A Novel Multitemporal Deep Fusion Network (MDFN) for Short-Term Multitemporal HR Images Classification

Yongjie Zheng, *Student Member, IEEE*, Sicong Liu, *Senior Member, IEEE*, Qian Du, *Fellow, IEEE*, Hui Zhao, *Student Member, IEEE*, Xiaohua Tong, *Senior Member, IEEE*, and Michele Dalponte, *Senior Member, IEEE*

***Abstract*—High-resolution (HR) satellite images, due to the technical constraints on spectral and spatial resolutions, usually contain only several broad spectral bands but with a very high spatial resolution. This provides rich spatial details of the objects on the Earth surface, while their spectral discrimination is relatively low. Recently, the increase of the satellite revisit times made it possible to acquire more frequent data coverage for finer classification. In this article, we proposed a novel multitemporal deep fusion network (MDFN) for short-term multitemporal HR images classification. Specifically, a two-branch structure of MDFN is designed, which includes a long short-term memory (LSTM) and a convolutional neural network (CNN). The LSTM branch is mainly used to learn the joint expression of different temporal-spectral features. For the CNN branch, the three-dimensional (3-D) convolution is firstly applied along the temporal and spectral dimensions to jointly learn the temporal-spatial and spectral-spatial information, respectively, and then the 2-D convolution is performed along the spatial dimension to further extract the spatial context information. Finally, features generated from the two different branches are fused to obtain the discriminative high-level semantic information for classification. Experimental results carried on two real multitemporal HR remote sensing datasets demonstrate that the proposed MDFN provides better classification performance over the state-of-the-art methods, and it also shows the potentiality to use short-term multitemporal HR images for more accurate land use/land cover mapping.**

***Index Terms*—Convolutional neural network (CNN), deep feature fusion, land use/land cover (LULC) classification, long short term memory (LSTM), multitemporal images.**

Manuscript received August 31, 2021; revised September 20, 2021; accepted October 7, 2021. Date of publication October 14, 2021; date of current version November 1, 2021. This work was supported in part by the National Key R&D Program of China under Grant 2018YFB0505000, in part by the Natural Science Foundation of China under Grant 42071324, and in part by the Shanghai Rising-Star Program under Grant 21QA1409100. *(Corresponding author: Sicong Liu.)*

Yongjie Zheng, Sicong Liu, Hui Zhao, and Xiaohua Tong are with the College of Surveying and Geoinformatics, Tongji University, Shanghai 200092, China (e-mail: yongjie@tongji.edu.cn; sicong.liu@tongji.edu.cn; zhaohui@tongji.edu.cn; xhtong@tongji.edu.cn).

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

Michele Dalponte is with the Department of Sustainable Agro-Ecosystems and Bioresources, Research and Innovation Centre Fondazione E. Mach, 38010 San Michele all'Adige, Italy (e-mail: michele.dalponte@fmach.it).

Digital Object Identifier 10.1109/JSTARS.2021.3119942

I. Introduction

With the rapid development of the earth observation (EO) technology, high resolution or even very-high-resolution (VHR) satellites [e.g., Pleiades, Gaofen (GF), and Worldview series] have been launched with a revisit time about 1 to 5 days (see Table I). Such high temporal resolution allows the adequate analysis of monotemporal or multitemporal images within a given period (e.g., short-term, medium-term, or long-term) [1], [2]. The increasing availability of HR remote sensing images offer great potential and opportunities for the applications, such as land use/land cover (LULC) mapping [3]–[5], forest monitoring [6], disaster evaluation [7] and urban management [8], etc.

Unlike hyperspectral (HS) images that provide a high spectral resolution, HR images usually have a relatively coarse spectral resolution but rich spatial details. Thus, it is quite difficult to accurately identify the rough spectral difference of different objects, especially for subcategories corresponding to the same major-category (e.g., grass, farmland, and trees that belong to the vegetation category) [9]. In addition, the phenomenon of misclassification is severe as the fact that these optical images are susceptible to clouds, shadows, illumination, and atmospheric reflection conditions, etc. In the latest multispectral HR satellite missions, the revisit frequency has improved leading to data with a high temporal resolution. Therefore, there is a very high potential in using short-term multitemporal information for HR image classification. How to effectively combine spatio-temporal-spectral information provided by such HR images becomes an open challenging question [4], [10], [11].

In the past decades, many spectral-spatial methods have been proposed for HR image classification by taking advantages of its rich context information. For example, popular methods developed based on the direct spectral-spatial feature extraction (e.g., Gabor filtering [12], edge-preserving filtering [5], [13], extended morphological profiles [14], and extended attribute profiles [15]), the multiple kernel learning [16], [17], the Markov random fields [18], [19] and the superpixel processing [20] achieve successful results in various applications in the literature. Despite their effectiveness in combining spectral-spatial features, these traditional machine learning methods usually generate many misclassification errors especially in edge details

TABLE I
EXAMPLES OF THE MOST WIDELY-KNOWN HR SATELLITES

Satellites	Number of Bands	Ground sample distance (m)		Repetition Cycle (d)
		Panchromatic	Multispectral	
Ziyuan-3/01	4	-	6	<3
Ziyuan-3/02	4	-	5.8	
Gaofen-2	4	0.8	3.2	5
Gaofen-1	4	2	8	4
Gaofen-6	4	2	8	4
Pleiades-1	4	0.5	2	1
Pleiades-2	4	0.5	2	
WordView-2	8	0.5	1.8	1.1
WordView-3	8	0.31	1.24	1
WordView-4	4	0.31	1.24	1
PlanetScope	4	-	3	1

and objects homogeneity due to limited feature representation and generalization abilities [21].

Recently, deep learning (DL) based methods have demonstrated their outstanding performance for remote sensing image classification [22]–[24]. Among them, the convolutional neural network (CNN) and the recurrent neural network (RNN) have been widely used. By taking into account the characteristics of remote sensing images, researchers have designed several spectral-spatial or shallow-deep feature fusion networks, such as the fast dense spectral–spatial convolution network (FDSSC) [25], the spectral-spatial unified network (SSUN) [26], and the spectral–spatial residual network (SSRN) [27], etc. With the robust capability of automatic feature learning, DL-based methods can adaptively exploit more complex and effective high-level features to enhance the classification performance [28].

Despite the success of the aforementioned classification methods, they were all designed for monotemporal remote sensing images. In reality, optical HR images are frequently affected by complex acquisition situations such as clouds contamination, illumination conditions and image bad strips, which leads to inaccurate mapping of ground materials. On the other hand, the limited spectral information in HR images leads to a difficulty with separating certain LULC classes with similar spectral signatures. Even for the common DL-based methods, most of them usually rely on high-level features in the last layer for classification. For example, information for the fully connected layer in the CNN has often undergone many down-sampling operations, which may not be suitable to accurately describe small objects details. The lack of a dominant spatial information with coarse spectral information would greatly reduce the applicability of multispectral HR image data.

Considering the above problems in monotemporal multispectral HR image classification, a joint classification of multitemporal HR images expands the spectral representation towards the temporal domain, increasing the possibility for more accurate LULC classification. Moreover, for typical ground objects (e.g., buildings, roads, trees, and water), they are usually stable within a certain period, which also makes the short-term multitemporal HR images useful.

In the literature, there are existing works on the classification of multitemporal HR images. As an example, in [29], multitemporal texture features (pseudocross variogram) were directly combined with the original bands for multitemporal images classification. In [30], some bands and some temporal normalized difference vegetation index were extracted from Sentinel-2 multispectral images time series for land and crop classification using the random forest (RF) classifier. In [31], a manifold alignment framework was proposed to leverage prior knowledge while exploiting spectral similarities in the underlying manifolds of two multitemporal HS images. Similarly, [32] introduced a novel semisupervised kernel manifold alignment (KEMA) method, which successfully applied KEMA to multitemporal and multisource VHR classification tasks. In [33], a two-stage classifier was proposed for classifying cloud- and snow-contaminated multitemporal images. The proposed classifier mainly combined multitemporal information to improve the missing data and then performed classification. In addition, some researchers also proposed DL-based methods for multitemporal images classification. For example, a temporal-attention CNN and gated recurrent unit (CNN-GRU) approach was proposed in [34] to distinguish subtle crop differences. A DL-based architecture namely twin neural networks for sentinel data (TWINNS) was proposed in [11] to boost the LULC classification task using radar and optical satellite images.

Despite the many existing studies on multitemporal classification methods there are still several problems and challenges.

1) Most existing approaches unitized traditional medium resolution images with a relatively long revisit time. There are few works on the short-term multitemporal HR images. The complementarity between spatio-temporal and spectral dependencies had not been fully considered in these methods, which will lead to the presence of many misclassifications.
2) Most studies still relied on traditional hand-crafted features, which are not able to extract the high-level discriminative feature representation. In addition, there were many misclassification errors because the fact that they did not further explore the invariant temporal-spectral features to suppress the abrupt or abnormal changes on each monotemporal image.

To overcome these drawbacks, in this article, a novel framework named multitemporal deep fusion network (MDFN) is proposed for dealing with the short-term multitemporal HR images classification, where long short-term memory (LSTM) and CNN branches are combined to extract and fuse rich spatio-temporal-spectral features. The main contributions of this article are summarized as follows.

