

Research Article

Delineating Mixed Urban “Jobs-Housing” Patterns at a Fine Scale by Using High Spatial Resolution Remote-Sensing Imagery

Yao Yao,^{1,2} Chen Qian,³ Ye Hong,⁴ Qingfeng Guan¹, Jingmin Chen¹, Liangyang Dai,¹ Zhangwei Jiang,² and Xun Liang¹

¹School of Geography and Information Engineering, China University of Geosciences, Wuhan 430078, Hubei, China

²Department of Data Technology and Products, Alibaba Group, Hangzhou 311121, Zhejiang, China

³Department of Engineering Systems and Environment, University of Virginia, Charlottesville, VA 22904, USA

⁴Institute of Cartography and Geoinformation, ETH Zurich, Stefano-Francini-Platz 5, CH-8093 Zürich, Switzerland

Correspondence should be addressed to Qingfeng Guan; guanqf@cug.edu.cn and Jingmin Chen; jingmin.cjm@alibaba-inc.com

Received 15 February 2020; Revised 15 March 2020; Accepted 20 March 2020; Published 14 April 2020

Guest Editor: Jun Yang

Copyright © 2020 Yao Yao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The spatial distribution pattern of jobs and housing plays a vital role in urban planning and traffic construction. However, obtaining the jobs-housing distribution at a fine scale (e.g., the perspective of individual jobs-housing attribute) presents difficulties due to a lack of social media data and useful models. With user data acquired from a location-based service provider in China, this study employs a deep bag-of-features network (BagNet) to classify remote-sensing (RS) images into various jobs-housing types. Considering Wuhan, one of the fastest developing cities in China, as a case study area, three jobs-housing types (i.e., only working, only living, and both working and living) at the land-parcel level are obtained. We demonstrate that the multiscale random sampling method can reduce the influence of image noise, increase the utilization of training data, and reduce network overfitting. By altering the network structure and the training strategy, BagNet achieved excellent fitting accuracy for identifying each jobs-housing type (overall accuracy > 0.84 and kappa > 0.8). For the first time, we demonstrate that urban socioeconomic characteristics can be obtained from high-resolution RS images using deep learning techniques. Additionally, we conclude that the total level of mixing within Wuhan is not high at present; however, Wuhan is continuously improving the mixture of jobs and housing. This study has reference value for extracting urban socioeconomic characteristics from RS images and could be used in urban planning as well as government management.

1. Introduction

Since the reform and opening up of China’s housing system, a large number of urban residents have chosen to purchase newly built commercial houses [1–3], which has caused the collapse of the urban jobs-housing space structure in the planned economy period [4]. The freedom of residence facilitated by the housing system reform has made the separation of mixed jobs-housing patterns increasingly common in China [5].

Many studies have shown that the separation of the urban jobs-housing spatial structure is conducive to the effective concentration of businesses and has the advantages of an agglomeration economy [6–9]. However, this separation has also produced many urban issues, such as

excessive commuting time, transportation costs, and increased environmental burden [10–16]. The increasingly severe urban traffic congestion and environmental degradation have become urgent problems. Therefore, a study of urban mixed jobs-housing patterns can explain the internal residential space structure and provide a reference for understanding urban complexity and optimizing urban spatial layouts.

Some scholars have researched the jobs-housing spatial structure based on census or household interview data [4, 17, 18]. With household interviews, Wang and Chai [4] conducted a study of changes in the jobs-housing relationship and the traditional unit housing system in Beijing. Zhou et al. [18] investigated changes in the jobs-housing space and commuting structure in Xi’an with sampling

survey data. The spatial resolution of these studies employs the basic unit of the census, such as administrative districts and streets. Thus, the generated results cannot reflect the distribution of jobs and housing at a fine scale, which means mixed jobs-housing patterns cannot be distinguished due to a lack of social media data from the perspective of individual attributes.

Because of the popularity of location-based services (LBS), many spatiotemporal data sets have evolved. These data sets record the trajectory of human activities and can be applied to describe and understand the urban jobs-housing space structure [19, 20]. For example, a city smart card system with location information is considered to be an effective way to analyze the personal data required for city commuting, and it has already been extensively employed to explain a city's jobs-housing space structure and commuting trajectory [21]. In addition to the city smart card system, signaling data from a mobile phone are also used in urban jobs-housing research [22]. These studies show that it is possible to employ LBS data to understand the jobs-housing space structure; however, a useful model is still lacking to make full use of these data to figure out the mixed jobs-housing pattern.

In recent years, scholars have begun to apply deep learning (DL) models to remote-sensing (RS) images for the extraction of economic activity characteristics. In RS images, the object scale variation can lead to weak feature representation for some scenes, influencing the classification results. To solve the problem of multiscale effects, Zhong et al. [23] combined a multiscale random sampling method with the large patch convolutional neural network (LPCNN) model for land use classification and obtained highly accurate land use classification results. Jean et al. [24] proposed an economic situation simulation model in an impoverished African country with data deficits using convolutional neural networks (CNNs), revealing that economic activity features can be extracted from RS images and applied to describe economic situations.

DL models can effectively mine social media data. Yao et al. [25] employed the word2vec model for extracting features in point-of-interest (POI) data. Incorporating the random forest model, they explored the spatial distribution of urban land use. Compared with the spatial density of RS image data, which covers an entire urban space, the distribution of social media data is sparse [26], which raises the question of what spatial scale should be considered when utilizing the data to analyze socioeconomic phenomena.

Studies also exist on extracting geometric information from multiple sources of data. Zhang and Du [27] found that information on urban scenes can be obtained automatically from high-resolution RS images. Liu et al. [28] introduced a probabilistic model to integrate multisource and geospatial big data to characterize urban mixed-use buildings. Song et al. [29] combined POIs with social properties and very high spatial resolution RS imagery with natural attributes to identify urban functions. Chen et al. and Shi et al. [30, 31] found that CNNs could be applied to extract building geometric information from high-resolution RS images. Nevertheless, these urban structure studies cannot explain

urban land use at a very fine scale and thus cannot delineate mixed jobs-housing patterns.

However, the above studies cannot present the urban structure from the perspective of individual attributes, especially in obtaining jobs-housing distribution because of the lack of media data and suitable models. In order to solve this problem, based on these studies, we conclude that the social and economic features of a city can be obtained from social media data and high-resolution RS images via DL models. This study addresses the question of whether this data combination can further explore urban mixed jobs-housing patterns. That is, do high-resolution RS images indicate high-level socioeconomic characteristics that reflect a mixed jobs-housing pattern?

We introduce a DL model to perform detailed simulations of the urban mixed jobs-housing pattern. The overall accuracy (OA) and kappa coefficient are used to evaluate the reliability of the model, and several case areas are chosen to analyze its reasonability. A comparison is conducted between the proposed method and several state-of-the-art DL models. As a result, false-color RGB images and the entropy index (EI) are used to visualize the global fitting map of jobs-housing types (only working, only living, and both working and living) in Wuhan, China. In this study, we explore whether there is a relationship between high-resolution RS images and mixed jobs-housing patterns.

2. Study Area and Data

The study area, Wuhan (Figure 1), the provincial capital of Hubei Province, is located in central China, and it has a total area of 8,494.41 square kilometers and 11.081 million residents. In 2018, the regional GDP was 1,448.729 billion Yuan (<http://www.wh.gov.cn/>). The downtown area of Wuhan includes Jiang'an, Jianghan, Qiaokou, Hanyang, Wuchang, Qingshan, and Hongshan. The GDP of Jianghan, Wuchang, and Jiang'an exceeds 100 billion Yuan.

The most critical data employed in this study comprise LBS user data and geographic location information. The data set is provided by one of the largest Internet companies in China, with a maximum user penetration rate of more than 85%. In large cities, such as Beijing, Shanghai, Guangzhou, Shenzhen, and Wuhan, the penetration rate exceeded 90%. This data set contains the trajectory information of approximately 500,000 random anonymous users within three months (2018.3.1–2018.5.30) in 2626 communities of Wuhan and represents the working and living locations of these users.

