

Article

Semantic-Enhanced Graph Convolutional Neural Networks for Multi-Scale Urban Functional-Feature Identification Based on Human Mobility

Yuting Chen ^{1,2} , Pengjun Zhao ^{1,2,3,*} , Yi Lin ⁴, Yushi Sun ⁴ , Rui Chen ^{1,3}, Ling Yu ^{1,3}  and Yu Liu ⁵ 

¹ Department of Urban Planning and Design, Shenzhen Graduate School, Peking University, Shenzhen 518055, China; yutingchen.ai@pku.edu.cn (Y.C.)

² Key Laboratory of Earth Surface System and Human-Earth Relations of Ministry of Natural Resources of China, Shenzhen 518055, China

³ School of Urban and Environmental Sciences, Peking University, Beijing 100091, China

⁴ Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, China

⁵ Institute of Remote Sensing and Geographical Information System, School of Earth and Space Sciences, Peking University, Beijing 100091, China

* Correspondence: pengjun.zhao@pku.edu.cn; Tel.: +86-156-116-85282

Abstract: Precise identification of spatial unit functional features in the city is a pre-condition for urban planning and policy-making. However, inferring unknown attributes of urban spatial units from data mining of spatial interaction remains a challenge in geographic information science. Although neural-network approaches have been widely applied to this field, urban dynamics, spatial semantics, and their relationship with urban functional features have not been deeply discussed. To this end, we proposed semantic-enhanced graph convolutional neural networks (GCNNs) to facilitate the multi-scale embedding of urban spatial units, based on which the identification of urban land use is achieved by leveraging the characteristics of human mobility extracted from the largest mobile phone datasets to date. Given the heterogeneity of multi-modal spatial data, we introduced the combination of a systematic data-alignment method and a generative feature-fusion method for the robust construction of heterogeneous graphs, providing an adaptive solution to improve GCNNs' performance in node-classification tasks. Our work explicitly examined the scale effect on GCNN backbones, for the first time. The results prove that large-scale tasks are more sensitive to the directionality of spatial interaction, and small-scale tasks are more sensitive to the adjacency of spatial interaction. Quantitative experiments conducted in Shenzhen demonstrate the superior performance of our proposed framework compared to state-of-the-art methods. The best accuracy is achieved by the inductive GraphSAGE model at the scale of 250 m, exceeding the baseline by 25.4%. Furthermore, we innovatively explained the role of spatial-interaction factors in the identification of urban land use through the deep learning method.

Keywords: graph convolutional neural networks; semantic embedding; feature fusion; urban land use; human mobility; spatial interaction



Citation: Chen, Y.; Zhao, P.; Lin, Y.; Sun, Y.; Chen, R.; Yu, L.; Liu, Y. Semantic-Enhanced Graph Convolutional Neural Networks for Multi-Scale Urban Functional-Feature Identification Based on Human Mobility. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 27. <https://doi.org/10.3390/ijgi13010027>

Academic Editors: Wolfgang Kainz, Jiangfeng She, Jun Zhu and Min Yang

Received: 27 October 2023

Revised: 26 December 2023

Accepted: 4 January 2024

Published: 11 January 2024



Copyright: © 2024 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

Investigating urban land use is pivotal for building sustainable and habitable cities [1], and the measured influence of the environment on travel behavior in human mobility is shown to be sensitive to land use [2]. In this research field, one of the most longstanding and far-reaching geographic issues is the modifiable areal unit problem (MAUP), concerned with the spatial scale and the zoning sub-problem [3,4]. Moreover, appropriate spatial-unit zoning is significant for urban planning, urban governance, and synthetic studies of urban geography [5]. Given the spatial heterogeneity of urban spatial units, predicting unknown functional features of spatial units from the effective representation of units'

relationships is treated as an extensive challenge in urban science [6]. In addition, the various features of the connections among spatial units are always disorganized in the social sensing process, making it arduous to adopt characteristics of spatial interaction in an urban geographic-prediction issue [7,8]. The rapid development of deep learning mechanics has made remarkable advances in geospatial artificial intelligence (GeoAI) research [9,10]. The deep neural network brings a new direction to the novel representation of urban spatial units, which encodes a spatial unit into high-dimensional embedding for the complete preservation of various types of geospatial information [9,11]. Graph convolutional neural network (GCNN)-related methods can represent spatial structures graphically, while traditional neural network approaches are not powerful enough to interpret spatial dynamics [12]. With the potency of capturing different spans of connection relationships and transforming them into the weights of neural networks, GCNNs can further learn a deep representation of spatial units' functional features from complicated geographic contexts involving human activities [13].

A comprehensive embedding of urban spatial units consists of physical feature vectors and human feature vectors [14,15]. The physical vector directly conveys information on the built-environment attributes of the individual urban spatial unit. The human vector, concerned with social-economic activities, implies estimable information on how citizens generate and operate urban spaces with their daily demands [16,17]. In other words, human activities carried out in different functional urban regions may reshape a place's land-use type [18]. Spatial interaction represented by human mobility as flows of population [19] can capture complementary dependencies in different ranges [15,20,21]. A profound understanding of spatial interaction favors optimizing crisis management, mitigating epidemic spreading, forecasting traffic status, and many other applications [22]. On the city scale, spatial interaction can dynamically indicate functional land use and urban structure through spatial-temporal human mobility patterns from a collective perspective [16,23,24]. Moreover, the urban land-use type is prone to diverge from official plans, due to the sophistication of urban evolutions; thus, it is of great necessity to identify and monitor functional land use for urban infrastructure construction, transportation modernization, resource assignment, ecology protection, and disaster assessment [25,26].

The identification of urban land use is an interdisciplinary problem related to the prediction of spatial units' functional features, which relies on the observed characteristics of the individual unit and the connected spatial units. Early studies that explicate the information of such kinds of geographic context as predefined mathematical functions cannot decently model the intricate essence of the urban system [8]. While most existing studies apply GCNNs to the embedding of overall patterns and flow rates or amounts of spatial interaction, fewer of them have in-depth discussions on integrating the semantics of mobility patterns and socioeconomic attributes into the GCNN framework for the better prediction of spatial-unit features. One of the major reasons is the limitation of the range and contents of their datasets. The deliberate arrangement of spatial units will considerably increase the prediction accuracy in GCNN-based deep-learning tasks; however, the spatial scale effect has not been validated in the identification of urban land use from spatial interaction with graph embedding [5]. In addition, due to the deficiency of transparency for most deep learning models, explainability continues to be a major concern in attaining greater insight into the characteristics underlying mobility flows [27].

This study aims to investigate the scale effect and semantic characteristics of spatial interaction in identifying functional features of urban spatial units. We developed a semantically enhanced GCNN framework to learn the multi-scale embedding of urban spatial units and the dynamic identification of functional urban land use by leveraging the geographic context of spatial interaction. The key contributions of this study are fourfold. First, we introduced a hierarchical data-alignment method for the spatial segmentation and mapping of urban land-use data and spatial interaction with the assemblage of individual human mobility data. Second, we primarily proposed a heterogeneous feature-fusion method to extract multi-dimensional attributes of spatial interaction from a massive amount of

mobile phone data. Third, we expanded the graph-construction method by aggregating the distributed spatial interaction information to their origination and destination at different spatial scales. Fourth, we achieved the goal of explaining the features' contribution to the deep-learning tasks of land-use identification. Quantitative experiments comparing the performance of various backbone GCNN models empirically demonstrate that the semantic-enhanced GCNN framework is more powerful than the state-of-the-art methods.

2. Related Work

2.1. Place Embedding with Graph Convolutional Neural Networks

The embedding of urban spatial units can be taken as a vector representation of their spatiotemporal characteristics, to manifest the similarity among urban spatial units in the vector space [28]. In the initial phase of embedding, unsupervised text encoding algorithms broadly applied in Natural Language Processing (NLP) tasks, such as Word2Vec [29,30] and Doc2Vec [31], and homologous subjects bottomed on graph theory, such as Node2Vec [32], Place2Vec [33], and Traj2Vec [34], have fostered significant development in place-embedding research. However, an enormous loss of information happens when spatial units are transformed into word sequences in a document by the Word2Vec model, because spatial units usually possess multi-dimensional features, whereas word sequences are dispensed in one-dimensional space [35]. In other words, the aforementioned methods cannot currently involve node features in graph embeddings. To overcome this limitation, a notable increase in interest has arisen in graph convolutional neural networks (GCNNs), as they can master spatial-unit embedding even when the scope of unit connections shifts [36].

The backbone models of GCNNs integrate a graph convolutional network (GCN), a relational graph convolutional network (R-GCN), inductive and directed graph sample and aggregate (GraphSAGE), and a graph attention network (GAT) [37–40]. GCN utilizes a competent layer-wise propagation rule for neural network models that directly deal with graphs [37]. On the strength of GCN, relation-specific transformations are brought to R-GCN to proceed with the highly multi-relational features of knowledge bases [38]. In preference to building discrete embeddings per node, GraphSAGE samples and aggregates the features from a node's local space to produce its embedding [39,41]. While GraphSAGE and GCN are learned in an unsupervised way, where nodes co-occurring is made along with short random walks in the embedding space, attributed network embedding (Attribute to Vector, Attri2Vec) is learned by capturing the context node similarity over a network structure-guided transformation in the original attribute space [42].

Thanks to the emergence of geospatial big data collected from location-based services (LBSs) and the Internet of Things (IoT), an increasing number of researchers have taken advantage of GCNNs to investigate urban issues, e.g., traffic prediction [43,44], urban land-use recognition [24,45], urban scene classification [25], urban security perception [46], public health evaluation [47], weather forecasting [48], and cultural association mining [49]. Nevertheless, the implicit semantics and contextual information of geospatial big data are underexploited in the identification of urban functional features. Regarding the application of GCNNs in embedding urban spatial units, an important issue to be addressed is the effective representation of multi-dimensional features that can be adaptive to most backbone models.

2.2. Advanced Data and Methods for Urban Land-Use Identification

Research in identifying functional urban regions or land-use types has evolved from rule-guided to data-driven, with a bottom-up strategy, over the past decades [50]. The data sources applied in this field vary from static data to dynamically recorded data. The static data mainly include remote sensing images, street view images, points of interest (POIs), building footprints, and traditional survey statistics [51–54]. The dynamically updated data consist of mobile phone signal data, social media data, taxi traces, smartcard records of metros or buses, and other human activity data [55–59]. High-resolution remote

sensing images and street view images are effective for deriving urban land covers, yet they struggle to continuously acquire the social-economic characteristics of urban spatial units, due to the accessibility of data and the shadows of abundant skyscrapers in high-density cities [60,61]. POI data is popular in urban studies for its facilitated accessibility, but the problem of widespread noise should not be disregarded [62]. Compared with static data sources, dynamically updated human mobility data have the bulge on extensive data coverage, diverse spatiotemporal resolution, and high sensitivity to social functional attributes [60,63,64].

