FINE-GRAINED BUILDING ATTRIBUTES MAPPING BASED ON DEEP LEARNING AND A SATELLITE-TO-STREET VIEW MATCHING METHOD

Dairong Chen[*], *Jinhua Yu*[*], *Weijia Li*[†]

School of Geospatial Engineering and Science, Sun Yat-sen University,
Zhuhai, Guangdong, 519082, China

ABSTRACT

Street view images (SVIs) are a kind of data with rich semantic information, which have unique advantages in the fine-grained recognition of building land use. Compared with seamless dense remote sensing image data, SVIs are sparse and unevenly distributed in space, which brings many challenges to the application of SVIs for urban mapping. To solve this problem, this study proposes a satellite-to-street data matching method between SVIs and building footprint data. This method first performed dense sampling on the nodes of building footprint vectors, then designed a constraint based on the spatial relationship of cross-view data to match the buildings recognized in SVIs with their corresponding building footprints. Based on the matching results, large-scale building scale land use mapping was conducted in the validation area. The experimental results show that the accuracy of matching can reach more than 80%. The building land use classification in the mapping result reaches an accuracy of 62.15%, 56.41%, and 0.535 for overall accuracy, F1-score, and Kappa coefficient, respectively. This study provides a new technical means for fine-grained urban land use recognition and mapping, which can effectively improve the efficiency of acquiring fine-grained attribute information of urban buildings.

***Index Terms*—** street view images (SVIs), building, land use, instance segmentation

1. INTRODUCTION

Urbanization is a key feature of contemporary social development. It involves the expansion and transformation of urban land use, which affects the environment and society in various ways [1]. Therefore, it is essential to accurately and finely measure the quantity and quality of urban land resources to identify the problems and challenges of urbanization, and to find solutions for scientific urban planning and sustainable development [2]. One way to measure urban land resources is to analyze the attributes and spatial distribution of buildings, which are one of the symbols of modern cities. Buildings reflect the functions, characteristics, and dynamics of urban land use. Their attributes and spatial distribution provide valuable insights for large-scale city understanding [3]. Driven by the rapid development of information technology and the demands of smart city construction, large-scale precise building attribute recognition and mapping has become a current research challenge.

With accelerated urbanization, attributes of urban buildings have changed significantly over time [4]. Many building attribute data are missing or laborious to be updated in time. High-resolution remote sensing data perform well in building segmentation and detection tasks from a vertical perspective, but challenges still exist due to the physical property and the limited features that can be obtained from the remote sensing data. Consequently, most studies are still oriented towards coarse feature types [5, 6], such as building clusters, roads, green space, water bodies, etc. By contrast, street view images (SVIs) have the advantage of extracting building facade features, which shows great potential for achieving the task of identifying fine-grained building attributes [7]. However, a cross-view matching problem between the perspective of street and satellite arises when using orthophoto data (such as GIS data) and SVIs at the same time [5].

In this study, as shown in Figure 1, a novel framework is proposed to achieve building-level land use mapping, which supports SVIs from OmniCity [8] and building footprint data from OpenStreetMap (OSM) as input. The main contributions of this study are as follows: (1) Fine-grained land use attributes of individual buildings were extracted from the facades of SVIs using instance segmentation models; (2) A satellite-to-street matching algorithm was proposed to integrate the attributes extracted from SVIs and the building footprints with high accuracy and efficiency; (3) Large-scale mapping of building-level land use type was conducted for the validation area by the satellite-to-street matching algorithm.

2. METHODOLOGY

2.1. Building Instance Segmentation in SVIs

Instance segmentation algorithms are widely studied computer vision methods that aim to identify and classify objects

[*]Equal Contribution.
[†]Corresponding author.

