

Residual HSRCNN: Residual Hyper-Spectral Reconstruction CNN from an RGB Image

Xian-Hua Han
Graduate School of Science
and Technology for Innovation,
Yamaguchi University, Japan
Email: hanxhua@yamaguchi-u.ac.jp

Boxin Shi
Institute of Digital Media,
School of EECS,
Peking University, China
Email: shiboxin@pku.edu.cn

Yinqiang Zheng
Digital Content and Media Sciences
Research Division, National Institute
of Informatics, Japan
Email: yqzheng@nii.ac.jp

Abstract—Hyper-spectral imaging has great potential for understanding the characteristics of different materials in many applications ranging from remote sensing to medical imaging. However, due to various hardware limitations, only low-resolution hyper-spectral and high-resolution multi-spectral or RGB images can be captured at video rate. This study aims to generate a hyper-spectral image via enhancing spectral resolution of an RGB image, which might be easily obtained by a commodity camera. Motivated by the success of deep convolutional neural network (DCNN) for spatial resolution enhancement of natural images, we explore a spectral reconstruction CNN for spectral super-resolution with an available RGB image, which predicts the high-frequency content of the fine spectral wavelength in narrow band interval. Since the lost high-frequency content can not be perfectly recovered, by leveraging on the baseline CNN, we further propose a novel residual hyper-spectral reconstruction CNN framework to estimate the non-recovered high-frequency content (Residual) from the output of the baseline CNN. Experiments on benchmark hyper-spectral datasets validate that the proposed method achieves promising performances compared with the existing state-of-the-art methods.

I. INTRODUCTION

Hyper-spectral imaging is an emerging technique for simultaneously obtaining a set of images of the same scene at a large number of narrow band wavelengths. The rich spectra significantly benefit the characterization of the imaged scene and greatly enhance performance in different computer vision tasks, including object recognition, classification, tracking and segmentation [15], [27], [32], [33], [39]. Furthermore, the characterization of hyper-spectral images has contributed to disease diagnosis in medical imaging [40] and land resource management/planning in remote sensing [7], [3]. Although hyper-spectral imaging can offer high spectral resolution, it imposes a severe limitation on the temporal resolution compared with RGB or multispectral images from commodity cameras, since a longer exposure time is usually necessary to simultaneously collect the large number bands of spectra within a narrow wavelength window, such that the signal-to-noise ratio is sufficient. This drawback can be partially resolved via collecting photons in a much larger spatial region than the visible spectrum cameras, resulting in much lower spatial resolution. Despite of the potential benefits of the hyper-spectral images in different vision tasks, the limited spatial resolution due to the spatial and temporal tradeoff

restricts its performances for scene analysis and understanding, where spectral mixture of different materials happens. The intuitive way to obtain a high-resolution hyper-spectral (HR-HS) image is to enhance the spatial resolution from the available low-resolution hyper-spectral (LR-HS) image, where the enhanced spatial factor is limited.

On the other hand, the increasing variety of visual sensors enables to collect immense amount of information from the environment. Most common visual sensors can only record a limited bands of wavelength, often containing the standard red, green, and blue, from the visible spectrum to match the trichromatic perception of the human visual system. Yet, they usually achieve higher spatial resolution in contrast to the hyper-spectral cameras. With the large amount of available RGB images, one research line is to learn the spectral structure of the visual world, and use it as a prior to predict hyper-spectral images with finer spectral resolution from a standard RGB image, which is called spectral reconstruction for HSI super-resolution (SR).

There is a large body of literature on single natural image super-resolution for spatial resolution enhancement, which have also been applied for HSI SR with an expanding factor from 2 to 4 [10], [37], [12]. Since the spatial resolution of the available HS image is considerably low compared with the commonly observed RGB image, the expanding factor of spatial resolution in hyper-spectral image is required to be large enough, for example, more than 10 in horizontal and vertical directions, respectively. Thus the enhanced spatial resolution with acceptable quality of the recovered HR image usually can not reach the required resolution for different applications.

With the benefit of the simultaneous availability of the LR-HS and HR-RGB images in some specific tasks, some work focused on generating HR-HS images via fusion of the available LR-HS and HR-RGB images [17], [19]. In this scenario, both spatial down-sampling operator to generate the LR-HS image from the HR-HS and spectral response function to generate the HR-RGB for the HR-HS image are assumed to be known, and the observed LR-HS and HR-RGB image are also needed to be precisely aligned, to obtain the accurate estimation of HR-HS images. With the above assumed known factors and some designed prior knowledge (constraints) about

the required HR-HS image, the fusion methods usually give impressive recovery performance [17], [19], [11], [18]. However, in most real applications, both spatial down-sampling operator and spectral response function are unknown, and it is difficult to accurately align the observed LR-HS and HR-RGB images.

