Robust Capsule Network Based on Maximum Correntropy Criterion for Hyperspectral Image Classification

Heng-Chao Li, *Senior Member, IEEE*, Wei-Ye Wang, Lei Pan, Wei Li, *Senior Member, IEEE*, Qian Du, *Fellow, IEEE*, and Ran Tao, *Senior Member, IEEE*

***Abstract*—Recently, deep learning-based algorithms have been widely used for classification of hyperspectral images (HSIs) by extracting invariant and abstract features. In our conference paper presented at IEEE International Geoscience and Remote Sensing Symposium 2018, 1-D-capsule network (CapsNet) and 2-D-CapsNet were proposed and validated for HSI feature extraction and classification. To further improve the classification performance, the robust 3-D-CapsNet architecture is proposed in this article by following our previous work, which introduces the maximum correntropy criterion to address the noise and outliers problem, generating a robust and better generalization model. As such, discriminative features can be extracted even if some samples are corrupted more or less. In addition, a novel dual channel framework based on robust CapsNet is further proposed to fuse the hyperspectral data and light detection and ranging-derived elevation data for classification. Three widely used hyperspectral datasets are employed to demonstrate the superiority of our proposed deep learning models.**

***Index Terms*—Capsule network (CapsNet), deep learning, feature extraction (FE), hyperspectral image (HSI) classification.**

I. Introduction

With advanced technologies on sensors and imaging systems, a hyperspectral image (HSI) contains abundant spectral information in hundreds of narrow and contiguous bands, and also presents rich contextual structure information of imaged scenes. HSIs have been widely applied in many fields, such as agriculture [1], mineralogy [2], and environment [3]. However, as the number of spectral bands increases, the problem of the curse of dimensionality [4] inevitably occurs. In order to address this issue, dimensionality reduction (DR) [5], [6] has been proved an effective technique for hyperspectral data processing, which usually includes feature extraction (FE) [7]–[9] and feature selection (FS) [10]–[12]. Different from FS that finds a subset of the original spectral bands, FE is to transform the original features into a low-dimensional feature space.

Generally, FE methods can be spectral-based and spatial-spectral-based. As far as the spectral-based FE techniques are concerned, principal component analysis (PCA) [13] is the most common one due to its simplicity. PCA is not designed to improve important discriminative information. To overcome this problem, linear discriminant analysis (LDA) [14] was proposed, which is further improved by local Fisher's discriminant analysis (LFDA) [15]. Considering the large spatial variability of spectral signature introduced by light-scattering mechanisms, the hyperspectral data usually presents a nonlinear characteristic [16], [17]. In this case, the aforementioned FE methods based on linear transformation may not be adequate for subsequent analysis.

Recently, manifold learning was successfully introduced to extract intrinsic features of hyperspectral data, such as locally linear embedding (LLE) [18], locality preserving projection (LPP) [19], and neighborhood preserving embedding (NPE) [20]. In addition, kernel techniques [21], [22] also succeed in solving the nonlinear problem by projecting the original data into a higher dimensional kernel-induced feature space, e.g., kernel principal component analysis (KPCA) [23]. By further combining the kernel techniques and graph theory, some advanced and promising FE methods have been developed, such as kernel Laplican regularized collaborative graph-based discriminant analysis (KLapCGDA) [24], kernel sparse and low-rank graph-based discriminant analysis (KSLGDA) [25].

However, in the case where only spectral information is considered for FE, classification performance is not so promising. Actually, spatial information has encouraged much better performance. For example, Huang *et al.* [26] proposed multiple morphological profiles extracted from multicomponent base images to enhance classification accuracy. In [27], a novel principal component analysis-based edge-preserving features (PCA-EPFs) method was developed to extract discriminative features. From the perspectives of graph construction and tensor representation, Pan *et al.* [28] proposed a tensor sparse and low-rank graph-based discriminant analysis method for the FE and classification of HSIs.

In view of the aforementioned FE methods, only shallow appearance features are extracted for classification. Recent studies

Manuscript received May 9, 2019; revised November 22, 2019; accepted January 10, 2020. Date of publication February 5, 2020; date of current version February 21, 2020. This work was supported by the National Natural Science Foundation of China under Grant 61871335. *(Corresponding author: Wei-Ye Wang.)*

H.-C. Li, W.-Y. Wang, and L. Pan are with the Sichuan Provincial Key Laboratory of Information Coding and Transmission, Southwest Jiaotong University, Chengdu 610031, China (e-mail: lihengchao_78@163.com; wwy@my.swjtu.edu.cn; mapan.lei@163.com).

W. Li and R. Tao are with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: liwei089@ieee.org; rantao@bit.edu.cn).

Q. Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS, 39762 USA (e-mail: du@ece.msstate.edu).

Digital Object Identifier 10.1109/JSTARS.2020.2968930

on deep learning [29]–[31] have opened a new topic for the analysis of HSIs. By designing two or more training layers, deep learning-based models, such as stacked autoencoder (SAE) [32], deep belief network (DBN) [33], and convolutional neural network (CNN) [34], [35], aim to simulate the biological process from the retina to the cortex of human beings. Depending on the high-level features learned by these deep architectures, better performance is certainly yielded.

Recently, for the first time, Chen *et al.* [36] proposed a hybrid framework that combines PCA and logistic regression simultaneously with the deep architecture of SAE, providing competitive performance. Subsequently, an improved autoencoder, called spatial updated deep autoencoder (SDAE), was proposed by integrating contextual information when updating features in [37]. Again, from the perspective of DBN, Chen *et al.* [38] developed a novel deep architecture that combines spatial-spectral FE and classification together to obtain better classification performance. However, the aforementioned deep models involve many parameters to be trained. Moreover, due to the vector-based representation of spatial information, SAE-based and DBN-based learning models fail to extract spatial information efficiently. Fortunately, the CNN has demonstrated advantages in effective extraction of spatial information and significant reduction of training parameters. In [39], Romero *et al.* proposed a deep CNN model for HSI processing by using greedy layerwise unsupervised pretraining. In order to improve the discriminative ability of extracted features, a supervised deep CNN architecture containing five layers was developed for HSI classification [40]. To effectively extract spectral and spatial features simultaneously, Chen *et al.* [41] further developed a 3-D-CNN model, along with the strategy for solving the overfitting problem. Recently, several deep CNN structures, such as residual network (ResNet) [42], [43], dense convolutional network (DenseNet) [44], [45], and dual path network (DPN) [46], have also been constructed for uncovering the highly discriminative spectral–spatial features of the HSI data.

However, the input and output of CNN are both scalars, which leads to a low representative ability. Besides, the CNN adopts the max-pooling strategy to offer invariance, ignoring spatial relationship of significant features that can improve the discriminative ability. Recently, a multilayer deep learning model based on the concept of so-called capsule has been proposed in [47]. In that work, Hinton *et al.* [47] pointed out that an activity vector containing the information of position, orientation, deformation, and texture in a capsule is better at revealing and learning the discriminative features. Deng *et al.* [48] proposed a modified two-layer capsule network (CapsNet) with limited training samples for HSI processing. To improve the network architecture and present a reasonable initialization strategy, Yin *et al.* [49] proposed a new architecture with using only three shallow layers and presenting the elaborated initialization strategy for the classification of HSIs. In order to extract more discriminative features, a supervised deep CapsNet architecture containing five layers was developed for HSI classification [50]. Aiming to overcome the deficiencies of the existing CNN-based models when classifying the HSI, Wang *et al.* [51] proposed a hybrid method based on CapsNet and TripleGANs models. However, the HSI usually contains the noise and outliers introduced in the process of data measurement and acquisition [52]. Although the aforementioned deep models can obtain good performance, this problem has still not been effectively addressed.

In recent years, it has been found that multisource data, e.g., light detection and ranging (LiDAR), can provide a source of complementary information to further improve classification performance. The elevation and object height information contained in LiDAR data is helpful with improvement on the discrimination of different classes. Therefore, compared with an individual source [53]–[55], the joint use of LiDAR and HSI can promote higher discrimination power. For instance, support vector machines (SVMs) and Gaussian maximum likelihood were adopted to investigate the joint classification for the HSI and LiDAR [55]. In order to accurately classify land covers, the CNN was developed to fuse the features extracted from the HSI and LiDAR in [56]. Furthermore, Xu *et al.* [57] developed a two-branch convolution neural network to fuse features from HSIs and LiDAR data. However, the CNN exploits only the significant differences in samples considered from the point of the pooling operation, ignoring the relative positions of significant differences in samples, or sample-related shapes, textures, and other location information. Meanwhile, the pooling operation loses exact location and relative spatial information.

In our previous conference paper [58], we proposed the 1-D-CapsNet and 2-D-CapsNet models for HSI FE and classification. Following this work and considering the aforementioned drawbacks of the existing deep models, the maximum correntropy criterion (MCC)-based robust 3-D-CapsNet architecture is proposed in this article. In our proposed work, the MCC is applied to control the negative impact introduced by the noise and outliers, generating a more robust and representative deep architecture. In order to efficiently exploit useful information from multisource data and further overcome the shortcomings of CNN for multisource remote sensing data classification, a novel dual channel framework based on CapsNet is proposed. In the CapsNet model, the dynamic routing agreement replaces the pooling operation in CNN to make full use of the position information of samples, resulting in a better representation and higher classification performance.

To sum up, the main contributions of our proposed work lie in the following aspects.

1) The MCC is introduced into the proposed 3-D-CapsNet model for robust feature learning. To the best of our knowledge, it is the first time that the MCC is used in CapsNet for addressing the noise and outlier problem in HSIs.
2) A novel MCC-based dual channel robust CapsNet framework is proposed to fuse multisource remote sensing data, e.g., hyperspectral data and LiDAR data, in which the spatial–spectral information of HSI and the elevation information of LiDAR can be efficiently fused to extract more discriminative features for the classification.

