A Multidimensional Sequential Convolutional Neural Network-Based Method for Hyperspectral Image Classification

Qiongdan Huang, Jiapeng Wang, Liang Li, and Shilin Kang

Abstract—Convolutional Neural Networks (CNNs) have emerged as a leading model in hyperspectral image (HSI) classification, owing to their unique features such as parameter sharing and localized receptive fields. Despite these advantages, current CNN-based approaches face significant challenges, including complex model architectures, the inability to fully extract deep features and the underutilization of CNNs kernels. This study introduces a novel multidimensional sequential CNN model, termed the 3D-1D-2D SCNN, specifically designed for HSI classification. The model begins with Principal Component Analysis (PCA) to reduce the dimensionality of the HSI data. It then sequentially applies multidimensional convolutions to extract a diverse array of features from the HSI data. The classification process is finalized using a linear classifier, ensuring precise categorization of the hyperspectral data. Simulation results demonstrate that this model outperforms traditional methods in classification accuracy across four widely-used HSI datasets.

Index Terms—Image processing, Hyperspectral, Image classification, SCNN.

I. Introduction

HYPERSPECTRAL images (HSIs) [1] are characterized by their exhaustive spectral details and information beyond the visible wavelength spectrum. These images enable the analysis of correlations among adjacent spectral bands and the response values of specific wavelengths, facilitating the classification of objects that are indistinguishable using conventional visible light imaging techniques. As a result, HSI classification has become increasingly prominent across various fields such as agriculture, healthcare, food sciences, and mineral exploration [2–5].

In the realm of HSI classification, convolutions across three distinct dimensions are extensively employed to extract features [6]. The one-dimensional convolutional neural network (1D-CNN) is predominantly utilized to capture the

Qiongdan Huang is a professor at the School of Communication and Information Engineering, Xi'an University of Post and Telecommunications, Xi'an, 710121, China. (corresponding author, phone: +8602988166381; e-mail: limitless010@163.com).

Jiapeng Wang is a postgraduate student at the School of Communication and Information Engineering, Xi'an University of Post and Telecommunications, Xi'an, 710121, China. (e-mail: 1075861687@qq.com).

Liang Li is a postgraduate student at the School of Communication and Information Engineering, Xi'an University of Post and Telecommunications, Xi'an, 710121, China. (e-mail: liliang3790@163.com).

Shilin Kang is a postgraduate student at the School of Communication and Information Engineering, Xi'an University of Post and Telecommunications, Xi'an, 710121, China. (e-mail: 2027365196@qq.com).

spectral features of HSIs, treating the spectral information of each pixel as a separate channel input for processing. Hu *et al.* [7] were pioneers in applying 1D-CNN to HSI classification, offering novel insights into this technique. Following this foundational work, researchers have concentrated on enhancing the 1D-CNN algorithm. For instance, Lu *et al.* [8] significantly reduced the computational demands and processing time of the 1D-CNN by optimizing the calculation of gradient descent and learning rates. Addressing the limited ability of the 1D-CNN to discriminate features, Gao *et al.* [9] introduced a modified 1D-CNN architecture featuring multi-size feature fusion. Through optimizing the parameters of the sampling and pooling layers, this approach enabled effective integration of features at various depths within the 1D-CNN, thereby yielding a more comprehensive representation of the feature information. The two-dimensional convolutional neural network (2D-CNN) is primarily utilized for extracting spatial features from HSIs. In this approach, spectral features of target pixels and their surrounding pixels within a specified neighborhood are processed as input samples. Historically, the significance of spatial information in HSI classification was underappreciated. Makantasis *et al.* [10] addressed this by applying PCA to reduce HSI dimensionality and employing a 2D-CNN with two layers of uniform convolutional kernels to extract spatial features. However, the use of fixed-sized kernels often leads to suboptimal feature extraction. In response, Chen *et al.* [11] developed convolutional kernels of varying shapes to enhance the extraction of spatial features, considering the translation and scaling invariance of these features. Furthermore, the three-dimensional convolutional neural network (3D-CNN), known for its efficacy in processing video sequences [12], has been adapted for HSI classification due to the analogous nature of HSIs and video data [13]. The unique kernels of the 3D-CNN allow for the integration of spectral and spatial information, yielding positive results [14]. Despite these advantages, the high computational demands of the 3D-CNN limit its independent use. Recent studies [15] have sought to optimize model training by fine-tuning parameters, thus reducing computational costs and enhancing efficiency.

The aforementioned approaches leverage CNNs for feature extraction from HSIs. However, standalone 2D-CNNs often fail to adequately capture spectral information from spatial features, leading to subpar classification accuracy. Similarly, standalone 3D-CNNs, while detailed, suffer from computational complexity and diminished performance in classifying categories with similar spectral textures. To mitigate these challenges, various hybrid CNN models have been developed. RoySK *et al.* [16] pioneered the HybridSN

model, which sequentially integrates three-dimensional and two-dimensional convolutional networks. This model adeptly extracts both spectral-spatial and deep spatial features from HSIs. Nevertheless, it does not adequately account for the extensive spectral features in HSIs, resulting in significant spectral feature loss. To improve network robustness and accelerate convergence, Ghaderizadeh *et al*. [17] incorporated 3D depth-wise separable convolution modules and fast convolution modules into the HybridSN framework. Positioned between the 3D and 2D convolutional layers, these modules have shown exceptional performance in scenarios with adversarial noise and limited training data. In a novel approach, Wang *et al*. [18] introduced a multidimensional parallel convolutional neural network (3D-2D-1D PCNN), enabling simultaneous feature extraction from multiple dimensions and linear feature fusion across different dimensions. However, this parallel network structure adds an additional layer of complexity to the model.

In response to the complexity of existing neural network-based approaches for HSI classification, the inability to fully extract deep features and the underutilization of CNNs kernels, this paper introduces a novel method termed the Multidimensional Sequential 3D-1D-2D Algorithm based on CNNs. This algorithm begins with the application of PCA [19] to reduce the dimensionality of raw HSI data. Following this reduction, a 3D-CNN module captures joint spatial-spectral information at multiple scales. This is followed by a 1D-CNN module designed to extract abstract spectral features from the previously captured spatial-spectral data. Additionally, a 2D-CNN module refines the extraction of spatial features following the 1D-CNN module. The process culminates with the classification of the HSI data using a linear classifier, aimed at achieving accurate and efficient categorization.

II. PROPOSED

A. Hyperspectral image preprocessing

HSIs encompass a wide array of spectral bands, many of which contain irrelevant data. Dimensionality reduction is a critical process that selectively retains essential and representative features while eliminating minor and redundant information [20]. This process not only simplifies the feature space but also enhances the efficiency of classification algorithms [21]. Additionally, HSIs are often affected by various types of noise, including atmospheric interference and instrumental noise. Dimensionality reduction helps to mitigate the effects of such noise, thereby improving the signal-to-noise ratio and enhancing the accuracy of image classification. PCA is a commonly used technique for dimensionality reduction in HSIs. This method transforms the HSI data into a matrix format, where each row corresponds to an individual pixel and each column represents a spectral band. The HSI data is then standardized, with the mean of each spectral band adjusted to zero and its variance normalized to one. Following standardization, the covariance matrix of the transformed data is computed. Eigenvalue decomposition is then performed on the covariance matrix to extract eigenvalues and their corresponding eigenvectors. These eigenvectors are arranged in descending order according to the magnitude of the eigenvalues, thereby organizing the principal components by their significance. The magnitude of the eigenvalues

embodying empirical knowledge, guides the selection of the primary components to be retained. The application of PCA effectively reduces the spectral dimension of the HSI data while preserving the most critical and representative information. This reduction simplifies data processing and enhances computational efficiency. Initially, the original HSI is denoted as $I \in R^{H \times W \times M}$, where H represents the image height, the width W, and M the number of spectral bands. After applying PCA, the initial L principal components along the spectral dimension are retained, transforming the reduced HSI into $X \in R^{H \times W \times L}$.

