An SOE-Based Learning Framework Using Multisource Big Data for Identifying Urban Functional Zones

Ying Feng, Zhou Huang, Yaoli Wang, Lin Wan, Yu Liu, Yi Zhang, and Xv Shan

***Abstract*—Identifying urban functional zones is of great significance for understanding urban structure and urban planning. The rapid growth and open accessibility of multisource big data, including remote sensing imagery and social sensing data, lead to a new way for dynamic identification of urban functional zones. In this article, we propose an SOE (scene–object–economy) based learning framework which integrates scene features from remote sensing imagery, object features from building footprints, and economy features from POIs (points of interest). From these three perspectives, rich information hidden in urban zone is excavated for function identification. Convolutional neural networks are used to extract high-level scene information from remote sensing images with different resolutions. Object features comprising a series of building indicators are constructed by measuring the area, perimeter, floor number, and year of the building. Moreover, we extract socioeconomic characteristics from POIs, which reflect different types of human activities in the urban zone. Last, random forest is used to identify functional zones based on SOE features. We apply the SOE-based framework to Shenzhen datasets and achieve 90.8% in accuracy with remote sensing images of 0.3-m spatial resolution. The experimental results show that the predicting performance of SOE-based framework is significantly better than other traditional methods, and the quantitative contribution of SOE factors is also revealed in determining functionality of urban zones.**

***Index Terms*—Convolutional neural network (CNN), multisource big data, remote sensing, scene-object-economy (SOE), urban functional zone.**

Manuscript received March 3, 2021; revised May 19, 2021; accepted June 15, 2021. Date of publication June 23, 2021; date of current version August 2, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFE0196100; and in part by the National Natural Science Foundation of China under Grant 41771425, Grant 41830645, and Grant 41625003. *(Corresponding author: Zhou Huang.)*

Ying Feng, Yaoli Wang, Yu Liu, and Yi Zhang are with the Institute of Remote Sensing and Geographical Information Systems, School of Earth and Space Sciences, Peking University, Beijing 100871, China (e-mail: 1901210143@pku.edn.cn; wang.yaoli@yahoo.com; liuyu@urban.pku.edu.cn; zy@pku.edu.cn).

Zhou Huang is with the Institute of Remote Sensing and Geographical Information Systems, School of Earth and Space Sciences, Peking University, Beijing 100871, China, and also with the State Key Laboratory of Media Convergence Production Technology and Systems, Beijing 100871, China (e-mail: huangzhou@pku.edu.cn).

Lin Wan is with the School of Geography and Information Engineering, China University of Geosciences, Wuhan 430078, China (e-mail: wanlin@cug.edu.cn).

Xv Shan is with the State Key Laboratory of Media Convergence Production Technology and Systems, Beijing 100871, China (e-mail: sklmcpts@xinhua.org).

Digital Object Identifier 10.1109/JSTARS.2021.3091848

I. Introduction

With the accelerating process of urbanization and rapid expansion of the scale of cities, a series of urban problems have appeared one after another, such as traffic congestion and housing shortage. Especially for some provincial capitals or metropolises, urban problems are particularly serious. Emergence of these problems has a strong connection with the improper regional planning within city. In detail, urban functional zones are areas where a series of economic activities occur, such as commerce, public, industry, and residence.

Identifying urban functional zones and understanding their spatial distribution and interaction laws are of great significance for supporting scientific planning. Traditionally, urban functional zone identification relies on planning maps of land use and questionnaire surveys. However, the survey-based method is expensive. In addition to the manpower and time costs, the reliability is also severely affected by subjective factors such as time, location, and personal experience of the investigator. Analyzing function regions with pictures or text asks for a huge amount of data from various sources which is limited by difficulties in the process of data acquisition. Moreover, the information obtained from a single data source is limited and not comprehensive. With development of remote sensing technology and improvement of multisource data accessibility in big data era, many researchers have used high-resolution remote sensing imagery or social sensing data such as points of interest (POIs), check-ins, and GPS trajectories to identify urban functional zones and achieve good results [1]–[5]. Therefore, how to integrate multisource big data to efficiently identify urban functional zones has become one of the research hotspots.

High-resolution remote sensing images are rich in physical representation information of landform and surface objects [6]. In recent years, in recognition of land use and land cover, the extraction method based on high-resolution remote sensing images have been widely used [7]–[11]. The analysis of urban land use is often performed in three types of spatial units, i.e., pixel, object, and scene. Pixel-based and object-based units are usually used to evaluate land cover and land use [12]–[15]. Scene-based units are usually used to identify urban functional zones and recognize urban land use patterns [16], [17], which usually utilize road networks to segment land for basic analysis units. Many studies have applied scene-based classification methods such as Latent Dirichlet Allocation (LDA) to extract physical features (such

as spectrum, texture, and SIFT features of ground components) from high-resolution remote sensing images [18]–[20]. However, it mines low-level semantic information of a parcel, ignoring connection of spatial distribution between ground objects. In high-density cities, many parcels might have the same physical properties but different functional properties. Some traditional classification methods are also difficult to combine classes of land use with actual functional zones of the city. In fact, the function of urban land is not only related to physical properties of parcels, but also affected by human activities.

Fusion of remote sensing images with social sensing data is a new alternative direction. Recent studies have shown that it performs well in identifying urban functional zones and solving high-density urban land use problems. With emergence of social sensing data, such as POIs, taxi trajectories, mobile phone data, social media data, and street view data, many new methods have been developed to understand urban systems [21]–[25]. For example, by combing remote sensing images with mobile phone positioning data, landscape and activities indicators were calculated to clustering for functional zones [26]. LSTM is used to extract characteristics of user's time series signature data and integrated satellite images for functional zones recognition [27]. Nighttime remote sensing images, satellite images, POIs, and mobile phone data are combined to draw a nationwide land use map [28]. Characteristics of human economic activities are extracted from Tencent data, POIs. And texture, spectrum, and scale invariant features are captured from remote sensing images. They are minded for functional zones classification. These widely used urban data can reflect changes in spatial patterns in time, human mobility, materials, and information. The above researches show that each type of geospatial data has its own advantages. Fusing multi-source big data is a promising solution for identifying urban land use. Nonetheless, the main challenge is that existing researches lack depth and comprehensive mining of scene, object, and socioeconomic features. And there is no framework to make a quantitative analysis in the contribution of features. In particular, fusion of remote sensing imagery and socioeconomic data is common, but joint use of three-dimensional (3-D) building data is ignored in feature mining. In fact, understanding similarities and differences between distribution patterns of 3-D buildings in urban zones is also of great significance to functional differentiation [29].

To fill research gap, we propose a scene–object–economy (SOE)-based learning framework for recognizing urban functional zones, which identifies functions in three dimensions, i.e., scenes from remote sensing imagery, objects from building footprints, and socioeconomic characteristics from POIs. There are three main parts of our contribution. First, SOE-based framework not only detects hidden scene information from the imagery, but also discloses connections and differences between buildings based on objects in a zone. Our framework introduces height of buildings in feature space, which is hard to obtain from imagery analysis. Second, a novel ensemble learning model is introduced to identify urban functional zones, combining DCNN (deep convolutional neural network) with traditional classifier models. Results show that prediction performance is greatly improved if features extracted by SOE model are used. Third, we further quantitatively compare weights of 3-D SOE features in different function scenarios. In addition, our method has been trained and tested in Beijing, achieving excellent performance. Training based on Shenzhen, the model generalizes well in Guangzhou.

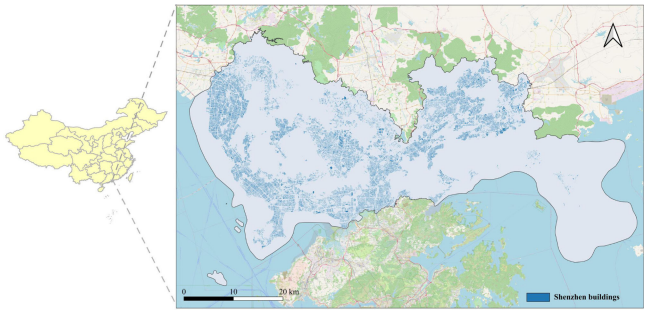

Fig. 1. Study area (Shenzhen city).

II. Study Area and Datasets

Our research area is Shenzhen, China, locating in southern Guangdong and east bank of the Pearl River Estuary (Fig. 1). By the end of 2019, the city has nine districts with a total area of 1997.47 square kilometers. It has a permanent population of 13.34 million, as the first total urbanized city in China. For sustainable development of Shenzhen, reasonable planning of urban functional zones should be taken into consideration. Four types of functional zones are involved in the study: commerce, public, industry, and residence. Among them, the commerce includes buildings such as companies, shopping centers, office buildings, and banks. The public includes public facilities, scientific research and education, public services, and other buildings. The industry includes buildings such as industrial parks and warehousing. The residence includes buildings where people live.

