RESEARCH ARTICLE

An Exploration on the Influence Factors of the Optimal Sample Width for Hyperspectral Remote Sensing Image Classification

NINGYANG LI[1], **ZHAOHUI WANG**[1], (Member, IEEE),
FAOUZI ALAYA CHEIKH[2], (Senior Member, IEEE),
AND MOHIB ULLAH[2], (Member, IEEE)

[1]Faculty of Computer Science and Technology, Hainan University, Haikou 570228, China
[2]Department of Computer Science, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology, 2815 Gjøvik, Norway

Corresponding author: Zhaohui Wang (william_hig@163.com)

ABSTRACT Spectral-spatial classification of hyperspectral image (HSI) has made enormous achievements in many applications. One of the critical attributes that affects classification accuracy is the width of HSI cube/patch. To seek the optimal sample width, most researches enumerate all possible widths and verify them with the corresponding widths of HSI cubes in turn, which will require model to be particular for each width and consume plenty of time and computing power. In this article, the influential factors of the optimal sample width are studied from the perspectives of model architecture and data set for spectral-spatial classification of HSI. Specifically, to investigate the influence factors from model architecture, diverse numbers of filters and kernel sizes are applied in models for classification. The potential influence factors from data set are reflected mainly in the spatial distribution of land-cover, which can be described with short, long, and average edges. Moreover, according to the number of samples, five kinds of spatial distributions from the fewest category, fewer categories, larger categories, the largest category, and whole data set are considered. To explore the relationships between them and the optimal sample width, samples from the corresponding categories are expanded via a centralized spectral-spatial sample expansion method for classification. Experimental results show that the most possible influence factor of the optimal sample width primarily focuses on the neutral short edge of whole data set.

INDEX TERMS Hyperspectral image classification, optimal sample width, influence factor, neutral short edge, centralized sample expansion.

I. INTRODUCTION

Tremendous advancements of the spectral and spatial resolutions of imaging sensors enable hyperspectral remote sensing image to gain fine spectral reflectance characteristics and abundant spatial distributions from the surface of the earth [1], [2]. Such rich information provides solid support for

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

better analysis of hyperspectral image (HSI). Classification of HSI has drawn growing attention of researchers from various fields, such as precise agriculture [3], military reconnaissance [4], mineral exploration [5], urban planning [6], etc.

In the early days, conventional machine learning algorithms, including random forest [7], k-nearest neighbor [8], support vector machine [9], extended morphological profile [10], sparse representation [11], and so on, are exploited to deal with HSI classification. Although they received

good classification accuracy, there are still some deficiencies which are difficult to mitigate. On one hand, these methods are not promising at modeling the deep feature representation which is important for precise recognition of the HSI which contains hundreds of bands [12]. On the other hand, some of them process spectral information or spatial information selectively, which curbs the complementation of different types of features.

During the past decade, with the continuous upgrade of computational hardware, deep learning algorithms have reaped unprecedented achievement and become the dominant methods in the fields of natural language processing [13] and image analysis [14]. Earlier deep learning models used for HSI classification, such as artificial neural network [15], stacked autoencoder [16], deep brief network [17], recurrent neural network [18], long short-term network [19], and etc., were utilized to extract the high-level spectral features from each pixel and take advantage of the spectral correlations between adjacent bands. However, due to the scarcity of spatial information, the classification results of these models are still unsatisfactory. With the proposal of convolutional neural network (CNN) [20], the spatial features of HSI can be depicted by 2-D CNN while preserving the original spatial layout [21], [22], [23]. To integrate the spectral and spatial information, the models adopting the HSI cube/patch composed of the center pixel and its neighborhoods have been the popular choices. 3-D CNN has become the most classic architecture to describe the spectral-spatial features with many 3-D convolutional kernels [21]. During this stage, various techniques, such as residual block [24], dense connection [25], capsule unit [26], were embedded to improve the spectral-spatial feature representation from different perspectives for classification [27], [28], [29]. Recently, to focus on the meaningful contents, attention mechanism [30] was introduced to stress the salient bands and pixels in HSI cubes. The classic attention modules, such as self-attention (SA) module [31], squeeze and excitation (SE) module [32], convolutional block attention module (CBAM) [33], and etc., can model the correlations between the input and the critical information via the multiple dot-product similarity [34], [35], [36], [37], the global compressed representation [38], [39], and the large-scale dependency [40], [41], [42], [43]. More recently, some researchers attempted to deduce attention maps using the spectral similarities between the center pixel and its neighborhoods, which results in the predominant role of the center pixel [44], [45], [46], [47]. Attention-based methods pick out local bands and areas in an HSI cube to promote feature extraction, which changes the effective band ranges and spatial areas indirectly.

From the traditional methods to the deep learning methods, researchers are increasingly concerned about the expression of spatial information. It is known that an HSI cube is composed of the center pixel and its neighborhoods. How many neighborhoods should be considered depends on a specific range, also called sample width, which is of great importance on spatial feature extraction and classification results. The

larger the width is, the richer spatial information the sample contains. However, some interfering pixels having different labels from the center pixel may be involved, which is adverse to the extraction of discriminating features. Hence, to confirm the optimal sample width, it is almost indispensable for the aforementioned works to prepare a lot of samples with different widths according to enumeration rule and then perform a great deal of experiments on different data sets. This process usually not only requires model to own an appropriate architecture for each sample width, but also costs massive computational and time consumptions, especially for the complicated models.

To mitigate this issue, this article studies the most possible influence factors of the optimal sample width from the points of model and data set for the spectral-spatial classification of HSI. Specifically, the models with different numbers of filters and kernel sizes are first used to explore the effects of the first type of influence factor, i.e., model architecture. Another type of influence factor focuses on the spatial distributions of data set. It can be interpreted as the short edge, long edge, and average edge of each land-cover region. Considering that the numbers of samples of each category and even each land-cover region are generally unequal, a data set could be divided into five parts, including the fewest category, fewer categories, larger categories, the largest category, and all categories (whole data set). Moreover, the spatial distributions of the five parts can be neutral or weighted with the sample proportions. To explore the influences of different kinds of spatial distributions, samples from the five parts are expanded in multiple scales by a centralized sample expansion algorithm. The spatial distributions in expanded data sets can be altered as need. Then, the type of influential expanded data set, the importance of neutral and weight spatial distributions, and the most influential edge are investigated successively.

The contributions of this article are as follows.

1) The most possible influence factor of the optimal sample width for the spectral-spatial classification of HSI is deduced. It does not come from the model architecture, but the neutral short edge of whole data set. According to the experimental analyses, the suggested optimal sample widths of the popular HSI data sets are provided.

2) A centralized spectral-spatial sample expansion method is proposed to change the spatial distributions of HSI cubes. It takes the spectral and spatial correlations between the real neighboring pixels into full account when inferring the virtual pixels. The new expanded sample is centered on the center pixel, which can maintain the consistency of number of samples to ensure the fairness of the experiments on five kinds of expanded data sets. All neighboring pixels can participate in the expansion process whether they are labeled or not. Besides, the proposed sample expansion method is also a practical way for the data augment of HSI.

The remainder of this article is organized as follows. Section II introduces the fundamental concepts and workflow in detail. In Section III, experimental results and analyses on two types of influence factors are presented. Finally, the summary of this article is given in Section IV.

II. EXPLORATION METHOD
Before the exploration, it is essential to clarify the definition of the optimal sample width. Depending on it, two types of influence factors, model architecture and data set are explored. Moreover, to investigate the influences factors from data set, a new centralized spectral-spatial sample expansion algorithm is proposed. Finally, the details of flowchart of the exploration are introduced.

A. CLARIFICATION OF THE OPTIMAL SAMPLE WIDTH
In current HSI classification researches, there are generally two kinds of principles used to decide the optimal sample width (OSW). One called the peak based optimal sample width (P-OSW), which is the sample width resulting in the highest classification accuracy [21], [22], [23], [28], [29], [42], [46]. The other, named the trend based optimal sample width (T-OSW), determines the optimal sample width relying on the trend of accuracy [27], [34], [39], [40], [44], [45], [47].

Suppose that a group of accuracy $a = \{a_3, a_5, \ldots, a_{2n+1}\}$ represents the classification results of model using different sample widths $w = \{3, 5, \ldots, 2n+1\}$, where n denotes the number of sample widths. In this article, overall accuracy (OA) a, which is a popular metric to measure the ratio of the numbers of the correctly labeled pixels to total pixels, is utilized to evaluate the classification performance of models.

For the P-OSW, the highest accuracy is acquired first and then the corresponding sample width w_A will be selected as the optimal option. This process can be described as follow

$$w_P = I(\max(a)) \tag{1}$$

where $I(\cdot)$ is the mapping function used to acquire corresponding width from accuracy a.

The T-OSW focuses mainly on the part with relative stable accuracy. Results in this part fluctuate slightly and have been close to the highest accuracy. Mathematically, this situation can be summarized as

$$\max(a) - a_i \leq \beta \tag{2}$$

where β denotes the fluctuation parameter. If the difference between current accuracy a_i and the highest accuracy is not larger than the value of β, it will be screened as high accuracy a_h. From them, the minimum width w_T will be selected then as the optimal one.

$$w_T = \min(I(a_h)) \tag{3}$$

The comparison between the two kinds of OSWs is shown in Fig. 1. According to Equa. (1), the P-OSW $w_P = 15$ can be obtained. For the T-OSW, when the value of fluctuation parameters β is set to 5, five data points will be selected as
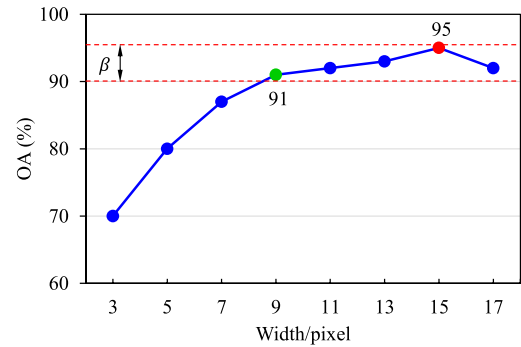

FIGURE 1. Comparison between P-OSW (red marker) and T-OSW (green marker). Where β is set to 5.

high accuracy $a_h = \{a_9, a_{11}, a_{13}, a_{15}, a_{17}\}$. The data point with minimum width $w_T = 9$ is the optimal one. Although such strategy may discard the highest classification accuracy, it can save a large amount of resource costs.

Besides, some researchers also try to choose the OSWs empirically [36], [37], [38], [41], [43]. But this solution is usually difficult to be applied in real scenarios.

In this exploration, the P-OSW is adopted as the evaluation metric of each influence factor. The reasons are as follows. First, the T-OSW depends on the fluctuation parameter which is a relatively subjective definition. Using the P-OSW without this parameter can ensure the fairness of comparison between data sets to some extent. Second, models are generally required to be trained sufficiently for stable classification performances, or the T-OSW may be identical to the P-OSW.

