Exploring Land Use Mapping with Multimodal Data Fusion and Convolutional Neural Network

Juan Gu [1,2], YongGuo Wang [1,2], Yun Gong [1,2], Tingting Jin [1,2], Jun Wu [1,2]

[1] Department of Natural Resources Investigating and Monitoring, Beijing Institute of Surveying and Mapping, Beijing, China;
[2] Beijing Key Laboratory of Urban Spatial Intelligent Sensing and Digital Governance, Beijing, China

Keywords: Time series, Land use classification, Fine-grained mapping, Multimodal data, Convolutional neural networks

Abstract

Accurate and efficient land-use mapping provides intuitive spatial information, which helps to rationalize the planning and deployment of land resources, and provides a basis for urban planning, agricultural development, environmental protection and other aspects. This study utilizes the Google Earth Engine platform and the Resnet-50 method to explore the spatial distribution of land use in Daxing District, Beijing in 2023, by combining point of interest (POI) data, nighttime light data, Sentinel-1 data, and Sentinel-2 data. The results of the study show that the accuracy of land use mapping using different data is different, and the accuracy of the Resnet-50 method is better than that of the Random Forest method. Making full use of the band features and index features of Sentienl-1 data and Sentinel-2 data, nighttime light data and POI data can improve the accuracy of land use mapping results. Among them, the land use mapping accuracy of the proposed method is the highest, with an OA of 88.11% and a Kappa coefficient of 0.83. Ranking the importance of different features found that VH band in January-March has the most important effect on the land use mapping results and the residential land in the POI data has the least important effect on the land use mapping results. This study provides a feasible reference program for efficiently and accurately obtaining land use mapping data for a large study area.

1.Introduction

Land use refers to the different ways of utilizing and developing land by human beings under certain natural and socio-economic conditions, based on the characteristics and needs of land resources [1]. It reflects the functional zoning, type of use and spatial layout of land resources, and involves a variety of forms, such as cultivated land, forested land, land for commercial services, industrial and mining land, residential land, land for public administration and public services, and land for transportation [2]. Rational planning of land use contributes to improving land use efficiency, achieving sustainable urban development, and effective urban planning [3].

A number of studies have been carried out to realize land use / land cover mapping to provide data support for exploring the spatial patterns and dynamics of the Earth surface information [4]. For example, mapping products with spatial resolution ranging from 100m to 1km (e.g., Global Land Cover Characterization (GLCC) data [5], Global Land Cover 2000 (GLC2000) dataset [6], etc.), spatial resolution ranging from 10m to 100m (e.g., GlobalLand30 [7], FROM-GLC30 [8], etc.), and spatial resolution less than 10m (e.g., 3 m land cover map of China [9], 1 m land cover map of China [10], etc.). However, most of the above studies focus on land cover type mapping, and research on fine land use mapping is limited. Unlike land cover data, land use contains socio-economic attributes that are more challenging to map [11]. Initially, land use mapping was obtained through on-site surveys, which was a time-consuming and labor-intensive method that did not allow rapid access to land use mapping results. With the development of remote sensing technology, remote sensing image data of different resolutions are widely used in land use classification research. These methods pay more attention to the physical attributes of the land cover itself, and do not perform well for land use mapping that includes social-economic attributes. With the advent of the geographic big data era, cell phone data, trajectory data, and other social sensing data have been used to infer the type of urban land use, but most of these data are confidential and cannot be used for large scale and fine-grained land use mapping. Open access point-of-interest (POI) data can provide potential socio-economic attributes, making it possible to accurately and quickly obtain urban land use mapping [12]. However, existing land use studies integrating multimodal remote sensing and social sensing data have mainly explored land use types in built-up areas, and relatively few studies have been conducted on large-scale all-land type areas [13]. Therefore, it is necessary to make full use of remote sensing data and social sensing data, while considering the physical attributes and socio-economic features of land cover, in order to realize the accurate and efficient acquisition of land use mapping studies for large-scale all-land type areas.

