

Article

Multi-Scale Recursive Identification of Urban Functional Areas Based on Multi-Source Data

Ting Liu, Gang Cheng * and Jie Yang

College Surveying & Land Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China; 212104020087@home.hpu.edu.cn (T.L.); yangjie@hpu.edu.cn (J.Y.)

* Correspondence: chenggang@hpu.edu.cn

Abstract: The study of urban functional area identification is of great significance for urban function cognition, spatial planning, and economic development. In the identification of urban functional areas, most studies considered only a single data source and a single division scale, the research results have problems such as low update frequency or incomplete information in a single data set, and overfitting or underfitting in a single spatial resolution. Aiming at the above problems, this paper proposes a multi-scale recursive recognition method based on interactive validation for urban functional areas using taxi trajectory data and point of interest (POI) data as the main data sources. First, the dynamic time warping (DTW) algorithm generates a time series similarity matrix, a CA-RFM model combining the clustering algorithm and random forest model is constructed. The model extracts significant feature regions as inputs through a K-medoid clustering algorithm, which are imported into the random forest model for urban functional zone (UFZ) identification. Then, to overcome the shortcomings of a single scale in expressing urban structural characteristics, a recursive model of different levels of urban road networks is established to classify multi-scale functional areas. Finally, cross-validation using the CA-RFM model and POI quantitative identification method obtains the final identification results of urban functional areas. This paper selects Shenzhen as the study area, the results show that the combination of clustering algorithm and random forest model greatly reduces the error of manual selection of training samples. In addition, the study demonstrates the superiority of the proposed method in two aspects, namely, faster delineation and improved accuracy in urban functional area identification.

Keywords: urban functional zone; CA-RFM model; multi-scale recursive recognition; POI quantitative identification



Citation: Liu, T.; Cheng, G.; Yang, J. Multi-Scale Recursive Identification of Urban Functional Areas Based on Multi-Source Data. *Sustainability* **2023**, *15*, 13870. <https://doi.org/10.3390/su151813870>

Academic Editors: Antonio De Nicola, Maria Luisa Villani and Raquel Sanchis

Received: 6 August 2023

Revised: 7 September 2023

Accepted: 13 September 2023

Published: 18 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As urbanization continues and the urban area expands, the functional urban area type differs from what was envisioned in earlier plans [1]. Understanding changes in urban functional areas is essential for effective urban development planning, natural resource allocation, and ecosystem management [2]. However, accurately identifying an urban functional zone is challenging due to the complexity and comprehensiveness of urban functions [3]. Most traditional studies rely on existing land-use information, field surveys or thematic data on functional zoning, which lack objective tests and are time consuming and labor intensive [4–6]. In recent years, high-quality ultra-high-resolution remote sensing images have shown certain advantages in representing urban functional zones due to their large coverage, rich image information and wide availability [7]. However, satellite imagery can only monitor the physical characteristics of the urban surface, which is not sufficient to identify social functions and characterize the spatial and temporal patterns of human mobility. The emergence of massive urban big data creates new urban computing and analytics opportunities. The use of points of interest location check-in data, public transportation data and cellular signaling data produces better results for the dynamic

description of the city or functional zoning of the city, and provides a rich means of identifying functional zones of the city [8–11], and can fill the gaps in missing semantic information regarding the functional space of remote sensing images.

The development and formation of functional urban areas depend largely on how residents interact with their surroundings. However, the explanation of UFZ has varied in previous research. Berry identified UFZ as the interconnection within areas through the distribution of activities and by flows of commodities between zones [12]. Karlsson et al. identified the UFZ by measuring the economic activities and intra-regional transportation infrastructure that existed within the region, and the modes of transport of interconnection that existed between regions [13]. Yuan et al. defined the UFZ as the areas developed to meet specific socioeconomic needs [14]. Although previous studies have varied in their interpretations of urban functional areas, the researchers characterize UFZ by their zoning characteristics and activity characteristics. The zoning characteristics are used to define the zone boundaries, while the activity characteristics are used to identify the zone functions.

As mentioned above, different studies have defined UFZ differently. Among those studies, zoning plays a critical role in UFZ identification, which mainly focuses on how to partition the urban area into several spatial units where diverse socioeconomic activities take place. The spatial segmentation methods mainly include grid-based segmentation, road network-based segmentation and image-based segmentation [15]. The grid-based segmentation method provides more granular results, as the segmented spatial units are generally smaller than those from other segmentation methods, with grid cell resolution ranging from 30 m per pixel to 10 km per pixel [16–18]. The road network-based segmentation method mainly includes defining segment boundaries from city-designated transportation zones or mapped road segments [19–22]. The commonly used image segmentation techniques are based on the spatial distribution characteristics of the image objects and the homogeneity of the functional types to generate urban functional area units [23]. Roads, as boundaries of the area, carry the daily activities of the residents and have a high degree of integration with the city. However, when using road network data for regional segmentation, there is no regulation on which level of the road should be selected as the standard of urban functional zone division [24–26]. To effectively express the hierarchical semantic information of urban functional zones, this paper will explore the significance of multi-scale functional zone division based on road network data.

In recent years, machine learning algorithms have been widely used in identifying the zone functions, including support vector machine, k-nearest neighbor algorithm, naive bayes and random forest. Support vector machine is a pattern recognition method based on statistical learning theory, the purpose of which is to find a hyperplane that makes it possible to correctly separate two types of data points as much as possible, and many scholars have applied this algorithm in urban functional area recognition. Deng proposed a polygonal Voronoi diagram method to divide urban areas, generate fine spatial analysis units and categorize the themes of spatial analysis units by support vector machine algorithm [27]. The k-nearest neighbor classification algorithm is a simple algorithm in data mining classification counting, which mainly relies on limited adjacent samples as the basis of classification. Liu uses dynamic time warping based k-nearest neighbor classification algorithm to classify and identify urban functional areas and uses POI data to assist analysis to obtain the final functional layout of Chengdu [28]. Naive bayes algorithm is a simple and effective classification algorithm, which is based on bayes theorem and feature independence assumption. Lefulebe uses planescope images and naive Bayes algorithms to classify and detect changes in urban land use and land cover in cape town [29]. Random forest algorithm is an integrated learning algorithm based on the decision tree algorithm, which is trained on the dataset by using multiple decision trees at the same time and obtains the final prediction through a voting mechanism or averaging. Grippa identifies urban land-use classifications at the block level using OpenStreetMap data and random forest algorithms [30]. Yao extracted the high-dimensional feature vector of POIs through the World2 Vec model and trained the feature vector through the random forest algorithm to

obtain the urban functional zone with high classification accuracy. Compared with the traditional K-means algorithm, the effectiveness of the random forest algorithm for urban functional zoning is verified [31]. Random forest algorithm increases the diversity of classification trees and improves the performance of individual classification or regression trees by back sampling and randomly changing the combinations of predictor variables, and it has the advantages of low computational cost, high model performance, high robustness and low risk of overfitting when processing data. Among the above machine learning algorithms, the random forest algorithm is better able to handle multiple classification problems with functional areas of data from multiple sources. In the process of using machine learning algorithms, it is extremely important to establish training samples, and the quality of the training samples is directly related to the training performance of the model. However, from the view of previous studies, most of the studies produce training samples manually labeled manually, this method is time consuming and laborious, and the accuracy of the training samples is not guaranteed.

Taxi trajectory data are widely used in the study of urban functional areas. Qian proposed an integrated model for recognizing urban functional zones using satellite images and taxi trajectories [32]. Gao used taxi trajectory data and applied a Gaussian mixture model to classify the inflow and trip count characteristics of regions, and based on these typical characteristics, these urban regions were regrouped by using the Pearson correlation coefficient clustering method [33]. Liu et al. extracted time series from taxi trajectory data to categorize and identify urban functional zones, and with the aid of POI assisted analysis, the final functional layout of Chengdu was derived [28]. However, most time series mining studies have used direct clustering methods, i.e., unsupervised learning, and there may be many inaccuracies in the definition of categories in unsupervised learning due to the lack of accurate data labels for the training samples.

Therefore, aiming at the above problems, this paper adopts different levels of road data to divide the urban space and form a multi-scale road network. Dynamic taxi trajectory data and static POI data are fused together, and the experiments are conducted from the perspectives of “dynamic” and “static”, and the CA-RFM model, which combines the clustering algorithm and random forest classification, is used for data mining. This paper provides a new idea for the application of trajectory data mining in urban function identification.

2. Materials and Methods

2.1. Study Area

The study area is located in Shenzhen, China (Figure 1), including Futian District, Luohu District, Nanshan District, Baoan District, Longgang District, another nine administrative regions and Dapeng New District, and is China’s special economic zone, national economic center city and international city. The total area of the study area is 1996.78 square kilometers, and the resident population is 10.7789 million. As one of the national economic centers and international cities, Shenzhen has complex and diverse urban landscapes with rich urban functional zoning. Shenzhen’s urban spatial structure is very complex, and the distributions of urban functions are interlaced with each other, although the distribution of functional areas such as residential areas and commercial areas shows a certain regularity in general, the functions of many areas are not single, but mixed functional areas formed by the interaction of multiple functions.

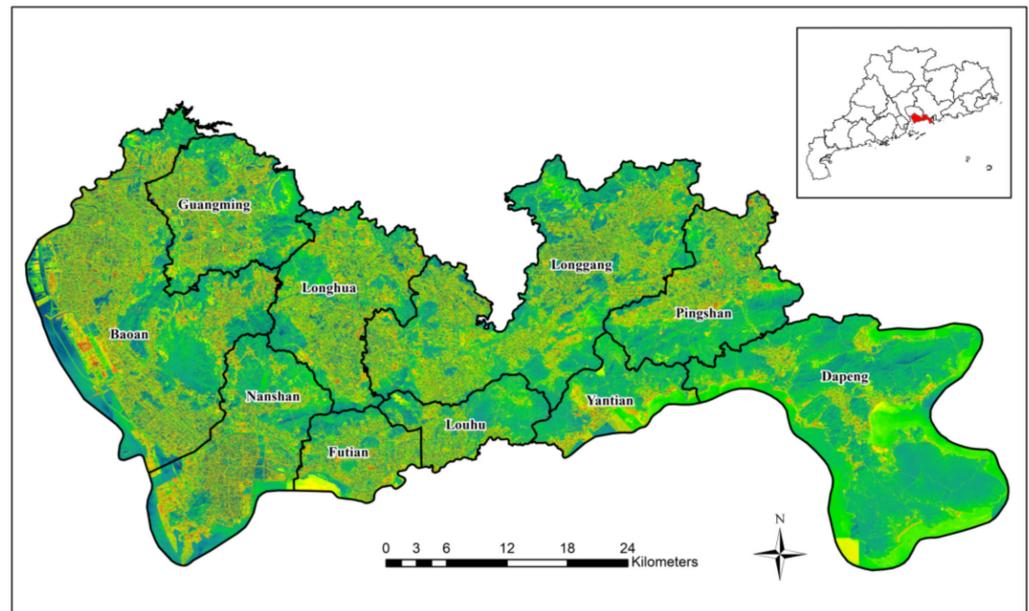


Figure 1. Study area.