B. INFLUENCE FACTORS FROM MODEL ARCHITECTURE
The deep learning-based models have been applied widely to extract spectral-spatial features for HSI classification. Taking the classic CNN as an example, it portrays various kinds of features from HSI cubes with its complex convolutional kernels (i.e., filters). Also, feature dimension can be compressed using pooling layers. Parameters in the two kinds of components, such as the number of filters, kernel size, pooling size, strides, the number of layers, and padding pattern, may have different effects on the OSW. Comparing with the others, the number of filters and kernel size play a more important role in feature extraction [20], [24]. Hence, this study discusses the influence factors from model based on the two kinds of parameters.

The number of filters, $\mathcal{F}$, corresponds to the number of feature maps. Each filter represents an abstract feature subspace. Thus, the capability of model to fit the spectral-spatial distribution of HSI data depends on the number of filters to a great extent. To discover the association between the OSW and the number of filters, models with different numbers of filters $\mathcal{F} = \{4, 8, 16, 32, 64, 128, 256\}$ are applied to classify the samples with different widths.

The kernel size, $\mathcal{K}$, indicates the receptive fields of pixels in feature map. Small kernel size is good at capturing local details while large kernel size can abstract global semantic features. This difference may cause different classification

performances when applying these kernels in different widths of HSI cubes. Similarly, multiple kernels with the spatial sizes of $\mathcal{K} = \{3, 5, 7, 9, 11, 13, 15\}$ are set in model to probe the relationship between the OSW and kernel size. The spectral sizes of these kernels remain the original settings.

C. INFLUENCE FACTORS FROM DATA SET

Data set is a crucial participant in classification. HSI data sets are generally acquired from different scenarios, such as farmland, downtown, forest, and etc. Land covers in these scenarios have diverse spatial structures and spectral characteristics. For example, the identical crop tends to be dense, buildings have regular corners and high reflectivity, and rivers are long and narrow. They often possess inconsistent spectral reflectivity in both visible and infrared bands. In an HSI cube, sample width decides on the abundance of spatial structures directly [34], [40], [41], [44], but has little impact on spectral properties. Therefore, the influence factors from data set are defined in this section from the perspective of the spatial distributions of land covers.
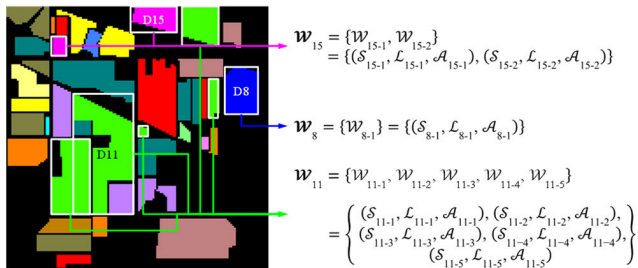

FIGURE 2. Spatial distributions of some categories. Spatial distribution of each sub-region (white rectangles) is represented with a triple $\mathcal{W}$ which contains short edge $\mathcal{S}$, long edge $\mathcal{L}$, and average edge $\mathcal{A}$.

1) NEUTRAL SPATIAL DISTRIBUTION

Assume that an HSI data set $\mathcal{H} \in \mathbb{R}^{r \times c \times b}$ is composed of D categories, where r, c, and b denote number of rows, number of columns, and number of bands, separately. There are N labeled pixels in ground truth map $\mathcal{G} \in \mathbb{R}^{r \times c}$. Taking the Indian Pines data set as an example, as shown in Fig. 2, there may be no less than one sub-region for each category. Let $\mathcal{R} = \{\mathcal{R}_1, \mathcal{R}_2, \cdots, \mathcal{R}_D\}$ denotes the numbers of sub-regions of all categories. In this figure, there are 2, 1, and 5 sub-regions for "D15", "D8", and "D11" categories, respectively. To compute the spatial sizes from these sub-regions with uneven shapes, an external rectangle (white rectangles in Fig. 2) is used to encompass each sub-region. This rectangle is represented with a triple $\mathcal{W}$ which contains a short edge $\mathcal{S}$, a long edge $\mathcal{L}$, and an average edge $\mathcal{A}$ (mean of short edge and long edge). The spatial distribution of ith category $\mathcal{W}_i^m$ can be interpreted by the mean of external rectangles of all sub-regions

$$\mathcal{W}_i^m = \left(\mathcal{S}_i^m, \mathcal{L}_i^m, \mathcal{A}_i^m \right) = \frac{1}{\mathcal{R}_i} \sum_{j=1}^{\mathcal{R}_i} \mathcal{W}_{ij}$$
$$= \frac{1}{\mathcal{R}_i} \sum_{j=1}^{\mathcal{R}_i} \left(\mathcal{S}_{ij}, \mathcal{L}_{ij}, \mathcal{A}_{ij} \right) \quad (4)$$

where $\mathcal{W}_{ij}$ is the external rectangle of jth sub-region of ith category. The operation of mean is conducted in each type of edge. After this, the overall spatial distribution of data set $\mathcal{W}_{all}^m$ can be obtained from the mean of the spatial distributions of all categories as follow.

$$\mathcal{W}_{all}^m = \left(\mathcal{S}_{all}^m, \mathcal{L}_{all}^m, \mathcal{A}_{all}^m \right) = \frac{1}{D} \sum_{i=1}^{D} \mathcal{W}_i^m \quad (5)$$

2) WEIGHTED SPATIAL DISTRIBUTION

For a data set, the numbers of samples of different categories or sub-regions are generally unequal. It is necessary to focus on this difference to improve the description of the spatial distribution. To this end, different weights are assigned for sub-regions and categories. From the ground truth map, the numbers of samples of all categories $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_D\}$ can be acquired, where the numbers of the samples in each sub-regions of ith category are $\mathcal{C}_i = \{\mathcal{C}_{i1}, \mathcal{C}_{i2}, \cdots, \mathcal{C}_{i\mathcal{R}_i}\}$. The proportions, i.e., weights, of samples of ith category $\mathcal{P}_i$ and its jth sub-regions $\mathcal{P}_{ij}$ are computed by

$$\mathcal{P}_i = \frac{\mathcal{C}_i}{N} = \frac{1}{N} \sum_{j=1}^{\mathcal{R}_i} \mathcal{C}_{ij} \quad (6)$$

$$\mathcal{P}_{ij} = \mathcal{P}_i \cdot \frac{\mathcal{C}_{ij}}{\mathcal{C}_i} = \frac{\mathcal{C}_{ij}}{N} \quad (7)$$

where N denotes the number of labeled pixels of data set.

Then, the weighted spatial distributions of ith category $\mathcal{W}_i^{wm}$ and whole data set $\mathcal{W}_{all}^{wm}$ are obtained by

$$\mathcal{W}_i^{wm} = \left(\mathcal{S}_i^{wm}, \mathcal{L}_i^{wm}, \mathcal{A}_i^{wm} \right) = \sum_{j=1}^{\mathcal{R}_i} \mathcal{P}_{ij} \cdot \mathcal{W}_{ij} \quad (8)$$

$$\mathcal{W}_{all}^{wm} = \left(\mathcal{S}_{all}^{wm}, \mathcal{L}_{all}^{wm}, \mathcal{A}_{all}^{wm} \right) = \sum_{i=1}^{D} \mathcal{P}_i \cdot \mathcal{W}_i^{wm}. \quad (9)$$

This kind of weighted spatial distribution takes the proportions of sub-regions and categories into consideration, which is a preferable description compared to neutral spatial distribution. In this exploration, they are all adopted as the potential influence factors of the OSW.
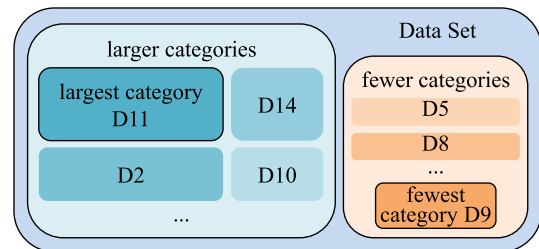
FIGURE 3. An HSI data set (taking Indian Pines as an example) is divided into five parts (rectangles with black edges) according to the number of samples, including the largest category, larger categories, the fewest category, fewer categories, and all categories.

3) SPATIAL DISTRIBUTIONS OF DIFFERENT CATEGORIES

However, it might still be difficult to locate the most possible influence factors of the OSW because the two kinds of descriptions are established on the hierarchy of data set (all categories). To find out the impact of the number of

samples on the selection of the OSW, an HSI data set is first divided into fewer category part and larger category part by a threshold $\mathcal{T}$. Categories have fewer numbers of samples ($\mathcal{P}_i \leq \mathcal{T}$) are in the first set whereas the rest ($\mathcal{P}_i > \mathcal{T}$) belong to the second one. Moreover, the category has the fewest number of samples (the fewest category) and the category has the largest number of samples (the largest category) are chosen to ascertain the influence of some extreme categories on the OSW further. Fig. 3 shows the relationship among the five parts. The neutral and weighted spatial distributions of them are as follows.

$$\mathcal{W}^m_{fest} = \left(\mathcal{S}^m_{fest}, \mathcal{L}^m_{fest}, \mathcal{A}^m_{fest}\right) = \mathcal{W}^m_{\mathcal{J}fest} \tag{10}$$

$$\mathcal{W}^{wm}_{fest} = \left(\mathcal{S}^{wm}_{fest}, \mathcal{L}^{wm}_{fest}, \mathcal{A}^{wm}_{fest}\right) = \sum_{j=1}^{\mathcal{R}_{\mathcal{J}fest}} \mathcal{P}_{\mathcal{J}fest\,j} \cdot \mathcal{W}_{\mathcal{J}fest\,j} \tag{11}$$

$$\mathcal{W}^m_{fer} = \left(\mathcal{S}^m_{fer}, \mathcal{L}^m_{fer}, \mathcal{A}^m_{fer}\right) = \frac{1}{D^{fer}} \sum_{i=1}^{D^{fer}} \mathcal{W}^m_{\mathcal{J}^{fer}_i} \tag{12}$$

$$\mathcal{W}^{wm}_{fer} = \left(\mathcal{S}^{wm}_{fer}, \mathcal{L}^{wm}_{fer}, \mathcal{A}^{wm}_{fer}\right) = \sum_{i=1}^{D^{fer}} \mathcal{P}_{\mathcal{J}^{fer}_i} \cdot \mathcal{W}^{wm}_{\mathcal{J}^{fer}_i} \tag{13}$$

$$\mathcal{W}^m_{ler} = \left(\mathcal{S}^m_{ler}, \mathcal{L}^m_{ler}, \mathcal{A}^m_{ler}\right) = \frac{1}{D^{ler}} \sum_{i=1}^{D^{ler}} \mathcal{W}^m_{\mathcal{J}^{ler}_i} \tag{14}$$

$$\mathcal{W}^{wm}_{ler} = \left(\mathcal{S}^{wm}_{ler}, \mathcal{L}^{wm}_{ler}, \mathcal{A}^{wm}_{ler}\right) = \sum_{i=1}^{D^{ler}} \mathcal{P}_{\mathcal{J}^{ler}_i} \cdot \mathcal{W}^{wm}_{\mathcal{J}^{ler}_i} \tag{15}$$

$$\mathcal{W}^m_{lest} = \left(\mathcal{S}^m_{lest}, \mathcal{L}^m_{lest}, \mathcal{A}^m_{lest}\right) = \mathcal{W}^m_{\mathcal{J}lest} \tag{16}$$

$$\mathcal{W}^{wm}_{lest} = \left(\mathcal{S}^{wm}_{lest}, \mathcal{L}^{wm}_{lest}, \mathcal{A}^{wm}_{lest}\right) = \sum_{j=1}^{\mathcal{R}_{\mathcal{J}lest}} \mathcal{P}_{\mathcal{J}lest\,j} \cdot \mathcal{W}_{\mathcal{J}lest\,j} \tag{17}$$

where $D^{fer}, D^{ler}, \mathcal{J}^{fer}, \mathcal{J}fest, \mathcal{J}^{ler}$, and $\mathcal{J}^{lest}$ indicate the numbers of fewer and larger categories, the indices of fewer, fewest, larger, and largest categories, separately.