Road networks, high-resolution remote sensing images, POIs, and building footprints are involved in study, as shown in Fig. 2.

Previous studies usually use road networks to divide cities into several polygonal parcels as basic units [30]–[32]. Those parcels are considered spatially independent, heterogeneous, and highly pure. Parcels in EULUC-China map are segmented using OSM road network [28], and we use them as basic units to identify functional zones in Fig. 2(a). An overlay analysis is used with parcels and Shenzhen's land use map in 2019 provided by Shenzhen Municipal Bureau of Planning and Natural Resources. And a semi-supervised method is used to select parcels with a purity higher than 0.6. It excludes examples of mixed functional regions and obtains representative lands. In the end, a total of 1150 parcels are sampled: 165 for the commerce, 176 for the public, 128 for the industry, and 681 for the residence.

We obtain high-resolution remote sensing imagery from Google Earth for the year 2019 which have three bands and two different spatial resolutions: 1.1 and 0.3 m.

Building footprints are obtained from Baidu Inc., 2019, containing 635 177 buildings in Shenzhen. Area, length, floor, and year of each building are measured to construct relevant features,

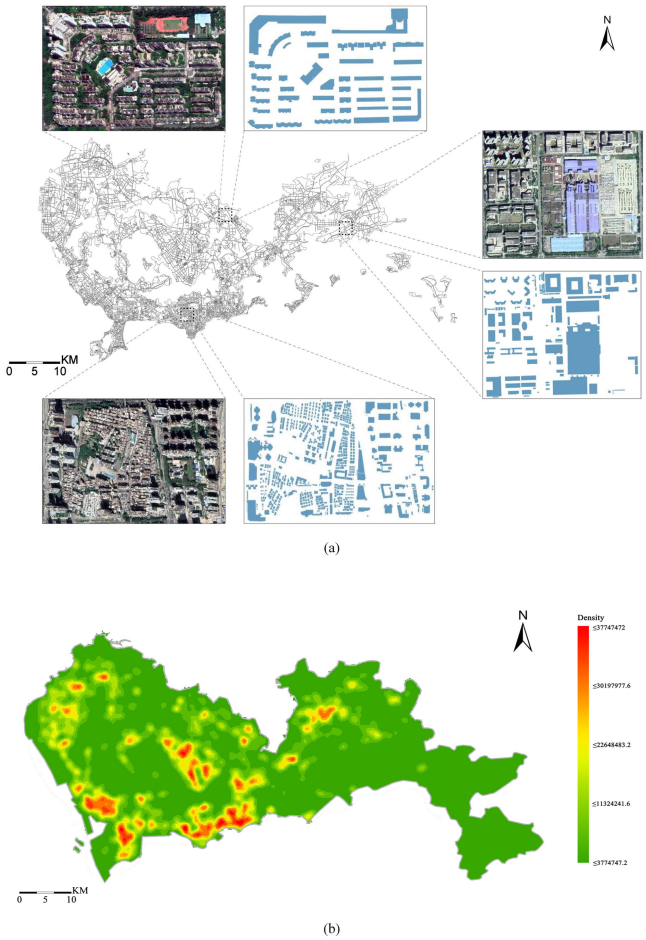

Fig. 2. (a) Research units segmented by OpenStreetMap network, remote sensing images, and building footprints in some sample units. (b) Kernel density map generated by POIs in Shenzhen.

which show physical properties and similarity of buildings. And they also reflect differences between four types of zones.

We collect POIs of Shenzhen from Baidu Map API, and obtain 1 044 411 in total, as shown in Fig. 2(b). POIs have been proven to play an important role in functional zone recognition [33]–[35]. It is not generated by physical information on land surface, but attribute tags and geographic points by human economic activities. To a certain extent, it reflects types of activities people perform at specific place.

III. Methodology

A. Machine Learning Framework Based on SOE

Two types of data sources are involved in this article: image data (remote sensing imagery) and nonimage data (building footprints and POIs). Traditional models perform terribly in dealing with this multimodal data fusing problem. For example, convolutional neural networks (CNNs) have achieved excellent results in image recognition tasks since they are good at extracting image features. However, they perform purely in extracting features from nonimage data. In this article, suitable methods are used to separately process features. Classification is performed

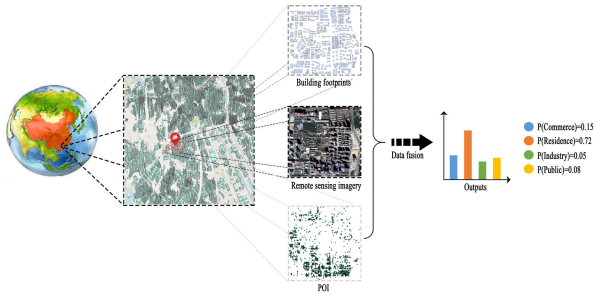

Fig. 3. SOE-based framework, fusing remote sensing imagery, building footprints, and POIs in classification of functional zones (commerce, residence, industry, and public).

at the fusion stage. In this article, we propose a new framework to extract features from different modal of data with different extractors and fuse those features at classification stage. CNNs are used to extract remote sensing image features. Statistical methods are employed to measure building footprints and POIs features. Then, they are combined for the final classification task.

We propose the concept of SOE and integrate multisource data to fully extract spatially implied semantic information. The SOE-based learning framework is shown in Fig. 3. Based on object, physical properties of buildings are measured. Area, perimeter, floor, year, and structure ratio are measured for differences between buildings in a zone. It includes totally 17 characteristics, e.g., sum, average, standard deviation of five attributes, as well as the density and number of buildings. Socioeconomic characteristics are extracted from POIs, composed of 14 different indicators. Hidden spatial information in remote sensing images is detected based on scene through CNNs. The parameter of the penultimate fully connected layer is changed to 64 for achieving balance between the number of features in other two data sources. After fusing them, we obtain 95-length feature vector. Six common classifiers including Random Forest (RF), XGBoost, Naive Bayes, SVM, LSTM, and dense layer are used to predict functional zones. The methodology flowchart is illustrated in Fig. 4.

In addition, we also analyze if there exists certain relationship between these three data sources. Analysis weights of different features for classification task and output them, making clear which data source has played a dominant role in classification.

B. Extracting Image Features

In our study, CNNs are utilized to extract features from high-resolution remote sensing images. In recent years, it has been confirmed that CNNs achieve excellent results in image recognition, object detection, and other tasks in the field of computer vision [36]–[39]. CNNs extract high-level hidden features, which are different from low-level semantic features such as texture, spectrum, and SIFT of scene's topography. Although we have no idea about meaning of features extracted by CNNs, these advanced features can greatly improve accuracy of classification. In our work, several CNNs based neural networks that have achieved good results in computer vision tasks are selected for comparative experiments, such as VGG16,

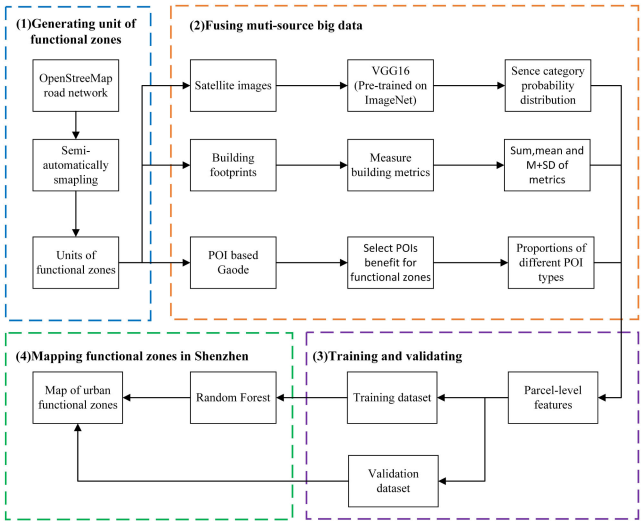

Fig. 4. Workflow of SOE-based methodology.

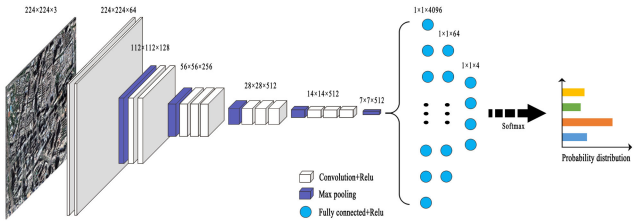

Fig. 5. Network structure of VGG16, the fully connected layer's parameters are modified to 64 and 4.

