A CNN with Noise Inclined Module and Denoise Framework for Hyperspectral Image Classification

Zhiquiang Gong, Ping Zhong, Senior Member, IEEE, Jiahao Qi, and Panhe Hu

Abstract—Deep Neural Networks have been successfully applied in hyperspectral image classification. However, most of prior works adopt general deep architectures while ignore the intrinsic structure of the hyperspectral image, such as the physical noise generation. This would make these deep models unable to generate discriminative features and provide impressive classification performance. To leverage such intrinsic information, this work develops a novel deep learning framework with the noise inclined module and denoise framework for hyperspectral image classification. First, we model the spectral signature of hyperspectral image with the physical noise model to describe the high intra-class variance of each class and great overlapping between different classes in the image. Then, a noise inclined module is developed to capture the physical noise within each object and a denoise framework is then followed to remove such noise from the object. Finally, the CNN with noise inclined module and the denoise framework is developed to obtain discriminative features and provides good classification performance of hyperspectral image. Experiments are conducted under two commonly used real-world datasets and the experimental results show the effectiveness of the proposed method. The implementation of the proposed method and other compared methods could be accessed at https://github.com/shendu-sw/noise-physical-framework.

Index Terms—Hyperspectral Image Classification, Convolutional Neural Networks (CNNs), Noise Inclined Module, Denoise framework, Diversity.

I. INTRODUCTION

Hyperspectral images by airborne or spaceborne sensors have played an important role in land cover detection, resource management, quality control, and others [1], [2]. The hundreds of spectral channels in the image can provide plentiful of spectral information to discriminate different objects of interest in classification task which tries to assign the unique land-cover label to each pixel. In spite of the high spectral resolution in hyperspectral image classification, there exists great noisy in each pixel which would make the great intra-class variance and inter-class similarity. This multiplies the difficulty to extract discriminative features from the image and leads to the degradation of the classification performance. Therefore, exploring an effective and proper model to remove noise from the image is urgent to perform effective image classification.

Nowadays, deep learning based methods have been introduced into hyperspectral image classification and contributed to dramatic improvements [3]. Deep learning networks (DNNs), especially convolutional neural networks (CNNs), can capture both the local and global information and thus provide discriminative features. To further improve the potential representational ability of CNNs to extract spatial and spectral features for hyperspectral image classification, most of prior works commonly focus on deepen or widen their network architectures [4] through the utilization of multi-scale information, spatial information, rotation variance, and others. For instance, [5] utilizes the multi-scale information and proposes a multiscale covariance maps for improvement of hyperspectral image classification. Similarly, [3] also focuses on embedding the multi-scale information into the deep model. [6] and [7] tries to leverage the spatial and spectral information of the hyperspectral image. [8] proposes a pyramidal bottleneck residual blocks which gradually increases the convolutional filters of the network to improve the spectral-spatial features extracted by the deep model. These prior methods pursue good performance over hyperspectral image classification by improving the model on spectral and spatial features while ignore the intrinsic physical characteristics of the image itself.

To utilize these intrinsic physical characteristics, some prior works utilize such physical prior to construct the training loss and supervise the training process of the deep model. As a representative work, [9] leverages the statistical properties of each class of hyperspectral image and construct the statistical loss for better representation. Similarly, [10] directly utilizes the high dimension property and models the hyperspectral image as nonlinear data manifold, and then constructs the training loss under such manifold assumption. These methods consider the class distributions of the image and construct the corresponding training loss which can supervise the training process and embed such class information with back propagation of the network. Thus good performance can be obtained with these additional information. However, these methods mainly use the class distribution from the view of training loss while ignore the potential of network module to extract these intrinsic physical characteristics.

As for hyperspectral images, each pixel is generated by radiance signals combined with the sensor noise as the adjacency radiance. These noises bring statistical variance to the spectral signature of pixels from the same class and therefore different classes in the image present great overlapping of the spectral curves. Such noise characteristics would negatively affect the deep model to discriminate different classes. To
solve the problem, motivated by [11], this work proposes a CNN to remove these noise and provide clean features from the image for better classification performance. Our method includes a noise inclined module to represent the generated noise from the image and a denoise framework to remove such noise. Such framework can provide clean features from the hyperspectral image and significantly improve the performance of the deep model.

To be concluded, the contributions of this work are twofold. First, this work develops a noise inclined module to extract noise information from the hyperspectral image. The module utilizes the noise space to represent the noise from the image. The diversity regularization and the large scale of noise space can guarantee a good representation of the noise from the image. Then, this work proposes a CNN framework to remove the extracted noise by the noise inclined module from the features, and provide clean and discriminative features to improve the classification performance.

