# SELF-SUPERVISED CALIBRATION OF THE DENOISING NETWORKS FOR HSI

*Orhan Torun*[*], *Seniha Esen Yuksel*
Hacettepe University
Dept. Elect. & Electron. Eng.
Beytepe, Ankara, Türkiye

*Erkut Erdem*
Hacettepe University
Dept. Comp. Eng.
Beytepe, Ankara, Türkiye

*Aykut Erdem*
Koç University
Dept. Comp. Eng.
TR-34450 Istanbul, Türkiye

## ABSTRACT

Typically, neural networks are trained using supervised learning (SL) and evaluated on unseen data. This type of training relies on a substantial amount of data, including clean images. However, in the case of hyperspectral images (HSIs), acquiring a large number of images along with clean versions can be challenging and expensive. This study proposes a two-stage learning strategy to train the model for HSI data with previously unseen noise patterns. The first stage involves supervised learning to train the model on noisy and clean data pairs. The second stage incorporates self-supervised calibration using only noisy data to adapt the model to specific noise patterns. For the latter, to estimate the middle spectral band, we leverage the information from its neighboring band as a target. To ensure the network learns meaningful relationships rather than merely copying the input, we strategically create a blind spot by excluding the target band from the input data. Therefore, our self-supervised learning technique is named as Blind Band Self-Supervised (BBSS) Learning. Our approach has been shown to improve the accuracy of the model for noisy HSIs, even when the network did not previously encounter the specific noise patterns in SL.

***Index Terms***— self-supervised, learning, calibration, unseen data, HSI

## 1. INTRODUCTION

Hyperspectral imaging is a technique that captures detailed spectral information by recording a large number of spectral bands in an image. This method is commonly used in applications such as agriculture and environmental monitoring, but the high dimensionality of resulting data can pose challenges for data processing and analysis. One significant challenge is the presence of noise in hyperspectral images (HSIs), which can decrease the performance of subsequent analysis tasks.

Recently, the use of neural networks for HSI denoising has become increasingly popular due to their ability to capture spatial and spectral features of the data effectively. However, one of the critical aspects of HSI denoising is the training of the network. One common approach is to use supervised learning (SL), where the network is trained using a set of clean and noisy pairs. This enables the network to learn from clean images and develop an understanding of the correct output for a given input. Initially, techniques for grayscale or RGB image denoising were adapted for HSI by modifying the input and output filter sizes and treating them as a single band [1, 2]. This approach is referred to as single-to-single (S2S) learning. However, these 2D filter-based approaches fail to fully exploit the abundant spectral information present in HSIs. Subsequently, networks specifically designed for HSI have been developed to leverage both the spectral and spatial information [3–8]. These models make use of multiple spectral bands from both clean and noisy pairs to train deep networks, a methodology which we refer to as multi-to-multi (M2M) learning. Recently, a highly performing SL method [9–12] has been developed, which uses multiple bands to estimate the middle band (multi-to-single, M2S). This enables the model to capture more information and enhance prediction performance.

Another approach is to employ self-supervised learning, where the network is solely trained using noisy data and is able to identify patterns and relationships within the data on its own. Nguyen *et. al.* [13] proposed using Stein's unbiased risk estimation (SURE) as a loss for training a CNN. SURE is an unbiased estimate of the mean-squared-error (MSE) and can be calculated using only noisy HSI. This approach allows for self-supervised training of the CNN without the need for clean data. However, since SURE is designed for Gaussian noise [14], its performance on complex noise is limited. In [15, 16], different generative networks were proposed and trained based on deep image prior (DIP) strategy [17] for HSI denoising. In [18], the subspace representation coefficients (referred to as eigenimages) of the HSI are used to propose a learning approach for generating pairs of noisy-noisy training eigenimages from noisy eigenimages, without relying on clean data during network training. However, the success of these studies depends on having prior knowledge, such as the number of endmembers, for different HSIs.