The rest of this article is organized as follows. Section II briefly reviews 3-D-CNN and CapsNet. Then, two proposed classification frameworks are introduced in detail in Section III,

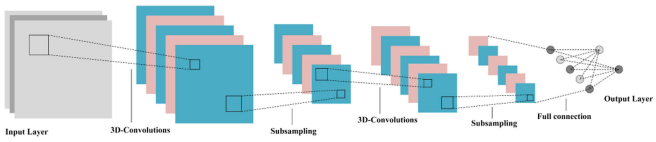


Fig. 1. Example of 3-D-CNN architecture, which has two convolutional layers, two pooling layers, and one fully connected layer.

including the robust 3-D-CapsNet and the dual channel robust CapsNet. The experimental results and discussion are presented in Section IV. Finally, Section V concludes this article.

II. Related Work

A. Convolutional Neural Network (CNN)

The CNN plays a significant role in processing visual-based problems, which was first proposed in [59] and has multifarious combinations of convolutional layers and pooling layers, and finally ends with a fully connected layer. By using the local connectivity among the neurons of adjacent layers, CNN can exploit the locally spatial information. In order to effectively investigate the spatial and intrinsic structure information from the 3-D data, 3-D-CNN has been proposed to extract the contextual structure features by 3-D convolutional kernels. A typical instance of 3-D-CNN is presented in Fig. 1, which includes two 3-D-convolutional layers, two pooling layers, and one fully connected layer. First, the feature extractors (3-D-convolutional kernels) of the convolutional layer sweep over the topology and transform the input into feature maps. Then, the pooling layer is adopted to enlarge the perception area and increase the invariance property of features. After that, the feature maps are flattened into a feature vector, followed by the fully connected layer. Finally, the softmax is utilized as the output layer activation to generate the prediction probabilities.

B. Capsule Network (CapsNet)

To overcome the shortcomings of CNN and make it closer to the cerebral cortex activity structure, Hinton [47] proposed a high dimensional vector called "capsule" to represent an entity (an object or a part of an object) by a group of neurons rather than a single neuron. The activities of the neurons within an active capsule represent various properties of a particular entity that is presented in the image. Each capsule learns an implicit definition of a visual entity that outputs the probability of an entity and a set of "instantiated parameters," including the precise pose (position, size, orientation), deformation, velocity, albedo, hue, texture, etc.

The architecture of CapsNet is different from other deep learning models. The results of inputs and outputs of CapsNet are vectors, whose norm and direction represent the existence probability and various attributes of the entity, respectively. The same level of a capsule helps to predict the instantiation parameters of a higher-level capsule through a transformation matrix and, subsequently, dynamic routing is adopted to make the prediction consistent. When multiple predictions are consistent, the higher-level of one capsule will become active.

A simple CapsNet architecture is shown in Fig. 2, in which the architecture is shallow with only two convolutional layers (Conv1, PrimaryCaps) and one fully connected layer (EntityCaps). Specifically, Conv1 is the standard convolutional layer, which converts images to primary features and outputs to PrimaryCaps through a convolution filter with a size of $13 \times 13 \times 256$. In the case where the original image is not suitable for the input of the first layer of the CapsNet, the principal features after convolutions are adopted.

The second convolutional layer constructs the corresponding vector structure as the input of the capsule layer. The traditional convolution of each output is a scalar, but the convolution of PrimaryCaps is different from the traditional one. It can be regarded as a 2-D convolution of eight different weights for the input of $15 \times 15 \times 256$. Each time the implementation takes 32 sizes of 11×11 steps to two convolved, and outputs $5 \times 5 \times 8 \times 32$ vector structure input. The third layer (EntityCaps) is the output layer, which contains nine standard capsules corresponding to 9 different classes.

Recalling the original formulation in [47], a layer of a capsule network is divided into multiple computational units named capsules. Assume that the capsule i outputs activity vector $\boldsymbol{u}_i$ from the PrimaryCaps i, it is provided to the capsule j to generate activity level $\boldsymbol{v}_j$ of EntityCaps. Propagation and updating are conducted using vectors between PrimaryCaps and EntityCaps.

Matrix processing is utilized for scalar input in each layer of traditional neural network, which is essentially a linear combination of output. The capsule processing input is divided into two stages: linear combination and routing. The linear combination here refers to the idea of processing scalar input by the neural network, which means to process the relationship between two objects in the scene through the visual transformation matrix while preserving their relative relation. Specifically, the linear combination is formulated as

$$\hat{\boldsymbol{u}}_{j|i} = \boldsymbol{u}_i \boldsymbol{W}_{ij} \tag{1}$$

where $\hat{\boldsymbol{u}}_{j|i}$ is a prediction vector produced by transforming the output $\boldsymbol{u}_i$ of a capsule in the layer below by a weight $\boldsymbol{W}_{ij}$. Then, in the routing stage, the input vector $\boldsymbol{s}_j$ of capsule j can be defined as

$$\boldsymbol{s}_j = \sum_i c_{ij} \hat{\boldsymbol{u}}_{j|i} \tag{2}$$

where c_{ij} is the coupling coefficient determined by the iterative dynamic routing process. The routing part is actually a weighted sum of $\hat{\boldsymbol{u}}_{j|i}$ with the coupling coefficient. The vector output of capsule j is calculated by applying a nonlinear squashing function that can ensure short vectors to be shrunk to almost zero length and long vectors get shrunk to a length slightly below one as

$$\boldsymbol{v}_j = \frac{\|\boldsymbol{s}_j\|^2}{1 + \|\boldsymbol{s}_j\|^2} \frac{\boldsymbol{s}_j}{\|\boldsymbol{s}_j\|}. \tag{3}$$

Obviously, the capsule's activation function actually suppresses and redistributes vector lengths. Its output can be used as the probability of the entity represented by the capsule in the current category.

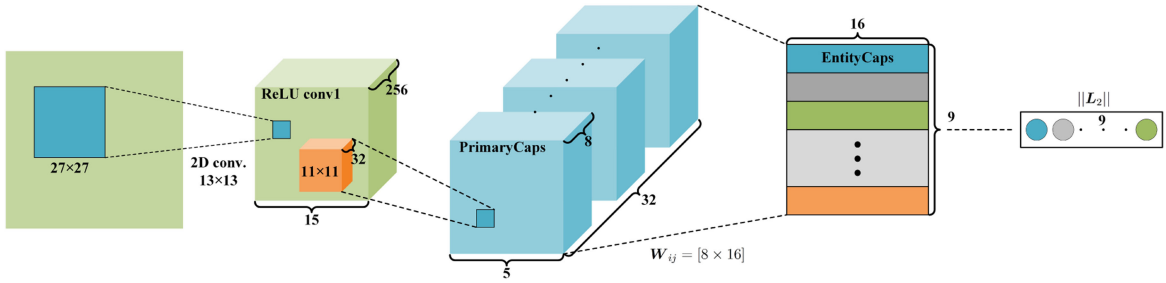


Fig. 2. Illustration of CapsNet classification framework.

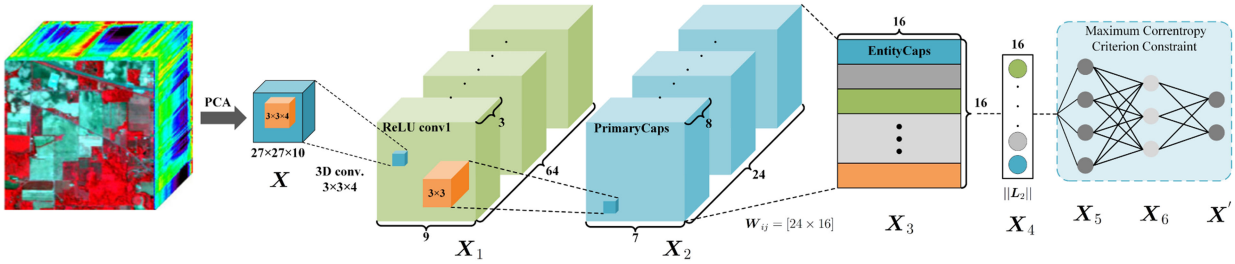


Fig. 3. Architecture of the proposed robust 3-D-CapsNet.

The total loss function of the original CapsNet is a weighted summation of marginal loss and reconstruction loss. The mean square error (MSE) is used in the original reconstruction loss function, which degrades the model significantly when processing noisy data. In the next section, the MCC [60] is presented to address this problem.

III. Proposed Robust Deep Frameworks

A. MCC

Recently, correntropy-based information theoretic learning (ITL) shows robust performance in face recognition [61]. Correntropy is a generalized similarity measure between two random variables A and B, defined as

$$V(\mathrm{A},\mathrm{B}) = E[\kappa_\sigma(\mathrm{A}-\mathrm{B})] \tag{4}$$

where $E(\cdot)$ is the expectation operator and $\kappa_\sigma(\cdot)$ is a kernel function. The most widely used kernel in correntropy is the Gaussian kernel $\kappa_\sigma(\mathrm{e}) = \exp\{-\|\mathrm{e}\|_2^2/\sigma^2\}$ with a width parameter σ. In practice, the joint probability density function of A and B is often unknown and only a finite number of empirical data $\{(a_i, b_i)\}_{i=1}^n$ are given, resulting in the sample estimator of correntropy

$$\hat{V}_{n,\sigma}(\mathrm{A},\mathrm{B}) = \frac{1}{n}\sum_{i=1}^{n}\kappa_\sigma(a_i - b_i). \tag{5}$$

The correntropy of the error between a_i and b_i can be used as a cost function for adaptive system training and referred to as the MCC. The optimization cost under MCC is thus

$$\max \frac{1}{n}\sum_{i=1}^{n}\kappa_\sigma(e_i) = \max \frac{1}{n}\sum_{i=1}^{n}\kappa_\sigma(a_i - b_i) \tag{6}$$

where $e_i = a_i - b_i$ is the error variable. Although the noise and outliers can introduce large errors when using the MSE loss function, this can be effectively controlled with the help of the MCC.