The remainder of this paper is arranged as follows. Section II details the construction of the proposed framework for hyperspectral image classification. Experimental results and analysis over two real-world datasets are summarized in Section III. Section IV concludes this paper with some future directions.

II. PROPOSED METHOD

Denote $X = \{x_1, x_2, \ldots, x_N\}$ as the set of training samples of the hyperspectral image where $y_i$ is the corresponding label of the sample $x_i$ and $N$ is the number of training samples.

A. Physical Modelling

The hyperspectral image can be physically modelled by at-sensor radiance signal model. Generally, the total radiation signal reaching the sensor consists of three components, namely the reflected radiance (radiation reflected from the pixel of interest), the adjacency radiance (radiation reflected from the surface surrounding the pixel of interest and scattered by the air volume into the sensor) and other radiance from sensor noise. The reflected radiance generates the spectral signature of a certain object and the other radiances formulate the noise imposed on the object.

Therefore, for hyperspectral image processing, the spectral characteristics of each object can be defined by a single class spectrum or a target subspace combined with noise from the neighbors, sensors, and others [12], [13]. Let $x_{\alpha}^m$ denote the object of interest from $m$-th class and it can be formulated as

$$x_{\alpha}^m = s^m + n(x_{\alpha}^m)$$

where $s^m$ is the spectral signature of the $m$-th class, and $n(x_{\alpha}^m)$ is the random additive noise from the neighbors and sensors imposed on this specific object.

To represent these noises imposed on the image, this work denotes the noise space $S = \{n_1, n_2, \ldots, n_k\}$ where $n_i$ stands for $i^{th}$ base noise of the space. Then, the noise can be seen as the linear weighted of noises from multiple sources in noise space and it can be calculated as

$$n(x_{\alpha}^m) = \lambda_1(x_{\alpha}^m)n_1 + \lambda_2(x_{\alpha}^m)n_2 + \cdots + \lambda_k(x_{\alpha}^m)n_k$$

where $n_1, n_2, \ldots, n_k$ represents the base noise from different sources and $\lambda_i(x_{\alpha}^m)$ denotes the weight parameter of the noise $n_i$ ($i = 1, 2, \ldots, k$).

B. Noise Inclined Module

This subsection proposes the noise inclined module to learn these base noises adaptively, estimate the weight parameters, and provide the representation of the noise of each object in the noise space. The noise inclined module consists of three parts, namely the noise extraction module to extract noise from the features, the noise reconstruction module to reconstruct the noise in the noise space and the self-supervised module to update the noise space.

1) Noise Extraction Module: The noise extraction module is performed by the noise mapping operation. In general imaging process, the noise $n(x_{\alpha}^m)$ of $x_{\alpha}^m$ is highly related to the signature $s^m$, which can be represented as $n(x_{\alpha}^m) \propto s^m$. Based on Eq. 1, the noise is also related to the noised feature $x_{\alpha}^m$.

$$n(x_{\alpha}^m) \propto x_{\alpha}^m$$

Considering the imaging mechanism, $n(x_{\alpha}^m)$ can be calculated through the non-linear transformation of $x_{\alpha}^m$, which is implemented by the fully-connected layer in this work. Denote $\phi(x_{\alpha}^m)$ as the learned features from the CNN where $\phi(\cdot)$ represents the mapping function of the CNN. Then, the extracted noise can be calculated by

$$n_f(x_{\alpha}^m) = W\phi(x_{\alpha}^m) + b$$

where $W, b$ represent the weight and the bias of the layer, respectively.

2) Noise Reconstruction Module: The key process for noise reconstruction is to reconstruct the noise in the noise space based on the extracted noise from noise extraction module. The noise reconstruction process can be seen as the following optimization,

$$\min_{\lambda_1, \ldots, \lambda_k} L_u = L_r + L_s + \alpha L_d$$

where $L_r = \|n_f(x_{\alpha}^m) - \sum_{i=1}^k \lambda_i n_i\|$ denotes the reconstruction term, $L_s = \sum_{i=1}^k |\lambda_i(x_{\alpha}^m)|$ is the regularization term to enforce the unique reconstruction solution, and $L_d$ is the noise diversity promoting augment and $\alpha$ is a constant that controls tradeoff between task objective and diversity. This paper defines a angle-based difference measure for diversity, that is,

$$L_d = \frac{1}{k(k-1)} \sum_{i \neq m} n_i^T n_m$$

It should be noted that $L_d$ mainly performs for the update of $n_i$.