As a complementary problem, some researchers have tried to increase the spectral resolution of the input image beyond the coarse RGB channels, and demonstrated the feasibility of learning hyper-spectral structure of the visual world from the observed RGB image only [28], [6], [16], [1].

Recently deep convolutional neural network (DCNN) has been applied for spatial resolution enhancement of natural images, and manifested promising performance [12], [13], [23]. The intuitive way to apply DCNN for HSI SR is to enhance the spatial resolution from a LR-HS image. This study proposes a CNN-based spectral resolution enhancement method from a RGB image for HSI SR, and learns the hyper-spectral structure of different materials in visual world, which is called hyper-spectral reconstruction CNN (HSRCNN). The estimated HR-HS image with the HSRCNN only is perfectly recovered as the ground truth HR-HS image even in training procedure, which results in a non-recovered residual image (high-frequency content). We overwrite the same CNN architecture on the baseline CNN to further estimate the non-recovered residual image from the output of the baseline CNN, and propose a novel residual hyper-spectral reconstruction CNN framework (Residual HSRCNN). Experiments on benchmark hyper-spectral datasets validate that the proposed method achieves promising performances compared with the existing state-of-the-art methods.

II. RELATED WORK

High-resolution HS images are useful in various application fields ranging from remote sensing to medical imaging, but it is difficult to simultaneously achieve high-resolution in both spatial and spectral domains due to technique and budget constraints [19]. Thus it has attracted much research attention to generate high resolution HS images via image processing and machine learning techniques based on the available LR-HS and HR-HS images. In remote sensing field, a high resolution panchromatic image is usually available accompanying with the low resolution multi-spectral or HS image and the fusion of these two images is generally known as the pansharp-ening operation [11], [18], [2], [26], [41]. In this scenario, most popular approaches concentrated on reliable illumination restoration by means of intensity substitution and projection with the explored sue saturation and principle component analysis [11], [18], which generally cause spectral distortion in the recovered image [8]. Recently, the HS image super-resolution is mainly based on the fusion method of the observed LR-HS and HR-RGB images, and it estimates the HR-HS image as an optimization problem with prior knowledge, such as sparsity representation and spectral physical properties as constraints [8], which have limited applicability.

Inspired by the success of the deep convolutional neural networks for different image processing and computer vision applications, such as object detection and segmentation [31], face recognition [30], image denoising [24], CNN has also been applied to natural image superresolution and achieved promising performance. Dong et al. [12] proposed a 3-layer CNN architecture (SRCNN), which manifests about better improvement and lower computational cost compared with the popularly used sparse-based methods, and extended the CNN architecture to deal with the available LR image without upsampling as input (Fast SRCNN) [13]. Kim et al. [22] exploited a very deep CNN architecture based on VGG-network [29], and only learned the lost high-frequency image (residual image) to speed up the training procedure. Ledig et al. [23] combined GAN network for estimating much sharper HR image, and sometime led to unreliable detailed structure. On the other hand, Li et al. [25] applied the CNN architecture of SRCNN for HS image superresolution from the LR-HS image and achieved acceptable performance. All the above CNN architectures for image superresolution (SR) aim to enhance spatial resolution with the low spatial resolution image as input, and the expanding factor of spatial resolution enhancement is usually limited to be lower than 4 in horizontal and vertical directions, respectively. On the other hand, several work explored CNN-based methods with variant backbone architectures to expand the spectral resolution with the available HR-RGB image, and manifested promising performance [16], [1], [5], [36]. The topic of this study is to learn spectral structure based on CNN, and propose a residual hyper-spectral reconstruction CNN to estimate the non-recovered high-frequency spectral content.

