3-D Sharpened Cosine Similarity Operation for Hyperspectral Image Classification

Xin Qiao, *Graduate Student Member, IEEE*, Swalpa Kumar Roy, and Weimin Huang, *Senior Member, IEEE*

***Abstract*—Due to the advantage of high spectral resolution, hyperspectral imaging techniques have been extensively used in a variety of fields. Hyperspectral images (HSIs) classification is one of the fundamental tasks and attracts significant research interest. HSIs classification is pivotal as it facilitates precise identification of objects, providing invaluable insights for Earth observation tasks, such as resource management and land cover analysis. In existing studies, convolutional operations have been broadly applied for HSIs classification, especially 3-D convolution, which has shown its effectiveness in extracting spectral–spatial features from the raw HSIs. However, HSIs exhibit the characteristic of high dimensionality and pose challenges in extracting more discriminative features. In order to enhance the capability of capturing discriminative spectral–spatial features, in this article, a novel and effective 3-D sharpened cosine similarity (SCS) operation is proposed, serving as a replacement for conventional 3-D convolutional operation in HSIs classification and enhancing the classification accuracy. The 3-D SCS operation calculates and sharpens the cosine similarity between kernels and HSI input data. Based on the 3-D SCS operation, a 3-D SCS neural network is developed for HSIs classification tasks. To evaluate the effectiveness of 3-D SCS operation, experiments are conducted on three real-world HSIs datasets, including the University of Pavia, the University of Trento, and the University of Houston. Quantitative and qualitative experimental results illustrate that the SCS operation can effectively extract discriminative spectral-spatial features, achieving superior performance over the CNNs under the same model configuration.**

***Index Terms*—3-D convolutional operation, 3-D sharpened cosine similarity (SCS) operation, classification, hyperspectral images, neural network.**

I. Introduction

HYPERSPECTRAL images (HSIs) typically contain hundreds of narrow and continuous spectral bands located in the visible and near-infrared ranges [1], [2]. These spectral bands provide abundant spectral information and are beneficial for distinguishing objects with similar visual appearances. Therefore, HSIs have broad applications in various Earth Observation (EO) tasks, such as agricultural analysis, vegetation monitoring, and urban planning [3], [4], [5].

While the abundant spectral information in HSIs is good for accurate target identification, it also poses significant challenges due to high dimensionality. Besides, the interrelationship among these continuous spectral bands may lead to the curse of dimensionality [6], [7]. As a result, it is challenging but crucial to effectively capture spatial–spectral information from raw HSIs.

In earlier studies, band selection and band extraction have been commonly used to extract informative spectral information and solve the problem of high dimensionality [8], [9], [10], [11]. Band selection identifies a few informative bands from entire spectral bands to preserve the main spectral information. On the other hand, band extraction employs methods, such as principal component analysis [12] or independent component analysis (ICA) [13] to convert high-dimensional spectral data into a low-dimensional feature space. Moreover, traditional classifiers, such as linear regression, support vector machine (SVM), and K-nearest neighbors (KNN) [14] are used for classification. Chen et al. [15], proposed an ensemble SVM method for HSIs classification and achieved good classification performance by combining multiple SVM models. Bo et al. [16] designed a spectral–spatial KNN model based on spectral–spatial information and the KNN algorithm. Their method determined the category for a given test sample by calculating the distance between itself and the training samples.

However, the abovementioned traditional classifiers employing handcrafted features require prior knowledge and exhibit limited generalization capacity with different datasets. Over the past few years, deep learning-based feature extraction and classification methods have been extensively applied for HSIs classification tasks and have achieved remarkable results [17], [18], [19], [20]. In contrast to traditional manual feature extraction methods, deep learning models can automatically capture features from the original HSIs through hierarchical neural networks. According to the distinct architecture of the network, deep learning-based classification methods could be roughly divided into convolutional neural networks (CNNs) [21], [22], recurrent neural networks (RNNs) [23], [24], graph neural network [25], [26], and self-attention-based vision transformers (ViT) [27], [28], [29], [30].

CNNs are able to discriminate diverse objects through their inherent local feature extraction capability and are the first deep learning-based methods for HSIs classification. Hu et al. [31] employed a straightforward CNN framework with a convolutional layer, a max pooling layer, and a fully connected layer

Manuscript received 8 September 2023; revised 3 November 2023; accepted 23 November 2023. Date of publication 28 November 2023; date of current version 8 December 2023. The work was supported by the Natural Sciences and Engineering Research Council of Canada Discovery under Grant NSERC RGPIN-2017-04508. *(Corresponding author: Weimin Huang.)*

Xin Qiao and Weimin Huang are with the Department of Electrical and Computer Engineering, Memorial University of Newfoundland, St. John's A1B 3X5, Canada (e-mail: xqiao@mun.ca; weimin@mun.ca).

Swalpa Kumar Roy is with the Department of Computer Science and Engineering, Alipurduar Government Engineering and Management College, Alipurduar 736206, India (e-mail: swalpa@cse.jgec.ac.in).

Digital Object Identifier 10.1109/JSTARS.2023.3337112

to extract spectral features and classify HSIs in the spectral domain. Yue et al. [32] utilized 1-D CNN and 2-D CNN to extract spectral features and spatial representations, respectively. Subsequently, these features were concatenated and fed into a fully connected layer for classification. Given the correlation between neighboring pixels, lots of methods based on 3-D convolution have been developed in order to capture spectral–spatial features simultaneously [33], [34], [35]. Spectral–spatial classification models take hypercubes comprising the central pixel and its surrounding pixels as input and are able to capture spectral–spatial features, improving the classification performance. For example, Zhong et al. [36] designed a spectral–spatial residual network that used a stack of 3-D convolutions in residual architecture to explore discriminative features. The residual connection within the spectral and spatial residual blocks could effectively alleviate the challenge of decreasing accuracy associated with increasing depth of the network.

The aforementioned CNNs assume that each spectral band and each spatial location of the input data have equal importance for classification. In order to make the model focus on more discriminative features and alleviate the effect of inference pixels, the attention mechanism is integrated into CNNs and obtains improved classification performance [37], [38], [39], [40]. Sun et al. [41] exploited a spectral–spatial attention network (SSAN) to emphasize discriminative features and reduce the impact of interfering pixels. Li et al. [42] proposed a double-branch dual-attention network to classify HSIs. In their method, a model with two branches was constructed first. Then, spectral attention and spatial attention were embedded in the two branches, respectively, to capture important spectral and spatial features. In order to establish dependencies among different dimensions, Qiao et al. [43] presented a cross-dimensional residual network for HSIs classification. Hang et al. [44] designed an attention-aided CNN for HSI spectral–spatial feature extraction and classification. The attention modules were incorporated into the CNNs and emphasized more useful spectral bands and spatial positions.

In addition to CNNs, RNNs have also been employed for HSIs classification. RNNs regard the spectral information as sequential data and capture spectral features effectively. Mou et al. [45] took hyperspectral pixels as sequential input and extracted features using RNNs for the first time. A modified gated recurrent unit was utilized to process the input sequential data. Experimental results demonstrated the efficiency of the sequential-based RNNs for the HSIs classification task. Mei et al. [46] designed a bidirectional long-short-term memory network to explore the bidirectional spectral correlation of HSIs. They also introduced a spatial–spectral attention mechanism to the proposed model to focus on effective information.

Moreover, as a new emerging neural network architecture, self-attention-based transformers have also been increasingly explored in the HSIs remote sensing community [47], [48], [49]. ViT could address the limitation of local receptive fields in CNNs and establish global dependencies. Hong et al. [50] first employed a transformer architecture for HSIs classification and proposed a transformer-based backbone called SpectralFormer to classify HSIs from a sequential perspective without any feature preprocessing. Sun et al. [27] proposed a spectral–spatial feature tokenization transformer (SSFTT) to extract informative features. The low-level features were first captured using hybrid 3-D and 2-D convolutional layers. Later, the high-level semantic features were extracted via a transformer encoder module with a Gaussian-weighted feature tokenizer. Wang et al. [51], utilized transformers to capture regional and global spatial contexts. The former context was located within homogeneous areas and the latter encompassed relationships between different regions.

Among various network architectures discussed above, CNNs stand out as the most extensively used models. The convolutional operation calculates the dot product between the input data and kernel, serving as the cornerstone in most CNNs. However, the inherent characteristic being unbounded in the dot product increases the variance of the model and diminishes its generalization capability. To alleviate this issue, a 3-D sharpened cosine similarity (SCS) operation is proposed here and implemented as a replacement for the 3-D convolutional operation to capture spectral–spatial features from the original HSIs data. Different from the convolutional layer, which slides the kernel on the image and calculates the dot product between the kernel and image, the SCS operation computes the cosine similarity between the kernel and image. Moreover, the standard cosine similarity is further sharpened through exponentiation operation (i.e., raising its value to a power of p), while retaining its original sign. This exponentiation operation is implemented when calculating the output of the 3-D SCS layer. Based on 3-D SCS operation, a simple but effective 3-D sharpened cosine similarity neural network (SCS-NN) is constructed for HSIs classification. Our preliminary work [52] initially proved the effectiveness of the 3-D SCS operation. In this article, we further extend our research by conducting more extensive experiments, such as comparing the 3-D SCS operation with more different methods and additional two datasets. The main contributions of this article are as follows.

