Enhancing Urban Land Use Identification Using Urban Morphology

Chuan Lin †, Guang Li †, Zegen Zhou, Jia Li, Hongmei Wang * and Yilun Liu *

Abstract: Urban land use provides essential information about how land is utilized within cities, which is critical for land planning, urban renewal, and early warnings for natural disasters. Although existing studies have utilized multi-source perception data to acquire land use information quickly and at low cost, and some have integrated urban morphological indicators to aid in land use identification, there is still a lack of systematic discussion in the literature regarding the potential of three-dimensional urban morphology to enhance identification effectiveness. Therefore, this paper aims to explore how urban three-dimensional morphology can be used to improve the identification of urban land use types. This study presents an innovative approach called the UMH–LUC model to enhance the accuracy of urban land use identification. The model first conducts a preliminary classification using points of interest (POI) data. It then improves the results with a dynamic reclassification based on floor area ratio (FAR) measurements and a variance reclassification using area and perimeter metrics. These methodologies leverage key urban morphological features to distinguish land use types more precisely. The model was validated in the Pearl River Delta urban agglomeration using random sampling, comparative analysis and case studies. Results demonstrate that the UMH–LUC model achieved an identification accuracy of 81.7% and a Kappa coefficient of 77.6%, representing an 11.9% improvement over a non-morphology-based approach. Moreover, the overall disagreement for UMH–LUC is 0.183, a reduction of 0.099 compared to LUC without urban morphology and 0.19 compared to EULUC-China. The model performed particularly well in identifying residential land, mixed-use areas and marginal lands. This confirms urban morphology’s value in supporting low-cost, efficient land use mapping with applications for sustainable planning and management.

Keywords: land use identification; social sensing data; urban morphology; Pearl River Delta urban agglomeration

1. Introduction

Urban land use encompasses fundamental details regarding the utilization of land within a city, to some extent reflecting the socio-economic functions of the land. At a refined scale, such information plays a crucial role in various aspects, including land planning, urban renewal [1,2], evaluation of environmental vulnerability [3,4], and natural disaster early warning [5,6]. It not only serves as an essential reference for governmental decision-making in land management [7,8], but also forms the foundation for all societal stakeholders to comprehend and utilize land resources for socio-economic activities [9–11].

However, conducting surveys to obtain urban land use information poses significant challenges. On the one hand, the functional aspects of land use in reality are highly


Academic Editor: Monika Kopecká

Received: 17 April 2024
Revised: 21 May 2024
Accepted: 27 May 2024
Published: 28 May 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/).
complex. Not only does land often serve multiple purposes, but land use changes frequently, making it difficult to categorize using a single land use type [12]. On the other hand, the acquisition of detailed urban land use information at a conventional scale primarily relies on manual interpretation of remote sensing imagery or on-site investigations, necessitating the mobilization of various resources and substantial efforts [13–15]. For instance, in the United States, the National Land Use and Land Cover Survey is conducted once every five years. This survey is a collaborative effort between the U.S. Geological Survey (USGS) and the U.S. Department of Agriculture (USDA) [16,17]. In China, the frequency of land use surveys is once every ten years. These surveys are organized and implemented through a collaborative effort involving multiple departments, under the leadership of a working group established by the State Council of China (2017, 2019). Until now, China has conducted three national land use surveys, with the most recent one starting in 2017 and concluding in 2020. The process of conducting the survey and publishing the resulting data took four years. The human and material resources mobilized for this endeavor are difficult to quantify. From this perspective, the acquisition of urban land use information at fine scale presents challenges of extensive workload, high economic costs, and low efficiency. Moreover, it encompasses significant temporal spans and exhibits strong lag effects, resulting in slow information update rates. Consequently, it becomes arduous to meet the dynamic monitoring needs of urban land use changes and assist urban land planning and management within the context of rapid economic development and accelerated urban expansion [18]. Therefore, a central research concern in the domain of urban planning and development pertains to the effective, efficient, and reliable identification of urban land use patterns through the utilization of contemporary scientific technology and information data. This pursuit is essential in addressing the demands of urban planning and development [19,20].

2. Related Literature

The advancement of urban land use identification research has been driven by both practical needs and technological development. From the perspective of data sources and identification methods, existing studies can be categorized into three types: remote sensing image classification method, social sensing Geo-data identification method, and multi-source data fusion identification method.

The first method involves utilizing remote sensing imagery to classify land use functions, supplemented by manual sampling verification. This method is widely applied in practical land use management. With the advancement of remote sensing technology, numerous high-resolution remote sensing images can be processed to extract and identify urban land use types or functions. For example, Hu [21] pioneered the use of decision tree methods to automatically identify fine-grained urban land use types with manually labeled areas such as office, industrial, municipal utilities, and transportation sectors. Mialhe [22] combines participatory activities with remote sensing analysis into an integrated methodology to describe and explain land-cover changes. However, remote sensing imagery is constrained by pixel size, resulting in spatially discontinuous classification results that do not accurately represent the actual land use patterns [23,24]. Additionally, remote sensing imagery is limited to observing optical characteristics of the horizontal surface of the urban, and relying solely on surface physical spectral information is insufficient for detailed classification of urban land types [25,26].

The second method involves social sensing (or crowdsourcing) Geo-data to identify land use functions [27]. With the widespread use of smartphones and the normalization of location-based services in daily life, a large amount of data with Geo-information has been generated, such as Point of Interest (POI) data, population density data, and more [28,29]. These data contain rich spatial socio-economic information and can serve as factors for representing urban land use characteristics [30–32]. For example, Liu and Long [33] employed a method for automatic identification and description of land parcels, which is based on OpenStreetMap (OSM) and POI data. This method enables the rapid
delineation of parcel boundaries, generation of various parcel attributes, and detection of urban functionality, development density, and mixed land use. Zhao [34] conducted a study that integrated multiple sources of geospatial data, including Weibo check-ins, taxi tracking, building surveys, as well as natural landscape features such as water, vegetation, and urban open spaces, to map land use types in Shenzhen City.

Compared to manual survey methods, the social sensing Geo-data identification method offers a relatively cost-effective and efficient means of acquiring socio-economic information on land use, avoiding labor-intensive field surveys. However, the accessibility and reliability of social sensing are issues that require careful consideration. On the one hand, with the strengthening of personal privacy awareness and the establishment of laws and regulations, departments and companies that have access to the data are strengthening information confidentiality management, making it more challenging to obtain such data [35,36]. On the other hand, due to the diverse and open nature of social sensing data, the quality or representativeness of the data varies. Different data sources have different scales, standards, and bias [37,38]. It takes time and effort to assess the authenticity and credibility of the data, as well as to organize and summarize.

