Accurate Estimation of the Proportion of Mixed Land Use at the Street-Block Level by Integrating High Spatial Resolution Images and Geospatial Big Data

Jialyu He, Xia Li, Penghua Liu, Xinxin Wu, Jinbao Zhang, Dachuan Zhang, Xiaojuan Liu, and Yao Yao

***Abstract*— Mixed land use has been widely used as a planning tool to improve the functionality of cities. However, depicting mixed land use is rather difficult due to its complexities. Previous studies have decomposed urban land areas using either remote sensing images or geospatial big data. Few studies have combined these two data sources because of the lack of methodologies. This article proposed an end-to-end two-stream convolutional neural network (CNN) for combining features (CF-CNN) to estimate the proportion of mixed land use by integrating high spatial resolution (HSR) images and geospatial big data of real-time Tencent user density (RTUD) data. Two deep learning networks, one for image information extraction and other for human activity-related information extraction, are used to construct two branches of CF-CNN. The mixed land use can be described by calculating the proportions of each land use type at the street-block level. Compared with methods for using single-source data, CF-CNN obtained the highest classification accuracy. We further applied the Shannon diversity index (SHDI) to quantify the agglomerated urban mixed land use. The Spearman correlation coefficients among the SHDI, community distance, and neighborhood vibrancy were calculated to verify the effectiveness of the mixed land use composition. Our framework provided an alternative way of identifying mixed land use structures by integrating multisource data.**

***Index Terms*— Deep learning, high spatial resolution (HSR) images, mixed land use, real-time Tencent user density (RTUD), remote sensing.**

I. Introduction

DUE to the rapid development of cities and universal urban renewal projects, especially in mega cities, single land use cannot satisfy the growing demands of human living. Instead, mixed land use has gradually become a desirable choice to ensure urban habitability [1]. In urban planning, cities are usually divided into land parcels of various types according to their local geographical conditions and human activities [2]. Mixed land use is composed of two or more types of land use, such as industrial zones, commercial zones, and residential districts, in a land parcel to simultaneously provide services for different groups [3].

As the key feature of mega cities, mixed land use can support urban livability by reducing the commuting distances, promoting the neighborhood vibrancy, and increasing the walking-to-driving ratio [4]–[6]. Nevertheless, mixed land use can also introduce some problems to urban development. For instance, a mixed land parcel that combines commercial zones, industrial zones, and residential districts may produce excessive noise, traffic congestion, or environmental pollution [7]–[9]. In contemporary urban planning, the spatial distribution of mixed land use can help in understanding the spatial pattern of the city. The information about the internal components of a city is generally quantified to assess their impact on urban development [10]. To measure the mixed degree of land use, urban planners introduced various indicators, such as the Shannon or Simpson indexes [11]. The assessment results can also provide support for urban planning and policy making. Hence, the spatial distribution and composition of urban mixed land use are essential tools for urban planners. The most common methods for mapping urban land are field surveys and manual interpretation using satellite images, which are both time consuming and laborious [12]. In particular, the implementation of the manual approach is considerably more challenging when the land parcels have mixed land use types. Thus, the urban mixed land use depiction has attracted extensive attention from urban planners.

However, research teams disregarded the phenomenon of urban mixed land use due to its complexity. The quantitative measurement of mixed land use is a challenging task,

Manuscript received February 25, 2020; revised April 22, 2020, June 14, 2020, and August 15, 2020; accepted September 26, 2020. This work was supported in part by the Key National Natural Science Foundation of China under Grant 41531176 and in part by the National Key Research and Development Program of China under Grant 2017YFA0604402. *(Corresponding author: Xia Li.)*

Jialyu He, Xinxin Wu, Jinbao Zhang, and Dachuan Zhang are with the Guangdong Provincial Key Laboratory of Urbanization and Geosimulation, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China (e-mail: hesysugis@foxmail.com; wuxx33@mail2.sysu.edu.cn; kampau@foxmail.com; zhangdchgis@foxmail.com).

Xia Li and Xiaojuan Liu are with the Key Laboratory of Geographic Information Science (Ministry of Education), School of Geographic Sciences, East China Normal University, Shanghai 200241, China (e-mail: lixia@geo.ecnu.edu.cn; liuxj58@mail2.sysu.edu.cn).

Penghua Liu is with the Alibaba Group and Ant Group, Hangzhou 311121, China (e-mail: penghua.lph@alibaba-inc.com).

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

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

Digital Object Identifier 10.1109/TGRS.2020.3028622

especially without fine-scale ground truth validation data. Existing studies have attempted to portray mixed land use using methodologies that were modified from conventional land use classification frameworks [13]. To reveal the spatial structure of cities, these studies usually segmented cities into blocks by community boundaries or road networks [14], [15]. The components of various types of land use were obtained at the block level to describe the mixed land use. In these studies, remote sensing images or geospatial big data (traditional remote sensing images collected by satellites or unmanned aerial vehicles are not included in geospatial big data in this article [16]) are still the primary data sources. The natural physical properties, socioeconomic features, and human activity-related information contained in multisource data are highly relevant to urban functional types [17]. Since existing studies on the identification of mixed land use are mainly based on single-source data, the integration of multisource data has the potential to enhance the performance of mixed land use depiction.

It has been a challenge to fuse these features extracted from multisource data to enhance the recognition power of the models for ground objects. A variety of models, such as hierarchal clustering [18], semantic information fusion [19], and deep learning techniques [20], have been introduced to integrate multisource data. The advantage of multisource data-based models is that they can simultaneously leverage the visual information of remote sensing images and the human activity-related information of geospatial big data to predict the land use at an urban-object level. Deep learning techniques have attracted a substantial amount of attention due to their strong capabilities for processing spatiotemporal data [21]. As the prototype of deep learning networks, convolutional neural networks (CNNs) have been rapidly developed and have achieved better performance than other baseline models (e.g., logistic regression, random forest, and support vector machine models) in various fields, including image processing [22], text analysis [23], and semantic segmentation [24]. Features extracted by CNNs are also practical for the representation of ground objects [25], [26].

Inspired by these advantages of CNNs, this article proposes an end-to-end two-stream CNN for combining features (CF-CNN) to estimate the accurate proportion of mixed land use by simultaneously processing high spatial resolution (HSR) images and geospatial big data. Specifically, two CNN architectures were employed for visual feature extraction of HSR images and human activity-related feature extraction of geospatial big data. These two types of CNN features were concatenated to provide CF-CNN with the ability to process multisource data. To verify the effectiveness of CF-CNN, we also separately classified land use using two deep learning networks for comparison. Our experiments were carried out in four districts of Guangzhou, which is one of the mega cities in China. The street-block data were utilized to divide the study area into 2931 units to calculate the proportions of land use types. Based on these proportions, we quantitatively measured the mixed land use pattern using the Shannon diversity index (SHDI) and analyzed its contribution to Guangzhou.

II. Related Work

Most land use classification projects can be completed by separately applying two primary data sources: remote sensing images and geospatial big data. For a particular ground object, remote sensing images can provide features related to its natural physical properties, including the spectrum, texture, and shape. Conversely, geospatial big data can provide various features related to socioeconomic environments and human activities [27]. Both types of features have an important role in land use classification. Still, the approaches to processing remote sensing images or geospatial big data are different due to their various data structures. Therefore, this section provides an overview of the methodologies based on single-source data (remote sensing images and geospatial big data) and multisource data, as well as the mixed land use depiction.

A. Remote Sensing Image-Based Methods

With a continuous breakthrough in remote sensing technology, land use classification based on remote sensing images is undergoing an unprecedented development. Previous studies have identified land use by feeding the natural physical features extracted from remote sensing images into classifiers [28]–[30]. The variety of remote sensing images has spawned numerous data processing methods. Traditional methods tend to extract the spectral, texture, or other attribute features of each research unit, thus generating feature vectors through different feature construction modes, such as feature concatenation [31], low-rank representation [32], and semantic models [33]. The primary research units are gradually converted from pixels to objects and scenes to better depict land parcels in a city. HSR images are available to enhance the performance of land use classification. However, extracting these handcrafted features require a considerable amount of engineering skill and domain expertise [34]. In particular, the large volume of HSR images requires high efficiency and performance of the model [35]. To solve these issues, more advanced and automated algorithms have been applied in recent years, the most attractive of which is the deep learning technique due to its ability to automatically extract higher level feature representations [36]–[39]. Many researchers have adequately trained deep learning models to classify the land use by building remote sensing databases with high quality, such as UC-Merced [40], Whu-SIRI [41], and AID [42]. In the absence of high-quality remote sensing databases, transferring pretrained networks to remote sensing researches has been shown to be effective and time saving due to their high performance for classifying traditional image databases. In addition, the function of deep learning models has been improved by adding other mechanisms, i.e., domain adaptation [43], atrous convolutions [44], and attention mechanism [45].

B. Geospatial Big Data-Based Methods

Geospatial big data are extensively employed in land use classification tasks due to their rich and diverse characteristics [27]. Carriers of geospatial big data, such as mobile phones, global positioning system devices, and laptops, can provide information about socioeconomic environments and human activities [46]. Previous studies have utilized a variety of geospatial big data to retrieve information about urban spatial structures. Yuan *et al.* [47] extracted human mobility patterns from floating car trajectories as a word to infer the functions of urban land parcels by applying a topic-based inference model. Pei *et al.* [48] proposed a semisupervised clustering algorithm to identify land use using standardized hourly call volume and total call volume obtained from mobile phone data. Chen *et al.* [49] applied a dynamic time warping distance-based k-medoids method to cluster the curves of population density to infer urban functions. Similar to remote sensing image-based methods, deep learning techniques have achieved a high level of performance in the geospatial big data processing. The difference is that deep learning frameworks based on big data are more diversified due to the diversification of big data structures. The variants of deep learning techniques have been proven to be effective in processing a wide range of geospatial big data, such as points of interest [46], air quality data [50], and spatial trajectories [51].

