Unsupervised Multispectral and Hyperspectral Image Fusion with Deep Spatial and Spectral Priors

Zhe Liu[1], Yinqiang Zheng[2], and Xian-Hua Han[1][0000−0002−5003−3180]

[1] Graduate School of Science and Technology for Innovation, Yamaguchi University, Japan
b602vz@yamaguchi-u.ac.jp
hanxhua@yamaguchi-u.ac.jp
[2] National Institute of Informatics, Tokyo, Japan
yqzheng@nii.ac.jp

Abstract. Hyperspectral (HS) imaging is a promising imaging modality, which can simultaneously acquire various bands of images of the same scene and capture detailed spectral distribution helping for numerous applications. However, existing HS imaging sensor can only obtain images with low spatial resolution. Thus fusing a low resolution hyperspectral (LR-HS) image with a high resolution (HR) RGB (or multispectral) image into a HR-HS image has received much attention. Conventional fusion methods usually employ various hand-crafted priors to regularize the mathematical model formulating the relation between the observations and the HR-HS image, and conduct optimization for pursuing the optimal solution. However, the politic prior would be various for different scenes and is difficult to hammer out for a specific scene. Recently, deep learning-based methods have been widely explored for HS image resolution enhancement, and impressive performance has been validated. As it is known that deep learning-based methods essentially require large-scale training samples, which are hard to obtain due to the limitation of the existing HS cameras, for constructing the model with good generalization. Motivated by the deep image prior that network architecture itself sufficiently captures a great deal of low-level image statistics with arbitrary learning strategy, we investigate the deep learned image prior consisting both spatial structure and spectral attribute instead of hand-crafted priors for unsupervised multispectral (RGB) and HS image fusion, and propose a novel deep spatial and spectral prior learning framework for exploring the underlying structure of the latent HR-HS image with the observed HR-RGB and LR-HS images only. The proposed deep prior learning method has no requirement to prepare massive triplets of the HR-RGB, LR-HS and HR-HS images for network training. We validate the proposed method on two benchmark HS image datasets, and experimental results show that our method is comparable or outperforms the state-of-the-art HS image super-resolution approaches.

1 Introduction

In recent decades of years, imaging technique has been witnessed significant progress for providing high-definition images in different applications from agriculture, astronomy to surveillance and medicine, to name a few. Although the acquired images with the existing imaging systems can provide the high-definition information in a specific domain such as spatial-, temporal- or spectral- domain according to the application requirement, it is difficult to simultaneously offer all-possible required detail distribution in all domains such as the high-resolution hyperspectral (HR-HS) images to meet the demand of the resolution enhancement in both spatial- and spectral- domains. It is well known that hyperspectral (HS) imaging employs both traditional two-dimensional imaging technology and spectroscopic technology for obtaining a three-dimensional cubic data for a scene, and enriches greatly the spectral information for being successfully applied in remote sensing [1], [2], medical image analysis [3], and many computer vision tasks, such as object recognition and classification [4], [5], [6], tracking [7], segmentation [8]. However, the detail distribution in spectral domain (high spectral resolution) implies less radiant energy being able to be collected for each band of narrow spectrum. For guaranteeing acceptable signal-to-noise ratio, photo collection has to be performed in a much larger spatial region via sacrificing the spatial resolution. On the other hand, ordinary RGB cameras usually produce RGB images with high-resolution in spatial domain. Thus fusing the low-resolution HS image (LR-HS) with a corresponding high-resolution RGB (HR-RGB) image to generate a HR-HS image (called as multispectral and hyperspectral image fusion) has attracted remarkable attention.

Multispectral and hyperspectral image fusion is a challenging task due to its ill-posed nature in reality. Most existing methods mainly employ various hand-crafted priors to regularize the mathematical model formulating the relation between the observations and the HR-HS image, and conduct optimization for pursuing the optimal solution. Therein, one research line explores different spectral representation methods according to physical property of the observed spectrum such as matrix factorization and spectral unmixing motivated by the fact that the HS observations can be formulated as a weighted linear combination of the reflectance function basis and their corresponding fraction coefficients [8]. On the other hand, many work investigated sparse-promoted representation [9] as the prior knowledge for modeling the spatial structure and local spectral characteristic based on a dictionary trained on the observed HR-RGB and LR-HS images, and proved feasibility for HR-HS image reconstruction. Beside sparse constraint on spectral representation, low-rank technique has also been exploited to encode the intrinsic spectral correlation prior on the underlying HR-HS image for reducing spectral distortion [10]. There are also several work to explore the global spatial structure and local spectral similarity priors for further boosting the performance of the HS image reconstruction [11], [12]. Although the promising performance with the hand-crafted priors such as mathematical sparsity, physical property of spectral unmixing, low-rank and similarity has been achieved, different scenes with highly diverse configurations both along space

and across spectrum should have various effective priors for modeling and to figure out a proper prior for a specific scene is still difficult.

