Deep Residual Network-Based Fusion Framework for Hyperspectral and LiDAR Data

Chiru Ge , *Member, IEEE*, Qian Du , *Fellow, IEEE*, Weiwei Sun , *Member, IEEE*,
Keyan Wang , *Member, IEEE*, Jiaojiao Li , *Member, IEEE*, and Yunsong Li

Abstract—This article presents a deep residual network-based fusion framework for hyperspectral and LiDAR data. In this framework, three new fusion methods are proposed, which are the residual network-based deep feature fusion (RNDFF), the residual network-based probability reconstruction fusion (RNPRF) and the residual network-based probability multiplication fusion (RNPMF). The three methods use extinction profile (EP), local binary pattern (LBP), and deep residual network. Specifically, EP and LBP features are extracted from two sources and stacked as spatial features. For RNDFF, the deep features of each source are extracted by a deep residual network, and then the deep features are stacked to create the fusion features which are classified by softmax classifier. For RNPRF, the deep features of each source are input to the softmax classifier to obtain the probability matrices, and then the probability matrices are fused by weighted addition to producing the final label assignment. For RNPMF, the probability matrices are fused by array multiplication. Experimental results demonstrate that the classification performance of the proposed methods significantly outperform existing methods in hyperspectral and LiDAR data fusion.

Index Terms—Deep residual network, extinction profile, Goddard 's LiDAR, hyperspectral, hyperspectral and thermal (G-LiHT) data, image fusion, local binary pattern (LBP), probability fusion, light detection and ranging (LiDAR).

Chiru Ge is with the School of Information Science and Engineering, Shandong Normal University, Jinan 47856, China (e-mail: gechirumsu@gmail.com).

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

Weiwei Sun is with the Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo 315211, China (e-mail: nb-sww@outlook.com).

Keyan Wang, Jiaojiao Li, and Yunsong Li are with the State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, Xi'an, Shaanxi 710071, China (e-mail: kywang@mail.xidian.edu.cn; jjli@xidian.edu.cn; ysli@mail.xidian.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3054392

I. INTRODUCTION

URBAN mapping is a challenging task in the field of remote sensing. The accuracy and efficiency of numerous classification methods for obtaining land-use information in urban are insufficient to meet the requirements of real-world applications such as land management and urban planning [1]. With the development of sensor technologies, detailed information about various objects on the urban area can be acquired from multisource images, e.g., hyperspectral images (HSI) with spectral information, LiDAR data describing the 3-D information of object surface. The fusion of HSI and LiDAR has great potential in improving the accuracy of urban mapping. However, automatic interpretation of these data remains difficult [2].

Hyperspectral imagers gather data from hundreds of narrow spectral bands in infrared and visible portions of the electromagnetic spectrum. HSI can present the spectral information of the observed scenes. Hyperspectral system relies on the reflection of sunlight to collect data and belongs to the passive sensing. Many algorithms have been designed for hyperspectral applications such as classification, feature extraction, and segmentation, etc., [3]–[12]. A LiDAR system uses the pulsed laser to measure distances, which belongs to the active remote sensing. For a LiDAR system, the coherent light pulses are transmitted, reflected by objects on the ground, and caught by a receiver. LiDAR data products mainly contain LiDAR point cloud, digital surface model (DSM), digital terrain model (DTM), and canopy height model (CHM). LiDAR images provide shape and height information of the scenes. LiDAR also has high precision and flexibility because it is less sensitive to weather conditions, can be operated at any time, and has adjustable system parameters such as scan angle, flight speed/height, scan rate, and pulse rate. Similarly, numerous algorithms have been proposed for LiDAR applications such as feature extraction and object detection, etc. [13]–[17]. Although algorithms are separately proposed for LiDAR and HSI, no one type of sensor can always meet the requirements of reliable image interpretation. For instance, hyperspectral images cannot distinguish targets comprised of the same material, such as roads and roofs with the same pitch material. On the other hand, LiDAR data alone cannot be used to discriminate targets with the same elevation, such as houses with the same height but had concrete or solar panel roofs. Therefore, it is necessary to jointly use the information in HSI and LiDAR images to provide more intelligent processing.

HSI and LiDAR have been used in combination in a large number of applications favorably, such as microclimate modeling [18], biomass estimation [19], fuel type mapping [20], and mapping plant richness [21]. Besides, for HSI and LiDAR image fusion classification [2], [22]–[29], the combined use of HSI and LiDAR image can lead to higher classification accuracies and better discrimination ability in urban and forests areas than using each source separately. HSI and LiDAR image fusion methods can be divided into feature-level fusion and decision-level fusion. Feature-level fusion combines hyperspectral and LiDAR features, and fusion features are classified to produce category labels by classifiers. In feature level fusion, the traditional fusion methods can be divided into direct stacking, extracting spatial features and stacking, extracting spatial features and stacking, and dimensionality reduction. The information extraction of the direct stacking method is not sufficient, resulting in low classification accuracy [29]. The extracting spatial features and stacking fusion strategy overcomes the shortcomings of the direct stacking fusion strategy and can extract sufficient information. However, the stacked features tend to have high dimensions, and with small samples in various real applications, the curse of dimensionality problem is prone to occur, resulting in the risk of overfitting the training data [23]. Thus, many work in the literature include dimensionality reduction [23], [24], [27]–[29]. However the best dimension for dimensionality reduction is unknown. For the decision-level fusion, features of hyperspectral and LiDAR are classified individually, and classification results are fused to produce category labels. In the decision level fusion, the traditional fusion method is max voting fusion strategy [25], but voting can result in rough results. Recently, the residual fusion strategy of the collaborative representation-based classifier is proposed [30], with two sensitive parameters adjusted during the fusion process.

The deep residual network is an improved network of traditional convolutional neural networks (CNNs) [31], [32], and it can alleviate the problem that classification accuracy of the CNNs models decreases when the network becomes deeper [31], [32]. Therefore, the deep residual network-based fusion framework for hyperspectral and LiDAR data[1] is proposed. Using the framework, we present one feature-level fusion method called residual network-based deep feature fusion (RNDFF) and two decision-level fusion methods named residual network-based probability reconstruction fusion (RNPRF) and residual network-based probability multiplication fusion (RNPMF). For the feature-level fusion, the deep features fusion strategy is utilized to extract high-level and low-dimensional features of hyperspectral and LiDAR images. The fusion features can be obtained by stacking of deep features [33], [34] or composite kernels [35]. In RNDFF, the deep features of each source are extracted by a deep residual network and are stacked to generate fusion features which are classified by softmax classifier [36]. RNDFF fuses the features in the hidden layer of deep residual networks, which is different from feature stack fusion strategy in [29]. RNDFF can overcome the dimension selection problem and offer better classification performance. For the probability

fusion framework, the deep features of each source are input to the softmax classifier to obtain the probability matrices. In RNPRF, the probability matrices are fused by weighted addition to produce the final label assignment. In RNPMF, the probability matrices are fused by array multiplication. RNPRF and RNPMF can alleviate the shortcomings in traditional feature and decision-level fusion methods and have no parameter to be adjusted during the fusion process. Note that the RNPRF and the RNPMF belong to probability fusion which is newly proposed for HSI and LiDAR fusion classification.

In the proposed fusion methods, the spatial features of hyperspectral and LiDAR images are first extracted. For HSI, it has been proven that spatial context information is helpful to improve classification accuracy of HSI [37]. For LiDAR images, shape and texture information are extracted by spatial feature extraction algorithms. Here, extinction profile (EP) and local binary pattern (LBP) are extracted as spatial features due to their effectiveness [30]. Recently, the spectral-spatial residual network (SSRN) is proposed for HSI classification [38]. The network structure of the SSRN is utilized because it can extract excellent high-level features from 3D data and be more robust and deeper than traditional 3D CNN used in [39]. In order to make our proposed fusion frameworks for actual data (G-LiHT data), calibration method [40] is used because it includes five kinds of data, i.e., hyperspectral images (1 m), Google Earth RGB (0.25 m), LiDAR CHM (1 m), LiDAR DTM (1 m), and LiDAR slope (1 m).

The main contributions of this article are listed as follows.

1) This article proposes three novel methods for the fusion of HSI and LiDAR image based on the EP, LBP, residual network, and probability fusion. The proposed fusion algorithms, i.e., RNPRF, RNDFF, and RNPMF, have no additional parameters to choose, easy implementation, and excellent classification performance for the LiDAR and HSI fusion.
2) A deep residual network-based fusion framework for HSI and LiDAR data is proposed.
3) The proposed fusion frameworks can be extended to any spatial features and any deep learning network structures for multisensor data fusion.

The rest of the article is organized as follows. Section II presents the methodology. Section III introduces experiment data, the experiment setup, and experimental results. Section IV concludes this article.

II. METHODOLOGY

Fig. 1 illustrates the proposed residual network-based fusion framework for classification of HSI and LiDAR images.

Fig. 2 shows the proposed probability reconstruction strategy for RNPRF and RNPMF. All data are separated into three datasets, which is a training sample set, validation sample set, and testing sample set. The framework has two parts: The training part and the testing part.

In the training part, spatial features are first extracted from hyperspectral images and LiDAR images. EP features and LBP features are stacked as the spatial features, which are abbreviated
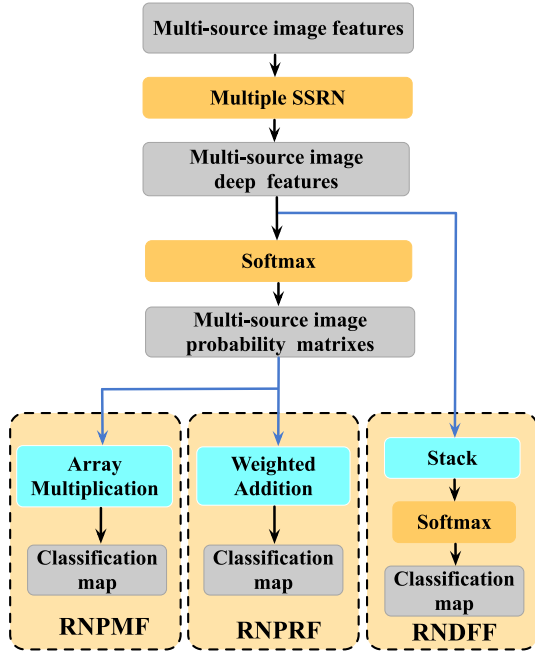
Fig. 1. Proposed residual network-based fusion framework.

as EPLBP. Then we have three input sources: HSI, HSI_EPLBP, and LiDAR_EPLBP. For each source, the training samples and their ground-truth labels update the parameters of the SSRN model through multiple iterations. In each iteration, an interim model is produced, and the validation samples and their corresponding labels are used to supervise the training process by calculating the classification accuracy of interim models on the validation set. Finally, the interim model with the highest classification accuracy for the validation set is selected as the final model. At the end of the training part, three final models are obtained.