C. Multisource Data-Based Method

Solving problems by integrating multisource data has been a research hotspot in recent years [52]. Compared with single-source data, multisource data can contain more information and achieve complementary advantages, which is conducive to making decisions about the same problem on multiple scales and perspectives [16]. Feature fusion is the most commonly applied strategy in multisource data integration; in this strategy, feature engineering is critical because it affects the quality of feature concatenation. For example, Zhang *et al.* [53] applied Weibo and points of interest to enhance the performance of classification based on remote sensing images. The density of geospatial big data was calculated to combine with the images. Liu *et al.* [19] constructed a framework to classify the urban land use by extracting semantic-level features from HSR images and geospatial big data. The semantic-level features are concatenated to obtain the final representation of the urban land parcel. However, these methods still have a common deficiency: although integrating multisource data through feature engineering can improve the accuracy of classification, the inconsistent data structures and feature dimensions might increase the complexity of the entire framework, thus requiring substantial expert knowledge and experimental time. Therefore, multistream CNNs have become increasingly attractive due to their strong unsupervised feature learning ability for multisource data. High-dimensional and abstract features can be automatically extracted. The integration of these CNN features is effective for classification tasks, whether the CNNs are built as an end-to-end model or individually serialized to construct the final classifier [22], [54].

D. Mixed Land Use Depiction

The quantitative measurement of a mixed degree is usually carried out based on the proportion of each component in a fixed unit. Therefore, how to obtain the ratio of various types of land use is a keypoint that needs to be solved urgently.

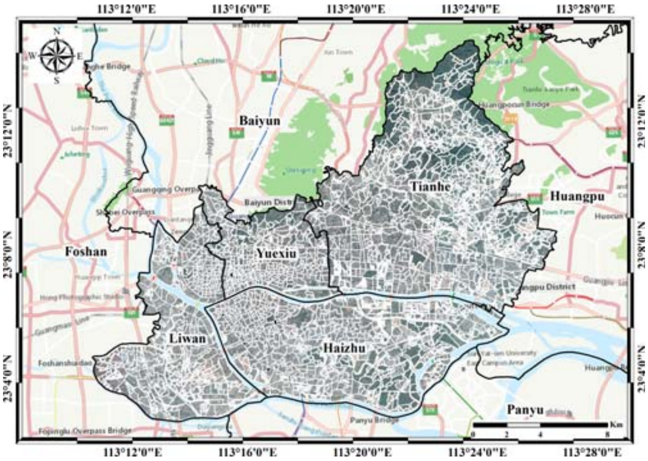


Fig. 1. Study area with street blocks in Guangzhou.

Huang *et al.* [13] applied a skeleton-based decomposition method that integrates deep learning techniques to map the urban land use with HSR images to calculate the mixed composition of each land parcel. Zhang and Du [15] decomposed urban scenes in land parcels by modifying the pixel unmixing method. Wu *et al.* [2] proposed an urban function base curve for decomposing urban mixed land use based on the temporal activity signatures extracted from check-in data. Xing *et al.* [14] constructed a dynamic human activity-driven model that integrates a massive amount of Twitter messages to assess the mixed land use patterns. All these studies have attempted to extract the information from a single data source, such as HSR images or check-in data, to identify the composition of various land uses and measure the mixed degree on the scale of urban land parcels. However, they neglected to integrate these multiperspective informations to improve the representational ability of the model. Thus, it is worthwhile to put forward a method to depict the mixed land use based on multisource data.

III. Study Area and Data Sets

A. Study Area and Data Collection

As the capital of Guangdong Province, Guangzhou is located in the Pearl River Delta, a fast-growing region in China [55]. Constrained by a limited amount of available land and a large number of inhabitants, Guangzhou is characterized by complex land use patterns. Some land parcels in Guangzhou are highly mixed and difficult to decompose [11]. In this article, we chose several central urban regions of Guangzhou as the study area, namely, Haizhu district, Yuexiu district, Tianhe district, and Liwan district (refer to Fig. 1). These four districts cover a total area of 303.27 km^2. An HSR image of the study area in 2018 was downloaded from Google Earth; the image contains three spectral bands. The spatial resolution of the image is approximately 1.16 m, and the size of the image is $23\,756 \times 19\,548$ pixels. To obtain the mixed composition of the land parcels, the street-block data were employed to segment our study area. In total, there are 2931 street blocks.

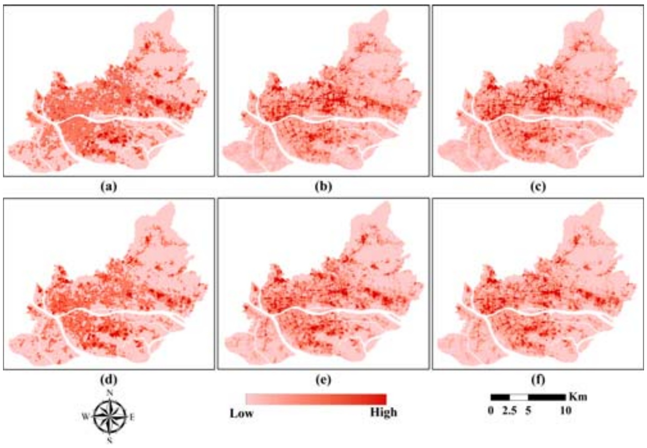


Fig. 2. RTUD data in the study area. (a) 8:00 A.M., (b) 12:00 P.M., and (c) 6:00 P.M. on a workday and (d) 8:00 A.M., (e) 12:00 P.M., and (f) 6:00 P.M. on a holiday.

Real-time Tencent user density (RTUD) data were selected as the geospatial big data to provide human activity-related information. Our RTUD data were collected from Tencent, one of the largest Internet enterprises in China [49]. More than 800 million people have accounts on the platform. User locations (addresses) are recorded when they are using location-related services such as Tencent Map or WeChat, thus enabling the RTUD data to store the population distribution of the current period. Compared with other conventional population density maps, the RTUD data have a higher spatiotemporal resolution. Tencent's platform marks each user position per hour in a fixed range, with a spatial resolution of approximately 27.05 m, rendering it ideal for fine-scale urban studies. Hence, we applied a web crawler tool to grab the RTUD data for a week as raster images. The population density raster images have 24 bands, representing 24 h of the day. According to previous studies using population density data, the trajectory of human activity significantly varies under the influence of weekdays and holidays [49], [56]. To effectively represent the dynamic human mobility, we divided the RTUD data into two parts: the population density on weekdays and the population density on holidays. Fig. 2 shows the raster images of the RTUD data at three different times. The pixel value of each image represents the average number of people at the corresponding time within the pixel extent, obtained by averaging the data of weekdays and holidays.

B. Samples for Training the Deep Learning Networks

Guangzhou has seven major types of land use: public management services, industrial zones, green land, commercial zones, residential districts, urban villages, and vacant land. We estimate the proportion of mixed land use by counting the number of image blocks of each type. The image block size of each pure scene was set to 128 pixels (approximately 150 m) based on our HSR image according to previous studies [13]. The label of each image block was tagged by manual interpretation [57]. We simultaneously obtained the samples from the HSR image and RTUD data. Figs. 3 and 4 show the sample images and sample curves of each land use type. We selected 2100 samples, including 300 samples per land use class for training.[1] Likewise, we selected another 350 samples, comprising 50 samples per land use type, to test our trained model. Considering the difference in the spatial resolution between the HSR image and the RTUD data, we selected the corresponding image block of the same area based on each sampling point and took the average population density of each time period in the block as the RTUD feature. Fig. 4 shows the averaged time-series population density curves of the training sample points for each land use type. The heterogeneity of the temporal feature curves is evidence of the effectiveness of the RTUD data for land use classification. We set each sample of the RTUD data to $D(w) = \{P_{w1}, P_{w2}, \ldots, P_{wn}\}$ for weekdays and $D(h) = \{P_{h1}, P_{h2}, \ldots, P_{hn}\}$ for holidays, where P_{wj} and P_{hj} are the population density at time j on weekday and holiday, respectively. Unlike traditional images, the RTUD data composed of $D(w)$ and $D(h)$ will be input into the deep learning networks as a vector with 48 dimensions, which represents the 24 h of weekdays and holidays.

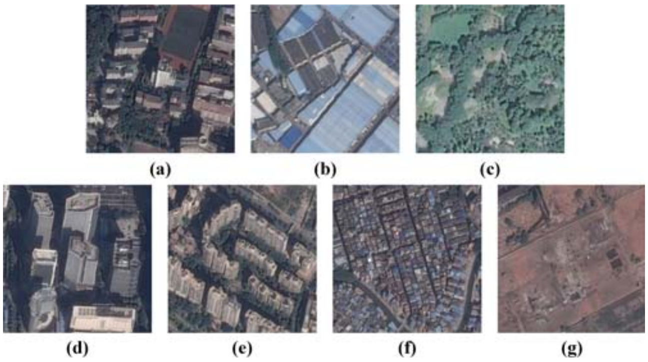


Fig. 3. Sample images collected in Guangzhou. (a) Public management services. (b) Industrial zones. (c) Green land. (d) Commercial zones. (e) Residential districts. (f) Urban villages. (g) Vacant land.