2.2. Data and Processing

• Taxi Trajectory Data

As a kind of public transportation in the city, taxis attract passengers with their convenience and are readily available. To make the research results reflect the travel situation of residents on working days and rest days, this paper selects the whole week trajectory data from 7 December 2015 (Monday) to 13 December 2015 (Sunday) for analysis, with a total of 419,258,185 records. Each record contains vehicle number, recording time, latitude and longitude, speed, direction and vehicle status, and the data format is shown in Table 1 below.

Table 1. Taxi trajectory data.

| ID | Time | Lon | Lat | Speed | Direction | Status |
|--------|---------------|-----------|------------|-------|-----------|--------|
| C124E2 | 1,448,934,913 | 22.579636 | 114.132820 | 62 | 53 | 1 |
| C2AXHP | 1,448,951,588 | 22.577946 | 114.130936 | 48 | 28 | 1 |
| ... | ... | ... | ... | ... | ... | ... |
| C685AD | 1,449,016,722 | 22.594633 | 114.044900 | 56 | 109 | 1 |
| CAEDHP | 1,447,927,156 | 22.597000 | 114.040520 | 78 | 119 | 1 |
| CDTISQ | 1,443,498,723 | 22.599183 | 144.039636 | 39 | 215 | 1 |
| ... | ... | ... | ... | ... | ... | ... |

In the original taxi trajectory data, the start point (point O) and end point (point D) of each trajectory are extracted based on the unique identifier of the taxi and the chronological order of the trajectory. The OD data are extracted according to the taxi ID, passenger status and timestamp. The passenger status is 1, indicating that the passenger is loaded, and 0 indicates that there is no passenger, then the point in a continuous trajectory where the passenger status changes from 0 to 1 is the starting point and the point where the passenger status changes from 1 to 0 is the destination point. Group the records by vehicle ID and arrange each group in ascending order by time, extract the record whose passenger status becomes 1 as the departure point data and the record whose passenger status becomes 0 as the destination point data, and get the coordinates of the start and end points as well as the time of the start and end points of each order. The above process was repeated for all records within the date, and 3.86 million OD data were extracted from the 419 million GPS records.

- POI Data

The POI data are derived from the Gaode Map Open Platform, which records the types of activities of urban residents in a certain location. In this paper, we collected the 2018 point-of-interest data within Shenzhen, and the original POI data had a wide range of hierarchical classifications, with the major categories covering a variety of sub-categories, and there were problems of repetition and crossover in different classifications. Therefore, it is necessary to reclassify the POI data. According to the “urban land-use classification and Planning and construction land-use standards”, considering the types and attributes of urban functional areas, this study reclassified POIs into public management public service facilities land, commercial service facilities land, residential land, industrial land, green space and square land. The total number of cleaned POI data is 450,591 records, and the classification table is shown in Table 2.

Table 2. POI data.

| ID | The Primary Classification | The Secondary Classification |
|----|--|--|
| 1 | Land for public administration and public service facilities | Public Facilities, science education and culture, sports leisure, government agencies and social organizations, medical care, etc. |
| 2 | Commercial service facility land | Catering services, shopping services, financial services, accommodation services, life services, etc. |
| 3 | Residential land | Business housing, tenement buildings, etc. |
| 4 | Industrial land | incorporated business, agricultural and fishery base, etc. |
| 5 | Green space and square land | Scenic spots, park squares, etc. |

- Road Network Data

The road network data are obtained from the official website of OpenStreetMap. The irregular grid composed of road network data is the basic unit representing the socio-economic functions of urban management and planning. Different levels of road networks divide the whole city into different traffic analysis zones (TAZ). The road grades selected in this study include expressways, trunk roads, main roads, secondary roads, ordinary branches, residential roads, service roads, etc. To ensure data quality, operations such as removing overhanging points in roads, extending independent road lines to connect with adjacent roads, and finally removing unnecessary internal roads by hand are performed on the road network data.

2.3. Method

In this study, the trajectory data are transformed into a time feature sequence, and the information is mined to achieve the purpose of identifying functional areas. The workflow of urban main functional area identification is shown in Figure 2, including the following three steps. Firstly, the K-medoid clustering algorithm based on DTW is used to cluster the time feature sequence, and the preliminary results of block clustering in the study area are obtained. Secondly, an ensemble method (CA-RFM model) combining the clustering algorithm with the random forest model is proposed. This method uses clustering algorithm to extract significant feature regions as input, effectively integrates time point features and POI point features and uses the random forest model to automatically identify UFZ. Finally, using up-bottom functional zoning identification, combining the semantic features of the city represented by POIs, the functional zoning categories of multi-scale block units are finely identified level by level.

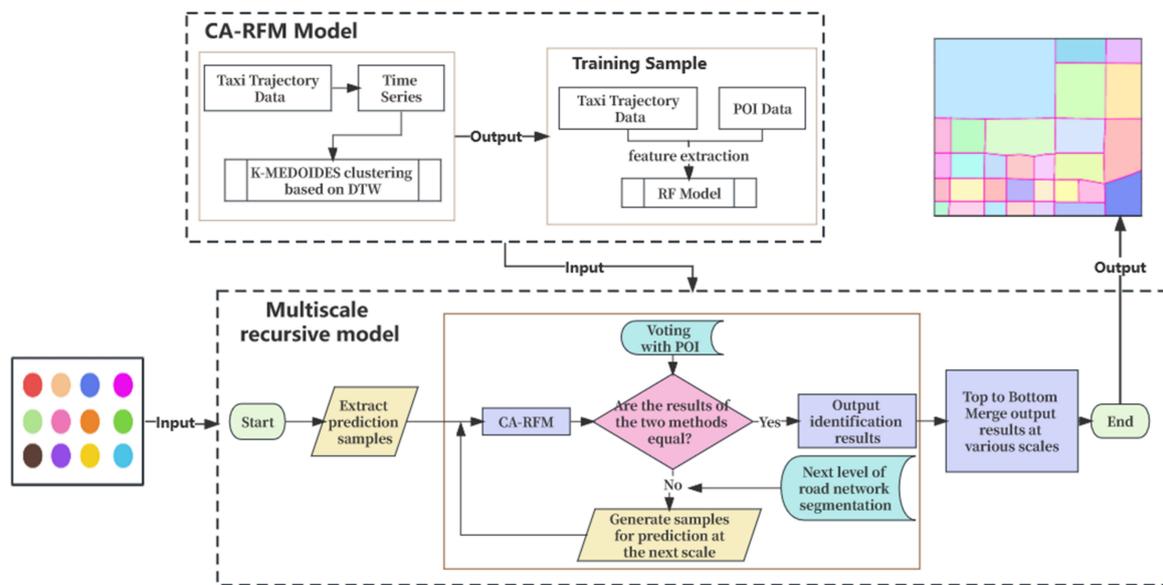


Figure 2. Workflow of urban functional area identification.

2.3.1. Methods of Time Series Generation

To understand the trip patterns of residents as a whole, the total number of passengers per hour on working days and rest days in TAZ was counted. On this basis, the average number of hours per hour on working days and rest days is counted.

Figure 3 shows that there are large differences in the travel patterns of residents on workdays and weekends, so workdays and weekends should be treated separately. The daily data of the pickup and the drop-off point on the workdays and weekends intersect with the TAZ data. Then, the pick-up and drop-off numbers within each hour and each TAZ were counted. Finally, the average passenger numbers over 24 h a day on workdays and weekends were calculated. We obtained 4 sets of data in TAZ.

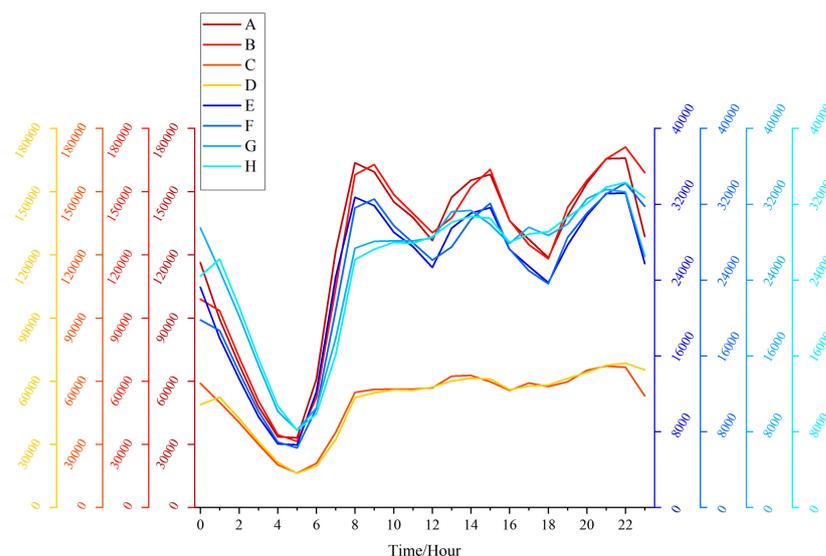


Figure 3. Characteristics of resident trips in the study area. A represents the number of pickups on workdays; B represents the number of pickups on weekends; C represents the number of drop-offs on workdays; D represents the number of drop-offs on weekends; E represents the average outflow of workdays; F represents the average inflow of working days; G represents the average outflow of weekends; H represents the average inflow of weekends.

