Zero-Shot Hyperspectral Image Denoising With Separable Image Prior

Ryuji Imamura
z8mca005@eng.kitakyu-u.ac.jp

Tatsuki Itasaka
a9mca003@eng.kitakyu-u.ac.jp

Masahiro Okuda
okuda-m@kitakyu-u.ac.jp

The University of Kitakyushu

Abstract

Supervised learning with a convolutional neural network is recognized as a powerful means of image restoration. However, most such methods have been designed for application to grayscale and/or color images; therefore, they have limited success when applied to hyperspectral image restoration. This is partially owing to large datasets being difficult to collect, and also the heavy computational load associated with the restoration of an image with many spectral bands. To address this difficulty, we propose a novel self-supervised learning strategy for application to hyperspectral image restoration. Our method automatically creates a training dataset from a single degraded image and trains a denoising network without any clear images. Another notable feature of our method is the use of a separable convolutional layer. We undertake experiments to prove that a separable network allows us to acquire the prior of a hyperspectral image and to realize efficient restoration. We demonstrate the validity of our method through extensive experiments and show that our method has better characteristics than those that are currently regarded as state-of-the-art.

1. Introduction

Hyperspectral image (HSI) restoration, whereby a degraded image is processed to produce a clear image, is an essential technique in HSI applications [3, 13, 14, 16] given that hyperspectral images are prone to being adversely affected by noise, which prevents the precise extraction of useful information in applications such as classification, unmixing, and target detection [4, 18, 23]. Various HSI restoration methods have been proposed. Optimization-based methods with image priors have constituted one of the mainstreams in the HSI reconstruction, and have been the subject of considerable research. Some methods have exploited the sparsity, low-rank properties, and non-local correlation of HSIs to restore images by solving optimization problems based on the priors [1, 8, 20, 21, 22]. Most of these methods can be regarded as convex optimization problems and can produce clear images as global optimums.

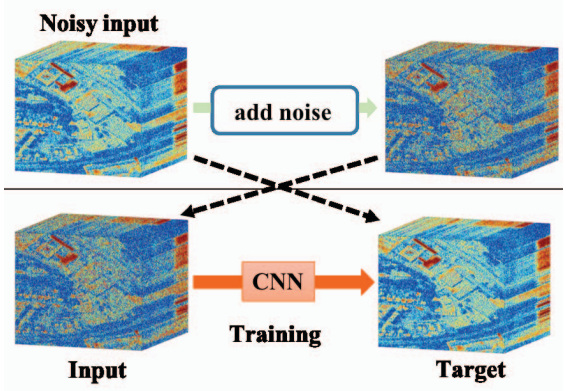


Figure 1. Self-supervised restoration: (top) We add noise to an noisy input. (bottom) We train CNN by setting the input as a target and the image with the added noise as an input for CNN.

Aside from the methods that rely on designed priors, deep learning has been successfully applied to various computer vision and image processing tasks, while image restoration based on deep neural networks (DNNs) has recently been the subject of research. It has been shown that restoration with a high degree of precision can be attained [5, 6, 15, 24] through the use of large supervised datasets. Although deep learning becomes integrated into image-restoration tasks, the high performance is limited to the applications of grayscale/color images. This is a result of it being extremely difficult to collect large datasets required for HSIs for supervised training, because special HSI sensors are required, and it often takes much time and effort to capture an HSI.

Another problem is that, to fully exploit the features of an HSI through the application of deep learning, a large-scale DNN is required. An HSI typically has more than a hundred spectral bands and, thus, to exploit the correlation between an HSI in the spectral domain, we need intermediate feature maps with many more channels in the DNN. For example, DnCNN [24] uses 17–20 layers with 64 filter kernels in each layer to train the denoising network to handle

color images. Therefore, based on the DnCNN settings, it would be preferable to use $64 \times N/3$ filters in each layer to capture the spectral information of an HSI with N spectral bands. Therefore, any increase in the number of parameters makes it difficult to train a network.

To circumvent these difficulties, we propose a novel DNN-based learning scheme for HSI restoration by introducing a new self-supervising strategy. Our method creates input/target patches from a single deteriorated image and trains denoising networks without any clear images (Figure 1). Another feature of our method is that, considering the anisotropic property of HSIs, we can utilize a separable convolutional network. We show that the proposed network with the separable convolution efficiently captures an HSI prior, through experiments involving some image-restoration tasks. To the best of our knowledge, this is the first attempt to apply self-supervised learning to HSI restoration with DNNs. The results of our experiments showed that our method is comparable to or outperforms conventional methods based on model-based methods.