The calculation processes of the neutral and weighted spatial distributions of fewer and larger categories, $\mathcal{W}^m_{fer}$, $\mathcal{W}^{wm}_{fer}$, $\mathcal{W}^m_{ler}$, and $\mathcal{W}^{wm}_{ler}$, are similar with those of data set, i.e., Equa. (5) and (9). The neutral and weighted spatial distributions of the fewest and largest categories, $\mathcal{W}^m_{fest}$, $\mathcal{W}^{wm}_{fest}$, $\mathcal{W}^m_{lest}$, and $\mathcal{W}^{wm}_{lest}$, are selected via Equa. (4) and (8).

In summary, ten influence factors, including the neutral and weighted spatial distributions of the fewest category, fewer categories, larger categories, the largest category, and whole data set, $\mathcal{W}^m_{fest}$, $\mathcal{W}^{wm}_{fest}$, $\mathcal{W}^m_{fer}$, $\mathcal{W}^{wm}_{fer}$, $\mathcal{W}^m_{ler}$, $\mathcal{W}^{wm}_{ler}$, $\mathcal{W}^m_{lest}$, $\mathcal{W}^{wm}_{lest}$, $\mathcal{W}^m_{all}$, and $\mathcal{W}^{wm}_{all}$ are designed to quest the association between them and the OSW.

D. CENTRALIZED SPECTRAL-SPATIAL SAMPLE EXPANSION

Experiments on original data sets cannot realize the goal of previous section because the ten influence factors are fixed, which results in that the effects of them on the OSW cannot be observed. There are two solutions, to shrink or to expand the data set. Considering that the inherent spatial structures of each category play an important role in deciding on the ten influence factors, the spatial structures should be preserved as well as possible in generated samples. Meanwhile,
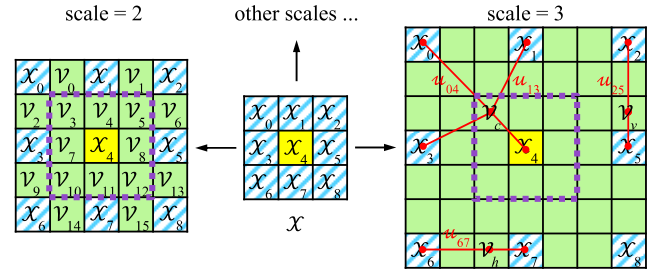

FIGURE 4. Expansion processes of an HSI cube ($\mathcal{w} = 3$) using the proposed algorithm. Where yellow pixels, pixels with blue diagonal, and green pixels indicate the center pixels, neighboring pixels, and virtual pixels, respectively. The red solid lines denote the spatial correlations between the virtual pixel and its referenced pixels. Pixels in purple dotted boxes form the new expanded samples.

the numbers of samples of each category should also be invariant, or the proportions of each sub-region and category will change. By the first solution, the shrinkage operation will erode the crucial spatial structures and decrease the number of available pixels. It may disturb the fairness of the comparison between original data set and shrunk data set. In contrast, the expansion strategy, which can remain initial spatial structures, is adopted. To keep the consistency of sample proportions among categories, the expansion operation is conducted on HSI cube merely instead of data set.

Previous work [48] attempts to generate new samples based upon available labeled pixels. The virtual (new) pixel is produced from two real (existing) pixels. During this process, the affinity is applied to assign unequal weights for two real pixels. But due to the lack of spatial correlation, the affinity between two pixels is identical no matter how far they are. That means the same virtual pixels will be generated in HSI cubes with different expansion scales, which may result in discontinuous spatial textures. To resolve this problem, a centralized spectral-spatial sample expansion algorithm is proposed to generate new samples. It takes spectral similarities and spatial correlations between real pixels into full account during generation process.

Let $\mathcal{X} \in \mathbb{R}^{\mathcal{w} \times \mathcal{w} \times b}$ represents an original HSI cube, where $\mathcal{w} = 3$ and b denote width and the number of bands, separately. As shown in Fig. 4, the expanded HSI cubes are composed of the pixels from original HSI cube and virtual pixels $\mathcal{V} \in \mathbb{R}^{1 \times 1 \times b}$. There are three kinds of virtual pixels, crossing pixel $\mathcal{V}_c$, horizontal pixel $\mathcal{V}_h$, and vertical pixel $\mathcal{V}_v$, in each expanded HSI cube. For crossing pixel $\mathcal{V}_c$, four closest real pixels (top-left $\mathcal{X}_0$, top-right $\mathcal{X}_1$, bottom-left $\mathcal{X}_3$, and bottom-right $\mathcal{X}_4$) are chose as the references for generation. While for the other two, $\mathcal{V}_h$ and $\mathcal{V}_v$, four closest real pixels (left $\mathcal{X}_6$, right $\mathcal{X}_7$) and (top $\mathcal{X}_2$, bottom $\mathcal{X}_5$) are chose, respectively.

To improve the reality of virtual pixel, the spectral similarity $\mathcal{s}$ and spatial correlation $\mathcal{u}$ between virtual pixel and real pixels are considered in this algorithm. Cosine distance [45], as a distance invariant metric, is utilized to measure the spectral similarity. The spatial correlation is evaluated based

upon the spatial positions of virtual pixel and real pixels. The two kinds of attributes can be described with

$$s_{ij} = \frac{\sum \mathcal{X}_i \cdot \mathcal{X}_j}{\sqrt{\sum \mathcal{X}_i^2} \cdot \sqrt{\sum \mathcal{X}_j^2}} \tag{18}$$

$$u_{ij} = 1 - \sqrt{\frac{(r_v - r_i)^2 + (c_v - c_i)^2}{(r_j - r_i)^2 + (c_j - c_i)^2}} \tag{19}$$

where $\mathcal{X}_i$, $\mathcal{X}_j$, r_v, and c_v denote ith and jth real pixels, the row and column of virtual pixel, separately.

Then, the virtual pixel V_{ij} is generated with the assistance of spectral similarity and spatial correlation

$$V_{ij} = \frac{u_{ij} + u_{ij} \cdot s_{ij}}{2} \cdot \mathcal{X}_i + \frac{(1 - u_{ij}) + (1 - u_{ij}) \cdot s_{ij}}{2} \cdot \mathcal{X}_j. \tag{20}$$

During the generation of virtual pixel, spectral similarity is weighted with the operation "$u \cdot s$", which allocates different spectral similarity for the pixels in different positions thereby relieving the issue that same similarity will cause unnatural spatial textures. The spectral similarity and spatial correlation are combined with the operation "$u + u \cdot s$" to make full use of the benefits for the generation.

As shown in the right part of Fig. 4, there are some disparities among the generation processes of three kinds of virtual pixels. For the crossing virtual pixel V_c, its corresponding two spectral similarities, s_{04}, s_{13}, and two spatial correlations, u_{04}, u_{13}, are obtained first. Then the mean of virtual pixels, V_{04} and V_{13}, is set to the final result of V_c. For the horizontal and vertical virtual pixels, V_h and V_v, they can be deduced directly by Equa. (18)-(20). The only difference is that their spatial correlations, u_{67} and u_{25}, are simplified as there are two closest real pixels only participate in respective generation process. The definitions of simplified spatial correlations are as follows.

$$u_{67} = 1 - (c_v - c_6) / (c_7 - c_6) \tag{21}$$

$$u_{25} = 1 - (r_v - r_2) / (r_5 - r_2) \tag{22}$$

In an expanded HSI cube, the center pixel and its neighborhoods in a region with size of $w \times w$ (purple dotted rectangles in Fig. 4) are cropped merely as a new sample, which increases the efficiency of algorithm in practice.

Certainly, there are also other generation strategies, such as pixel rearrangement [49]. It arranges the pixels belong to the same category in horizontal or crossing ways. Because of the limitation of sample width, it is hard for this solution to vary the values of the influence factors from data set. Moreover, the rearrangement operation will destroy initial spatial structures. The proposed centralized sample expansion method is thus the preferable scheme to study the influence factors from data set.

E. WORKFLOW

After clarifying the fundamental concepts, the workflow of this exploration is presented in this section. As shown in
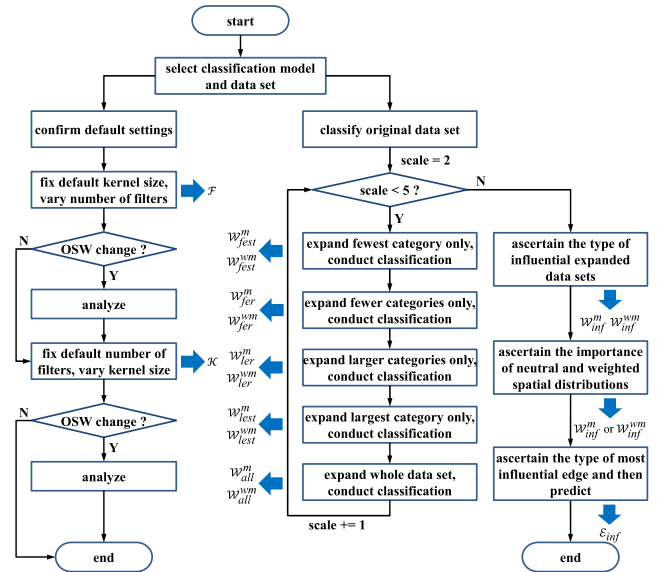

FIGURE 5. Flowchart of the exploration. In left part, two influence factors from model architecture are explored. In right part, ten influence factors from data set are investigated.

Fig. 5, after selecting a classification model and a data set, the two types of influence factors are explored in two pathways.

Before exploring the influence factors from model architecture, the default number of filters $\mathcal{F}^d$ and default kernel size $\mathcal{K}^d$ of model are confirmed from the related article [21]. Then, by setting kernel size to a default value, models with different numbers of filters $\mathcal{F}$ are trained and tested. If the OSW is changed, the further analysis will be conducted to quantify the effect of number of filters. If not, next influence factor, kernel size $\mathcal{K}$, will be investigated. In this process, kernel size is variable but number of filters is pined. Similarly, further analysis will be done if the OSW changes. Or the exploration of the first kind of influence factors is end.

