Bi-branch Neural Network for Urban Functional Zone Mapping: Combining Remote Sensing Imagery and Point-of-Interest Data

Liangzhong Ying[2], Xiana Chen[1,3,4,5*], Juejing Zhao[6], Yuzhou Zhang[2], Hua Sun[2], Wei Tu[1,3,4,5]

1 Guangdong Key Laboratory of Urban Informatics, Shenzhen University, Shenzhen, Guangdong 518060, China - chenxiana2019@email.szu.edu.cn, tuwei@szu.edu.cn
2 Popsmart Technology(Zhejiang)Co., Ltd, Ningbo, 31500, China - yingliangzhong@popsmart.cn, zhangyuzhou@popsmart.cn, sunhua@popsmart.cn
3 Key Laboratory for Geo-Environmental Monitoring of Great Bay Area, Ministry of Natural Resources, Shenzhen University, Shenzhen, Guangdong 518060, China
4 Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen, Guangdong 518060, China
5 School of Architecture and Urban Planning, Shenzhen University, Shenzhen, Guangdong 518060, China
6 Ningbo Synsense Technology Co., Ltd., Ningbo, Zhejiang 315000, China - juejing.zhao@synsense.ai

Keywords: Urban functional zone, neural network, multi-source data fusion

Abstract

Urban functional zone (UFZ) classification is essential for understanding city dynamics, supporting urban planning, and enabling effective resource allocation. Traditional approaches rely heavily on remote sensing imagery, which often lacks the contextual information necessary for distinguishing between zones with similar visual features but different functions. This study proposes a novel multi-modal bi-branch deep learning model, named BibDL, which integrates remote sensing imagery with Point-of-Interest (POI) data for UFZ classification. The BibDL model leverages the complementary strengths of these data sources: remote sensing provides spatial and structural information, while POI data offers insights into human activities and land use patterns. Experimental results demonstrate that the BibDL model significantly outperforms a baseline model trained only on imagery, achieving higher F1 scores of 0.975 and Kappa coefficients of 0.953 across multiple UFZ categories. In particular, the BibDL model shows improved performance in challenging zones such as Commerce, Public, and Academia, which are often misclassified when using imagery alone. An ablation study highlights the substantial accuracy gains achieved by incorporating POI data, underscoring the value of a multi-modal approach for UFZ classification. The findings suggest that combining remote sensing imagery with contextual POI density image offers a powerful framework for more precise, context-aware UFZ classification, with implications for urban planning, smart city development, and sustainable resource management.

[*] Corresponding author

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-1647-2025 | © Author(s) 2025. CC BY 4.0 License.

1. Introduction

Rapid urbanization and city expansion have led to a range of urban issues, including traffic congestion and housing shortages. Urban functional zones (UFZs), reflecting significant urban social functions and economic activities, are extensively used as fundamental spatial units to analyze the spatial and social structures of urban environments (Tu et al., 2018; Tu et al., 2024). An optimal UFZ spatial layout helps mitigate urbanization challenges. Identifying UFZs and understanding their spatial distribution and interaction patterns are of great significance in supporting scientific planning (Cao et al., 2025).

With the rapid development of remote sensing (RS) high-resolution images have gradually shown potential in the task of UFZ recognition. RS images provide rich physical detail, capturing essential characteristics such as spectral signatures, textures, shapes, and angles, which are invaluable for identifying landscape compositions and urban morphological structures. For example, leveraging state-of-the-art transformer architecture, Wang et al. (2022) developed a U-shaped transformer network to analyze high-resolution urban scene imagery, achieving promising results across four challenging datasets. Du et al. (2021) utilized RS images to map large-scale, fine-grained UFZs by applying a multi-scale semantic segmentation network combined with an object-oriented approach, achieving significant improvements in classification results. These studies demonstrate the efficacy of RS imagery in generating detailed urban classification maps, as confirmed by various applications (Li et al., 2016). However, while RS images excel at capturing static physical characteristics such as building layouts and urban spatial structures, they often lack the capability to reflect dynamic changes and human-centric activities. This limitation becomes particularly evident when distinguishing between UFZs that, despite being visually similar, exhibit distinct functional attributes tied to human activities. Consequently, relying exclusively on RS imagery for UFZ classification presents notable challenges, especially in complex urban environments where subtle differences in human interactions and socio-economic functions are critical (Du et al., 2021; Lu et al., 2022).