III. RESIDUAL HYPER-SPECTRAL RECONSTRUCTION CNN

The intuitive way to apply CNN for image super-resolution is to learn the HR-HS image directly from the available LR-HS images, and the expanding factor of spatial resolution is limited to be lower than 8 in horizontal and vertical directions, respectively. We call the spatial resolution enhancement CNN as Spatial-CNN. In the scenario of the HSI SR, the complementary problem to the spatial resolution enhancement is to expand the spectral resolution for recovering the finer spectra of narrow band from a HR-RGB image. The goal of this study is to estimate a high resolution hyper-spectral image $\mathbf{Z} \in \mathbb{R}^{W \times H \times L}$, where W and H denote the spatial dimensions and L is the spectral band number, from a HR-MS (RGB) image $\mathbf{Y} \in \mathbb{R}^{W \times H \times l}$ ($l \ll L$). In our experiments, the available HR-MS image is an RGB image with spectral band number $l = 3$. An image observation model for depicting the relationship between the desired HR-HS and the input HR-RGB images can be formulated as

$$\mathbf{Y} = \mathbf{Z} *^{Spec} \mathbf{R} \downarrow + \mathbf{n} \quad (1)$$

where \mathbf{R} represents the spectral transformation matrix (a one-dimensional spectral-directional filter) decided by camera design, which maps the HR-HS image \mathbf{Z} to the HR-RGB

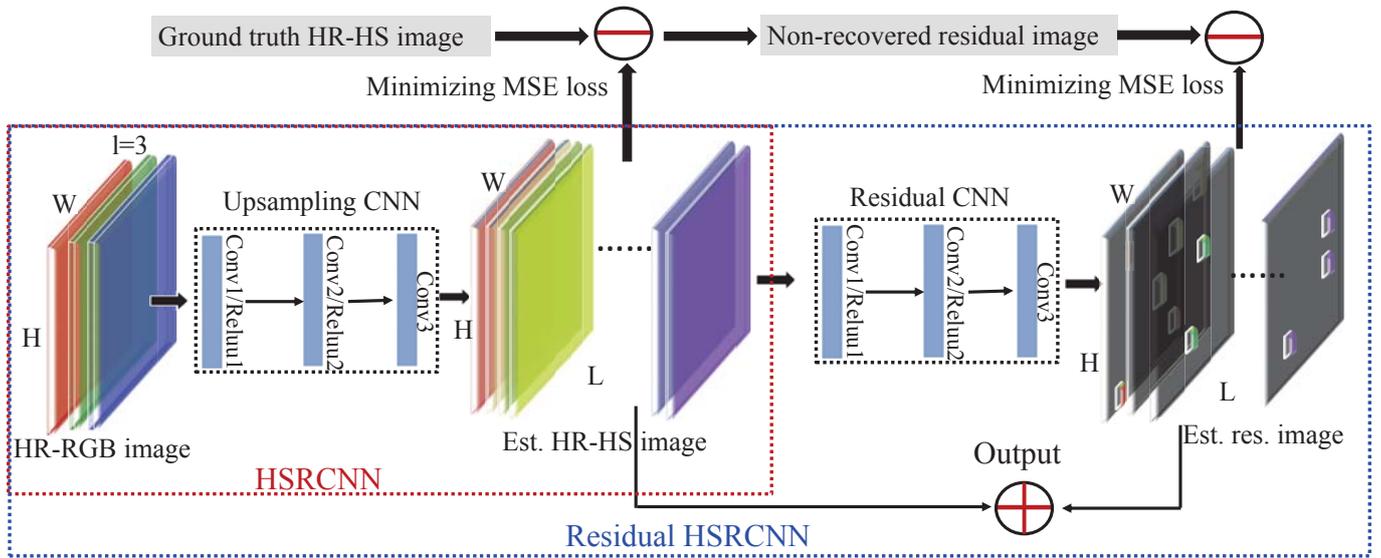


Figure 1. The schematic concept of the HSRCNN and residual HSRCNN.

image \mathbf{Y} , \ast^{Spec} denotes the convolutional operation in spectral domain and \downarrow is the down-sampling operation. This study constructs an end-to-end CNN architecture to estimate the HR-HS image \mathbf{Z} from the observed HR-RGB image \mathbf{Y} , named as HSRCNN. Furthermore, in order to obtain the non-recovered high-frequency spectral content (Residual) with the HSRCNN, we overwrite the baseline CNN architecture (Three convolutional layers as shown in the ‘Upsampling CNN’ part of Fig. 1), to update the residual component.

A. Hyper-Spectral Reconstruction CNN: HSRCNN

Motivated by its success for image superresolution and its compact structure, we follow the SRCNN as in [12], [25], which mainly consists of three convolutional layers that conduct three operations in the mapping process from LR images to HR images following the schematic concept in sparse coding-based SR: patch extraction and representation, non-linear mapping, and reconstruction. Instead of only the illumination (Y component in YCbCr color space) component as input and output in the conventional SRCNN framework, the input in our HSRCNN is the high spatial resolution RGB image, and the output is the high resolution image in both spatial and spectral domain (HR-HS) image. Patch extraction and representation extracts some overlapping patches from the input color image, and explains each color patch as a high dimensional vector. The convolution and RELU layers in CNN acts as a non-linear function which maps a high-dimensional vector (conceptually, the representation of the patches in the input) to another high-dimensional vector (the feature map in the middle-layer of CNN). Reconstruction process combines the mapped CNN features into the final multiple-bands of HR image.