Recently, deep learning (DL) based methods have popularly been applied for the HS image reconstruction, and evolved into three research directions: 1) conventional spatial resolution enhancement with the observed LR-HS image, 2) traditional spectral resolution enhancement with the observed HR-RGB image, 3) fusion method with both LR-HS and HR-RGB images. Compared with the traditional prior-promoted methods, DL based methods do not need to rely on any assumption on the prior knowledge and can automatically capture the intrinsic characteristics of the latent HS images via data-driven learning. However, the DL based methods are generally used in a fully supervised way, and it is mandatory to previously collect large amount of training triplets consisting of the observed LR-HS and HR-RGB images, and their corresponding HR-HS images for learning optimal network parameters [13], [14], [15]. It is known that in the HS image reconstruction scenario, it is extremely hard to obtain large-scale training samples especially the HR-HS images as the label samples. In spite of the prospected advantage, the fully supervised DL scheme suffers from less generalization in real applications due to small number of the available training triplets. On one hand, Ulyanov et al. [16] advocated that the architecture of a generator network itself can capture quite a lot of low-level image priors with arbitrary learning strategy, and proposed deep image prior (DIP) learning with the deep network. The DIP method has successfully been applied to different natural image restoration tasks and manifested excellent results without any additional training samples.

Motivated by the fact that the deep network architecture itself carries large amount of low-level prior knowledge as explored in the DIP work [16], we propose a novel deep spatial and spectral prior (DSSP) learning farmework for HS image reconstruction. With a random noisy input, we attempt to learn a set of optimal parameters via searching the network parameter space to recover the latent HR-HS image, which is capable of approximating the observed HR-RGB and LR-HS images under a degradation procedure. In the network training step, we leverage both observed LR-HS and HR-RGB images of the under-studying scene to formulate the loss functions for capturing the underlying priors of the latent HR-HS image. Via employing the deep learned spatial and spectral priors, our proposed DSSP method can effectively recover the underlying spatial and spectral structure of the latent HR-HS image even only with the observed HR-RGB and LR-HS images, and it is not mandatory to prepare massive triplets of the HR-RGB, LR-HS and HR-HS images.

The main contributions of this work are three-fold:

1) We propose a novel unsupervised framework for fusing the observed LR-HS and HR-RGB (multispectral: MS) images to generate a HR-HS image, called as MS/HS fusion, in deep learning scenario.

2) We propose a deep spatial and spectral prior learning network for the MS/HS fusion, which is expected to effectively characterize the spatial structure

and the spectral attribute in the latent HR-HS image without manually analysis of the content in the under-studying scene.

3) We leverage both modality data of the observed LR-HS and HR-RGB images, and construct the loss functions of our proposed DSSP network for learning more reliable priors in the latent HR-HS image.

We validate our method on two benchmark HS image datasets, and experimental results show that our method is comparable or outperforms the state-of-the-art HS image super-resolution approaches.

The rest of this paper is organized as follows. Section 2 surveys the related work including traditional pan-sharpening and prior-promoted methods and deep learning based methods. Section 3 presents the proposed deep spatial and spectral prior learning framework for HS image reconstruction. Extensive experiments are conducted in Sec. 4 to compare our proposed framework with state-of-the-art methods on two benchmark datasets. Conclusion is given in Sec. 5.

2 Related Work

2.1 Traditional Methods

Multispectral and hyperspectral (MS/HS) image fusion is closely related multispectral (MS) image pan-sharpening which aims at merging a low-resolution MS image with a high-resolution wide-band panchromatic image [17], [18], [19]. There are many developed methods for MS pan-sharpening, which can be mainly divided into two categories: component substitution [18] and multiresolution analysis. Although MS/HS image fusion can intuitively be treated as a number of pan-sharpening sub-problems with each band of HR-MS (RGB) image as a panchromatic image, it cannot make full use of the spectral correlation and always suffers from the high spectral distortion.

Recently, many methods formulate MS/HS image fusion as an inverse optimization problem, and leverage the hand-crafted priors in the latent HR-HS image for boosting reconstruction performance. How to design the appreciate priors plays a key role in finding the feasible solutions for the optimization problem. The existing methods extensively investigated the prior knowledge for spatial and spectral representation such as physical spectral mixing, sparsity, low-rank, and manifest impressive performance. Yokoya et al. [14] proposed coupled nonnegative matrix factorization (CNMF) to fuse a pair of HR-MS and LR-HS images and gave a convincing spectral recovery result. Lanaras et al. [15] exploited the coupled spectral unmixing method for HS image reconstruction, and utilized near-end alternating linearization method to optimize. The other research effort concentrated the sparsity promoting approaches via imposing sparsity constraints on the representative coefficients [20]. Grohnfeldt et al. [21] employed a joint sparse representation via firstly learning the corresponding HS and MS (RGB) patch dictionary, and then using the sparse coefficients in each individual band image to reconstruct the spatial local structure (patch). Akhtar et al. [22]

conducted another sparse coefficient estimation algorithm and designed the generalized simultaneous orthogonal matching pursuit (G-SOMP) by assuming that the same atoms are used to reconstruct the spectrum of pixels in the local grid region. In order to use the prior more effectively in the inherent structure of the HR-HS image, Dong et al. [23] investigated a non-negative structured sparse representation (NSSR) method, whose principle is to use spectral similarity in local regions to limit sparse representation learning in order to restore HR-HS images closer to the real. Han et al. [24] extended to employ both local spectral and global structure similarity in the sparse-promotion scenario for further improving the robust recovery of HR-HS images. For now, although these hand-crafted prior algorithms have already achieved promising performance, seeking the suitable prior for a specific scene is still a challenging task.