1) A 3-D SCS operation is proposed and employed for HSIs feature extraction and classification. The implemented 3-D SCS operation utilizes the value of cosine similarity rather than the dot product to capture spectral–spatial features. The extracted features are further amplified via exponentiation operation.
2) A simple but effective 3-D SCS-NN is built based on the 3-D SCS operation for HSIs classification. Moreover, it is proved that the 3-D SCS operation can easily replace the conventional 3-D convolutional operation in existing networks to further improve their classification performance.
3) In order to validate the performance of the 3-D SCS operation, a series of comparison experiments on three benchmark HSIs datasets are carried out and the results illustrate the effectiveness of the developed method for HSIs classification tasks.

The rest of this article is organized as follows. Section II introduces the 3-D SCS operation and the constructed 3-D SCS-NN. Section III describes the three datasets used for validation, the experiment settings, and results, and Section IV describes the corresponding discussion. Finally, Section V concludes this article.

II. Proposed Method

A. 3-D Sharpened Cosine Similarity

Convolutional operations are the foundation of most existing HSIs classification models. In HSIs classification tasks, it is essential to capture both spectral information from various spectral bands and spatial features. The spatial features refer to the patterns related to the spatial distribution of pixels within HSIs, such as the texture features and relationships between neighboring pixels. The 3-D operation takes the 3-D data cube as input and integrates both spectral and spatial information. As a result, 3-D convolutional operations are the most commonly used method to capture spectral–spatial features. The 3-D convolution is implemented by convolving the 3-D kernel and the input feature cube. Formally, the output value at the given position p_0 on the output feature cubes $\mathbf{y}$ could be expressed as follows:

$$\mathbf{y}(p_0) = \mathbf{w} \cdot \mathbf{x} = \sum_{p_n \in R} \mathbf{w}(p_n) \cdot \mathbf{x}(p_0 + p_n) \tag{1}$$

where $\mathbf{x}$ and $\mathbf{y}$ represent the input and output feature cubes of 3-D convolution, respectively. $\mathbf{w}$ is the 3-D kernel and R denotes the receptive field. The kernel is a filter (or array) that slides over hyperspectral input data to extract relevant features. The elements in the kernel are initialized randomly and updated during the training process. p_0 stands for the current position on both feature cubes $\mathbf{x}$ and $\mathbf{y}$. $p_n \in R$ enumerates all the positions within the receptive field. Take the receptive field of $(3 \times 3 \times 3)$ as an example, $p_n = \{(-1,-1,-1),(-1,-1,0),\ldots,(1,1,0),(1,1,1)\}$.

3-D convolutions obtain the output value via the dot product operation, as shown in (1). However, the dot product operation inherently suffers from the problem of being unbounded, which can potentially increase the variance of the model. To alleviate this issue, in this article, a novel 3-D SCS operation is proposed for the HSIs feature extraction and classification. Unlike 3-D convolution, 3-D SCS calculates the value of cosine similarity, which is bounded in the range of -1 to 1. Moreover, the cosine similarity values are further sharpened using an exponentiation operation (i.e., raising the values to a power of p), while reserving the original sign. This exponentiation operation enables a better discrimination between different features. Formally, the calculation of 3-D SCS is shown as follows:

$$\mathbf{v}(p_0) = \text{sign}(\mathbf{w} \cdot \mathbf{x}) \left(\frac{\mathbf{w} \cdot \mathbf{x}}{\max(\|\mathbf{w}\|, \epsilon) \cdot \max(\|\mathbf{x}\|, \epsilon)} \right)^p \tag{2}$$

where $\mathbf{v}$ is the output feature cube of 3-D SCS operation. Sign$(\mathbf{w} \cdot \mathbf{x})$ means keeping the original sign of the dot production operation, which is defined in (1). $\|\mathbf{w}\|$ and $\|\mathbf{x}\|$ denote the norm of $\mathbf{w}$ and $\mathbf{x}$, respectively. It should be noted that in order to avoid being divided by zero, the larger value between the corresponding norm and a small value ϵ will be kept. The cosine similarity output is additionally raised to a power of p, and this power value is dynamically updated during the training process. In our experiments, ϵ is set to 10^{-6}, and p is initialized as 2.

TABLE I
Hyperparameters of the Proposed 3-D SCS-NN Model

Blocks	Composition	Kernel size	Padding	Stride
Block 1	3-DSCS BN ReLU	24, (7, 3, 3)	(0, 0, 0)	(2, 1, 1)
Block 2	3-DSCS BN ReLU	32, (7, 3, 3)	(0, 0, 0)	(2, 1, 1)
Block 3	3-DSCS BN ReLU	32, (7, 3, 3)	(0, 0, 0)	(2, 1, 1)
Avg Pool	Average Pool3-D	(1, 2, 2)	(0, 0, 0)	(1, 1, 1)
FC	Linear1 Linear2	128 Num Classes	/	/

B. 3-D SCS-NN

To evaluate the effectiveness of the proposed 3-D SCS operation in extracting discriminative spectral–spatial features, a simple neural network called 3-D SCS-NN is developed based on the 3-D SCS operation for HSIs classification. Suppose the HSI dataset is denoted by $\mathbf{H} \in \mathbf{R}^{h \times w \times b}$, where h, w, and b represent height, width, and the number of spectral bands, respectively. Within the dataset, there are N labeled pixels represented as $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\} \in \mathbb{R}^{1 \times 1 \times b}$. $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N\} \in \mathbb{R}^{1 \times 1 \times c}$ stands for their corresponding labels and c is the number of land used and land cover categories. In order to fully utilize both spectral and spatial information, the small HSI patches comprising the central pixel and its neighboring pixels are extracted. The generated HSI patches are denoted as $\mathbf{Z} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N\} \in \mathbb{R}^{s \times s \times b}$, where s refers to the patch size. The label of the central pixel is regarded as the label for the whole patch.

Fig. 1 shows the overall architecture of the developed 3-D SCS-NN. The 3-D SCS-NN model is composed of three main blocks. The choice of a three-layer structure is based on the balance between model complexity and classification performance. Each block incorporates a 3-D SCS layer, followed by a 3-D batch normalization (BN) layer and a nonlinear ReLU layer. The hyperparameters, including the kernel size, padding, and stride are listed in Table I. The criteria for choosing these hyperparameters includes the tradeoff between model complexity and computation efficiency, spatial resolution, and spectral bands. For example, kernel size is related to the spatial distribution. Since a pixel and its neighboring pixels do not always belong to the same class, a small kernel size is used to incorporate features of the neighboring pixel while reducing interference. The stride on spectral dimension is set to 2 to reduce the number of spectral bands. All these three layers capture spectral and spatial features. The former includes wavelength variation and the latter refers to the spatial distribution. However, those extracted features also exhibit some differences. First, the features extracted by neural networks are hierarchical, meaning the first layer captures simple patterns and deeper layers capture more abstract high-level features. Second, the sizes of the captured features for each layer are also different. Take the University of Pavia (UP) dataset as an example, for the input patch with a size of (103,11,11), the sizes of output features for these three layers are (49,9,9), (22,7,7),

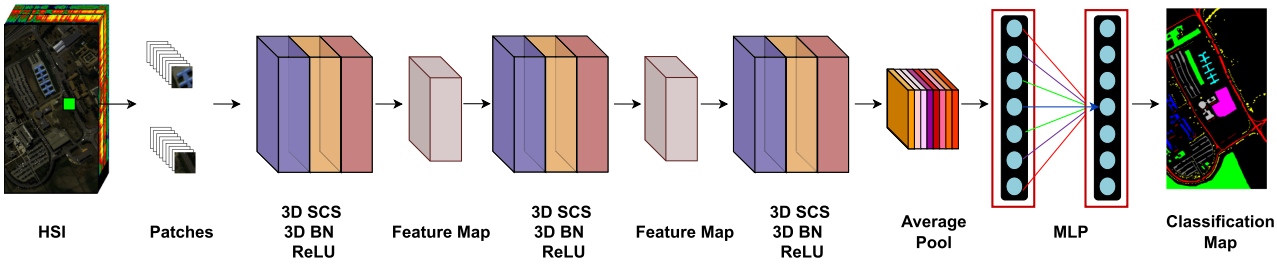


Fig. 1. Overall architecture of the constructed 3-D SCS-NN.

and (8,5,5), respectively. These processes could be defined as follows:

$$\mathbf{X}^{k+1} = \mathrm{SCS}(\mathbf{X}^k; \mathbf{W}^{k+1}, \mathbf{b}^{k+1}) \tag{3}$$

$$\mathrm{BN}(\mathbf{X}^{k+1}) = \frac{\mathbf{X}^{k+1} - \mathrm{E}(\mathbf{X}^{k+1})}{\mathrm{Var}(\mathbf{X}^{k+1})} \tag{4}$$

where SCS is the 3-D SCS layer based on the 3-D SCS operation. $\mathbf{X}^k$ and $\mathbf{X}^{k+1}$ are the input and output feature cubes of the $(k+1)$th layer. $\mathbf{W}^{k+1}$ and $\mathbf{b}^{k+1}$ denote the weights and the bias of the $(k+1)$th layer. $\mathrm{E}()$ and $\mathrm{Var}()$, respectively, represent the expectation and variance of the input feature cubes.