In the testing part, the testing samples of HSI, HSI_EPLBP, LiDAR_EPLBP are input into three final models. The probability matrix of each source is obtained. Then three probability matrices are reconstructed by weighted addition. The reconstruction parameters are calculated by validation samples and validation labels. To produce the reconstruction parameters, the validation samples of HSI, HSI_EPLBP, LiDAR_EPLBP are input into three final models. Three probability matrices of the validation set are obtained and are fitted to the ground-truth label of the validation set by weighted addition. The weighted parameters are used for the reconstruction of three testing probability matrices. Finally, the reconstructed probability matrix of the testing set is utilized to produce the final label assignment.

Fig. 3 shows the proposed deep feature fusion strategy of RNDFF. HSI samples, HSI_EPLBP samples, and LiDAR_EPLBP samples are the input of the network. For each source, deep features are extracted by SSRN. Then deep features of three sources are stacked as the fusion feature. Finally, the fusion feature is the input of dense softmax to produce the probability matrix. In the deep feature fusion strategy, features are fused in the hidden layer of the deep learning network.

A. Spatial Features

The spatial features of HSI contain texture and shape information, and the spatial features of LiDAR data contain elevation, texture, and shape information, which benefit to the classification task. Since DTM and CHM are images in the form of 2D matrices, deep feature extraction cannot be performed using a 3D residual network on LiDAR images. The spatial features of HSI are also extracted as one data source for the deep residual network because the spatial contexture information benefit to improve the classification accuracy of HSI [37]. EP and LBP are selected as spatial feature extraction algorithms.

1) Extinction Profile: The extinction profile (EP) [41], [42] is made up of a sequence of thickening and thinning transforms acquired by a set of extinction filters (EFs) which preserve the relative image extrema. The EP for the input grayscale image H is described as [41]

$$\mathrm{EP(H)} = \{\underbrace{\phi^{P_{\lambda_1}}(\mathrm{H}), \phi^{P_{\lambda_2}}(\mathrm{H}), \cdots \phi^{P_{\lambda_S}}(\mathrm{H})}_{\text{thickening profile}}, \mathrm{H}$$

$$\underbrace{\gamma^{P_{\lambda_S}}(\mathrm{H}), \gamma^{P_{\lambda_{S-1}}}(\mathrm{H}), \cdots \gamma^{P_{\lambda_1}}(\mathrm{H})}_{\text{thinning profile}}\} \quad (1)$$

where the terms γ and ϕ are thinning and thickening transforms, and $P_{\lambda_S} : \{P_{\lambda_i}\}(i = 1, \ldots, S)$, a set of S ordered predicates (i.e., $P_{\lambda_i} \subseteq P_{\lambda_k}, i \leq k$) which determine the number of extrema in EP.

To obtain the EP features from HSI, independent component analysis (ICA) is used to extract relatively independent and the most informative components at first. Then, three independent components (ICs) are used as base images to create the EP [41]: $\mathrm{EP(H)} = \{\mathrm{EP(IC_1), EP(IC_2), EP(IC_3)}\}$. $\mathrm{EP(IC)}$ contains multiple attributes (e.g., volume, height, diagonal of bounding box, area, and standard deviation).

In order to obtain the EP features from the DTM and the CHM, EP with multiple attributes are directly extracted from the CHM and the DTM.

2) Local Binary Pattern: LBP [43] is one of the grayscale and rotation-invariant texture operators. LBP features which are unlike the EP features focus on the texture information. LBP features can virtually be utilized for the classification of HSI [44]. Therefore, LBP features are used together with EP features as spatial features.

For a center pixel y_c, a binary value is assigned to each neighbor of y_c in a local window. If the coordinate of y_c is $(0, 0)$, the coordinates of m neighbors $\{y_i\}_{i=0}^{m-1}$ are $(-r \sin(2\pi i/m), r \cos(2\pi i/m))$.

The LBP is an m-bit binary code obtained by thresholding the neighbors $\{y_i\}_{i=0}^{m-1}$ as

$$\mathrm{LBP}_{m,r}(y_c) = \sum_{i=0}^{m-1} U(y_i - y_c) \times 2^i \quad (2)$$

where $y_i - y_c \geq 0$, $U(y_i - y_c) = 1$ and $y_i - y_c \leq 0$, $U(y_i - y_c) = 0$.

To extract the LBP features for HSI, principal component analysis (PCA) [45] is first utilized to hyperspectral images. Next 2 is applied to the principal components of HSI. Since the
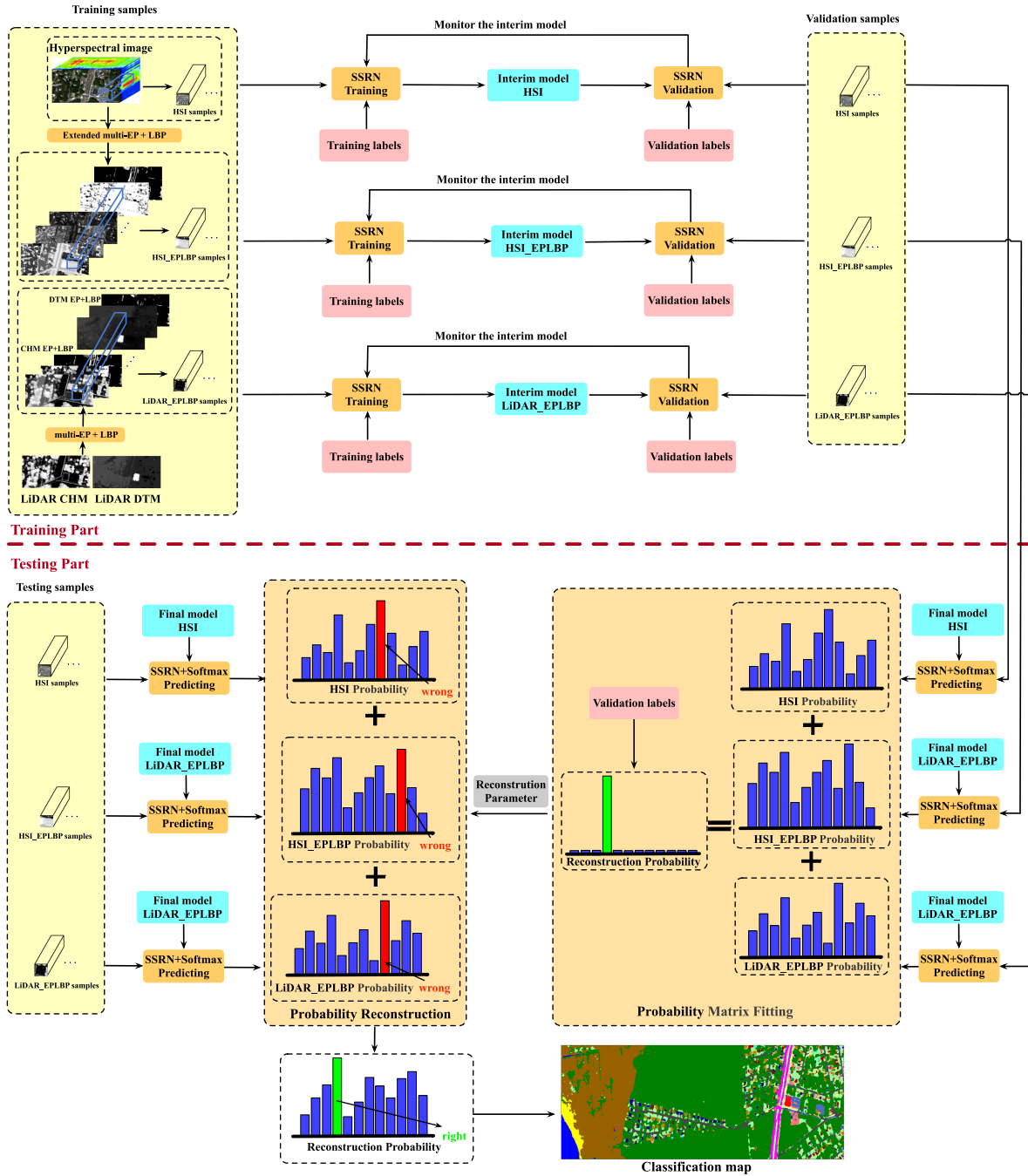
Fig. 2. Framework of the proposed probability reconstruction fusion strategy applies to G-LiHT real data. The framework shows a testing sample of the fourth class, which was misclassified by three classifiers utilizing each single source, but correctly classified by the fusion method.

CHM and the DTM have only one band, 2 is applied to the CHM and DTM to calculate the LBP code. After calculating the LBP code, an occurrence histogram is acquired as a nonparametric statistical estimate on the local patch. Then, a binning procedure is required to guarantee that the histogram features have the same dimension. Finally, LBP features are stacked with EP features.

Fig. 2 shows that HSI_EPLBP is the stacked features of the EP and the LBP features for HSI, while LiDAR_EPLBP is also the stacked features which can be described as [CHM_EP;CHM_LBP;DTM_EP;DTM_LBP].