B. Robust 3-D-CapsNet Based on MCC

In this subsection, a robust 3-D-CapsNet architecture is developed to extract the spatial–spectral features effectively, whose architecture is shown in Fig. 3. Here, only special parts of the proposed CapsNet are explained in detail.

The neighboring pixels may have similar spatial contexts, which can provide supporting information for classification. The $K \times K$ neighbors of a pixel are exploited as the input of this CapsNet model. In the first layer, a convolutional kernel with the size of $3 \times 3 \times 4$ is used for 3-D convolution to combine spatial and spectral information, providing expressive features for the second layer and passing the spatial–spectral information through the whole network. In order to make the underlying capsules more perceptive, the stride of size $3 \times 3 \times 3$ is chosen to expand the perceived area. Since a single convolutional layer is unable to extract appropriate features, another convolutional capsule layer (24 convolution channels, eight filters per convolution) is added. Unlike traditional convolutional layer, each channel of convolutional capsule layer generate eight (i.e., the number of neural units contained in each capsule) feature maps, which can be understood as a matryoshka doll. The third layer (EntityCaps) is the output layer, which contains 16 standard capsules corresponding to 16 different classes.

The loss function of the proposed model can be formulated as a combination of two terms: Margin loss term and reconstruction constraint term

$$L = L_m + \lambda L_r \tag{7}$$

where λ is a balance coefficient to assign an appropriate weight to the reconstruction loss. The margin loss function of the framework can be expressed as

$$L_m = T_c \max\left(0, m^+ - \| \boldsymbol{v}_c \|\right)^2 + \lambda_1 (1 - T_c) \max\left(0, \| \boldsymbol{v}_c \| - m^-\right)^2 \tag{8}$$

where c is the classification category and λ_1 is a free parameter that needs to be tuned empirically, T_c is the indicator of the classification (c exists as 1, c does not exist as 0), m^+ is the upper boundary, and m^- is the lower boundary. In addition, $\|\boldsymbol{v}_c\|$ is the length of the activity vector (i.e., the probability). For each category, there is a separate loss function as the objective for model optimization.

The MSE is usually used in the original reconstruction loss function, which depends heavily on the Gaussian and linear assumptions. However, in presence of non-Gaussian noise and large outliers, the effectiveness of the MSE-based loss function is significantly deteriorated. To address this issue, the MCC is used to deal with the noise and outliers in HSIs. The correntropy function is adopted as a simple and robust cost function that may achieve much better performance in practical applications, particularly when the data contains outliers. Thus, the reconstruction constraint loss function of the framework can be expressed as

$$L_r = 1 - \exp\left\{ - \frac{\|X - X'\|_2^2}{\sigma^2} \right\} \tag{9}$$

where X is the original input data and X' is the reconstructed output data. MCC maps the input space into a high-dimensional space by using the Gaussian kernel function. It is not difficult to see that the MCC cost in (9) with Gaussian kernel reaches its minimum if $X = X'$. Compared with the MSE, the MCC is more robust for outliers. The MCC is only sensitive in an area of small residual errors, which are controlled by the kernel bandwidth. With this constraint, extremely erroneous samples have less influence on the model. Hence, using MCC to replace the MSE-based loss function may achieve much better performance, especially when the data contains the noise and outliers.

In the proposed algorithm, the weight is iteratively updated by minimizing the joint loss function in (7). During the optimization, the parameters of the neural network are fixed and only the weight is updated through the gradient back-propagation (BP). The proposed loss function can be optimized through the BP algorithm, which involves two parts: forward weight update and reconstruction weight update.

The $\{(X_1, Y_1), \ldots, (X_k, Y_k)\}$ are minibatches of training samples. With the input sample X_k in hand, the feature map $\boldsymbol{v}_c$ and the reconstructed sample X'_k can be obtained from the forward pass of CapsNet. $W_{i(=1,2,3)}$ and $W_{j(=4,5,6)}$ correspond to the weight parameters of forward layer $i = \{1, 2, 3\}$ and reconstruction layer $j = \{4, 5, 6\}$, respectively. In layer j of feature reconstruction, the loss function L is optimized to obtain the gradient $\frac{\partial L}{\partial X_k}$. The gradient $\frac{\partial L}{\partial X_k}$ is back-propagated to update the W_j, which can be shown as

$$\begin{aligned} \frac{\partial L}{\partial X_k} &= \frac{\partial L_m}{\partial X_k}\frac{\partial X_k}{\partial W_j} + \lambda \frac{\partial L_r}{\partial X_k}\frac{\partial X_k}{\partial W_j} \\ &= 0 + \lambda \frac{\partial L_r}{\partial X_k}\frac{\partial X_k}{\partial W_j} \\ &= \lambda \frac{\partial L_r}{\partial X_k}\frac{\partial X_k}{\partial W_j} \end{aligned} \tag{10}$$

$$W_j(t) = W_j(t-1) - \eta \sum_k \lambda \frac{\partial L_r}{\partial X_k}\frac{\partial X_k}{\partial W_j} \tag{11}$$

where t is the number of iterations and η is the learning rate. For example, the weight of the last layer is updated by

$$W_6(t) = W_6(t-1) - \eta \sum_k \frac{\lambda}{\sigma^2} 2(X_k - X'_k) X_6(k) \times \exp\left\{ - \frac{\|X_k - X'_k\|_2^2}{\sigma^2} \right\} \tag{12}$$

where W_6 and $X_6(k)$ are the weight and the input of the last layer, respectively.

In layer i of forward feature extraction, the proposed CapsNet optimization is implemented to minimize the term $L_m + \lambda L_r$ to obtain the gradient $\frac{\partial L_m}{\partial X_k}$ and $\frac{\partial L_r}{\partial X_k}$. Finally, both gradients $\frac{\partial L_m}{\partial X_k}$ and $\frac{\partial L_r}{\partial X_k}$ are back-propagated to update W_i, which can be shown as

$$\frac{\partial L}{\partial X_k} = \frac{\partial L_m}{\partial X_k}\frac{\partial X_k}{\partial W_i} + \lambda \frac{\partial L_r}{\partial X_k}\frac{\partial X_k}{\partial W_i} \tag{13}$$

$$W_i(t) = W_i(t-1) - \eta \sum_k \left(\frac{\partial L_m}{\partial X_k}\frac{\partial X_k}{\partial W_i} + \lambda \frac{\partial L_r}{\partial X_k}\frac{\partial X_k}{\partial W_i} \right). \tag{14}$$

For example, the weight of the third layer is updated by

$$W_3(t) = W_3(t-1) - \eta \sum_k \left(\frac{\partial L_m}{\partial X_k}\frac{\partial X_k}{\partial W_3} + \lambda \frac{\partial L_r}{\partial X_k}\frac{\partial X_k}{\partial W_3} \right). \tag{15}$$

$$\frac{\partial L_m}{\partial X_k}\frac{\partial X_k}{\partial W_3} = \begin{cases} 2T_c \left(A^T A W_3 - \frac{m^+ A^T A W_3}{\|AW_3\|} \right) + 2\lambda_1 (1 - T_c) \times \left(A^T A W_3 - \frac{m^- A^T A W_3}{\|AW_3\|} \right), & \text{if } m^+ > \|v_c\|, \|v_c\| > m^- \\ 2\lambda_1 (1 - T_c) \left(A^T A W_3 - \frac{m^- A^T A W_3}{\|AW_3\|} \right), & \text{if } m^+ \le \|v_c\|, \|v_c\| > m^- \\ 2T_c \left(A^T A W_3 - \frac{m^+ A^T A W_3}{\|AW_3\|} \right), & \text{if } m^+ > \|v_c\|, \|v_c\| \le m^- \\ 0, & \text{if } m^+ \le \|v_c\|, \|v_c\| \le m^- \end{cases} \tag{16}$$

$$\frac{\partial L_r}{\partial X_k}\frac{\partial X_k}{\partial W_3} = \frac{1}{\sigma^2} 2(X_k - X'_k) A W_4 W_5 W_6 \times \exp\left\{ - \frac{\|X_k - X'_k\|_2^2}{\sigma^2} \right\} \tag{17}$$

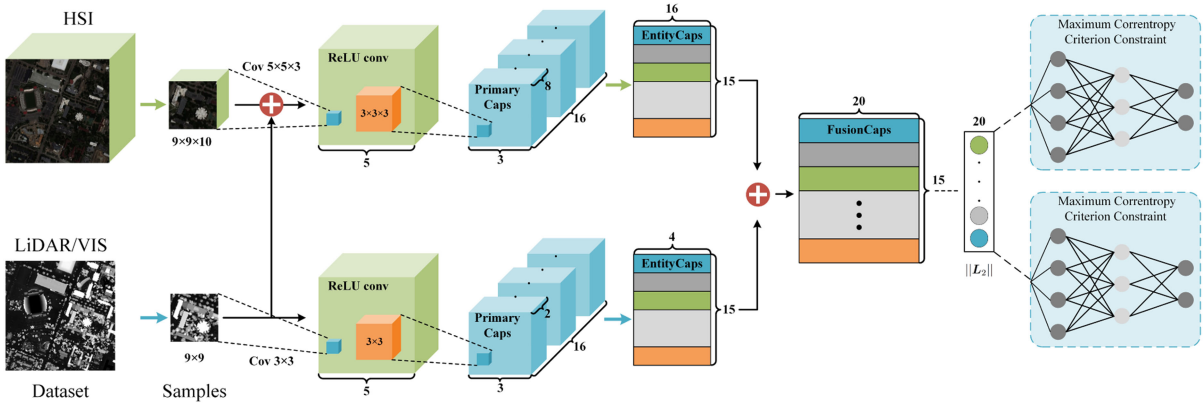


Fig. 4. Architecture of the proposed MCC-based dual channel robust CapsNet for multisource remote sensing data classification. Note that the input data are a local patch around its center pixel.