2.2 Deep learning based methods

Motivated by the success of deep learning in the field of nature RGB image enhancement, deep convolutional neural network has been applied for MS/HS image fusion, and does not need to model the hand-crafted prior. Han et al. [25] conducted a pilot study to use a simple 3-layer CNN for fusing the LR-HS and HR-RGB images with large difference of spatial structures, and further extended to more complex CNN for pursuing better performance. Palsson et al. [26] proposed a 3D-CNN based MS/HS fusion method by using PCA to reduce the computational cost. Dian et al. [27] proposed to combine the optimization- and CNN- based methods together, and validated promising HR image reconstruction results. All the above deep learning based methods are implemented under a fully supervised way, and require to previously prepare a lot of training triplets including the LR-HS, HR-RGB (MS) and the label HR-HS images for network training. However, large amount of training samples especially the HR-HS images in the HS image reconstruction scenario are difficult to be collected. Thus, Qu et al. [28] investigated an unsupervised encoder-decoder architecture to solve the MS/HS fusion problem which dose not need for any training by using a HS image dataset. Although the prospected applicability using a CNN-based end-to-end network in an unsupervised way, this method needs to be carefully optimized for the two subnetworks in an alternating way, and still has much potential for performance improvement. This study also aims at proposing an unsupervised MS/HS fusion network via automatically learning both the spatial and spectral priors in an end-to-end learning way.

3 Proposed Method

In this part, we firstly describe the formula expression for the problem of the MS/HS image fusion, and then investigate the proposed unsupervised MS/HS image fusion with the deep spatial and spectral priors (DSSP) including the generator network architecture, which automatically learns the underlying priors of the latent HR-HS image from the observed image pair of LR-HS and HR-RGB images only, and the constructed loss function for network training.

3.1 Problem Formulation

Given an observed image pair: a LR-HS image $\mathbf{X} \in \mathbb{R}^{w \times h \times L}$ and a HR-RGB image $\mathbf{Y} \in \mathbb{R}^{W \times H \times 3}$, where w, h and L stands for the width, height and the spectral channel number of the LR-HS image, W and H denotes the width and height of the HR-RGB image, our goal is to reconstruct HR-HS image: $\mathbf{Z} \in \mathbb{R}^{W \times H \times L}$ via merging $\mathbf{X}$ and $\mathbf{Y}$. The degraded model of the observed $\mathbf{X}$ and $\mathbf{Y}$ from the latent $\mathbf{Z}$ can be mathematically formulated as:

$$\mathbf{X} = \mathbf{ZDB} + \mathbf{n}, \mathbf{Y} = \mathbf{CZ} + \mathbf{n} \tag{1}$$

where $\mathbf{B}$ and $\mathbf{D}$ stand for the spatial blurring filter and down-sampling function to transform $\mathbf{Z}$ to $\mathbf{X}$, and $\mathbf{C}$ denotes the spectral sensitivity function (CSF) of a RGB sensor and $\mathbf{n}$ represents the observed noise. The heuristic approach to utilize the observed $\mathbf{X}$ and $\mathbf{Y}$ for estimating $\mathbf{Z}$ is usually to minimize the following reconstruction errors:

$$\mathbf{Z}^{\star} = \arg\min_{\mathbf{Z}} \|\mathbf{X} - \mathbf{ZBD}\|_F^2 + \|\mathbf{Y} - \mathbf{CZ}\|_F^2 \tag{2}$$

where $\|\cdot\|_F$ represents the Frobenius norm. Eq. (2) trys to find out an optimized $\mathbf{Z}^{\star}$ which can minimize the reconstruction error of the observations. The terms in Eq. (2) rely on the observed data. According to the degradation procedure of the observed images $\mathbf{X}$ and $\mathbf{Y}$, it is known that the total number of unknown variables in $\mathbf{Z}$ is much more than the known variables in $\mathbf{X}$ and $\mathbf{Y}$, and thus results in ill-posed nature in this task. To address this problem, most existing methods popularly explores various hand-crafted priors for modeling the underlying structure of the HR-HS image to regularize the reconstruction error minimization problem, which is formulated as a regularization term:

$$\mathbf{Z}^{\star} = \arg\min_{\mathbf{Z}} \|\mathbf{X} - \mathbf{ZBD}\|_F^2 + \|\mathbf{Y} - \mathbf{CZ}\|_F^2 + \phi R(\mathbf{Z}) \tag{3}$$

where ϕ represents hyper-parameter to make a balance between the contribution of the regularization term and the reconstruction error. As we know that seeking an appropriate prior for a specific scene is difficult technically. This study advocates that a large amount of low-level image statistics can be captured by the deep network architecture itself, and it is prospected to generate a more plausible image according to the possessed low-level priors in the deep network. In the HS image scenario, we employ a deep network architecture to automatically learn the spatial and spectral priors in the latent HR-HS image, and then reconstruct a reliable HR-HS image constrained by the learned priors.

