

RESEARCH ARTICLE OPEN ACCESS

Quantitative Identification of Mixed Urban Functions: A Probabilistic Approach Based on Physical and Social Sensing Data

Yan Zhang^{1,2}  | Mei-Po Kwan^{1,2}  | Boting Yu²  | Yanjun Liu³ | Liuyi Song²  | Nengcheng Chen⁴ 

¹Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong SAR, China | ²Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Hong Kong SAR, China | ³College of Transportation Engineering, Tongji University, Shanghai, China | ⁴School of Geography and Information Engineering, China University of Geosciences, Wuhan, China

Correspondence: Mei-Po Kwan (mpk654@gmail.com)

Received: 4 June 2024 | **Revised:** 19 October 2024 | **Accepted:** 22 October 2024

Funding: This work was supported by the Hong Kong Productivity Council and Transport Department (Grant PSRI/44/2208/PR); the Hong Kong Research Grants Council (General Research Fund Grants 14605920, 14611621, 14606922, 14603724); Collaborative Research Fund (Grant C4023-20GF); Research Matching Grants RMG (Grants 8601219, 8601242); the Research Committee on Research Sustainability of Major Research Grants Council Funding Schemes (Grant 3133235) of the Chinese University of Hong Kong.

Keywords: data fusion | quantitative diversity identification | social sensing | street view images | urban functional zones

ABSTRACT

This paper proposes a diversity identification method based on information fusion for quantitatively identifying mixed urban functional zones (UFZ), addressing the critical need for better city planning and management. This method integrates both social and physical sensing data, considering the frequency of urban functional occurrences and the intensity of human activity. Specifically, we extract “dynamic” human activity features from crowdsourced smart device data and “static” visual features from street view images. Based on the fused multi-modal data, our method infers the large-scale distribution of UFZs more accurately. We also create a standardized mixed UFZ dataset for model training and testing, which includes residential, commercial, public services, industrial, and ecological categories. In general, the method transforms the functional label recognition task into a probability distribution recognition task. It addresses complex land use distributions rather than simply assigning a single label to each zone. The result shows that our method could achieve a Cosine similarity of (0.542 ± 0.143) , the lowest Chebyshev of (0.785 ± 0.043) , and $L1$ distances of (0.264 ± 0.080) , indicating more accurate and consistent predictions and closer match to true distributions.

1 | Introduction

Occupying only 2% of the global land area, cities accommodate 50% of the population, consume 75% of energy, and emit 85% of carbon dioxide (Y. Zhang, N. Chen, et al. 2021). According to the Chinese population census, China's urbanization rate has increased from 36.22% in 2000% to 63.89% in 2020. This means that approximately 400 million people have migrated from rural

areas to cities (CSY 2018), which indicates the continuous urban expansion in China during the past two decades.

Associated with urban expansion is the constant change in urban functional zones (UFZs), reflecting evolving socioeconomic functions and human activities (Hu et al. 2021; Lu et al. 2022). Meanwhile, actual urban development often deviates from early urban plans, which is difficult to update frequently (urban plans

are typically redrawn every 5 or 10 years). The inconsistency between the prescribed land use in urban plans and the current functions also leads to a constant demand for monitoring UFZs (Liu et al. 2019).

The spatial clustering of the same type of socioeconomic activity in cities produces UFZs, which refer to the allocation and use of land resources for specific economic, social, and environmental functions. Each functional area is dominated by one or more functions (such as residential, educational, or commercial). For example, an area in which most land resources are used for industrial purposes has a strong industrial functional attribute; if the land resources are primarily used for residential buildings, then that area has a strong residential function. UFZs are essential components of cities (Gong et al. 2018). Unlike the typical land use and land cover (LULC) classification task, which emphasizes the physical characteristics of the surface, the identification of UFZs focuses more on socioeconomic functional attributes.

Most of the existing urban function research is to assign each zone with a detailed label. However, a functional zone is rarely occupied by single functions, and there are many “mixed-use blocks” (H. Liu et al. 2020). Quantitatively describing the weights (dominant functions) of various functions in urban areas remains a challenging task. Further, a block is often composed of buildings with different heights (it means different human activity densities) and buildings of different functional types. In other words, different buildings may have different importance due to the different populations and frequency of occurrence (Lu et al. 2020). For instance, a train station may occupy a relatively small area within a region, but it could still be considered the “dominant function” of that region due to its high level of human activity intensity and functional frequency. This is because train stations often serve as transportation hubs that connect people from different parts of the city, making them a crucial component of urban infrastructure.

Therefore, the research gap lies in how to incorporate human activity intensity and functional frequency into the consideration of mixed land-use, with a particular focus on the role of dominant functions. In other words, there is still no unified approach to quantify and identify this mixed-use state. To effectively deal with the effort posed by mixed-use diversity and the intensity of human activities, it is essential to achieve a comprehensive and detailed understanding of the urban environment (Wu et al. 2020). Existing urban environment sensing methods can

be broadly divided into two categories: social sensing and physical sensing, which exhibit significant differences in sensing precision, sensing coverage, and sensing dynamic. Social sensing methods have a wide coverage and provide timely feedback. In contrast, physical sensing methods have the advantages of a uniform format and higher sensing quality (Liu et al. 2018). In recent years, physical sensing (based on sensor networks) and social sensing technologies have made great progress (Cowie, Arthur, and Williams 2018). Geo-tagged big data have achieved widespread coverage in urban spaces. Ubiquitous sensing devices, such as smartphones, wearable devices, surveillance cameras, and vehicles, have greatly promoted the development of sensing capabilities (F. Zhang et al. 2019). Every person can be regarded as a “sensor,” providing feedback on events and entities in the real space (Liu et al. 2015), continuously generating large amounts of multi-modal sensing data (Zhang, Liu, and Biljecki 2023; Zhang, Kwan, and Ma 2024). These data include not only quantitative data but also natural language text, images, and videos (Zhang et al. 2020). Such data have greatly deepened our understanding of cities and brought significant changes to the study of UFZs. The main experimental approach of this article is how to combine these data to better address existing research problems (Table 1) in identifying UFZs.

In this study, we first construct a quantitative data set of urban functions. We take into account the frequency of function occurrence and the intensity of human activities. Subsequently, we use a diversity identification method of UFZs based on information fusion. We identify the probabilities of five major urban functions: residential, commercial, public services, industrial, and ecological and explore the mixed state of urban functions.

The rest of the paper is organized as follows. In Section 2, we review related work on the identification of urban functions. In Section 3, we introduce the method to construct the urban function distribution data set and the UFZ distribution identification method. Section 4 presents the results of the identification of urban function distribution. The final section summarizes the article and discuss future prospects.

2 | Related Work

Existing research on UFZs can be broadly classified into three types. The first is based on remote sensing imagery data, using convolutional neural networks (CNNs) and their variants for

TABLE 1 | Advantages and disadvantages of social sensing and physical sensing.

Sensing type	Social sensing	Physical sensing
Sensing approach	Weibo, smartphones, etc.	Probe vehicles, probe backpacks, etc.
Sensing accuracy	Massive data, lower accuracy	Very high accuracy
Sensing range	Wide coverage, high observation density in urban areas	Uniform coverage, low observation density
Sensing source	Human behavior patterns	Geographic entity status
Sensing characteristics	Low data density, high entropy	High data density, low entropy

recognition (Zhou et al. 2020). The second relies mainly on GPS trajectories (Wang et al. 2018; Yu et al. 2019; Zhang et al. 2016, 2018), and cell phone signaling data (Lane et al. 2010; Yuan, Zheng, and Xie 2012). These studies mainly used trajectory similarity matching or origin–destination (OD) complex network analysis methods to identify UFZs. The third utilizes natural language processing (NLP) algorithms to extract the latent geographic features. The most commonly used mining algorithms are term frequency-inverse document frequency (TF-IDF) (Aizawa 2003), Latent Dirichlet Allocation (LDA) topic models (Blei, Ng, and Jordan 2003), and Probabilistic Latent Semantic Analysis (pLSA) (Bosch, Zisserman, and Munoz 2006). These unsupervised clustering methods can automatically determine the number of UFZs (topics) (Gao, Janowicz, and Couclelis 2017). With the introduction of Google's Word2vec model, this word embedding method has gained widespread acclaim (K. Liu et al. 2020; Zhai et al. 2019). A standard processing approach is to analogize POI categories (such as the commercial category) as “words” and regions as “documents” (Yao et al. 2017; Huang, Wang, and Cong 2024). Subsequently, improved algorithms like Place2vec emerged, not only considering the spatial relationships among POIs but also obtaining denser embedding representations (Zhai et al. 2019).

As shown in Table 2, some recent urban function zone-related studies are described in terms of their main data sources, analysis scale levels, key classification quantities, and primary research methods. Based on data sources, the methods can be broadly divided into those based on high-resolution remote sensing imagery (Huang et al. 2021) and those based on urban spatiotemporal big data (social sensing) (Liu and Long 2016).

