbf{v_f}} \|\mathbf{x}_{k+1,\mathbf{f}} - \mathbf{v_f}\|^2 + \mu'\|\Delta_{\mathcal{D}}(\mathbf{f})(\mathbf{v_f} - \tilde{\mathbf{x}}_\mathbf{f})\|^2 + \nu'\|\mathbf{E}_0(\mathbf{v_f} - \tilde{\mathbf{x}}_\mathbf{f})\|^2 \tag{16}$$

Table 1. Averaged RMSE, PSNR, SAM, ERGAS and SSIM of different methods on the CAVE and Harvard data sets.

Methods	CAVE data set					Harvard data set				
	RMSE	PSNR	ERGAS	SAM	SSIM	RMSE	PSNR	ERGAS	SAM	SSIM
UAL	1.854	44.656	0.196	4.33	0.9910	1.833	45.807	0.323	3.58	0.9832
UAL + Ours	**1.587**	**45.939**	**0.171**	**4.08**	**0.9917**	**1.784**	**46.034**	**0.316**	**3.54**	**0.9833**
NSSR	2.236	43.439	0.244	5.22	0.9849	1.874	45.540	0.363	3.73	0.9821
NSSR + Ours	**2.068**	**44.044**	**0.230**	**5.19**	**0.9854**	**1.844**	**45.649**	**0.357**	**3.69**	**0.9822**
LTTR	2.300	43.277	0.249	5.50	0.9848	1.914	45.251	0.375	3.81	0.9813
LTTR + Ours	**2.235**	**43.613**	**0.243**	**5.27**	**0.9851**	**1.887**	**45.392**	**0.374**	**3.77**	**0.9915**

with $\mu' = \mu/\rho$ and $\nu' = \nu/\rho$. The complex vectors $\mathbf{v_f}, \tilde{\mathbf{x}}_\mathbf{f}, \mathbf{x}_{k+1,\mathbf{f}}$ and the complex diagonal matrix $\Delta_\mathcal{D}$ are defined as follows:

$$\mathbf{v_f} = \{\underline{\mathcal{V}}_\ell(\mathbf{f}), \ell = 1, \ldots, B\}$$
$$\tilde{\mathbf{x}}_\mathbf{f} = \{\underline{\tilde{\mathcal{X}}}_\ell(\mathbf{f}), \ell = 1, \ldots, B\}$$
$$\mathbf{x}_{k+1,\mathbf{f}} = \{\underline{\mathcal{X}}_{k+1,\ell}(\mathbf{f}), \ell = 1, \ldots, B\} \quad (17)$$
$$\Delta_\mathcal{D}(\mathbf{f}) = \text{diag}\{\underline{\mathcal{D}}_\ell(\mathbf{f}), \ell = 1, \ldots, B\}$$

For each $\mathbf{f}$, the solution of (16) can be computed as:

$$\mathbf{v_f} = \mathbf{T_f}^{-1}(\mathbf{x}_{k+1,\mathbf{f}} + \mu'\Delta_\mathcal{D}(\mathbf{f})^*\Delta_\mathcal{D}(\mathbf{f})\tilde{\mathbf{x}}_\mathbf{f} + \nu'\mathbf{E}_0^*\mathbf{E}_0\tilde{\mathbf{x}}_\mathbf{f}) \quad (18)$$

where * denotes the complex conjugate and $\mathbf{T_f}$ is the real tri-diagonal matrix of size $B \times B$ given by:

$$\mathbf{T_f} = (\mathbf{I}_B + \mu'\Delta_\mathcal{D}(\mathbf{f})^*\Delta_\mathcal{D}(\mathbf{f}) + \nu'\mathbf{E}_0^*\mathbf{E}_0) \quad (19)$$

Finally, we can obtain $\mathbf{v}_{k+1}$ by separately calculating the inverse 2D DFT of each $\underline{\mathcal{V}}_\ell$ with $\ell = 1, \ldots, B$. This procedure is summarized in Algorithm 2.

5. EXPERIMENTS

We shall now validate the proposed strategy with experimental results. The code is made available at github.com/xiuheng-wang.

Two public HSI data sets, namely, the CAVE data set [19] and the Harvard data set [20], were used for our experiments. The CAVE data set is composed of 32 HSIs with a spatial dimension 512×512, and 31 channels in the spectral domain, covering the visible spectrum from 400 nm to 700 nm. The Harvard data set contains 50 HSIs consisting of 1392×1040 pixels in the spatial domain, and 31 channels ranging from 420 nm to 720 nm. For the Harvard data set, the top left 1024×1024 pixels were cropped and extracted.

The HSIs of the two data sets were scaled to range $[0, 1]$ and served as the ground truth for $\mathbf{X}$. The LR HSI $\mathbf{Y}$ was generated according to (1) where $\mathbf{B}$ is a uniform blurring operator over non-overlapping blocks of size 32×32, and $\mathbf{S}$ is a down-sampling operator with the down-sampling factor $s = 32$. The HR RGB $\mathbf{Z}$ image was obtained with (1), with $\mathbf{R}$ the response of a Nikon D700 camera.

