Integrating Zhuhai-1 Hyperspectral Imagery With Sentinel-2 Multispectral Imagery to Improve High-Resolution Impervious Surface Area Mapping

Xiaoxiao Feng, Zhenfeng Shao, Xiao Huang, Luxiao He, Xianwei Lv, and Qingwei Zhuang

***Abstract*—Mapping impervious surface area (ISA) in an accurate and timely manner is essential for a variety of fields and applications, such as urban heat islands, hydrology, waterlogging, and urban planning and management. However, the large and complex urban landscapes pose great challenges in retrieving ISA information. Spaceborne hyperspectral (HS) remote sensing imagery provides rich spectral information with short revisit cycles, making it an ideal data source for ISA extraction from complex urban scenes. Nevertheless, insufficient single-band energy, the involvement of modulation transfer function (MTF), and the low signal-to-noise ratio (SNR) of spaceborne HS imagery usually result in poor image clarity and noises, leading to inaccurate ISA extraction. To address this challenge, we propose a new deep feature fusion-based classification method to improve 10 m resolution ISA mapping by integrating Zhuhai-1 HS imagery with Sentinel-2 multispectral (MS) imagery. We extract deep features that include spectral and spatial features, respectively, from MS and HS imagery via a 2-D convolutional neural network (CNN), aiming to increase feature diversity and improve the model's recognition capability. The Sentinel-2 imagery is used to enhance the spatial information of the Zhuhai-1 HS image, improving the urban ISA retrieval by reducing the impact of noises. By combining the deep spatial features and deep spectral features, we obtain joint spatial-spectral features, leading to high classification accuracy and robustness. We test the proposed method in two highly urbanized study areas that cover Foshan city and Wuhan city, China. The results reveal that the proposed method obtains an overall accuracy of 96.72% and 96.75% in the two study areas, 18.78% and 8.66% higher than classification results with only HS imagery as input. The final ISA extraction overall accuracy is 95.42% and 95.50% in the two study areas, the highest among the comparison methods.**

***Index Terms*—Convolutional neural network (CNN), feature fusion, impervious surface area (ISA) mapping, sentinel-2 imagery, Zhuhai-1 spaceborne hyperspectral (HS) imagery.**

Manuscript received August 9, 2021; revised October 30, 2021 and January 4, 2022; accepted March 5, 2022. Date of publication March 8, 2022; date of current version March 25, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 42090012, in part by 03 Special Research and 5G Project of Jiangxi Province in China under Grant 20212ABC03A09, in part by the Zhuhai Industry University Research Cooperation Project of China under Grant ZH22017001210098PWC, and in part by the Key R&D project of Sichuan Science and Technology Plan under Grant 2022YFN0031. *(Corresponding author: Zhenfeng Shao.)*

Xiaoxiao Feng, Zhenfeng Shao, Luxiao He, Xianwei Lv, and Qingwei Zhuang are with the State Key Laboratory of Information Engineering in Surveying, Mapping, Remote Sensing, Wuhan University, Wuhan 430079, China (e-mail: fengxxalice2018@gmail.com; shaozhenfeng@whu.edu.cn; heluxiao@foxmail.com; xianweilv@whu.edu.cn; zhuangqingwei@whu.edu.cn).

Xiao Huang is with the Department of Geosciences, University of Arkansas, Fayetteville, AR 72701 USA (e-mail: xh010@uark.edu).

Digital Object Identifier 10.1109/JSTARS.2022.3157755

I. Introduction

IMPERVIOUS surface area (ISA) is usually defined as natural or artificial surfaces (e.g., roads, parking lots, roofs made of cement concrete, glass, asphalt, plastic, tiles, metal, etc.) covering in cities that prevent water from penetrating into the ground [1]. The rapid progress of urbanization inevitably leads to tremendous changes in land use and land cover types. ISA is a key indicator in evaluating the urban ecological environment and usually poses notable negative impacts on the urban environment [2], climate [3], [4], and hydrology [5]–[7]. Therefore, the evaluation of ISA distribution should focus on not only its spatial expansion, but also its environmental consequences. Furthermore, it is of great significance for the sustainable development strategy of urban planning and management to obtain accurate ISA information in a timely manner and investigate the impact of its dynamic changes on the environment.

Remote sensing technology has been widely used in ISA monitoring, thanks to its extensive spatial coverage and high temporal frequency. Early studies on ISA were mostly based on medium-resolution multispectral (MS) satellites such as Landsat Thematic Mapper (TM) [8], [9] and Enhanced Thematic Mapper (ETM+) [10]. However, the complexity of urban landscapes and broadband reflectance data pose great challenges in ISA classification, as many urban materials cannot be distinguished accurately. Besides, ISA maps with coarse resolutions are limited in potential applications, e.g., distinguishing urban functional areas [1]. In comparison, fine-resolution ISA maps allow for more spatial-explicit studies such as investigating the impact of urbanization on energy, water, carbon cycles, vegetation phenology, and surface climate [11]. Hyperspectral (HS) imagery can provide not only spatial information of features, but also rich spectral information that can accurately reflect heterogeneous spectral characteristics of features, leading to fine identification and classification. Most of the existing ISA studies utilized classic HS data captured by the airborne HS sensors [12], such as the simulated Environmental Mapping and Analysis Program (EnMAP) [13], the Hyperspectral Digital Imagery Collection Experiment (HYDICE) [14], and Reflective Optics System Imaging Spectrometer (ROSIS) [15]. Signal-to-noise ratio (SNR) describes the quality of a measurement. In charge-coupled device (CCD) imaging, SNR refers to the ratio of the measured signal to the overall measured noise (frame-to-frame) at that pixel. High SNR is particularly important in applications requiring precise measurement. The advantage gained from the

fine spectral information obtained from HS sensors can be offset by the lower SNR when compared to MS sensors because of the fewer number of photons captured by each detector due to the narrower width of the spectral channels. Compared to the spaceborne HS data, images from airborne HS sensors are characterized by high spatial resolution and high SNR. However, the airborne HS data is limited in synoptic coverage at urban scales, which limits their use for systematically mapping urban land cover of arbitrary cities around the world [13]. This study marks a pioneering effort to integrate the spaceborne HS data and spaceborne MS data (Sentinel-2) for accurate and fine-grained (10 m) ISA mapping.

The classification-based ISA extraction methods aim to first extract spatial and spectral features and feed them into classifiers to obtain the ISA distribution map. Traditional classification methods include maximum likelihood estimation (MLE) [16], support vector machine (SVM) [17], [18], random forest (RF) [19], and their derivations [20]–[22]. Among them, SVM is superior to MLE, as it can solve the nonlinear classification problem. Further, parallel SVM (PSVM) [23] has been developed to solve the computational complexity problem, and the hierarchical PSVM method is designed based on sequential minimal optimization (SMO) [24] and SVM. In addition, kernel methods combined with SVM are widely used in HS image classification to improve separability [25]. Recently, improved sparse representation, e.g., synchronous orthogonal matching pursuit (SOMP) [26] and synchronous subspace pursuit (SSP) [27], was applied to HS image classification and achieved great classification results. In the aforementioned methods, training samples are used to learn the sparse representation dictionary, where the test samples in HS images are sparsely represented. The representation residuals are further compared to find the best representation to determine the label of samples.

However, traditional classification methods largely rely on expertise and are dependent on parameter settings, leading to their low automation and low generalization. The deep learning networks, such as stacked automatic encoder (SAE) [28], deep belief network (DBN) [29], [30], and deep convolutional neural network (DCNN) [31], [32], are different from traditional feature extraction methods. Compared with other networks, CNN uses local connections to extract features with shared weights. Such a design facilitates effective information retrieval and reduces the number of parameters needed to be trained. Chen *et al.* [33] applied a self-coding network to classify the reduced HS images and achieved decent results. They further found that CNN can extract the spatial and spectral features of the objects in images in a more effective manner, thus leading to better classification results.

After reviewing relevant literature, we identified the following challenges in ISA retrieval based on HS images:

1) The low SNR and modulation transfer function (MTF) of spaceborne HS data lead to defective spatial information, evidenced by the low-quality spectral information of ground objects.
2) The spectral-based CNN methods fail to integrate the spatial information of ground objects, which results in salt-and-pepper noises in classified results, thus leading to reduced classification accuracy.

To address these challenges, we propose a novel approach to improving the ISA extraction accuracy by integrating Sentinel-2 MS data and Zhuhai-1 HS data. The first strategy is to first fuse HS and MS images and then obtain the ISA results using classifiers. Commonly used HS-MS image fusion methods can be roughly classified into pan-sharpening and subspace-based methods. Pan-sharpening-based methods include component substitution, multiresolution analysis, and sparse representation [34]. The latter category, e.g., Bayesian method-based methods and spectral unmixing-based methods, focuses on the inherent spectral characteristics of scenes. The other strategy is to first fuse the features extracted from HS and MS, and further obtain the classification results. Comparing these two strategies, the former one highly relies on the fused image, and the accuracy of classification results based on pixel-level fusion images depends on the spectral fidelity of the fusion algorithm. Therefore, in this article, we use CNNs to extract the features from HS and MS images, respectively, and fuse the features to obtain the classification results.

The proposed integration process is achieved by fusing the spectral and spatial deep features extracted from HS and MS images, thus potentially improving the accuracy of the final ISA map. As HS imagery contains abundant spectral information while MS data contains detailed spatial information, we extract spectral and spatial deep features from HS imagery and MS imagery, respectively. In this study, we utilize two-dimensional (2-D) CNN to extract the deep features and further enhance features by fusing extracted spectral and spatial deep features. To deal with salt-and-pepper noises of classification results, the object-based image analysis (OBIA) [35] method is a commonly used approach. However, the OBIA classification method is mainly for images with very high spatial resolution. For this study, the spatial resolution of HS satellite images used in this article is 10 m, which is not ideal for the application of OBIA. Therefore, we use a 2-D CNN network to extract the spatial information of images and further perform impervious surface classification.

The main contributions of this article are summarized as follows.