3.2 The Proposed Deep Spatial and Spectral Priors (DSSP)

The deep learning based methods such as DCGAN [29] and its variants verified that high-definition and high-quality images with a specific concept can be generated from a random noise, which means that to search the network parameter

space from the initial random state can learn the inherent structure (prior) in the latent image of a specific concept. In addition, DIP [16] explored image prior possessing capability of network architecture for different restoration tasks of natural RGB images, and manifested impressive results. This study investigates the deep learned prior in the latent HR-HS image including HR spatial structure and spectral attribute, and aims at generating the HR-HS image with the observed LR-HS and HR-RGB images only. We design an hourglass network architecture consisting of encoder and decoder subnets, each with 4-blocks (levels). The network schematic in detail is shown in Fig. 1. The input of the network is a noise cube $\mathbf{n} \in \hat{\mathbb{R}}^{W \times H \times L}$ with the same size of the required HR-HS image, and we expect that the network output: $f_\theta(\mathbf{n})$ (θ: network parameters) should approach the required HR-HS image: $\mathbf{Z}$. The goal of this work is to search the

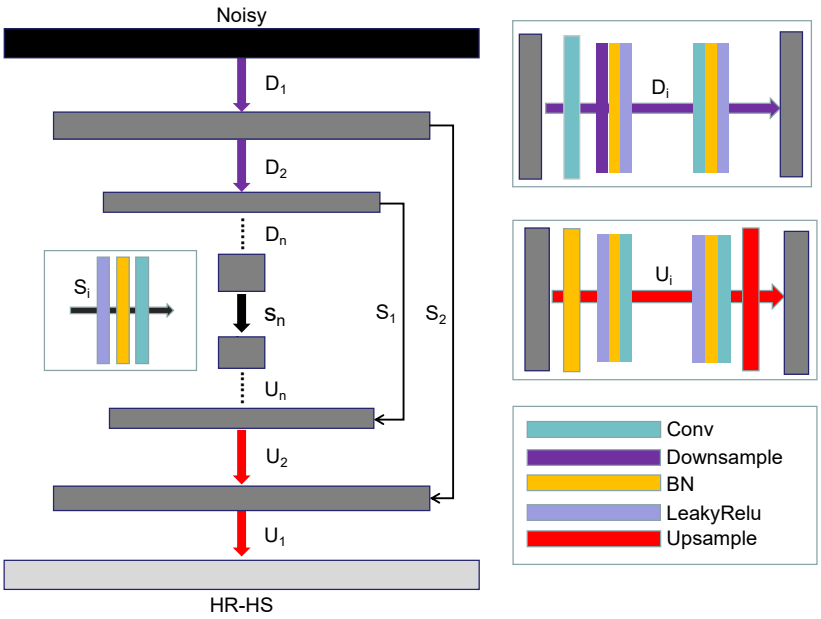


Fig. 1. Our generative network with an hourglass architecture, which generates the latent HR-HS image from a noisy input via automatically exploring the underlying spatial and spectral priors.

network parameter space to pursue a set of optimal parameters for satisfying the above criteria. However, due to the unknown $\mathbf{Z}$, it is impossible to construct quantitative criteria directly using $\mathbf{Z}$ for this task.

With the availability of the LR-HS and HR-RGB images, we turn to $\mathbf{X}$ and $\mathbf{Y}$ to formulate quantitative criteria (loss function) for network learning. Since,

as in Eq. (1), the LR-HS image $\mathbf{X}$ is a blurred down-sampled version of $\mathbf{Z}$, and HR-RGB image $\mathbf{Y}$ is a transformed version of $\mathbf{Z}$ in channel direction using CSF: $\mathbf{C}$, we implement the two operations as two convolutional layers with pre-defined weights (non-trainable) after the output layer of the baseline hourglass-like network. The convolutional layer for blurring/down-sampling operator has the kernel size and stride according to the spatial expanding factor W/w and the kernel weights are pre-calculated according to Lanczos2 filter. The output of this layer is denoted as $\hat{f}_{\mathbf{BD}}(f_\theta(\mathbf{n}))$, which has the same size and should be approximated to $\mathbf{X}$. Thus according to $\mathbf{X}$, the first loss function is formulated as:

$$L_1(\mathbf{n}, \mathbf{X}) = \left\| \mathbf{X} - \hat{f}_{\mathbf{BD}}(f_\theta(\mathbf{n})) \right\|_F^2 \tag{4}$$

While the spectral transformation operation (from $\mathbf{Z}$ to $\mathbf{Y}$) is implemented as the convolutional layer with 1×1 kernel size, input and output channels: L and 3, where the kernel weight is fixed as the CSF: $\mathbf{C}$ according to the used RGB camera. Then the output of this layer should be an optimal approximation of $\mathbf{Y}$. Denoting it as $\hat{f}_{\mathbf{C}}(f_\theta(\mathbf{N}))$, the second loss function is formulated as:

$$L_2(\mathbf{n}, \mathbf{Y}) = \left\| \mathbf{Y} - \hat{f}_{\mathbf{C}}(f_\theta(\mathbf{n})) \right\|_F^2 \tag{5}$$

Via combining the L_1 and L_2 loss functions, we finally minimize the following total loss for searching a set of network parameter from the initialed random state:

$$L(\mathbf{n}, \mathbf{X}, \mathbf{Y}) = \arg\min_\theta L_1(\mathbf{n}, \mathbf{X}) + L_2(\mathbf{n}, \mathbf{Y}) \tag{6}$$

From Eq. (6), it can be seen the network is learned with the available observations only without any additional training samples. After completing training, the baseline network output: $f_\theta(\mathbf{n})$ is our required HR-HS image.