In summary, the time series of each ultimately generated research unit as follows:

$$\{O_0, O_1, \dots, O_{23}, D_0, D_1, \dots, D_{23}, O'_0, O'_1, \dots, O'_{23}, D'_0, D'_1, \dots, D'_{23}\}$$

where $O_0 - O_{23}$ represents the average outflow of workdays, $D_0 - D_{23}$ represents the average inflow of workdays, $O'_0 - O'_{23}$ represents the average outflow of weekends, $D'_0 - D'_{23}$ represents the average inflow of weekends.

2.3.2. Dynamic Time Warping

Dynamic time warping algorithm finds the best correspondence between two observation sequences by regularizing the time dimension with certain constraints, which can explore the similarity and difference of time series with maximum flexibility, and it is the most commonly used quantitative method to measure the similarity of time series.

Given time series $P = [r_1 \dots r_i \dots r_n]$ and $Q = [s_1 \dots s_j \dots s_m]$, construct an $n \times m$ matrix grid, where each element of the matrix represents the distance between points r_i and s_j , the DTW algorithm needs to ensure the minimum difference when aligning P and Q . Build path $W = [w_1 \dots w_r \dots w_k]$, among them $\max(m, n) \leq K \leq m + n - 1$, it is necessary to satisfy three conditions: boundary conditions, continuity conditions and monotone conditions. The boundary condition is that the starting point of the path is the lower left corner element $(1, 1)$ of the matrix, and the endpoint is the upper right corner element (n, m) . The continuity condition means that, except for the start and end points, each element on the path must have two points around it that are adjacent to each other; The monotonicity condition requires that the next element on the path must lie to the right of or above the previous element, while not spanning two elements. Among all the paths that satisfy the above three constraints, the one with the smallest d_{DTW} is selected as the output result, that is, the path with the smallest distance between P and Q is measured:

$$d_{DTW}(i, j) = MAR(i, j) + \min(d_{DTW}(i, j - 1), d_{DTW}(i - 1, j), d_{DTW}(i - 1, j - 1)) \quad (1)$$

where $1 \leq i \leq 96, 1 \leq j \leq 96$, $d_{DTW}(i, j)$ is the minimum cumulative distance of the current matrix element $MAR(i, j)$ with $d_{DTW}(0, 0) = 0, d_{DTW}(0, j) = d_{DTW}(i, 0) = \infty$.

In this study, DTW eliminates phenomena such as shifting in the time trend when measuring the variability of the 96-dimensional OD point time series in TAZ, and the time comparisons based on the DTW distance calculations are good.

2.3.3. K-Medoid Clustering Algorithm

For a large amount of data without labels, semi-supervised learning usually adopts manual methods to mark a small number of data labels with typical characteristics as the training samples to train most of the remaining data without labels [34]. In this paper, training samples are generated by combining unsupervised learning with manual labeling, which greatly improves the accuracy of the experiment.

The DTW algorithm can be used to obtain the plot distance matrix, that is, the similarity matrix of the time series of taxi traffic volume of the block unit, based on which the clustering analysis can distinguish the differences between different plot types. In the phase of generating training samples, the clustering method used in this study is the K-medoid algorithm. K-medoid clustering is the preferred method in large-scale data clustering analysis and is less affected by outliers, which makes it more suitable for this study [35].

To evaluate the reliability of the results of different clustering numbers, this study introduced the silhouette to evaluate the clustering quality of each cluster. In the context of the K-medoid algorithm, assuming that in an existing clustering result, where $a(i)$ represents the mean value of the DTW distance between sample point i and other sample points in the

same cluster and $b(i)$ represents the mean value of the minimum DTW distance between sample point i and other clusters, then there are:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2)$$

If $s(i)$ is close to 1, it means that the sample point i matches well with the existing clustering results. If $s(i)$ is close to -1 , it means that the sample point i should belong to its neighbor clusters. The higher the mean value of $s(i)$ of all points means the better the clustering results.

2.3.4. CA-RFM Model

Random Forest is essentially a collection of many decision trees, and multiple trees are integrated through an integrated learning concept based on traditional decision tree algorithms, which ultimately results in a final prediction based on multiple tree voting. The randomness of the random forest is reflected in the fact that the training samples of each decision tree are randomly selected, and the splitting attributes of each node in the tree are also randomly selected. Therefore, the accuracy of random forest classification results greatly depends on the accuracy of training samples. The clustering algorithm obtains the preliminary division result of the functional area by directly clustering the time series data, and some areas have an inaccurate division. To make the classification results more credible, this study takes the regions with significant features of each category in the clustering results as the training sample regions of the random forest model and combines the clustering algorithm and the random forest model to construct the CA-RFM model. This combination of supervised and unsupervised learning to select samples increases the accuracy of the training samples to some extent and improves the precision of the experiment.

Given that several studies have confirmed that different urban functional areas have different time statistical features and POI point features, but these two types of characteristics are seldom fully integrated and used for functional area classification, this study used the CA-RFM model to fuse these two types of characteristics, and the model was used to classify the functional areas.

- Extraction of time statistical features;

Taxi OD data reflect the mobile information of passengers in different regions. Different functional areas provide different social functions, and the number of taxi pick-ups and drop-offs will change with time. Extracting 48-dimensional time statistical features of taxi pick-up and drop-off point data for one week in each basic analysis unit are extracted for functional area identification of each unit. Considering the differences in the travel of residents on workdays and weekends, the total number of taxis getting on and off per hour for 24 h per day on workdays and weekends were counted separately in each of the basic units of analysis to generate a 96-dimensional data feature. The calculation method of the average statistics of the getting on and off points of each basic analysis unit is as follows:

1. Average statistics of getting on points of taxi trajectory data

The calculation method of the 24 h average statistics of the getting on point of each basic analysis unit on workdays is shown in formula (3):

$$V_{on} = \frac{\sum_{d_k \in S_h} M_{on}(u_{:,d_k})}{W_d} \quad (3)$$

where V_{on} is the statistics of getting on points; $u_{:,d_k}$ is the 24-dimensional vector of the number of getting on points on the k th working day; $M_{on}(v)$ is the mean form of the daily getting on point statistic; S_h is a one-week taxi trajectory experimental dataset; W_d is the total number of workdays in a week.

The calculation method of the 24 h average statistics of getting on points of each basic analysis unit on weekends is as above, and the number of days on the workdays can be replaced by the number of days on weekends.

2. Average statistics of getting off points of taxi trajectory data

The calculation method of the 24 h average statistics of the getting off point of each basic analysis unit on workdays is as shown in formula (4):

$$V_{off} = \frac{\sum_{e_k \in S_h} M_{off}(u_{:,e_k})}{W_d} \quad (4)$$

where V_{off} is the statistics of getting off points; $u_{:,e_k}$ is the 24-dimensional vector of the number of getting off points on the k th working day; $M_{off}(v)$ is the mean form of the daily getting off point statistic; S_h is a one-week taxi trajectory experimental dataset; W_d is the total number of weekends in a week.

Considering each functional area's different area, the density of each face OD point is calculated as a feature to make up for the OD point flow information lost by normalization processing. In summary, the time statistical characteristics of each research unit are finally generated as 97 dimensions, that is:

$$\{O_0, O_1, \dots, O_{23}, D_0, D_1, \dots, D_{23}, O'_0, O'_1, \dots, O'_{23}, D'_0, D'_1, \dots, D'_{23}, Den\}$$

where $O_0 - O_{23}$ is the average outflow characteristics of the workdays, $D_0 - D_{23}$ is the average inflow characteristics of the workdays, $O'_0 - O'_{23}$ is the average outflow characteristics of the weekends, $D'_0 - D'_{23}$ is the average inflow characteristics of the weekends and Den is the point density characteristics, a total of 97 dimensions.

- Extraction of POI point features.

The number of POI points reflects the absolute value difference of different types of interest points in the functional area, which can be used to assist in judging the actual functional attributes of the functional area. However, the absolute value of the POI may also cover the actual dominant attribute information in the region, so the point density and enrichment index of the POI are introduced as auxiliary discriminant information. Twelve representative types of POIs are selected from the general category of the POI, which are catering service, scenic spot, company and enterprise, shopping service, finance and insurance, science and education and cultural service, housing, life service, sports and leisure service, medical and health service, government agencies and social organizations and accommodation service. For each plot divided, the point density and enrichment index of each type of POI point in the plot are calculated [36].

The density of POI points is expressed as:

$$Density_{POI(i,j)} = \frac{Num_{POI(i,j)}}{Area_j} \quad (5)$$

where $Density_{POI(i,j)}$ is the density of type i POIs in the functional area of type j ; $Num_{POI(i,j)}$ is the number of type i POIs in the class j functional area; $Area_j$ is the total area of the class j functional area.

The POI enrichment index is expressed as:

$$F_{i,j} = \frac{n_{i,j}/n_j}{N_i/N} \quad (6)$$

where $F_{i,j}$ is the enrichment index of the class i POIs in the class j functional area; $n_{i,j}$ is the number of type i POIs in the class j functional area; n_j is the total number of POIs in the class j functional area; N_i is the total number of type i POIs; N is the total number of all POIs in the entire study area. The higher the F indicates the higher the enrichment index of type i POIs in the class j functional area.

In summary, the final 24-dimensional features of POI points for each research unit were generated, i.e.,

$$\{D_1, D_2, \dots, D_{12}, F_1, F_2, \dots, F_{12}\}$$

where $D_1 - D_{12}$ is the densities of the 12 types of POIs, and $F_1 - F_{12}$ is the enrichment index of the 12 types of POIs. These two indexes are used as POI point features, and a total of 24-dimensional POI point features are extracted.

2.3.5. Quantitative Identification of a POI

The POI contains a large amount of semantic information about urban functions and is a way to quantitatively identify functional areas. Considering the large difference in the amount of POI data between different categories and the differences in the geographic entities they represent and the public awareness, this study introduces the two indicators of frequency density (FD) and category ratio (CR) to determine the functional attributes, and the calculation formulas are as follows [37].

$$FD_i = \frac{n_i}{N_i} (i = 1 \dots 5) \quad (7)$$

$$CR_i = \frac{F_i}{\sum_{i=1}^5 F_i} \times 100\% (i = 1 \dots 5) \quad (8)$$

where i represents the i -th of the five POI types; n_i is the number of the i -th type of POI in the block unit; N_i is the total number of i -th type of POI; FD_i is the frequency density of i -th type of POI in the block unit to the total number of POIs of that type. CR_i is the ratio of the frequency density of i -th type of POI to the frequency density of all types of POIs in the block unit.