In HSI classification, individual pixels typically provide limited information, such as spectral or color values, which may be insufficient for distinguishing between distinct categories. Additionally, single pixels are susceptible to noise, occlusions, and variations, leading to unstable classification outcomes. In this context, the pixels surrounding the target pixel often contain crucial features relevant to the target. By incorporating neighboring pixels as samples, a broader context is captured, enhances the model's stability. This approach leverages the shared characteristics of multiple pixels, thereby mitigating the influence of individual pixel anomalies on the classification results [22]. The process begins by expanding the boundaries of the input image to accommodate edge pixels. Subsequently, a $K \times K$ neighborhood around each target pixel is selected as samples, assuming that these adjacent pixels share identical labels with the central pixel. Following these preprocessing steps, the original hyperspectral image, initially with dimensions $H \times W \times M$, is transformed into N three-dimensional image blocks (patches) of size $K \times K \times L$. Here, where N denotes the total number of samples.

B. Multidimensional sequential SCNN model

The proposed multidimensional sequential SCNN model, as depicted in Fig. 1, employs a single-branch structure. This design facilitates the sequential processing of samples through the feature extraction module, thereby enhancing the capability to extract deep, abstract features. Initially, the model employs a 3D-CNN module to extract joint spatial-spectral features from the HSIs. The activation score of a neuron in a 3D-CNN convolution at location (x, y, z) in the i-th layer's j-th feature map is given by:

$$v_{i,j}^{x,y,z} = \phi\left(\sum_{\tau} \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} \sum_{r=0}^{R-1} \omega_{i,j,\tau}^{p,q,r} v_{(i-1),\tau}^{(x+p),(y+q),(z+r)} + b_{i,j}\right) \quad (1)$$

where ϕ denotes the activation function, $b_{i,j}$ is the bias of the ith layer's jth feature map, τ represents feature map from the i-1th layer that is currently connected, the heigh and width of the convolution kernels are represented by P and Q, and the depth of kernel along spectral dimension is represented by R. The weight $w_{i,j,\tau}^{p,q,r}$ connecting to the τth feature map is indexed by (p, q, r). In contrast to the standard 2D-CNN, where convolution kernels are constrained to the spatial dimensions (width and height), the kernels in 3D-CNN further extend to the channel dimension. This enables 3D-CNN to concurrently learn joint spatial-spectral features from HSI data.

Subsequently, the model employs a 1D-CNN to capture abstract spectral features from the joint features while min-
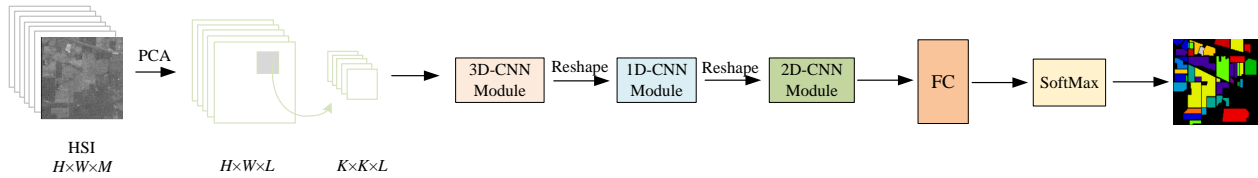
Fig. 1. The procedure of our multidimensional sequential CNN (SCNN) framework.

imizing significant information loss. The activation value of the neuron at position z in the j-th feature map of the i-th convolutional layer in a 1D-CNN is given by:

$$v_{i,j}^z = \phi\left(\sum_{\tau}\sum_{p=0}^{P-1} \omega_{i,j,\tau}^p v_{(i-1),\tau}^{(z+p)} + b_{i,j}\right) \qquad (2)$$

The parameters used are identical to those described in (1). Finally, a 2D-CNN module is employed to extract advanced abstract spatial features. The joint spatial-spectral features initially extracted by the 3D-CNN module preserve the global information inherent in the raw data. Building on this foundation, the spectral features captured by the 1D-CNN module are more localized and detailed. The 2D-CNN module subsequently processes these meaningful features, effectively filtering out irrelevant patterns and avoiding redundant features, thereby enhancing the compactness and efficiency of the model. Additionally, while the 1D-CNN primarily encodes spectral information, which may exhibit spatial similarities, the 2D-CNN leverages these spectral details to learn spatial combinations and shape patterns. This enhances the model's discriminative capability, crucial for improving classification accuracy.

This model leverages the intrinsic properties of convolutional kernels in CNNs and the rich spectral and spatial information available in HSI. By employing a sequential methodology, it effectively integrates diverse features, thereby enhancing its ability to extract abstract features. This approach ensures high precision in HSI classification while maintaining balanced computational efficiency. The dual focus on accuracy and efficiency significantly enhances the model's practicality, making it a robust solution for real-world applications in HSI.

As depicted in Fig. 2, the feature extraction module consists of three distinct CNNs: 3D-CNN, 1D-CNN, and 2D-CNN. The 3D-CNN module includes two layers of three-dimensional convolution followed by a normalization layer, arranged sequentially. After processing, the output from the 3D-CNN module is fed into the 1D-CNN module, which features a single layer of one-dimensional convolution designed to extract spectral features from the joint spatial-spectral data. The features extracted by the 1D-CNN are then passed to the 2D-CNN module, which employs two layers of two-dimensional convolution to analyze spatial relationships within the samples. These operations are performed in parallel, and the resultant features are subsequently combined. A final 2D convolution and a dropout layer are applied to consolidate the features of each sample.

Compared to methods that directly extract spectral, spatial, and joint features in parallel from the original image, the sequential feature extraction approach proposed in this paper

is more effective. It leverages the inherent data information within the HSI more efficiently. Low-level features extracted in earlier stages form the basis for the high-level features, while different levels of features complement each other. This synergy enables the acquisition of semantically rich and comprehensive feature descriptions, enhancing the model's overall performance in HSI classification.
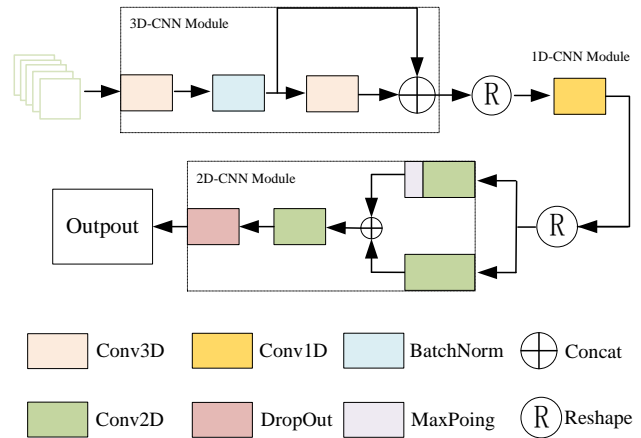

Fig. 2. The details of the feature extraction module.

The algorithm proposed in this paper opts not apply the 2D-CNN module directly to extract abstract spatial information following the use of the 3D-CNN module. This decision stems from the capability of two-dimensional convolution kernels to compress high-dimensional HSI data, condensing multi-channel information into single-channel feature maps, potentially resulting in the loss of many unextracted features. When conducting feature extraction on input images, 2D-CNN convolution layers require the number of channels in the convolution kernels to match that of the input data. The computational process is defined as follows:

$$v_{i,j}^{x,y} = \phi\left(\sum_{\tau}\sum_{p=0}^{P-1}\sum_{q=0}^{Q-1} \omega_{i,j,\tau}^{p,q} v_{(i-1),\tau}^{(x+p),(y+q)} + b_{i,j}\right) \qquad (3)$$

where the parameters used are identical to those denoted in Equation (1). The structure of 2D-CNN convolution kernels sharply compresses the information in HSI data after undergoing 2D-CNN convolutions, impeding subsequent feature extraction. Despite this substantial loss of data information, the importance of spatial information for HSI classification cannot be overstated. Therefore, the initial stage of this study employs 3D-CNN and 1D-CNN modules to extract spatial-spectral and spectral features, respectively. Only after these stages is the 2D-CNN module used to capture abstract spatial features. This strategy helps minimize the loss of valuable

information and allows for the comprehensive extraction of various feature types, thereby enhancing the overall effectiveness of the classification process.