The progression of research methods conducted in urban land-use identification contains two stages: one focuses on density analysis and cluster analysis, while the other concentrates on data mining and deep learning. Most density-analysis and cluster-analysis approaches are generally acceptable for their hallmark of lightweight computation and time-saving functions [60]. However, for certain urban regions with spatial heterogeneity, the foregoing methods using shallow features of data may have limited identification accuracy [60]. Accordingly, scholars seek to employ data-mining methods like topic models, which can generate latent semantics for delineating functional urban zones [65]. Further, a growing group of deep-learning methods has been widely applicable in urban functional-zone identification, including the Place2Vec model and the convolutional neural network (CNN)-based models, such as large-patch CNN and deep-feature CNN [66,67]. Those experiments prove that deep learning's migration capability and scalability have unlimited potential in urban land-use identification and smart cities [60,68]. In summary, a method to combine dynamically updated data and graph-based deep learning techniques for urban functional-feature identification is long-awaited. And the spatial scale effect, previously underappreciated, needs to be deeply explored in the current research stage.

3. Methodology

3.1. Overview

In this study, a semantic-enhanced GCNN framework (Figure 1) to investigate the relationship between spatial dynamics represented by spatial interaction and urban functional features is proposed. We start with the alignment of multi-modal datasets, including mobile phone signal data that contain user locations and mobile phone usage data, land use data, and city boundary data within the research area. As the datasets are of different spatial scales, we conduct spatial segmentation on the study area and map multi-modal data to urban spatial units. Three kinds of features are extracted and fused under feature engineering, to enrich the semantic properties of urban spatial units. We then build up the spatial interaction matrix, based on the aggregated trips extracted from mobile phone data. With the combination of the feature fusion matrix and spatial interaction matrix, a variety of graphs involving the information of geographic semantics and spatial topological structure are constructed. Finally, we introduce a series of backbone models under the GCNN framework to test their performance in urban spatial-unit embedding, presenting the fact that they are competitive with state-of-the-art models. Furthermore, we analyze the contribution of multiple features in deep-learning classification tasks.

3.2. Graph Representation

3.2.1. Building the Grid-Travel Corpus

Grid cells are typically regarded as a metric function for coding space for their utility in delivering multi-scale spatial representations in urban computing [69]. The entire urban region is divided into uniform grids, as the urban spatial unit at different scales for the research demands. The human mobility involving travel between origin–destination (OD) grids is systematically documented on an hourly basis, encompassing both the travel and social attributes outlined in Table 1. Travel attributes primarily delineate the physical characteristics of mobility, while social attributes provide a comprehensive depiction of individuals engaged in travels. These attributes within the mobility data or OD data vividly

showcase the diverse nature of human mobility, indicating a significant correlation between social structures and urban spaces [70,71].

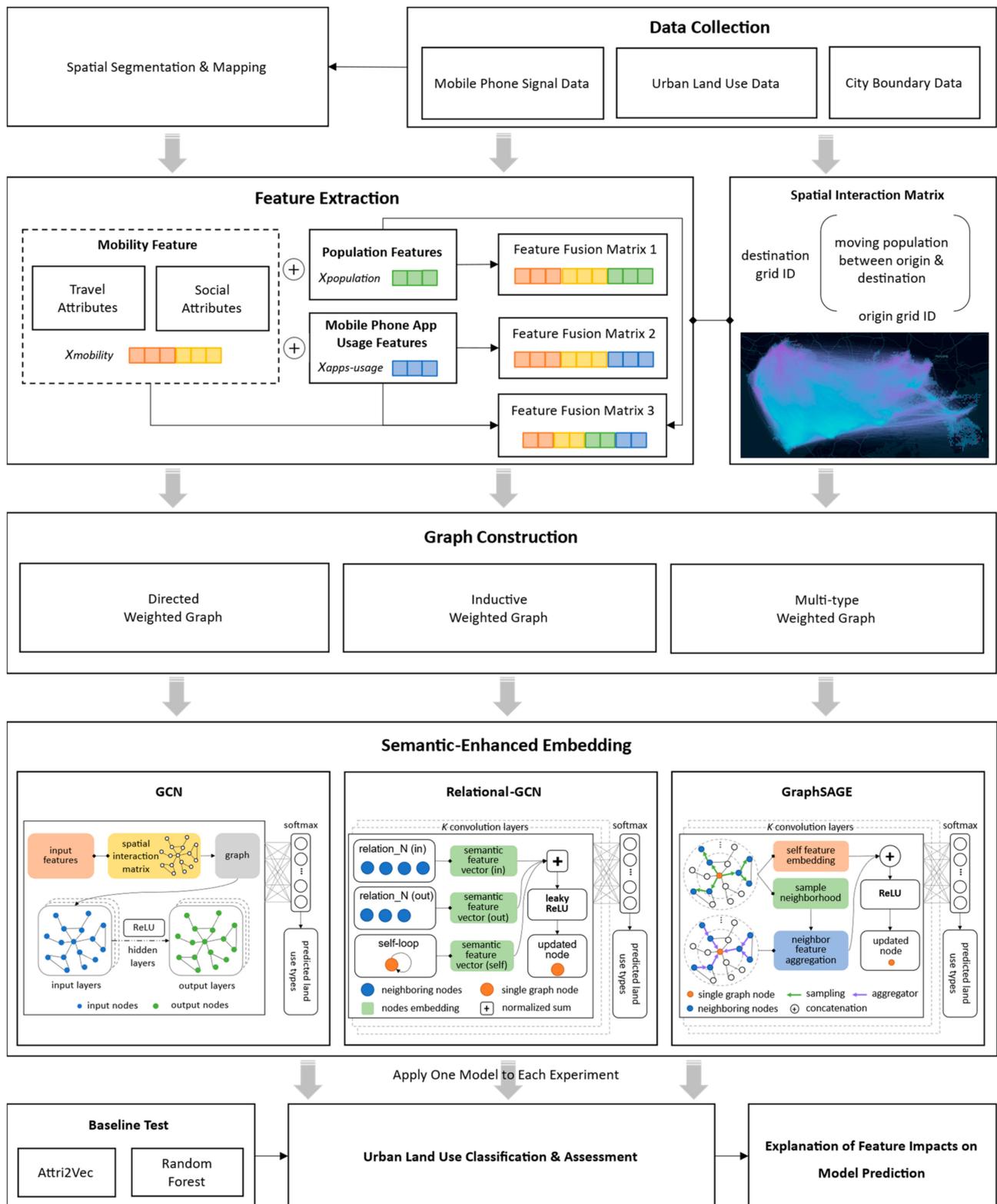


Figure 1. Overview of semantic-enhanced GCNN framework.

Table 1. Travel and social attributes for characterizing human mobility.

	Name	Connotation	Examples
Travel Attribute	Travel Type	The purpose of the travel	commuting, dwelling, etc.
	Transport Mode	The means of transportation	subway, highway, airplane, train, etc.
	Travel Time	The exact hour of departure and arrival time and the total time (h) spent on the trip	departure time: 14:00 arrival time: 16:00 total time: 2 h
	Distance	The distance (km) between the origin and destination	15 km, 30 km, etc.
	Speed	The average speed (km/h) over the total distance	40 km/h, 70 km/h, etc.
	Date	The date and the type of day that the traveling happens	5 November 2019, workday 23 November 2019, weekend
Social Attribute	Gender	Gender group of travelers	Male, Female
	Age	The age group of travelers	Youth, Middle-aged, The elderly, etc.
	Occupation	Occupational classification of travelers	Financier, Teacher, Doctor, Administrator, Farmer, etc.
	Education	Educational qualification of travelers	B.D., M.D., Ph.D., etc.
	Income	Income level of travelers	Ten levels, from 1 to 10
	Probability of Owning Cars	The probability level that a traveler owns a car	Five levels: high, middle, low, etc.
	Probability of Owning Houses	The probability level that a traveler owns a house	Five levels: high, middle, low, etc.

To infer urban functional features with human mobility patterns, the traveling population flows as the link between urban spatial units, and its attributes should be aggregated to the origin and destination grids. Given the data heterogeneity of travel attributes and social attributes, a self-adapting data-normalization method is proposed to support the training of the GCNN models. Above all, those attributes are classified as continuous attributes and discrete attributes (categorical attributes). The continuous attributes include travel time, distance, and speed. The average and standard deviation of these data are calculated based on the origin grid (O-grid) and destination grid (D-grid), separately. The discrete attributes include travel type, transport mode, date of travel, people's age, gender, occupation, education, income, and the probability of owning a car or a house. The frequency at which the value of those discrete attributes appears is primarily calculated as the ratio. Since the attribute value is also important for feature construction, we multiply the ratio with the attribute value to obtain the weight of the attribute. The average and standard deviation of attribute ratio and weight are then obtained from the O-grid and D-grid, respectively.

Apart from the property of human mobility as an intrinsic attribute of urban spatial interaction, the intensity and frequency of spatial interaction also play a significant role in quantifying spatial interaction and delineating urban functional features. The intensity of urban spatial interaction can be defined as the volume of flows between urban spatial units, which is obtained by summarizing the number of travelers per hour. The number of travelers departing from the O-grid is calculated as "out-population", while the number of travelers arriving at the D-grid is calculated as "in-population"; the difference between "out-population" and "in-population" of each grid is defined as "stay-population". The frequency of urban spatial interaction can be defined as the ratio of flows between urban spatial units, which is acquired by computing the ratio of travel numbers of origin-

destination pairs compared to the total travel records. Specifically, the travel ratio of O-grids is calculated as “out-frequency”, while the travel ratio of D-grids is calculated as “in-frequency”.

3.2.2. Generating a Feature Matrix from Multi-Modal Data Fusion

To probe how the semantic features of urban spatial units influence the performance of identifying urban functional features, the grids are further characterized by fusing the OD data with the other two types of human activity data derived from mobile phone use records. The first type is the monthly population (Pop) of each grid, which covers the population of working, dwelling, and visiting activities. A series of social attributes are extracted to present a comprehensive population portrait of urban spatial units, such as gender, age, occupation, education, etc. With similar social attributes of OD data, Pop data shares the same aforementioned normalization method. The second type of human activity data is the monthly mobile phone data usage of different mobile apps (Flux) located in each grid. These mobile apps are classified into six categories, including communication, entertainment, business, news, tools, and domestic services. The total mobile phone data usage of each category is calculated based on separate grids. Utilizing OD data as the primary features of urban spatial units, the normalized Pop data and Flux data are individually integrated with the OD data. Consequently, three distinct node feature matrices (OD and Pop, OD and Flux, and OD and Pop and Flux) are derived through multi-dimensional data fusion, serving as the foundation for graph construction in the context of urban spatial analysis.

3.2.3. Constructing Semantic-Enhanced Graph

A graph is constructed with nodes and with edges linking between nodes. While the nodes refer to urban spatial units, the edges correlate with the spatial interactions between nodes. The grid features are taken as node features, and the intensity of spatial interaction between grids is taken as edge weights in the graph. Three kinds of graphs are constructed in this study: directed weighted, inductive weighted, and multi-type weighted. In the directed graph, the origin and destination grids are taken as the source node and target node, respectively. The inductive weighted graph is undirected, and its nodes are treated as the same type. The multi-type weighted graph is a heterogeneous directed graph, where various types of edges symbolize different relationships between nodes. In our multi-type graph, two types of edges are constructed by, respectively, in-flow and out-flow, the weight of which is represented by in-population and out-population calculated in Section 3.2.1.