Social sensing and human activities are increasingly recognized as effective methods for dynamically identifying urban areas. Researchers have leveraged social sensing data, such as Point-of-Interest (POIs), check-ins, and GPS trajectories, to delineate functional areas within cities, yielding promising results (Xing et al., 2024). Unlike RS images, which captures the physical characteristics of urban environments, social sensing data stem from human activities and often carry temporal dimensions, offering a rich tapestry of socioeconomic insights (Du et al., 2024). For instance, using LJ1-01 NTL remote sensing satellite data and mobile big data, Zhou et al. extracted UFZs at the street-level scale (2019). Xu et al. (2022) proposed a Multi-dimension Feature Learning (MDFL) model that integrates high-dimensional geospatial big data with RS images for urban region function recognition, demonstrating the potential of combining these modalities. Among the various forms of social sensing data, POIs stand out as the most significant static data source. They are not only easy to access but also provide detailed land-use information, reflecting human activities and their geographic distribution (Lu et al., 2022; S. Xu et al., 2020). Recent advancements, such as the work by Yu et al. (2023), underscore the value of unified deep learning frameworks that seamlessly combine visual features from RS imagery with social features from POI data. These frameworks have proven effective in enhancing UFZ classification accuracy. This integration approach emphasizes the importance of a holistic understanding of urban dynamics, one that captures both the physical structure and the socio-functional fabric of urban environments.

Despite these advancements, existing studies often fail to fully integrate multimodal semantics for classification tasks, resulting in complex multistage processes that can compromise the quality and reliability of UFZ mapping. To fill these gaps, this study further proposed a bi-branch different deep neural network (DNN), namely, the bi-branch deep learning (BibDL) model, which utilizes two different neural network branches to comprehensively learn features of RS images and POI data and then fuses these features to map the UFZ more accurately.

2. Study Area and Data

2.1 Study Area

The research was carried out in Shenzhen, Guangdong Province, China, as shown in Figure 1. By the end of 2023, the city has ten districts under its jurisdiction, with a total area of 1997.47 square kilometers. With a resident population of 17,790,100, it is the first fully urbanized city in China. To achieve sustainable development, Shenzhen must rationally plan the functional zoning of the city. This study involves nine types of functional zones: residence, industry, commerce, public, academia, road, water, green land, and farmland. Their definitions are detailed in Table 1.

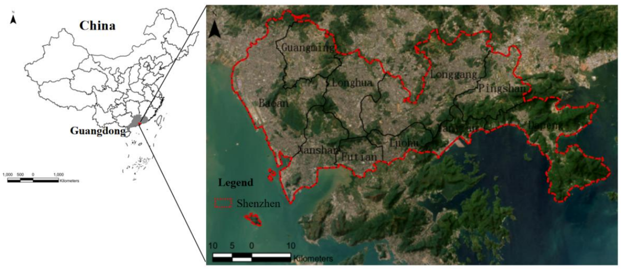

Figure 1. The study area in China.

2.2 Study Data

High-resolution RS images and POIs are involved in the study. We obtained high-resolution RS imagery from Google Earth for the year 2019 which has three bands and a spatial resolution of 1.0 m. The POI data obtained from the year 2023 from Baidu Map contained 19 categories: Road auxiliary facilities, food service, car repair service, motorcycle service, company and enterprise, shopping service, education and culture service, sports and leisure services, transportation facilities services, financial insurance services, accommodation services, scenic attractions, car sales, government agencies and social organizations, life services, healthcare services, commercial housing, public facilities, and car service. To a certain extent, it reflects the types of activities people perform at specific places.

Types	Definitions
Residence	Areas designated for housing and living purposes, including apartments, single-family homes, and other types of dwellings where people reside.
Industry	Zones are primarily allocated for manufacturing, production, and industrial operations, including factories, warehouses, and other facilities associated with large-scale

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-1647-2025 | © Author(s) 2025. CC BY 4.0 License.

	industrial activities.
Commerce	Businesses and services operate, such as retail stores, shopping malls, restaurants, and offices, serving the economic and consumer needs of the population.
Public	Areas dedicated to public infrastructure and services, such as government buildings, hospitals, cultural institutions, and facilities providing essential services to the general public.
Academia	Zones where educational institutions are concentrated, including schools, universities, research centers, and other facilities dedicated to academic and educational activities.
Traffic	Transport infrastructure including streets, highways, airports, etc.
Water	Bodies of water within the urban landscape, including rivers, lakes, reservoirs, etc.
Green land	Urban green spaces, such as parks, gardens, and recreational areas, where vegetation is dominant.
Farmland	Agricultural zones on the outskirts or within the city where crops are grown and livestock is raised.

