IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 July 2023, accepted 20 July 2023, date of publication 24 July 2023, date of current version 2 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3298447

RESEARCH ARTICLE

Zero-Shot Hyperspectral Image Denoising Using Self-Completion With 3D Random Patterned Masks

TATSUKI ITASAKA, (Student Member, IEEE), AND MASAHIRO OKUDA, (Senior Member, IEEE)
Faculty of Science and Engineering, Doshisha University, Kyotanabe, Kyoto 610-0394, Japan
Corresponding author: Tatsuki Itasaka (cyjg1101@mail4.doshisha.ac.jp)

ABSTRACT Hyperspectral images (HSIs) have higher spectral resolution than RGB images and are used in various tasks. However, HSIs are prone to degradation due to noise generated during imaging, making it difficult to obtain non-degraded images. Additionally, supervised learning, which relies on pairs of degraded and non-degraded images, is often challenging to apply to HSI restoration because of the high cost of imaging and the need to prepare large amounts of data. To overcome these limitations, recent advances in self-supervised learning have led to the development of learning-based image restoration methods that do not require non-degraded images. However, these methods have limitations, including low accuracy and the need to estimate the noise distribution. In this paper, we propose a zero-shot HSI deep denoising method based on self-supervised image restoration. The proposed method achieves zero-shot recovery by repeatedly predicting blind-spots in 3D blocks during the learning process. Notably, our method does not require training or clean images, nor does it rely on noise distribution information. Numerical experiments and ablation studies confirmed that the restoration accuracy of the proposed method is comparable to or better than that of conventional zero-shot methods.

INDEX TERMS Hyperspectral image, denoising, zero-shot learning, self-supervised learning.

I. INTRODUCTION

Hyperspectral images (HSIs) are used for various tasks including remote sensing [1], [2], [3], [4], computer vision [5], [6], and medical diagnosis [7], [8]. However, noise degradation often occurs during HSI capture, making denoising essential for practical application in these tasks.

Image restoration methods can be broadly classified into optimization-based and model-based approaches. Owing to the challenges involved in obtaining non-degraded HSIs and the high cost of imaging, it is often difficult to create training datasets that include pairs of degraded and non-degraded HSIs, which are required for general image restoration using supervised learning. Therefore, model-based methods that use prior information as a regularization term, or patch-based methods that use self-similarity, are often used instead of learning-based methods [9], [10].

Recent advancements in self-supervised learning have led to the development of several self-supervised RGB image restoration methods [11], [12], [13]. These methods rely solely on degraded training images and do not require non-degraded RGB images, as they learn from the degraded images during training. For example, Deep Image Prior [11] restores images by learning a transformation from randomly sampled noise to a single observed degraded image. This method exploits the CNN's bias to preferentially learn smooth components to obtain a restored image by stopping the training early before restoring the noise. However, determining the timing of the early stopping as a hyperparameter in advance is not an easy task.

Noise2Void [12] addresses the limitation of Deep Image Prior, which requires the early stopping timing to be determined in advance. This method uses a blind-spot strategy

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Liu.

in which some pixel values are missing and the missing pixel values are predicted from the surrounding pixel values. This method employs a blind-spot strategy in which randomly selected pixel values are replaced with the surrounding pixel values, and the missing pixel values are predicted to restore the original image. This approach learns from a large dataset of degraded images to obtain an image restoration model. The recently proposed RGB image restoration method Blind2Unblind [13] is based on a blind-spot strategy. The method learns from the output obtained by this strategy, which causes the pixel values to be missing along specific patterns, and the output is obtained by inputting the degraded image into the model. Similar to Noise2Void, this method learns from a large dataset of degraded images. Although there are many conventional self-supervised learning methods for RGB images, these are difficult to be applied to HSIs. This is because the spectral resolution of HSI cameras varies based on sensor characteristics, and it is prohibitively costly to prepare a substantial number of degraded HSIs with matching spectral conditions. In addition, the direct extensions of zero-shot or few-shot RGB image restoration methods do not take into account spectral information specific to HSIs well, resulting in poor performance.

To address this problem, several zero-shot HSI restoration methods have been proposed that learn only from an observed degraded HSI. Examples include Deep HS Prior [14], which is an extension of Deep Image Prior that can be applied to HSI, and a method proposed by Imamura [15] et al., which takes as input a further degraded form of an original degraded HSI and learns using the original degraded HSI as the target data. However, these methods require early stopping and prior information on the degradation process, respectively. Deep HS Prior has a similar learning strategy to Deep Image Prior and requires early stopping. The method of Imamura et al. has a drawback in that the performance improvement of the model is contingent upon the specific degradation process employed during the learning phase. If the degradation process applied to the non-degraded HSI differs significantly from the degradation process used during the training phase, it becomes hard to achieve satisfactory restoration performance. In addition to these methods, several approaches combining model-driven and data-driven methods have been proposed for HSI restoration [16], [17], [18], [19].

This paper presents a deep-learning-based approach for zero-shot HSI denoising, which is useful when few well-conditioned degraded HSIs are available, and non-degraded HSIs are unavailable. The conventional methods have the limitations, such as the need for early stopping in Deep HS Prior and the reliance on the specific degradation process employed in Imamura et al.'s approach. The proposed method overcomes these limitations by introducing a blind-spot strategy using a 3D patterned mask with randomly varying combinations of blind pixels. Our method has an advantage in that it does not require any training or clean images, nor does it require noise distribution information. The blind-spot strategy bypasses learning identity mapping and remains independent from information about the degradation process. The use of 3D patterned masks enhances restoration accuracy by incorporating information from both the spatial and spectral directions. Randomizing blind pixel combinations improves generalization performance by augmenting the training data. We demonstrate that our method achieves higher restoration accuracy than conventional methods for certain noise types and effectively suppresses the overfitting during zero-shot training.

II. SELF-SUPERVISED IMAGE RESTORATION

A. PROBLEM SETTING

The image degradation process due to noise is generally expressed as follows:

$$\mathbf{y} = \mathbf{x} + \mathbf{n}. \tag{1}$$

Here, $\mathbf{x}$ is a non-degraded image, $\mathbf{n}$ is the noise, and $\mathbf{y}$ is an observed degraded image. The objective of image restoration is to obtain an estimated image $\hat{\mathbf{x}}$ that closely resembles the non-degraded image $\mathbf{x}$ from the observed degraded image $\mathbf{y}$. In image restoration using supervised learning, the weights Θ of the deep learning model $f_\Theta(\cdot)$ are adjusted so that $\mathbf{x}_n = f_\Theta(\mathbf{y}_n)$, using a large training dataset $\mathbf{x}_n, \mathbf{y}_n (n = 1 \ldots N)$ consisting of pairs of non-degraded and degraded images. Conventional self-supervised learning methods [12], [13] learn from a large number of degraded images $\mathbf{y}_n (n = 1, \ldots, N)$ and then perform inference on another test image to be denoised. In contrast, the zero-shot learning approach described in this paper relies on a single input image for both training and testing, without the need for additional training data. Thus, the same input image $\mathbf{y}$ is used throughout the denoising process.