A directed graph is comprised of urban spatial interactions reflected by human mobility flows (m) between spatial units, which are aggregated from the travel trajectories of a crowd of people:

$$G_m = (V_m, E_m) \quad (1)$$

where V_m signifies a variety of graph nodes, and each node $v_i \in V_m$ represents an urban spatial unit $u_i \in U$, composed of origin nodes N_o and destination nodes N_d of trajectories:

$$V_m = N_o \cup N_d \quad (2)$$

E_m represents a set of edges, and each edge $e_{ij} = (v_i, v_j, a_{ij}) \in E_m$ stands for a directed mobility trajectory between two nodes v_i and v_j in end-to-end spatial interaction. The interaction intensity between urban spatial units is thus encoded as edge weights a_{ij} . Edge directions are consistent with trajectory directions.

For a multi-type graph, the edge type that represents the relations between nodes is defined as R in the multi-type graph:

$$G_m = (V_m, E_m, R_m) \quad (3)$$

and then each edge is refined as $e_{ij} = (v_i, r, v_j, a_{ij}) \in E_m$ where $r \in R_m$ is a relation type. Moreover, E_m in the multi-type graph is defined as the aggregation of bidirectional edges, namely in-edge (E_{in}) and out-edge (E_{out}):

$$E_{m(\text{heterogeneous})} = E_{in} \cup E_{out} \quad (4)$$

For each node $v_i \in V_m$ in graph G_m , the local attributes of urban spatial units and the spatial interaction attributes are encoded as node features $x_k \in X_v$. Specifically, X_v is a subset of the fusion of the feature matrix, which is the concatenation (\oplus) of the urban mobility features, local population features, and mobile phone data usage of various apps:

$$X_v \subseteq X = X_{\text{mobility}} \oplus X_{\text{population}} \oplus X_{\text{apps-usage}} \quad (5)$$

3.3. Graph Convolution

In this study, the problem of identifying urban land-use type from the characteristics of urban spatial units and the spatial interaction between them can be defined in a straightforward form:

$$Z = J(X_v, G_m) \quad (6)$$

where J is the general notation of the GCN model designed for the node-classification task in graph-embedding issues. The propagation model and loss function of the GCNN backbones we used will be further discussed in the following sub-sections.

3.3.1. GCN

In the GCN framework, the embedding of the urban spatial unit is conducted in K convolution layers; thus, each node v_i in the graph has K node embeddings through hidden state $h_i^{(k)} \in H^{(K)}$. While the node features X_v are applied to the initial hidden layer $H^{(0)}$, the prediction of node attributes is obtained from the embedding $H^{(K)}$ in the last layer. A layer-wise propagation rule is followed by the multi-layer GCN [72].

$$H^{(k+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}} \tilde{S} \tilde{D}^{-\frac{1}{2}} H^{(k)} W^{(k)}\right) \quad (7)$$

In particular, $\tilde{S} = S + I_N$ is a weighted spatial interaction matrix $S \in \mathbb{R}^{N \times N}$, with added self-connections. It denotes the topological structure and edge weights of graph G_m . I_N is the identity matrix, and N typifies the node number of graph G_m . $W^{(k)}$ is the learnable weight matrix specified by layers, and $\tilde{D}_{ii} = \sum_j \tilde{S}_{ij}$ is a diagonal degree matrix. Following the layer-wise propagation rule discussed in Equation (7), we apply a two-layer GCN model to our study for supervised node classification, which is an instantiated form of Equation (6):

$$Z = f(X_v, S) = \text{softmax}\left(\hat{S} \text{ReLU}\left(\hat{S} X W^{(0)}\right) W^{(1)}\right) \quad (8)$$

Here, X is the input data that serve as $H^{(0)}$ in Equation (7). $\hat{S} = \tilde{D}^{-\frac{1}{2}} \tilde{S} \tilde{D}^{-\frac{1}{2}}$ is pre-calculated, based on the symmetric weighted matrix \tilde{S} . $W^{(0)} \in \mathbb{R}^{C \times H}$ is a weight matrix for an input layer to a hidden layer with C input channels and H feature maps. $W^{(1)} \in \mathbb{R}^{H \times F}$ is a weight matrix for a hidden layer to an output layer with F classes. The activation functions softmax and ReLU are applied row-wise. Given that the identification of urban land-use types belongs to supervised multi-class node classification, we estimate the categorical cross-entropy error for all nodes with land-use labels:

$$\mathcal{L} = - \sum_{v \in V_m} \sum_{f=1}^F Y_{vf} \log Z_{vf} \quad (9)$$

where Y is denoted as node labels. The weights of GCN are trained using gradient descent, and the loss function above is applied to tune the weights to predict better results.

3.3.2. Relational GCN

Motivated by the architecture of GCN, in the “ $k + 1$ ”-th convolutional layer of a relational multi-type graph, a node represented by v_i is updated by the forward pass, and the propagation model for calculating it is defined as the following:

$$h_i^{(k+1)} = \sigma \left(\sum_{r \in R_m} \sum_{j \in N_i^r} \frac{1}{c_{i,r}} W_{j,r}^{(k)} h_j^{(k)} + W_0^{(k)} h_i^{(k)} \right) \quad (10)$$

where N_i^r depicts the group of node indices under relation $r \in R_m$ as the neighbors of node i . Moreover, $c_{i,r}$ is a learnable constant for task-oriented normalization. In contrast to the regular GCN model, the R-GCN model conducts transformations for individual relations, depending on the type and direction of its edge.

A critical problem of the aforementioned propagation model is that the increase in relation number in the graph will lead to a sharp rise in the number of parameters, which often becomes a primary contributor to overfitting. To address this issue, we apply basis-decomposition to regularize the weights of R-GCN layers in this study, where each $W_r^{(k)}$ is defined as follows:

$$W_r^{(k)} = \sum_{b=1}^B a_{rb}^{(k)} V_b^{(k)} \quad (11)$$

This is a linear combination of basis transformation $V_b^{(k)} \in \mathbb{R}^{d^{(k+1)} \times d^{(k)}}$ with coefficients $a_{rb}^{(k)}$, which depend on r relation type.

3.3.3. GraphSAGE

The main purpose of GraphSAGE is to produce an approximate representation of adjacent nodes by aggregating feature information from the local node’s neighborhood; meanwhile, the representations of dissimilar nodes are quite distinctive. The loss function designed for the output representations is shown in Equation (12).

$$\mathcal{L}(z_u) = -\log \left(\sigma \left(z_u^T z_v \right) \right) - Q \cdot \mathbb{E}_{v_n \sim P_n(v)} \log \left(\sigma \left(-z_u^T z_{v_n} \right) \right) \quad (12)$$

Specifically, v and u are nodes that coexist, and they can reach each other on a random walk of definite length; σ is the nonlinear activation function for the representations of vertex z_u ($\forall u \in V_m$). P_n denotes the distribution of negative sampling, and the negative samples are counted by Q . Incompatible with preceding embedding methods, a constant number of neighbors $N(v)$ are evenly sampled from the set $[u \in V_m : (u, v) \in E_m]$ in GraphSAGE. Moreover, the erratic node neighborhood features are operated by aggregator functions with symmetry properties.

$$\text{AGGREGATE}_{E_k}^{\text{pool}} = \text{mean} \left(\left[\sigma \left(W_{\text{pool}} h_{u_i}^k + b \right), \forall u_i \in N(v) \right] \right) \quad (13)$$

Here, *mean* denotes the element-wise mean operator, and b is an optional bias toward each node’s aggregated embedding in the given graph. The reason we employ a mean pooling aggregator is to efficiently capture alternative aspects of the node’s neighborhood.

4. Implementation and Results

4.1. Study Area and Data Description

The study area Shenzhen (Figure 2) is one of the special economic zones of China. This port city is located on the east coast of the Pearl River Delta, bordering Hong Kong to the south. Two types of data are utilized in this study: mobile phone data and land use data. The mobile phone data were formally collected from China Unicom (China United Network Communications Group Co., Ltd. in Beijing, China), one of the largest

national communication operators. Those long-sequence mobile phone data convey three types of information. The first is users' hourly intracity mobility data between OD grids, characterized by travel attributes and social attributes. The second is the population data of dwelling, working, and visiting in each grid. The third is the mobile phone usage data of different apps per grid. The mobile phone data covers three months of 2019: March, September, and December. Regarding the label category of urban spatial units, the ground truth of land-use data was collected from the third official National Land Resource Survey, entitled "Classification of Land Use Status", which was launched by Shenzhen State Council in September 2018 (GBT 21010-2017). An approximation strategy for labeling a spatial unit is to select the land-use type with the largest area in the spatial unit as the unique identifier. The spatial resolution was set as the grid of 250 m, 500 m, and 1000 m, respectively. In addition, the WGS84 geographic coordinate system and UTM 50 N project system were adopted in this study. Figure 3 illustrates the spatial distribution of Pop data and Flux data in the study area. Based on those data, a set of heterogeneous graphs was systematically constructed across three distinct spatial scales, as delineated in Table 2.

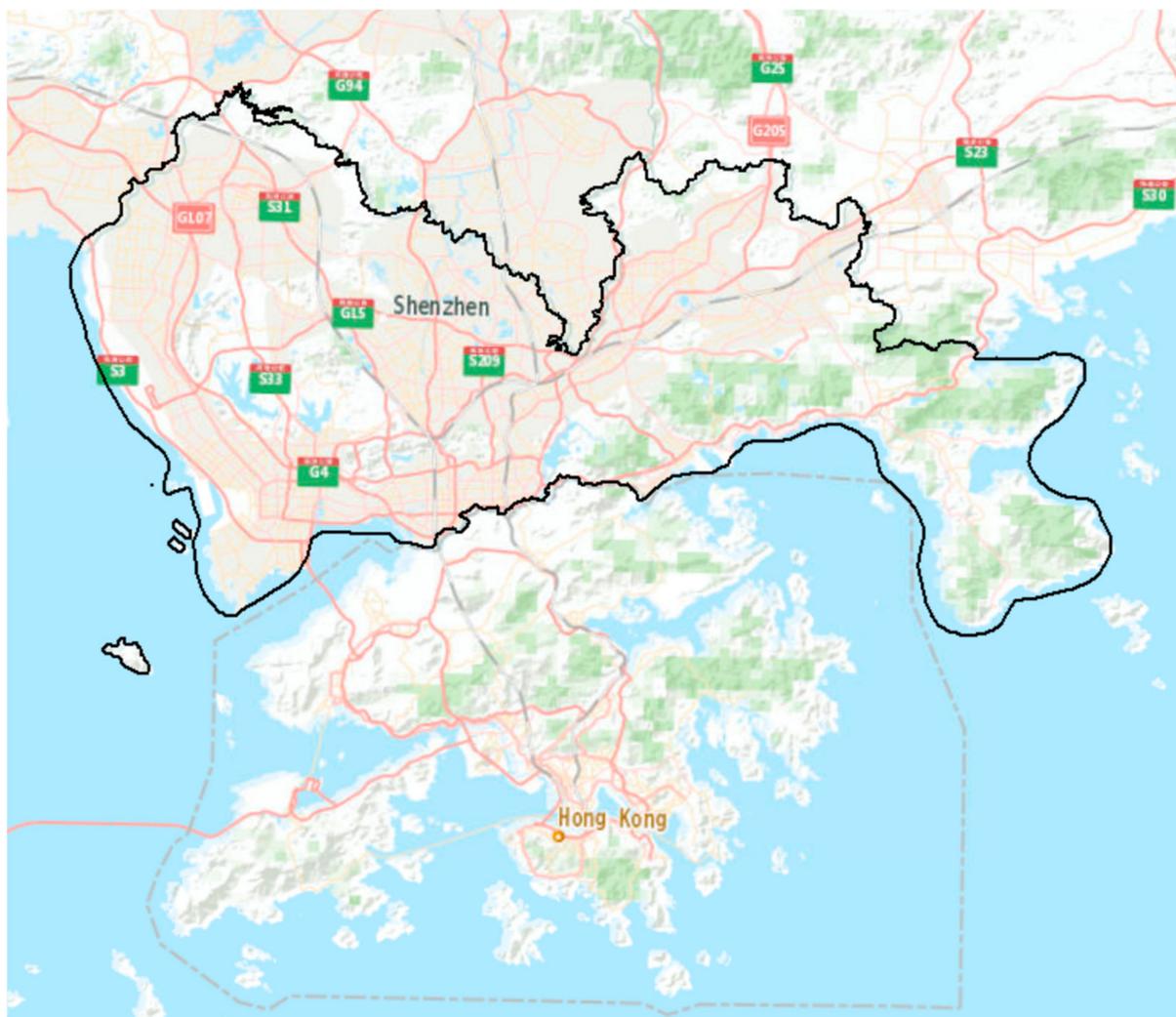


Figure 2. Location of the study area.