Instead of obtaining the values of $\lambda_i(x_{\alpha}^m)$ ($i = 1, \ldots, k$) through optimization methods, this work tries to estimate these values directly using the correlation between the base noise
and the noise features. Generally, the base noises $n_1, \ldots, n_k$ are expected to be independent to each other, therefore, this work tries to use the cosine similarity to estimate the values and impose the independent regularization over the base noises. 

As [11], we compute the cosine similarity $s_j(x_m^A)$ of the $j$-th base noise $n_j$ and the extracted noise $n_f(x_m^A)$.

$$s_j(x_m^A) = \frac{n_j n_f(x_m^A)}{\|n_j\|\|n_f(x_m^A)\|}$$  (7) 

Then, the pre-reconstructed noise is calculated by

$$n' = \sum_{j=1}^{k} s_j(x_m^A)n_i$$  (8) 

To preserve the energy of the noise, the final estimation of $\lambda_j$ given $x_m^A$ can be measured by

$$\lambda_j(x_m^A) = \frac{\|n_f(x_m^A)\|}{\|n'\|} s_j(x_m^A)$$  (9) 

Then, the final reconstructed noise can be formulated as

$$n_{res}(x_m^A) = \sum_{j=1}^{k} \lambda_j(x_m^A)n_j$$  (10) 

3) Noise Self-Supervised Module: These base noises in noise space are updated through self-supervised learning based on the gradient descent of Eq. 5 and it can be calculated as

$$\frac{\partial L_u}{\partial n_i} = \frac{\partial L_s}{\partial n_i} + \alpha \frac{\partial L_d}{\partial n_i}$$  (11) 

Based on derivation rule,

$$\frac{\partial L_s}{\partial n_i} = -2\lambda_i(x_m^A)(n_f(x_m^A) - \sum_{j=1}^{k} \lambda_j(x_m^A)n_j)$$  (12) 

Besides, $\frac{\partial L_d}{\partial n_i}$ is used to decrease the redundancy and make independent between different base noises, and it can be calculated by

$$\frac{\partial L_d}{\partial n_i} = \frac{2}{k(k-1)} \sum_{l \neq i} n_l$$  (13) 

In addition, this work uses the decay way for self-supervised update. The final self-supervised update of the base noise can be formulated as

$$n_j^{(k+1)} = \beta n_j^{(k)} + (1 - \beta) \times \frac{\partial L_u}{\partial n_i}$$  (14) 

where $\beta \in [0, 1]$ is the decay rate. This strategy is also the self-supervised updating just as [11].

C. The Proposed Framework

This work proposes a novel framework with noise inclined module and denoise operation to extract clean and discriminative features from the hyperspectral image. The noise inclined module is used to extract the physical noise from the image, reconstruct the noise in a noise space, and finally update the noise space under self-supervised way. 

Then, the denoise operation is followed to provide the clean features from the image, and the denoise operation can be implemented by

$$f_{clean}(x_m^A) = \phi(x_m^A) - n_{res}(x_m^A)$$  (15) 

The obtained clean features are then used for the classification of different objects in the image. The flowchart of the proposed method is shown in Fig. 1.

This work utilizes the center loss [14] for the training of the deep model to increase the inter-class variance of the learned clean features, and the loss can be formulated as

$$L_C = \frac{1}{2} \sum_{x_m^A \in X} \|f_{clean}(x_m^A) - c_m\|^2$$  (16)
where \( c_m \) denotes the class center in the loss.

III. EXPERIMENTAL RESULTS

A. Experimental Setups
To validate the effectiveness of the proposed method, we conduct experiments over two real-world hyperspectral image datasets, namely Pavia University dataset and Houston 2013 dataset.

Pavia University data [15] was collected by the reflective optics system imaging spectrometer (ROSIS-3) sensor ranging from 0.43 to 0.86 \( \mu m \) with a spatial resolution of 1.3m per pixel. The data contains 610 \( \times \) 340 pixels with 115 bands where 12 bands have removed due to noise and absorption and the remaining 103 channels are used. 42776 samples from 9 classes have been labeled for experiments.

Houston 2013 data [16] was acquired over the University of Houston campus and the neighboring urban area with a resolution of 2.5 m/pixel, and provided as part of the 2013 IEEE Geoscience and remote Sensing Society data fusion contest. It consists of 349 \( \times \) 1905 pixels of which a total of 15011 labeled samples divided into 15 classes have been chosen for experiments. Each pixel denotes a sample and contains 144 spectral bands ranging from 0.38 to 1.05 \( \mu m \).