The influence factors from data set are analyzed based upon the classification results on the original and expanded data sets. The fewest category, fewer categories, larger categories, the largest category, and all categories are expanded so that the impacts of different neutral and weighted spatial distributions, $\mathcal{W}_{fest}^m$, $\mathcal{W}_{fest}^{wm}$, $\mathcal{W}_{fer}^m$, $\mathcal{W}_{fer}^{wm}$, $\mathcal{W}_{ler}^m$, $\mathcal{W}_{ler}^{wm}$, $\mathcal{W}_{lest}^m$, $\mathcal{W}_{lest}^{wm}$, $\mathcal{W}_{all}^m$, and $\mathcal{W}_{all}^{wm}$, on the OSW can be discovered, respectively. For convenience, the five kinds of expanded data sets are denoted as "ex-fest", "ex-fer", "ex-ler", "ex-lest", and "ex-all". To ensure the robustness of final conclusion, three scales (2, 3, and 4) of the expanded data sets are prepared. The original data set can be seen as the special expanded data set with the expansion scale of 1. Classification results on it are the benchmark for those of other five kinds of expanded data sets.

After all experiments, the influential expanded data sets will be found via quantitative analysis. Then the comparison between neutral and weighted spatial distributions, $\mathcal{W}_{inf}^m$ and $\mathcal{W}_{inf}^{wm}$, is further conducted to ascertain the more important one. Finally, the correlations between the OSW and three

kinds of edges of resulting spatial distribution are compared and confirmed. On the basis of the influential factor $\mathcal{E}_{inf}$, the suggested OSWs of other HSI data sets are predicted.

III. EXPERIMENTAL RESULTS AND ANSLYSES

In this section, data sets for verification are presented first. Then the summaries of popular models for HSI classification are described. Finally, the influence factors from model and the influence factors from data set are analyzed in sequence.

A. DATA SETS

In this exploration, four public HSI data sets [50], [51], including Indian Pines, Botswana, Dioni and Loukia, which were collected over dense farmland, scattered delta, and mountainous region. This is to reduce the deviation of results and ensure the reliability of conclusion.

The Indian Pines data set was gathered by the Airborne Visible/Infrared Imaging Spectrometer sensor over Indian Pines test site in 1992. It contains 145×145 pixels and 200 bands. The spatial resolution of data set is 20 meters/pixel (m/p). There are 16 categories in 10249 labeled pixels.

The Botswana data set is collected by the Hyperion sensor mounted on the Earth Observing-1 (EO-1) satellite over the Okavango Delta, Botswana in 2001. It consists of 1476×256 pixels and 145 bands. The spatial resolution is 30 m/p. There are 14 categories in 3248 labeled pixels.

The Dioni and Loukia data sets were also obtained by the Hyperion sensor mounted on the EO-1 satellite. They are the HyRANK benchmark datasets, which were developed in the framework of the ISPRS Scientific Initiatives in 2018. The spatial sizes of the two data sets 1376×250 and 945×249, respectively. Both have 176 spectral bands. There are 12 categories in 20025 labeled pixels and 14 categories in 13503 available pixels for Dioni and Loukia data sets, separately.

The false-color images and the ground-truth maps are shown in Fig. 6. As presented in Tables 1-4, (15% and 85%) of the labeled pixels are selected randomly as the training and test sets for the four data sets. In this exploration, the threshold $\mathcal{T}$ used to separate fewer categories from larger categories is set to 5%.

B. MODELS FOR EXPLORATION

To ensure the effectiveness and completeness of the research, four popular deep learning models, including 2-D CNN [21], 3-D CNN [21], spectral spatial residual network (SSRN) [27], residual spectral spatial attention network (RSSAN) [40], spectral-similarity-based spatial attention module (S^3AM) [45], and multi-head transformer-based attention with CNN-wavelet (MT-CW) [37] are chose. The brief descriptions are as follows.

2-D CNN: It takes the HSI cube as the input and extracts the spatial features with several 2-D convolutional layers. The label is deduced by a fully-connected layer.
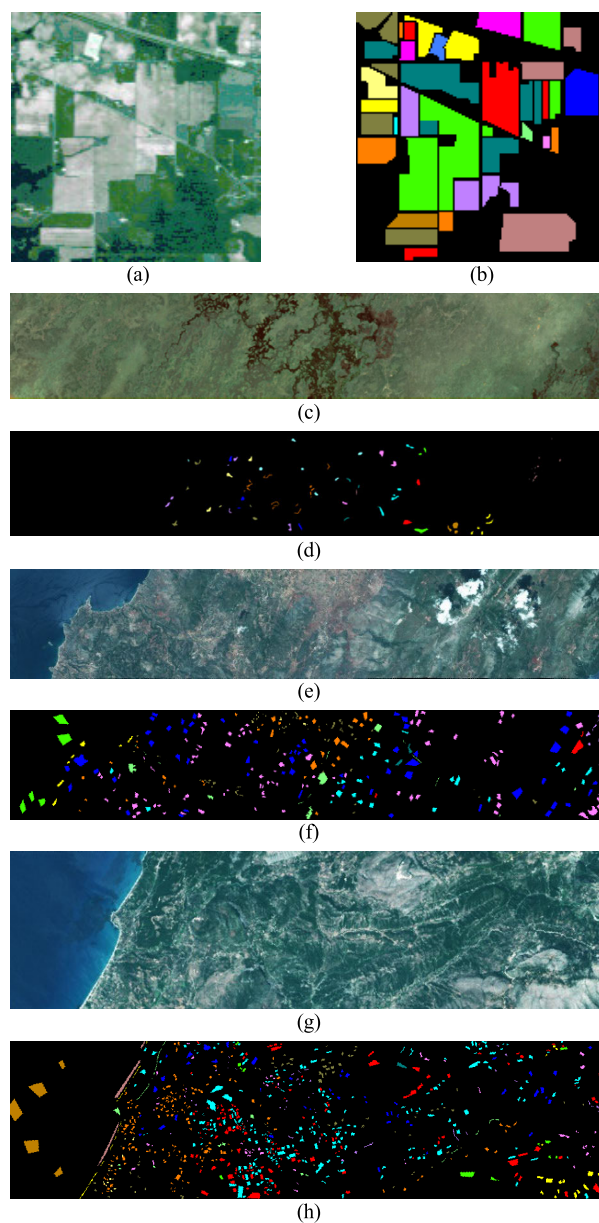
FIGURE 6. False color images and ground truth maps of four data sets. (a) and (b): Indian Pines, (c) and (d): Botswana, (e) and (f): Dioni, (g) and (h): Loukia.

3-D CNN: Its architecture is similar to that of 2-D CNN. With the 3-D convolutional layers, the spectral-spatial features can be extracted from HSI cubes.

SSRN: It is a deep 3-D CNN-based framework which adopts the residual connections to improve the backpropagation of gradients and alleviate the phenomenon of declining-accuracy.

RSSAN: It uses 2-D residual network as baseline to extract the spectral and spatial features and the convolutional block attention module to select useful spectral bands and enhance "good" pixels and weaken "bad" pixels in HSI cube.

S^3AM: It infers the spatial attention mask based on the spectral similarity between the center pixel of HSI cube and

TABLE 1. Numbers of samples in Indian pines data set.

No.	Land-cover type	Training	Test	Total	Proportion
1	Alfalfa	7	39	46	0.45%
2	Corn-notill	214	1214	1428	13.93%
3	Corn-mintill	124	706	830	8.10%
4	Corn	36	201	237	2.31%
5	Grass-pasture	72	411	483	4.71%
6	Grass-trees	109	621	730	7.12%
7	Grass-pasture-mowed	4	24	28	0.27%
8	Hay-windrowed	72	406	478	4.66%
9	Oats	3	17	20	0.20%
10	Soybean-notill	146	826	972	9.48%
11	Soybean-mintill	368	2087	2455	23.95%
12	Soybean-clean	89	504	593	5.79%
13	Wheat	31	174	205	2.00%
14	Woods	190	1075	1265	12.34%
15	Building-Grass-Trees-Drives	58	328	386	3.77%
16	Stone-Steel-Towers	14	79	93	0.91%
	Total	1537	8712	10249	100%

TABLE 2. Numbers of samples in Botswana data set.

No.	Land-cover type	Training	Test	Total	Proportion
1	Water	41	229	270	8.31%
2	Hippo grass	15	86	101	3.11%
3	Floodplain grasses1	38	213	251	7.73%
4	Floodplain grassed2	32	183	215	6.62%
5	Reeds1	40	229	269	8.28%
6	Riparian	40	229	269	8.28%
7	Firescar2	39	220	259	7.97%
8	Island interior	31	172	203	6.25%
9	Acacia woodland	47	267	314	9.67%
10	Acacia shrublands	37	211	248	7.64%
11	Acacia grasslands	46	259	305	9.39%
12	Short mopane	27	154	181	5.57%
13	Mixed mopane	40	228	268	8.25%
14	Exposed soils	14	81	95	2.92%
	Total	487	2761	3248	100%

TABLE 3. Numbers of samples in Dioni data set.

No.	Land-cover type	Training	Test	Total	Proportion
1	Dense urban fabric	189	1073	1262	6.30%
2	Mineral extraction sites	30	174	204	1.02%
3	Non irrigated arable land	92	522	614	3.07%
4	Fruit trees	22	128	150	0.75%
5	Olive groves	265	1503	1768	8.83%
6	Coniferous forest	54	307	361	1.80%
7	Dense sclerophyllous vegetation	756	4280	5036	25.14%
8	Sparse sclerophyllous vegetation	956	5418	6374	31.83%
9	Sparcely vegetated areas	263	1491	1754	8.76%
10	Rocks and sand	73	419	492	2.46%
11	Water	241	1371	1612	8.05%
12	Coastal water	59	339	398	1.99%
	Total	3000	17025	20025	100%

TABLE 4. Numbers of samples in Loukia data set.

No.	Land-cover type	Training	Test	Total	Proportion
1	Dense urban fabric	44	244	288	2.13%
2	Mineral extraction sites	11	56	67	0.50%
3	Non irrigated arable land	82	460	542	4.01%
4	Fruit trees	12	67	79	0.59%
5	Olive groves	211	1190	1401	10.38%
6	Broad-leaved forest	34	189	223	1.65%
7	Coniferous forest	75	425	500	3.70%
8	Mixed forest	161	911	1072	7.94%
9	Dense sclerophyllous vegetation	569	3224	3793	28.09%
10	Sparse sclerophyllous vegetation	421	2382	2803	20.76%
11	Sparcely vegetated areas	61	343	404	2.99%
12	Rocks and sand	74	413	487	3.61%
13	Water	209	1184	1393	10.32%
14	Coastal water	68	383	451	3.34%
	Total	2032	11471	13503	100%

total iterations is 200. To accelerate convergence, the learning rate is decayed by 50% per 40 epochs. Each experiment is performed ten times and the average results are reported so as to relieve the impact of the randomness caused by different initialization of parameters and samples.