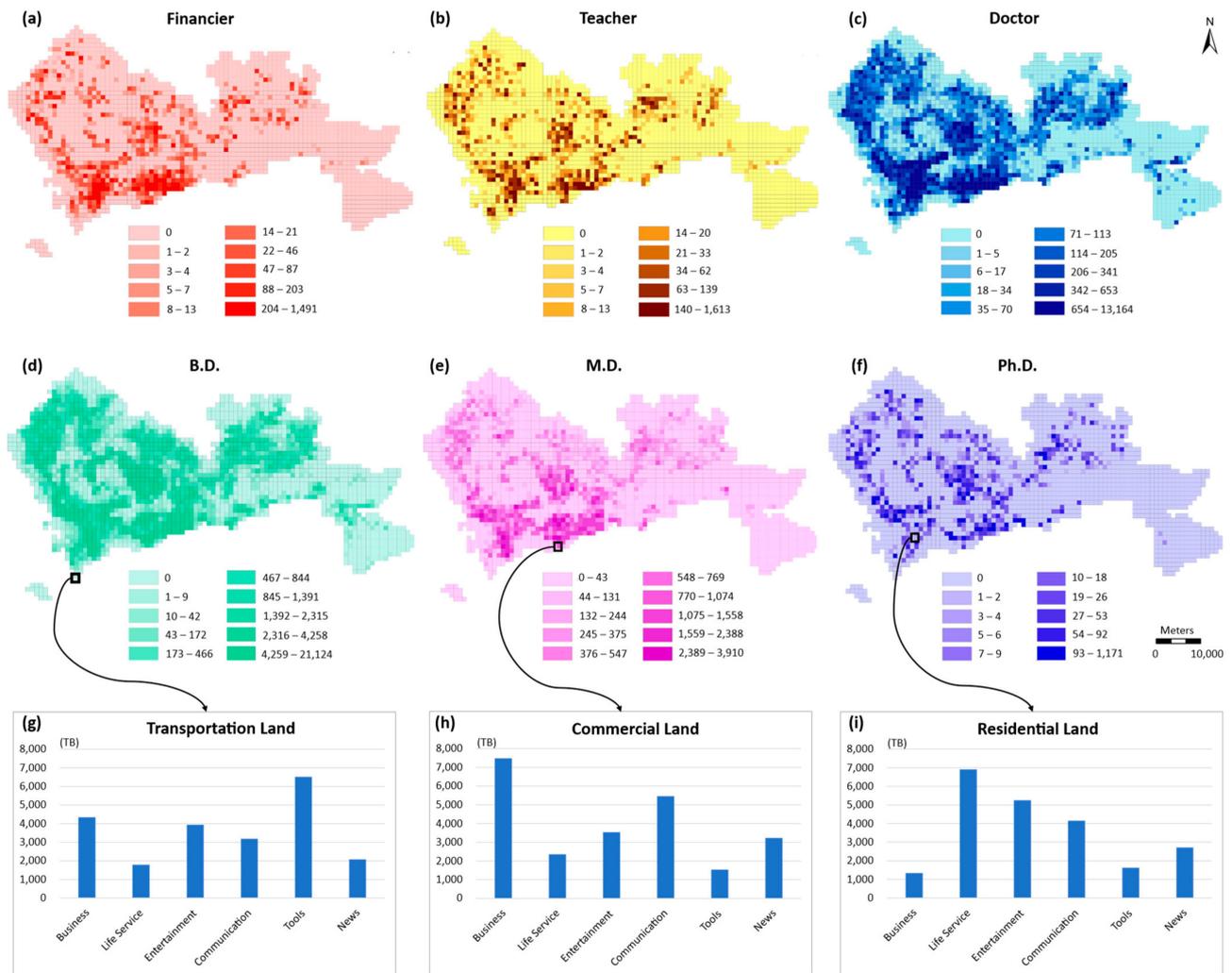


Figure 3. Examples of diverse demographic distribution on the spatial scale of 1000 m. (a–c): Demographic distribution of occupational classifications in Shenzhen; (d–f): Demographic distribution of educational qualifications in Shenzhen; (g–i): Mobile phone usage data of various apps in the spatial unit with different land-use types.

Table 2. Description of heterogeneous graph structures.

Spatial Scale	Node Number	Edge Number	Data Source	Feature Dimension	Edge Weighted	Direction Type
250 m	28,180	3,753,124	OD	112	✓	Directed and Undirected
			OD and Flux	118	✓	
			OD and Pop	158	✓	
			OD and Pop and Flux	164	✓	
500 m	7838	2,376,382	OD	112	✓	Directed and Undirected
			OD and Flux	118	✓	
			OD and Pop	158	✓	
1000 m	2133	835,408	OD and Pop and Flux	164	✓	Directed and Undirected
			OD	112	✓	
			OD and Flux	118	✓	
			OD and Pop	158	✓	
			OD and Pop and Flux	164	✓	

4.2. Geo-Semantic Embedding and Prediction

The identification of urban functional features such as land use according to spatial graphs is intrinsically a node-classification task in geo-semantic embedding and prediction issues. To contextualize the empirical results of the proposed method, we compared it against the state-of-the-art baselines including Random Forest and Attri2Vec, and four backbone GCNN models: GCN, relational-GCN, inductive GraphSAGE, and directed GraphSAGE. All models were implemented in the TensorFlow framework with the Stelargraph library. In terms of the experimental setup, 80% of the dataset was prepared for training, while the remaining 20% was used for validation and testing. For the urban land-use-identification task, we applied a three-layer model in all GCNN frameworks introduced in this study. In the GCN framework, the model works with a layer size of 128 for the first two layers, followed by a rectified linear unit (ReLU) activation for those hidden layers, as well as a softmax activation for the last output layer. In the GraphSAGE framework, the model executes with a batch size of 256 in both inductive and directed GraphSAGE. The number of in-samples and out-samples is set as 64 for each hidden layer in the directed GraphSAGE, while the same number is set for the undirected samples in the inductive GraphSAGE. In the R-GCN framework, the first two layers are activated by leaky rectified linear units (Leaky ReLU), whereas the output layer is normally activated by softmax for node classification. As R-GCN uses basis-decomposition for the regularization of layer weights, we chose 30 as the number of basis-decomposition, based on the validation set performance. The architectural hyperparameters of those tasks had been optimized on randomly sampled datasets. We set the drop rate as 0.3 for all layers, and imposed the bias to True. All models were trained using the Adam optimizer, to minimize categorical cross-entropy on the training nodes, implemented by an initial learning rate of 1×10^{-3} and a decay of 1×10^{-4} . We used an early stopping strategy with a patience of 200 epochs on both the accuracy and the cross-entropy loss.

4.3. Results of Urban Functional Feature Identification

4.3.1. The Performance of Multi-Modal Data Fusion

The semantic information provided by node features plays a significant role in the GCN-based node-classification task. Assuming that the performance of urban functional feature identification will improve with the increase in the amount of semantic information for urban spatial units, we conducted experiments on three multi-modal data fusion matrices, using Random Forest as the baseline and GCN as the representation of GCNN models. Results abbreviated in Table 3 convey overall classification accuracy (OA) and Kappa coefficient in percentages to evaluate the performance of the model and dataset. The results indicate that the data fusion matrix of population data and mobile phone app usage data outperformed the semantic feature matrix generated from single human mobility data in tasks executed by both the Random Forest and the GCN model. We can conclude that the integration of diverse human activity data will provide valuable information to the GCN model for identifying urban functional features.

Regarding the performance in various urban spatial scales, the scale of 250 m, with an OA of 82.59%, outperformed the scale of 500 m and 1000 m over four multi-modal data-fusion scenarios. In particular, the scale of 250 m outperformed the scale of 500 m by 4.9%, on average, while it outperformed the scale of 1000 m by 15.2%, on average. It turns out that the performance of urban functional-feature identification is negatively related to the spatial scale. Additionally, we found that the classification results of different land-use types are correlated with their proportions in spatial units. Taking the spatial scale of 1000 m as an example in Figure 4, residential land and industrial land with a large proportion of spatial units have higher F1-scores in the classification results. Non-construction land accounts for the largest proportion of spatial units, but its F1-score does not change drastically with variation in data-fusion scenarios, like other land types. An essential reason is that human activities have little interaction with non-construction land.

This further illustrates the fact that the semantic information on human activities plays a significant role in identifying urban functional features.

Table 3. Model performance in four data-fusion scenarios at different spatial scales.

		Random Forest		GCN	
		OA (%)	Kappa (%)	OA (%)	Kappa (%)
250 m	OD	63.16	49.54	74.79	60.51
	OD and Pop	65.31	51.85	77.36	62.72
	OD and Flux	67.36	52.93	79.48	65.83
	OD and Pop and Flux	69.78	54.24	82.59	69.25
500 m	OD	60.95	45.93	70.16	56.19
	OD and Pop	63.81	48.18	72.71	58.04
	OD and Flux	64.97	49.16	74.59	60.93
	OD and Pop and Flux	67.02	51.63	78.14	64.01
1000 m	OD	55.72	39.84	63.48	47.23
	OD and Pop	57.29	40.52	65.61	49.56
	OD and Flux	59.25	41.96	67.34	52.59
	OD and Pop and Flux	62.98	44.35	70.79	55.74

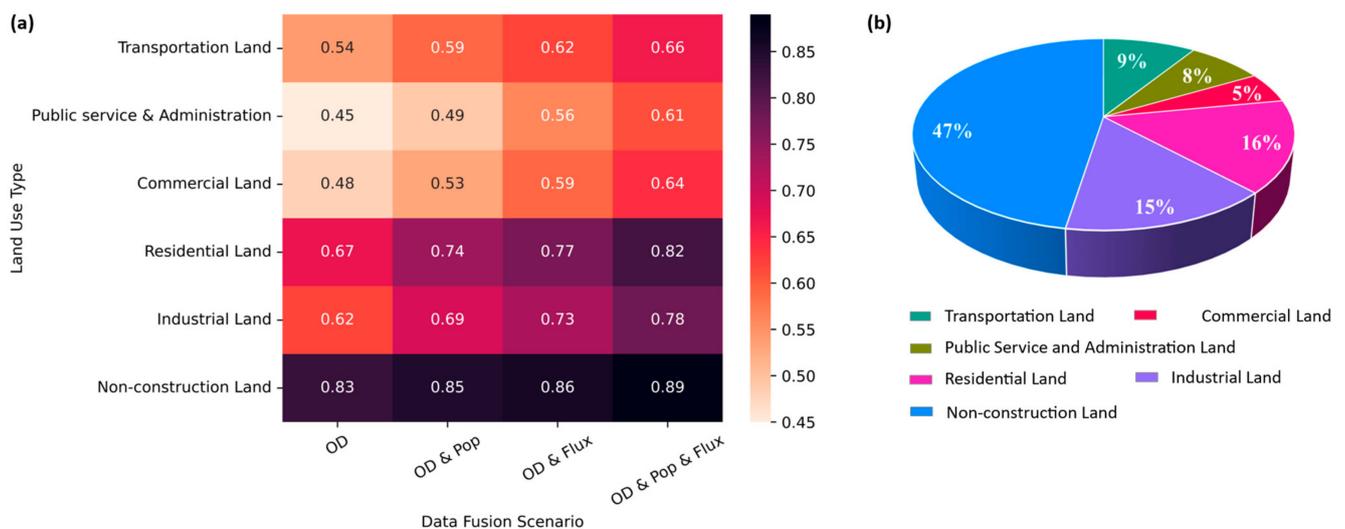


Figure 4. (a) F1-score of land-use types predicted by the GCN model in different data-fusion scenarios at the spatial scale of 1000 m. The F1-score distribution of the other two spatial scales is similar to the scale of 1000 m. (b) The proportion of units of each land-use type in the total number of spatial units.

