PCCN-MSS: Parallel Convolutional Classification Network Combined Multi-Spatial Scale and Spectral Features for UAV-Borne Hyperspectral With High Spatial Resolution Imagery

Linhuan Jiang [ID], Zhen Zhang [ID], Bo-Hui Tang [ID], *Senior Member, IEEE*, Lehao Huang [ID], and Bingru Zhang [ID]

Abstract—Hyperspectral remote sensing images with high spatial resolution (H^2 imagery) have an abundant spatial-spectral information, holding tremendous potential for remote sensing fine-grained monitoring and classification. However, challenges such as high spatial heterogeneity, severe intra-class spectral variability, and poor signal-to-noise ratio especially in unmanned aerial vehicle (UAV) hyperspectral imagery constrain and hinder the performance of fine-grained classification. Convolutional neural network (CNN) emerges as a formidable and excellent tool for image mining and feature extraction, offering effective utility for land cover classification. In this article, a parallel convolutional classification network model based on multimodal filters [including independent component analysis (ICA)-two-dimensional (2-D)-FPN and spectral attention (SA)-3-D-CNN branching structures] PCCN-MSS is proposed for precise H^2 imagery classification. The ICA-2-D-FPN branch integrates ICA into 2-D-CNN to extract the multispatial scale and spectral information of H^2 imagery by feature pyramid networks, meanwhile, the SA-3-D-CNN branch is designed to extract the spatial and spectral information by combining SA mechanism and 3-D-CNN. Taking hyperspectral imagery of UAVs containing vegetation and artifactual material ground as an example, the proposed PCCN-MSS model achieves an overall accuracy of 78.18%, which outperforms by 9.58% to the compared methods. The proposed PCCN-MSS method can mitigate the classification issues of severe salt-and-pepper noise and inaccurate boundary, delivering more satisfactory classification results with robust classification performance and remarkable advantages for H^2 imagery.

I. INTRODUCTION

HYPERSPECTRAL remote sensing images (HSIs) encompass an extensive range of continuous spectral information, which can distinguish subtle differences across diverse ground surfaces. Compared with the space-borne and aircraft-borne platforms HSI, the H^2 imagery acquired by the unmanned aerial vehicle (UAV)-borne hyperspectral sensor possesses superior spatial resolution (centimeter level) [1]. In addition, UAV hyperspectral observation platforms are cost-effectiveness, increased flexibility, and timely operational efficiency compared with satellite-based and airborne hyperspectral observation platforms. Consequently, UAV H^2 imagery are applied extensively in military surveying, agricultural production, environmental monitoring, and urban planning [2], [3], [4], [5].

Compared with traditional HSI, H^2 imagery provides a continuous spectral reflectance curve containing a continuous spectral reflectance curve with an abundance of detailed information about ground objects. This richness of data significantly enhances the accuracy and reliability of image classification. Nevertheless, the notable characteristic of H^2 imagery is a profound spectral heterogeneity of ground objects with the improvement of spatial resolution. Consequently, this leads to a substantial different spectrum in the same objects, which grievously exacerbates the intra-class variance of the features. This intra-class variance poses a significant challenge for fine-grained classification tasks, as it complicates the accurate discrimination of objects within the same class. For UAV H^2 imagery classification, traditional image classification methods based on machine learning, such as support vector machine [6], [7], random forest [8], decision tree [9], [10], and maximum likelihood [11], are revealing under-fitting to high-dimensional complex sample, susceptibility to over-fitting, and large amounts of mixed noise. Thereby, the limitations of these algorithms restrict the potential applications and fields of HSI classification.

Deep learning (DL) possesses remarkable capabilities in learning from training samples, coupled with excellent

Linhuan Jiang, Zhen Zhang, Lehao Huang, and Bingru Zhang are with the Faculty of Land Resource Engineering, Kunming University of Science and Technology, Kunming 650093, China, and also with the Key Laboratory of Plateau Remote Sensing, Yunnan Provincial Department of Education, Kunming 650093, China (e-mail: linhuanjiang@stu.kust.edu.cn; zhangzhen@kust.edu.cn; lehaohaung@stu.kust.edu.cn; bingruzhang@stu.kust.edu.cn).

Bo-Hui Tang is with the Faculty of Land Resource Engineering, Kunming University of Science and Technology, Kunming 650093, China, also with the Key Laboratory of Plateau Remote Sensing, Yunnan Provincial Department of Education, Kunming 650093, China, and also with the State Key Laboratory of Resources and Environment Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China (e-mail: tangbh@kust.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3370632

data-driven adaptability and portability. Which can extract complex samples by excavating data laws and stacking multilayer neural structures, which have performed satisfactorily in HSI classification. Semantic segmentation and convolutional neural networks (CNNs) are among the most widely used and influential DL algorithms in image classification [12], [13], [14], [15], [16], [17], [18]. Their proficiency in precisely segmenting and classifying images based on learned features has rendered them indispensable in a wide range of computer vision tasks. CNN excels in minimizing parameters and model complexity by effectively combining locally perceived regions with weight sharing. This optimization refines the network structure, ensuring the invariance to shifts and deviations, ultimately enhancing the model's generalization capability and stability, while avoiding overfitting. Pixel-level and patch-level CNN classification methods in HSI are excellently proficient at distinguishing inter-class differences and precisely delineating the boundaries of the ground objects, which is particularly suitable for artificial architecture. The prevalent CNN models can be primarily categorized into one-dimensional (1-D)-CNN, 2-D-CNN, and 3-D-CNN [19]. While 1-D-CNN effectively processes one-dimension vector data, it tends to ignore the rich spatial information during effectively extracting the spectral information of images. Correspondingly, spatial neighborhood information is sufficiently utilized in 2-D-CNN. Researchers like Liu et al. [20], Li et al. [21], and Yu et al. [22] have successfully applied 2-D-CNN to enhance HSI classification, proposing innovative approaches for spectral-spatial classification and noise mitigation. Certainly, combining 1-D-CNN and 2-D-CNN can also improve the classification performance for HSI data to a certain extent. Naji et al. [23] have constructed an imbalance data HSI generation and classification model based on DL. That utilizes 1-D-CNN for the extraction of spectral features and 2-D-CNN to extract spatial features, effectively integrating both types of information for improved classification performance. The 3-D-CNN architecture extends the 2-D-CNN by adding a third dimension to the convolutional kernel, enabling the simultaneous convolution of spatial neighborhood information along with the spectral dimension. Researchers have proposed various 3-D-CNN models tailored for HSI classification, incorporating features like wavelet coefficients prediction [24], deep feature extraction [25], spatial-spectral pyramid networks [26], and multiscale three-branch feature fusion [27]. These models leverage the unique capabilities of 3-D-CNN to extract discriminative features from complex hyperspectral data, resulting in state-of-the-art classification performance. To address the challenges posed by limited samples and the enormous computation for CNN-based hyperspectral image classification, Bhatti et al. [28] have introduced a multifeature fusion approach using 3-D-CNN and graph attention network, which combines the strengths of both techniques to achieve improved accuracy and efficiency. Summarily, 2-D-CNN and 3-D-CNN models with the analyzing data ability are popularly applied and get consistent unanimous praise. However, to mitigate the impact of the severe spectral variability and spatial heterogeneity of H^2 imagery, several classification algorithms tailored for H^2 imagery have emerged by "spectral patching" approach [29] and encoder-decoder architecture [30], [31], which preserve the global spatial information and nearly all spectral information of the original hyperspectral image. In addition, innovative methods such as incorporating adaptive weighting of each channel, pixel, and scale-awareness of feature maps [32], jointly utilizing spatial geometric and spectral information for multispectral point cloud superpoint segmentation [33], assembling 3-D-CNN and 2-D-CNN layers [34], and developing 3-D-Deep feature extraction CNN model which uses both spectral and spatial information [35] have been proposed for fine-grained classification of H^2 imagery. Scholars have employed a range of modules to enhance the capabilities of their models. For instance, the CNN-enhanced incorporates CNN and graph convolutional network branches to conduct feature learning [36], the kernel tensor sparse coding model can enhance linear separation by utilizing the kernel tensor representation mechanism [37], and the SSAN combines 2-D-CNN spatial extraction and RNN exploitation of spectral information [38].

However, the increased intra-class variance and pronounced spatial heterogeneity in UAV H^2 imagery complicate the statistical distribution of spectral features. This complexity pyramidally diminishes the distinguishability of spectral information, posing a challenge for accurate analysis and interpretation. Existing algorithms may struggle to accurately differentiate between various categories of human-made structures or precisely identify individual buildings, which are complex structures with the potential for overlap or proximity. This limitation significantly impairs the analysis and understanding of urban landscapes, infrastructure, and other crucial components of H^2 images that encompass human-made constructions. To bridge this gap, this study introduces PCCN-MSS, a cutting-edge parallel convolutional classification network combined multispatial scale and spectral feature. The proposed PCCN-MSS model incorporates two distinct branches to comprehensively extract spatial and spectral information from H^2 imagery. The first branch combines a 2-D-CNN with feature pyramid networks (FPNs) and independent component analysis (ICA) to effectively capture the multispatial scale and spectral information (ICA-2-D-FPN). In spectral attention (SA)-3-D-CNN branch, augments a 3-D-CNN branch with SA with an enhanced focus on spectral nuances to extract the spatial and spectral information. The proposed model, PCCN-MSS, aims to maximize the utilization of spatial-spectral information by leveraging the strengths of its constituent models. The parallel convolutional classification network model makes significant contributions in three crucial areas.

1) *SA Mechanism in 3-D-CNN:* The integration of an SA mechanism into 3-D-CNN enhances the extraction of spectral information while considering spatial neighborhood features. This enhancement not only improves the robustness of the network but also deepens the understanding of the spectral features in H^2 imagery with greater precision.
2) *FPN in ICA-2-D-CNN Branch:* The utilization of FPN within the ICA-2-D-CNN branch, which leveraging the high resolution of lower level features and the rich semantic information of higher level features simultaneously.

This simultaneous leveraging of multiscale feature maps is crucial for detecting and recognizing small targets.