IV. Methodology

The proposed framework (CF-CNN) consists of a simplified residual network (SRes-Net) and a modified Visual Geometry Group network (PVGG-Net). The main goal of the CF-CNN is to estimate the proportion of urban mixed land use by fusing features extracted from multisource data. Our framework can be divided into five parts: 1) we simplified the conventional residual neural network (ResNet) as the SRes-Net and used it to classify the HSR images; 2) due to the particularity of the RTUD data, we introduced a 1-D PVGG-Net, which was modified based on a VGG16 network; 3) we randomly selected samples to train these two deep learning networks, and the street-block data were utilized to segment the study area into land parcels; 4) we employed the two independently trained deep learning networks as feature extractors to construct a two-stream CF-CNN by concatenating the extracted features before the last classifier; and 5) at the street-block level,

[1] The code and dataset are downloadable at https://github.com/SysuHe/MultiSourceData_CFCNN

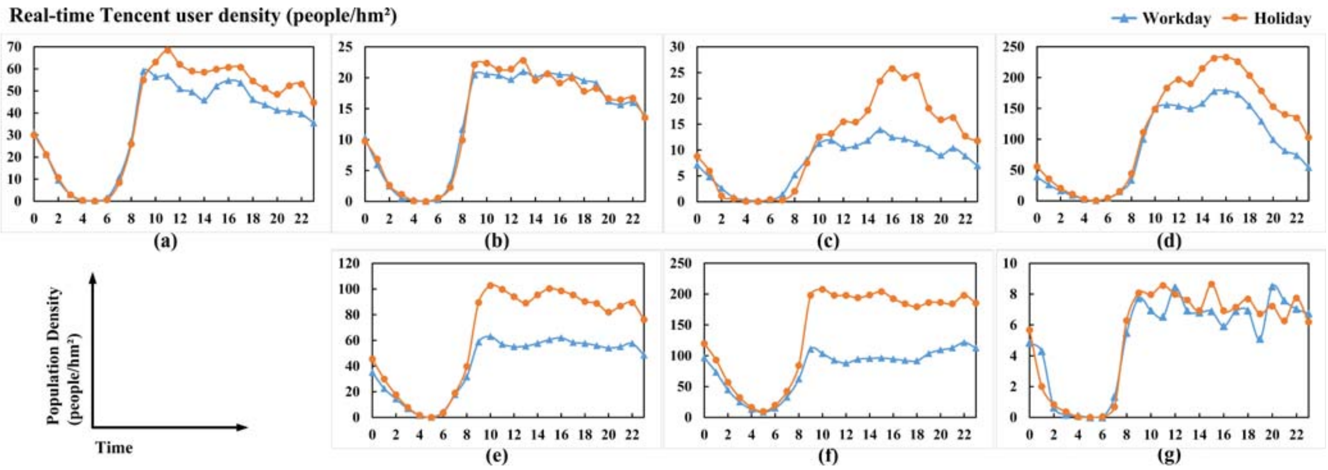


Fig. 4. RTUD curves of the samples collected in Guangzhou. (a) Public management services. (b) Industrial zones. (c) Green land. (d) Commercial zones. (e) Residential districts. (f) Urban villages. (g) Vacant land.

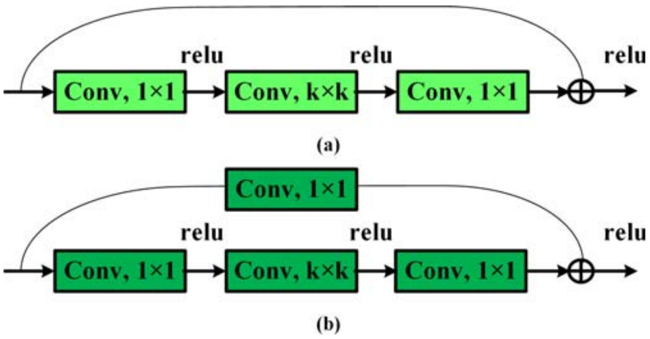


Fig. 5. Two types of residual blocks. (a) Identity block in Fig. 6. (b) Convolutional block in Fig. 6.

we counted the number of land blocks for each category so that the percentage can be eventually calculated as the final mixed proportion maps.

A. Simplified ResNet Network

ResNets were introduced by He *et al.* [58] to ease the training process of deep learning networks. The core idea of ResNet is to establish the skip connections between a layer and its subsequent layer, which are called residual blocks (refer to Fig. 5). The residual blocks allow the network to capture more abstract image information with a deeper structure. Generally, ResNet is designed to include as many layers as possible to enhance the representation ability of the nonlinear characteristics of the network. Nevertheless, a network with so many layers is not needed, because it is still too deep for our image data set. Thus, we adopted the SRes-Net based on the simplest ResNet model, rendering it more suitable for our data set. As shown in Fig. 6, the structure of the SRes-Net contains one convolutional layer, one max pooling layer, six residual blocks, two dropout layers, a seven-way fully connected layer, and a softmax classifier. Two main types of residual blocks are shown in Fig. 5: the identity block and the convolutional block. The identity block can reduce the number of parameters for network training using a parameter-free identity shortcut, while the convolutional block is used for the matching dimension [59]. To reduce the number of parameters for training, we applied these two types of residual blocks in our structure. Every residual block contains three convolutional layers with sizes of 1 × 1, 3 × 3, and 1 × 1.

B. 1-D Deep Learning Network

Given the significant modal differences between the RTUD data and the HSR image, the methods for extracting their features should also be different. Modified from VGG-16 [60], we applied a 1-D deep learning network for processing the curves of population density (PVGG-Net). The advantage of the PVGG-Net is that it can extract the temporal information about human activity from the RTUD data through multiple 1-D convolutional layers, which are designed to process 1-D temporal data [61]. Concerning the structure of WaveNet [62], we introduced atrous convolutional layers into the PVGG-Net to increase the receptive field of the convolution kernel without losing the time sequence information. As shown in Fig. 7, the structure of the PVGG-Net contains five 1-D convolutional layers, two fully connected layers, and a softmax classifier. The size of all the convolution kernels was set to 3 to obtain a more abstract feature. The dilation rate of the atrous convolutional layers is set to 2. Another three dropout layers were added to avoid overfitting caused by a massive amount of parameters and interdependence among neuron nodes. The probability of neuron inactivation was set to 25% and 50% in different layers. At the end of the network, we adopted a seven-way fully connected layer and a softmax classifier to obtain the classification results.

C. Estimating the Proportion of Urban Mixed Land Use

Both previously described models have very similar bottleneck layers, followed by a seven-way fully connected layer. As shown in Fig. 8, we can concatenate these two CNN bottleneck layers since they have been verified as a good representation of the input [26], [63]. For the ground object of the same location, we simultaneously input the HSR image blocks and the population density curves into the network. Both networks (SRes-Net and PVGG-Net) merely keep the

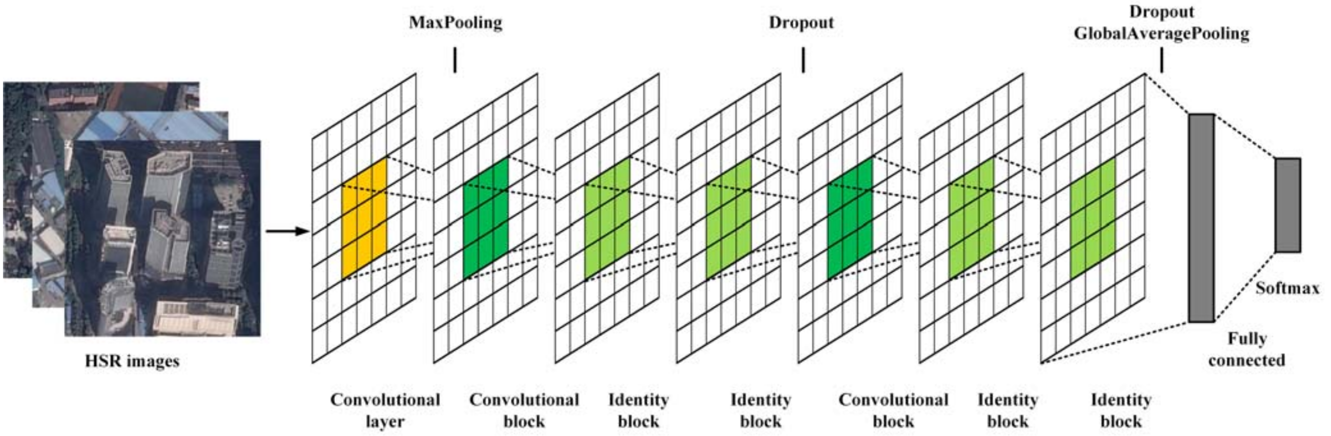


Fig. 6. Structure of the SRes-Net with one convolutional layer, one max pooling layer, six residual blocks (see Fig. 5), a seven-way fully connected layer, and a softmax classifier.