AlexNet, GoogLeNet, and ResNet [40]–[43]. These four CNNs have their own characteristics. AlexNet, winner of the 2012 ImageNet competition, is a relatively shallow network. In the 2014 ImageNet Challenge, GoogLeNet won the first place and VGG16 won the second place, both characterized by a deeper level. However, GoogLeNet has fewer parameters, one-twelfth of AlexNet. Unlike the three, ResNet has a deeper network layer with residual blocks, reaching hundreds of layers.

VGG16 uses a uniform size of 3*3 convolution kernel, simplifying the neural network structure. It achieves a deeper level and extracts more advanced features. In the experiment, in order achieve a balance between the remote sensing imagery and other data sources, the parameter of 4096 in the second fully connected layer fc2 is changed to 64. The parameter of 1000 categories in the third fully connected layer fc3 is also changed to 4 to match our classification task. Then feature vector in fc2 is output after training. The structure of network is shown in Fig. 5. ResNet50 contains 18 residual blocks. One residual block contains two 1*1 and a 3*3 convolution kernels. 1*1 convolution kernel is only to change the number of output channels. 3*3 convolution kernel may change the size of tensors, which is related to the parameter stride. Last, the last fully connected layer parameter is modified from 1000 to 4. In AlexNet and GoogleNet, same modifications are made in layers.

In this stage, we use images as input of CNNs and softmax conduct classification directly. Then, probability distribution of four categories is obtained. With RF as final classifier, we also compare performance of image features extracted by different neural networks.

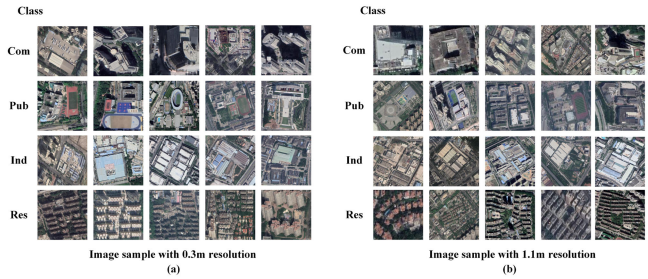

Fig. 6. Remote sensing images with different resolutions: (a) 0.3 m resolution, and (b) 1.1 m resolution.

C. Training and Verification of CNN

In this article, remote sensing images with two resolutions are used; one with a spatial resolution of 0.3 m and the other is 1.1 m. The imagery data are visualized in Fig. 6. Spatial coverage areas, coordinate system, and other parameters of two data are completely consistent. We use a 4:1 ratio to divide image data randomly into training set and test set. For images in training set, a series of data augmentation methods are adopted to avoid overfitting in training of models. First, center crop is carried out for the size of 224*224, fitting input of network. Then, images are flipped horizontally and vertically with a probability of 0.5, which can expand richness of samples and avoid difference caused by orientation. Next, brightness, contrast, saturation, and hue are adjusted to meet different needs for scene under different lighting. These four operations are all performed with a probability of 0.5, where jitter value is randomly set from the maximum to the minimum value of attributes in image pixel array.

To train a network with strong generalization ability, a large amount of data is required. In this article, for two kinds of remote sensing images with different resolutions, training set has 918 samples and test set has 232 samples. In order to solve problem that the amount of data is too small to fully train entire network, transfer learning is adopted. Our dataset is used to fine-tune the pretrained weights on ImageNet [44]. Model parameters need to be updated iteratively for minimizing loss between predictions and ground truth during feature extraction.

During training, we utilize three different ways: 1) training all parameters of the neural network, 2) training parameters of high-level neural layers, and 3) only training parameters of the fully connected layer. Through experiments, the most suitable way for training can be found out. For several CNNs, the parameter of final fully connected layer is modified to 4.

D. Extraction of Building Features

Buildings and land parcels relate closely to each other. Usually, differences between physical properties of buildings can reflect functional attributes of a zone.

Area represents actual floor area of a building. We calculate total area, average area, and standard deviation of area of all buildings within zone. For commercial parcels, buildings vary greatly in size, such as shopping malls occupy a large area but office buildings occupy small area. However, the residential parcel is often highly uniform, areas of buildings inside are similar. So the standard deviation is relatively small. There are differences in fluctuation range of the average area in different functional zones.

Length of the building outline is represented by building perimeter. This attribute measures to observe difference in length between buildings in a zone. The sum of perimeters, average perimeter, and standard deviation of perimeters of buildings are calculated in parcels. These characteristics can better distinguish the residence from zones in other types. The edge lengths of residential buildings have a high degree of similarity and conform to regular distribution.

Building structure ratio refers to the ratio of perimeter to building area. Buildings with a larger perimeter of the same area usually have more complex shape. The complexity of shape facilitates distinction between regular rectangular buildings and irregularly shaped buildings. For example, the commercial and public usually have buildings with round and irregular polygonal appearances, such as stadiums, shopping centers, etc. This is helpful to distinguish residential buildings. We measure structural ratio of each building and calculate its sum, average, and standard deviation.

Floor represents total number of floors of a building. This indicator is introduced because buildings have different heights to meet the needs of functional zones. Office buildings usually locate in the center of zones and have up to 30 floors which distinguish greatly from ordinary residential buildings. Buildings on the industry are different from other three in that their heights are lower. Here, considering that buildings have different uses and heights, the total floor, average floor, and standard deviation are calculated.

Year refers to the time when a building was built. This attribute can reflect newness of buildings in zones. For the industry, the new and remote vacant land is constantly being built for factories or warehousing logistics, and some of its buildings are newer and uneven. The same problem exists in the commerce, and there are inconsistencies in building years. However, the residence is completely opposite, buildings of which varies little in built years, so it is easier to distinguish them from other three. The degree of heterogeneity in zones is measured with three indicators: the sum of the construction years, average of the years, and standard deviation of years in buildings.

The density and total number of buildings in parcels are counted. Density can reflect sparseness of objects in a zone, revealing distribution of buildings in land parcels.

Seventeen indicators are measured in a parcel. As shown in Fig. 7, regulars are explored in three aspects, such as spatial distribution of individuals, connection between individuals, and relationship between individuals and scenes. First, spatial join method in ArcGIS is deployed to connect buildings with the intersecting parcel, obtaining all information of buildings within a parcel. Then, indicators are calculated based on statistical methods. Finally, the feature vector is generated to represent building information.

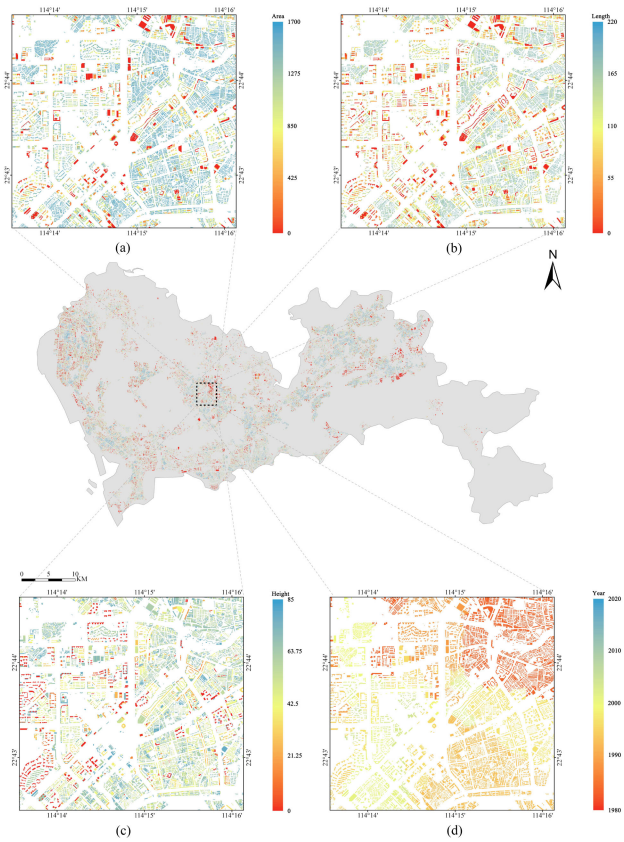

Fig. 7. Thematic maps of building footprints. (a) Area of buildings. (b) Perimeter of buildings. (c) Height of buildings. (d) Year of buildings.

E. Extraction of Socioeconomic Features

Types of functional zones are not only related to geography, but also closely connected to a series of economic activities. POIs are used to extract socioeconomic characteristics for enriching information about land parcels. First, beneficial for classification, POIs are reclassified into 14 categories, including public facilities, catering services, education and cultural services, shopping services, companies and enterprises, medical care services, accommodation services, commercial residences, life services, landscapes, transportation facilities services, financial insurance services, sports and leisure services, government agencies, and social organizations. Next, the spatial join method in ArcGIS is implemented to connect POIs with its intersecting parcel. Third, proportions of 14 types of indicators are calculated based on statistical methods, representing socioeconomic characteristics of land parcels.