The trajectory information is obtained from anonymous users who have granted permission of collecting Global Positioning System (GPS) data during the process of using the LBS application. The main area of human activities during the 3 months is located, with a buffer zone set as 500 meters in width [22, 32]. We focus on occupants in an age range of 18 to 65 years old (excluding students, freelancers, and retirees). The geographic location information was obtained from the user data set, which is shown in Table 1. Particular attention should be paid to the student population, as in China, the permanent home for students is

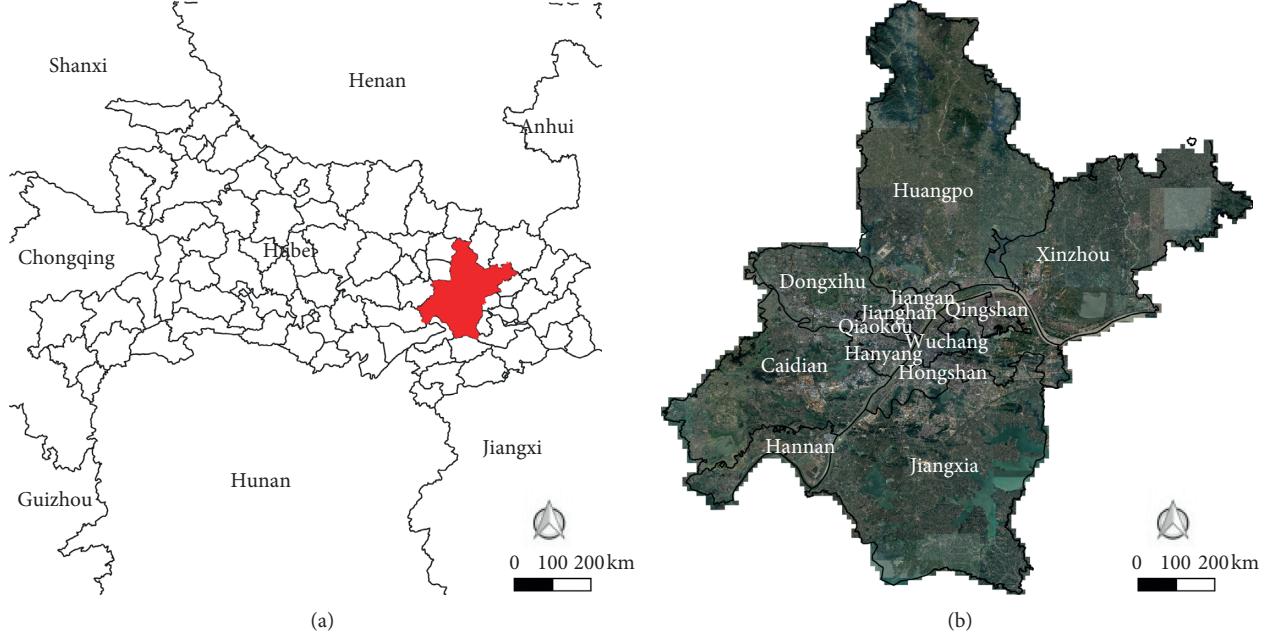


FIGURE 1: Case study area: (b) Wuhan, which is located in (a) Hubei Province.

TABLE 1: The user data, including geographic location information and three different types of jobs-housing attribute populations. OW: only working, OL: only living, and WL: working and living.

FID	Name	Lat	Lng	WL	OL	OW
0	Guangrongfang	30.5927	114.2824	0.1306	0.8110	0.0584
1	West Peach Blossom Island City Garden	30.5500	114.2175	0.0035	0.9965	0.0000
2	Renxin Olin Court	30.5013	114.1585	0.0536	0.9356	0.0108
3	Wuhan Construction School	30.5692	114.2127	0.0071	0.9857	0.0072
4	Changyang Lijiang City	30.5549	114.2925	0.0800	0.8589	0.0611
5	Wuhan Sports School	30.5731	114.2130	0.0838	0.8682	0.0480
:	:	:	:	:	:	:

at school, which leads to the result that 90% of the population of schools has the attribute of only living at the school location. Hence, to analyze the distribution of faculty, the student population is excluded from the data set.

The user data set estimates the proportion of three types of jobs-housing attribute populations at the land-parcel scale. This data set includes the rates of the attributes only working (OW), only living (OL), and both working and living (WL) together with the location information. In the original data, the sum of OW, OL, and WL residents within each parcel is not equal to 1. The potential reason is that some users do not open the function of GPS during the process of using the LBS application. We transform these three rates of jobs-housing attributes to make the sum of them equals to 1, which could reveal the relationship between three attributes and be convenient for future calculation and visualization. WL residents represent people who work and live in a parcel with a buffer zone set to 500 meters, including self-employed residents, operators of small-scale companies, and industrial park dormitory workers. The distributions of three rates of jobs-housing attributes are shown in Figure 2.

We also collect the land-parcel data of Wuhan from Gaode Maps, which is one of the largest map service companies in China, as the basis for the division of land parcels in the study area. The urban area of Wuhan contains 8,257 land parcels with an average area of 100 square meters per parcel. All of these land parcels are urban functional zones that do not include vegetation, water bodies, soils, or roads. The spatial distribution of land parcel is shown in Figure 3.

Figure 1 shows an RS image downloaded from Google Earth at the level of 16, which contains RGB bands with a size of $32,512 \times 32,768$. According to the research of Yao et al. [25], the shadows in the background have little influence on extracting functional zones from Google Earth RS imagery and the slight date difference between geospatial data and RS image is tolerable on integration. User data could be related to the RS image via geographic location information. Based on the latitude and longitude of each land parcel, we cut the RS image into small pieces (estimated as 100×100 pixels) and obtained the RS image data set. The RS images of three typical jobs-housing types in Wuhan are shown in Figure 4.

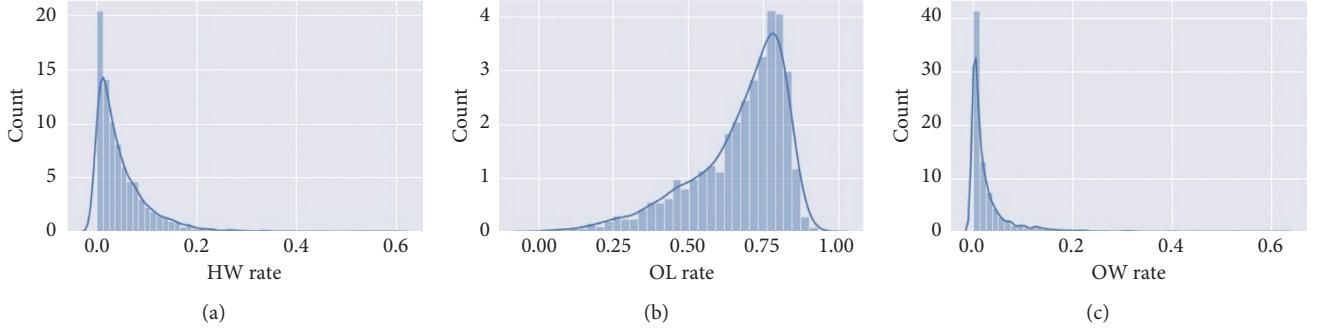


FIGURE 2: Distributions of three rates of jobs-housing attributes. (a) Distribution of HW rate. (b) Distribution of OL rate. (c) Distribution of OW rate.