Algorithm 1: Training Algorithm.

Input: Minibatches $\{(X_1, Y_1), \ldots, (X_k, Y_k)\}$ of training samples, learning rate η, number of iterations $t = 0$, hyperparameters λ, σ, initialized parameters W in the network

Output: The trained network parameters W.

while not converge do
  $t = t + 1$,
  for layer $i = \{1, 2, 3\}$ of forward layer **do**,
    Compute the joint loss L by (7),
    Compute the backpropagation error $\frac{\partial L}{\partial X_k}$ for each k by (13),
    Update the parameters W_i by (14),
  end for
  for layer $j = \{4, 5, 6\}$ of reconstruction layer **do**,
    Compute the reconstruction loss L_r by (9),
    Compute the backpropagation error $\frac{\partial L_r}{\partial X_k}$ for each k by (10),
    Update the parameters W_j by (11),
  end for
end while

where $A = C_3 X_2(k)$, W_3 and $X_2(k)$ are the weight and the input of the third layer, respectively. C_3 is the coupling coefficient of the third layer determined by the iterative dynamic routing process. The main steps of the proposed robust CapsNet optimization process are summarized in Algorithm 1.

C. Dual Channel Robust CapsNet Based on MCC

LiDAR can provide the elevation information of height and shape with respect to the sensor. LiDAR contains full of altitude information, which is valuable for better describing the same scenario obtained by the light sensor. Because different data sources have different characteristics, a variety of classification fusion strategies are proposed to combine multiple characteristics of different data sources. A novel MCC-based dual channel robust CapsNet framework is proposed for pixelwise classification with fusing multisource remote sensing data, e.g., HSI and LiDAR.

The main procedure of the proposed classification framework is shown in Fig. 4, in which a MCC-based dual channel robust CapsNet for HSI and LiDAR is included. There are four parts in the framework: a 3-D-CapsNet channel, a 2-D-CapsNet channel, a fusion network, and a fully connected deep neural network. The 3-D-CapsNet is designed to extract the spatial–spectral features of HSI, the 2-D-CapsNet is designed to extract the elevation features of LiDAR data, the fusion connected network is designed to fuse the extracted features, and the fully connected deep neural network is given to reconstruct input from HSI and LiDAR.

Due to abundant spectral information, HSIs can be used to distinguish and detect ground targets with high diagnostic ability. However, the poor spatial resolution of HSI remains a major concern while the LiDAR data has more accurate elevation resolution. Therefore, the HSI and LiDAR images are firstly merged as the input for the 3-D-CapsNet. Meanwhile, spatial neighborhoods as the input of the 2-D-CapsNet are individually generated from LiDAR. Next, normalized data are fed into the first layer followed by a convolutional operation with a spatial kernel of $5 \times 5 \times 3$ and 5×5 correspond to 3-D-CapsNet and 2-D-CapsNet. In the next layer, the convolutional operation is applied with a spatial kernel of $3 \times 3 \times 3$ and 3×3 to perform a local feature detection to obtain high-level features in each channel, which is used as the input of primary capsule. In 2-D-CapsNet, each primary capsule adopts the convolutional operation (16 convolution channels, two filters per convolution) containing two convolutional units with 3×3 kernel and a stride of 2×1. The output of 2-D-CapsNet is a capsule of a 4-D vector of size $3 \times 3 \times 2 \times 16$. In 3-D-CapsNet, each primary capsule adopts the convolutional operation (16 convolution channels, eight filters per convolution) containing eight convolutional units with $3 \times 3 \times 3$ kernel and a stride of 8×1. The output of 3-D-CapsNet is a capsule of a 4-D vector of size $3 \times 3 \times 8 \times 16$. The last layer, i.e., the output layer of each channel, is a set of 15 standard capsules, with each capsule representing a

category. Through 2-D-CapsNet and 3-D-CapsNet, the elevation information and spatial–spectral information can be extracted, respectively. Feature stacking is used for the fusion of spectral, spatial, and elevation features. The new features contain different features of the same objects, i.e., each entity capsule. Through splicing operation, the features extracted from shallow networks are aggregated into the main entity. To this end, the design concept of the capsule network can be better utilized for fusing different features. After fusion, all of the features are integrated at the third layer and used as input to the reconstruction layer. Then, the MCC is applied to reconstruct the loss function of 2-D-CapsNet and 3-D-CapsNet, respectively. Algorithm 1 shows the optimization process for the proposed framework. In conclusion, the dual channel CapsNet can effectively overcome the disadvantageous impact introduced by the noise and outliers, generating a more robust and representative deep architecture.

IV. Experiments and Discussion

To demonstrate the effectiveness of our proposed CapsNet-based FE frameworks, some existing deep learning models, such as SAE [36], CNN [41], CapsNet [50], and dual-tunnel CNN [57], are used for comparison purpose. The SVM [55] classifier with a radial basis function (RBF) kernel is used as the baseline method. The class-specific accuracy, overall accuracy (OA), average accuracy (AA), and kappa coefficient (κ) are presented for quantitative assessment after ten runs. All experiments are implemented on an Intel Core i7 $-$ 8700 CPU with 8G RAM and a Nvidia GeForce GTX 1080. The TensorFlow library is adopted for the design of deep learning architecture.

A. Datasets

The first hyperspectral dataset was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines test site in northwest Indiana, in 1992. The image represents an agricultural scenario with 145 $\times$ 145 pixels and consists of 220 bands that covers the spectral range from 0.2 to 2.4 μm with a spatial resolution of 20 m. After removing 20 water absorption bands, 200 bands are preserved for subsequent analysis. A total of 10 249 samples from 16 different classes are used in the experiments, from which 10% samples per class are chosen for training and the rest for testing. The specific numbers of training and testing samples are listed in Table I. The ground truth and false color composition of three bands are shown in Fig. 5.

The second hyperspectral dataset was acquired by the AVIRIS instrument over Kennedy Space Center (KSC), Florida, on March 23, 1996. The KSC dataset has an altitude of approximately 20 km, with a spatial resolution of 18 m. The dataset includes 176 bands used for the analysis after removing water absorption and low-signal-to-noise-ratio bands. For classification purpose, 13 classes are defined for the site. The ground truth and false color composition of three bands are shown in Fig. 6. The specific number of training and testing samples is listed in Table II.

The last dataset was acquired using an ITRES Compact Airborne Spectrographic Imager 1500 hyperspectral imager over

TABLE I
Class Labels and the Number of Training and Tesing Samples for Indian Pines Dataset

#	Class	Train	Test
1	Alfalfa	5	41
2	Corn-notill	142	1286
3	Corn-mintill	83	747
4	Corn	24	213
5	Grass-pasture	48	435
6	Grass-trees	73	657
7	Grass-pasture-mowed	3	25
8	Hay-windrowed	48	430
9	Oats	2	18
10	Soybean-notill	97	875
11	Soybean-mintill	245	2210
12	Soybean-clean	59	534
13	Wheat	21	184
14	Woods	126	1139
15	Build-Grass-Trees-Drives	39	347
16	Stons-Steel-Towers	9	84
	Total	1024	9225

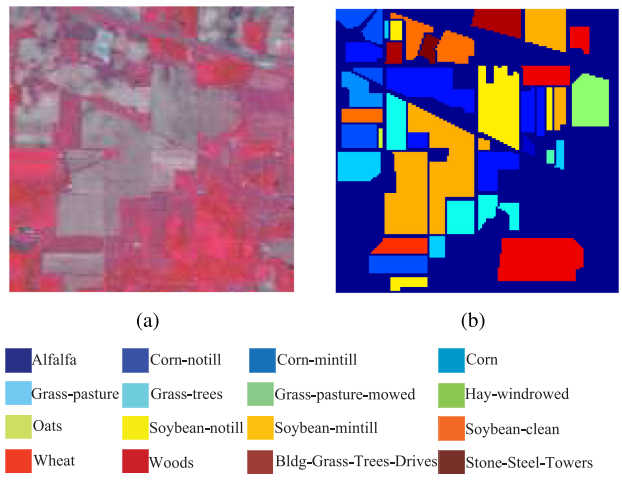


Fig. 5. Indian Pines dataset. (a) Pseudocolor image. (b) Ground truth map.

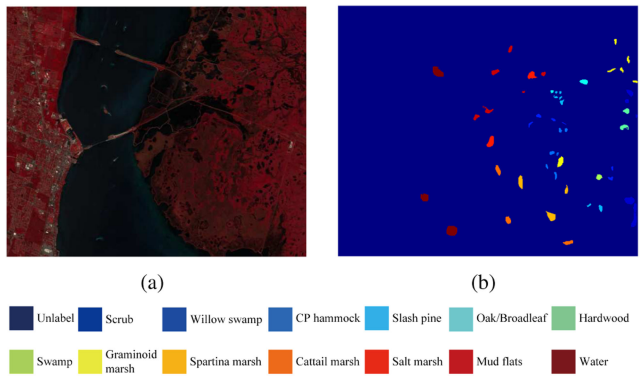


Fig. 6. Kennedy Space Center dataset. (a) Pseudocolor image. (b) Ground truth map.