3) *Parallel Network Structure for Spatial-Spectral Information Extraction:* Parallel convolutional classification network structure, which combines 2-D-FPN with ICA and 3-D-CNN with SA mechanism, simultaneously extracts spatial and spectral information from images.

The rest section of this article is structured as follows. Section II describes the methodology employed in this article, detailing the proposed classification method for UAV H^2 imagery. The experimental results based on the parallel neural network structure for comparison are presented in Section III. Section IV discusses about experimental results and Section V concludes this article.

II. METHOD

A. Prerequisite

ICA is a computational method that employs linear transformation to decompose multivariate statistical signals into their constituent additive subcomponents, which is presumed to exhibit statistical independence and non-Gaussian, enabling ICA to effectively disentangle complex data mixtures and reveal the underlying independent sources. ICA remains relevant even when perfect statistical independence is unachievable, and it facilitates the independent examination of observed signal statistics and deeper exploration of data characteristics. ICA effectively reduces the dimensionality of HSI data and eliminates extraneous and useless information while preserving crucial details. Numerous scholars have capitalized on ICA in hyperspectral image classification, building upon its theoretical foundations as outlined in [39], [40], [41], and [42].

The FPN architecture, designed for tasks such as target detection and semantic segmentation, tackles the intricacies of detecting and segmenting objects at different scales [43]. By constructing multiscale feature pyramids and training them for an "end-to-end" approach, FPN effectively handles various scale targets. This process involves two pivotal steps: bottom-up feature extraction, which captures the finer details and top-down feature extraction for focusing on the coarser context. The prediction results are then seamlessly combined via lateral connections to yield the ultimate prediction, ensuring both precision and contextual awareness.

Attention mechanisms endow different weights for various positions to emphasize critical information, thereby enhancing model performance, robustness, and generalization. In the context of adjusting spectral information, this mechanism is referred to as the SA mechanism. Involving recalibration of the spectrum features. Numerous scholars have successfully integrated attention mechanisms into spectral similarity measurements, effectively aggregating similar spectra and optimizing cognitive performance of CNNs [44], [45], [46], [47].

B. Parallel Convolutional Classification Network Combined Multispatial Scale and Spectral Features

In our study, the application of ICA effectively extracts high-level statistical information and accomplishes dimensionality reduction. This improves the efficiency of the data utilization, significantly reducing the computation burden on the 2-D-CNN. In addition, FPNs are incorporated into this branch to facilitate learning of the image features by different spatial scales. One notable strength of PCCN-MSS lies in its meticulously crafted SA design. This design recalibrates the original data by amplifying critical bands while suppressing redundant or unnecessary ones, ensuring optimal utilization of spectral information. 3-D-CNN incorporates the SA mechanism, which significantly enhances model performance and classification accuracy by efficiently attending to critical spectral features. Finally, the outputs from the two cotraining branches are comprehensively evaluated through a weighted summation approach, to obtain the precise and reliable final classification results. The complete architecture of the proposed model, PCCN-MSS, is depicted in Fig. 1. The innovative model comprises an ICA-2-D-FPN branch that utilizes 2-D multiscale spatial feature pyramids with ICA for dimensionality reduction and SA-3-D-CNN spatial-spectral information extractor with an SA mechanism. The network configuration for PCCN-MSS model is meticulously outlined in Table I. Specifically, the ICA-2-D-FPN branch is delineated on the left and SA-3-D-CNN branch is presented on the right.

In the ICA-2-D-FPN branch, the top 10-band processed by ICA are retained to optimize both computational efficiency and memory utilization. A patch neighborhood P, with dimensions of $m \times m \times c$ is selected from the ICA-processed features. Feature maps at different scales are extracted from the input image through a bottom-up feature extraction stage, using three 2-D separable convolutional layers. The higher level feature maps undergo two upsampling operations to align their dimensions with the corresponding lower level feature maps. Subsequently, these maps are seamlessly fused through element-wise summation, layer by layer. Lateral connections are then employed to compute the ultimate probability for each category.

Concurrently, the SA-3-D-CNN branch is fed with original image data without dimensionality reduction. Recognizing that the original bands often contain a considerable amount of redundant information, some of which may be unless, the branch adopted the SA mechanism. This mechanism serves to recalibrate the original spectral data, emphasizing salient features while suppressing unnecessary or noisy band information. The recalibrated data is expanded dimension and fed into 3-D-CNN, which are then learn by flattened and fully connected. Finally, the Softmax function is employed to classify these features.

Critical to PCCN-MSS is the synthesis and analysis of the results from the two branches mentioned above. In acquiring the ultimate prediction results, PCCN-MSS utilizes a weighted summation approach, the mathematical formulation of which is presented in the following equation:

$$P = \text{argmax}\,(a \times P_1 + b \times P_2) \qquad (1)$$

where P represents the final output of the model, which is the prediction result of the parallel network PCCN-MSS at the pixel level. The weights assigned to P_1 and P_2 are represented by "a" and "b" respectively. Here, P_1 and P_2 correspond to the prediction result of the ICA-2-D-FPN branch and the SA-3-D-CNN branch, respectively. The argmax function is exploited to pinpoint the variable value associated with the maximum
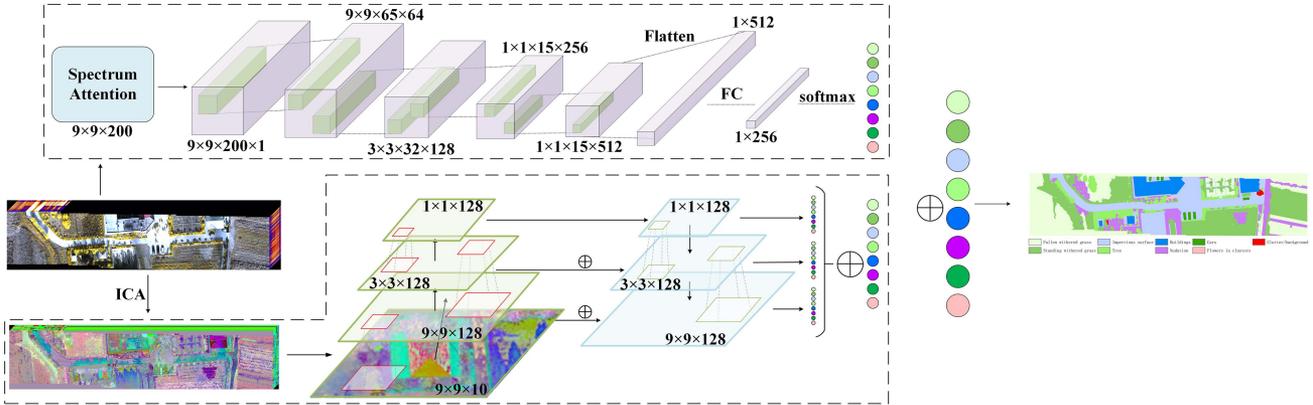
Fig. 1. Model architecture of parallel CNN PCCN-MSS. The 2-D-FPN branch with ICA and the SA 3-D-CNN branch are followed by a weighted summation that is used to aggregate the spatial and spectral information.

TABLE I
NETWORK SETUP OF PCCN-MSS DESIGNED BASED ON ICA-2-D-FPN BRANCH AND SA-3-D-CNN BRANCH

ICA-2D-FPN branch					SA-3D-CNN branch					
Layer	IS	OS	C	A	S	Layer	IS	OS	C	A	S
ICA	(, ,200)	(, ,10)	/	/	/	SA	(5,5,200)	(5,5,200)	/	/	/
Conv1	(9,9,128)	(3,3,128)	128×5×5	R	(1,1)	Conv0	(5,5,200,1)	(5,5,65,64)	64×5×1×8	R	(1,1,3)
Conv2	(3,3,128)	(1,1,128)	128×3×3	R	(1,1)	Conv1	(5,5,65,64)	(1,5,32,128)	128×1×5×3	R	(1,1,2)
Upsample1	(1,1,128)	(3,3,128)	×3	/	(1,1)	Conv2	(1,5,32,128)	(1,1,15,256)	256×1×1×3	R	(1,1,2)
Upsample2	(3,3,128)	(9,9,128)	×3	/	(1,1)	Conv3	(1,1,15,256)	(1,1,1,512)	512×1×1×15	R	(1,1,1)
						Flatten	(1,1,512)	512	/	/	/
IS: Input shape	C: Convolutional parameters		S: Strides			FC	512	256	256	/	/
OS: Output shape	A: Activation		R: Rule			Softmax	256	8	8	/	/

function value, thereby identifying the prediction category with the greatest reliability.

In addition, Loss function is adopted in our proposed model to quantitatively evaluate the discrepancy or error between the predicted results of the model and the ground truth (GT). This approach enables a more nuanced understanding of the performance of model, pinpointing specific areas where improvements are needed. The experiments conduct a comprehensive evaluation of Loss function for both branches by the weighted equation, which takes into account the contributions, and the Loss function of PCCN-MSS is presented in the following equation:

$$\text{Loss} = a \times \text{Loss}_1 + b \times \text{Loss}_2 \tag{2}$$

where Loss denotes the final loss function in our model, Loss_1 and Loss_2 represent the output of the loss function for ICA-2-D-FPN and SA-3-D-CNN branch respectively. Both Loss_1 and Loss_2 are cross-entropy loss functions. The weighted coefficient of Loss_1 and Loss_2 are denoted by "a" and "b" respectively. Particularly, both "a" and "b" above are self-defined, and in this article, "a" and "b" is setted by 1/3 and 2/3, respectively.