B. CONVENTIONAL SELF-SUPERVISED LEARNING METHODS

Recently, several self-supervised and zero-shot RGB image restoration methods have been proposed [11], [12], [13]. In Deep Image Prior [11], the input image is a randomly sampled noisy image $\mathbf{n}_{\text{in}}$, and the model used for image restoration is obtained by minimizing the following loss function:

$$\mathbf{L}(\mathbf{y}, \mathbf{n}_{\text{in}}) = \frac{1}{2} \|\mathbf{y} - f_\Theta(\mathbf{n}_{\text{in}})\|_2^2. \tag{2}$$

The CNN has an inherent bias towards learning low-frequency components first. As noise primarily consists of high-frequency components, early stopping halts learning before these high-frequency components are restored, meaning that they cannot be recovered through the denoising process, thus obtaining a denoised image. However, the timing of early stopping must be predetermined to prevent the deep learning model $f_\Theta(\cdot)$ from simply learning to map the input $\mathbf{n}_{\text{in}}$ to $\mathbf{y}$ as the training progresses, without considering the characteristics of the observed degraded image $\mathbf{y}$.

Self-supervised image restoration methods that use blind-spot strategies, such as Noise2Void [12], have been successful. Blind-spot strategies are methods of obtaining a

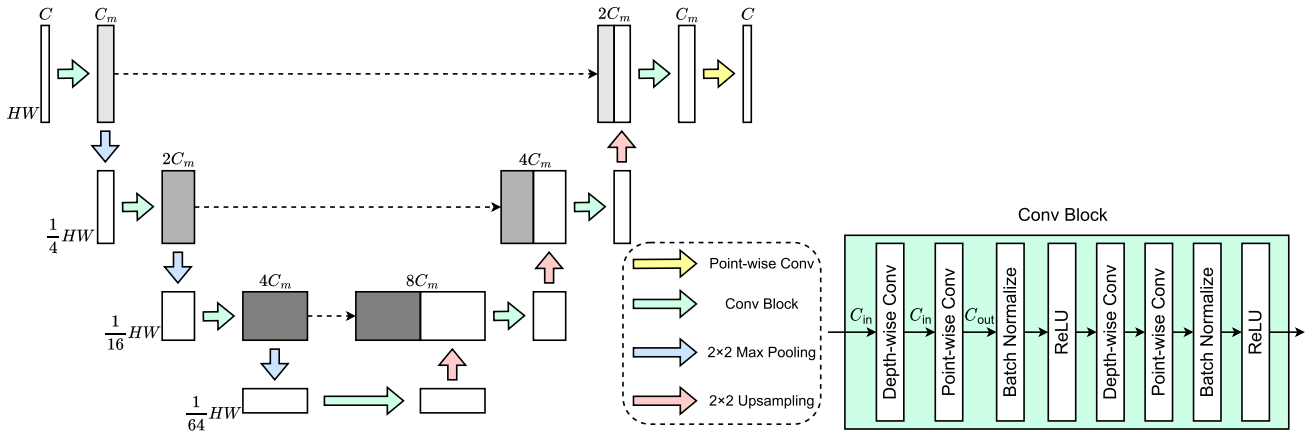

FIGURE 1. The model architecture of proposed method. The model structure is based on U-Net, and the number of parameters is reduced by using the separable convolution without compromising restoration accuracy.

denoising model by predicting some intentionally hidden "blind pixels" from surrounding unblinded pixels in the learning process. This strategy overcomes the limitation of outputting the observed degraded image $\mathbf{y}$ as is, which is a drawback of Deep Image Prior, and enables learning from a large number of degraded images. Noise2Void learns to replace randomly selected pixels in a degraded image with pixel values from the surrounding area and estimates the original pixel values from the replaced pixels.

Blind2Unblind [13], a recent self-supervised denoising method, also employs a blind-spot strategy. In this method, blind-spots are regularly placed in a 2×2 region, and multiple inputs with different blind-spots are provided to the model to obtain the restored images corresponding to each input. The output can then be obtained by extracting and combining only the pixels corresponding to the respective blinded pixels from the obtained restored images. During the training process, the loss function is the weighted sum of the output obtained by stitching the pixel values corresponding to the blind-spots and the output obtained from the degraded image, where the weight of the output obtained from the observed degraded image gradually increases. After training, restoration is performed by directly inputting the observed degraded image into the model.

These methods, with the exception of Deep Image Prior, rely on large numbers of degraded images, making it challenging to apply them directly to HSIs, where the cost of preparing large amounts of data under the same conditions (e.g., number of bands and objects to be imaged) is prohibitively high. To address this problem, several zero-shot HSI restoration methods based on self-supervised RGB image restoration methods that only learn from an observed degraded HSI have been proposed. Deep HS Prior [14] is an extension of Deep Image Prior, as described above, that can be applied to the observed degraded HSI and is trained by the loss function in (2). Because it uses the architecture of Deep Image Prior, it still suffers from the drawback of outputting an observed degraded HSI as is if the learning process is too advanced. Imamura et al.'s method [15] addresses this problem by using information on the degradation process. The model is trained through supervised learning, using the observed degraded HSI as the target data and further degraded HSIs along the degradation process of the observed degraded HSI as input data. Because a large number of doubly degraded HSIs can be generated for a single observed degraded HSI, a large training dataset can be created and trained during the zero-shot learning process. However, the denoising performance deteriorates when information on the degradation process concerning the observed degraded HSI, such as the noise distribution, cannot be accurately estimated.

Our proposed method addresses the aforementioned issues through the utilization of a blind-spot strategy tailored for zero-shot HSI restoration, as well as a deep-learning model employing separable convolution. By employing the blind-spot strategy during training, we effectively suppress identity mapping and eliminate the necessity to make assumptions about the degradation process. Furthermore, the incorporation of random blind-spot changes and a deep learning model with separable convolution helps mitigate overfitting caused by limited data availability. These enhancements enable us to achieve highly accurate restoration not only for synthetic noise, such as Gaussian noise, but also for data containing unknown information about the degradation process, such as HSI affected by real noise.