III. SIMULATION ANALYSIS

A. Simulation data set

Experiments were conducted to simulate the algorithm using Pavia University (PaviaU), Indian Pine (IP), Salinas Valley (SV), and Botswana (BS) datasets.

The Pavia University dataset comprises 610×340 pixels and includes 9 different classes. It contains 103 distinct spectral bands ranging from 430 nm to 860 nm.

The Indian Pines dataset comprises 200 contiguous spectral bands, spanning the spectral range from 0.4 to 2.5 micrometers. The dataset encompasses 145×145 pixels and comprises 16 different land cover classes, including various types of vegetation, soil, and buildings.

The Salinas dataset also contains 224 contiguous spectral bands covering the spectral range from 0.4 to 2.5 microns. Due to the absorption of water vapour, some of the bands have been excluded, resulting in a total of approximately 204 bands being used. The dataset comprises 512×217 pixels and 16 feature classes.

The Botswana dataset contains 242 spectral bands, spanning the range from 0.4 to 2.5 microns. Due to the presence of noise and water vapour absorption, some bands were excluded, resulting in the actual number of bands used being 145. The dataset contains 1476×256 pixels and comprises 14 feature classes.

B. Simulation environment and training process

The experimental framework was established using hardware that includes a 12-core Intel® Xeon® Platinum 8352V CPU operating at 2.10 GHz, complemented by an NVIDIA RTX 4090 graphics card. The computational environment was configured with Python 3.8, utilizing the Pytorch framework for model development and testing. To optimize the training process, a learning rate of 0.001 was initially set, with a decay rate of 0.000001 to adjust the learning rate gradually over time. The Adam optimizer was employed for updating model parameters, and the categorical_crossentropy function was used as the loss metric. Model performance was assessed using the accuracy metric during both training and evaluation phases. The training regimen consisted of a batch size of 32 and was conducted over 200 iterations, covering all training samples. The results, including the number of samples utilized for both training and testing across these datasets, are detailed in Table I.

TABLE I: Information about the samples used in training and testing phases for each HSI dataset.

Class	Datasets			
	PaviaU	IP	SV	BS
Number of Classes	9	16	16	14
Total Numer of samples	42776	10249	54129	3248
Training set	849	510	533	42
Test set	41927	9739	53596	3206

C. Selection of spectral bands

In the proposed algorithm, PCA is employed to select an appropriate number of spectral bands before the hyperspectral data is fed into the network model. PCA not only significantly reduces the data dimensionality and computational complexity but also maximizes the retention of useful information from the original data. However, the selection of the number of spectral bands is a crucial factor, as it directly influences the classification accuracy and efficiency. To determine the optimal number of spectral bands, comparative experiments were conducted on four datasets using varying numbers of bands. The results were analyzed based on overall accuracy (OA) and the total number of model parameters (Total Params), as shown in Fig. 3.

From the perspective of OA, it was observed that for the PaviaU and SV datasets, the OA values stabilized once the number of bands reached a certain threshold. Further increasing the number of bands did not significantly enhance accuracy and could even lead to a decrease due to the introduction of noise. Conversely, the classification accuracy on the IP and BS datasets was notably affected by the number of spectral bands. Analyzing from the perspective of model parameters, a positive correlation between the number of bands and the total number of parameters was evident. Specifically, for the PaviaU and SV datasets, the classification accuracy stabilized at approximately 98% and 97%, respectively, around 55 spectral bands. For the IP and BS datasets, the highest accuracy was achieved at 55 bands. To balance high classification accuracy and model complexity, 55 spectral bands were selected for all four datasets.

D. Selection of spatial window size

In HSI classification, to mitigate the influence of individual pixels on classification outcomes, it is common practice to select a neighborhood of pixels around a central pixel as a sample. This approach leverages the collective features of multiple pixels to enhance model stability. Consequently, an appropriate spatial window is crucial for both classification accuracy and efficiency. An excessively large window may introduce irrelevant background information, diminishing classification accuracy, while an overly small window may fail to capture sufficient spatial information, resulting in suboptimal accuracy. To determine the optimal spatial window size, comparative experiments were conducted on four datasets using varying window sizes. The results were analyzed based on OA as shown in Fig. 4.

As the spatial window size increases, the OA for the BS and IP datasets initially decreases and then increases. The SV dataset maintains relatively stable accuracy across different window sizes. The PaviaU dataset shows an initial increase in accuracy followed by a decline. When the window size exceeds 5×5, the classification accuracy begins to decline across all datasets, with this trend being particularly pronounced for the BS and IP datasets. Simultaneously, the computational burden increases as the window size grows. To balance classification accuracy and computational cost, a window size of 5×5 was selected for subsequent experiments.
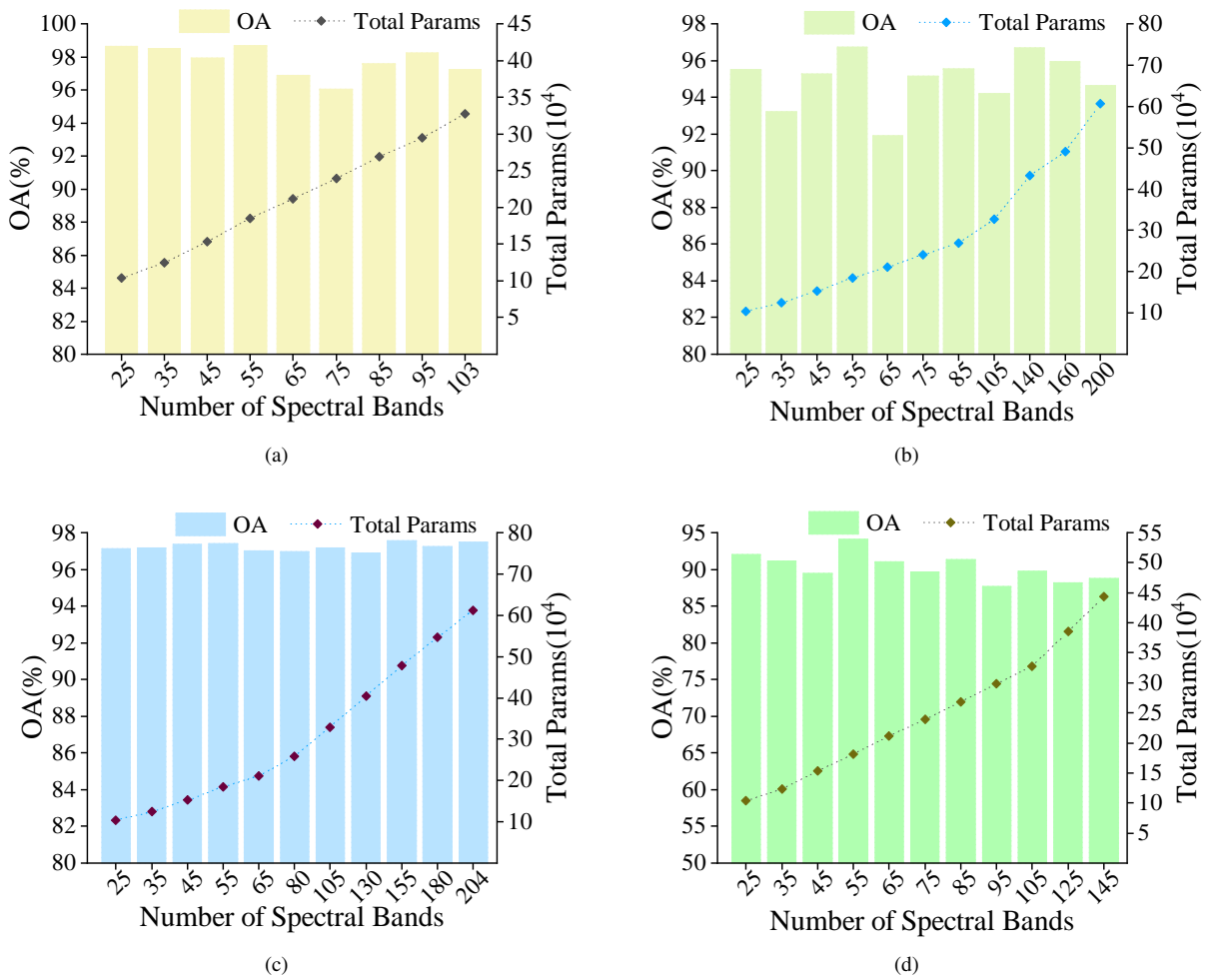
Fig. 3. OAs and model parameter quantity of SCNN with different numbers of spectral bands on different datasets. (a) PaviaU dataset. (b) IP dataset. (c) SV dataset. (d) BS dataset.
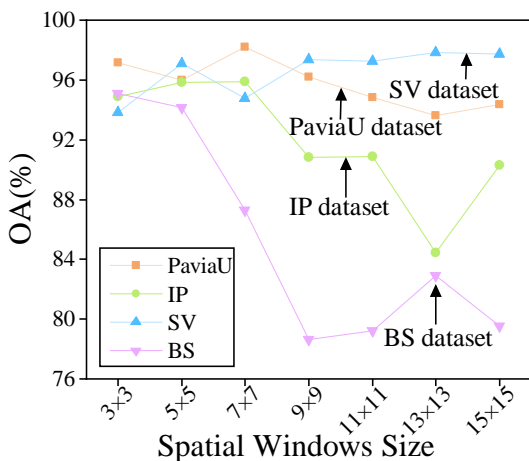