1) By integrating LSTM and CNN branches, the unified spatio-temporal-spectral features are extracted and fused at different layers, and in this way the invariant temporal-spectral features combined with multiple short-term temporal images can be greatly benefited. Furthermore, rich and complex high-level information is generated for the classification by fusing features at different layers without the pooling or other down-sampling operations.
2) The three-dimensional (3-D) convolutions toward the spectral dimension and temporal dimension are designed to learn the rich spectral-spatial and temporal-spatial correlation information, and then reduce the misclassification errors caused by clouds, shadows, illumination, and

atmospheric reflection conditions on each monotemporal image.

3) The 2-D convolutions based on concatenating two types of 3-D convolution features are designed to capture the spatial context information. This guarantees a strong descriptive capability for the invariant spatio-temporal-spectral features that contribute to the improved classification.

The rest of this article is organized as follows. Section II introduces the related techniques including LSTM and CNN. Section III provides a detailed description of the proposed MDFN framework. Section IV presents the experimental results obtained on two real multitemporal HR image datasets. Finally, Section V draws the conclusion and discusses future research.

II. Related Work

A. Long Short-Term Memory

LSTM is a special structure of RNN and it is capable to learn long-term dependencies and deal with the gradient vanishing or the exploding problems present in the traditional RNN [35], [36]. It works well to solve a large variety of problems in the temporal or spectral domains, and is successfully used for HS classification. The neurons of the LSTM not only receive information from other neurons, but also receive their own information with feedback loops. Therefore, the LSTM with "memory" makes it more suitable for time series analysis compared with other networks such as CNN.

The key concepts of the LSTM are the cell state and the gating mechanism. In particular, the data transmission and processing in LSTM are realized by three key gate units: the forget gate f_t, the input gate i_t, and the output gate o_t, which are used for implementing information protection and control [36]:

1) Forget Gate f_t: It is mainly used to control the amount of information that needs to be forgotten respect to the previous moment. Its mathematical model can be formulated as follows

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{1}$$

where W_f and b_f represent the weight and bias of f_t, respectively, x_t is the candidate state at current time t, and σ is the nonlinear activation function, such as Tanh, Sigmoid, and ReLU.

2) Input Gate i_t: It is used to control the amount of information that needs to be saved in x_t. The specific formulas are defined as

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{2}$$

$$C_t{}' = \tanh(W_C[h_{t-1}, x_t] + b_C) \tag{3}$$

where W_i and b_i represent the weight and bias of i_t, respectively, $C_t{}'$ is the new candidate value generated by the tan$h$ function, and W_C and b_C represent its weight and bias, respectively.

3) Output Gate o_t: It represents the initial output, and h_t is the final output which is obtained by multiplying the C_t compressed to the $[-1, 1]$ state with o_t. These mathematical formulas of o_t and h_t can be defined as

$$C_t = f_t \times C_{t-1} + i_t \times C_t{}' \tag{4}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t \times \tan h(C_t) \tag{6}$$

where C_t is the new cell state information at time t, which can be considered as an intermediate state saved by LSTM, W_o and b_o represent the weight and bias of o_t, respectively.

B. Convolutional Neural Network

The CNN is a popular type of neural network, which uses at least one convolution layer in the framework [37], [38]. In general, the CNN mainly consists of four components, i.e., the convolution layer, the pooling layer, the batch normalization layer and the fully connected layer. According to the dimension of the convolution layer, CNNs can be divided into 1-D-CNN, 2-D-CNN and 3-D-CNN.

1) 1-D-CNN: it is usually designed to use the pixel vector along the radiometric dimension to extract deep features, which can be conceptually considered as a spectral-based classification approach [39], [40].

2) 2-D-CNN: the convolution kernels of a 2-D-CNN move along the height and width directions of an image, and extract multilevel features (e.g., shallow, medium, and deep features) from specified local neighborhoods. The shallow layers usually extract some low-level features like edges and textures, while the deep layers generate more abstract and discriminative high-level features. For a 2-D convolution layer, the convolution value Y of the given pixel (i, j) on the kth channel in the lth convolution layer can be obtained as

$$Y_{i,j}^{k,l} = f\left[(W_{k,l} * X^{l-1})_{i,j} + b_{k,l}\right] \tag{7}$$

where b is the bias of the convolutional layer, f is the nonlinear activation function, (i, j) represents the coordinate of a given pixel, and X^{l-1} represents the set of input sensors (and is also the set of output in the l-1th layer [41]).

2) 3-D-CNN: different from the 2-D-CNN that extracts image features from the spatial direction, a 3-D-CNN extends the 2-D convolution into a 3-D convolution along the channel direction. It performs convolution along the height, width, and channel dimensions of the input image. In case of multiframe videos and multitemporal images, the 3-D convolution is capable to learn time-series dependencies. For a 3-D convolution layer, the output value Y at the pixel location (i, j, r) on the kth channel in the lth convolution layer can be generated as

$$Y_{i,j,r}^{k,l} = f\left[(W_{k,l} * X^{l-1})_{i,j,r} + b_{k,l}\right] \tag{8}$$

III. Proposed MDFN Framework

The technical framework of the proposed MDFN is shown in Fig. 1. Its two-branch structure includes: the LSTM branch that is used for modeling the spectral-temporal dependencies of short-term multitemporal HR images; and the CNN branch with CONV3_T, CONV3_S and CONV2 modules. It is used for learning the rich spatio-temporal-spectral information at different levels. Note that in order to avoid the down-sampling effect that impacts on the classification of ground objects, there is no pooling layer in the CNN branch.

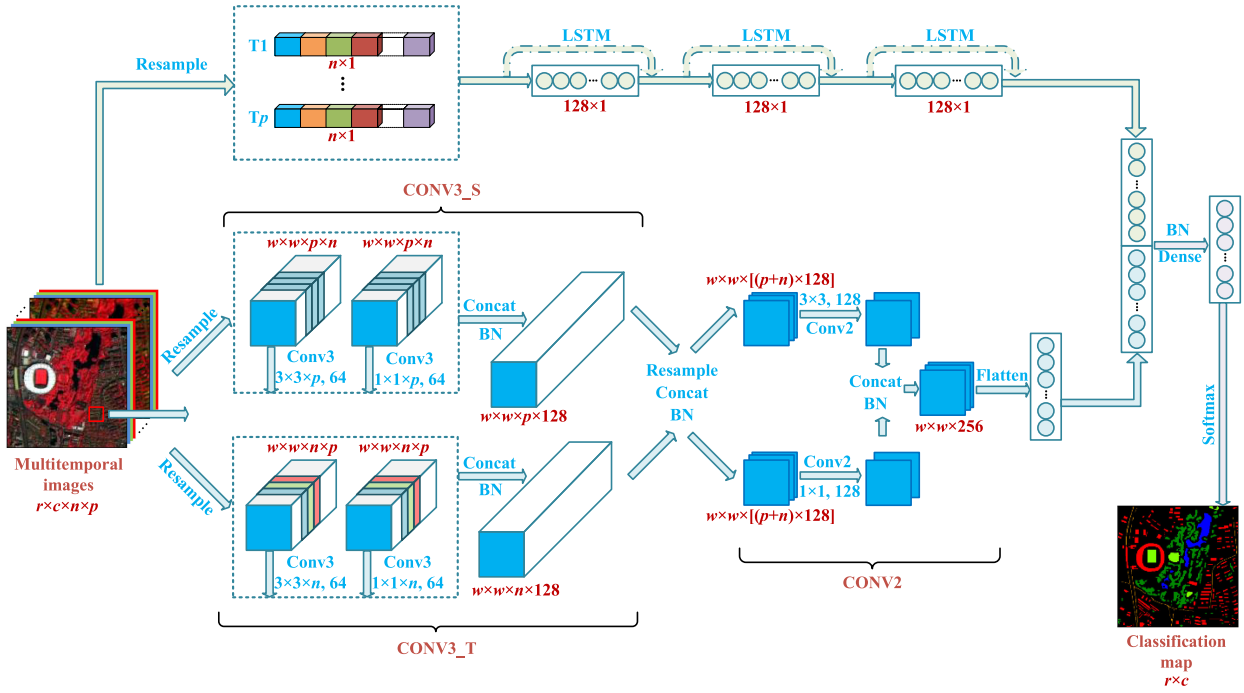

Fig. 1. Flowchart of the proposed MDFN for short-term multitemporal HR images classification.

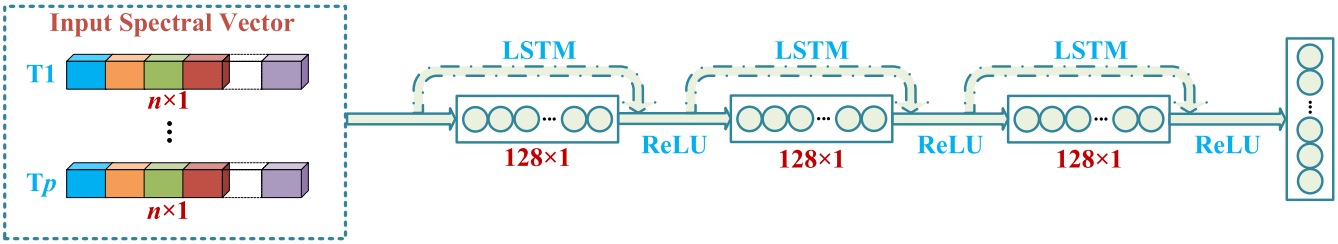

Fig. 2. Structure of LSTM branch in the proposed MDFN.

