Abstract—With the increasing availability and resolution of satellite sensor data, multispectral (MS) and panchromatic (PAN) images are the most popular data that are used in remote sensing among applications. This article proposes a novel cross-resolution hidden layer feature fusion (CRHFF) approach for joint classification of multiresolution MS and PAN images. In particular, shallow spectral and spatial features at a global scale are first extracted from an MS image. Then, deep cross-resolution hidden layer features extracted from MS and PAN are fused from patches at a local scale according to an autoencoder (AE)-like deep network. Finally, the selected multiresolution hidden layer features are classified in a supervised manner. By taking advantage of integrated shallow-to-deep and global-to-local features from the high-resolution MS and PAN images, the cross-resolution latent information can be extracted and fused in order to better model imaged objects from the multimodal representation and finally increase the classification accuracy. Experimental results obtained on three real multiresolution datasets covering complex urban scenarios confirm the effectiveness of the proposed approach in terms of higher accuracy and robustness with respect to literature methods.

Index Terms—Classification, feature-level fusion, multiresolution images, remote sensing, shallow and deep features.

I. INTRODUCTION

NOWADAYS, due to the increasing satellite sensor data availability and quality, Earth’s land-cover/use change detection and classification at a fine resolution have received more attention [1]–[4]. Many emerging new applications in urban, agriculture, disaster, and forestry fields require the full use and fusion of multisource or multitemporal remote sensing images in order to exploit complementary information and promote identification accuracy [5]–[10]. In the current scenario, there are many Earth observation satellites that can simultaneously acquire multimodal images in the same scene, among which the multispectral (MS) and panchromatic (PAN) images are the most widely used data for fusion in the practical applications. Different from the traditional moderate-resolution satellites, such as Landsat or SPOT families, newly launched satellites can acquire high-resolution (HR) and even very-high-resolution (VHR) MS and PAN images. This provides a great opportunity as well as challenges to properly integrate the multiresolution data, promoting their applications at a fine level. Table I lists some examples of these satellites (with a spatial resolution of MS image higher than 5 m) and their corresponding parameters.

<table>
<thead>
<tr>
<th>Country</th>
<th>Satellites</th>
<th>MS (meter/pixel)</th>
<th>PAN</th>
<th>Launch Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>IKONOS</td>
<td>4</td>
<td>1</td>
<td>1999</td>
</tr>
<tr>
<td></td>
<td>QuickBird</td>
<td>2.44</td>
<td>0.61</td>
<td>2001</td>
</tr>
<tr>
<td></td>
<td>GeoEye-1</td>
<td>1.65</td>
<td>0.41</td>
<td>2008</td>
</tr>
<tr>
<td></td>
<td>WorldView-2</td>
<td>1.84</td>
<td>0.46</td>
<td>2009</td>
</tr>
<tr>
<td></td>
<td>WorldView-3,-4</td>
<td>1.24</td>
<td>0.31</td>
<td>2014, 2016</td>
</tr>
<tr>
<td></td>
<td>Planet Labs</td>
<td>5</td>
<td>3</td>
<td>2014</td>
</tr>
<tr>
<td>Spain</td>
<td>DEIMOS-2</td>
<td>4</td>
<td>1</td>
<td>2013</td>
</tr>
<tr>
<td>France</td>
<td>Pleiades-1,-2</td>
<td>2</td>
<td>0.5</td>
<td>2011, 2012</td>
</tr>
<tr>
<td>UK</td>
<td>TripleSat</td>
<td>3.2</td>
<td>0.8</td>
<td>2015</td>
</tr>
<tr>
<td>South Korea</td>
<td>KOMPAGT-3</td>
<td>2.8</td>
<td>0.7</td>
<td>2012</td>
</tr>
<tr>
<td>Korea</td>
<td>KOMPSAT-3A</td>
<td>2.2</td>
<td>0.55</td>
<td>2015</td>
</tr>
<tr>
<td>China</td>
<td>Gaofen-2</td>
<td>4</td>
<td>1</td>
<td>2014</td>
</tr>
<tr>
<td></td>
<td>Gaofen-6</td>
<td>8</td>
<td>2</td>
<td>2018</td>
</tr>
<tr>
<td></td>
<td>Gaofen-7</td>
<td>3.2</td>
<td>0.8</td>
<td>2019</td>
</tr>
<tr>
<td></td>
<td>Super-View-1</td>
<td>2</td>
<td>0.5</td>
<td>2016, 2018</td>
</tr>
<tr>
<td></td>
<td>JiLin-1</td>
<td>2.88</td>
<td>0.72</td>
<td>2015</td>
</tr>
</tbody>
</table>

TABLE I
EXAMPLES OF EO SATELLITES THAT SIMULTANEOUSLY ACQUIRE MS AND PAN IMAGES
Usually, MS sensors collect data in red, green, blue, and near-infrared four bands, with relatively lower spatial resolution compared to PAN sensors, due to physical limitations and technical constraints of onboard storage and bandwidth transmission [11]. In contrast, a PAN image with only a single broadband has a much higher spatial resolution [12]. According to the MS image is usually used for identifying different types of land objects, whereas the PAN image can accurately describe the geometrical properties of objects, which is of great benefit to image interpretation at HR. The fusion of these two images can utilize both spatial and spectral information to further increase identification capability [13]. However, with the unprecedentedly increased spatial resolution, especially for the HR and VHR PAN/MS images that reach a submeter/meter level, their effective fusion becomes a very important yet challenging task.

In general, methods for fusing MS and PAN images include pixel-level fusion that is known as pan-sharpening (PS) techniques and feature-level fusion methods. For the former, traditional algorithms include component substitution (CS), multiscale decomposition-based method, hybrid method, and model-based algorithms [14]. In [13], five PS algorithms (i.e., Gram–Schmidt (GS), principal component analysis, high-pass filter, wavelet transform, and generalized intensity–hue–saturation) and decision fusion were designed, and their impacts on the performance of change detection were compared and analyzed. In [15], eight advanced PS methods, including various state-of-the-art and advanced deep learning (DL) methods, were studied through the task of anomaly detection. In recent years, DL methods have also been used for PS [16]–[18]. The basic idea is to train a PS model between the fused image and the observations based on a DL architecture, and then, the model is used to construct the final fused MS image. However, the PS process inevitably introduces spectral and spatial distortions in the resultant fused MS image, which influences the final detection or classification results [19]. Despite a DL-based PS method may achieve desirable results with less spectral distortion, it requires more prior data to train a robust network [20]. For the latter, feature-level fusion methods first extract representative features from MS and PAN images and then integrate these features via a robust fusion model for further classification or detection. Accordingly, they are more straightforward for applications without producing a pan-sharpened image and avoid the limitations of pixel-level fusion methods to some extent. In [21], texture features (i.e., homogeneity, contrast, and entropy) were extracted from PAN images using the co-occurrence matrix, and spectral features (i.e., normalized band values) were calculated from MS images. Then, object-based classification using the standard nearest neighbor was applied as a fusion analysis for forest-type classification. In [22], a graph cut method was combined with the linear mixture model, and MS and PAN data were integrated to generate a context classification map. In [23], a unified Bayesian framework was presented to iteratively discover semantic segments from a PAN image and inferring cluster labels for the segments from an MS image to obtain the classification maps.

