A Hybrid Capsule Network for Land Cover Classification Using Multispectral LiDAR Data

Yongtao Yu📧, *Member, IEEE*, Haiyan Guan📧, *Senior Member, IEEE*, Dilong Li📧, Tiannan Gu, Lanfang Wang, Lingfei Ma📧, and Jonathan Li📧, *Senior Member, IEEE*

Abstract—Land cover mapping is an effective way to quantify land resources and monitor their changes. It plays an important role in a wide range of applications. This letter proposes a hybrid capsule network for land cover classification using multispectral light detection and ranging (LiDAR) data. First, the multispectral LiDAR data were rasterized into a set of feature images to exploit the geometrical and spectral properties of different types of land covers. Then, a hybrid capsule network composed of an encoder network and a decoder network is trained to extract both high-level local and global entity-oriented capsule features for accurate land cover classification. Quantitative classification evaluations on two data sets show that the overall accuracy, average accuracy, and kappa coefficient of over 97.89%, 94.54%, and 0.9713, respectively, are obtained. Comparative studies with five existing methods confirm that the proposed method performs robustly and accurately in land cover classification using the multispectral LiDAR data.

Index Terms—Capsule network, feature image, land cover classification, multispectral light detection and ranging (LiDAR), point cloud.

I. INTRODUCTION

WITH the rapid development of society, the processes of urbanization and industrialization are getting accelerated globally. Although these progresses can bring social and economic benefits, they also affect the stability and sustainability of the environment and accelerate the changes in land covers. Timely and precisely updating land cover information is significant for environment protection, land management, landscape pattern analysis, and so on. The collection of land cover information is primarily based on field surveys and analysis of remote sensing data. The field surveys can provide accurate land cover details; however, they are labor-intensive, time-consuming, and even inoperable in some

Y. Yu, T. Gu, and L. Wang are with the Faculty of Computer and Software Engineering, Huaiyin Institute of Technology, Huaian 223003, China (e-mail: allennessy.yu@gmail.com).

H. Guan is with the School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: guanhy.nj@nuist.edu.cn).

D. Li is with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430072, China (e-mail: scholar.dll@gmail.com).

L. Ma and J. Li are with the Department of Geography and Environmental Management, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: l53ma@uwaterloo.ca; junli@uwaterloo.ca).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2019.2940505

conditions. Comparatively, the acquisition of high-resolution remote sensing data covering large areas of the earth surface can be easily and efficiently carried out using remote sensing sensors, such as satellite sensors, unmanned aerial vehicle (UAV) sensors, and light detection and ranging (LiDAR) sensors. Thus, remote sensing techniques have been widely used for land cover mapping.

Land cover mapping using remote sensing images is a classic topic and has been intensively studied. The remote sensing images can provide detailed texture information, which favors to distinguish among different land cover types. In [1], a multilevel deep learning architecture composed of a set of convolutional neural networks (CNNs) was designed for land cover classification using multitemporal satellite images. Likewise, the CNN was also used in [2] for land cover classification using 2-D spectral curve graphs. Based on the UniBagging model, time-series Landsat images were exploited in [3] for monitoring land cover changes. In [4], high-resolution synthetic aperture radar (SAR) images were considered for land cover classification. Due to the large number of bands that reflect diverse spectral properties of land covers, hyperspectral images have become a promising alternative for land cover classification. For instance, capsule networks [5], sparse representation [6], generative adversarial networks [7], graph convolutional networks [8], CNNs [9], and so on were proposed for land cover classification using hyperspectral images. Moreover, land cover classification by fusing multisource images has also been exploited to improve classification accuracy [10], [11].

Recent development in LiDAR technologies has burst a great number of applications. The LiDAR sensors can collect both 3-D geospatial information and reflectance properties of the earth surface. Consequently, land cover mapping using the LiDAR data has also been carried out. In [12], a geographic object-based approach based on the elevation and intensity features of the LiDAR points was proposed for land cover classification. Extinction profiles and composite kernel support vector machines (SVMs) were suggested in [13] using LiDAR-derived features and digital surface models. In [14], the morphological profiles and CNN were proposed for LiDAR data classification. A spatial transformation network combined with the morphological profiles was designed in [15]. In [16], the LiDAR points were classified into diverse land cover types by integrating ant colony optimization and object-based analysis. To extract heterogeneous features, a multi-kernel sparse representation classification pipeline was proposed in [17]. In addition, land cover mapping by fusing the LiDAR data and remote sensing

images has also been studied. For instance, discriminative graph-based fusion [18], discriminant correlation analysis [19], multi-kernel learning [20], and so on were proposed by fusing features from hyperspectral images and LiDAR data. Moreover, CNN [21], Bayesian network [22], random forest (RF) [23], and so on were exploited for land cover classification by fusing the LiDAR data and optical images.