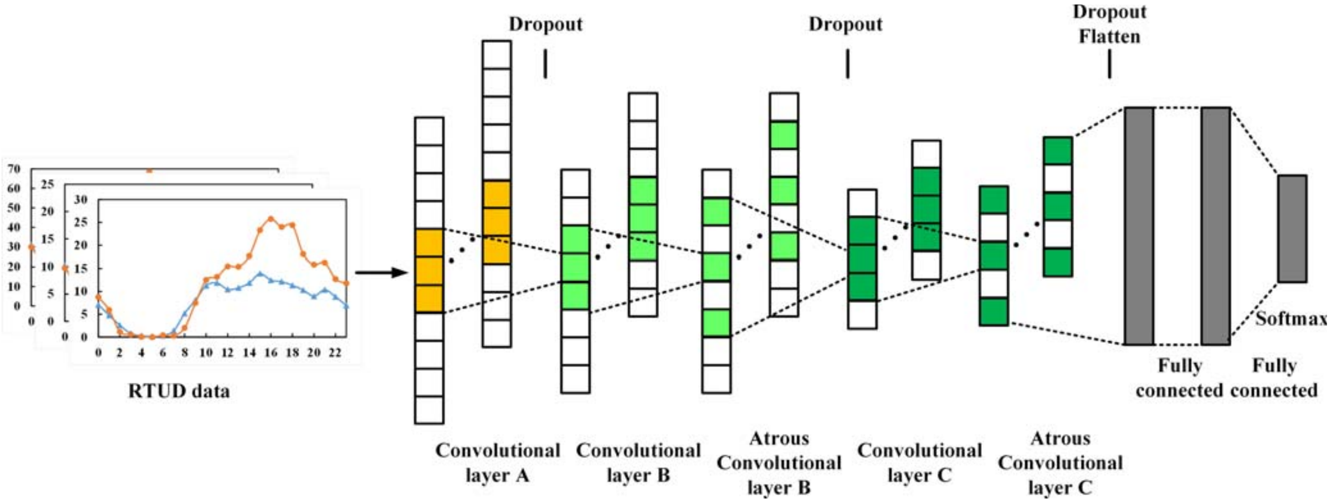


Fig. 7. Structure of the 1-D deep learning network (PVGG-Net) with five convolutional layers, two fully connected layers, and a softmax classifier.

part of feature extraction by discarding the last layer. The fully connected layers in each network were retained to extract the features from multisource data. Later, the extracted features were concatenated into the final high-dimensional feature vector. A fully connected layer and a softmax classifier were applied to obtain the classification result by feeding the final high-dimensional feature vector. Based on the classification results of each block, we can obtain the proportion maps by calculating the percentages of all the land use categories. The equation of the proportion is expressed as

$$p_{ik} = \frac{n_k}{\sum_k^{K_i} n_k} \quad i = 1, 2, \ldots, N \tag{1}$$

where p_{ik} represents the proportion of the kth category in the ith land parcel, and n_k is the total number of land blocks of the kth category.

To further characterize the urban mixed use pattern in the land parcels, we applied entropy indices to the measurement of the land use diversity. Generally, most of the indicators that characterize the mixed land use pattern are derived from landscape ecology. Considering the similarity among the landscape indices, we chose the SHDI [64]. The function corresponding to the SHDI is expressed as

$$S_i = -\sum_k^{K_i} p_{ik}\ln(p_{ik}) \quad i = 1, 2, \ldots, N \tag{2}$$

where K_i is the total number of urban land use categories in the ith land parcel, and p_{ik} represents the proportion of the kth category. When there is only one category of land inside the local land parcel, the value of S_i is 0. Otherwise, the value of S_i will be larger if the land parcel contains multiple categories of land use.

V. Results

In our experiments, 2100 selected samples were used for model training, while the other 350 samples were used for model testing. Both deep learning networks of the CF-CNN were set as the feature extractor to extract the features from the HSR images and RTUD data. The dimensions of the

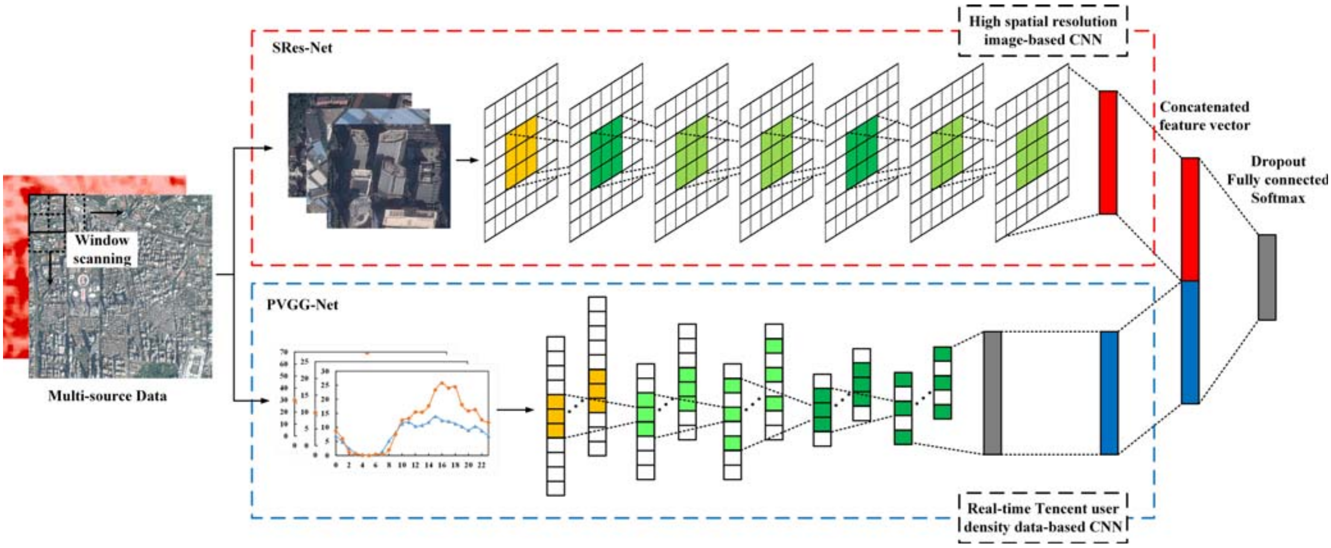

SRes-Net
High spatial resolution image-based CNN
Concatenated feature vector
Dropout
Fully connected
Softmax
Window scanning
Multi-source Data
PVGG-Net
Real-time Tencent user density data-based CNN

Fig. 8. Structure of the proposed CF-CNN (constructed by the SRes-Net and PVGG-Net). The extracted features are concatenated into the final dropout, seven-way fully connected layer, and softmax classifier to classify the label of input image block.

two CNN features were fixed at 1024. We then concatenated these two feature vectors into one 2048-D CNN feature. The 2048-D feature vector was regarded as the representation of each sample based on multisource data integration. A seven-way fully connected layer and a softmax classifier were added at the end of the CF-CNN to obtain the final classification result. To compare with the single-source data-based method, we conducted the experiments merely using the SRes-Net or PVGG-Net. The HSR images and RTUD data were applied as input data for these two comparative experiments. In addition to comparing the accuracy curves of all deep learning models, we also employed the confusion matrix, overall accuracy, and Kappa coefficient as the criteria for model evaluation. To depict the mixed land use of the study area, we utilized the scanning window to obtain the image blocks. All image blocks were entered into the trained CF-CNN for classification. The mixed land use was revealed through the statistics of various components in each land parcel. Furthermore, we applied the SHDI to characterize the mixed land use and conducted a Spearman correlation analysis among the SHDI, community distance, and neighborhood vibrancy to illustrate the effectiveness of our results.

A. Validation of the CF-CNN Framework

The 2100 training samples were divided into two parts: 80% of the samples were used for model training, and the remaining 20% were used for model validation. The learning rate of the training step, maximum number of iterations, and batch size of each training step for the SRes-Net were set to 0.0004, 100, and 16, respectively. The learning rate of the training step, maximum number of iterations, and batch size of each training step for the PVGG-Net were set to 0.0005, 100, and 16, respectively. When we trained the CF-CNN with multisource data, the learning rate of the training step, maximum number of iterations, and batch size of each training step were set to 0.005, 100, and 16, respectively. All experiments were run on a server running Windows and using the GeForce GTX 1060 GPU. We selected the Keras library in Python 3.5 for network construction and training. For 100 iterations, the training time was 22–25 min for the CF-CNN, 18–20 min for the SRes-Net, and 3–4 min for the PVGG-Net.

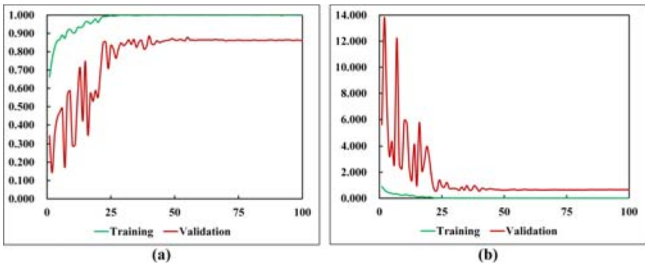


Fig. 9. Curves of accuracy (y-axis) and cross-entropy loss (y-axis) with the number of iterations (x-axis) for the SRes-Net. (a) Accuracy. (b) Cross-entropy loss.

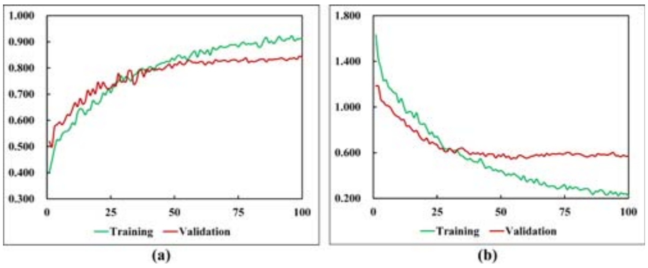


Fig. 10. Curves of accuracy (y-axis) and cross-entropy loss (y-axis) with the number of iterations (x-axis) for the PVGG-Net. (a) Accuracy. (b) Cross-entropy loss.