The third approach integrates remote sensing and social sensing data and employs machine learning methods to address the high-dimensional non-linear relationships between land use functions and recognition factors [39]. For instance, Liu [40] and Gong [41] utilized high-resolution remote sensing imagery, OSM, nighttime light, POI, and Tencent positioning service data to achieve urban land use function recognition. This research method has opened up a new avenue for identifying land use functions in urban areas. Subsequently, Chen [42] developed a plot-based urban land use classification and recognition model. The model employs a multi-layer stacking learning approach to automatically learn and integrate various sources of spatial data during the training process. However, persistent challenges remain. Firstly, the high dimensionality of multi-source data often necessitates the utilization of machine learning-based classification algorithms, such as random forests and neural networks, in these approaches. Nevertheless, a common drawback is the lack of interpretability regarding the effects and influences of various factors on the classification results. Secondly, since these recognition models heavily depend on a substantial volume of sample data, the generalization performance of the models has yet to be validated. When applying a model developed based on a specific urban sample to other cities, the accuracy often falls short.

In general, existing methods have made beneficial attempts to integrate remote sensing and social sensing data sources for urban land use investigations. However, these aforementioned studies primarily face two key issues. Firstly, regarding remote sensing data sources, they solely capture the optical features of the two-dimensional urban surface, overlooking the potential of urban spatial morphology for characterizing land use types. Numerous studies have indicated that urban spatial morphology, which represents the spatial manifestation of urban socio-economic development [43,44], is closely intertwined with land use functions [45]. Generally, different land use functions exhibit distinct spatial morphologies [46]. For instance, the disparities in form between residential and industrial land are evident [47]. Residential areas often feature densely arranged buildings with consistent morphological patterns, such as similar building heights, floor area ratios, and distribution patterns (Figure 1a). On the other hand, industrial areas typically have lower building heights, larger individual building footprints, and wider spacing between structures (Figure 1b). Therefore, by incorporating the morphological characteristics of buildings into the identification of land use functions based on urban spatial morphology, it is theoretically feasible to enhance the accuracy of identification [48,49].
Secondly, social sensing Geo-data present challenges of representativeness and bias. Social sensing data often revealing only partial land functions and overlooking substantive functional types in specific usage scenarios. For instance, it could misclassify land use types, like mixed-use commercial and residential complexes as exclusively commercial land [40,50,51]. When it comes to residential land, reliance solely on urban POI or nighttime light data for identification can lead to significant deviations. The primary reason is that POI predominantly reflect commercial and fundamental service infrastructure and lack representativeness of the distribution of residential functions [41,51]. Meanwhile, nighttime light data, due to spatial resolution limitations and the diffusion effect of light, can easily confound residential and commercial land use. However, despite integrating remote sensing data and social sensing data, this method still encounters challenges, including limited generalization performance due to reliance on sample data and a lack of interpretability in the recognition results [52,53].

Theoretically, the introduction of urban morphological features can effectively rectify deviations in the identification of residential land [54,55]. Residential land is characterized by a uniform and homogeneous arrangement of buildings, typically exhibiting similar sizes, shapes, and heights. This orderly and organized disposition is particularly evident in medium to large residential neighborhoods. By extracting such features from urban morphological data, it becomes possible to accurately capture the residential attributes within different land use types. Likewise, the incorporation of urban morphological information can enhance the differentiation between various land use types and improve the accuracy of identification results. Although some studies have integrated indicators of urban morphology to aid in land use identification, the existing literature lacks a systematic discussion on the potential of three-dimensional spatial morphology in enhancing identification effectiveness [45]. Therefore, the objective of this paper is to explore how urban three-dimensional morphology can be utilized to improve urban land use type identification.

Section 3 of this paper will elucidate the methodology for incorporating urban morphology into land use identification. In Section 4, we will present a case study to analyze and validate the accuracy of the identification method and showcase the research results. This will be followed by a dedicated discussion in Section 5. Finally, the conclusion will be provided in Section 6.

3. Methodology

3.1. Basic Idea

To enhance the identification of urban land use categories using urban morphology, this study proposes a two-stage approach for urban land use classification. In the initial
stage, a preliminary identification of urban land use result will be established with urban road networks, impervious surface, and POI data. In the subsequent stage, the preliminary identification will be improved by incorporating vertical and horizontal urban morphology factors. The vertical morphology will encompass factors such as building height and floor area ratio, while the horizontal morphology will consider the shape index and area variance of building footprints. Through the analysis of variance and other statistical indicators derived from these morphology features, our aim is to refine and enhance the land use identification. Based on the above idea, we have developed the Urban Morphology based Land Use Classification (UMH–LUC) model, whose framework is illustrated in Figure 2. The UMH–LUC consists of two modules: Module 1—Preliminary identification based on multiple social sensing data; and Module 2—Identification improvement based on urban morphology.

Figure 2. The framework of Urban Morphology based Land Use Categories identification (UMH–LUC) model.

3.2. Preliminary Land Use Type Identification Based on POI Data

In the module 1, a POI area weighted method is adopted for the preliminary land use identification. Before identification, the built-up lands are segmented as blocks (identification units) based on city road networks. To reduce interference and focus on built-up areas, non-building areas are excluded by overlaying with an urban impervious surface product. Furthermore, the latest road network data are obtained from the OpenStreetMap website. Buffer zones are created based on the road surface width of different levels of roads. Finally, the built-up land vector image is intersected with the road network data to generate identification units that align with the actual land use conditions.

To establish connections between the attributes reflected by the POI data and the corresponding land use types, the POI categories are reclassified based on urban land use types. The method for identifying land use types using POI data involves calculating the quantity of different types of POIs within a block and assigning weights to them [56]. The dominant attribute type of POI within the block represents the primary land use type. The commonly used approach to determine these weights is based on the reference
land area of various types of POI data. The reference land area for different types of POIs is determined based on the study area to ensure the accuracy of the preliminary identification results. It should be particularly noted that the reference land areas of different types of POI can be obtained from official urban planning standards in relevant areas to ensure that the reference standards used are adapted to the geographical characteristics of the study area [57].