Due to the increasing band number in the input and output images in our HSRCNN, we adjust the spatial filter size in

three convolutional layers as 3×3 , 3×3 , 5×5 with full connection in spectral domain instead of the used sizes 9×9 , 1×1 , 5×5 in the conventional spatial SRCNN. The baseline CNN architecture for HSRCNN and Residual HSRCNN is shown in the ‘Upsampling CNN’ part of Fig. 1, and the schematic concept of HSRCNN is the part of the red rectangle frame in Fig. 1. With the proposed HSRCNN, we want to estimate a HR-HS image \mathbf{Z} from the observed RGB image \mathbf{Y} , and the objective function is formulated as the following

$$\hat{\mathbf{Z}} = \arg \min_{\theta} \|\mathbf{Z} - f_{HSRCNN}(\mathbf{Y}, \theta)\|^2, \quad (2)$$

where $f_{HSRCNN}(\cdot)$ denotes the transformation function from the input RGB image \mathbf{Y} to the HR-HS image \mathbf{Z} with the filter parameters θ of three convolutional layers, and $\hat{\mathbf{Z}}$ is the estimated HR-HS image via the HSRCNN network.

B. Residual HSRCNN

Since the original HSRCNN needs to not only reconstruct low-frequency content but also estimate high-frequency spectral content lost in the input HR-RGB image, it is unavoidable that some high-frequency spectral content can not be recovered, and thus produce the non-recovered residual image: $\mathbf{Z}^{Res} = \mathbf{Z} - \hat{\mathbf{Z}} = \mathbf{Z} - f_{HSRCNN}(\mathbf{Y}, \theta)$. In order to further learn \mathbf{Z}^{Res} , this study overwrites the baseline CNN architecture of HSRCNN to construct a residual HSRCNN for estimating the non-recovered high-frequency spectral content from the $\hat{\mathbf{Z}}$. We combine the original HR-HS image: \mathbf{Z} and the residual component: \mathbf{Z}^{Res} learning procedure to construct an end-to-end residual component estimation network, and the objective function of the residual HSRCNN is formulated as:

$$\{\hat{\mathbf{Z}}, \hat{\mathbf{Z}}^{Res}\} = \arg \min_{\theta, \theta^{Res}} \omega_1 \|\mathbf{Z} - f_{HSRCNN}(\mathbf{Y}, \theta)\|^2 + \omega_2 \|\mathbf{Z}^{Res} - f_{Res}(f_{HSRCNN}(\mathbf{Y}, \theta), \theta^{Res})\|^2, \quad (3)$$

where $f_{Res}(\cdot)$ denotes the transformation function from the estimated HR-HS image $f_{HSRCNN}(\mathbf{Y}, \theta)$ to the residual component $\hat{\mathbf{Z}}^{Res}$ with the filter parameters θ^{Res} of three convolutional layers, and $\hat{\mathbf{Z}}^{Res}$ is the estimated residual component via the Residual HSRCNN network. ω_1 and ω_2 are the weights of the reconstruction errors on the first estimation of \mathbf{Z} and the residual component estimation of \mathbf{Z}^{Res} . The final estimation of the HR-HS image is the element-wise summation of $\hat{\mathbf{Z}}$ and $\hat{\mathbf{Z}}^{Res}$: $\hat{\mathbf{Z}}^{Final} = \hat{\mathbf{Z}} + \hat{\mathbf{Z}}^{Res}$. The schematic concept of the proposed Residual HSRCNN is shown in Fig. 1.

IV. EXPERIMENTAL RESULTS

We evaluate the proposed approach using two publicly available hyper-spectral imaging database: the CAVE dataset [38] with 32 indoor images including paintings, toys, food, and so on, captured under controlled illumination, and the Harvard dataset [9] with 50 indoor and outdoor images recorded under daylight illumination. The dimensions of the images from the CAVE dataset are 512×512 pixels, with 31 spectral bands of 10 nm wide, covering the visible spectrum from 400 to 700 nm; the images from the Harvard dataset have the dimensions of 1392×1040 pixels with 31 spectral bands of width 10 nm, ranging from 420 to 720 nm, from which we extract the top left 1024×1024 pixels in our experiments.