Fig. 4. OAs of SCNN with different sizes of spatial windows.

E. Comparing and evaluating the outcomes of the simulation

In order to verify the performance and efficiency of the proposed algorithm on the HSI classification task, the algorithm of this paper is compared and experimented with Support Vector Machine (SVM) [23], CDCNN [24], SSRN [25], DBDA [26], and FDSSC [27]. The parameters for each comparison method were configured according to specifications outlined in their respective publications.

The effectiveness of each algorithm was assessed based on three primary criteria: average accuracy of classification (AA), OA, and the statistical kappa coefficient (Kappa). These metrics provide a comprehensive overview of each model's classification performance. Additionally, the training and inference times for each model were meticulously recorded to evaluate computational efficiency.

The experimental outcomes for various algorithms on the PaviaU dataset are detailed in Tables II and III, and illustrated in Fig. 5. Analysis of Table II and Fig. 5 indicates that the traditional method utilizing SVM achieves the lowest classification accuracy, recording only 91.17%. In contrast, algorithms leveraging CNNs demonstrate significantly superior performance compared to SVM, underscoring the substantial advantages of CNNs in HSI classification. Particularly, the CDCNN algorithm, which is based on 2D-CNN, does not exploit the joint spectral-spatial features inherent in HSIs, resulting in performance that is less optimal compared to other CNN-based approaches. Conversely, algorithms that incorporate 3D-CNNs, such as SSRN, FDSSC, and DBDA, exhibit superior accuracy across all categories and the three evaluative metrics (AA, OA, and Kappa), outperforming both SVM and CDCNN. However, these algorithms still do not

fully capitalize on the potential synergies between different features within HSIs. The algorithm introduced in this study not only achieved the highest overall classification accuracy and Kappa but also demonstrated outstanding performance in category-specific accuracy, with each category achieving at least a 93% accuracy rate. According to Table III, the proposed algorithm also shows more rapid convergence and the shortest inference time, indicating superior computational efficiency. These results robustly support the conclusion that the proposed algorithm effectively utilizes convolutional kernels of varying dimensions within CNNs to integrate diverse features, thereby producing a richer and more comprehensive feature representation.

TABLE II: The categorized results of different network models on the PaviaU dataset. **Boldfaced** numbers highlight the best and the second best are underlined.

Class	Color	SVM	CDCNN	SSRN	FDSSC	DBDA	SCNN
1		89.98	90.78	<u>97.31</u>	96.38	**99.71**	96.90
2		93.03	97.13	99.05	99.13	**99.57**	<u>99.34</u>
3		81.47	78.96	92.57	90.44	**99.86**	<u>94.44</u>
4		96.12	98.81	**99.68**	98.41	99.40	<u>99.48</u>
5		93.51	**100.00**	<u>99.96</u>	99.36	99.84	99.87
6		91.69	88.19	96.38	**99.33**	98.74	<u>99.15</u>
7		88.22	95.45	94.82	<u>98.82</u>	**100.00**	96.42
8		81.92	87.81	88.29	**95.63**	85.14	<u>93.15</u>
9		**100.00**	<u>99.77</u>	99.49	98.46	**100.00**	99.42
OA		91.17	93.55	96.91	97.75	<u>98.06</u>	**98.09**
AA		90.66	92.95	96.39	97.33	**98.03**	<u>97.57</u>
Kappa		88.16	91.14	95.91	97.01	<u>97.43</u>	**97.47**

TABLE III: Training and testing time of different network models on PaviaU dataset. **Boldfaced** numbers highlight the best and the second best are underlined.

Dataset	Algorithm	Training Times (s)	Testing Times (s)
PaviaU	SVM	<u>26.13</u>	7.60
	CDCNN	**19.03**	<u>4.17</u>
	SSRN	77.81	9.74
	FDSSC	67.29	10.11
	DBDA	76.49	13.43
	SCNN	40.69	**3.78**

The experimental results for the IP dataset are detailed in Tables IV and V, and visually represented in Fig. 6. An examination of Table IV alongside Fig. 6 indicates that while the algorithm proposed in this study did not achieve the highest metrics, it performs comparably to the leading DBDA algorithm, with only marginal differences 0.05 in OA and 0.06 in the Kappa. Excluding the DBDA algorithm, the proposed model maintains a leading position among the evaluated algorithms. Specifically, within the IP dataset, the proposed algorithm excelled in several categories, achieving a classification accuracy of 96.59% for the second category and 99.00% for the sixth category. Table V further illustrates the computational efficiency of the proposed algorithm, which demonstrates a shorter model training time compared to the DBDA algorithm, indicative of faster convergence. Fig. 6 highlight the limitations of SVM, which exhibits numerous noise points in the classification effect diagram due to their reliance solely on spectral features of HSIs without leveraging additional feature information. In contrast, the use of algorithms based on 3D-CNN and

2D-CNN significantly reduces the number of noise points in the classification results. The model developed in this study shows superior performance relative to other algorithms, as evidenced by a clearer classification effect diagram and fewer noise points. This improved performance underscores the effectiveness of the proposed algorithm in integrating diverse feature information and achieving more distinct and accurate classification outcomes.

TABLE IV: The categorized results of different network models on the IP dataset. **Boldfaced** numbers highlight the best and the second best are underlined.