After weighting the different types of POI data, the POI attribute type with the highest weight is selected as the dominant type within the block, enabling the determination of urban land use function. Equation (1) is employed, utilizing the proportion of different functions within each block, as well as the quantity and average reference area of the POI points:

\[
S_k = \frac{\sum_{i=1}^{n} P_i \times A_i}{\sum_{i=1}^{m} P_i \times A_i}
\]

where \(S_k\) represents the proportion of the \(k\)th function within the block. \(i\) represents the \(i\)th class of POI for the \(k\)th function. \(n\) denotes the number of secondary classes of POI points for the \(k\)th function, while \(m\) represents the total number of all types of POI, including the \(k\)th function. \(P_i\) and \(A_i\) correspond to the quantity and average reference area of the \(i\)th class of POI within the block. \(P_j\) and \(A_j\) refer to the quantity and average reference area of the \(j\)th class of POI within the block. By applying this formula, we calculate the proportions of different functions within each block. The primary land use type classification result for the block is determined by labeling the function with the highest proportion of POIs.

3.3. Improvement of the Preliminary Identification Based on the Urban Spatial Morphology

In Module 2, the improvement of the preliminary identification results is carried out based on the vertical and horizontal spatial morphology. A vertical morphology dissimilarity index (VI) and a horizontal morphology variance index (HI) are proposed.

3.3.1. Identification Results Improvement with Vertical Morphological Feature

The vertical morphology-based reclassification procedure is an iterative process where each iteration involves adjusting the classification of each block and each land use type until all blocks have been processed. Initially, the VI is calculated with Equation (2) for each block. The comparison is conducted by evaluating the block’s FAR against the average FAR of all blocks belong to the same land use type. This step identifies blocks that may have been misclassified if the VI is greater than the Z-Score threshold \(t_{\text{threshold}}\) of particular land use \(L\). The \(t_{\text{threshold}}\) can be dynamically adjusted to suit specific conditions in different cities or regions.

Next, the identified samples that are potentially misclassified undergo reclassification attempts into other land use types. This step involves comparing the sample’s FAR with the FAR distribution of other land use types. If the sample’s FAR aligns with the distribution of a particular land use type, it is reclassified accordingly. If a sample cannot be matched with any land use type, it is classified as mixed-use land. It is noteworthy that mixed-use land is not classified in the primary identification based on POI data. This is because POI data predominantly emphasize commercial attributes, making it challenging to showcase other significant features. Artificially reducing the weight of commercial attributes in the POI data would distort the data and undermine their original connotation. Therefore, identifying mixed-use land in this step ensures that it considers both the commercial attributes attributed by the POI data and other attributes derived from the vertical morphology of the city.

\[
VI_i = \left| \frac{X_i - X_L}{S_L} \right|
\]
where $X_l$ represents the FAR of the sample block, $\bar{X}_L$ represents the average FAR for particular land use $L$, and $S_L$ represents the standard deviation of the FAR for particular land use $L$.

### 3.3.2. Identification Results Improvement with Horizontal Morphological Feature

The horizontal morphology-based reclassification procedure quantifies the similarity of building footprint within a block by evaluating their horizontal morphological differences. A lower value of the building footprint dissimilarity index indicates a higher level of similarity in horizontal morphology. Based on this analysis, blocks with buildings can be classified into three categories: highly similar, moderately similar, and dissimilar, according to their building morphology specific to different land use types. Consequently, basing the analysis on urban horizontal morphological characteristics can improve the preliminary identification accuracy of residential and mixed-use land, which is reduced due to the bias inherent in POI data.

The model pre-processes the building patch vector data to calculate various parameters, such as building number, area, and perimeter. One of the horizontal morphological features used in the analysis is the number of buildings within a block. This feature provides information about the complexity of the block. If the number of buildings within a block exceeds a certain threshold (denoted as $thod_{num1}$), it indicates that the block’s horizontal morphological feature is relatively complex and requires further analysis and adjustment of land use classification. Consequently, the blocks selected through this process will be designated as mixed-use land.

In addition to the number of buildings, the model also calculates the HI for each block that surpasses the threshold for the number of buildings. The HI is calculated using Equation (3), determined by the variances of building footprint’s area and perimeter. Based on the characteristic of high urban morphological similarity in residential areas, a selection of blocks is made using HI. If the number of buildings and the HI within a block exceeds a certain threshold (denoted as $thod_{num2}$ and $thod_{HI}$), it indicates that these blocks, initially misclassified as different land use types, are subsequently reclassified as residential land.

$$HI_k = \frac{w_1}{n} \frac{1}{\sqrt{n}} \frac{\sum_{i=1}^{n} (A - \bar{A})^2}{\sum_{i=1}^{n} (A - \bar{A})^2} + w_2 \frac{1}{n} \frac{1}{\sqrt{n}} \frac{\sum_{i=1}^{n} (L - \bar{L})^2}{\sum_{i=1}^{n} (L - \bar{L})^2}$$

In the equation, $HI$ represents the horizontal morphology variance of the Kth block. $A$ denotes the area of each building patch, and $\bar{A}$ represents the average area of all building patches within the Kth block. $L$ denotes the perimeter of each building patch, and $\bar{L}$ represents the average perimeter of all building patches within the Kth block. $w_1$ and $w_2$ represent the respective weight values, and it is assumed that $w_1 + w_2 = 1$.

Since the focus of this paper is not to discuss the perimeter and area which have a greater influence on the morphological urban characteristics, the weight of the two is equal, i.e., $w_1 = w_2$.

Work in computing VI and HI needs to be achieved jointly through ArcGIS Pro 3.0.2 and Python 3.9. The specific flow chart of the model is shown in Figure 3.
3.4. Verification Method

To validate the performance of the presented method and the effectiveness of incorporating urban spatial morphology, the land use identification results will be compared to two other datasets: the preliminary identification result without urban spatial morphology information, and the EULUC-China [41] dataset, which is identified using multi-source fusion data and a machine learning model. EULUC-China is one of the high-accuracy methods for fine-scale land use identification and shares the same study area as this study; however, it does not incorporate urban spatial morphology into the identification process.

According to Pontius Jr. [58–60], compared to the Kappa coefficient, the cross-tabulation matrix provides a more precise evaluation and comparison of classifications. Therefore, this paper uses two approaches in combination to assess the accuracy of the results. The validation process involves random sampling from the identified blocks, and a confusion matrix is used to assess several key indicators, including Producer’s Accuracy, User’s Accuracy, Quantity disagreement, Allocation disagreement, Overall Accuracy, and Kappa coefficient. These indicators provide insights into the model’s performance and its ability to accurately classify land use types.

To validate the results, two reference datasets are selected for cross-checking: (1) land use survey data from the Third National Land Survey, which was conducted during the period of 2017–2020, and serve as a reliable reference for land use classification and helps evaluate the accuracy of the model’s predictions; (2) manually identified data
with high-resolution remote sensing images and street view images, which provide a comprehensive and accurate reference for cross-checking the model’s results.