In our study, we introduce a two-stage learning strategy for HSI analysis, overcoming limitations in noise adaptation. The first stage employs a M2S supervised learning method,

[*]This project is funded by TUBITAK Project 123E385, and the BAGEP awards from the Science Academy to SEY.

crucial for the model to recognize dense noise patterns. Traditionally, training pairs are generated by adding synthetic noise [3–5, 10], but this often fails to mimic real-world noise distributions, leading to suboptimal performance. To address this, the second stage introduces a novel self-supervised calibration approach, drawing from Noise2Noise (N2N) [19, 20] and Noise2Void (N2V) [21] strategies. This method uses the intrinsic structure of the data, allowing the model to learn without clean target data. In this self-supervised phase, we estimate a middle spectral band using its neighboring bands. A key feature is the creation of a 'blind spot' by excluding the target band from the input, preventing the model from simply replicating the input and encouraging it to learn complex patterns within the spectral data. Hence, we call our model blind band self-supervised (BBSS) learning. This two-stage approach starts with pre-training the model on dense noise data, followed by self-supervised calibration on test data to adapt to specific sparse noise patterns. This strategy enhances the model's performance and efficiency, effectively improving accuracy in various types of HSIs.

## 2. THE PROPOSED DENOISING METHOD

In this study, we suggest a two-step learning method for denoising HSIs given in Fig. 1. As mentioned in Sec. 1, the first stage involves using a supervised learning technique called M2S to train a deep neural network model for denoising individual bands of the HSI. This model is trained using a dataset consisting of pairs of noisy and clean single bands, and it considers both multi-scale spatial and multi-scale spectral adjacent bands in order to perform denoising. In the second stage, we propose a self-supervised learning approach for calibrating the proposed network on unseen noisy data without the need for clean data. Our approach is inspired by N2N [22] and N2V [21] and involves using the neighboring band as the target while creating a blind spot in the input by removing the target band. This allows us to calibrate the network and improve its performance on unseen noisy data. In general, N2N learning is a method for denoising images in which only noisy observations are available. It is based on the idea that if two different noisy observations of the same image are given, the difference between the two observations can be used to estimate the underlying clean image. However, this method requires two different images with the same content and independently corrupted by noise.

Our approach, named *blind band self-supervised* (BBSS), is specifically designed for calibrating the proposed model using a single HSI, taking into account the fact that HSIs have many bands with high-resolution spectral information ($\leq$ 10 nm) and assuming that each band is corrupted independently. In this context, the goal of the training task is to find the set of parameters $\theta$ that minimize the loss between the output of the network function $D_\theta(\mathbf{Y}_i)$ and the target spectral band $\mathbf{X}_i$ for a given input spectral band $\mathbf{Y}_i$, with the assumption of S2S training. This can be written as:

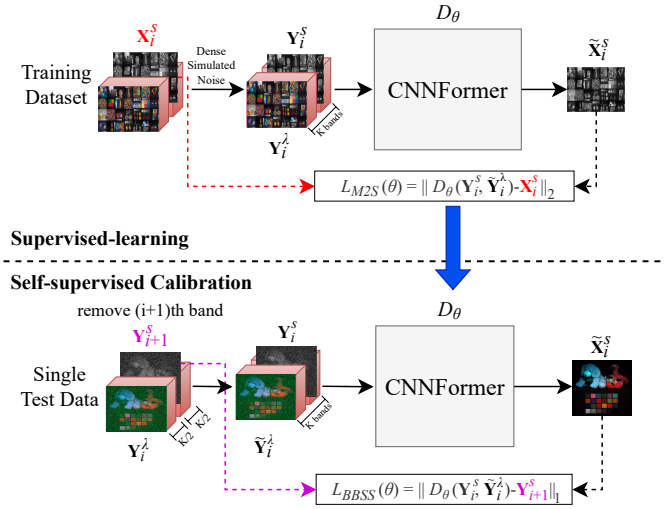


**Fig. 1**. The method overview of a two-stage learning scheme.

$$\text{argmin}_\theta \mathbb{E}_{(\mathbf{Y}_i, \mathbf{X}_i)} \{ L(D_\theta(\mathbf{Y}_i), \mathbf{X}_i) \} \tag{1}$$