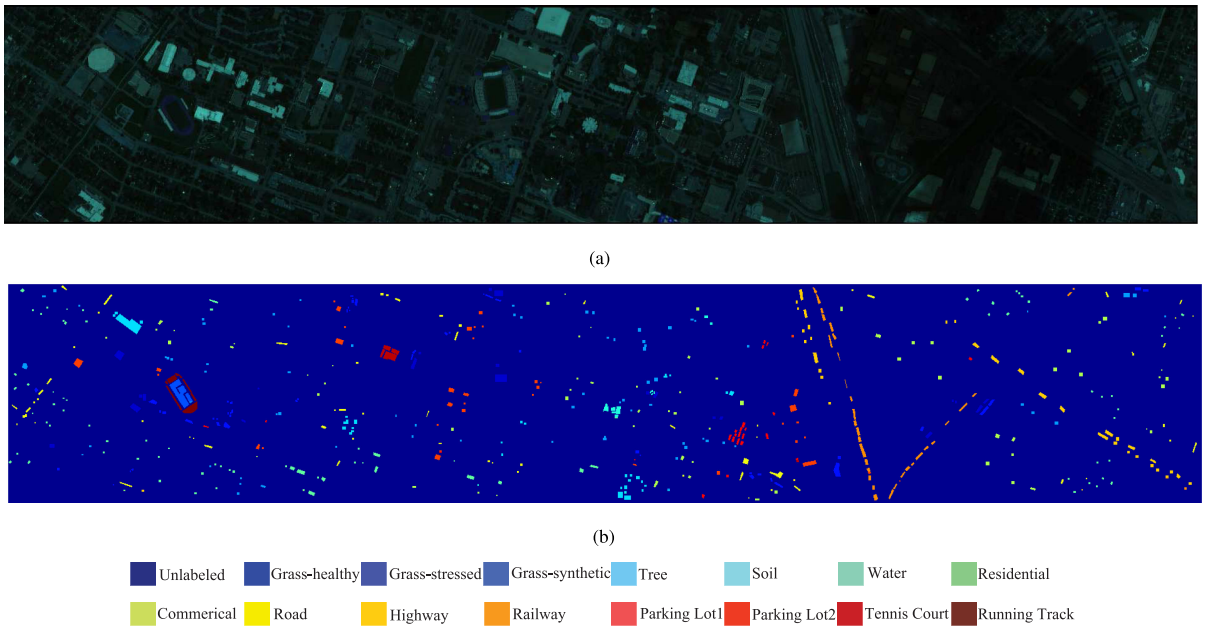


Fig. 7. University of Houston dataset. (a) Pseudocolor image. (b) Ground truth map.

TABLE II
CLASS LABELS AND THE NUMBER OF TRAINING AND TESING SAMPLES FOR KSC DATASET

#	Class	Train	Test
1	Scrub	77	684
2	Willow swamp	25	218
3	CP hammock	26	230
4	Slash pine	26	226
5	Oak/Broadleaf	17	144
6	Hardwood	23	206
7	Swamp	11	94
8	Graminoid marsh	44	387
9	Spartina marsh	52	468
10	Cattail marsh	41	363
11	Salt marsh	42	377
12	Mud flats	51	452
13	Water	93	834
	Total	528	4683

TABLE III
CLASS LABELS AND THE NUMBER OF TRAINING AND TESING SAMPLES FOR HOUSTON DATASET

#	Class	Train	Test
1	Health grass	198	1053
2	Stressed grass	190	1064
3	Synthetic grass	192	505
4	Tress	188	1056
5	Soil	186	1056
6	Water	182	143
7	Residential	196	1072
8	Commercial	191	1053
9	Road	193	1059
10	Highway	191	1036
11	Railway	181	1054
12	Parking lot 1	192	1041
13	Parking lot 2	184	285
14	Tennis court	181	247
15	Running track	187	473
	Total	2832	12197

the University of Houston campus and the neighboring urban area, which includes a hyperspectral dataset and a rasterized LiDAR dataset that are geographically coregistered. There are 144 spectral bands that range from 0.38 to 1.05 μm. This image presents an area with the size of 1905 $\times$ 349 pixels and the spatial resolution of 2.5 m. A total of 15 different classes are included in this data. The ground truth and false color composition of three bands are shown in Fig. 7. The specific number of training and testing samples is listed in Table III.

B. Parameter Tuning

According to the framework shown in Figs. 3 and 4, two deep networks are need to be trained. The experimental data is first normalized, and a neighborhood size of 27 $\times$ 27 is adopted. Considering the limitation of input and memory, only two convolutional layers and one fully connected layer are applied in the proposed frameworks. For the learning rate, 0.01 is an appropriate choice. During the training process, the minibatch size is set to 100, and the training epoch number is 3000. For the SVM and extreme learning machine (ELM) classifiers, the SVM with an RBF kernel is used in this work, in which the hyperplane parameters C and σ are tuned in the range $\{0.001, 0.01, \ldots, 10, 100, 1000\}$ and $\{2^{-8}, 2^{-7} \cdots, 2^{7}, 2^{8}\}$ by fivefold cross-validation, respectively. Then, ELM with the sigmoid activation function is adopted, where the hidden layer parameters a_i and b_i are randomly generated based on uniform

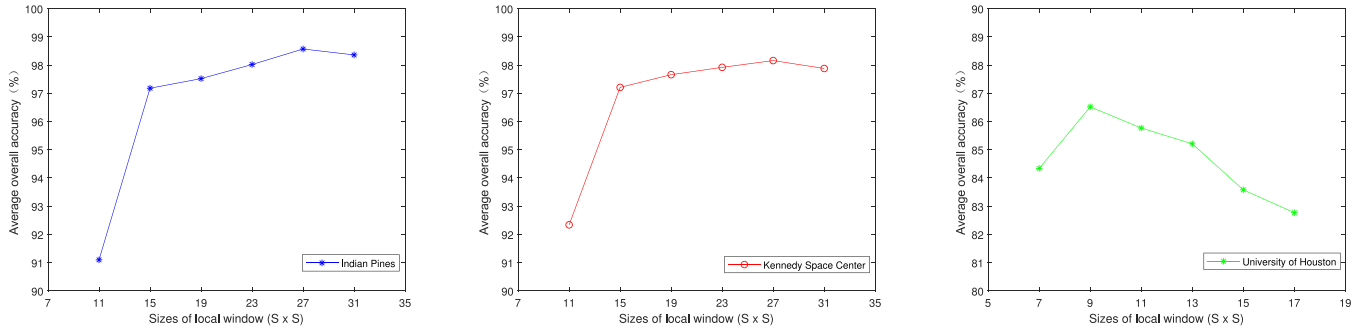


Fig. 8. Classification accuracies (OA) of the proposed method on the three datesets with different sizes of local window. (a) Indian Pines. (b) Kennedy Space. (c) University of Houston.

TABLE IV
SUMMARY OF THE PARAMETERS IN EACH LAYER OF THE TOPOLOGY OF THE PROPOSED 3-D-CAPSNET

Data set	Indian Pines	KSC
Conv1	$3 \times 3 \times 4 \times 64$	$3 \times 3 \times 156 \times 64$
PrimaryCaps	$3 \times 3 \times 8 \times 24$	$5 \times 5 \times 8 \times 56$
EntityCaps	16×16	16×13
Reconstruction1	512	512
Reconstruction2	1024	1024

TABLE V
SUMMARY OF THE PARAMETERS IN EACH LAYER OF THE TOPOLOGY OF THE PROPOSED DUAL CHANNEL CAPSNET

Layer	3-D-CapsNet	2-D-CapsNet
Conv1	$5 \times 5 \times 3 \times 128$	$5 \times 5 \times 32$
PrimaryCaps	$3 \times 3 \times 8 \times 16$	$3 \times 3 \times 2 \times 16$
EntityCaps	16×15	4×15
FusionCaps	20×15	
Reconstruction1	512	512
Reconstruction2	1024	1024

distribution from the range $[-1, 1]$. Tables IV and V summarizes the configuration parameters for each layer, which have been demonstrated to be a good choice for obtaining promising results with testing HSI and LiDAR. The effect of different sizes ($s \times s$) of local window on classification performance is further studied, whose experimental results in Fig. 8 indicate that the proposed 3-D-CapsNet can yield the satisfactory performance when s is fixed to 27 for Indian Pines and Kennedy Space Center dataset. For the University of Houston dataset, 9 is a reasonable value for s in dual channel robust CapsNet. The reason may be that the formers are rural/vegetation scenarios, which contain large spatial homogeneity. The University of Houston dataset includes small homogeneous regions since it is obtained from an urban area.

C. Comparison of Classification Performance

Tables VI–VIII show the class-specific accuracy, OA, AA, and κ under different methods for three experimental datasets.

TABLE VI
CLASSIFICATION WITH SPATIAL–SPECTRAL FEATURES ON THE INDIAN PINES DATASET

Classifier	SVM	SAE	CNN	CapsNet	Proposed
OA(%)	95.39	93.81	97.27	98.03	**98.57**
AA(%)	90.21	89.79	**97.78**	93.04	95.99
$\kappa \times 100$	94.73	92.92	98.03	97.76	**98.38**
1	78.04	92.68	**97.56**	87.80	87.80
2	91.29	92.76	95.95	96.26	**97.12**
3	98.25	90.49	99.46	98.79	**100.00**
4	82.15	78.87	99.06	91.54	**99.53**
5	97.93	91.03	**99.08**	98.85	98.39
6	99.68	99.08	97.56	**99.23**	98.39
7	80.00	80.00	96.00	96.00	**100.00**
8	99.76	100.00	100.00	99.06	99.30
9	44.44	72.22	**88.88**	33.33	61.11
10	91.08	93.02	97.02	**98.17**	97.82
11	96.24	95.79	98.46	98.55	**98.64**
12	93.44	88.20	**98.12**	97.19	97.19
13	97.28	96.73	98.36	96.73	**98.91**
14	98.77	98.41	99.91	99.82	99.91
15	97.40	86.45	99.13	99.42	**99.71**
16	97.61	80.95	100.00	97.61	100.00

Bold entities indicate the optimal value in each category.