Figs. 9–11 show the curves of accuracy and cross-entropy loss during training process of three deep learning networks

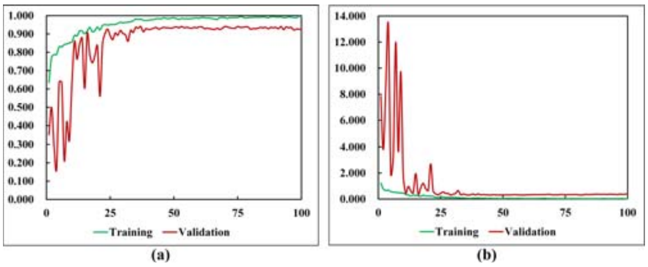


Fig. 11. Curves of accuracy (y-axis) and cross-entropy loss (y-axis) with the number of iterations (x-axis) for the CF-CNN. (a) Accuracy. (b) Cross-entropy loss.

TABLE I

CONFUSION MATRICES FOR THE TESTING DATA SET VIA DIFFERENT CLASSIFICATION STRATEGIES

PVGG-Net	**Pub**	**Ind**	**Gre**	**Com**	**Res**	**Urv**	**Vac**
Pub	43	2	0	2	1	1	1
Ind	2	45	3	0	0	0	0
Gre	0	1	47	0	2	0	0
Com	0	1	0	45	4	0	0
Res	0	1	1	1	44	2	1
Urv	0	0	0	1	0	49	0
Vac	0	1	7	0	0	0	42
	Overall accuracy=0.900			**Kappa=0.883**			
SRes-Net	**Pub**	**Ind**	**Gre**	**Com**	**Res**	**Urv**	**Vac**
Pub	38	2	2	2	3	0	3
Ind	0	48	2	0	0	0	0
Gre	0	0	50	0	0	0	0
Com	0	0	0	40	9	0	1
Res	0	1	0	3	46	0	0
Urv	1	0	0	0	0	49	0
Vac	0	0	0	0	0	0	50
	Overall accuracy=0.917			**Kappa=0.903**			
CF-CNN	**Pub**	**Ind**	**Gre**	**Com**	**Res**	**Urv**	**Vac**
Pub	43	0	0	3	3	0	1
Ind	0	49	1	0	0	0	0
Gre	0	0	49	1	0	0	0
Com	0	0	0	43	6	0	1
Res	2	0	0	0	47	0	1
Urv	0	0	0	0	1	49	0
Vac	0	0	0	0	0	0	50
	Overall accuracy=0.943			**Kappa=0.933**			

Types of land use: Pub = Public management services, Ind = Industrial zones, Gre = Green land, Com = Commercial zones, Res = Residential districts, Urv = Urban villages, Vac = Vacant land.

(SRes-Net, PVGG-Net, and CF-CNN). With the iterations of the models, the training accuracy of the SRes-Net and the CF-CNN approached 1.000, while the training accuracy of the PVGG-Net reached 0.923 after approximately 50 iterations. The validation accuracy of the three models also reached 0.940 (CF-CNN), 0.886 (SRes-Net), and 0.845 (PVGG-Net) among which the CF-CNN obtained the highest validation accuracy. All the curves eventually tended toward stability. Furthermore, by tracking the cross-entropy metric, the value of the training loss and validation loss rapidly decreased in the initial stage and converged after approximately 40 iterations. There was no obvious overfitting during the training process. To evaluate the effectiveness of the fused CNN features, the testing accuracies of the SRes-Net and PVGG-Net were compared with the testing accuracy of the CF-CNN. Table I illustrates the confusion matrices for the testing data set, including the overall accuracy and Kappa coefficient, under the different classification strategies. The proposed strategy of our framework achieved the highest classification performance, with an overall accuracy of 0.943 and a Kappa of 0.933. For comparison, the SRes-Net, being specialized in processing the HSR images, reached an overall accuracy of 0.917 and a Kappa of 0.903, while the PVGG-Net, being specialized in processing the RTUD data, reached an overall accuracy of 0.900 and a Kappa of 0.883.

As shown in Table I, several land use categories with typical features, such as industrial zones with scattered factories, green land, and vacant land with open space, were easier to identify using the SRes-Net. For other categories where the features were not easily distinguished, such as commercial zones, the higher classification accuracy was obtained using the PVGG-Net due to the distinguishable human activity features reflected in the RTUD data. However, there are also several land use categories where the classification performance is unstable, especially with numerous one-sample or two-sample misclassifications among a total of 50 testing samples. Public management services, likely including schools, hospitals, or research institutes, have a complex and diverse internal land structure, leading to the easy misclassification of one or two samples into other categories using either the SRes-Net or PVGG-Net [13]. The classification performance of residential districts and vacant land achieved by the PVGG-Net is relatively poor. The classification results of residential districts also produced many one-sample misclassifications, which indicate that the population density among various residential districts is different [65]. Vacant land is similar to green land because it has fewer human activities, resulting in the misclassification of some testing samples. Conversely, commercial zones have easily distinguishable human activity features, leading to a better classification performance, which is hard to discriminate using the HSR images due to the similar appearance between commercial zones and residential districts [20].

We further presented some examples of the classification results based on the feature of single-source data versus the fused features of multisource data in Fig. 12. For regions with image features, such as A and B, which are hard to distinguish, the range and trend of their RTUD curves match those of public management services and commercial zones, respectively, in Fig. 4. In this case, the PVGG-Net can avoid the misclassifications produced by the SRes-Net. However, for regions with distinguishable image features, the SRes-Net can obtain an accurate classification result merely based

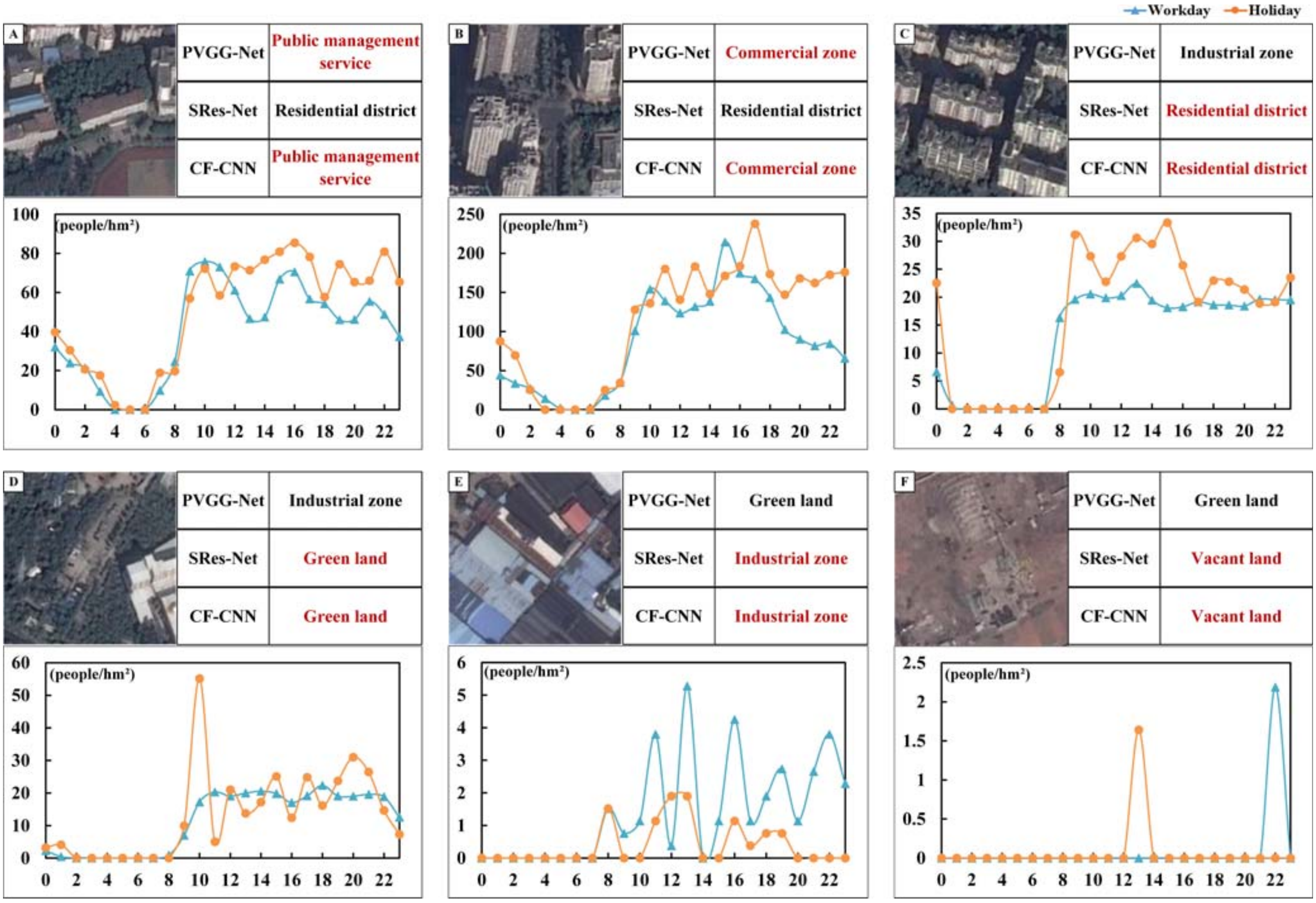


Fig. 12. Examples of the land use classification results. PVGG-Net denotes the 1-D network applied for classifying the RTUD data. SRes-Net denotes the simplified residual network applied for classifying the HSR images. CF-CNN denotes the framework proposed in this article. The ground truth of examples. (a) Public management services. (b) Commercial zones. (c) Industrial zones. (d) Green land. (e) Residential districts. (f) Vacant land. Red font indicates that the classification result of this method is correct.

on image features. For example, although both the RTUD curves of C and D have relatively distinct features that are similar to that of the industrial zone, the vast differences from the RTUD curves of their actual land use categories produced misclassifications. Besides, the RTUD curves of other misclassified regions produced by the PVGG-Net, such as E and F, are chaotic, reflecting a sparse population that is likely caused by the quality of the data and the collection time. In a word, the misclassified examples using single-source data-based methods are caused by various reasons, but all examples could be correctly identified after integrating the features by the CF-CNN. These results demonstrate that the integration of multisource data can reduce the error rate of the model and produce a more reliable land use map.