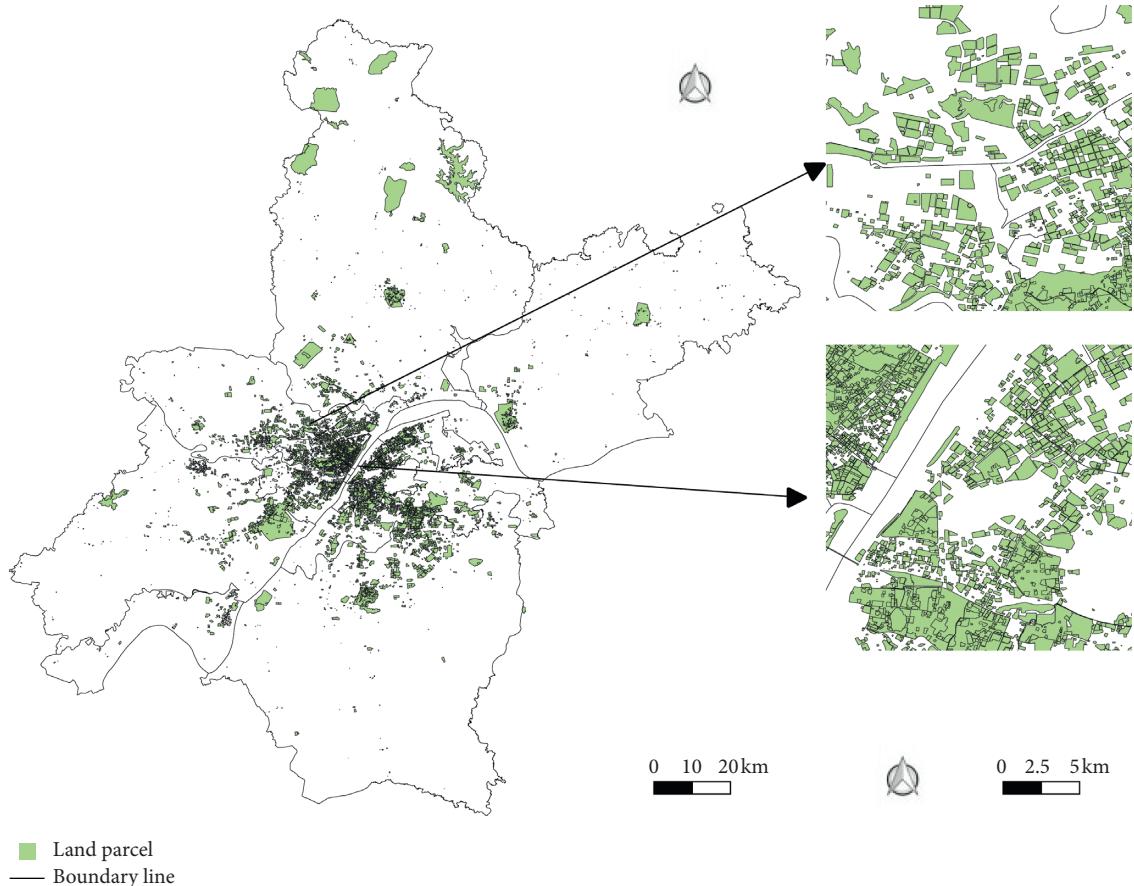


FIGURE 3: Spatial distribution of land parcels in Wuhan (the green shapes represent land parcels).

3. Methodology

The workflow of this study, as shown in Figure 5, can be summarized as follows: (1) The user data set is classified to form the multiscale spatial data set by a multiscale sampling method after data preprocessing. (2) A state-of-the-art DL model is adopted for identifying the relationship between the spatial data set and the mixed jobs-housing index. (3) With the trained model, three distributions of mixed jobs-housing types within Wuhan were estimated and evaluated by the OA, kappa coefficient, and entropy index.

3.1. Data Preprocessing. In this study, because the sum of three jobs-housing attribute rates in the original data is not equal to 1, we cannot apply one model to predict three attributes. Classification is an application of CNN; therefore, we choose to build three CNN models for WL, OL, and OW separately; however, the user data in this study are a continuous data set and need to be discretized [24]. Additionally, the user data are transformed to a normal distribution by oversampling [33]. Referring to the discretization process of Yao et al. [34], this study calculates the mean μ and standard deviation σ of WL, OL, and OW. The



FIGURE 4: Three types of typical areas in Wuhan: typical OL areas include (a) Wuhan Ship Vocational and Technical College ($OL = 0.8707$), (b) Jianghan University College of Arts and Sciences ($OL = 0.9654$), (c) Jin He Yuan ($OL = 0.9530$), and (d) the Oriental Garden ($OL = 0.9966$). Typical OW areas include (e) Lingxiuzhongnan ($OW = 0.7433$), (f) Shangheju ($OW = 0.5106$), (g) Hubei Post Apartments ($OW = 0.8717$), and (h) the Geological Yard ($OW = 0.8607$). Typical WL areas include (i) the Hubu Lane Community ($WL = 0.6642$), (j) Wuhan Institute of Engineering Science and Technology ($WL = 0.6568$), (k) Meidi Employee Living Quarters ($WL = 0.7034$), and (l) Guoyao Apartments ($WL = 0.6487$).

user data are discretized in the range $[\mu - 3\sigma, \mu + 3\sigma]$, with steps of 0.5σ .

Based on the studies of Ren et al. [35], the surrounding environments have a potential influence on the function of a community. As described above, the user data is obtained by setting a buffer zone at 500 meters. This could consider the influence of surrounding environments. In this study, we need to make segmentations on the land parcels which could contain the communities and the surrounding environments.

Studies have shown that multiscale problems may occur in RS images due to variations in the resolution of the RS images [36, 37]. In this study, although the multiscale problem is not a large issue, it still influences fitting accuracy [23]. The small and imbalanced sample size of the original RS image data set in this study may induce overfitting issues during the process of training. Therefore, we employ the multiscale random sampling method proposed by Zhong et al. [23]. The method is as follows: first, according to the user data, the latitude and longitude can be obtained to locate the parcel of the user data on the RS image; second, the length of the sampling window is set to W (W is set as the size that ensures the parcel of communities can be completely covered); then, based on the parcel of user data, a certain number of samples are randomly considered with

length s ($0.75 W \leq s \leq W$). Sampling each parcel in this way ensures that a sufficient number of multiscale spatial data sets are obtained. Figure 6 shows an example of multiscale random sampling, which is automatically obtained from the RS images based on location information. This study combines oversampling and multiscale random sampling to build the training data set.

3.2. Mixed Jobs-Housing Pattern Extraction Based on BagNet. BagNet is a CNN model conducted by Brendel and Bethge [38] that can fully utilize each part of an image and obtain complete information about the image. The structure of BagNet is shown in Figure 7. BagNet intercepts an input image with specific pixel dimensions and then uses a 1×1 convolutional layer on each image block to obtain a class vector. All class vectors of image blocks are summed together, and predictions are made based on the most significant class vector.

The BagNet structure differs from the structure of traditional CNN models, which always uses an entire image to calculate a class vector without obtaining the summation. In this study, superfluous information exists in the borders of an image, which does not contain a parcel of communities. But