Following these three blocks, a 3-D average pooling layer is employed to fuse the extracted spectral–spatial features. A multilayer perceptron consisting of two linear layers is used to obtain the final classification result finally.

III. Experiments

In this section, the performance of the proposed 3-D SCS operation is evaluated on three benchmark HSI datasets. First, the selected datasets, including the UP, the University of Houston (UH), and the University of Trento (UT) are described. Then, experiment settings are introduced. Finally, the quantitative and qualitative experimental results are presented, along with a corresponding discussion.

A. Datasets

1) UP: The UP dataset was captured via a Reflective Optics System Imaging Spectrometer sensor over the UP located in Northern Italy. The scene consists of 610 × 340 pixels with a spatial resolution of 1.3 m. After removing 12 noisy bands from the original 115 bands, there are 103 spectral bands ranging from 430 to 860 nm. This dataset contains nine land cover classes in total. Fig. 2 displays the spatial distribution of both training and test data for each class.

2) UH: IEEE Geoscience and Remote Sensing Society (GRSS) released the UH dataset for the 2013 IEEE GRSS data fusion contest. Since then, the UH dataset has been one of the widely used benchmark datasets for HSI classification. The UH dataset was collected over the UH campus and its surrounding communities by an airborne sensor. There are 349 × 1905 pixels in spatial dimension and 144 bands in spectral dimension. The spatial resolution is 2.5 m and the spectral bands range from 380 to 1050 nm. 15 different classes are involved in the ground truth. The spatial distribution of training and test data are shown in Fig. 3.

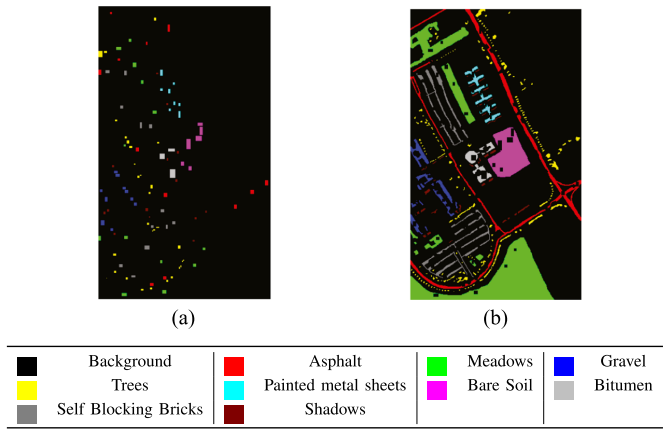


Fig. 2. UP dataset. (a) Training data. (b) Test data.

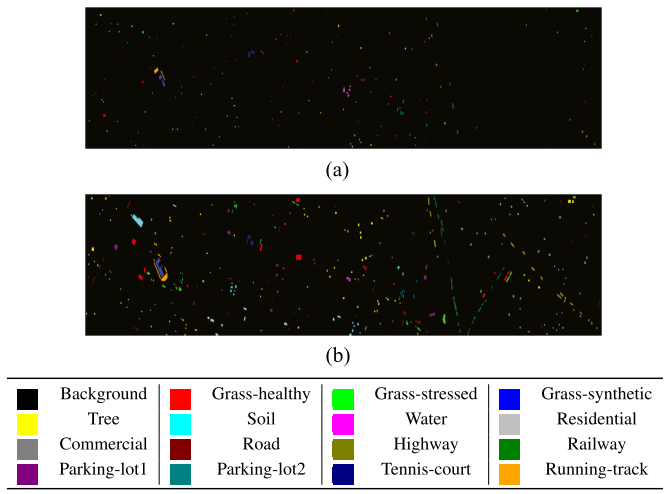


Fig. 3. UH dataset. (a) Training data. (b) Test data.

3) UT: The UT dataset was captured over the UT campus in Italy through the airborne imaging spectrometer for application Eagle sensor. Six classes are distributed in a spatial size of 600 × 166, along with a resolution of 1 m. The UT dataset includes 63 spectral bands over 402–989 nm. All the labeled pixels are disjointedly divided into training and test data, which are displayed in Fig. 4.

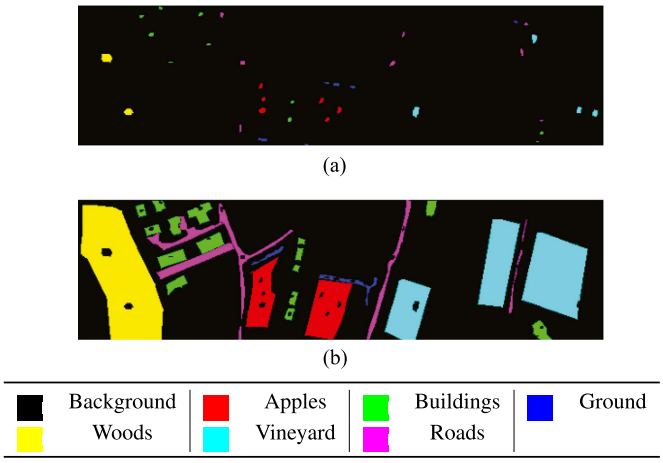


Fig. 4. UH dataset. (a) Training data. (b) Test data.

It should be noted that different from the random sampling method used in our preliminary work [52], in this article, the disjointed sampling method is adopted. In the disjointed sampling method, the training data are limited to certain small areas, which could effectively alleviate the problem of potential spatial overlapping between training and test samples caused by random sampling.

B. Experiments Setup

To validate the effectiveness of the 3-D SCS operation proposed in this article, a series of comparison experiments with various methods are conducted. First, since 3-D SCS is a fundamental operation aiming at enhancing the capability of capturing discriminative spectral–spatial features, comparative analysis with two other fundamental operations, the common vanilla convolution and the central difference convolution (CDC), are carried out. Vanilla convolution is the most widely used operation, which calculates the dot production between the kernel and the input data. CDC [53], on the other hand, calculates the difference between the central elements and their neighboring elements. For the vanilla convolution, the 3-D SCS-NN is compared with several established methods with similar complexity, including 2-D CNN, HybridSN, and 3-D CNN. For the CDC, a similar architecture called 3-D CDC-NN, which replaces the 3-D SCS operation with the 3-D CDC operation is constructed. The padding strategy is adopted in the 3-D CDC-NN model. Moreover, the performance of the 3-D SCS operation is compared with one of the state-of-the-art models, i.e., the SSFTT [27]. In the original SSFTT, 3-D convolutions are employed to extract low-level features, which are fed into the transformer encoder later. In the modified model, denoted as SCS-SSFTT, the 3-D convolution operations are replaced by the 3-D SCS operations. These comparative experiments provide a comprehensive assessment of the 3-D SCS operation.

The code is implemented in Python and all the networks are constructed using the PyTorch framework. All the comparison experiments are performed on the Digital Research Alliance of Canada Cedar cluster equipped with a GPU V100 with 32 GB. The processor model is Intel Silver 4216 Cascade Lake with a frequency of 2.1 GHz and the operating system is Linux. In total, 64 GB of RAM is allocated during the training. The whole labeled datasets are disjointedly separated into training and test data, as illustrated in Figs. 2–4. In addition, during the training process, in order to select the best model, a subset of the training data is chosen as the validation data. Specifically, for the whole training data, 90% are used to optimize and update the parameters in the model, and the other 10% are utilized for validation. It should be noted that in the experiments of evaluating the performance with smaller sizes of training data, the portions of training samples are reduced to 3%, 5%, 7%, and 9%, and the validation data percentages are correspondingly adjusted to 97%, 95%, 93%, and 91%. In this study, the cross-entropy is employed as the loss function. The training process is configured to run 500 epochs and an early stopping strategy is adopted. The training process would be stopped if the loss remains unreduced for 50 consecutive epochs for the validation data. This strategy can effectively avoid potential overfitting. Various classification metrics including classwise accuracies, overall accuracy (OA), average accuracy (AA), and the Kappa coefficient (Kappa) are calculated. To ensure reliability, each method is tested five times. The accuracies are presented as the mean values along with their corresponding standard deviations.

Moreover, HSIs exhibit the characteristic of high dimensionality, which requires preprocessing in order to stabilize the training process. In this study, the max–min standardization is applied to the spectral dimension, normalizing it to the range of 0–1.