Pytorch is used as the deep learning framework and all the experiments are carried out on a PC with an Intel@ Xeon(R) Gold 6226R CPU and Nvidia Quadro RTX 6000 GPU. For pavia university data, 200 samples per class are randomly selected for training and the remainder are for testing while for Houston2013 data, the training and the testing set are following the 2013 GRSS datafusion contest. The dimension of the learned features is set to 400. The learning rate is set to \( 1e^{-4} \) and the training batch is set to 4. The hyperparameter \( \alpha, \beta \) is set to 1, 0.9, respectively.

B. General Performance
At first, we present the general performance of the proposed method for hyperspectral image classification. In this set of experiments, the number of base noise in the noise space is set to 1024. The backbone of the proposed method is the 3-D CNN.

As showed in table I and II, the proposed method can improve the performance of vanilla CNNs by introducing the physical information in the training process. As for the Pavia University data and the Houston2013 data, the CNNs with the proposed framework can significantly improve the CNNs’ performance. For Pavia University, the accuracy can increase from 94.97\% \( \pm \) 0.69\% to 96.56\% \( \pm \) 0.47\% with the proposed framework. For Houston2013 data, the accuracy can be improved by about 6\%. This indicates that the physical information is effective for the training of the deep CNNs in the hyperspectral image classification literature.

Furthermore, Fig. 3 and 4 shows the classification maps of different datasets. It also demonstrates that the proposed method using the physical information can remarkably decrease the classification errors and further improve the classification performance.

C. Performance with Different Number of Base Noise
In this set of experiments, we test the performance of the proposed framework under different number of base noise in the noise space. The number of base noise in the noise space denotes the representational ability of the noise space. Generally, a larger number of base noise can provide a better representation of the noise in the signature characteristics of the hyperspectral image and thus the clean feature can be more discriminative. However, excessive number of base noise increases the learned parameters and makes redundancy between the base noise which makes the degradation of the classification performance. Figure 5 shows the tendencies of the performance under different number of base noise. As for Pavia University data, the classification performance achieves the best with base noise of 2048 while for Houston2013 data, the classification performance achieves best under 1024 base noise. The tendencies of the classification performance over the two datasets validate the conclusions before.

D. Performance with Different Sizes of neighbors
In this set of experiments, we present the performance under different sizes of neighbors. Fig 6 shows the tendencies with different neighbor sizes over the two datasets. As the figure shows, the 3-D CNN is not very friendly to extract features from hyperspectral image with larger neighbor sizes. Generally, spectral-spatial classification with larger neighbor size contains more spatial information and thus provide more discriminative features for classification. While due to the limitation of the backbone model, the classification performance even decreases with excessive large spatial sizes (e.g. \( 9 \times 9 \)).

More importantly, as Fig. 6 shows, the classification performance can be significantly improved with the proposed method by using physical information under different sizes of neighbors. Especially, for Pavia University data, the classification accuracy can be increased by 1.72\% under \( 5 \times 5 \) neighbor and 2.32\% under \( 9 \times 9 \) neighbor. While for Houston2013 data, the classification accuracy can even be increased by about 6\% under \( 5 \times 5 \) neighbor.

E. Comparisons with Other Methods
To thoroughly evaluate the effectiveness of the proposed method, this work selects Support Vector Machine(SVM), 3-D CNN [1], PResNet [17], HybridSN [18], SSFTTNet [19] as baselines. Table I and II show the comparisons over the two data sets, respectively. The experimental results in each table are with the same experimental setups and \( 5 \times 5 \) neighbors are used as the spatial information in the experiments.

From table I, we can find that the proposed method which achieves 96.56\% \( \pm \) 0.47\% OA outperforms these baseline methods over Pavia University data where the best provides an OA of 94.97\% \( \pm \) 0.69\% by 3-D CNN. As for Houston2013 data, it can be noted that the proposed method also provides an superior performance (84.41\% \( \pm \) 0.73\%) by proposed method vs 80.96\% \( \pm \) 1.44\% by SSFTTNet). Furthermore, the proposed method can provide a better performance by using a different backbone CNN. This indicates that the proposed method can
provide an comparable or better performance than other recent methods.