B. Mixed Land Composition

To estimate the proportions of mixed land use in our study area, we used a rectangular window to scan the entire area and split it into blocks. The fixed size of the scanning windows was set to 128 × 128, identical to the sampling window, and adjacent image blocks overlapped by 64 pixels to avoid the loss of spatial information [39]. We obtained approximately 55 000 image blocks by scanning the study area. The category of each image block was identified by the CF-CNN. The street-block data were used to divide the study area into land parcels, which were the basic processing units of Guangzhou. We, therefore, counted the proportions of the land use categories based on the land parcels.

Fig. 13 illustrates the proportions of mixed land use categories at the street-block level. The distribution of each land use category is spatially heterogeneous. As the study area is located in the central urban area of Guangzhou, the proportion of vacant land is the lowest. The proportion of green land is also relatively small and mainly distributed in the northern part of Tianhe district and the southern part of Haizhu district, where Furnace Forest Park and Haizhu Wetland Park are located. The places for people to live in always have a high demand in cities, so residential districts account for the highest proportion and are evenly distributed in the study area. Close to the edges of the suburbs, the main components are industrial zones and green land, as shown in Fig. 13(b) and (c). The city has not yet expanded into these areas so that the proportion of residential districts is relatively small. Most street blocks with a high proportion of public management services are located in the northern part of Tianhe and Haizhu districts, where

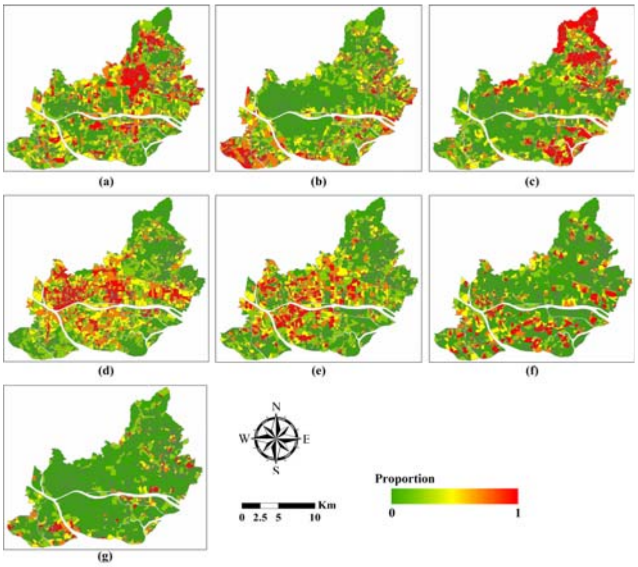


Fig. 13. Proportions of urban mixed land use categories in the study area. (a) Public management services. (b) Industrial zones. (c) Green land. (d) Commercial zones. (e) Residential districts. (f) Urban villages. (g) Vacant land.

there are many universities, such as Sun Yat-sen University, Guangzhou, China, South China University of Technology, Guangzhou, and Jinan University, Guangzhou. Due to the rapid development and numerous urban renewal projects in Guangzhou, many factories have been relocated to the suburbs to enhance the quality of life in the city center. Consequently, the street blocks with a large proportion of industrial zones are now distributed at the edge of the city. The problems of urban villages are the same as those of factories, with ongoing demolition and renovation, which are also the reason why the spatial distribution of urban villages is similar to that of industrial zones. Conversely, commercial zones are mainly distributed in the center of the study area, especially in Tianhe and Yuexiu districts. Tianhe district has the largest central business district in Guangzhou, while Yuexiu district has two famous commercial pedestrian streets, attracting a large number of people every weekend.

C. Measurement of Mixed Land Use Based on the SHDI

The proportion of land use is only a preliminary result, and it is impossible to intuitively express the mixed degree of each land parcel. To quantitatively illustrate the urban mixed land use pattern, we embedded the proportions of various land use types into the SHDI [11], [49]. Fig. 14 shows the spatial distribution of the SHDI at the street-block level. We conducted a Spearman correlation analysis between the SHDIs and the area of land parcels to assist in analyzing the distribution pattern of mixed land use. The Spearman correlation coefficient ($r = 0.476$, $p < 0.01$) indicates that the SHDI has a moderate positive correlation with the area of land parcels. To be specific, some relatively large land parcels are more likely to have a high SHDI due to their broad extent. Conversely, the land parcels with relatively small areas easily have a low SHDI. Most of the low-SHDI-valued land parcels with a small area are densely distributed in the mountains and parks of the study area, while other parcels are densely distributed in the city center, thus exhibiting agglomerated distributions. The results reveal that land parcels with low SHDI values do not usually represent undeveloped regions. The newly developed regions through urban renewal projects might be upscale residential districts or modern commercial plazas with low SHDIs.

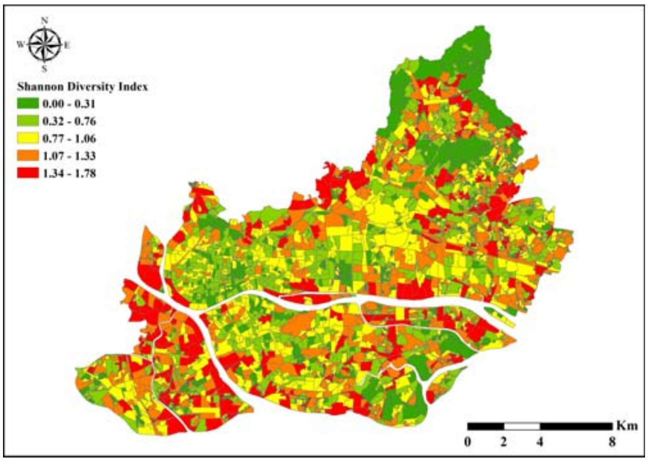


Fig. 14. Spatial distribution of the SHDI.

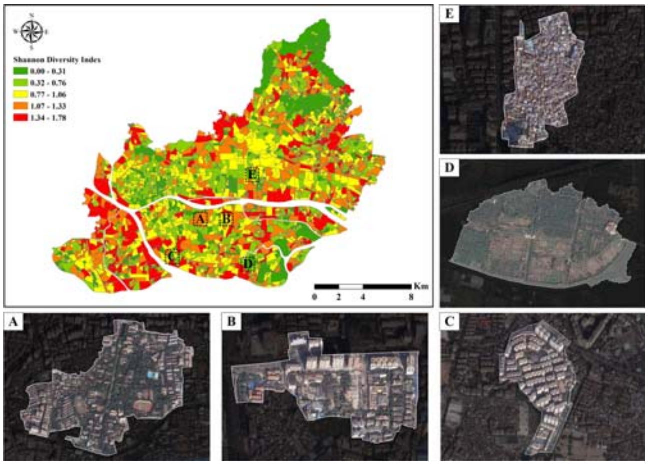


Fig. 15. Locations and satellite images of the sample zones. (a) Sun Yat-sen University. (b) Kecun. (c) Poly Garden. (d) Guangdong Haizhu National Wetland Park. (e) Shipai Community.

We selected several typical land parcels with high- and low-SHDI values for further comparison to more intuitively analyze the mixed land use patterns. As shown in Fig. 15, we selected land parcels A and B with high-SHDI values. The value of land parcel A is 1.126, which is higher than the value of 1.006 for land parcel B. We also selected land parcels C (0.224), D (0.165), and E (0.000), which have low SHDI values. Land parcel A is the main campus of Sun Yat-sen University. Although the highest proportion of land use is public management services, some areas inside land parcel A are used for living, dining, and shopping. Land parcel B is

TABLE II
SPEARMAN CORRELATION ANALYSIS FOR THE COMMUNITY DISTANCE AND NEIGHBORHOOD VIBRANCY

Variable		**Community Distance**	**Shannon Diversity Index**
Community Distance	Spearman Correlation	1	-0.103**
	Sig.(2-tailed)		0.000
	N	0	2931
		Neighborhood Vibrancy	**Shannon Diversity Index**
Neighborhood Vibrancy	Spearman Correlation	1	0.315**
	Sig.(2-tailed)		0.000
	N	0	2931

**Correlation is significant at the 0.01 level (2-tailed test)

Kecun in Haizhu district, a popular commercial zone that was redeveloped in the urban village. The parcel contains a small proportion of residential districts and urban villages, which is the main cause of the high SHDI. Unlike land parcels A and B, the SHDIs of land parcels C, D, and E are less than 0.300. The low SHDIs means that the parcels are mainly composed of one type of land use category, and the proportions of other types of land use category are very small. The main compositions of land parcels C, D, and E are residential buildings, green land, and urban villages, respectively. The previously mentioned comparison of these cases indicates that the SHDI value based on the type of urban land use within the land parcels is reasonable.