With the rapid development of deep learning technology, convolutional neural networks (CNN) have become more and more prominent in various data processing tasks, especially in the fields of image, time series, and signal processing. Compared with traditional machine learning methods such as Random Forest, CNN shows stronger feature extraction and pattern recognition capabilities. In addition, CNN is more efficient and flexible when dealing with large-scale data, and is able to utilize a large amount of training data to improve the generalization ability of the model, thus surpassing the performance of traditional methods in certain tasks.

This study takes Daxing District of Beijing as an example, and makes full use of the Google Earth Engine platform and multimodal data such as time series Sentinel-1 data, Sentinel-2 data, nighttime light data, POI data, combined with the Resnet-50 method, to explore the spatial distribution of land use in Daxing District of Beijing in 2023, as well as to explore the influence of different multimodal data on the accuracy and importance of land use mapping.

2.Material And Methods

2.1 Study area and datasets

Daxing District, between latitude 39°26'-39°50'N and longitude 116°13'-116°43'E, is located in the southern suburbs of Beijing, with a total area of 1,036.33 square kilometers. As of 2022, Daxing District has a resident population of 1,186,000, with 8 streets and 14 towns. Its land use types include: cultivated land, forest land, commercial and service land, industrial and mining land, residential land, public administration and public service land, and transportation land.

In this study, Sentinel-1 data and Sentinel-2 data for January-December 2023 were utilized, and seasonal synthesis images were performed for January-March, April-June, July-September, and October-December in order to efficiently avoid the effects of clouds and other factors. Among them, the Sentinel-1 data utilized the VV and VH bands, and the Sentinel-2 data utilized the B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12 bands. On this basis, nighttime light data from January-December 2023 and POI data from 2020 were combined to provide support data for land use mapping. The nighttime light data reflects the brightness of the ground light source, which can help us to speculate the economic activities, population distribution, urban expansion and other information in the region. POI data refers to the specific geographic location of a site or area, which represent specific points of interest related to human activities, and can help to refine the classification and update of the existing land use types. In this study, 3000 samples were manually selected for land use mapping with 70% of the samples and 30% of the samples were evaluated for accuracy.

2.2Method

Fig. 1 shows the process of land use mapping proposed in this study. First, Sentinel-2 data were used to calculate Normalized Difference Vegetation Index (NDVI), Normalized Difference Built-up Index (NDBI) and Normalized Difference Water Index (NDWI), which reflect the characteristics of different features, especially vegetation, water bodies and urban construction land. Second, the kernel density estimation method was utilized to calculate the POI data based on 2020 to analyze commercial and service land, industrial and mining land, residential land, public administration and public service land, and transportation land, and reveals the spatial distribution density of these land use types. Finally, the Resnet-50 method was used to characterize the multimodal data (Sentinel-1 data, Sentinel-2 data, nighttime light data, and POI data) for land use mapping, and to explore the influence and importance of different data on the accuracy of land use mapping.

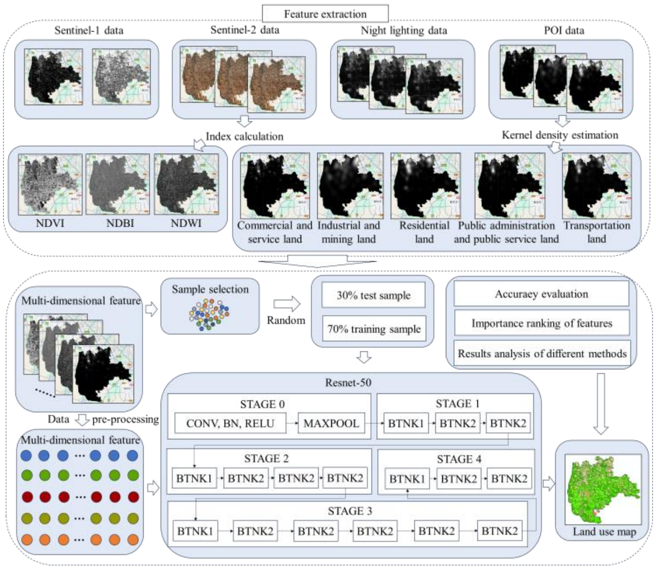

Fig. 1. Land use mapping framework proposed in this study

NDVI, NDBI and NDWI indexes were calculated using Sentinel-2 data [14]. Among them, NDVI is a remote sensing index that reflects the health of vegetation, NDBI is an index that identifies urban buildings and other human-made structures by analyzing the characteristics of surface reflectance, and NDWI is a remote sensing index that is used to extract water bodies, which is able to effectively differentiate between water bodies and other surface cover types (e.g., vegetation, bare soil, etc.). The specific equations are shown below.