The FD and CR of each type of POI within each functional area unit are calculated according to the formula. Referring to the research of Chi Jiao et al., and through multiple adjustment tests, the CR value of 30% is determined as the standard to judge the nature of the functional area of the unit [38]. That is, when the proportion of a certain type of POI type is greater than 30%, the unit is judged to be a single functional area; when the proportion of all POI types in the unit does not exceed 30%, the area is determined to be a mixed functional area, and the mixed type depends on the two most dominant POI types in the unit; three and more than three mixed cases are not considered in this study.

2.3.6. Multi-Scale Recursive Recognition Method Based on Cross-Validation

The auxiliary data used in the delineation of urban functional zones vary, while the block unit formed by the road network is closer to the boundary of urban functional zones, easy to obtain and is the most widely used data in the delineation of functional zones. The road network-based method can better estimate the actual distribution of urban roads, and the use of multilevel urban road networks divided into functional district block units can better meet the scientific management of urban planning departments and assist decision making. For this reason, this study proposes a multi-level research unit division method based on road grade, i.e., using highways, trunk roads and main roads as the first-level unit demarcation line, adding ordinary street roads based on the first-level demarcation line as the second-level demarcation line, and adding service roads based on the second-level demarcation line as the third-level demarcation scale, to obtain the third-level scale research unit.

Based on multi-level road network division, a multi-scale recursive identification method based on cross-validation is proposed by combining the results of CA-RFM model extraction and POI quantitative identification, as shown in Figure 4. The CA-RFM model is used to determine the urban functional area category of the block unit at each scale. In this process, the POI-based voting was used to verify the identification based on the CA-RFM model and the results of the validation determined which blocks would be divided into sub-blocks at the next scale. Whether the block unit is divided into the next scale depends

on the consistency of the extraction results of the two methods, to realize the top-down hierarchical division from large-scale road network to small-scale road network.

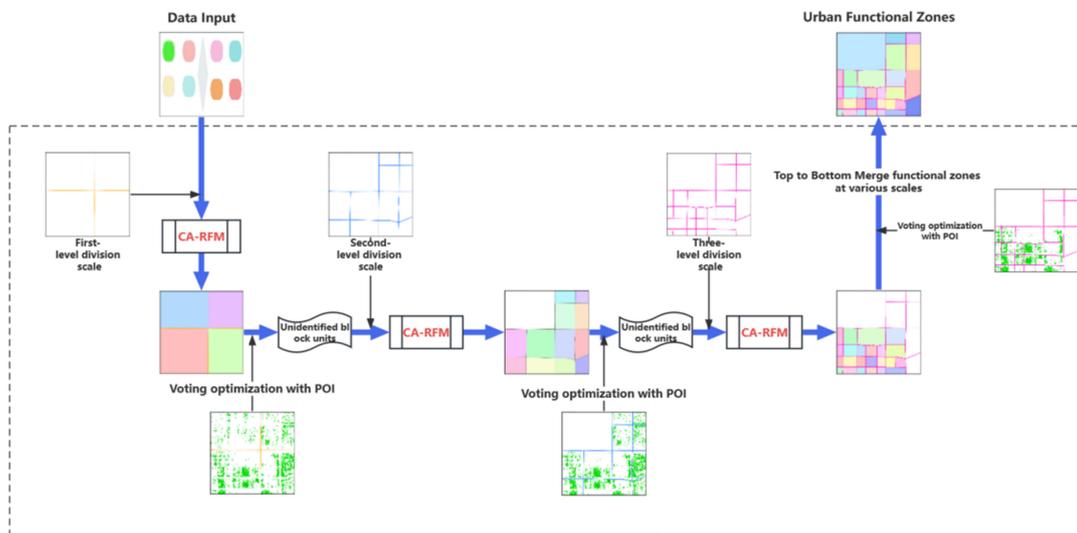


Figure 4. Multi-scale recursive recognition based on cross-validation.

This study formulates the principles that the method needs to follow. Firstly, after calculating the CR values of all POI types in each block unit, the attribute similarity between the quantitative identification of POI based on CR judgment and the identification based on the CA-RFM model is calculated to determine whether the urban functional area is divided and the attributes of the block unit.

- (1) For the unit with a CR value greater than 30% of POI type, if the functional attributes determined by CR are consistent with the functional identification results of the CA-RFM model, the functional area attributes of the block unit are determined and the block unit is no longer divided. If the functional attributes determined by CR are inconsistent with the identification results of the CA-RFM model, the block unit is further divided until the functional attributes of the two methods are consistent.
- (2) For all units with CR values of POI types less than 30%, if the functional attributes determined by CR are consistent with the functional identification results of the CA-RFM model, the results are retained and the unit will not be divided. If the functional attributes determined by CR are inconsistent with the functional identification results of the CA-RFM model, the block unit is further divided until the functional attributes of the two methods are consistent.
- (3) For the unit that does not contain POI (CR is a null value), it is called a null value unit. The recognition result of the CA-RFM model will be the terminal functional area category of the block unit and will not be divided. For the unit that does not contain trajectory data or the number of time statistical features of 0 exceeds 80% of the total number of features, the functional attributes determined by the CR value are the terminal functional area category of the unit and will not be divided. For the block unit with inconsistent attribute results obtained by the two methods in the third level division, the functional attribute determined by the CR value is the final functional area category of the unit; for units that contain neither POI data nor trajectory data, they are referred to as no-value unit and are not used as discriminatory regions.

3. Results

3.1. Training Sample Generation of CA-RFM Model

For the training of the CA-RFM model, training samples with labels are essential. To obtain the training samples, the K-medoids algorithm was utilized to cluster the pre-processed time series data, the reliability of the number of clusters is evaluated by the

silhouette coefficient. The change in the silhouette value with the number of clusters K is shown in Figure 5a–c. The larger the contour coefficient, the smaller the number of clusters. Too small silhouette values lead to over-categorization, which produces too many irrelevant categories; too large silhouette values lead to under-categorization, which makes it difficult to separate different categories with similar properties. Therefore, we choose the point where the concave–convex nature of the function changes, i.e., the inflection point. From the graph, it can be seen that there are inflection points at each level when the number of clusters is 7, 6 and 6, respectively. Considering the change of silhouette with K and the size of the data volume, the number of clusters at each level of the road network is finally determined to be 7, 6 and 6.

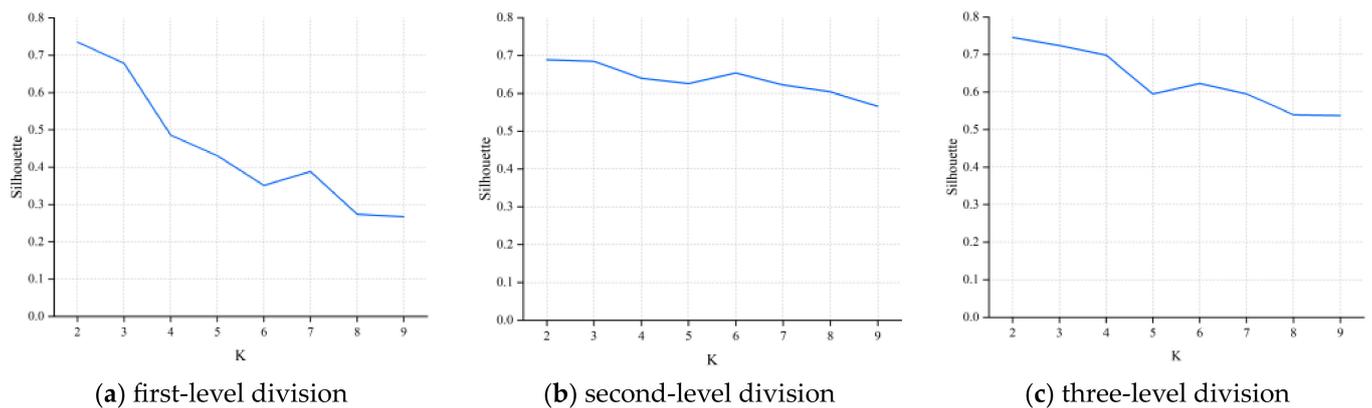
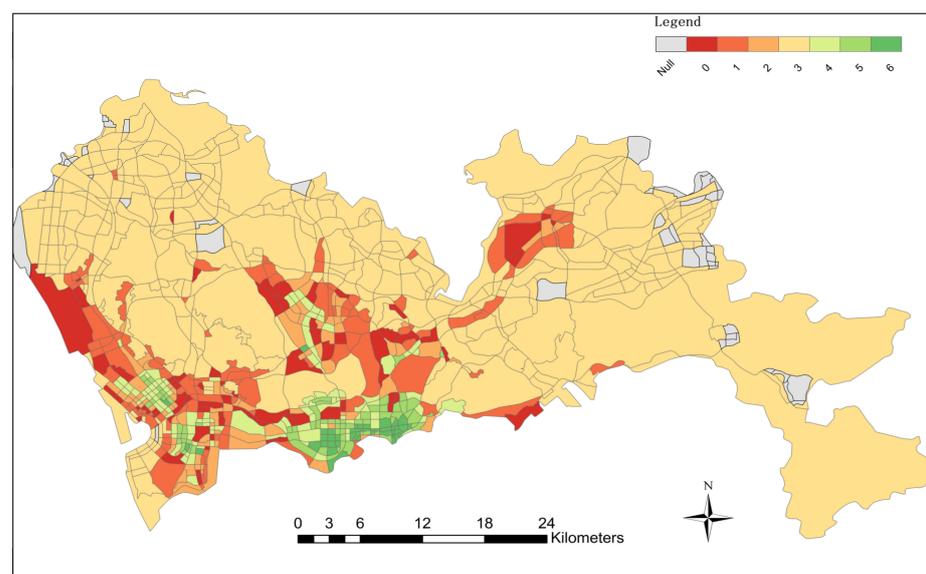