Class	Color	SVM	CDCNN	SSRN	FDSSC	DBDA	SCNN
1		25.92	0.00	<u>92.30</u>	75.67	**97.67**	73.52
2		62.35	70.73	92.33	94.18	<u>95.85</u>	**96.02**
3		69.33	69.95	89.57	93.79	**97.11**	<u>96.59</u>
4		59.13	74.07	<u>97.29</u>	92.62	**99.02**	96.44
5		86.00	95.35	97.23	<u>98.97</u>	97.83	**99.00**
6		84.52	93.90	96.50	<u>98.12</u>	95.73	**99.01**
7		84.61	0.00	**91.66**	81.73	72.72	<u>89.53</u>
8		87.55	86.60	<u>97.87</u>	97.46	**100.00**	96.70
9		53.84	36.36	**100.00**	<u>77.78</u>	**100.00**	73.86
10		72.70	82.84	<u>93.35</u>	93.03	**95.56**	93.18
11		71.44	73.62	93.00	96.84	**99.06**	<u>98.20</u>
12		64.92	48.70	87.86	95.35	<u>97.01</u>	**97.11**
13		90.35	92.00	99.13	<u>99.83</u>	**100.00**	99.31
14		90.54	89.35	**98.79**	97.54	94.43	<u>97.77</u>
15		70.29	85.61	83.33	94.14	**97.72**	<u>96.15</u>
16		<u>98.41</u>	**98.83**	89.23	98.33	95.50	96.68
OA		74.74	78.17	93.51	95.81	**96.97**	<u>96.92</u>
AA		73.24	68.62	<u>93.72</u>	92.84	**95.95**	93.69
Kappa		70.97	74.90	92.59	95.22	**96.55**	<u>96.49</u>

TABLE V: Training and testing time of different network models on IP dataset. **Boldfaced** numbers highlight the best and the second best are underlined.

Dataset	Algorithm	Training Times (s)	Testing Times (s)
IP	SVM	<u>28.80</u>	<u>2.25</u>
	CDCNN	**13.20**	**1.28**
	SSRN	51.36	4.14
	FDSSC	52.81	2.63
	DBDA	85.17	3.76
	SCNN	64.32	8.74

The experimental analysis of various algorithms on the SV dataset is detailed in Tables VI and VII, and visually represented in Fig. 7. A comparative analysis between Table VI and Fig. 7 indicates that SVM outperforms the CDCNN algorithm. This superiority is attributed to the diverse sample types present in the SV dataset, which pose challenges for the CDCNN algorithm due to its limited ability to handle datasets with extensive category variability. The results from the different algorithms reveal that while the method proposed in this study did not achieve the top metrics, it demonstrated more effective classification results compared to those algorithms that rely solely on either 2D-CNN or 3D-CNN, such as CDCNN, SSRN, and FDSSC. As illustrated in Table VII, although the classification accuracy of the proposed algorithm slightly trails that of the DBDA algorithm, it significantly reduces the time required for model training and testing by 41.63 seconds and 12.34 seconds, respectively. This efficiency suggests that the proposed method is particularly adept at balancing computational efficiency with classification accuracy, offering a compelling solution for scenarios where both are critical.

Fig. 5. Classification maps of different network models on PaviaU dataset. (a) False color image(FCI). (b) Ground-truth(GT). (c)-(h) The classification maps with different network models.
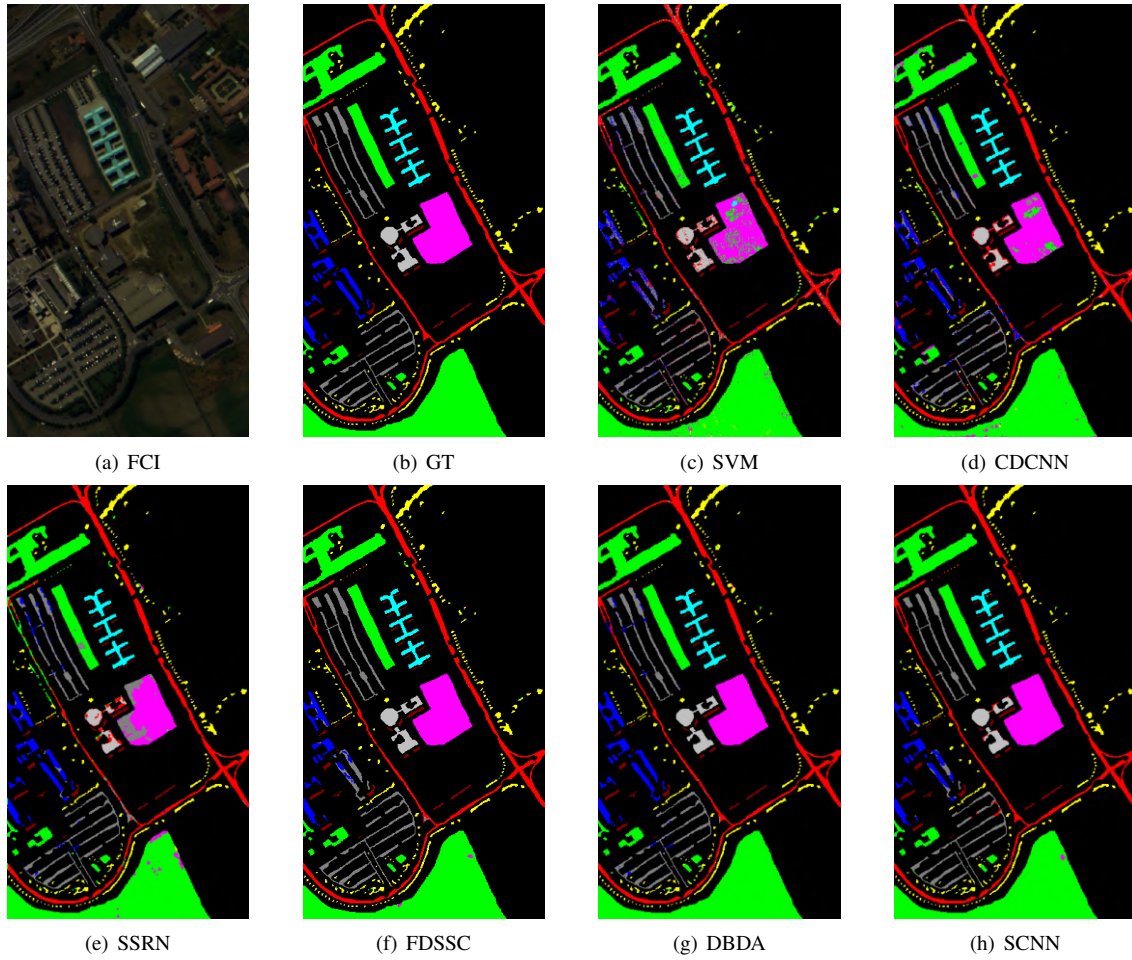

Fig. 6. Classification maps of different network models on IP dataset. (a) False color image(FCI). (b) Ground-truth(GT). (c)-(h) The classification maps with different network models.