Table 1. Definition of the nine types of functional regions.

3. Methodology

An innovative data fusion framework is proposed, integrating RS imagery and POI data to improve the representation and classification of UFZs. The workflow, illustrated in Figure 2, introduces a specialized division of labor between two neural network branches: the RS-branch and the POI-branch, each designed to exploit the unique strengths of its respective data source. The process begins by generating a regular grid overlaid with land use maps to derive UFZ labels. RS imagery is partitioned into grid cells, capturing the spatial and spectral characteristics of urban landscapes, while POI data undergoes kernel density estimation, producing density maps that reveal the intensity and spatial distribution of human activities and urban services. These distinct data sources are fed into separate branches of the network. The RS-branch extracts high-level visual features such as texture, shape, and spectral patterns from raw imagery, which are vital for identifying physical structures and land cover. Simultaneously, the POI-branch focuses on extracting social and functional attributes, leveraging POI data to capture the socio-economic dynamics and functional diversity of urban environments. This dual-branch architecture is a core innovation, allowing each branch to focus on complementary dimensions of UFZ classification. The outputs of the two branches are seamlessly fused in the network's final layers, producing a unified probability distribution for UFZ labels by synthesizing both visual and functional information. This fusion strategy enables the model to effectively capture the intricate interplay between the physical form of urban spaces and their associated social functions. The details of this integrated approach are discussed in depth in the following subsections.

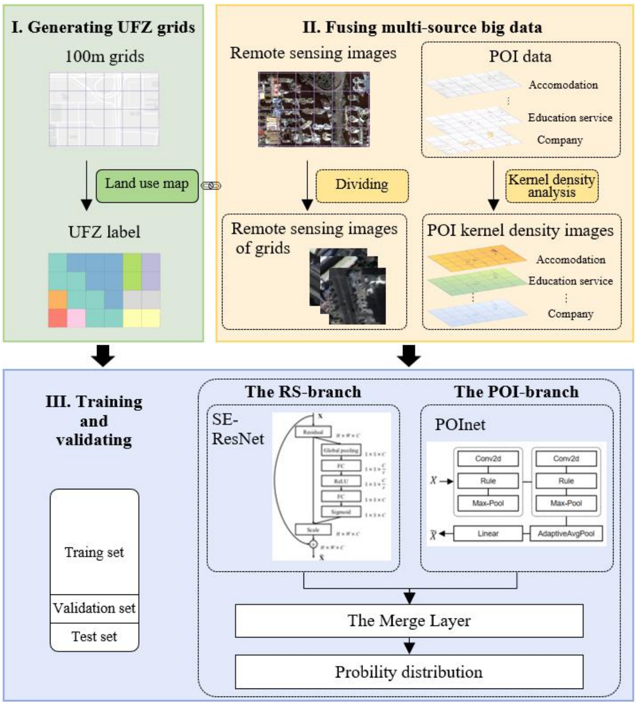

Figure 2. Overview of methodology.

3.1 Generating UFZ Grids

A grid-based approach is adopted to delineate UFZs at a 100m x 100m spatial resolution. This resolution balances detailed analysis with computational efficiency, enabling effective mapping of urban functionalities across diverse landscapes. Each grid cell is assigned a dominant UFZ label based on the predominant land use type within its boundaries, as determined from the land use map. To ensure compatibility, both RS images and POI data are preprocessed to align with this 100m x 100m grid resolution.

3.2 RS-branch

The RS-branch leverages a pre-trained SE-ResNeXt101-32x4d network to extract high-level features from high-resolution RS images of 100m grids. This network architecture incorporates the benefits of ResNeXt, which employs grouped convolutions and a bottleneck design to efficiently capture multi-scale information with fewer parameters. To enhance feature representation, the Squeeze-and-Excitation (SE) module is integrated into the ResNeXt101-32x4d model. This module adaptively assigns weights to different feature channels, emphasizing the most discriminative spatial information in RS images (Hu et al., 2018).