$$NDVI = \frac{(NIR-Red)}{(NIR+Red)} \quad (1)$$

$$NDWI = \frac{(Green-NIR)}{(Green+NIR)} \quad (2)$$

$$NDBI = \frac{(MIR-NIR)}{(MIR+NIR)} \quad (3)$$

where NIR is the near-infrared band of Sentinel-2 images; Red is the red band of Sentinel-2 images; Green is the green band of Sentinel-2 images; MIR is the medium-infrared band of Sentinel-2 images.

Using the kernel density estimation method can effectively analyze the spatial distribution density of land uses based on POI data, especially in urban planning, land use classification and spatial analysis [15]. Through the kernel density estimation method, the spatial distribution characteristics of different types of land can be calculated. The kernel density estimation method is a nonparametric statistical method for estimating the density distribution of data points in space. In land use analysis, the kernel density estimation method can help determine the intensity of the distribution of a particular type of feature in a given area.

The Resnet-50 method is a deep residual network whose main innovation is the introduction of residual blocks, the core idea of which is to solve the problem of gradient vanishing or gradient explosion in deep neural networks by directly adding inputs to outputs through jump connections. Unlike traditional machine learning methods, the Resnet-50 method is able to automatically learn complex features of data, and especially excels on high-dimensional data such as images. It extracts features layer by layer through convolutional layers in deep neural networks, and is able to effectively capture features from low-level features (e.g. edges, textures) to high-level features (e.g. objects).

Feature importance ranking helps to understand which features are most critical to the model's predictions. The Random Forest method is a typical integrated learning method that consists of multiple decision trees that can be used to calculate the importance of features. Feature importance ranking helps to simplify the model, improve interpretability, and plays an important role in feature selection [16].

3.Results And Discussion

In order to prove the effectiveness of the proposed method and explore the influence of different data on mapping accuracy, land use mapping of Daxing District of Beijing was carried out using different data, and the corresponding accuracy evaluation was performed.

As shown in Table I, the results of land use mapping are not the same for different datasets. The OA of the S2-band-RF method is 81.67% and the Kappa coefficient is 0.74. The OA of the S2-band-index-RF method is 82.56% and the Kappa coefficient is 0.75. The OA of the S2-band-index-S1-RF method is 86.33 % and the Kappa coefficient is 0.81. The OA of the S2-band-index-S1-night-RF method is 86.44% and the Kappa coefficient is 0.81. The OA of the S2-band-index-S1-night-poi-RF method is 86.67% and the Kappa coefficient is 0.81. The OA of the S2-band-index-S1-night-poi-Resnet-50 method is 88.11% and the Kappa coefficient is 0.83. The accuracy of the S2-band-index-S1-night-poi-Resnet-50 method is significantly better than the other methods.

Table1 Accuracy Evaluation Results Of Different Methods

	Method					
	S2-band-RF [1]	***S2-band-index-RF*** [2]	***S2-band-index-S1-RF*** [3]	***S2-band-index-S1-night-RF*** [4]	***S2-band-index-S1-night-poi-RF*** [5]	***S2-band-index-S1-night-poi-Resnet-50*** [6]
OA/%	81.67	82.56	86.33	86.44	86.67	88.11
Kappa	0.74	0.75	0.81	0.81	0.81	0.83

Notes 1: The S2-band-RF method is the land use mapping obtained by using the band information of Sentinel-2 data combined with the Random Forest method.

Notes 2: The S2-band-index-RF method is a land use mapping obtained by utilizing the band information and index information of Sentinel-2 data combined with the Random Forest method.

Notes 3: The S2-band-index-S1-RF method is obtained by utilizing the band information and index information of Sentinel-2 data and the band information of Sentinel-1 data combined with the Random Forest method for land use mapping.