C. ICA-2-D-FPN Branch

The input to 1-D-CNN is restricted to 1-D vector data. 1-D-CNN utilizes 1-D convolution, 1-D pooling, and fully connected for feature extraction, followed by Softmax for classification. Contrary to 1-D-CNN, 2-D-CNN extracts spatial neighborhood information using 2-D convolution and 2-D pooling. Reshaping

images are unnecessary and can be directly fed into the 2-D-CNN network. The following equation presents the computation process of 2-D-CNN:

$$\text{v}_{l,j}^{xy} = f \left(\sum_m \sum_{h=0}^{H_l-1} \sum_{w=0}^{W_l-1} k_{l,j,m}^{hw} v_{(l-1),m}^{(x+h)(y+w)} + b_{l,j} \right). \tag{3}$$

Depthwise separable convolution represents a distinctive subcategory within the realm of 2-D convolutions, encompassing SeparableConv2D and DepthwiseConv2D. The latter solely executes the initial phase of spatial convolution along the depth axis, while SeparableConv2D can implement the entire gamut of depth-separated convolution work. SeparableConv2D combines spatial convolution in the depth direction with the point-by-point convolution in the output channel. In this article, the 2-D-CNN block utilizes SeparableConv2D, which has fewer trainable weight parameter. This judicious approach not only mitigates computational burden but also enhances model performance. For each input channel, spatial convolution can be convolved using SeparableConv2D and then the output channels are mixed by point-wise convolution (1×1 convolution).

FPN extracts feature through both Bottom-up and Top-down processes. In the bottom-up block, each layer systematically produces a feature map "A," and feature map "B" will be derived from the preceding feature map through Upsampling2D in the top-down block. Feature map "A" and feature map "B" are seamlessly merged through lateral connections for elementwise. As illustrated in Fig. 2, the ICA-2-D-FPN branch generates
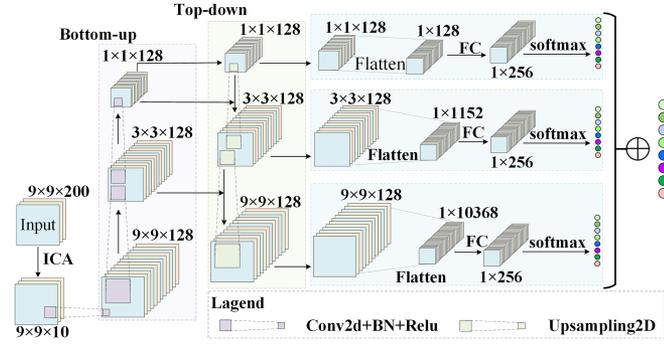
Fig. 2. Model architecture of ICA-2-D-FPN in PCCN-MSS involves an initial processing of the original H^2 data using ICA to reduce its dimensionality. And the neighborhoods of pixels resulting from the ICA processing are fed into the 2-D-FPN for extraction of multispatial scale features.

probability results at three distinct spatial scales. These outputs are subsequently averaged to arrive at the ultimate prediction. The employment of the FPN architecture is pivotal in extracting features from images at different spatial scales, thereby substantially bolstering the classification performance of the model.

However, the original data is often high-dimensions and redundancy, leading to significant computational and memory demands. ICA can separate data or signals into linear combinations of statistically independent non-Gaussian sources, effectively reducing the dimensionality while preserving the essential features. Consequently, combining 2-D-FPN construction with ICA can alleviate the curse of dimensionality, tackle computational and memory challenges, and yield superior results. The ICA-2-D-FPN branch model structure is depicted in Fig. 2. In this structure, the data processed by ICA is segmented into multiple neighborhoods P of size $m \times m \times c$ centered around a pixel. These blocks are then used as inputs for the 2-D-FPN. The ICA-2-D-FPN branch incorporates three stacked 2-D convolution blocks. Each convolutional block comprises a convolutional layer, a batch normalization layer, and a ReLU nonlinear activation function. Pooling and dropout have not been utilized in the ICA-2-D-FPN branch. The feature map of the current layer is combined with the one derived by applying the Upsampling2D function to the preceding layer, through lateral connections that facilitate the integration of features across layers. The three results are then passed to a flattened layer and a fully connected layer. The fully connected layer uses the Softmax function to calculate results. Ultimately, the ICA-2-D-FPN branch yields its results through the utilization of an averaging technique.

D. SA-3-D-CNN Branch

The 3-D-CNN extends the principles of 1-D-CNN and 2-D-CNN to analyze 3-D input data directly. By employing 3-D convolution and 3-D pooling, this model effectively extracts both spatial and spectral information from the input image. As depicted in Fig. 3, the architectural design of the 3-D-CNN bears some resemblance to that of the 2-D-CNN. Nevertheless, the utilization of both spatial and spectral information in the 3-D-CNN requires a larger memory capacity and longer processing time compared to the 2-D-CNN. But its deeper analysis ability

can extract more informative features and then typically yield superior classification results

$$\mathrm{v}_{l,j}^{xyz} = f\left(\sum_m \sum_{h=0}^{H_l-1} \sum_{w=0}^{W_l-1} \sum_{r=0}^{R_l-1} k_{l,j,m}^{hwr} v_{(l-1),m}^{(x+h)(y+w)(z+r)} + b_{l,j}\right).$$
(4)

Compared to conventional 3-D-CNN models, PCCN-MSS introduces an SA block, which is designed to decipher distinct spectral information and augment feature learning. In our SA mechanism, the input data is first subjected to global convolution and then applies (5) to derive the output

$$z_c = x \times (-y_c)$$
(5)

where x represents the input data utilized for the SA mechanism. In addition, y_c denotes the feature result obtained from the cth global convolution of the original data. Equation (5) introduces the variable z_c, which can be interpreted as a comprehensive mapping of the spectral sequence within a defined spatial proximity. This variable effectively encapsulates both the spectral and spatial attributes inherent in the images, serving as a unified representation. By considering the cth spectral representation z_c alongside the corresponding cth spectral band x, a linear mapping is obtainable, as outlined in the following equation:

$$h_c = x \times z_c = x \times x \times (-y_c).$$
(6)

However, when tackling complex UAV H^2 imagery with rich spectral and spatial information, simple linear mapping is insufficient for extracting effectively. Therefore, in this article, (6), which is the initial SA block of PCCN-MSS, is updated and presented in (7), and the UAV H^2 imagery is adequately extracted using nonlinear mapping

$$f = \exp(z_c) - \exp(-z_c))/(\exp(z_c) + \exp(-z_c)).$$
(7)

By integrating (5) and (7), we can summarize the SA mechanism adopted in this article, leading to the formulation of the following equation:

$$f = (\exp(-x \times y_c) - \exp(x \times y_c))/ \\ \times (\exp(-x \times y_c) + \exp(x \times y_c)).$$
(8)

III. RESULTS

A. Dataset and Experimental Setup

1) Dataset: Field-acquired H^2 imagery is utilized for experimental purposes in this article. Fig. 4 presents the false composite image along with its corresponding GT. The data is acquired using Hyspex sensor equipped on the UAV, offering a spatial resolution of 5 cm and 200 bands available for the experiment. Bands cover visible and near-infrared spectral ranges from 380 to 1000 nm. In conjunction with field-collected photographs, the experiment establishes the ground-truth label mapping depicted in Fig. 4. The GT contains a total of 450 × 1000 pixels, with labeled pixels classified into 8 categories. These categories are Fallen withered grass, Standing withered
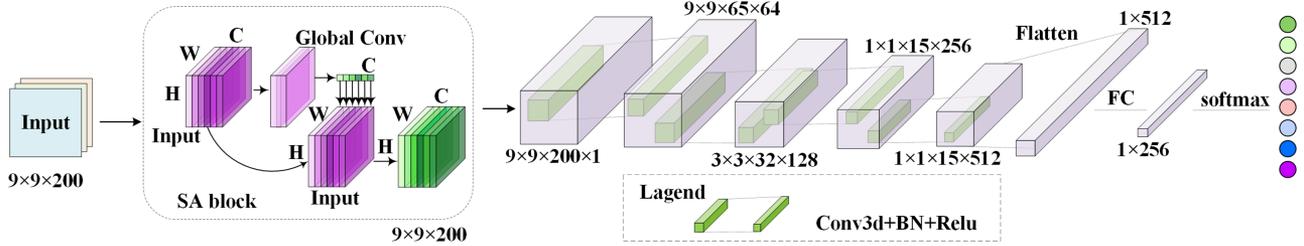
Fig. 3. Model architecture of SA-3-D-CNN in PCCN-MSS. The SA block is the SA mechanism module that is placed before the first convolution block to enhance the model's performance. It achieves this by applying spectral recalibrating for the original image neighborhood data.
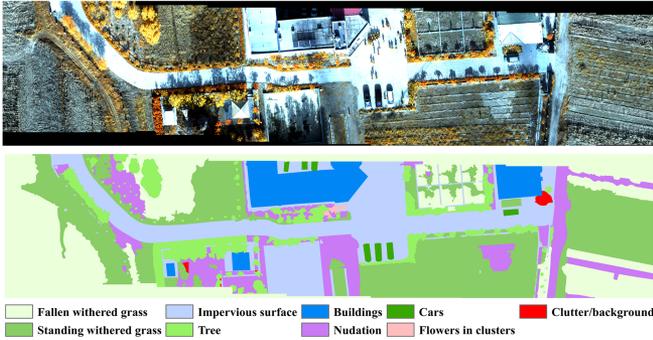

Fig. 4. False composite image of UAV H^2 imagery (R:621 nm, G:715 nm, B: 829 nm) and GT.

TABLE II
CLASSES NAME AND SAMPLES NUMBER

Classes r	Classes Name	Training Samples	Testing Samples
1	Fallen withered grass	300	1 303 084
2	Standing withered grass	300	920 326
3	Impervious surface	300	347 651
4	Tree	300	301 661
5	Buildings	300	479 870
6	Nudation	300	28 540
7	Cars	300	5182
8	Flowers in clusters	300	10 361
	Total	2400	3 396 675

grass, Impervious surface, Tree, Buildings, Nudation, Cars, and Flowers in clusters, as detailed in Table II.

2) Training and Testing Setup: As depicted in Table II, 300 samples are randomly selected from each category for model training, while all remaining labeled samples are reserved for testing the classification accuracy of the model. Notably, the number of training samples is a mere fraction of the total, accounting for only 0.071% of the overall labeled samples.