F. Classification and Verification of SOE Features at the Fusion Stage

In this article, six different classifiers are utilized for the final feature fusion and classification. Four categories of functional zones are involved, including commerce, public, industry, and

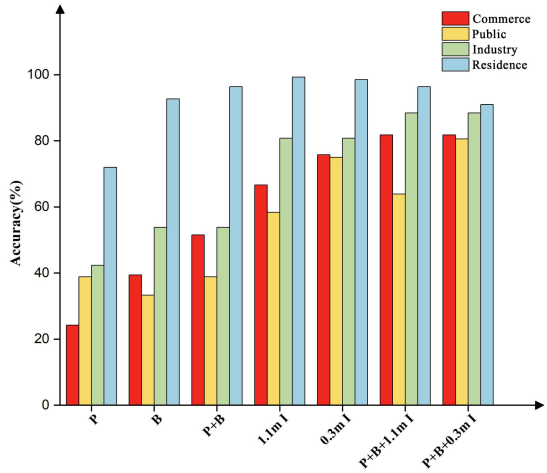

Fig. 8. Result of classification in functional zones with combining different data sources. P (POIs), B (building footprints), P+B (POIs + building footprints), 1.1 m (the remote sensing imagery with 1.1 m resolution), 0.3 m (the remote sensing imagery with 0.3 m resolution), P+B+1.1 m I (POIs + building footprints + the remote sensing imagery with 1.1 m resolution), P+B+0.3 m I (POIs + building footprints + the remote sensing imagery with 0.3 m resolution).

residence. First of all, samples are preprocessed and labeled. Based on Shenzhen's land use map in 2019, a semi-supervised method is used for labeling. Proportions of these four land use are counted in a parcel. If the proportion of a certain land use is greater than 0.6, it will be considered as a relatively pure land and labeled with this particular land use. On the contrary, if less than 0.6, it is considered as a mixed functional zone. These samples are excluded. After screening a round of samples, we perform manual selecting to check whether there is error in the truth of land. Finally, the selected samples are divided randomly into training set and test set at a ratio of 4:1.

After comparison experiments with five classifiers (i.e., XGBoost, Naive Bayes, SVM, LSTM, and dense layer), RF is chosen as the final classifier at fusion stage, which has achieved good results in a number of classification studies [45]–[47]. RF uses multiple trees to train and predict samples, finally voting for decision. A category with the most votes is a predicted functionality.

IV. Results and Discussion

A. Classification Results With Different SOE Features

We put forward SOE, combining concepts of scenes, objects, economy, and digging deep into rich features of functional zones. Remote sensing images can extract high-level semantic features of regional spatial distribution based on scenes. Some open source data can extract features of human economic activities. They all reflect the use of functional zones from a certain perspective. However, knowledge form topographic and human beings in 2-D exists limit, but deployment of buildings reflects functions in 3-D based on object. By changing the input of SOE features, different results can be obtained based on our framework, shown in Fig. 8. When using single POIs or building

TABLE I
Classification Results for Each Category and Related Metrics

Input data	Commerce	Public	Industry	Residence	Accuracy	Kappa	F1 score
POIs	24.24 %	38.89 %	42.31 %	97.81 %	71.98 %	0.476	0.531
Buildings	39.39 %	33.33 %	53.85 %	92.70 %	71.55 %	0.491	0.561
POIs+Buildings	51.51 %	38.89 %	53.85 %	96.35 %	76.29 %	0.581	0.617
1.1m Images	66.67 %	58.33 %	80.77 %	99.27 %	86.21 %	0.760	0.781
0.3m Images	75.76%	75.00%	80.77%	98.54%	89.65%	0.829	0.825
POIs+Buildings+1.1m Images	81.81%	63.89%	88.46%	96.35%	88.36%	0.805	0.818
POIs+Buildings+0.3m Images	81.81%	80.55%	88.46%	96.35%	90.95%	0.849	0.858

footprints, accuracy has reached 70%. However, there is an imbalance between categories. For only POIs, Kappa is 0.476 and F1 score is 0.531. For building footprints, Kappa is 0.491 and F1 score is 0.561. For both, accuracy of the commerce and public are relatively low. When POIs and building footprints are integrated, accuracy of the commerce is increased about 11%, and overall accuracy is increased about 4.3%; Kappa is 0.581 and F1 score is 0.617. Relatively, remote sensing imagery achieves best results in single-source classification. The accuracy of remote sensing images with 1.1 m resolution reaches 86.21%, with Kappa 0.76 and F1 score 0.781. However, accuracy of the commerce and public are 20–30% lower compared with other two types. For 0.3 m resolution images, overall accuracy has increased by 3.44%, the commerce has increased by 9.09%, and the public has increased by 11.67%. Kappa value is 0.829 and F1 score is 0.825. So we believe that when the resolution is clearer than 1.1 m, there are limitations in optimization of overall classification, but problem of category imbalance can be resolved to a certain extent. When integrating POIs, building footprints and remote sensing images with 0.3 m resolution, difference of accuracy between categories have been greatly improved. Among them, the commerce increases by 5.35%, the public increases by 5.55%, the industry increases by 7.66%, and overall accuracy is 90.95%. Kappa is 0.849 and F1 score is 0.858. Classification results for each category are shown in Table I. Therefore, fusion of multisource data based on SOE can further improve classification, of which the map is shown in Fig. 9.

Especially, we count differences of features based on building level, including building area, perimeter, floor, year, and other indicators, which always have been ignored in previous studies. Based on scene, we count density and total number of buildings. Functional zones in different types vary in characteristics, the detail is in Fig. 10. Mean floor of the commerce, public, industry, and residence are 10.83, 6.30, 5.43, and 7.67, respectively. This reflects floors of the commerce and residence's buildings are relatively high, which have a certain deviation from public and industry. And the standard deviation of floors in four categories is 6.84, 2.65, 2.24, and 3.56. For example, the commerce may

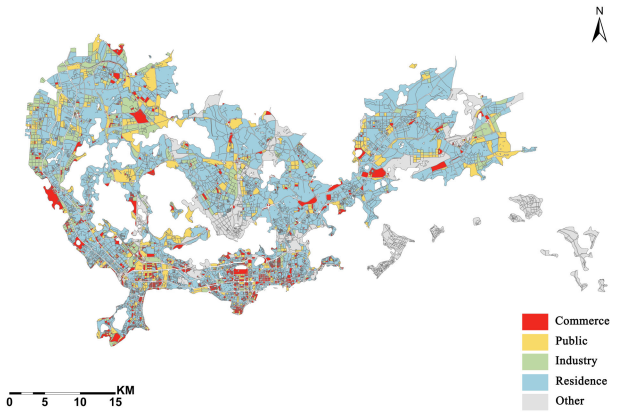

Fig. 9. Map of Shenzhen's functional zones.

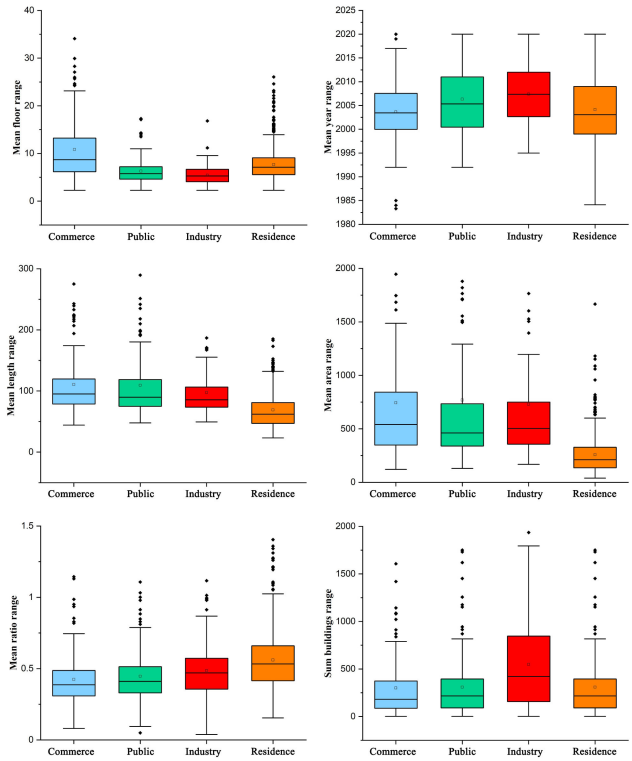

Fig. 10. Visualization of building features in different functional regions.

contain buildings with different uses such as supermarkets, shopping centers, small shops, and office buildings. There are certain restrictions on floors of supermarkets, shopping malls, generally within seven floors. Even for some small shops, personal retail stores, etc., floors are lower, controlled within two floors. But office buildings are usually high, even reaching 30 floors. This makes heights in a zone greater diversity.