**IV. Conclusions**

In this work, a novel CNN with noise inclined module and denoise framework is developed for hyperspectral image classification. First, the physical noise model is constructed to represent the noise information. Then, the noise inclined module is developed to adaptively learn the noise from each object. Finally, the denoise framework is followed to remove the extracted noise information and provide clean features to better discriminate different objects. The experimental results over two challenging datasets have demonstrated the effectiveness of the proposed method.

In future work, it is suggested investigating other effective architecture using the intrinsic information. Another important future avenue of exploration is other intrinsic structures of the hyperspectral image for effective learning of deep models, such as the manifold structure of class distribution.
Fig. 3. Classification mapping by different methods over Pavia University (Overall accuracies). (a) Ground truth; (b) 3-D CNN (95.42%); (c) HybridSN(46.92%); (d) PResnet(94.16%); (e) SSFTTnet(92.72%); (f) Proposed Method(97.24%).

Fig. 4. Classification mapping by different methods over Houston2013 (Overall accuracies). (a) Ground truth; (b) 3-D CNN (77.28%); (c) HybridSN(79.31%); (d) PResnet(81.12%); (e) SSFTTnet(81.98%); (f) Proposed Method(85.46%).

Fig. 5. Classification performance with different number of base noise in the noise space over (a) Pavia University; (b) Houston2013.

REFERENCES

[1] A. B. Hamida, A. Benoit, P. Lambert, and C. B. Amar, “3-d deep learning approach for remote sensing image classification,” IEEE Transactions on Geoscience and Remote Sensing, vol. 56, no. 8, pp. 4420–4434, 2018.

[2] S. Shabbir and M. Ahmad, “Hyperspectral image classification—traditional to deep models: A survey for future prospects,” arXiv preprint arXiv: 2101.06116, 2021.

[3] Z. Gong, P. Zhong, Y. Yu, and et al, “A cnn with multiscale convolution and diversified metric for hyperspectral image classification,” IEEE TGRS, vol. 57, no. 6, pp. 3599–3618, 2019.

[4] Y. Yuan, C. Wang, and Z. Jiang, “Proxy-based deep learning framework for spectral–spatial hyperspectral image classification: Efficient and robust,” IEEE TGRS, vol. 60, pp. 1–15, 2021.

[5] N. He, M. E. Paoletti, J. M. Haut, L. Fang, S. Li, A. Plaza, and J. Plaza, “Feature extraction with multiscale covariance maps for hyperspectral image classification,” IEEE TGRS, vol. 57, no. 2, pp. 755–769, 2018.

[6] H. Lee and H. Kwon, “Going deeper with contextual cnn for hyperspectral image classification,” IEEE TIP, vol. 26, no. 10, pp. 4843–4855, 2017.

[7] B. Xi, J. Li, Y. Li, R. Song, Y. Xiao, Y. Shi, and Q. Du, “Multi-direction networks with attentional spectral prior for hyperspectral image classification,” IEEE TGRS, vol. 60, pp. 1–15, 2022.

[8] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, “Deep pyramidal residual networks for spectral–spatial hyperspectral image classification,” IEEE TGRS, vol. 57, no. 2, pp. 740–754, 2018.

[9] Z. Gong, P. Zhong, and W. Hu, “Statistical loss and analysis for deep learning in hyperspectral image classification,” IEEE TNNLS, vol. 32, no. 1, pp. 322–333, 2021.

[10] Z. Gong, W. Hu, X. Du, P. Zhong, and P. Hu, “Deep manifold embedding for hyperspectral image classification,” IEEE Transactions on Cybernetics, 2021.
Fig. 6. Classification performance with different neighbor sizes over (a) Pavia University; (b) Houston2013.

[16] C. Debes, A. Merentitis, R. Heremans, J. Hahn, N. Frangiadakis, T. Van Kasteren, and et al., “Hyperspectral and lidar data fusion: Outcome of the 2013 grss data fusion contest,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 7, no. 6, pp. 2405–2418, 2014.

[17] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, “Deep pyramidal residual networks for spectral–spatial hyperspectral image classification,” IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 2, pp. 740–754, 2018.

[18] S. K. Roy, G. Krishna, S. R. Dubey, and B. B. Chaudhuri, “Hybridsn: Exploring 3-d-2-d cnn feature hierarchy for hyperspectral image classification,” IEEE Geoscience and Remote Sensing Letters, vol. 17, no. 2, pp. 277–281, 2019.

[19] L. Sun, G. Zhao, Y. Zheng, and Z. Wu, “Spectral-spatial feature tokenization transformer for hyperspectral image classification,” IEEE Transactions on Geoscience and Remote Sensing, 2022.