We treat the original images in the databases as ground truth \mathbf{Z} , and simulate to produce the observed HR-RGB images \mathbf{Y} by integrating the ground truth over the spectral channels using the spectral response \mathbf{R} of a Nikon D7006 camera. In the HSI SR scenario using CAVE and Harvard datasets, most work attempted to recover the HR-HS image from observed LR-HS and HR-RGB images, and the used expanding factor of spatial resolution is generally set as 32 for horizontal and vertical directions, respectively. In order to validate the recovery performance of the HR-HS image from LR-HS image with the usually used spatial expanding factors, we also conducted the experiments using spatial-CNN with the same unsampling network structure as shown in Fig. 1. The input LR-HS images in the Spatial-CNN are generated via down-sampling the groundtruth with a factor of 32 to create 16×16 images, which is implemented by averaging over 32×32 pixel blocks as done in [21], [4].

For preparing samples for CNN training, we extract 15×15 overlapped patches as input data from the input HR-RGB, and the corresponding center patches with size $11 \times 11 \times L$ as output from the ground-truth HR-HS images of training dataset. The simple Euclidean distance between the estimated output and the ground-truth patches is minimized to learn the HSRCNN and Residual HSRCNN network parameters. Our network, implemented with Caffe [20], is trained from scratch, using the SGD optimizer. We use a minibatch size of 128 in training

Table I
THE AVERAGED RMSE, PSNR, SAM AND ERGAS OF SPATIAL-CNN AND SPECTRAL-CNNs: HSRCNN, RESIDUAL HSRCNN ON CAVE DATASET.

	RMSE	PSNR	SAM	ERGAS
Spatial-CNN	28.41	19.66	19.77	2.98
HSRCNN	5.27	34.31	7.81	0.53
Residual HSRCNN	4.78	35.07	7.37	0.49

Table II
THE AVERAGED RMSE, PSNR, SAM AND ERGAS OF HSRCNN, RESIDUAL HSRCNN ON HARVARD DATASET.

	RMSE	PSNR	SAM	ERGAS
HSRCNN	3.48	38.72	5.37	0.49
Residual HSRCNN	3.33	38.93	5.31	0.35

procedure, and train the network for 3000000 iterations with the fixed learning rate 0.0001. Our model parameters are initialized according to Gaussian distribution with standard deviation 0.001.

We have randomly selected 20 HSIs from CAVE database to train HSRCNN and Residual HSRCNN models, and the remainder are used for validation of the performance of the proposed HSRCNN method. For Harvard database, 10 HSIs have been randomly selected for CNN model training, and the remainder 40 HSIs are as test for validation. To evaluate the quantitative accuracy of the estimated HS images, four objective error metrics including root-mean-square error (RMSE), peak-signal-to-noise ratio (PSNR), relative dimensionless global error in synthesis (ERGAS) [34], and spectral angle mapper (SAM) [35] are used. The metrics: (ERGAS) [34] calculates the average amount of specific spectral distortion normalized by intensity mean in each band as defined below:

$$\text{ERGAS} = 100 \times \frac{N}{M} \sqrt{\frac{1}{L} \sum_{l=1}^L \frac{\text{RMSE}(i)}{\mu_i}} \quad (4)$$

where $\frac{N}{M}$ is the ratio between the pixel sizes of the available HR-RGB and LR-HS images, μ_i is the intensity mean of the i -th band of the LR-HS image, and L is the number of LR-HS bands. The smaller ERGAS, the smaller the spectral distortion. The SAM [35] measures the spectral distortion between the ground-truth and estimated HR-HS images, and the distortion of two spectral vectors \mathbf{z}_n and $\hat{\mathbf{z}}_n$ is defined as following:

$$\text{SAM}(\mathbf{z}_n, \hat{\mathbf{z}}_n) = \arccos \left(\frac{\langle \mathbf{z}_n, \hat{\mathbf{z}}_n \rangle}{\|\mathbf{z}_n\|_2 \cdot \|\hat{\mathbf{z}}_n\|_2} \right) \quad (5)$$

The overall SAM is finally obtained by averaging the SAMs computed from all image pixels. Note that the value of SAM is expressed in degrees and thus belongs to $(-90, 90]$. The smaller the absolute value of SAM, the less important the spectral distortion.