3) Parameters Setting and Configuration of Operating Environment: In addition to the network sample number settings outlined in Table II, the other training parameters of the model are as follows. The initial learning rate (lr) is established at 0.001, with an exponential decay rate of 0.1 applied every 1000 steps. The training process involves a batch size of 100 and runs

for 30 000 epochs. During the experiment, 9×9 neighborhood cubes with all spectral data of H^2 imagery are fed into the model. Furthermore, testing is performed in batches of 100 samples.

The programming language used in this experiment is Python 3.6, installed on the Ubuntu 20.04 system. The configured environment contains CUDA 11.4, CUDNN 7.6, and essential packages such as Tensorflow 1.14, Scipy 1.5, and Numpy 1.19.

B. Classification Results

In this section, our attention is on assessing the efficacy of the proposed model for UAV H^2 imagery dataset. To determine its accuracy, three metrics including overall accuracy (OA), average accuracy (AA), and Kappa coefficient are used for accuracy evaluation are employed. The experiments compare PCCN-MSS against traditional machine learning methods such as SVM, and popular DL methods including 1-D-CNN, 2-D-CNN, 3-D-CNN, and parallel networks SSAN. The classification outcomes achieved by these various methods when applied to UAV H^2 imagery are showcased in Fig. 5. Specifically, Fig. 5(a)–(f) presents the results obtained by SVM, 1-D-CNN, 2-D-CNN, 3-D-CNN, SSAN, and the proposed PCCN-MSS, respectively. Fig. 5(g) depicts the visual representation of the GT mapping, serving as a benchmark for assessing the classification accuracy of the various methods.

SVM [see Fig. 5(a)] exhibits acceptable classification results, distinguishing various categories well, and successfully distinguishing most categories, it nevertheless falls short in some instances. Notably, the model partially misclassifies impervious surfaces as buildings, highlighting a limitation in its ability to differentiate between these two closely related categories. However, the centimeter-level UAV H^2 imagery with unprecedented detail utilized in the experiment presents significant challenges. Objects that may appear similar or overlapping in such imagery are difficult to distinguish solely on spectral information without incorporating comprehensive spatial information. Fig. 5(b) reveals that the 1-D-CNN model faces considerable difficulties in accurately classifying certain categories, which struggles to distinguish between standing withered grass, tree, and nudation due to its reliance on spectral information, resulting in noticeable salt-and-pepper noise. This issue can be attributed to the model's heavy reliance on spectral information alone, without adequately incorporating spatial features. Furthermore, the presence of numerous shadows in the imagery poses a major challenge
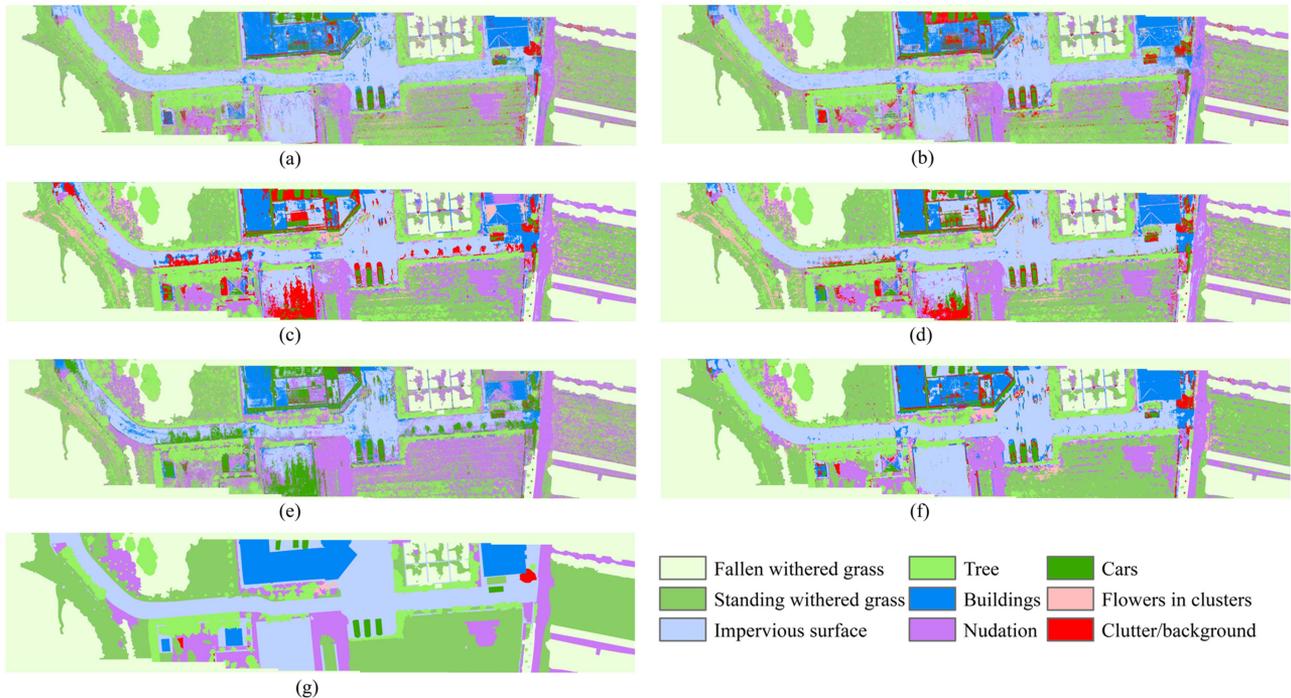
Fig. 5. Classification results of UAV H^2 imagery based on six methods. Classification results of (a) SVM, (b) 1-D-CNN, (c) 2-D-CNN, (d) 3-D-CNN, and (e) SSAN. (f) Parallel model PCCN-MSS classification result proposed in this article. (g) GT reference images.

to classification algorithms, especially 2-D-CNN. As shown in Fig. 5(c), this model has difficulty handling shadow interference, resulting in the misclassification of impervious surfaces as clutter/background. This misclassification highlights a significant weakness in 2-D-CNN's ability to accurately interpret complex visual information in the presence of shadowing effects. Fig. 5(d) demonstrates the improvement achieved by utilizing 3-D-CNN, which integrates spatial and spectral information, to significantly alleviate noise and enhance overall classification performance. Indeed, a noticeable portion of the noise observed in Fig. 5(b) is mitigated in Fig. 5(d), and the misclassification in Fig. 5(c) is conspicuously alleviated. These improvements demonstrate the effectiveness of 3-D-CNN incorporating multidimensional data in enhancing classification outcomes. Contrarily, the SSAN parallel network depicted in Fig. 5(e) plagues significant misclassification issues resulting in a notably poor classification result. The parallel convolutional classification networks PCCN-MSS presented in this article, as illustrated in Fig. 5(f), exhibit a remarkable decrease in misclassification compared to previously mentioned algorithms, underscoring its improved accuracy and effectiveness in handling complex classification tasks. In Fig. 5(f), the classification result of PCCN-MSS reveals a notable decrease in misclassification compared with 2-D-CNN. Furthermore, an enhancement in the classification results of each category depicted in Fig. 5(d) shows improvement when compared to those obtained using the 3-D-CNN. Despite the significant advancements with 3-D-CNN, there remains a substantial scope for further improvement. The incorporation of a parallel network enhances the performance of the model, delivering superior classification results. Specifically, it effectively avoids

the misidentification of impervious surfaces, demonstrating its superiority in accurately distinguishing various categories.

In addition to analyzing the classification results, quantitative analysis is paramount in assessing model performance. Table III presents the accuracies achieved on the UAV H^2 dataset using the five aforementioned models (SVM, 1-D-CNN, 2-D-CNN, 3-D-CNN, SSAN). Notably, the OA of the SVM, 1-D-CNN, and 3-D-CNN classification results is approximately similar, with values hovering around 68.60%, 68.18%, and 65.33%, respectively. In contrast, the SSAN classification result exhibits the most pronounced misclassification problems, culminating in the lowest OA at 50.13%. Conversely, the parallel networks PCCN-MSS emerge victorious for other models with an outstanding OA, AA, and Kappa of 78.18%, 80.80%, and 0.716, respectively. Compared to these models, PCCN-MSS demonstrates significant improvements in OA, surpassing them by a margin of at least 9.58%. Moreover, the proposed network exhibits exemplary performance across all ground classes. Notably, the accuracy of seven classes exceeds 70%, comprising predominantly of Fallen withered grass, standing withered grass, impervious surface, tree, nudation, cars, and flowers in clusters. In conclusion, both qualitative results and quantitative evaluations emphasize that PCCN-MSS attains optimal classification results for the UAV H^2 imagery when compared to traditional machine learning and DL methods.

IV. DISCUSSION

This chapter comprehensively encompasses three distinct discussions designed to demonstrate the performance of the model

TABLE III
CLASSIFICATION RESULTS FOR UAV H^2 IMAGERY USING VARIOUS MODELS AND THE PROPOSED METHOD

class number	SVM	1-D-CNN	2-D-CNN	3-D-CNN	SSAN	PCCN-MSS
1	0.630750	0.644052	0.528981	0.622441	**0.831501**	0.773223
2	0.764018	0.743550	0.684178	0.688764	**0.867561**	0.815785
3	0.758814	0.809462	0.846588	0.818756	0.502664	**0.885990**
4	**0.728796**	0.575475	0.485626	0.491196	0.444007	0.719037
5	0.602276	0.647991	0.697734	0.653067	0.341648	**0.699232**
6	0.755028	0.613176	0.685402	0.600312	0.065120	**0.809050**
7	0.698468	0.658519	0.766144	0.804998	0.01212	**0.912441**
8	0.705187	0.532220	0.538880	0.621611	0.041784	**0.849545**
OA	68.60%	68.18%	62.53%	65.33%	50.13%	**78.18%**
AA	70.54%	65.31%	65.42%	66.26%	38.83%	**80.80%**
Kappa	0.599	0.594	0.530	0.561	0.393	**0.716**

With the highest precision highlighted in bold font.

proposed in this article: ablation study, parametric analysis, and analysis of model stability. Ablation study, methodically evaluates the ramifications of eliminating specific components from the model, thereby elucidating their respective contributions. The parametric analysis delves into the intricacies of how alterations in pivotal parameters influence the performance of PCCN-MSS. Ultimately, the analysis of model stability investigates the robustness and consistency of the model across repeated empirical endeavors.