4.3.2. The Performance of Different Embedding Models

Given that the semantic feature of the graph network constructed from integrating three kinds of human activity data outperformed other data-fusion scenarios in the task of urban functional-feature identification, we employed this graph network to test the performance of different embedding models at three urban spatial scales. Results archived in Table 4 connote OA and Kappa coefficient in percentages. The performance of all those models at small urban spatial scales was superior to those at large urban spatial scales. For the fine spatial scale of 250 m, the mean accuracy of all models was 80.19%. The GraphSAGE model strongly outperformed other GCNN models and baselines, and exceeded the mean accuracy by 9.1%. The inductive one, with an accuracy of 87.49%, shows a modest improvement compared to the directed one. For larger spatial scales

of 500 m and 1000 m, the mean accuracy was 76.41% and 69.88%, respectively. At the scale of 500 m, the directed GraphSAGE model with an accuracy of 83.17% significantly outperformed other GCNN models and baselines, and exceeded the mean accuracy by 8.9%. At the scale of 1000 m, the R-GCN model, with an accuracy of 75.72%, outperformed other models and exceeded the mean accuracy by 8.3%. In summary, the GCNN models that emphasize the direction and relational type of graph edges are more suitable for identifying land-use types at middle and large spatial scales. The GCNN models that strengthen local characteristics of graph nodes are more competent in the identification of land-use types at small spatial scales.

Table 4. Performance of backbone GCNN models and baselines.

	250 m		500 m		1000 m	
	OA (%)	Kappa (%)	OA (%)	Kappa (%)	OA (%)	Kappa (%)
Random Forest	69.78	54.24	67.02	51.63	62.98	44.35
Attri2Vec	71.35	55.82	69.36	53.19	64.54	45.93
GCN	82.59	69.25	78.14	64.01	70.79	55.74
Relational GCN	83.71	70.56	81.04	67.15	75.72	61.92
Directed GraphSAGE	86.20	74.72	83.17	69.73	73.91	58.07
Inductive GraphSAGE	87.49	76.28	79.72	65.28	71.33	56.85

This result also brings conceptual insights into the embedding of urban spatial units. Both the directed GraphSAGE model and the R-GCN model exploit the edge direction that traces the directionality of spatial interactions. While the edge in directed GraphSAGE maps the one-directional relationship between the origin and the destination, the edge with bilateral types in R-GCN explicitly maps the bi-directional relationship between spatial units. The performance variations in directed GraphSAGE and R-GCN indicate that graph embeddings with the directionality of spatial interaction have a crucial impact on large-scale urban functional-feature identification. Regarding the inductive GraphSAGE model that achieves superior performance at the 250 m scale, it recognizes the node's local role in the neighborhood by incorporating adjacent nodes' features and topological structures. This leads to another noteworthy conclusion: the graph embedding with the adjacency of spatial interaction makes a distinctive contribution to identifying small-scale urban functional features.

4.3.3. Geospatial Distribution of Urban Functional Features

To investigate the geospatial distribution of urban functional features, we took the land-use identification results output from the GCN model at the scale of 1000 m as an example (Figure 5). The regions near the southern border of Shenzhen are economically prosperous, due to their superior geographical conditions and high population density. The land-use types and human activities in these regions are more complex for GCNN embedding. The prediction result in the north of Shenzhen is better than that in the south of the city. The results also show that residential land is the most distinguishable with our semantic-enhanced GCNN identification framework. Inferred from the average computation based on prediction results, over 56.1% of industrial land was successfully identified. Around 27.9% of commercial land was misclassified as residential land, for which an example is demonstrated by the middle arrow in Figure 5. One of the major reasons is that many commercial areas are located near residential areas, and there were more instances of residential land use in our dataset. Further supplementary information was required to precisely discriminate them. A noticeable issue is that no sound quality within the input features is sufficiently effective for recognizing public-service and administration land, which occurs around areas such as schools, hospitals, parks, etc. Most of these areas were misclassified as residential land instead, of which an instance is shown by the right arrow in Figure 5. Similarly, there were a small number of transportation land-use areas in our

dataset, and a number of industrial parks were located around the transportation land, to facilitate logistics. Thus, 16.8% of them were misclassified as industrial land, with more occurrences. An example is displayed by the left arrow in Figure 5. Finally, the highest accuracy appeared in the identification of non-construction land at all three spatial scales, because the sample size of this land-use type accounts for the largest proportion. The potential reason will be discussed in Section 5.

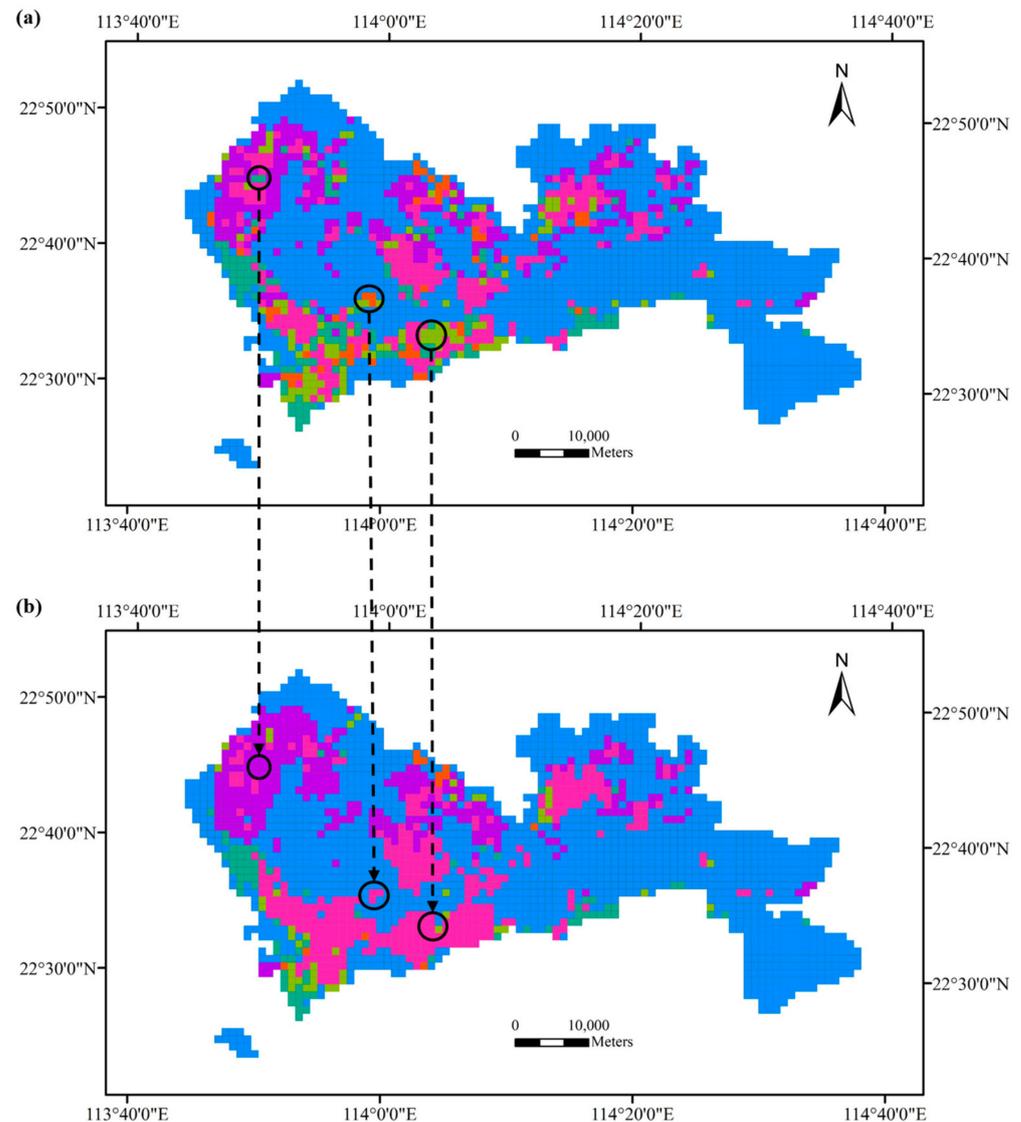


Figure 5. The identification of urban land use achieved by the GCN model at the spatial scale of 1000 m. (a) Ground truth of Shenzhen land use; (b) Identification results of Shenzhen land use.

4.3.4. Explanation of Feature Impacts on Model Prediction

Understanding why a model makes certain predictions is of great importance for results interpretation, the explanation of differences between models, and the extension of our understanding of the phenomenon under analysis [73,74]. We used Shapley Addictive exPlanations (SHAP) [75] to figure out how the input features of human mobility contributed to determining the output urban land use of the GCNN model. According to game theory, SHAP enumerates optimal Shapley values to measure the contribution of individual features. It indicates how the occurrence of certain features influences the model function, compared to the normal performance for the complete dataset. With the support of SHAP, we demonstrated global interpretations of feature impacts at three urban spatial

scales (Figure 6) and the corresponding local interpretations for six classes of urban land use at the spatial scale of 1000 m (Figure 7).