III. PROPOSED METHOD

A. MODEL ARCHITECTURE

In this section, we present the model structure, learning method, and loss function used in the proposed method. The proposed model architecture, as shown in FIGURE 1, is based on U-Net [20] and employs the separable convolution [21] method. The separable convolution method decomposes the conventional 2D convolution into depth-wise (channel-wise) convolution and 1×1 (point-wise) convolution, which reduces the number of required parameters while minimizing the reduction in the expressive capability of the convolution layer. When the kernel size for 2D convolution is $q \times q$ and the number of input and output channels is the same, the separable convolution reduces the number of parameters to

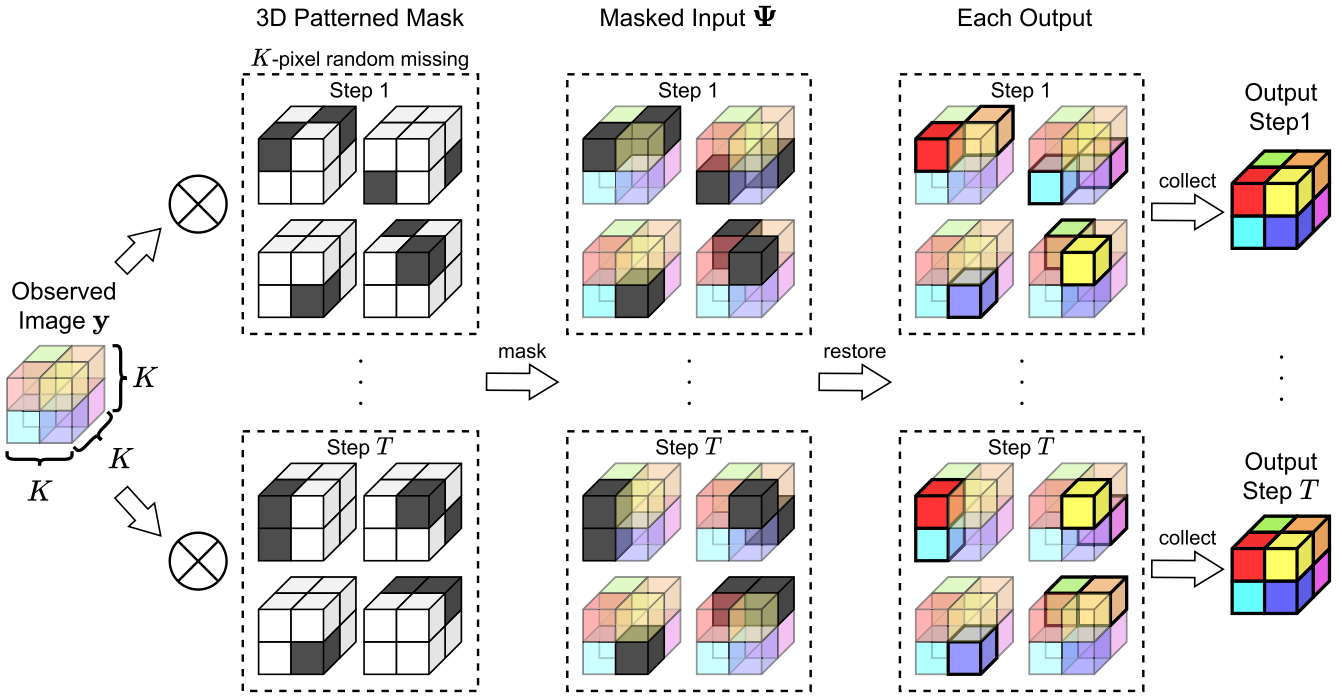

FIGURE 2. The diagram illustrates the model for image restoration using 3D random patterned masks. At each step, black pixels are randomly selected as blind-spots, and the element-wise product of these multiple masks and the input is taken to obtain multiple masked inputs. The output is obtained by combining the restored blind pixels.

approximately $1/q^2$. Because HSI provides dense information in the band direction, building a deep learning model requires a vast number of parameters, and overfitting is a common issue. HSIs tend to have a lower diagonal correlation between bands than other 3D data types such as video or MRI (Magnetic Resonance Imaging). This is due to the nature of HSI being composed of a continuous sequence of 2D images with similar spatial structures. Previous work has shown that separable convolution works well in zero-shot HSI restoration, although it loses the ability to exploit diagonal correlation across the bands [22]. We adopt this technique in our approach to improve the accuracy of HSI restoration. For more information on the effectiveness of separable convolution for HSI, see the paper [22]. We discuss the validity of model architecture later in Section IV-C1.

In FIGURE 1, H, W, and C represent the sizes of the input $\mathbf{x} \in \mathbb{R}^{H\times W\times C}$, and C_m is the number of channels of the intermediate features in the hidden layer of the model. The subscript next to each box in FIGURE 1 indicates the change in the size of the intermediate features of the model. Each process in the model is described in the center of FIGURE 1: 2×2 max pooling or 2×2 upsampling changes the spatial size of the intermediate outputs, and the first point-wise convolution in the conv block changes the number of channels in the intermediate features from the input channels C_{in} to the output channels C_{out}.

B. BLIND-SPOT STRATEGY

FIGURE 2 illustrates the proposed method's blind-spot selection process. The proposed method for selecting blind-spots uses 3D patterned masks with K^2 variations, where K blind-spots are randomly selected from each $K \times K \times K$ region in the input image. For example, when $K = 2$, four 3D patterned masks are used in one training step, and each mask contains two blind-spots. When selecting the positions of the blind-spots of the 3D patterned masks, it is necessary to ensure that they do not overlap with other masks. Each resulting 3D patterned mask is then replicated to match the size of the input and multiplied element-wise with the observed degraded HSI to create the masked inputs Ψ. Moreover, the 3D patterned masks are updated at each training step, randomly altering the combination of blind-spot pixels. To obtain the output, each masked input is fed into the deep-learning model $f_\Theta(\cdot)$, and the masked pixels are extracted and combined. By making the mask 3D, it becomes possible to perform restoration that considers the spatial and spectral directional information possessed by the HSI. In addition, by randomly selecting blind-spots, the amount of training data is augmented to suppress overfitting. Increasing the value of K improves the restoration accuracy, but increases the computational cost owing to an increase in the number of masked inputs to be restored. We discuss the validity of mask modification later in Section IV-C2.

This method aims to minimize the following loss function that uses masked and unmasked inputs as in [13]:

$$\begin{aligned}\mathbf{L}(\mathbf{y}) = &\|g\{f_\Theta(\Psi)\} + \alpha\hat{f}_\Theta(\mathbf{y})\text{-}(\alpha+1)\mathbf{y}\|_2^2 \\ &+ \beta\|g\{f_\Theta(\Psi)\} - \mathbf{y}\|_2^2. \end{aligned} \tag{3}$$

Here, $g(\cdot)$ is a function for extracting and combining blind-spot pixels from the model output, $\hat{f}_\Theta(\cdot)$ is a function

for obtaining the model output without backpropagation, and α and β are hyperparameters. The right term $\beta\|g\{f_\Theta(\Psi)\} - \mathbf{y}\|_2^2$ in (3) is a regularization term to stabilize learning, and β determines the strength of this regularization. In this paper, we opt not to tune α and β but instead employ the default values of $\alpha = \beta = 1$. The restored image is obtained from the weighted sum of the masked inputs and the restored output from the degraded image, as described in [13]. Unlike Blind2Unblind, the proposed method does not change the hyperparameter α during training.

In Equation (3), the model is trained such that the output obtained from the masked image and that obtained from the observed degraded image are close to the observed degraded image $\mathbf{y}$. When the level of image degradation is severe, there is an increased likelihood of overfitting in the model. To address this issue, the ℓ_2 weight decay is applied to the model parameters to prevent overfitting.