4 Experiment Result

We validate our proposed DSSP network on 32 indoor HS images in CAVE dataset and 50 indoor and outdoor HS images in Harvard dataset. The images in CAVE dataset including paintings, toys, food, and so on, are captured under controlled illumination, and their dimensions are 512×512 pixels, with 31 spectral bands of 10 nm wide, covering the visible spectrum from 400 to 700nm. The Harvard dataset has 50 images under daylight illumination, both outdoors and indoors, using a commercial hyperspectral camera (Nuance FX, CRI Inc.). The images in Harvard dataset are of a wide variety of real-world materials and objects. We firstly took the top-left 1024×1024 regions from the raw HS images, and down-sampled them to size 512×512 as the ground-truth HS images. For experiment conducting, we synthesized the low-resolution HS image with the down-sampling factors of 8 and 32 using the Bicubic interpolation, and then the observed LR-HS images have sizes of $64 \times 64 \times 31$ and $16 \times 16 \times 31$, respectively. We generated the HR-RGB images via multiplying the spectral channels of the

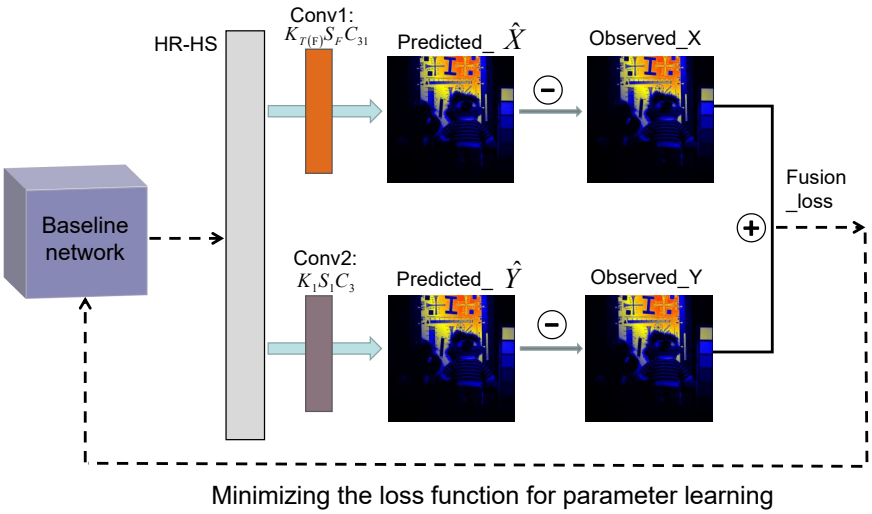


Fig. 2. The schematic concept of our proposed DSSP framework, which leverages the observed LR-HS and HR-RGB image to formulate the loss function for network training. The spatial down-sampling operation is implemented using a convolutional layer: 'Conv1: $K_{T(F)} \times S_F \times C_{31}$', with kernel size, stride and kennel number: T(F), F (spatial expanding factor) and 31, respectively while the spectral linear transformation from the HR-HS image to the HR-is image, is implemented using the convolutional layer: 'Conv2: $K_1 \times S_1 \times C_3$', with the kernel size, stride and kennel number: 1, 1 and 3, respectively.

ground truth HR-HS images with the spectral response function of Nikon D700 camera.

We conducted experiments with our proposed deep spatial and spectral prior (DSSP) framework using the observed LR-HS and HR-RGB images only, which means that the combined loss function in Eq. (6) is used for network parameter learning. Since our proposed method is evolved from the deep image prior for natural image restoration problems, which was further extended to enhance hyperspectral image for deep spatial prior (DSP) learning with the loss function in Eq. (4) [30], we compare the experimental results with our DSSP framework and the conventional DSP method. The experiments were conducted under the same experimental setting with 12000 iterations and learning rate 0.001 for both DSSP and DSP learning. Further, to avoid that the predicted HR-HS image drops down a local minimized point, we added a vibrated random noise with much smaller deviation to the network noise input on each iteration in the experiments. We

evaluate experimental results with five commonly-used quantitative metrics: root mean square error (RMSE), peak signal to noise ration (PSNR), structure similarity (SSIM), spectral angle mapper (SAM) and relative dimensional global error (ERGAS). We calculate the mean metric values of all images in the CAVE and Harvard datasets for comparison.