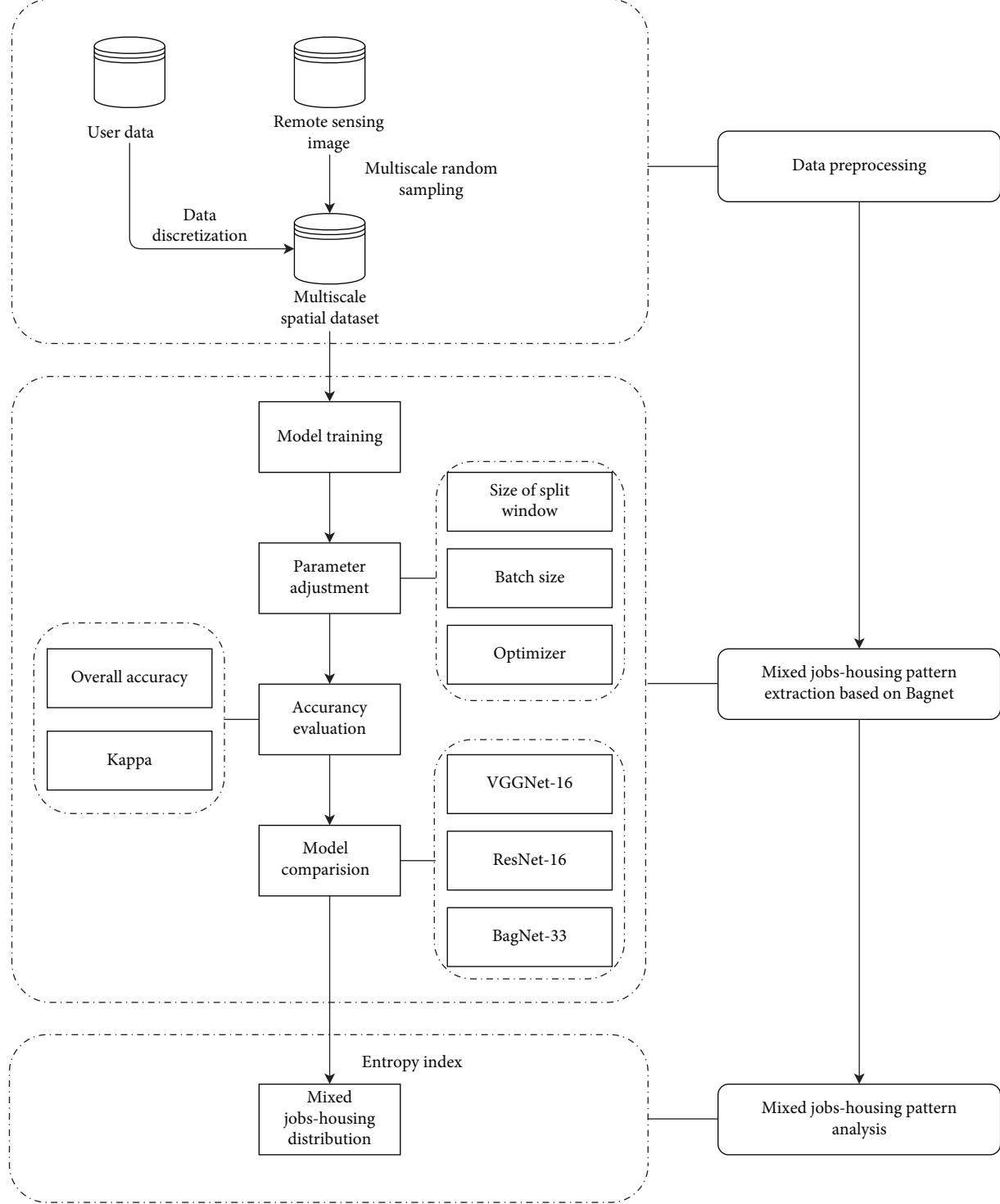


FIGURE 5: The workflow of jobs-housing mixed analysis based on BagNet.

the borders cannot be ignored because the surrounding environments may have a potential influence on the function of communities. Since BagNet can analyze how each part of an image affects classification, it can fully utilize the complete information about an image. Even though several parts in the borders of an image may predict a wrong result, the majority of parts in the center of an image could obtain sufficient

information to make a correct classification. The final result of the vote depends on the majority of parts; therefore, BagNet can learn useful spatial information about LBS user data and place less emphasis on useless information.

The study selected the classic VGGNet [39] and ResNet [40] models to compare with the BagNet model. VGGNet is a CNN model developed by researchers from the Visual

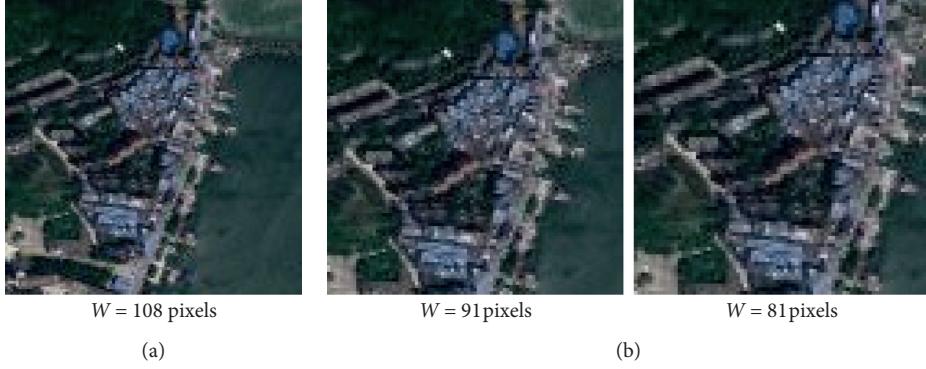


FIGURE 6: An example of multiscale random sampling. (a) Original image, (b) Multiple random samples.

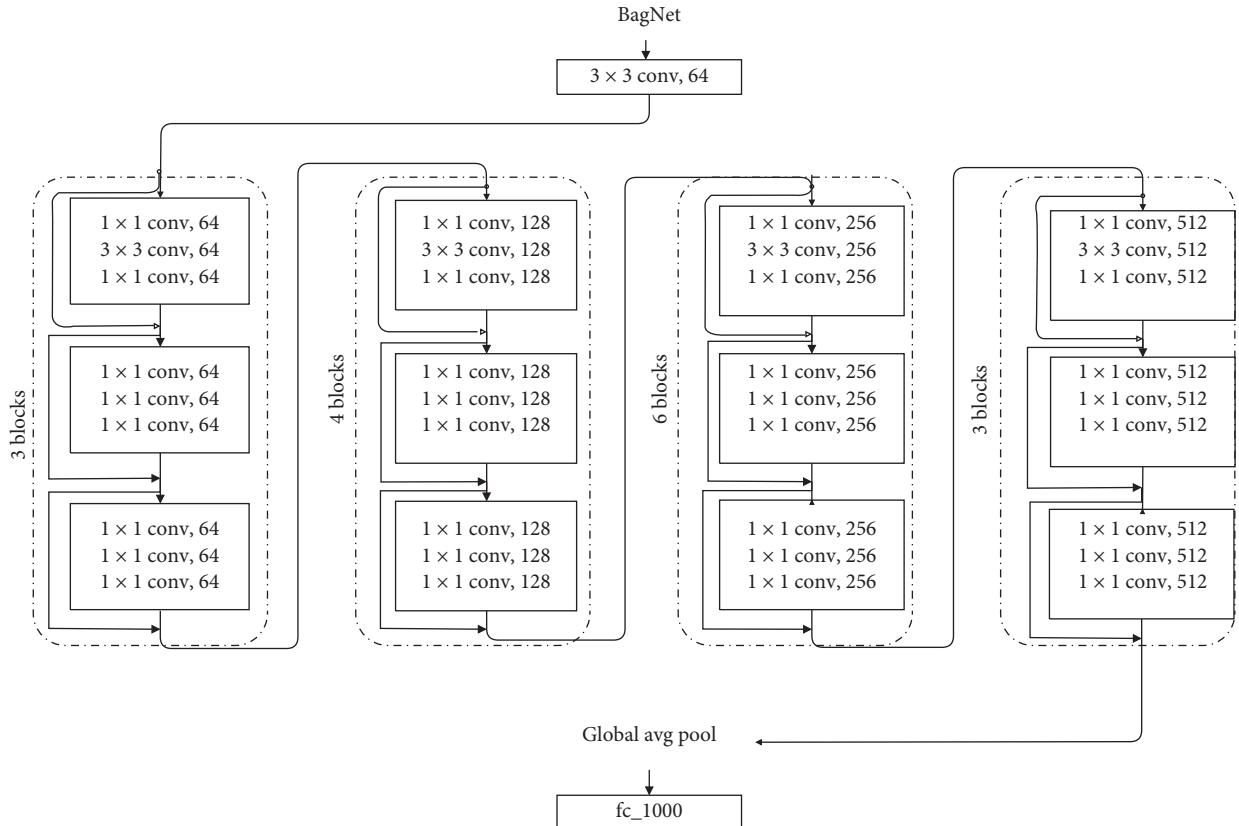


FIGURE 7: The BagNet architecture is almost equivalent to the ResNet architecture except for a few changes in the strides and the replacement of most 3×3 convolutions with 1×1 convolutions (the dotted arrow shows the direction of the training process) [38].