The same situation also appears in area of buildings, which depends on its purpose. Zones in other three types, all of which are above 700 square meters, are much larger than the residence, 257.61 square meters. The area of a standard supermarket is generally 500–1500 square meters. That is, the area of large shopping malls and small retail stores, 20 square meters, may

TABLE II
RESULTS OF SORTING WEIGHTING FACTORS WHEN ONLY POIS IS USED

The features of POIs	Factor of weight
Life service	0.0126
Public facilities	0.0299
Medical care services	0.0436
Sports and leisure services	0.0489
Financial insurance services	0.0507
Government agencies and social organizations	0.0569
Catering services	0.0656
Transportation facilities services	0.0708
Education and cultural services	0.0833
Accommodation services	0.0852
Companies and enterprises	0.0966
Commercial and residential dual-purpose blocks	0.0984
Attractions	0.1082
Shopping services	0.1494

differ by dozens of times. The residence often contains buildings, regularly distributed, neatly arranged, having the same purpose. So the standard deviation of its characteristic is relatively small.

Four categories zones are about 2003, 2006, 2007, and 2004 in mean year. Buildings on industrial land are built in a relatively new year. The mean perimeter of residential buildings, 68.98 m, is quite different from other three categories with 110.55, 109.37, and 97.58. The perimeter of residential buildings is smaller than other three types, but individuals change little. But for mean structure ratio, the order from largest to smallest is the residence, industry, public, and commerce. This reflects complexity of building shape; commercial buildings often have special shapes. According to statistics, there are the largest number of industrial buildings and the smallest number of commercial buildings in a zone. Therefore, we suggest that buildings should be taken into consideration in future functional region recognition based on these differences.

B. Evaluating the Factor Contribution of Different Data Sources

Different features are combined by RF for both classification and factor contribution evaluation. Only using POIs can achieve accuracy of 71.9%. The weights of features in classification are illustrated in Table II. Shopping services, attractions, commercial, and residential dual-purpose blocks have a greater contribution to classification. The shopping services include large supermarkets, shopping malls, characteristic commercial streets, etc., which can infer that the zone is likely to be the commerce. However, some convenience stores, personal supplies stores, cosmetics stores, etc. are included, opened near residential communities. Commercial and residential dual-purpose blocks include office buildings, travel agencies, and talent markets, favoring business and finance. However, it also includes beauty salons, telecommunications business halls, lottery sales points, and residence. These are biased toward the residential land. These POIs indicators can provide information for different categories of urban zones, so lack of uniqueness leads to relatively low accuracy.

Only using building footprints, accuracy can reach 71.5%. Weights of 17 indicators are listed in Table III. The standard

TABLE III
RESULTS OF SORTING WEIGHTING FACTORS WHEN ONLY BUILDING FOOTPRINTS IS USED

The features of the building	Factor of weight
Building year standard deviation	0.0278
Mean building ratio	0.0390
Mean building year	0.0395
Sum of building area	0.0412
Sum of buildings	0.0416
Building ratio standard deviation	0.0421
Sum of building length	0.0488
Sum of building floor	0.0504
Sum of building ratio	0.0516
Mean building length	0.0559
Sum of building year	0.0596
Building length standard deviation	0.0624
Mean building floor	0.0642
Mean building area	0.0699
Building floor standard deviation	0.0727
Density of buildings	0.0914
Building area standard deviation	0.1417

TABLE IV
TOP 10 OF WEIGHTING FACTORS WHEN USING POIS AND BUILDING FOOTPRINTS

Features	Factor of weight
Building floor standard deviation	0.0366
Mean building floor	0.0370
Education and cultural services	0.0398
Landscapes	0.0419
Mean building area	0.0437
Commercial and residential dual-purpose blocks	0.0454
Companies and enterprises	0.0552
Density	0.0671
Shopping services	0.0699
Building Area standard deviation	0.0853

deviation of area plays the most important role, and its weight is 0.1417. Moreover, the density and standard deviation of floors also contribute a lot, which are 0.0914 and 0.0727, respectively. Results indicate that the introduced height of buildings plays an important role in classification. Similarly, spatial patterns of buildings in different functional zones are quite different, which make the standard deviation of building as a decisive factor in classification.

The accuracy obtains 76.2%, with combining POIs and building footprints data. And we find that addition of buildings based on POIs is benefit to identification of the commerce and industry, which improves 27.27%, 11.54%. Factor weights have changed slightly and are shown in Table IV. The standard deviation of area still plays the largest role. Because from the architectural point of view, the commerce and industry usually have larger areas, which makes standard deviation of area the largest. And the shopping services also have a certain influence. Among the top ten features of importance, there are five building indicators and 5 POIs indicators. POIs are mainly biased toward the commerce, the public, and the residence and with low spatial coverage of the industry. Among building features, area, density, and floor still contribute the most, while standard deviation, mean of area, and height best reflect differences of architectural style in a zone.

Furthermore, when remote sensing image features are added, factor weights change a lot. Results reveal that image features play a decisive role in classification. The total weight proportion

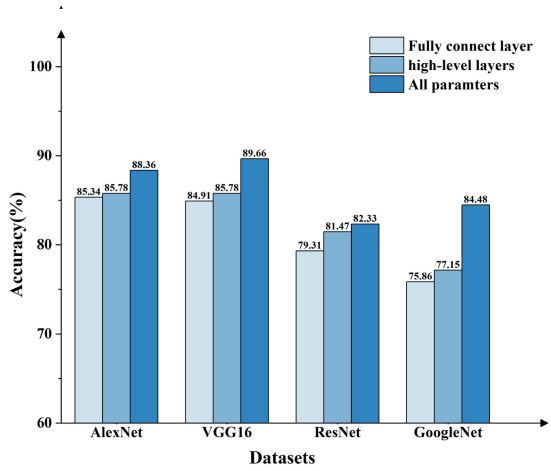

Fig. 11. Results of CNNs with different training ways.

reaches about 0.92. That indicates DCNN-based image features provide rich and diverse information about urban zones. This is in line with our expectations. When only remote sensing images are used, accuracy of 89.65% is achieved. However, it is no doubt that POIs and buildings also play a greater auxiliary role. Using single data source, even the remote sensing imagery cannot solve problem of uneven accuracy between categories. Addition of POIs and buildings not only improves accuracy, but also reduces the gap between categories. The commerce increases by 6.05%, the public increases by 5.55%, the industry increases by 7.69%, and the residence reduces by 2.19%. The gap between categories has decreased, of which accuracy reaches more than 80%. Therefore, we believe that based on SOE framework, integrating features of each dimension is a feasible way to effectively identify functional zones.

C. Identification Effects With Remote Sensing Images of Different Resolutions

The CNN are trained in different ways with images of 0.3 m resolution. The results obtained by using softmax layer are displayed in Fig. 11. Training with parameters of different neural layer, outputs are slightly different. Generally speaking, there is such a phenomenon that the method of all-parameters training is better than the partial neural layers and the nearly fully connected layer. The accuracy of partial neural layers and the fully connected layer are within the range of 2%. However, the training effect of all-parameters is quite different from the previous two ways. The gap can even reach up to 9%. AlexNet, VGG16, ResNet, and GoogLeNet are optimized when all layers are opened for training, of which results are 88.36%, 89.66%, 82.33%, and 84.48%. Four CNNs are utilized for recognition of functional zones. As shown in Fig. 11, it is observed that there are certain differences between effects of CNNs. VGG16 achieves the best result of 89.66%, which is 7.33% better than ResNet, 5.18% better than GoogleNet, and 1.3% better than AlexNet. Therefore, training all-parameters with VGG16 can achieve relatively optimal results.

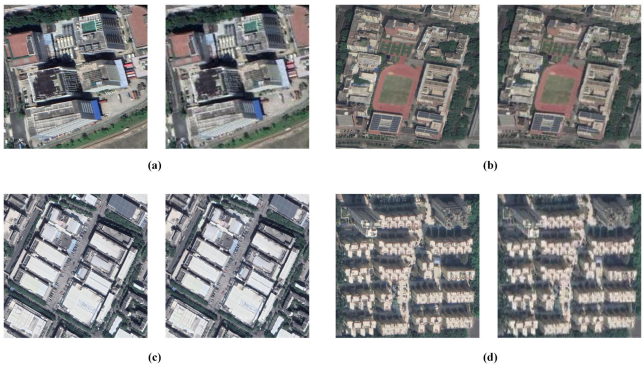

Fig. 12. Comparison of different resolution images. (a) Commerce. (b) Public. (c) Industry. (d) Residence.