The RS image is initially processed by the initial convolutional layers of the pre-trained network to extract low-level features, such as edges, textures, and shapes. As the network deepens, subsequent layers extract increasingly complex features, including building structures, road networks, and vegetation cover. To adapt the network to the specific task of remote sensing image analysis, the original 1000-class classification layer is replaced with a custom fully connected layer. This layer reduces the feature dimensionality to 256 dimensions. Adaptive Average Pooling is then applied to downsample the feature maps, ensuring a fixed-size input for the fully connected layer.

3.3 POI-branch

3.3.1 POI Kernel Density Generation: POI categories can be viewed as virtual words that reflect socioeconomic properties. Therefore, the number and distribution of POIs in each UFZ indicate the land use patterns and socioeconomic functions. Specifically, we first convert them into the corresponding 2m resolution kernel density heatmap according to the number of POI categories and input it into the convolutional neural network, as CNN is better at processing two-dimensional continuous image data.

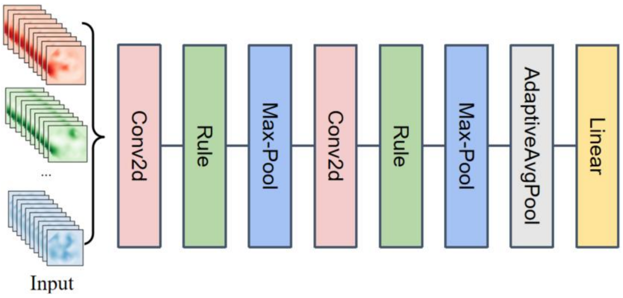

Figure 3. Architecture of the proposed POI-branch.

3.3.2 POI-branch Model: To effectively handle multiple POI categories, our model stacks 19 normalized POI density images into a multi-channel input tensor. Each channel corresponds to a specific POI type, representing its spatial distribution. The input POI data is structured as a 3D tensor of shape [H, W, C], where H and W are the height and width of the image (224 x 224 pixels), and C is the number of POI density channels (19).

The POI-branch shown in Figure 3 is designed to extract category-specific spatial distribution features from each channel and capture inter-category interactions through subsequent convolutional operations. The POI density images are first processed by a two-layer convolutional neural network. The initial convolutional layer takes the 19-channel POI density map as input and applies a convolution with 64, 3x3 kernels, producing a 64-channel feature map. The ReLU activation function is used to introduce non-linearity, and max pooling is applied to downsample the feature maps, reducing computational complexity and increasing the receptive field of the convolutional features.

The pooled feature maps are then fed into a second convolutional layer with 128 kernels, generating a higher-dimensional feature representation. Similar to the RS-branch, the output feature maps from the conventional layers are subjected to adaptive average pooling to reduce them to a fixed-size output. Finally, a fully connected layer is used to further compress the pooled feature vectors into a 256-dimensional representation, capturing the most salient characteristics of the POI density data.

3.4 Fusion Layer

The fusion layer integrates the output vectors from both the RS-branch and POI-branch, producing the predicted UZF category. Specifically, the 256-dimensional feature vectors extracted from the RS images and POI kernel density images are concatenated into a single 512-dimensional feature vector. To enhance the representational power of the fused features, we introduce a weighted fusion strategy. The weight parameters are learned through backpropagation, allowing the model to dynamically adjust the importance of RS image features and POI density map features based on their contributions during training. The resulting 512-dimensional feature vector is then fed into a fully connected layer to reduce its dimensionality to 128. This compressed representation is subsequently passed to a final fully connected layer, which acts as a classifier. The classifier outputs a probability distribution over the nine UZF categories, indicating the model's confidence in each class for the given input. The predicted UZF category corresponds to the class with the highest probability.