A. LSTM Branch

Based on the characteristics of the multitemporal images, the LSTM branch is designed to model the temporal correlation of the multitemporal spectral information. This branch makes full use of the spectral-temporal dependencies of the short-term multitemporal HR images, which is essentially conducive to a high-precision identification of different ground objects. Fig. 2 shows the structure of the LSTM branch in the proposed MDFN. By considering the number of HR images, three LSTM layers are set to extract the discriminative and invariant temporal-spectral features. In addition, the number of filter kernels is set to 128, and the effective ReLU is selected as its nonlinear activation function.

B. CNN Branch

To reduce the spectral inconsistence due to abnormal changes caused by complex acquisition situations, such as clouds contamination, illumination conditions and striping noise occurred in the monotemporal images, the CNN branch is designed to jointly use 3-D–2-D convolution layers. Specifically, 3-D convolution layers include two paralleled modules, i.e., the 3-D convolution along the multitemporal direction (denoted as CONV3_T) and the 3-D convolution along the spectrum direction (denoted as CONV3_S). CONV3_S is effective to capture the invariant information of the same object in the spatio-temporal-spectral dimension, and CONV3_T can learn the discriminative information between different or similar objects in the spatio-spectral-temporal dimension. This results in a more comprehensive and reasonable feature representation in the multitemporal classification domain. Finally, results of the CONV3_T and CONV3_S modules are imported to the 2-D convolution module (denoted as CONV2) to further learn the spatial context information based on different receptive fields.

1) *CONV3_T*: as shown in Fig. 3, the CONV3_T module includes two 3-D convolution layers: 64 $3 \times 3 \times n$ and 64 $1 \times 1 \times n$ convolution kernels to capture the discriminative temporal-spatial information at two spatial scales (*i.e.*, 3×3 and 1×1). Let the raw multispectral HR images be

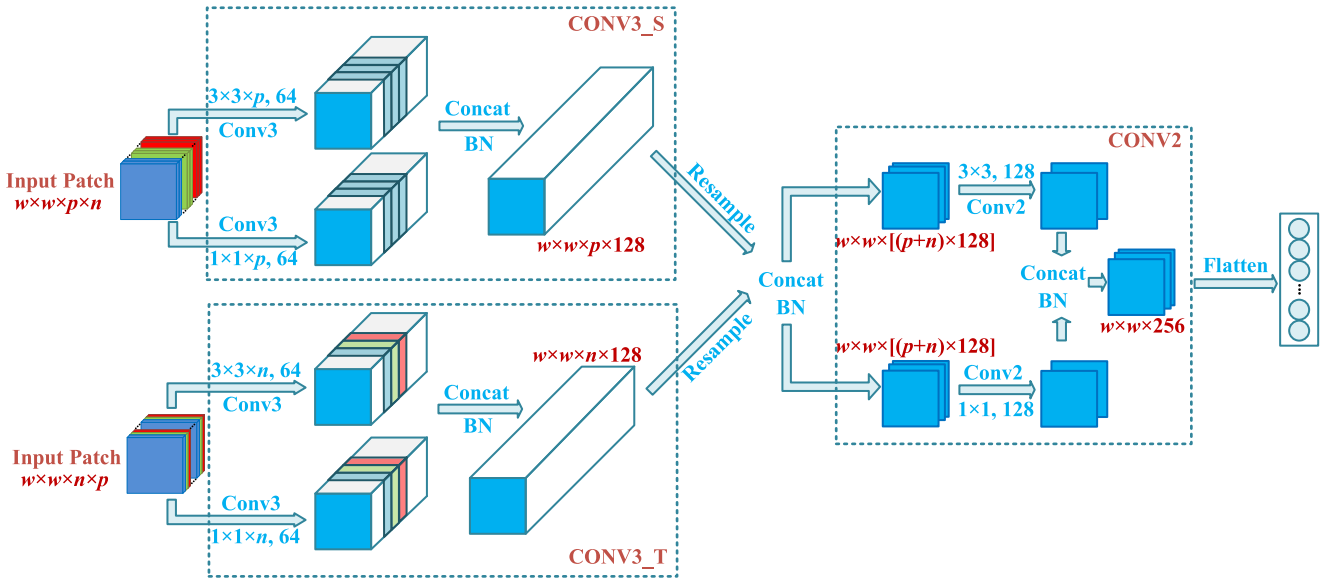

Fig. 3. Structure of CNN branch in the proposed MDFN.

$I_{mul} \in \mathbb{R}^{r\times c\times n\times p}$, where $r\times c$ denotes the size of these images, n represents the number of channels for a monotemporal image, and p is the number of the phases for the multitemporal images. It is worth noting that the input of CONV3_T is resampled to$I_{\mathrm{CONV3_T}} \in \mathbb{R}^{r\times c\times n\times p}$. With the 3-D convolution along the spectral-temporal direction, it is beneficial to learn the discriminative temporal-spatial information.

2) *CONV3_S*: similar to the CONV3_T, the CONV3_S module also includes two 3-D convolution layers: 64 $3\times3\times p$ and 64 $1\times1\times p$ convolution kernels (see Fig. 3). However, the input form of CONV3_S is resampled to $I_{\mathrm{CONV3\_S}} \in \mathbb{R}^{r\times c\times p\times n}$. The main difference between CONV3_T and CONV3_S is that the former directly stacks each monotemporal image according to the time sequence, and the latter stacks the multitemporal data based on the spectral sequence. Therefore, the defined CONV3_S module performs the 3-D convolution along the temporal-spectral direction. It is useful to model the stable spectral-spatial features on the time series, and then reduce the spectral variations or even abrupt temporal changes caused by the influence of clouds, shadows, and other external environmental conditions.

3) *CONV2*: As illustrated in Fig. 3, the CONV2 module is performed based on the concatenated of 3-D convolution features to further enhance the local spatial context information extraction at different scales (i.e., 3×3 and 1×1). This further improves the strong descriptive capability for the invariant spatio-temporal-spectral features that contribute to the final classification.

C. Multilevel Feature Fusion and Classification

In this step, the final spatio-temporal-spectral features generated by the LSTM and the CNN branches are fused together by the concatenation layer. The stacked multilevel features are then imported into to the fully connected layer to obtain the higher-level semantic information. Finally, the discriminative features are input into the Softmax classifier to realize an end-to-end automatic multitemporal HR images classification.

IV. Experimental Results and Analysis

A. Datasets Descriptions

In this article, two real HR multitemporal datasets are used to evaluate the performance of the proposed and reference methods.

1) *Ponte Arche (PA)*: The first dataset is built on three PlanetScope images, which were collected over the PA area, in the Autonomous Province of Trento in Italy. PlanetScope satellite constellation consists of more than 130 small satellites called Doves. These Dove satellites launched in groups greatly improved the revisit times. So PlanetScope can provide once daily images of every location on Earth. Three short-term multitemporal images characterized by four spectral bands (blue, green, red and near-infrared) were acquired over the PA area on April 2, 17, 22, and 2018, respectively. This dataset has a size of 304×361 pixels with a ground resolution of 3 meters. False color composite images for the three dates and the ground reference (GR) map are shown in Fig. 4. There exist five LULC classes in the analyzed area (i.e., buildings, roads, trees, farmland and soil). From three monotemporal images, it can be seen that there are various radiometrical changes due to illumination variations, topographical relief, and measurement noise conditions *(e.g.,* see the forest and farmland areas in Fig. 4).

2) *Shanghai (SH)*: The second dataset is made up of four images acquired over the city center of Houkou district, Shanghai, China, by the GF-1/6 satellite sensors. The

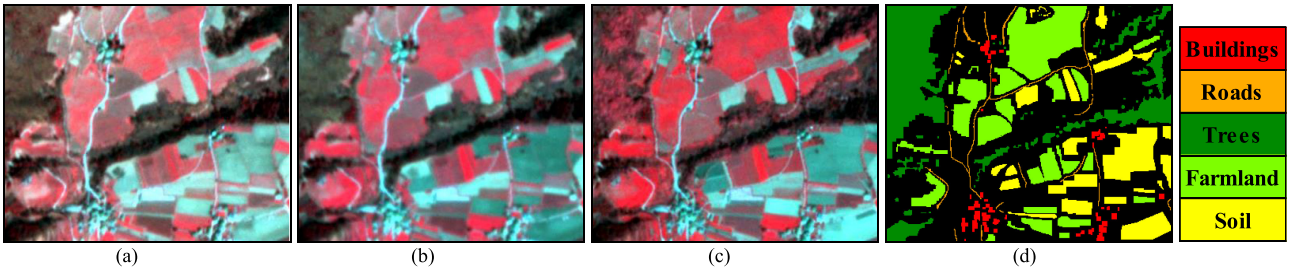

Fig. 4. False color composite images and the GR map of the PA dataset. (a) April 2, 2018 (T1). (b) April 17, 2018 (T2). (c) April 22, 2018 (T3). (d) GR map.