C. Quantitative Comparison

1) Classification Accuracies: Tables II–IV list the classification accuracies of different methods on the three selected datasets. All results are shown in a format of mean ± standard deviation of five repetitions, with the best results highlighted in bold.

The classification results on the UP dataset are given in Table II. Compared with the 3-D CNN and 3-D CDC-NN, the constructed 3-D SCS-NN achieves superior results with an OA of 88.41%, AA of 88.26%, and Kappa of 84.53%. The OA is 0.93% higher than that of 3-D CNN, and 8.94% higher than that of 3-D CDC-NN. Besides, SCS-SSFTT exhibits a higher AA value and competitive OA and Kappa values compared with the original SSFTT model. These results illustrate the effectiveness of the proposed 3-D SCS operation in capturing discriminative features for the HSI classification task.

For the UH dataset, 2-D CNN obtains a relatively lower accuracy, as presented in Table III. This could be attributed to the fact that 2-D CNN only considers the spatial information while neglecting the spectral features, which are important for HSI classification. The lower spatial resolution of the UH dataset also leads to a worse performance for 2-D CNN. Besides, under a similar model complexity, 3-D SCS-NN outperforms both 3-D CNN and 3-D CDC-NN in terms of OA, AA, and Kappa. Specifically, the OA, AA, and kappa values are improved by 1.54%, 1.26%, and 1.58% compared with 3-D CNN, and by 2.16%, 2.93%,

TABLE II
CLASSIFICATION RESULTS ON UP DATASET

	2-D CNN	3-D CNN	HybridSN	3-D CDC-NN	3-D SCS-NN	SSFTT	SCS-SSFTT
1	82.41 ± 1.53	83.03 ± 1.55	76.65 ± 1.22	83.70 ± 2.11	89.05 ± 1.73	84.74 ± 1.59	**90.49 ± 1.67**
2	87.70 ± 2.10	91.99 ± 1.26	84.15 ± 2.52	69.67 ± 2.29	90.03 ± 2.88	**94.57 ± 1.44**	92.76 ± 1.88
3	58.56 ± 2.66	53.85 ± 2.45	64.99 ± 3.84	54.09 ± 1.78	64.10 ± 3.77	70.34 ± 2.73	**81.39 ± 5.55**
4	97.16 ± 0.32	98.19 ± 0.63	97.16 ± 0.46	83.54 ± 6.09	98.29 ± 0.68	97.42 ± 0.96	**98.90 ± 0.59**
5	99.26 ± 0.49	99.08 ± 0.36	98.99 ± 1.08	99.32 ± 0.46	**99.50 ± 0.29**	99.10 ± 0.64	99.03 ± 0.41
6	53.23 ± 7.60	71.56 ± 5.13	52.70 ± 3.98	**99.55 ± 0.43**	74.75 ± 6.14	88.77 ± 2.11	78.87 ± 2.90
7	64.61 ± 6.11	80.18 ± 2.80	79.71 ± 2.11	78.78 ± 1.79	84.63 ± 2.01	85.65 ± 7.07	**88.91 ± 3.98**
8	92.29 ± 2.01	**98.12 ± 0.32**	96.37 ± 0.79	97.15 ± 0.54	97.00 ± 0.69	97.21 ± 0.79	94.80 ± 1.16
9	92.20 ± 0.99	96.70 ± 0.49	92.33 ± 2.58	95.14 ± 0.72	96.96 ± 0.65	**97.46 ± 0.53**	95.55 ± 1.00
OA	82.53 ± 0.68	87.48 ± 0.87	80.94 ± 1.31	79.47 ± 1.37	88.41 ± 0.65	**91.65 ± 0.66**	91.05 ± 0.48
AA	80.83 ± 1.39	85.86 ± 0.82	82.56 ± 0.94	84.55 ± 0.81	88.26 ± 0.63	90.58 ± 0.76	**91.19 ± 0.74**
Kappa	76.51 ± 0.93	83.30 ± 1.14	74.71 ± 1.62	73.94 ± 1.63	84.53 ± 0.71	**88.81 ± 0.86**	87.98 ± 0.60

TABLE III
CLASSIFICATION RESULTS ON UH DATASET

	2D CNN	3D CNN	HybridSN	3D CDC-NN	3D SCS-NN	SSFTT	SCS-SSFTT
1	82.34 ± 0.44	82.91 ± 0.13	82.32 ± 0.23	81.50 ± 0.38	82.32 ± 1.51	82.60 ± 0.40	**86.86 ± 3.24**
2	82.82 ± 0.43	83.44 ± 0.14	83.33 ± 0.90	83.83 ± 0.91	83.38 ± 0.30	84.47 ± 0.62	**84.76 ± 0.38**
3	47.72 ± 2.44	72.20 ± 5.84	80.00 ± 2.05	67.17 ± 3.79	72.40 ± 3.63	**87.60 ± 7.68**	86.73 ± 3.55
4	93.86 ± 2.70	98.48 ± 1.92	87.46 ± 1.54	88.92 ± 1.92	98.03 ± 1.61	96.69 ± 3.41	**96.97 ± 1.40**
5	96.55 ± 0.87	99.15 ± 0.69	99.02 ± 0.45	97.58 ± 0.57	99.00 ± 0.78	**99.89 ± 0.07**	97.08 ± 1.55
6	83.50 ± 3.90	92.59 ± 3.55	91.33 ± 1.05	90.91 ± 3.96	93.71 ± 3.65	87.41 ± 4.92	**96.92 ± 1.80**
7	85.07 ± 1.71	86.75 ± 1.33	84.07 ± 1.11	86.49 ± 1.43	**92.74 ± 1.05**	83.68 ± 2.76	92.44 ± 1.16
8	49.23 ± 4.89	67.73 ± 1.12	66.02 ± 6.98	60.68 ± 2.13	69.86 ± 5.37	**70.75 ± 3.19**	66.42 ± 3.63
9	63.61 ± 4.46	**85.48 ± 0.67**	74.49 ± 2.89	81.59 ± 2.99	81.44 ± 2.37	84.76 ± 1.66	84.51 ± 1.70
10	54.96 ± 3.93	53.76 ± 1.65	44.96 ± 0.49	56.31 ± 1.59	57.93 ± 3.18	51.24 ± 3.50	**79.92 ± 6.68**
11	51.14 ± 2.44	61.27 ± 1.89	68.56 ± 1.52	71.35 ± 5.20	65.62 ± 3.48	77.86 ± 2.52	**84.84 ± 4.10**
12	64.36 ± 7.77	77.81 ± 2.91	83.19 ± 4.68	**91.74 ± 0.51**	85.19 ± 2.62	86.92 ± 2.04	85.28 ± 3.50
13	68.42 ± 3.47	91.23 ± 1.46	68.63 ± 9.37	90.67 ± 0.98	94.74 ± 1.22	89.96 ± 1.43	**93.19 ± 1.98**
14	80.24 ± 5.39	92.23 ± 1.74	96.76 ± 3.22	82.91 ± 3.38	91.98 ± 2.16	97.65 ± 4.10	**100.0 ± 0.00**
15	77.34 ± 5.29	84.61 ± 7.63	81.95 ± 6.86	73.02 ± 1.84	80.30 ± 10.0	**88.16 ± 9.62**	58.44 ± 4.96
OA	71.83 ± 1.06	80.30 ± 0.55	78.04 ± 1.62	79.68 ± 1.19	81.84 ± 0.91	82.99 ± 1.02	**85.48 ± 0.86**
AA	72.08 ± 1.25	81.98 ± 0.91	79.47 ± 2.00	80.31 ± 1.15	83.24 ± 1.05	84.64 ± 1.08	**86.29 ± 0.76**
Kappa	69.48 ± 1.13	78.70 ± 0.59	76.61 ± 1.72	78.02 ± 1.29	80.28 ± 0.99	81.61 ± 1.10	**84.24 ± 0.93**