Notes 4: The S2-band-index-S1-night-RF method is a land-use mapping obtained by utilizing the band information, index information of the Sentinel-2 data, the band information of the Sentinel-1 data, and the nighttime light data combined with the Random Forest method.

Notes 5: The S2-band-index-S1-night-poi-RF method is a land use mapping obtained by utilizing the band information of Sentinel-2 data, the index information, the band information of Sentinel-1 data, the nighttime light data and the POI data combined with the Random Forest method.

Notes 6: The S2-band-index-S1-night-poi-Resnet-50 method is a land use mapping obtained by utilizing the band information of Sentinel-2 data, the index information, the band information of Sentinel-1 data, the nighttime light data and the POI data combined with the Resnet-50 method.

As shown in Fig. 2, there are subtle differences in land use mapping between the methods of the different data sources. The spatial distribution of land use in Daxing District of Beijing is basically the same. Commercial and service land is mainly concentrated in the northern part of Daxing District as well as in villages and towns, and forest land is mainly distributed in the peripheral areas of Daxing District and around commercial and service land. Cultivated land is mainly distributed around villages and towns. Residential land is mainly located in the vicinity of villages and towns and commercial and service land. The spatial distribution of public administration and public service land is relatively fragmented.

As shown in Table II, with the gradual integration of multi-source data and the upgrading of classification models, the classification accuracy for various land cover types shows an overall upward trend. Specifically, the improvement in accuracy for Residential land was the most significant; from the S2-band-RF to the ResNet-50 method, the PA increased from 85.12% to 94.32%, and the UA rose from 72.54% to 90.93%. In particular, the inclusion of S1 data led to a sharp rise in UA, indicating that radar data makes a significant contribution to the identification of residential land. Transportation land and commercial and service land are the two land cover classes with the lowest producer accuracy. Forest land and cultivated land exhibit some misclassification and confusion. The ResNet-50 method achieved the best or second-best accuracy performance across all land cover classes, validating the superiority of deep learning models in land use classification.

As shown in Fig. 3, different features have different degrees of influence on land use mapping. Among them, the VH band in January-March is the most important in the results of land use mapping, followed in order by the VV band in July-September, the NDBI in October-December, the NDBI in July-September, the NDVI in April-June, and so on. The residential land in the POI data are the least important in the results of land use mapping, followed in order by the October-December B11 band, January-March B4 band, January-March B7 band, April-June B11 band, and April-June B4 band.