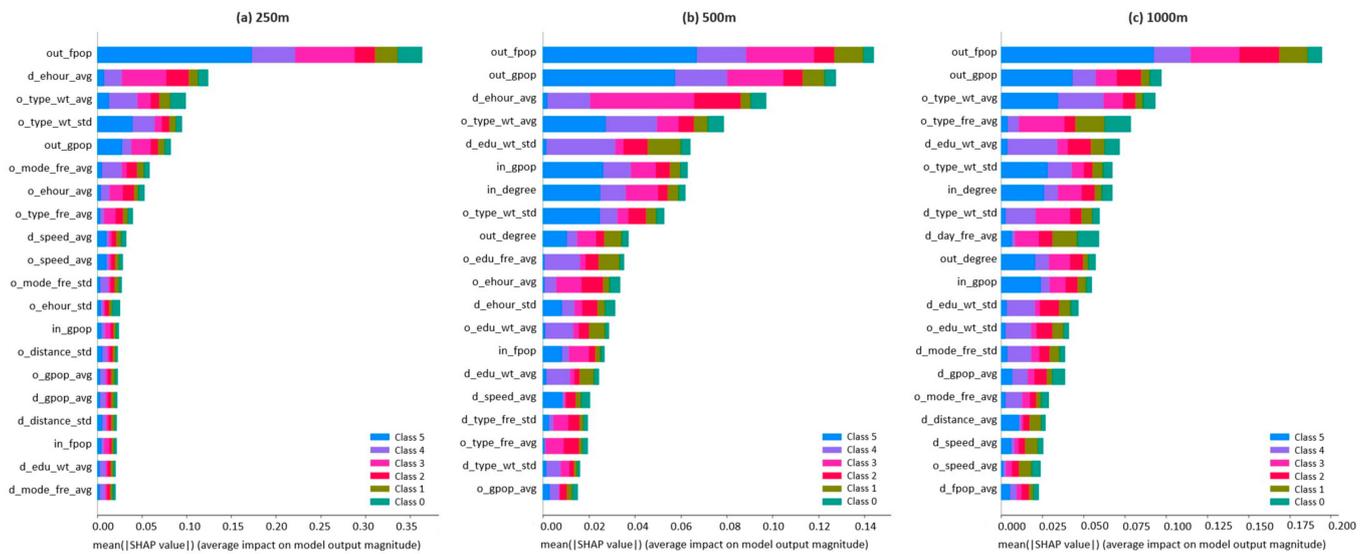


Figure 6. Global summary of distributed SHAP values for the top 20 contributed features of urban land-use identification by GCN model. Class 0–5 in the legend corresponds to the six classes of land-use types. Class 0: Transportation land; Class 1: Public service and administration land; Class 2: Commercial land; Class 3: Residential land; Class 4: Industrial land; Class 5: Non-construction land. Names with the prefix “o” and “d” are features extracted from the origin and the destination, respectively. The feature name with the suffix “avg” is the average of that feature value. The feature name with the suffix “std” is the standard deviation of that feature value. The feature names with the abbreviations “fre” and “wt” are, respectively, the ratio and the weight of the attribute mentioned in Section 3.2.1.

From a global perspective, the features extracted from the travel attributes contributed more to the identification of urban land use at small spatial scales, like 250 m, while the features extracted from the social attributes contributed more to the identification of urban land use at middle and large spatial scales, such as 500 m and 1000 m. In addition, the features for describing the graph network of urban mobility, such as the frequency of spatial interaction between urban spatial units (the out_degree and in_degree of the urban spatial unit), were of similar importance to the model prediction for three spatial scales. Two of the most relevant features were the full-weighted population (out_fpop) and gender-weighted population (out_gpop), which flows out of the urban spatial unit, which contributed more to the identification of non-construction land, residential land, and industrial land than that of commercial land, transportation land, and public service and administration land. While other features only brought a marginal contribution to the global prediction, the average weight of transport type (o_type_wt_avg) from the origin and the average arrival time (d_ehour_avg) to the destination showed more relevance for the identification of commercial land and residential land.

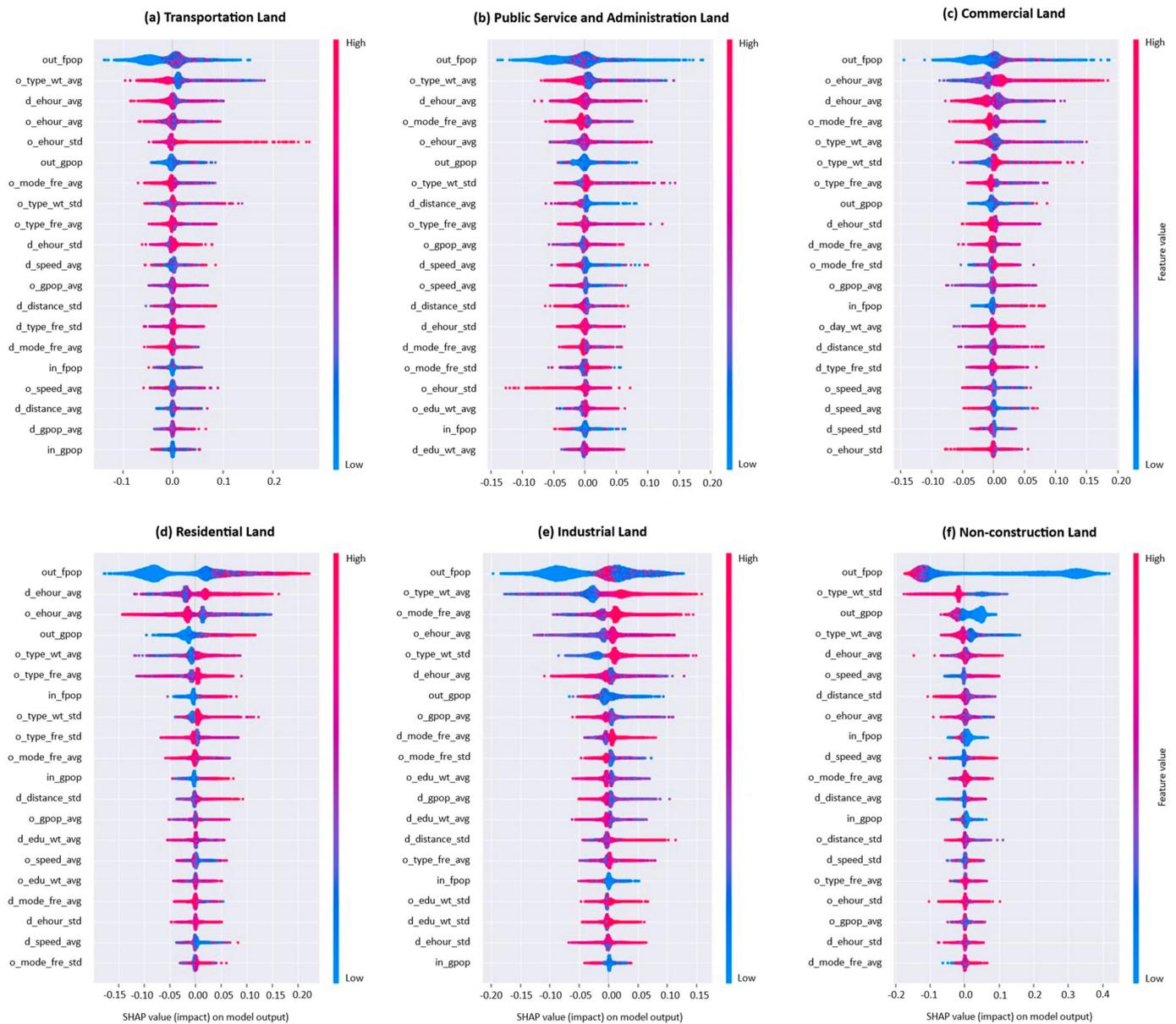


Figure 7. Local summary of distributed SHAP values related to the top 20 contributed features for identifying six urban land-use classes. The spot on the horizontal axis depicts the feature’s Shapely value for that land-use class, indicating how the feature influences the rise and fall of the probability of identifying that land-use class. The blue spot and red spot typify the features that have a low and high value, respectively.

From a local perspective (Figure 7), we take features’ Shapley value for the land-use identification at the spatial scale of 1000 m as an example. Most of the features had a mixed effect on the identification probability of transportation land. For public-service and administration land, a decrease in the identification probability of this land-use type is provoked by a vast distance between origin and destination, which leads, contrarily, to an increase in the probability of identifying residential land. A possible explanation is that people tend to choose nearby public facilities, yet they must return home, no matter how far away they live. Also, the average speed across a destination was inversely proportional to the identification probability of administration land and commercial land, which coincides with the fact that there are speed limits in most administrative and commercial areas. Apart from the spatial features, the temporal features also showed a strong impact on the model output. The average departure time from the origin was conducive to an increase in the

identification probability of commercial land, which resulted in a reduction in residential land. A rational explanation is that people leave residential areas early for commuting, and are inclined to visit commercial areas late in the day, after leaving work. In contrast to the features of transport attributes, features of social attributes revealed the social impact on land-use identification. For instance, the educational qualification of people from the origin had a positive effect on the identification probability of public-service and administration land; however, the opposite effect was shown in predicting industrial land. A possible explanation is that most of the people working in the administrative sector are highly educated, while industrial parks have no evident requirement concerning education level. Compared to other features of social attributes, such as age, gender, income, occupation, etc., educational qualification was the only one that appeared in the top 20 relevant features for the identification of urban land use.

5. Discussion

In this study, we introduced a spatial-hierarchical method in which a series of backbone GCNN models for urban functional-feature identification were operated at different spatial scales. The results substantiate a negative correlation between prediction accuracy and spatial scale. Notably, it is suggested that an optimal size of 250 m should be deemed suitable for urban spatial units in land-use studies, in conjunction with spatial interaction analyses. This observation is underpinned by the notion that large-scale spatial units tend to amalgamate subtle distinctions in land-use characteristics, compounded by the escalating data occasionality associated with a higher volume of spatial interaction flows at larger scales [76,77]. In addition, the downsizing of spatial resolution provides an augmented set of training samples for GCNN models, thereby enhancing their capacity to discern intricate relationships between spatial interaction patterns and urban land-use dynamics. Furthermore, we applied heterogeneous GCNN models to explore how spatial interactions affect the identification task of urban functional features. An important finding is that the directionality of spatial interaction makes a greater contribution to large-scale tasks, while the adjacency of spatial interaction has a more prominent impact on small-scale tasks. As the spatial scale increases, the intensity of spatial interaction shows an upward trend, enlarging the edge weights in GCNN models. The edge direction distinguishes the spatial interactions with similar intensities, and enriches the semantic information of the graph structure [78], thereby improving the accuracy of urban functional-feature identification. For the small spatial scale, according to the first law of geography [79], the local attributes of adjacent graph nodes have higher similarity, which enhances the representation of spatial dependency for identifying urban functional features.

In the second place, to probe the model performance in different human-activity scenarios, we extended the graph feature matrix by fusing the mobility data with population data and mobile phone usage data, respectively. The best model performance was derived from using the integration of three kinds of human-activity datasets. Concerning geographic context, human mobility data generally have wider coverage in both space and time, providing useful insights to represent social-economic activities between urban spatial units. The traffic attributes and social attributes of human mobility further characterize the diversification of relationships between spatial units. Additionally, the demographic data containing various population structures depict the long-term routine of human activity [80], revealing their stable usage of urban functions. The spatial pattern of mobile phone usage data may illustrate the short-term routine of human activity, reflecting the urban vitality in spatial units with different functions [81]. Therefore, when the spatial interaction traced by human mobility patterns is taken as a vital factor for inferring urban functional features, other types of human activity are also worthy of academic attention. While mobile phone data are gaining popularity in spatial interaction research, their availability varies in different countries and regions. A potential alternative is to integrate human mobility features using other location-based services, such as GPS data from taxis or Uber, public transportation data, social media data, etc.

6. Conclusions

This study proposes a novel semantic-enhanced GCNN framework for the multi-scale embedding of urban spatial units, and applies it to the identification of urban functional features. It is the first time that spatial interaction features derived from mobile phone data have been taken as the only data source for urban land-use identification. Above all, our work enriches the theory of multi-scale spatial interaction based on human mobility flows, and promotes the technology of feature-driven node classification in GCNNs.