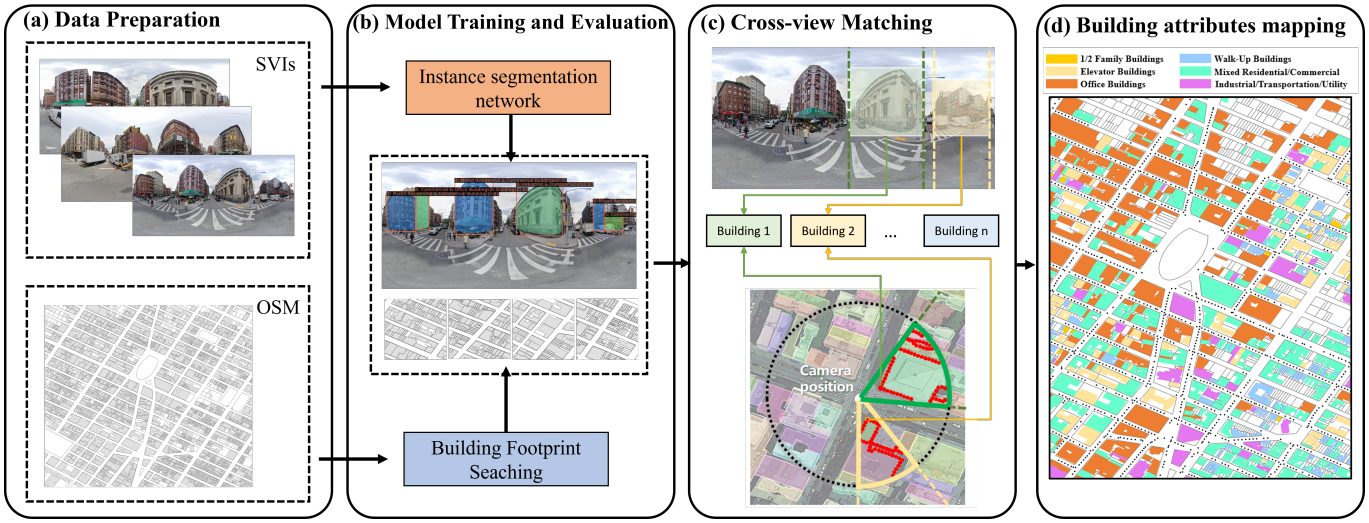

Fig. 1. The flowchart of the proposed framework, including data preparation, model training and evaluation, cross-view matching and building attributes mapping.

in images or videos by drawing bounding boxes and masks around them and giving the corresponding object classification results. This study employed Mask R-CNN [9] to extract the geometric boundary of buildings in SVIs and obtain building land use attribute classification results, which will be used in satellite-to-street matching and building land use attribute mapping.

2.2. Satellite-to-street Matching

This study proposes a satellite-to-street building matching algorithm that can accurately match the building attribute recognition results on SVIs with the building footprint vector data. This algorithm utilizes the geometric spatial relationships of cross perspective data to set spatial distance constraints for matching. The specific details are as follows.

At a given camera position $C_k = [lon_k, lat_k]$, we performed bilinear interpolation on the original nodes of the building footprint vectors for n times to obtain a denser set of nodes, which makes the representation of building footprint vectors more accurate and ensures the efficiency of the algorithm at the same time. Assuming that there are c buildings in the SVIs, where the i^{th} building footprint vector originally contains m nodes, after the multiple bilinear interpolations, the number of nodes in the building footprint vectors increased to $m_{dense} = m \times 2^n$. Thus, we generated a dense set of edge nodes for the i^{th} building footprint vector, as shown in equation (1).

$$M_i = [lon_i^j, lat_i^j], i \in [0, c), j \in [0, m \times 2^n) \quad (1)$$

The distances between each point in M_i and the camera position C_k were calculated using equation (2).

$$D_i = \sqrt{(M_i - C_k)^2} \quad (2)$$

We used the values in D_i as the distance constraint (J) and tried to minimize it to obtain the best matching node, as shown in equation (3).

$$J = argmin(D) \quad (3)$$

The attribute obtained from SVIs was assigned to the building footprint that contains the node with the lowest matching constraint (J).

After the above matching process, each building footprint was assigned with a unique land use attribute. Based on the building attributes and footprint vectors, the large-scale building land use mapping is conducted for the validation area.

3. EXPERIMENTS

3.1. Experimental Settings

Our study is based on an open-source dataset named OmniCity [8]. We select two sub-datasets (in panorama and mono views) that are collected from Google Street View Images. The SVIs in panorama view were used for the experiment, which contain 14,400 SVIs for training and 3,600 for validation, with a ratio of 4:1.

The experiments are mainly based on mmdetection [10] and PyTorch [11] framework, and the hyperparameter setting used by the OmniCity benchmark model [8] is adopted for our experiments. We use ResNet-50 and FPN pre-trained on ImageNet as the backbone for the instance segmentation models.