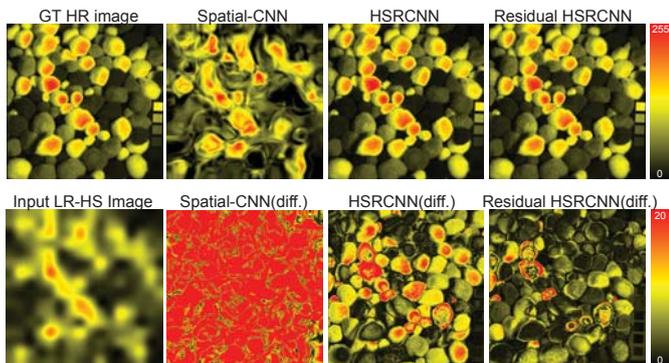


Figure 2. The ‘pompons’ image example from the CAVE database. The first row shows the ground-truth HR image and the recovered images by Spatial-CNN, HSRCNN and the proposed Residual HSRCNN, respectively. The second row gives the corresponding input band image in the LR-HS image for Spatial-CNN, the absolute difference images between the ground-truth image and the recovered HR-HS images in the first row.

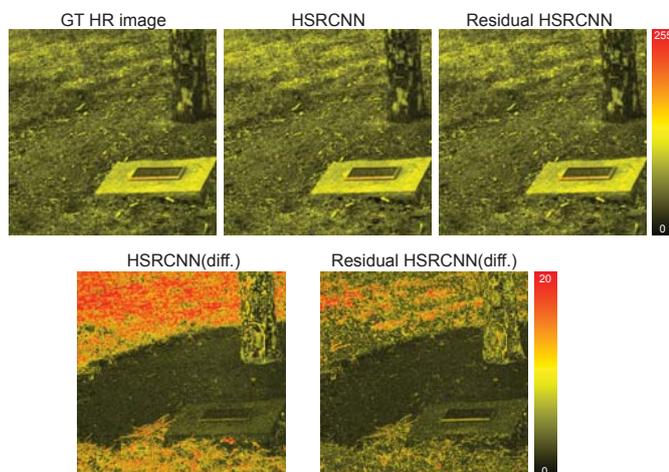


Figure 3. The ‘imgb6’ image example from the CAVE database. The first row shows the ground-truth HR image and the recovered images by HSRCNN and the proposed Residual HSRCNN, respectively. The second row gives the absolute difference images between the ground-truth image and the recovered HR-HS images in the first row.

a) Comparison of CNN resolution enhancement models in different domains: As we introduced in section III, the CNN based method can be used for recovering the HR-HS image from either of the available LR-HS, HR-RGB images, which are named as Spatial-CNN, Spectral-CNN (HSRCNN) and an extended residual estimation version of HSRCNN: Residual HSRCNN. We conducted experiments using the Spatial-CNN for spatial resolution enhancement with 32 expanding factor for horizontal and vertical directions, HSRCNN and the proposed Residual HSRCNN for spectral resolution enhancement with about 10 expanding factor (from RGB: 3 to 31-bands spectra), and calculate the average values of the evaluation metrics; RMSE, PSNR, SAM and ERGAS of the 12 test images in CAVE database. The compared results are shown in Table I, which manifests much better results of the Spectral-CNN than Spatial-CNN due to the smaller

Table III
THE COMPARED PERFORMANCE OF OUR PROPOSED RESIDUAL HSRCNN AND THE STATE-OF-THE-ART METHOD ON CAVE DATASET.

	SP-based	Shallow A+	VDCNN	Our
RMSE	5.4	6.70	4.76	4.78
SAM	N.A.	N.A.	12.10	7.37

expanding factor in spectral-domain (about 10 from 3 to 31) than spatial domain (32 from 16 to 512 for horizontal and vertical directions, respectively), and further performance improvement using the proposed Residual HSRCNN model. One recovered HS image example and the corresponding residual images with the ground-truth HR images from CAVE database is visualized in Fig. 2 using different CNN models.

Since the HS image recovery performance with Spectral-CNN (HSRCNN and the Residual HSRCNN) can be significantly improved compared Spatial-CNN model, for Harvard dataset we only provide the compared results with the CNN model for spectral enhancement in Table II. From Table II, it can be seen that the proposed Residual HSRCNN provides better performances than the conventional spectral-resolution enhancement method: HSRCNN for all four evaluation metrics. One recovered HS image example and the corresponding residual images with the ground-truth HR image from Harvard database is visualized in Fig. 3 using HSRCNN and Residual HSRCNN models.