1) The extraction of the spectral and spatial deep features from HS and MS images, respectively, and their fusion contribute to better feature retrieval from the ground objects in images, thus leading to improved classification accuracy.
2) The fusion of spectral and spatial deep features improves the model's robustness and reduces noises in classified results assisted by the supplement of spatial information.
3) Zhuhai-1 HS data (2-day revisiting cycle) and Sentinel-2 MS data (5-day revisiting cycle) have a considerably high temporal resolution. Therefore, their combination realizes a high-temporal fine-grained ISA mapping, providing the basis for future time series ISA analysis and timely supports in urban land management and construction planning.

The rest of this article is organized as follows. Section II introduces related works and the proposed method. Section III describes the study areas and experimental datasets. Section IV presents and analyzes the experimental results. Section V

discusses the effectiveness of the proposed method compared with a single feature classification network and the effect of different patch sizes on the ISA extraction. Finally, Section VI concludes this article.

II. Methodology

A. Convolutional Neural Network

CNN has received wide attention in recent years and achieved great performances in classification, detection, and many other tasks. CNN has two characteristics: local connection and shared weights. In each convolution layer, feature maps are generated by multiple learnable filters, which can be expressed as

$$y_j^l = \sum_{i=1}^{d} f\left(x_i^{l-1} * w_{ij}^l + b_j^l\right) \tag{1}$$

where x_i^{l-1} denotes the ith feature map of $l-1$ layer, y_j^l denotes the jth feature map of l layer, and d is the number of the input feature maps. w_{ij}^l and b_j^l are the randomly initialized weights and bias, respectively. $*$ denotes the convolutional operator, and f denotes the nonlinear activation functions, such as Sigmoid, Tanh, and rectified linear unit (ReLU) [36]. In this article, we use parametric ReLU (PReLU) [37], which can be formulated as

$$\text{PReLU}\left(y_i\right) = \begin{cases} y_i & if\ y_i > 0 \\ a_i y_i & if\ y_i \le 0 \end{cases} \tag{2}$$

where y_i is the input of ith channel, and a_i is a coefficient that controls the slope. After convolution operations, a max-pooling layer is used to downsample the feature maps [38]. In this way, the output size and the number of parameters can be reduced, effectively avoiding overfitting. The max-pooling operation can be formulated as

$$y_{r,c} = \max_{0 \le g,n \le h} \left(x_{r+g,c+h}\right) \tag{3}$$

where $y_{r,c}$ is the neuron value at (r, c) in the output layer. g and h are the pixel position around the center neuron at (r, c) within the image patch.

The network training procedure consists of forward and backward propagations, aiming to reduce the gap between the predicted labels and the ground truth labels by updating model parameters. The loss/cost is calculated by the differences between the predicted values and the ground-truthing values in the forward propagation. The purpose of backpropagation is to reduce loss by adjusting the parameters. In this study, we use the softmax cross-entropy loss

$$c = -\frac{1}{m}\sum_{i=1}^{m}\left[x_i \log\left[z_i\right] + \left[1 - x_i\right]\log\left[1 - z_i\right]\right] + \frac{\lambda}{2\,m}\sum_{j=1}^{N} w_j^2 \tag{4}$$

where m is the size of the image batch, x_i and z_i denote the ith predicted label and the ground truth label, respectively. N means the number of weights. λ is the parameter to adjust the proportion between the former term (original loss function) and the regularization term (the latter term) in (4). Besides, we set λ to $\frac{1}{2}$ to simplify the process of derivation. Studies have proved that l_2 regularization term can avoid models from overfitting [31], [39].

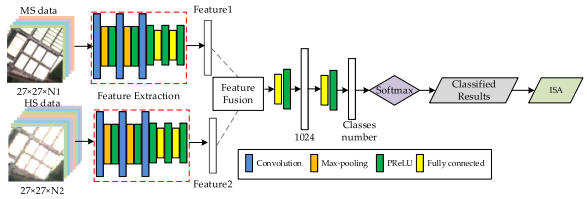

Fig. 1. Workflow of ISA extraction by fusing spectral-spatial deep features using 2-D CNN.

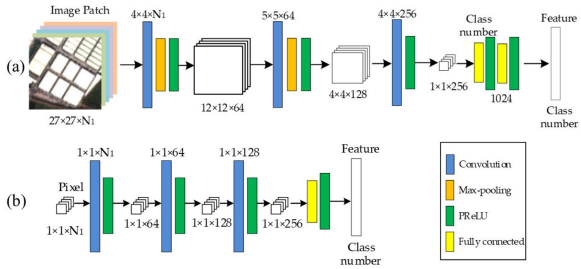

Fig. 2. Workflow of the 2-D CNN-based deep features extraction. (a) the workflow of patchwise feature extraction; (b) the workflow of the pixelwise feature extraction.

The classification approaches of HS images based on CNN can be grouped into three categories, i.e., spectral feature-based, spatial feature-based, and spatial-spectral feature-based methods [40]. Spectral feature-based classification methods apply one-dimension (1-D) CNN to extract the deep spectral features for classification [41], [42]. In comparison, spatial feature-based classification methods apply 2-D CNN to extract the spatial information for classification [43]. The main difference between 1-D CNN and 2-D CNN is the dimensionality of the convolution operation. In this study, we use 2-D CNN to extract the spectral features from HS images and spatial features from MS images and further fuse the obtained spectral-spatial deep features to improve the classification accuracy of ISA.

B. Extraction and Fusion of Spectral and Spatial Features via 2-D CNN

HS images contain rich spectral information that benefits accurate descriptions of the spectral characteristics of ground objects. Given the spaceborne nature, HS images are with limited SNR and spatial resolution. In contrast, MS images with the same spatial resolution are characterized by high SNR. Therefore, to simultaneously obtain the spectral and spatial information, we use 2-D CNN to extract the deep spectral features from HS images and the deep spatial features from MS images. We further fuse the extracted deep features for land cover classification and eventually map the fine-grained ISA distribution. The specific workflow is shown in Fig. 1.

1) Deep Features Extraction From HS and MS Datasets: In this article, the spatial and spectral deep features of images are extracted by three convolution layers and a fully connected layer [see Fig. 2(a)]. The first layer takes the 27×27 image patch with

N_1 channels and calculates 64 feature maps using 4×4 receptive field and a nonlinear activation PReLU. The second layer takes the 12×12 image patch with 64 channels and calculates 128 feature maps using 5×5 receptive field and PReLU. The third layer takes the 4×4 image patch with 128 channels and calculates 256 feature maps using 4×4 receptive field and a nonlinear activation PReLU. The calculating process of these three convolution layers can be expressed in (5). Finally, the full connected layer takes the 1×1 vector with 256 feature maps to derive the classification results. The workflow of the patchwise feature extraction is shown in Fig. 2(a), and the workflow of pixelwise feature extraction is shown in Fig. 2(b).

$$\begin{cases} f_1(x) = \max\left(0, b_1 + w_1 * x\right) \\ w_1 : 64 \times (4 \times 4 \times N_1), b_1 : 64 \times 1 \\ f_2(x) = \max\left(0, b_2 + w_2 * f_1(x)\right) \\ w_2 : 128 \times (5 \times 5 \times 64), b_2 : 128 \times 1 \\ f_3(x) = \max\left(0, b_3 + w_3 * f_2(x)\right) \\ w_3 : 256 \times (4 \times 4 \times 128), b_3 : 256 \times 1. \end{cases} \tag{5}$$

To compare the classification performance between the pixel-based 1-D CNN and image patch-based 2-D CNN, we use two workflows to obtain the results [see Fig. 2(a) and (b)]. The pixel-based 1-D CNN extracts only spectral features, while the patch-based 2-D CNN extracts both spectral and spatial features from images. We further conduct experiments to analyze how such workflow selection influences classification accuracy.

2) Spectral and Spatial Deep Features Fusion: Multisource images contain diverse information, while single-source images may not achieve the best classification performance due to their lack of feature diversity. We use concatenation to fuse the features to enhance feature discrimination ability, which denote as FF-C

$$\mathbf{H}_C^{(l+1)} = \left[\mathbf{H}_{HS}^{[l]}, \mathbf{H}_{MS}^{[l]}\right] \tag{6}$$

where $[\bullet, \bullet]$ denote the elementwise addition, elementwise multiplication, and concatenation operations, respectively. $\mathbf{H}_{\mathrm{HS}}^{[l]}$ and $\mathbf{H}_{\mathrm{MS}}^{[l]}$ denote the lth layer features extracted from HS and MS images, respectively.

III. Study Areas and Datasets

This study includes two study areas that cover parts of Foshan city in Guangdong Province, China and Wuhan city in Hubei Province, China, respectively. The HS dataset is derived from the Zhuhai-1 Orbita HS satellite, while the MS dataset is derived from the Sentinel-2 satellite.

A. Zhuhai-1 OHS HS Datasets

The second batch of Zhuhai-1 microsatellites was successfully launched on April 26, 2018, including four Orbita HS satellites (referred to as OHS-A, OHS-B, OHS-C, and OHS-D) and one video satellite (OVS-2 A). The spatial resolution of OHS data is 10 m, with an imaging range of 150 km, the spectral resolution of 2.5 nm, and the spectrum from 400 to 1000 nm (see Table I). A single HS satellite has 15–16 daily orbits, and the single data acquisition time of each orbit is less than 8 min. At present, the revisiting cycle of the four satellites is two days.

TABLE I
Center Wavelength of OHS HS Data

Channel	Central wavelength (nm)	Channel	Central wavelength (nm)	Channel	Central wavelength (nm)	Channel	Central wavelength (nm)
B01	466	B09	596	B17	716	B25	836
B02	480	B10	610	B18	730	B26	850
B03	500	B11	626	B19	746	B27	866
B04	520	B12	640	B20	760	B28	880
B05	536	B13	656	B21	776	B29	896
B06	550	B14	670	B22	790	B30	910
B07	566	B15	686	B23	806	B31	926
B08	580	B16	700	B24	820	B32	940

TABLE II
Center Wavelength and Spatial Resolution of Sentinel-2 Imagery

Channel	Central wavelength (nm)	Spatial resolution (m)
Band1 - Coastal	443	60
Band2 - Blue	490	10
Band3 - Green	560	10
Band4 - Red	665	10
Band5 - Red Edge	705	20
Band6 - Red Edge	740	20
Band7 - Red Edge	783	20
Band8 - Near infrared	842	10
Band8a - Red Edge	865	20
Band9 - Water vapor	945	60
Band10 - SWIR-Cirrus	1,373	60
Band11 - SWIR1	1,610	20
Band12 - SWIR2	2,190	20

The OHS satellite is characterized by its small size, high spatial resolution, large breadth, and short revisit period. It is expected to benefit various tasks that include ecological environment monitoring, urban construction management, agricultural production, disaster prediction, and assessment.