TABLE IV
CLASSIFICATION RESULTS ON UT DATASET

	2-D CNN	3-D CNN	HybridSN	3-D CDC-NN	3-D SCS-NN	SSFTT	SCS-SSFTT
1	96.32 ± 0.60	99.74 ± 0.26	96.95 ± 1.12	**99.82 ± 0.14**	99.58 ± 0.15	98.54 ± 0.44	98.62 ± 0.26
2	89.68 ± 2.17	87.21 ± 1.21	85.65 ± 5.32	**87.71 ± 0.86**	81.92 ± 3.84	87.38 ± 3.35	82.15 ± 5.08
3	**98.77 ± 1.06**	89.95 ± 4.77	91.82 ± 4.44	80.00 ± 2.65	87.17 ± 1.94	77.22 ± 5.24	76.68 ± 4.98
4	95.60 ± 0.60	97.24 ± 1.20	96.10 ± 2.44	94.59 ± 0.56	98.66 ± 0.56	95.47 ± 1.83	**98.88 ± 1.08**
5	91.13 ± 2.77	99.50 ± 0.08	98.08 ± 1.76	98.90 ± 0.77	99.08 ± 0.25	99.81 ± 0.15	**99.91 ± 0.02**
6	67.45 ± 7.55	76.08 ± 4.01	70.68 ± 10.3	84.69 ± 3.21	82.79 ± 4.69	87.37 ± 5.13	**87.86 ± 3.43**
OA	90.68 ± 0.95	95.13 ± 0.64	93.23 ± 1.87	94.93 ± 0.28	95.56 ± 0.17	95.56 ± 0.57	**96.20 ± 0.63**
AA	89.82 ± 0.88	91.62 ± 0.85	89.88 ± 3.06	90.95 ± 0.36	91.53 ± 0.56	**90.96 ± 1.17**	90.68 ± 1.21
Kappa	87.54 ± 1.26	93.48 ± 0.85	90.94 ± 2.51	93.24 ± 0.38	94.03 ± 0.23	94.06 ± 0.76	**94.91 ± 0.84**

and 2.26% compared with 3-D CDC-NN. Similar results could also be observed for SCS-SSFTT, which obtains the highest OA (85.48%), AA (86.29%), and Kappa (84.24%), outperforming SSFTT by 2.49%, 1.65%, and 2.63%, respectively. Moreover, SCS-SSFTT exhibits the best classification accuracies in most categories, even achieving 100% for class 14, i.e., tennis court. These enhanced classification accuracies illustrate that the SCS operation could effectively extract spectral–spatial features and boost the classification performance when replacing the convolution operation.

In terms of the UT dataset, 3-D SCS-NN attains a better OA score than 3-D CNN and 3-D CDC-NN, as reported in Table IV. Furthermore, 3-D SCS-SSFTT also surpasses SSFTT in both OA and the majority classwise accuracies, demonstrating the superior performance of the SCS operation.

2) Computational Cost: In addition to quantitative classification accuracies, quantitative computational cost comparisons are also investigated, as displayed in Table V. Since the early stop strategy is adopted during the training process, the final training epochs are different. To make a fair comparison, following [54], the average training time per epoch is reported. All these values are obtained on a computer with an NVIDIA RTX 3070 GPU. It should be noted that 2-D CNN and 3-D CNN have different numbers of convolution kernels and kernel sizes. As a result, 2-D CNN has more parameters than 3-D CNN on the UP and UT datasets. For the proposed 3-D SCS operation, the introduction of the exponentiation operation results in more parameters compared with the common 3-D convolution operation. The number of increased parameters is decided by the number of 3-D kernels. Although the 3-D SCS operation results in additional parameters

TABLE V
COMPARISON OF PARAMETERS AND INFERENCE TIME FOR DIFFERENT METHODS

Dataset	Computational Cost	Methods						
		2-D CNN	3-D CNN	HybridSN	3-D CDC-NN	3-D SCS-NN	SSFTT	SCS-SSFTT
UP	Number of parameters	264,909	247,033	1,879,289	1,447,073	247,121	489,153	489,225
	Training time per epoch (s)	0.23	0.62	1.07	1.83	1.08	0.41	0.62
	Inference time (s)	5.47	5.98	6.22	7.14	6.61	5.90	6.52
UH	Number of parameters	316,765	329,727	2,635,775	1,959,847	329,815	678,471	678,543
	Training time per epoch (s)	0.19	0.63	1.07	1.79	1.09	0.35	0.56
	Inference time (s)	1.64	2.07	2.28	2.87	2.54	2.00	2.24
UT	Number of parameters	214,606	164,726	1,141,622	934,686	164,814	304,638	304,710
	Training time per epoch (s)	0.04	0.09	0.14	0.24	0.15	0.07	0.11
	Inference time (s)	1.99	2.30	2.36	3.17	2.75	2.24	2.44

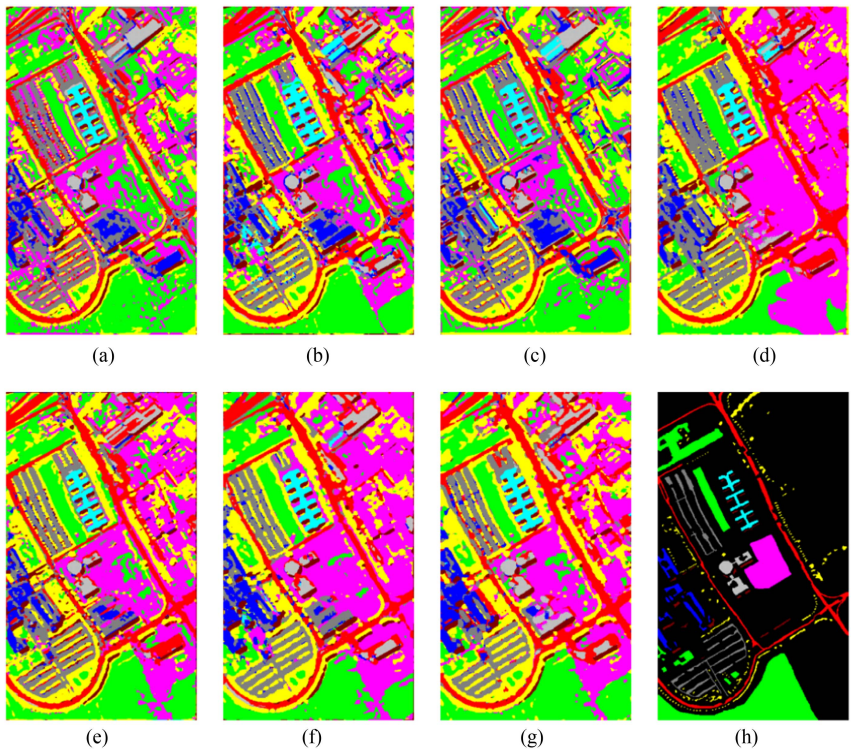


Fig. 5. Classification maps for the UP dataset generated by various methods. (a) 2-D CNN. (b) 3-D CNN. (c) HybridSN. (d) 3-D CDC-NN. (e) 3-D SCS-NN. (f) SSFTT. (g) SCS-SSFTT. (h) Ground Truth.

and increases both training and inference time, this tradeoff is acceptable given the improved accuracies it produces.

D. Visualization Comparison

To qualitatively compare the classification performance, the pixel-level classification maps obtained by different methods, i.e., 2-D CNN, 3-D CNN, HybridSN, 3-D CDC-NN, 3-D SCS-NN, SSFTT, and SCS-SSFTT on the UP, UH, and UT datasets are depicted in Figs. 5–7, respectively. These qualitative classification maps are in line with the quantitative accuracies in Table II–IV. The maps generated by 2-D CNN, take Fig. 6(a) as an example, exhibit more misclassified pixels compared with those based on spectral–spatial features. This is due to that 2-D CNN misses the spectral information and only considers the spatial features. In addition, the input patches could incorporate pixels belonging to different categories, further challenging the spatial-based classification methods. Compared with the 3-D convolution-based method, i.e., 3-D CNN and SSFTT, the 3-D SCS operation-based methods yield better classification maps, especially for the region highlighted by the red box in Fig. 6(g), demonstrating its effectiveness in extracting discriminative spectral–spatial features and enhancing HSI classification performance.

Figs. 8–10 show the distribution of the spectral–spatial features extracted by different methods using T-distributed stochastic neighbor embedding (t-SNE) [55]. For 3-D SCS-NN, although there is a certain degree of mixing, such as class 2 (buildings) and class 6 (road) in the UT dataset, it can be observed that features extracted from the same category are clustered together, while those generated from different categories are separated from each other. Moreover, SCS-SSFTT also generates a more tightly distribution, e.g., class 1 (apples) in Fig. 10(g). Therefore, it can be inferred that the SCS operation can capture distinctive features from different categories.