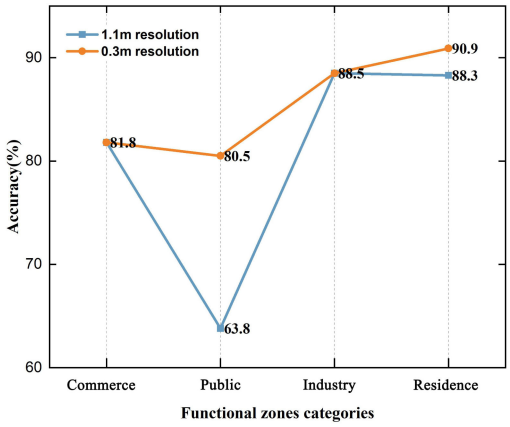

Fig. 13. Classification results with remote sensing images of 1.1 m and 0.3 m.

We think that an improvement of VGG16 over AlexNet is to use continuous 3 ×x 3 convolution kernel instead of larger convolution kernel. So convolution kernels of the same size can be used repeatedly to extract more complex and expressive features in functional regions. And the deeper layers of neural network or the more complex model is no better in our task. It may have the problem that gradient dispersion and loss cannot converge, such as GoogleNet and ResNet. Identifying urban functional regions for us, the result show that VGG16 has more advantages.

The impact of different resolution images for classification is also explored. The above optimal training way is applied to experiments on remote sensing images with resolutions of 0.3 and 1.1 m. In the case of different resolutions, the same functional region contains different pixels. From a visual point of view, differences are obvious. For example, the commerce with 0.3 m resolution contains 627 324 pixels, which is clearer than the 1.1 m resolution containing 39 560 pixels in Fig. 12(a). The public has 865 700 pixels with 0.3 m resolution, 54 175 pixels with 1.1 m resolution in Fig. 12(b). The industry involves 4 297 161 pixels with 0.3 m resolution, 267 795 pixels with 1.1 m resolution in Fig. 12(c). And the residence includes 607 185 pixels in 0.3 m resolution, 38 048 pixels in 1.1 m resolutions in Fig. 12(d). More pixels the image contains, the clearer it will be. Pixels of remote sensing images with 0.3 m resolution are about 16 times of 1.1 m resolution. The outline of buildings and plots are more detailed. These two image data are used for training.

Prediction results are illustrated in Fig. 13. Training with 1.1 m resolution remote sensing images, VGG16 with softmax layer reaches accuracy of 88.3%, including the commerce 81.8%, public 63.8%, industry 88.5%, and residence 96.3%. Except for the public accounts for about 60%, other three categories have reached more than 80%. In general category of the public, it contains more complex types, such as hospitals, scientific research institutions, schools, public services, etc. So buildings with different uses have no uniform style, irregular distribution, and similar characteristics. But using images with 0.3 m resolution, total accuracy gets 90.9%, with the commerce 81.8%, the public 80.5%, the industry 88.5%, and the residence 96.3%. Especially for the public, accuracy increases about 17%, compared with the previous one. This indicates that identification of the public parcels are more dependent on microscopic scene features. Accuracy of the residence is significantly higher than other types, illustrating that characteristics of residential parcels are more prominent and consistent. Results reveal that the higher resolution of image data, the better performance in CNNs classification. We speculate that CNN extracts more advanced features for the public in higher resolution images, which can be distinguished from other types. It seems that high-resolution images still play a significant role in recognition tasks. Even if the overall accuracy is limited in certain, it can reduce differences between different categories.

D. Identification Effects With Different Classifiers at the Fusion Stage

The article proposes the ensemble framework to identify functional zones. In order to verify optimal of our model selection, different classifiers are used for supplementary experiments, including XGBoost, SVM, Naive Bayes, LSTM, and dense layer. They are roughly divided into four types, including decision tree fusion, optimal plane segmentation, probability statistics, and neural network. And results are illustrated in Fig. 14. It is observed that the ensemble learning model based on decision tree has achieved better results. The accuracy of RF reaches 90.94% and XGBoost obtains 89.22%, much higher than other types classifiers. The model based on decision tree is to select an optimal eigenvalue for division. RF and XGBoost perform better, which also suggests features of functional zones with a high dimension are quite different in four types. Classification results of SVM, LSTM, and dense layer are all around 60%. But Naive Bayes based on probability statistics only achieves an accuracy of 48%. SVM combines multiple binary classifiers when dealing with multiple classification problems. This requires pairwise combination of samples in different categories. For example, when the commerce is classified with mixed data in other three types, there is no obvious optimal plane for segmentation. Characteristics of them are similar in some respects. It is expected that LSTM is often used for time series data, while dense layer is suitable for data with more feature dimensions. Naive Bayes assumes that attributes are independent of each

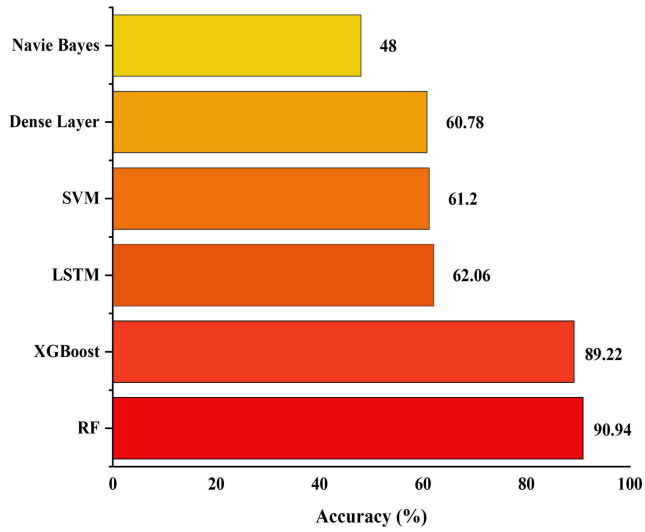

Fig. 14. Results by different classifiers.

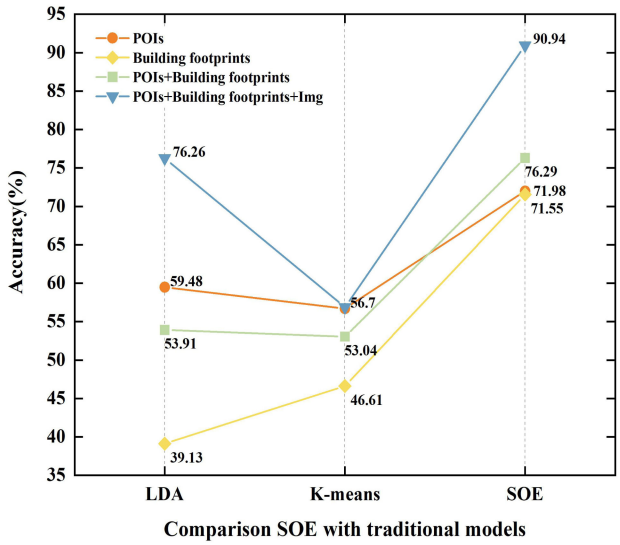

Fig. 15. Result of different method.

other. But there is a greater correlation between features of SOE dimensions. Moreover, our data lacks a prior probability, leading to a decrease in accuracy.

E. Comparison With Traditional Approaches

Two traditional models widely used in land use mapping are compared with our framework, LDA and K-means. Unlike models introduced above, both of them are based on unsupervised. Using two models with our data, we conduct four groups of experiments: 1) single POIs, 2) single building footprints, 3) combining POIs and building footprints, 4) combining POIs, building footprints, and remote sensing images. In Fig. 15, the result show that 1) only using POIs, LDA achieves accuracy of 59.48% and K-means achieves 56.70%, 2) only using building footprints, LDA gets 39.13% and K-means gets 46.61%, 3) combining POIs with building footprints, LDA and K-means obtain 53.91%, 53.04%, 4) LDA and K-means are 76.26%, 56.87% based on SOE features, lower than our framework with 90.94% accuracy. And it seems that they cannot extract features of POIs and buildings well. After comparison, we believe that our framework based SOE is more suitable and has achieved good results.

In the experiment, parcels of Shenzhen in EULUC-China map are used as research unit. For Level I (residence, entertainment, transportation (not participate in verification), industry, and office) classification, Peng Gong used samples with more than 70% of the dominant land use to train and verify, achieving accuracy of 58.9% [28]. In our study of identifying Shenzhen functional zones, samples with more than 60% of the dominant land use are employed for training and verification. The result reaches accuracy of 90.94%. This shows that the ensemble framework we proposed can effectively extract features of functional zones for classification.

F. Generalization to Other Cities

In order to evaluate applicability of our method, we conduct additional experiments in Beijing and Guangzhou.