A. Ablation Study

This article meticulously conducted three sets of ablation experiments to thoroughly investigate the impact of FPN, SA mechanisms, and ICA on the performance of PCCN-MSS. These experiments are specifically designed to delineate the contributions of these individual components to the overall model performance, elucidating their respective roles and significance in augmenting classification results.

1) Impact of the Construction of FPN in ICA-2-D-CNN Branch: The FPN module is superior for extracting features at different spatial scales in the image. A comparison of the classification performance between the PCCN-MSS network without FPN and with FPN for H^2 imagery is illustrated in Fig. 6. Specifically, Fig. 6(a) depicts the performance of PCCN-MSS without the inclusion of FPN, whereas Fig. 6(b) presents PCCN-MSS with FPN. The visual results vividly illustrate the disparities between the two sets of experiments.

Fig. 6(b) illustrates the superior performance of PCCN-MSS when incorporating the FPN for UAV H^2 imagery classification. Fig. 6(a) shows the model without FPN, whereas Fig. 6(b) demonstrates the improved PCCN-MSS. Fig. 6(c) serves as the GT benchmark for evaluation. The result maps exhibit reduced salt-and-pepper noise, indicating a more refined and accurate classification. In addition, the adoption of the FPN structure has led to a notable decrease in misclassifications caused by shading effects. The OA of PCCN-MSS with FPN surpasses that of its counterpart without FPN by approximately 1.23% (as presented in Table IV). The improvement can be attributed to the FPN to facilitate the model's learning of feature information across various spatial scales, thereby enabling it to extract and utilize
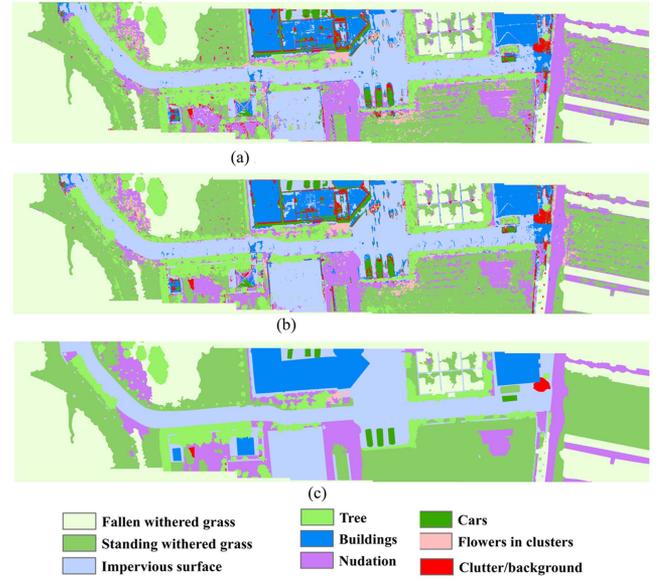

Fig. 6. Comparative analysis highlights the impact of integrating the FPN structure in the ICA-2-D-CNN branch. (a) PCCN-MSS without FPN. (b) PCCN-MSS. (c) GT.

TABLE IV
CLASSIFICATION RESULTS ABOUT ICA-2-D-CNN BRANCH OF PCCN-MSS ADOPTS FPN STRUCTURE

class number	PCCN-MSS without FPN	PCCN-MSS
1	0.749066	0.773223
2	0.793461	0.815785
3	0.873706	0.885990
4	0.750140	0.719037
5	0.711467	0.699232
6	0.793031	0.809050
7	0.908245	0.912441
8	0.818497	0.849545
OA	76.95%	**78.18%**
AA	79.97%	**80.80%**
Kappa	0.702	**0.716**

multispatial information more effectively from UAV H^2 imagery for classification purposes.

2) Impact of the Addition of SA Mechanisms: To validate the effectiveness of the SA mechanisms proposed in this study,
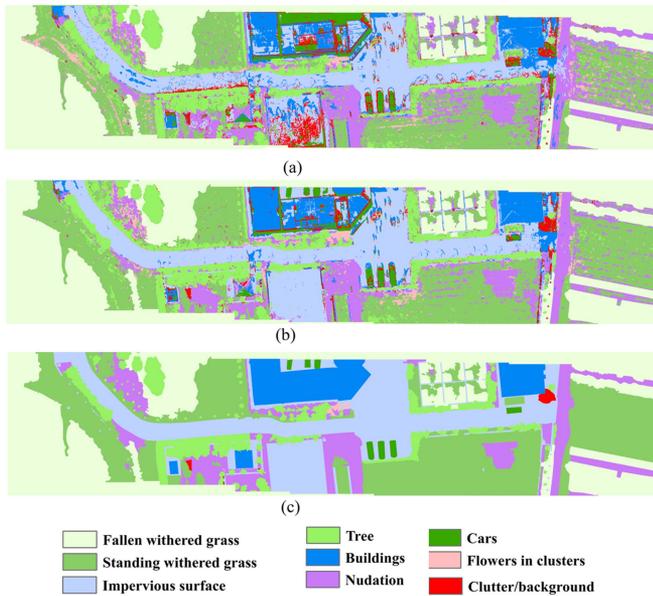
Fig. 7. Comparative analysis is conducted to assess the results achieved by incorporating SA under different conditions. (a) PCCN-MSS without SA. (b) PCCN-MSS. (c) GT.
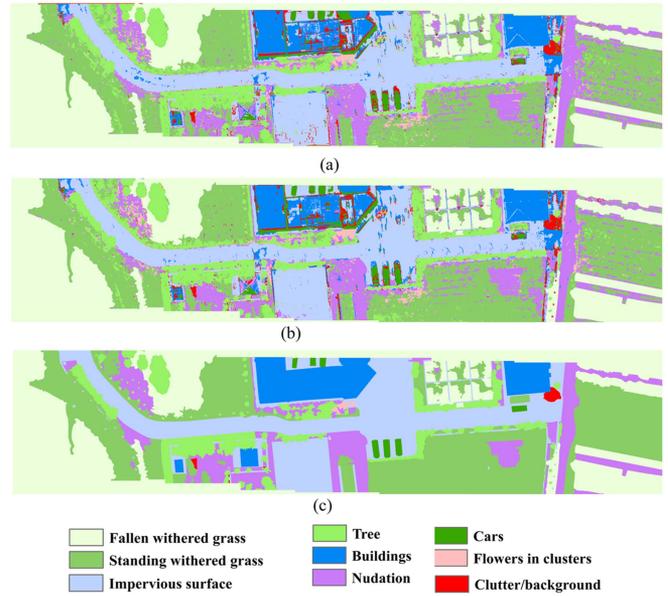
TABLE V
EVALUATION OF CLASSIFICATION ACCURACY OF SA MECHANISMS

class number	PCCN-MSS without SA	PCCN-MSS
1	0.669378	0.773223
2	0.648472	0.815785
3	0.866177	0.885990
4	0.556271	0.719037
5	0.685965	0.699232
6	0.735194	0.809050
7	0.811565	0.912441
8	0.607917	0.849545
OA	67.68%	**78.18%**
AA	69.76%	**80.80%**
Kappa	0.590	**0.716**

two distinct sets of attention mechanism combination experiments were conducted. Specifically, the PCCN-MSS model was implemented without SA blocks, as illustrated in Fig. 7(a), featuring a parallel configuration of ICA-2-D-FPN and 3-D-CNN branches. Additional experiments integrated SA blocks into the PCCN-MSS architecture and evaluated their impact on model performance, with the results presented in Fig. 7(b). Fig. 7(c) depicts the GT mapping for UAV H^2 imagery. To offer a comprehensive quantitative assessment, Table V presents a detailed overview of the classification accuracy achieved in each configuration, emphasizing the significant role played by SA blocks in enhancing model performance.

Fig. 7 illustrates a discernible improvement in performance boost for PCCN-MSS when the SA block is integrated, as compared to the parallel network model with SA. This is evident in the performance of the model to effectively mitigate the impact of shadows on the imagery. Furthermore, Table V provides unequivocal evidence that the integration of the SA block into the parallel network results in a progressive enhancement in the

Fig. 8. Impact of ICA of 2-D-FPN for UAV H^2 imagery classification. (a) PCCN-MSS without ICA. (b) PCCN-MSS. (c) GT.

accuracy of the classification model. Prominently, the model incorporating SA blocks exhibits a remarkable 10.50% increase in OA in comparison to the one without SA. These findings underscore the efficacy of the proposed SA block in extracting significant spatial and spectral information from images, thereby significantly improving model performance.

3) Impact of the Addition ICA: This section investigates the influence of ICA on the performance of the model through conducting two sets of experiments. A comparative analysis assessed the impact of implementing the ICA algorithm on the performance of PCCN-MSS. These experiments aim to explore both the effects and the overall performance gains achieved through the integration of the ICA algorithm on the 2-D-CNN branch. The results of these investigations are visually represented in Fig. 8, with Fig. 8(a) depicting the model without ICA and Fig. 8(b) showcasing PCCN-MSS, which incorporate the ICA algorithm on the 2-D-FPN branch and introducing SA in the 3-D-CNN branch of the parallel network. The accuracy of H^2 imagery pertaining to the ablation experiment is clearly illustrated in Table VI.

As evident from Fig. 8 and Table VI, the performance of Fig. 8(b) conspicuously surpasses that of Fig. 8(a). Therefore, it can be deduced that the integration of the ICA algorithm in the 2-D-CNN branch of the parallel convolutional classification network, which fuse multispatial scale and spectral feature, effectively enhances overall network performance for UAV H^2 image classification. The experimental figures clearly demonstrate that the classification performance of PCCN-MSS undergoes substantial improvement when incorporating the ICA algorithm in the 2-D-FPN branch. This enhancement is predominantly reflected in its proficiency to precisely identify buildings and effectively differentiate nudation and standing with withered grass. ICA separates data or signals as statistically independent linear combinations of non-Gaussian signal sources by linear

TABLE VI
RESULTS OF THE PARALLEL NETWORK BEFORE AND AFTER ADDING ICA IN 2-D-CNN BRANCH

class number	PCCN-MSS without ICA	PCCN-MSS
1	0.756021	0.773223
2	0.815049	0.815785
3	0.893393	0.885990
4	0.714622	0.719037
5	0.732491	0.699232
6	0.781449	0.809050
7	0.905509	0.912441
8	0.822249	0.849545
OA	77.97%	**78.18%**
AA	80.26%	**80.80%**
Kappa	0.714	**0.716**

transformation, ultimately elevating the classification accuracy of UAV H^2 images.