3.5 Evaluation Metric

We used the precision to measure the accuracy of each category, the F1 score to evaluate the overall categorization accuracy, and the Kappa coefficient (Kappa) to evaluate the overall performance of each model, as shown in the following equation:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{1}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{F1} = \frac{2 * \text{precision} * \text{recall}}{(\text{precision} + \text{recall})} \tag{3}$$

$$p_e = \frac{a_1 * b_1 + a_2 * b_2 + \ldots + a_c * b_c}{n * n} \tag{4}$$

$$Kappa = \frac{p_o - p_e}{1 - p_e} \tag{5}$$

where TP = number of pixels correctly categorized into positive categories

TN = the number of pixels correctly categorized into negative categories

FP = the number of pixels incorrectly categorized into positive categories

FN = the number of pixels incorrectly categorized into negative categories

p_o = the overall classification accuracy

a_1, a_2, ... a_c = the number of true samples in each category

b_1, b_2, ... b_c = the number of predicted samples in each category

n = the total number of samples.

4. Results and Analysis

4.1 Results of BibDL Model for UFZ Classification

The results of the presented BibDL model are shown in Table 2. The BibDL model has been demonstrated high effectiveness in classifying several UFZ categories, with an overall model F1 score of 0.975 and a kappa value of 0.953. The performance metrics, precision, recall, and F1-score, were calculated for each of the nine functional zones to assess the model's accuracy, as shown in Figure 4. The BibDL model exhibits high accuracy in classifying most UFZs, particularly in Residential, Industrial, Water, Traffic, and Green land, with an F1-score over 0.9, respectively, indicating near-perfect classification accuracy. The model's performance for commercial and public areas is relatively balanced but lower compared to other zones, indicating room for improvement.

Model	F1	Kappa
BibDL	0.975	0.953
RS-branch	0.927	0.867
U-Net DL (Yang et al., 2021)	0.857	0.747
AlexNet (W. Zhou et al., 2020)	0.786	0.623

Table 2. The results of models.

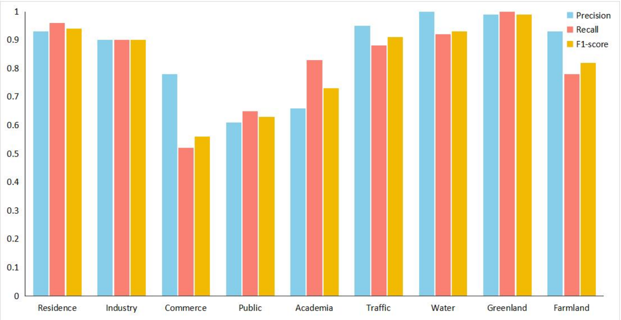

Figure 4. Precision, recall, and F1-score for each category.

4.2 Urban Functional Zone Identification Results Based on BibDL

To evaluate the fine-grained UFZ mapping capabilities of BibDL, detailed UFZ maps for four areas in Shenzhen are presented in Figure 5. These areas represent typical urban development types, including urban center, sub-center, transit region, and suburbs, each with distinct functional characteristics and spatial distribution.

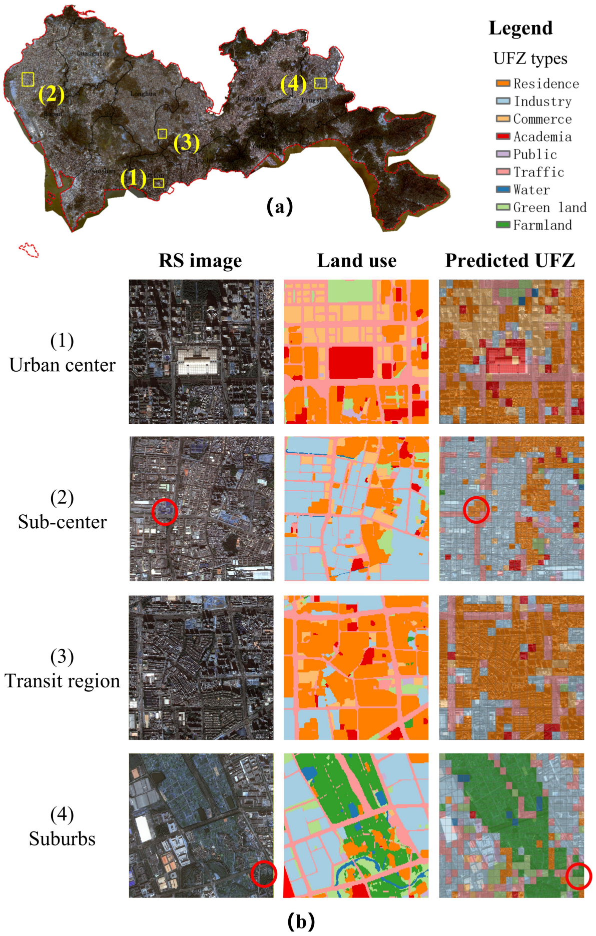

Figure 5. Representative urban blocks. (a) Locations of urban blocks; (b) RS images, land use of 1m resolution, and predicted UFZ.