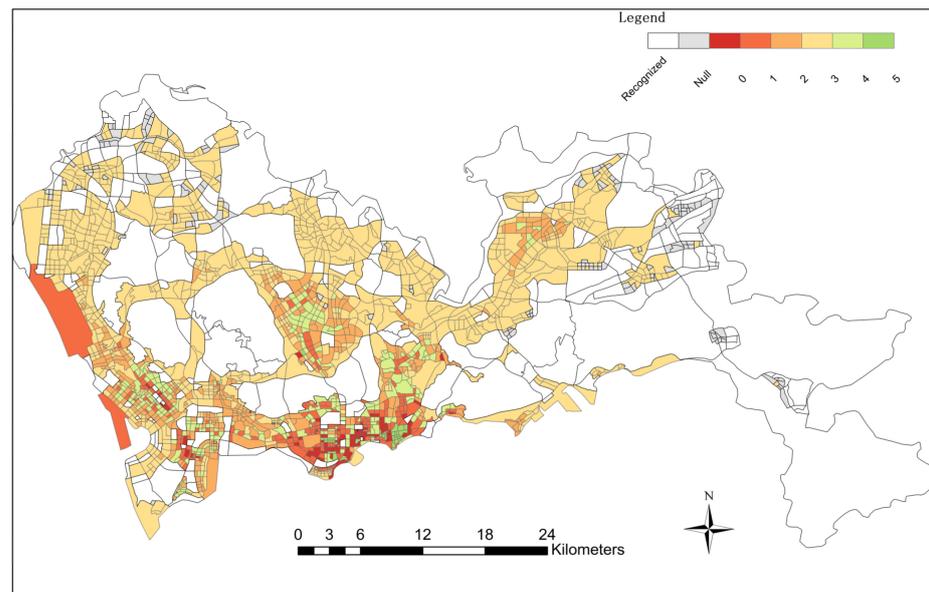
Figure 5. The changes in silhouette values with different numbers of cluster.

According to the overall planning of Shenzhen City, the POI enrichment index of each type in each block unit and the category of urban functional areas marked by high-definition remote sensing images, this study selects a certain amount of significant feature areas from Figure 6a–c as the input of the CA-RFM model and generates training samples for training the model. The sample sizes of industrial and commercial mixed area (C1), green scenic spot (C2), life and recreation mixed area (C3), mature commercial area (C4), industrial/public service mixed area (C5), public and commercial mixed area (C6) and urban residential area (C7) are all 75.

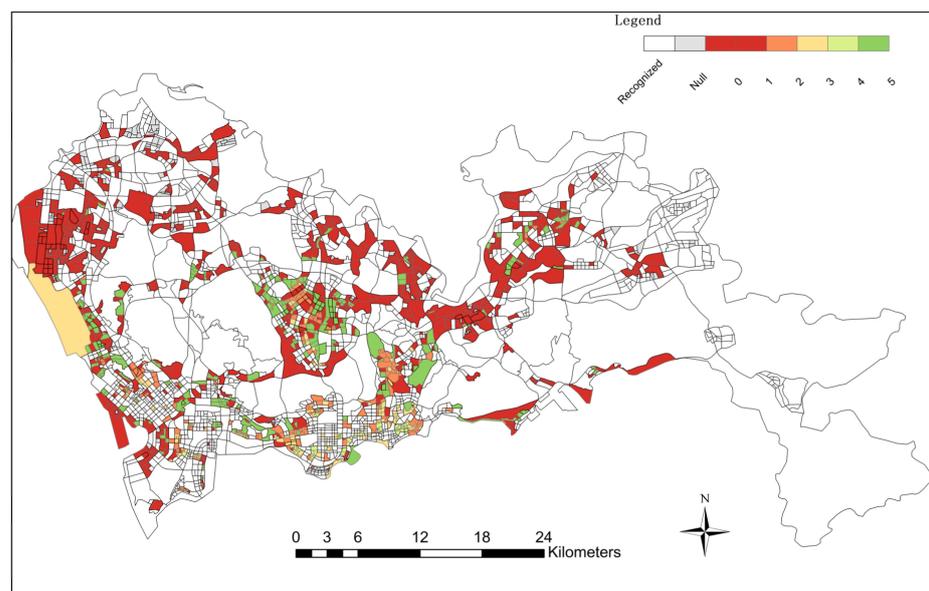


(a) first-level clustering

Figure 6. Cont.



(b) second-level clustering



(c) three-level clustering

Figure 6. K-MEDIODS clustering results.

3.2. Multi-Scale Recursive Urban Functional Area Identification Results

Figure 7 shows the identification results of functional areas with highways, trunk roads and main roads as the first-level division scale. The road grade used in the first-level division scale is mainly responsible for the long-distance and fast transportation services of the city. It can be used as a landmark road of a city, and its zoning scale is relatively large. The study area was divided into 919 first-level block units, of which only 270 units were successfully identified by the multi-scale model, and the remaining 649 units did not reach the threshold of similarity calculation. The results show that the functional attributes of these 649 units are highly heterogeneous, and there are multiple categories of functional areas within the block units. These first-level block units need to be subdivided on the next scale.

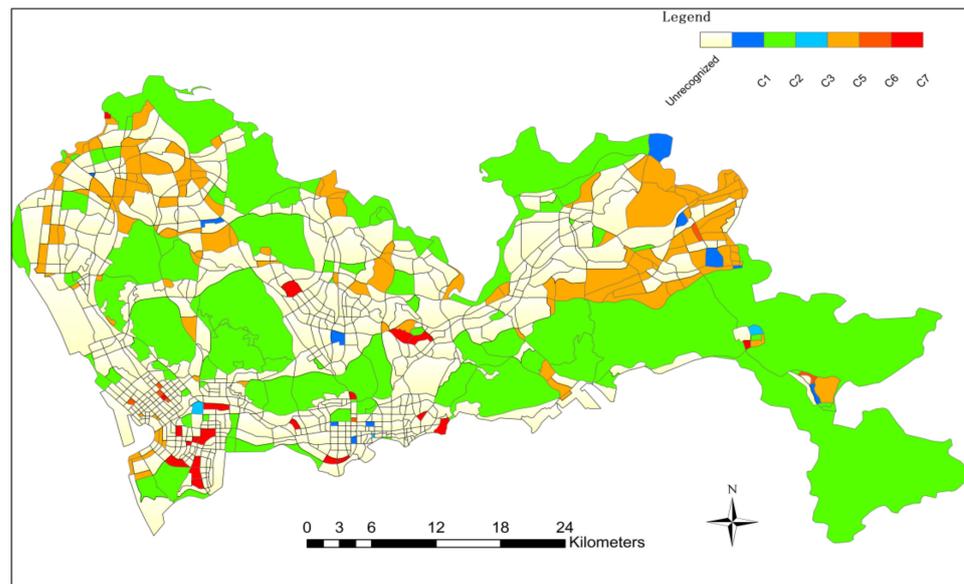


Figure 7. Identification results of the first level of segmentation.

Taking ordinary streets as the dividing boundary, a total of 2071 secondary block units are divided. The similarity algorithm is used to calculate the recognition results of the CA-RFM model. A total of 1308 secondary block units are successfully identified, and the remaining 763 secondary block units need to be divided at the next scale (Figure 8). In the secondary division, the number of identified functional areas has soared, especially in the mixed area of life and recreation and the mixed area of industry/public service. This also shows that, in urban planning and design, many factories, public service areas and residential areas are designed with ordinary streets as the boundary. In addition, ordinary streets are used to connect most areas of the city. Residential areas, industrial areas and public service areas are generally located near convenient streets.

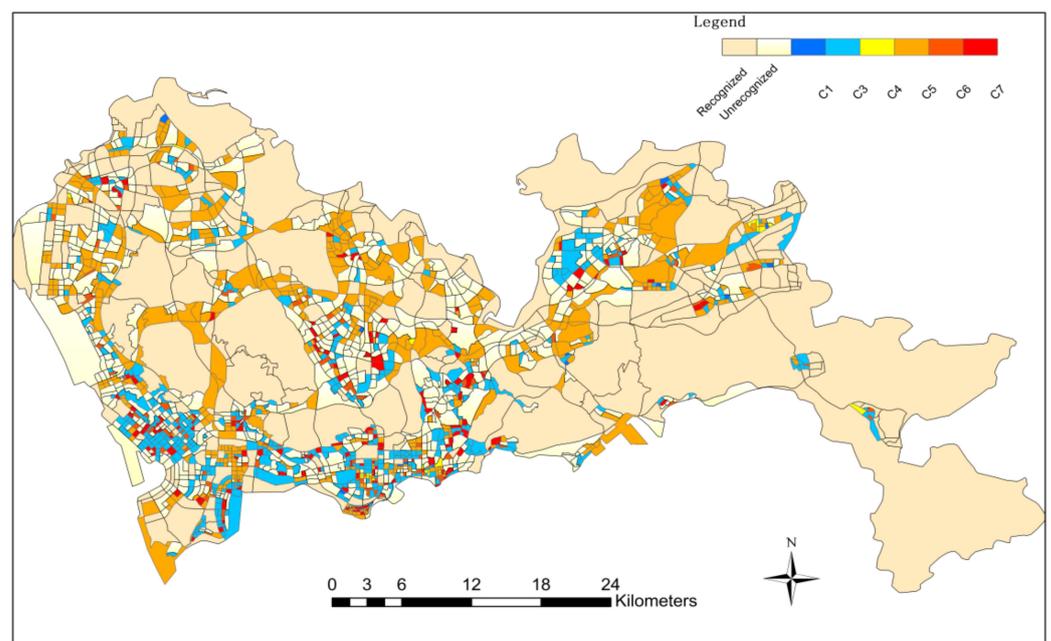


Figure 8. Identification results of secondary segmentation.

Figure 9 shows the results of the final level of block unit identification using service roads, with a total of 1510 tertiary block units identified, and some smaller-scale mixed

public and commercial areas and urban residential areas identified in large numbers. At the same time, functional areas mainly based on industrial mixed functions (industrial-green mixed areas, industrial-residential mixed areas) and functional areas mainly based on public-service mixed functions (public-green mixed areas, public-residential mixed areas) are also identified. In the last block unit, 12 types were identified, namely: C1-industrial and commercial mixed, C2-green scenic spot, C3-life and recreation mixed area, C4-mature commercial area, C5-industrial/public service mixed area, C6-public commercial mixed area, C7-urban residential area, C8-industrial and green mixed area, C9-public residential mixed area, C10-public green mixed area, C11-industrial and residential mixed area, C12-green residential mixed area. Among them, the mixed area mainly composed of industrial and public service mainly includes some small office areas, small factories and factories, etc., which are relatively small in area, so it is necessary to divide the functional zoning unit of the minimum scale road. It can be seen that the land-use types in Shenzhen are mainly mixed with residential land, industrial land and public land.