For example, Liu et al. (2017) integrated remote sensing and social media data using traffic analysis zones as spatial scale and employing support vector machines to classify seven categories of urban functions. Huang et al. (2021) used taxi trajectories to describe the interaction status between urban streets in order to capture long-range spatial dependence. According to statistics, 50.98% of urban function or land use identification studies used only a single data source, whereas 49.02% used two or more data sets (Liu et al. 2017; Yin et al. 2021). Due to the ease of access and rich geographic meaning of POIs, they became the most widely used data, being adopted by 74.51% of the studies.

Chen et al. (2021) argued that the main analysis scales in existing urban function research are the pixel, object, and parcel levels. Pixel-based methods mainly use the spectral and texture features of remote sensing imagery to classify urban functions. On the contrary, blocks representing relatively homogeneous UFZs are more compatible with the basic units of urban planning and management. Based on this rationale, scholars have proposed the EULUC-China data set (Gong et al. 2020). This data set, based on multiple data sources, such as 10-m satellite imagery, OpenStreetMap (OSM), night light remote sensing, POIs, and Tencent mobility data, uses a decision tree model to generate functional distribution products for 27 major cities in China, including 5 primary classification labels and 12 secondary classification labels, with an overall classification accuracy of 60%. The UFZ delineation scale and labels in this paper are also implemented based on this scheme.

3 | Methodology for Urban Functional Zone Distribution Identification

As shown in Figure 1, we conducted an experiment designed to detect UFZs. The purpose of the experiment is to distill the essence of urban dynamism and functionality into quantifiable measures and construct a mixed distribution data set for model learning. Additionally, we have built an information fusion network to recognize the land use condition.

3.1 | A Quantitative Description Approach of Urban Functional Zones

This paper seeks to combine multi-source data and deep learning methods to comprehensively detect and evaluate mixed urban functions. The first issue we address is the quantification of the “mixed” state of UFZs, and considering both the frequency of UFZs and the intensity of population activities. The most commonly used data to quantify the mixed state is POI, which includes functional attributes and allows for the calculation of mixed degrees. However, since POI data are point-based, it does not adequately reflect the intensity of population activities. Therefore, we also adopt area of interest (AOI) data, which includes functional information and could be used to calculate the population activities. Population activities are measured by Tencent Mobility data, which is identified as heat points.

As shown in Figure 2, we display the distribution of the AOI data in the study area, which is located in Wuhan City, China and contains more than 10,000 entities.

The quantitative methods of UFZ need to consider some factors. First, the total human activity of an AOI area is assessed by the dynamic population density weight pop : the higher the value, the more important the area. Second, the scarcer that type of AOI is, and the more significant the impact of that AOI on the UFZ function (the example of a train station in the intro section). We select the TF-IDF method to highlight the importance of less frequently occurring types (Y. Zhang, Z. Chen, et al. 2021). It is a statistical measure that can determine the importance of a word in a document or a collection of documents, considering the commonality and rarity of the word. In this paper, it is used to measure the importance of AOI function in a UFZ. The formula is as follows:

$$\text{TF} - \text{IDF}(t, d, D) = \text{TF}(t, d) \times \text{IDF}(t, D) \quad (1)$$

where t is an AOI in a UFZ, d is a UFZ, and D is the collection of all UFZs (total study area). $\text{TF}(t, d)$ represents the frequency of AOI t in the UFZ d , and $\text{IDF}(t, D)$ represents the inverse document frequency of AOI t in the collection D . $\text{TF}(t, d)$ can be calculated with the following formula:

$$\text{TF}(t, d) = \frac{n_t}{\sum_{w \in d} n_w} \quad (2)$$

where n_t represents the number of times AOI t appears in functional zone d , and $\sum_{w \in d} n_w$ represents the total number of all AOIs

TABLE 2 | Data and methods of recent research about urban function zones.

Reference	Data source	Scale level	Functional category	Research method
Gao, Janowicz, and Couclelis (2017)	POI	Delaunay triangulation	Unsupervised	Topic modeling
Yao et al. (2017)	POI	Traffic analysis zone	Unsupervised	Word2vec
Liu et al. (2017)	Remote sensing and social media	Traffic analysis zone	7 categories	Support vector machine
Wang et al. (2018)	Trajectory and POI	Spatial grid	3 categories	Semantic model
Xing and Meng (2018)	Street view	Administrative division	4 categories	Semantic model
Zhai et al. (2019)	POI	Administrative division	8 categories	Place2vec
Gong et al. (2020)	POI and night-time remote sensing	Traffic analysis zone	12 categories	Random forest
Du et al. (2020)	Trajectory	Traffic analysis zone	4 categories	Topic modeling
Wu et al. (2020)	Social media	Traffic analysis zone	Unsupervised	Linear decomposition
Cao et al. (2020)	Remote sensing and social sensing	Image grid	9 categories	Fusion model
Hu et al. (2021)	Taxi trajectory	Streets	3 categories	Graph convolution
Y. Zhang, Z. Chen, et al. (2021)	POI	Spatial grid	4 categories	Semantic model
Qian et al. (2021)	Social media and POI	Urban area	10 categories	Hierarchical clustering
Li et al. (2021)	POIs and remote sensing, etc.	Spatial grid	6 categories	Random forest
Huang et al. (2021)	High-resolution remote sensing images	Parcels	12 categories	Convolutional neural network
Li et al. (2021)	Remote sensing and POIs	Administrative division	12 categories	Ensemble learning
Du et al. (2021)	High-resolution remote sensing images	Multi-scale polygons	12 categories	Convolutional neural network
Fang et al. (2022)	Street view images	Parcels	5 categories	Graph convolution
Song et al. (2022)	Mobile signaling and POI	Building level	9 categories	Dynamic time warping
Huang et al. (2022)	POI	Parcels	10 categories	Self-supervised methods
Yang et al. (2022)	POI	Administrative division	4 categories	Graph convolution

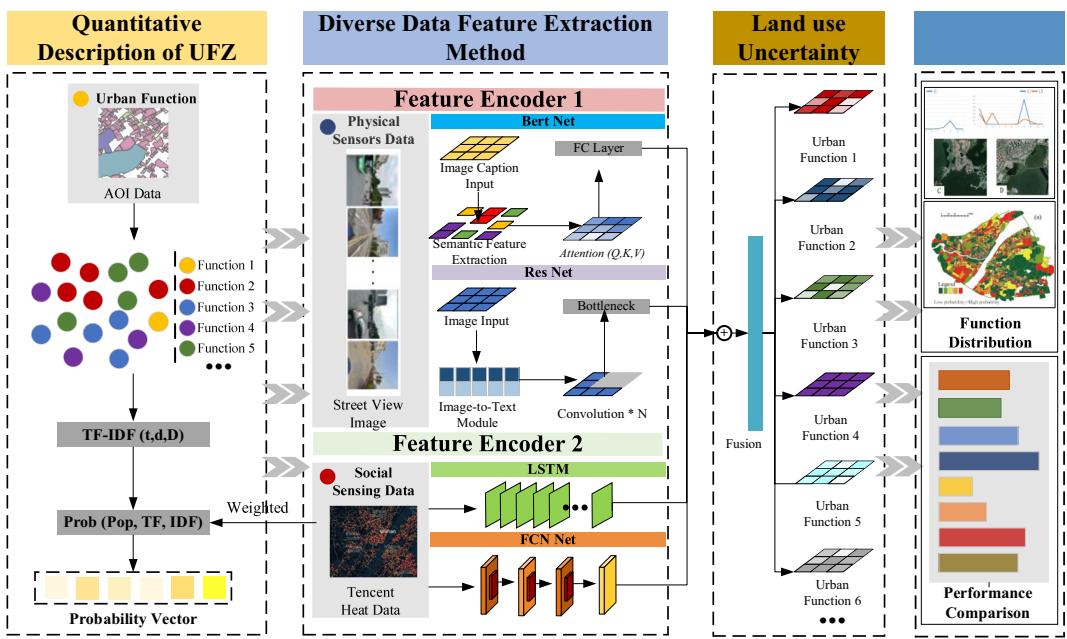


FIGURE 1 | Schematic diagram of this research.