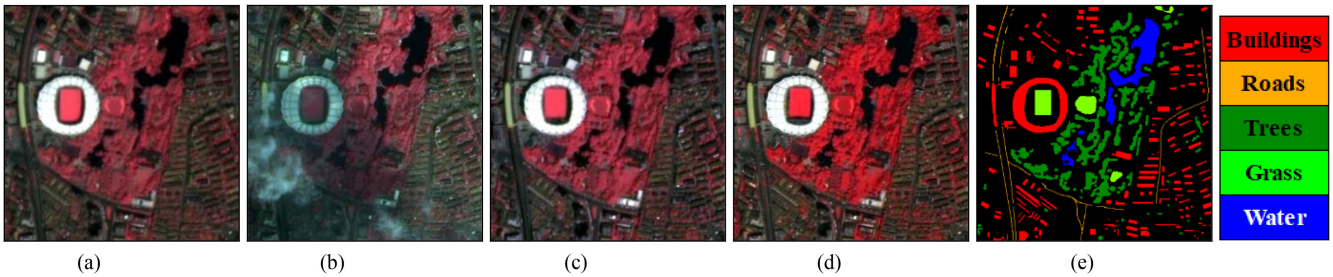

Fig. 5. False color composite images and the GR map of the SH dataset. (a) GF-1B: April 8, 2019 (T1). (b) GF-1: April 12, 2019 (T2). (c) GF-1D: April 15, 2019 (T3). (d) GF-6: April 18, 2019 (T4). (e) GR map.

acquisition times for the four images are April 8, 12, 15 and 18, 2019, respectively. It is worth noting that GF-1 and GF-6 are the high-resolution optical EO satellites of China National Space Administration (CNSA) with a four-day repetition cycle. Based on the acquired raw images, several image preprocessing (e.g., orthorectification, pansharpening, and image registration) were carried out. These images contain four spectral bands (i.e., blue, green, red and near-infrared) with a spatial resolution of 2 m after the pansharpening operation. The size of the final selected image subset for the experiment is 400×400 pixels. Fig. 5 presents the four dates false color composite images and the GR map. There are five land-cover classes (i.e., buildings, roads, trees, grass, and water) in different colors are shown in Fig. 5(e). In addition, the quality of monotemporal images changes at different times, especially it can see clearly that clouds contamination and shadows in T2 image [see Fig. 5(b)] and different illumination conditions in other images [see Fig. 5(a)–(d)].

Note that the GR maps in the two datasets were made according to a careful image interpretation. Details of the reference samples for each class in the two datasets are given in Table II.

TABLE II
DETAILED REFERENCE SAMPLE INFORMATION IN TWO DATASETS

Classes	Number of samples (pixels)	
	PA	SH
Buildings	1749	13400
Roads	1413	1358
Trees	16535	13954
Farmland	16671	-
Grass	-	2582
Soil	10775	-
Water	-	4318

B. Experimental Setup and Parameter Settings

To demonstrate the effectiveness of the proposed MDFN approach, five reference methods were implemented and compared in the experiments including two popular classifiers, i.e., support vector machine (SVM) and RF, and three state-of-the-art classification networks: FDSSC [25]; SSUN [26]; and SSRN [27]. The FDSSC, SSUN and SSRN are three excellent state-of-the-art DL-based frameworks. Specifically, the FDSSC framework uses different 3-D convolutional kernel sizes to extract spectral and spatial features separately, and the 3-D densely-connected structures was used for deep learning of features, leading to extremely accurate classification. The SSUN structure also includes the LSTM and CNN branches that is similar to the proposed MDFN. But it is designed for monotemporal HS image classification, where LSTM is used to learn the spectral group information and CNN for extracting the spatial context information. The SSRN is designed with 3-D residual blocks and 3-D convolutional layers, potentially more suitable to learn the spatio-temporal-spectral information.

Parameter setting and evaluation were carried out in the experiment, in order to analyze the classification performance of the proposed MDFN and the compared methods. For the SVM classifier, the radial basis function was selected as kernel function. For the RF classifier, the number of decision trees was set to 500. For different networks used in this article, detailed parameter settings are given in Table III. In general, a larger window size (w) will help to increase the classification accuracy.

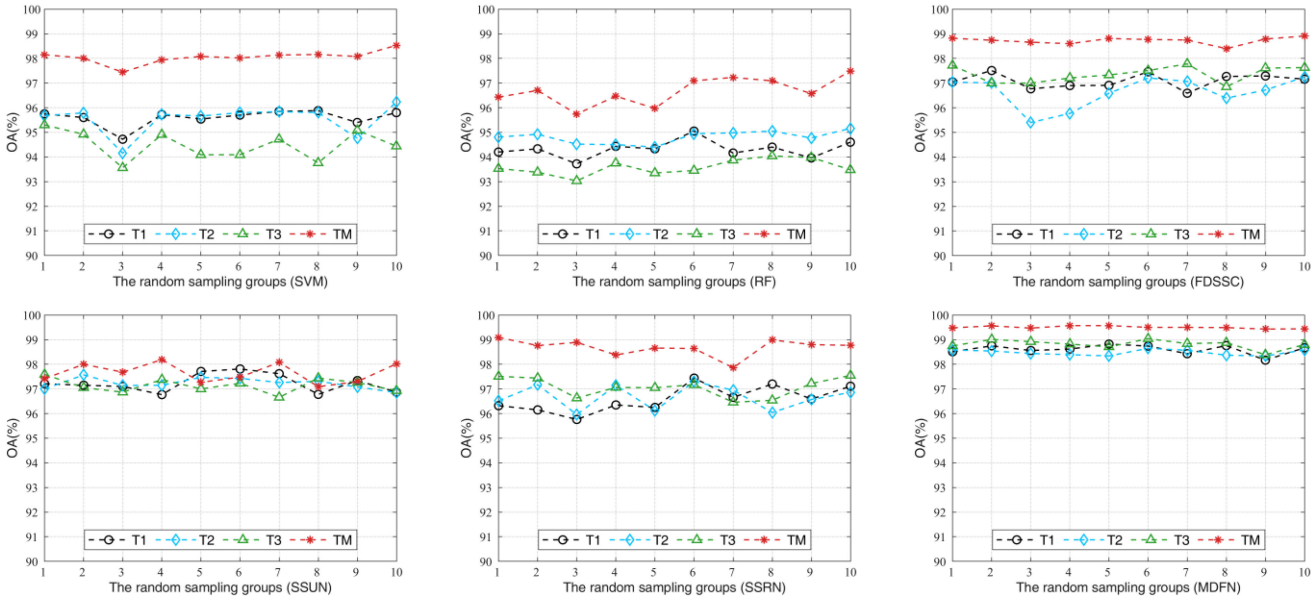

Fig. 6. Classification accuracies obtained by different methods with ten runs of random sampling under 1% training samples (PA dataset). T1–T3 represent three monotemporal image results, respectively, and TM represents the multitemporal image results.

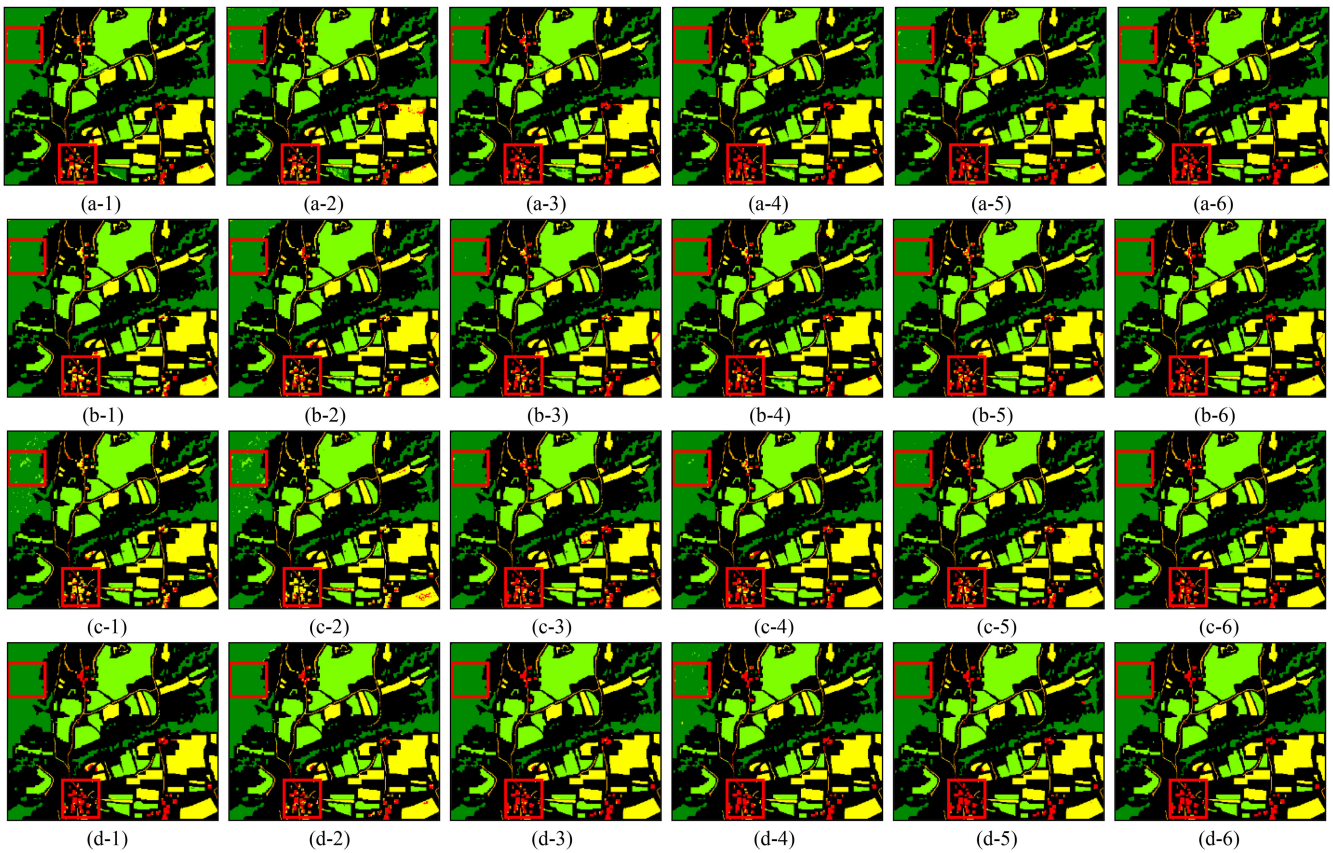

Fig. 7. Classification maps obtained by different methods on the PA dataset. (a)–(c) Three monotemporal image classification maps, respectively. (d) Multitemporal classification maps; and column 1–6 represent the classification maps obtained by SVM, RF, FDSSC, SSUN, SSRN, and MDFN methods, respectively.