2. Separable CNN Capturing Structural Prior

2.1. Architecture

A non-separable CNN with dense coefficients is inherently better able to capture the features of images than a separable CNN with the same number of layers, but it is often hard to train it for an image cube with many channels like a HSI, because we may need more kernels for each layer to capture the features in the spectral direction. This would make its optimization very difficult.

On the other hand, it is well known that a measured spectrum in a hyperspectral pixel can be approximated by a small fraction of spectral signatures, called endmembers. In an aerial scene, endmembers correspond to familiar macroscopic materials. In general, endmembers in a neighbor tends to be highly correlated, which leads to the low-rankness of the HSI, especially in the spectral direction. Considering the property of HSIs, we can safely ignore diagonal correlations in the spatial-spectral plains because less information is carried by the neighboring pixels in the diagonal direction, pointing to the advantage of using a separable network.

In this section, we introduce the architecture with the separable CNN and examine, through self-supervised image restoration, that the separable architecture not only reduces the training complexity but also encloses the structural advantage to represent an HSI.

Figure 2 illustrates our network that includes separable convolutional layers, batch normalization units, and ReLU. Separable convolution [7] was originally introduced to reduce the computational complexity of CNN by approximating 3D convolution to separate convolutions in the spatial and channel dimensions. Convolution in the spatial dimension is performed independently in each channel (depth-wise convolution), while convolution in the channel direction is performed using a 1D kernel (point-wise convolution).

2.2. Performance of Separable CNN

We now demonstrate that a network with a separable convolution can efficiently capture the structural prior of the HSIs by considering a simple example of image hole-filling. Let the linear operator $\mathbf{A} \in \{0, 1\}^{N \times N}$ denote a known random-sampling matrix, where N is the total number of pixels in an HSI. Given only the pixels corresponding to the sampling points of $\mathbf{A}$, we consider the problem of recovering an entire image $\mathbf{x}$.

Here, we train a network by minimizing the mean squared error only for those pixels masked by $\mathbf{A}$, that is,

$$\min_{\Phi} \|\mathbf{A}\mathbf{x} - \mathbf{A}\Phi(\mathbf{A}\mathbf{x})\|_2^2, \tag{1}$$

where $\mathbf{A}\mathbf{x}$ is a set of measured pixels, and we denote a neural network by Φ. The aim of the task is to estimate an original image $\mathbf{x}$.

We trained two networks by minimizing (1). One of these was a non-separable CNN with tightly coupled coefficients. It had eight layers, with each layer having $3 \times 3 \times M \times L$ parameters, where M, L are the channel lengths of the input and output of a convolution, respectively. The other was a separable CNN like that shown in Figure 2. This also has eight layers, with each layer having $3 \times 3 \times 1 \times M$ parameters for the depth-wise convolution and $1 \times 1 \times M \times L$ parameters for the point-wise convolution. In the experiment, we divided an HSI into $20 \times 20 \times \{\text{number of spectral bands}\}$ blocks and used them as inputs. The training was carried out by carefully tuning the hyper-parameters to attain the best performance. Note that this training with (1) differs from DIP in that we used the image as is, while DIP starts with random samples to generate a clear image.

Figure 3 illustrates an example of the hole-filling task. Figure 3 (a) and (b) show the original image and an observation, respectively. The results obtained with the non-separable and separable CNN are shown in Figure 3 (c) and (d), respectively. Both of the networks interpolates the holes even though the cost is evaluated only on the sampling points indicated by $\mathbf{A}$ by virtue of smoothing capability of the CNNs. It should be noted that even though the separable CNN has the same number of layers and fewer parameters, the performance is superior to that of the non-separable CNN.