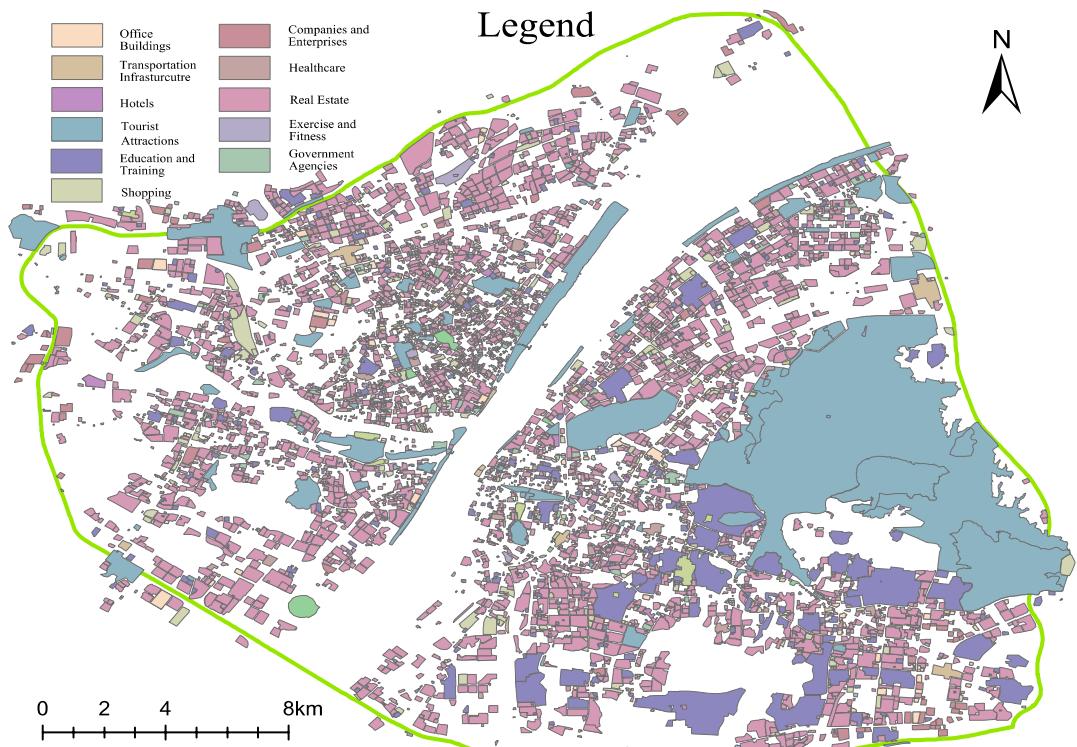


FIGURE 2 | Spatial distribution of the AOI area data within the Wuhan Three Ring Road.

in functional zone d . $IDF(t, D)$ can be calculated with the following formula:

$$IDF(t, D) = \log \frac{|D|}{|d \in D: t \in d|} \quad (3)$$

where $|D|$ represents the total number of functional zones in the study area D , and $|d \in D: t \in d|$ represents the number of functional zones in the study area D that contain AOI t . The product

of $TF(t, d)$ and $IDF(t, D)$ is the TF-IDF weight of a functional type (AOI class) in a UFZ, which can be used to measure the land use uncertain condition.

At this time, TF-IDF weight did not take human activity into consideration. In order to address this, we use the formula of $pop \times TF \times IDF$ to weight each functional type. The pop_{ij} (the weighted factor of i th functional category within the j th UFZ) is calculated by the following formula:

$$pop_{i,j} = \sum_{k=1}^n (heat_{i,j,k} \times strength_{i,j,k}) \quad (4)$$

where $heat_{i,j,k}$ and $strength_{i,j,k}$ ¹ represent human activity frequency and strength in the j th UFZ, respectively. This UFZ contains n AOIs. The more human activities, the higher the relative importance of that functional category.

Assuming the population ratio occupied by the i th function in that UFZ (j) is $pop_{i,j}$, setting the normalization (proportional scaling calculation) operation as $Norm$, then the final weighted factors of function i in zone j is

$$\text{Prob}_{\text{land}_i} = Norm(pop_{i,j} \times TF_{i,j} \times IDF_{i,j}) \quad (5)$$

Then we could obtain the normalized probabilities of each function i in zone j . At this point, the task of UFZ identification can be described as a label distribution learning problem (Geng 2016). For zone i , the goal is to learn the conditional distribution of each function label $\text{Prob}_{\text{land}_i}$, that is, estimating the function distribution $y_i = \{\text{Prob}_{\text{land}_1}, \text{Prob}_{\text{land}_2}, \dots, \text{Prob}_{\text{land}_i}\} = 1$.

3.2 | Development of the Mixed Functions Identification Model

As mentioned in the Introduction Section, the social sensing and physical sensing methods both have their advantages and disadvantages. In general, multi-source data are superior to single-source data in urban observations because they could complement each other. Deep learning models can judge which source's features are more important (Huang et al. 2018; Y. Zhang et al. 2019).

This paper mainly targets social sensing data (time series record of Tencent Mobility) and physical sensing data (street view image). For the former data, we use long short-term memory (LSTM) neural networks to extract "dynamic" features. For the

latter one, we extract the "static" features based on the Residual Neural Networks (ResNet). Image semantic features are also extracted based on the method provided by the study (Zhang, Liu, and Biljecki 2023; Zhang, Zhang, and Chen 2022). Then we could use the fused information features to predict the proportion of various functions.

3.2.1 | Extracting Time Series Features for Urban Dynamics

LSTM is a variant of the Recurrent Neural Network (RNN) (Sak, Senior, and Beaufays 2014), based on historical information memory, allowing the model to predict long-distance feature output. It has been widely used in language models, time series prediction, classification, translation, and sequence generation. LSTM was developed to solve the "gradient vanishing" and "gradient explosion" problems. Its structure is shown in Figure 3.

LSTM consists of forget gates, input gates, and output gates. It can control the in-flow and out-flow information, allowing the network to selectively remember or forget information as needed. Each LSTM cell's input includes the hidden state h_{t-1} and cell state C_{t-1} at the previous moment, and the input x_t at time t . The output of each LSTM cell includes the hidden state h_t and the cell state C_t at time t , with the detailed calculation formulas of each state shown in formula (6).

$$\begin{aligned} i_t &= \sigma(W_{ii}x_t + b_{ii} + W_{hi}h_{t-1} + b_{hi}) \\ f_t &= \sigma(W_{if}x_t + b_{if} + W_{hf}h_{t-1} + b_{hf}) \\ g_t &= \tanh(W_{ig}x_t + b_{ig} + W_{hg}h_{t-1} + b_{hg}) \\ o_t &= \sigma(W_{io}x_t + b_{io} + W_{ho}h_{t-1} + b_{ho}) \\ c_t &= f_t \odot c_{t-1} + i_t \odot g_t \\ h_t &= o_t \odot \tanh(c_t) \end{aligned} \quad (6)$$

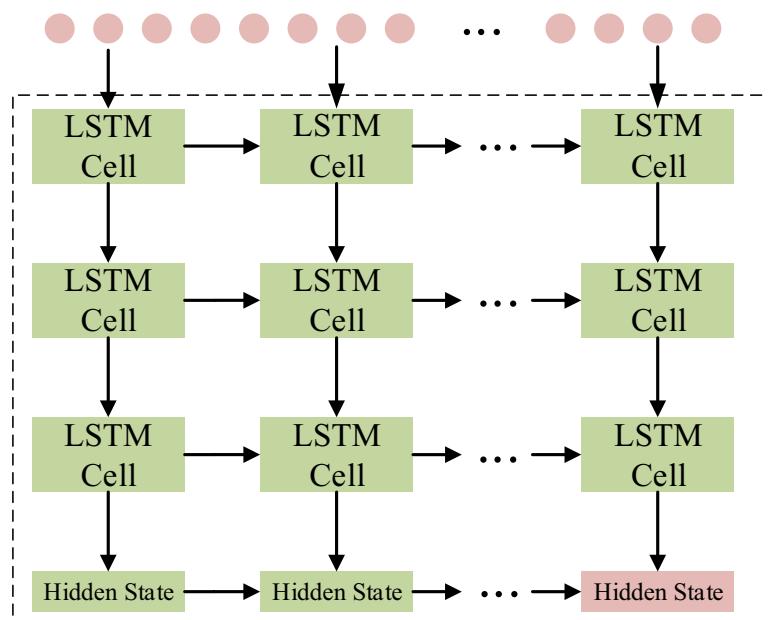


FIGURE 3 | The structure of the LSTM model.

In the formula, σ is the sigmoid activation function, \odot is the Hadamard product, W_{ii} represents the weight matrix from hidden state to hidden state, and W_{hi} represents the weight matrix between input and hidden states, which are the main objects of model learning. In this paper, the output of the last hidden state of the model is considered the extracted “dynamic” feature.

3.2.2 | The Information Fusion Approach for Integrating Diverse Data Sources

Information fusion refers to the process of combining or merging different features (Xiao 2019). It includes feature concatenation, feature addition, max pooling, and min pooling methods. Feature concatenation refers to simply stitching multiple features horizontally (vertically) together without any other processes. Feature addition refers to adding the numerical values of multiple aligned embeddings. Max pooling refers to taking the maximum value from multiple features as the fused feature. In contrast, min pooling refers to taking the minimum value from multiple features, which could reduce data noise.

The specific implementations of these four methods are illustrated in Figure 4, where (a–d) represent the feature concatenation, feature addition, max pooling, and min pooling methods,

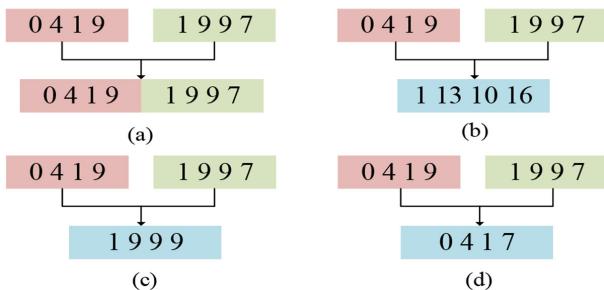


FIGURE 4 | Different methods of data fusion (a–d represent the feature concatenation, feature addition, max pooling, and min pooling methods, respectively).

respectively. To showcase the richness of multi-source data more, we adopt the feature concatenation method.