The compared experimental results on CAVE dataset using our proposed DSSP and the conventional DSP methods are given in Table 1 for both expanding factors 8 and 32 in spatial domain. From the results of Table 1, it can be seen that our proposed method can greatly outperform the conventional DSP method on all five metrics. Table 2 manifests the compared results on Harvard dataset, which also demonstrates much better performance of our DSSP method.

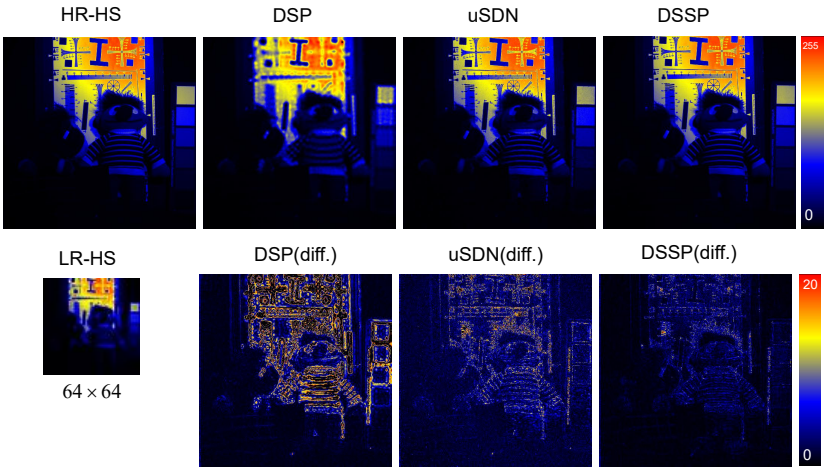
Table 1. Quantitative comparison results of our proposed DSSP framework and the conventional DSP method [30] on the CAVE dataset.

Factor	Method	RMSE	PSNR	SSIM	SAM	ERGAS
8	DSP [30]	7.5995	31.4040	0.8708	8.2542	4.2025
	DSSP	**2.0976**	**42.5251**	**0.9780**	**5.2996**	**1.1190**
32	DSP [30]	16.0121	24.7395	0.7449	13.0761	8.5959
	DSSP	**3.1279**	**39.0291**	**0.9619**	**7.6520**	**1.6366**

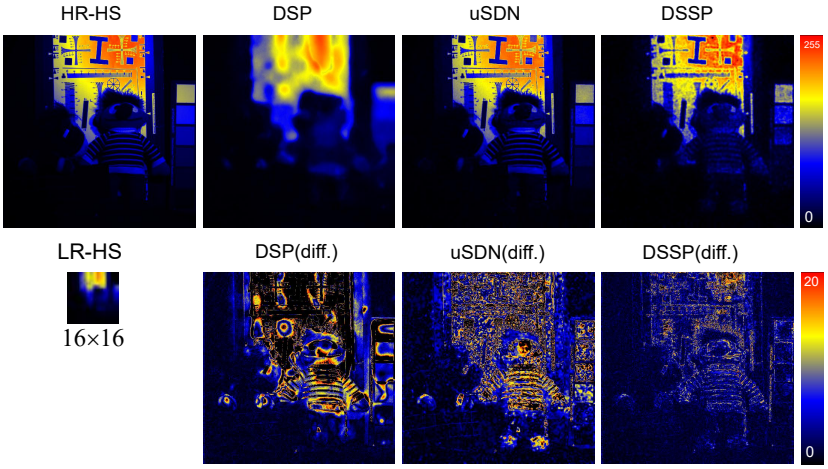
Table 2. Quantitative comparison results of our proposed DSSP framework and the conventional DSP method [30] on the Harvard dataset.

Factor	Method	RMSE	PSNR	SSIM	SAM	ERGAS
8	DSP [30]	7.9449	30.8609	0.8029	3.5295	3.1509
	DSSP	**2.1472**	**42.6315**	**0.9736**	**2.3204**	**1.0089**
32	DSP [30]	13.2507	26.2299	0.7186	5.6758	5.6482
	DSSP	**2.8366**	**40.3152**	**0.9602**	**3.5171**	**1.5809**

Finally, we compare the experimental results with other state-of-the-art methods. Since our DSSP method is an unsupervised MS/HS fusion strategy, for fair comparison we provide the results of the unsupervised methods including optimization based approaches: MF [31], CMF [14], BSR [32] and unsupervised deep learning based method: uSDN [28] in Table 3 with the expanding factor 32 of spatial domain for both CAVE and Harvard datasets, which manifests the promising performance using our proposed DSSP framework. The visual examples in both CAVE and Harvard datasets with our DSSP method, the conventional DSP and the uSDN methods are shown in Fig. 3 and Fig. 4.