From a theoretical perspective, predicting the unknown attributes of urban spatial units, such as land-use type, from spatial interaction is a challenging issue in urban geography, as the unknown attributes of a geographic unit have previously been inferred from its observed local attributes, rather than from the relationships between geographic units. Instead of using auxiliary information (e.g., POIs, remote sensing images, building footprints) as the observed attributes, as in existing studies, we exploited in depth the latent semantic information of spatial interaction as the geographic context, and aggregated the decentralized attributes to urban spatial units. This work leads to the high accuracy of GCNN models for urban land-use identification compared to state-of-the-art methods, showing the great power of spatial interaction data in precisely predicting the attributes of urban spatial units. Additionally, the scale effect as an important issue in the field of urban geography has rarely been examined in the existing studies. We explored the scale effect of spatial interaction and GCNN-based urban land-use identification in three typical spatial scales for urban studies. Our results prove that the directionality of spatial interaction has considerable influence on large-scale urban functional-feature identification, while the adjacency of spatial interaction exerts a profound effect on the same issue at a small scale.

From a technological perspective, while most related studies have focused on the effective fusing of multi-source data, we propose the combination of a systematic data-alignment method and a feature-generation method for the robust construction of heterogeneous graphs with multi-modal spatiotemporal data, providing an adaptive solution for improving the GCNNs' performance in geographic problems. As far as we can tell, the mobile phone dataset we collected in the current study is the largest and most current dataset used to date in the research of urban land-use identification through convolutional deep learning. In utilizing this large dataset, the generalization ability of our model was improved by augmenting the training samples. We conducted extensive experiments on this dataset with backbone GCNN models at different spatial scales, offering a scientific approach of scale-oriented model prioritization to the relevant research.

In terms of our contributions to the explainability of deep-learning methods, we used quantitative techniques to explain the role of spatial-interaction characteristics in land use identification, strengthening our understanding of the human–land relationship in the urban environment. By identifying the functional features of urban spatial units through AI methods, our study can provide references to the quick examination of urban spatial structure, timely verification of land-use planning, and synthetical optimization of urban functional form. Meanwhile, the study acknowledges specific limitations. The model's performance is contingent upon the quality of the data, wherein the spatial-interaction pattern derived from human mobility data typically adheres to a long-tail distribution. Effectively mitigating the inherent sparsity within human mobility data poses a formidable challenge for GCNNs. Concurrently, achieving a nuanced equilibrium between prediction accuracy and computational efficiency across diverse scales remains a complex endeavor. In future research, our paramount focus will be the development of a self-adaptive GCNN framework, which is designed to exhibit superior performance in identifying urban functional features characterized by sparse data, across a spectrum of spatial scales.

Author Contributions: Conceptualization, Yuting Chen and Pengjun Zhao; methodology, Yuting Chen; formal analysis, Yuting Chen; investigation, Yuting Chen, Yi Lin and Yushi Sun; validation, Yuting Chen, Yi Lin and Yushi Sun; resources, Yuting Chen and Pengjun Zhao; data curation, Yuting

Chen; writing—original draft preparation, Yuting Chen; writing—review and editing, Yuting Chen, Pengjun Zhao, Yi Lin, Yushi Sun and Yu Liu; visualization, Rui Chen and Ling Yu; supervision, Pengjun Zhao; project administration, Pengjun Zhao; funding acquisition, Pengjun Zhao and Yuting Chen. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (41925003, 42130402), the Shenzhen Science and Technology Innovation Program (RCBS20221008093330064), the Shenzhen Science and Technology Program (JCYJ20220818100810024), and the Introduction Project of Postdoctoral International Exchange Program (YJ20210217), the Guangdong Basic and Applied Basic Research Foundation (2022A1515010696).

Data Availability Statement: The data presented in this study are available on request from the authors due to privacy restrictions.

Acknowledgments: The authors would like to thank Hui Lin, Mai-Po Kwan, Zhaoya Gong, Yizhen Yan, Zhuoyi Zhao, as well as the editors and anonymous reviewers for the constructive comments, which helped improve this paper.

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

References

- Comtois, C.; Slack, B. *The Geography of Transport Systems*; Routledge: London, UK, 2009.
- Boarnet, M.; Crane, R. The influence of land use on travel behavior: Specification and estimation strategies. *Transp. Res. Part A Policy Pract.* **2001**, *35*, 823–845. [[CrossRef](#)]
- Wong, D.W. The modifiable areal unit problem (MAUP). In *WorldMinds: Geographical Perspectives on 100 Problems: Commemorating the 100th Anniversary of the Association of American Geographers 1904–2004*; Springer: Dordrecht, The Netherlands, 2004; pp. 571–575.
- Balsa-Barreiro, J.; Menendez, M.; Morales, A.J. Scale, context, and heterogeneity: The complexity of the social space. *Sci. Rep.* **2022**, *12*, 9037. [[CrossRef](#)] [[PubMed](#)]
- Jing, C.; Zhang, H.; Xu, S.; Wang, M.; Zhuo, F.; Liu, S. A hierarchical spatial unit partitioning approach for fine-grained urban functional region identification. *Trans. GIS* **2022**, *26*, 2691–2715. [[CrossRef](#)]
- Tao, H.; Wang, K.; Zhuo, L.; Li, X. Re-examining urban region and inferring regional function based on spatial-temporal interaction. *Int. J. Digit. Earth* **2019**, *12*, 293–310. [[CrossRef](#)]
- Liu, Y.; Liu, X.; Gao, S.; Gong, L.; Kang, C.; Zhi, Y.; Chi, G.; Shi, L. Social Sensing: A New Approach to Understanding Our Socioeconomic Environments. *Ann. Assoc. Am. Geogr.* **2015**, *105*, 512–530. [[CrossRef](#)]
- Zhu, D.; Zhang, F.; Wang, S.; Wang, Y.; Cheng, X.; Huang, Z.; Liu, Y. Understanding Place Characteristics in Geographic Contexts through Graph Convolutional Neural Networks. *Ann. Assoc. Am. Geogr.* **2020**, *110*, 408–420. [[CrossRef](#)]
- Mai, G.; Janowicz, K.; Hu, Y.; Gao, S.; Yan, B.; Zhu, R.; Cai, L.; Lao, N. A review of location encoding for GeoAI: Methods and applications. *Int. J. Geogr. Inf. Sci.* **2022**, *36*, 639–673. [[CrossRef](#)]
- Pei, T.; Sobolevsky, S.; Ratti, C.; Shaw, S.; Li, T. Machine learning and deep learning. In *Elgar Encyclopedia of Technology and Politics*; Edward Elgar Publishing: Cheltenham, UK, 2022; Volume 11, p. 113.
- Gao, R.; Xie, J.; Zhu, S.C.; Wu, Y.N. Learning grid cells as vector representation of self-position coupled with matrix representation of self-motion. *arXiv* **2018**, arXiv:1810.05597.
- Klemmer, K.; Xu, T.; Acciaio, B.; Neill, D.B. SPATE-GAN: Improved Generative Modeling of Dynamic Spatio-Temporal Patterns with an Autoregressive Embedding Loss. *Proc. AAAI Conf. Artif. Intell.* **2022**, *36*, 4523–4531. [[CrossRef](#)]
- Defferrard, M.; Bresson, X.; Vandergheynst, P. Convolutional neural networks on graphs with fast localized spectral filtering. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 3844–3852.
- Liu, X.; Chen, M.; Claramunt, C.; Batty, M.; Kwan, M.-P.; Senousi, A.M.; Cheng, T.; Strobl, J.; Cöltekin, A.; Wilson, J.; et al. Geographic information science in the era of geospatial big data: A cyberspace perspective. *Innovation* **2022**, *3*, 100279. [[CrossRef](#)] [[PubMed](#)]
- Huang, W.; Zhang, D.; Mai, G.; Guo, X.; Cui, L. Learning urban region representations with POIs and hierarchical graph infomax. *ISPRS J. Photogramm. Remote. Sens.* **2023**, *196*, 134–145. [[CrossRef](#)]
- Yang, X.; Stewart, K.; Tang, L.; Xie, Z.; Li, Q. A Review of GPS Trajectories Classification Based on Transportation Mode. *Sensors* **2018**, *18*, 3741. [[CrossRef](#)] [[PubMed](#)]
- Zhang, F.; Wu, L.; Zhu, D.; Liu, Y. Social sensing from street-level imagery: A case study in learning spatio-temporal urban mobility patterns. *ISPRS J. Photogramm. Remote. Sens.* **2019**, *153*, 48–58. [[CrossRef](#)]
- Goodchild, M.F. Formalizing place in geographic information systems. In *Communities, Neighborhoods, and Health: Expanding the Boundaries of Place*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 21–33.
- Roy, J.R.; Thill, J.-C. Spatial interaction modelling. *Pap. Reg. Sci.* **2003**, *83*, 339–361. [[CrossRef](#)]
- Zhong, C.; Arisona, S.M.; Huang, X.; Batty, M.; Schmitt, G. Detecting the dynamics of urban structure through spatial network analysis. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 2178–2199. [[CrossRef](#)]