b) Comparison with the state-of-the-art spectral-resolution enhancement methods: This section provides the compared results with the recent state-of-the-art HSI SR methods for spectral resolution enhancement. Several work for spectral resolution enhancement from an RGB image have been explored and mainly include the sparse-promoting method (SP-based) [6], the shallow A+ method [1] and the very-deep CNN-based model (VDCNN) [16]. The state-of-the-art studies have usually evaluated the HS recovery performance with the RMSE metric and VDCNN [14] also provided the SAM results on CAVE dataset. Table III manifests the compared results using our methods and the recently published work on CAVE dataset. Our proposed Residual HSRCNN model can over-perform the other methods and give the comparable RMSE value with VDCNN [16], but can greatly reduce the spectral distortion compared with VDCNN method [16]. In addition, the residual component estimation strategy can also be applied to more deep CNN architecture, and is prospected to further improve HS recovery performance.

V. CONCLUSION

This study proposed to estimate a hyper-spectral image via enhancing spectral resolution from an RGB image obtained using a commodity camera. Motivated by the success of deep convolutional neural network for spatial resolution enhancement of natural images, we explored a spectral reconstruction CNN for spectral super-resolution with an available RGB image, which predicts the high-frequency content of the fine

spectral wavelength in narrow band interval via learning the hyper-spectral structure of different materials in visual world, called as hyper-spectral reconstruction CNN. Since the lost high-frequency content with the HSRCNN can not be perfectly recovered, we overwrite the same CNN architecture on the baseline CNN to further estimate the non-recovered high-frequency content (residual) from the output of the baseline CNN, and thus produce a non-recovered residual image (high-frequency content). This leads to a novel residual hyper-spectral reconstruction CNN framework to recover more accurate HS images, which is called Residual HSRCNN. Experiments on benchmark hyper-spectral datasets validated that the proposed method achieved promising performances compared with the existing state-of-the-art methods.

VI. ACKNOWLEDGMENT

This research is supported by the open collaborative research program at National Institute of Informatics (NII) Japan (FY2018).