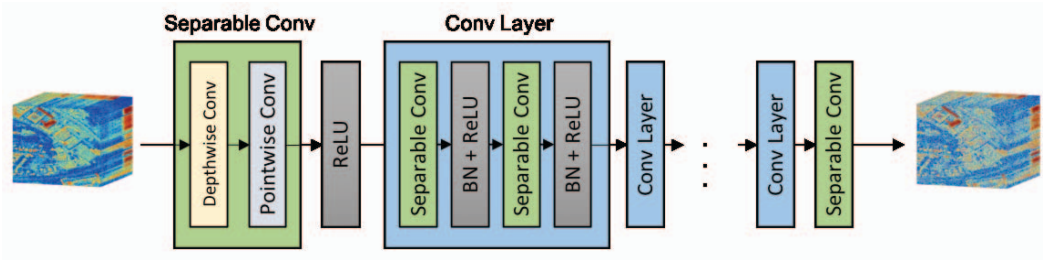
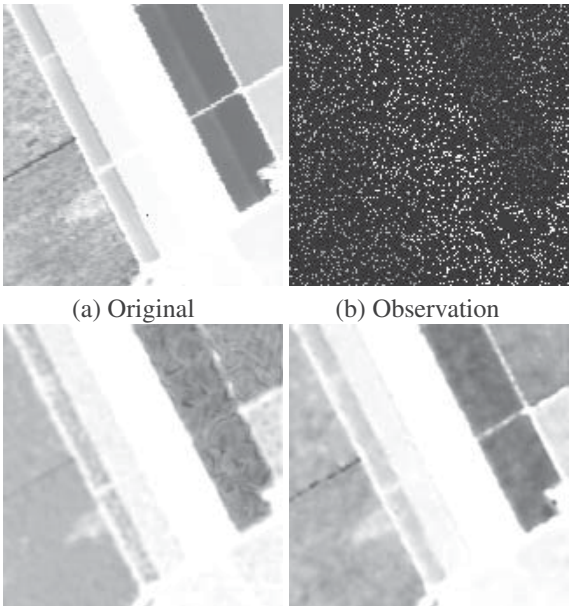


Figure 2. Architecture: We use a simple separable CNN composed of depth-wise and point-wise convolutional layers. We use M kernels for a single depth-wise layer, and its M outputs are concatenated with respect to the channel direction.



(a) Original (b) Observation

(c) PSNR: 37.88 (d) PSNR: 39.13

Figure 3. Self-supervised hole-filling: (a) original image, (b) observation, (c) result obtained with dense network, (d) results obtained with separable network (PSNR values indicate the mean of PSNR for all bands)

3. Self-supervised Training for Gaussian Denoising

A CNN, composed of linear filters, is inherently able to represent a signal with a high correlation. Ulyanov et al. [19] exploited this property using Deep Image prior (DIP), proposing a generative network that maps a random noise vector to a natural image. They demonstrated that the structure of the CNN itself encloses the statistical prior of the images and exhibits excellent performance with some restoration tasks, such as denoising, super-resolution, and inpainting. Lehtinen et al. [11] used a noisy image pair to learn the "*noise to noise*" relationship between the two measurements, showing that it is possible to produce a clear image using the smoothing property of the CNNs. They experimentally demonstrated that a clear image, which is often hard to obtain, is not needed to train a denoising network. Krull et al. [10] built on this idea to develop self-supervised learning and proposed a method called Noise2Void (N2V), in which an input image is predicted using a convolutional kernel without a center coefficient. DIP and N2V do not require us to make any assumptions for the degradation process of an input image. However, the use of DIP often gives rise to over-smoothing, especially when applied to the restoration of severely degraded images, while the performance of N2V becomes worsen when an image does not have a high correlation in a local region, such as images with many holes. Shocher et al. [17] proposed another approach to self-supervised learning for image super-resolution. They assumed that the statistical properties of an image do not change with the scale, and that the relationship between an original image and its downsized version is learned using a smaller image pair that is made based on the measurements.

Our work was inspired by these methods. In the next section, we propose a novel self-supervised strategy and introduce a separable convolutional network for HSI restoration. We show that the separable structure is highly suited to the HSI restoration tasks, achieving high-fidelity restoration by our self-supervised learning.

We consider a degradation model with additive noise:

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \tag{2}$$

where $\mathbf{n}$ denotes the Gaussian noise with a standard deviation σ. Our goal is to estimate a latent clear image $\mathbf{x}$, given a degraded observation $\mathbf{y}$.

Our denoising method builds upon the fact that an HSI has rich information even when it is degraded by noise, and that the separable convolution captures the latent low-rank structure of an HSI. Our method involves finding a separable CNN that can remove noise by learning only with a degraded input. In the training step, we automatically create a target image and minimize the error between the input and the target (Figure 1).