The above feature-level fusion methods mainly focus on artificial features that require a domain expert’s knowledge. On the contrary, DL-based techniques that can automatically learn abstract and robust deep features from the original data are becoming a very promising way for dealing with the fusion of MS and PAN images at feature level. Within this context, in [24], a superpixel-based multiple local network model was proposed to classify the MS image; then, a PAN image was used to fine-tune the classification results. In [25], a stacked autoencoder (AE) was used to extract the spectral features from an MS image, and a convolutional neural network (CNN) was used to extract spatial features from a PAN image; then, spectral and spatial features were concatenated to obtain final classification results. In [26], two CNN modules inspired by the VGG model were designed for MS and PAN images at their original resolution; then, they were combined to perform land-cover classification. In [27], a novel framework was proposed via 3-D and 2-D adaptive multiscale convolutional networks and a perceptual loss function for MS and PAN images classification. In [28], based on a data-driven DL, a spatial attention module (SA-module) for PAN images and a channel attention module (CA-module) for MS images were designed to extract the features that were then fused. In [29], a local spatial attention module (LSA-module) for the PAN image and a global channel attention module (GCA-module) for the MS image were designed; then, an interaction module effectively reduced the differences in the characteristics obtained by the PAN branch and the MS branch. Then, the GCA-module was used to further enhance feature representation from the fused features for classification.

The above existing feature-level fusion methods are proven to potentially outperform pixel-level fusion methods. However, there are still some open issues that require further investigation.

1) In the current DL-based methods, the patch is fixed as a rectangle; thus, the spatial integrity and connectivity of land objects may be mishandled.
2) In the two-branch deep network fusion methods (e.g., in [25], [26], [28], and [29]), high-level abstract features in the last layer of each branch are concatenated. Thus, the cross-resolution representation in the middle-layer features is ignored. How to properly extract and utilize the intermediate cross-resolution information has not been fully investigated.
3) In feature-level fusion methods (e.g., in [25]), the multi-resolution MS and PAN images require a resampling operation, which will inevitably introduce interpolation errors and increase data processing burden.

By considering the aforementioned open issues, in this article, we propose a cross-resolution hidden layer feature fusion (CRHFF) approach to HR/VHR MS and PAN image classification. To the best of our knowledge, there is no similar work in the literature that deals with the same task. The main contributions of this article are highlighted as follows.

1) By taking advantage of shallow-to-deep integration and global-to-local features in the proposed CRHFF approach, the inconsistent feature representation...
problem of the local patches can be solved, where the objects can be modeled in a more comprehensive way while increasing the classification accuracy. Moreover, shallow-to-deep feature extraction procedure is designed in an unsupervised and automatic fashion, which makes it very interesting for practical applications.

2) The spatial and spectral information in MS and PAN images can be neutralized through the novel AE-like deep network. Intermediate hidden layer features at different resolutions are fused using a multibranch CNN. This leads to a more detailed and precise cross-resolution feature representation than the traditional PS or features stacking, thus further enhancing the classification performance.

3) Different from the conventional operation where an MS image is first upsampled, in this work, the proposed architecture is built by considering the low-resolution shallow features from the MS image as input and the HR PAN image as output. The cross-resolution conversion is made during the process of network training, where the bias and computational burden introduced by the upsampling process are significantly reduced.

The rest of this article is organized as follows. The proposed CRHFF approach is described in detail in Section II. Datasets used in experiments are introduced in Section III. Experimental results and the related analysis are presented in Section IV. Finally, Section V draws the conclusions.

II. PROPOSED FEATURE-LEVEL FUSION APPROACH

The proposed CRHFF approach aims to extract and fuse the global-to-local and shallow-to-deep features hidden in multiresolution MS and PAN images for classification. Fig. 1 shows its block diagram that mainly consists of three steps: 1) shallow spectral–spatial feature extraction at a global scale; 2) deep multihidden layer feature extraction at a local scale; and 3) cross-resolution feature fusion and classification.

A. Step 1: Shallow Spectral–Spatial Feature Extraction

Features used for image classification can be shallow features (i.e., artificial features extracted from the original image by some specific image processing operations) or deep features (i.e., extract from a deep network by DL approaches) [30]. In order to take full advantages of the context information in the HR/VHR images, shallow spatial features are usually considered, such as the extended multivariate profile (EAP) [31], [32], extinction profile [33], edge-preserving filtering features [34], Gabor features [35], and superpixel-guided filter features [36]. In the proposed CRHFF approach, we selected EAP as an example of shallow spatial features, which are combined with the original MS bands as extended shallow spectral features. Note that such shallow spectral–spatial features focus on the global representation of image objects and consider their integrity and connectivity, which will benefit the deep local information extraction and fusion in the deep feature generation step. Other effective shallow features can also be integrated in the proposed framework.

As shown in Step 1 in Fig. 1, we define the size of input MS and PAN images as $H \times D \times c$ and $nH \times nD \times 1$, respectively, where $H$ and $D$ represent the height and weight of the MS image, respectively, $c$ is the number of MS bands, $n$ is the resolution ratio between PAN and MS images, and $t$ represents the number of EAP features based on all MS bands, which are described in detail as follows.

Attribute profiles (APs) are an extension of the widely used morphological profiles (MPs). The AP operation replaces the structural elements of traditional morphological operation with general attribute criteria, which can reflect the structural characteristics of objects more effectively. In particular, APs are obtained by processing a scalar grayscale image $a_c$, according to a criterion $T$, with $m$ attribute thickening ($\gamma^T_m$) operators and $m$ attribute thinning ($\gamma^T_m$) operators, instead of the conventional morphological filters by reconstruction [32]

$$AP(a) = \{\phi^T_m(\alpha), \phi^T_{m-1}(\alpha), \ldots, \phi^T_1(\alpha), \alpha, \gamma^T_1(\alpha), \ldots, \gamma^T_{m-1}(\alpha), \gamma^T_m(\alpha)\}.$$ (1)

Extended attribute profiles (EAPs) are built based on APs, and EAP is the combination of different EAPs [31], [32]. In this work, we compute the APs on each band of the MS image, so the corresponding EAP can be expressed as

$$EAP = \{AP(g_1), AP(g_2), \ldots, AP(g_c)\}$$ (2)

where $g_1, g_2, \ldots, g_c$ are the MS bands. In particular, in this work, the following four attributes are selected in APs: 1) area of the regions $a$; 2) length of the diagonal of the box bounding the region $d$; 3) first moment invariant of Hu, moment of inertia $i$; and 4) standard deviation of the gray-level values of the pixels in the regions $s$. For each individual AP, EAP can be expressed as $EAP_a$, $EAP_d$, $EAP_i$, and $EAP_s$. Then, the final EAP can be formulated as

$$EAP = \{EAP_a, EAP_d, EAP_i, EAP_s\}.$$ (3)

B. Step 2: Deep Multihidden Layer Feature Extraction

Differently from the simple concatenation of MS and PAN features based on a two-branch structure deep network used in the literature, we extract cross-resolution latent features of MS and PAN images through an end-to-end deep network using an AE architecture. The AE network was first proposed to reduce data dimensionality [37]. The architecture of an AE involves an encoder and a decoder. The former converts the input into a hidden representation that only keeps the most representative information, and the latter recovers the input data from the hidden representation. Accordingly, the hidden representation can be viewed as the input features for the reconstruction. In recent years, the AE networks have been widely applied to image super-resolution [38], [39] and PS [40].