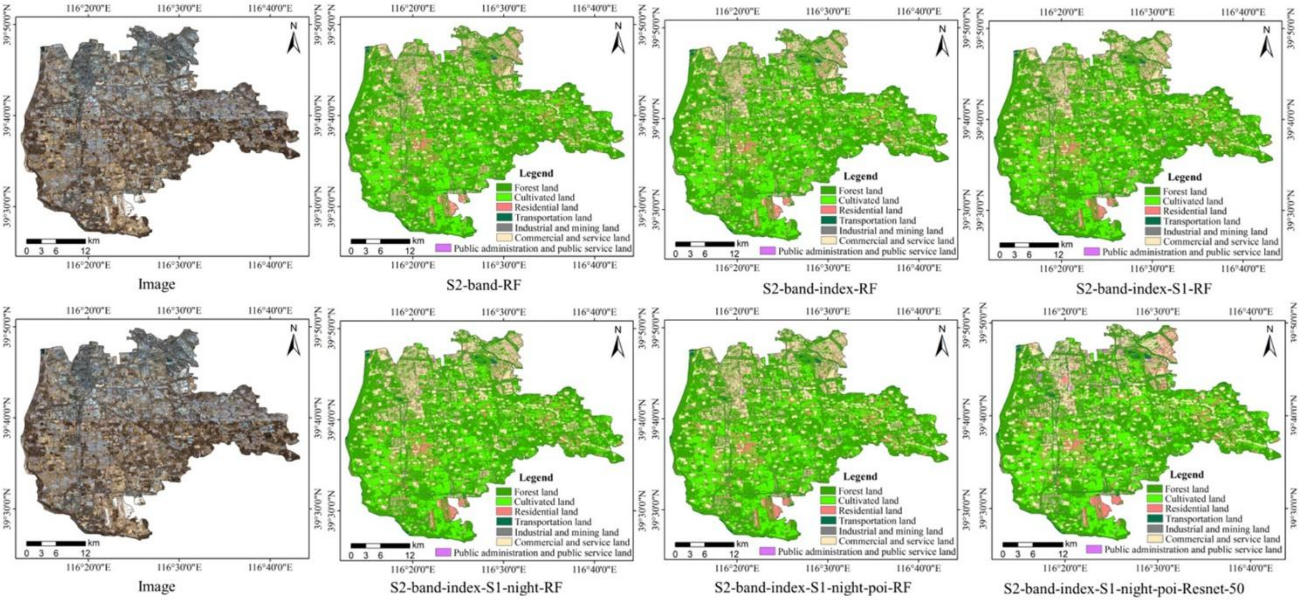

Fig. 2. Results of different methods of land-use mapping

TABLE I. CLASSIFICATION ACCURACY RESULTS FOR DIFFERENT METHODS

Method	Types	Forest land	Cultivated land	Residential land	Transportation land	Industrial and mining land	Commercial and service land	Public administration and public service land
S2-band-RF	PA/%	70.34	97.55	85.12	48.99	78.05	48.24	83.00
	UA/%	93.75	80.06	72.54	79.67	77.00	94.11	99.67
S2-band-index-RF	PA/%	71.89	97.14	87.18	44.21	81.43	59.00	82.33
	UA/%	94.35	81.22	69.80	83.62	80.33	94.45	99.00
S2-band-index-S1-RF	PA/%	72.11	96.20	91.55	64.12	95.65	68.43	80.33
	UA/%	94.69	80.86	89.86	91.23	80.91	97.00	97.00
S2-band-index-S1-night-RF	PA/%	72.56	96.73	94.27	64.12	94.30	74.14	80.33
	UA/%	95.84	80.37	90.04	93.08	88.14	97.00	97.00
S2-band-index-S1-night-poi-RF	PA/%	72.56	97.47	93.36	61.38	95.65	74.14	80.33
	UA/%	94.14	80.71	89.98	92.92	87.12	97.00	80.33
S2-band-index-S1-night-poi-Resnet-50	PA/%	73.45	97.58	94.32	62.29	96.71	75.11	81.33
	UA/%	95.17	81.83	90.93	93.89	88.06	98.00	81.33