Based on the N2N [19], we propose using neighboring band as target to minimize the loss between the output of the network function $D_\theta(\mathbf{Y}_i)$ and the adjacent spectral band $\mathbf{Y}_{i+1}$ for a given input spectral band $\mathbf{Y}_i$ as follows:

$$\text{argmin}_\theta \mathbb{E}_{(\mathbf{Y}_i, \mathbf{Y}_{i+1})} \{ L(D_\theta(\mathbf{Y}_i), \mathbf{Y}_{i+1}) \} \tag{2}$$

By following the M2S learning approach, we can modify the above expression for denoising each band to include both the spectral information from the neighboring bands and the spatial information as input:

$$\text{argmin}_\theta \mathbb{E}_{(\mathbf{Y}_i, \mathbf{Y}_{i+1})} \{ L(D_\theta(\mathbf{Y}_i, \mathbf{Y}_i^\lambda), \mathbf{Y}_{i+1}) \} \tag{3}$$

where $\mathbf{Y}_i^\lambda$ is the set of $K$ neighboring spectral bands of $\mathbf{Y}_i$, including the target spectral band $\mathbf{Y}_{i+1}$. However, since the target is provided as input, the network will simply learn the identity mapping, as expected. To prevent this, we create a blind spot in the spectral bands by removing the target band from the input, similar to what the N2V [21] method does for a single pixel. Therefore, Eq. 3 can be rewrite as:

$$\text{argmin}_\theta \mathbb{E}_{(\mathbf{Y}_i, \mathbf{Y}_{i+1})} \{ L(D_\theta(\mathbf{Y}_i, \tilde{\mathbf{Y}}_i^\lambda), \mathbf{Y}_{i+1}) \} \tag{4}$$

where $\tilde{\mathbf{Y}}_i^\lambda$ is obtained by removing the target band $\mathbf{Y}_{i+1}$ from the set of adjacent spectral bands, resulting in a blind spot in the input data that prevents the network from learning the identity mapping.

In our work, we utilize a modified hybrid CNN-Transformer architecture, adapted from [23], which we named CNNFormer for HSI data analysis. In order to showcase the effectiveness

(a) (b) (c) (d) (e) (f)

**Fig. 2**. A visual representation demonstrating the effectiveness of the two-stage learning approach. (a) False-color original CAVE image with bands (30,15,10), (b) Noisy image where each band is polluted by GN and one-third of bands are randomly chosen to add IN with intensity ranged from 10% to 70%, (c) Result of QRNN3D pre-training on GN, (d) Result of TRQ3D pre-training on GN, (e) Result of CNNFormer pre-training on GN, (f) Result of CNNFormer using blind band self-supervised learning after 34 epochs.

of two-stage learning, we conducted BBSS training with previously unseen noise during the training phase and obtained highly successful outcomes. We consider different loss functions for each stage of our two-stage learning strategy. In the SL stage, given a M2S training set $\{(\mathbf{Y}_i^s, \mathbf{Y}_i^\lambda), \mathbf{X}_i^s\}_{i=1}^N$ where $N$ is the number of training patches, $\mathbf{X}_i^s$ is a single-band clean patch of noisy low-quality patch $\mathbf{Y}_i^s$, and $\mathbf{Y}_i^\lambda$ is the noisy $K$ adjacent spectral bands of $\mathbf{Y}_i^s$. The loss function of the proposed denoiser ($D_\theta$) with the parameter set $\theta$ is:

$$\mathcal{L}_{M2S}(\theta) = \frac{1}{2N}\sum_{i=1}^{N} \|D_\theta(\mathbf{Y}_i^s, \mathbf{Y}_i^\lambda) - \mathbf{X}_i^s\|_2 \qquad (5)$$