B. Sentinel-2 MS Datasets

Sentinel-2 is an Earth observation mission from the Copernicus Programme (operated by the European Space Agency) that systematically acquires optical imagery at high spatial resolution (10–60 m) over land and coastal waters. The mission supports a broad range of services and applications such as agricultural monitoring, emergency management, land cover classification, and water quality monitoring. Sentinel-2 has a 5-day revisiting cycle. We select the bands of 10 m spatial resolution for feature extraction (see Table II). The dataset is downloaded from the USGS website from the collection of Level-1 C products.

C. Study Area

The first study area (344.84 km^2), located at $113^\circ 4'$–$113^\circ 15'$E, $22^\circ 48'$–$22^\circ 59'$N, covers part of Foshan city in South China's Guangdong province (see Fig. 3). Lying in the middle of the Pearl River delta plain, Foshan city has a high degree of urbanization and owns a large number of scattered hills, rivers, and water networks, including navigation, irrigation, aquaculture, and other functional areas. The second study area (388.09 km^2), located at $114^\circ 6'$–$114^\circ 20'$ E, $30^\circ 22'$–$30^\circ 34'$N, covers part of Wuhan city in Hubei province (see Fig. 4). The Wuhan city is located in the east of Jianghan Plain and on the middle reaches of the Yangtze River at the intersection of the Yangtze and Han rivers.

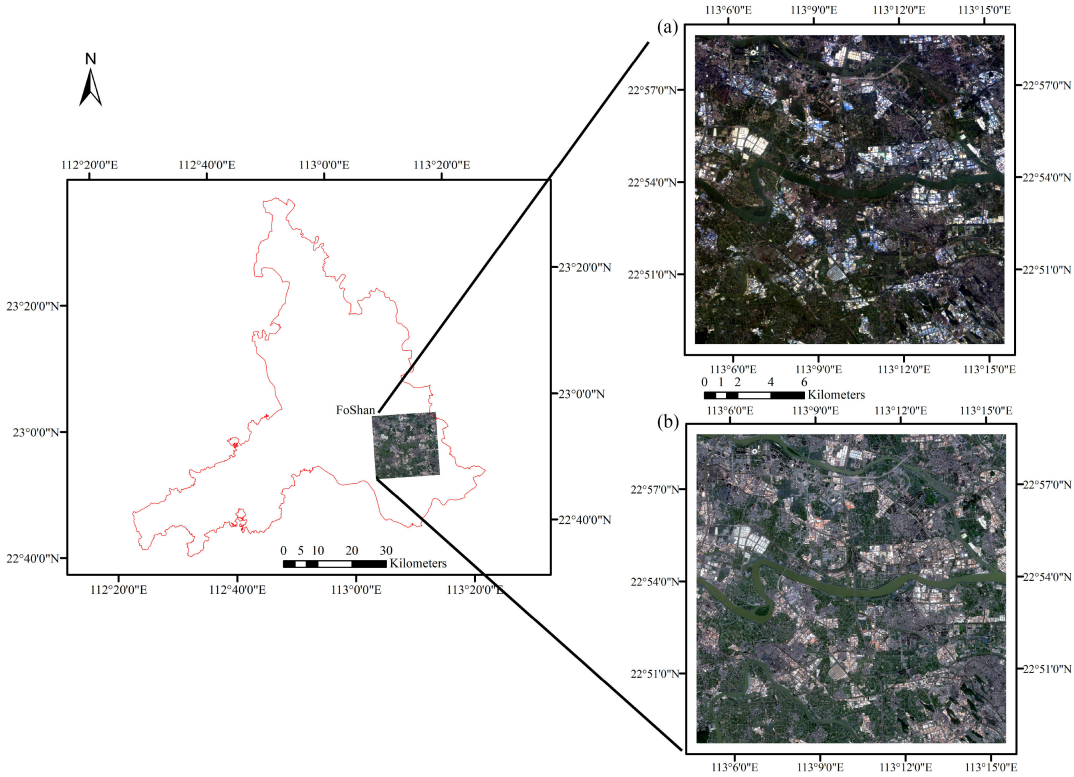

Fig. 3. Study area in Foshan city (1855 × 1855 pixels). (a) Zhuhai-1 HS image (shown in bands 12, 6, 1 as RGB) acquired on November 9, 2019; (b) Sentinel-2 MS image (shown in bands 4, 3, 2 as RGB) acquired on November 11, 2019).

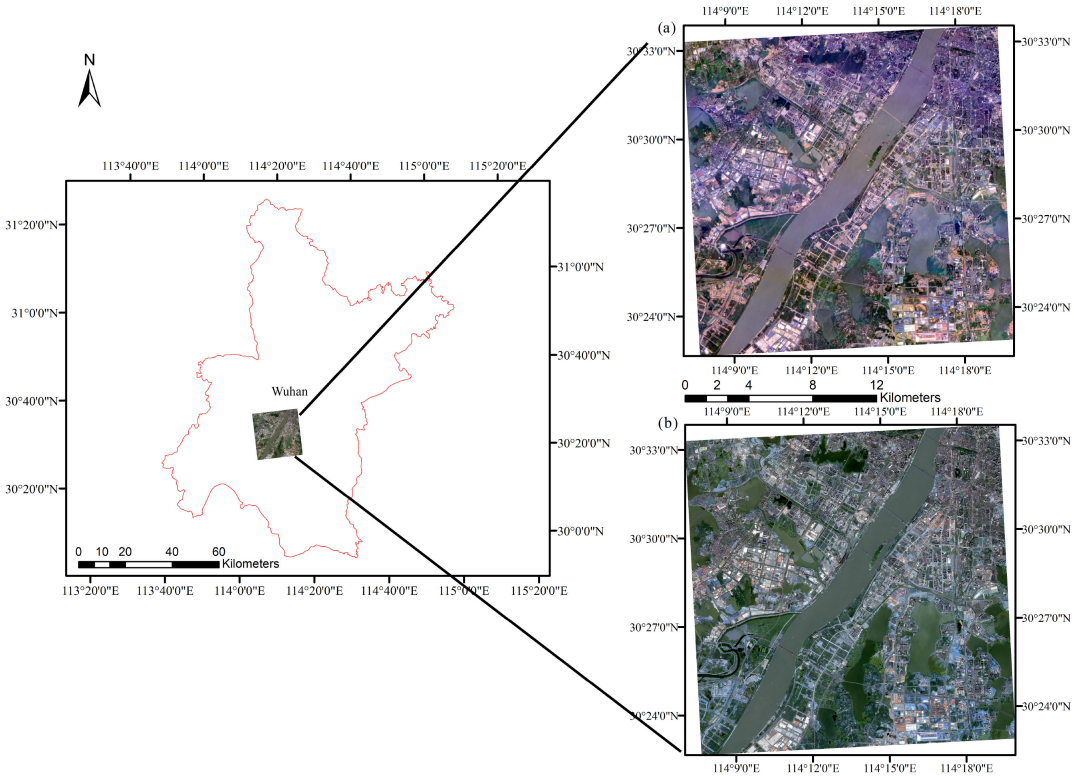

Fig. 4. Study area in Wuhan city (1970 × 1970 pixels). (a) Zhuhai-1 HS image (shown in bands 12, 6, 1 as RGB) acquired on September 21, 2019; (b) Sentinel-2 MS image (shown in bands 4, 3, 2 as RGB) acquired on September 22, 2019).

Fig. 5. Examples of 8 different land cover types in the Foshan study area. (a) Google Earth image; (b) Zhuhai-1 HS image; (c) Sentinel-2 MS image.

Both study areas are characterized by high-level urbanization and dense water networks, making them prone to waterlogging issues, especially after frequent and intensive rain. The fine-resolution ISA distribution can provide the basis for the investigation of urban resilience. Monitoring of ISA distribution dynamics plays a vital role in urban environmental impact analysis and planning management. Both Zhuhai-1 HS image and Sentinel-2 MS image were captured under clear-sky conditions to illustrate the effectiveness of the proposed classification workflow that integrates these two images for an improved 10 m ISA mapping. Given the short time intervals between HS and MS images in the two study areas, we believe there exist no significant changes in ground features. The image SNR is estimated by the number of object types in the study area [44]. The SNR values of the HS and MS images in the Foshan study area are 28.45 and 149.69 dB, respectively. The SNR values of HS and MS images in the Wuhan study area are 35.36 and 149.66 dB, respectively.

IV. Experiments and Results

In this section, we detail our experimental settings and present the results along with the analysis. Section IV-A details the sample selection procedure, Section IV-B shows the experimental setup. Section IV-C shows the effectiveness of the proposed fusion algorithm that integrates HS and MS deep features, and Section IV-D shows the comparison of classification results obtained from the proposed method and other state-of-art classification methods.

A. Sample Selection

Before the sample selection, the HS and MS images were geometric registered manually by selecting correspondence points. The training, validation, and testing samples used in this study were all selected from Sentinel-2 images via human interpretation against Google Earth imagery. We first randomly select the sample points (pixels) from the image, making sure the sampling points are evenly distributed on the image. Then, we classify the points into their corresponding types. Finally, we divide sample points into training, validation, and testing samples according to the ratio of 8:1:1. For the pixelwise input, samples are the central pixels, while for the patchwise input, samples are the patches with different sizes centered around the central pixels. The Foshan study area has eight land cover types, i.e., vegetation, roof, asphalt road, river, dense building, bright ISA, pound, and soil (see Fig. 5). The number of samples for each land cover type can be found in Table III. The total samples of the Foshan study area contain 473 302 pixels. For the eight derived land cover types, the land cover types that include roof, asphalt road, dense building, and bright ISA are classified as ISA.