Fig. 7. Classification maps of different network models on SV dataset. (a) False color image(FCI). (b) Ground-truth(GT). (c)-(h) The classification maps with different network models.
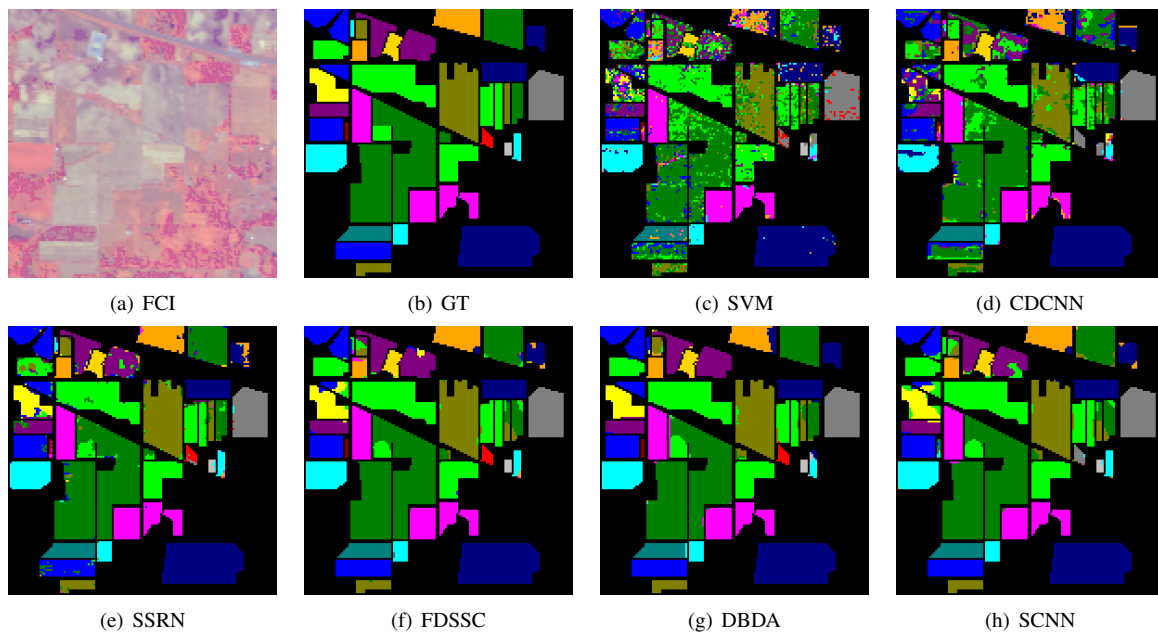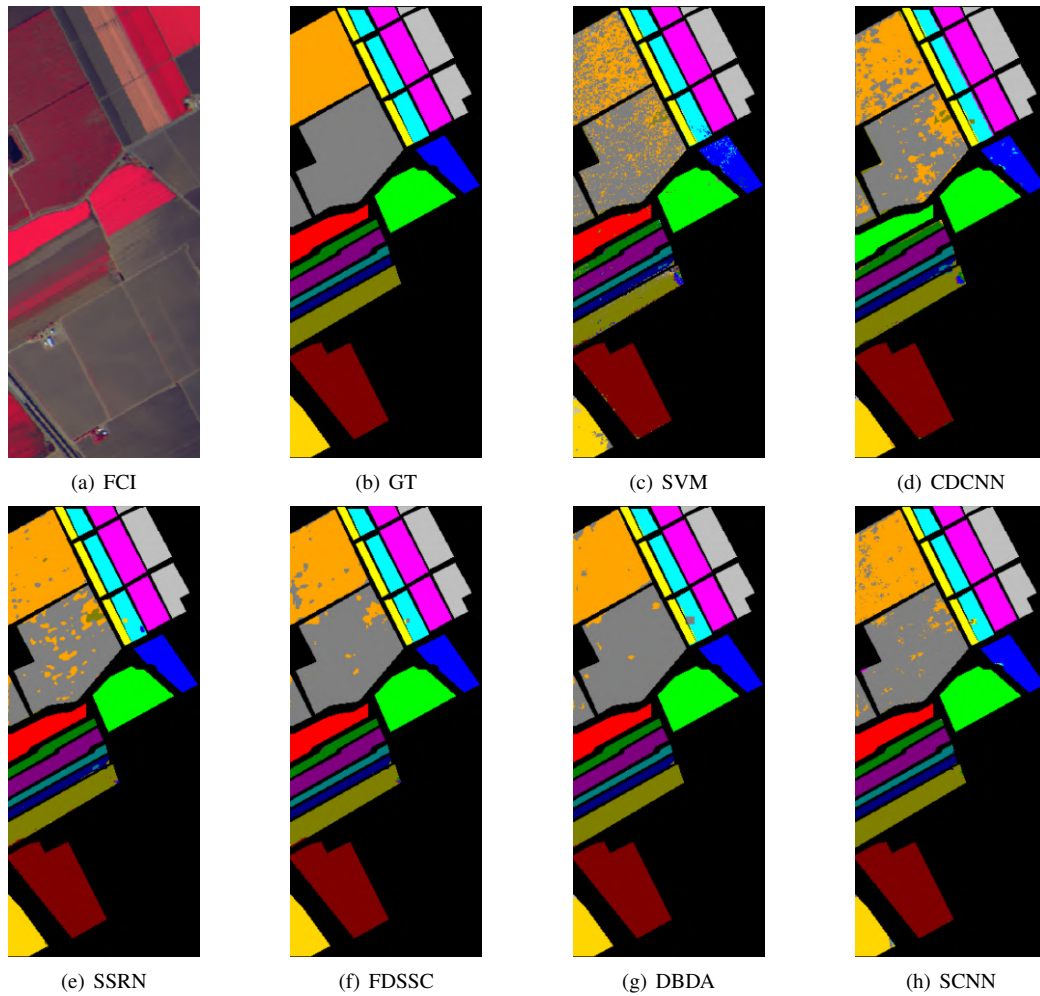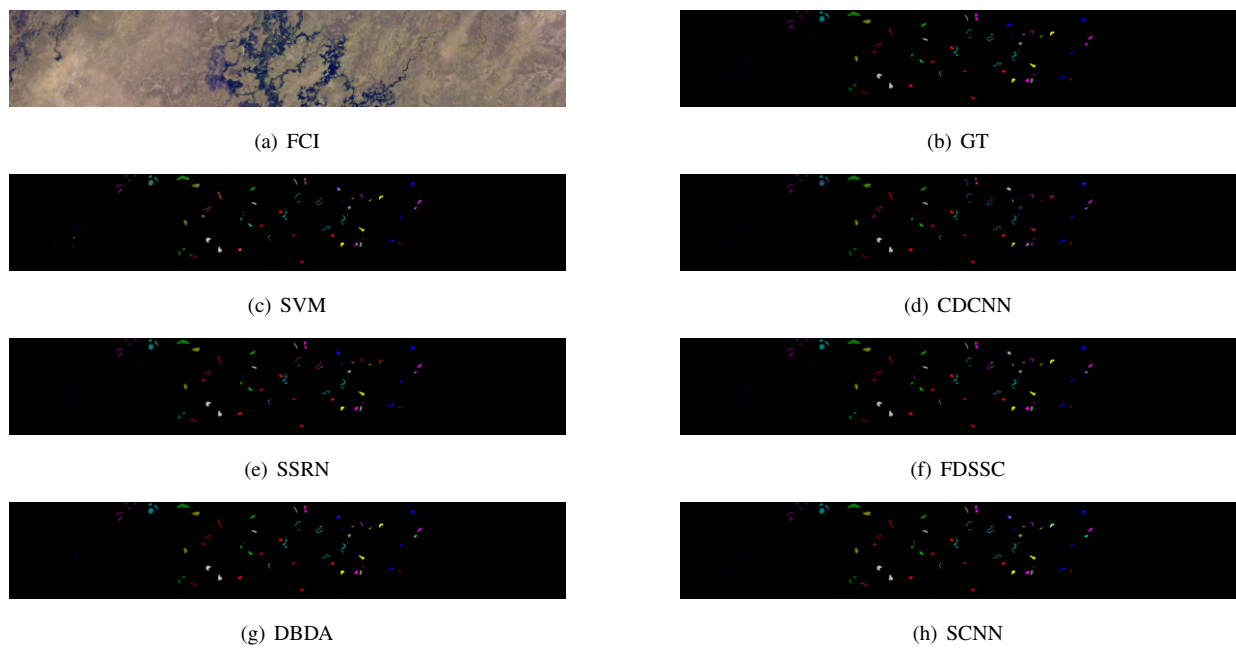

Fig. 8. Classification maps of different network models on BS dataset. (a) False color image(FCI). (b) Ground-truth(GT). (c)-(h) The classification maps with different network models.

The experimental results for various algorithms on the BS dataset are detailed in Tables VIII and IX, and illustrated in Fig. 8. An analysis of Table IX and Fig. 8 reveals that algorithms employing 3D-CNNs outperform those based on 2D-CNNs and traditional methods in terms of classification accuracy. This underscores the crucial role of joint spectral-spatial feature extraction performed by 3D-CNNs in the classification of HSIs. While the algorithm proposed in this study did not achieve the best results across all metrics, it consistently demonstrated a high accuracy rate, exceeding 93%. Table VIII shows that although the classification accuracy of the proposed algorithm is slightly below that of the leading algorithms, it records the shortest model training and testing times. This underscores the computational efficiency of the proposed method, highlighting its ability to balance performance with speed, thereby making it a viable option for applications requiring rapid processing without a significant compromise in accuracy.

TABLE VI: The categorized results of different network models on the SV dataset. **Boldfaced** numbers highlight the best and the second best are underlined.