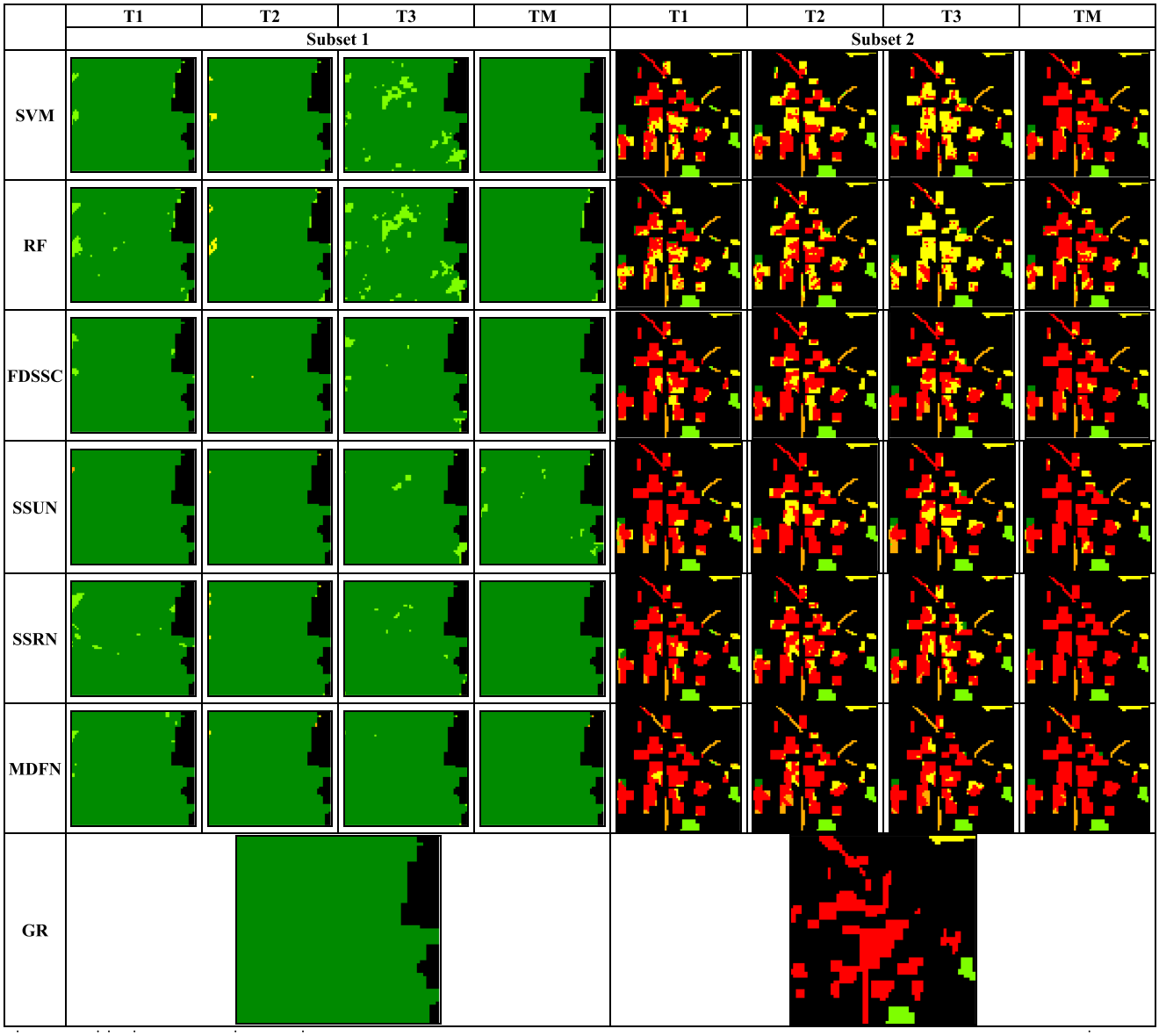

Fig. 8. Classification maps obtained by different methods at local subsets on the PA dataset. Columns 1–4 represent three monotemporal and the multitemporal classification results of subset 1; columns 5–8 are the classification results of subset 2 (corresponding to the highlighted subsets in Fig. 7); and rows 1–6 represent the local classification maps obtained by SVM, RF, FDSSC, SSUN, SSRN, and MDFN methods, respectively, and row 7 shows the GR maps of two subsets.

TABLE III
PARAMETERS SETTINGS IN DIFFERENT METHODS

Type	FDSSC	SSUN	SSRN	MDFN
Window size (w)	4	8	4	4
Batch size	32	128	16	128
Epoch	80	500	200	20

However, this also increases the possibility of the over-smoothing and time consumption. Therefore, considering the tradeoff between the accuracy and the computational cost, we set the size of the input block as four for the FDSSC, SSRN and MDFN methods, and eight for the SSUN due to the fact that its network includes three average pooling layers. For the batch size and epoch, they are set according to the literature work and multiple trials in our experiments, which are provided as given in Table III.

In addition, considering the Adam optimizer can solve sparse gradients on noisy problems, so the Adam with a learning rate of 0.0001 was selected as the optimization algorithm of the proposed MDFN framework. In addition, in order to make consistence with the multitemporal input in the proposed MDFN method, we stacked p-times of the monotemporal image as the input of the proposed MDFN.

The training samples of two datasets are set as 1% randomly, and the rest (99%) are used for test. Quantitative experiments are performed with ten runs in order to eliminate the errors caused by random samplings. Four indices are utilized to evaluate the

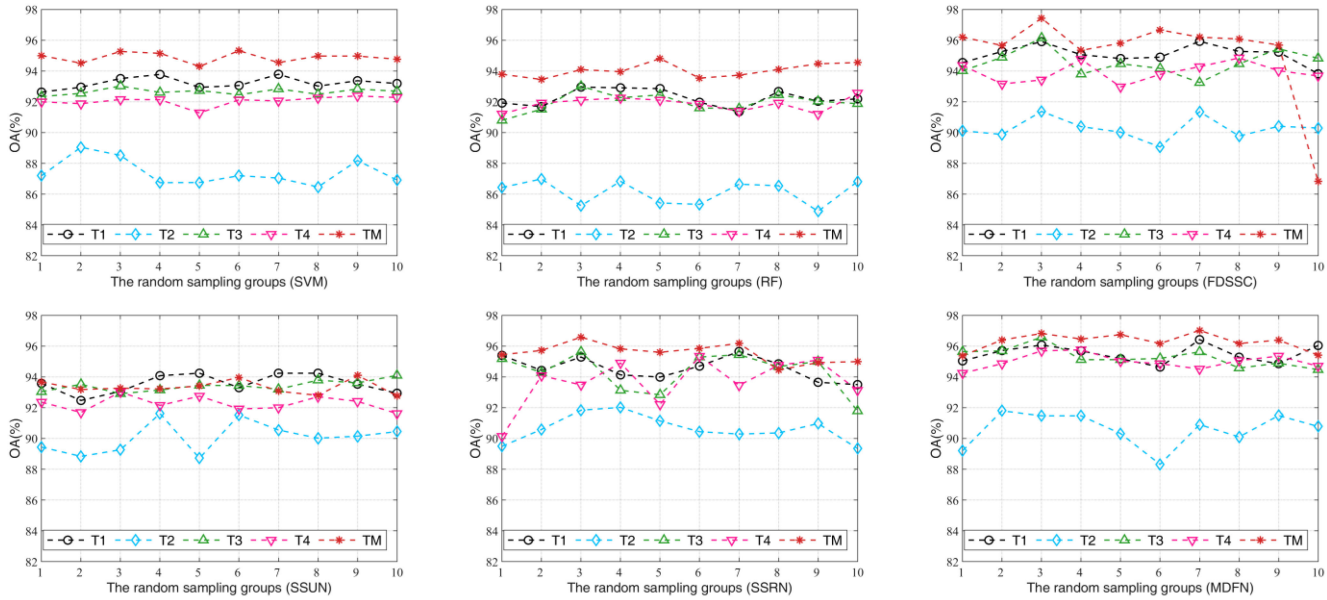

Fig. 9. Classification accuracies obtained by different methods with ten runs of random sampling under 1% training samples (SH dataset). T1–T3 represent three monotemporal image results, respectively, and TM represents the multitemporal image results.

performance of the considered classification methods, including the class accuracy (CA), the overall accuracy (OA), the kappa coefficient (κ), and the computing cost (T).

All experiments were carried out on a computer with Intel (R) Core (TM) i7-6700 CPU, 3.41 GHz, NVIDIA Quadro K620, RAM 32.0 GB.