B. Residual Network

A part of Fig. 3 in the yellow block shows the network structure of the SSRN. SSRN continuously extracts the z-axis features and the xy-axis features for the 3D input data. For HSI, the z-axis features represent spectral features and the xy-axis features represent spatial features. For EPLBP features, the z-axis feature represents the spatial features of different attributes under the same block, and the xy-axis feature represents the spatial features of the data block under the same attribute. SSRN is applied to extract useful features of the 3D data for
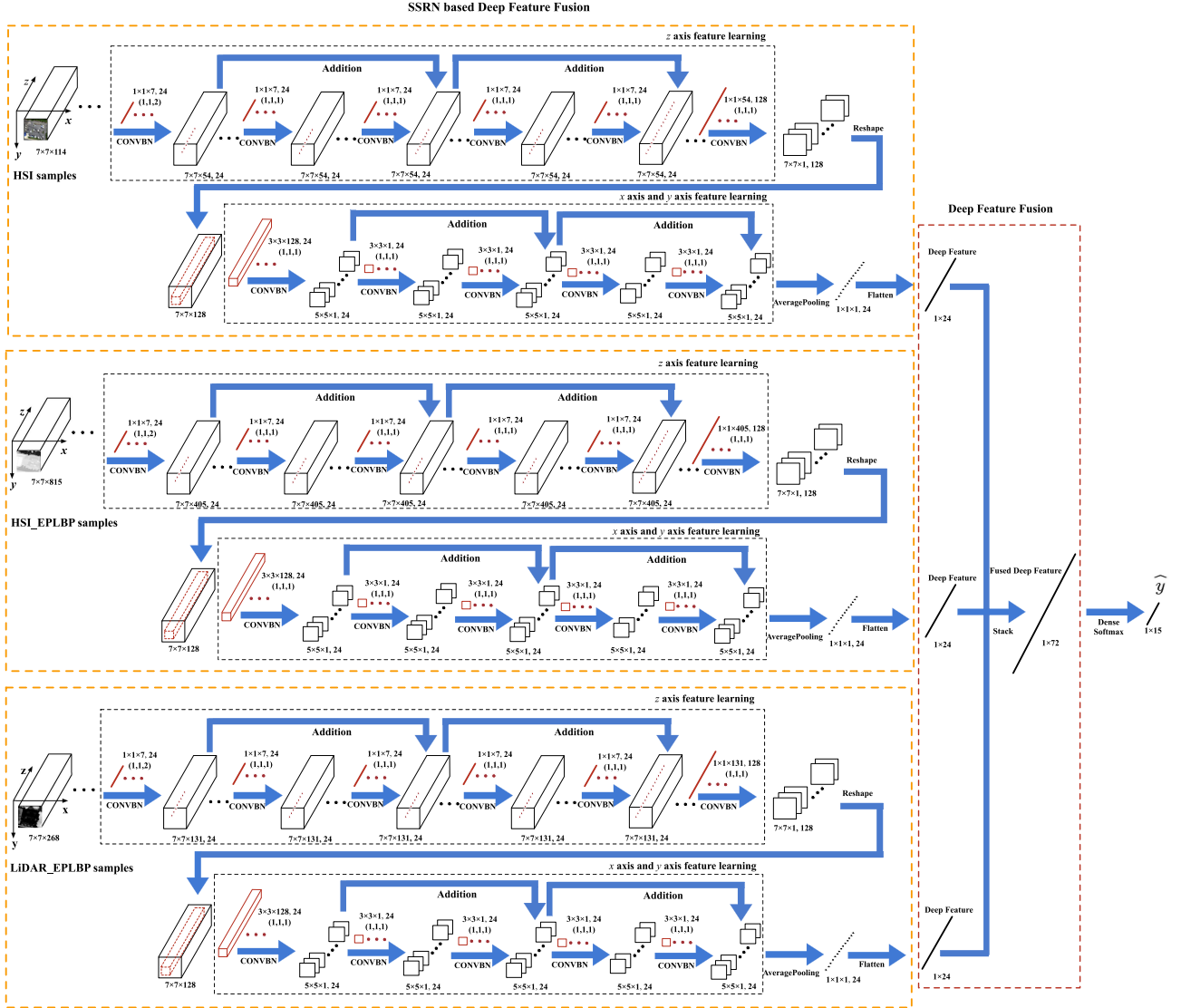
Fig. 3.   Framework of the SSRN-based deep feature fusion.

classification automatically [38]. We take the 3D samples of the G-LiHT HSI with a size of $7 \times 7 \times 114$ as an example to illustrate the SSRN which is shown in the top yellow block of the Fig. 3.

For the z-axis feature learning, 24 $1 \times 1 \times 7$ kernels with a subsampling stride of $(1, 1, 2)$ convolve the input tensor of HSI to produce 24 $7 \times 7 \times 54$ feature cubes in the first convolutional layer. Because the raw input data with high dimension has redundant information, the purpose of this layer is to remove some of the redundant information. Then, two consecutive residual blocks learn the deep z-axis feature. Each residual block contains two convolutional layers and one identity mapping with 24 $1 \times 1 \times 7$ vector kernels. Padding is utilized in the residual block to ensure that the output cubes are the same size as the input data. In order to maintain the discriminative spectral features, 128 $1 \times 1 \times 54$ vector kernels are used, and the resulting $7 \times 7 \times 128$ cube is utilized as the input to extract the xy-axis features.

For the xy-axis feature learning, to extract low-level xy-axis features, 24 $3 \times 3 \times 128$ kernels with a subsampling stride of $(1, 1, 1)$ convolve the input data to produce 24 $5 \times 5 \times 1$ feature cubes in the first convolutional layer. Next, similar to their z-axis counterparts, two residual blocks are used to learn deep xy-axis features. For each convolutional in the residual block, 24 $3 \times 3 \times 1$ kernels and padding are used. Finally, an average pooling layer converts the extracted 24 $5 \times 5 \times 1$ deep feature to a 24 $1 \times 1 \times 1$ feature. Through the flatten layer, deep 1×24 features for classification are obtained.

C. Fusion Strategies

Three fusion strategies based on residual networks are proposed, which is probability reconstruction, deep feature fusion, and probability multiplication.

1) Probability Reconstruction: For the deep residual network, an obtained deep feature is input into the dense softmax

layer to obtain the probability vector. In the probability vector, the category corresponding to the maximum value is the category that the sample most likely belongs to. Therefore, we use probability reconstruction to fuse the probability matrix from HSI, HSI_EPLBP, and LiDAR_EPLBP. Such a fusion strategy draws on the ideas in [46] and [30], because the probability and the residual can also represent categories indirectly.

Let $p(\mathbf{y}) = [p_1(\mathbf{y}), p_2(\mathbf{y}), \ldots, p_l(\mathbf{y}), \ldots, p_C(\mathbf{y})]$ represent the probability vector for a testing sample y, where C stands for the total number of the classes and $p_l(\mathbf{y})$ is the output of the dense softmax layer and indicates the probability that the sample y belongs to the l class. The probability reconstruction is described as

$$p(\mathbf{y}_i) = abp_{HSI}(\mathbf{y}_i) + b(1-a)p_{HSI\_EPLBP}(\mathbf{y}_i)$$
$$+ (1-b)p_{LiDAR\_EPLBP}(\mathbf{y}_i) \quad (3)$$

where $p_{HSI}(\mathbf{y}_i)$, $p_{HSI\_EPLBP}(\mathbf{y}_i)$, and $r_{LiDAR\_EPLBP}(\mathbf{y}_i)$ stand for the probability vectors from HSI spectral features, stacked EP and LBP features of HSI, and stacked EP and LBP features of LiDAR, respectively. The parameters a and b control weights of three probability vectors. Finally, category label is computed using

$$\text{class}(\mathbf{y}_i) = \arg\max_{l=1,\ldots,C}(p(\mathbf{y}_i)). \quad (4)$$

The weighting parameters a and b can be calculated by the validation set and the fused probability vector can be obtained as

$$p(\mathbf{v}_i) = abp_{HSI}(\mathbf{v}_i) + b(1-a)p_{HSI\_EPLBP}(\mathbf{v}_i)$$
$$+ (1-b)p_{LiDAR\_EPLBP}(\mathbf{v}_i) \quad (5)$$

where v_i is the ith sample in the validation set, $p_{HSI}(\mathbf{v}_i)$, $p_{HSI\_EPLBP}(\mathbf{v}_i)$, and $r_{LiDAR\_EPLBP}(\mathbf{v}_i)$ stand for the probability vectors from HSI, HSI_EPLBP, and LiDAR_EPLBP, respectively. The prediction label of the validation set is calculated as

$$\text{class}(\mathbf{v}_i) = \arg\max_{l=1,\ldots,C}(p(\mathbf{v}_i)). \quad (6)$$

Because the ground-truth label of the validation set is known, all the values of a, b in steps of 0.01 from 0 to 1 can be traversed to find the max OA for the validation set. The parameter a, b corresponding to the max OA is used in (3). As shown in Fig. 2, a testing sample of the fourth class which was misclassified by three classifiers utilizing each single source, but correctly classified by the fusion method.

2) Deep Feature Fusion: Fig. 3 displays the network of the proposed deep feature fusion strategy. The testing samples of HSI, HSI_EPLBP, and LiDAR_EPLBP are the input into the SSRN network. Then, deep features of each source are extracted. The size of the deep feature is 1×24. The three kinds of deep features are stacked to obtain the fusion features whose size is 1×72. In the end, the fusion features are input into the dense softmax layer to obtain the probability matrix. The probability matrix can be converted to the labels by (4). No additional parameters are required for the deep feature fusion strategy.

3) Probability Multiplication: Inspired by the Bayesian soft fusion [47], we proposed the probability multiplication fusion. The advantage of the probability multiplication fusion is that no additional parameters are needed in the fusion process, and the probability multiplication fusion does not require an additional validation set compared to the probability reconstruction method.

Assume that the corresponding labels of the testing sample y under the HSI, HSI_EPLBP, and LiDAR_EPLBP features are W_i, V_i, and Q_i. The ideal classification result is

$$\max_{W_i, V_i, Q_i} \{P(W_i, V_i, Q_i | y, y, y)\}. \quad (7)$$

Since the HSI, HSI_EPLBP spatial features, and LiDAR_EPLBP spatial features belong to different types of features or different image sources, the probability of the three classification labels for the same coordinate y_i point is independent. Thus, the fusion equations can be designed as

$$P(W_i, V_i, Q_i | y, y, y) = P(W_i | y) \cdot P(V_i | y) \cdot P(Q_i | y) \quad (8)$$
$$P(y_i) = P_{HSI}(y_i). * P_{HSI\_EPLBP}(y_i)$$
$$. * P_{LiDAR\_EPLBP}(y_i). \quad (9)$$

Finally, class label is assigned according to (4).

III. EXPERIMENTAL RESULTS

In the experimental part, two HSI and LiDAR datasets were used. The parameter selection process of our methods and the compared methods were introduced. The proposed methods were evaluated by visual comparison and classification metrics like average accuracy (AA), overall accuracy (OA), and Kappa coefficient.