Figure 5 shows the detailed design of our method. The first module is LSTM. The stacked LSTM has been verified to increase its capability to capture dynamic features (Cao et al. 2020). This paper designed a three-layer stacked feature extraction architecture, taking the output of the last state as this module output. The second module is the Fully Connected Network (FCN), which is also used for processing time series data and can provide another data perspective. The third module is Bert, which is used for extracting visual semantic information in street views. The fourth module is ResNet 101, which is used for extracting static visual information from street views. The multi-modal information fusion is input into the prediction module.

The loss function of deep learning models determines the direction of model optimization. As seen in Figure 5, our goal is to calculate the functional distributions. The traditional cross-entropy loss function ($-\sum y_i \log(\hat{y}_i)$) for training classifiers is no longer applicable. Our predictions are not in the form of one-hot encoding like $[0, 0, 1, \dots, 0]$, but in a format of probability distributions list. The loss function should enable us to minimize the difference between the predicted probability distribution and the real probability distribution as much as possible, which could help the model better fit the training data. We therefore used the Kullback–Leibler divergence (KL divergence) as the loss function. It can measure the symmetrical distance between two probability distributions. The mathematical formula is as follows:

$$KL(\hat{y} \| y) = \sum_{c=1}^M \hat{y}_c \log \frac{\hat{y}_c}{y_c} \quad (7)$$

$$JS(\hat{y} \| y) = \frac{1}{2} \left(KL\left(y \| \frac{y+\hat{y}}{2}\right) + KL\left(\hat{y} \| \frac{y+\hat{y}}{2}\right) \right) \quad (8)$$

In the formula, \hat{y} and y are two probability distributions, the KL divergence calculates the difference between them. \hat{y}_c is the c th element of the \hat{y} distribution, and y_c is the c th element of the y distribution. KL divergence is a nonnegative number and it is

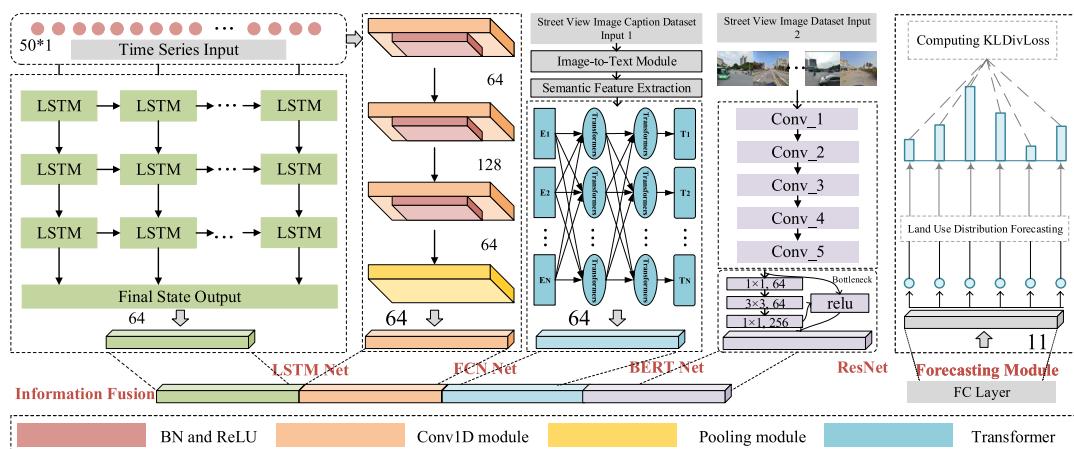


FIGURE 5 | A identification method of mixed UFZs based on information fusion.

zero when the two distributions are completely the same. This means that the larger the KL divergence, the greater the difference between the two distributions.

Since the KL divergence is not symmetrical, meaning $\text{KL}(\hat{y} \parallel y) \neq \text{KL}(y \parallel \hat{y})$. The standardized and symmetric Jensen-Shannon divergence described in formula (8), which could result in a smoother result. Based on this divergence calculation

TABLE 3 | Parameters and data sizes of each part of the model.

Data name	Component	Data format
Input data	Street view image	800*500*3
	Time series sequence	(50, 1, 64)
x_1	Conv1d (kernel size = 8)	(64, 128, 43)
x_1	Conv1d (kernel size = 5)	(64, 256, 39)
x_1	Conv1d (kernel size = 3)	(64, 64, 37)
x_1	Average pooling	(64, 64)
x_2	LSTM	(64, 50, 64)
x_2	Last state output	(64, 64)
x_3	Bert	(64, 384)
x_3	FC (384, 64)	(64, 64)
x_4	ResNet 101	(64, 1000)
x_4	Average pooling	(64, 1000)
x_4	FC (1000, 128)	(64, 128)
x_f	Fusion	(64, 320)
x_f	FC (320, 5)	(64, 5)

Note: x_1-x_4 indicate different module inputs, x_f for fused input, and the batch size is 64.

method, the model's pointwise loss calculation formula is as follows:

$$\text{Loss}(y_{\text{pred}}, y_{\text{true}}) = y_{\text{true}} \cdot \log \frac{y_{\text{true}}}{y_{\text{pred}}} = y_{\text{true}} \cdot (\log y_{\text{true}} - \log y_{\text{pred}}) \quad (9)$$

The specific parameters of each step of the model are shown in Table 3. The first part processes the time series data in the format of (50, 1, 64) through three convolutional full connection treatments (Conv1d), respectively, obtaining data sizes of (64, 128, 43), (64, 256, 39), and (64, 64, 37). The second part extracts LSTM features from time series data, taking the output in the form of (64, 50, 64) at the last time step. The third part extracts semantic features from image caption data using Bert, the feature size is (64, 384), and after entering the fully connected layer (FC), obtaining an output size of (64, 64). The fourth part uses the Pre-Trained ResNet 101 to extract the average pooling image features of size (64, 1000).

Finally, the features of the four above parts are fused, resulting in a feature x_f shape of (64, 320), entered into the FC layer. The distribution of each function contained in the region is calculated using the softmax function.

3.3 | Metrics for Assessing Probability Distribution Accuracy

Common accuracy evaluation metrics used in fitting regression tasks include mean squared error, root mean squared error, and mean absolute error (MAE). They could be used in regression tasks but not could be used in our label distribution learning task (Cha 2007; Huang et al. 2023). This article thus elected the following suitable evaluation metrics for comparison:

$$1. L1 \text{ distance (L1)} \downarrow: \sum_{k=1}^m |\hat{y}_i^{f_k} - y_i^{f_k}| \quad (10)$$

$$2. \text{ Chebyshev distance (Chebyshev)} \downarrow: \max_k |\hat{y}_i^{f_k} - y_i^{f_k}| \quad (11)$$

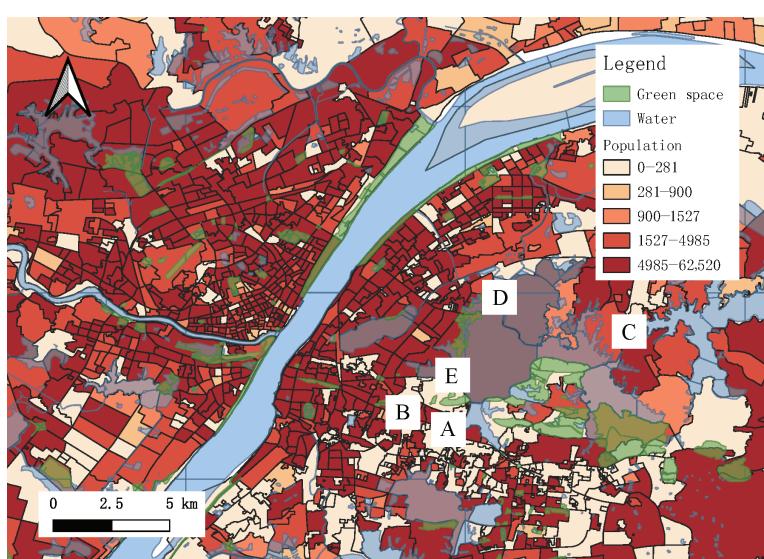


FIGURE 6 | Scales and illustration of urban functional zones.



FIGURE 7 | Human activity changes at different times in the study area. It both includes workdays and weekends.

TABLE 4 | Detailed fields of the e-map AOI data.