The Wuhan study area has 10 land cover types, including soil, bright ISA, concrete road, vegetation, dense building, lake, asphalt road, algae, roof, and river (see Fig. 6). The number of samples for each classified land cover type is listed in Table IV. The total samples of the Wuhan study area contain

TABLE III
Number of Training, Validation, and Testing Samples Used in the Foshan Study Area

No.	Class	Training	Validation	Testing
1	Vegetation	39,832	4,929	4,938
2	Roof	12,187	1,508	1,564
3	Asphalt Road	24,535	2,956	3,065
4	River	139,079	17,406	17,431
5	Dense Building	48,186	6,035	6,012
6	Bright ISA	44,944	5,595	5,573
7	Pound	58,010	7,402	7,305
8	Soil	11,869	1,499	1,442
Total		378,642	47,330	47,330

TABLE IV
Number of Training, Validation, and Testing Samples Used in the Wuhan Study Area

No.	Class	Training	Validation	Testing
1	Soil	12,691	1,545	1,608
2	Bright ISA	23,054	2,854	2,767
3	Concrete Road	9,326	1,185	1,127
4	Vegetation	50,745	6,304	6,367
5	Dense Building	11,040	1,357	1,413
6	Lake	27,327	3,344	3,472
7	Asphalt Road	45,293	5,690	5,628
8	Algae	12,551	1,612	1,546
9	Roof	15,241	1,959	1,969
10	River	28,291	3,595	3,548
	Total	235,559	29,445	29,445

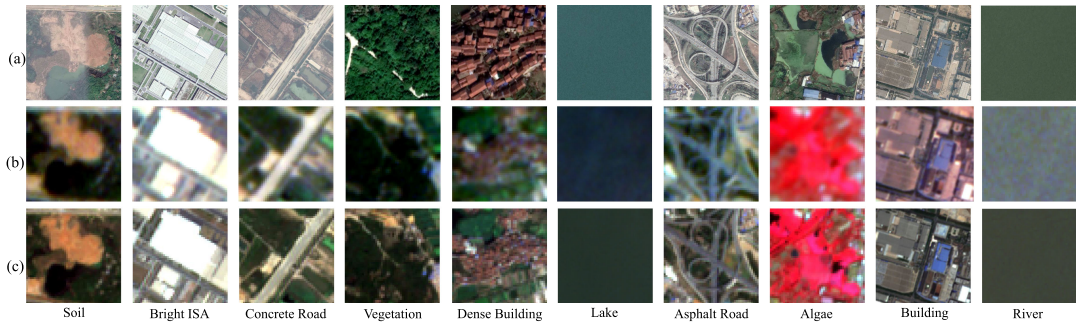

Fig. 6. Examples of eight different land cover types in the Foshan study area. (a) Google Earth image; (b) Zhuhai-1 HS image; (c) Sentinel-2 MS image.

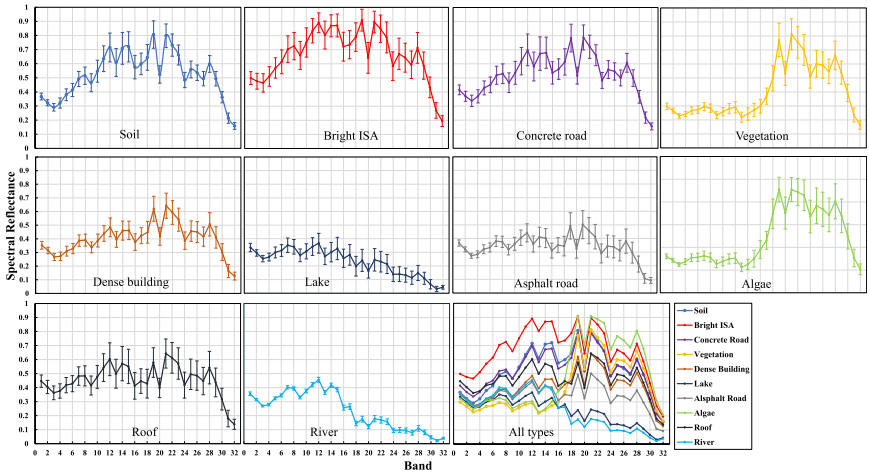

Fig. 7. Spectral reflectance curves of each land cover type from Zhuhai-1 HS image for the training samples in the Wuhan study area. The curve type is obtained by calculating the average value and standard deviation of the spectral reflectance of all training samples for each land cover type and for each given band.

294 449 pixels. Land cover types that include bright ISA, concrete road, dense building, asphalt road, and roof are classified as ISA.

To illustrate that different land cover types have distinct spectral characteristics in HS images, which is beneficial for the classification, we present the spectral reflectance curves in Fig. 7. Using the Wuhan study area as an example. For each land cover type in the Wuhan study area, we calculate the average and standard deviation of all training samples for each land cover type and for each band, following [45]. Fig. 8(a) and (b) shows the ground-truth distribution of land cover types in the Foshan and Wuhan study areas.

B. Experimental Setup

1) Implementation Details: The proposed network in this study is implemented on the Pytorch platform with Adam optimizer [46]. In the network training, we set the maximum number of epochs to 30, the batch size in the training phase to 100, and the learning rate to 0.001. The input data is normalized into [0, 1]. According to the number of training samples in the two study areas, we set the number of training batches to 3786 and 2355 in Foshan and Wuhan study areas, respectively. To reduce overfitting and to stabilize the network during the training phase, we set the l_2 norm regularization to 0.01.

2) Comparison With Baseline Methods: The competing methods are the classic classification methods with the following parameter settings.

1) RF: 200 decision trees are used in the classifier.

2) SVM: The kernel is the radial basis function with two optimal hyperparameters σ and λ, set to 0.1 and 0.01, respectively.

3) Multinomial Logistic Regression (MLR): We choose the l_2 regularization as the penalty (set to 0.01) and "lbfgs" as the solver.

4) Multilayer Perceptron (MLP): We set the batch size to 100, the max epoch to 30, the l_2 norm regularization to 0.01, activation function to ReLU, and the optimizer to Adam.

5) Vanilla Recurrent Neural Network (RNN).

6) RNN with gated recurrent units (GRU).

7) RNN with long short term memory (LSTM). The code for these competing methods is available in [47]. All methods use the same training, validation, and testing samples.

The performances of the classification results are assessed based on three indicators that include the overall accuracy (OA), average accuracy (AA), and Kappa coefficient (Kappa). The OA measures the ratio between correctly classified testing samples and the total number of testing samples. The AA measures the average percentage of correctly classified samples for an individual class. The Kappa measures the percentage agreement corrected by the level of agreement that can be expected by chance alone. Each land cover type is assessed based on two

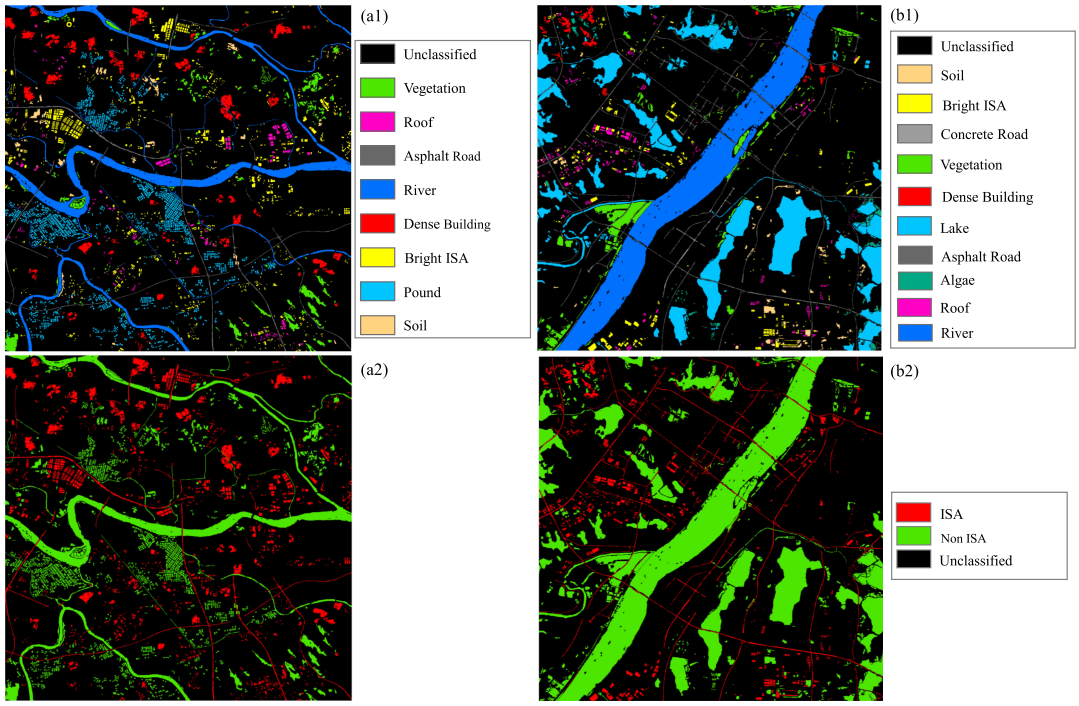

Fig. 8. (a1) and (a2), respectively, show the land cover and ISA in the Foshan study area; (b1) and (b2), respectively, show the land cover and ISA in the Wuhan study area.