In the original AE, the input and recovered data are exactly the same. In the proposed AE-like deep network, the input data are the patches derived from the EAP, while the recovered data are the patches derived from the PAN image. In the process of training AE-like deep network, low-resolution patches of EAP [denoted as $x$(Patch$_{1}$,EMAP)] are automatically aligned to the size of HR patches of PAN image [denoted
Fig. 1. Block diagram of the proposed CRHFF approach.

as $x(Patches_{PAN})$. In Step 2 of Fig. 1, the conversion between the multiresolution $x(Patches_{EMAP})$ and $x(Patches_{PAN})$ is illustrated in detail. Each pair of $x(Patches_{EMAP})$ ($R \times R \times t$) and $x(Patches_{PAN})$ ($nR \times nR \times 1$) is acquired over the same area to ensure the high correlation between the two types of source data. Let us assume $\hat{x}(Patches_{EMAP})$ denotes the reconstructed data of $x(Patches_{EMAP})$ through convolutional layers and upsampling layers. The energy function of the
reconstruction error is defined as

$$J = \frac{1}{K} \sum_{j=1}^{k} \left[ (x(Patches_{SPAN}))^j - (x(Patches_{EMAP}))^j \right]^2$$  \hspace{1cm} (4)$$

where $k$ is the number of $x(Patches_{EMAP})$.

Specifically, there are six conv_block, two upsampling layers, and one convolutional layer in Step 2, where conv_block contains the convolutional layer, the batch normalization (BN), and an activation function. Parameters of kernel and feature maps are $[3 \times 3]:128$, which means that the kernel size of the convolution is $3 \times 3$, and 128 feature maps are generated. Note that resolutions are different between $x(Patches_{EMAP})$ and $x(Patches_{SPAN})$, and thus, two upsampling layers are used after conv_block with an upsampling factor for rows and columns of $2 \times 2$. Thus, $x(Patches_{EMAP})$ are resampled into a unified spatial resolution of PAN image, whereas the hidden layers contain different resolution features.

There are two kinds of convolution kernels (i.e., $3 \times 3$ and $5 \times 5$). In the low- and medium-resolution feature maps, a $3 \times 3$ convolution kernel is used, while a $5 \times 5$ convolution kernel is used in the last HR feature maps. This is because usually low-resolution feature maps have small size patches, whereas HR feature maps have larger size patches. Since the dimensionality of feature maps from the last convolutional layer is different from the single-band PAN image, the last convolutional layer with one feature map is used to make them consistent.

This deep multihidden layer feature extraction method is unsupervised. Here, the number of training samples of the network is also the number of patches, rather than labeled samples of classes. In order to enlarge the number of training samples, each pixel in EMAP is used to generate training patches, and zero values are filled up for boundaries of EMAP.

C. Step 3: Cross-Resolution Feature Fusion and Classification

The deep hidden layer features extracted between $x(Patches_{EMAP})$ and $x(Patches_{SPAN})$ in Step 2 represent distinct latent features cross different scales in two data. Lower resolution feature maps close to $x(Patches_{EMAP})$ contain more homogeneous spectral–spatial information, which is beneficial to identify pixels in the same object. Higher resolution feature maps close to $x(Patches_{SPAN})$ contain more detailed spatial features, which are useful for fine classification. In order to take advantage of multi-resolution deep hidden features, they are fused in Step 3, as shown in Fig. 1. In particular, we choose three hidden layers, i.e., layers 2, 5, and 8, of size $R \times R$, $2R \times 2R$, and $4R \times 4R$, respectively. Three-parallel CNN modules are designed to extract deep features at different resolutions. Then, the three deep feature sets are connected followed by two dense layers, and the SoftMax classifier is applied to perform the final classification.

The convolution process for the parallel CNN modules is described as follows. Let $F_{\text{hidden},u}(u = 1, 2, 3)$ be the input features of the three CNN modules. The output in the $l$th layer can be written as

$$Z_l = \begin{cases} W_l \otimes F_{\text{hidden},u} + b^l, & l = 1 \\ W_l \otimes Z_{l-1} + b^l, & l = 2, \ldots, p \end{cases}$$  \hspace{1cm} (5)$$

where $W$ is the weight, $b$ is the bias, and $\otimes$ denotes the convolution operation. Then, a BN layer to accelerate network convergence and mitigate gradient explosion or vanishing problem is added over the output $Z^l$; it can be denoted as $\text{BN}(Z^l)$. Before importing $\text{BN}(Z^l)$ into the next block, a rectifier linear unit (ReLU) activation function is implemented

$$\text{ReLU}(\text{BN}(Z^l)) = \max(0, \text{BN}(Z^l)).$$  \hspace{1cm} (6)$$

Output features in the last fully connected layer are then transformed into a probability distribution for specific categories, where the cross entropy is used to measure the prediction loss of the network. To minimize the loss function, the stochastic gradient descent algorithm is adopted to update parameters and to optimize the model.

It is worth noting that unlike Step 2 that is unsupervised, Step 3 is a supervised process. As shown in Fig. 1, three-parallel CNN modules are similar, and the only difference is the number of maxpooling layers. Considering different patch sizes of hidden features, there are no maxpooling layers in branch 1 (hidden layer 2), one maxpooling layer in branch 2 (hidden layer 5), and two maxpooling layers in branch 3 (hidden layer 8). Accordingly, after several convolutional layers, the size of the feature maps in different branches remains the same.

As in Step 2, conv_block contains a convolutional layer, a BN, and an activation function. Parameters of each layer are $[3 \times 3]:128$, which means that the kernel size of the convolution is $3 \times 3$, and 128 feature maps are produced. All maxpooling layers are implemented with a pooling size of $2 \times 2$ with a stride equal to 2. There is a global maxpooling layer followed by the last convolution layer of each branch. The global maxpooling operation extracts one feature from each feature map. It acts as a high-pass filter and reduces the number of parameters.