Given a BBSS training set $\{(\mathbf{Y}_i^s, \tilde{\mathbf{Y}}_i^\lambda), \mathbf{Y}_{i+1}^s\}_{i=1}^N$ where $\mathbf{Y}_{i+1}^s$ is an adjacent patch of low-quality patch $\mathbf{Y}_i^s$, and $\tilde{\mathbf{Y}}_i^\lambda$ is obtained by removing the target band ($\mathbf{Y}_{i+1}^s$) from adjacent spectral bands, which creates a blind spot in the resulting data to prevent the network from learning the identity mapping. The loss function of the proposed self-supervised training scheme is defined as follows:

$$\mathcal{L}_{BBSS}(\theta) = \frac{1}{2N}\sum_{i=1}^{N} \|D_\theta(\mathbf{Y}_i^s, \tilde{\mathbf{Y}}_i^\lambda) - \mathbf{Y}_{i+1}^s\|_1 \qquad (6)$$

## 3. EXPERIMENTS

In this section, we present both quantitative and qualitative results of our proposed method on simulated test data. First, we trained our networks on Gaussian-noisy data for 50 epochs, following the strategy described in [5,24]. Then, we tested the Gaussian noise-trained models on a scenario where we added non-i.i.d. Gaussian noise (GN) to all bands and randomly selected one-third of the bands to receive impulse noise (IN), with intensity ranging from 10% to 70%. Next, we initialize CNNFormer model with the weights from the 50th epoch and calibrate it using the proposed BBSS learning strategy on the noisy test data. This self-supervised training phase consisted of 34 epochs. In Table 1, we show the quantitative results of this experiment.

**Table 1**. Quantitative results on the CAVE dataset.

| Method | MPSNR↑ | MSSIM↑ |
|---|---|---|
| Noisy HSI | 16.421 | 0.137 |
| BM4D [25] | 27.951 | 0.592 |
| BCTF-HSI [26] | 30.773 | 0.759 |
| NMoG [27] | 26.711 | 0.608 |
| LLRGTV [28] | 31.468 | 0.800 |
| GLF [29] | 29.370 | 0.767 |
| QRNN3D-P[1] [5] | 25.119 | 0.599 |
| TRQ3D-P[1] [24] | 24.855 | 0.604 |
| CNNFormer-P[1] | 27.804 | 0.646 |
| CNNFormer + BBSS | **34.341** | **0.904** |

In Fig. 2, we present the original test data selected from the CAVE data, its noisy version along with the denoising results of QRNN3D [5], TRQ3D [24], and two versions of our CNNFormer model, one trained only on GN and the other one with BBSS calibration being applied. As the QRNN3D, TRQ3D and our CNNFormer were not trained with impulse noise during SL phase, the output results exhibit artifacts. The results indicate that the network models trained solely on GN was only partially able to reduce the impulse noise in the test data, leading to low accuracies and overall poor performance. This is likely due to the fact that the network models did not encounter similar noise during the supervised training phase. Networks trained on Gaussian noise focus on smoothing out small variations. However, they are not equipped to handle the drastic changes caused by impulse noise. On the other hand, the benefits of using the BBSS calibration stage to improve the performance of the CNNFormer network are clearly evident. According to the metric scores of both mean peak signal-to-noise ratio-MPSNR, and mean structural sim-

[1]-P stands for the Gaussian noise pre-trained models.

ilarity index-MSSIM, we have achieved highly successful results compared to the classical methods listed in Table 1. Additionally, we found the calibrating process to be fast and efficient, as it was completed quickly using a single dataset. Overall, the two-stage training process demonstrated strong performance and efficiency in reducing noise that the network had not previously encountered and increasing the adaptation of the proposed network.

## 4. CONCLUSION

To sum up, we demonstrated that our proposed two-stage training approach produces promising outcomes. By enhancing the network's supervised performance, which represents the highest achievable level for BBSS learning, we can surpass the performance of all existing state-of-the-art methods. Moving forward, our future efforts will involve dedicating efforts to improving the quality of the network's supervised learning. Additionally, we intend to investigate the applicability of the two-stage learning method on various datasets and real-world noisy data.