All experiments are conducted on a computer with an Intel Core i9-9900K at 3.60 GHz central processing unit × 16 with 64-GB RAM and a TITAN RTX graphical processing unit with 24-GB RAM. The operating system is the Ubuntu 18.04.6 LTS 64-bit. The deep learning frameworks are Tensorflow 2.0.0 and Keras 2.3.1.

C. ANALYSES OF THE INFLUENCE FACTORS FROM MODEL ARCHITECTURE

The influence factors, including the number of filters $\mathcal{F}$ and kernel size $\mathcal{K}$, are the common parameters of a CNN, which play an important role in spectral-spatial feature extraction. To clarify the impacts of the two factors on the OSW, models with different numbers of filters $\mathcal{F} = \{4, 8, 16, 32, 64, 128, 256\}$ as well as various kernel sizes $\mathcal{K} = \{3, 5, 7, 9, 11, 13, 15\}$ are built to classify the samples in different widths $w = \{3, 5, \ldots, 31\}$. Considering that the 2-D CNN and 3-D CNN are the representative feature extractors, so the classification results of them are reported.

1) ANALYSIS OF NUMBER OF FILTERS

From Fig. 7, there are three conclusions can be derived. First, for each number of filters, the OAs of models remain the growing trends roughly as sample width becomes larger. This is because a wider sample contains richer spatial information. Second, from the view of sample width, the corresponding curves have similar shapes, especially for the larger widths, e.g., 25, 27, 29, and 31. Third, the OSWs for all numbers of filters are almost all 31 (blue solid lines) according to Equa. (1). If the OSW can be affected by number of filters, the three characteristics will not valid. Hence there is weak association between the OSW and number of filters.

2) ANALYSIS OF KERNEL SIZE

From Fig. 8, there are also three conclusions can be obtained. First, for each kernel size, the OAs of models keep the growing trends as sample width becomes larger, which is similar

its neighborhoods. This module was installed before the 3-D residual network to refine the input.

MT-CW: It first uses multi-head attention module to select important bands and then extract the spatial features with CNN and Wavelet transform. The combined features are sent to classifier for the final prediction.

The architectures of all models are duplicated according to original articles during re-implementation. The parameters of all models are initialized by the *Xavier* normal distribution [52]. During training procedure, the RMSprop [53] optimizer is used to propagate gradients backward. The values of its parameters (*learning rate*, *beta1*, *beta2*) are set to (0.001, 0.9, 0.999). The batch size is 32 and the number of
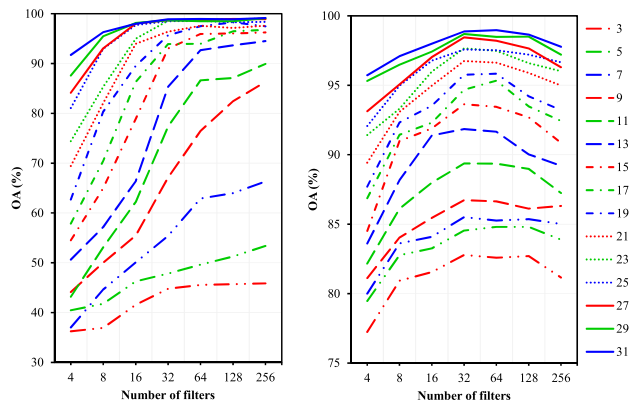
FIGURE 7. OAs of the 2-D CNN (left) and 3-D CNN (right) with different numbers of filters on different widths of samples of Indian Pines data set.
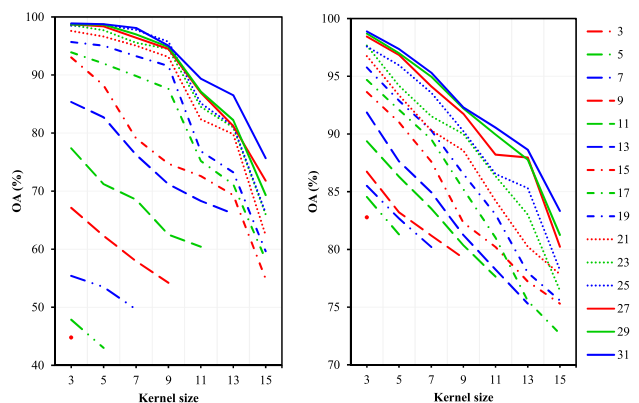

FIGURE 8. OAs of the 2-D CNN (left) and 3-D CNN (right) with different kernel sizes on different widths of samples of Indian Pines data set.

to the case of the number of filters. Second, from the angle of sample width, the curves of their OAs all represent the downward tendencies, which may be led by the abundant parameters in larger kernels. Last but not least, the OSWs for all kernel sizes are almost also 31 except for the 2-D CNN with the kernel size of 9. Based on these analyses, the relationship between the OSW and kernel size is also little.

D. ANALYSES OF THE INFLUENCE FACTORS FROM DATA SET

In this section, the impacts of the neutral and weighted spatial distributions of the fewest category, fewer categories, larger categories, the largest category, and all categories, $\mathcal{W}_{fest}^{m}$, $\mathcal{W}_{fest}^{wm}$, $\mathcal{W}_{fer}^{m}$, $\mathcal{W}_{fer}^{wm}$, $\mathcal{W}_{ler}^{m}$, $\mathcal{W}_{ler}^{wm}$, $\mathcal{W}_{lest}^{m}$, $\mathcal{W}_{lest}^{wm}$, $\mathcal{W}_{all}^{m}$, and $\mathcal{W}_{all}^{wm}$, are analyzed via the classification results of models between the original (scale = 1) and expanded (scale = 2, 3, 4) data sets. During experiments, the parameters of number of filters and kernel size of the selected models are set to the default options.

1) QUANTITATIVE COMPARISON

Fig. 9 shows the OAs of four models on different widths of samples from original and five kinds of expanded data

sets. It can be observed that the trends of accuracy polylines of RSSAN, S³AM, and MT-CW are more stable than those of SSRN. This is because the attention-based model which can recognize relevant areas well in HSI cubes for higher classification accuracy. Among all sub-figures, there are two similar attributes can be obtained. First, the OAs of models on the original data sets are higher than those on other five kinds of expanded data sets when sample width are smaller than 15. Moreover, the more the number of expanded samples (ex-fest < ex-fer < ex-lest < ex-ler < ex-all), the lower the OAs are. This is caused probably by the expansion operation which limits the spatial information in small ($w < 15$) HSI cubes. Second, the OAs of models on expanded data sets begin to increase gradually when sample width is larger than 15. In particular, the growths of OAs of models on "ex-all" and "ex-ler" data sets are more and more obvious with the increase of expansion scale. It is because partial lost spatial information may be supplemented in wider samples. As marked in Fig. 9, this also results in that the OSWs shift to the right with numbers of expanded samples rise, which is obvious when expansion scale is set to 4. But the OAs of corresponding OSWs in each sub-figure are almost equal.

The aforementioned phenomena demonstrate that there is a certain association indeed between the OSW and the ten kinds of spatial distributions. Moreover, the validity of the proposed spectral-spatial sample expansion method is also confirmed.

2) QUALITATIVE COMPARISON

Fig. 10 presents the classification maps of SSRN and RSSAN on original and ex-all Indian Pines data set. The corresponding OAs are 98.91%, 98.84%, 98.82%, 98.89%, 99.12%, 99.12%, 99.19%, and 99.09%, respectively. The OAs of models on expanded data sets are very close to those on original data set. Moreover, it can be seen that the eight prediction maps are similar to the groundtruth map shown in Fig. 6(b). The number of noises and incorrect corners in each sub-figure is same as others visually. This reveals that the centralized sample expansion algorithm can reserve the crucial information well.

3) ASCERTAIN THE TYPES OF INFLUENTIAL EXPANDED DATA SETS

Tables from 5 to 8 report the neutral/weighted spatial distributions $\mathcal{W}^{m}/\mathcal{W}^{wm}$ and OSWs w_{P} of models on original and five kinds of expanded data sets, respectively. To find out the importance of the five kinds of expanded data sets, three edges of neutral/weighted spatial distributions in three scales are combined with corresponding OSWs (marked with black rectangles in each table) to form 48 basic units for calculating Pearson correlation coefficient. In particular, when expansion scale is 1, the results on the original data set (long and black rectangles) are copied for other five kinds of expanded data set.

As shown in Table 9, the Pearson correlation coefficients $\wp$ between the spatial distribution and OSW increase when number of expanded samples n raises. Hence, the correlation
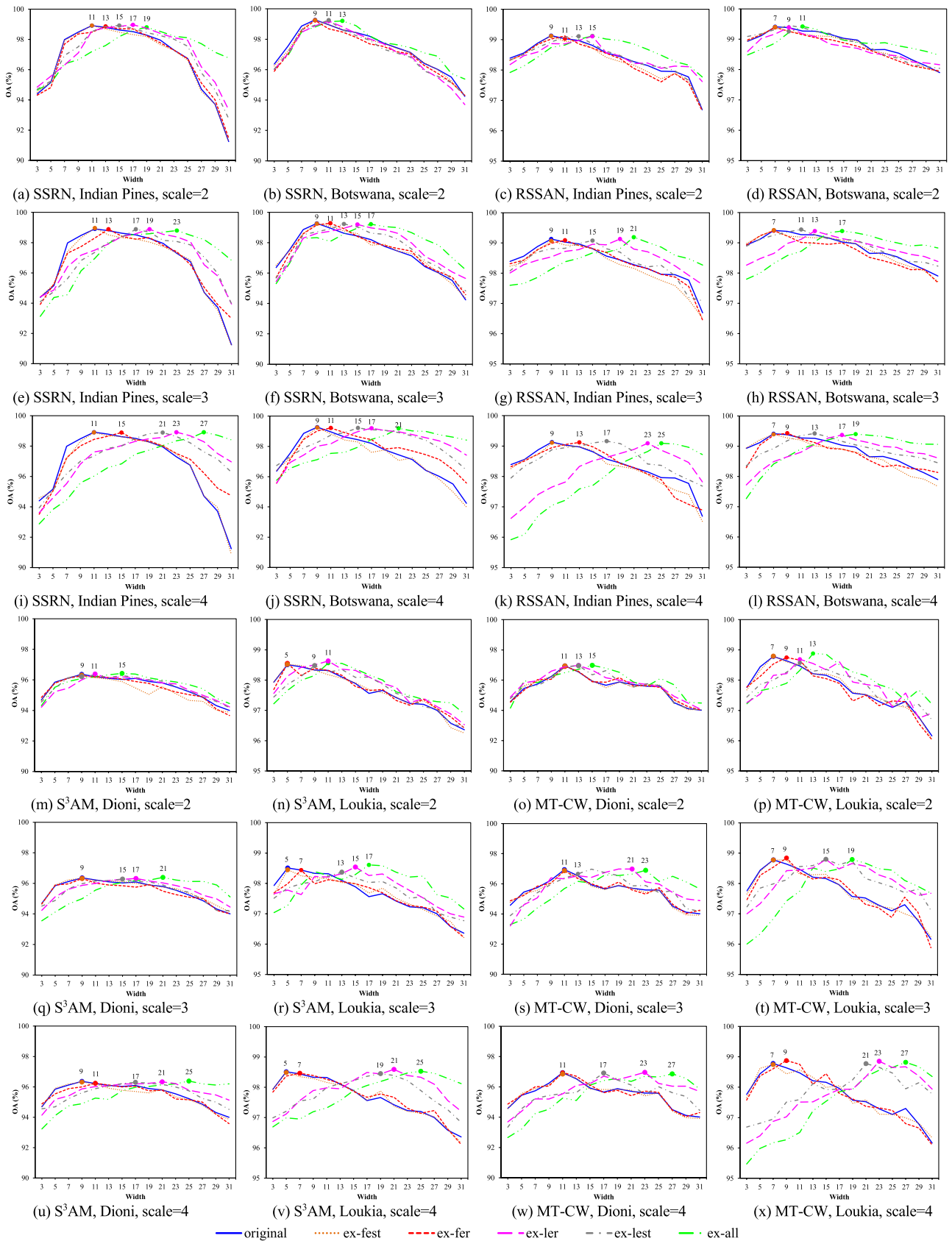
FIGURE 9. OAs of SSRN, RSSAN, S³AM, and MT-CW on different widths of samples from original (scale = 1) and expanded (scale = 2, 3, 4) Indian Pines, Botswana, Dioni, and Loukia data sets. Each sub-figure contains the comparison of original and five kinds (ex-fest, ex-fer, ex-ler, ex-lest, and ex-all) of expanded data sets. Where the solid data markers represent OSWs.
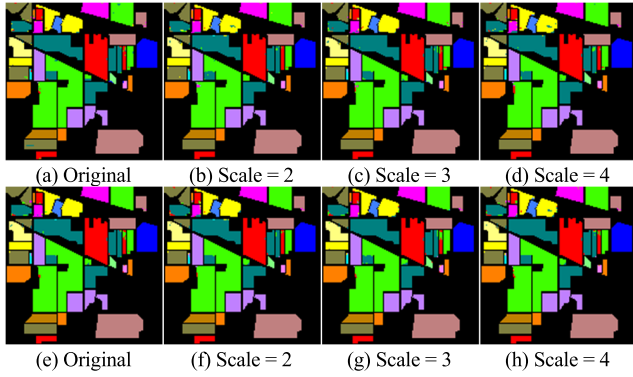
(a) Original   (b) Scale = 2   (c) Scale = 3   (d) Scale = 4