Class	Color	SVM	CDCNN	SSRN	FDSSC	DBDA	SCNN
1		99.74	0.00	99.91	99.99	**100.00**	99.88
2		99.64	64.97	99.87	99.99	**100.00**	99.94
3		92.47	97.34	96.72	99.17	**100.00**	99.77
4		97.76	93.36	**98.69**	98.23	95.81	97.22
5		95.40	98.52	99.43	99.52	**99.56**	99.41
6		**100.00**	95.47	99.95	99.98	100.00	99.52
7		97.59	99.43	99.75	99.99	99.74	**100.00**
8		71.67	84.48	93.27	94.92	**97.15**	93.48
9		98.76	99.73	99.66	99.61	99.77	**100.00**
10		89.50	85.27	97.94	98.95	**99.62**	99.19
11		94.21	91.49	96.13	96.21	**100.00**	99.17
12		96.27	98.22	99.00	98.96	**100.00**	100.00
13		96.81	91.55	98.28	99.73	**100.00**	99.87
14		94.79	98.57	98.76	98.66	**99.42**	98.45
15		68.35	73.04	87.60	88.12	**96.38**	91.24
16		99.87	99.30	99.65	99.75	**100.00**	99.42
OA		87.89	87.29	96.22	96.50	**98.70**	97.19
AA		93.30	85.67	97.79	98.26	**99.21**	98.53
Kappa		86.47	85.85	95.79	96.10	**98.56**	96.87

TABLE VII: Training and testing time of different network models on SV dataset. **Boldfaced** numbers highlight the best and the second best are underlined.

Dataset	Algorithm	Training Times(s)	Testing Times(s)
SV	SVM	18.78	10.63
	CDCNN	**17.09**	6.73
	SSRN	49.51	11.82
	FDSSC	75.95	13.11
	DBDA	70.79	17.77
	SCNN	29.16	**5.43**

TABLE VIII: Training and testing time of different network models on BS dataset. **Boldfaced** numbers highlight the best and the second best are underlined.

Dataset	Algorithm	Training Times(s)	Testing Times(s)
BS	SVM	2.27	0.60
	CDCNN	2.65	0.36
	SSRN	6.46	0.88
	FDSSC	6.88	0.97
	DBDA	3.42	1.28
	SCNN	**1.93**	**0.31**

TABLE IX: The categorized results of different network models on the BS dataset. **Boldfaced** numbers highlight the best and the second best are underlined.

| Class	Color	SVM	CDCNN	SSRN	FDSSC	DBDA	SCNN
1		**100.00**	99.62	98.13	96.33	96.33	97.04
2		70.70	55.36	**100.00**	**100.00**	**100.00**	98.93
3		84.10	89.53	**100.00**	**100.00**	**100.00**	100.00
4		65.95	**100.00**	94.93	91.73	94.61	99.46
5		82.62	90.00	**95.75**	93.43	83.03	76.67
6		65.71	52.09	72.64	80.87	**100.00**	87.75
7		78.77	69.88	99.21	79.55	**100.00**	91.66
8		65.87	45.45	97.02	98.98	**100.00**	93.33
9		75.18	61.71	77.48	96.56	95.06	**99.63**
10		69.82	90.19	93.48	96.06	**97.21**	92.77
11		95.49	71.42	**100.00**	**100.00**	99.66	**100.00**
12		93.10	51.01	96.70	**97.77**	95.65	94.62
13		76.25	81.32	**100.00**	**100.00**	**100.00**	96.87
14		90.41	49.06	**100.00**	**100.00**	**100.00**	100.00
OA		78.63	69.59	92.92	93.96	**96.77**	94.05
AA		79.57	71.90	94.67	95.09	**97.25**	94.91
Kappa		76.87	67.08	92.32	93.45	**96.50**	93.56
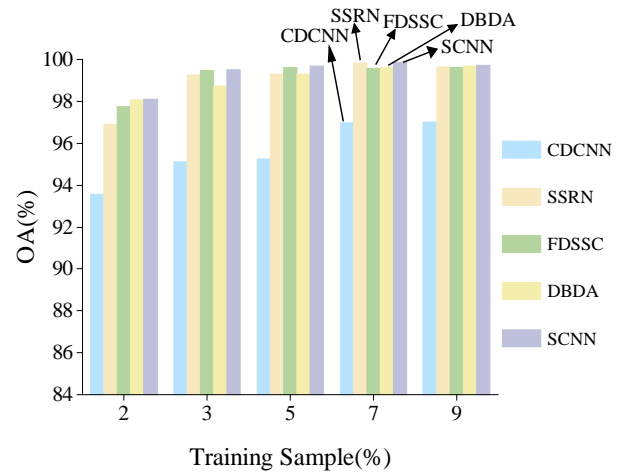
Fig. 9. OAs of different methods at various training sample proportions on PaviaU datasets.
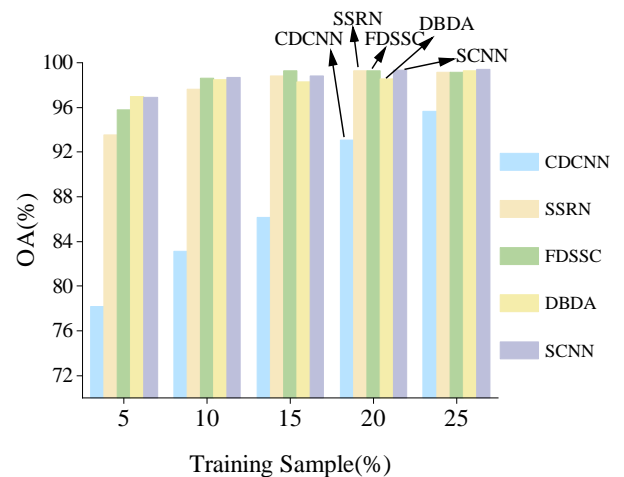

Fig. 10. OAs of different methods at various training sample proportions on IP datasets.
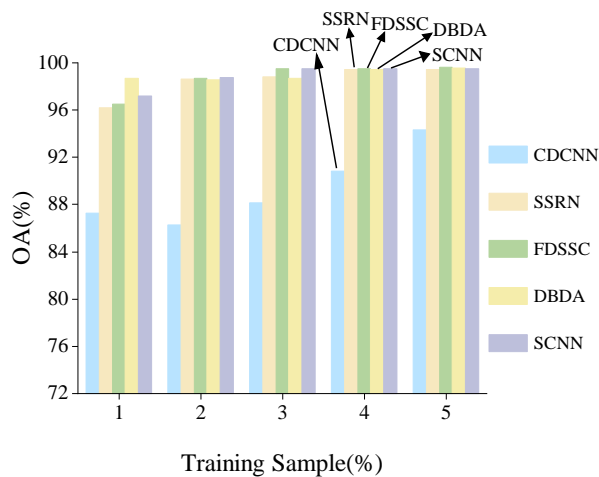
Fig. 11. OAs of different methods at various training sample proportions on SV datasets.
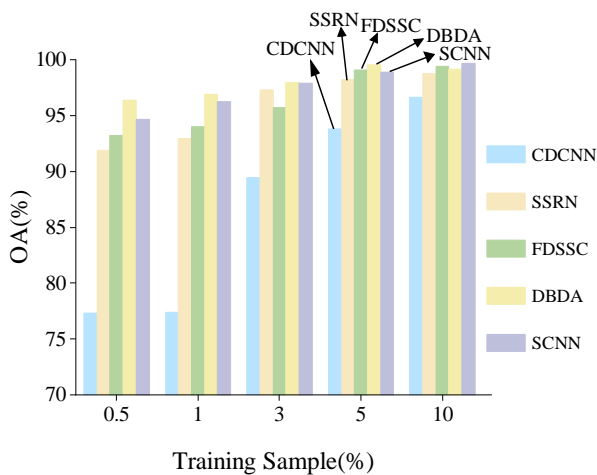

Fig. 12. OAs of different methods at various training sample proportions on BS datasets.