IV. EXPERIMENTS

A. EXPERIMENTAL CONDITIONS

In this section, we demonstrate the effectiveness of the proposed method through several denoising experiments. Furthermore, we conducted ablation studies to verify the effectiveness of the proposed model architecture and blind-spot strategy. The model architecture ablation study examined the difference in restoration accuracy between the standard and separable convolutions. In the blind-spot strategy ablation study, we compared the effectiveness of the blind-spot strategy for each channel of the conventional method Blind2Unblind [13], the blind-spot strategy that extends it to 3D, and the blind-spot strategy of our proposed method. We describe the denoising experiments and ablation studies in Sections IV-B and IV-C, respectively.

As test images, we used Frisco, Stanford, Pavia-C, and Pavia-U, which are commonly used in noise removal experiments, and normalized them to the range [0, 1]. Owing to the large size of the Pavia-C image, we cropped a portion of each image for use in the experiment. The sizes of each HSI used in the experiments are shown in TABLE 1.

TABLE 1. The size of the test data.

image	size
Frisco	$300 \times 300 \times 148$
Stanford	$290 \times 300 \times 148$
Pavia-C	$600 \times 400 \times 102$
Pavia-U	$610 \times 340 \times 103$

When training the proposed model, a $40 \times 40 \times C$ patch was repeatedly extracted from the observed degraded hyperspectral image $\mathbf{y} \in \mathbb{R}^{H \times W \times C}$ and used for training. In addition, we applied random left-right flipping and rotation by 0, 90, 180, and 270 degrees for data augmentation during training. Nearest neighbor upsampling was used in the process shown in FIGURE 1. The batch size during training was set to 24. The number of intermediate channels C_m in the proposed model was set to 200. As in the conventional methods, Adam was used as the optimizer. The initial learning rate was set to 0.001 and the number of epochs to 200. The learning rate was halved every 40 epochs. We set both $K = 3$ and $K = 2$, the latter of which was used in the proposed blind-spot strategy. Additionally, we set both α and β of the hyperparameter to 1. We use four NVIDIA RTX A4000 GPUs for distributed parallel learning.

The following four methods were utilized for comparison in our experiments. We employed BM4D [10], a Gaussian denoising method for general 3D data, and FastHyDe [9], a highly accurate HSI Gaussian denoising method, as optimization-based methods. As for deep-learning-based methods, we employed Deep HS Prior [14], which is a DIP-based HSI denoising method, and Imamura et al.'s method [15], which uses $\mathbf{y}$ as target data and degraded images with noise added to $\mathbf{y}$ as input data. For BM4D, FastHyDe, and Deep HS Prior, we modified some of the code provided by the authors and tuned the hyperparameters.

As evaluation metrics, PSNR and ERGAS [23] were used to evaluate the overall image error, SSIM [24] was used to evaluate the similarity of the spatial structure of the images, and SAM [25] was used to evaluate the similarity in the spectral direction.

B. DENOISING

This section presents the results of denoising experiments conducted using various noise types. We used Gaussian noise with mean 0 and standard deviation σ, Laplacian noise with mean 0 and standard deviation σ, and Poisson noise dependent on the pixel values for each denoising experiment. We set the standard deviation σ to 0.1 and 0.15. The up (or down) arrows next to the evaluation metrics in TABLE 2, 3, and 4 indicate whether higher or lower values in the metrics, respectively, represent higher accuracy. The bold values in the tables represent the best restoration accuracies for that experimental setup, whereas the underlined values represent the second-best restoration accuracies.

1) GAUSSIAN NOISE REMOVAL

In this experiment, observed images degraded by Gaussian noise with standard deviations of 0.1 and 0.15 were created, and the noise removal accuracy was compared. The experimental results are shown in TABLE 2. The results in TABLE 2 show that the proposed method has comparable or better restoration accuracy than the other zero-shot methods. With a standard deviation of 0.1, the proposed method outperformed the conventional method in all metrics for Frisco, three metrics for Pavia-C and Pavia-U, and two metrics for Stanford. In the Stanford data, the proposed method had the second-best restoration accuracy in SSIM, one of the evaluation metrics for which it did not have the best restoration accuracy. With a standard deviation of 0.15, the proposed method showed better restoration accuracy than the conventional method, except for the Stanford data. The proposed method outperformed the conventional method in

TABLE 2. Denoising result in Gaussian noise. The bold values represent the best restoration accuracies, whereas the underlined values represent the second-best restoration accuracies.

standard deviation σ		0.1				0.15			
metrics		PSNR↑	SSIM↑	SAM↓	ERGAS↓	PSNR↑	SSIM↑	SAM↓	ERGAS↓
Frisco	BM4D	37.52	0.9728	7.251e-2	44.15	35.52	0.9592	9.036e-2	55.26
	FastHyDe	41.80	0.9893	4.682e-2	27.64	39.64	0.9834	5.927e-2	35.23
	Deep HS Prior	39.17	0.9812	6.149e-2	36.24	36.32	0.9607	8.772e-2	49.68
	Imamura et al.	41.81	0.9896	4.484e-2	26.88	39.88	**0.9860**	5.354e-2	33.02
	Ours	**42.25**	**0.9903**	**3.486e-2**	**20.90**	**40.28**	0.9852	**4.103e-2**	**25.64**
Stanford	BM4D	37.84	0.9722	1.238e-1	60.41	35.81	0.9573	1.544e-1	76.04
	FastHyDe	42.22	0.9892	8.493e-2	36.27	**40.29**	0.9834	1.036e-1	45.30
	Deep HS Prior	39,59	0.9814	1.119e-1	48.73	37.46	0.9667	1.539e-1	62.55
	Imamura et al.	42.48	**0.9905**	7.656e-2	**35.07**	40.28	**0.9861**	9.955e-2	**44.40**
	Ours	**42.59**	0.9901	**4.704e-2**	40.75	40.18	0.9831	**5.985e-2**	53.81
Pavia-C	BM4D	35.56	0.9637	1.044e-1	46.08	33.46	0.9435	1.324e-1	58.69
	FastHyDe	37.33	0.9773	8.460e-2	37.53	35.50	0.9665	1.040e-1	46.31
	Deep HS Prior	35.58	0.9669	1.025e-1	45.80	33.93	0.9504	1.242e-1	55.42
	Imamura et al.	37.52	0.9772	8.296e-2	**36.77**	36.04	0.9694	9.563e-2	**43.50**
	Ours	**37.72**	**0.9781**	**8.213e-2**	41.70	**36.41**	**0.9726**	**8.245e-2**	46.82
Pavia-U	BM4D	35.67	0.9671	9.053e-2	37.99	33.62	0.9499	1.142e-1	48.10
	FastHyDe	37.28	0.9783	7.451e-2	31.53	35.34	0.9680	9.259e-2	39.39
	Deep HS Prior	35.69	0.9693	8.887e-2	37.75	33.80	0.9522	1.107e-1	46.95
	Imamura et al.	37.31	0.9773	7.447e-2	**31.50**	35.95	**0.9718**	8.412e-2	**36.77**
	Ours	**37.73**	**0.9786**	**6.012e-2**	35.56	**36.20**	0.9717	**6.544e-2**	41.01