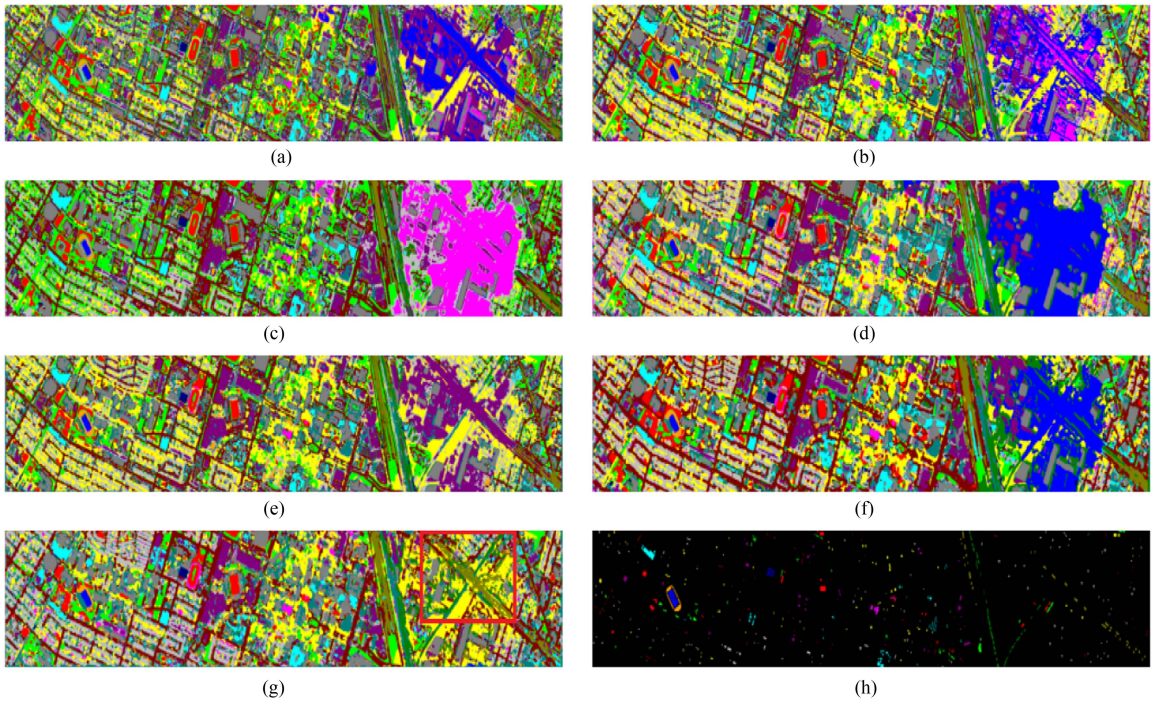


Fig. 6. Classification maps for the UH dataset generated by various methods. (a) 2-D CNN. (b) 3-D CNN. (c) HybridSN. (d) 3-D CDC-NN. (e) 3-D SCS-NN. (f) SSFTT. (g) SCS-SSFTT. (h) Ground Truth.

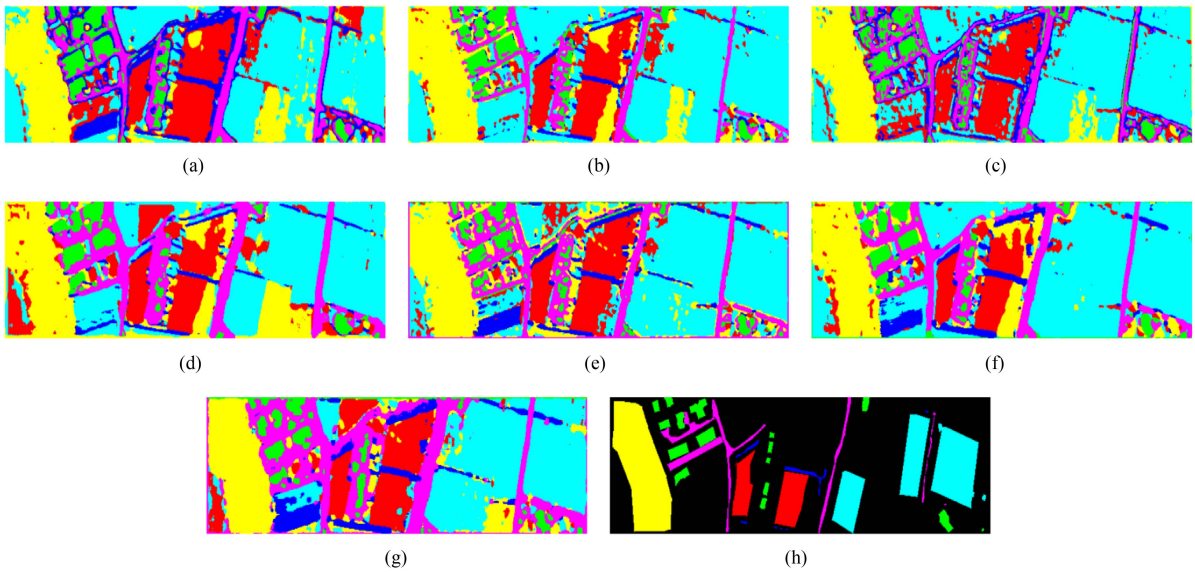


Fig. 7. Classification maps for the UT dataset generated by various methods. (a) 2-D CNN. (b) 3-D CNN. (c) HybridSN. (d) 3-D CDC-NN. (e) 3-D SCS-NN. (f) SSFTT. (g) SCS-SSFTT. (h) Ground truth.

E. Robustness Evaluation

In order to assess the robustness of the proposed operation, experiments using limited training samples are conducted in this section. Smaller training data are randomly selected from the original training data. Specifically, 3%, 5%, 7%, and 9% of the original training data are selected for training, while corresponding 97%, 95%, 93%, and 91% are used for validation during the training process. The test data remain unchanged. Methods with similar complexities are also tested for comparison, including 2-D CNN, 3-D CNN, HybirdSN, 3-D CDC-NN, and 3-D SCS-NN. The OA scores on different datasets are shown in Fig 11. It is observed that as the training data size decreases, the OA of various methods also decreases and 2-D CNN exhibits the biggest decline. The developed 3-D SCS-NN presents the best OA on the UP and UT datasets with limited training samples, demonstrating the superior generalization capability of the 3-D SCS operation. In terms of the UH dataset, 3-D SCS-NN obtains

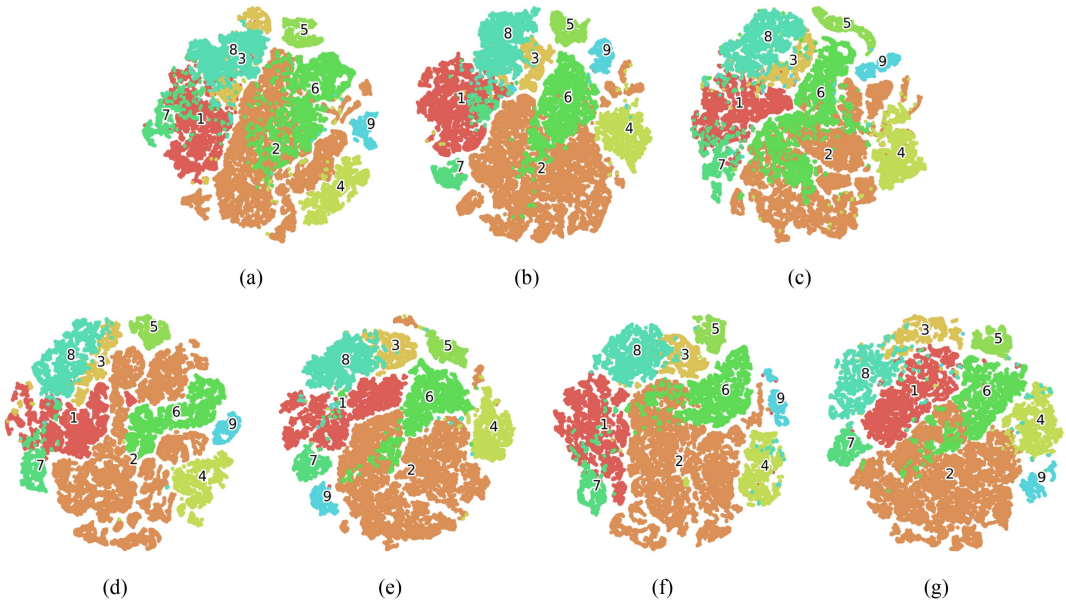


Fig. 8. T-SNE visualization for the UP datasets. (a) 2-D CNN. (b) 3-D CNN. (c) HybridSN. (d) 3-D CDC-NN. (e) 3-D SCS-NN. (f) SSFTT. (g) SCS-SSFTT.

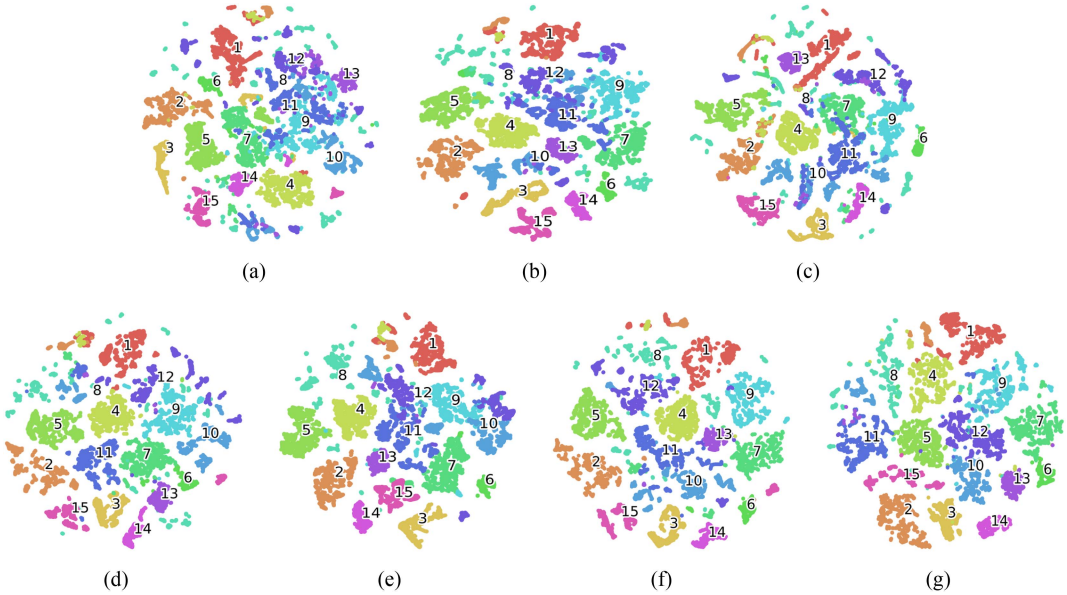


Fig. 9. T-SNE visualization for the UH datasets. (a) 2-D CNN. (b) 3-D CNN. (c) HybridSN. (d) 3-D CDC-NN. (e) 3-D SCS-NN. (f) SSFTT. (g) SCS-SSFTT.

the second-best performance. This could be attributed to the higher complexity of the UH dataset, which has more classes compared with the UP and UT datasets.