TABLE V
LAND COVER TYPE CLASSIFICATION RESULTS (UA, PA AND AA) IN THE FOSHAN STUDY AREA BASED ON FF-C

Classes	UA(%)	PA(%)	AA(%)
Vegetation	98.27	98.70	98.49
Roof	93.52	88.85	91.18
Asphalt Road	86.48	88.64	87.56
River	98.61	97.71	98.16
Dense Building	95.65	97.10	96.38
Bright ISA	99.02	99.18	99.10
Pound	95.98	96.97	96.47
Soil	93.88	91.47	92.67
OA(%)		96.72	

TABLE VI
LAND COVER TYPE CLASSIFICATION RESULTS (UA, PA AND AA) IN THE WUHAN STUDY AREA BASED ON FF-C (HS AND MS IMAGES)

Classes	UA(%)	PA(%)	AA(%)
Soil	97.71	96.85	97.28
Light impervious	98.14	96.38	97.26
Concrete road	92.87	91.30	92.09
Vegetation	97.10	97.51	97.31
Dense Building	95.24	94.83	95.04
Lake	97.94	97.85	97.90
Asphalt Road	95.23	96.37	95.80
Algae	94.49	95.05	94.77
Roof	94.90	94.66	94.78
River	99.72	99.52	99.62
OA(%)		96.75	

indicators that include the User's Accuracy (UA) and Producer's Accuracy (PA). UA represents the number of correctly classified samples divided by the total number of samples in the ground truth. PA represents the number of correctly classified samples divided by the total number of samples classified as the land cover type.

C. Classification Results

In this section, we compare the classification results of the proposed method and other state-of-art methods.

1) The Performance of the Feature Fusion Methods Classification: The feature fusion-based method is denoted as FF-C. We extract the image patches in size of 27×27 pixels, a small patch size to ensure a homogenous land cover type in each patch.

From Tables V and VI, it can be seen that the classification results of FF-C yields high accuracy in each type. In Table V, Bright ISA in the Foshan study area can be identified accurately with the highest PA. From Table VI, we notice that the land cover classification results from FF-C are satisfactory, which proves the strong capability of the feature fusion method. This is very helpful with the fine ISA distribution extraction.

2) Classification Results of FF-C and Comparison Methods: In this section, we compare the classification accuracy of each land cover type and OA, AA, Kappa obtained by different methods for the two study areas. Quantitative results are shown in Tables VII and VIII. Figs. 11 and 12 present classification maps in two study areas for visual comparison among different methods. Table VII indicates that the proposed FF-C method obtains the best AA, OA, and Kappa in the Foshan study area, higher than the best results among the comparison methods (obtained from RF) by 5.78%, 2.76%, and 3.55%, respectively. The proposed FF-C method presents the best classification accuracy for all land cover types, except for vegetation, roof, and river.

From Table VII, we notice that FF-C also yields the best AA, OA, and Kappa, higher than GRU that obtains the second-highest OA by 2.05%, 0.97%, and 1.21%, respectively. Comparing

TABLE VII
QUANTITATIVE COMPARISONS OF DIFFERENT METHODS IN TERMS OF OA, AA, AND KAPPA IN THE FOSHAN STUDY AREA

Classes	RF	MLP	MLR	SVM	GRU	LSTM	RNN	FF-C
Vegetation	98.89	97.45	98.1	98.54	98.68	97.93	**98.97**	98.70
Roof	89.90	69.50	78.26	80.31	77.11	**90.60**	85.81	88.85
Asphalt Road	72.85	47.44	63.49	66.2	72.66	63.82	65.61	**88.64**
River	**97.85**	96.12	96.96	96.90	**97.85**	94.99	96.29	97.71
Dense building	90.54	81.24	84.23	84.20	83.12	88.91	85.15	**97.10**
Bright ISA	98.47	96.77	97.58	97.63	98.76	97.74	98.40	**99.18**
Pound	95.56	92.02	92.84	92.98	92.64	96.78	95.35	**96.97**
Soil	73.16	43.34	61.10	60.82	71.78	65.74	62.41	**91.47**
AA (%)	89.65	77.98	84.07	84.70	86.57	87.06	86.00	**94.83**
OA (%)	94.12	88.17	91.02	91.30	92.26	92.07	91.89	**96.72**
Kappa	0.9258	0.8505	0.8867	0.8903	0.9023	0.9003	0.8978	**0.9587**

The bold numbers indicate the best values for accuracy assessment.

TABLE VIII
QUANTITATIVE COMPARISONS OF DIFFERENT METHODS IN TERMS OF OA, AA, AND KAPPA IN THE WUHAN STUDY AREA

Classes	RF	MLP	MLR	SVM	GRU	LSTM	RNN	FF-C
Soil	95.02	86.07	86.88	94.34	95.15	94.53	95.52	**96.85**
Bright ISA	**96.75**	92.59	92.92	95.74	96.21	95.81	95.63	96.38
Concrete road	86.34	56.79	64.42	81.99	87.13	85.98	87.13	**91.30**
Vegetation	**98.81**	95.78	96.03	97.97	98.52	97.99	97.99	97.51
Dense Building	84.01	52.72	58.67	86.06	87.19	89.38	85.49	**94.83**
Lake	98.21	94.24	95.62	99.02	98.39	98.24	**99.31**	97.85
Asphalt Road	**97.12**	91.58	93.37	95.42	96.59	95.58	92.08	96.37
Algae	88.75	77.94	77.10	87.58	89.13	89.97	92.37	**95.05**
Roof	90.50	82.07	79.18	88.37	93.75	93.09	89.39	**94.66**
River	99.58	99.24	98.25	99.61	98.93	**99.72**	99.46	99.52
AA (%)	93.51	82.90	84.24	92.61	94.10	94.03	93.44	**96.03**
OA (%)	95.84	88.97	89.83	95.02	95.99	95.75	94.95	**96.75**
Kappa	0.9518	0.872	0.882	0.9424	0.9536	0.9509	0.9417	**0.9624**

The bold numbers indicate the best values for accuracy assessment.

TABLE IX
OVERALL ACCURACY OF ISA DISTRIBUTION OF TWO STUDY AREAS FROM DIFFERENT METHODS

	Foshan		Wuhan	
Methods	ISA(%)	NonISA(%)	ISA(%)	NonISA(%)
RF	89.86	65.74	93.65	97.54
MLP	79.06	92.92	83.05	93.59
MLR	84.32	94.51	84.78	93.76
SVM	85.04	94.57	92.21	97.22
GRU	85.94	95.55	94.22	97.38
LSTM	87.36	94.52	93.73	97.33
RNN	86.07	94.92	91.27	97.82
FF-C	**95.42**	97.41	**95.5**	97.72

The bold numbers indicate the best accuracy of ISA extraction.

Tables VII and VIII, it can be seen that for both study areas, the performance of FF-C is generally superior to other classification methods. Due to the higher HS image quality in the Wuhan study area, land cover classification results in the Wuhan study area are better than those in the Foshan study area. The above results demonstrate that the integration of MS and HS data via feature fusion improves the classification accuracy of land cover types.

Figs. 9 and 10 present the land cover classification maps obtained by different methods in the Foshan and Wuhan study areas, respectively. A visual comparison reveals that pixelwise classification methods result in salt-and-pepper noises in classified land use types. In comparison, the proposed FF-C method yields smoother classification maps due to the combination of deep features from MS and HS images that further enhance the model's identification ability. Table IX shows the overall accuracy of extracted ISA from two study areas. The results suggest that FF-C obtains the highest accuracy of ISA. For the Foshan study area the OA from FF-C is higher than RF that obtains the second-highest OA by 6.19%. For the Wuhan study area the OA from FF-C is higher than GRU that obtains the second-highest OA by 1.36%.

The final ISA extraction results obtained by the proposed FF-C method in Foshan and Wuhan study areas are shown in Fig. 13 Table X shows the proportion of each land cover and ISA in two study areas. From Fig. 13 and Table X, we notice that, compared to the Foshan study area, pervious surfaces (e.g., green space and lakes) in the Wuhan study have more extensive coverage. Even though the proportion of water in the Foshan study area is larger, it is mostly used for aquaculture. The Foshan study area is located in Guangdong Province, one of the fastest-growing provinces in China, so its urbanization process is considerably faster than the Wuhan study area.

V. DISCUSSION

In this section, we analyze the effectiveness of the feature fusion strategy by comparing the 1-D CNN and 2-D CNN classification without performing HS and MS feature fusion in Section V-A. Section V-B shows the visual comparison of classification results obtained from different methods. Section V-C discusses the impact of different patch size on classification results.

A. Effectiveness HS and MS Data Feature Fusion

To verify the effectiveness of deep features fusion in land cover classification, we analyze the classification results