We used a state-of-the-art deep learning method UAL described in [12] to calculate the deep prior $\tilde{\mathbf{X}}$ for each $\mathbf{X}$. We set the hyper-parameters as follows: $\mu = 0.05$, $\nu = 0.001$ in (3), and $\rho = 0.001$ in (9). The number of iterations K in Algorithm 2 was set to 20, which was sufficient to ensure convergence. To assess the quality of reconstructed images, we considered the following metrics: the root mean-square error (RMSE), the peak-signal-to-noise-ratio (PSNR), the spectral angle mapper (SAM) [21], the error of relative global adimensional synthesis (ERGAS) [22] and the structural similarity
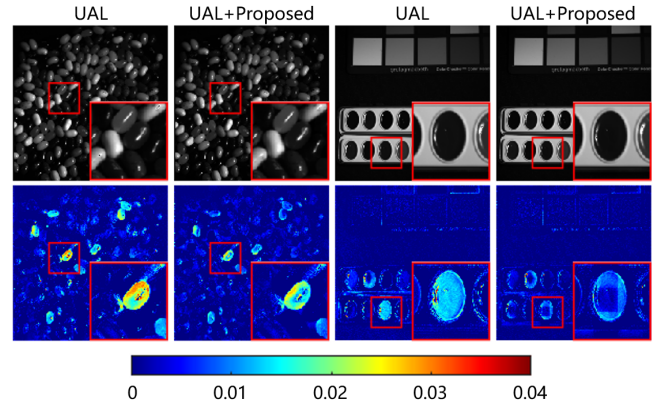

Fig. 2. Reconstructed images and corresponding error maps of two images from the CAVE data set in the 540 nm band.

(SSIM) [23]. Table 1 reports the performance of the UAL super-resolution algorithm [12], and of our algorithm which combines the UAL and the degradation model inversion (Algorithm 2). It can be observed that our algorithm significantly improved the performance of the UAL. Figure 2 confirms this observation by showing that our approach produced smaller reconstruction errors than the UAL.

For comparison purpose, we also considered other state-of-the-art super-resolution algorithms than the UAL to produce priors $\tilde{\mathbf{X}}$: the NSSR [5] based on sparse decompositions, and the LTTR [6] based on tensor factorizations. The rational was to show that coupling our approach with these algorithms improves their performance as it makes use of the linear degradation model and exploits the spectral-spatial smoothness of HSIs. Algorithm 2 was setup as described above. To setup NSSR and LTTR, we used the codes provided by their authors and fine-tuned all parameters to achieve the best super-resolution performance. Table 1 confirms that our approach allowed us to improve the performance of both NSSR and LTTR. The best performance was however achieved by using, with our algorithm, the deep priors provided by the UAL.

6. CONCLUSION

In this paper, we introduced an HSI super-resolution method which makes use of a degradation model in the data-fidelity term of the objective function and, on the other hand, utilizes the spectral-spatial gradient deviation of latent HSIs and the output of a convolutional neural network as a deep prior regularizer. Experiments showed the performance improvement achieved with this strategy compared with state-of-the-art methods.

7. REFERENCES

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

[2] S. Henrot, C. Soussen, and D. Brie, "Fast positive deconvolution of hyperspectral images," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 828–833, 2012.

[3] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J.-Y. Tourneret, M. Chen, and S. Godsill, "Multiband image fusion based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7236–7249, 2016.

[4] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Springer, 2014, pp. 63–78.

[5] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, 2016.

[6] 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.

[7] J. Chen, C. Richard, and P. Honeine, "Nonlinear estimation of material abundances in hyperspectral images with ℓ_1-norm spatial regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2654–2665, 2013.

[8] Y. Song, E.-H. Djermoune, J. Chen, C. Richard, and D. Brie, "Online deconvolution for industrial hyperspectral imaging systems," *SIAM J. Imaging Sci.*, vol. 12, no. 1, pp. 54–86, 2019.

[9] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-d-convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 639–643, 2017.

[10] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "Pannet: A deep network architecture for pan-sharpening," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 5449–5457.

[11] 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)*, 2019, pp. 1585–1594.

[12] 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)*, 2020, pp. 3073–3082.

[13] W. Xie, J. Lei, Y. Cui, Y. Li, and Q. Du, "Hyperspectral pansharpening with deep priors," *IEEE Trans. Neural Netw. Learn. Syst.*, 2019.

[14] X. Wang, J. Chen, Q. Wei, and C. Richard, "Hyperspectral image super-resolution via deep prior regularization with parameter estimation," *IEEE Trans. Circuits Syst, Video Technol*, 2021.

[15] M. Vella, B. Zhang, W. Chen, and J. F. C. Mota, "Enhanced hyperspectral image super-resolution via rgb fusion and tv-tv minimization," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2021, pp. 3837–3841.

[16] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, 1995.

[17] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation ax+ xb= c [f4]," *Commun. ACM.*, vol. 15, no. 9, pp. 820–826, 1972.

[18] 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.

[19] 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, 2010.

[20] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2011, pp. 193–200.

[21] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm," in *Proc. Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, vol. 1, pp. 147–149.

[22] L. Wald, "Quality of high resolution synthesised images: Is there a simple criterion?," in *Proc. 3rd conf. "Fusion of Earth Data"*. SEE/URISCA, 2000, pp. 99–103.

[23] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.