(e) Original   (f) Scale = 2   (g) Scale = 3   (h) Scale = 4

FIGURE 10. Classification maps of SSRN (a)-(d) and RSSAN (e)-(h) on the original and ex-all Indian Pines data sets when OSWs were adopted.

TABLE 5. Neutral/weighted spatial distributions and OSWs of SSRN and RSSAN on original and expanded Indian pines data set.

Sample	Expanded category	Neutral spatial distribution	Weighted spatial distribution	OSW (SSRN | RSSAN)	
Original (Scale = 1)	Fewest (9)	(2.00, 10.00, 6.00)	(2.00, 10.00, 6.00)	—	—
	Fewer (1, 4, 5, 7 - 9, 13, 15, 16)	(9.83, 16.97, 13.40)	(13.94, 22.18, 18.06)	—	—
	Larger (2, 3, 6, 10 - 12, 14)	(**13.23**, 24.43, 18.83)	(18.95, 35.52, 27.23)	—	—
	Largest (11)	(17.80, 31.80, 24.80)	(26.11, 48.21, 37.16)	—	—
	All	(**11.32**, 20.23, 15.78)	(17.98, 32.95, 25.46)	11	9
Expanded (Scale = 2)	Ex-fest	(**11.44**, 20.86, 16.15)	(17.99, 32.97, 25.48)	11	9
	Ex-fer	(16.85, 29.78, 23.32)	(20.67, 37.22, 28.95)	13	11
	Ex-ler	(17.10, 30.92, 24.01)	(33.27, 61.62, 47.45)	17	15
	Ex-lest	(12.43, 22.22, 17.33)	(24.23, 44.49, 34.36)	15	13
	Ex-all	(22.64, 40.46, 31.56)	(35.96, 65.90, 50.92)	19	15
Expanded (Scale = 3)	Ex-fest	(**11.57**, 21.48, 16.53)	(17.99, 32.99, 25.49)	11	9
	Ex-fer	(22.38, 39.33, 30.85)	(23.36, 41.50, 32.43)	13	11
	Ex-ler	(22.89, 41.61, 32.25)	(48.57, 90.28, 69.43)	19	19
	Ex-lest	(13.54, 24.21, 18.88)	(30.49, 56.04, 43.26)	17	15
	Ex-all	(33.96, 60.69, 47.34)	(53.94, 98.85, 76.38)	23	21
Expanded (Scale = 4)	Ex-fest	(**11.69**, 22.11, 16.90)	(17.99, 33.00, 25.50)	11	9
	Ex-fer	(27.91, 48.88, 38.39)	(26.04, 45.78, 35.91)	15	13
	Ex-ler	(28.68, 52.30, 40.49)	(63.86, 118.95, 91.41)	23	23
	Ex-lest	(14.66, 26.20, 20.43)	(36.74, 67.59, 52.17)	21	17
	Ex-all	(45.28, 80.92, 63.12)	(71.92, 131.8, 101.84)	27	25

TABLE 6. Neutral/weighted spatial distributions and OSWs of SSRN and RSSAN on original and expanded Botswana data set.

Sample	Expanded category	Neutral spatial distribution	Weighted spatial distribution	OSW (SSRN | RSSAN)	
Original (Scale = 1)	Fewest (14)	(2.91, 4.82, 3.86)	(4.02, 6.96, 5.49)	—	—
	Fewer (2, 14)	(4.95, 7.74, 6.35)	(5.92, 9.33, 7.63)	—	—
	Larger (1, 3 - 13)	(**8.08**, 12.50, 10.29)	(9.43, 15.05, 12.24)	—	—
	Largest (9)	(9.00, 15.40, 12.20)	(10.74, 17.88, 14.31)	—	—
	All	(**7.63**, 11.82, 9.73)	(9.22, 14.71, 11.96)	9	7
Expanded (Scale = 2)	Ex-fest	(**7.84**, 12.16, 10.00)	(9.34, 14.91, 12.12)	9	7
	Ex-fer	(8.34, 12.93, 10.63)	(9.58, 15.27, 12.42)	9	7
	Ex-ler	(14.56, 22.53, 18.55)	(18.08, 28.85, 23.47)	11	9
	Ex-lest	(8.28, 12.92, 10.60)	(10.26, 16.43, 13.35)	11	9
	Ex-all	(15.26, 23.64, 19.46)	(18.44, 29.42, 23.92)	13	11
Expanded (Scale = 3)	Ex-fest	(**8.05**, 12.51, 10.28)	(9.46, 15.11, 12.28)	9	7
	Ex-fer	(9.05, 14.03, 11.54)	(9.94, 15.83, 12.88)	11	7
	Ex-ler	(21.48, 33.25, 27.37)	(26.95, 42.99, 34.97)	15	13
	Ex-lest	(8.92, 14.02, 11.47)	(11.30, 18.16, 14.73)	13	11
	Ex-all	(22.89, 35.46, 29.19)	(27.66, 44.13, 35.88)	17	17
Expanded (Scale = 4)	Ex-fest	(**8.26**, 12.85, 10.55)	(9.57, 15.32, 12.44)	9	7
	Ex-fer	(9.76, 15.14, 12.45)	(10.29, 16.40, 13.34)	11	9
	Ex-ler	(28.41, 43.96, 36.19)	(35.81, 57.13, 46.47)	17	17
	Ex-lest	(9.56, 15.12, 12.34)	(12.33, 19.89, 16.11)	15	13
	Ex-all	(30.52, 47.28, 38.92)	(36.88, 58.84, 47.84)	21	19

coefficients between the spatial distributions of "ex-all" data set and OSWs are highest. Similarly, the "ex-ler" data set receives the second highest correlation. But for the "ex-fest" data set, its correlation coefficients are zero, which implies that expanding the fewest category almost has no effect on OSW. The correlation coefficients on "ex-fer" and "ex-lest" data sets are still in low levels because the number of expanded samples are still limited. Therefore, the spatial

TABLE 7. Neutral/weighted spatial distributions and OSWs of S^3AM and MT-CW on original and expanded Dioni data set.

Sample	Expanded category	Neutral spatial distribution	Weighted spatial distribution	P-OSW (S^3AM | MT-CW)	
Original (Scale = 1)	Fewest (4)	(2.35, 3.03, 2.69)	(3.32, 4.34, 3.83)	—	—
	Fewer (2 - 4, 6, 10, 12)	(7.61, 10.62, 9.11)	(11.85, 14.79, 13.32)	—	—
	Larger (1, 5, 7 - 9, 11)	(**10.78**, 14.73, 12.75)	(13.96, 17.99, 15.98)	—	—
	Largest (8)	(12.23, 15.46, 13.84)	(15.59, 19.13, 17.36)	—	—
	All	(**9.19**, 12.67, 11.93)	(13.72, 17.64, 15.68)	9	11
Expanded (Scale = 2)	Ex-fest	(9.39, 12.93, 11.16)	(13.75, 17.67, 15.71)	9	11
	Ex-fer	(13.00, 17.98, 15.49)	(15.04, 19.28, 17.16)	9	11
	Ex-ler	(14.58, 20.04, 17.31)	(26.14, 33.64, 29.89)	11	13
	Ex-lest	(10.21, 13.96, 12.09)	(18.69, 23.73, 21.21)	11	13
	Ex-all	(18.38, 25.34, 21.86)	(27.45, 35.28, 31.36)	15	15
Expanded (Scale = 3)	Ex-fest	(**9.58**, 13.18, 11.38)	(**13.77**, 17.70, 15.74)	9	11
	Ex-fer	(16.80, 23.29, 20.04)	(16.35, 20.92, 18.63)	9	11
	Ex-ler	(19.97, 27.40, 23.69)	(38.55, 49.64, 44.09)	17	19
	Ex-lest	(11.23, 15.25, 13.24)	(23.65, 29.81, 26.73)	15	15
	Ex-all	(27.58, 38.02, 32.80)	(41.17, 52.91, 47.04)	21	23
Expanded (Scale = 4)	Ex-fest	(**9.78**, 13.43, 11.60)	(**13.80**, 17.74, 15.77)	9	11
	Ex-fer	(20.60, 28.60, 24.60)	(17.66, 22.55, 20.11)	11	11
	Ex-ler	(25.36, 34.77, 30.06)	(50.96, 65.64, 58.30)	21	23
	Ex-lest	(12.25, 16.54, 14.39)	(28.61, 35.90, 32.26)	17	17
	Ex-all	(36.77, 50.69, 43.73)	(54.90, 70.55, 62.72)	25	27

TABLE 8. Neutral/weighted spatial distributions and OSWs of S^3AM and MT-CW on original and expanded Loukia data set.