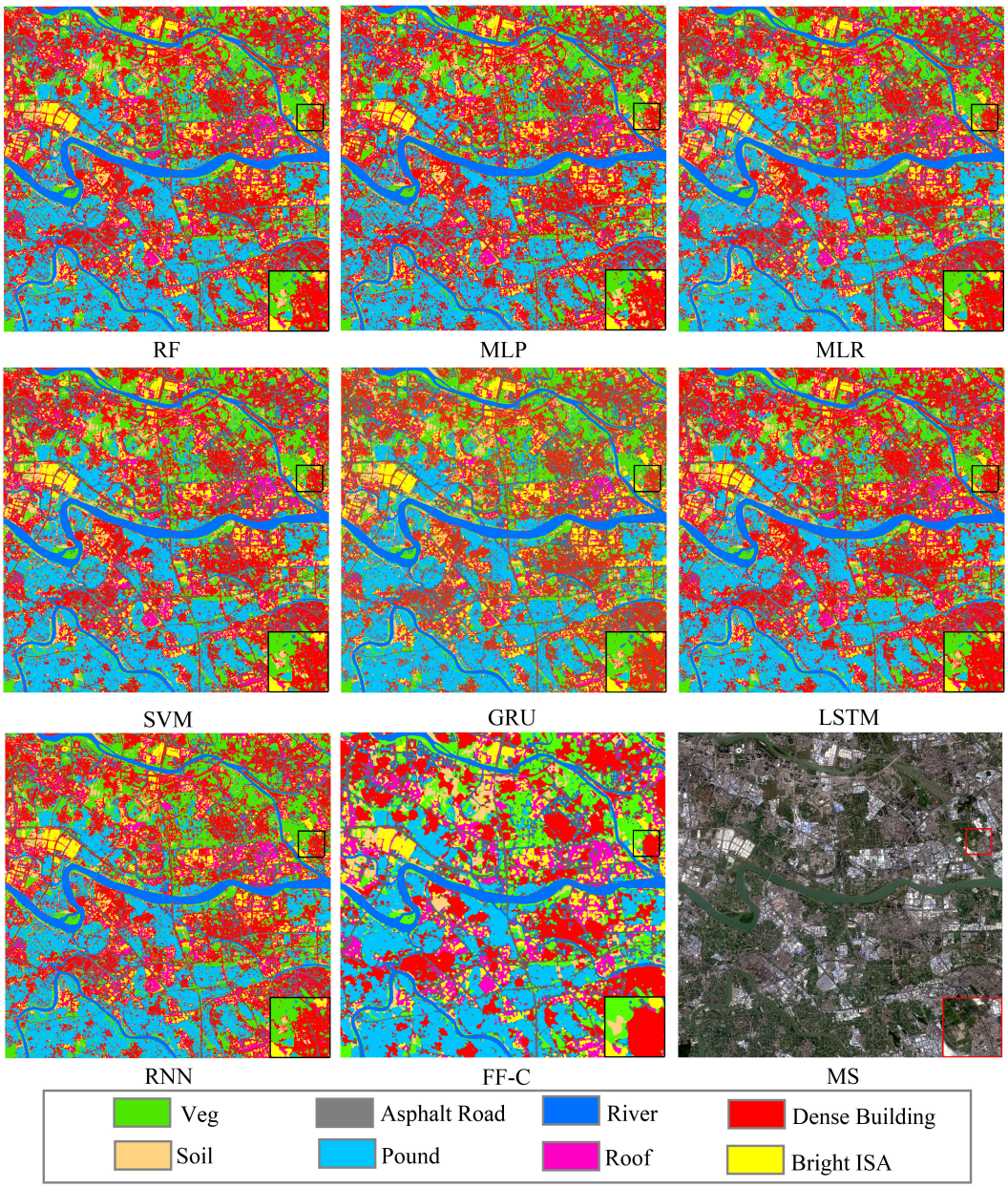

Fig. 9. MS image and classification maps from different methods in the Foshan study area with one demarcated areas zoomed in two times for easy observation.

TABLE X
PROPORTION OF EACH LAND COVER AND ISA IN TWO STUDY AREAS.

The Foshan study area				The Wuhan study area			
Classes	Percent (%)	Classes	Percent (%)	Percent(%)	Classes	Percent	Classes
Soil	5.52	Pervious Surface Area	19.23	31.22	Pervious Surface Area	4.98	Soil
Veg	13.71					22.35	Veg
						3.89	Algae
River	9.3	Water	27.43	21.44	Water	9.96	River
Pound	18.13					11.48	Lake
Asphalt Road	16.65					28.31	Asphalt Road
Roof	9.19					4.72	Concrete road
Dense building	20.67	ISA	53.33	51.22	ISA	7.47	Roof
Bright ISA	6.82					7	Dense building
						3.72	Bright ISA

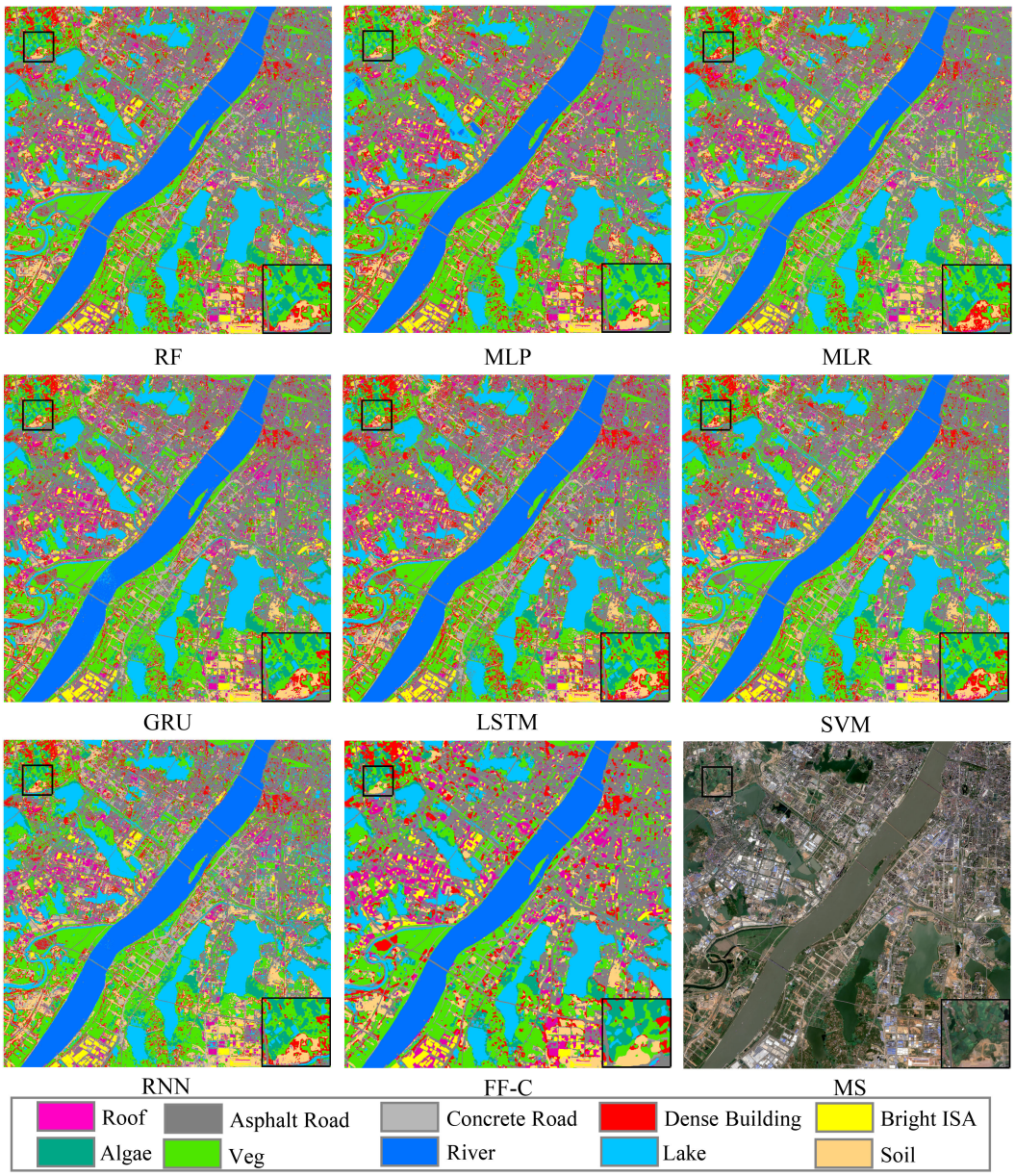

Fig. 10. MS image and classification maps from different methods in the Wuhan study area with one demarcated areas zoomed in two times for easy observation.

obtained by spectral and spatial deep features separately, as well as by the feature fusion method based on the same training, validation, and testing samples. For the deep feature fusion method, we first use 2-D CNN to, respectively, extract the deep features from the HS image and MS image and use fuse the features to explore whether spectral-spatial deep feature fusion can lead to a better land cover classification accuracy. We present the classification results using 1-D CNN-based methods (HS-1-D CNN and MS-1-D CNN) and 2-D CNN-based methods (HS-2-D CNN and MS-2-D CNN) for HS and MS image classification, respectively. 1-D CNN-based methods take pixelwise input, while 2-D CNN-based methods take patchwise input.

Fig. 12 presents the land cover type classification accuracy (AA, OA, and kappa) in the Foshan study area on HS and MS data under the FF-C fusion strategy. The results suggest that 2-D CNN-based methods obtain higher accuracy compared to 1-D CNN-based methods. Furthermore, the FF-C obtains the best classification results for all land cover types except bright ISA and soil. Although the accuracy of bright ISA and soil from FF-C fail to achieve the best results, it is very close to the optimal value. In addition, FF-C obtains the highest AA, OA, and kappa in the Foshan study area. The improvement curve (red lines) in Fig. 12 shows the notable improvement of FF-C in all land cover types, especially in the land cover type of road (an improvement of 0.79).

From Table XI, we observe that the integration of the deep features extracted from MS data leads to improved classification accuracy in all land cover types from the Foshan study area. This not only verifies the effectiveness of the feature fusion strategy on enhancing the feature representation, but also indicates that such an integration of MS and HS data can compensate for the quality deficiency in HS data.

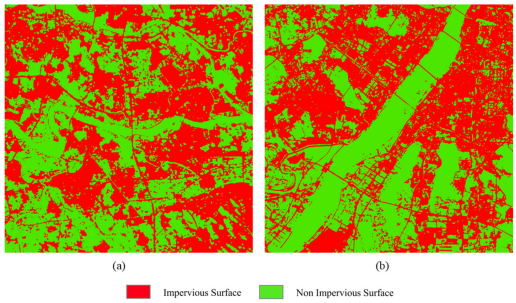

Fig. 11. ISA distribution (from FF-C) in two study areas.

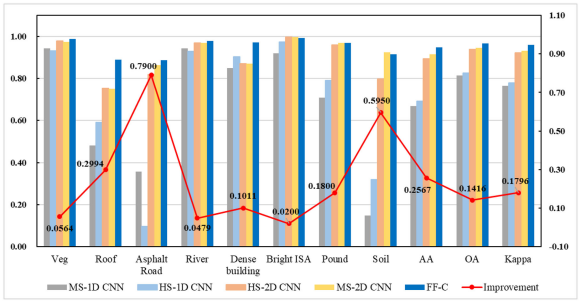

Fig. 12. Land cover type classification accuracy (AA, OA, and kappa) in the Foshan study area on HS and MS data under FF-C fusion strategy. The red curve reveals the improvement comparing the method that fuses HS and MS data to the method that uses HS data alone.