TABLE VII
CLASSIFICATION WITH SPATIAL–SPECTRAL FEATURES ON THE KENNEDY SPACE CENTER DATASET

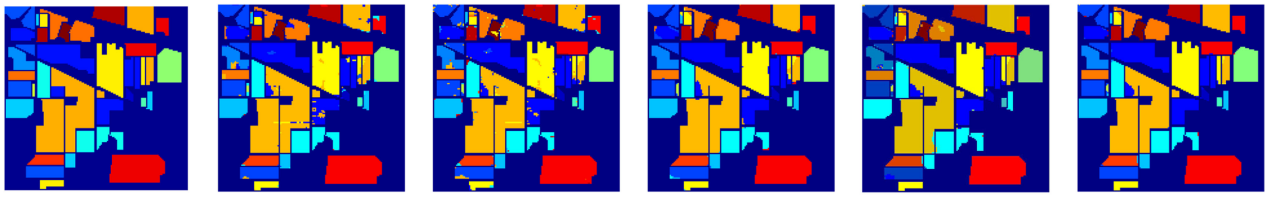
Classifier	SVM	SAE	CNN	CapsNet	Proposed
OA(%)	88.21	87.20	96.15	97.07	**98.16**
AA(%)	83.22	81.12	92.76	95.61	**97.78**
$\kappa \times 100$	86.87	85.72	95.71	96.73	**97.95**
1	88.15	90.05	**99.12**	97.51	98.24
2	82.56	77.98	97.70	99.08	**99.54**
3	79.13	86.07	92.17	97.82	**99.13**
4	64.60	61.50	80.97	81.41	**86.28**
5	87.50	75.69	79.86	95.83	**99.30**
6	82.30	73.78	92.33	**97.57**	96.60
7	45.74	48.93	73.40	79.78	**96.80**
8	81.39	93.54	**99.48**	94.05	99.22
9	87.17	94.23	100.00	100.00	99.57
10	93.93	74.10	100.00	100.00	99.44
11	97.34	87.79	93.89	**96.26**	94.96
12	92.25	92.25	97.12	**98.45**	97.12
13	100.00	100.00	100.00	100.00	100.00

Bold entities indicate the optimal value in each category.

TABLE VIII
CLASS SPECIFIC AND OVERALL CLASSIFICATION ACCURACY (%) OF DIFFRENT METHODS FOR THE HOUSTON DATA

Classifier	SVM($\mathcal{H}$)	SVM($\mathcal{H}+\mathcal{L}$)	ELM($\mathcal{H}$)	ELM($\mathcal{H}+\mathcal{L}$)	CNN($\mathcal{H}$)	CNN($\mathcal{H}+\mathcal{L}$)	Proposed($\mathcal{H}$)	Proposed($\mathcal{H}+\mathcal{L}$)
OA(%)	76.53	78.63	76.90	79.55	77.81	83.19	81.23	**86.52**
AA(%)	78.08	79.98	78.26	80.20	75.25	80.05	82.49	**87.66**
$\kappa \times 100$	74.74	77.00	74.60	77.79	75.93	81.72	79.62	**85.41**
1	82.34	82.72	81.96	**82.91**	81.95	82.43	80.07	81.10
2	81.39	81.39	83.55	83.08	84.68	**84.96**	82.71	81.02
3	95.84	95.45	92.23	**97.03**	94.25	96.83	93.66	96.44
4	91.29	91.95	91.48	91.76	91.66	92.42	**93.66**	88.35
5	96.97	98.20	98.86	99.15	98.57	99.71	99.34	**100.00**
6	79.02	79.72	79.02	79.02	74.12	72.08	88.81	**95.80**
7	72.95	74.25	77.89	78.92	53.35	79.47	82.56	**86.38**
8	38.08	45.96	35.52	47.77	66.38	79.01	70.09	**90.88**
9	75.64	77.24	73.94	76.49	73.56	81.30	71.48	**82.53**
10	63.51	62.84	59.36	59.65	56.46	64.57	62.74	**72.78**
11	77.13	78.18	76.57	83.11	**92.97**	88.61	74.95	88.99
12	68.01	78.19	72.72	80.50	85.39	**86.75**	82.61	83.09
13	60.70	62.46	60.00	57.89	2.10	2.45	66.67	**76.14**
14	93.12	93.93	91.90	91.90	86.23	91.90	**95.55**	93.93
15	95.14	95.77	93.87	93.87	87.10	**98.30**	91.75	97.46

Bold entities indicate the optimal value in each category.



(a) (b) (c) (d) (e) (f)

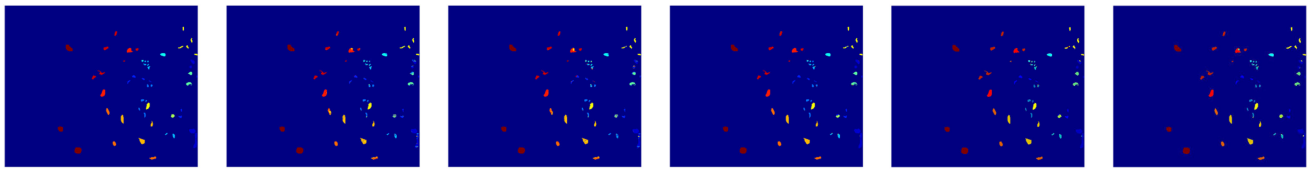
Fig. 9. Classification maps of different methods for the Indian Pines dataset. (a) Ground truth map. (b) Spatial–Spectral-SVM (95.39%). (c) Spatial–Spectral-SAE (93.81%). (d) 3-D-CNN (97.27%). (e) CapsNet (98.03%). (f) Proposed (98.57%).

Overall, the proposed CapsNet deep models provide better classification performance than other deep learning methods. As shown in Tables VI and VII, the proposed deep model performs better than the original CapsNet, which is due to that the MCC is introduced into the proposed model to handle the noise and outliers, leading to a more robust and representative deep architecture.

Tables VI–VIII present the performance of all the considered methods. From Table VI, it can be seen that the proposed framework is able to present the satisfying results when compared with other methods. Regarding the classification maps produced by the spatial–spectral classifiers in Fig. 9, the proposed method presents smoother results than the CapsNet, CNN, SVM, and SAE. For instance, we can see that the classification map produced by the proposed robust deep model [see Fig. 9(f)] exhibits less misclassified pixels than the corresponding map generated by the CapsNet [see Fig. 9(e)]. With the help of spatial and spectral information, as well as the MCC, discriminative features can still be extracted even if some samples are corrupted more or less.

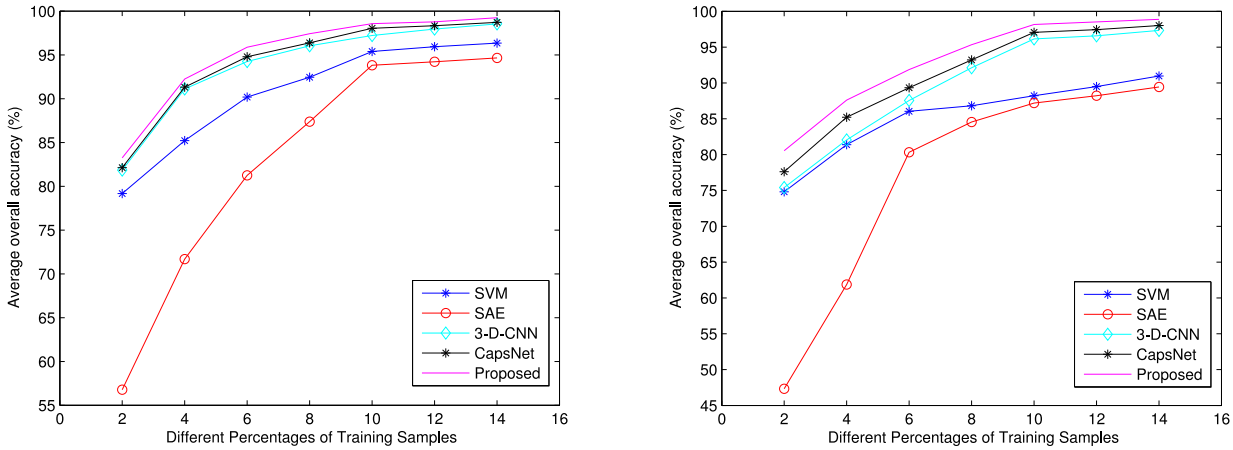
Table VII and Fig. 10 present the classification performance of all the considered methods on the KSC data. In this experiment, although the structures of deep learning models are similar, the classification results show great difference. Specifically, it can be observed that the proposed robust CapsNet model and the existing CapsNet model perform very well when compared with SVM, CNN, and SAE. The proposed model achieves the best performance, followed by the existing CapsNet method, the CNN method, and the SAE method. Specifically, the OA of the proposed model is 98.16%, which is 1.09% higher than that of the CapsNet (97.07%). In addition, the experiments under different percentages of training samples are also conducted to analyze the performance of the proposed method. The 2%, 4%, 6%, 8%, 10%, 12%, and 14% percentages samples of the Indian Pines and Kennedy Space Center datasets are randomly selected for training. The OA curves of all the considered methods are given in Fig. 11, from which it is obvious that the proposed robust 3-D-CapsNet can obtain the best classification performance under all percentages of training samples for these two HSI datasets.