C. Classification Performance

1) Results of the PA Dataset: Fig. 6 illustrates the quantitative comparison between the monotemporal and multitemporal images classification accuracies obtained by different methods after ten runs. From the fluctuations of monotemporal (T1–T3) and multitemporal (TM) curves, one can see that the OA of the multitemporal images classification is not only higher than the one of monotemporal image classification, but also relatively less affected by random sampling, exhibiting a more stable performance. By comparing multitemporal results obtained by different methods (see the red curves in Fig. 6), the proposed MDFN approach with the highest OA values and lowest fluctuating deviations exhibits the best performance. These results demonstrate that the proposed MDFN is robust under the random sampling condition, but also that it has a high descriptive capability of stable spatio-temporal-spectral features for short-term multitemporal HR images.

Table IV gives the quantitative assessment of the proposed MDFN and the reference methods based on the multitemporal images classification. The proposed MDFN approach produces the highest OA value with the lowest standard deviation value (i.e., 0.05). In particular, the MDFN (OA = 99.50%) achieves roughly 1% and 2% of OA increase than the three advanced DL-based methods FDSSC (98.72%), SSRN (98.69%) and SSUN (97.65%), respectively. Traditional machine learning methods SVM and RF classifiers present the low computing cost as they do not implement multitemporal feature fusion operations, but their accuracies are lower than the proposed MDFN.

TABLE IV
CLASSIFICATION ACCURACIES (%) PROVIDED BY DIFFERENT METHODS WITH 1% TRAINING SAMPLES (PA DATASET)

Classes	SVM	RF	FDSSC	SSUN	SSRN	MDFN
Buildings	75.33	59.88	87.50	78.86	89.02	94.42
Roads	76.38	69.61	83.05	78.36	82.07	94.06
Trees	99.95	99.71	99.96	99.43	99.93	99.97
Farmland	99.79	99.28	99.71	99.15	99.77	99.90
Soil	99.00	97.52	99.19	98.19	98.85	99.71
OA (%)	98.05 ±0.27	96.68 ±0.56	98.72 ±0.14	97.65 ±0.39	98.69 ±0.35	**99.50 ±0.05**
κ×100	97.20 ±0.39	95.21 ±0.80	98.17 ±0.20	96.62 ±0.57	98.11 ±0.50	99.28 ±0.07
T(s)	0.46 ±0.06	1.36 ±0.07	57.53 ±3.49	42.44 ±1.20	130.07 ±1.41	61.67 ±1.38

Fig. 7 visualizes the best classification maps obtained by different methods on each monotemporal image and multitemporal image. In order to better illustrate the classification performance at local scales, the subsets highlighted in red rectangles in Fig. 7 are further compared in Fig. 8. It is obvious that the qualitative results between different methods are in line with the quantitative results provided in Fig. 6 and Table IV. Among all results, the monotemporal image with only four broad spectral bands is quite prone to confuse between buildings and soil, and between trees and farmland (see in rows 1–3 in Fig. 7, and in columns T1–T3 in Fig. 8). After fusing three monotemporal images, the misclassification problems in buildings and trees are greatly improved (see in row 4 in Fig. 7, and in column T4 in Fig. 8). This demonstrates that the multitemporal information integration can significantly improve the separability of similar ground objects (e.g., trees and farmland). Furthermore, compared with the reference methods, the proposed MDFN results in the most accurate, regular and

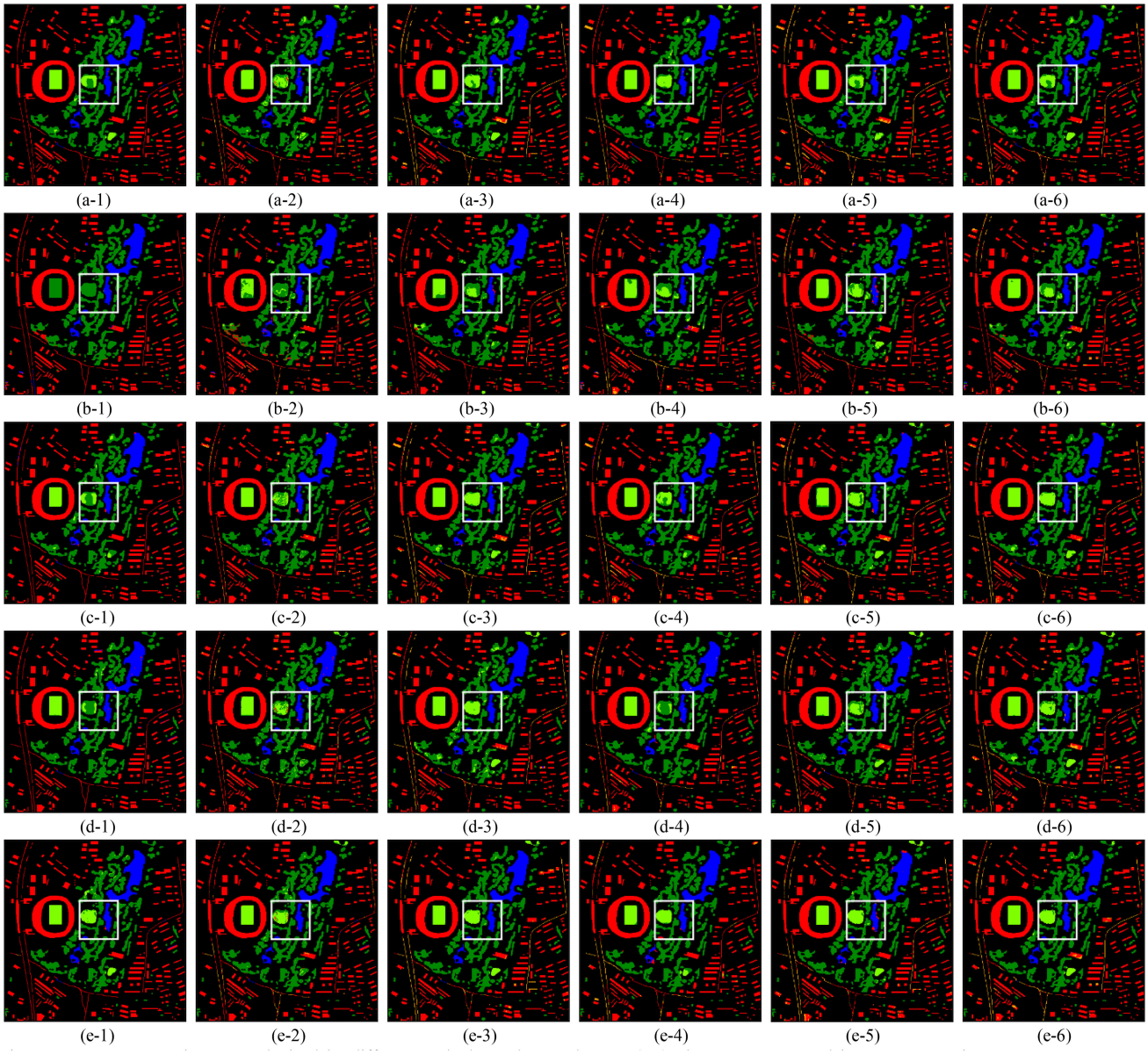

Fig. 10. Classification maps obtained by different methods on the SH dataset. (a)-(c). Three monotemporal image classification maps, respectively. (d) Multitemporal classification maps; and columns 1–6 represent the classification maps obtained by SVM, RF, FDSSC, SSUN, SSRN, and MDFN methods, respectively.

smooth classification map with OA = 99.57% (see the whole classification map in Fig. 7d-6 and the subset results in the row MDFN in Fig. 8). To sum up, it verifies that the proposed MDFN is more robust owing to the strong spatio-temporal-spectral complementary ability even under limited training samples. The good generalization capability and effective feature representation enables the MDFN to offer an outstanding performance for short-term multitempotal HR images classification.

2) Results on the SH Dataset: Fig. 9 illustrates the classification accuracies obtained on four monotemporal and multitemporal images by different methods. One can see that in all methods, the multitemporal classification is clearly superior to the ones using monotemporal images according to the higher OA values.

Monotemporal classification results showed different performances due to the variety in their data quality. The SVM, RF and MDFN methods have relatively smaller fluctuations affected by the random sampling compared with other DL-based methods. The proposed MDFN method presents the best multitemporal classification performance with the highest OA values and a better stability. In addition, one can observe that MDFN is also effective in monotemporal classification compared with the reference methods.