B. Parametric Analysis

The model performance of the PCCN-MSS model is intricately tied to its parameters. Hence, this section is dedicated to exploring the ramifications of adjusting the parameters within the PCCN-MSS model on image classification results. The parameters under examination encompass a range of SA mechanisms, dimensionality reduction techniques, the number of bands retained for ICA, the placement of the SA mechanisms proposed in this article, and the weights "*a*" and "*b*" in (1). The comprehensive discussion encompasses all experiments conducted using the PCCN-MSS parallel network, which comprises two distinct branches: the 2-D-FPN branch that leverages ICA to excavate multispatial scale features and facilitate reduction dimensionality, and the 3-D-CNN branch with SA to extract spatial-spectral information. By delving into these parameters, we aim to gain a deeper understanding of how they collectively influence the model's performance and ultimately contribute to improved image classification accuracy.

1) Impact of Different SA Mechanisms for PCCN-MSS Model: The initial ablation experiment has demonstrated a substantial improvement in the performance of the proposed model after incorporating the SA mechanism. Attention mechanisms, known for their ability to enhance network performance by focusing on either spatial or spectral information, have proven beneficial in prior studies. In previous studies, the self-attention (SE) mechanism [48] stands as a classical approach. To validate the efficacy of the SA block introduced in this study, a comparison was made between the SE mechanism and the SA block within the context of the PCCN-MSS model. Fig. 9 illustrates the performance of PCCN-MSS with SE or SA, and GT. Furthermore, Table VII presents a quantitative comparison of classification accuracy, providing further evidence of the effectiveness of the SA block.

In Fig. 9, when examining the classification of standing withered grass, PCCN-MSS equipped with the SA block achieves notably better results than when using the SE mechanism. Specifically, the model with SA block reduces a considerable

Fallen withered grass · Tree · Cars · Standing withered grass · Buildings · Flowers in clusters · Impervious surface · Nudation · Clutter/background

Fig. 9. Impact of different attention mechanisms for PCCN-MSS model. (a) PCCN-MSS model with SE mechanism. (b) PCCN-MSS model with SA block. (c) GT.

TABLE VII
CLASSIFICATION ACCURACIES OF PCCN-MSS WITH DIFFERENT ATTENTION MECHANISMS FOR UAV H^2 IMAGES CLASSIFICATION

class number	PCCN-MSS with SE	PCCN-MSS
1	0.694693	0.773223
2	0.802517	0.815785
3	0.876672	0.885990
4	0.776474	0.719037
5	0.717106	0.699232
6	0.827358	0.809050
7	0.885808	0.912441
8	0.840540	0.849545
OA	75.48%	**78.18%**
AA	80.26%	**80.80%**
Kappa	0.685	**0.716**

number of misclassifications, whereas the SE method tends to misidentify Standing withered grass as flowers in clusters. Table VII further illustrates this comparison: the left column presents the classification accuracy when the SE network is integrated into the PCCN-MSS, while the right column reflects the accuracy gained by adding the SA block to the 3-D-CNN branch of PCCN-MSS. Moreover, by incorporating the SA block proposed in this study into the parallel network, OA surpasses that of the SE network by 2.69% for the UAV H^2 imagery dataset. This underscores the superior ability of the SA block to preferentially learn more spectral features, leading to improved categorization of UAV H^2 image categorization.

2) Impact of Methods for Dimensionality Reduction in 2-D-FPN Branch: PCCN-MSS comprises two branches: the SA-3-D-CNN branch and the ICA-2-D-FPN branch. In the ICA-2-D-FPN branch, the input data undergoes processing and dimensionality reduction using ICA. Ablation experiments focusing on ICA reveal its beneficial impact on model performance. While principal component analysis (PCA), another

Fig. 10. Results of 2-D-CNN branch with (a) PCA or (b) ICA, and (c) GT mapping for UAV H^2 imagery.

TABLE VIII
IMPACT OF USING ICA OR PCA IN 2-D-FPN BRANCH FOR UAV H^2 IMAGE CLASSIFICATION

class number	PCCN-MSS with PCA	PCCN-MSS
1	0.741278	0.773223
2	0.762661	0.815785
3	0.883866	0.885990
4	0.724984	0.719037
5	0.725012	0.699232
6	0.750173	0.809050
7	0.916271	0.912441
8	0.806210	0.849545
OA	75.85%	**78.18%**
AA	78.88%	**80.80%**
Kappa	0.688	**0.716**

well-known dimensionality reduction technique, differs from ICA in its approach to extracting the principal component from the data. In this section, PCA is employed in place of ICA in the 2-D-FPN branch to investigate the effects of various dimensionality reduction methods on model performance. This comparison aims to provide insights into the relative effectiveness of ICA and PCA within the PCCN-MSS framework. The results of this series of experiments are visualized in Fig. 10. Specifically, Fig. 10(a) illustrates the performance of the PCA method within the 2-D-FPN branch, while Fig. 10(b) showcases PCCN-MSS, wherein the raw data has undergone processing via ICA. To quantitatively assess the impact of these dimensionality reduction techniques, Table VIII presents the accuracy metrics associated with each experiment.

Fig. 10 reveals a notable difference in the performance of PCA and ICA when applied to UAV H^2 imagery. While both methods aim to reduce dimensionality, their impact on classification results varies. In particular, the results obtained using PCA exhibit slightly worse performance compared to ICA. One

observable issue in the PCA classification is the presence of significant salt-and-pepper noise within the impervious surfaces. This noise is believed to be primarily caused by shadows, which interfere more prominently with PCA than with ICA. A closer examination of Table VIII further underscores this point. The accuracy achieved by PCCN-MSS, which incorporates ICA for data processing, marginally exceeds that of PCA by approximately 2.33%. This relatively small margin, however, can be crucial in certain applications where high classification accuracy is paramount. These findings suggest that ICA is considerably more robust to shadow interference than PCA. This enhanced robustness may stem from the ability of ICA to capture non-Gaussian and higher order statistical dependencies within the data, which are often missed by PCA. Consequently, when dealing with complex imagery such as UAV H^2 data, where shadows and other forms of noise are prevalent, the use of ICA may be preferred.

3) Impact of Retained Bands in ICA Data on the Performance of PCCN-MSS Model: The ablation experiments conducted in chapter three underscore the pivotal role of ICA within the PCCN-MSS model. Consequently, this section provides a detailed discussion on determining the optimal number of bands retained for ICA. The dataset utilized in this study comprises a total of 200 bands from the original UAV H^2 data. To gain a comprehensive understanding, the experiments with varying numbers of ICA bands are retained, specifically 10, 20, 30, 40, 50, 100, 150, and the full 200 bands.

The results presented in Fig. 11(a) offer valuable insights into the relationship between the number of retained bands and the classification accuracy achieved by the 2-D-FPN when fed with ICA-processed data. This indicates that the optimal classification accuracy is attained when retained ICA components are 10. However, beyond a certain point, retaining additional bands may not necessarily contribute to improved accuracy but could instead introduce redundancy and potential noise into the model.

4) Impact of the Placement of the SA Mechanisms for PCCN-MSS Model: As evident from the preceding discussion, the SA mechanism (SA block) introduced in this article demonstrates remarkable performance. To further investigate its impact on model performance, a series of experiments were conducted by varying the position of the SA block within the model architecture. Specifically, the SA block was positioned ahead of the first convolutional block (Conv0), the second (Conv1), the third (Conv2), and the fourth (Conv3). The experimental results, presented in Fig. 11(b), provide valuable information for the optimal positioning of the SA block.