Geometry Group, University of Oxford, and Google DeepMind. VGGNet explores the relationship between the depth of CNNs and their performance by repeatedly using 3×3 convolution kernels and a 2×2 max-pooling layer, and it successfully constructs a 16- to 19-layer CNN [39]. ResNet is built by a residual block and can effectively solve the problem of gradient disappearance [40].

The study employed a cross-entropy loss function, which is a common loss function on classification:

$$L = -x_{\text{label}} + \log \sum_{j=1}^N e^{x_j}. \quad (1)$$

In equation (1), x represents the input category, the label is the index value of the actual category, and N represents the number of categories. Based on the control variable method, this study adjusts the segmentation window, batch size, and optimizer and obtains the best-performing model. By comparing the results of the BagNet model and other DL models, the effectiveness of BagNet in this experiment was verified.

3.3. Accuracy Evaluation and Urban Mixed Functional Pattern Analysis. In the evaluation of RS image classification, a confusion matrix (Table 2) is usually applied to determine

TABLE 2: Confusion matrix.

		Prediction		
		Category 1	...	Category <i>i</i>
Actuality	Category 1	X_{11}	...	X_{1i}
	:	:	:	:
	Category <i>i</i>	X_{i1}	...	X_{ii}

the accuracy and reliability of the classification [41]. In this study, the classification results were evaluated by using the overall accuracy (OA) and kappa coefficient.

The OA is expressed as the percentage of the total number of correct predictions, that is, the sum of all values of the diagonal elements in the confusion matrix, divided by the total of all samples (equation fd2(2)). In the 1960s, Fleiss et al. [42] proposed the kappa indicator as an indicator of the extent to which the classification results outperform a random classification (equation fd3(3)). Kappa falls between 0 and 1, and a higher kappa value indicates better classification results.

$$OA = \frac{N \sum_{i=1}^n X_{ii}}{N}, \quad (2)$$

$$Kappa = \frac{N \sum_{i=1}^n X_{ii} - \sum_{i=1}^n (X_{i+} * X_{+i})}{N^2 - \sum_{i=1}^n (X_{i+} * X_{+i})}. \quad (3)$$

In these equations, n is the category, N is the sum of the number of categories, X_{ii} is a diagonal element of the confusion matrix, X_{i+} is the sum of the columns of a category, and X_{+i} is the sum of the rows of a category.

This study refers to the calculation of the entropy index in the landscape pattern index [43, 44]. The entropy index is used to quantitatively measure the mixing degree of each parcel of user data. This value is calculated by equation (4), and the value of the mixed entropy falls in (0, 1). The higher the mixed entropy value, the higher the mixing degree of the land parcel.

$$EI = -\frac{\sum_{i=1}^n p_i \times \ln p_i}{\ln n}. \quad (4)$$

In this equation, n is the total number of categories, and p_i is the proportion of attributes in the lot.

4. Results

4.1. Parameter Sensitivity Analysis. In this study, the sampling window needs to be set to ensure that the parcel can be entirely contained, and the multiscale sampling method is used to obtain the spatial data in the sampling window. The data set is recorded as D and contains a total of 26,260 sets of data. This study randomly fetches 80% of the data and makes data augmentation by combining oversampling and multiscale random sampling as training data D_{TR} , 10% of the data as verification data D_V , and 10% of the data as test data D_{TE} . In the network training process, D_{TR} is used for training and the size of the training data set is around 200,000, D_V is used for fine-tuning the parameters, and D_{TE} is used to evaluate the final result.

As shown in Table 3, this step applied 9×9 , 17×7 , and 33×33 -pixel image blocks to train the BagNet model. SGD

TABLE 3: Accuracy comparison of different sizes of image blocks.

	WL		OL		OW	
	OA	Kappa	OA	Kappa	OA	Kappa
BagNet-9	0.4952	0.4045	0.4857	0.4141	0.5231	0.4455
BagNet-17	0.7942	0.7822	0.8071	0.7947	0.8076	0.7766
BagNet-33	0.8242	0.8022	0.8271	0.8047	0.8357	0.8092

was selected as the optimizer algorithm, the batch size was set to 32, and the learning rate was set to 0.01. According to these results, the larger the size of an image block, the better the accuracy results since a larger amount of information about the images is obtained.

As shown in Table 4, we set the batch size = 8, 16, and 32 to train the BagNet model. The image block was set to 33×33 pixels, SGD was selected as the optimizer algorithm, the learning rate was set to 0.01 and the dropout rate was set to 0.4. Properly setting the batch size decreases the use of computer memory and accelerates training. According to these results, setting the batch size to 16 can decrease the training time and maintain satisfactory results.

As shown in Table 5, we used SGD, Momentum, and Adam for training the BagNet model. The image block was set to 33×33 pixels, the batch size was set to 16, the learning rate was set to 0.01 and the dropout rate was set to 0.4. According to this result, the accuracy of SGD is better than that of Adam. After adding Momentum, SGD could obtain a better accuracy result than other optimizers.

4.2. Comparing with Several State-of-the-Art CNN-Based Models. The VGGNet, ResNet, and BagNet network models were selected as base models for the experiment. The training strategy was set to a dropout rate of 0.4, a batch size of 16, and a learning rate of 0.01, and SGD + Momentum was selected as the optimizer. After training, neither VGGNet nor ResNet converged, but the BagNet training converged. An analysis of the original image data set revealed that each image contained edge noise. VggNet used 3×3 convolution and 2×2 max-pooling throughout, which meant that every part of the image was involved in the training. However, the large amount of noise data on the edges of images interferes in the process of adjusting parameters in training. Although ResNet increased the depth of the network, it could not solve the problem of noise interference. When a large amount of interference information is confused with useful information, the training cannot converge.

BagNet does not consider the spatial sorting method of an image, which means that BagNet focuses on each part of an image instead of the overall image [38]. BagNet classifies images according to small local features of the images. The constraints on local features can directly determine how each part of the image affects classification, which enables the algorithm to fully utilize the total information of the image and reduce the weight of useless information obtained from noise data. This means BagNet could obtain useful information from the center and borders of an image and reduce the influence of the noises because the final result of the vote depends on the majority of parts which have made

TABLE 4: Accuracy comparison of different batch sizes.

	WL		OL		OW	
Batch size	OA	Kappa	OA	Kappa	OA	Kappa
8	0.8302	0.8202	0.8413	0.8212	0.8456	0.8207
16	0.8401	0.8201	0.8456	0.8208	0.8462	0.8227
32	0.8324	0.8116	0.8452	0.8201	0.8462	0.8226

TABLE 5: Accuracy comparison of different optimizers.

	WL		OL		OW	
Optimizer	OA	Kappa	OA	Kappa	OA	Kappa
SGD	0.8401	0.8201	0.8456	0.8208	0.8462	0.8227
SGD + Momentum	0.8482	0.8207	0.8468	0.8219	0.8472	0.8227
Adam	0.8200	0.8024	0.8324	0.8010	0.8205	0.7916

correct predictions. Table 6 shows a comparison of the three models.

4.3. Mixed Jobs-Housing Pattern in the Case Study Area. Based on the previous comparison and analysis, this study adopted an improved model of BagNet-33 for the experiment. The training strategy was set to a dropout rate of 0.4, a batch size of 16, and a learning rate of 0.01, and SGD + Momentum was selected as the optimizer. The loss function during the training process of BagNet is shown in Figure 8. After obtaining the classification of three jobs-housing attributions, we sum the product of means and probabilities of each category to estimate the fitting result. In this study, the global fitting results of three types of mixed attributes and typical plots using RGB synthesis were visualized [45]. The red band indicates WL, the green band indicates OL, and the blue band indicates OW. The spatial distribution of the resulting population composition ratio is shown in Figure 9. The entropy calculation was performed according to the WL, OL, and OW attributes of each parcel, and the distribution result is shown in Figure 10.