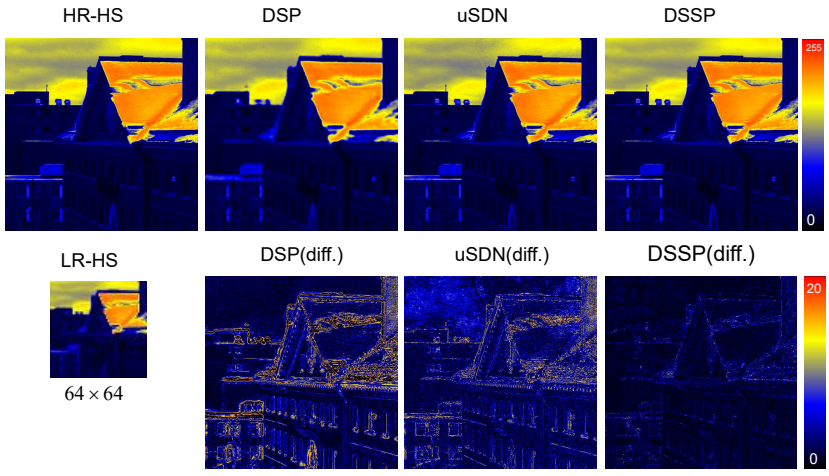


(a) Expanding factor: 8

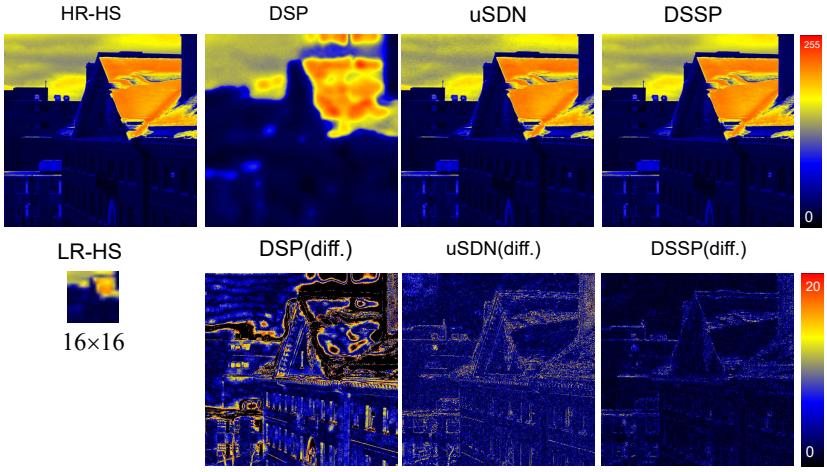


(b) Expanding factor: 32

Fig. 3. The predicted LR-HR image of 'chart and stuffed toy' sample in the CAVE dataset for both spatial expanding factors: 8 and 32, which visualizes the 16-th band image. The first column shows the ground truth HR image and the input LR image, respectively. The second to fourth columns show results from DSP [16], uSDN [28], our proposed method, with the upper part showing the predicted images and the lower part showing the absolute difference maps w.r.t. the ground truth.



(a) Expanding factor: 8



(b) Expanding factor: 32

Fig. 4. The predicted LR-HR image of 'img1' sample in the Harvard dataset for both spatial expanding factors: 8 and 32, which visualizes the 16-th band image. The first column shows the ground truth HR image and the input LR image, respectively. The second to fourth columns show results from DSP [16], uSDN [28], our proposed method, with the upper part showing the predicted images and the lower part showing the absolute difference maps w.r.t. the ground truth.

Table 3. The compared average RMSE, SAM and PSNR with the state-of-the-art unsupervised methods on both CAVE and Harvard datasets.

Method	CAVE			Harvard		
	RMSE	SAM	PSNR	RMSE	SAM	PSNR
MF [31]	3.47	8.29	38.61	2.93	3.99	40.02
CMF [14]	4.23	7.71	37.98	2.86	4.46	39.97
BSR [32]	3.79	9.12	35.25	3.7	4.26	38.52
uSDN [28]	3.89	7.94	37.46	3.02	3.98	38.08
DSSP	**3.1279**	**7.652**	**39.0291**	**2.8366**	**3.5171**	**40.3152**

5 Conclusion

This study proposed a deep unsupervised prior learning network for the fusion of multispectral and hyperspectral images. Motivated that a generative network architecture itself can capture large amount of low-level image statistics, we attempted to construct a simple network for learning the spatial and spectral priors in the latent HR-HS images. The proposed prior learning network can effectively leverage the HR spatial structure in HR-RGB images and the detailed spectral properties in LR-HS images to provide more reliable HS images reconstruction without any training samples. Experimental results on both CAVE and Harvard datasets showed that the proposed method has achieved impressive performance.

6 Acknowledgement

This research was supported in part by the Grant-in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) under the Grant No. 20K11867, and ROIS NII Open Collaborative Research 2020-20FC02.

References

1. Plaza, A., Benediktsson, J.A., Boardman, J.W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., et al.: Recent advances in techniques for hyperspectral image processing. Remote sensing of environment **113** (2009) S110–S122
2. Goetz, A.F.: Three decades of hyperspectral remote sensing of the earth: A personal view. Remote Sensing of Environment **113** (2009) S5–S16
3. Lu, G., Fei, B.: Medical hyperspectral imaging: a review. Journal of biomedical optics **19** (2014) 010901
4. Manolakis, D., Shaw, G.: Detection algorithms for hyperspectral imaging applications. IEEE signal processing magazine **19** (2002) 29–43