The findings depicted in Fig. 11(b) reveal a clear trend, indicating a gradual decline in model performance as the SA block is positioned deeper within the architectural design. One possible explanation for this trend is that the SA mechanism in this study employs global convolution to learn image features. As such, it is likely that the original data contains the richest and most relevant information for the SA block to learn effectively. Placing the SA block earlier in the architecture allows it to process this raw and unfiltered information, potentially leading to better feature extraction and ultimately superior model performance.
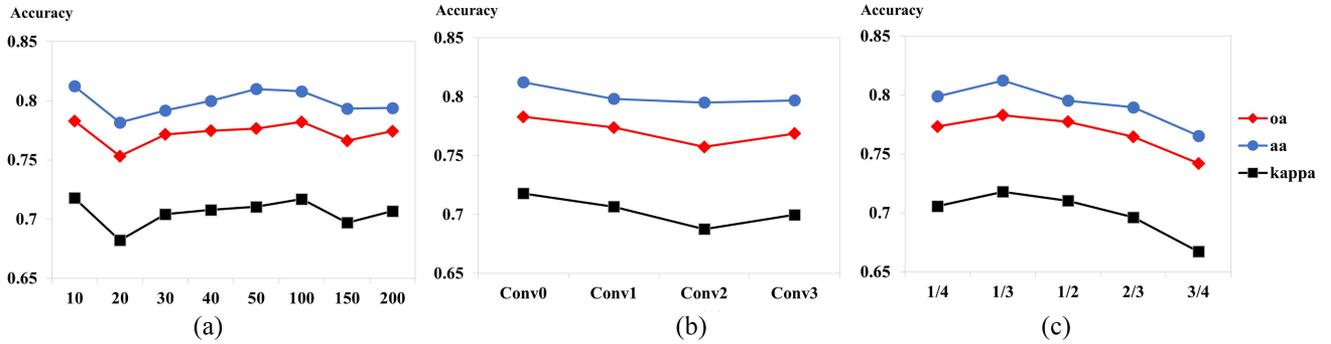
Fig. 11. Performance of the PCCN-MSS model is influenced by several key parameters. (a) Number of ICAs. (b) Positioning of the SA block. (c) Weights "*a*" and "*b*" utilized in (1).
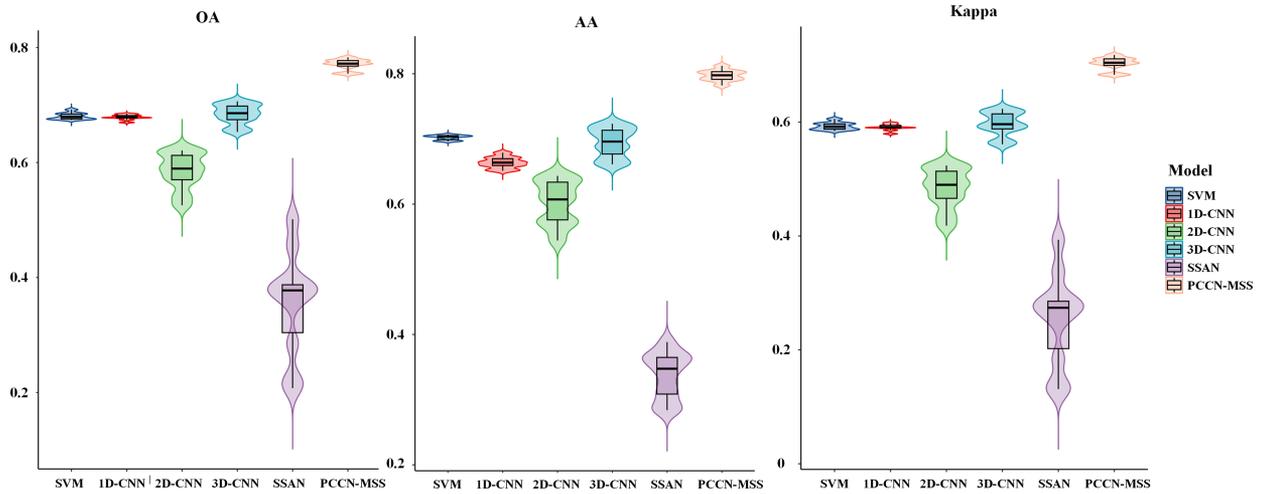

Fig. 12. Statistics of OA, AA, and Kappa accuracy obtained from 10 experiments respectively for the six models, SVM, 1-D-CNN, 2-D-CNN, and 3-D-CNN, and parallel network SSAN and PCCN-MSS, respectively.

*5) Impact of the Weights "*a*" and "*b*" in Two Branch Networks:* The model introduced in this article utilizes a weighted sum approach to integrate the predictions of the ICA-2-D-FPN branch and the SA-3-D-CNN branch. This integration method is mathematically represented in (1), which incorporates the weights "*a*" and "*b*." These weights play a pivotal role in determining the model's performance. Given the significance of these weights, this section delves into the impact of varying "*a*" and "*b*" on the performance of PCCN-MSS in UAV H^2 image classification. Specifically, we conducted experiments with different weight combinations, where "*a*" was set to 1/4, 1/3, 1/2, 2/3, and 3/4, while "*b*" was correspondingly calculated as 1-*a*. The resulting accuracy variations are graphically represented in Fig. 11(c).

Analysis of Fig. 11(c) reveals that the model achieves optimal performance when "*a*" is set to 1/3 and "*b*" to 2/3 for UAV H^2 imagery classification. This suggests that a balanced integration of the ICA-2-D-FPN and SA-3-D-CNN branches is crucial for achieving high classification accuracy. However, as the value of "*a*" increases beyond 1/3, the proportion of the ICA-2-D-FPN contribution of the branch to the weighted sum grows, leading to a gradual decrease in OA. This trend underscores the importance of carefully selecting the weights "*a*" and "*b*" to strike an effective balance between the two branches for optimal model performance.

C. Analysis of Model Stability

Ten experiments are conducted on SVM, 1-D-CNN, 2-D-CNN, 3-D-CNN, SSAN, and PCCN-MSS using identical initial parameters, i.e., epoch = 30 000, lr = 0.001, train number = 300, cube size = 9, and ICA number = 10. The statistics for the ten experiments are presented in Fig. 10, where from left to right indicates OA statistics, AA statistics, and Kappa statistics. The box plot, which includes the maximum, minimum, average values, and quartiles, effectively illustrates the data variability. This is specifically designated for the stability evaluation of the model in this experiment.

As shown in Fig. 12, SSAN exhibits the most significant fluctuations, followed by 2-D-CNN and 3-D-CNN. PCCN-MSS demonstrates a narrower OA fluctuation range of 75.45% to 78.29%, with an AA fluctuation of just 3.03%. However, the SSAN, which is also a parallel network, has an OA fluctuation range of 52.11% to 48.35%, with an AA fluctuation of 13.01%.

The PCCN-MSS model proposed in this article based on 2-D-FPN branch with ICA and the SA 3-D-CNN branch has relatively favorable stability.

V. CONCLUSION

HSIs with high spatial resolution hold tremendous potential for remote sensing fine-grained monitoring and classification. However, challenges such as high spatial heterogeneity, severe intra-class spectral variability and poor signal-to-noise ratio especially in UAV hyperspectral imagery constrain and hinder the performance of fine-grained classification. In this article, a parallel convolutional classification network model, PCCN-MSS model based on multimodal filters is proposed for precise H^2 imagery classification, which comprises two distinct branches, the ICA-2-D-FPN branch seamlessly integrates ICA with 2-D-CNN, leveraging FPN to effectively extract multispatial scale and spectral information from H^2 imagery. Meanwhile, the SA-3-D-CNN branch is meticulously designed to extract spatial and spectral information by synergistically combining SA block with 3-D-CNN.

Hyperspectral imagery is captured by UAVs, which contain vegetation and ground-based artifacts, characterized by its nanoscale spectral resolution and centimeter-level spatial resolution. The experimental results demonstrate that our proposed method achieves a satisfactory classification performance for H^2 imagery. Specifically, the PCCN-MSS model achieves impressive classification results with an OA of 78.18%, surpassing the compared methods by a significant margin of 9.58%. The ICA algorithm and SA mechanism of PCCN-MSS can significantly enhance the network's performance and improve classification accuracy. Furthermore, the OA of PCCN-MSS exhibits favorable stability, ranging between 75.45% and 78.29%, indicating a narrow fluctuation of only 3.03%.

By extracting global spatial-spectral information, the proposed PCCN-MSS method effectively tackles classification challenges related to the significant boundary inaccuracies and salt-and-pepper noise in traditional classification methods. This innovative approach delivers more satisfactory classification results with robust performance and remarkable advantages for H^2 imagery. The method presented in this article offers robust and reliable technical support for the precise classification and monitoring of ground features in upcoming widely used H^2 remote sensing images. Its versatility can extend to diverse downstream fields, including military surveying, agricultural production, environmental monitoring, and urban planning. In addition, this work lays a foundation for future quantitative inversion at a finer granularity of detail in remote sensing.

REFERENCES

[1] Y. Zhong, X. Hu, C. Luo, X. Wang, J. Zhao, and L. Zhang, "WHU-Hi: UAV-borne hyperspectral with high spatial resolution (H2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF," *IEEE Trans. Geosci. Remote Sens.*, vol. 250, pp. 112012–112029, 2020, doi: 10.1016/j.rse.2020.112012.

[2] L. P. Osco et al., "A review on deep learning in UAV remote sensing," *Int. J. Appl. Earth Observation Geoinf.*, vol. 102, pp. 102456–102476, 2021, doi: 10.1016/j.jag.2021.102456.

[3] E. Adam, O. Mutanga, and D. Rugege, "Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: A review," *Wetlands Ecol. Manage.*, vol. 18, pp. 281–296, 2009, doi: 10.1007/s11273-009-9169-z.

[4] V. E. Brando and A. G. Dekker, "Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 6, pp. 1378–1387, Jun. 2003, doi: 10.1109/tgrs.2003.812907.

[5] Y. Zhong et al., "Mini-UAV-borne hyperspectral remote sensing: From observation and processing to applications," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 4, pp. 46–62, Dec. 2018, doi: 10.1109/mgrs.2018.2867592.

[6] Y. N. Chen, T. Thaipisutikul, C. C. Han, T. J. Liu, and K. C. Fan, "Feature line embedding based on support vector machine for hyperspectral image classification," *Remote Sens.-Basel*, vol. 13, pp. 130–158, 2021, doi: 10.3390/rs13010130.

[7] G. Y. Chen, W. Xie, A. Krzyzak, and S.-E. Qian, "Hyperspectral image classification via principal component analysis, 2D spatial convolution, and support vector machines," *J. Appl. Remote Sens.*, vol. 15, no. 3, 2021, Art. no. 032202, doi: 10.1117/1.JRS.15.032202.

[8] S. Yuan et al., "MSLM-RF: A spatial feature enhanced random forest for on-board hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5534717, doi: 10.1109/tgrs.2022.3194075.

[9] B. Hu, Q. Li, and G. B. Hall, "A decision-level fusion approach to tree species classification from multi-source remotely sensed data," *ISPRS J. Photogramm.*, vol. 1, pp. 100002–100013, 2021, doi: 10.1016/j.ophoto.2021.100002.

[10] S. B. Kotsiantis, "Decision trees: A recent overview," *Artif. Intell. Rev.*, vol. 39, pp. 261–283, 2011, doi: 10.1007/s10462-011-9272-4.

[11] J. Y. Lee and T. A. Warner, "Segment based image classification," *Int. J. Remote Sens.*, vol. 27, pp. 3403–3412, 2007, doi: 10.1080/01431160600606866.

[12] Y. Dong, Q. Liu, B. Du, and L. Zhang, "Weighted feature fusion of convolutional neural network and graph attention network for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 31, pp. 1559–1572, 2022, doi: 10.1109/tip.2022.3144017.