D. Correlation Analysis

According to previous studies [6], [10], the mixture of urban land use can place living, working, and dining close together, thus shortening the community distance and enhancing the neighborhood vibrancy. To validate the effectiveness of the mixture obtained by the CF-CNN, we conducted a Spearman correlation analysis among the SHDI, community distance, and neighborhood vibrancy. The mobile phone data were employed to extract the dynamic trajectory of human activity. A total of 6 177 063 travel trajectories for one week were obtained in the study area. In each land parcel, the accumulated population and average traveled distance were calculated according to the origin and destination point of each trajectory; then, we regarded these values as the neighborhood vibrancy and community distance, respectively. The results of the Spearman correlation analysis are shown in Table II. It can be seen that the SHDI has a weak negative correlation with community distance ($r = -0.103$, $p < 0.01$) but a moderate positive correlation with neighborhood vibrancy ($r = 0.315$, $p < 0.01$), which indirectly verifies the effectiveness of the SHDI according to the empirical phenomenon. Although the relatively small r-value indicates that mixed land use is not the only factor affecting these two urban metrics, the bias of mobile phone data might affect the result of the Spearman correlation analysis.

VI. Conclusion and Discussion

Mixed land use is favored by urban planners due to its positive impacts on urban habitability. Therefore, it is necessary to accurately identify the mixed land use areas. With the availability of emerging data sources, this article first depicted urban mixed land use at the street-block level by integrating HSR images and geospatial big data (represented by the RTUD data). We adopted two CNN architectures, the SRes-Net and PVGG-Net, to extract the features from the HSR images and RTUD data, respectively. Finally, these two features were concatenated to generate the fused CNN feature in the CF-CNN. Compared with conventional methods using single-source data, the proposed CF-CNN had the highest classification performance (overall accuracy = 0.943, Kappa = 0.933). The Kappa coefficient increased by 0.030 and 0.050, which proved that the fused feature from multisource data is more beneficial to land use classification tasks.

To assess the degree of urban mixed land use, we calculated the SHDI using the proportion of each land use type at the street-block level. The spatial distribution of the SHDI revealed the complexity of the current urban structure in the study area. Some of the undeveloped marginal regions with a large parcel size were likely to have a high SHDI, while regions distributed in the city center with emerging commercial zones or residential districts were likely to have a low SHDI. In addition, the SHDI was utilized to analyze the Spearman correlations with community distance and neighborhood vibrancy. The Spearman correlation analysis confirmed empirical evidence that mixed land use can reduce the community distance and promote the neighborhood vibrancy, which also demonstrated the effectiveness of the mixed land use obtained by the proposed framework.

Although our framework can characterize the phenomenon of urban mixed land use, several factors might influence the performance. First, due to the accessibility of geospatial big data, this article selected only the RTUD data as a fine-scale proxy for human activities. Some other types of geospatial big data that could provide useful information about land use classification were not employed in our framework. Besides, the combination methods might also lead to different classification performance. Second, object-oriented classification for the HSR images always encounters the problem of poor segmentation. Previous studies have employed multiresolution segmentation methods, but the segmentation results were not ideal [66], [67]. Considering the integrity of a single land parcel, we applied the street-block data as a partitioning benchmark. However, as shown in some cases in Fig. 15, segmentation errors (such as wide roads) still existed at the street-block level, which would affect our mapping results.

Therefore, we can add other types of data into our framework to address more complicated situations in the future. The RTUD data mainly reflect the attributes of human time-series behaviors. Other types of geospatial big data, such as points of interest and Weibo check-in data, can also provide information for classification from other aspects. In addition, a more efficient and intelligent segmentation method should be considered to reliably split land parcels. Simultaneously, a minimum area criterion of the segmentation should also be applied.

References

[1] N. A. Nabil and G. E. A. Eldayem, "Influence of mixed land-use on realizing the social capital," *HBRC J.*, vol. 11, no. 2, pp. 285–298, Aug. 2015.

[2] L. Wu, X. Cheng, C. Kang, D. Zhu, Z. Huang, and Y. Liu, "A framework for mixed-use decomposition based on temporal activity signatures extracted from big geo-data," *Int. J. Digit. Earth*, vol. 13, no. 6, pp. 708–726, Dec. 2018.

[3] S. Abdullahi, B. Pradhan, S. Mansor, and A. R. M. Shariff, "GIS-based modeling for the spatial measurement and evaluation of mixed land use development for a compact city," *GIScI. Remote Sens.*, vol. 52, no. 1, pp. 18–39, Jan. 2015.

[4] R. Cervero, "Mixed land-uses and commuting: Evidence from the American housing survey," *Transp. Res. A, Policy Pract.*, vol. 30, no. 5, pp. 361–377, Sep. 1996.

[5] M. Stevenson *et al.*, "Land use, transport, and population health: Estimating the health benefits of compact cities," *Lancet*, vol. 388, no. 10062, pp. 2925–2935, Dec. 2016.

[6] Y. Yue, Y. Zhuang, A. G. O. Yeh, J.-Y. Xie, C.-L. Ma, and Q.-Q. Li, "Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy," *Int. J. Geograph. Inf. Sci.*, vol. 31, no. 4, pp. 658–675, Apr. 2017.

[7] H.-O. Kwon and S.-D. Choi, "Polycyclic aromatic hydrocarbons (PAHs) in soils from a multi-industrial city, South Korea," *Sci. Total Environ.*, vols. 470–471, pp. 1494–1501, Feb. 2014.

[8] J. C. Seong *et al.*, "Modeling of road traffic noise and estimated human exposure in Fulton County, Georgia, USA," *Environ. Int.*, vol. 37, no. 8, pp. 1336–1341, Nov. 2011.

[9] B. R. Sperry, M. W. Burris, and E. Dumbaugh, "A case study of induced trips at mixed-use developments," *Environ. Planning B, Planning Des.*, vol. 39, no. 4, pp. 698–712, Aug. 2012.

[10] L. Tian, Y. Liang, and B. Zhang, "Measuring residential and industrial land use mix in the peri-urban areas of China," *Land Use Policy*, vol. 69, pp. 427–438, Dec. 2017.

[11] Y. Zhuo, H. Zheng, C. Wu, Z. Xu, G. Li, and Z. Yu, "Compatibility mix degree index: A novel measure to characterize urban land use mix pattern," *Comput., Environ. Urban Syst.*, vol. 75, pp. 49–60, May 2019.

[12] F. Hu, G.-S. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sens.*, vol. 7, no. 11, pp. 14680–14707, Nov. 2015.

[13] 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, Sep. 2018.

[14] H. Xing, Y. Meng, and Y. Shi, "A dynamic human activity-driven model for mixed land use evaluation using social media data," *Trans. GIS*, vol. 22, no. 5, pp. 1130–1151, Sep. 2018.

[15] 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, Nov. 2015.

[16] X. Deng *et al.*, "Geospatial big data: New paradigm of remote sensing applications," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 10, pp. 3841–3851, Oct. 2019.

[17] 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, Nov. 2019, Art. no. 101374.

[18] W. Tu *et al.*, "Portraying urban functional zones by coupling remote sensing imagery and human sensing data," *Remote Sens.*, vol. 10, no. 1, p. 141, Jan. 2018.

[19] X. Liu *et al.*, "Classifying urban land use by integrating remote sensing and social media data," *Int. J. Geograph. Inf. Sci.*, vol. 31, no. 8, pp. 1675–1696, Aug. 2017.

[20] 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, May 2019.

[21] M. Reichstein *et al.*, "Deep learning and process understanding for data-driven Earth system science," *Nature*, vol. 566, no. 7743, p. 195, Feb. 2019.

[22] X. Huang, Z. Li, C. Wang, and H. Ning, "Identifying disaster related social media for rapid response: A visual-textual fused CNN architecture," '*Int. J. Digit. Earth*, vol. 13, no. 9, pp. 1017–1039, Jun. 2019.

[23] J. Li, H. Zhi, J. Plaza, S. Li, and L. Yu, "Social media: New perspectives to improve remote sensing for emergency response," *Proc. IEEE*, vol. 105, no. 10, pp. 1900–1912, Oct. 2017.

[24] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 1520–1528.

[25] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, vol. 353, no. 6301, pp. 790–794, Aug. 2016.

[26] Y. Yao, J. Zhang, Y. Hong, H. Liang, and J. He, "Mapping fine-scale urban housing prices by fusing remotely sensed imagery and social media data," *Trans. GIS*, vol. 22, no. 2, pp. 561–581, Mar. 2018.

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

[28] D. Lu and Q. Weng, "Use of impervious surface in urban land-use classification," *Remote Sens. Environ.*, vol. 102, nos. 1–2, pp. 146–160, May 2006.

[29] Q. Man, P. Dong, and H. Guo, "Pixel-and feature-level fusion of hyperspectral and lidar data for urban land-use classification," *Int. J. Remote Sens.*, vol. 36, no. 6, pp. 1618–1644, Mar. 2015.

[30] T. Van De Voorde, W. Jacquet, and F. Canters, "Mapping form and function in urban areas: An approach based on urban metrics and continuous impervious surface data," *Landscape Urban Planning*, vol. 102, no. 3, pp. 143–155, Sep. 2011.