A. Data

1) 2012 Houston Data: 2012 Houston data contain HSI and LiDAR DSM data which were gathered on June 2012 across the University of Houston campus and the neighboring urban area. The data can be downloaded via the web page.[2] The hyperspectral image consists of 349×1905 pixels and has 144 spectral bands with a wavelength range of 380 to 1050 nm. The corresponding DSM data consist of 349×1905 pixels. The spatial resolution of DSM and HSI is 2.5 m. The HSI was gathered on June 23, 2012, between 17:37:10 and 17:39:50 UTC with 5500 feet the average height of the sensor. The LiDAR data were gathered on June 22, 2012, between 14:37:55 and 15:38:10 UTC with 2000 feet the average height of the sensor. 2012 Houston data include 15 classes which were set by the DFTC via photointerpretation. Table I gives the number of training, validation, and testing samples. In the experiment, training samples and testing samples are standard and given, and validation samples are randomly chosen from testing samples in order to supervise the network and select parameters. Fig. 4 displays HSI, LiDAR DSM, and the areas of the training and

[2] [Online]. Available: http://openremotesensing.net/knowledgebase/matlab-codes-for-fusion-of-hyperspectral-and-lidar-data/

TABLE I
2012 HOUSTON: NUMBER OF TRAINING AND TESTING SAMPLES

	Class		Number of Samples		
No	Name		Training	Validation	Testing
1	Grass Healthy		198	100	1053
2	Grass Stressed		190	100	1064
3	Grass Synthetis		192	100	505
4	Tree		188	100	1056
5	Soil		186	100	1056
6	Water		182	100	143
7	Residential		196	100	1072
8	Commercial		191	100	1053
9	Road		193	100	1059
10	Highway		191	100	1036
11	Railway		181	100	1054
12	Parking Lot 1		192	100	1041
13	Parking Lot 2		184	100	285
14	Tennis Court		181	100	247
15	Running Track		187	100	473
	Total		2832	1500	12197

TABLE II
WERTHEIM: NUMBER OF TRAINING, VALIDATION, AND TESTING SAMPLES

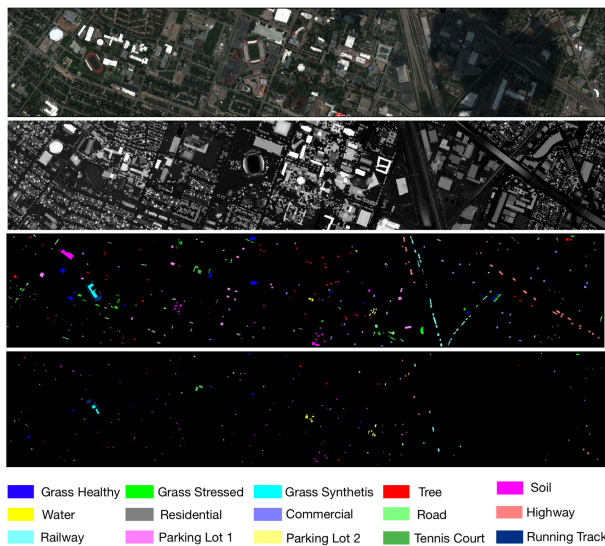|	Class		Number of samples		
No	Name		Training	Validation	Testing
1	Tree		811	969	177163
2	Sea		182	224	10416
3	Sand		130	135	4548
4	Commercial		96	95	1024
5	Pool/Water		50	44	191
6	Major thoroughfares		334	382	4604
7	Bare soil		326	328	6230
8	Road		372	370	6893
9	Parking lot		191	163	1458
10	Grass		296	303	2890
11	Residential grey		116	119	725
12	Residential coffee		144	144	1271
13	Solar panel		30	30	95
14	Car		116	113	420
15	Electric wire		138	140	1052
	Total		3332	3559	218980

Fig. 4. 2012 Houston data—From top to bottom: The HSI utilizing bands 30, 40, and 60, as blue, green, and red respectively; LiDAR DSM image; labeled training samples; labeled testing samples; and the legend of different classes.
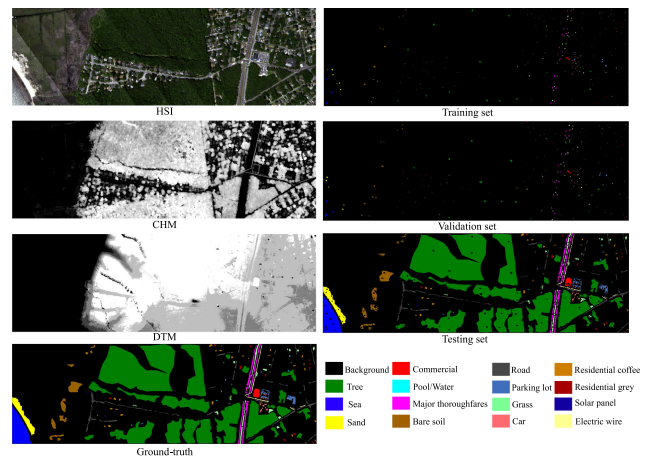

Fig. 5. Wertheim—The HSI data utilizing bands 20, 30, and 40, as blue, green, and red respectively; LiDAR CHM image; LiDAR DTM image; Ground-truth; labeled training samples; labeled validation samples; labeled testing samples; and the legend of different classes.

testing samples. It is worth noting that there is a cloudy area on the right side of the Houston data, which poses a challenge to the fusion algorithm design.

2) Wertheim Data: Wertheim data come from Goddard's LiDAR, Hyperspectral, and Thermal (G-LiHT) data. The G-LiHT Airborne Imager [48] supplies coregistered LiDAR DTM, LiDAR CHM, LiDAR Point Cloud, hyperspectral reflectance image at 1-m spatial resolution, and Google Earth overlay by keyhole markup language (KML) at 0.25-m spatial resolution. The G-LiHT data can be downloaded for free via the web page.[3]

Experiments are conducted on the HSI, CHM, and DTM data which were collected on June 2016 across the Wertheim region (the geographical coordinate is at $40°46'30''$ latitude, $-72°52'48''$ longitude). The hyperspectral image has 114 bands

with spectral range of 420 to 950 nm. The spatial resolution of HSI, CHM, and DTM is 1 m. The dataset of the entire scene contains 501×1523 pixels. According to the calibration method [40], ground-truth is obtained and 15 categories are in the Wertheim data. After obtaining the ground-truth, we manually select the samples from some regions in ground-truth as training and validation sets. Since training samples of the 2012 Houston data were selected by regions, we follow the same way. Table II shows these categories and the corresponding number of training, validation, and testing samples. Fig. 5 gives HSI, CHM, DTM, the labeled ground-truth, the positions of the training samples, the positions of the validation samples, the positions of the testing samples, and legend of 15 categories.

B. Comparison of Methods

HSI, HSI_EPLBP, $LiDAR\_EPLBP$ present the classification accuracies of HSI, HSI_EPLBP, LiDAR_EPLBP directly using SSRN in the experiment. Four fusion methods are

---

[3][Online]. Available: https://glihtdata.gsfc.nasa.gov

used for comparison due to their popularity, which are briefly described as follows.

Stacking means the direct stacked feature of HSI and LiDAR data. For Houston data, the direct stacked feature is HSI+DSM. For Wertheim, the stacked feature is HSI+CHM+DTM. CHM+DTM has better classification performance than DSM [40], but Houston data only provides DSM. Feature stacking (FS) means the stacked feature of HSI, HSI_EPLBP, LiDAR_EPLBP. The generalized graph-based fusion (GGF) [23] algorithm is utilized as a comparison algorithm. Max voting (MV) means HSI, HSI_EPLBP, LiDAR_EPLBP are fused by 10, where $L_{HSI}(x_i)$ means the category label of HSI sample, $L_{HSI\_EPLBP}(x_i)$ is the category label of HSI_EPLBP sample, and $L_{LiDAR\_EPLBP}(x_i)$ is the category label of Li-DAR_EPLBP sample. When the three labels are inconsistent, the label of HSI is selected as

$$L = \begin{cases} L_{HSI\_EPLBP}(x_i) \\ \qquad L_{HSI\_EPLBP}(x_i) = L_{LiDAR\_EPLBP}(x_i) \\ L_{HSI}(x_i) \quad \text{else.} \end{cases}$$
(10)

It is worth noting that the validation set is still used to supervise the training process. Therefore, the validation set cannot be used as additional samples of training for the MV strategy.

The four proposed methods to be investigated are: Residual network-based probability reconstruction fusion (RNPRF) using the framework shown in Fig. 2, residual network-based deep feature fusion (RNDFF) using the framework shown in Fig. 3, and residual network-based probability multiplication fusion (RNPMF) using the probability multiplication fusion strategy. Compared with RNPRF, RNPMF does not require additional parameters. RNPRF.*RNDFF is that the probability matrices of the RNPRF are combined with the probability matrices of the RNDFF by array multiplication as

$$P(y_i) = P_{RNPRF}(y_i). * P_{RNDFF}(y_i).$$
(11)

C. Parameter Settings

For parameter selection for generating EP features, α is set to 5, and s is set to 7 according to [29], [42]. For LBP features, the parameters (m, r) is selected as $(8, 1)$ and the size of the local patch is selected as 21×21 according to [43]. For the 2012 Houston data, the HSI_EPLBP has 225 EP features and 590 LBP features, and LiDAR_EPLBP has 75 EP features and 59 LBP features. For Wertheim data, the HSI_EPLBP has 225 EP features and 590 LBP features, and LiDAR_EPLBP has 150 EP features and 118 LBP features.

There are four parameters to be set for the residual network: The learning rate, the spatial size of the input cubes, the kernel number of convolutional layers, and the number of residual blocks, which control the training process and classification accuracy of the trained network. Since the training set is small, the batch size is set to 16. In the training process, the models with the highest classification performance in the validation set are retained, and all the results are generated by these optimal models. In each training process, 200 epochs are sufficient for the SSRN to converge [38].