In recent years, multispectral LiDAR systems, which can collect the LiDAR data with multiple channels simultaneously, have been developed. Multispectral LiDAR systems provide superior advantages than traditional single-channel LiDAR systems in land cover mapping. In [24], the object-based RF analysis was adopted for land cover classification based on multispectral LiDAR data. In [25], a hierarchical rule-based classifier was selected for multispectral LiDAR classification. The rules were derived based on the knowledge of the geometrical and neighborhood properties of the LiDAR points. Image-based classification using a maximum likelihood (ML) classifier was investigated in [26] for multispectral LiDAR classification. Using point features extracted from the elevation, intensity, and geometry properties, seeded region growing was developed in [27] for land/water discrimination. In [28], a comparative feature study of deep Boltzmann machine (DBM), principal component analysis, and RF was conducted for land cover classification using multispectral LiDAR data.

In this letter, we develop a hybrid capsule network for land cover classification using multispectral LiDAR data. The proposed method is performed on five types of feature images rasterized taking into account the geometrical and spectral properties of multispectral LiDAR data. Specifically, during network architecture design, we found that the network performance can be enhanced by connecting two conventional neural layers before the capsule layers for low-level feature extraction, as well as using large kernels in the convolutional layers and small kernels in the capsule layers. The main contributions include the following: 1) a novel hybrid capsule network for land cover classification; 2) two sibling subnetworks for extracting local and global features, respectively; and 3) a decoder network with a new loss function for avoiding overfitting.

II. STUDY AREA AND MULTISPECTRAL LiDAR DATA

In this letter, two multispectral LiDAR data sets were used to perform land cover classification. The first data set (WS) was collected in Whitchurch–Stoveville, Ontario, Canada. This data set contained 19 flying strips and covered an area of about 3.21 km^2. The second data set (TB) was collected in Tobermory, Ontario, Canada. This data set contained ten flying strips and covered an area of about 1.99 km^2. These two data sets were collected using the Teledyne Optech's Titan multispectral LiDAR system. The Titan system has three spectral channels that work in intermediate infrared (Channel 1), near infrared (Channel 2), and visible (Channel 3) wavelengths, respectively, and can collect the LiDAR data simultaneously. The wavelengths of Channels 1, 2, and 3 are 1550, 1064, and 532 nm, respectively. Different channels can obtain different reflectance properties of land covers. For instance, the soil tends to reflect most in Channel 1 and lowest in Channel 3. Thus, the multispectral LiDAR system can collect both the
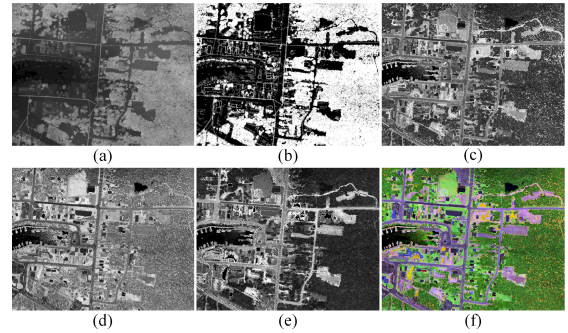

Fig. 1. Rasterized feature images with regard to (a) elevation, (b) number of returns, (c)–(e) three channels of spectral intensities, and (f) pseudo-color image by merging the three spectral intensity images.

geometrical features and multiple spectral features of the earth surface, which provides a promising data source for land cover classification.