AOI Name	Land use category	Address	Category	Subcategory
Wuhan Hongji Hospital	Public services	No. 6 Huiji Second Road, Jian Construction Avenue	Medical	Specialty hospital
Renmin Hospital	Public services	No. 5 Changtian Road, Qiaokou District	Medical	General hospital
...
Wuhan Huangpi District Education Bureau	Public services	215 Xiangyang Street, Huangpi District, Wuhan	Government agency	Administrative unit
Hubei Provincial Department of Water Resources	Public services	17 Zhongnan Road, Wuchang District, Wuhan	Government agency	Administrative unit
Hannan District High School, Wuhan	Public services	169 Chenxi Road, Hannan District, Wuhan	Education and training	High school
Institute of Hydrobiology, Chinese Academy of Sciences	Public services	7 Donghu South Road, Wuchang District, Wuhan, Hubei Province	Education and training	Research institution
Nanhu Gas Station	Public services	69 Wuliang Road, Hongshan District, Wuhan, Hubei Province	Transportation infrastructure	Gas station
Wuhan Tiandi One Mall (Zhongshan Avenue Branch)	Commercial	1515 Zhongshan Avenue, Jiang'an District, Wuhan	Shopping	Shopping mall

Note: The AOI data include 11 functional tags: transportation facilities, companies, medical, real estate, government agencies, education and training, tourist attractions, shopping, sports and fitness, hotels, and office buildings, which form the basis for the UFZ labels.

$$3. \text{ Cosine similarity (Cosine)} \uparrow : \frac{\left(\sum_{k=1}^m \hat{y}_i^{f_k} y_i^{f_k} \right)}{\left(\sqrt{\sum_{k=1}^m \hat{y}_i^{f_k}} \sqrt{\sum_{k=1}^m y_i^{f_k}} \right)} \quad (12)$$

where $\hat{y}_i^{f_k}$ is the estimated distribution of functional type f_k , and $y_i^{f_k}$ is the corresponding ground truth distribution. In the formula, \downarrow indicates that the smaller the metric, the result better, and \uparrow indicates that the larger the metric, the result better.

4 | Urban Function Zones Perception and Identification Experiment

4.1 | Data Set Construction

As shown in Figure 6, the study area contains 1000 communities, which are the smallest administrative units in the city. The scale of analysis for urban functional zones (UFZs) is at the community level.

TABLE 5 | Land mix-use situations at the community scale in the study area.

UFZ	Func1	Func2	Func3	Func4	Func5	Func6	Func 7	Func 8	Func9	Func 10	Func11	Area ratio
347	0.008	0	0	0	0.004	0.002	0.006	0.980	0	0.001	0	0.419
350	0	0	0	0	0	1.000	0	0	0	0	0	0.154
351	0.655	0	0	0	0	0	0	0	0	0.345	0	0.007
353	1.000	0	0	0	0	0	0	0	0	0	0	0.496
355	0.015	0.004	0	0.014	0	0	0.159	0.806	0	0	0.002	0.103
356	0.652	0	0	0	0	0	0.348	0	0	0	0	0.631
357	0.144	0.615	0	0	0	0	0	0.202	0	0.017	0.023	0.039
359	0	1.000	0	0	0	0	0	0	0	0	0	0.043
360	0.681	0.055	0	0	0	0	0	0.086	0.178	0	0	0.571
362	0.436	0	0.321	0	0	0	0.008	0.152	0.083	0	0	0.044
364	0.288	0	0	0	0	0	0	0.702	0	0.010	0	0.077
365	0	0	0	0	0.076	0	0	0.924	0	0	0	1.288
366	1.000	0	0	0	0	0	0	0	0	0	0	0.016
367	0	0.978	0	0	0	0	0	0.022	0	0	0	0.046
368	0.002	0	0	0.001	0	0	0	0.997	0	0	0	6.750
369	0	0.672	0	0	0	0	0.328	0	0	0	0	0.013
370	0.002	0.817	0	0	0.155	0	0.015	0	0.011	0	0	0.148
377	0	0	0	0	0	0.511	0	0.489	0	0	0	0.280

Note: The sum of probabilities for various functions is 1. The numbers 1–11 represent the 11 UFZ categories.

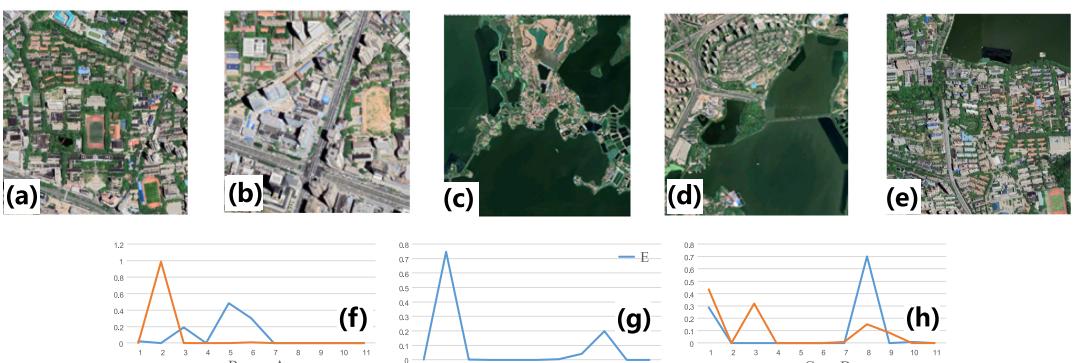


FIGURE 8 | Quantitative description results of mixed urban functions. Panels a-e show remote sensing images of regions A-E, while panels f-h display the probability distributions of different functions.

The first study data set is Tencent Mobility data, calculated from the real-time spatial locations of over 1 billion Tencent software users, stored in the form of points that include latitude and longitude information and heat value information. The heat value has been proven to effectively reflect human activities (Qian et al. 2021). Its spatial resolution is 25m, offering the advantages of high spatiotemporal resolution and low acquisition cost. We assembled data for 50 hours, including both workdays and weekends (October 25th, 26th, and 27th, 2019), totaling 7,348,690 heat points. Figure 7 shows the changes in human activity at different times.

The second data set consists of 75,628 street view images. In addition, we also collect approximately 10,000 AOIs, including 11 functional types, with detailed data fields as shown in Table 4.

As mentioned in Section 3.1, a weighted method considering multi-requirements was applied. The sum of probabilities of all categories in each zone equals 1 (100%). To avoid the interference of extreme values, the ratio of the total area of AOIs to the zone's total area is also calculated. If the value of the ratio is too low (<40%), it is considered an invalid sample. At last, we filtered out approximately 800 valid zones. The results are shown in Table 5 and Figure 8,

TABLE 6 | Cumulative importance of various urban functions in the data set.

Urban function	AOI number	TF – IDF _{SUM}
Transportation facilities	356	6.66013
Companies	619	19.6575
Medical	203	26.758
Real estate	5330	498.082
Government agencies	450	39.3432
Education and training	1510	96.2632
Tourist attractions	276	23.1291
Shopping	389	61.4024
Sports and fitness	49	11.6567
Hotels	430	23.2627
Office buildings	729	72.7854

**FIGURE 9** | Urban functional categories merged results based on the feature similarity.

where the first column represents the UFZ number, and columns 1–11 represent the weighted functional distribution ($\text{Prob}_{\text{land}_j}$).

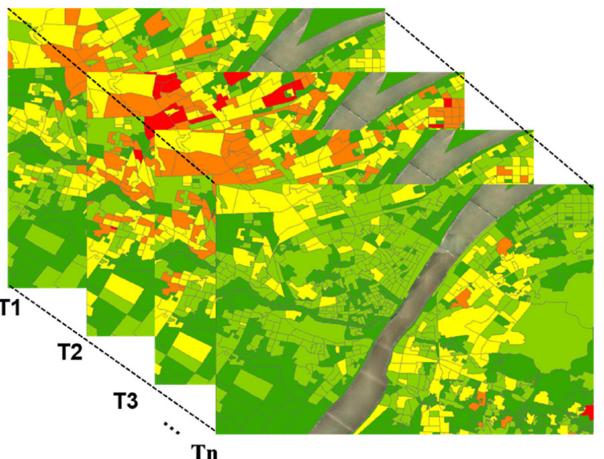
We selected five UFZ samples randomly and named them A, B, C, D, and E (they were marked in Figure 6), in order to more specifically illustrate the results of constructing our UFZ distribution data set. Then we plotted the functional distribution result in Figure 8. We can see that this data set can more accurately depict the distribution of the UFZs, and also provides a richer characterization of land use.

As shown in Table 6, it displays the number of AOIs and their sum weighted values $\text{TF} - \text{IDF}_{\text{SUM}}$, representing their cumulative importance in the data set. It could be calculated by the following formula:

$$\text{TF} - \text{IDF}_{\text{SUM}} = \sum_{j=0}^{j=\text{UFZNum}} \text{Prob}_{\text{land}_{ij}} \quad (13)$$

where UFZNum refers to the total UFZs in the study area.

The frequency of occurrence of 11 main AOI functions is different, with only 49 AOIs for sports and fitness. The most numerous are real estate, with more than 5,000 parcels. Faced with such great differences between the numbers of different AOIs, a simple number-weighted method would undoubtedly lead to lower accuracy.