We assume that a network that can recover a clear image $\mathbf{x}$ from an observation $\mathbf{y}$ also has the ability to recover $\mathbf{y}$ from $\mathbf{y} + \mathbf{n}$. Based on this assumption, we first estimate the

Table 1. Results for Gaussian noise removal

Methods	σ	PaviaC	PaviaU	Frisco	Stanford	IndianPine	Washington	Cupurite
BM4D [12]	0.05	38.54	38.53	40.41	40.74	38.01	39.59	38.14
	0.1	34.94	34.94	36.74	37.04	35.26	35.58	34.77
	0.15	32.83	32.71	34.75	35.04	33.69	33.41	32.99
	0.2	31.4	31.15	33.39	33.67	32.55	31.98	31.83
	0.25	30.35	30.00	32.33	32.65	31.65	30.89	31.00
FastHyDe [25]	0.05	**40.68**	**40.23**	**45.78**	**46.08**	38.96	**44.86**	39.57
	0.1	37.56	37.27	41.78	42.19	37.29	40.71	37.33
	0.15	35.76	35.35	39.57	40.20	36.04	38.38	36.18
	0.2	34.39	34.11	38.20	38.76	35.17	36.80	35.39
	0.25	33.46	33.05	37.13	37.39	34.48	35.71	34.81
DIP [19]	0.05	34.87	34.45	37.72	37.18	36.12	33.84	37.23
	0.1	32.74	32.85	33.82	32.43	31.65	30.56	36.90
	0.15	30.13	30.52	31.08	28.70	28.42	27.23	35.60
	0.2	27.02	27.91	27.61	25.65	25.99	24.71	33.79
	0.25	25.40	25.62	25.45	24.08	24.12	22.63	32.12
N2V [10]	0.05	32.32	31.32	35.57	37.20	33.42	35.30	33.68
	0.1	31.77	30.64	35.35	36.14	34.00	34.57	33.03
	0.15	31.38	30.36	35.26	35.57	33.81	33.78	32.72
	0.2	31.04	30.10	34.56	34.64	33.34	33.04	32.32
	0.25	30.73	29.71	34.06	34.11	32.90	32.73	32.06
Ours	0.05	38.84	38.66	42.62	45.78	**39.74**	44.09	**40.32**
	0.1	**37.60**	**37.65**	**41.81**	**42.29**	**37.66**	**41.43**	**38.94**
	0.15	**35.73**	**35.87**	**39.76**	**40.35**	**36.48**	**39.52**	**37.51**
	0.2	**34.87**	**34.81**	**38.72**	**39.02**	**35.55**	**37.85**	**36.72**
	0.25	**33.96**	**33.73**	**37.58**	**37.90**	**34.75**	**36.63**	**35.86**

standard deviation σ' of the noise from the noisy image $\mathbf{y}$ and further add noise with $\tilde{\sigma} = (1+\alpha)\sigma'$, $(\alpha << 1)$ to $\mathbf{y}$, as follows:

$$\tilde{\mathbf{y}} = \mathbf{y} + \tilde{\mathbf{n}}, \tag{3}$$

where $\tilde{\mathbf{n}} \sim \mathcal{N}(0,\, \tilde{\sigma}^2)$. We adopted the method proposed in [2] to estimate the standard deviation. Considering the estimation error, we randomly sample the value of α within a range of $[-0.1,\, 0.1]$ and created training sets $\{\mathbf{y}', \mathbf{y}\}$ during the training steps. We optimize the network with the training sets $\{\mathbf{y}', \mathbf{y}\}$ by minimizing a simple loss function with a ℓ_2 norm:

$$\min_{\Phi} \|\Phi(\tilde{\mathbf{y}}) - \mathbf{y}\|_2^2, \tag{4}$$

where Φ is the separable CNN shown in Figure 2. This strategy efficiently trains the network and yields sufficient performance, but we further enhance the performance by replacing the input image $\mathbf{y}$ with a restored image every a few hundred epochs.

4. Experiments

We collect a dataset of HSIs from web sites, which are commonly used for research purposes. The images have 100 to 225 spectral channels. We used four separable layers in the network. Each layer had 100 depth-wise kernels with a size of $(3 \times 3 \times 1)$, 400 point-wise kernels with a size of $(1 \times 1 \times 100)$, batch normalization, and ReLU. For training, data were augmented by rotating and flipping the input data and were then divided into $20 \times 20 \times \{\text{number of bands}\}$ blocks. We used a minibatch size of 32, and the network was trained using the Adam optimizer [9]. We started with a learning rate of 0.01, which was halved every dozens of epochs[1].