III. METHOD

A. Feature Image Generation

To take full advantage of the geometrical and spectral properties of multispectral LiDAR data, we rasterize the 3-D LiDAR points from the three channels into a set of feature images. To this end, first, we merge the three sets of the LiDAR points from the three channels into a single point cloud according to their geographical coordinates. Then, we vertically partition the merged point cloud into a grid structure with a grid size (spatial resolution) of r_g (e.g., 0.5 m in this letter). Finally, the LiDAR points in each grid are interpolated into a single pixel in the resultant feature image. The gray value of a pixel in a feature image is interpolated based on the properties of the LiDAR points in the associated grid using the inverse distance weighted interpolation method [29]. In this letter, we generate five types of feature images by taking into account the elevation, the number of returns, and three channels of spectral intensities. Specifically, for the three spectral intensity images, only the LiDAR points from the associated channel are used for interpolation. For the elevation and number of returns images, all the merged LiDAR points are used for interpolation. Fig. 1(a)–(e) illustrates the five types of rasterized feature images. Fig. 1(f) shows the pseudo-color image by merging the three spectral intensity images.

B. Hybrid Capsule Network for Land Cover Classification

To perform land cover classification from the rasterized feature images, we construct a hybrid capsule network. Unlike traditional CNNs that use scalar neurons to encode the probabilities of the existence of specific features, the capsule networks consist of vectorial capsules to encode entity features [30]. The instantiation parameters of a capsule represent a specific type of entity; the length of the capsule encodes the probability of the existence of that entity. The capsule networks have been proven to be more powerful and reliable in extracting the intrinsic features of the objects. Thus, to improve land cover classification accuracy, we construct a hybrid capsule network using the rasterized feature images.

At the training stage, we merge and convert the five types of feature images into a single multispectral image, where each pixel is composed of five spectral channels coming from the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

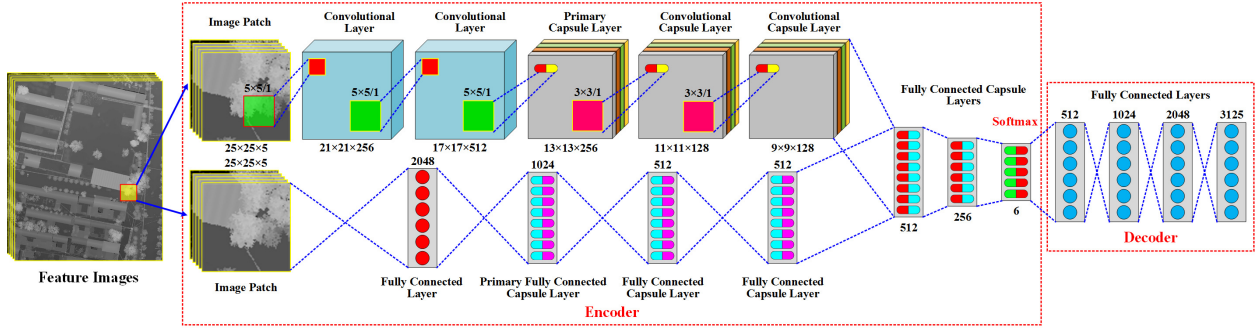YU *et al.*: HYBRID CAPSULE NETWORK FOR LAND COVER CLASSIFICATION

3

Fig. 2. Architecture of the hybrid capsule network. A capsule can be viewed as a set of neurons. All the capsules have the same dimension of 16.

associated gray values of the five feature images, respectively. Then, a set of training image patches with a size of $n \times n$ pixels are generated centered at the labeled pixels from the multispectral image. The class label of an image patch is assigned as the class of the central pixel. Finally, these labeled image patches are used to train the hybrid capsule network.

Fig. 2 presents the architecture of the hybrid capsule network, which contains an encoder network and a decoder network. The encoder network has two sibling subnetworks (a convolutional capsule subnetwork and a fully connected capsule subnetwork) for, respectively, extracting local and global features from the image patches. The classification is conducted by fusing the local and global features from the two sibling subnetworks. As shown in Fig. 2, the convolutional capsule subnetwork contains two conventional convolutional layers, a primary capsule layer, and two convolutional capsule layers. The first two convolutional layers function to extract low-level local features from the input image through convolution operations. These features are further encoded into high-order capsules to depict different levels of entities. For these two layers, the widely used rectified linear unit (ReLU) is adopted as the activation function.

The primary capsule layer converts the low-level scalar feature representations into high-order vectorial capsule representations. It can be constructed based on conventional convolution operations. Denote D_p as the number of feature maps and S_p as the capsule dimension in the primary capsule layer. Then, $D_p \times S_p$ kernels are performed on the second convolutional layer. After convolutions, the generated feature maps are grouped into D_p groups, each of which contains S_p feature maps. For each group, the S_p components at each position of the feature maps form an S_p-dimensional capsule.