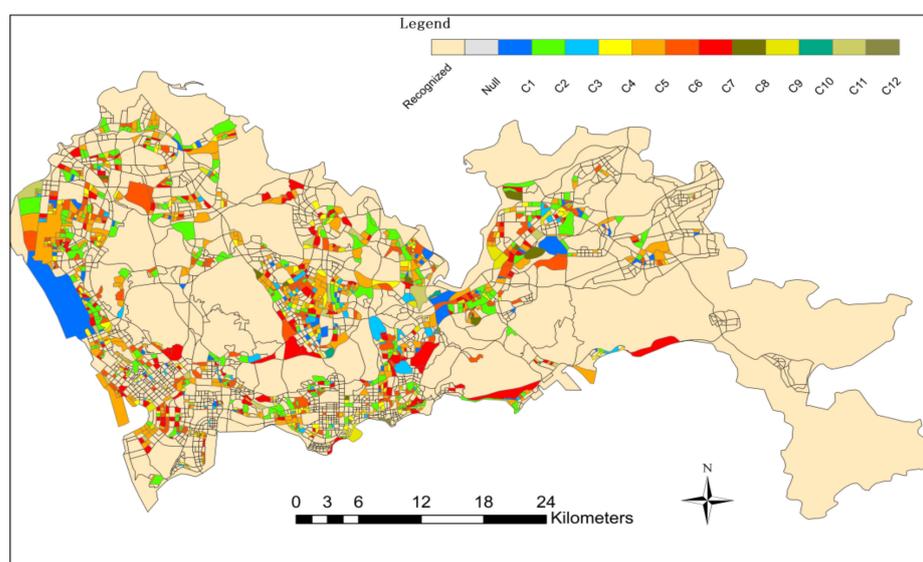


Figure 9. Identification results of the three levels of segmentation.

Finally, the classification results of the above three scales are combined to obtain the overall functional area identification results of the study area, as shown in Figure 10. Based on the division from the large-scale road network to the small-scale road network, this method realizes the identification of multi-scale urban functional areas from top to bottom. The study area is divided into 3088 block units. For each type of functional area type, a certain amount of block units in the classification results are extracted. The results are tested with the overall planning of Shenzhen City and the 'Mapping of Basic Urban Land Use Types in China: Preliminary Results in 2018' [39]. The calculation results of the confusion matrix are as follows (Figure 11), and the overall recognition accuracy is 0.874%. The above experiments demonstrate that the multiscale recursive recognition method combines the two methods organically. On the one hand, it realizes the mutual test of the two recognition results and improves the extraction accuracy of the functional area. The accuracy of the recognition results of the CA-RFM model is tested by using the functional semantic information implied by POIs. The CA-RFM solves the problem that there are no POI data in some units and POIs may have inaccurate data. On the other hand, it reduces the unnecessary division of some blocks and improves the operational efficiency of the model.

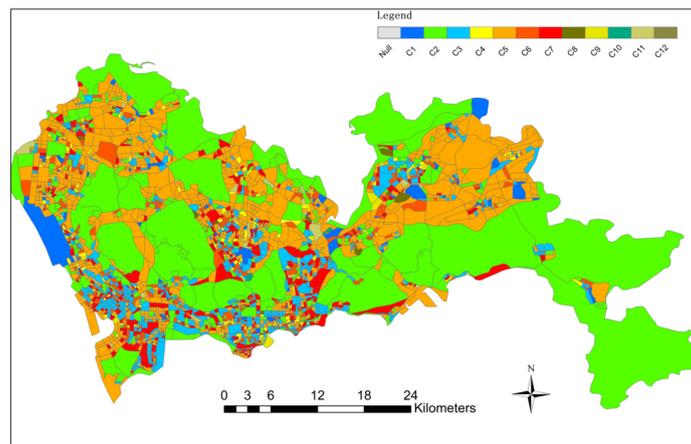


Figure 10. Final consolidation results of urban functional areas.

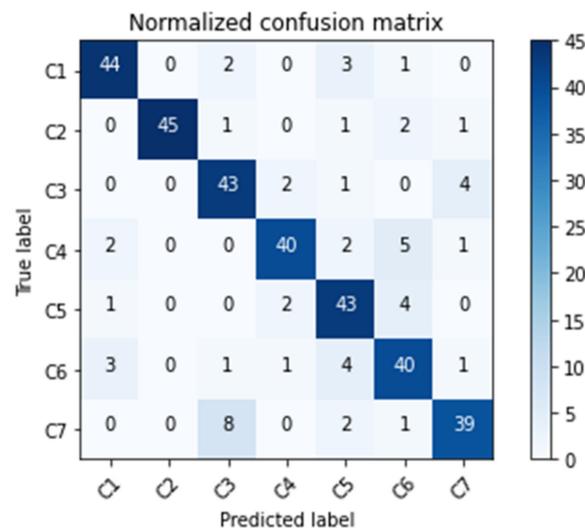


Figure 11. Confusion matrix results.

4. Contrast Experiment

To verify the performance of the multi-scale recursive identification method based on cross-validation in identifying urban functional areas, the functional area identification results of this paper’s method (E) and the single-scale POI quantitative identification method (A), the multi-scale POI quantitative identification method (B), the single-scale CA-RFM model (C) and the multi-scale CA-RFM model (D) are compared. To keep the variables constant, the block units of each layer obtained in Figure 10 are used as single-scale functional area constraint boundaries, and the three-level scales of this study are used as multi-scale functional area identification constraint boundaries. Table 3 shows the accuracy comparison of different examples.

According to the combination of different scales and methods, five groups of comparative experiments were generated. From the perspective of “scale”, based on the same method, the overall accuracy OA and Kappa of group B were higher than those of group A, at the same time, the multi-scale recognition results of group CD were also better than the single-scale recognition results. From the perspective of “method”, the POI quantitative identification method is better than the CA-RFM model based on the same single scale in the two groups of AC, and the CA-RFM model is better than the POI quantitative identification based on the same multi-scale in the two groups of BD. On the whole, the multi-scale recognition results are better than the single-scale recognition results. In this case, the method of this study (group E) obtained the highest OA and Kappa coefficients: OA-0.874, Kappa-0.853. In contrast, the method proposed in this paper has the best recognition effect.

Table 3. Accuracy evaluation of identification results of urban functional areas in different examples is a table.

| ID | Scale | | Method | | OA | Kappa |
|----|--------------|-------------|------------------------------------|--------------|-------|-------|
| | Single-Scale | Multi-Scale | Quantitative Identification of POI | CA-RFM Model | | |
| A | ✓ | | ✓ | | 0.672 | 0.617 |
| B | | ✓ | ✓ | | 0.746 | 0.703 |
| C | ✓ | | | ✓ | 0.647 | 0.588 |
| D | | ✓ | | ✓ | 0.757 | 0.717 |
| E | | ✓ | ✓ | ✓ | 0.874 | 0.853 |

5. Discussion

POI data can reflect the spatial distribution of features, provide rich socioeconomic information and have better spatial timeliness than traditional data. Therefore, using POI data to identify urban functional area attributes is simpler and more efficient than traditional methods. However, POI data also have limitations; POI data cannot reflect dynamic information. Therefore, combining cab trajectory data and POI data to analyze urban functional areas better meets the current requirements. In this paper, urban neighborhoods divided by multilevel road networks are used as the unit of study to fit the urban form more accurately and make the identification results more accurate. In order to validate the recognition results of the method, the results were compared with Google Earth images, Golder maps, and real photos of landmark areas. The comparison results for some typical areas are shown in Table 4, and Google images from 2016 were chosen to be as close as possible to the time of the trajectory data and POI data.

Table 4. Comparison and evaluation of functional area identification results.

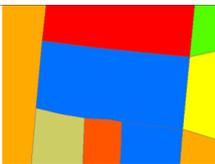
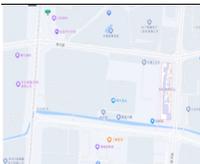
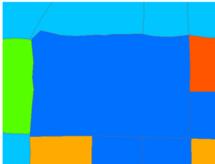
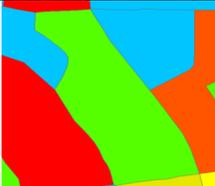
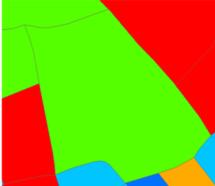
| Function Area | No. | Results of Identification | Google Earth Image | Gaode Map | Real Photos of Landmark Site |
|--|-----|---|--|---|---|
| C1: industrial and commercial mixed | 1 |  |  |  |  |
| | 2 |  |  |  |  |
| C2: green scenic spot | 3 |  |  |  |  |
| | 4 |  |  |  |  |

Table 4. Cont.

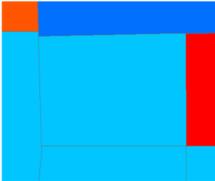
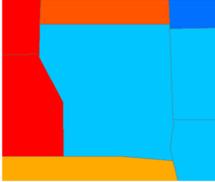
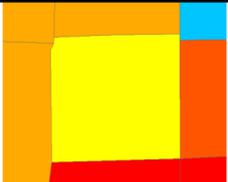
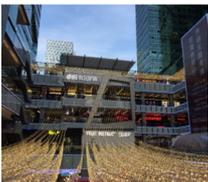
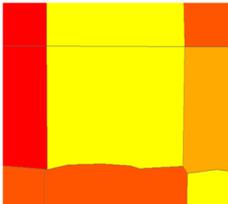
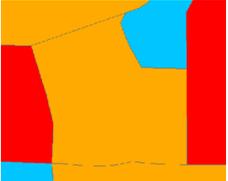
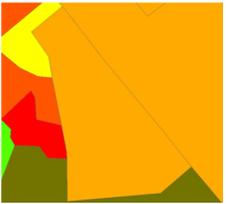
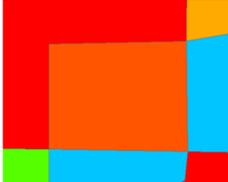
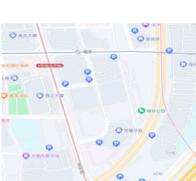
| Function Area | No. | Results of Identification | Google Earth Image | Gaode Map | Real Photos of Landmark Site |
|---|-----|---|--|---|---|
| C3: life and recreation mixed area | 5 |  |  |  |  |
| | 6 |  |  |  |  |
| C4: mature commercial area | 7 |  |  |  |  |
| | 8 |  |  |  |  |
| C5: industrial/public service mixed area | 9 |  |  |  |  |
| | 10 |  |  |  |  |
| C6: public commercial mixed area | 11 |  |  |  |  |
| | 12 |  |  |  |  |

Table 4. Cont.