Sample	Expanded category	Neutral spatial distribution	Weighted spatial distribution	P-OSW (S^3AM | MT-CW)	
Original (Scale = 1)	Fewest (2)	(10.00, 16.00, 13.00)	(10.00, 16.00, 13.00)	—	—
	Fewer (1 - 4, 6, 7, 11, 12, 14)	(5.45, 8.39, 6.92)	(9.47, 15.16, 12.31)	—	—
	Larger (5, 8 - 10, 13)	(**7.00**, 9.30, 8.15)	(7.96, 11.12, 9.54)	—	—
	Largest (9)	(4.00, 5.65, 4.82)	(6.21, 9.04, 7.63)	—	—
	All	(**6.00**, 8.72, 7.36)	(8.30, 12.03, 10.17)	5	7
Expanded (Scale = 2)	Ex-fest	(**6.72**, 9.86, 8.29)	(8.35, 12.11, 10.23)	5	7
	Ex-fer	(9.51, 14.11, 11.81)	(10.43, 15.45, 12.94)	5	9
	Ex-ler	(8.50, 12.04, 10.27)	(14.47, 20.65, 17.56)	11	11
	Ex-lest	(6.29, 9.12, 7.70)	(10.05, 14.57, 12.31)	9	11
	Ex-all	(12.00, 17.44, 14.72)	(16.60, 24.06, 20.33)	11	13
Expanded (Scale = 3)	Ex-fest	(**7.43**, 11.00, 9.22)	(8.40, 12.19, 10.30)	5	7
	Ex-fer	(13.01, 19.51, 16.26)	(12.57, 18.86, 15.71)	7	9
	Ex-ler	(11.00, 15.36, 13.18)	(20.64, 29.27, 24.96)	15	17
	Ex-lest	(6.57, 9.52, 8.05)	(11.79, 17.11, 14.45)	13	15
	Ex-all	(18.01, 26.15, 22.08)	(24.91, 36.10, 30.50)	17	19
Expanded (Scale = 4)	Ex-fest	(**8.15**, 12.15, 10.15)	(**8.45**, 12.27, 10.36)	5	7
	Ex-fer	(16.51, 24.90, 20.71)	(14.70, 22.27, 18.49)	7	11
	Ex-ler	(13.50, 18.68, 16.09)	(26.81, 37.89, 32.35)	21	23
	Ex-lest	(6.86, 9.93, 8.39)	(13.53, 19.65, 16.59)	19	21
	Ex-all	(24.01, 34.87, 29.44)	(33.21, 48.13, 40.67)	25	27

TABLE 9. Pearson correlation coefficients of neutral/weighted spatial distributions and OSWs on five kinds of expanded data sets.

		ex-fest	ex-fer	ex-ler	ex-lest	ex-all
Indian Pines	n	20	1976	8273	2455	**10249**
	p	0	0.4976	0.6167	0.3675	**0.7502**
Botswana	n	95	196	3052	314	**3248**
	p	0	0.1566	0.8431	0.3728	**0.8696**
Dioni	n	150	2219	17806	6374	**20025**
	p	0	0.2696	0.7271	0.4819	**0.8745**
Loukia	n	67	3041	10462	3793	**13503**
	p	0	0.4863	0.6827	0.3729	**0.8693**

distributions, $\mathcal{W}_{ler}^m$, $\mathcal{W}_{ler}^{wm}$, $\mathcal{W}_{all}^m$, and $\mathcal{W}_{all}^{wm}$, of "ex-all" and "ex-ler" data sets have stronger correlations with OSW.

4) ASCERTAIN THE IMPORTANCE OF NEUTRAL AND WEIGHTED SPATIAL DISTRIBUTIONS

In this section, the relative importance of neutral and weighted spatial distributions of "ex-ler" and "ex-all" data sets are evaluated. The mean absolute error (MAE) and root mean square error (RMSE) are employed to measure the difference between three kinds of edges of neutral/weighted spatial distribution and OSW. From Tables 10 and 11, it can be seen that the MAEs and RMSEs of each edge of neutral

TABLE 10. MAEs and RMSEs between neutral/weighted spatial distributions and OSWs on "Ex-ler" data sets.

Data set	Metric	Neutral			Weighted		
		short	long	average	short	long	average
Indian Pines	MAE	**3.00**	19.27	11.13	23.92	58.95	41.44
	RMSE	**3.62**	20.61	12.05	26.96	64.99	45.97
Botswana	MAE	**6.11**	15.64	10.71	10.27	23.67	16.97
	RMSE	**7.25**	17.83	12.51	12.16	26.71	19.42
Dioni	MAE	**2.23**	8.22	5.25	16.84	26.14	21.49
	RMSE	**2.57**	9.04	5.77	19.23	29.31	24.26
Loukia	MAE	4.25	**2.02**	2.71	3.81	11.21	7.51
	RMSE	5.16	**2.39**	3.47	4.03	11.84	7.92

TABLE 11. MAEs and RMSEs between neutral/weighted spatial distributions and OSWs on "Ex-all" data sets.

Data set	Metric	Neutral			Weighted		
		short	long	average	short	long	average
Indian Pines	MAE	**9.55**	32.08	20.95	26.45	63.88	45.15
	RMSE	**11.78**	36.13	23.98	29.98	70.92	50.43
Botswana	MAE	**5.17**	15.30	10.08	8.80	22.53	15.65
	RMSE	**6.31**	17.59	11.92	10.54	25.48	17.97
Dioni	MAE	**5.18**	13.43	9.33	16.06	25.85	20.95
	RMSE	**6.37**	15.68	11.01	18.54	29.21	23.87
Loukia	MAE	**1.25**	6.30	2.90	5.26	14.58	9.92
	RMSE	**1.41**	6.82	3.23	5.70	15.84	10.75

TABLE 12. MAEs and RMSEs between neutral spatial distribution (larger categories and all categories) of original data sets and OSWs.

Data set	Metric	Larger categories			All categories		
		short	long	average	short	long	average
Indian Pines	MAE	3.23	14.43	8.83	**1.32**	10.23	5.78
	RMSE	3.38	14.46	8.89	**1.67**	10.28	5.87
Botswana	MAE	**1.00**	4.50	2.29	**1.00**	3.82	1.73
	RMSE	**1.00**	4.61	2.50	1.07	3.95	2.00
Dioni	MAE	**1.00**	4.73	2.75	**1.00**	2.67	1.93
	RMSE	**1.27**	4.83	2.93	1.29	2.85	2.17
Loukia	MAE	**1.00**	3.30	2.15	**1.00**	2.72	1.36
	RMSE	1.41	3.45	2.37	**1.00**	2.90	1.69

TABLE 13. Suggested sample widths of some classic HSI data sets.

Data set	Neutral short edge of data set	Suggested sample widths
Indian Pines [50]	11.32	11
Botswana [50]	9.00	9
Pavia University [50]	4.39	5
Pavia Center [50]	4.91	5
Salinas [50]	6.87	7
Kennedy Space Center [50]	10.35	11
Loukia [51]	6.00	7, 5
Dioni [51]	9.19	9

spatial distributions are lower than those of weighted spatial distributions. This phenomenon may be caused by the weighting operation which enlarges the adverse influence of the category has more samples on final spatial distributions. In addition, it can be also observed that the MAEs and RMSEs of short edges, long edges, and average edges of neutral spatial distributions on the two expanded data sets are lowest, highest, middle, separately. This means adopting the HSI cubes with smaller width is prone to acquire higher accuracy. Such HSI cubes contain less irrelevant spatial structures. However, it is not sufficient to verify the strong correlation between the OSW and the neutral short edge because some error values of neutral long and average edges are smaller than those of the weighted short edges (marked with solid black rectangles). As a result, three edges of the neutral spatial distributions $\mathcal{W}^m_{ler}$, and $\mathcal{W}^m_{all}$ are all remained in subsequent exploration.

5) ASCERTAIN THE MOST INFLUENTIAL EDGE AND PREDICTION

To analyze the correlation between the neutral spatial distribution and OSW further, the MAEs and RMSEs between the three kinds of edges of larger categories and all categories of original data sets (bold values in Tables 5-8) and OSWs are presented in Table 12. It can be observed that the MAEs of neutral short edges are lower than those neutral long and average edges. That means there is a closer relationship between the OSW and neutral short edge. Compared with larger categories, the MAEs and RMSEs between the OSW and the neutral short edge of all categories are better. Besides, the neutral short edge of data set is more practical than that of larger categories. Since the threshold $\mathcal{T}$, which depends on the

properties of data set to select the fewer and larger categories is not easy to be determined. Therefore, the neutral short edge of data set $\mathcal{S}^m_{all}$ is the most influential factor of OSW.

Based on this conclusion, the recommended sample widths for other HSI data sets are provided in Table 13. In addition, previous analyses reveal that the optimal number of filters and kernel size are 32 and 3, respectively. With these suggestions, the training procedure and architecture design of some HSI classification methods, such as CNN-based models, attention mechanism-based models, and etc., can be handled efficiently and conveniently.

IV. CONCLUSION

In this article, the influence factors from model architecture and data set are explored to find out the determinant of the OSW for the spectral-spatial classification of HSI. For the influence factors from model architecture, models with different numbers of filters and kernel sizes are deployed for classification. For the influence factors from data set, five kinds of neutral and weighted spatial distributions are designed to evaluate their respective impacts on the OSW. Analyses of the classification results on four public HSI data sets reveal that the most possible influence factor of the OSW does not come from model architecture, but locates in the neutral short edge of all data set. Depending on this conclusion, the suggested OSWs of several HSI data sets are provided.

Future work will focus on how to describe the spatial distributions of data set preciously. The external rectangle used in this exploration is proper for the land-cover with orderly and fragmented regions. Some deviations may be resulted when encountering some narrow and inclined land-covers.

REFERENCES

[1] M. Ahmad, S. Shabbir, S. K. Roy, D. Hong, X. Wu, J. Yao, A. M. Khan, M. Mazzara, S. Distefano, and J. Chanussot, "Hyperspectral image classification—Traditional to deep models: A survey for future prospects," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 968–999, 2022.

[2] M. A. Moharram and D. M. Sundaram, "Land use and land cover classification with hyperspectral data: A comprehensive review of methods, challenges and future directions," *Neurocomputing*, vol. 536, pp. 90–113, Jun. 2023.

[3] A. Khan, A. D. Vibhute, S. Mali, and C. H. Patil, "A systematic review on hyperspectral imaging technology with a machine and deep learning methodology for agricultural applications," *Ecol. Informat.*, vol. 69, Jul. 2022, Art. no. 101678.

[4] M. Ebrahim Aghili, M. Imani, and H. Ghassemian, "Clustering based background learning for hyperspectral anomaly detection," *Egyptian J. Remote Sens. Space Sci.*, vol. 26, no. 3, pp. 477–489, Dec. 2023.

[5] S. Lorenz, P. Ghamisi, M. Kirsch, R. Jackisch, B. Rasti, and R. Gloaguen, "Feature extraction for hyperspectral mineral domain mapping: A test of conventional and innovative methods," *Remote Sens. Environ.*, vol. 252, Jan. 2021, Art. no. 112129.