IV. Discussion

3-D convolution is the most commonly used operation in existing HSIs classification models. This article introduces a novel 3-D SCS operation as an alternative to 3-D convolution in the field of HSIs classification. In order to evaluate its effectiveness and superiority, a series of experiments have been conducted. First of all, a simple but effective model named 3-D SCS-NN is constructed based on the designed 3-D SCS operation. The qualitative and quantitative experimental results demonstrate that 3-D SCS-NN could achieve higher classification accuracies than 3-D CNN and 3-D CDC-NN under similar model architectures.

Second, the performance of the 3-D SCS operation is further assessed through SSFTT, a transformer-based state-of-the-art model. The experiments illustrate that when the 3-D convolution is replaced with the 3-D SCS operation in SSFTT for feature extraction, the performance could be improved. Besides, considering the lack of sufficient labeled data is a common challenge in practical applications, experiments with small sizes of training data are conducted. The proposed 3-D SCS operation again obtains superior results, illustrating its robustness under small

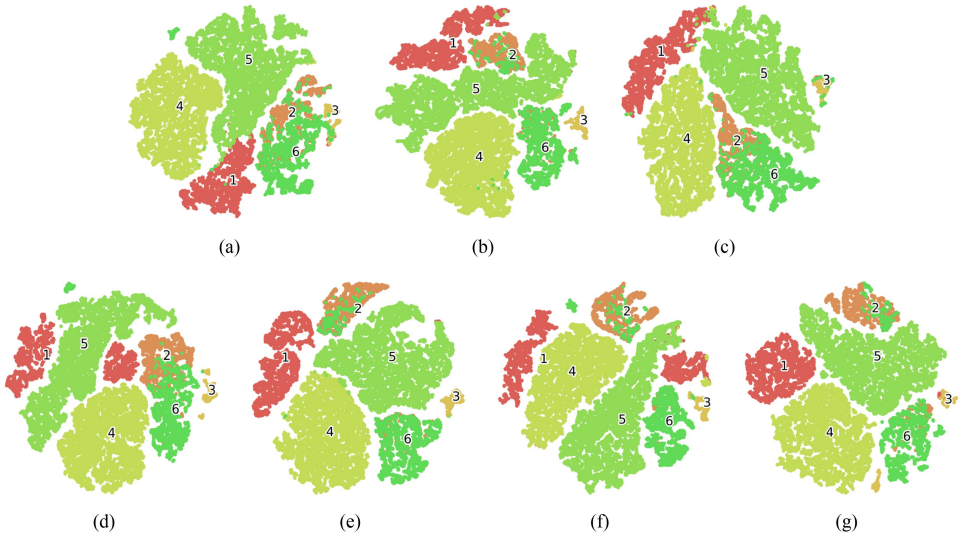


Fig. 10. T-SNE visualization for the UT datasets. (a) 2-D CNN. (b) 3-D CNN. (c) HybridSN. (d) 3-D CDC-NN. (e) 3-D SCS-NN. (f) SSFTT. (g) SCS-SSFTT.

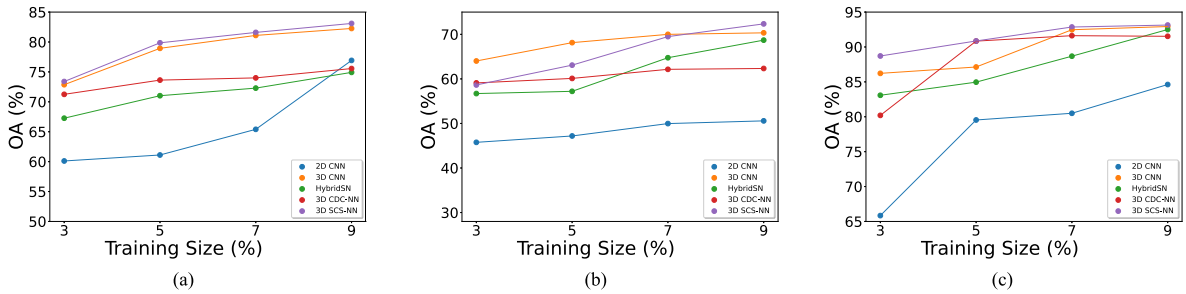


Fig. 11. Classification accuracies under limited training samples. (a) UP. (b) UH. (c) UT.

training samples. However, it should also be noted that since the 3-D SCS operation calculates the cosine similarity value and introduces additional parameters, the computation cost of the 3-D SCS is higher than that of conventional the 3-D convolution operation.

V. Conclusion

HSIs classification plays a crucial role in object identification. In this article, a novel and effective 3-D SCS operation is proposed to capture discriminative spectral features and those related to spatial distribution of pixels within HSIs. The value of cosine similarity rather than the dot product between the input and the kernels is calculated and forwarded in the network. Besides, in order to better distinguish different features, the value of cosine similarity is further sharpened through an exponentiation operation while keeping the original sign unchanged. Based on the implemented 3-D SCS operation, a 3-D SCS-NN model is developed for HSIs classification and obtains better classification performance compared with a convolution-based 3-D CNN model with the same structure. Specifically, 3-D SCS-NN achieves OAs of 88.41%, 81.84%, and 95.56% on the UP, UH, and UT datasets, improved by 0.93%, 1.54%, and 0.43%, respectively, compared with 3-D CNN. Moreover, the 3-D SCS operation can be used as a replacement for the 3-D convolution operation and experiments prove that the classification performance could be further enhanced with the replacement. The 3-D SCS operation also presents good robustness under limited training samples with relatively simple datasets.

Compared with 3-D convolution, the 3-D SCS operation also exhibits some limitations, primarily in terms of computational cost. Since cosine similarity involves more computation compared with dot production and exponentiation introduces additional parameters, models based on the 3-D SCS operation typically require more computational resources, which could be justified given the improved classification performance.

Acknowledgment

The author would like to thank the Digital Research Alliance of Canada for providing the computational resources.

References

[1] A. Plaza et al., "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, Sep. 2009.

[2] X. Qiao and W. Huang, "A dual frequency transformer network for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 10344–10358, Nov. 2023, doi: 10.1109/JSTARS.2023.3328115.

[3] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[4] M. Ahmad et al., "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, Jan. 2022.

[5] X. Tao et al., "A new deep convolutional network for effective hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6999–7012, 2022.

[6] B. Kumar, O. Dikshit, A. Gupta, and M. K. Singh, "Feature extraction for hyperspectral image classification: A review," *Int. J. Remote Sens.*, vol. 41, no. 16, pp. 6248–6287, 2020.

[7] B. Rasti et al., "Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, Dec. 2020.

[8] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 118–139, Jun. 2019.

[9] H. Yang, Q. Du, H. Su, and Y. Sheng, "An efficient method for supervised hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 138–142, Jan. 2011.

[10] N. Audebert, B. Le Saux, and S. Lefèvre, "Deep learning for classification of hyperspectral data: A comparative review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 159–173, Jun. 2019.

[11] S. K. Roy, S. Das, T. Song, and B. Chanda, "DARecNet-BS: Unsupervised dual-attention reconstruction network for hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 12, pp. 2152–2156, Dec. 2021.

[12] M. P. Uddin, M. A. Mamun, and M. A. Hossain, "PCA-based feature reduction for hyperspectral remote sensing image classification," *IETE Tech. Rev.*, vol. 38, no. 4, pp. 377–396, 2021.

[13] A. Villa, J. Chanussot, C. Jutten, J. A. Benediktsson, and S. Moussaoui, "On the use of ICA for hyperspectral image analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2009, pp. IV-97–IV-100.

[14] Y. Guo, S. Han, Y. Li, C. Zhang, and Y. Bai, "K-nearest neighbor combined with guided filter for hyperspectral image classification," *Proc. Comput. Sci.*, vol. 129, pp. 159–165, 2018.

[15] Y. Chen, X. Zhao, and Z. Lin, "Optimizing subspace SVM ensemble for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1295–1305, Apr. 2014.