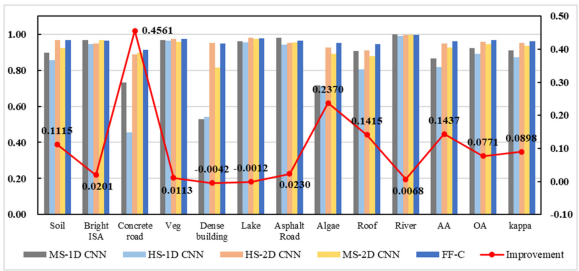

Fig. 13. Land cover type classification accuracy (AA, OA, and kappa) in the Wuhan study area on HS and MS data under FF-C fusion strategy. The red curve reveals the improvement comparing the method that fuses HS and MS data to the method that uses HS data alone.

Fig. 13 presents the land cover type classification accuracy (AA, OA, and kappa) in the Wuhan study area on HS and MS data under the FF-C fusion strategy. Table XI shows the land cover type classification results (UA and PA) in the Wuhan study area based on 1-D/2-D CNN (HS images alone) and feature fusion strategies (HS and MS images). The results reveal that the land cover type of concrete road achieves the greatest improvement in accuracy by 0.4561. The experimental results from our two study areas classification demonstrate the effectiveness of integrating MS data with HS data when performing land cover classification. The deep features from MS data might enhance the spatial information from HS data, thus leading to better classification performance when MS and HS data are fused.

TABLE XI
LAND COVER TYPE CLASSIFICATION RESULTS (UA AND PA) IN THE FOSHAN STUDY AREA BASED ON 1D/2D CNN (HS IMAGES ALONE) AND FF-C (HS AND MS IMAGES)

Method	HS-1D CNN		HS-2D CNN		FF-C	
Classes	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)
Vegetation	91.17	93.18	97.75	97.77	98.27	98.7
Roof	82.9	59.04	92.53	75.38	93.52	88.85
Asphalt Road	53.47	9.64	75.48	81.86	86.48	88.64
River	94.87	92.92	98.22	96.94	98.61	97.71
Dense Building	57.73	90.24	92.41	86.99	95.65	97.1
Bright ISA	94.12	97.26	94.63	99.58	99.02	99.18
Pound	75.12	78.97	93.84	95.95	95.98	96.97
Soil	66.71	31.97	72.03	79.82	93.88	91.47
OA (%)	82.56		93.69		**95.87**	

TABLE XII
LAND COVER TYPE CLASSIFICATION RESULTS (UA AND PA) IN THE WUHAN STUDY AREA BASED ON 1D/2D CNN (HS IMAGES ALONE) AND FF-C (HS AND MS IMAGES)

Method	HS 1D CNN		HS 2D CNN		FF-C	
Classes	UA (%)	PA (%)	UA (%)	PA (%)	UA (%)	PA (%)
Soil	88.03	85.7	95.99	96.6	97.71	96.85
Bright ISA	86.13	94.38	97.34	94.96	98.14	96.38
Concrete road	58.06	45.7	91.67	88.82	92.87	91.3
Vegetation	92.38	96.38	96.4	97.33	97.1	97.51
Dense Building	66.99	54.14	91.75	95.26	95.24	94.83
Lake	97.43	95.59	96.79	97.97	97.94	97.85
Asphalt Road	85.51	94.07	94.09	95.14	95.23	96.37
Algae	89.41	71.35	94.98	92.48	94.49	95.05
Roof	88.98	80.51	95.41	91.04	94.9	94.66
River	98.54	98.84	99.46	99.52	99.72	99.52
OA (%)	89.04		95.89		**96.24**	

Comparing results from these two study areas, we notice that the improvement in the Wuhan study area is not as notable as the one in the Foshan study area. This is because, with similar SNRs of the MS data, the Wuhan study area has a higher quality HS image (SNR = 35.36 dB) than that in the Foshan study area (SNR = 28.45 dB). Thus, the designed feature enhancement model has a less notable impact in the Wuhan study area. We notice that the SNRs in MS of both study areas are around 150 dB, twice higher than that of HS data. This means MS data contains more spatial information of ground objects than HS data. Given that the 2-D CNN can extract the deep features from images by considering the contextual information in both spatial and spectral domains, fusing the spectral and spatial deep features extracted from HS image and MS image is able to improve classification performance.

B. Visual Comparison

This section presents the details in classification maps from different methods in the two study areas. For the Foshan study area (see Fig. 14), the highlighted black rectangle is dominated by bare soil and vegetation, belonging to the pervious surface; while the classified category from the comparison method is dense building, belonging to impervious surface. Such misclassification leads to reduced ISA extraction accuracy and overestimation of ISA. For the Wuhan study area (see Fig. 15), the highlighted black rectangle is dominated by bare soil, which is wrongly classified into dense buildings and concrete roads in the comparison method. Overall, it can be seen that the classification from the proposed FF-C method can improve the accuracy of features recognition, thus obtaining more accurate ISA distribution information.

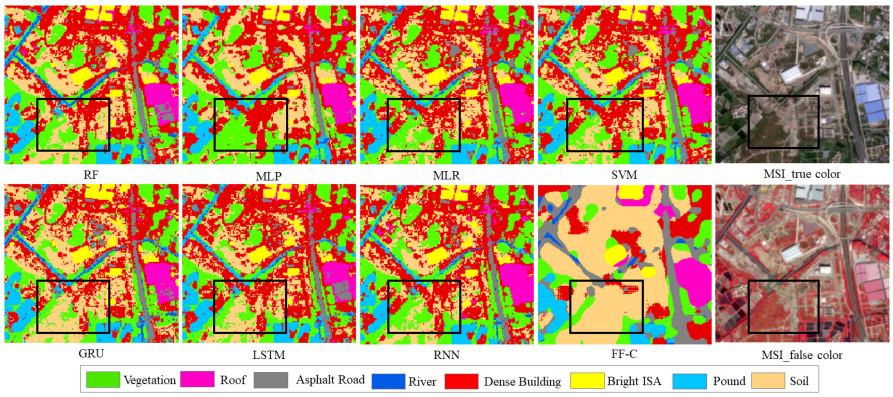

Fig. 14. Selected classification results of the Foshan study area.

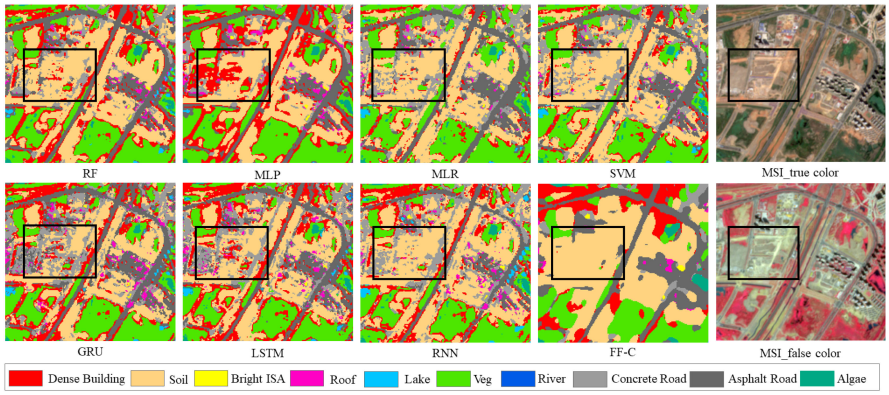

Fig. 15. Selected classification results of the Wuhan study area.

TABLE XIII
LAND COVER TYPE CLASSIFICATION RESULTS IN THE FOSHAN STUDY AREA BASED ON FF-C WITH DIFFERENT PATCH SIZE

Classes	13	15	17	19	21	23	25	27	29	31
Vegetation	98.33	98.41	98.37	98.53	98.45	98.74	98.74	98.70	98.45	98.10
Roof	86.67	87.44	87.12	87.63	88.08	88.65	88.08	88.85	88.14	84.87
Asphalt Road	87.89	86.27	87.89	87.00	87.23	87.52	86.50	88.64	85.39	85.12
River	98.10	98.04	97.99	97.92	97.73	97.74	97.71	97.71	97.25	96.99
Dense building	93.55	93.82	93.95	93.97	93.79	95.15	94.74	97.10	96.82	93.75
Bright ISA	98.96	99.20	98.86	98.93	98.93	98.98	99.07	99.18	98.75	98.26
Pound	96.82	97.02	96.44	96.46	96.50	97.05	97.12	96.97	97.21	96.51
Soil	75.24	76.63	80.51	80.10	83.29	84.81	86.48	91.47	90.36	77.95
AA (%)	91.94	92.1	92.64	92.57	93.00	93.58	93.55	**94.83**	94.05	91.44
OA (%)	95.70	95.75	95.82	95.77	95.81	96.19	96.12	**96.72**	96.20	95.01
Kappa	0.9458	0.9464	0.9474	0.9468	0.9472	0.9520	0.9511	**0.9587**	0.9404	0.9372

C. Patch Size Analysis

The size of input patches is an important parameter that determines, to a certain degree, the classification accuracy of the model. To explore the influence of patch sizes on the classification performance, we conduct additional experiments in the Foshan study area. Table XIII shows the land cover classification results in the Foshan study area based on FF-C with different patch sizes. We observe improved classification accuracy with the increase in the patch sizes from 13 to 27, especially for soil (an improvement by 16.23%).

Fig. 16 presents the OA of classification results corresponding to different patch sizes. It can be seen that when the patch size reach 27 pixel, the accuracy is the highest. Therefore, the patch size is set to 27 in our experiments.

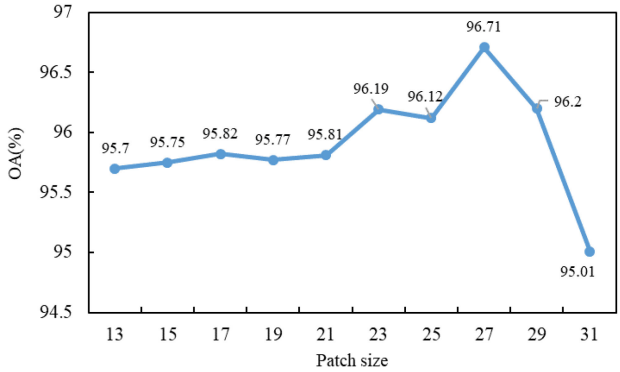

Fig. 16. Overall accuracy (%) with different patch sizes in the Foshan study area.