TABLE III
OA (%) OF THE RESIDUAL NETWORK WITH DIFFERENT INPUT SIZES

OA	Input Size	HSI	HSI_EPLBP	LiDAR_EPLBP
Houston	3x3	87.24	85.75	57.56
	5x5	87.55	86.37	61.23
	7x7	88.18	87.19	63.33
	9x9	87.58	87.76	62.39
	11x11	**88.98**	**87.77**	**66.29**
Wertheim	3x3	98.58	99.19	97.83
	5x5	98.70	99.31	98.12
	7x7	99.19	**99.46**	98.19
	9x9	**99.39**	99.34	98.26
	11x11	99.05	99.30	**98.48**

TABLE IV
OA (%) OF THE RESIDUAL NETWORK WITH DIFFERENT KERNEL NUMBERS

OA	Kernel Numbers	HSI	HSI_EPLBP	LiDAR_EPLBP
Houston	8	86.37	78.25	47.00
	16	**90.75**	87.81	61.74
	24	88.98	87.77	66.29
	32	89.11	**88.16**	**67.63**
Wertheim	8	98.89	97.04	97.46
	16	99.39	98.93	97.99
	24	**99.39**	**99.46**	**98.48**
	32	99.25	99.35	97.14

The learning rate controls convergence speed. In particular, inappropriate learning rate settings will result in divergence or slow convergence. In the experiment, the grid search method is used, and the optimal learning rates are selected from $\{0.01, 0.003, 0.001, 0.0003, 0.0001, 0.00003\}$ for each dataset. According to the classification result, the learning rates of Houston HSI and HSI_EPLBP are selected to be 0.0003, and the learning rates of Houston LiDAR_EPLBP is set to 0.0001. For Wertheim data, the learning rates are selected to be 0.0003 for all three features.

Table III shows OA (%) of the residual network with different input sizes. For Houston data, as the sample size increases, the classification accuracy increases. For Wertheim data, the input size of HSI with the highest classification accuracy is 9×9, the input size of HSI_EPLBP with the highest classification accuracy is 7×7, and the input size of LiDAR_EPLBP with the highest classification accuracy is 11×11. A larger sample input size may not guarantee higher classification accuracy, because larger input samples may increase interference from other classes of samples.

Fig. 3 shows that the SSRN has the same kernel numbers in each convolutional layer of the residual blocks. In Table IV, the classification accuracy of the residual network under three features is tested, where the kernel number is from 8 to 32, with an interval of 8. In Table IV, for the Houston data, the models with 16 kernels in each convolutional layer of the residual blocks offer the highest classification accuracy for HSI, and the models with 32 kernels produce the highest classification accuracy for HSI_ELPBP and LiDAR_EPLBP. For the Wertheim data, the models with 24 kernels achieve the best performance for the three features.

The number of residual blocks represents the depth of the residual network and describes the representation capability of the residual network. Table IV shows the OA (%) of the residual
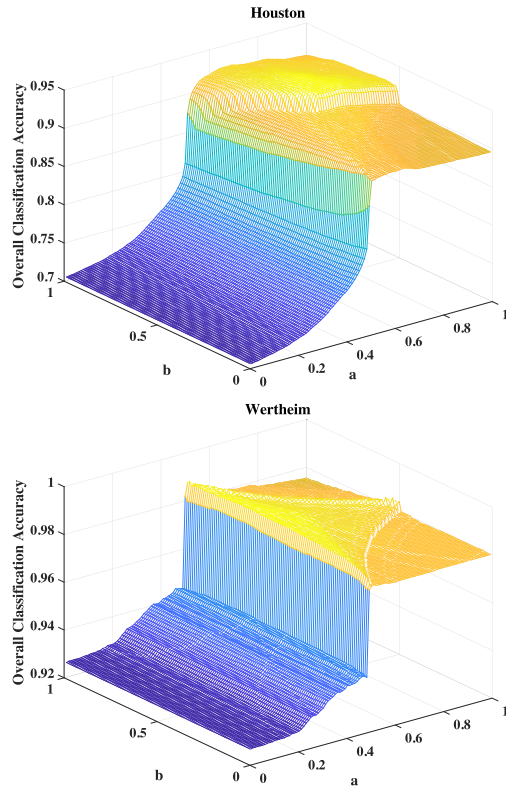
Fig. 6. Parameter a and b selection result for RNPRF.

network with different numbers of residual blocks. Table IV shows that for Houston's HSI, HSI_EPLBP, LiDAR_EPLBP features, the models with the highest classification accuracy contain $2+2$, $3+2$, $3+2$ residual blocks, respectively. For Wertheim's HSI, HSI_EPLBP, LiDAR_EPLBP features, the models with the highest classification accuracy contain $2+2$, $2+3$, $2+2$ residual blocks, respectively. The model with the highest classification accuracy of HSI_EPLBP features requires more residual blocks than that of HSI because the dimension of HSI_EPLBP features (815) is higher than the dimension of HSI (114).

For the RNPRF, Fig. 6 describes the parameter selection process for (5) and (6). For Houston data, the parameter a and b are selected by validation set as 0.61 and 0.72, respectively, while for Wertheim data, they are 0.65 and 0.51, respectively.

For the GGF, the numbers of nearest neighbors of features were selected as 150. For Wertheim data, EP+LBP CHM features and EP+LBP DTM features are stacked as one of the inputs to the GGF. The GGF uses the support vector machine (SVM) classifier, whose parameters (i.e., c and g) are tuned via cross-validation. For Houston data, after the GGF, the dimension of the fusion features with the highest classification accuracy is 38. For Wertheim data, the dimension of the fusion features with the highest classification accuracy is 74.

D. Experimental Results and Discussion

Tables VI and VII show the classification result of HSI, HSI_EPLBP, and LiDAR_EPLBP achieved by the SSRN for

TABLE V
OA (%) OF THE RESIDUAL NETWORK WITH DIFFERENT NUMBERS OF RESIDUAL BLOCKS. THE $\alpha+\beta$ MEANS AN SSRN USING α Z-AXIS RESIDUAL BLOCKS AND β X-AXIS AND Y-AXIS RESIDUAL BLOCKS

OA	Residual Block Numbers	HSI	HSI_EPLBP	LiDAR_EPLBP
	1+1	87.56	86.79	67.11
	1+2	89.74	87.83	67.46
	2+1	88.72	88.28	67.23
Houston	2+2	**90.75**	88.16	67.63
	3+2	87.23	**89.46**	**70.42**
	2+3	90.10	86.69	67.72
	3+3	88.51	86.81	66.88
	1+1	99.20	99.18	98.19
	1+2	99.31	99.14	97.74
	2+1	99.17	99.28	98.12
Wertheim	2+2	**99.39**	99.46	**98.48**
	3+2	99.36	99.30	98.47
	2+3	99.13	**99.54**	97.94
	3+3	98.69	99.51	98.37

TABLE VI
HOUSTON DATA: CLASSIFICATION RESULT OF HSI, HSI_EPLBP, AND LiDAR_EPLBP ACHIEVED BY THE SSRN

	HSI	HSI_EPLBP	$LiDAR\_EPLBP$
1	82.24	79.39	27.35
2	85.15	84.87	65.70
3	99.80	98.61	84.55
4	93.09	96.21	68.56
5	100.00	100.00	88.73
6	95.80	100.00	79.02
7	85.26	91.70	79.38
8	84.14	83.29	84.71
9	88.57	83.95	53.82
10	91.80	100.00	74.23
11	89.09	95.54	100.00
12	98.75	71.28	59.08
13	88.42	80.70	72.63
14	100.00	100.00	100.00
15	97.67	94.93	41.44
OA	90.75	89.46	70.42
AA	91.99	90.70	71.95
Kappa	0.8996	0.8857	0.6794

TABLE VII
WERTHEIM DATA: CLASSIFICATION RESULT OF HSI, HSI_EPLBP, AND LiDAR_EPLBP ACHIEVED BY THE SSRN

	HSI	HSI_EPLBP	$LiDAR\_EPLBP$
1	99.98	99.92	99.87
2	99.90	100.00	98.17
3	100.00	100.00	99.30
4	64.45	61.43	80.27
5	100.00	100.00	100.00
6	96.31	98.07	88.12
7	100.00	100.00	97.67
8	93.75	99.33	89.37
9	99.04	94.17	77.30
10	99.69	99.97	69.97
11	87.17	98.21	94.76
12	100.00	100.00	100.00
13	100.00	100.00	100.00
14	93.81	95.48	99.52
15	81.84	79.09	97.62
OA	99.39	99.54	98.48
AA	94.40	95.04	92.80
Kappa	0.9820	0.9865	0.9553

TABLE VIII
HOUSTON DATA: CLASSIFICATION RESULTS ACHIEVED BY EIGHT
FUSION METHODS

	Stacking	*FS*	*MV*	*GGF*
1	82.15	78.73	78.63	81.48
2	85.15	85.15	85.15	84.21
3	99.01	97.82	98.61	100.00
4	98.77	92.61	93.09	92.80
5	100.00	100.00	100.00	97.63
6	95.80	100.00	95.80	95.80
7	82.00	98.97	90.21	92.72
8	95.54	79.01	87.08	90.12
9	95.00	99.15	91.69	87.82
10	63.90	90.93	95.66	67.08
11	98.67	100.00	96.30	95.45
12	90.59	92.51	99.23	94.04
13	90.18	74.74	89.82	78.95
14	100.00	100.00	100.00	100.00
15	100.00	99.58	97.25	99.58
OA	90.35	92.15	92.36	89.39
AA	91.78	92.61	93.24	90.51
Kappa	0.8954	0.9147	0.9171	0.8848
	RNPMF	*RNPRF*	*RNDFF*	*RNPRF.* ∗ *RNDFF*
1	78.63	78.63	79.39	79.49
2	84.96	85.15	82.61	85.06
3	91.68	98.42	98.02	98.22
4	95.27	95.93	92.52	93.37
5	100.00	100.00	100.00	100.00
6	100.00	100.00	93.01	95.80
7	91.79	91.70	91.79	95.24
8	93.83	95.54	84.24	90.22
9	81.78	95.66	83.76	89.24
10	96.33	97.68	91.12	97.30
11	100.00	100.00	98.01	100.00
12	85.59	99.42	99.62	100.00
13	85.61	89.82	73.68	89.12
14	100.00	100.00	100.00	100.00
15	95.56	97.25	100.00	100.00
OA	91.20	**94.36**	90.83	93.55
AA	92.07	**95.01**	91.18	94.20
Kappa	0.9045	**0.9388**	0.9004	0.9299