[31] X. Huang, H. Liu, and L. Zhang, "Spatiotemporal detection and analysis of urban villages in mega city regions of China using high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3639–3657, Jul. 2015.

[32] Q. Wang, X. He, and X. Li, "Locality and structure regularized low rank representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 911–923, Feb. 2019.

[33] Q. Zhu, Y. Zhong, S. Wu, L. Zhang, and D. Li, "Scene classification based on the sparse homogeneous–heterogeneous topic feature model," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2689–2703, May 2018.

[34] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.

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

[36] M. Mahdianpari, B. Salehi, M. Rezaee, F. Mohammadimanesh, and Y. Zhang, "Very deep convolutional neural networks for complex land cover mapping using multispectral remote sensing imagery," *Remote Sens.*, vol. 10, no. 7, p. 1119, Jul. 2018.

[37] W. Song, S. Li, L. Fang, and T. Lu, "Hyperspectral image classification with deep feature fusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3173–3184, Jun. 2018.

[38] Y. Zhong, X. Han, and L. Zhang, "Multi-class geospatial object detection based on a position-sensitive balancing framework for high spatial resolution remote sensing imagery," *ISPRS J. Photogramm. Remote Sens.*, vol. 138, pp. 281–294, Apr. 2018.

[39] Y. Zhong, Q. Zhu, and L. Zhang, "Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6207–6222, Nov. 2015.

[40] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geograph. Inf. Syst.*, New York, NY, USA, 2010, pp. 270–279.

[41] B. Zhao, Y. Zhong, G.-S. Xia, and L. Zhang, "Dirichlet-derived multiple topic scene classification model for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 2108–2123, Apr. 2016.

[42] G.-S. Xia *et al.*, "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3965–3981, Jul. 2017.

[43] Q. Shi *et al.*, "Domain adaption for fine-grained urban village extraction from satellite images," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1430–1434, Aug. 2020.

[44] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*. [Online]. Available: http://arxiv.org/abs/1706.05587

[45] Q. Wang, S. Liu, J. Chanussot, and X. Li, "Scene classification with recurrent attention of VHR remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1155–1167, Feb. 2019.

[46] Y. Yao *et al.*, "Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model," *Int. J. Geograph. Inf. Sci.*, vol. 31, no. 4, pp. 825–848, Apr. 2017.

[47] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and POIs," in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, London, U.K., 2012, pp. 186–194.

[48] T. Pei, S. Sobolevsky, C. Ratti, S.-L. Shaw, T. Li, and C. Zhou, "A new insight into land use classification based on aggregated mobile phone data," *Int. J. Geograph. Inf. Sci.*, vol. 28, no. 9, pp. 1988–2007, May 2014.

[49] Y. Chen *et al.*, "Delineating urban functional areas with building-level social media data: A dynamic time warping (DTW) distance based k-medoids method," *Landscape Urban Planning*, vol. 160, pp. 48–60, Apr. 2017.

[50] Y. Liang, S. Ke, J. Zhang, X. Yi, and Y. Zheng, "GeoMAN: Multi-level attention networks for geo-sensory time series prediction," in *Proc. 27th Int. Joint Conf. Artif. Intell. (IJCAI)*, Stockholm, Swede, Jul. 2018, pp. 3428–3434.

[51] J. Zhang *et al.*, "The Traj2 Vec model to quantify residents' spatial trajectories and estimate the proportions of urban land-use types," *Int. J. Geograph. Inf. Sci.*, pp. 1–19, Feb. 2020.

[52] J. Liu, T. Li, P. Xie, S. Du, F. Teng, and X. Yang, "Urban big data fusion based on deep learning: An overview," *Inf. Fusion*, vol. 53, pp. 123–133, Jan. 2020.

[53] Y. Zhang, Q. Li, H. Huang, W. Wu, X. Du, and H. Wang, "The combined use of remote sensing and social sensing data in fine-grained urban land use mapping: A case study in Beijing, China," *Remote Sens.*, vol. 9, no. 9, p. 865, Aug. 2017.

[54] S. Srivastava, J. E. Vargas-Muñoz, and D. Tuia, "Understanding urban landuse from the above and ground perspectives: A deep learning, multimodal solution," *Remote Sens. Environ.*, vol. 228, pp. 129–143, Jul. 2019.

[55] X. Liang, X. Liu, D. Li, H. Zhao, and G. Chen, "Urban growth simulation by incorporating planning policies into a CA-based future land-use simulation model," *Int. J. Geograph. Inf. Sci.*, vol. 32, no. 11, pp. 2294–2316, Aug. 2018.

[56] N. Niu *et al.*, "Integrating multi-source big data to infer building functions," *Int. J. Geograph. Inf. Sci.*, vol. 31, no. 9, pp. 1871–1890, May 2017.

[57] W. Li, H. Fu, L. Yu, and A. Cracknell, "Deep learning based oil palm tree detection and counting for high-resolution remote sensing images," *Remote Sens.*, vol. 9, no. 1, p. 22, Dec. 2016.

[58] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[59] P. Liu *et al.*, "Building footprint extraction from high-resolution images via spatial residual inception convolutional neural network," *Remote Sens.*, vol. 11, no. 7, p. 830, Apr. 2019.

[60] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[61] Y. Xu, J. Cheng, L. Wang, H. Xia, F. Liu, and D. Tao, "Ensemble one-dimensional convolution neural networks for skeleton-based action recognition," *IEEE Signal Process. Lett.*, vol. 25, no. 7, pp. 1044–1048, Jul. 2018.

[62] A. van den Oord *et al.*, "WaveNet: A generative model for raw audio," 2016, *arXiv:1609.03499*. [Online]. Available: http://arxiv.org/abs/1609.03499

[63] J. He, X. Li, Y. Yao, Y. Hong, and Z. Jinbao, "Mining transition rules of cellular automata for simulating urban expansion by using the deep learning techniques," *Int. J. Geograph. Inf. Sci.*, vol. 32, no. 10, pp. 2076–2097, Jun. 2018.

[64] Y. Song, L. Merlin, and D. Rodriguez, "Comparing measures of urban land use mix," *Comput., Environ. Urban Syst.*, vol. 42, pp. 1–13, Nov. 2013.

[65] Y. Yao *et al.*, "Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data," *Int. J. Geograph. Inf. Sci.*, vol. 31, no. 6, pp. 1220–1244, Jun. 2017.

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

[67] X. X. Zhu *et al.*, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

Jialyu He received the B.S. degree in geographic information system from Sun Yat-sen University, Guangzhou, China, in 2016, where he is pursuing the Ph.D. degree in cartography and geographic information system with the School of Geography and Planning.

His research interests include scene classification, land use simulation, and urban computing.

Xia Li is a Professor with the School of Geographic Sciences, East China Normal University, Shanghai, China. He has developed GeoSOS and FLUS models, which have been widely used for urban and land use simulation. He has authored more than 200 academic articles, of which many appeared in top international journals, such as *Nature Communications*, *Nature Sustainability*, *Remote Sensing of Environment*, the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and *International Journal of Geographical Information Science*. His major research interests include land-use change analysis, urban simulation and spatial optimization, and global land use modeling and the impact analysis.

Dr. Li is on the editorial boards of the international journals of *International Journal of Geographical Information Science* and *Computers, Environment and Urban Systems*.

Penghua Liu received the B.S. degree in geographic information science and the M.S. degree in cartography and geography information system from Sun Yat-sen University, Guangzhou, China, in 2017 and 2020, respectively.

He is working with the Alibaba Group and Ant Group, Hangzhou, China. His research interests include machine learning and deep learning in geospatial big data mining and intelligent understanding of remote sensing images.

Xinxin Wu received the M.S. degree in geographic information science from Sun Yat-sen University, Guangzhou, China, in 2020, where she is pursuing the Ph.D. degree in cartography and geography information system.

Her research interests include multisource spatial data mining and urban structure understanding.

Jinbao Zhang received the B.S. degree in geographic information science from Sun Yat-sen University, Guangzhou, China, in 2017, where he is pursuing the Ph.D. degree in cartography and geographic information system.

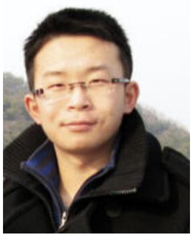
His research interests include machine learning and deep learning in geospatial data mining and intelligent understanding of remote sensing images.

Dachuan Zhang received the B.S. degree in remote sensing science and technology from the College of Information Engineering, China University of Geosciences, Wuhan, China, in 2015. He is pursuing the Ph.D. degree in cartography and geography information system with Sun Yat-sen University, Guangzhou, China.

He is adept at using machine learning algorithms in geoanalysis and geosimulations. His main research interests comprise multisource big data mining and land use simulations in urban planning.

Xiaojuan Liu received the M.S. degree in geographic information science from Sun Yat-sen University, Guangzhou, China, in 2020. She is pursuing the Ph.D. degree in cartography and geographic information systems with East China Normal University, Shanghai, China.

Her specific interests are machine learning and remote sensing techniques in ecosystem services and biodiversity conservation.



Yao Yao received the B.S. degree in surveying and mapping engineering and the M.S. degree in geodesy and surveying engineering from Wuhan University, Wuhan, China, in 2008 and 2011, respectively, and the Ph.D. degree in cartography and geographic information system from Sun Yat-sen University, Guangzhou, China, in 2017.

He is an Associate Professor with the School of Geography and Information Engineering, China University of Geosciences, Wuhan. His research interests include the application of geospatial big data and urban computing.