VI. Conclusion

In this study, we propose a 2-D CNN-based method to improve the accuracy of ISA extraction at 10 m spatial resolution by combining Sentinel-2 MS data and Zhuhai-1 HS data. We test our proposed approach in two study areas that cover Foshan and Wuhan city, China. We first utilize 2-D CNN to extract the spatial and spectral deep features of MS data and HS data, then fuse the extracted deep features via a fully connected layer for the final classification. To investigate the influence of the fusion method on the final results, we compare the feature fusion strategies with other comparison methods. The results prove the superiority of feature fusion methods compared to nonfusion methods. In the future, we plan to explore the impact of model depths on image feature extraction and develop more advanced fusion modules to take full advantage of the detailed spectral information from HS images and detailed spatial information from MS images. In addition, we plan to test the proposed method in other regions to further investigate its generalizability.

Acknowledgment

The authors would like to the anonymous reviewers, for their valuable suggestions and comments that helped us improve this article significantly.

References

[1] Q. Weng, "Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends," *Remote Sens. Environ.*, vol. 117, pp. 34–49, 2012.

[2] C. Li, Z. Shao, L. Zhang, X. Huang, and M. Zhang, "A comparative analysis of index-based methods for impervious surface mapping using multiseasonal sentinel-2 satellite data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3682–3694, Mar. 2021.

[3] A. J. Arnfield, "Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island," *Int. J. Climatol., A J. Roy. Meteorological Soc.*, vol. 23, no. 1, pp. 1–26, 2003.

[4] P. Coseo and L. Larsen, "How factors of land use/land cover, building configuration, and adjacent heat sources and sinks explain urban heat islands in Chicago," *Landscape Urban Plan.*, vol. 125, pp. 117–129, 2014.

[5] K. Conway, J. Barrie, P. Hill, W. Austin, and K. Picard, "Mapping sensitive benthic habitats in the strait of Georgia, coastal British Columbia: Deep-water sponge and coral reefs," *Geol. Surv. Can.*, vol. 2, pp. 1–6, 2007.

[6] H. Du *et al.*, "Influences of land cover types, meteorological conditions, anthropogenic heat and urban area on surface urban heat island in the yangtze river delta urban agglomeration," *Sci. Total Environ.*, vol. 571, pp. 461–470, 2016.

[7] Z. Shao, H. Fu, D. Li, O. Altan, and T. Cheng, "Remote sensing monitoring of multi-scale watersheds impermeability for urban hydrological evaluation," *Remote Sens. Environ.*, vol. 232, 2019, Art. no. 111338.

[8] X.-P. Song, J. O. Sexton, C. Huang, S. Channan, and J. R. Townshend, "Characterizing the magnitude, timing and duration of urban growth from time series of landsat-based estimates of impervious cover," *Remote Sens. Environ.*, vol. 175, pp. 1–13, 2016.

[9] L. Zhang, Q. Weng, and Z. Shao, "An evaluation of monthly impervious surface dynamics by fusing landsat and modis time series in the Pearl river delta, China from 2000 to 2015," *Remote Sens. Environ.*, vol. 201, pp. 99–114, 2017.

[10] D. Lu and Q. Weng, "Spectral mixture analysis of the urban landscape in indianapolis with landsat ETM imagery," *Photogrammetric Eng. Remote Sens.*, vol. 70, no. 9, pp. 1053–1062, 2004.

[11] X. Huang, D. Wen, J. Li, and R. Qin, "Multi-level monitoring of subtle urban changes for the megacities of China using high-resolution multi-view satellite imagery," *Remote Sens. Environ.*, vol. 196, pp. 56–75, 2017.

[12] S. Roessner, K. Segl, U. Heiden, and H. Kaufmann, "Automated differentiation of urban surfaces based on airborne hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1525–1532, Jul. 2001.

[13] A. Okujeni, S. van der Linden, and P. Hostert, "Extending the vegetation-impervious-soil model using simulated enmap data and machine learning," *Remote Sens. Environ.*, vol. 158, pp. 69–80, 2015.

[14] B. Feng and J. Wang, "Constrained nonnegative tensor factorization for spectral unmixing of hyperspectral images: A case study of urban impervious surface extraction," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 4, pp. 583–587, Apr. 2019.

[15] F. Chen, K. Wang, T. Van de Voorde, and T. F. Tang, "Mapping urban land cover from high spatial resolution hyperspectral data: An approach based on simultaneously unmixing similar pixels with jointly sparse spectral mixture analysis," *Remote Sens. Environ.*, vol. 196, pp. 324–342, 2017.

[16] A. H. Strahler, "The use of prior probabilities in maximum likelihood classification of remotely sensed data," *Remote Sens. Environ.*, vol. 10, no. 2, pp. 135–163, 1980.

[17] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[18] V. Vapnik, *The Nature of Statistical Learning Theory*. Cham, Switzerland: Springer, 2013.

[19] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[20] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.

[21] S. Schulter, P. Wohlhart, C. Leistner, A. Saffari, P. M. Roth, and H. Bischof, "Alternating decision forests," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 508–515.

[22] E. Tuv, A. Borisov, G. Runger, and K. Torkkola, "Feature selection with ensembles, artificial variables, and redundancy elimination," *J. Mach. Learn. Res.*, vol. 10, pp. 1341–1366, 2009.

[23] P. Peng, Q.-L. Ma, and L.-M. Hong, "The research of the parallel SMO algorithm for solving SVM," in *Proc. Int. Conf. Mach. Learn. Cybern.*, 2009, vol. 3, pp. 1271–1274.

[24] P.-H. Chen, R.-E. Fan, and C.-J. Lin, "A study on SMO-type decomposition methods for support vector machines," *IEEE Trans. Neural Netw.*, vol. 17, no. 4, pp. 893–908, Jul. 2006.

[25] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[26] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[27] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.

[28] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.

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

[30] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 609–616.

[31] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.

[32] P. Ghamisi, Y. Chen, and X. X. Zhu, "A self-improving convolution neural network for the classification of hyperspectral data," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 10, pp. 1537–1541, Oct. 2016.

[33] Y. Chen, Y. Wang, Y. Gu, X. He, P. Ghamisi, and X. Jia, "Deep learning ensemble for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1882–1897, Jun. 2019.

[34] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, Jun. 2017.

[35] T. Blaschke, "Object based image analysis for remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 65, no. 1, pp. 2–16, 2010.
[36] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
[37] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1026–1034.
[38] Z. Zuo *et al.*, "Learning contextual dependence with convolutional hierarchical recurrent neural networks," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 2983–2996, Jul. 2016.
[39] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.
[40] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6690–6709, Sep. 2019.
[41] J. M. Haut, M. E. Paoletti, J. Plaza, J. Li, and A. Plaza, "Active learning with convolutional neural networks for hyperspectral image classification using a new Bayesian approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6440–6461, Nov. 2018.
[42] X. Yang, Y. Ye, X. Li, R. Y. Lau, X. Zhang, and X. Huang, "Hyperspectral image classification with deep learning models," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5408–5423, Sep. 2018.
[43] L. Jiao, M. Liang, H. Chen, S. Yang, H. Liu, and X. 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.
[44] J. M. Nascimento and J. M. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898–910, Apr. 2005.
[45] W. Li, R. Dong, H. Fu, J. Wang, L. Yu, and P. Gong, "Integrating google earth imagery with landsat data to improve 30-m resolution land cover mapping," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111563.
[46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[47] M. Paoletti, J. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 279–317, 2019.

Xiaoxiao Feng received the bachelor's degree in surveying and mapping from Southeast University, Nanjing, China, in 2014, and the master's degree in earth exploration and information technology from the China University of Geosciences, Wuhan, China, in 2017, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China in 2021.

He is currently a Ph.D. student with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University. Her research interests include hyperspectral image processing and urban impervious surface extraction.

Zhenfeng Shao received the bachelor's in surveying engineering and master's degrees in cartography and geographical information system from Wuhan Technical University of Surveying and Mapping, Wuhan, China, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China.

He is currently a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China. His research interest mainly focuses on urban remote sensing applications. The specific research directions include high-resolution remote sensing image processing and analysis, key technologies and applications from digital cities to smart cities and sponge cities.

Xiao Huang received the bachelor's degree in remote sensing and information engineering from Wuhan University of China, Wuhan, in 2015, the master's degree in city planning and architecture from the Georgia Institute of Technology China, Shenzhen, China, in 2016, and the Ph.D. degree in geography from the University of South Carolina, Columbia, SC, USA, in 2020.

He is currently an Assistant Professor with the Department of Geosciences, University of Arkansas, Fayetteville, AR, USA. His research interests include remote sensing and GIS in natural hazards, data-driven visualization and advanced data fusion flood models, big social data mining, regional geospatial analysis, remote sensing, and GeoAI.

Luxiao He received the bachelor's degree in geo-information science and technology and the master's degree in earth exploration and information technology from the China University of Geosciences, Wuhan, China, respectively, in 2014 and 2017, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China in 2021.

He is currently working toward the Ph.D. degree with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China.

His research interests include high spatial resolution image processing and application.

Xianwei Lv received the bachelor's degree in geographic information science from the East China University of Science and Technology, Nanchang, China, in 2016 and the master's degree in surveying and mapping from the China University of Geosciences, Beijing, China, in 2019. He is currently working toward the Ph.D. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China.

He does research in deep learning for very high-resolution image processing and applications.

Qingwei Zhuang received the bachelor's degree in surveying and mapping from Henan Polytechnic University, Henan, China, in 2017, and the master's degree in surveying and mapping from the University of Chinese Academy of Sciences, Beijing, China, in 2020. He is currently working toward the Ph.D. degree in photogrammetry and remote sensing with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China.

His research interest mainly focuses on remote sensing applications. The specific research directions include remote sensing image processing and analysis, key technologies and applications in urban ecosystem.