Finally, a fine-tuning strategy is employed to avoid the difficulty of simultaneous parameters optimization in the three branches. A pretrained model is required before fine-tuning. Therefore, as described in Step 3 (see Fig. 1), three branches are first trained with a large learning rate separately. Then, the layers after GlobalMaxPool (i.e., Dense and SoftMax layer with gray shading and dotted box marking in Step 2 of Fig. 1) of each pretrained model are removed and the pretrained parameters of remaining layers are fixed. We denote the GlobalMaxPool features of the three CNN branches as $F_{\text{branch},u}(u = 1, 2, 3)$, which are concatenated as

$$F_{\text{Merge}} = f(W \otimes (F_{\text{branch},1} \| F_{\text{branch},2} \| F_{\text{branch},3} ) + b)$$  \hspace{1cm} (7)$$

where $\|$ means concatenating the GlobalMaxPool features of the three CNN branches and $f$ is a nonlinear activation function. Then, two dense layers and SoftMax classifier layer are used to generate the final classification map.
III. EXPERIMENTAL RESULTS

Experiments were conducted on three real multiresolution remote sensing datasets, which were acquired by three different satellite sensors (QuickBird, DEIMOS-2, and GaoFen-2). Ground reference maps are built according to careful image interpretation, where the spatial resolution of the reference maps is fixed as the same as that of the corresponding PAN image.

Algorithms were implemented by using MATLAB and Python, and the DL networks were built using Tensorflow with the high-level API Keras, which is a simplified interface to Tensorflow. In particular, experiments based on DL networks were carried out on the Ubuntu 18.04.5, with Intel Xeon Gold 6130 CPUs at 2.10 GHz, 159-GB RAM, and GPU of NVIDIA GRID P40-24Q, 22 GB.

A. Description of Datasets

1) Xuzhou Dataset (XZ): This dataset was acquired by the QuickBird satellite over urban area of Xuzhou city, China. The PAN image has a size of $1132 \times 1516$ pixels with a spatial resolution of 0.6 m, and the MS image has a size of $283 \times 379$ pixels with a spatial resolution of 2.4 m. This scene contains seven land-cover classes, including buildings1 (with red roofs), buildings2 (with bluish roofs), buildings3 (with gray roofs), playground, roads, vegetation, and water; 200 training samples for each class were selected, and the rest were used for testing. Fig. 2(a) shows the true-color composite image of the MS image, Fig. 2(b) shows its corresponding PAN image, and Fig. 2(c) shows the ground reference map. Fig. 2(d) shows the zoom of the portion of the image highlighted in the yellow box in (c). Table II lists the number of training and test samples used in the experiments.

2) Vancouver Dataset (VC): The 2016 IEEE Geoscience and Remote Sensing Society (GRSS) Data Fusion Contest [41] offered MS and PAN images that were acquired on March 31 and May 30, 2015, over Vancouver city, Canada, from the DEIMOS-2 satellite. A subset of the whole image was selected for experiments. The spatial resolutions of the PAN and the MS images (with blue, green, red, and near-infrared bands) are 1 and 4 m, respectively. The size of the MS and PAN image has a size of $1132 \times 1516$ pixels with a spatial resolution of 0.6 m, and the MS image has a size of $283 \times 379$ pixels with a spatial resolution of 2.4 m. This scene contains seven land-cover classes, including buildings1 (with red roofs), buildings2 (with bluish roofs), buildings3 (with gray roofs), playground, roads, vegetation, and water; 200 training samples for each class were selected, and the rest were used for testing. Fig. 2(a) shows the true-color composite image of the MS image, Fig. 2(b) shows its corresponding PAN image, and Fig. 2(c) shows the ground reference map. Fig. 2(d) shows the zoom of the portion of the image highlighted in the yellow box in Fig. 2(c). Note that the reference map is made according to a careful manual image interpretation of VHR images and Google maps. Table II lists the number of training and test samples used in the experiments.

<table>
<thead>
<tr>
<th>Classes</th>
<th>Number of samples (pixels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>Name</td>
</tr>
<tr>
<td>1</td>
<td>Buildings1</td>
</tr>
<tr>
<td>2</td>
<td>Buildings2</td>
</tr>
<tr>
<td>3</td>
<td>Buildings3</td>
</tr>
<tr>
<td>4</td>
<td>Playground</td>
</tr>
<tr>
<td>5</td>
<td>Roads</td>
</tr>
<tr>
<td>6</td>
<td>Vegetation</td>
</tr>
<tr>
<td>7</td>
<td>Water</td>
</tr>
</tbody>
</table>

Fig. 2. XZ dataset: (a) true-color composite of the MS image, (b) PAN image, (c) ground reference map, and (d) zoom of the portion of the image highlighted in the yellow box in (c).

PAN image has a size of $1132 \times 1516$ pixels with a spatial resolution of 0.6 m, and the MS image has a size of $283 \times 379$ pixels with a spatial resolution of 2.4 m. This scene contains seven land-cover classes, including buildings1 (with red roofs), buildings2 (with bluish roofs), buildings3 (with gray roofs), playground, roads, vegetation, and water; 200 training samples for each class were selected, and the rest were used for testing. Fig. 2(a) shows the true-color composite image of the MS image, Fig. 2(b) shows its corresponding PAN image, and Fig. 2(c) shows the ground reference map. Fig. 2(d) shows the zoom of the portion of the image highlighted in the yellow box in Fig. 2(c). Note that the reference map is made according to a careful manual image interpretation of VHR images and Google maps. Table II lists the number of training and test samples used in the experiments.

2) Vancouver Dataset (VC): The 2016 IEEE Geoscience and Remote Sensing Society (GRSS) Data Fusion Contest [41] offered MS and PAN images that were acquired on March 31 and May 30, 2015, over Vancouver city, Canada, from the DEIMOS-2 satellite. A subset of the whole image was selected for experiments. The spatial resolutions of the PAN and the MS images (with blue, green, red, and near-infrared bands) are 1 and 4 m, respectively. The size of the MS and PAN

Authorized licensed use limited to: Qian Du. Downloaded on July 03,2022 at 02:22:26 UTC from IEEE Xplore. Restrictions apply.
images are $345 \times 219$ and $1380 \times 876$ pixels, respectively. There are mainly six classes in the scene, including buildings1 (with brown roofs), buildings2 (with white roofs), roads, railways, trees, and water. Two hundred training samples for each class were selected, and the others were considered as test samples. Fig. 3 presents the true-color composite images of the MS and the PAN images, and their ground reference map. The training and testing samples are listed in Table III. Fig. 4 shows the true-color composite of the MS and the PAN images, and the corresponding ground reference map.

3) Shanghai Dataset (SH): This dataset was made up of a pair of MS and PAN images acquired by the Chinese GaoFen-2 satellite over Shanghai, China, on January 2, 2015. The spatial resolution of MS (with blue, green, red, and near-infrared bands) and PAN images are 4 and 1 m, respectively. The corresponding image sizes are $300 \times 305$ and $1200 \times 1220$ pixels. There are five classes in this image scene, and the training and testing samples used in experiments are listed in Table IV.