In our experiments, we added Gaussian noise with a standard deviation of $\sigma = [0.05, 0.1, 0.15, 0.2, 0.25]$ and evaluated the performance by averaging the PSNR for all the bands. We compared our method with the conventional model-based methods, namely, BM4D [12] and FastHyDe [25], and self-supervised DNN methods, that is, DIP [19] and N2V [10]. We used dense networks for DIP and N2V with the same number of layers shown in their papers, and in each layer we appropriately extended the number of channels. With our method, we take a noisy image and estimate the standard deviation of the noise from the input using the technique described in [2]. Using the estimated standard

[1]The code is available on https://github.com/separable-image-prior/self-supervised-hyperspectral-image-restoration

deviation, we further added Gaussian noise and used it as a target image, as described in Section 3.

Table 1 lists the average PSNR values for some images. From the table, we can confirm the superiority of the proposed method at high noise levels over FastHyDe, which is known as state-of-the-art in HSI denoising. When the noise level is low, FastHyDe significantly performs well. Focusing on the DNN-based method, the performance of DIP is inferior to the optimization-based method, but N2V shows the same performance as the optimization-based method at high noise levels.

Acknowledgement. This work was partially supported by the Ministry of Internal Affairs and Communications of Japan under the Strategic Information and Communications R&D Promotion Programme, No. 3620.

References

[1] H. K. Aggarwal and A. Majumdar. Hyperspectral image denoising using spatio-spectral total variation. *IEEE Geosci. Remote Sens. Lett.*, 13(3):442–446, 2016.

[2] J. M. Bioucas-Dias and J. M. Nascimento. Hyperspectral subspace identification. *IEEE Transactions on Geoscience and Remote Sensing*, 46(8):2435–2445, 2008.

[3] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and remote sensing magazine*, 1(2):6–36, 2013.

[4] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE journal of selected topics in applied earth observations and remote sensing*, 5(2):354–379, 2012.

[5] H. C. Burger, C. J. Schuler, and S. Harmeling. Image denoising: Can plain neural networks compete with bm3d? In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2392–2399. IEEE, 2012.

[6] Y. Chen and T. Pock. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE transactions on pattern analysis and machine intelligence*, 39(6):1256–1272, 2017.

[7] F. Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[8] W. He, H. Zhang, L. Zhang, and H. Shen. Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration. *IEEE Trans. Geosci. Remote Sens.*, 54(1):178–188, 2016.

[9] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[10] A. Krull, T.-O. Buchholz, and F. Jug. Noise2void-learning denoising from single noisy images. *arXiv preprint arXiv:1811.10980*, 2018.

[11] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila. Noise2noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189*, 2018.

[12] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE Trans. Image Process.*, 22(1):119–133, 2013.

[13] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, et al. Recent advances in techniques for hyperspectral image processing. *Remote sensing of environment*, 113:S110–S122, 2009.

[14] M. Rizkinia, T. Baba, K. Shirai, and M. Okuda. Local spectral component decomposition for multi-channel image denoising. *IEEE Trans. Image Process.*, 25, Issue:7:3208–3218, July 2016.

[15] U. Schmidt and S. Roth. Shrinkage fields for effective image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2774–2781, 2014.

[16] G. Shaw and D. Manolakis. Signal processing for hyperspectral image exploitation. *IEEE Signal Process. Magazine*, 19(1):12–16, 2002.

[17] A. Shocher, N. Cohen, and M. Irani. “zero-shot” super-resolution using deep internal learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3118–3126, 2018.

[18] D. W. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker. Anomaly detection from hyperspectral imagery. *IEEE signal processing magazine*, 19(1):58–69, 2002.

[19] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018.

[20] Y. Wang, J. Peng, Q. Zhao, Y. Leung, X.-L. Zhao, and D. Meng. Hyperspectral image restoration via total variation regularized low-rank tensor decomposition. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(4):1227–1243, 2018.

[21] Q. Yuan, L. Zhang, and H. Shen. Hyperspectral image denoising employing a spectral-spatial adaptive total variation model. *IEEE Trans. Geosci. Remote Sens.*, 50(10):3660–3677, 2012.

[22] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan. Hyperspectral image restoration using low-rank matrix recovery. *IEEE Trans. Geosci. Remote Sens.*, 52(8):4729–4743, 2014.

[23] H. Zhang, J. Li, Y. Huang, and L. Zhang. A nonlocal weighted joint sparse representation classification method for hyperspectral imagery. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.*, 7(6):2056–2065, 2014.

[24] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.

[25] L. Zhuang and J. M. Bioucas-Dias. Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(3):730–742, 2018.