5. Makantasis, K., Karantzalos, K., Doulamis, A., Doulamis, N.: Deep supervised learning for hyperspectral data classification through convolutional neural networks. In: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IEEE (2015) 4959–4962
6. Manolakis, D., Marden, D., Shaw, G.A., et al.: Hyperspectral image processing for automatic target detection applications. Lincoln laboratory journal **14** (2003) 79–116
7. Treado, P., Nelson, M., Gardner Jr, C.: Hyperspectral imaging sensor for tracking moving targets (2012) US Patent App. 13/199,981.
8. Veganzones, M.A., Tochon, G., Dalla-Mura, M., Plaza, A.J., Chanussot, J.: Hyperspectral image segmentation using a new spectral unmixing-based binary partition tree representation. IEEE Transactions on Image Processing **23** (2014) 3574–3589
9. Chen, Y., Nasrabadi, N.M., Tran, T.D.: Hyperspectral image classification using dictionary-based sparse representation. IEEE transactions on geoscience and remote sensing **49** (2011) 3973–3985
10. Zhao, Y.Q., Yang, J.: Hyperspectral image denoising via sparse representation and low-rank constraint. IEEE Transactions on Geoscience and Remote Sensing **53** (2014) 296–308
11. Pu, H., Chen, Z., Wang, B., Jiang, G.M.: A novel spatial–spectral similarity measure for dimensionality reduction and classification of hyperspectral imagery. IEEE Transactions on Geoscience and Remote Sensing **52** (2014) 7008–7022
12. Yu, H., Gao, L., Liao, W., Zhang, B.: Group sparse representation based on nonlocal spatial and local spectral similarity for hyperspectral imagery classification. Sensors **18** (2018) 1695
13. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence **38** (2015) 295–307
14. Yokoya, N., Yairi, T., Iwasaki, A.: Coupled non-negative matrix factorization (cnmf) for hyperspectral and multispectral data fusion: Application to pasture classification. In: 2011 IEEE International Geoscience and Remote Sensing Symposium, IEEE (2011) 1779–1782
15. Lanaras, C., Baltsavias, E., Schindler, K.: Hyperspectral super-resolution by coupled spectral unmixing. In: Proceedings of the IEEE international conference on computer vision. (2015) 3586–3594
16. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2018)
17. Zhu, X.X., Bamler, R.: A sparse image fusion algorithm with application to pansharpening. IEEE transactions on geoscience and remote sensing **51** (2012) 2827–2836
18. Choi, J., Yu, K., Kim, Y.: A new adaptive component-substitution-based satellite image fusion by using partial replacement. IEEE Transactions on Geoscience and Remote Sensing **49** (2010) 295–309
19. Dhore, A., Veena, C.: A new pan-sharpening method using joint sparse fi image fusion algorithm. International Journal of Engineering Research and General Science **2** (2014) 447–55
20. Liang, H., Li, Q.: Hyperspectral imagery classification using sparse representations of convolutional neural network features. Remote Sensing **8** (2016) 99
21. Zhu, X.X., Grohnfeldt, C., Bamler, R.: Exploiting joint sparsity for pansharpening: The j-sparsefi algorithm. IEEE Transactions on Geoscience and Remote Sensing **54** (2015) 2664–2681

22. Akhtar, N., Shafait, F., Mian, A.: Sparse spatio-spectral representation for hyperspectral image super-resolution. In: European conference on computer vision, Springer (2014) 63–78
23. Meng, G., Li, G., Dong, W., Shi, G.: Non-negative structural sparse representation for high-resolution hyperspectral imaging. In: Optoelectronic Imaging and Multimedia Technology III. Volume 9273., International Society for Optics and Photonics (2014) 92730H
24. Han, X.H., Shi, B., Zheng, Y.: Self-similarity constrained sparse representation for hyperspectral image super-resolution. IEEE Transactions on Image Processing **27** (2018) 5625–5637
25. Han, X.H., Chen, Y.W.: Deep residual network of spectral and spatial fusion for hyperspectral image super-resolution. In: 2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM), IEEE (2019) 266–270
26. Palsson, F., Sveinsson, J.R., Ulfarsson, M.O.: Multispectral and hyperspectral image fusion using a 3-d-convolutional neural network. IEEE Geoscience and Remote Sensing Letters **14** (2017) 639–643
27. Dian, R., Li, S., Guo, A., Fang, L.: Deep hyperspectral image sharpening. IEEE transactions on neural networks and learning systems (2018) 1–11
28. Qu, Y., Qi, H., Kwan, C.: Unsupervised sparse dirichlet-net for hyperspectral image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition. (2018) 2511–2520
29. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434 (2015)
30. Sidorov, O., Hardeberg, J.Y.: Deep hyperspectral prior: Denoising, inpainting, super-resolution. CoRR **abs/1902.00301** (2019)
31. Kawakami, R., Matsushita, Y., Wright, J., Ben-Ezra, M., Tai, Y.W., Ikeuchi, K.: High-resolution hyperspectral imaging via matrix factorization. In: CVPR 2011, IEEE (2011) 2329–2336
32. Wei, Q., Bioucas-Dias, J., Dobigeon, N., Tourneret, J.Y.: Hyperspectral and multispectral image fusion based on a sparse representation. IEEE Transactions on Geoscience and Remote Sensing **53** (2015) 3658–3668