The average mixed jobs-housing entropy of Wuhan is 0.1982. Wuhan has a typical large central group structure, which indicates that more resources are focused on the central area for the development of the economy. The city has a center-focused developmental spatial structure [46]. The working attributes gradually weaken from the central area to the surrounding area, while the residential attributes strengthen and tend to slowly become mixed work and residential attributes.

The working centers of several remote urban areas, such as Caidian, Jiangxia, Huangpi, Xinzhou, Dongxihu, and Hannan, are located in the area close to the central city. In general, Wuhan's development is focused on the central urban area. Few areas exist with a single working or living attribute. The jobs-housing properties of most areas are mixed, which indicates that Wuhan is developing toward increasing the level of mixed land use [47].

This study selects three typical cases to prove the reliability of the analysis. Figures 9(A) and 10(A) are typical university education areas in Wuhan, including residential buildings and related living facilities, such as Nanwang Villa

TABLE 6: Accuracy comparison of different models.

	WL		OL		OW	
	OA	Kappa	OA	Kappa	OA	Kappa
VGGNet-16	0.2546	0.2342	0.2352	0.2141	0.2454	0.2257
ResNet-17	0.4657	0.4423	0.4843	0.4651	0.4742	0.4548
BagNet-33	0.8482	0.8207	0.8468	0.8219	0.8472	0.8227

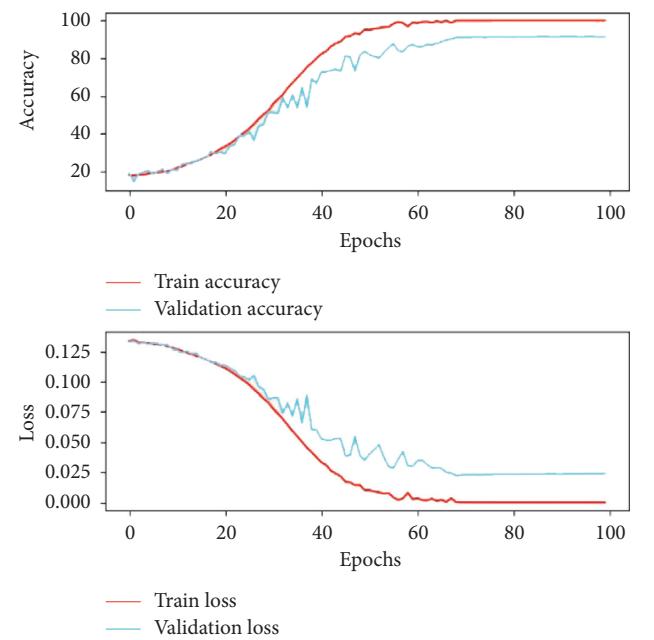


FIGURE 8: The loss function during the process of training BagNet model.

and the Sunshine Community. The residential area accounts for a relatively high proportion of residents; the OL attribute of the area is significantly higher than that of other regions, and the jobs-housing mixing level is relatively low ($WL = 0.0124$, $OL = 0.9781$, $OW = 0.0095$, $EI = 0.1095$).

Figures 9(B) and 10(B) show typical working areas that are famous scenic spots in Wuhan and compose a mixed administrative, medical, and cultural area. The OW attribute of this area is significantly higher than that of most other areas, and the jobs-housing mixing level is very high

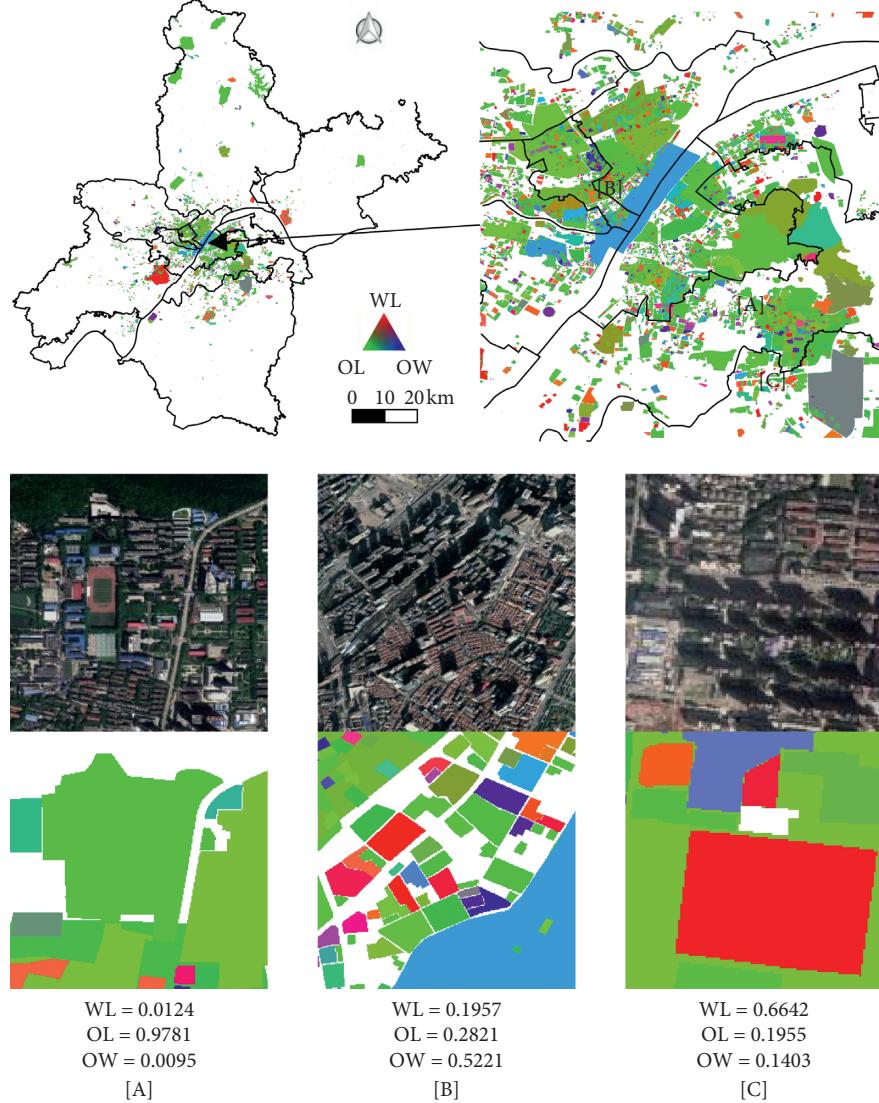


FIGURE 9: Fitting results of the jobs-housing model in Wuhan: (A) China University of Geosciences, (B) Hankou River Beach, and (C) Guanggu Software Park.

($WL = 0.1957$, $OL = 0.2821$, $OW = 0.5221$, $EI = 0.9243$). Figures 9(C) and 10(C) show a typical industrial park with an occupational residence, including a large number of software enterprises and staff quarters. Therefore, the region has higher WL attributes and a higher jobs-housing mixing level ($WL = 0.6642$, $OL = 0.1955$, $OW = 0.1403$, and $EI = 0.7886$). The case area analysis shows that the BagNet model can effectively and reasonably extract and quantify the urban mixed jobs-housing pattern.

5. Discussion

This study determines whether a correlation exists between RS images and an urban jobs-housing pattern on a relatively fine scale. We combine user data and RS image data and employ a multiscale random sampling method to address the multiscale issues in the image and the limited data problem. This study segments the land parcels containing the communities and surrounding environments

instead of using the border of communities as a sample, which could consider the potential influence of environments on the function of land parcels. This study introduces the BagNet model and adjusts the parameters, by selecting an appropriate segmentation window size and applying the dropout mechanism, which effectively improved the fitting accuracy. The BagNet-33 model in this study produced excellent fitting results, which indicated that the DL model can be effectively applied to the analysis of urban mixed land use.