**FIGURE 10** | Time-series heat aggregation method according to zone units.

This paper also merged functional categories, in order to improve the accuracy and robustness of the UFZ identification model. Most current UFZ classification models typically assign a binary label (yes or no) to each zone, whereas in this research, we should assign multiple labels. Retaining all function labels would essentially require running multiple regression models simultaneously, as we also need to calculate the

intensity for each urban function label. This makes our task inherently more complex, as we are calculating a probability distribution rather than assigning a single label. Merging categories that were similar in nature could increase the model's

ability to generalize and still could capture the key urban functions. As shown in Figure 9, we saved five main functional categories: residential,² commercial,³ public services,⁴ industrial,⁵ and ecological.⁶

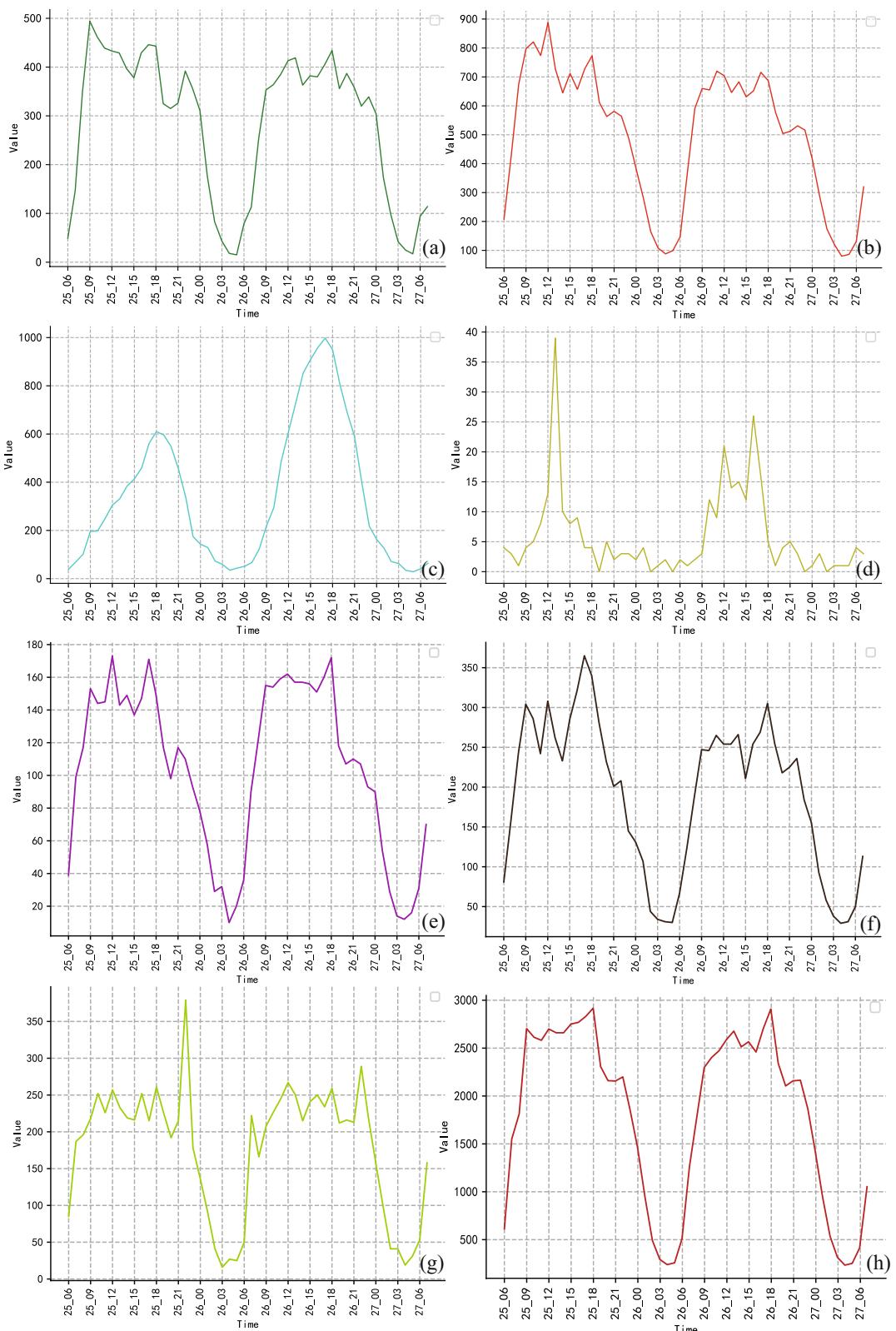


FIGURE 11 | Time series heat of different UFZs with the same square. Panels a-e illustrate the model-predicted probabilities for zones containing residential, commercial, industrial, and public service functions, respectively. Panels a-h present time series data showing the intensity of human activities across different functional zones.

4.2 | The Identification Result of Mixed Urban Functions

The Tencent mobility data set we utilized encompasses records spanning 50 h, which depict human activities. The data set is primarily represented as heat points, as illustrated in Figure 7 ($n=50$). For each UFZ, a time series is generated capturing varying intensity levels at different times, as shown in Figure 10.

Based on the aggregation method shown in Figure 10, we draw Figure 11 (i.e., the human activity change curves of various UFZs). It records the population changes in the same area size for a University (a), a residential area (b), the commercial area (c), the tourist area (d), the company clusters (e), sports venues (f), Theater and surroundings (g), and the Hospital (h). In this representation, the horizontal axis denotes time, whereas the vertical axis reflects the relative number of people, which indicates the intensity of human activity in each area.

From Figure 11, it can be seen that the human activity sequences under different urban function types exhibit significant time and intensity differences. The commercial area has the highest absolute intensity, reaching its peak at 17:00 on the holiday (October 26th). The tourist zone has the lowest spatial density, while the university's population density is lower than that of residential areas, and their time series changes have certain similarities. In other words, it is difficult to distinguish the sequence of residential areas from others. This means that it is possible to identify different urban functions through social sensing "dynamic" features, but relying solely on it still has shortcomings. The time series changes in Figure 11a,b are similar, necessitating the addition of physical sensing (street view) data to distinguish between them. This is also the motivation for proposing multimodal data-coupled observation for us.

Based on the method proposed in the former Figure 5, comparisons were made between using only social sensing data (Tencent Mobility), only physical sensing data (street view data), and multimodal fusion data. Experiments were carried out in functional zones with different population structures, income levels, and occupational distributions in the study area, randomly divided into training and testing sets according to an 80% and 20% rule. The experimental results are shown in Table 7 (mean \pm standard deviation).

Table 7 shows the results of our experiment. In this table, Multimodal Fusion1 and Multimodal Fusion2, respectively, represent distribution predictions with and without considering the spatial relationships between zones (Zhang, Liu, and

Biljecki 2023). As expected, the multimodal fusion method achieved better performance, showing some improvements in all indicators relative to individual modal, as mean Cosine value (0.539, 0.542) and L1 mean value (0.277, 0.264). The smallest similarity distance, which means that the model-inferred distribution was close to the actual distribution. Our method outperform competitive baselines in most evaluation measures.

Our results also indicate that the information-fused feature embeddings are more meaningful and consist of rich urban function information. This demonstrates the significant prospects of multimodal information fusion in urban analysis, where combining different types of data sources can improve the accuracy and effectiveness of urban downstream tasks. Compared to individual modal methods, the identification accuracy of our multimodal method increased by about 5% in the urban functional identification task.

To better illustrate the computational results of the model, we plotted the probabilities of different zones in the study area being predicted as various categories. Figure 12a-d each show the predicted probabilities of zones containing residential, commercial, industrial, and public service functions, respectively. Compared to traditional urban function identification methods, the results can better quantify the degree of urban function mixing, and capture the "mixed habitation" diversity state of urban functions.

5 | Discussion and Conclusion

This paper introduces a new framework for quantifying mixed urban functions and creates a ground truth dataset about their distribution. The framework makes several important contributions to quantifying and understanding mixed urban functions at a fine-grained scale. First, by incorporating both functional scarcity (TF-IDF weighting) and human activity intensity (population weighting), the distribution more accurately reflects the relative importance of different urban functions within each zone, going beyond simple presence/absence measures. Conventional UFZ mapping has largely treated areas as having a single dominant function, failing to capture the inherent mixing of activities that characterize urban environments.

Second, the fusion of social sensing data tracking human dynamics with physical sensing data capturing the built environment provides a more comprehensive view compared to using either modality alone. The temporal signatures of location-based services reveal fluctuations in human activities tied to different

TABLE 7 | Accuracy of different data modalities in identifying UFZ distribution.

Data modality	Cosine \uparrow	Chebyshev \downarrow	L1 \downarrow
Social sensing	0.528 ± 0.161	0.757 ± 0.005	0.326 ± 0.071
Physical sensing	0.507 ± 0.158	0.831 ± 0.047	0.285 ± 0.089
Multimodal Fusion1	0.539 ± 0.150	0.832 ± 0.051	0.277 ± 0.081
Multimodal Fusion2	0.542 ± 0.143	0.785 ± 0.043	0.264 ± 0.080

Note: Due to the presence of multiple functional types, accuracy evaluation is in the form of mean \pm standard deviation.