The quantitative assessment results of different methods for multitemporal images classification are given in Table V. Compared with the five reference methods, the proposed MDFN achieved the highest classification accuracy (i.e., OA = 96.30%) with a small standard deviation value (i.e., 0.55%). The SSRN

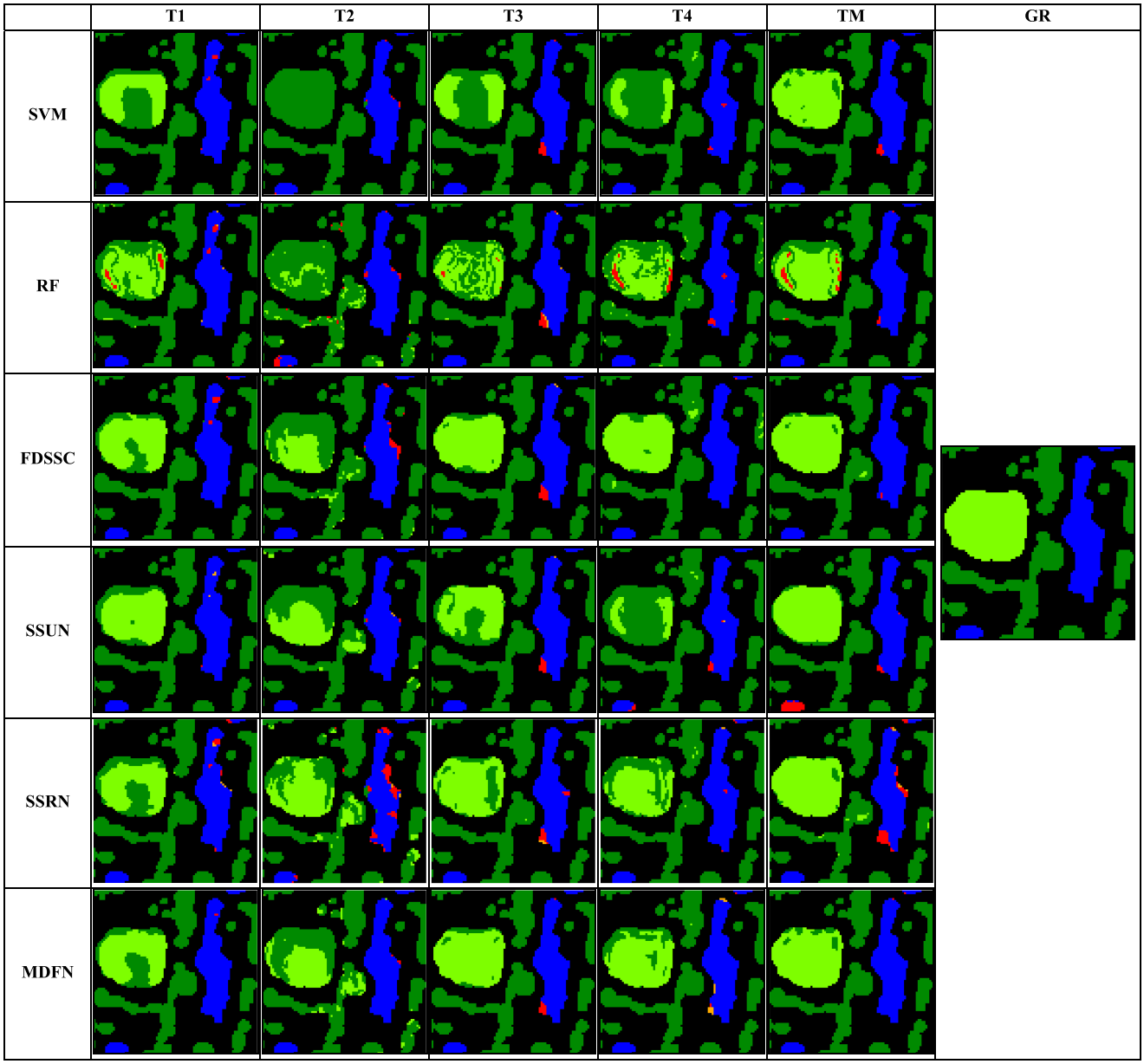

Fig. 11. Classification maps obtained by different methods at local subsets on the SH dataset. Columns 1-4 represent three monotemporal and the multitemporal classification results of the subset highlighted in Fig. 10; column 5 shows the GR maps of the local subset; and rows 1–6 represent the local classification maps obtained by SVM, RF, FDSSC, SSUN, SSRN, and MDFN methods, respectively.

method achieved the highest OA (95.55%) among five reference methods. However, its computing cost is about three times higher than MDFN. The performance of three DL-based methods is reduced since the limited number of training samples were considered in this case, so it is very difficult to meet the network training requirements. This also demonstrates that the proposed MDFN approach effectively integrates the multitemporal information, while shows a strong robustness even in a small-sample case.

Fig. 10 visualizes the best classification maps obtained by different methods on each monotemporal and multitemporal image. In order to better illustrate the classification performance at local scales, subsets highlighted in white rectangles in Fig. 10 are further compared in Fig. 11. In consistent with the above quantitative analysis results, from the qualitative analysis, it can be seen that the joint classification of the multitemporal HR images can effectively reduce the miss and false alarms in some typical ground objects. For example, the classification maps of column 5 in Fig. 11 (i.e., TM) present less misclassification between trees (in dark green) and grass (in bright green) than in the first columns (T1–T4). Especially, classification errors in T2 image affected by clouds and shadows (see Fig. 10 column 2) are eliminated, thus they are correctly classified in TM results (see Fig. 10 column 6). FDSSC resulted in good classification results as shown in Fig. 10e-3 and Fig. 11 row FDSSC, comparing with the other reference methods. However, it fluctuates greatly

TABLE V
CLASSIFICATION ACCURACIES (%) PROVIDED BY DIFFERENT METHODS WITH 1% TRAINING SAMPLES (SH DATASET)

Classes	SVM	RF	FDSSC	SSUN	SSRN	MDFN
Buildings	99.61	98.71	93.93	97.82	96.11	97.79
Trees	98.74	99.04	99.06	97.61	98.80	99.52
Water	99.42	99.70	99.70	99.31	98.61	99.91
Roads	6.64	21.08	59.02	24.99	59.01	49.85
Grass	88.24	71.83	92.32	73.03	89.25	89.51
OA (%)	94.88	94.05	95.19	93.34	95.55	**96.30**
	±0.34	±0.45	±2.99	± 0.45	±0.63	**±0.55**
κ×100	92.36	91.11	93.03	90.11	93.48	94.53
	±0.51	±0.67	±4.14	±0.66	±0.92	±0.81
T (s)	1.06	2.14	50.14	34.27	102.31	32.09
	±0.02	±0.07	±2.48	±0.85	±1.54	±1.23

and its average accuracy with the highest standard deviation value (2.99%) under ten runs of random sampling (see Fig. 9 and Table V). Therefore, the proposed MDFN with the higher robustness and excellent classification performance is more suitable for multitemporal HR images classification.

V. CONCLUSION

In this article, a novel multitemporal deep fusion and classification network MDFN has been developed based on short-term multitemporal HR images. The two branches of LSTM and CNN are designed in MDFN to learn the discriminative information from spectral, spatial and temporal dimensions. Experimental results obtained on two real multitemporal HR datasets validated the effectiveness of the proposed MDFN. Compared with traditional and state-of-the-art methods, MDFN exhibits three main advantages: the discriminative spatio-temporal-spectral feature extraction and fusion; high model stability under few-shot learning; and high descriptive capability for the temporal-invariant ground objects. With a robust end-to-end learning process, the proposed MDFN not only efficiently learns the multitemporal information, but also improves the spectral stability and the discrimination of typical ground objects. It can be also applicable to the short-term multiple UAV/airborne image classification.

ACKNOWLEDGMENT

The authors would like to thank Planet Labs for providing the Dove images, and also thank the China Centre for Resources Satellite Data and Application for providing the GF images.

REFERENCES

[1] S. Liu, D. Marinelli, L. Bruzzone, and F. Bovolo, "A review of change detection in multitemporal hyperspectral images: Current techniques, applications, and challenges," *IEEE Geosci. Remote Sens. Mag.,* vol. 7, no. 2, pp. 140–158, Jun. 2019.

[2] S. Liu, Q. Du, X. Tong, A. Samat, and L. Bruzzone, "Unsupervised change detection in multispectral remote sensing images via spectral-spatial band expansion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 12, no. 9, pp. 3578–3587, Sep. 2019.

[3] Y. Wang, H. Shi, Y. Zhuang, Q. Sang, and L. Chen, "Bidirectional grid fusion network for accurate land cover classification of high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 13, pp. 5508–5517, Sep. 2020, doi: 10.1109/JSTARS.2020.3023645.

[4] S. Liu, Y. Zheng, Q. Du, A. Samat, X. Tong, and M. Dalponte, "A novel feature fusion approach for VHR remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 14, pp. 464–473, Dec. 2021, doi: 10.1109/JSTARS.2020.3041868.

[5] S. Liu *et al.*, "A multi-scale superpixel-guided filter feature extraction and selection approach for classification of very-high-resolution remotely sensed imagery," *Remote Sens.,* vol. 12, no. 5, Mar. 2020, Art. no. 865.

[6] T. Yamada, E. C. Pedrino, M. D. C. Nicoletti, and L. E. Moschini, "A high resolution image based approach for estimating the canopy cover of a semi-deciduous Brazilian atlantic forest fragment," *IEEE Latin Amer. Trans.,* vol. 19, no. 10, pp. 1657–1664, Oct. 2021.

[7] D. Brunner, G. Lemoine, and L. Bruzzone, "Earthquake damage assessment of buildings using VHR optical and SAR imagery," *IEEE Trans. Geosci. Remote Sens.,* vol. 48, no. 5, pp. 2403–2420, May 2010.

[8] Y. Han, F. Bovolo, and L. Bruzzone, "Fine co-registration of VHR images for multitemporal urban area analysis," in *Proc. 8th Int. Workshop Anal. Multitemporal Remote Sens. Images*, 2015, pp. 1–4.

[9] D. Hong *et al.*, "Endmember-Guided unmixing network (EGU-Net): A general deep learning framework for self-supervised hyperspectral unmixing," *IEEE Trans. Neural Netw. Learn. Syst.*, May 2021, doi: 10.1109/TNNLS.2021.3082289.