21. Qiao, S.; Huang, G.; Yeh, A.G.-O. Mobility as a Service and urban infrastructure: From concept to practice. *Trans. Urban Data, Sci. Technol.* **2022**, *1*, 16–36. [[CrossRef](#)]
22. Alessandretti, L.; Aslak, U.; Lehmann, S. The scales of human mobility. *Nature* **2020**, *587*, 402–407. [[CrossRef](#)]
23. Zheng, Y.; Capra, L.; Wolfson, O.; Yang, H. Urban computing: Concepts, methodologies, and applications. *ACM Trans. Intell. Syst. Technol. (TIST)* **2014**, *5*, 1–55. [[CrossRef](#)]
24. Hu, S.; Gao, S.; Wu, L.; Xu, Y.; Zhang, Z.; Cui, H.; Gong, X. Urban function classification at road segment level using taxi trajectory data: A graph convolutional neural network approach. *Comput. Environ. Urban Syst.* **2021**, *87*, 101619. [[CrossRef](#)]
25. Xu, Y.; Jin, S.; Chen, Z.; Xie, X.; Hu, S.; Xie, Z. Application of a graph convolutional network with visual and semantic features to classify urban scenes. *Int. J. Geogr. Inf. Sci.* **2022**, *36*, 2009–2034. [[CrossRef](#)]
26. Zhang, Y.; Li, Q.; Tu, W.; Mai, K.; Yao, Y.; Chen, Y. Functional urban land use recognition integrating multi-source geospatial data and cross-correlations. *Comput. Environ. Urban Syst.* **2019**, *78*, 101374. [[CrossRef](#)]
27. Simini, F.; Barlacchi, G.; Luca, M.; Pappalardo, L. A Deep Gravity model for mobility flows generation. *Nat. Commun.* **2021**, *12*, 6576. [[CrossRef](#)] [[PubMed](#)]
28. Tian, C.; Zhang, Y.; Weng, Z.; Gu, X.; Chan, W.K. Learning Fine-grained Location Embedding from Human Mobility with Graph Neural Networks. In Proceedings of the 2022 International Joint Conference on Neural Networks (IJCNN), Padova, Italy, 18–23 July 2022.
29. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781.
30. Church, K.W. Word2Vec. *Nat. Lang. Eng.* **2017**, *23*, 155–162. [[CrossRef](#)]
31. Le, Q.; Mikolov, T. Distributed representations of sentences and documents. In Proceedings of the International Conference on Machine Learning, PMLR, Beijing, China, 22–24 June 2014.
32. Grover, A.; Leskovec, J. node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
33. 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](#)]
34. Zhang, J.; Li, X.; Yao, Y.; Hong, Y.; He, J.; Jiang, Z.; Sun, J. The Traj2Vec model to quantify residents' spatial trajectories and estimate the proportions of urban land-use types. *Int. J. Geogr. Inf. Sci.* **2021**, *35*, 193–211. [[CrossRef](#)]
35. Xu, Y.; Zhou, B.; Jin, S.; Xie, X.; Chen, Z.; Hu, S.; He, N. A framework for urban land use classification by integrating the spatial context of points of interest and graph convolutional neural network method. *Comput. Environ. Urban Syst.* **2022**, *95*, 101807. [[CrossRef](#)]
36. Bruna, J.; Zaremba, W.; Szlam, A.; LeCun, Y. Spectral networks and locally connected networks on graphs. *arXiv* **2013**, arXiv:1312.6203.
37. Kipf, T.N.; Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv* **2016**, arXiv:1609.02907.
38. Schlichtkrull, M.; Kipf, T.N.; Bloem, P.; Van Den Berg, R.; Titov, I.; Welling, M. Modeling relational data with graph convolutional networks. In *The Semantic Web, Proceedings of the 15th International Conference, ESWC 2018, Heraklion, Greece, 3–7 June 2018*; Springer: Berlin/Heidelberg, Germany, 2018; Volume 15.
39. Hamilton, W.; Ying, Z.; Leskovec, J. Inductive representation learning on large graphs. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
40. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv* **2017**, arXiv:1710.10903.
41. Hajibabae, P.; Malekzadeh, M.; Heidari, M.; Zad, S.; Uzuner, O.; Jones, J.H. An empirical study of the graphsage and word2vec algorithms for graph multiclass classification. In Proceedings of the 2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 27–30 October 2021.
42. Zhang, D.; Yin, J.; Zhu, X.; Zhang, C. Attributed network embedding via subspace discovery. *Data Min. Knowl. Discov.* **2019**, *33*, 1953–1980. [[CrossRef](#)]
43. Zhao, L.; Song, Y.; Zhang, C.; Liu, Y.; Wang, P.; Lin, T.; Deng, M.; Li, H. T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 3848–3858. [[CrossRef](#)]
44. Yu, B.; Lee, Y.; Sohn, K. Forecasting road traffic speeds by considering area-wide spatio-temporal dependencies based on a graph convolutional neural network (GCN). *Transp. Res. Part C: Emerg. Technol.* **2020**, *114*, 189–204. [[CrossRef](#)]
45. Fang, F.; Zeng, L.; Li, S.; Zheng, D.; Zhang, J.; Liu, Y.; Wan, B. Spatial context-aware method for urban land use classification using street view images. *ISPRS J. Photogramm. Remote. Sens.* **2022**, *192*, 1–12. [[CrossRef](#)]
46. Chen, T.; Bowers, K.; Zhu, D.; Gao, X.; Cheng, T. Spatio-temporal stratified associations between urban human activities and crime patterns: A case study in San Francisco around the COVID-19 stay-at-home mandate. *Comput. Urban Sci.* **2022**, *2*, 1–12. [[CrossRef](#)] [[PubMed](#)]
47. Jin, R.; Xia, T.; Liu, X.; Murata, T.; Kim, K.-S. Predicting Emergency Medical Service Demand With Bipartite Graph Convolutional Networks. *IEEE Access* **2021**, *9*, 9903–9915. [[CrossRef](#)]
48. Yu, X.; Shi, S.; Xu, L. A spatial-temporal graph attention network approach for air temperature forecasting. *Appl. Soft Comput.* **2021**, *113*, 107888. [[CrossRef](#)]
49. Wang, H.; Zhang, H.; Tang, G.; Zhou, L.; Jiang, S. Inter-city association pattern recognition by constructing cultural semantic similarity network. *Trans. GIS* **2022**, *26*, 2225–2243. [[CrossRef](#)]

50. Du, Z.; Zhang, X.; Li, W.; Zhang, F.; Liu, R. A multi-modal transportation data-driven approach to identify urban functional zones: An exploration based on Hangzhou City, China. *Trans. GIS* **2020**, *24*, 123–141. [[CrossRef](#)]
51. 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](#)]
52. Li, X.; Zhang, C.; Li, W. Building block level urban land-use information retrieval based on Google Street View images. *GIScience Remote. Sens.* **2017**, *54*, 819–835. [[CrossRef](#)]
53. Ye, C.; Zhang, F.; Mu, L.; Gao, Y.; Liu, Y. Urban function recognition by integrating social media and street-level imagery. *Environ. Plan. B Urban Anal. City Sci.* **2021**, *48*, 1430–1444. [[CrossRef](#)]
54. Andrade, R.; Alves, A.; Bento, C. POI Mining for Land Use Classification: A Case Study. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 493. [[CrossRef](#)]
55. Tu, W.; Cao, J.; Yue, Y.; Shaw, S.-L.; Zhou, M.; Wang, Z.; Chang, X.; Xu, Y.; Li, Q. Coupling mobile phone and social media data: A new approach to understanding urban functions and diurnal patterns. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 2331–2358. [[CrossRef](#)]
56. Mawuenyegah, A.; Li, S.; Xu, S. Exploring spatiotemporal patterns of geosocial media data for urban functional zone identification. *Int. J. Digit. Earth* **2022**, *15*, 1305–1325. [[CrossRef](#)]
57. 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](#)]
58. 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.
59. Huang, X.; Zhao, Y.; Ma, C.; Yang, J.; Ye, X.; Zhang, C. TrajGraph: A Graph-Based Visual Analytics Approach to Studying Urban Network Centralities Using Taxi Trajectory Data. *IEEE Trans. Vis. Comput. Graph.* **2015**, *22*, 160–169. [[CrossRef](#)]
60. Liu, B.; Deng, Y.; Li, M.; Yang, J.; Liu, T. Classification Schemes and Identification Methods for Urban Functional Zone: A Review of Recent Papers. *Appl. Sci.* **2021**, *11*, 9968. [[CrossRef](#)]
61. Cao, R.; Tu, W.; Yang, C.; Li, Q.; Liu, J.; Zhu, J.; Zhang, Q.; Li, Q.; Qiu, G. Deep learning-based remote and social sensing data fusion for urban region function recognition. *ISPRS J. Photogramm. Remote. Sens.* **2020**, *163*, 82–97. [[CrossRef](#)]
62. Psyllidis, A.; Gao, S.; Hu, Y.; Kim, E.-K.; McKenzie, G.; Purves, R.; Yuan, M.; Andris, C. Points of Interest (POI): A commentary on the state of the art, challenges, and prospects for the future. *Comput. Urban Sci.* **2022**, *2*, 20. [[CrossRef](#)]
63. Chen, Z.; Yeh, A.G.-O. Delineating functional urban areas in Chinese mega city regions using fine-grained population data and cellphone location data: A case of Pearl River Delta. *Comput. Environ. Urban Syst.* **2022**, *93*, 101771. [[CrossRef](#)]
64. Pei, T.; Sobolevsky, S.; Ratti, C.; Shaw, S.-L.; Li, T.; Zhou, C. A new insight into land use classification based on aggregated mobile phone data. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1988–2007. [[CrossRef](#)]
65. Ríos, S.A.; Muñoz, R. Land Use detection with cell phone data using topic models: Case Santiago, Chile. *Comput. Environ. Urban Syst.* **2017**, *61*, 39–48. [[CrossRef](#)]
66. Zhong, Y.; Fei, F.; Zhang, L. Large patch convolutional neural networks for the scene classification of high spatial resolution imagery. *J. Appl. Remote. Sens.* **2016**, *10*, 25006. [[CrossRef](#)]
67. Bao, H.; Ming, D.; Guo, Y.; Zhang, K.; Zhou, K.; Du, S. DFCNN-Based Semantic Recognition of Urban Functional Zones by Integrating Remote Sensing Data and POI Data. *Remote Sens.* **2020**, *12*, 1088. [[CrossRef](#)]
68. Lin, H.; Xu, B.; Chen, Y.; Jing, Q.; You, L. The virtual geographic environments: More than the digital twin of the physical geographical environments. In *New Thinking in GIScience*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 17–28.
69. Mai, G.; Janowicz, K.; Yan, B.; Zhu, R.; Cai, L.; Lao, N. Multi-scale representation learning for spatial feature distributions using grid cells. *arXiv* **2020**, arXiv:2003.00824.
70. Wang, J.; Kong, X.; Xia, F.; Sun, L. Urban human mobility: Data-driven modeling and prediction. *ACM SIGKDD Explor. Newsl.* **2019**, *21*, 1–19. [[CrossRef](#)]
71. Williams, A.M.; Foord, J.; Mooney, J. Human mobility in functional urban regions: Understanding the diversity of mobilities. *Int. Rev. Sociol.* **2012**, *22*, 191–209.
72. Mohan, A.; Pramod, K.V. Network representation learning: Models, methods and applications. *SN Appl. Sci.* **2019**, *1*, 1–23. [[CrossRef](#)]
73. Hofman, J.M.; Watts, D.J.; Athey, S.; Garip, F.; Griffiths, T.L.; Kleinberg, J.; Margetts, H.; Mullainathan, S.; Salganik, M.J.; Vazire, S.; et al. Integrating explanation and prediction in computational social science. *Nature* **2021**, *595*, 181–188. [[CrossRef](#)] [[PubMed](#)]
74. Burkart, N.; Huber, M.F. A Survey on the Explainability of Supervised Machine Learning. *J. Artif. Intell. Res.* **2020**, *70*, 245–317. [[CrossRef](#)]
75. Lundberg, S.M.; Erion, G.G.; Lee, S.-I. Consistent individualized feature attribution for tree ensembles. *arXiv* **2018**, arXiv:1802.03888.
76. Arbia, G.; Petrarca, F. Effects of scale in spatial interaction models. In *Spatial Econometric Interaction Modelling*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 85–101.
77. Yang, X.; Fang, Z.; Yin, L.; Li, J.; Lu, S.; Zhao, Z. Revealing the relationship of human convergence–divergence patterns and land use: A case study on Shenzhen City, China. *Cities* **2019**, *95*, 102384. [[CrossRef](#)]
78. Rossi, E.; Charpentier, B.; Di Giovanni, F.; Frasca, F.; Günemann, S.; Bronstein, M. Edge Directionality Improves Learning on Heterophilic Graphs. *arXiv* **2023**, arXiv:2305.10498.
79. Tobler, W. On the first law of geography: A reply. *Ann. Assoc. Am. Geogr.* **2004**, *94*, 304–310. [[CrossRef](#)]

-
80. Chen, N.; Akar, G. Effects of neighborhood types & socio-demographics on activity space. *J. Transp. Geogr.* **2016**, *54*, 112–121.
 81. Liu, S.; Zhang, L.; Long, Y.; Long, Y.; Xu, M. A New Urban Vitality Analysis and Evaluation Framework Based on Human Activity Modeling Using Multi-Source Big Data. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 617. [[CrossRef](#)]

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.