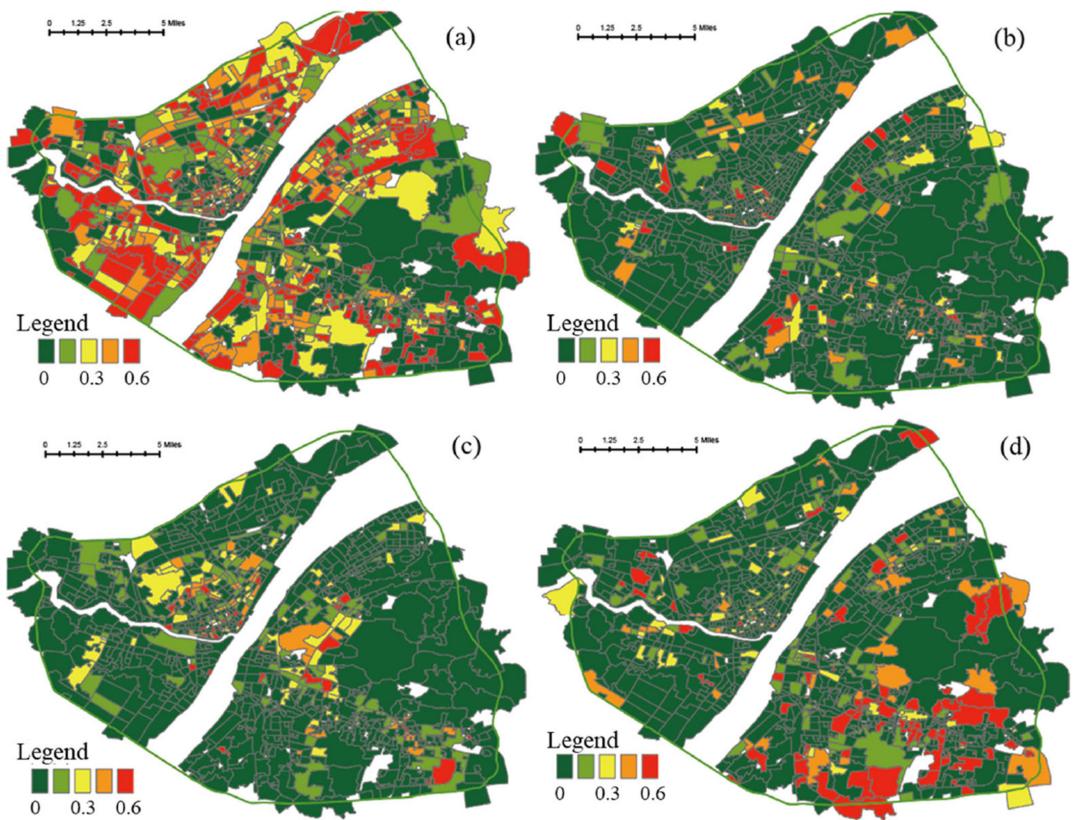


FIGURE 12 | Distribution of various UFZs identified by our model. Panels a-d illustrate the model-predicted probabilities for zones containing residential, commercial, industrial, and public service functions, respectively.

urban functions, such as residential, commercial, or industrial areas. Meanwhile, street view imagery conveys semantic cues about the physical urban landscape. Combining these complementary data streams allows the model to disambiguate communities that may exhibit similar activity patterns but differ in their visual/physical characteristics.

Third, defining urban functional mapping as a probability distribution recognition task enables a more nuanced understanding beyond conventional classification, explicitly quantifying the mixing proportions of different functions within zones. This aligns with the reality that most urban areas are not composed of a single monofunctional use but rather integrated mixes of residential, commercial, and other activities interwoven together.

With this framework, planners can optimize resource allocation. By designing the distribution of commercial, residential, and public service facilities, they can reduce peak-hour traffic congestion and commuting times. Additionally, optimizing public transportation routes and medical facilities based on population density and functional diversity can improve healthcare service coverage. Understanding the mixed land use state and distribution enables planners to co-locate schools, parks, and shops, enhancing livability. Properly allocating residential and entertainment facilities in commercial areas meets recreational needs and stimulates economic development, boosting regional vitality.

Due to the limited scale of observation, we focus on only a few types of urban functions, while future studies should expand

the number of urban function labels. Another significant limitation of our study is the potential for overfitting, particularly because the deep learning framework incorporates only approximately 1,000 communities. This relatively small sample size may lead to a model that performs well on the training data but struggles to generalize effectively to unseen data. In the future study, we will aggregate more data, thereby increasing the diversity and quantity of input samples, which can help in better capturing the underlying patterns within the data. Additionally, employing suitable penalized loss functions and higher quality of street view images may also improve its generalization.

Acknowledgments

This research was supported by a grant from the Smart Traffic Fund from the Hong Kong Productivity Council and Transport Department (Grant No: PSRI/44/2208/PR). It was also supported by grants from the Hong Kong Research Grants Council (General Research Fund Grant Nos. 14605920, 14611621, 14606922; Collaborative Research Fund Grant No. C4023-20GF; Research Matching Grants RMG 8601219, 8601242), and a grant from the Research Committee on Research Sustainability of Major Research Grants Council Funding Schemes (Grant No. 3133235) of the Chinese University of Hong Kong. The funders had no role in the study design, data collection, data analysis, decision to publish, or preparation of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Endnotes

¹Strength value is generated by Tencent trip data. It dynamically measures the relative population activity intensity in a given area. The higher the population density in the area at a specific time, the higher the strength value.

²Residential category for residential use, this category includes areas, such as housing complexes, apartments, and villas. These zones typically offer essential facilities like supermarkets, schools, and parks, catering to the daily needs of residents.

³Commercial category covers areas dedicated to commercial activities, including shopping malls, office buildings, retail centers, and dining establishments. These zones serve as economic hubs, providing goods and services, and are usually located in city centers or busy districts.

⁴Public services category provides essential public services, including schools, hospitals, government offices, and cultural facilities like museums and libraries. They play a crucial role in offering education, healthcare, and administrative services to the public.

⁵Industrial category includes areas designated for industrial production, such as factories, warehouses, and logistics parks. These zones are typically located away from residential areas to minimize pollution and noise, serving as centers for manufacturing and economic production.

⁶Ecological category comprises natural reserves, wetlands, parks, and forests, while does not only serve environmental conservation purposes but also function as tourist attractions. With their scenic beauty and rich ecological resources, they are ideal for outdoor activities and recreation. Ecological areas help promote tourism and raise environmental awareness, contributing to both economic growth and environmental protection.

References

- Aizawa, A. 2003. "An Information-Theoretic Perspective of TF-IDF Measures." *Information Processing & Management* 39: 45–65.
- Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3: 993–1022.
- Bosch, A., A. Zisserman, and X. Munoz. 2006. "Scene Classification via pLSA." In *European Conference on Computer Vision*, 517–530.
- Cao, R., W. Tu, C. Yang, et al. 2020. "Deep Learning-Based Remote and Social Sensing Data Fusion for Urban Region Function Recognition." *ISPRS Journal of Photogrammetry and Remote Sensing* 163: 82–97.
- Cha, S.-H. 2007. "Comprehensive Survey on Distance/Similarity Measures Between Probability Density Functions." *City* 1: 1.
- Chen, B., Y. Tu, Y. Song, et al. 2021. "Mapping Essential Urban Land Use Categories With Open Big Data: Results for Five Metropolitan Areas in The United States of America." *ISPRS Journal of Photogrammetry and Remote Sensing* 178: 203–218.
- Cowie, S., R. Arthur, and H. T. Williams. 2018. "@ choo: Tracking Pollen and Hayfever in the UK Using Social Media." *Sensors* 18: 4434.
- CSY. 2018. *China Statistical Yearbook*. Beijing: China Statistical Publishing House.
- Du, S., S. Du, B. Liu, and X. Zhang. 2021. "Mapping Large-Scale and Fine-Grained Urban Functional Zones From VHR Images Using a Multi-Scale Semantic Segmentation Network and Object Based Approach." *Remote Sensing of Environment* 261: 112480.
- Du, Z., X. Zhang, W. Li, F. Zhang, and R. Liu. 2020. "A Multi-Modal Transportation Data-Driven Approach to Identify Urban Functional Zones: An Exploration Based on Hangzhou City, China." *Transactions in GIS* 24: 123–141.
- Fang, F., L. Zeng, S. Li, et al. 2022. "Spatial Context-Aware Method for Urban Land Use Classification Using Street View Images." *ISPRS Journal of Photogrammetry and Remote Sensing* 192: 1–12.
- Gao, S., K. Janowicz, and H. Couclelis. 2017. "Extracting Urban Functional Regions From Points of Interest and Human Activities on Location-Based Social Networks." *Transactions in GIS* 21: 446–467.
- Geng, X. 2016. "Label Distribution Learning." *IEEE Transactions on Knowledge and Data Engineering* 28: 1734–1748.
- Gong, F. Y., Z. C. Zeng, F. Zhang, X. Li, E. Ng, and L. K. Norford. 2018. "Mapping Sky, Tree, and Building View Factors of Street Canyons in a High-Density Urban Environment." *Building and Environment* 134: 155–167.
- Gong, P., B. Chen, X. Li, et al. 2020. "Mapping Essential Urban Land Use Categories in China (EULUC-China): Preliminary Results for 2018." *Science Bulletin* 65: 182–187.
- Hu, S., S. Gao, L. Wu, et al. 2021. "Urban Function Classification at Road Segment Level Using Taxi Trajectory Data: A Graph Convolutional Neural Network Approach." *Computers, Environment and Urban Systems* 87: 101619.
- Huang, W., L. Cui, M. Chen, D. Zhang, and Y. Yao. 2022. "Estimating Urban Functional Distributions With Semantics Preserved POI Embedding." *International Journal of Geographical Information Science* 36: 1905–1930.
- Huang, W., J. Wang, and G. Cong. 2024. "Zero-Shot Urban Function Inference With Street View Images Through Prompting a Pretrained Vision-Language Model." *International Journal of Geographical Information Science* 38: 1–29.
- Huang, W., D. Zhang, G. Mai, X. Guo, and L. Cui. 2023. "Learning Urban Region Representations With POIs and Hierarchical Graph Infomax." *ISPRS Journal of Photogrammetry and Remote Sensing* 196: 134–145.
- Huang, X., T. Hu, J. Li, Q. Wang, and J. A. Benediktsson. 2018. "Mapping Urban Areas in China Using Multisource Data With a Novel Ensemble SVM Method." *IEEE Transactions on Geoscience and Remote Sensing* 56: 4258–4273.
- Huang, X., J. Yang, J. Li, and D. Wen. 2021. "Urban Functional Zone Mapping by Integrating High Spatial Resolution Nighttime Light and Daytime Multi-View Imagery." *ISPRS Journal of Photogrammetry and Remote Sensing* 175: 403–415.
- Lane, N. D., E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell. 2010. "A Survey of Mobile Phone Sensing." *IEEE Communications Magazine* 48: 140–150.
- Li, X., T. Hu, P. Gong, et al. 2021. "Mapping Essential Urban Land Use Categories in Beijing With a Fast Area of Interest (AOI)-Based Method." *Remote Sensing* 13: 477.
- Liu, H., Y. Xu, J. Tang, et al. 2020. "Recognizing Urban Functional Zones by a Hierarchical Fusion Method Considering Landscape Features and Human Activities." *Transactions in GIS* 24: 1359–1381.
- Liu, X., J. He, Y. Yao, et al. 2017. "Classifying Urban Land Use by Integrating Remote Sensing and Social Media Data." *International Journal of Geographical Information Science* 31: 1675–1696.
- Liu, X., and Y. Long. 2016. "Automated Identification and Characterization of Parcels With OpenStreetMap and Points of Interest." *Environment and Planning B: Planning and Design* 43: 341–360.
- Liu, Y., X. Liu, S. Gao, et al. 2015. "Social Sensing: A New Approach to Understanding Our Socioeconomic Environments." *Annals of the Association of American Geographers* 105: 512–530.