F. Analysis of the impact of training sample proportion on classification performance

The number of training samples significantly impacts the performance of deep learning-based HSI classification methods. Consequently, this study compares the OA results of various methods under different training sample ratios. As illustrated in Fig. 9, Fig.10, Fig.11 and Fig.12, the PaviaU dataset was trained with 2%, 3%, 5%, 7%, and 9% of samples, the IP dataset with 5%, 10%, 15%, 20%, and 25% of samples, the SV dataset with 1%, 2%, 3%, 4%, and 5% of samples and the BS dataset with 0.5%, 1%, 3%, 5%, and 10% of samples.

The experimental results indicate that the classification accuracy of all methods improves with an increase in the number of training samples, and the accuracy differences among various methods gradually diminish. Specifically, the CDCNN method exhibits lower classification accuracy compared to other methods. The SSRN method shows a continuous improvement in classification accuracy with more training samples, but it does not outperform other algorithms. FDSSC and DBDA achieve the best classification results

under certain training sample ratios. Comparatively, SCNN attains the highest accuracy across most training sample ratios and maintains a high classification accuracy even with fewer training samples.

IV. CONCLUSION

A novel multi-dimensional sequential CNN model called SCNN is designed for HSI classification in this paper. The SCNN leverages a serial concatenation strategy to integrate diverse features, thereby enhancing the representation capability of abstract features. This approach ensures high-precision classification of HSIs while maintaining a balance in computational efficiency. The SCNN model comprehensively utilizes convolutional kernels of varying dimensions within the network, enabling the extraction of a broad spectrum of features from HSIs and minimizing information loss during the feature extraction process. The designed network model was validated through simulations on various HSI datasets, demonstrating favorable results. Simulation outcomes indicate that, compared to classifications based solely on 2D-CNN or 3D-CNN algorithms, the proposed SCNN algorithm not only significantly improves the classification accuracy of HSIs but also considers computational efficiency. This balance of accuracy and efficiency renders the proposed method particularly suitable for HSI classification tasks.

REFERENCES

[1] M. A. Bendoumi, Mingyi He, and Shaohui Mei, "Hyperspectral Image Resolution Enhancement Using High-Resolution Multispectral Image Based on Spectral Unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 10, pp. 6574–6583, 2014.

[2] A. Benelli, C. Cevoli, and A. Fabbri, "In-field hyperspectral imaging: An overview on the ground-based applications in agriculture," *Journal of Agricultural Engineering*, vol. 51, no. 3, pp. 129–139, 2020.

[3] L.-J. S. Jong, A. L. Post, D. Veluponnar, F. Geldof, H. J. Sterenborg, T. J. Ruers, and B. Dashtbozorg, "Tissue classification of breast cancer by hyperspectral unmixing," *Cancers*, vol. 15, no. 10, p. 2679, 2023.

[4] M. Zhu, D. Huang, X. Hu, W. Tong, B. Han, J. Tian, and H. Luo, "Application of hyperspectral technology in detection of agricultural products and food: A Review," *Food Science & Nutrition*, vol. 8, no. 10, pp. 5206–5214, 2020.

[5] S. Peyghambari and Y. Zhang, "Hyperspectral remote sensing in lithological mapping, mineral exploration, and environmental geology: an updated review," *Journal of Applied Remote Sensing*, vol. 15, no. 3, pp. 031 501–031 501, 2021.

[6] Q. Yi, Y. H. Zhang, and Y. Zong, "A survey of hyperspectral image classification algorithms based on convolutional neural networks," *Electronics Optics & Control*, vol. 30, no. 3, pp. 70–77, 2023.

[7] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *Journal of Sensors*, vol. 2015, no. 1, p. 258619, 2015.

[8] Y. Lu, L. D. Wu, and J. Zhu, "Classification of hyperspectral images based on convolutional neural networks," *Computer Engineering and Design*, vol. 39, no. 9, pp. 2836–2841, 2018.

[9] W. Q. Gao and Z. Y. Xiao, "Multi-scale feature fusion 1d-cnn potato plant hyperspectral data feature classification and element identification," *Journal of Chinese Agricultural Mechanization*, vol. 43, no. 3, pp. 111–119, 2022.

[10] K. Makantasis, K. Karantzalos, A. Doulamis, and N. Doulamis, "Deep supervised learning for hyperspectral data classification through convolutional neural networks," in *2015 IEEE international geoscience and remote sensing symposium (IGARSS)*, 2015, pp. 4959–4962.

[11] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE transactions on geoscience and remote sensing*, vol. 54, no. 10, pp. 6232–6251, 2016.

[12] J. Xiao, L. Li, D. Xu, C. Long, J. Shao, S. Zhang, S. Pu, and Y. Zhuang, "Explore video clip order with self-supervised and curriculum learning

for video applications," *IEEE Transactions on Multimedia*, vol. 23, pp. 3454–3466, 2020.

[13] X. Zhang, "Research on hyperspectral remote sensing image classification based on three-dimensional convolutional neural network," Ph.D. dissertation, University of Chinese Academy of Sciences (Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences), 2021.

[14] Q. Mu, "Research on hyperspectral image classification based on 3d convolutional neural network," Ph.D. dissertation, Hebei University of Technology, 2022.

[15] Y. Kou and X. Liu, "Hsi shadow enhancement based on d-dsr and classification with parameter optimization," *IAENG International Journal of Applied Mathematics*, vol. 53, no. 3, pp. 1012–1027, 2023.

[16] 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, 2020.

[17] S. Ghaderizadeh, D. Abbasi-Moghadam, A. Sharifi, N. Zhao, and A. Tariq, "Hyperspectral image classification using a hybrid 3d-2d convolutional neural networks," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 7570–7588, 2021.

[18] J. X. Wang, Y. Ren, S. Y. Wang, X. W. Gao, and Y. Ye, "Research on remote sensing images classification method based on multi-dimensional features," *Journal of Optoelectronics·Laser*, vol. 33, no. 8, pp. 807–814, 2022.

[19] M. Ye, C. Ji, H. Chen, L. Lei, H. Lu, and Y. Qian, "Residual deep PCA-based feature extraction for hyperspectral image classification," *Neural Computing and Applications*, vol. 32, no. 18, pp. 14 287–14 300, 2020.

[20] Z. Wang, S. Liang, L. Xu, W. Song, D. Wang, and D. Huang, "Dimensionality reduction method for hyperspectral image analysis based on rough set theory," *European Journal of Remote Sensing*, vol. 53, no. 1, pp. 192–200, 2020.

[21] S. A. Medjahed, T. A. Saadi, A. Benyettou, and M. Ouali, "Binary cuckoo search algorithm for band selection in hyperspectral image classification," *IAENG International Journal of Computer Science*, vol. 42, no. 3, pp. 183–191, 2015.

[22] M. Li, Y. Liu, G. Xue, Y. Huang, and G. Yang, "Exploring the Relationship Between Center and Neighborhoods: Central Vector Oriented Self-Similarity Network for Hyperspectral Image Classification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 4, pp. 1979–1993, 2023.

[23] Y. Guo, X. Yin, X. Zhao, D. Yang, and Y. Bai, "Hyperspectral image classification with SVM and guided filter," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, p. 56, 2019.

[24] H. Lee and H. Kwon, "Going deeper with contextual cnn for hyperspectral image classification," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4843–4855, 2017.

[25] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–spatial residual network for hyperspectral image classification: A 3-d deep learning framework," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 847–858, 2017.

[26] R. Li, S. Zheng, C. Duan, Y. Yang, and X. Wang, "Classification of hyperspectral image based on double-branch dual-attention mechanism network," *Remote Sensing*, vol. 12, no. 3, p. 582, 2020.

[27] W. Wang, S. Dou, Z. Jiang, and L. Sun, "A fast dense spectral–spatial convolution network framework for hyperspectral images classification," *Remote sensing*, vol. 10, no. 7, p. 1068, 2018.