[16] C. Bo, H. Lu, and D. Wang, "Spectral-spatial K-nearest neighbor approach for hyperspectral image classification," *Multimedia Tools Appl.*, vol. 77, pp. 10419–10436, 2018.

[17] 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.

[18] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.

[19] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS J. Photogrammetry Remote Sens.*, vol. 158, pp. 279–317, Dec. 2019.

[20] C. Zhang, G. Li, and S. Du, "Multi-scale dense networks for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 9201–9222, Nov. 2019.

[21] B. Liu, X. Yu, P. Zhang, A. Yu, Q. Fu, and X. Wei, "Supervised deep feature extraction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 1909–1921, Apr. 2018.

[22] M. Zhao, S. Shi, J. Chen, and N. Dobigeon, "A 3-D-CNN framework for hyperspectral unmixing with spectral variability," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2022, Art. no. 5521914.

[23] W. Hu, H. Li, L. Pan, W. Li, R. Tao, and Q. Du, "Spatial–spectral feature extraction via deep convLSTM neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4237–4250, Jun. 2020.

[24] X. Zhang, Y. Sun, K. Jiang, C. Li, L. Jiao, and H. Zhou, "Spatial sequential recurrent neural network for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4141–4155, Nov. 2018.

[25] Y. Ding et al., "Multi-feature fusion: Graph neural network and CNN combining for hyperspectral image classification," *Neurocomputing*, vol. 501, pp. 246–257, 2022.

[26] G. Li and M. Ye, "Spatial-spectral hyperspectral classification based on learnable 3D group convolution," 2023, *arXiv:2307.07720*.

[27] L. Sun, G. Zhao, Y. Zheng, and Z. Wu, "Spectral–spatial feature tokenization transformer for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Mar. 2022, Art. no. 5522214.

[28] X. Yang, W. Cao, Y. Lu, and Y. Zhou, "Hyperspectral image transformer classification networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, May 2022, Art. no. 5528715.

[29] Z. Han, D. Hong, L. Gao, J. Yao, B. Zhang, and J. Chanussot, "Multimodal hyperspectral unmixing: Insights from attention networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Apr. 2022, Art. no. 5524913.

[30] S. K. Roy, A. Deria, D. Hong, B. Rasti, A. Plaza, and J. Chanussot, "Multimodal fusion transformer for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Jul. 2023, Art. no. 5515620.

[31] W. Hu, Y. Huang, L. Wei, F. Zhang, and H. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, pp. 1–2, 2015.

[32] J. Yue, W. Zhao, S. Mao, and H. Liu, "Spectral–spatial classification of hyperspectral images using deep convolutional neural networks," *Remote Sens. Lett.*, vol. 6, no. 6, pp. 468–477, May 2015.

[33] Y. Li, H. Zhang, and Q. Shen, "Spectral–spatial classification of hyperspectral imagery with 3 D convolutional neural network," *Remote Sens.*, vol. 9, no. 1, 2017, Art. no. 67.

[34] A. B. Hamida, A. Benoit, P. Lambert, and C. B. Amar, "3-D deep learning approach for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4420–4434, Aug. 2018.

[35] 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.

[36] 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.

[37] G. Li and C. Zhang, "Faster hyperspectral image classification based on selective kernel mechanism using deep convolutional networks," 2022, *arXiv:2202.06458*.

[38] Z. Xue, M. Zhang, Y. Liu, and P. Du, "Attention-based second-order pooling network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9600–9615, Nov. 2021.

[39] 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.

[40] M. E. Paoletti et al., "Parameter-free attention network for spectral-spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5516817.

[41] 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.

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

[43] X. Qiao, S. K. Roy, and W. Huang, "Rotation is all you need: Cross dimensional residual interaction for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 5387–5404, Jun. 2023.

[44] 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.

[45] 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.

[46] S. Mei, X. Li, X. Liu, H. Cai, and Q. Du, "Hyperspectral image classification using attention-based bidirectional long short-term memory network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5509612.

[47] X. He, Y. Chen, and Z. Lin, "Spatial-spectral transformer for hyperspectral image classification," *Remote Sens.*, vol. 13, no. 3, Jan. 2021, Art. no. 498.

[48] X. Qiao, S. K. Roy, and W. Huang, "Multiscale neighborhood attention transformer with optimized spatial pattern for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, Sep. 2023, Art. no. 5523815.

[49] X. Qiao and W. Huang, "Spectral-spatial-frequency transformer network for hyperspectral image classification," in *Proc. IEEE Sens. Appl. Symp.*, 2023, pp. 1–6.

[50] D. Hong et al., "SpectralFormer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Feb. 2022, Art. no. 5518615.
[51] D. Wang, J. Zhang, B. Du, L. Zhang, and D. Tao, "DCN-T: Dual context network with transformer for hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 32, pp. 2536–2551, May 2023.
[52] X. Qiao, S. K. Roy, H. Wu, and W. Huang, "Hyperspectral image classification based on 3 D sharpened cosine similarity operation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 7669–7672.
[53] Z. Yu et al., "Searching central difference convolutional networks for face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 5295–5305.
[54] S. Jia et al., "A semisupervised Siamese network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5516417.
[55] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.

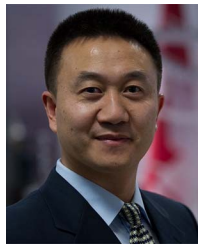
Xin Qiao (Graduate Student Member, IEEE) received the B.Eng. degree in navigation from the Dalian Maritime University, Dalian, China, in 2018, and the M.Eng. degree in naval architecture and ocean engineering from the Zhejiang University, Zhejiang, China, in 2021. He is currently working toward the Ph.D. degree in electrical engineering with the Memorial University of Newfoundland, St. John's, NL, Canada.

His research interest focuses on hyperspectral image classification.

Swalpa Kumar Roy received the bachelor's and the master's degrees in computer science and engineering from the West Bengal University of Technology, Kolkata, India, in 2012, and the Indian Institute of Engineering Science and Technology, Shibpur, (IIEST Shibpur), Howrah, India, in 2015, respectively, and the Ph.D. degree in computer science and engineering from the University of Calcutta, Kolkata, in 2021.

From 2018 to 2023, he was an Assistant Professor with the Department of Computer Science and Engineering, Jalpaiguri Government Engineering College, Jalpaiguri, India, and from 2015 to 2016, he was a Project Linked Person with the Optical Character Recognition Laboratory, Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata. He is currently working as an Assistant Professor with the Department of Computer Science and Engineering, Alipurduar Government Engineering and Management College, Alipurduar, India. His research interests include computer vision, deep learning, and remote sensing.

Dr. Roy was nominated for the prestigious Indian National Academy of Engineering (INAE) engineering teachers mentoring fellowship program by INAE Fellows in the years 2021 and 2023. He was the recipient of the Outstanding Paper Award in the second Hyperspectral Sensing Meets Machine Learning and Pattern Analysis at the Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing in 2021. He is an Associate Editor of the Journal of Springer *Nature Computer Science* 2022 onward. He was also a Reviewer for IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.



Weimin Huang (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in radio physics from Wuhan University, Wuhan, China, in 1995, 1997, and 2001, respectively, and the M.Eng. degree in electrical engineering from the Memorial University of Newfoundland, St. John's, NL, Canada, in 2004.

From 2008 to 2010, he was a Design Engineer with Rutter Technologies, St. John's. Since 2010, he has been with the Faculty of Engineering and Applied Science, Memorial University of Newfoundland, where he is currently a Professor. He has authored or coauthored more than 260 research articles. His research interests include the mapping of oceanic surface parameters via high-frequency ground wave radar, X-band marine radar, synthetic aperture radar, and global navigation satellite systems.

Dr. Huang has been a Technical Program Committee Member. He is the General Co-Chair and Technical Program Co-Chair for the IEEE Oceanic Engineering Society 13th Currents, Waves, and Turbulence Measurement Workshop. He is an Editor of the book Ocean Remote Sensing Technologies: High Frequency, Marine, and GNSS-Based Radar. He is also an Area Editor for the IEEE CANADIAN JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING, an Associate Editor for IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, IEEE JOURNAL OF OCEANIC ENGINEERING, *Remote Sensing*, Frontiers in Marine Science, and he is a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING and five other journals. He is a Reviewer for more than 130 international journals and a Reviewer for many IEEE international conferences, such as *RadarCon*, *International Conference on Communications*, IEEE GLOBAL COMMUNICATIONS CONFERENCE, and *IEEE International Geoscience and Remote Sensing Symposium, and Oceans*. From 2018 to 2021, he was a Member and Co-Chair of the Electrical and Computer Engineering Evaluation Group for Natural Sciences and Engineering Research Council of Canada Discovery Grants. He was the recipient of a Postdoctoral Fellowship from the Memorial University of Newfoundland, the Discovery Accelerator Supplements Award from NSERC in 2017, and the IEEE Geoscience and Remote Sensing Society 2019 Letters Prize Paper Award as well as some other teaching and research awards.