- Liu, Y., X. Zhang, X. Kong, R. Wang, and L. Chen. 2018. "Identifying the Relationship Between Urban Land Expansion and Human Activities in the Yangtze River Economic Belt, China." *Applied Geography* 94: 163–177.
- Liu, Y., A. X. Zhu, J. Wang, W. Li, G. Hu, and Y. Hu. 2019. "Land-Use Decision Support in Brownfield Redevelopment for Urban Renewal Based on Crowdsourced Data and a Presence-and-Background Learning (PBL) Method." *Land Use Policy* 88: 104188.
- Lu, J., V. Goswami, M. Rohrbach, D. Parikh, and S. Lee. 2020. "12-in-1: Multi-Task Vision and Language Representation Learning." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10437–10446.
- Lu, W., C. Tao, H. Li, J. Qi, and Y. Li. 2022. "A Unified Deep Learning Framework for Urban Functional Zone Extraction Based on Multi-Source Heterogeneous Data." *Remote Sensing of Environment* 270: 112830.
- Qian, J., Z. Liu, Y. Du, et al. 2021. "Quantify City-Level Dynamic Functions Across China Using Social Media and POIs Data." *Computers, Environment and Urban Systems* 85: 101552.
- Sak, H., A. W. Senior, and F. Beaufays. 2014. "Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition." *arXiv Preprint arXiv:1402.1128*.
- Song, Z., H. Wang, S. Qin, et al. 2022. "Building-Level Urban Functional Area Identification Based on Multi-Attribute Aggregated Data From Cell Phones—A Method Combining Multidimensional Time Series With a SOM Neural Network." *ISPRS International Journal of Geo-Information* 11: 72.
- Wang, Y., Y. Gu, M. Dou, and M. Qiao. 2018. "Using Spatial Semantics and Interactions to Identify Urban Functional Regions." *ISPRS International Journal of Geo-Information* 7: 130.
- Wu, L., X. Cheng, C. Kang, D. Zhu, Z. Huang, and Y. Liu. 2020. "A Framework for Mixed-Use Decomposition Based on Temporal Activity Signatures Extracted From Big Geo-Data." *International Journal of Digital Earth* 13: 708–726.
- Xiao, F. 2019. "Multi-Sensor Data Fusion Based on the Belief Divergence Measure of Evidences and the Belief Entropy." *Information Fusion* 46: 23–32.
- Xing, H., and Y. Meng. 2018. "Integrating Landscape Metrics and Socioeconomic Features for Urban Functional Region Classification." *Computers, Environment and Urban Systems* 72: 134–145.
- Yang, M., B. Kong, R. Dang, and X. Yan. 2022. "Classifying Urban Functional Regions by Integrating Buildings and Points-of-Interest Using a Stacking Ensemble Method." *International Journal of Applied Earth Observation and Geoinformation* 108: 102753.
- Yao, Y., X. Li, X. Liu, et al. 2017. "Sensing Spatial Distribution of Urban Land Use by Integrating Points-of-Interest and Google Word2Vec Model." *International Journal of Geographical Information Science* 31: 825–848.
- Yin, J., J. Dong, N. A. S. Hamm, et al. 2021. "Integrating Remote Sensing and Geospatial Big Data for Urban Land Use Mapping: A Review." *International Journal of Applied Earth Observation and Geoinformation* 103: 102514.
- Yu, B., T. Lian, Y. Huang, et al. 2019. "Integration of Nighttime Light Remote Sensing Images and Taxi GPS Tracking Data for Population Surface Enhancement." *International Journal of Geographical Information Science* 33: 687–706.
- Yuan, J., Y. Zheng, and X. Xie. 2012. "Discovering Regions of Different Functions in a City Using Human Mobility and POIs." In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 186–194.
- Zhai, W., X. Bai, Y. Shi, Y. Han, Z. R. Peng, and C. Gu. 2019. "Beyond Word2vec: An Approach for Urban Functional Region Extraction and Identification by Combining Place2vec and POIs." *Computers, Environment and Urban Systems* 74: 1–12.
- Zhang, D., J. Wan, Z. He, et al. 2016. "Identifying Region-Wide Functions Using Urban Taxicab Trajectories." *ACM Transactions on Embedded Computing Systems* 15: 1–19.
- Zhang, J., Z. Yin, P. Chen, and S. Nicelle. 2020. "Emotion Recognition Using Multi-Modal Data and Machine Learning Techniques: A Tutorial and Review." *Information Fusion* 59: 103–126.
- Zhang, X., W. Li, F. Zhang, R. Liu, and Z. Du. 2018. "Identifying Urban Functional Zones Using Public Bicycle Rental Records and Point-of-Interest Data." *ISPRS International Journal of Geo-Information* 7: 459.
- Zhang, Y., N. Chen, W. Du, Y. Li, and X. Zheng. 2021. "Multi-Source Sensor Based Urban Habitat and Resident Health Sensing: A Case Study of Wuhan, China." *Building and Environment* 198: 107883.
- Zhang, Y., Z. Chen, X. Zheng, N. Chen, and Y. Wang. 2021. "Extracting the Location of Flooding Events in Urban Systems and Analyzing the Semantic Risk Using Social Sensing Data." *Journal of Hydrology* 603: 127053.
- Zhang, Y., M. P. Kwan, and H. Ma. 2024. "Sensing noise exposure and its inequality based on noise complaint data through vision-language hybrid method." *Applied Geography* 171: 103369.
- Zhang, Y., Q. Li, W. Tu, K. Mai, Y. Yao, and Y. Chen. 2019. "Functional Urban Land Use Recognition Integrating Multi-Source Geospatial Data and Cross-Correlations." *Computers, Environment and Urban Systems* 78: 101374.
- Zhang, Y., P. Liu, and F. Biljecki. 2023. "Knowledge and Topology: A Two Layer Spatially Dependent Graph Neural Networks to Identify Urban Functions With Time-Series Street View Image." *ISPRS Journal of Photogrammetry and Remote Sensing* 198: 153–168.
- Zhang, Y., F. Zhang, and N. Chen. 2022. "Migratable Urban Street Scene Sensing Method Based on Vision Language Pre-Trained Model." *International Journal of Applied Earth Observation and Geoinformation* 113: 102989.
- Zhou, W., D. Ming, X. Lv, K. Zhou, H. Bao, and Z. Hong. 2020. "SO-CNN Based Urban Functional Zone Fine Division With VHR Remote Sensing Image." *Remote Sensing of Environment* 236: 111458.