The CNN derivative model BagNet is innovatively introduced to improve the accuracy of the results. This attempt was effective in applying the DL method in the mixed analysis of mixed urban jobs-housing patterns. Compared with VGGNet and ResNet, BagNet is more suitable for extracting socioeconomic information from RS images, namely, the spatial distribution of jobs-housing patterns.

This study identified a strong correlation between high-resolution RS images and urban jobs-housing patterns.

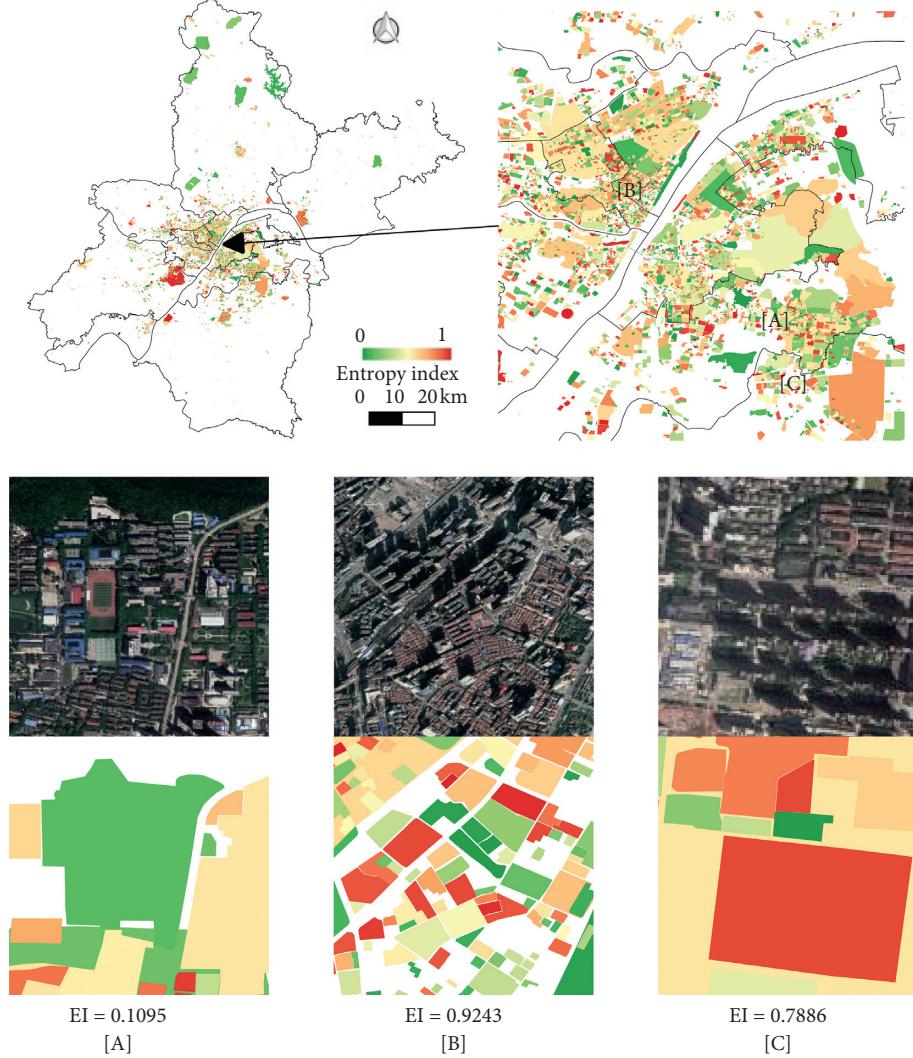


FIGURE 10: Mixed jobs-housing results for Wuhan: (A) China University of Geosciences, (B) Hankou River Beach, and (C) Guanggu Software Park.

Using DL to mine high-level semantic information in high-resolution RS image data, this study revealed a strong relationship between this semantic information and urban socioeconomic features. The mixed jobs-housing pattern was obtained with the constructed fitting model, showing that two different modes of observation, namely, “bottom-up” (social perception) [26] and “top-down” (satellite remote-sensing) [48], are effective in representing urban socioeconomic characteristics.

In this study, the fitting results of user jobs-housing data at the parcel scale in Wuhan were obtained from the BagNet model and analyzed with an entropy calculation. The jobs-housing mixing level in OL areas is low, while that in OW and WL areas is high. Moreover, the rationality of the fitting results was demonstrated in the case area analysis. As the level of mixed land use is closely related to urban development [47], further development at the level of mixed land use is needed to promote economic growth. Furthermore, this model could also be applied to analyze

the distribution of residents and the user portrait of communities, which would be helpful in urban planning and urban design.

Despite the strategic contribution to supporting urban development, this study has areas that can be improved. The data employed in this study comprise RS images and a user data set. Studies have suggested that urban socioeconomic information can be explored by coupling multisource social media data [49]. In the future, we could potentially improve the ability to infer mixed jobs-housing patterns by coupling multisource social media data. Also, during the training process, the effect of each parameter on the results is not quantifiable. The training strategy is designed based on experience and repeated experiments. Moreover, the computation time of training BagNet model is around 20 hours, and this study does not make a comparison on computation time between the proposed models because there is not a time requirement. Future works would focus on improving efficiency.

6. Conclusions

This study designed a DL model based on accurate mining of semantic information in high-resolution RS images, which reflected the mixed spatial distribution of a city. We determined that social perception data and RS images can be combined to reflect urban socioeconomic characteristics and further obtain a mixed jobs-housing pattern. Considering Wuhan as the study area, we show that the mixture level is relatively low. The government should plan and construct additional mixed functional areas to increase the level of mixed land use and to stimulate economic development. This study is conducive to understanding urban complexity and optimizing the urban spatial structure and could be used for urban planning and governmental management.

Data Availability

Data are available on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Yao Yao and Chen Qian contributed equally to this work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (41801306, 41671408, 41901332, 61773383); Fundamental Research Funds for National University, China University of Geosciences (Wuhan) (CUG190606); Open Fund of State Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University (18S01); National Key R&D Program of China (2017YFB0503804); the Natural Science Foundation of Hubei Province (2017CFA041).