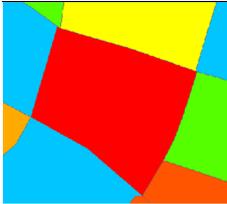
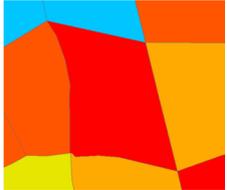
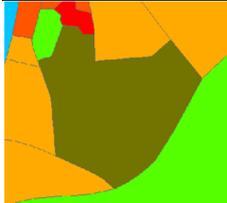
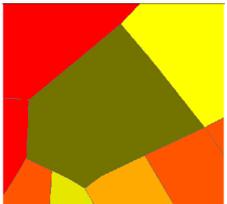
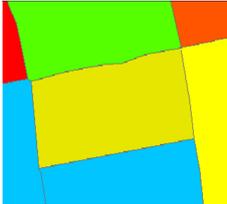
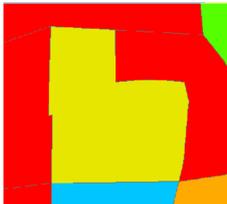
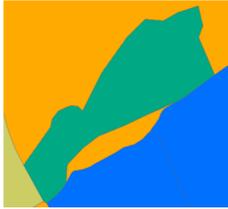
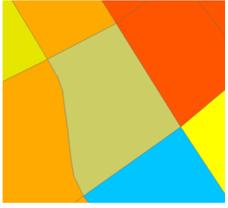
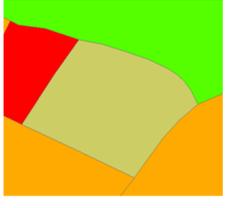
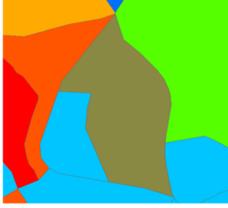
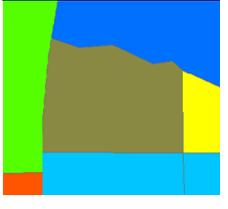
| Function Area | No. | Results of Identification | Google Earth Image | Gaode Map | Real Photos of Landmark Site |
|--|-----|---|--|---|---|
| C7: urban residential area | 13 |  |  |  |  |
| | 14 |  |  |  |  |
| C8: industrial and green mixed area | 15 |  |  |  |  |
| | 16 |  |  |  |  |
| C9: public residential mixed area | 17 |  |  |  |  |
| | 18 |  |  |  |  |
| C10: public green mixed area | 19 |  |  |  |  |

Table 4. Cont.

| Function Area | No. | Results of Identification | Google Earth Image | Gaode Map | Real Photos of Landmark Site |
|---|-----|---|--|---|---|
| C10: public green mixed area | 20 |  |  |  |  |
| C11: industrial and residential mixed area | 21 |  |  |  |  |
| | 22 |  |  |  |  |
| C12: green residential mixed area | 23 |  |  |  |  |
| | 24 |  |  |  |  |

The landmark area in Group 1 is the Tianhong Shopping Center, which is one of the more well-known shopping centers in Shenzhen, with many specialty stores and brand stores of famous brands; the area in Group 2 is the Bauhinia City Shopping Plaza, and there are also a number of medium-sized shopping malls such as the Qunxing Plaza Shopping Center and the Maoye Department Store in this area, with a wide variety of specialty themed food and beverage and themed merchandise stores, and with notable commercial functions. In addition, there are a large number of tall office buildings in the vicinity, where the nature of the work is evident, and the mix of the two main functions is in line with the urban function of “mixed industry and commerce”. The landmark areas in the comparison of Groups 3 and 4 are Shiang Mee Park and Baoan Park, respectively. Shiang Mee Park is a comprehensive municipal park integrating culture, leisure and experience, and Baoan Park is also a good place for citizens’ fitness, ecological sightseeing and leisure, which is in line with the positioning of its functional area as a “Green Scenic Spot”. The landmark areas in the comparison of Groups 5 and 6 are Huangpu Nga Yuan and Jingtian South 3rd Street Park, which are not only densely populated with residential buildings but also have places for people to relax and have fun, which is in line with their function

as a “Mixed Lifestyle and Recreation Area”. The landmark areas in the comparison of Groups 7 and 8 are the Joyo INTOWN Shopping Center and Star River COCO Par. These two large shopping centers are included in the CBD business circle of Futian District, which is a large-volume, composite and diversified commercial agglomeration area, and the commercial function of these two areas occupies most of the area, with a small number of buildings for other functions in the surrounding area, which is consistent with its functional classification as a “mature commercial area”. The two areas occupy the majority of the area for commercial functions, with fewer buildings for other functions in the surrounding area, and are therefore in line with their function as “mature commercial districts”. The landmark areas in the comparison of Groups 9 and 10 are the Mok Mo Wan Primary School and the Shenzhen Baoxing Hospital, of which the Shenzhen Baoxing Hospital is a Grade II general hospital. Both areas contain large industrial areas, such as the Jishengchang Industrial Area and the Maadi Industrial Area, which have a medium density of small- and medium-sized firms, which is consistent with their functional positioning of a “Mixed Industrial/Public Service Area”. This is in line with the “mixed industrial/public service district” functional designation. The landmark areas in the comparison of Groups 11 and 12 are the Shenzhen Documentation Service Center and the Huanggang Community Library. Group 11 contains small- and medium-sized shopping malls, led by the Excellence Shopping Center, while Group 12 contains shopping and leisure venues such as the Huanggang Commercial City and Times Square, which are in line with the positioning of the “public-commercial mixed zone” functional area. In the comparison of Groups 13 and 14, there is a large number of higher-density and better-arranged residential buildings, which is consistent with its functional area positioning as an “urban residential area”. The landmark areas of Groups 15 and 16 are Civic Square and Longhua Park, where a large number of science and technology parks and wholesalers are located, which is in line with the positioning of the functional area as an “industrial and green mixed zone”. Groups 17 and 18 are characterized by the Donghai Experimental Primary School Kindergarten and Nanshan Foreign Language Kewa School, while Group 17 contains a large number of training institutes and medium-density residential areas, and Group 18 contains neighborhoods such as Rhine Garden and Mangrove Garden, which are in line with the positioning of the area as a “Mixed Residential and Communal District”. The landmark areas of Groups 19 and 20 are the Bonjour Monastery and the Pak Nai Hang Park, which contain public service facilities such as schools, sports and recreation facilities, and is in line with the positioning of the functional area as a “Mixed Use Public and Green Area”. The landmark areas of Groups 21 and 22 are Longhua Industrial Zone No. 3 in Bao’an District and Shenzhen Yanguang Middle School. Group 21 contains many small factories and companies, as well as a certain density of residential areas, which is in line with its positioning as a “Mixed Industrial and Residential Area” functional area. Group 22, which was originally a mixed industrial/public service district, was misclassified as “mixed industrial/residential”. The landmark areas of Groups 23 and 24 are the Plaza of the Unified Building in Jixia Village and the Community Park in Kwun Lung Village, which contain more neighborhoods and recreational plazas and is in line with the positioning of the functional area as a “mixed green and residential area”.

From the identification results, we can find that the mixed living and recreational areas with mainly residential functions and urban residential areas have better identification results; the mature commercial areas in the study area can be identified, especially those located in the location of urban business districts can be identified more accurately; The mixed industrial/public service areas were identified with high accuracy, but some of them were misclassified as mixed public housing and mixed industrial–housing areas, which is related to the quality of the data. In addition, large green areas can be accurately identified, and some ecological areas of the city, large squares and parks can be identified. Some small areas of parks and squares are mistaken for public service facilities, and mixed with residential areas are misclassified as other mixed areas, with poor single identification results. Although the methodology proposed in this study achieves the expected results

and provides a basis for fast and accurate identification of urban functional zones, there are still some limitations. For example, the taxi trajectory data used in this study have some positioning errors and such data do not completely cover the study area due to regional transportation, economic and infrastructure constraints, which leads to bias in the identification process; secondly, the actual boundaries of the delineated study units and functional areas are different, which can also make the identification results inaccurate. In the future, richer data and more precise divisions can be added and chosen to explore the functional areas of the city to improve recognition accuracy.

6. Conclusions

With the deepening of the urbanization process, the urban structure presents complex and regular characteristics, and this paper analyzes the urban spatial structure from the perspective of big data mining. In the era of big data, the emergence of massive data has added new data sources to the identification of urban functional areas. However, single data have inevitable defects in the identification of functional areas. Therefore, this paper uses a combination of multi-source data to improve the accuracy and reliability of functional area identification. Combined with taxi trajectory data, POI data and multi-scale road network data, a multi-scale recursive identification method of urban functional areas based on POI frequency density analysis and the CA-RFM model is proposed. Experiments and comparisons show the feasibility and superiority of the method. The method can provide a theoretical basis for urban land planning, administrative division adjustment, urban resource allocation and other fields, and has auxiliary and guiding value for the scientific integration of land use and urbanization. The contribution of this study is mainly manifested in two aspects:

- (1) The time series data are clustered and analyzed using DTW based K-MEDOIDS clustering, and the raw output of the clusters is used as the input to the CA-RFM model, which improves the accuracy and efficiency of the sample region selection using this auxiliary method. The overall accuracy of the experiment is 87.4%, which can be improved by up to 20% compared to the other control experiments in this paper, and the UFZ classification results also show the effectiveness of these sample zone selections.
- (2) Using multilevel road networks to decompose block unit level by level, combined with POI quantitative identification and CA-RFM model, a multi-scale recursive identification method of urban functional areas based on interactive validation is proposed to realize the fine extraction of functional areas from top to bottom, which avoids the shortcomings of the use of a single road network. The interactive validation of the two methods improves the overall classification accuracy. In addition, the recognition results of the joint use of CA-RFM model and CR can alleviate the negative impacts when there are no POI data, no taxi trajectory data and too little trajectory data in some blocks.