4. Case Study

4.1. Study Area

The Pearl River Delta urban agglomeration, located in the central-southern part of Guangdong Province, China is chosen as the study area. The agglomeration comprises nine prefecture-level cities: Guangzhou, Shenzhen, Foshan, Dongguan, Zhuhai, Zhongshan, Jiangmen, Huizhou, and Zhaoqing. The location are shown in Figure 4.

The total area of the Pearl River Delta region is 55,368.7 km², and it had a GDP of 10.5 trillion RMB in 2022. The permanent population of the region is 78.29 million. The Pearl River Delta is known for its high level of economic development and is considered a pioneer zone for China’s reform and opening-up. It serves as an important regional economic center and a hub for scientific and technological innovation and research and development in China.

The region’s economic development has led to the expansion of the built-up area, making it suitable for verifying the effectiveness of the proposed model for large-scale land use identification. The central area, particularly near the Guangzhou–Foshan–Shenzhen–Dongguan urban belt, exhibits a highly concentrated land use pattern, which can be analyzed using the model. Additionally, outer cities like Huizhou, Jiangmen, and Zhaoqing, have moderate levels of economic development, providing further insights into land use identification in diverse areas.

The research results obtained from the Pearl River Delta urban agglomeration are representative and applicable to other countries and regions. The region’s characteristics, including its economic development, diverse land use patterns, and urban complexity, make it a suitable case study for evaluating the proposed model’s effectiveness in large-scale land use identification.

![Figure 4. Study area and example of segmentation for identification units.](image)

4.2. Data Sources

This study utilized five datasets: POI data, Road network, Urban impervious surfaces, Urban building footprints data, and Urban heights data:
The POI data were collected from Amap (https://lbs.amap.com/, accessed on 1 December 2020), covering 62 types of institutions, such as government agencies, businesses, schools, shopping centers, and hospitals.

Road network data were obtained from OSM (https://www.openstreetmap.org, accessed on 7 December 2020), which were used for segmentation of identification units. The data include road level attributes, such as urban highways, main roads, and secondary roads. These data were used for hierarchical buffering analysis to obtain identification units.

The urban impervious surface data were obtained from the ESRI official website (https://livingatlas.arcgis.com, accessed on 25 June 2023), specifically the 2021 land use cover data for the study area.

The urban building footprints data were also obtained from Amap. They contained 2,129,478 building footprints within the study area.

The urban height data were sourced from a model developed by Wu et al. [61]. This model utilizes all-weather satellite observation data and considers factors such as shadows to estimate building heights. A random forest model was used to establish and optimize the estimation model, resulting in the calculation of building height grid data for the year 2020. Results show that building height simulation has a strong correlation with real observations at the national scale (RMSE of 6.1 m, MAE = 5.2 m, R = 0.77).

4.3. Land Use Identification

4.3.1. Experimentation and Parameter Setting

The experimental process and parameter settings can be succinctly described in two steps.

1. Preliminary Identification Based on POI Data: The impervious surface vector data were extracted from the 2021 land use cover data within the study area. Subsequently, the OSM road network data were used to create buffer zones based on road network classifications. The width of each road network level was determined according to the China urban road construction standards. The impervious surface vector data were then clipped using the buffer zones, resulting in the identification units utilized in this study.

To develop a novel POI classification system aligned with urban land classification, diverse POI points in the study area were reclassified. Equation (1) was employed to calculate the quantity of each POI type within individual blocks. The reference land area of each land use type was determined based on the criteria established by Xu [56], which are listed in Table 1.

<table>
<thead>
<tr>
<th>Classification of Land Use</th>
<th>Type of POI</th>
<th>Reference Land Area (ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public Management and Service</td>
<td>Government Agency</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Social Group</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Library and Exhibition Facility</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>Cultural Activity</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Higher Education Institution; Secondary School; Primary School</td>
<td>10;3;2</td>
</tr>
<tr>
<td></td>
<td>Research Institution</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>Public Gymnasium; Sports Ground</td>
<td>1;0.1</td>
</tr>
<tr>
<td></td>
<td>General Hospital; Specialty Hospital; Clinics and Pharmacy</td>
<td>6;2;0.1</td>
</tr>
<tr>
<td></td>
<td>Nursing Home; Welfare Home</td>
<td>1</td>
</tr>
<tr>
<td>Commercial Services</td>
<td>Catering Industry</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Large shopping mall; Supermarket; Convenience store</td>
<td>5;0.3;0.01</td>
</tr>
</tbody>
</table>
(2) Results Improvement with Morphological Feature: For vertical improvement, statistical analysis was conducted to determine the total building height for each block. Due to the difficulty of obtaining precise floor height data, we make an assumption about floor heights. According to the General Specifications for Civil Buildings GB 55031-2022 of China [62], the minimum clear floor height in areas intended for regular human activity should not be less than 2.00 m. Given that this specification represents a limit condition, and considering that various types of construction land typically include redundant space in actual construction, to ensure the universality of the research findings and approximate real-world conditions as closely as possible, this study sets the theoretical floor height at 3 m. A unit pixel area representing the building footprint (10 m × 10 m), and a theoretical floor height of 3 m was assumed to calculate FAR for each block. In this study, the $\text{thod}_L$ was established at 1.55. A proportion of 94.1% of the sample set corresponding to each land use type was retained, while the remaining 5.9% of misclassified blocks (also known as overflow blocks) were excluded and reassigned to their appropriate land use types.

Regarding horizontal improvement, the $\text{thod}_{num}$ was set at 10. This threshold takes into account the fact that commercial and public management and service lands typically have a moderate to low number of buildings, while residential and mixed-use lands tend to have a higher number. Additionally, for the threshold of HI ($\text{thod}_H$), a value of 200 was determined through calibration. This threshold effectively distinguishes between land use types with similar and dissimilar urban morphologies.