References

- [1] Y. P. Wang and A. Murie, "Commercial housing development in urban China," *Urban Studies*, vol. 36, no. 9, pp. 1475–1494, 1999.
- [2] Y. P. Wang and A. Murie, "Social and spatial implications of housing reform in China," *International Journal of Urban and Regional Research*, vol. 24, no. 2, pp. 397–417, 2000.
- [3] F. Wu, "Residential relocation under market-oriented redevelopment: the process and outcomes in urban China," *Geoforum*, vol. 35, no. 4, pp. 453–470, 2004.
- [4] D. Wang and Y. Chai, "The jobs-housing relationship and commuting in Beijing, China: the legacy of Danwei," *Journal of Transport Geography*, vol. 17, no. 1, pp. 30–38, 2009.
- [5] N. Ta, Y. Chai, Y. Zhang, and D. Sun, "Understanding job-housing relationship and commuting pattern in Chinese cities: past, present and future," *Transportation Research Part D: Transport and Environment*, vol. 52, pp. 562–573, 2017.
- [6] R. Dekle and J. Eaton, "Agglomeration and land rents: evidence from the prefectures," *Journal of Urban Economics*, vol. 46, no. 2, pp. 200–214, 1999.
- [7] S. S. Rosenthal and W. C. Strange, "The determinants of agglomeration," *Journal of Urban Economics*, vol. 50, no. 2, pp. 191–229, 2001.
- [8] H. Wang, H. Schandl, G. Wang, L. Ma, and Y. Wang, "Regional material flow accounts for China: examining China's natural resource use at the provincial and national level," *Journal of Industrial Ecology*, vol. 23, no. 6, pp. 1425–1438, 2019.
- [9] H. Wang, G. Wang, J. Qi et al., "Scarcity-weighted fossil fuel footprint of China at the provincial level," *Applied Energy*, vol. 258, Article ID 114081, 2020.
- [10] M. Charron, "From excess commuting to commuting possibilities: more extension to the concept of excess commuting," *Environment and Planning A: Economy and Space*, vol. 39, no. 5, pp. 1238–1254, 2007.
- [11] A. Guo, J. Yang, X. Xiao, J. Xia, C. Jin, and X. Li, "Influences of urban spatial form on urban heat island effects at the community level in China," *Sustainable Cities and Society*, vol. 53, Article ID 101972, 2020.
- [12] Z. Qiao, L. Liu, Y. Qin, X. Xu, B. Wang, and Z. Liu, "The impact of urban renewal on land surface temperature changes: a case study in the main city of Guangzhou, China," *Remote Sensing*, vol. 12, no. 5, p. 794, 2020.
- [13] V. Van Acker and F. Witlox, "Commuting trips within tours: how is commuting related to land use?" *Transportation*, vol. 38, no. 3, pp. 465–486, 2011.
- [14] J. Yang, Y. Wang, X. Xiao, C. Jin, J. Xia, and X. Li, "Spatial differentiation of urban wind and thermal environment in different grid sizes," *Urban Climate*, vol. 28, Article ID 100458, 2019.
- [15] Z. Qiao, C. Wu, D. Zhao et al., "Determining the boundary and probability of surface urban heat island footprint based on a logistic model," *Remote Sensing*, vol. 11, no. 11, p. 1368, 2019.
- [16] J. Yang, S. Jin, X. Xiao et al., "Local climate zone ventilation and urban land surface temperatures: towards a performance-based and wind-sensitive planning proposal in megacities," *Sustainable Cities and Society*, vol. 47, Article ID 101487, 2019.
- [17] Y.-H. Tsai, "Quantifying urban form: compactness versus sprawl," *Urban Studies*, vol. 42, no. 1, pp. 141–161, 2005.
- [18] J. Zhou, C. Zhang, X. Chen, W. Huang, and P. Yu, "Has the legacy of Danwei persisted in transformations? the jobs-housing balance and commuting efficiency in Xi'an," *Journal of Transport Geography*, vol. 40, pp. 64–76, 2014.
- [19] R. Ahas and Ü. Mark, "Location based services-new challenges for planning and public administration?" *Futures*, vol. 37, no. 6, pp. 547–561, 2005.
- [20] Y. Long and J.-C. Thill, "Combining smart card data and household travel survey to analyze jobs-housing relationships in Beijing," *Computers, Environment and Urban Systems*, vol. 53, pp. 19–35, 2015.
- [21] L. Liu, A. Biderman, and C. Ratti, "Urban mobility landscape: real time monitoring of urban mobility patterns," in *Proceedings of the 11th International Conference on Computers in Urban Planning and Urban Management*, pp. 1–16, Hong Kong, China, June 2009.
- [22] N. Wan and G. Lin, "Life-space characterization from cellular telephone collected GPS data," *Computers, Environment and Urban Systems*, vol. 39, pp. 63–70, 2013.
- [23] Y. Zhong, F. Fei, and L. Zhang, "Large patch convolutional neural networks for the scene classification of high spatial resolution imagery," *Journal of Applied Remote Sensing*, vol. 10, pp. 1–20, 2016.

- [24] 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, 2016.
- [25] Y. Yao, X. Li, X. Liu et al., "Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model," *International Journal of Geographical Information Science*, vol. 31, no. 4, pp. 825–848, 2017.
- [26] Y. Liu, "Revisiting several basic geographical concepts: a social sensing perspective," *Acta Oceanologica Sinica*, vol. 4, 2016.
- [27] X. Zhang and S. Du, "A linear Dirichlet mixture model for decomposing scenes: application to analyzing urban functional zonings," *Remote Sensing of Environment*, vol. 169, pp. 37–49, 2015.
- [28] X. Liu, N. Niu, X. Liu et al., "Characterizing mixed-use buildings based on multi-source big data," *International Journal of Geographical Information Science*, vol. 32, pp. 738–756, 2018.
- [29] J. Song, T. Lin, X. Li, and A. V. Prishchepov, "Mapping urban functional zones by integrating very high spatial resolution remote sensing imagery and points of interest: a case study of Xiamen, China," *Remote Sensing*, vol. 10, no. 11, p. 1737, 2018.
- [30] Y. Chen, L. Tang, X. Yang, M. Bilal, and Q. Li, "Object-based multi-modal convolution neural networks for building extraction using panchromatic and multispectral imagery," *Neurocomputing*, In press, 2019.
- [31] Y. Shi, Q. Li, and X. X. Zhu, "Building segmentation through a gated graph convolutional neural network with deep structured feature embedding," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 159, pp. 184–197, 2020.
- [32] N. Wan and G. Lin, "Classifying human activity patterns from smartphone collected GPS data: a fuzzy classification and aggregation approach," *Transactions in GIS*, vol. 20, no. 6, pp. 869–886, 2016.
- [33] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, pp. 1263–1284, 2009.
- [34] Y. Yao, X. Liu, X. Li et al., "Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data," *International Journal of Geographical Information Science*, vol. 31, pp. 1220–1244, 2017.
- [35] M. Ren, Y. Lin, M. Jin, Z. Duan, Y. Gong, and Y. Liu, "Examining the effect of land-use function complementarity on intra-urban spatial interactions using metro smart card records," *Transportation*, pp. 1–23, 2019.
- [36] J. Qin and N. H. C. Yung, "Scene categorization via contextual visual words," *Pattern Recognition*, vol. 43, no. 5, pp. 1874–1888, 2010.
- [37] L. Zhou, Z. Zhou, and D. Hu, "Scene classification using a multi-resolution bag-of-features model," *Pattern Recognition*, vol. 46, no. 1, pp. 424–433, 2013.
- [38] W. Brendel and M. Bethge, "Approximating cnns with bag-of-local-features models works surprisingly well on imagenet," 2019, <http://arxiv.org/abs/1904.00760>.
- [39] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <http://arxiv.org/abs/1409.1556>.
- [40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, pp. 770–778, Las Vegas, NV, USA, June 2016.
- [41] R. G. Congalton, "A review of assessing the accuracy of classifications of remotely sensed data," *Remote Sensing of Environment*, vol. 37, no. 1, pp. 35–46, 1991.
- [42] J. L. Fleiss, J. Cohen, and B. S. Everitt, "Large sample standard errors of kappa and weighted kappa," *Psychological Bulletin*, vol. 72, no. 5, pp. 323–327, 1969.
- [43] S. G. Plexida, A. I. Sfougaris, I. P. Ispikoudis, and V. P. Papanastasis, "Selecting landscape metrics as indicators of spatial heterogeneity-a comparison among Greek landscapes," *International Journal of Applied Earth Observation and Geoinformation*, vol. 26, pp. 26–35, 2014.
- [44] M. G. Turner, R. H. Gardner, R. V. O'Neill, and R. V. O'Neill, *Landscape Ecology in Theory and Practice*, Vol. 401, Springer, Berlin, Germany, 2001.
- [45] K. Dovey and E. Pafka, "What is functional mix? an assemblage approach," *Planning Theory & Practice*, vol. 18, no. 2, pp. 249–267, 2017.
- [46] H. Huang, Y. Dennis Wei, and Dennis, "Intra-metropolitan location of foreign direct investment in Wuhan, China: institution, urban structure, and accessibility," *Applied Geography*, vol. 47, pp. 78–88, 2014.
- [47] Y. Liu, T. Luo, Z. Liu, X. Kong, J. Li, and R. Tan, "A comparative analysis of urban and rural construction land use change and driving forces: implications for urban-rural coordination development in Wuhan, Central China," *Habitat International*, vol. 47, pp. 113–125, 2015.
- [48] X. Liu, J. He, Y. Yao et al., "Classifying urban land use by integrating remote sensing and social media data," *International Journal of Geographical Information Science*, vol. 31, no. 8, pp. 1675–1696, 2017.
- [49] S. Jiang, X. Qian, T. Mei, and Y. Fu, "Personalized travel sequence recommendation on multi-source big social media," *IEEE Transactions on Big Data*, vol. 2, no. 1, pp. 43–56, 2016.