[13] Y. Qing, W. Liu, L. Feng, and W. Gao, "Improved transformer net for hyperspectral image classification," *Remote Sens.-BASEL*, vol. 13, pp. 2216–2236, 2021, doi: 10.3390/rs13112216.

[14] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 639–643, May 2017, doi: 10.1109/LGRS.2017.2668299.

[15] H. Yu, L. Gao, W. Liao, B. Zhang, A. Pizurica, and W. Philips, "Multiscale superpixel-level subspace-based support vector machines for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, pp. 2142–2146, Nov. 2017, doi: 10.1109/lgrs.2017.2755061.

[16] S. Mei, J. Ji, Y. Geng, Z. Zhang, X. Li, and Q. Du, "Unsupervised spatial–Spectral feature learning by 3D convolutional autoencoder for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, pp. 6808–6820, Sep. 2019, doi: 10.1109/tgrs.2019.2908756.

[17] L. Mou, X. Lu, X. Li, and X. X. Zhu, "Nonlocal graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, pp. 8246–8257, Dec. 2020, doi: 10.1109/tgrs.2020.2973363.

[18] S. Bera, V. K. Shrivastava, and S. C. Satapathy, "Advances in hyperspectral image classification based on convolutional neural networks: A review," *Comput. Model. Eng. Sci.*, vol. 133, pp. 219–250, 2022, doi: 10.32604/cmes.2022.020601.

[19] Q. Wang, C. Yin, H. Song, T. Shen, and Y. Gu, "UTFNet: Uncertainty-guided trustworthy fusion network for RGB-thermal semantic segmentation," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 7001209, doi: 10.1109/lgrs.2023.3322452.

[20] X. Liu, H. Wang, Y. Meng, and M. Fu, "Classification of hyperspectral image by CNN based on shadow area enhancement through dynamic stochastic resonance," *Access*, vol. 7, pp. 134862–134870, 2019, doi: 10.1109/access.2019.2941872.

[21] X. Li, M. Ding, and A. Pižurica, "Group convolutional Neural networks for hyperspectral image classification," in *Proc. Int. Conf. Image Process.*, pp. 639–643, 2019, doi: 10.1109/ICIP.2019.8803839.

[22] Z. Yu, Y. Cui, C. Shao, S. Gao, and C. Wang, "Pyramidal and conditional convolution attention network for hyperspectral image classification using limited training samples," *Int. J. Remote Sens.*, vol. 43, pp. 2885–2914, 2022, doi: 10.1080/01431161.2022.2074808.

[23] H. A. H. Naji, T. Li, Q. Xue, and X. Duan, "A hypered deep-learning-based model of hyperspectral images generation and classification for imbalanced data," *Remote Sens.-Basel*, vol. 14, pp. 6406–6427, 2022, doi: 10.3390/rs14246406.

[24] J. Yang, Y. Zhao, C. Chan, and L. Xiao, "A multi-scale wavelet 3D-CNN for hyperspectral image super-resolution," *Remote Sens.-Basel*, vol. 11, pp. 1557–1578, 2019, doi: 10.3390/rs11131557.

[25] M. Kanthi, T. H. Sarma, and C. S. Bindui, "A 3D deep CNN based feature extraction and hyperspectral image classification," in *Proc. IEEE India Geosci. Remote Sens. Symp.*, 2020, pp. 229–232, doi: 10.1109/InGARSS48198.2020.9358920.

[26] J. Zhou, S. Zeng, G. Gao, Y. Chen, and Y. Tang, "A novel spatial–Spectral pyramid network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5519314, doi: 10.1109/tgrs.2023.3303338.

[27] M. Q. Alkhatib et al., "Tri-CNN: A three branch model for hyperspectral image classification," *Remote Sens.-Basel*, vol. 15, pp. 316–334, 2023, doi: 10.3390/rs15020316.

[28] U. A. Bhatti et al., "MFFCG – Multi feature fusion for hyperspectral image classification using graph attention network," *Expert Syst. Appl.*, vol. 229, pp. 120496–120509, 2023, doi: 10.1016/j.eswa.2023.120496.

[29] X. Hu et al., "SPNet: Spectral patching end-to-end classification network for UAV-borne hyperspectral imagery with high spatial and spectral resolutions," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5503417, doi: 10.1109/tgrs.2021.3049292.

[30] S. K. Roy, S. Manna, T. Song, and L. Bruzzone, "Attention-based adaptive spectral–Spatial kernel ResNet for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7831–7843, Sep. 2021, doi: 10.1109/tgrs.2020.3043267.

[31] Z. Zhang, T. Jiang, C. Liu, and L. Zhang, "An effective classification method for hyperspectral image with very high resolution based on encoder–Decoder architecture," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 1509–1519, 2021, doi: 10.1109/jstars.2020.3046245.

[32] X. Hu, X. Wang, Y. Zhong, and L. Zhang, "S3ANet: Spectral-spatial-scale attention network for end-to-end precise crop classification based on UAV-borne H2 imagery," *ISPRS J. Photogramm.*, vol. 183, pp. 147–163, 2022, doi: 10.1016/j.isprsjprs.2021.10.014.

[33] W. QingWang et al., "Multispectral point cloud superpoint segmentation," *Sci. China Technological Sci.*, pp. 183–198, 2023, doi: 10.1007/s11431-023-2528-8.

[34] 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 Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 277–281, Feb. 2020, doi: 10.1109/lgrs.2019.2918719.

[35] M. Kanthi, T. H. Sarma, and C. S. Bindu, " A 3D-deep CNN based feature extraction and hyperspectral image classification," in *Proc. IEEE India Geosci. Remote Sens. Symp.*, 2020, pp. 229–232, doi: 10.1109/InGARSS48198.2020.9358920.

[36] Q. Liu, L. Xiao, J. Yang, and Z. Wei, "CNN-enhanced graph convolutional network with pixel- and superpixel-level feature fusion for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8657–8671, Oct. 2021, doi: 10.1109/tgrs.2020.3037361.

[37] L. Yang, R. Zhang, Y. Bao, S. Yang, and L. Jiao, "Kernel tensor sparse coding model for precise crop classification of UAV hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 5511405, doi: 10.1109/lgrs.2023.3326452.

[38] X. Mei et al., "Spectral-spatial attention networks for hyperspectral image classification," *Remote Sens.-BASEL*, vol. 11, 2019, doi: 10.3390/rs11080963.

[39] D. Lupu, I. Necoara, J. L. Garrett, and T. A. Johansen, "Stochastic higher-order independent component analysis for hyperspectral dimensionality reduction," *IEEE Trans. Comput. Imag.*, vol. 8, pp. 1184–1194, 2022, doi: 10.1109/tci.2022.3230584.

[40] G. Salimi-Khorshidi, G. Douaud, C. F. Beckmann, M. F. Glasser, L. Griffanti, and S. M. Smith, "Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers," *Neuroimage*, vol. 90, pp. 449–468, 2014, doi: 10.1016/j.neuroimage.2013.11.046.

[41] M. Peng et al., "Mapping land subsidence and aquifer system properties of the Willcox Basin, Arizona, from InSAR observations and independent component analysis," *Remote Sens. Environ.*, vol. 271, pp. 112894–112908, 2022, doi: 10.1016/j.rse.2022.112894.

[42] R. J. Johnson, J. P. Williams, and K. W. Bauer, "AutoGAD: An improved ICA-based hyperspectral anomaly detection algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3492–3503, Jun. 2013, doi: 10.1109/tgrs.2012.2222418.

[43] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid Networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2117–2125, doi: 10.48550/.arXiv.1612.03144.

[44] Y. Liu, K. Cao, R. Wang, M. Tian, and Y. Xie, "Hyperspectral image classification of brain-inspired spiking neural network based on attention mechanism," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6009405, doi: 10.1109/lgrs.2022.3172410.

[45] L. Mou and X. X. Zhu, "Learning to pay attention on spectral domain A spectral attention module based convolutional network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, pp. 110–122, Jan. 2020, doi: 10.1109/TGRS.2019.2933609.

[46] J. He, L. Zhao, H. Yang, M. Zhang, and W. Li, "HSI-BERT: Hyperspectral image classification using the bidirectional encoder representation from transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 165–178, Jan. 2020, doi: 10.1109/tgrs.2019.2934760.

[47] M. E. Paoletti, S. Moreno-Álvarez, Y. Xue, J. M. Haut, and A. Plaza, "AAtt-CNN: Automatic attention-based convolutional neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5511118, doi: 10.1109/tgrs.2023.3272639.

[48] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

Linhuan Jiang received the B.S. degree in surveying and mapping engineering from Yuxi Normal University, Yunnan, China, in 2018. She is currently working toward the M.S. degree in photogrammetry and remote sensing with the Kunming University of Science and Technology, Yunnan, China.

Zhen Zhang received the B.S. degree in remote sensing science and technology, and the M.S., and Ph.D degrees in photogrammetry and remote sensing from the Shandong University of Science and Technology, Qingdao, China, in 2012, 2016, and 2021, respectively.

He is currently an Instructor with the Kunming University of Science and Technology, Yunnan, China. His research interests include hyperspectral remote sensing and wetland remote sensing.

Bo-Hui Tang (Senior Member, IEEE) received the B.S. degree from Wuhan University, Wuhan, China, in 1999, the M.S. degree from the China Remote Sensing Satellite Ground Station, Chinese Academy of Sciences, Beijing, China, in 2004, and the Ph.D. degree from the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China, in 2007, all in cartography and geographical information system.

He is currently a Research Fellow with the Kunming University of Science and Technology and the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences. His main research interests include the retrieval and validation of land surface net radiation, and land surface temperature and emissivity.

Lehao Huang received the B.S. degree in surveying and mapping engineering from Minjiang University, Fujian, China, in 2019. He is currently working toward the M.S. degree in surveying and mapping engineering with the Kunming University of Science and Technology, Yunnan, China.

Bingru Zhang received the B.S. degree in surveying and mapping engineering from Kunming University of Science and Technology, Yunnan, China, in 2019. She is currently working toward the M.S. degree in surveying and mapping engineering with the Kunming University of Science and Technology, Yunnan, China.