TABLE 3. Denoising result in Laplacian noise. The bold values represent the best restoration accuracies, whereas the underlined values represent the second-best restoration accuracies.

standard deviation σ		0.1				0.15			
metrics		PSNR↑	SSIM↑	SAM↓	ERGAS↓	PSNR↑	SSIM↑	SAM↓	ERGAS↓
Frisco	BM4D	30.46	0.8561	1.734e-1	95.72	28.00	0.7690	2.278e-1	125.5
	FastHyDe	39.90	0.9845	5.755e-2	34.10	37.96	0.9765	7.068e-2	42.66
	Deep HS Prior	37.57	0.9737	7.366e-2	43.64	35.54	0.9588	9.328e-2	54.88
	Imamura et al.	40.61	**0.9881**	5.011e-2	31.22	**38.61**	**0.9806**	6.320e-2	39.39
	Ours	**40.65**	0.9864	**3.921e-2**	**24.60**	38.63	0.9784	**4.611e-2**	**30.95**
Stanford	BM4D	30.91	0.8628	3.207e-1	133.8	28.51	0.7794	4.136e-1	176.3
	FastHyDe	40.58	0.9844	1.002e-1	**43.78**	38.49	0.9758	1.237e-1	**55.56**
	Deep HS Prior	37.72	0.9713	1.408e-1	60.57	35.66	0.9554	1.770e-1	76.77
	Imamura et al.	40.12	0.9841	1.033e-1	45.89	38.40	**0.9790**	1.237e-1	55.69
	Ours	**40.72**	**0.9851**	**5.497e-2**	50.00	**38.60**	0.9783	**5.578e-2**	58.54
Pavia-C	BM4D	28.98	0.8355	2.196e-1	98.27	26.70	0.7494	2.828e-1	127.7
	FastHyDe	35.75	0.9683	1.011e-1	45.00	33.87	0.9532	1.252e-1	**55.92**
	Deep HS Prior	34.24	0.9548	1.194e-1	53.44	32.25	0.9296	1.498e-1	67.23
	Imamura et al.	36.55	0.9726	9.202e-2	**41.08**	33.31	0.9363	1.345e-1	59.73
	Ours	**36.62**	**0.9737**	**8.118e-2**	46.03	**34.47**	**0.9590**	**9.623e-2**	57.24
Pavia-U	BM4D	29.01	0.8419	1.943e-1	81.90	26.67	0.7550	2.526e-1	107.2
	FastHyDe	35.59	0.9697	8.994e-2	38.27	33.89	0.9566	1.091e-1	**46.52**
	Deep HS Prior	34.03	0.9553	1.079e-1	45.67	32.12	0.9302	1.343e-1	57.02
	Imamura et al.	36.29	0.9730	8.237e-2	**35.38**	33.82	0.9497	1.091e-1	47.08
	Ours	**36.53**	**0.9736**	**6.348e-2**	39.70	**34.37**	**0.9600**	**7.467e-2**	50.00

three metrics for Frisco and Pavia-C, two metrics for Pavia-U, and one metric for Stanford. The proposed method for Frisco and Pavia-U had the second-highest restoration accuracy in terms of SSIM. Although the restoration accuracy was slightly higher in terms of SAM, the excellent restoration accuracy in both the SSIM and SAM measures suggests that the proposed method achieved balanced denoising results without learning to excessively reduce errors in the spatial or spectral directions.

2) LAPLACIAN NOISE REMOVAL

In this experiment, the accuracy of noise removal in the observed images degraded by Laplacian noise with standard deviations of 0.1 and 0.15 was compared, and the experimental results are presented in TABLE 3. The results in the table demonstrate that the proposed method outperformed the conventional methods. Specifically, using a standard deviation of 0.1, the proposed method outperformed the conventional method in three evaluation metrics for all data, and Frisco's restoration results for the proposed method had the second-best restoration accuracy in terms of SSIM. Using a standard deviation of 0.15, the results showed superior restoration accuracy for Pavia-C and Pavia-U in three metrics and for Frisco and Stanford in two metrics. Additionally, the results for Frisco showed the second-best restoration accuracy in terms of PSNR and SSIM, and those for Stanford showed the second-best results in terms of SSIM.

TABLE 4. Denoising result in Poisson noise. The bold values represent the best restoration accuracies, whereas the underlined values represent the second-best restoration accuracies.

noise strength constant s		15				6			
metrics		PSNR↑	SSIM↑	SAM↓	ERGAS↓	PSNR↑	SSIM↑	SAM↓	ERGAS↓
Frisco	FastHyDe	40.14	0.9895	4.866e-2	30.96	35.66	0.9797	6.512e-2	49.15
	Deep HS Prior	38.88	0.9808	6.228e-2	36.50	36.53	0.9686	8.152e-2	47.72
	Imamura et al.	41.76	**0.9915**	4.405e-2	26.13	39.10	0.9840	5.670e-2	35.16
	Ours	**41.96**	0.9906	**3.042e-2**	**20.87**	**39.69**	**0.9842**	**3.673e-2**	**26.69**
Stanford	FastHyDe	41.63	0.9909	7.250e-2	38.24	37.70	0.9824	9.266e-2	59.87
	Deep HS Prior	40.22	0.9848	9.696e-2	45.20	37.45	0.9727	1.293e-1	61.79
	Imamura et al.	42.08	0.9908	7.213e-2	36.02	39.30	0.9844	8.075e-2	50.52
	Ours	**42.84**	**0.9925**	**3.537e-2**	**31.82**	**40.05**	**0.9867**	**4.200e-2**	**43.30**
Pavia-C	FastHyDe	36.46	0.9773	8.554e-2	41.30	33.34	0.9618	1.091e-1	59.07
	Deep HS Prior	35.38	0.9666	1.034e-1	46.80	33.24	0.9463	1.304e-1	59.86
	Imamura et al.	37.11	0.9784	8.016e-2	**38.40**	34.27	0.9608	1.035e-1	53.30
	Ours	**37.66**	**0.9801**	**7.005e-2**	40.37	**35.15**	**0.9672**	**8.666e-2**	**51.86**
Pavia-U	FastHyDe	36.10	0.9765	7.722e-2	35.99	32.79	0.9609	9.947e-2	52.68
	Deep HS Prior	34.97	0.9648	9.434e-2	40.98	32.76	0.9433	1.207e-1	52.85
	Imamura et al.	36.59	0.9762	7.218e-2	**34.08**	33.56	0.9536	9.389e-2	48.37
	Ours	**37.23**	**0.9783**	**5.641e-2**	35.93	**34.78**	**0.9653**	**6.937e-2**	**46.56**