### Table III

<table>
<thead>
<tr>
<th>Classes</th>
<th>Number of samples (pixels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>Name</td>
</tr>
<tr>
<td>1</td>
<td>Buildings1</td>
</tr>
<tr>
<td>2</td>
<td>Buildings2</td>
</tr>
<tr>
<td>3</td>
<td>Roads</td>
</tr>
<tr>
<td>4</td>
<td>Railways</td>
</tr>
<tr>
<td>5</td>
<td>Trees</td>
</tr>
<tr>
<td>6</td>
<td>Water</td>
</tr>
</tbody>
</table>

### Table IV

<table>
<thead>
<tr>
<th>Classes</th>
<th>Number of samples (pixels)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>Name</td>
</tr>
<tr>
<td>1</td>
<td>Buildings</td>
</tr>
<tr>
<td>2</td>
<td>Roads</td>
</tr>
<tr>
<td>3</td>
<td>Water</td>
</tr>
<tr>
<td>4</td>
<td>Trees</td>
</tr>
<tr>
<td>5</td>
<td>Grass</td>
</tr>
</tbody>
</table>

B. Parameter Tuning

The proposed feature-level fusion architecture represents a proper definition of the parameter values to enhance the classification performance. For shallow features, optimal parameters of EMAP were selected after multiple trials: parameters were set as $a = 2000$, $d = 200$, $i = 0.5$, and $s = 10$. In order to analyze and validate in detail the proposed CRHFF approach, the obtained classification results are compared after parameter tuning according to different patch sizes and different layers.

1) Multiscale Comparison: The performance of different patch sizes in the deep network is compared. Results obtained based on PAN image patch sizes of $20 \times 20$, $24 \times 24$, $28 \times 28$, $32 \times 32$, $36 \times 36$, $40 \times 40$, and $44 \times 44$ are shown in Fig. 5. We can see that the patch sizes resulting in the highest classification accuracies are $28 \times 28$ for both the VC and the SH datasets and $36 \times 36$ for the XZ dataset. Patches with different sizes contain spatial features at different scales. Since the classification of large-scale objects may contain isolated noise, large patches are usually preferred. However, using large patches is always time-consuming and may increase the misclassification of small objects. An optimal patch is determined to reach a compromise between image resolution and object size.
TABLE V

<table>
<thead>
<tr>
<th>Class</th>
<th>SVM</th>
<th>F Stack</th>
<th>RF</th>
<th>F Stack</th>
<th>VGG-Like</th>
<th>DMIL</th>
<th>MultiResoLCC</th>
<th>CRHFF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buildings1</td>
<td>65.43</td>
<td>68.57</td>
<td>69.87</td>
<td>74.46</td>
<td>96.39±2.21</td>
<td>98.08±0.76</td>
<td>95.43±0.72</td>
<td>98.78±0.13</td>
</tr>
<tr>
<td>Buildings2</td>
<td>59.20</td>
<td>62.39</td>
<td>56.26</td>
<td>61.56</td>
<td>89.72±9.65</td>
<td>91.92±7.22</td>
<td>90.98±1.46</td>
<td>96.69±2.45</td>
</tr>
<tr>
<td>Buildings3</td>
<td>50.19</td>
<td>56.35</td>
<td>49.27</td>
<td>51.37</td>
<td>89.64±6.56</td>
<td>89.97±4.73</td>
<td>84.22±1.31</td>
<td>95.7±1.83</td>
</tr>
<tr>
<td>Playground</td>
<td>92.30</td>
<td>93.80</td>
<td>91.28</td>
<td>94.39</td>
<td>99.84±0.49</td>
<td>99.85±0.43</td>
<td>99.89±0.07</td>
<td>99.97±0.06</td>
</tr>
<tr>
<td>Roads</td>
<td>58.18</td>
<td>55.79</td>
<td>54.14</td>
<td>61.97</td>
<td>92.39±6.18</td>
<td>93.01±1.30</td>
<td>90.68±1.20</td>
<td>89.66±5.48</td>
</tr>
<tr>
<td>Vegetation</td>
<td>97.55</td>
<td>98.68</td>
<td>97.09</td>
<td>98.11</td>
<td>98.89±0.63</td>
<td>98.02±1.22</td>
<td>97.71±0.35</td>
<td>96.80±0.67</td>
</tr>
<tr>
<td>Water</td>
<td>99.54</td>
<td>99.62</td>
<td>98.72</td>
<td>98.61</td>
<td>99.78±0.16</td>
<td>99.55±0.91</td>
<td>99.07±0.26</td>
<td>99.58±0.30</td>
</tr>
<tr>
<td>OA</td>
<td>69.00</td>
<td>71.18</td>
<td>69.14</td>
<td>73.35</td>
<td>94.78±0.89</td>
<td>95.71±0.47</td>
<td>93.60±0.81</td>
<td>96.45±1.68</td>
</tr>
<tr>
<td>AA</td>
<td>74.63</td>
<td>77.03</td>
<td>73.80</td>
<td>77.21</td>
<td>95.25±4.58</td>
<td>95.74±4.07</td>
<td>94.00±5.65</td>
<td>96.71±3.50</td>
</tr>
<tr>
<td>Kappa</td>
<td>62.18</td>
<td>64.81</td>
<td>62.11</td>
<td>67.03</td>
<td>93.42±1.11</td>
<td>94.56±1.87</td>
<td>91.93±1.01</td>
<td>95.52±2.10</td>
</tr>
<tr>
<td>OA</td>
<td>69.00</td>
<td>71.18</td>
<td>69.14</td>
<td>73.35</td>
<td>94.78±0.89</td>
<td>95.71±0.47</td>
<td>93.60±0.81</td>
<td>96.45±1.68</td>
</tr>
<tr>
<td>AA</td>
<td>74.63</td>
<td>77.03</td>
<td>73.80</td>
<td>77.21</td>
<td>95.25±4.58</td>
<td>95.74±4.07</td>
<td>94.00±5.65</td>
<td>96.71±3.50</td>
</tr>
<tr>
<td>Kappa</td>
<td>62.18</td>
<td>64.81</td>
<td>62.11</td>
<td>67.03</td>
<td>93.42±1.11</td>
<td>94.56±1.87</td>
<td>91.93±1.01</td>
<td>95.52±2.10</td>
</tr>
</tbody>
</table>

C. Classification Performance Evaluation and Analysis

In order to validate the effectiveness of the proposed CRHFF approach, five reference methods were also considered for comparison purposes. We selected two state-of-the-art feature-level fusion methods, i.e., deep multiple instance learning (DMIL) [25] and MultiResolution Land Cover Classification (MultiResoLCC) [26], and three popular methods, i.e., support vector machine (SVM), random forest (RF), and VGG-Like based on PS and feature stacking fusion strategies. Specifically, SVM was implemented using the libsvm toolbox, where the radial basis function was selected as the kernel function. RF was defined with 200 trees. VGG-Like contains three fewer Maxpool layers compared to the original VGG16 [43] limited by the patch size of input data and adds a fully connected layer with ClassNum (the number of classes) neurons to the last layer. It is important to note that final results were generated based on the average of ten times running of each method in order to test the robustness of the methods.

For convenience of expression, the following abbreviations are defined: $F_{PS}$ means the pan-sharpened results of MS and PAN images using the GS PS algorithm, $F_{Stack}$ represents the concatenation of MS (after upsampling) and PAN images, $F_{MS\times PAN}$ and $F_{EMAP\times PAN}$ represent the feature-level fusion based on the MS and PAN images and on the EMAP and PAN images in the proposed CRHFF approach, respectively.