[10] S. Saha, L. Mou, C. Qiu, X. X. Zhu, F. Bovolo, and L. Bruzzone, "Unsupervised deep joint segmentation of multitemporal high-resolution images," *IEEE Trans. Geosci. Remote Sens.,* vol. 58, no. 12, pp. 8780–8792, Dec. 2020.

[11] D. Ienco *et al.*, "Combining sentinel-1 and sentinel-2 satellite image time series for land cover mapping via a multi-source deep learning architecture," *ISPRS J. Photogramm. Remote Sens.,* vol. 158, pp. 11–22, Sep. 2019.

[12] O. Rajadell, P. García-Sevilla, and F. Pla, "Spectral–spatial pixel characterization using Gabor filters for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.,* vol. 10, no. 4, pp. 860–864, Jul. 2013.

[13] G. Moser and S. B. Serpico, "Edge-preserving classification of high-resolution remote-sensing images by Markovian data fusion," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2009, pp. IV–765–IV–768.

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

[15] M. D. Mura, J. A. Benediktsson, and L. Bruzzone, "Classification of hyperspectral images with extended attribute profiles and feature extraction techniques," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2010, pp. 76–79.

[16] S. Niazmardi, A. Safari, and S. Homayouni, "Similarity-Based multiple kernel learning algorithms for classification of remotely sensed images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 10, no. 5, pp. 2012–2021, May 2017.

[17] Y. Gu *et al.*, "Multiple kernel learning for hyperspectral image classification: A review," *IEEE Trans. Geosci. Remote Sens.,* vol. 55, no. 11, pp. 6547–6565, Nov. 2017.

[18] U. B. Gewali and S. T. Monteiro, "Spectral angle based unary energy functions for spatial-spectral hyperspectral classification using Markov random fields," in *Proc. 8th Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, 2016, pp. 1–6.

[19] A. Wang, P. Liu, and C. Xie, "Urban land use classification from high-resolution SAR images based on multi-scale Markov random field," in *Proc. 24th Int. Conf. Geoinformatics*, 2016, pp. 1–4.

[20] A. K. Neves, T. S. Körting, C. D. G. Neto, A. R. Soares, and L. M. G. Fonseca, "Hierarchical classification of Brazilian Savanna physiognomies using very high spatial resolution image, superpixel and Geobia," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 3716–3719.

[21] Z. Zhang, D. Liu, D. Gao, and G. Shi, "S^3Net: Spectral-spatial-semantic network for hyperspectral image classification with the multiway attention mechanism," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–17, Apr. 2021, doi: 10.1109/TGRS.2021.3067356.

[22] D. Hong *et al.*, "Multimodal remote sensing benchmark datasets for land cover classification with a shared and specific feature learning model," *ISPRS J. Photogramm. Remote Sens.,* vol. 178, pp. 68–80, Aug. 2021.

[23] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.,* vol. 59, no. 5, pp. 4340–4354, May 2021.

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

[25] W. Wang *et al.*, "A fast dense spectral–spatial convolution network framework for hyperspectral images classification," *Remote Sens.,* vol. 10, no. 7, pp. 1068, Jul. 2018.

[26] Y. Xu, L. Zhang, B. Du, and F. Zhang, "Spectral–Spatial unified networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.,* vol. 56, no. 10, pp. 5893–5909, Oct. 2018.

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

[28] Z. Zheng and J. Cao, "Fusion high-and-low-Level features via ridgelet and convolutional neural networks for very high-resolution remote sensing imagery classification," *IEEE Access,* vol. 7, pp. 118472–118483, Sep. 2019.

[29] H. Jin, P. Li, and W. Fan, "Land cover classification using multitemporal CHRIS/PROBA images and multitemporal texture," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2008, pp. 742–745.

[30] A. Khaliq, L. Peroni, and M. Chiaberge, "Land cover and crop classification using multitemporal Sentinel-2 images based on crops phenological cycle," in *Proc. IEEE Workshop Environ., Energy, Struct. Monit. Syst.*, 2018, pp. 1–5.

[31] H. L. Yang and M. M. Crawford, "Manifold alignment for classification of multitemporal hyperspectral data," in *Proc. 3rd Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, 2011, pp. 1–4.

[32] D. Tuia, D. Marcos, and G. Camps-Valls, "Multi-temporal and multi-source remote sensing image classification by nonlinear relative normalization," *ISPRS J. Photogramm. Remote Sens.,* vol. 120, pp. 1–12, Jul. 2016.

[33] A. Salberg, "Land cover classification of cloud-contaminated multitemporal high-resolution images," *IEEE Trans. Geosci. Remote Sens.,* vol. 49, no. 1, pp. 377–387, Jan. 2011.

[34] Z. Li, G. Chen, and T. Zhang, "Temporal attention networks for multitemporal multisensor crop classification," *IEEE Access,* vol. 7, pp. 134677–134690, Sep. 2019.

[35] J. Feng *et al.*, "Attention multibranch convolutional neural network for hyperspectral image classification based on adaptive region search," *IEEE Trans. Geosci. Remote Sens.,* vol. 59, no. 6, pp. 5054–5070, Jun. 2021.

[36] W. Hu *et al.*, "Spatial–Spectral feature extraction via deep ConvLSTM neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.,* vol. 58, no. 6, pp. 4237–4250, Jun. 2020.

[37] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*: Cambridge, MA, USA: MIT Press, 2016.

[38] X. Hao, G. Zhang, and S. Ma, "Deep learning," *Int. J. Semantic Comput.,* vol. 10, no. 3, pp. 417–439, 2016.

[39] C. Yu, R. Han, M. Song, C. Liu, and C.-I. Chang, "A simplified 2D-3D CNN architecture for hyperspectral image classification based on spatial–spectral fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 13, pp. 2485–2501, Apr. 2020, doi: 10.1109/JSTARS.2020.2983224.

[40] Y. Chen *et al.*, "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.

[41] B. Praveen and V. Menon, "Study of spatial–spectral feature extraction frameworks with 3-D convolutional neural network for robust hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.,* vol. 14, pp. 1717–1727, Dec. 2021, doi: 10.1109/JSTARS.2020.3046414.

Yongjie Zheng (Student Member, IEEE) received the B.S. degree in remote sensing science and technology from Henan Polytechnic University, Jiaozuo, China, in 2018, and the M.S. degree in photogrammetry and remote sensing from Tongji University, Shanghai, China, in 2021.

Her current research interests include deep learning, feature extraction and fusion, and multispectral/hyperspectral image classification/change detection.

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 interests include multitemporal data analysis, change detection, multispectral/hyperspectral remote sensing and planetary remote sensing.

Dr. Liu was the recipient of (ranked as third place) of 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 Multitemporal Remote Sensing Images (MultiTemp 2019). He is the Program Committee Member for SPIE Remote Sensing Symposium: Image and Signal Processing for Remote Sensing XXVII (2020-2021), and also served as the Session Chair for many international conferences, such as International Geoscience and Remote Sensing Symposium (2017–2019). He is/was a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING and *Remote Sensing*.

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland, Baltimore County, Baltimore, MD, USA, in 2000.

She is currently a Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. Her research interests include hyperspectral remote sensing image analysis and applications, and machine learning.

Dr. Du was the recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society. She was a Co-Chair for the Data Fusion Technical Committee of the IEEE GRSS from 2009 to 2013, the Chair for the Remote Sensing and Mapping Technical Committee of International Association for Pattern Recognition from 2010 to 2014, and the General Chair for 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 for 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. She is currently a Member of the IEEE Periodicals Review and Advisory Committee and SPIE Publications Committee. She is a Fellow of SPIE-International Society for Optics and Photonics.

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 Jiangxi University of Science and Technology, Ganzhou, China, in 2017. She is currently working toward the Ph.D. degree in surveying and mapping with Tongji University, Shanghai, China.

Her current research interests include multi-source remote sensing data fusion in the field of earth observation and planetary exploration.

Xiaohua Tong (Senior Member, IEEE) received the Ph.D. degree from Tongji University, Shanghai, China, in 1999.

He was a Postdoctoral Researcher with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan, China, between 2001 and 2003. He was a Research Fellow with The Hong Kong Polytechnic University in 2006, and a Visiting Scholar with the University of California, Santa Barbara, CA, USA, between 2008 and 2009. His current research interests include remote sensing, GIS, uncertainty and spatial data quality, image processing for high resolution, and hyperspectral images.

Dr. Tong is currently the Vice-Chair of the Commission on Spatial Data Quality of the International Cartographical Association, and the Co-Chair of the ISPRS working group (WG II/4) on Spatial Statistics and Uncertainty Modeling.

Michele Dalponte (Senior Member, IEEE) received the M.Sc. degree in telecommunications engineering and the Ph.D. degree in information and communication technologies from the University of Trento, Trento, Italy, in 2006 and 2010, respectively.

He was a Postdoctoral Researcher with the Norwegian University of Life Sciences, Oslo, Norway, and with the University of Cambridge, U.K. He is currently a Researcher with the Forest Ecology and Biogeochemical Cycles Group, Research and Innovation Center, Edmund Mach Foundation, San Michele all'Adige, Italy. His work has been published in international journals and presented at international conferences. He is a Reviewer for many remote sensing journals. His research interests include the field of remote sensing, in particular the analysis of hyperspectral, multispectral, and LIDAR data for forest monitoring.