4.3.2. Land Use Identification Results

The identification results are shown in Figure 5, with a total of 26,389 identified land parcels. Among these, there are 8250 residential blocks, accounting for 31.26% of the total; 6311 commercial blocks, accounting for 23.92%; 4616 mixed-use land blocks, accounting for 17.49%; 3651 public management and service blocks, accounting for 13.84%; and 3469 industrial blocks, accounting for 13.15%.
To assess the accuracy of the identification results, a random sample of 1000 blocks was selected from the identified results for comparison, and a confusion matrix was computed. The confusion matrix is shown in Figure 6.
4.3.3. Results’ Validation and Comparisons

Figure 6 is a confusion matrix heatmap, which presents the accuracy assessment of land use identification for different land use types. In the plot, the diagonal cells are those identifying the correct categories, while also calculating the information for PA (producer accuracy) and UA (user accuracy) for each land type. Figure 7 displays the quantity disagreement and allocation disagreement of UMH–LUC. Quantity disagreement illustrates the disparity in the proportion of identified versus actual land categories, while allocation disagreement reveals the rates of commission and omission in the identification results. Figure 7a illustrates the intensity (proportion) of both types of disagreement across various land use categories, allowing for visual discernment of the inconsistency emphasis for different land types. Figure 7b displays the proportion of disagreement for different land use categories compared to actual land use types. From these figures, we analyze the reasons for the discrepancies in identification accuracy across different land use types.

Firstly, for blocks with relatively single-use functions, such as public management and service, storage, transportation, and special use areas (including religious land and tourism land), high recognition accuracy is achieved. The overall disagreement rates for these categories are generally lower than the overall disagreement (Figure 7b blue line) across all land types. This can be attributed to the POI data providing clear indications of these specific land use types.

Secondly, land use types with residential functions, such as ‘mixed-use land’ and ‘residential’, demonstrate favorable identification results. Particularly, residential land has an overall disagreement rate of 0.133, with quantity disagreement at 0.05 and allocation disagreement at 0.08. This effectiveness stems from their distinct urban morphology features, which facilitate precise differentiation. Notably, mixed-use land exhibits higher errors in quantity allocation, reaching 0.2, due to sparse building data in some suburban areas and less pronounced floor area ratio characteristics, resulting in overall disagreement rates exceeding those of all land types.

However, challenges persist in the accurate identification of certain land use types. Notably, the allocation disagreement for commercial land is exceptionally high at 0.31, indicating a frequent misclassification of other land types as commercial. This high error rate in identifying commercial land is largely due to the bias of POI data towards commercial attributes, which may lead to misclassifications.

Industrial land also experiences a significant level of confusion. Its overall disagreement rate has reached 0.41, with a quantity disagreement of 0.17 and an allocation disagreement of 0.24. On one hand, industrial land is typically located in suburban areas, away from city centers, where sparse road networks result in larger identification units. These larger units inevitably encompass varying land use functions, leading to misclassifications due to the scale of identification units. On the other hand, industrial land manifests in various forms. In the Pearl River Delta region, for example, it can appear in the form of vast industrial parks or as factories established through village collective investments. The former, with large areas and well-developed infrastructure, including dormitory and commercial zones, tends to be misclassified as mixed-use or commercial land. The latter, usually found in towns and surrounded by rural homesteads and self-built houses, is easily misidentified as residential land.

Overall, these challenges highlight the complexity of accurately identifying certain land use types and the influence of various factors, including data biases and the diverse characteristics of different land use categories.
The identification results were compared with the preliminary identification results (denoted as LUC without urban morphology) and the EULUC-China classification product. The accuracy comparisons of three land use identifications are presented in Table 2 and Figure 8.
Table 2. Comparison of the accuracy of the three identification results.

<table>
<thead>
<tr>
<th>Type</th>
<th>LUC without Urban Morphology</th>
<th>UMH–LUC</th>
<th>EULUC-China</th>
</tr>
</thead>
<tbody>
<tr>
<td>OA</td>
<td>69.81</td>
<td>81.70</td>
<td>62.70</td>
</tr>
<tr>
<td>Kappa</td>
<td>66.58</td>
<td>77.56</td>
<td>62.49</td>
</tr>
</tbody>
</table>

Table 2 showcases that the UMH–LUC can achieve the highest comparative accuracy among the three methods in distinguishing different land use types. The UMH–LUC model achieved an identification accuracy of 81.70% with a Kappa coefficient of 77.56, which was an 11.89% improvement compared to LUC without urban morphology methods and a 20.50% improvement compared to EULUC-China.

Figure 8 showcases more carefully the difference in the identification results of the three methods through the quantity disagreement and allocation disagreement. The quantity disagreement in UMH–LUC is 0.046, which is significantly lower than those observed in the other two methods, suggesting that this model’s classification results most closely mirror real-world proportions. The allocation disagreement is 0.137, which is similar to that of the EULUC-China method but lower than that observed in LUC without urban morphology, due primarily to higher errors in specific land categories (commercial and industrial). However, overall, this indicates a better accuracy of internal land categories within the model. The overall disagreement for UMH–LUC is 0.183, a reduction of 0.099 compared to LUC without urban morphology and 0.19 compared to EULUC-China.

4.3.4. Situ Validation

To further illustrate and compare the differences in land use identification results, case situ validations have been conducted on three selected regions: Huaqiangbei Commercial District in Shenzhen, Tianhe Road Commercial District in Guangzhou, and Xiaowan Neighborhood in Zhuhai. These cases aim to visually demonstrate the variations and discrepancies in land use identification among the three methods, providing a comprehensive understanding of their performance and limitations.

(1) Huaqiangbei Commercial District in Shenzhen Case Study

The first case study focuses on the Huaqiangbei Commercial District, which is situated in the central urban district of Shenzhen, as shown in the Figure 9. Huaqiangbei is
renowned as the leading market for electronic products in Shenzhen and is considered one of the largest electronic markets globally. The district primarily specializes in the wholesaling and retailing of electronic goods and encompasses various subsidiary industries, including electronic components and computer accessories. As of December 2020, the district had 717 commercial establishments along its streets, featuring more than 20 large-scale shopping centers, each exceeding 10,000 square meters in business area. Impressively, Huaqiangbei housed over 10,000 individual taxpayers, generated annual sales revenue of approximately 23 billion yuan, and employed a workforce of 130,000.

Figure 9. Shenzhen Huaqiang North Business District case: (a1) UMH–LUC result; (a2) EULUC-China result; (a3) LUC without urban morphology result; (b) Study Area; (c) Remote sensing image of the case area; (d) Street view of the case area.

Huaqiangbei is characterized by a diverse industrial composition, dense commercial arrangement, and a unique integration of residential land uses. It features a mix of land use types, including residential zones, office spaces, commercial outlets, and public service amenities, creating a comprehensive business ecosystem. This ecosystem seamlessly combines wholesale and retail activities based on electronic products with commercial, office, hotel, cultural, entertainment, and residential functionalities. In terms of land use classification, both the EULUC-China and the LUC without urban morphology methods predominantly categorize Huaqiangbei as a commercial zone with ancillary residential and public management and service spaces. However, the UMH–LUC identification proposed in this study depicts Huaqiangbei as mixed-use land, which more accurately reflects the observed reality on the ground.