1) Results on the XZ Dataset: The OA, the average accuracy (AA), the Kappa coefficient (Kappa), and the class-by-class accuracies are listed in Table V. To further compare the stability of methods, standard deviation is also calculated. One can see that the proposed CRHFF ($F_{EMAP\times PAN}$) approach is superior to other reference methods providing the highest OA = 98.28%, with an improvement of 2.58%, 4.68%, and 1.83% with respect to VGG-Like ($F_{Stack}$), DMIL, and MultiResoLCC methods, respectively. Moreover, it also outperforms the classical SVM ($F_{Stack}$) and RF ($F_{Stack}$) by sharply increasing the OA values of approximately 27.1% and 24.93%, respectively. The proposed CRHFF ($F_{MS\times PAN}$) approach is also superior to other reference methods in terms of class accuracies on buildings2, buildings3, roads, and water. By combination with EMAP, the proposed CRHFF ($F_{EMAP\times PAN}$) greatly improves the classification performance, especially for roads.

Authorized licensed use limited to: Qian Du. Downloaded on July 03,2022 at 02:22:26 UTC from IEEE Xplore. Restrictions apply.
Fig. 7. Classification maps obtained by different methods on the XZ dataset: (a) SVM (\(F_{PS}\)) (69.00%), (b) SVM (\(F_{Stack}\)) (71.18%), (c) RF (\(F_{PS}\)) (69.14%), (d) RF (\(F_{Stack}\)) (73.35%), (e) VGG-Like (\(F_{PS}\)) (94.98%), (f) VGG-Like (\(F_{Stack}\)) (95.72%), (g) DMIL (93.59%), (h) MultiResoLCC (96.48%), (i) proposed CRHFF (\(F_{MS}\ast PAN\)) (97.45%), and (j) proposed CRHFF (\(F_{EMAP}\ast PAN\)) (98.38%).

Fig. 8. Classification maps obtained by different methods at local scale on the XZ dataset: (a) SVM (\(F_{PS}\)), (b) SVM (\(F_{Stack}\)), (c) RF (\(F_{PS}\)), (d) RF (\(F_{Stack}\)), (e) VGG-Like (\(F_{PS}\)), (f) VGG-Like (\(F_{Stack}\)), (g) DMIL, (h) MultiResoLCC, (i) proposed CRHFF (\(F_{MS}\ast PAN\)), and (j) proposed CRHFF (\(F_{EMAP}\ast PAN\)).

buildings3, and vegetation compared to the CRHFF (\(F_{MS}\ast PAN\)). Furthermore, for the \(F_{Stack}\) fusion strategy, the obtained OA values of SVM, RF, and VGG-Like are 71.18%, 73.35%, and 95.70%, respectively, with an improvement of 2.18%, 4.21%, and 0.92% on the \(F_{PS}\) fusion strategy. This also demonstrates the advantage of feature-level fusion strategies compared with the pixel-level fusion ones.

For a qualitative evaluation, classification maps associated with the average OA values among ten runs are shown in Fig. 7 (which corresponds to the quantitative results in Table V). In order to better evaluate the classification performance, Fig. 8 shows the classification results at a local scale. We can see that the proposed CRHFF (\(F_{MS}\ast PAN\)) provides more reliable classification maps than the reference methods. Besides, by joining the EMAP information, CRHFF (\(F_{EMAP}\ast PAN\)) further enhanced the classification performance, particularly in preserving inner class homogeneity for roads and buildings. Moreover, compared to the DL-based methods (i.e., VGG-Like, DMIL,
TABLE VI

COMPARISON OF THE CLASSIFICATION ACCURACIES (%) PROVIDED BY DIFFERENT METHODS (VC DATASET)

<table>
<thead>
<tr>
<th>Class</th>
<th>SVM</th>
<th>RF</th>
<th>VGG-Like</th>
<th>DMIL</th>
<th>MultiResoLCC</th>
<th>CRHFF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F_{PS}$</td>
<td>$F_{Stack}$</td>
<td>$F_{MS}$</td>
<td>$F_{Stack}$</td>
<td>$F_{EMAP}$</td>
<td>$F_{PAN}$</td>
</tr>
<tr>
<td>Buildings1</td>
<td>51.45</td>
<td>57.47</td>
<td>54.61</td>
<td>57.10</td>
<td>82.56±6.95</td>
<td>82.75±6.96</td>
</tr>
<tr>
<td>Buildings2</td>
<td>94.34</td>
<td>94.41</td>
<td>95.10</td>
<td>95.16</td>
<td>98.40±0.95</td>
<td>94.95±9.57</td>
</tr>
<tr>
<td>Roads</td>
<td>64.48</td>
<td>59.01</td>
<td>75.10</td>
<td>78.43</td>
<td>93.01±1.85</td>
<td>91.84±3.00</td>
</tr>
<tr>
<td>Railways</td>
<td>85.23</td>
<td>91.32</td>
<td>82.44</td>
<td>89.66</td>
<td>97.17±2.88</td>
<td>98.83±0.87</td>
</tr>
<tr>
<td>Trees</td>
<td>97.75</td>
<td>97.72</td>
<td>97.95</td>
<td>97.70</td>
<td>96.69±3.77</td>
<td>96.24±2.95</td>
</tr>
<tr>
<td>Water</td>
<td>99.91</td>
<td>99.35</td>
<td>99.36</td>
<td>99.69</td>
<td>97.49±5.33</td>
<td>99.31±1.21</td>
</tr>
<tr>
<td>OA</td>
<td>88.20</td>
<td>88.80</td>
<td>89.36</td>
<td>90.37</td>
<td>94.56±4.78</td>
<td>95.65±1.51</td>
</tr>
<tr>
<td>AA</td>
<td>81.69</td>
<td>83.21</td>
<td>84.09</td>
<td>86.29</td>
<td>94.22±0.61</td>
<td>93.99±6.14</td>
</tr>
<tr>
<td>Kappa</td>
<td>77.37</td>
<td>78.56</td>
<td>79.69</td>
<td>81.56</td>
<td>89.49±8.04</td>
<td>91.65±2.84</td>
</tr>
</tbody>
</table>

Fig. 9. Classification maps obtained by different methods on the VC dataset: (a) SVM ($F_{PS}$) (88.20%), (b) SVM ($F_{Stack}$) (88.80%), (c) RF ($F_{PS}$) (89.36%), (d) RF ($F_{Stack}$) (90.37%), (e) VGG-Like ($F_{PS}$) (94.95%), (f) VGG-Like ($F_{Stack}$) (95.85%), (g) DMIL (92.43%), (h) MultiResoLCC (97.42%), (i) proposed CRHFF ($F_{EMAP\cdotPAN}$) (98.20%), and (j) proposed CRHFF ($F_{EMAP\cdotPAN}$) (98.58%). The first and third rows represent the whole classification maps at global scale, whereas the second and fourth rows represent the subsets at local scale.

MultiResoLCC, and CRHFF), classification maps obtained by the SVM and RF methods exhibit more salt-and-pepper noises, showing commission errors mainly on roads and buildings.