3) POISSON NOISE REMOVAL

The experiment utilized the following equation to obtain the observed degraded image:

$$\mathbf{y} = P(s \cdot \mathbf{x})/s. \tag{4}$$

Here, $P(s \cdot \mathbf{x})$ is a function that applies noise based on the Poisson distribution of the mean $s \cdot \mathbf{x}$, and s is a constant that determines the strength of the noise. As indicated in (4), the noise varies depending on the image. The noise strength constants used in this experiment were $s = 15$ and 6, and the results are presented in TABLE 4. In this table, BM4D, which was used as a conventional method in other experiments, was excluded because of its significantly lower restoration accuracy compared to the other methods. The table shows that the proposed method has excellent restoration accuracy for all test data and experimental conditions. For $s = 15$, our method outperformed the conventional methods in all evaluation metrics, except for Frisco in SSIM, and Pavia-C and Pavia-U in ERGAS. For $s = 6$, our method outperformed the conventional method in all evaluation metrics. Notably, our method significantly outperformed the conventional method in terms of PSNR and SAM, indicating that the restoration accuracy of the spectral information was superior to that of the conventional method. These results demonstrate the effectiveness of the proposed method for restoring images with additive noise or noise that depends on pixel values.

4) REAL-WORLD NOISE REMOVAL

In this experiment, we compared the performance of the proposed method and the conventional methods by applying them to real-world noise. Since non-degraded HSI data is not available, objective evaluation is not used. For this experiment, we used four HSI data (Urban, Botswana, Loukia, Nefeli) with real-world noise. Due to the difficulty of determining the optimal timing for early stopping in advance with Deep HS Prior, we used the default setting of 1800 epochs for our experiment. To estimate the standard deviation of the assumed Gaussian noise distribution in Imamura et al.'s method, we followed the approach described in [26]. BM4D and FastHyDe were excluded from the experimental results as they did not yield better restoration results compared to the other methods.

The results of real-world noise removal are presented in FIGURE 3. To enhance visualization, we extracted a single channel from the denoised HSI, cropped it, normalized it, and displayed it. The enlarged images framed by the red frame in the lower right corner of the first and second lines in FIGURE 3 provide a closer view of the respective areas. The denoised images demonstrate that Deep HS Prior tends to produce blurry results, while Imamura et al.'s and the proposed method generate sharper images. Examining the images individually, we observe that in the first row, Deep HS Prior and Imamura et al.'s method fails to restore the black hole-like object in the upper left of the enlarged image, which is replaced by the surrounding pattern. In the second row, the road appears blurred in the Deep HS Prior restoration result, although the horizontal stripe noise in the input was successfully removed by all methods. Comparing Imamura et al.'s method and the proposed method in the third and fourth lines, we notice that Imamura et al.'s method exhibits a collapsed texture compared to the proposed method. These results indicate that the proposed method effectively removes noise while preserving the texture of the HSI, outperforming conventional methods.

C. ABLATION STUDIES

In this section, we describe the ablation studies performed to verify the efficacy of the model architecture and the blind-spot strategy in the proposed method. The ablation study of the model architecture explains the difference between the standard convolution (convolution with dense kernels) and separable convolution. The study of the blind-spot strategy compared three blind-spot strategies: Blind2Unblind, a 3D extension of Blind2Unblind's

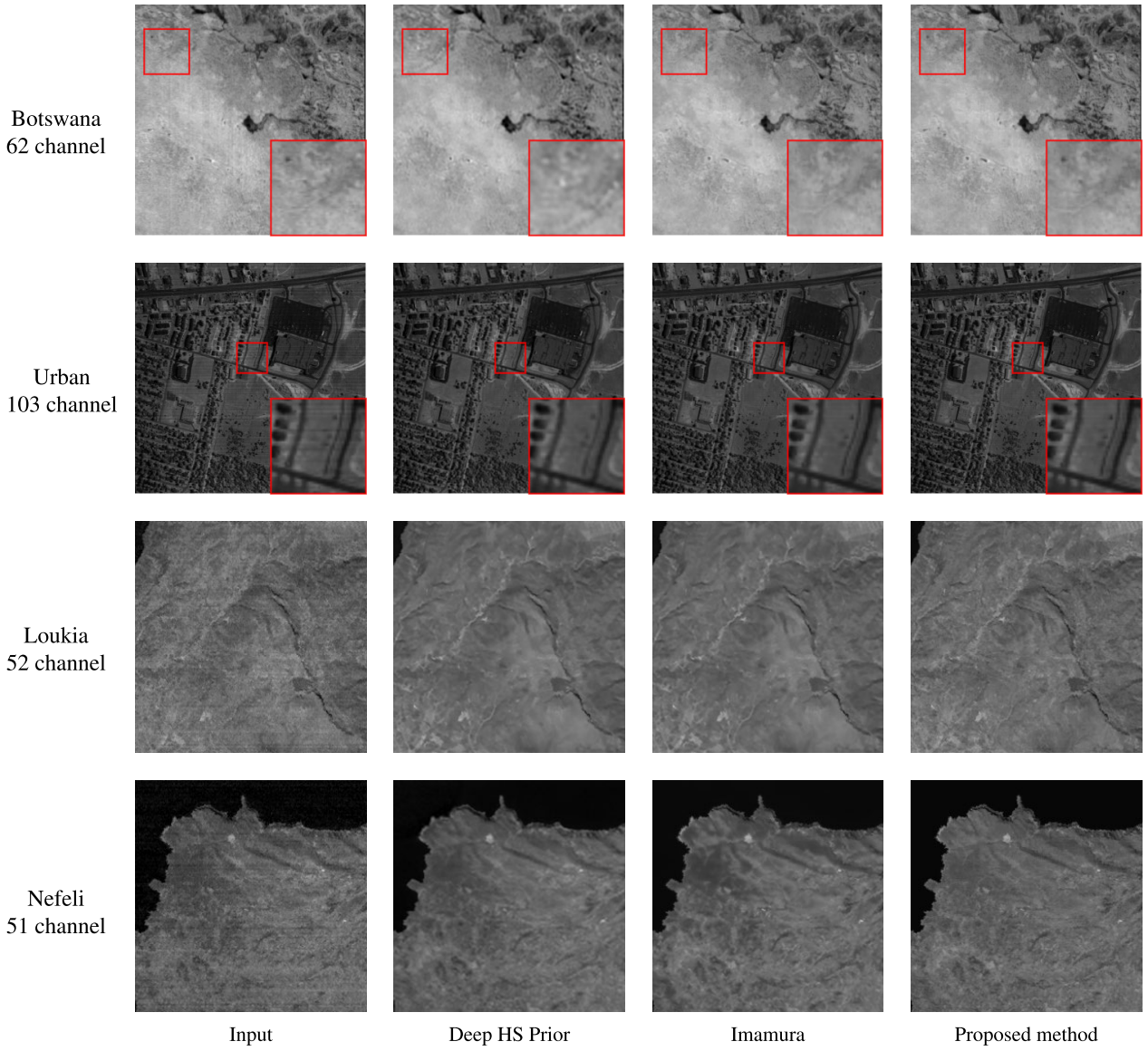

FIGURE 3. Results for Real-world noise: where one channel is extracted from HSI, crop, normalized, and displayed.

blind-spot strategy, and the proposed method. The ablation study uses the experimental setup for Gaussian noise removal described in Section IV-B1, and the four data (Frisco, Pavia-C, Pavia-U, Stanford). The values in TABLE 5 to 7, which show the experimental results, represent the averages of the results for the respective test data used in Section IV-B.