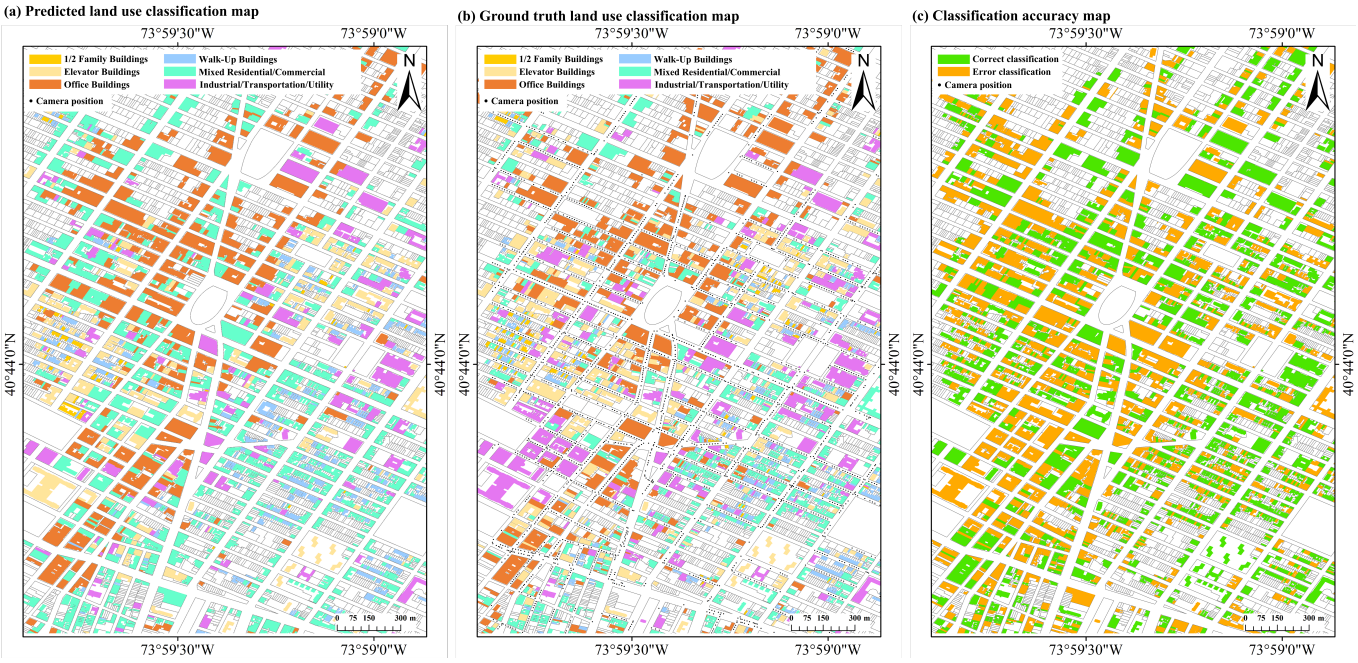

Fig. 2. Mapping results of building-scale fine-grained land use types (a) prediction map; (b) ground truth map; (c) classification accuracy map

The models were trained on NVIDIA RTX 2080 GPU for 12 epochs, with a batch size of 16, a learning rate starting from 0.02 and decreasing by a factor of 0.1 from the 8^{th} to 11^{th} epoch, and the stochastic gradient descent (SGD) optimizer with a weight decay of 10^4 and a momentum of 0.9.

3.2. Evaluation Metrics

The Satellite-to-street matching results are categorized into four types: correctly matched (CM), incorrectly matched (IM), unlabeled (UL), and undetected (UD). Three quantitative metrics are used to measure the matching results: accuracy rate, error rate, and missing rate. These metrics represent the fractions of correctly matched, incorrectly matched, and undetected buildings in the annotated data, as shown in equations (4), (5), and (6).

$$Accuracy\ rate = \frac{CM}{CM + IM + UD} \tag{4}$$

$$Error\ rate = \frac{IM}{CM + IM + UD} \tag{5}$$

$$Missed\ rate = \frac{UD}{CM + IM + UD} \tag{6}$$

We evaluate the building attribute classification in the mapping results using overall accuracy, F1-score and Kappa coefficient as overall metrics, and using precision, recall and F1-score as category metrics.

3.3. Satellite-to-street Matching Results

As shown in Table 2, the satellite-to-street matching achieved a matching accuracy of 82.30% and had the lowest matching error rate and miss rate of 16.32% and 1.38%. The experimental results demonstrate that nodes interpolation could guarantee the coverage of the target building footprint nodes by the detection field of view and improve the satellite-to-street matching results.