The two convolutional capsule layers aim at extracting high-order capsule features through capsule convolution operations. For the capsules in these layers, the total input to a capsule j is a weighted sum over all predictions from the capsules in the convolution kernel in the lower layer

$$C_j = \sum_i a_{ij} \cdot U_{j|i} \tag{1}$$

where C_j is the total input to capsule j; a_{ij} is the coupling coefficient indicating the degree of contribution that capsule i in the layer below activates capsule j; $U_{j|i}$ is the prediction from capsule i to capsule j and it is computed as follows:

$$U_{j|i} = \mathbf{W}_{ij} \cdot U_i \tag{2}$$

where U_i is the output of capsule i and $\mathbf{W}_{ij}$ is the network weight matrix. Specifically, the coupling coefficients between capsule i and all its connected capsules in the layer above sum to 1 and are determined by a dynamic routing process [30]. Recall that we use the length of a capsule to predict the probability of the existence of an entity. Thus, we use the nonlinear "squashing" function [30] as the activation function to ensure that capsules with short vectors result in low probability predictions and capsules with long vectors result in high probability predictions. The squashing function is defined as follows:

$$U_j = \frac{\|C_j\|^2}{1 + \|C_j\|^2} \cdot \frac{C_j}{\|C_j\|}. \tag{3}$$

By such a conversion, the short capsules are shrunk toward zero length and the long capsules approach a length close to 1.

As shown in Fig. 2, the fully connected capsule subnetwork consists of a conventional fully connected layer, a primary fully connected capsule layer, and two fully connected capsule layers. The fully connected layer functions to extract low-level global features from the input image. These features are further encoded into high-order capsules through the primary fully connected capsule layer. Similarly, the primary fully connected capsule layer can be constructed based on conventional fully connected operations. Then, the resultant neurons are equally partitioned into groups to form a set of capsules. The two fully connected capsule layers aim to extract high-order capsule features with a global perspective. Likewise, dynamic routing among capsules is used to cast predictions and activate capsules, and the squashing function is used to normalize capsule outputs.

The local and global capsule features from the two sibling subnetworks are fused by flattening and concatenation and further fed into the last three fully connected capsule layers to conduct classification. The last layer is a softmax layer composed of a set of class-oriented capsules for encoding different types of land covers.

The decoder network consists of four conventional fully connected layers that take as input the capsule outputs in the softmax layer of the encoder network to reconstruct the input image, thereby encouraging the capsules to encode the most relevant and intrinsic instantiation parameters of the input image. To this end, we mask out all but the capsule whose class label corresponds to the input image patch. Then, the instantiation parameters of this capsule are fed into the decoder network for reconstruction.

The parameters in the hybrid capsule network are iteratively refined through the error backpropagation process. To effectively train the hybrid capsule network, we use the following multi-task loss function:

$$L = \sum_{i=1}^{N} \sum_{k=1}^{V} L_{\text{cls}}^{k} + \eta \sum_{i=1}^{N} L_{\text{rec}}^{i} \quad (4)$$

where L_{cls} and L_{rec} are the classification and reconstruction losses, respectively; N and V are the numbers of training image patches and class-oriented capsules in the softmax layer, respectively; η is a regularization factor to balance the two terms. For class k, the classification loss is defined as follows:

$$L_{\text{cls}}^{k} = T_k \cdot \max(0, m^+ - \|U_k\|)^2$$
$$+ \lambda(1 - T_k) \cdot \max(0, \|U_k\| - m^-)^2 \quad (5)$$

where $T_k = 1$ if a training patch belongs to class k; otherwise, $T_k = 0$. m^+ and m^- are, respectively, the lower bound for the probability of a training patch being an instance of class k and the upper bound for the probability of a training patch not belonging to class k. They are set as $m^+ = 0.92$ and $m^- = 0.08$. λ is a regularization factor. The reconstruction loss is defined as the robust smooth-L_1 loss [31] between the outputs of the neurons in the decoder network and the input image patch.

At the classification stage, we first rasterize the multispectral LiDAR data into five types of feature images, which are further fused and converted into a multispectral image. Then, an image patch with a size of $n \times n$ pixels is generated centered at each pixel to be classified. Finally, these generated patches are fed into the encoder network to mark them into different land cover types. The class label of a patch is determined as follows:

$$T^* = \arg\max_{k} \|U_k\| \quad (6)$$

where U_k is the capsule output in the softmax layer. This class label is assigned to the central pixel of the patch as its predicated land cover type.