For Beijing, 1535 samples are obtained by using semi-supervised sampling method. And training set and test set are also divided by a ratio of 4:1. For training set, a series of feature extraction methods are used to obtain images, buildings, POIs features, and then RF is used to classify vector features. Among 304 functional regions in test set, 42 commercial financial regions, 59 public facilities regions, 19 industrial storage regions, and 184 residential regions are included. In test set, the overall accuracy rate reaches 86.51%, 69.04% for commerce, 76.27% for public, 52.63% for industry, and 97.28% for residence. This confirms that our method is also applicable to other cities. The result of Beijing is little lower than that of Shenzhen because different cities have its own styles and land forms. There are great differences in distribution of their internal functional regions.

For Guangzhou, we validate the model trained in Shenzhen with multisource big data. Similarly, through semi-supervision, we select plots with dominant land accounting for more than 0.6. We utilize 758 functional regions, including 67 commercial regions, 130 public facilities regions, 145 industrial storage regions, and 416 residential regions. The result shows that overall accuracy rate of functional regions is 86.41%, including 77.61% for commercial, 76.92% for public, 83.44% for industrial, and 91.82% for residential. Considering there is no continuity in space of test samples, we show results in an area where samples are densely distributed in Fig. 16. These functional regions are concentrated in Haizhu district and Tianhe district. Fig. 16(a) is the prediction type of functional region and Fig. 16(b) is the real type of functional regions. Through comparison, it is found that classification error in the public functional regions is obvious. As mentioned above, the public includes many subcategories, such as educational institutions, parks, government agencies, service buildings, and facilities. So it is more challenging but the overall performance is good.

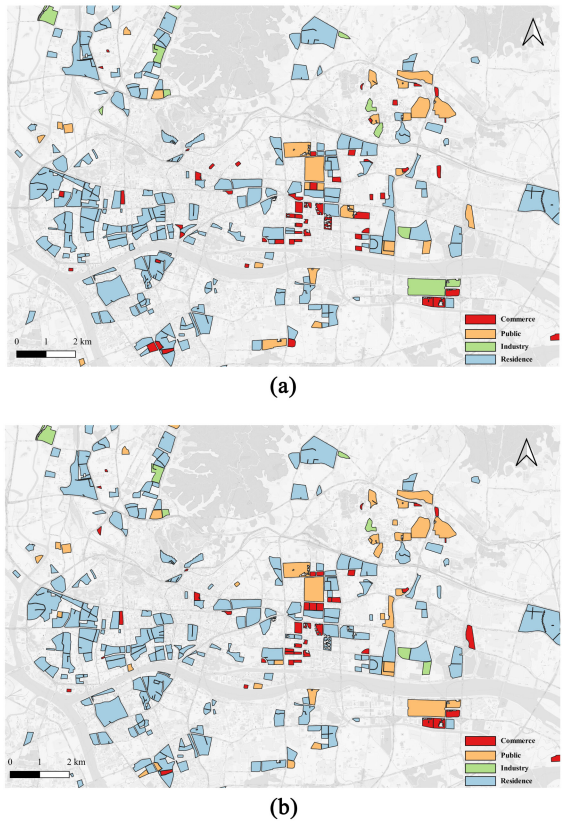

Fig. 16. Comparison of predicted values and the truth of some functional regions in Guangzhou. (a) Prediction values. (b) Ground truth.

Based on training on one city, it is applied to another city, showing the generalization ability of our model. The fusion method of SOE based to identify functional regions can effectively combine connections between scene, object, and economy, and deeply mine rich features.

V. Conclusion

Identifying urban functional zones plays a significant role in urban planning, designing, and managing. It is very important to find an efficient method replacing traditional survey approaches. The emergency of big data provides a new perspective for solving this problem. In recent years, fusion of multisource big data has become a research trend in recognition of functional zones to improve accuracy. Some studies integrate remote sensing data with socioeconomic features, combining topographic information and human activity characteristics. However, few experiments have considered characteristics of buildings. Our research integrates remote sensing, building footprints, and socioeconomic data based on the concept of SOE, conducting spatial information mining from three perspectives: scenes, objects, and economic. The height is first introduced to supplement 3-D information of buildings. This can make up for the insufficiency of extracting features from plane and analyzing architectural rules of building floors in functional zone. As expected, the height of buildings varies greatly to meet needs of uses. Mining more dimensional features can enrich information of parcels. In addition, an ensemble learning framework is proposed to better integrate SOE features. This is inseparable from selection of methods and fusion of features, both of which are conducive to the classification accuracy. Results show that our framework VGG16+RF is an excellent combination. It achieves the best results in a series of supplementary experiments, with an accuracy of 90.94%. Other types of classifiers such as LSTM, dense layer, SVM, and Naive Bayes are inferior for this task. We believe that features of functional regions have a threshold for node segmentation, so it is more suitable for decision tree type classifiers.

In fusion of multisource data, VGG16 extracts semantic information from remote sensing images and 64-D features in the penultimate fully connected layer are as output vector. In more detail, three different ways are tried to train the neural network. Training of all parameters in network can better fit our data. And results show that the higher the resolution, the better the classification performance. In addition, 17 building indicators and 14 socioeconomic characteristics are constructed. RF is used to automatically combine features for the final classification. In the research, different data sources are combined for experiments, which prove that the integration based on SOE is more helpful for identification of functional zones.

More importantly, contributing factors of different source data and differences in distribution patterns of buildings have been analyzed. Results show that remote sensing images play a decisive factor in classification. High-level features extracted by CNNs are extremely discriminative. On the influence of image resolution, we find that higher image resolution will lead to better performance. It is undeniable that area and height of buildings are also differential, which provides us with ideas for exploring their laws. Experimental results in Beijing and Guangzhou show great potential of our method.

In future research, improvements will be made in following aspects. First, identification of mixed regions will be considered. We will expand samples of mixed functions for further experiment. Second, the classification system of function region will be refined. Public category can be subdivided into research institutions, schools, hospitals, public facilities, institutions, etc. The refined land use data should be collected to complete labeling of parcels and expand the sample size of each parcel type. Third, the scope of current research will be expanded. And SOE-based methodology will be applied to more cities. Subsequently, differences in distribution patterns between cities can be deeper explored and application scenarios of model migration between cities will be developed.

References

[1] Y. Liu *et al.*, "Social sensing: A new approach to understanding our socioeconomic environments," *Ann. Assoc. Amer. Geography*, vol. 105, no. 3, pp. 512–530, May 2015.

[2] Y. Zhong *et al.*, "Open-source data-driven urban land-use mapping integrating point-line-polygon semantic objects: A case study of Chinese cities," *Remote Sens. Environ.*, vol. 247, 2020, Art. no. 111838.

[3] C. Zhang *et al.*, "An object-based convolutional neural network (OCNN) for urban land use classification," *Remote Sens. Environ.*, vol. 216, pp. 57–70, 2018.

[4] 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.