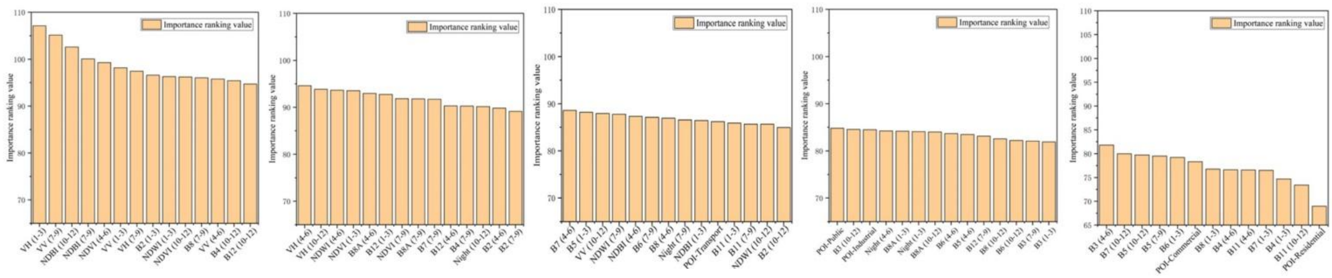

Fig. 3. Importance ranking of different features

4.Discussion

4.1 The impact of different data combinations on classification accuracy

By comparing the classification performance of various data combinations, this study demonstrates the effectiveness of multi-source data fusion in land-use mapping. The experimental results indicate that the S2-band-RF method performs relatively robustly as a baseline model. With the introduction of additional features, the overall accuracy of the S2-band-index-RF method improves, suggesting that spectral indices can effectively enhance the distinguishability of land-use classes. Following the further integration of S1 radar data, the overall accuracy of the S2-band-index-S1-RF method improved significantly, indicating that radar data plays an important complementary role to optical remote sensing data. Radar data can penetrate cloud cover and provide scattering information sensitive to surface structure and moisture, offering unique advantages in areas where optical data is limited or where land cover structural features are pronounced. The gradual integration of night-time lighting (NTL) and POI data did not yield the expected sustained improvement. The overall accuracy of the S2-band-index-S1-night-RF method and the S2-band-index-S1-night-poi-RF method showed only a slight improvement compared to the previous combination. This phenomenon may stem from uncertainties in the spatial resolution, update frequency, and correspondence with land use types of the night-time light and POI data; their contribution is, to some extent, overshadowed by the advantages of the remote sensing data. It is worth noting that with the introduction of deep learning methods, the S2-band-index-S1-night-poi-ResNet-50 method achieved the highest overall accuracy. ResNet-50, through its residual architecture, is able to effectively extract deep spatial features, outperforming traditional random forest methods that rely on manually constructed features, thereby confirming the potential of deep learning in land use classification tasks.

From the use of the S2 band alone to the gradual integration of spectral indices, radar data, night-time lighting and POI data, classification accuracy showed an overall upward trend, validating the rationale behind the use of multi-source data. The ResNet-50 method significantly outperformed the Random Forest method, consistent with recent research findings in the field of remote sensing image classification and in line with expectations regarding the feature extraction capabilities of deep learning. The significant improvement in accuracy achieved by radar data when combined with optical remote sensing aligns with the expected sensitivity of radar to surface structure and moisture. Furthermore, when building upon existing S2 and S1 data, the accuracy improvements derived from night-time lighting and POI data were modest and lower than anticipated. This may be related to the spatial heterogeneity of the urban-rural transition zone in Daxing District, where significant differences in the density of night-time lighting and POIs between urban and rural areas make it difficult for the model to uniformly learn effective features. Residential land in the POI data had the least impact on the classification results, which was somewhat unexpected. A possible reason is that residential land exhibits spectral similarity to certain commercial, service, or industrial and mining land types in remote sensing imagery, and the semantic features of POIs struggle to effectively distinguish residential land from other mixed-use land types.

4.2 Analysis of study limitations

The multi-source data fusion framework proposed in this study is methodologically universal; provided that POI data sources exist within the study area, this method can be applied. When applied in different countries, the method requires adaptation based on the locally available POI data sources. Although this study achieved satisfactory classification results, the following limitations remain: Firstly, the spatial resolution of Sentinel-2 and Sentinel-1 still suffers from the problem of mixed pixels when identifying land use at a fine scale, making it difficult to accurately delineate the boundaries of small features and linear features. Secondly, the spatial resolution of night-time light data is relatively low, and the presence of light spillover effects makes it difficult to precisely match feature boundaries, which may introduce noise. Furthermore, POI data reflect facility types rather than actual land-use characteristics. A single facility may correspond to multiple land-use categories (for example, a catering-related POI may be located on commercial, residential or mixed-use land), thereby leading to semantic ambiguity. There is also a discrepancy between the timing of POI data and that of remote sensing imagery. Finally, although multi-temporal data have been utilised, the absence of data covering critical periods of phenological change may affect the accuracy of identifying seasonal features such as crops.

5.CONCLUSIONS

This study utilizes the Google Earth Engine platform and the Resnet-50 method, combined with multimodal data, such as Sentienl-1 data, Sentinel-2 data, nighttime light data and POI data, to perform land use mapping in Daxing District, Beijing, in the year 2023, and to analyze the situation of the importance of different features to land use mapping. In land use mapping, the accuracy of land use mapping increases with the increase of data sources. The addition of band and index information of Sentinel-1 data and Sentinel-2 data, nighttime light data and POI data can improve the accuracy of land use mapping. The accuracy of the Resnet-50 method is significantly better than that of the Random Forest method. The land use mapping accuracy of the proposed method is the highest with OA of 88.11% and Kappa coefficient of 0.83. The VH band in January-March has the most important effect on the land use mapping results, and the residential land in the POI data has the least important effect on the land use mapping results. This study provides a feasible reference basis for accurately and quickly obtaining land use mapping data for a large study area.