IV. RESULTS AND DISCUSSION

A. Land Cover Classification

Through classification performance and computational efficiency analysis, the optimal network configurations are illustrated in Fig. 2. At the training stage, we randomly initialized all layers of the hybrid capsule network by drawing parameters from a zero-mean Gaussian distribution with a standard deviation of 0.01. The learning rate and the maximum epoch were set as 0.001 and 2000, respectively. For the two data sets with feature images of 4086×3143 and 2523×3154 pixels, respectively, rasterized with a spatial resolution of 0.5 m, 40% of the labeled data were randomly selected for training the hybrid capsule network. The patch size was configured as 25×25 pixels. Once the hybrid capsule network was constructed, we applied it to the remaining 60% of the labeled data in each data set to evaluate its accuracy on land cover classification. In this letter, we focused on six types of land covers: 1) water; 2) vegetation; 3) road; 4) soil; 5) building; and 6) other impervious surface. To quantitatively assess the classification results, we adopted the following three metrics: overall accuracy (OA), average accuracy (AA), and kappa
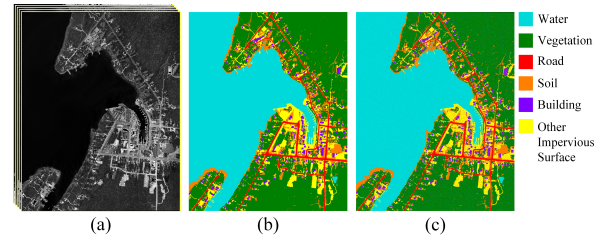

Fig. 3. (a) Rasterized feature images of the TB data set. (b) Ground-truth land cover map. (c) Land cover classification result.

TABLE I

LAND COVER CLASSIFICATION RESULTS ON THE WS DATA SET

Type	Proposed	CapsNet	RF	Rule-based	ML	DBM
1	99.11±0.12	98.34±0.13	92.77±1.56	94.91±0.89	94.76±0.82	98.10±0.13
2	94.53±0.18	93.15±0.21	86.61±1.77	88.13±0.98	87.99±1.07	92.42±0.24
3	93.72±1.17	91.71±1.19	82.53±3.01	82.80±2.41	82.92±2.62	83.63±1.31
4	92.91±0.41	91.62±0.53	84.97±2.12	86.27±1.42	85.84±1.35	88.71±0.57
5	92.23±1.95	91.45±2.18	82.38±4.08	87.71±3.44	88.63±3.37	88.95±2.43
6	94.76±1.78	93.83±2.14	81.22±3.87	91.32±3.12	90.25±3.19	93.54±2.32
OA(%)	97.89±0.15	95.79±0.21	90.64±2.04	91.57±1.13	91.23±1.12	94.36±0.27
AA(%)	94.54±0.17	93.35±0.28	85.08±2.15	88.52±1.32	88.40±1.29	90.89±0.32
$\kappa \times 100$	97.13±0.14	95.02±0.24	84.96±2.22	90.22±1.25	89.67±1.23	93.78±0.29

TABLE II

LAND COVER CLASSIFICATION RESULTS ON THE TB DATA SET

Type	Proposed	CapsNet	RF	Rule-based	ML	DBM
1	99.34±0.11	98.63±0.12	93.42±1.44	95.13±0.87	95.34±0.76	98.35±0.12
2	95.25±0.16	94.21±0.21	86.88±1.74	89.20±0.95	88.92±1.02	93.51±0.22
3	94.17±1.16	92.92±1.18	83.35±2.89	83.69±2.31	83.17±2.52	88.26±1.28
4	93.52±0.39	92.74±0.51	85.24±2.06	87.71±1.33	88.05±1.28	90.13±0.55
5	92.41±1.87	91.53±2.07	82.57±3.79	87.92±3.36	88.21±3.29	89.92±2.39
6	95.02±1.66	93.91±2.05	82.26±3.76	91.55±3.08	91.74±3.10	93.81±2.24
OA(%)	98.34±0.09	96.32±0.18	90.96±1.92	91.89±1.12	92.15±1.08	95.13±0.24
AA(%)	94.95±0.14	93.99±0.23	85.62±2.07	89.20±1.29	89.24±1.22	92.33±0.27
$\kappa \times 100$	97.76±0.11	95.84±0.22	85.31±2.14	91.03±1.21	91.27±1.06	94.24±0.25

coefficient (κ). For the two data sets, we conducted ten Monte Carlo runs and reported the mean and standard deviation of these metrics. Tables I and II detail the land cover classification results on the two data sets. For visual inspections, Fig. 3 illustrates the rasterized feature images, the ground-truth land cover map, and the land cover classification result of the TB data set.