[5] Z. Huang, H. Qi, C. Kang, Y. Su, and Y. Liu, "An ensemble learning approach for urban land use mapping based on remote sensing imagery and social sensing data," *Remote Sens.*, vol. 12, no. 19, 2020, Art. no. 3254.
[6] X. Xing *et al.*, "Mapping human activity volumes through remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5652–5668, 2020, doi: 10.1109/JSTARS.2020.3023730.
[7] D. Ienco, R. Interdonato, R. Gaetano, and D. Ho Tong Minh, "Combining Sentinel-1 and Sentinel-2 satellite image time series for land cover mapping via a multi-source deep learning architecture," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 11–22, 2019.
[8] F. Pacifici, M. Chini, and W. J. Emery, "A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1276–1292, 2009.
[9] S. W. Myint, P. Gober, A. Brazel, S. Grossman-Clarke, and Q. Weng, "Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery," *Remote Sens. Environ.*, vol. 115, no. 5, pp. 1145–1161, 2011.
[10] F. Chen, K. Wang, T. Van de Voorde, and T. F. Tang, "Mapping urban land cover from high spatial resolution hyperspectral data: An approach based on simultaneously unmixing similar pixels with jointly sparse spectral mixture analysis," *Remote Sens. Environ.*, vol. 196, pp. 324–342, 2017.
[11] G. Cheng, X. Xie, J. Han, L. Guo, and G.-S. Xia, "Remote sensing image scene classification meets deep learning: challenges, methods, benchmarks, and opportunities," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3735–3756, 2020, doi: 10.1109/JSTARS.2020.3005403.
[12] P. Gong and P. Howarth, "The use of structural information for improving land-cover classification accuracies at the rural-urban fringe," *Photogramm. Eng. Remote Sens.*, vol. 56, pp. 67–73, Jan. 1990.
[13] L. Ma, M. Li, X. Ma, L. Cheng, P. Du, and Y. Liu, "A review of supervised object-based land-cover image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 130, pp. 277–293, 2017.
[14] M. Li, A. Stein, and W. Bijker, "Urban land use extraction from very high resolution remote sensing images by bayesian network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 3334–3337.
[15] B. Zhao, Y. Zhong, and L. Zhang, "A spectral–structural bag-of-features scene classifier for very high spatial resolution remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 116, pp. 73–85, 2016.
[16] X. Zhang and S. Du, "A linear dirichlet mixture model for decomposing scenes: Application to analyzing urban functional zonings," *Remote Sens. Environ.*, vol. 169, pp. 37–49, 2015.
[17] W. Zhao, Y. Bo, J. Chen, D. Tiede, T. Blaschke, and W. J. Emery, "Exploring semantic elements for urban scene recognition: Deep integration of high-resolution imagery and OpenStreetMap (OSM)," *ISPRS J. Photogramm. Remote Sens.*, vol. 151, pp. 237–250, 2019.
[18] D. Blei, A. Ng, and M. Jordan, "Latent dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, May 2003.
[19] X. Liu *et al.*, "Classifying urban land use by integrating remote sensing and social media data," *Int. J. Geographical Inf. Sci.*, vol. 31, no. 8, pp. 1675–1696, Aug. 2017.
[20] Y. Zhang, Q. Li, W. Tu, K. Mai, Y. Yao, and Y. Chen, "Functional urban land use recognition integrating multi-source geospatial data and cross-correlations," *Comput. Environ. Urban Syst.*, vol. 78, 2019, Art. no. 101374.
[21] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2012, pp. 186–194.
[22] N. Niu *et al.*, "Integrating multi-source big data to infer building functions," *Int. J. Geographical Inf. Sci.*, vol. 31, no. 9, pp. 1871–1890, Sep. 2017.
[23] C. Zhong, X. Huang, S. Müller Arisona, G. Schmitt, and M. Batty, "Inferring building functions from a probabilistic model using public transportation data," *Comput. Environ. Urban Syst.*, vol. 48, pp. 124–137, 2014.
[24] X. Zhang, S. Du, and Q. Wang, "Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data," *ISPRS J. Photogramm. Remote Sens.*, vol. 132, pp. 170–184, 2017.
[25] X. Liu, C. Kang, L. Gong, and Y. Liu, "Incorporating spatial interaction patterns in classifying and understanding urban land use," *Int. J. Geographical Inf. Sci.*, vol. 30, no. 2, pp. 334–350, Feb. 2016.
[26] W. Tu *et al.*, "Portraying urban functional zones by coupling remote sensing imagery and human sensing data," *Remote Sens.*, vol. 10, no. 1. 2018, Art. no. 141.
[27] R. Cao *et al.*, "Deep learning-based remote and social sensing data fusion for urban region function recognition," *ISPRS J. Photogramm. Remote Sens.*, vol. 163, pp. 82–97, 2020.
[28] P. Gong *et al.*, "Mapping essential urban land use categories in China (EULUC-China): Preliminary results for 2018," *Sci. Bull.*, vol. 65, no. 3, pp. 182–187, 2020.
[29] H. Xing and Y. Meng, "Integrating landscape metrics and socioeconomic features for urban functional region classification," *Comput. Environ. Urban Syst.*, vol. 72, pp. 134–145, 2018.
[30] T. Hu, J. Yang, X. Li, and P. Gong, "Mapping urban land use by using landsat images and open social data," *Remote Sens.*, vol. 8, no. 2, 2016, Art. no. 151.
[31] Y. Long and X. Liu, "Automated identification and characterization of parcels (AICP) with OpenStreetMap and Points of Interest," *Environ. Planning B Planning Des.*, vol. 43, pp. 341–360, Nov. 2013.
[32] W. Zhang, W. Li, C. Zhang, D. M. Hanink, X. Li, and W. Wang, "Parcel-based urban land use classification in megacity using airborne LiDAR, high resolution orthoimagery, and Google Street View," *Comput. Environ. Urban Syst.*, vol. 64, pp. 215–228, 2017.
[33] S. Gao, K. Janowicz, and H. Couclelis, "Extracting urban functional regions from points of interest and human activities on location-based social networks," *Trans. GIS*, vol. 21, no. 3, pp. 446–467, Jun. 2017.
[34] Y. Yao *et al.*, "Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model," *Int. J. Geographical Inf. Sci.*, vol. 31, no. 4, pp. 825–848, Apr. 2017.
[35] S. Xu, L. Qing, L. Han, M. Liu, Y. Peng, and L. Shen, "A new remote sensing images and point-of-interest fused (RPF) model for sensing urban functional regions," *Remote Sens.*, vol. 12, no. 6, 2020, Art. no. 1032.
[36] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, Mar. 2015, pp. 815–823, doi: 10.1109/CVPR.2015.7298682.
[37] W. Liu *et al.*, "SSD: Single shot MultiBox detector BT," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.
[38] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
[39] 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.
[40] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9.
[41] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, Jan. 2017.
[42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Comput. Sci.*, Sep. 2014.
[43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
[44] D. Marmanis, M. Datcu, T. Esch, and U. Stilla, "Deep learning Earth observation classification using ImageNet pretrained networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 105–109, Jan. 2016.
[45] V. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 67, pp. 93–104, Jan. 2012.
[46] S. Oliveira, F. Oehler, J. San-Miguel-Ayanz, A. Camia, and J. M. C. Pereira, "Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest," *Forest Ecol. Manage.*, vol. 275, pp. 117–129, 2012.
[47] A. Stumpf and N. Kerle, "Object-oriented mapping of landslides using Random Forests," *Remote Sens. Environ.*, vol. 115, no. 10, pp. 2564–2577, 2011.

Ying Feng received the bachelor's degree in information engineering from the School of Information Engineering, China University of Geosciences, Wuhan, China, in 2019. She is currently working toward the master's degree in geographical information systems (GIS) at the Institute of Remote Sensing and Geographic Information System, School of Earth and Space Sciences, Peking University.

Her research interests include social perception, spatial analysis, and deep learning technology.

Zhou Huang received the B.Sc. degree in geographical information systems (GIS) and the Ph.D. degree in cartography and GIS from Peking University, Beijing, China, in 2004 and 2009, respectively.

He is currently an Associate Professor of GI-Science with the Institute of Remote Sensing and Geographical Information Systems, Peking University. In addition, he serves as the Deputy Director with the Institute of Remote Sensing and GIS, Peking University; the Beijing Key Laboratory of Spatial Information Integration and Its Applications; and the Engineering Research Center of Earth Observation and Navigation, Ministry of Education, China. He has authored and coauthored more than 50 academic papers in international journals or conferences. His research interests include big geodata, high-performance geocomputation, distributed geographic information processing, spatial data mining, and spatial database.

Dr. Huang was selected for the Youth Talent Innovation Plan in Remote Sensing Science and Technology, in 2015, funded by the Ministry of Science and Technology of China.

Yaoli Wang received the bachelor's degree in geographical information system from the Peking University, Beijing, China, in 2012, the master's degree in geography from the University of Georgia, Athens, GA, USA, in 2014, and the Ph.D. degree in infrastructure engineering from the University of Melbourne, Parkville, VIC, Australia.

She is currently a Postdoc in Geographic Information Science with Peking University. Her research interests include inferring human behaviors from social and spatial network analysis, and on-demand urban transportation systems.

Lin Wan received the bachelor's degree in computer science and technology from Wuhan University of Technology, Wuhan, China, in 2002, the master's degree in computer software and theory from the China University of Geosciences, Wuhan, China, in 2009, and the doctor's degree in geographical information systems from the School of Information Engineering, China University of Geosciences, Wuhan, China, in 2012.

His research interests include deep learning, and big data analysis and mining.

Yu Liu received the B.S. and M.S. degrees from the School of City and Environment in 1994 and 1997, respectively, and the Ph.D. degree from the School of Computer Science and Technology, Peking University, Beijing, China, in 2003.

He is currently a Professor with the Institute of Remote Sensing and Geographical Information Systems, Peking University. His research interests include humanities and social science based on big geodata.

Yi Zhang received the bachelor's degree from the School of Computer Science, Beijing University of Technology, Beijing, China, in 1994, and the master's and doctor's degree in geographic information science from Peking University, Beijing, China, in 2001 and 2011, respectively.

His research interests include qualitative geographic knowledge representation, geographic information retrieval, text spatiotemporal semantic processing, and tourism big data.

Xv Shan received the graduate degree from Heilongjiang University, Harbin, China, in 2004.

He is a staff of State Key Laboratory of Media Convergence Production Technology and Systems, also serves as a Deputy Director Space News Lab of Xinhua News Agency. He is dedicated to exploring the integration and innovation of Big Data technology, Satellite Remote Sensing Technology and Media. He has organized and implemented several news reports and research topics based on spatial–temporal big data mining and analyzing.

Mr. Shan was the recipient of the Xinhua News Agency Innovation Award for many times.