Much work remains to be done in this area of research in the future. In this paper, taxis are used as a representative of residents' traveling and other residents' traveling modes are ignored. In addition, location service big data, such as cell phone check-in data and microblog check-in data, are important references for the interpretation and classification of urban land use. Therefore, multi-source urban big data should be integrated to measure urban morphology in future research to make the classification results more detailed and reliable. In addition, after obtaining reliable classification results of urban functional zones, the spatial structure of each functional zone and its correlation can be analyzed. In turn, the degree of rational utilization of urban space can be assessed and effective optimization suggestions can be attempted.

Author Contributions: Conceptualization, T.L. and G.C.; methodology, T.L. and G.C.; validation, T.L.; formal analysis, T.L.; investigation, T.L.; writing—original draft preparation, T.L.; writing—review and editing, G.C.; visualization, T.L.; supervision, G.C.; project administration, G.C. and J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Fundamental Research Funds for the Universities of Henan Province (NSFRF180329), the MOE (Ministry of Education in China) Project of Humanities and Social Sciences (15YJCZH018) and the Science and Technology Project of Henan Province (162102210063).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study can be obtained from the first author by 15639116331@163.com with reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Antikainen, J. The concept of functional urban area. Findings of the ESPON project 1.1.1. *Informationen Zur Raumentwickl.* **2005**, *7*, 447–456.
2. Salkin, P.E. The politics of land use reform in New York: Challenges and opportunities. *John's Law Rev.* **1999**, *73*, 1041.
3. Long, Y.; Shen, Z.; Long, Y.; Shen, Z. Discovering functional zones using bus smart card data and points of interest in Beijing. In *Geospatial Analysis to Support Urban Planning in Beijing*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 193–217.
4. Yang, S. *A Study on Population Distributing and Function Area in Shanghai*; Capital University of Economics and Business: Beijing, China, 2007.
5. Tian, G.; Wu, J.; Yang, Z. Spatial pattern of urban functions in the Beijing metropolitan region. *Habitat Int.* **2010**, *34*, 249–255. [[CrossRef](#)]
6. Wang, H. Rise of new special development zones and polarization of socio-economic space in Xi'an. *Acta Geogr. Sin.-Chin. Ed.* **2006**, *61*, 1024.
7. Lienou, M.; Maitre, H.; Datcu, M. Semantic annotation of satellite images using latent Dirichlet allocation. *IEEE Geosci. Remote Sens. Lett.* **2009**, *7*, 28–32. [[CrossRef](#)]
8. Zhong, Y.; Zhu, Q.; Zhang, L. Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6207–6222. [[CrossRef](#)]
9. Zhong, Y.; Zhao, B.; Zhang, L. Multiagent object-based classifier for high spatial resolution imagery. *IEEE Trans. Geosci. Remote Sens.* **2013**, *52*, 841–857. [[CrossRef](#)]
10. Zhai, W.; Bai, X.; Shi, Y.; Han, Y.; Peng, Z.-R.; Gu, C. Beyond Word2vec: An approach for urban functional region extraction and identification by combining Place2vec and POIs. *Comput. Environ. Urban Syst.* **2019**, *74*, 1–12. [[CrossRef](#)]
11. Wang, M.; Zhang, X.; Niu, X.; Wang, F.; Zhang, X. Scene classification of high-resolution remotely sensed image based on ResNet. *J. Geovisualization Spat. Anal.* **2019**, *3*, 16. [[CrossRef](#)]
12. Berry, B.J.L. *Interdependency of Spatial Structure and Spatial Behavior: A General Field Theory Formulation*; Papers of the Regional Science Association; Springer: Berlin/Heidelberg, Germany, 1968; Volume 21, pp. 205–227.
13. Karlsson, C. *Clusters, Functional Regions and Cluster Policies*; JIBS and CESIS Electronic Working Paper Series; Centre of Excellence for Science and Innovation Studies: Stockholm, Sweden, 2007.
14. Yuan, N.J.; Zheng, Y.; Xie, X.; Wang, Y.; Zheng, K.; Xiong, H. Discovering urban functional zones using latent activity trajectories. *IEEE Trans. Knowl. Data Eng.* **2014**, *27*, 712–725. [[CrossRef](#)]
15. Huang, X.; Wang, C.; Li, Z.; Ning, H. A 100 m population grid in the CONUS by disaggregating census data with open-source Microsoft building footprints. *Big Earth Data* **2021**, *5*, 112–133. [[CrossRef](#)]
16. Meng, Y.; Hou, D.; Xing, H. Rapid detection of land cover changes using crowdsourced geographic information: A case study of Beijing, China. *Sustainability* **2017**, *9*, 1547. [[CrossRef](#)]
17. Jongman, B.; Wagemaker, J.; Revilla Romero, B.; De Perez, E.C. Early flood detection for rapid humanitarian response: Harnessing near real-time satellite and Twitter signals. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 2246–2266. [[CrossRef](#)]
18. Hou, H. Research on Urban Functional Area Recognition Method Based on Multi-Source Data. Master's Thesis, Henan University of Finance and Economics and Law, Zhengzhou, China, 2020. [[CrossRef](#)]
19. García-Palomares, J.C.; Salas-Olmedo, M.H.; Moya-Gomez, B.; Condeço-Melhorado, A.; Gutiérrez, J. City dynamics through Twitter: Relationships between land use and spatiotemporal demographics. *Cities* **2018**, *72*, 310–319. [[CrossRef](#)]
20. Hu, T.; Yang, J.; Li, X.; Gong, P. Mapping urban land use by using landsat images and open social data. *Remote Sens.* **2016**, *8*, 151. [[CrossRef](#)]
21. Liu, X.; Long, Y. Automated identification and characterization of parcels with OpenStreetMap and points of interest. *Environ. Plan. B Plan. Des.* **2016**, *43*, 341–360. [[CrossRef](#)]

22. Cheng, J.; Liu, J.; Gao, Y. Analyzing the spatial and temporal characteristics of cab trips in Beijing based on the time series clustering method. *J. Earth Inf. Sci.* **2016**, *18*, 1227–1239.
23. Zhou, W.; Ming, D.; Lv, X.; Zhou, K.; Bao, H.; Hong, Z. SO-CNN based urban functional zone fine division with VHR remote sensing image. *Remote Sens. Environ.* **2020**, *236*, 111458. [[CrossRef](#)]
24. Song, J.; Lin, T.; Li, X.; Prishchepov, A.V. Mapping urban functional zones by integrating very high spatial resolution remote sensing imagery and points of interest: A case study of Xiamen, China. *Remote Sens.* **2018**, *10*, 1737. [[CrossRef](#)]
25. Shen, Y.; Karimi, K. Urban function connectivity: Characterisation of functional urban streets with social media check-in data. *Cities* **2016**, *55*, 9–21. [[CrossRef](#)]
26. Zhang, F.; Du, B.; Zhang, L. Scene classification via a gradient boosting random convolutional network framework. *IEEE Trans. Geosci. Remote Sens.* **2015**, *54*, 1793–1802. [[CrossRef](#)]
27. Deng, Y.; He, R. Refined Urban Functional Zone Mapping by Integrating Open-Source Data. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 421. [[CrossRef](#)]
28. Liu, X.; Tian, Y.; Zhang, X.; Wan, Z. Identification of urban functional regions in chengdu based on taxi trajectory time series data. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 158. [[CrossRef](#)]
29. Lefulebe, B.E.; Van der Walt, A.; Xulu, S. Fine-scale classification of urban land use and land cover with planetscope imagery and machine learning strategies in the city of Cape Town, South Africa. *Sustainability* **2022**, *14*, 9139. [[CrossRef](#)]
30. Grippa, T.; Georganos, S.; Zarougui, S.; Bognounou, P.; Diboulo, E.; Forget, Y.; Lennert, M.; Vanhuysse, S.; Mboga, N.; Wolff, E. Mapping urban land use at street block level using openstreetmap, remote sensing data, and spatial metrics. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 246. [[CrossRef](#)]
31. Yao, Y.; Li, X.; Liu, X.; Liu, P.; Liang, Z.; Zhang, J.; Mai, K. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 825–848. [[CrossRef](#)]
32. Qian, Z.; Liu, X.; Tao, F.; Zhou, T. Identification of urban functional areas by coupling satellite images and taxi GPS trajectories. *Remote Sens.* **2020**, *12*, 2449. [[CrossRef](#)]
33. Gao, Q.; Fu, J.; Yu, Y.; Tang, X. Identification of urban regions' functions in Chengdu, China, based on vehicle trajectory data. *PLoS ONE* **2019**, *14*, e0215656. [[CrossRef](#)] [[PubMed](#)]
34. Zhu, X.; Goldberg, A.B. *Introduction to Semi-Supervised Learning*; Springer Nature: Berlin/Heidelberg, Germany, 2022.
35. Zhu, C.; Cheng, G.; Wang, K. Big data analytics for program popularity prediction in broadcast TV industries. *IEEE Access* **2017**, *5*, 24593–24601. [[CrossRef](#)]
36. Yao, T.; Zhang, Y.; Guan, Q.; Mai, K.; Zhang, J. Identifying Multi-Level Urban Functional Structures Using Time-Series Taxi Trajectories. *J. Wuhan Univ.* **2019**, *44*, 875–884. [[CrossRef](#)]
37. Hu, Y.; Han, Y. Identification of urban functional areas based on POI data: A case study of the Guangzhou economic and technological development zone. *Sustainability* **2019**, *11*, 1385. [[CrossRef](#)]
38. Jiao, C.; Jiao, L.; Dong, T.; Gu, Y.; Ma, Y. Quantitative Identification of Urban Functional Areas Based on POI Data and Its Visualization. *Surv. Mapp. Geogr. Inf.* **2016**, *41*, 68–73. [[CrossRef](#)]
39. Gong, P.; Chen, B.; Li, X.; Liu, H.; Wang, J.; Bai, Y.; Chen, J.; Chen, X.; Fang, L.; Feng, S.; et al. Mapping essential urban land use categories in China (EULUC-China): Preliminary results for 2018. *Sci. Bull.* **2020**, *65*, 182–187. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.