## 5. REFERENCES


[1] Y. Chang et al., "Hsi-denet: Hyperspectral image restoration via convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 667–682, 2019. 1

[2] K. Zhang et al., "Beyond a Gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, 2017. 1

[3] E. Pan et al., "Sqad: Spatial-spectral quasi-attention recurrent network for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022. 1, 2

[4] Z. Wang et al., "Sscan: A spatial–spectral cross attention network for hyperspectral image denoising," *IEEE Geosci. Remote. Sens. Lett.*, vol. 19, pp. 1–5, 2022. 1, 2

[5] K. Wei et al., "3-d quasi-recurrent neural network for hyperspectral image denoising," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 363–375, 2021. 1, 2, 3

[6] F. Xiong et al., "Mac-net: Model-aided nonlocal neural network for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022. 1

[7] A. Dixit et al., "Unfold: 3d u-net, 3d cnn and 3d transformer based hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, 2023. 1

[8] M. Li et al., "Spectral enhanced rectangle transformer for hyperspectral image denoising," in *CVPR*, 2023, pp. 5805–5814. 1

[9] Q. Shi et al., "Hyperspectral image denoising using a 3-d attention denoising network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10348–10363, 2021. 1

[10] Q. Yuan et al., "Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1205–1218, 2019. 1, 2

[11] Q. Zhang et al., "Hybrid noise removal in hyperspectral imagery with a spatial–spectral gradient network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7317–7329, 2019. 1

[12] O. Torun et al., "Hyperspectral image denoising via self-modulating convolutional neural networks," *Signal Processing*, vol. 214, pp. 109248, 2023. 1

[13] H. Nguyen et al., "Hyperspectral image denoising using sure-based unsupervised convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3369–3382, 2020. 1

[14] M. Zhussip et al., "Extending stein's unbiased risk estimator to train deep denoisers with correlated pairs of noisy images," *NeurIPS*, vol. 32, 2019. 1

[15] Y. Miao et al., "Hyperspectral denoising using unsupervised disentangled spatiospectral deep priors," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022. 1

[16] O. Sidorov and J. Yngve Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *ICCVW*, Oct 2019. 1

[17] D. Ulyanov et al., "Deep image prior," in *CVPR*, 2018, pp. 9446–9454. 1

[18] L. Zhuang et al., "Eigenimage2eigenimage (e2e): A self-supervised deep learning network for hyperspectral image denoising," *IEEE Trans. Neural Netw. Learn. Syst.*, 2023. 1

[19] J. Lehtinen et al., "Noise2noise: Learning image restoration without clean data," *ICML*, p. 2965–2974, 2018. 2

[20] Y. Qian et al., "Hyperspectral image restoration with self-supervised learning: A two-stage training approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2021. 2

[21] A. Krull et al., "Noise2void-learning denoising from single noisy images," in *CVPR*, 2019, pp. 2129–2137. 2

[22] B. Rasti et al., "Noise reduction in hyperspectral imagery: Overview and application," *Remote Sensing*, vol. 10, no. 3, pp. 482, 2018. 2

[23] L. Chen et al., "Simple baselines for image restoration," *arXiv preprint arXiv:2204.04676*, 2022. 2

[24] L. Pang et al., "Trq3dnet: A 3d quasi-recurrent and transformer based network for hyperspectral image denoising," *Remote Sensing*, vol. 14, no. 18, pp. 4598, 2022. 3

[25] M. Maggioni et al., "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, 2013. 3

[26] K. Wei and Y. Fu, "Low-rank bayesian tensor factorization for hyperspectral image denoising," *Neurocomputing*, vol. 331, pp. 412 – 423, 2019. 3

[27] Y. Chen et al., "Denoising hyperspectral image with non-iid noise structure," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 1054–1066, 2017. 3

[28] W. He et al., "Hyperspectral image denoising using local low-rank matrix recovery and global spatial–spectral total variation," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 3, pp. 713–729, 2018. 3

[29] L. Zhuang and J.M. Bioucas-Dias, "Hyperspectral image denoising based on global and non-local low-rank factorizations," in *ICIP*, 2017, pp. 1900–1904. 3