Figure 5(1) shows the urban center, a predominantly commercial area. However, some grids were misclassified as Industry due to their industrial-like appearance in RS imagery and slightly sparse POI data. This highlights the model's reliance on POI data to accurately classify areas with ambiguous visual features. Figure 5(2) depicts a sub-center transitioning from industrial to high-tech and residential. Despite similar RS imagery to industrial areas, the presence of dense residential and service POI enabled accurate classification as highlighted in the red block. This demonstrates the importance of POI data in disambiguating areas with similar visual characteristics. Figure 5(3) focuses on a transit region with a mix of industrial and residential uses and ongoing construction. Misclassifications of traffic zones as Industry occurred due to the industrial-like appearance of the region and sparse traffic-related POI. Figure 5(4) showcases the suburbs, characterized by industrial parks, green spaces, and farmland. The model accurately classified these land uses, even distinguishing between Green land and Farmland, which can be challenging based solely on RS imagery. In summary, the BibDL model demonstrates robust UFZ classification capabilities across diverse urban development types. However, its performance varies depending on the availability and quality of POI data and the distinctiveness of RS imagery features. While the model excels in areas with clear functional differentiation, it faces challenges in zones with mixed functions or sparse POI distributions, emphasizing the need for complementary data sources in such contexts.

4.3 Impact of Multi-source Data Fusion on UFZ Classification

To assess the contribution of multi-data sources in UFZ classification, an ablation experiment comparing four models was conducted. The results of the ablation experiment are summarized in Table 2, comparing the classification performance of four models using the F1-score and Kappa coefficient as evaluation metrics. Among the models, the proposed BibDL model, which integrates remotely sensed imagery and POI density data, achieved the highest performance showcasing the effectiveness of multi-modal data integration. The RS-branch we introduced, which relies solely on remotely sensed imagery, also performed well but with slightly lower metrics, achieving an F1-score of 0.927 and a Kappa coefficient of 0.867, illustrating the limitations of using only image data for UFZ classification. The benchmark U-Net model (Yang et al., 2021), widely recognized for its image-based urban functional area classification, achieved an F1-score of 0.857 and a Kappa coefficient of 0.747, performing lower than the RS-branch model due to its simpler structure and lack of multi-modal data. Finally, the SO-CNN model (W. Zhou et al., 2020), which utilizes super objects as mapping units, showed the lowest performance with an F1-score of 0.786 and a Kappa coefficient of 0.623, reflecting the challenges in capturing complex urban functional zone characteristics. These results underline the significant improvements achieved by the BibDL model through the integration of complementary data sources.

Figure 6 and Figure 7 illustrate the comparison between the confusion matrices for the BibDL model and the RS-branch model we proposed, highlighting the significant contribution of multi-modal data to UFZ classification accuracy. The BibDL model achieves superior performance across most classes, particularly in distinguishing finer-grained functional zones such as Commerce, Public, and Academia. For example, in the BibDL model, the Commerce category shows a markedly higher accuracy (71.19%) compared to the RS-branch model (58.18%), indicating the added value of POI features in capturing complex functional characteristics. Similarly, the Public class is perfectly identified (73.33% accuracy) by the BibDL model, while the RS-branch model struggles with this category (54.90% accuracy), misclassifying it into other urban types such as Academia. This improvement can be attributed to the detailed functional information captured by POI data, which helps differentiate Academia zones from visually similar categories like Public and Residence. The RS-branch model, in contrast, shows higher confusion between these classes due to the overlapping spectral and textural features in RS imagery. Another notable improvement is observed in the classification of Traffic zones, where the BibDL model achieves an accuracy of 95.38%, outperforming the RS-branch model's 80.49%. It highlights the contribution of POI data in clarifying ambiguous zones that might be visually similar to Industry in RS imagery. For categories like Green land and Farmland, which share similar spectral characteristics in RS imagery, the BibDL model achieves nearly perfect accuracy of 99.78% and 93.18%, respectively. These two types of areas may have less POI distribution, but their remote sensing imagery is relatively well differentiated, so both BibDL and RS-branch perform better.