Table 1. Quantitative evaluation of satellite-to-street matching results.

Nodes interpolation	Accuracy	Error	Miss
No	77.78%	20.44%	1.79%
Yes	82.30%	16.32%	1.38%

3.4. Classification and Mapping of Building Land Use

After the satellite-to-street matching, each building in the validation set was assigned a unique land use attribute, and the results were quantitatively evaluated. The overall accuracy, F1-score and Kappa coefficient of the classification results were 62.15%, 56.41% and 0.535, respectively. The classification accuracy of each land use category is shown in Table 2, in which C1-C6 denote 1/2 Family Buildings, Walk-Up Buildings, Elevator Buildings, Mixed Residential/Commercial, Office Buildings, and Industrial/Transportation/Utility, respec-

tively. It was observed that C2, C4 and C5 have better performance compared with C1, C3 and C6 (with F1-score less than 0.5).

Table 2. Quantitative evaluation of the building land use classification results.

	Precision (%)	Recall (%)	F1-score (%)
C1	46.28	30.94	37.09
C2	60.66	67.72	64.00
C3	57.75	39.14	46.66
C4	64.48	81.79	72.11
C5	62.72	49.80	55.52
C6	64.16	35.58	45.78

Figure 2 shows the mapping results of building land use. The upper left part of the validation area corresponds to the business district of the city, where there are many office buildings, while the lower right part corresponds to the residential area, where there are more mixed-use, walk-up and elevator buildings. It is thus clear that the model prediction results can well present the differentiation of building attributes within the validation area, and show good consistency with the ground truth for most regions.

4. CONCLUSION

By leveraging street view images as a novel data source, this study employs instance segmentation methods to identify the fine-grained land use attributes of buildings from SVIs, and develops a satellite-to-street view matching method that fuses the building attributes and its corresponding footprint data with promising accuracy. This method provides a new technical means for fine-grained urban land resource recognition and mapping, and can effectively enhance the efficiency of obtaining fine-grained attribute information of urban buildings.

5. ACKNOWLEDGMENTS

This work was supported partially by National Natural Science Foundation of China (No. 42201358).

6. REFERENCES

[1] W. Feng, Y. Liu, and L. Qu, "Effect of land-centered urbanization on rural development: A regional analysis in china," *Land Use Policy*, vol. 87, p. 104072, 2019.

[2] X. Liu, Y. Huang, X. Xu, X. Li, X. Li, P. Ciais, P. Lin, K. Gong, A. D. Ziegler, A. Chen, *et al.*, "High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015," *Nature Sustainability*, vol. 3, no. 7, pp. 564–570, 2020.

[3] Z. Shao, N. S. Sumari, A. Portnov, F. Ujoh, W. Musakwa, and P. J. Mandela, "Urban sprawl and its impact on sustainable urban development: a combination of remote sensing and social media data," *Geospatial Information Science*, vol. 24, no. 2, pp. 241–255, 2021.

[4] X. Yuan and V. Sarma, "Automatic urban water-body detection and segmentation from sparse alsm data via spatially constrained model-driven clustering," *IEEE Geoscience and Remote Sensing Letters*, vol. 8, no. 1, pp. 73–77, 2010.

[5] F. Fang, Y. Yu, S. Li, Z. Zuo, Y. Liu, B. Wan, and Z. Luo, "Synthesizing location semantics from street view images to improve urban land-use classification," *International Journal of Geographical Information Science*, vol. 35, no. 9, pp. 1802–1825, 2021.

[6] F. Fang, L. Zeng, S. Li, D. Zheng, J. Zhang, Y. Liu, and B. Wan, "Spatial context-aware method for urban land use classification using street view images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 192, pp. 1–12, 2022.

[7] Y. Zhu, X. Deng, and S. Newsam, "Fine-grained land use classification at the city scale using ground-level images," *IEEE Transactions on Multimedia*, vol. 21, no. 7, pp. 1825–1838, 2019.

[8] W. Li, Y. Lai, L. Xu, Y. Xiangli, J. Yu, C. He, G.-S. Xia, and D. Lin, "Omnicity: Omnipotent city understanding with multi-level and multi-view images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17397–17407, 2023.

[9] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

[10] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, *et al.*, "Mmdetection: Open mmlab detection toolbox and benchmark," *arXiv preprint arXiv:1906.07155*, 2019.

[11] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.