Acknowledgment

The authors would like to thank the editors and the anonymous reviewers for their constructive comments and suggestions, which greatly helped to improve the quality of the manuscript.

References

[1] F. Zheng, C. Xiao, and Z. Feng, “Impact of armed conflict on land use and land cover changes in global border areas,” *L. Degrad. Dev.*, vol. 34, no. 3, pp. 873–884, Feb. 2023.

[2] C. Zhang *et al.*, “Spatial-temporal characteristics of carbon emissions from land use change in Yellow River Delta region , China,” *Ecol. Indic.*, vol. 136, p. 108623, 2022.

[3] H. Xie, Y. He, J. Zou, and Q. Wu, “Spatio-temporal difference analysis of cultivated land use intensity based on emergy in the Poyang Lake Eco-economic Zone of China,” *J. Geogr. Sci.*, vol. 26, no. 10, pp. 1412–1430, 2016.

[4] Z. Li, B. Chen, S. Wu, M. Su, J. M. Chen, and B. Xu, “Deep learning for urban land use category classification: A review and

experimental assessment,” *Remote Sens. Environ.*, vol. 311, no. July, 2024.

[5] T. R. Loveland *et al.*, “Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data,” *Int. J. Remote Sens.*, vol. 21, no. 6, pp. 1303–1330, 2000.

[6] X. Liu *et al.*, “High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform,” *Remote Sens. Environ.*, vol. 209, no. February, pp. 227–239, 2018.

[7] J. Chen, X. Cao, S. Peng, and H. Ren, “Analysis and applications of GlobeLand30: A review,” *ISPRS Int. J. Geo-Information*, vol. 6, no. 8, pp. 230–247, 2017.

[8] H. Liu, P. Gong, J. Wang, X. Wang, G. Ning, and B. Xu, “Production of global daily seamless data cubes and quantification of global land cover change from 1985 to 2020 - iMap World 1.0,” *Remote Sens. Environ.*, vol. 258, p. 112364, 2021.

[9] R. Dong, W. Fang, H. Fu, L. Gan, J. Wang, and P. Gong, “High-Resolution Land Cover Mapping Through Learning With Noise Correction,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022.

[10] Z. Li, W. He, M. Cheng, J. Hu, G. Yang, and H. Zhang, “SinoLC-1: the first 1 m resolution national-scale land-cover map of China created with a deep learning framework and open-access data,” *Earth Syst. Sci. Data*, vol. 15, no. 11, pp. 4749–4780, Oct. 2023.

[11] Z. Lina and L. Lixun, “Spatial distribution and format difference of large-scale retail business facilities: a case study of Guangzhou based on POI data,” *Trop. Grography*, vol. 39, no. 5, pp. 88–100, 2019.

[12] Y. Su, Y. Zhong, Y. Liu, and Z. Zheng, “A graph-based framework to integrate semantic object/land-use relationships for urban land-use mapping with case studies of Chinese cities,” *Int. J. Geogr. Inf. Sci.*, vol. 37, no. 7, pp. 1582–1614, 2023.

[13] S. Du *et al.*, “Mapping urban functional zones with remote sensing and geospatial big data: a systematic review,” *GIScience Remote Sens.*, vol. 61, no. 1, p. 2404900, Dec. 2024.

[14] L. Zhu, Z. Guo, H. Xing, and W. Sun, “A coupled temporal-spectral-spatial multidimensional information change detection framework method: a case of the 1990-2020 Tianjin, China,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 16, pp. 5741–5758, 2023.

[15] X. yi Guo, H. yan Zhang, Y. qiao Wang, J. jun Zhao, and Z. xiang Zhang, “The driving factors and their interactions of fire occurrence in Greater Khingan Mountains, China,” *J. Mt. Sci.*, vol. 17, no. 11, pp. 2674–2690, 2020.

[16] J. Hua, G. Chen, L. Yu, Q. Ye, H. Jiao, and X. Luo, “Improved Mapping of Long-Term Forest Disturbance and Recovery Dynamics in the Subtropical China Using All Available Landsat Time-Series Imagery on Google Earth Engine Platform,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 2754–2768, 2021.