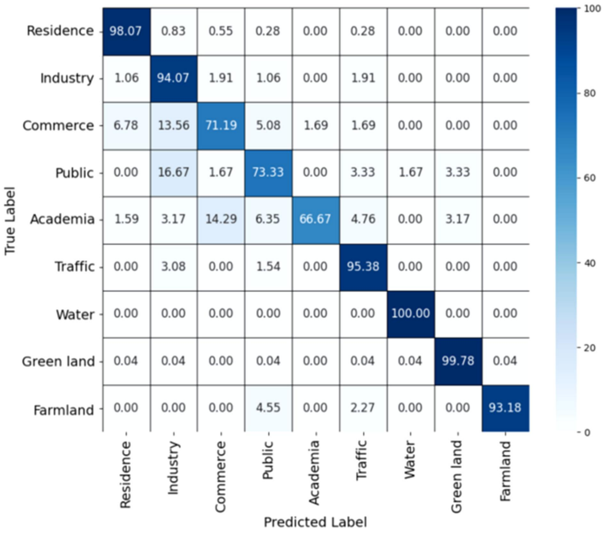

Figure 6. Confusion matrix for the BibDL model with RS image features and POI features.

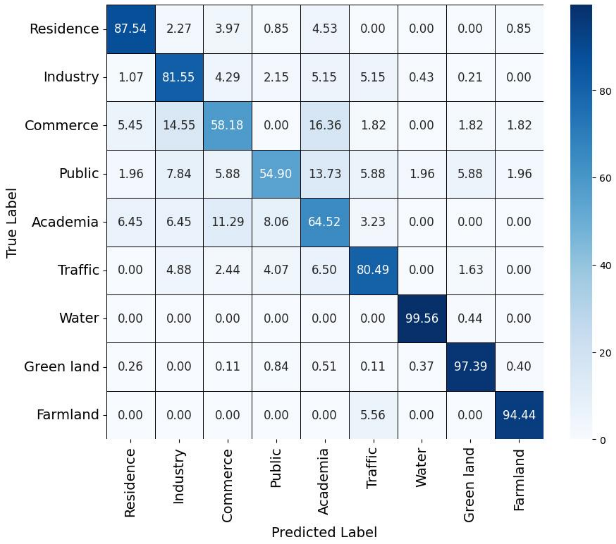

Figure 7. Confusion matrix for the RS-branch model with only RS image features.

In conclusion, the BibDL model demonstrates that integrating multi-modal data, particularly RS imagery and POI features, enhances UFZ classification accuracy. POI data provides functional and contextual information that resolves ambiguities in visually similar classes and improves the model's robustness in diverse urban contexts. These findings emphasize the critical role of multi-modal approaches in advancing UFZ mapping, especially for heterogeneous urban landscapes.

5. Conclusion

This study developed a multi-modal deep learning approach, the BibDL model, for UFZ classification, leveraging both RS imagery and POI data. By integrating these data sources, the BibDL model achieved significant improvements in classification accuracy, particularly for complex and overlapping urban zones where visual features alone are insufficient. The model demonstrated high F1-scores of 0.975 and Kappa coefficients of 0.953 across various UFZ categories, including challenging classes such as Academia, Commerce, and Public. The ablation experiment highlights the value of contextual information from POI data in enhancing model performance.

While the proposed method demonstrates significant improvements in UFZ classification, several limitations and potential areas for future research can be identified. One limitation lies in RS imagery and POI data quality and availability. For future work, integrating other data sources, such as street-level imagery, social media data, or temporal data reflecting changes over time, could enhance the model's contextual understanding and adaptability. Advanced techniques like transfer learning or domain adaptation may also help generalize the model to new cities or regions with distinct urban layouts and POI distributions.

Acknowledgment

This study is supported and funded by Science and Technology Program of Ningbo, China (2023S171) , the Open Fund of The Innovation Center of Spatial-temporal Information and Equipment of Smart Cities, MNR (STIEIC-KF202307), The Innovation Team of the Department of Education of Guangdong Province (2024KCXTD013), the Key Project of Shenzhen Commission of Science and Technology (No.JCYJ20220818 100200001), and The Center for Scientific Research and Development in Higher Education Institutes, MOE (2024HT013).