(2) Tianhe Road Commercial District in Guangzhou Case Study

The second case focuses on the Tianhe Road Commercial District in Guangzhou, which is considered the central business hub of the city. Situated in the Tianhe District...
and positioned on the city’s newly established central axis, this district covers a commercial expanse of 1.4 million square meters. It is home to over 20 substantial shopping centers, 12 international five-star hotels, and more than 800 global brands. The district experiences high daily foot traffic, with over 1.5 million individuals visiting on a regular day, which nearly quadruples during holiday periods. Its annual sales have surpassed one trillion yuan, demonstrating its economic vibrancy. The district’s growth has also stimulated the development of nearby real estate sectors. Along Tianhe Road, numerous residential communities have emerged, accompanied by a comprehensive range of infrastructural amenities such as educational institutions, healthcare facilities, and government offices. As a result, the district has evolved into a highly functional area characterized by a dense and integrated mix of commercial and residential land uses.

One distinctive feature of the Tianhe Road Commercial District is the prevalent “ground-floor economy” mode, where commercial activities dominate street-level spaces, while residential units occupy the upper stories. This mixed-use land pattern, particularly pronounced in this district due to its robust economic vibrancy, exemplifies a sophisticated urban land use strategy. For example, the Liuyun residential community, highlighted by the red circle in the accompanying image (Figure 10d), is well-known for its vital ground-floor economy. However, accurately recognizing this intricate land use configuration poses a significant challenge during the identification process. Within the framework of the EULUC-China classification, this community is primarily classified as residential land. On the other hand, the LUC without urban morphology method identifies it as a combination of commercial and public management and services land. Ultimately, the UMH–LUC method categorizes this area as mixed-use land, providing a more accurate representation of the complex reality observed on the ground.

Figure 10. Guangzhou Tianhe Road business circle case: (a1) UMH–LUC result; (a2) EULUC-China result; (a3) LUC without urban morphology result; (b) Study Area; (c) Remote sensing image of the case area; (d) Street view of the case area.
The case study of the Tianhe Road Commercial District highlights the challenges in accurately identifying and categorizing mixed-use land areas, particularly those characterized by a prevalent ground-floor economy. It further demonstrates the effectiveness of the UMH–LUC approach in capturing the intricate land use composition and providing a more accurate classification in such complex urban settings.

(3) Xiawan Neighborhood in Zhuhai Case Study

The third case focuses on the Xiawan Neighborhood in Zhuhai, which is situated near Zhuhai’s Gongbei Port and is less than two kilometers away from Macau, as shown in the Figure 11. It is characterized by a dense population and primarily functions as a residential zone. However, in the land use identification results of both the EULUC-China and the LUC without urban morphology methods, certain residential areas within Xiawan are mistakenly identified as commercial zones. This misclassification arises from the presence of small shops and guesthouses, which are a result of the area’s advantageous transportation location and which enrich the commercial POI data. During the land-use identification process, the prominence of commercial characteristics tends to overshadow residential attributes, leading to an inaccurate categorization of these areas as commercial land.

In contrast, the identification results derived from the UMH–LUC method utilize morphological features extracted from urban morphology to reinforce the recognition of residential attributes, thereby improving the accuracy in differentiating residential land. This approach more accurately represents the true nature of the land use in the area, ensuring alignment with the actual conditions observed on-site.

The Xiawan case highlights the challenges faced in accurately identifying and distinguishing residential areas that may have commercial features. It underscores the effectiveness of the UMH–LUC approach in leveraging urban morphology to enhance the accuracy of land use identification, particularly in complex urban environments where mixed land use patterns are present.
The discernible precision contrast and the aforementioned validation cases underscore substantial disparities among the three methods’ results. When comparing the land use identification results of EULUC-China and LUC without urban morphology, it is evident that the actual identification outcomes are largely similar, with minor discrepancies attributed to the timeliness of the POI data. The granularity of the road network primarily affects the size of the identification units but does not alter the recognized land use functions.

The critical distinguishing aspect between the identification results obtained using the UMH–LUC approach and those derived from the LUC without urban morphology method, as well as EULUC-China, lies in the integration of urban morphological factors. This integration forms the core proposition of this study, which suggests that incorporating urban morphology enhances land use identification. Through rigorous accuracy validation and data analysis, it has been demonstrated that the inclusion of urban morphology significantly improves the effectiveness of land use identification. Moreover, the study validates that the developed model for extracting urban morphology features adeptly captures the spatial morphological characteristics of urban environments, resulting in optimal identification results.
5. Discussion

5.1. Distribution and Sensitivity of the Urban Morphological Factors

The proposed UMH–LUC model incorporates a morphology-based improvement method that relies on two important parameters: VI and HI. The impact of these parameters on the identification results is discussed in this section. Figure 12 presents a Sankey plot that visually illustrates the roles played by the vertical and horizontal reclassification models in the UMH–LUC approach.

Figure 12. Sankey plot of preliminary identification results reclassification based on urban morphological improvement.

5.1.1. The Impact of VI on the Identification Results

Figure 13a,b shows the distribution of VI and FAR for different land use types. It can be observed that the VI values exhibit relatively small differences among different land use types, except for mixed-use land, which has lower VI values due to the discreteness of the data. On the other hand, FAR values exhibit distinct range intervals for different land use types. Mixed-use land, characterized by its complex and diverse functions, has significantly higher FAR values. Residential land also tends to have higher FAR values, indicating a correlation between residential attributes and high FAR. Industrial land typically has lower FAR values due to safety requirements and the need for low-density spaces. Transportation land (stations), storage land (warehouses, logistics centers), and special land (temples, historical buildings) have lower FAR values compared to other land use types, reflecting their specific functions and characteristics.
The effectiveness of the model depends on the threshold setting. The VI threshold $\theta_{VI}$ determines whether a sample needs to be reclassified. Setting $\theta_{VI}$ too highly, especially close to or exceeding 2, may not result in significant reclassification since it would allow mixed-use land to retain its original land use type, which should have been screened out. Conversely, setting $\theta_{VI}$ close to 1 or lower may lead to a large number of mixed-use land occurrences, as many original land use types would meet the reclassification criteria. Therefore, in practical applications, it is crucial to set $\theta_{VI}$ reasonably, based on the overall distribution structure of FAR in the study area and the range of FAR values for different land use types.