Table VIII presents the classification performance of the University of Houston dataset, which indicates that the joint use of spatial, spectral, and elevation information derived from the HSI and LiDAR data could lead to better classification performance than the individual use of a single data source. Several traditional classifiers, e.g., SVM and ELM [62], and the recently developed two-channel CNN are used to compare with our proposed method. For the clarity, some notations are defined hereafter: HSI data are represented symbolically by $\mathcal{H}$, HSI and LiDAR data concatenated together are represented as $\mathcal{H}+\mathcal{L}$. From Table VIII, it can be seen that the proposed robust fusion framework could obtain the best classification performance for Houston datasets. As for the Houston data that presents a complex urban area, the proposed method can also produce the best results, offering 3.33% gain in OA when compared to the two-channel CNN. For qualitative evaluation of the classification performance, visual maps are illustrated in Fig. 12. The proposed method produces the most accurate and noiseless classification maps. There are two main reasons for this



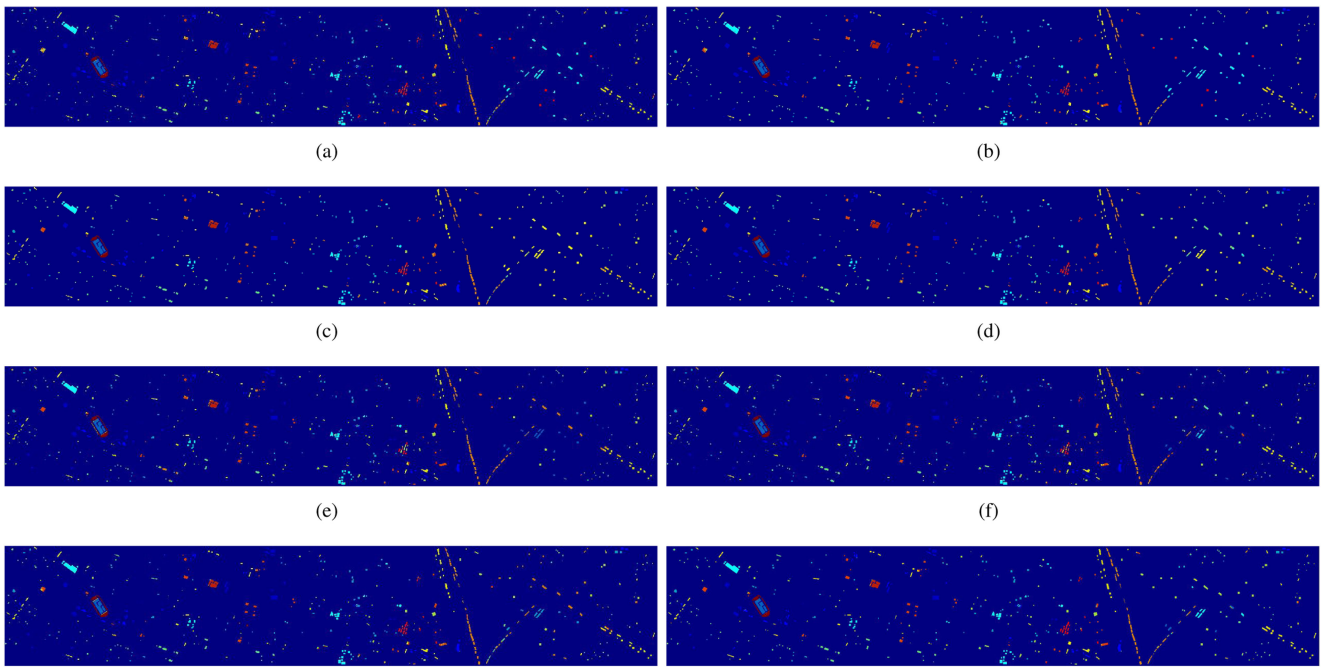
(a) (b) (c) (d) (e) (f)

Fig. 10. Classification maps of different methods for the Kennedy Space Center dataset. (a) Ground truth map. (b) Spatial–Spectral-SVM (88.21%). (c) Spatial–Spectral-SAE (87.20%). (d) 3-D-CNN (96.15%). (e) CapsNet (97.07%). (f) Proposed (98.16%).



(a) (b)

Fig. 11. Classification maps of all the considered methods with different percentages of training samples for the two datasets. (a) Indian Pines. (b) Kennedy Space Center.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 12. Classification maps of different methods for the Houston dataset. (a) SVM($\mathcal{H}$) (76.50%). (b) SVM($\mathcal{H}+\mathcal{L}$) (78.60%). (c) ELM($\mathcal{H}$) (77.30%). (d) ELM($\mathcal{H}+\mathcal{L}$) (79.20%). (e) CNN($\mathcal{H}$) (77.77%). (f) CNN($\mathcal{H}+\mathcal{L}$) (83.19%). (g) Proposed($\mathcal{H}$) (81.23%). (h) Proposed($\mathcal{H}+\mathcal{L}$) (86.52%).

superior performance. On the one hand, this improvement could be due to the benefit of invariant features learned by high-level CapsNet FE on HSI and LiDAR, which are important for the following fusion and classification steps. On the other hand, our dual channel fusion scheme takes advantage of MCC to handle the noise and outliers and further to fuse multisensor features in a more robust and effective way.

V. Conclusion

In this article, a robust 3-D-CapsNet architecture has been proposed for HSI classification, which introduces the MCC to address the noise and outliers problem, generating a robust and strong generalization model. Moreover, a novel MCC-based dual channel robust CapsNet framework has also been proposed for pixelwise classification with fusing multisource remote sensing data, e.g., HSI and LiDAR. The proposed dual channel CapsNet model contains two different channels which consists of the same architecture of 2-D-CapsNet and 3-D-CapsNet to extract the elevation information and spatial–spectral information, respectively. Experimental results have demonstrated the superiority of our proposed deep learning models when compared with SVM, SAE, CNN, and CapsNet.

Acknowledgment

The authors would like to thank the Editor-in-Chief, the Associate Editor, and the anonymous reviewers for their careful reading and valuable comments, which greatly helped them to improve technical quality and presentation of this article.

References

[1] Q. Y. Xie *et al.*, "Leaf area index estimation using vegetation indices derived from airborne hyperspectral images in winter wheat," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3586–3594, Aug. 2014.

[2] R. J. Murphy, S. Schneider, and S. T. Monteiro, "Consistency of measurements of wavelength position from hyperspectral imagery: Use of the ferric iron crystal field absorption at 900 nm as an indicator of mineralogy," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2843–2857, May 2014.

[3] J. M. B.-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[4] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[5] W. W. Sun *et al.*, "Nonlinear dimensionality reduction via the ENH-LTSA method for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 2, pp. 375–388, Feb. 2014.

[6] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, Apr. 2012.

[7] N. H. Ly, Q. Du, and J. E. Fowler, "Sparse graph-based discriminant analysis for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 7, pp. 3872–3884, Jul. 2014.

[8] W. He, H. Y. Zhang, L. P. Zhang, W. Philips, and W. Z. Liao, "Weighted sparse graph based dimensionality reduction for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 686–690, May 2016

[9] Z. H. Xue, P. J. Du, J. Li, and H. J. Su, "Simultaneous sparse graph embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6114–6132, Nov. 2015.

[10] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, Oct. 2008.

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

[12] X. H. Cao, T. Xiong, and L. C. Jiao, "Supervised band selection using local spatial information for hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 329–333, Mar. 2016.

[13] I. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer-Verlag, 1986.

[14] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.

[15] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, pp. 1027–1061, May 2007.

[16] W. W. Sun *et al.*, "UL-Isomap based nonlinear dimensionality reduction for hyperspectral imagery classification," *ISPRS J. Photogrammetry Remote Sens.*, vol. 89, pp. 25–36, Mar. 2014.

[17] Y. Xu, Q. Du, W. Li, C. Chen, and N. H. Younan, "Nonlinear classification of multispectral imagery using representation-based classifiers," *Remote Sens.*, vol. 9, 2017, Art. no. 662.

[18] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[19] X. F. He and P. Niyogi, "Locality preserving projections," in *Proc. Adv. Neural Inf. Process. Syst.*, 2003, Vancouver, BC, Canada, vol. 16, pp. 234–241.

[20] X. F. He, D. Cai, S. C. Yan, and H. J. Zhang, "Neighborhood preserving embedding," in *Proc. IEEE 10th Int. Conf. Comput. Vision*, Beijing, China, 2005, pp. 1208–1213.

[21] J. T. Peng, Y. C. Zhou, and C. L. P. Chen, "Region-kernel-based support vector machines for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4810–4824, Sep. 2015.

[22] W. Li, S. Prasad, and J. E. Fowler, "Decision fusion kernel-induced spaces for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3399–3411, Jun. 2014.

[23] J. Yang, A. F. Frangi, J.-Y. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: A complete kernel Fisher discriminant framework for feature extraction and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 230–244, Feb. 2005.

[24] W. Li and Q. Du, "Laplacian regularized collaborative graph for discriminant analysis of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7066–7076, Dec. 2016.

[25] L. Pan, H.-C. Li, W. Li, X.-D. Chen, G.-N. Wu, and Q. Du, "Discriminant analysis of hyperspectral imagery using fast kernel sparse and low-rank graph," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6085–6098, Nov. 2017.

[26] X. Huang *et al.*, "Multiple morphological profiles from multicomponentbased images for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4653–4669, Dec. 2014.

[27] X. D. Kang, X. L. Xiang, S. T. Li, and J. A. Benediktsson, "PCA-based edge-preserving features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7140–7151, Dec. 2017.

[28] L. Pan, H.-C. Li, Y.-J. Deng, F. Zhang, X.-D. Chen, and Q. Du, "Hyperspectral dimensionality reduction by tensor sparse and low-rank graph-base discriminant analysis," *Remote Sens.*, vol. 9, 2017, Art. no. 452.

[29] B. Chen, G. Polatkan, G. Sapiro, D. Blei, D. Dunson, and L. Carin, "Deep learning with hierarchical convolutional factor analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1887–1901, Aug. 2013.

[30] H. Goh, N. Thome, M. Cord, and J.-H. Lim, "Learning deep hierarchical visual feature coding," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 12, pp. 2212–2225, Dec. 2014.

[31] T.-H. Chan, K. Jia, S. H. Gao, J. W. Lu, Z. N. Zeng, and Y. Ma, "PCANet: A simple deep learning baseline for image classification?" *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5017–5032, Dec. 2015.

[32] B. Du, W. Xiong, J. Wu, L. F. Zhang, L. P. Zhang, and D. C. Tao, "Stacked convolutional denoising auto-encoders for feature representation," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1017–1027, Apr. 2017.