As reflected in Tables I and II, the proposed method achieved the OA, AA, and κ values of 97.89% ± 0.15%, 94.54% ± 0.17%, and 0.9713 ± 0.0014 on the WS data set and 98.34% ± 0.09%, 94.95% ± 0.14%, and 0.9776 ± 0.0011 on the TB data set, respectively. Specifically, for the two data sets, the best classification performance was obtained for water because of its homogeneous and distinctive properties, whereas a relatively lower accuracy was obtained for building because of the diverse geometrical and spectral properties. In addition, the misclassification generally appeared at the boundaries of two different land cover types. For instance, some pixels belonging to soil were falsely labeled as other impervious surface.

B. Comparative Study

To further evaluate the performance of the proposed hybrid capsule network, we conducted a set of comparative tests with the following five methods: CapsNet [5], RFs [24], rule-based classifier [25], ML classifier [26], and DBM [28]. We applied these methods to the two multispectral LiDAR data

sets to evaluate their land cover classification performance. The detailed classification results are shown in Tables I and II. Comparatively, the CapsNet and DBM outperformed the other three methods. This is because instead of extracting low-level features from the input data like in the RF, rule-based, and ML methods, the CapsNet and DBM exploit high-level, deep features for land cover classification. These features are more intrinsic and distinct. In addition, the RF method relies on a preprocessing of homogeneous region segmentation. Thus, it performed unsatisfactorily for building and other impervious surface. The superior performance obtained by the CapsNet is due to the principal component analysis strategy for salient feature selection. On the whole, benefited from the capabilities in extracting both high-level local and global features, our proposed hybrid capsule network achieved promising land cover classification results and outperformed the other methods with regard to classification accuracy. However, the proposed hybrid capsule network had no absolute advantages with regard to computational complexity compared with the CapsNet and DBM, which are less lightweight with less network layers.

V. CONCLUSION

This letter has presented a hybrid capsule network for land cover classification using multispectral LiDAR data. The multispectral LiDAR data were rasterized into a set of feature images to exploit the geometrical and spectral properties of different types of land covers. The hybrid capsule network took full advantage of vectorial capsules to extract both local and global entity-oriented features from the rasterized feature images. Land cover classification evaluations on two data sets showed that the proposed method achieved the OA, AA, and κ values of over 97.89%, 94.54%, and 0.9713, respectively. Comparative studies with five existing methods confirmed that the proposed method was more accurate and outperformed the other five methods in land cover classification using multispectral LiDAR data.

REFERENCES

[1] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778–782, May 2017.

[2] M. Kim *et al.*, "Convolutional neural network-based land cover classification using 2-D spectral reflectance curve graphs with multitemporal satellite imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 4604–4617, Dec. 2018.

[3] J. Xiao, H. Wu, C. Wang, and H. Xia, "Land cover classification using features generated from annual time-series Landsat data," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 739–743, May 2018.

[4] C. O. Dumitru, G. Schwarz, and M. Datcu, "SAR image land cover datasets for classification benchmarking of temporal changes," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 5, pp. 1571–1592, May 2018.

[5] J. Yin, S. Li, H. Zhu, and X. Luo, "Hyperspectral image classification using CapsNet with well-initialized shallow layers," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1095–1099, Jul. 2019.

[6] T. Dundar and T. Ince, "Sparse representation-based hyperspectral image classification using multiscale superpixels and guided filter," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 246–250, Feb. 2019.

[7] Y. Zhan, D. Hu, Y. Wang, and X. Yu, "Semisupervised hyperspectral image classification based on generative adversarial networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 212–216, Feb. 2018.

[8] A. Qin, Z. Shang, J. Tian, Y. Wang, T. Zhang, and Y. Y. Tang, "Spectral–spatial graph convolutional networks for semisupervised hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 241–245, Feb. 2019.