2) Results of VC Dataset: Table VI lists the obtained classification accuracy. We can see that the OA of the proposed CRHFF ($F_{EMAP\cdotPAN}$) is 98.59%, which is 2.94%, 6.19%, and 1.22% higher than those of the VGG-Like ($F_{Stack}$), the DMIL, and the MultiResoLCC methods, respectively. We can also observe from the confusion matrix that buildings1 and roads, and building2 and railways are more likely to be misclassified due to the similar spectral representations. However, in the proposed CRHFF ($F_{EMAP\cdotPAN}$), accuracies of buildings1 and roads have improvements of 2.38% and 4.15% with respect to those of MultiResoLCC, which are the highest among all the reference methods. The proposed CRHFF ($F_{EMAP\cdotPAN}$) further improves the accuracy of the above-mentioned misclassified classes, by extracting more spatial information with the EMAP at a global scale.

Moreover, the proposed CRHFF approach outperformed the classical SVM ($F_{Stack}$) and RF ($F_{Stack}$) by approximately increasing of 9.79% and 8.22% OA values, respectively. Similar to the previous dataset, for the SVM, the RF, and the VGG-Like methods, feature-level fusion ($F_{Stack}$) strategy is superior to pixel-level fusion ($F_{PS}$) strategy. Fig. 9 shows the obtained classification maps for a detailed qualitative analysis at global and local scales. The proposed CRHFF ($F_{EMAP\cdotPAN}$) preserves the inner class homogeneity and achieves an accuracy on boundaries superior to the others, especially for
3) Results of SH Dataset: Table VII shows the qualitative classification results. Similar conclusions can be drawn as in the previous two datasets. The highest OA value was achieved by the proposed CRHFF (i.e., 98.12%), which is 4.53%, 9.33%, and 2.52% higher than those of the VGG-Like roads and buildings that are difficult to be distinguished in the HR images. The CRHFF (Fخماباس) integrating the EMAP information can better preserve inner class homogeneity for roads and buildings, thus further enhancing the classification performance.

Fig. 10. Classification maps obtained by different methods on the SH dataset: (a) SVM (Fس) (81.14%), (b) SVM (Fب) (82.29%), (c) RF (Fس) (83.43%), (d) RF (Fب) (85.27%), (e) VGG-Like (Fس) (92.42%), (f) VGG-Like (Fب) (94.01%), (g) DMIL (88.91%), (h) MultiResoLCC (95.71%), (i) proposed CRHFF (Fخماباس) (96.49%), and (j) proposed CRHFF (Fخماباس) (98.13%). The first and third rows represent the whole classification maps at global scale, whereas the second and fourth rows represent the subsets at local scale.

### Table VII
Comparison of the Classification Accuracies (%) Provided by Different Methods (SH Dataset)

<table>
<thead>
<tr>
<th>Class</th>
<th>SVM</th>
<th>RF</th>
<th>VGG-Like</th>
<th>DMIL</th>
<th>MultiResoLCC</th>
<th>CRHFF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Fس</td>
<td>Fب</td>
<td>Fس</td>
<td>Fب</td>
<td>Fس</td>
<td>Fب</td>
</tr>
<tr>
<td>Buildings</td>
<td>66.93</td>
<td>66.80</td>
<td>74.40</td>
<td>76.07</td>
<td>89.77</td>
<td>94.47</td>
</tr>
<tr>
<td>Roads</td>
<td>89.64</td>
<td>94.85</td>
<td>83.58</td>
<td>88.33</td>
<td>91.86</td>
<td>95.43</td>
</tr>
<tr>
<td>Trees</td>
<td>98.43</td>
<td>99.03</td>
<td>97.16</td>
<td>97.40</td>
<td>99.06</td>
<td>99.72</td>
</tr>
<tr>
<td>Grass</td>
<td>97.35</td>
<td>98.54</td>
<td>96.30</td>
<td>95.99</td>
<td>99.88</td>
<td>99.99</td>
</tr>
<tr>
<td>OA</td>
<td>81.14</td>
<td>82.29</td>
<td>83.43</td>
<td>85.27</td>
<td>93.06</td>
<td>93.92</td>
</tr>
<tr>
<td>AA</td>
<td>90.46</td>
<td>91.84</td>
<td>90.26</td>
<td>91.54</td>
<td>96.03</td>
<td>96.64</td>
</tr>
<tr>
<td>Kappas</td>
<td>73.31</td>
<td>75.01</td>
<td>76.11</td>
<td>78.70</td>
<td>89.72</td>
<td>90.55</td>
</tr>
</tbody>
</table>

Authorized licensed use limited to: Qian Du. Downloaded on July 03,2022 at 02:22:26 UTC from IEEE Xplore. Restrictions apply.
IV. CONCLUSION

In this article, a CRHFF approach is proposed for joint classification of MS and PAN images. It fills the gap in the traditional ways that fusion of MS and PAN images mainly relies on the PS or individual feature extraction followed by stacking. In particular, to alleviate the degradation of spatial integrity and connectivity of land objects by local patches, we first extract spatial features from the MS image at a global scale, and then, deep hidden layer features are extracted from MS and PAN images and fused from patches at a local scale with an AE-like deep network. Moreover, different scale information of land objects is considered by means of cross-resolution latent features, without implementing upsampling/downsampling operations. Experimental results obtained on three real multiresolution datasets acquired by QuickBird, DEIMOS-2, and GaoFen-2 satellites confirmed the effectiveness of the proposed approach, compared with the state-of-the-art methods in terms of higher classification accuracy and robustness.

In future work, we will explore other cross-resolution fusion networks that could further improve the multiresolution data classification efficiency and accuracy.

ACKNOWLEDGMENT

The authors would like to thank Deimos Imaging for acquiring and providing the VC dataset and the IEEE GRSS Image Analysis and Data Fusion Technical Committee. They would also like to thank the China Centre for Resources Satellite Data and Application for providing the Gaofen-2 images.

REFERENCES

[20] X. Meng et al., “A large-scale benchmark data set for evaluating pan-
sharpening performance: Overview and implementation,” IEEE Geosci.

sification using data fusion of multispectral and panchromatic high-

classification of panchromatic and multispectral images by Markov

eralized metaphor of Chinese restaurant franchise to fusing both
panchromatic and multispectral images for unsupervised classification,”


classification for PAN and MS imagery,” IEEE Trans. Geosci.

architecture for land cover classification of PAN and MS imagery,”

perceptual loss for panchromatic and multispectral images classification,”

attention fusion deep network for multiresolution remote-sensing image

[29] W. Ma et al., “A spatial-channel collaborative attention network for
enhancement of multiresolution classification,” Remote Sens., vol. 13,

[30] B. Rasti et al., “Feature extraction for hyperspectral imagery: The
evolution from shallow to deep: Overview and toolbox,” IEEE Geosci.

profiles with morphological attribute filters for the analysis of hyper-
spectral data.” Int. J. Remote Sens., vol. 31, no. 22, pp. 5975–5991,

[32] M. Dalla Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and
L. Bruzzone, “Classification of hyperspectral images by using extended
morphological attribute profiles and independent component analysis.”