References

Cao, R., Wei, T., Chen, D., & Zhang, W. (2025). Mapping urban villages in China: Progress and challenges. *Computers, Environment and Urban Systems*.

Du, S., Du, S., Liu, B., & Zhang, X. (2021). Mapping large-scale and fine-grained urban functional zones from VHR images using a multi-scale semantic segmentation network and object based approach. *Remote Sensing of Environment*, 261, 112480.

Du, S., Zhang, X., Lei, Y., Huang, X., Tu, W., Liu, B., Meng, Q., & Du, S. (2024). Mapping urban functional zones with remote sensing and geospatial big data: A systematic review. *GIScience & Remote Sensing*, 61(1), 2404900. doi.org/10.1080/15481603.2024.2404900.

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-Excitation Networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7132–7141. doi.org/10.1109/ CVPR.

2018.00745.

Li, M., Stein, A., Bijker, W., & Zhan, Q. (2016). Urban land use extraction from Very High Resolution remote sensing imagery using a Bayesian network. *ISPRS Journal of Photogrammetry and Remote Sensing*, 122, 192–205.

Lu, W., Tao, C., Li, H., Qi, J., & Li, Y. (2022). A unified deep learning framework for urban functional zone extraction based on multi-source heterogeneous data. *Remote Sensing of Environment*, 270, 112830. doi.org/10.1016/j.rse.2021.112830

Tu, W., Gao, W., Li, M., Yao, Y., He, B., Huang, Z., Zhang, J., & Guo, R. (2024). Spatial cooperative simulation of land use-population-economy in the Greater Bay Area, China. *International Journal of Geographical Information Science*, 38(2), 381–406. doi.org/10.1080/13658816.2023.2285459.

Tu, W., Hu, Z., Li, L., Cao, J., Jiang, J., Li, Q., & Li, Q. (2018). Portraying urban functional zones by coupling remote sensing imagery and human sensing data. *Remote Sensing*, 10(1), 141.

Wang, L., Li, R., Zhang, C., Fang, S., Duan, C., Meng, X., & Atkinson, P. M. (2022). UNetFormer: A UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 190, 196–214.

Xing, X., Yu, B., Kang, C., Huang, B., Gong, J., & Liu, Y. (2024). The Synergy Between Remote Sensing and Social Sensing in Urban Studies: Review and perspectives. *IEEE Geoscience and Remote Sensing Magazine*, 2–31. doi.org/10.1109/ MGRS.2023.3343968.

Xu, S., Qing, L., Han, L., Liu, M., Peng, Y., & Shen, L. (2020). A New Remote Sensing Images and Point-of-Interest Fused (RPF) Model for Sensing Urban Functional Regions. *Remote Sensing*, 12(6), Article 6.doi.org/10.3390/rs12061032.

Xu, W., Wang, J., & Wu, Y. (2022). Multi-Dimension Geospatial Feature Learning for Urban Region Function Recognition. *IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium*, 5832–5835. doi.org/10.1109/IGARSS46834.2022.9884450.

Yang, Y., Wang, D., Yan, Z., & Zhang, S. (2021). Delineating Urban Functional Zones Using U-Net Deep Learning: Case Study of Kuancheng District, Changchun, China. *Land*, 10(11), Article 11. doi.org/10.3390/land10111266.

Yu, M., Xu, H., Zhou, F., Xu, S., & Yin, H. (2023). A Deep-Learning-Based Multimodal Data Fusion Framework for Urban Region Function Recognition. *ISPRS International Journal of Geo-Information*, 12(12), Article 12. doi.org/10.3390/ijgi1212 0468.

Zhou, Q., Zhang, Y., Gao, D., & Sun, B. (2019). Recognition of urban functional regions at street scale based on lj1-01 night-time light remote sensing and mobile big data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *IV-4/W9*, 119–124. doi.org/10.5194/isprs-annals-IV-4-W9-119-2019.

Zhou, W., Ming, D., Lv, X., Zhou, K., Bao, H., & Hong, Z. (2020). SO–CNN based urban functional zone fine division with VHR remote sensing image. *Remote Sensing of Environment*, 236, 111458. doi.org/10.1016/j.rse.2019.111458.