REFERENCES

- [1] J. Aeschbacher, J. Wu, and R. Timofte. In defense of shallow learned spectral reconstruction from rgb images. *In Workshop of The IEEE International Conference on Computer Vision (ICCV)*, 2017.
- [2] B. Aiazzi, S. Baronti, F. Lotti, and M. Selva. A comparison between global and context-adaptive pansharpening of multispectral images. *IEEE Geosci. Remote Sens. Lett.*, 6(2):302–306, 2009.
- [3] N. Akhtar, F. Shafait, and M. A. Sungp. A greedy sparse approximation algorithm for hyperspectral unmixing. *ICPR*, pages 3726–3731, 2014.
- [4] N. Akhtar, F. Shafait, and A. Mian. Sparse spatio-spectral representation for hyperspectral image super-resolution. *ECCV*, pages 63–78, 2014.
- [5] A. Alvarez-Gila, J. van de Weijer, and E. Garrote. Adversarial networks for spatial context-aware spectral image reconstruction from rgb. *IEEE International Conference on Computer Vision Workshop (ICCVW 2017)*, 2017.
- [6] B. Arad and O. Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. *In European Conference on Computer Vision*, pages 19–34, 2016.
- [7] J. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci. Remote Sens. Mag.*, 1(2):6–36, 2013.
- [8] M. Cetin and N. Musaoglu. Merging hyperspectral and panchromatic image data: Qualitative and quantitative analysis. *Int. J. Remote Sens.*, 30(7):1779–1804, 2009.
- [9] A. Chakrabarti and T. Zickler. Statistics of real-world hyperspectral images. *CVPR*, pages 193–200, 2011.
- [10] H. Chang, D. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. *CVPR*, 2004.
- [11] P. Chavez, S. Sides, and J. Anderson. Comparison of three different methods to merge multiresolution and multispectral data: Landsat tm and spot panchromatic. *Photogramm. Eng. Rem. S.*, 30(7):1779–1804, 1991.
- [12] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 38(2):295–307, 2015.
- [13] C. Dong, C. C. Loy, and X. Tang. Accelerating the super-resolution convolutional neural network. *Proceedings of European Conference on Computer Vision (ECCV)*, 2016.
- [14] 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 Transaction on Image Processing*, 25(3):2337–2352, 2016.
- [15] M. Fauvel, Y. Tarabalka, J. Benediktsson, J. Chanussot, and J. Tilton. Advances in spectral-spatial classification of hyperspectral images. *Proc. IEEE*, 101(3):652–675, 2013.
- [16] S. Galliani, C. Lanaras, D. Marmanis, E. Baltsavias, and K. Schindler. Learned spectral super-resolution. *arXiv preprint arXiv:1703.09470*, 2017.
- [17] C. Grohnfeldt, X. X. Zhu, and R. Bamler. Jointly sparse fusion of hyperspectral and multispectral imagery. *IGARSS*, 2013.
- [18] R. Haydn, G. Dalke, J. Henkel, and J. Bare. Application of the ihs color transform to the processing of multisensor data and image enhancement. *Int. Symp on Remote Sens. of Env.*, 1982.
- [19] B. Huang, H. Song, H. Cui, J. Peng, and Z. Xu. Spatial and spectral image fusion using sparse matrix factorization. *IEEE Trans Geosci. Remote Sens.*, 52(3):1693–1704, 2014.
- [20] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*, 2014.
- [21] R. Kawakami, J. Wright, Y.-W. Tai, Y. Matsushita, M. Ben-Ezra, and K. Ikeuchi. High-resolution hyperspectral imaging via matrix factorization. *CVPR*, pages 2329–2336, 2011.
- [22] J. Kim, J. K. Lee, and K. M. Lee. Accurate image super-resolution using very deep convolutional networks. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [23] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. *Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [24] H. Li. Deep learning for image denoising. *International Journal Signal Processing, Image Processing and Pattern Recognition*, 7(3):171–180, 2014.
- [25] Y. Li, J. Hua, X. Zhao, W. Xie, and J. Li. Hyperspectral image super-resolution using deep convolutional neural network. *Neurocomputing*, 266:29–41, 2017.
- [26] A. Minghelli-Roman, L. Polidori, S. Mathieu-Blanc, L. Loubersac, and F. Cauneau. Spatial resolution improvement by merging meris-etm images for coastal water monitoring. *IEEE Geosci. Remote Sens. Lett.*, 3(2):227–231, 2006.
- [27] H. Nguyen, A. Benerjee, and R. Chellappa. Tracking via object reflectance using a hyperspectral video camera. *CVPRW*, pages 44–51, 2010.
- [28] R. M. Nguyen, D. K. Prasad, and M. S. Brown. Training-based spectral reconstruction from a single rgb image. *In European Conference on Computer Vision*, pages 186–201, 2014.
- [29] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.
- [30] Y. Sun, X. Wang, and X. Tang. Deep learning face representation by joint identification-verification. *Adv. Neural Inf. Process. Syst.*, 27:1988–1996, 2014.
- [31] C. Szegedy, S. Reed, D. Erhan, D. Anguelov, and S. Ioffe. Scalable, high-quality object detection. *arXiv: 1412.1441v3*, pages 1–10, 2015.
- [32] Y. Tarabalka, J. Chanussot, and J. Benediktsson. Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers. *IEEE Trans. Syst., Man, Cybern., Syst.*, 40(5):1267–1279, 2010.
- [33] M. Uzair, A. Mahmood, and A. Mian. Hyperspectral face recognition using 3d-dct and partial least squares. *BMVC*, pages 57.1–57.10, 2013.
- [34] L. Wald. Quality of high resolution synthesized images: Is there a simple criterion? *Proc. of Fusion Earth Data*, pages 99–103, 2000.
- [35] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J. Toureret. Hyperspectral and multispectral image fusion based on a sparse representation. *IEEE Trans Geosci. Remote Sens.*, 53(7):3658–3668, 2015.
- [36] Z. Xiong, Z. Shi, H. Li, L. Wang, D. Liu, and F. Wu. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. *The IEEE International Conference on Computer Vision (ICCV)*, pages 518–525, 2017.
- [37] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution as sparse representation of raw image patches. *CVPR*, pages 1–8, 2008.
- [38] F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar. Generalized assorted pixel camera: Post-capture control of resolution, dynamic range and spectrum. *IEEE Transaction on Image Processing*, 19(9):2241–2253, 2010.
- [39] D. Zhang, W. Zuo, and F. Yue. A comparative study of palmprint recognition algorithm. *ACM Comput. Surv.*, 44(1):2:1–2:37, 2012.
- [40] Y. Zhou, H. Chang, K. Barner, P. Spellman, and B. Parvin. Classification of histology sections via multispectral convolutional sparse coding. *CVPR*, pages 3081–3088, 2014.
- [41] R. Zurita-Milla, J. Clevers, and M. E. Schaepman. Unmixing-based landsat tm and meris fr data fusion. *IEEE Geosci. Remote Sens. Lett.*, 5(3):453–457, 2008.