[33] W. H. Diao, X. Sun, X. W. Zheng, F. Z. Dou, H. Q. Wang, and K. Fu, "Efficient saliency-based object detection in remote sensing images using deep belief networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 2, pp. 137–141, Feb. 2016.

[34] S. H. Mei, J. Y. Ji, J. H. Hou, X. Li, and Q. Du, "Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4520–4533, Aug. 2017.

[35] L. C. Jiao, M. M. Liang, H. Chen, S. Y. Yang, H. Y. Liu, and X. H. Cao, "Deep fully convolutional network-based spatial distribution prediction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5585–5599, Oct. 2017.
[36] Y. S. Chen, Z. H. Lin, X. Zhao, G. Wang, and Y. F. 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.
[37] X. R. Ma, H. Y. Wang, and J. Geng, "Spectral-spatial classification of hyperspectral image based on deep auto-encoder," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4073–4085, Sep. 2016.
[38] Y. S. Chen, X. Zhao, and X. P. 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.
[39] A. Romero, C. Gatta, and G. Camps-Valls, "Unsupervised deep feature extraction for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1349–1362, Mar. 2016.
[40] W. Hu, Y. Y. Huang, W. Li, F. Zhang, and H. C. Li, "Deep convolutional neural networks for hyperspectral image classification," *J. Sensors*, vol. 2015, doi: 10.1155/2015/258619.
[41] Y. S. Chen, H. L. Jiang, C. Y. Li, X. P. 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.
[42] 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.
[43] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. J. Plaza, and F. Pla, "Deep pyramidal residual networks for spectral-spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 740–754, Feb. 2019.
[44] M. E. Paoletti, J. M. Haut, J. Plaza, and A. Plaza, "Deep&dense convolutional neural network for hyperspectral image classification," *Remote Sens.*, vol. 10, no. 9, 2018, Art. no. 1454.
[45] 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.
[46] X. Kang, B. Zhuo, and P. Duan, "Dual-path network-based hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 447–451, Mar. 2019.
[47] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. 31st Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 1–11.
[48] F. Deng, S. L. Pu. Ran, X. H. Chen, Y. S. Shi, T. Yuan, and S. Y. Pu, "Hyperspectral image classification with capsule network using limited training samples," *Sensors.*, vol. 18, no. 9, pp. 3153–3175, Sep. 2018.
[49] J. H. Yin, S. Li, H. M. Zhu, and X. Y. Luo, "Hyperspectral image classification using CapsNet with well-initialized shallow layers," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1095–1099, Jul. 2019.
[50] M. E. Paoletti *et al.*, "Capsule networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2145–2160, Apr. 2019.
[51] X. Wang, K. Tao, and Y. Chen, "CapsNet and triple-gans towards hyperspectral classification," in *Proc. 5th Int. Workshop Earth Observ. Remote Sens. Appl.*, Xi'an, 2018, pp. 1–4, doi: 10.1109/EORSA.2018.8598574.
[52] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 178–188, Jan. 2016
[53] C. Debes *et al.*, "Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, Nov. 2014.
[54] J. T. Mundt, D. R. Streutker, and N. F. Glenn, "Mapping sagebrush distribution using fusion of hyperspectral and LiDAR classifications," *Photogramm. Eng. Remote Sens.*, vol. 72, no. 1, pp. 47–54, Nov. 2006.
[55] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of hyperspectral and LIDAR remote sensing data for classification of complex forest areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1416–1427, May. 2008.
[56] P. Ghamisi, B. Hofle, and X. X. Zhu, "Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 3011–3024, Jun. 2017.
[57] X. D. Xu, W. Li, Q. Ran, Q. Du, L. Gao, and B. Zhang, "Multisource remote sensing data classification based on convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 937–949, Feb. 2018.
[58] W.-Y. Wang, H.-C. Li, L. Pan, G. Yang, and Q. Du, "Hyperspectral image classification based on capsule network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 3571–3574.
[59] K. Fukushima, "Neocognitron: A hierarchial neural network capable of visual pattern recognition," *Neural Netw.*, vol. 1, no. 2, pp. 119–130, 1988.
[60] W. Liu, P. P. Pokharel and J. C. Principe, "Correntropy: Properties and applications in non-gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.
[61] R. He, W.-S. Zheng, and B.-G. Hu, "Maximum correntropy criterion for robust face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1561–1576, Aug. 2011.
[62] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3681–3693, Jul. 2015.

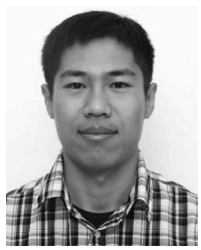
Heng-Chao Li (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in information and communication engineering from Southwest Jiaotong University, Chengdu, China, in 2001 and 2004, respectively, and the Ph.D. degree in information and communication engineering from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2008.

From 2013 to 2014, he was a Visiting Scholar with Prof. W. J. Emery with the University of Colorado at Boulder, Boulder, CO, USA. He is currently a Professor with the Sichuan Provincial Key Laboratory of Information Coding and Transmission, Southwest Jiaotong University. His research interests include the statistical analysis of synthetic aperture radar images, remote sensing image processing, and signal processing in communications.

Dr. Li was the recipient of several scholarships or awards, including the Special Grade of the Financial Support from the China Postdoctoral Science Foundation, in 2009, and the New Century Excellent Talents in University from the Ministry of Education of China, in 2011. He serves as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING. He has been a Reviewer for several international journals and conferences, such as the IEEE TRANSACTIONS GEOSCIENCE AND REMOTE SENSING, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, the IEEE TRANSACTIONS ON IMAGE PROCESSING, *IET Radar, Sonar & Navigation*, *IET Signal Processing*, *IET Image Processing*, *Pattern Recognition*, the *International Journal of Remote Sensing*, *Remote Sensing*, and the *Canadian Journal of Remote Sensing*.

Wei-Ye Wang received the B.Sc. degree in computer science and technology in 2014 from Southwest Jiaotong University, Chengdu, China, where he is currently working toward the Ph.D. degree in signal and information processing with the School of Information Science and Technology.

His research interests include hyperspectral image classification, machine learning, and deep learning.



Lei Pan received the B.Sc. degree in communication engineering from the Shandong University of Science and Technology, Qingdao, China, in 2010, the M.Sc. degree in communication and information system in 2013, and the Ph.D. degree in signal and information processing from the School of Information Science and Technology from the Southwest Jiaotong University, Chengdu, China, in 2019.

His research interests include remote sensing image analysis and processing.

Wei Li (Senior Member, IEEE) received the B.Sc. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2007, the M.Sc. degree in information science and technology from Sun Yat-sen University, Guangzhou, China, in 2009, and the Ph.D. degree in electrical and computer engineering from Mississippi State University, Starkville, MS, USA, in 2012.

Subsequently, he spent one year as a Postdoctoral Researcher with the University of California, Davis, CA, USA. He was a Professor with the College of Information Science and Technology at Beijing University of Chemical Technology, Beijing, China, from 2013 to 2019. He is currently a Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing. His research interests include hyperspectral image analysis, pattern recognition, and data compression.

Dr. Li was the recipient of the 2015 Best Reviewer Award from IEEE Geoscience and Remote Sensing Society (GRSS) for his service for IEEE IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS) and the Outstanding Paper Award at IEEE International Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (Whispers), 2019. is currently serving as an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS and the IEEE JSTARS. He served as Guest Editor for special issue of *Journal of Real-Time Image Processing*, *Remote Sensing*, and IEEE JSTARS.

Qian Du (Fellow, IEEE) received the B.S. and M.S. degrees from the Beijing Institute of Technology, and the M.S. and Ph.D. degrees from the University of Maryland at Baltimore, Baltimore, MD, USA, in 1998 and 2000, respectively, all in electrical engineering.

She is currently the Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. She is also an Adjunct Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks.

Dr. Du is a fellow of the SPIE-International Society for Optics and Photonics. She was the recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society (GRSS). She has served as the Co-Chair for the Data Fusion Technical Committee of the IEEE GRSS from 2009 to 2013. She was the Chair with the Remote Sensing and Mapping Technical Committee of International Association for Pattern Recognition, from 2010 to 2014. She was the General Chair for the fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Shanghai, in 2012. She served as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the *Journal of Applied Remote Sensing*, and the IEEE SIGNAL PROCESSING LETTERS. Since 2016, she has been the Editor-in-Chief for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Ran Tao (Senior Member, IEEE) received the B.Sc. degree from the Electronic Engineering Institute of PLA, Hefei, China, in 1985, and the M.Sc. and Ph.D. degrees from the Harbin Institute of Technology, Harbin, China, in 1990 and 1993, respectively. He has been a Senior Visiting Scholar with the University of Michigan, Ann Arbor, MI, USA, and the University of Delaware, Newark, DE, USA, in 2001 and 2016, respectively. He has authored or coauthored three books and more than 100 peer-reviewed journal articles. He is currently a Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China. His current research interests include fractional Fourier transform and its applications, theory, and technology for radar and communication systems.

Dr. Tao is a fellow with the Institute of Engineering and Technology and the Chinese Institute of Electronics. He was the recipient of the National Science Foundation of China for Distinguished Young Scholars, in 2006, and a Distinguished Professor of Changjiang Scholars Program, in 2009. He was a recipient of the First Prize of Science and Technology Progress in 2006 and 2007, and the First Prize of Natural Science in 2013, both awarded by the Ministry of Education. He has been a Chief-Professor of the Creative Research Groups with the National Natural Science Foundation of China, since 2014, and he was a Chief-Professor of the Program for Changjiang Scholars and Innovative Research Team in university, during 2010–2012. He is currently the Vice-Chair of the IEEE China Council and the International Union of Radio Science (URSI) China Council and a member of Wireless Communication and Signal Processing Commission, URSI.