[9] N. Ouyang, T. Zhu, and L. Lin, "Convolutional neural network trained by joint loss for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 457–461, Mar. 2019.

[10] C. Sukawattanavijit, J. Chen, and H. Zhang, "GA-SVM algorithm for improving land-cover classification using SAR and optical remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 3, pp. 284–288, Mar. 2017.

[11] Z. Xu, J. Chen, J. Xia, P. Du, H. Zheng, and L. Gan, "Multisource earth observation data for land-cover classification using random forest," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 789–793, May 2018.

[12] X. Zhou and W. Li, "A geographic object-based approach for land classification using LiDAR elevation and intensity," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 669–673, May 2017.

[13] P. Ghamisi and B. Höfle, "LiDAR data classification using extinction profiles and a composite kernel support vector machine," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 659–663, May 2017.

[14] A. Wang, X. He, P. Ghamisi, and Y. Chen, "LiDAR data classification using morphological profiles and convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 774–778, May 2018.

[15] X. He, A. Wang, P. Ghamisi, G. Li, and Y. Chen, "LiDAR data classification using spatial transformation and CNN," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 125–129, Jan. 2019.

[16] M. I. Sameen, B. Pradhan, H. Z. M. Shafri, M. R. Mezaal, and H. B. Hamid, "Integration of ant colony optimization and object-based analysis for LiDAR data classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 2055–2066, May 2017.

[17] Y. Gu, Q. Wang, and B. Xie, "Multiple kernel sparse representation for airborne LiDAR data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 1085–1105, Feb. 2017.

[18] 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.

[19] F. Jahan, J. Zhou, M. Awrangjeb, and Y. Gao, "Fusion of hyperspectral and LiDAR data using discriminant correlation analysis for land cover classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 10, pp. 3905–3917, Oct. 2018.

[20] Y. Gu, Q. Wang, X. Jia, and J. A. Benediktsson, "A novel MKL model of integrating LiDAR data and MSI for urban area classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5312–5326, Oct. 2015.

[21] Z. Xu, K. Guan, N. Casler, B. Peng, and S. Wang, "A 3D convolutional neural network method for land cover classification using LiDAR and multi-temporal Landsat imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 144, pp. 423–434, Oct. 2018.

[22] Z. Kang, J. Yang, and R. Zhong, "A Bayesian-network-based classification method integrating airborne LiDAR data with optical images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 4, pp. 1651–1661, Apr. 2017.

[23] Q. Wu, R. Zhong, W. Zhao, K. Song, and L. Du, "Land-cover classification using GF-2 images and airborne LiDAR data based on random forest," *Int. J. Remote Sens.*, vol. 40, nos. 5–6, pp. 2410–2426, Mar. 2019.

[24] L. Matikainen, K. Karila, J. Hyyppä, P. Litkey, E. Puttonen, and E. Ahokas, "Object-based analysis of multispectral airborne laser scanner data for land cover classification and map updating," *ISPRS J. Photogramm. Remote Sens.*, vol. 128, pp. 298–313, Jun. 2017.

[25] N. Ekhtari, C. Glennie, and J. C. Fernandez-Diaz, "Classification of airborne multispectral LiDAR point clouds for land cover mapping," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 6, pp. 2068–2078, Jun. 2018.

[26] S. Morsy, A. Shaker, and A. El-Rabbany, "Multispectral LiDAR data for land cover classification of urban areas," *Sensors*, vol. 17, no. 5, p. 9581, Apr. 2017.

[27] S. Morsy, A. Shaker, and A. El-Rabbany, "Using multispectral airborne LiDAR data for land/water discrimination: A case study at Lake Ontario, Canada," *Appl. Sci.*, vol. 8, no. 3, p. 349, Feb. 2018.

[28] S. Pan, H. Guan, Y. Yu, J. Li, and D. Peng, "A comparative land-cover classification feature study of learning algorithms: DBM, PCA, and RF using multispectral LiDAR data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1314–1326, Apr. 2019.

[29] Y. Yu, J. Li, H. Guan, C. Wang, and J. Yu, "Automated detection of road manhole and sewer well covers from mobile LiDAR point clouds," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 9, pp. 1549–1553, Sep. 2014.

[30] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 3856–3866.

[31] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Santiago, Chile, Dec. 2015, pp. 1440–1448.