TABLE 5. Comparison of convolution at $K = 2$.

σ	metrics	PSNR	SSIM	SAM	ERGAS
0.1	standard	**40.04**	**0.9853**	**5.080e-2**	**33.84**
	separable	39.59	0.9826	5.631e-2	36.06
0.15	standard	36.24	0.9624	9.300e-2	45.06
	separable	**38.06**	**0.9775**	**6.160e-2**	**42.19**

1) ABLATION STUDY OF MODEL ARCHITECTURE

In this section, we examine the effectiveness of the separable convolution. The results of the Gaussian noise removal experiments at $K = 2$ and $K = 3$ are shown in TABLE 5 and TABLE 6, and the PSNR plots at a standard deviation of 0.15 are shown in FIGURE 4.

TABLE 6. Comparison of convolution at $K = 3$.

σ	metrics	PSNR	SSIM	SAM	ERGAS
0.1	standard	**40.28**	**0.9857**	**5.084e-2**	**33.22**
	separable	40.07	0.9843	5.596e-2	34.70
0.15	standard	35.66	0.9539	2.995e-1	66.56
	separable	**38.27**	**0.9781**	**6.208e-2**	**41.77**

The results in TABLE 5 and TABLE 6 show that the standard convolution achieved comparable or slightly better accuracy than the separable convolution when the standard deviation was 0.1. However, the model with separable convolution outperformed the model with standard convolution when the standard deviation was 0.15. Notably, the standard convolution exhibited a significant decrease in accuracy for the standard deviation of 0.15.

The results in FIGURE 4 demonstrate that the model using standard convolution tended to overfit when the standard deviation was 0.15. One possible reason for this is that the standard convolution model has a significantly higher number of parameters, approximately 160M at $C_m = 200$, compared to the separable convolution model, which only has approximately 18M parameters. Although standard convolution shows high restoration accuracy at low standard deviations, the proposed method uses separable convolution due to robustness to severely degraded HSI. The effectiveness of the separable convolution is confirmed by the fact that the standard convolution only slightly improves at low standard deviations, while the separable convolution achieves considerably better accuracy at high standard deviations.

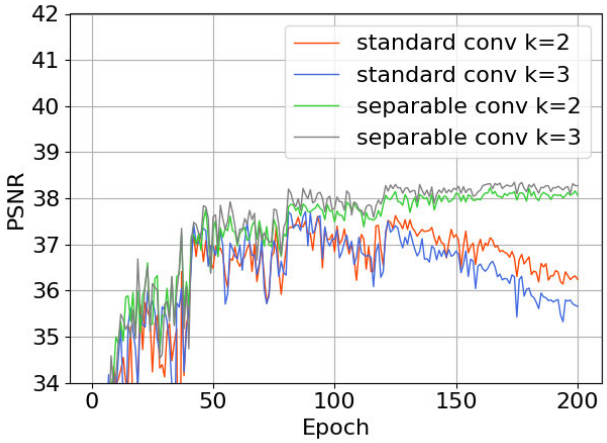

FIGURE 4. PSNR plots for each convolution with Gaussian noise removal with the standard deviation of 0.15.

2) ABLATION STUDY OF BLIND-SPOT STRATEGY

In this section, we evaluate the effectiveness of the proposed blind-spot strategy by comparing it to Blind2Unblind and a strategy that extends Blind2Unblind's strategy to 3D data. Blind2Unblind masks each pixel in a 2×2 region of each channel of the observed degraded HSI, reconstructs the value of each pixel, and combines them. We apply the same process as in Blind2Unblind to each pixel in the $2 \times 2 \times 2$ regions using the 3D extended blind-spot strategy.

TABLE 7. Comparison of blind-spot strategies.

σ	metrics	PSNR	SSIM	SAM	ERGAS
0.1	B2U [13]	37.81	0.9711	6.798e-2	45.89
	Pixel-wise	39.58	0.9816	6.175e-2	37.10
	Ours ($K = 2$)	39.59	0.9826	5.631e-2	36.06
	Ours ($K = 3$)	**40.07**	**0.9843**	**5.596e-2**	**34.70**
0.15	B2U [13]	36.69	0.9715	6.285e-2	47.82
	Pixel-wise	38.14	0.9776	6.332e-2	42.48
	Ours ($K = 2$)	38.06	0.9775	**6.160e-2**	42.19
	Ours ($K = 3$)	**38.27**	**0.9781**	6.208e-2	**41.77**

The experimental results are presented in TABLE 7 and FIGURE 5. In TABLE 7, B2U represents Blind2Unblind, whereas the "Pixel-wise" method is Blind2Unblind's blind-spot strategy extended to 3D, and Ours ($K = 2$) and Ours ($K = 3$) denote our cases with $K = 2$ and $K = 3$, respectively, as shown in FIGURE 2. The results in TABLE 7 indicate that our method achieved the best scores. The restoration accuracy of the Pixel-wise method was higher than that of B2U, indicating that the use of a 3D mask and the incorporation of HSI spectral information contributed to the improvement of restoration accuracy. Moreover, the proposed method is superior in terms of computational cost because the Pixel-wise method uses K^3 masks (8 masks for $K = 2$), whereas the proposed method utilizes K^2 masks (4 masks for $K = 2$ and 9 masks for $K = 3$).

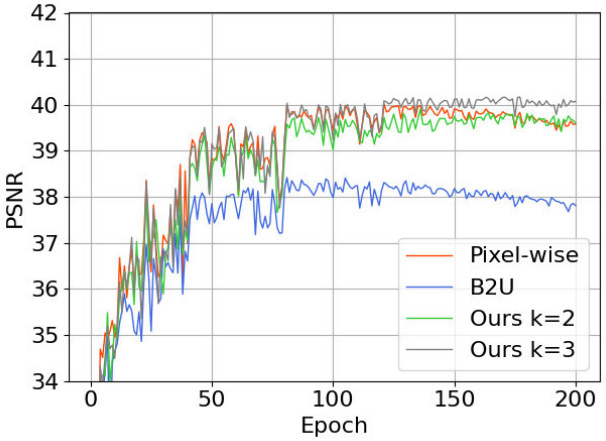

FIGURE 5. PSNR plots for each blind-spot strategy in Gaussian noise removal with a standard deviation of 0.1.

In addition, as shown in FIGURE 5, the PSNR of the Blind2Unblind and Pixel-wise methods gradually decreased after approximately 120 epochs, indicating that they were overfitting. However, the PSNR of the proposed method converged, suggesting that the generalization performance was improved by incorporating randomness. These results demonstrate the effectiveness of the proposed blind-spot strategy.