TABLE IX
WERTHEIM DATA: CLASSIFICATION RESULTS ACHIEVED BY
EIGHT FUSION METHODS

	Stacking	*FS*	*MV*	*GGF*
1	99.97	100.00	99.99	99.97
2	100.00	100.00	100.00	97.89
3	100.00	100.00	100.00	100.00
4	70.41	98.93	73.24	87.01
5	100.00	100.00	100.00	90.05
6	97.81	97.74	98.96	99.44
7	100.00	100.00	100.00	100.00
8	96.69	94.47	98.24	97.40
9	94.44	76.95	96.57	95.82
10	99.72	99.62	99.72	99.86
11	98.90	100.00	97.38	92.00
12	99.84	100.00	100.00	94.41
13	91.58	100.00	100.00	96.84
14	94.76	99.52	96.67	97.62
15	94.20	99.90	86.22	99.81
OA	99.60	99.61	99.68	99.62
AA	95.89	97.81	96.47	96.54
Kappa	0.9883	0.9886	0.9906	0.9887
	RNPMF	*RNPRF*	*RNDFF*	*RNPRF.* ∗ *RNDFF*
1	100.00	100.00	100.00	100.00
2	100.00	100.00	99.99	100.00
3	100.00	100.00	100.00	100.00
4	81.64	80.27	91.50	84.28
5	100.00	100.00	100.00	100.00
6	99.57	99.04	98.33	99.13
7	100.00	100.00	100.00	100.00
8	99.25	98.16	96.84	98.42
9	97.05	93.07	99.38	99.52
10	99.41	97.44	99.86	99.93
11	99.17	98.76	98.90	99.72
12	100.00	100.00	100.00	100.00
13	100.00	100.00	100.00	100.00
14	97.14	98.57	98.33	98.81
15	83.08	90.68	95.82	96.20
OA	99.76	99.70	99.79	**99.83**
AA	97.09	97.07	**98.60**	98.40
Kappa	0.9930	0.9911	0.9939	**0.9951**

the Houston data and the Wertheim data, respectively. The classification accuracy of Wertheim data is higher than the Houston data because in the Wertheim testing samples, the tree samples are the majority and tree samples in three features are easily recognized by the classifiers. Therefore, for Wertheim data, AA is much lower than OA. In addition, the spatial resolution of Wertheim data is 1 m and the spatial resolution of Houston data is 2.5 m. Therefore, Wertheim's sample is less likely to contain multiple objects and easy to be classified.

Table VIII shows the per-class accuracy, AA, OA, and Kappa coefficient of our proposed methods and the compared methods for the Houston data. As we can see, FS offers higher accuracy than $Stacking$, which means that the EP+LBP features are valid for the used deep residual network. The performance of $RNPRF$ is higher than MV because RNPRF fuses the probability matrices, not the category labels, and even if the results of the three classifiers are wrong, RNPRF can correct the result, but the MV method cannot. Therefore, RNPRF can achieve more accurate decision-level fusion. The GGF method is a typical representative algorithm for stacking plus dimension reduction fusion strategy. The RNPRF.*RNDFF is the combination of the RNPRF and the RNDFF. In Table VIII, the proposed $RNPRF$, $RNDFF$, $RNPMF$, and $RNPRF. ∗ RNDFF$ outperform the benchmark methods $Stacking$ and GGF, which demonstrates the effectiveness of proposed methods.

Table IX lists the accuracy of each class, AA, OA, and Kappa coefficient of our proposed methods and the compared methods for the Wertheim data. The classification accuracy of FS is higher than $Stacking$ for Wertheim data, which means that the EP+LBP features are valid for the deep residual network. The proposed $RNPRF$, $RNDFF$, and $RNPMF$ again outperform the compared methods $Stacking$, FS, MV, and GGF. As shown in Table IX, $RNPRF. ∗ RNDFF$ offers better performance than $RNPRF$ and $RNDFF$, which means that the combination of proposed fusion frameworks may further improve classification performance.

To show statistical significance in performance improvement, we apply the nonparametric McNemar test [49], [50] to evaluate Kappa accuracy improvement with the proposed methods for Wertheim data. For two methods to be compared, f_{12} denotes the number of samples that the first method misclassifies but the second method can correctly classify. f_{21} is the number of samples that the second method misclassifies but the first method can correctly classify. Then, the McNemar's test for two methods can be defined as

$$z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}}. \quad (12)$$

For a 5% level of significance, the corresponding z value is 1.96. If z value is greater than 1.96, the second method is better than the
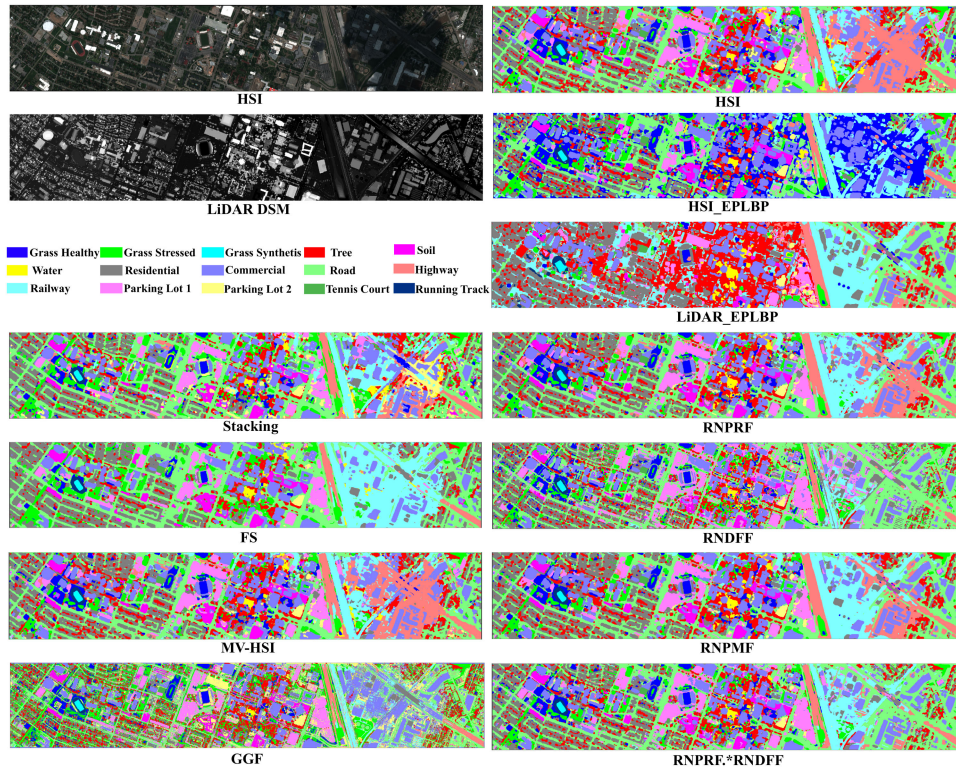
Fig. 7. Comparison of the mapping results of Houston data.

TABLE X
z VALUES (AND SIGNIFICANCE) IN THE McNEMAR'S TEST FOR WERTHEIM
DATA (THE 5% LEVEL OF SIGNIFICANCE IS SELECTED)

	$RNPMF$	$RNPRF$	$RNDFF$	$RNPRF.*RNDFF$
$Stacking$	12.25(Yes)	6.84(Yes)	12.70(Yes)	17.11(Yes)
FS	9.58(Yes)	5.81(Yes)	12.35(Yes)	15.46(Yes)
MV	9.61(Yes)	1.86(No)	8.60(Yes)	14.45(Yes)
GGF	8.85(Yes)	4.75(Yes)	10.89(Yes)	14.10(Yes)

TABLE XI
z VALUES (AND SIGNIFICANCE) IN THE McNEMAR'S TEST FOR 2012 HOUSTON
DATA (THE 5% LEVEL OF SIGNIFICANCE IS SELECTED)

	$RNPMF$	$RNPRF$	$RNDFF$	$RNPRF.*RNDFF$
$Stacking$	2.99(Yes)	16.27(Yes)	1.60(No)	11.92(Yes)
FS	-3.87(No)	10.42(Yes)	-5.56(No)	6.78(Yes)
MV	-5.19(No)	14.33(Yes)	-6.67(No)	6.83(Yes)
GGF	4.91(Yes)	14.53(Yes)	3.96(Yes)	11.99(Yes)

first method, and the two methods have significant performance discrepancy. Table X shows the z values when the Stacking, FS, MV, and GGF methods are set as the location of the first method and the newly proposed RNPMF, RNPRF, RNDFF, and RNPRF.*RNDFF methods are set as the location of the second method in f_{12} and f_{21} For Wertheim data. The performance of the newly proposed methods is statistically better than benchmarking methods, except that the discrepancy between RNPRF and MV in the Wertheim data is statistically insignificant. For the 2012 Houston data in Table XI, RNPRF and RNPRF.*RNDFF perform the best.

Figs. 7 and 8 show the comparison of the mapping results of Houston data and Wertheim data, together with the false-color image of HSI, LiDAR data, and the colors of the corresponding category labels. Due to the limited number of testing samples, the evaluation results may be incomplete. Therefore, it is necessary to perform a full-image classification to more comprehensively compare the classification performance of the fusion methods especially for the unlabeled areas.

In Fig. 7, visual comparison between different methods is in agreement with quantitative comparison in Tables VI and VIII. As shown in Fig. 7, the GGF generated classification map includes high level of great noise for Houston data. Fig. 7 shows that the proposed RNPRF, RNDFF, RNPMF produce more accurate and smooth classification maps than GGF because the SSRN can fully extract information of the sample tensors, and the fusion strategies fully integrate the information in HSI and LiDAR features. In Fig. 7, for the cloudy area in Houston, the RNDFF and the corresponding combination algorithm RNPRF.*RNDFF generate more accurate classification maps.

In Fig. 8, the visual comparison between different methods matches the quantitative comparison in Tables VII and IX. As shown in Fig. 8, the GGF generated classification map includes obvious misclassification in the beach area, for instance, it misclassifies some sea samples as trees. However, our proposed method RNPRF, RNDFF, RNPMF, and their combination algorithms have more accurate and smooth classification maps than GGF.