To further elucidate the underlying mechanism of VI adjustments, we randomly selected 67 blocks pre-identified as mixed-use land within the study area, along with some blocks of other types nearby, totaling 197 blocks. Figure 14 illustrates the variations in the number of mixed-use areas of land identified by the vertical recognition model under different VI values.

**Figure 13.** Box plots of the four important parameters: (a) VI; (b) FAR; (c) Buildings Number; (d) HI.

**Figure 14.** Grouped bar chart of plot verification conditions in the sample area under different VI values.
This chart displays three data sets under various VI thresholds: the number of mixed-use land blocks within the sample area, the total number of blocks in the sample area, and the number of mixed-use lands identified by the model. The chart illustrates that the number of mixed-use lands recognized by the model decreases as the VI value increases, and significantly increases as the VI threshold decreases. Specifically, when the VI value is below 1.5, the number of mixed-use lands identified by the model exceeds the number determined by prior knowledge, which clearly contradicts the objective conditions of the real world, indicating that overly low VI values can lead to ‘contamination’ of blocks considered correctly identified. We have spatially visualized this type of ‘contamination’ using the Huaqiangbei area in Shenzhen and the Tianhe Pedestrian Commercial Circle in Guangzhou as examples, as shown in Figure 15.

![Figure 15](image)

**Figure 15.** Differences in identification results for different VI thresholds.

As can be seen from the chart, in both the Huaqiangbei area of Shenzhen and the Tianhe Pedestrian Commercial Circle in Guangzhou, a low VI results in a large number of plots being misclassified as mixed-use land. When the VI is adjusted to a reasonable range, the model effectively balances the extraction of mixed-use areas and the identification of other types of land. However, when the VI is too high, it results in a loss of accuracy in identifying mixed-use land.

5.1.2. The Impact of Building Number and HI on the Identification Results

The proposed horizontal morphology-based reclassification model in this study is based on a variance-weighted algorithm using the perimeter and area of buildings. It includes two important parameters: Building Numbers and HI. Figure 13c,d depicts the distribution of Building Numbers and HI for various land use types, highlighting their...
distinct characteristics. Residential land and industrial land have the highest building numbers (per block), but industrial land exhibits significant variation in HI, while residential land shows a smaller variation. This indicates that the HI can effectively differentiate between these two land use types. Transportation land and storage land have relatively fewer buildings, but they exhibit a larger variation in HI. This reflects the significant differences in scale, size, and design compared to other land use types. Special use areas have the fewest buildings, and their proportion is also minimal among all land use types, indicating a strong distinguishing factor.

Buildings number determines whether a block needs to be reclassified. Setting the $t_{hod\_num}$ too high may result in incorrectly identified blocks being excluded by the model, while setting it too low may include correctly identified blocks in the reclassification process. Regarding the $t_{hod\_HI}$, setting it too high may lead to the omission of land use types with high HI, while setting it too low may have the opposite effect.

Figure 13c indicates that the number of buildings on commercial, public management and service, and mixed-use land typically ranges from 1 to 30, with a median around 10. Considering real-world conditions, where residential areas generally have a higher count of buildings, setting a threshold of 10 for residential land can help filter out potential misclassifications as other types of land. Subsequently, the model uses the HI of the filtered blocks to determine the final results. For mixed-use land, the threshold is set at 300, based on a reasonable count of buildings within a plot, taking into account the significant numbers found in some schools and large residential complexes. Thus, if the number of buildings in a plot exceeds 300, it suggests a complex mix of functions, such as in urban villages, and will be reclassified as mixed-use land.

We have set certain presets for the HI thresholds. Initially, Table 3 shows that the average HI for residential land is the lowest, except for special use areas, which are not considered due to its high POI recognition accuracy and small sample size. Using the average HI of residential land as a reference, we explored the distribution of land use types within different HI intervals ranging from 0 to 700. We randomly selected 100 blocks samples for accuracy verification at intervals of 100 within the 100–700 threshold range, with the residential verification accuracies for different HI intervals presented in the following Figure 16. The sampling verification results indicate that setting the HI at 200 allows for correcting of more residential blocks with the most reliable accuracy.

Table 3. The mean HI of different land use types.

<table>
<thead>
<tr>
<th>LUC without Urban Morphology</th>
<th>HI (Mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage</td>
<td>3677.087</td>
</tr>
<tr>
<td>Industrial</td>
<td>1679.969</td>
</tr>
<tr>
<td>Public Management and Service</td>
<td>942.052</td>
</tr>
<tr>
<td>Mix-use Land</td>
<td>875.489</td>
</tr>
<tr>
<td>Transportation</td>
<td>1515.146</td>
</tr>
<tr>
<td>Residential</td>
<td>739.408</td>
</tr>
<tr>
<td>Commercial</td>
<td>1136.99</td>
</tr>
<tr>
<td>Special Use Area</td>
<td>502.06</td>
</tr>
</tbody>
</table>
Therefore, in this study, the $t_{holm1}$ was set to 300, the $t_{holm2}$ was set to 10, and the HI was set to 200. In practical applications, due to variations in urbanization levels, economic development, architectural styles, and forms across different study areas, the thresholds should be adjusted accordingly to achieve the best results.

5.2. Model Performance Changes in Central and Marginal Cities of Urban Agglomerations

The categorization of central and marginal cities is a concept that is comparatively relative, primarily hinging on the degree of urban economic development and administrative stature. Within the scope of this study, the Pearl River Delta urban agglomeration is a region with a distinctly hierarchical structure. Guangzhou, serving as the provincial capital and the administrative heart of Guangdong Province, in conjunction with Shenzhen, known for its status as an economic special zone and a trailblazer in openness and reforms, are the linchpins of this agglomeration. Both cities are recognized as part of the elite group of China’s four major first-tier cities, forming the core of the Pearl River Delta urban ensemble. Foshan and Dongguan, with their strong manufacturing bases, are considered second-tier cities in terms of economic scale within this cluster. The third tier includes other cities, which can be regarded as marginal in this context, such as Zhuhai, Zhongshan, Jiangmen, Zhaoqing, and Huizhou.

Existing literature does not delve into the discrepancies in land use identification accuracy between central and marginal cities. Theoretically, assuming uniformity in the application of the same identification model, the precision of land use identification within a region is predominantly swayed by its intrinsic characteristics rather than the influence of adjacent cities. Furthermore, marginal cities, with their relatively modest economic growth and smaller urban extents, ostensibly present a less complex identification challenge compared to their central counterparts. Contrarily, observations from identification outcomes reveal a markedly superior performance in central cities over marginal cities. Figures 17 and 18 shows the identification results of Shenzhen, while Figures 19 and 20 shows the identification results in Zhongshan.
Figure 17. Identification results of Shenzhen from EULUC-China.