[33] P. Ghamisi, R. Souza, J. A. Benediktsson, X. Z. Zhu, L. Rittner,
and R. A. Lotufo, “Extinction profiles for the classification of remote

[34] X. Kang, S. Li, and J. A. Benediktsson, “Spectral–spatial hyperspectral

[35] X. Kang, C. Li, S. Li, and H. Lin, “Classification of hyperspectral images
by Gabor filtering based deep network,” IEEE J. Sel. Topics Appl. Earth

[36] S. Liu et al., “A multi-scale superpixel-guided filter feature extrac-
tion and selection approach for classification of very-high-resolution
remotely sensed imagery,” Remote Sens., vol. 12, no. 5, p. 862,

of data with neural networks,” Science, vol. 313, no. 5786, pp. 504–507,


and F. Pla, “A new deep generative network for unsupervised remote

[40] C. Liu et al., “Band-independent encoder–decoder network for pan-

http://www.grss-ieee.org/community/technical-committees/data-fusion

tion,” in Proc. 3rd Int. Conf. Learn. Represent., San Diego, CA, USA,

large-scale image recognition,” in Proc. Int. Conf. Learn. Represent.,

Sicong Liu (Senior Member, IEEE) received the B.Sc. degree in geographical information system and the M.E. degree in photogrammetry and remote sensing from the China University of Mining and Technology, Xuzhou, China, in 2009 and 2011, respectively, and the Ph.D. degree in information and communication technology from the University of Trento, Trento, Italy, 2015. He is currently an Associate Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. His research inter-
ests include multitemporal data analysis, change detection, multispec-
tral/hyperspectral remote sensing, and planetary remote sensing.

Dr. Liu serves as the Program Committee Member for the SPIE Remote Sensing Symposium: Image and Signal Processing for Remote Sensing XXVI–XXXVIII for the term of 2020–2022. He was the winner (ranked as third place) of the Paper Contest of the 2014 IEEE GRSS Data Fusion Contest. He is the Technical Co-Chair of the Tenth International Workshop on the Analysis of Multispectral and Remote Sensing Imagery (MultiTemp 2019). He served as the Session Chair for many international conferences, such as the International Geoscience and Remote Sensing Symposium, from 2017 to 2019. He is/was a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS) and Remote Sensing.

Hui Zhao (Student Member, IEEE) received the B.S. degree in geomatics engineering from Chuzhou University, Chuzhou, China, in 2012, and the M.S. degree in geomatics engineering from the Jiangxi University of Science and Technology, Ganzhou, China, in 2017. She is currently pursuing the Ph.D. degree in surveying and mapping with Tongji University, Shanghai, China. Her research interests include multisource remote sensing data fusion in the field of Earth observation and planetary exploration.

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland at Baltimore, Baltimore, MD, USA, in 2000. She is currently a Bobby Shackoul Professor with the Department of Electrical and Computer Engineer-
ing, Mississippi State University, Starkville, MS, USA. Her research interests include hyperspectral remote sensing image analysis and applications, and machine learning. Dr. Du is a fellow of the SPIE—International Society for Optics and Photonics (SPIE). She is also a member of the IEEE Periodicals Review and the Advisory Committee and the SPIE Publications Committee. She was a recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society (GRSS). She was the Co-Chair of the Data Fusion Technical Committee of the IEEE GRSS from 2009 to 2013 and the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014 and the General Chair of the fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing held at Shanghai, China, in 2012. She was an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, Journal of Applied Remote Sensing, and IEEE SIGNAL PROCESSING LETTERS. From 2016 to 2020, she was the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.
Lorenzo Bruzzone (Fellow, IEEE) received the Laurea (M.S.) degree (summa cum laude) in electronic engineering and the Ph.D. degree in telecommunications from the University of Genoa, Genoa, Italy, in 1993 and 1998, respectively. He is currently a Full Professor of telecommunications at the University of Trento, Trento, Italy, where he teaches remote sensing, radar, and digital communications. He is also the Founder and the Director of the Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento. His research interests are in the areas of remote sensing, radar and synthetic aperture radar (SAR), signal processing, machine learning, and pattern recognition. He promotes and supervises research on these topics within the frameworks of many national and international projects. He is the Principal Investigator of many research projects. Among the others, he is currently the Principal Investigator of the radar for icy moon exploration (RIME) instrument in the framework of the JUpiter ICy moons Explorer (JUICE) mission of the European Space Agency (ESA) and the Science Lead for the High Resolution Land Cover project in the framework of the Climate Change Initiative of ESA. He is the author (or coauthor) of 294 scientific publications in refereed international journals (221 in IEEE journals), more than 340 papers in conference proceedings, and 22 book chapters. He is the editor/coeditor of 18 books/conference proceedings and one scientific book. His papers are highly cited, as proven from the total number of citations (more than 40 000) and the value of the H-index (92) (source: Google Scholar). Dr. Bruzzone has been a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) since 2009, where he has been the Vice-President for Professional Activities since 2019. He ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Seattle, July 1998. He was a recipient of many international and national honors and awards, including the recent IEEE GRSS 2015 Outstanding Service Award, the 2017 and 2018 IEEE IGARSS Symposium Prize Paper Awards, and the 2019 WHISPER Outstanding Paper Award. He is the Co-Founder of the IEEE International Workshop on the Analysis of Multi-Temporal Remote-Sensing Images (MultiTemp) Series and is currently a member of the Permanent Steering Committee of this series of workshops. Since 2003, he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. He has been the Founder of the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE, for which he has been Editor-in-Chief from 2013 to 2017. He was invited as a keynote speaker in more than 40 international conferences and workshops. He was a guest coeditor of many special issues of international journals. He is also an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was a Distinguished Speaker of the IEEE Geoscience and Remote Sensing Society from 2012 to 2016.

Alim Samat (Member, IEEE) received the B.S. degree in geographic information system from Nanjing University, Nanjing, China, in 2009, the M.S. degree in photogrammetry and remote sensing from the China University of Mining and Technology, Xuzhou, China, in 2012, and the Ph.D. degree in cartography and geographic information system from Nanjing University, in 2015. He is currently an Associate Researcher with the State Key Laboratory of Desert and Oasis Ecol- ogy, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Ürümqi, China. His research interests include polarimetric synthetic aperture radar (PolSAR) and optical remote sensing for land applications, image processing and pattern recognition, and machine learning. Dr. Samat serves as a Reviewer for several international journals, including IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and Pattern Recognition.

Xiaohua Tong (Senior Member, IEEE) received the Ph.D. degree from Tongji University, Shanghai, China, in 1999. He worked as a Post-Doctoral Researcher at the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China, from 2001 to 2003. He was a Research Fellow with The Hong Kong Polytechnic University, Hong Kong, in 2006, and a Visiting Scholar at the University of California at Santa Barbara, Santa Barbara, CA, USA, from 2008 to 2009. His research interests include remote sensing, geographic infor- mation science (GIS), uncertainty and spatial data quality, image processing for high resolution, and hyperspectral images. Dr. Tong serves as the Vice-Chair for the Commission on Spatial Data Quality of the International Cartographical Association and the Co-Chair for the ISPRS Working Group (WG II/4) on Spatial Statistics and Uncertainty Modeling.