Compared to conventional methods, the proposed method requires approximately two to three times the computation time. However, our method still has room for improvement in terms of optimal model selection and the choice of hyperparameters during training. We did not perform these hyperparameter optimization tasks at this time because they need huge computational costs.

V. CONCLUSION

In this paper, we proposed a deep-learning-based method for zero-shot HSI denoising. The model architecture is based on U-Net and incorporates the separable convolutions. In addition to spatial directional information, we introduced a novel blind-spot strategy that utilizes 3D randomly patterned masks for spectral directional information restoration. By randomly missing K pixels in each $K \times K \times K$ region using K^2 3D patterned masks without overlap, our method achieves high restoration accuracy and generalization performance. We conducted several experiments to demonstrate the superiority of the proposed method over other

optimization- and deep-learning-based zero-shot HSI restoration methods. We also performed an ablation study of the model structure and blind-spot strategy of the proposed method. The proposed model incorporates separable convolutions to suppress overfitting in cases of severe degradation during zero-shot HSI restoration. Our study demonstrates that 3D masks are effective in zero-shot HSI restoration and that random blind-spot selection contributes to improved generalization performance.

REFERENCES

[1] M. Borengasser, W. S. Hungate, and R. Watkins, *Hyperspectral Remote Sensing: Principles and Applications*. Boca Raton, FL, USA: CRC Press, 2007.
[2] L. Ojha, M. B. Wilhelm, S. L. Murchie, A. S. McEwen, J. J. Wray, J. Hanley, M. Massé, and M. Chojnacki, "Spectral evidence for hydrated salts in recurring slope lineae on Mars," *Nature Geosci.*, vol. 8, no. 11, pp. 829–832, Nov. 2015.
[3] H. Liu, T. Yu, B. Hu, X. Hou, Z. Zhang, X. Liu, J. Liu, X. Wang, J. Zhong, Z. Tan, S. Xia, and B. Qian, "UAV-borne hyperspectral imaging remote sensing system based on acousto-optic tunable filter for water quality monitoring," *Remote Sens.*, vol. 13, no. 20, p. 4069, Oct. 2021.
[4] F. Andika, M. Rizkinia, and M. Okuda, "A hyperspectral anomaly detection algorithm based on morphological profile and attribute filter with band selection and automatic determination of maximum area," *Remote Sens.*, vol. 12, no. 20, p. 3387, Oct. 2020.
[5] X. Cao, F. Zhou, L. Xu, D. Meng, Z. Xu, and J. Paisley, "Hyperspectral image classification with Markov random fields and a convolutional neural network," *IEEE Trans. Image Process.*, vol. 27, no. 5, pp. 2354–2367, May 2018.
[6] A. Sellami and S. Tabbone, "Deep neural networks-based relevant latent representation learning for hyperspectral image classification," *Pattern Recognit.*, vol. 121, Jan. 2022, Art. no. 108224.
[7] A. Bjorgan and L. L. Randeberg, "Towards real-time medical diagnostics using hyperspectral imaging technology," in *Proc. Eur. Conf. Biomed. Opt.* Washington, DC, USA: Optica Publishing Group, 2015, pp. 1–9, Paper 953712.
[8] G. Lu and B. Fei, "Medical hyperspectral imaging: A review," *J. Biomed. Opt.*, vol. 19, no. 1, Jan. 2014, Art. no. 010901.
[9] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 730–742, Mar. 2018.
[10] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, Jan. 2013.
[11] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.
[12] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2Void—Learning denoising from single noisy images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2129–2137.
[13] Z. Wang, J. Liu, G. Li, and H. Han, "Blind2Unblind: Self-supervised image denoising with visible blind spots," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2022, pp. 2027–2036.
[14] O. Sidorov and J. Y. Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 1–8.
[15] R. Imamura, T. Itasaka, and M. Okuda, "Zero-shot hyperspectral image denoising with separable image prior," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, 2019, pp. 1–5.
[16] Q. Zhang, Q. Yuan, J. Li, F. Sun, and L. Zhang, "Deep spatio-spectral Bayesian posterior for hyperspectral image non-i.i.d. noise removal," *ISPRS J. Photogramm. Remote Sens.*, vol. 164, pp. 125–137, Jun. 2020.
[17] Q. Zhang, Q. Yuan, M. Song, H. Yu, and L. Zhang, "Cooperated spectral low-rankness prior and deep spatial prior for HSI unsupervised denoising," *IEEE Trans. Image Process.*, vol. 31, pp. 6356–6368, 2022.
[18] Q. Zhang, Y. Zheng, Q. Yuan, M. Song, H. Yu, and Y. Xiao, "Hyperspectral image denoising: From model-driven, data-driven, to model-data-driven," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 6, 2023, doi: 10.1109/TNNLS.2023.3278866.
[19] Q. Zhang, Y. Dong, Q. Yuan, M. Song, and H. Yu, "Combined deep priors with low-rank tensor factorization for hyperspectral image restoration," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.
[20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*. Munich, Germany: Springer, Oct. 2015, pp. 234–241.
[21] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1251–1258.
[22] R. Imamura, T. Itasaka, and M. Okuda, "Self-supervised hyperspectral image restoration using separable image prior," 2019, *arXiv:1907.00651*.
[23] Q. Du, N. H. Younan, R. King, and V. P. Shah, "On the performance evaluation of pan-sharpening techniques," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 518–522, Oct. 2007.
[24] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[25] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm," in *Proc. JPL, Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, vol. 1, 1992, pp. 1–3.
[26] J. M. Bioucas-Dias and J. M. P. Nascimento, "Hyperspectral subspace identification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, Aug. 2008.

TATSUKI ITASAKA (Student Member, IEEE) received the B.E. and M.E. degrees, in 2019 and 2021, respectively. He is currently pursuing the Ph.D. degree with the Graduate School of Science and Engineering, Doshisha University. He was with the Faculty of Environmental Engineering, The University of Kitakyushu, from 2015 to 2021. His research interests include image restoration, high dynamic range imaging, and digital filter design.

MASAHIRO OKUDA (Senior Member, IEEE) received the B.E., M.E., and Dr.Eng. degrees from Keio University, Yokohama, Japan, in 1993, 1995, and 1998, respectively. From 1996 to 2000, he was a Research Fellow with the Japan Society for the Promotion of Science. He was with the University of California at Santa Barbara, Santa Barbara, CA, USA, and Carnegie Mellon University, Pittsburgh, PA, USA, as a Visiting Scholar, in 1998 and 1999, respectively. From 2000 to 2020, he was with the Faculty of Environmental Engineering, The University of Kitakyushu, Kitakyushu, Japan. He is currently a Professor with the Faculty of Science and Engineering, Doshisha University. His research interests include image restoration, high dynamic range imaging, multiple image fusion, and digital filter design. He received the SIP Distinguished Contribution Award, in 2013, the IE Award, and the Contribution Award from IEICE, in 2017.