Figure 18. Identification results of Shenzhen from UMH–LUC.
Figure 19. Identification results of Zhongshan from EULUC-China.

Figure 20. Identification results of Zhongshan from UMH–LUC.
In order to compare the accuracy difference between central cities and marginal cities, the 1000 samples extracted in the previous accuracy validation can be divided into three tiers based on the city level. The accuracy verification results are shown in Table 4.

Table 4. Precision comparison of different levels of cities.

<table>
<thead>
<tr>
<th>Type</th>
<th>UMH-LUC</th>
<th>LUC without Urban Morphology</th>
<th>EULUC-China</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central City</td>
<td>84.15%</td>
<td>73.33%</td>
<td>68.97%</td>
</tr>
<tr>
<td>Secondary City</td>
<td>76.92%</td>
<td>70.00%</td>
<td>56.52%</td>
</tr>
<tr>
<td>Marginal City</td>
<td>70.73%</td>
<td>62.50%</td>
<td>47.37%</td>
</tr>
</tbody>
</table>

The verification results indicate significant differences in accuracy between central cities and marginal cities. The reasons for this are as follows:

(1) Variations in Land Use Intensity: Central cities exhibit a heightened degree of economic development, which translates into more intensive and clearly delineated land use patterns. Conversely, marginal cities, characterized by comparatively lower economic growth, demonstrate more expansive and less organized land use, marked by reduced levels of intensity. The land in these marginal cities frequently displays a blend of functionalities.

(2) Discrepancies in Road Network Density: The identification units employed in this study hinge on city road networks, a methodology prevalent in land use identification research utilizing social sensing data. This approach is effective in central cities where road networks are notably dense. Nonetheless, it presupposes a sufficiently dense road network to facilitate practical subdivision into units. In marginal cities, characterized by less extensive urban development and smaller built-up areas, the principal road networks are sparser, leading to the creation of larger identification units. Employing roads of a smaller scale leads to two primary issues: firstly, the road network data becomes incomplete and, secondly, it results in a fragmented segmentation of the identification units, which does not accurately reflect the actual land use situation.

(3) Differences in Data Quality and Quantity: The richness of social sensing data, encompassing POI and building morphology information, is intrinsically tied to the degree of urban development. In more developed locales, these datasets are typically more robust and all-encompassing. As a result, central cities, with their heightened economic activities, possess ample and distinct social sensing data, facilitating accurate identification of predominant land use functions. Conversely, marginal cities, often a blend of urban and rural landscapes, exhibit a limited variety of POI data, predominantly linked to consumer activities. Furthermore, the availability and completeness of three-dimensional urban morphology data in these areas may be compromised, leading to gaps in current and relevant information.

Emphasis on functional recognition within marginal cities should be a pivotal focus in future land use identification studies. To augment the accuracy of identifying land uses in marginal cities, distinct datasets and indicator systems, tailored specifically for marginal and central cities, are warranted in large-scale analyses. For example, in marginal cities, the application of lower-tier road network data for delineating identification units might be more suitable. In regions where road networks are sparse and the array of diverse social sensing data is limited, further investigation is necessary to address data deficiencies. This includes exploring the integration of novel factors to refine and enhance the outcomes of identification processes.
6. Conclusions

This paper presents the UMH–LUC model for urban land use identification that incorporates multi-source social sensing data and urban morphology. The model follows a two-stage approach, with an initial preliminary identification using road networks, impervious surface identification, and POI data, followed by an optimization stage that incorporates vertical and horizontal urban morphology factors. The UMH–LUC model significantly improves the accuracy of land use identification by incorporating urban morphology features. Compared to the comparison methods, UMH–LUC demonstrates clear advantages in identifying residential land, mixed-use land and marginal areas. The model achieved an identification accuracy of 81.70% with a Kappa coefficient of 77.56, which is a 11.89% improvement compared to other methods. The quantity disagreement in UMH–LUC is 0.046, which is significantly lower than those observed in the other two methods, suggesting that this model’s classification results most closely project real-world proportions. The allocation disagreement is 0.137, which is similar to that of the EULUC-China method but lower than that observed in LUC without urban morphology. Overall, the overall disagreement for UMH–LUC is 0.183, a reduction of 0.099 compared to LUC without urban morphology and 0.19 compared to EULUC-China. By leveraging urban morphology, UMH–LUC provides more precise and robust identification results for different types of land use. The study also discovered that, compared to economically developed central cities, the identification of marginal cities poses greater challenges. Delving into how to overcome the limitations of current data and identification methods to enhance the precision of recognizing marginal cities is a subject worthy of in-depth exploration.

The constraints of the study include data dependency and quality, subjectivity in accuracy verification, and the time cost associated with parameter setting. Future research opportunities lie in exploring the inclusion and integration of a broader range of factors to enhance the accuracy of land use identification. Quantitative evaluation of factors’ contributions and dimension reduction analysis are also important for further investigation.

Overall, this paper contributes to the field of land use identification by incorporating urban morphology and social sensing data, leading to improved precision in distinguishing various land types. However, further research is needed to address the constraints and explore additional factors for more accurate depiction of urban land use.

Author Contributions: Conceptualization, G.L. and Y.L.; Data curation, C.L. and G.L.; Formal analysis, C.L. and G.L.; Funding acquisition, H.W. and Y.L.; Investigation, C.L. and G.L.; Methodology, C.L. and G.L.; Project administration, Y.L.; Resources, H.W. and Y.L.; Software, C.L. and G.L.; Supervision, Y.L.; Validation, Z.Z. and J.L.; Visualization, C.L.; Writing—original draft, C.L. and G.L.; Writing—review and editing, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Natural Science Foundation of China (Grants No. 42071356), Guangdong Provincial Department of Natural Resources Science and Technology Project (Grants No. GDZRZYKJ2024004), the Key Laboratory of Natural Resources Monitoring in Tropical and Subtropical Area of South China, Ministry of Natural Resources (Grants No. 2024NRMZ04), General Project of Guangdong Provincial Philosophy and Social Science Fund (Grants No. GD24CGL63), and Guangdong Geographical Science Data Center (Grants No. 2021B1212100003).

Data Availability Statement: Data are contained within the article.

Acknowledgments: We especially thank to the ‘Shared Cup’ Technology Resource Sharing and Service Innovation Competition organized by the National Science and Technology Infrastructure Center of China for providing data resource support for this study.

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


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.