For the fusion classification of LiDAR and HSI data, some methods have been proposed in recent years. But, many fusion algorithms are very complex. Since the 2012 Houston dataset is the standard dataset and given training and testing set, algorithm comparisons are convenient. In Table XII, compared
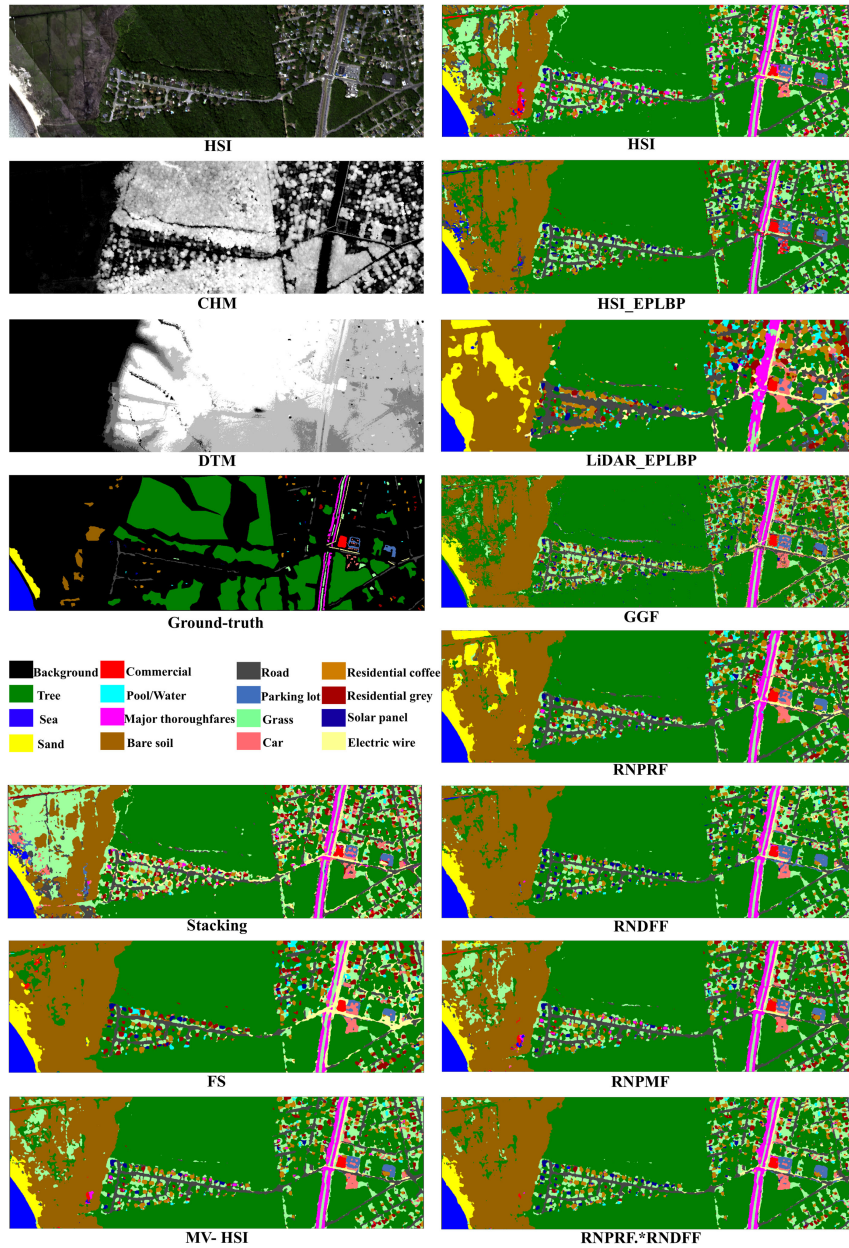
Fig. 8. Comparison of the mapping results of Wertheim data.

TABLE XII
HOUSTON DATA: CLASSIFICATION RESULTS ACHIEVED BY SOME EXISTING METHODS AND RNPRF

	Image-to-Image CNN[51]	PToP CNN[52]	CNN+ELM[53]	H+L CNN[54]	Proposed RNPRF
OA	86.81	92.48	92.57	87.98	94.36
AA	87.20	93.55	92.48	90.11	95.01
Kappa	0.8576	0.9187	0.9193	0.8698	0.9388

Table XII doesn't need the bold entities because it is not for result comparison and just show the classification accuracy of each data source.

with the results of image-to-image CNN [51], PToP CNN [52], CNN+ELM [53], H+L CNN [54] in their papers, our proposed method RNPRF has higher classification accuracy in the 2012 Houston data.

Table XIII provides the training and testing time of the residual network in various methods. We carried out all experiments on

an Intel Xeon E5-1620 v4 central processing unit and NVIDIA 2080 graphics processing unit machine with 64 GB of random access memory. It is worth noting that the proposed RNPRF and RNPMF methods can simultaneously use the SSRN to train the HSI, HSI_EPLBP, and LiDAR_EPLBP features, which can greatly save training time. In Table XIII, the slowest training

TABLE XIII
TRAINING AND TESTING TIME OF USING DIFFERENT METHODS

		HSI	HSI_EPLBP	LiDAR_EPLBP	Stacking	FS	RNDFF	RNPRF	RNPMF
Houston	Training Time(s)	2793.74	9563.04	2842.56	3526.13	8916.90	4969.12	9563.04	9563.04
	Testing Time(s)	21.37	45.10	9.38	38.05	49.59	15.78	45.10	45.10
Wertheim	Training Time(s)	2658.78	4616.27	3965.53	2472.42	8891.67	6600.07	4616.27	4616.27
	Testing Time(s)	129.93	126.23	118.32	139.09	260.68	31.54	129.93	129.93

and testing time of HSI, HSI_EPLBP, and LiDAR_EPLBP represent the training and testing time of RNPRF and RNPMF. Fortunately, the support of GPUs has greatly reduced the extra computational cost and shortened the training time.

It was noted that the selection of the validation samples is important for the performance of the proposed methods because both the model selection and the parameter selection of RNPRF depend on the validation samples.

IV. CONCLUSION

A new framework is proposed for the fusion of hyperspectral and LiDAR data based on EP, LBP, and deep residual network. Specifically, EP and LBP features are extracted from two sources. Then, the deep features of each source are extracted by SSRN and fused by RNDFF. While for probability fusion, the deep features of each source are extracted by the SSRN, and probability matrices can be got through the softmax classifier. Reconstruction probability matrices are fused by RNPRF and RNPMF to generate the final classification map. The proposed fusion algorithms, i.e., RNPRF, RNDFF, and RNPMF, have no additional parameters to tune, and offer good universality and excellent classification performance for the LiDAR and HSI fusion and can be used for actual G-LiHT data. Our proposed probability fusion strategy has been first applied to the fusion classification of HSI and LiDAR images in the field of remote sensing, and it is superior to some existing methods in performance.

REFERENCES

[1] B. Huang, B. Zhao, and Y. Song, "Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery," *Remote Sens. Environ.*, vol. 214, pp. 73–86, 2018.
[2] C. Debes et al., "Hyperspectral and lidar data fusion: Outcome of the 2013 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, Jun. 2014.
[3] Y. Xu, Q. Du, W. Li, and N. H. Younan, "Efficient probabilistic collaborative representation-based classifier for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 11, pp. 1746–1750, Nov. 2019.
[4] J. Peng and Q. Du, "Robust joint sparse representation based on maximum correntropy criterion for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7152–7164, Dec. 2017.
[5] T. Zhan, L. Sun, Y. Xu, G. Yang, Y. Zhang, and Z. Wu, "Hyperspectral classification via superpixel kernel learning-based low rank representation," *Remote Sens.*, vol. 10, no. 10, 2018 , Art. no. 1639.
[6] C. Yu, B. Xue, M. Song, Y. Wang, S. Li, and C. Chang, "Iterative target-constrained interference-minimized classifier for hyperspectral classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1095–1117, Apr. 2018.
[7] B. Xue et al., "A subpixel target detection approach to hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5093–5114, Sep. 2017.
[8] P. Fu, X. Sun, and Q. Sun, "Hyperspectral image segmentation via frequency-based similarity for mixed noise estimation," *Remote Sens.*, vol. 9, no. 12, 2017, Art. no. 1237.
[9] A. M. Saranathan and M. Parente, "Uniformity-based superpixel segmentation of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1419–1430, Mar. 2016.
[10] X. Zhang, J. Zhang, C. Li, C. Cheng, L. Jiao, and H. Zhou, "Hybrid unmixing based on adaptive region segmentation for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3861–3875, Jul. 2018.
[11] W. Xie, J. Lei, Y. Cui, Y. Li, and Q. Du, "Hyperspectral pansharpening with deep priors," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 31, no. 5, pp. 1529–1543, May 2019.
[12] W. Xie, Y. Shi, Y. Li, X. Jia, and J. Lei, "High-quality spectral-spatial reconstruction using saliency detection and deep feature enhancement," *Pattern Recognit.*, vol. 88, pp. 139–152, 2019.
[13] J. Sreevalsan-Nair and A. Jindal, "Using gradients and tensor voting in 3D local geometric descriptors for feature detection in airborne LiDAR point clouds in urban regions," in *Proc. IEEE Int. Geosci. Remote Sensing Symp., Conf. Proc.*, 2017, pp. 5881–5884.
[14] A. Y. Hata and D. F. Wolf, "Feature detection for vehicle localization in urban environments using a multilayer LiDAR," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 420–429, Feb. 2016.
[15] H. Guan, Y. Yu, J. Li, and P. Liu, "Pole-like road object detection in mobile LiDAR data via supervoxel and bag-of-contextual-visual-words representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 4, pp. 520–524, Apr. 2016.
[16] X. Hu, Y. Li, J. Shan, J. Zhang, and Y. Zhang, "Road centerline extraction in complex urban scenes from LiDAR data based on multiple features," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7448–7456, Nov. 2014.
[17] Y. Yu, J. Li, H. Guan, F. Jia, and C. Wang, "Learning hierarchical features for automated extraction of road markings from 3-D mobile LiDAR point clouds," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 2, pp. 709–726, Feb. 2015.
[18] E. N. Broadbent et al., "Linking rainforest ecophysiology and micro-climate through fusion of airborne LiDAR and hyperspectral imagery," *Ecosphere*, vol. 5, no. 5, pp. 1–37, 2014.
[19] G. V. Laurin et al., "Above ground biomass estimation in an African tropical forest with LiDAR and hyperspectral data," *ISPRS J. Photogrammetry Remote Sens.*, vol. 89, pp. 49–58, 2014.
[20] M. Huesca, D. Riao, and S. L. Ustin, "Spectral mapping methods applied to LiDAR data: Application to fuel type mapping," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 74, pp. 159–168, 2019.
[21] C. R. Hakkenberg, K. Zhu, R. K. Peet, and C. Song, "Mapping multi-scale vascular plant richness in a forest landscape with integrated LiDAR and hyperspectral remote-sensing," *Ecology*, vol. 99, no. 2, pp. 474–487.
[22] X. Xu, W. Li, Q. Ran, Q. Du, L. Gao, and B. Zhang, "Multisource remote sensing data classification based on convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 937–949, Feb. 2018.
[23] W. Liao, A. Pižurica, R. Bellens, S. Gautama, and W. Philips, "Generalized graph-based fusion of hyperspectral and LiDAR data using morphological features," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 3, pp. 552–556, Mar. 2015.
[24] Y. Gu and Q. Wang, "Discriminative graph-based fusion of HSI and LiDAR data for urban area classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 6, pp. 906–910, Jun. 2017.
[25] C. Zhang, M. Smith, and C. Fang, "Evaluation of Goddards LiDAR, hyperspectral, and thermal data products for mapping urban land-cover types," *GIScience Remote Sens.*, vol. 55, no. 1, pp. 90–109, 2018.