[6] R. Hänsch and O. Hellwich, "Fusion of multispectral LiDAR, hyperspectral, and RGB data for urban land cover classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 2, pp. 366–370, Feb. 2021.

[7] S. Amini, S. Homayouni, A. Safari, and A. A. Darvishsefat, "Object-based classification of hyperspectral data using random forest algorithm," *Geo-Spatial Inf. Sci.*, vol. 21, no. 2, pp. 127–138, Jan. 2018.

[8] K. Huang, S. Li, X. Kang, and L. Fang, "Spectral–spatial hyperspectral image classification based on KNN," *Sens. Imag.*, vol. 17, no. 1, pp. 1–13, Dec. 2016.

[9] O. Okwuashi and C. E. Ndehedehe, "Deep support vector machine for hyperspectral image classification," *Pattern Recognit.*, vol. 103, Jul. 2020, Art. no. 107298.

[10] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.

[11] Y. Zhang, Y. Ma, X. Dai, H. Li, X. Mei, and J. Ma, "Locality-constrained sparse representation for hyperspectral image classification," *Inf. Sci.*, vol. 546, pp. 858–870, Feb. 2021.

[12] D. L. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," *AMS Math Challenges Lect.*, vol. 1, p. 32, Aug. 2000.

[13] A. Bordes, X. Glorot, J. Weston, and Y. Bengio, "Joint learning of words and meaning representations for open-text semantic parsing," in *Proc. 15th Int. Conf. Artif. Intell. Statist. (AISTATS)*, 2012, pp. 127–135.

[14] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[15] E. Merényi, W. H. Farrand, J. V. Taranik, and T. B. Minor, "Classification of hyperspectral imagery with neural networks: Comparison to conventional tools," *EURASIP J. Adv. Signal Process.*, vol. 2014, no. 1, pp. 1–19, Dec. 2014.

[16] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.

[17] Y. Chen, X. Zhao, and X. Jia, "Spectral–spatial classification of hyperspectral data based on deep belief network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2381–2392, Jun. 2015.

[18] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2017.

[19] F. Zhou, R. Hang, Q. Liu, and X. Yuan, "Hyperspectral image classification using spectral–spatial LSTMs," *Neurocomputing*, vol. 328, pp. 39–47, Feb. 2019.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[21] 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 Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.

[22] A. Ari, "Multipath feature fusion for hyperspectral image classification based on hybrid 3D/2D CNN and squeeze-excitation network," *Earth Sci. Informat.*, vol. 16, no. 1, pp. 175–191, Mar. 2023.

[23] N. Li and Z. Wang, "Hyperspectral image ship detection based upon two-channel convolutional neural network and transfer learning," in *Proc. IEEE 5th Int. Conf. Signal Image Process. (ICSIP)*, Oct. 2020, pp. 88–92.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[25] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[26] S. Sabour, N. Frosst, and G. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2017, pp. 3859–3869.

[27] Z. Zhong, J. Li, Z. Luo, and M. Chapman, "Spectral–spatial residual network for hyperspectral image classification: A 3-D deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 847–858, Feb. 2018.

[28] J. Ren, M. Shi, J. Chen, R. Wang, and X. Wang, "Hyperspectral image classification using multi-level features fusion capsule network with a dense structure," *Appl. Intell.*, vol. 53, no. 11, pp. 14162–14181, Jun. 2023.

[29] D. Kumar and D. Kumar, "A spectral–spatial 3D-convolutional capsule network for hyperspectral image classification with limited training samples," *Int. J. Inf. Technol.*, vol. 15, no. 1, pp. 379–391, Jan. 2023.

[30] J. Su, J. Zeng, D. Xiong, Y. Liu, M. Wang, and J. Xie, "A hierarchy-to-sequence attentional neural machine translation model," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 26, no. 3, pp. 623–632, Mar. 2018.

[31] Z. Lin, M. Feng, C. N. Santos, M. Yu, B. Xiang, B. Zhou, and Y. Bengio, "A structured self-attentive sentence embedding," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2017, pp. 1–15.

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

[33] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, vol. 2018, pp. 3–19.

[34] H. Sun, X. Zheng, X. Lu, and S. Wu, "Spectral–spatial attention network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3232–3245, May 2020.

[35] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[36] D. Hong, Z. Han, J. Yao, L. Gao, B. Zhang, A. Plaza, and J. Chanussot, "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5518615.

[37] H. Tulapurkar, B. Banerjee, and K. M. Buddhiraju, "Multi-head attention with CNN and wavelet for classification of hyperspectral image," *Neural Comput. Appl.*, vol. 35, no. 10, pp. 7595–7609, Apr. 2023.

[38] L. Zhao, L. Yi, X. Li, W. Hu, J. Wu, and G. Zhang, "Compact band weighting module based on attention-driven for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9540–9552, Nov. 2021.

[39] H. Gao, Y. Miao, X. Cao, and C. Li, "Densely connected multiscale attention network for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2563–2576, 2021.

[40] M. Zhu, L. Jiao, F. Liu, S. Yang, and J. Wang, "Residual spectral–spatial attention network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 449–462, Jan. 2021.

[41] N. Li and Z. Wang, "Spectral–spatial fused attention network for hyperspectral image classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2021, pp. 3832–3836.

[42] W. Guo, H. Ye, and F. Cao, "Feature-grouped network with spectral–spatial connected attention for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5500413.

[43] R. Hang, Z. Li, Q. Liu, P. Ghamisi, and S. S. Bhattacharyya, "Hyperspectral image classification with attention-aided CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2281–2293, Mar. 2021.

[44] N. Li and Z. Wang, "Spatial attention guided residual attention network for hyperspectral image classification," *IEEE Access*, vol. 10, pp. 9830–9847, 2022.

[45] N. Li, Z. Wang, F. A. Cheikh, and M. Ullah, "S^3AM: A spectral-similarity-based spatial attention module for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 5984–5998, 2022.

[46] Z. Ge, G. Cao, Y. Zhang, X. Li, H. Shi, and P. Fu, "Adaptive hash attention and lower triangular network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5509119.

[47] Z. Zhao, D. Hu, H. Wang, and X. Yu, "Center attention network for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3415–3425, 2021.

[48] X. Kang, C. Li, S. Li, and H. Lin, "Classification of hyperspectral images by Gabor filtering based deep network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1166–1178, Apr. 2018.

[49] H. Sun, X. Zheng, and X. Lu, "A supervised segmentation network for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 30, pp. 2810–2825, 2021.

[50] *Hyperspectral Remote Sensing Scenes—Grupo de Inteligencia Computacional (GIC)*. Accessed: Jun. 28, 2022. [Online]. Available: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sens ing_Scenes

[51] K. Karantzalos, C. Karakizi, Z. Kandylakis, and G. Antoniou, "HyRANK hyperspectral satellite dataset I. (Version V001) [data set]," Tech. Rep., Apr. 2018.

[52] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural network," in *Proc. 13th Int. Conf. Artif. Intell. Statist. (AISTATS)*, 2010, pp. 249–256.

[53] T. Tieleman and G. Hinton, "Lecture 6.5-RMSPROP: Divide the gradient by a running average of its recent magnitude," *COURSERA, Neural Netw. Mach. Learn.*, vol. 4, no. 2, pp. 26–31, 2012.

NINGYANG LI received the B.S. degree in remote sensing science and technology from the Faculty of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo, China, in 2019, and the M.S. degree in software engineering from the Faculty of Computer Science and Technology, Hainan University, Haikou, China, in 2023. From March 2022 to September 2022, he was an Exchange Student with the Department of Computer Science, Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology, Gjøvik, Norway. His research interests include hyperspectral image processing and analysis and deep learning.

ZHAOHUI WANG (Member, IEEE) received the M.S. degree in image processing from the University of Derby, U.K., in 2004, and the Ph.D. degree from the University of Leeds, U.K., in 2008. Then, he joined the Norwegian Colour and Visual Computing Laboratory, Gjøvik, Norway, to work on visual computing, multispectral color imaging research projects. He joined Hainan University, China, in 2013. He is currently a Professor in computer science with the Faculty of Computer Science and Technology. His current research interests include hyperspectral image processing and analysis, remote sensing image processing and its applications, computer vision, and deep learning. His professional memberships include IS&T, SPIE, IEEE, and CCF.

FAOUZI ALAYA CHEIKH (Senior Member, IEEE) received the B.Sc. degree in electronics from l'Ecole Nationale d'Ingenieurs de Tunis, in 1992, the M.Sc. degree in signal processing, in 1997, and the Ph.D. degree in information technology from the Tampere University of Technology (TUT), Tampere, Finland, in April 2004. Since 1994, he has been a Researcher with the Signal Processing Algorithm Group, Tampere University of Technology. Since 2006, he has also been with the Department of Computer Science and Media Technology, Gjøvik University College, Norway, as an Associate Professor. Since January 2016, he has also been with the Norwegian University of Science and Technology (NTNU). He teaches courses on image and video processing and analysis and media security. His research interests include e-learning, 3D imaging, image and video processing and analysis, video-based navigation, biometrics, pattern recognition, embedded systems, and content-based image retrieval. In these areas, he has published more than 150 peer-reviewed journals and conference articles and supervised four postdoctoral researchers, nine Ph.D. and a number of M.Sc. thesis projects. He is currently a co-supervisor of nine Ph.D. students. He has been involved in several European and national projects among them ESPRIT, NOBLESS, COST 211Quat, HyPerCept, IQ-Med, and H2020 ITN HiPerNav. He is a member of NOBIM and Forskerforbundet (The Norwegian Association of Researchers-NAR). He is on the editorial board of the *IET Image Processing* journal and the editorial board of the *International Journal of Advanced Robotics and Automation* and the technical committees of several international conferences. He is an expert reviewer to a number of scientific journals and conferences related to the field of his research.

MOHIB ULLAH (Member, IEEE) received the bachelor's degree in electronic and computer engineering from Politecnico di Torino, Italy, in 2012, the master's degree in telecommunication engineering from the University of Trento, Italy, in 2015, and the Ph.D. degree in computer science from the Norwegian University of Science and Technology (NTNU), Norway, in 2019. He is currently a Postdoctoral Research Fellow with NTNU, where he is involved in research, management, teaching, and industrial projects. His research interests include medical imaging, crowd analysis, object segmentation, behavior classification, and tracking. In these research areas, he has published several high-impact peer-reviewed journals, conferences, and workshop articles. He served as a Program Committee Member for the International Workshop on Computer Vision in Sports (CVsports). He also served as the Chair for the Technical Program at the European Workshop on Visual Information Processing. He is a Guest Editor of the *Applied Sciences* journal. He is a Reviewer of many conferences and journals *Neurocomputing* (Elsevier), *Neural Computing and Applications* (Elsevier), *Multimedia Tools and Applications* (Springer), IEEE Access, the *Journal of Imaging*, IEEE CVPRw, IEEE ICIP, and IEEE AVSS.

• • •