[26] M. Dalponte, L. Bruzzone, and D. Gianelle, "Fusion of hyperspectral and LiDAR remote sensing data for classification of complex forest areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1416–1427, May 2008.

[27] B. Rasti, P. Ghamisi, J. Plaza, and A. Plaza, "Fusion of hyperspectral and LiDAR data using sparse and low-rank component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6354–6365, Nov. 2017.

[28] B. Rasti, P. Ghamisi, and R. Gloaguen, "Hyperspectral and LiDAR fusion using extinction profiles and total variation component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3997–4007, Jul. 2017.

[29] P. Ghamisi, B. Hofle, and X. X. Zhu, "Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 6, pp. 3011–3024, Jun. 2017.

[30] C. Ge, Q. Du, W. Li, Y. Li, and W. Sun, "Hyperspectral and LiDAR data classification using kernel collaborative representation based residual fusion," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1963–1973, Jun. 2019.

[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. Comput. Vis. Pattern Recognit.*, *Conf. Proc.*, 2016, pp. 770–778.

[32] K. He, X. Zhang S. Ren, J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.*, S. N. W. M. e. C. V. Leibe B., M. J., Ed., vol. 9908, 2016, pp. 630–645.

[33] M. Zhang, W. Li, Q. Du, L. Gao, and B. Zhang, "Feature extraction for classification of hyperspectral and LiDAR data using patch-to-patch CNN," *IEEE Trans. Cybern.*, vol. 50, no. 1, pp. 100–111, Jan. 2020.

[34] Y. Xu, B. Du, and L. Zhang, "Multi-source remote sensing data classification via fully convolutional networks and post-classification processing," in *Proc. IEEE Int. Geoscience and Remote Sensing Symp., Conf. Proc.*, 2018, pp. 3852–3855.

[35] H. Li, P. Ghamisi, U. Soergel, and X. X. Zhu, "Hyperspectral and LiDAR fusion using deep three-stream convolutional neural networks," *Remote Sens.*, vol. 10, no. 10, 2018, Art. no. 1649.

[36] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning 6.2.2.3 Softmax Units for Multinoulli Output Distributions*. Cambridge, MA, USA: MIT Press, 2016.

[37] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environment*, vol. 113, pp. S 110–S122, 2009.

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

[39] L. Wei, G. Wu, Z. Fan, and D. Qian, "Hyperspectral image classification using deep pixel-pair features," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 844–853, Feb. 2017.

[40] Y. Li, C. Ge, W. Sun, J. Peng, Q. Du, and K. Wang, "Hyperspectral and LiDAR data fusion classification using superpixel segmentation-based local pixel neighborhood preserving embedding," *Remote Sens.*, vol. 11, no. 5, 2019, Art. no. 550.

[41] P. Ghamisi, R. Souza, J. A. Benediktsson, X. X. Zhu, L. Rittner, and R. A. Lotufo, "Extinction profiles for the classification of remote sensing data," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5631–5645, Oct. 2016.

[42] P. Ghamisi, R. Souza, J. A. Benediktsson, L. Rittner, R. Lotufo, and X. X. Zhu, "Hyperspectral data classification using extended extinction profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 11, pp. 1641–1645, Nov. 2016.

[43] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[44] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3681–3693, Jul. 2015.

[45] I. T. Jolliffe, *Principal Component Analysis (Springer Series in Statistics)*. New York, NY, USA: Springer-Verlag, 1986.

[46] J. Zou, W. Li, C. Chen, and Q. Du, "Scene classification using local and global features with collaborative representation fusion," *Inf. Sci.*, vol. 348, pp. 209–226, 2016.

[47] C. Wu, B. Du, X. Cui, and L. Zhang, "A post-classification change detection method based on iterative slow feature analysis and bayesian soft fusion," *Remote Sens. Environ.*, vol. 199, pp. 241–255, 2017.

[48] B. Cook *et al.*, "Nasa Goddards LiDAR, hyperspectral and thermal (G-LiHT) airborne imager," *Remote Sens.*, vol. 5, no. 8, 2013, Art. no. 4045.

[49] G. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy: Photogrammetric engineering remote sensing," *Photogrammetric Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, 2004.

[50] H. Su, H. Yang, Q. Du, and Y. Sheng, "Semisupervised band clustering for dimensionality reduction of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 6, pp. 1135–1139, Nov. 2011.

[51] M. Zhang, W. Li, X. Wei, and X. Li, "Collaborative classification of hyperspectral and LiDAR data using unsupervised image-to-image CNN," in *Proc. 10th IAPR Workshop on Pattern Recognit. Remote Sensing.*, 2018, pp. 1–5.

[52] M. Zhang, W. Li, Q. Du, L. Gao, and B. Zhang, "Feature extraction for classification of hyperspectral and LiDAR data using patch-to-patch CNN," *IEEE Trans. Cybern.*, vol. 50, no. 1, pp. 100–111, Jan. 2020.

[53] H. Li, P. Ghamisi, U. Soergel, and X. Zhu, "Hyperspectral and LiDAR fusion using deep three-stream convolutional neural networks," *Remote Sens.*, vol. 10, no. 10, p. 1649, 2018.

[54] X. Xu, W. Li, Q. Ran, Q. Du, L. Gao, and B. Zhang, "Multisource remote sensing data classification based on convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 937–949, Feb. 2018.

Chiru Ge (Member, IEEE) received the B.E. degree in communication engineering from Shandong Normal University, Jinan, China, in 2012, and the Ph.D. degree in communication and information systems from Xidian University, Xi'an, China, in 2019.

From 2017 to 2018, he was with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA, as a Visiting Scholar with Prof. Q. Du, to study on hyperspectral and LiDAR image fusion. He is currently a Lecturer with Shandong Normal University, Jinan, China. His research interests include hyperspectral remote sensing image analysis and processing, hyperspectral, and LiDAR image fusion.

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of MarylandBaltimore 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, pattern classification, data compression, and neural networks.

Dr. Du is a Fellow of the SPIE International Society for Optics and Photonics. She was a Cochair of the Data Fusion Technical Committee of the IEEE GRSS from 2009 to 2013. She was the Chair of the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition from 2010 to 2014. She was the General Chair of the fourth IEEE Geoscience and Remote Sensing Society (GRSS) Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing held at Shanghai, China, in 2012. She was a recipient of the 2010 Best Reviewer Award from the IEEE GRSS. She has served as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS), the *Journal of Applied Remote Sensing*, and the IEEE SIGNAL PROCESSING LETTERS. Since 2016, she has been the Editor-in-Chief of the IEEE JSTARS.
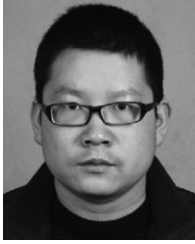
Weiwei Sun (Member, IEEE) received the B.S. degree in surveying and mapping and the Ph.D. degree in cartography and geographic information engineering from Tongji University, Shanghai, China, in 2007 and 2013, respectively.

From 2011 to 2012, he was with the Department of Applied Mathematics, University of Maryland at College Park, College Park, MD, USA, as a Visiting Scholar with the famous Prof. J. Benedetto to study on the dimensionality reduction of hyperspectral image. From 2014 to 2016, he held a Postdoctoral position with the State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, to study intelligent processing in hyperspectral imagery. He is currently a Professor with Ningbo University, Ningbo, China. He is also a Visiting Scholar with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. He has authored over 50 journal papers. His research interests include hyperspectral image processing with manifold learning, anomaly detection, and target recognition of remote sensing imagery using compressive sensing.

Jiaojiao Li (Member, IEEE) received the B.E. degree in computer science and technology, the M.S. degree in software engineering, and the Ph.D. degree in communication and information systems from Xidian University, Xi'an, China, in 2009, 2012, and 2016, respectively.

She was an exchange Ph.D. Student of Mississippi State University, Starkville, MS, USA, supervised by Dr. Qian Du. She is a Postdoctoral Researcher and a Lecturer with the School of Telecommunication, Xidian University. Her research interests include hyperspectral remote sensing image analysis and processing, pattern recognition, and data compression.

Keyan Wang (Member, IEEE) received the B.E. degree in communication engineering, the M.S. degree in telecommunication and information systems, and the Ph.D. degree in signal and information processing from Xidian University, Xian, China, in 2002, 2005, and 2008, respectively.

In 2004, she joined the School of Telecommunications Engineering, Xidian University, where she is currently an Associate Professor. From 2014 to 2015, she was with McMaster University, Hamilton, ON, Canada, as a Visiting Scholar. Her research interests include image and video transmission and processing, visible and multispectral remote sensing image compression coding and processing.

Yunsong Li received the M.S. degree in telecommunication and information systems and the Ph.D. degree in signal and information processing from Xidian University, Xian, China, in 1999 and 2002, respectively.

He joined the School of Telecommunications Engineering, Xidian University, in 1999, where he was a Professor. He is the Director of Image Coding and Processing Center, State Key Laboratory of Integrated Service Networks. His research interests focus on image and video processing, hyperspectral image processing, and high-performance computing.