Multimodal Fusion Framework for Urban functional zone change detection using Remote Sensing and Social Sensing Data

Bichen Fang [1,2,3,4], Xiana Chen [1,2,3,4], Hanxin Chen [5,6], Zejun Xiang [5,6], Bo Hu [5,6*], Zhongwen Hu [1,2,3,4], Wei Tu [1,2,3,4]

[1] School of Architecture and Urban Planning, Shenzhen University, Shenzhen 518060, China - fangbichen2020@email.szu.edu.cn, chenxiana2019@email.szu.edu.cn, zwhoo@szu.edu.cn, tuwei@szu.edu.cn
[2] Guangdong Key Laboratory of Urban Informatics, Shenzhen Key Laboratory of Spatial Smart Sensing and Service, Shenzhen 518060, China
[3] Key Laboratory for Geo-Environmental Monitoring of Great Bay Area, MNR, Shenzhen 518060, China
[4] Guangdong Laboratory of Artificial Intelligence and Digital Economy (Shenzhen), Shenzhen 518060, China
[5] Chongqing Academy of Surveying and Mapping, Chongqing 401147, China - chenhx@cqkcy.com, xiangzj@cqkcy.com, hubo@cqkcy.com
[6] Technology Innovation Center for Spatio-temporal Information and Equipment of Intelligent City, Ministry of Natural Resources, Chongqing 40000, China

Keywords： Change detection; Multimodal fusion; Urban functional zones (UFZs); Mobile positioning data

Abstract

Change detection reveals the shifts in land-use distribution and composition over time, providing valuable insights for urban management and socioeconomic analysis. Previous studies have focused primarily on single-modal data like high-resolution remote sensing (RS) imagery, overlooking the role of social sensing characteristics in determining urban functional zones (UFZs). In this study, we propose a novel multimodal dual-branch change detection (MDB-CD) deep learning framework, for detecting UFZ changes by combining RS imagery and social sensing data. It includes the RS-branch and the mobile positioning (MP) branch. For the RS-branch, RS imagery provides detailed spatial information about urban transformations; for the MP-branch, MP data as a typical social sensing data captures temporal patterns in human mobility, offering insights into functional shifts in urban spaces. By fusing these two complementary modalities, our approach allows for a more nuanced detection of urban functional zone changes. Empirical results in Shenzhen, China demonstrate that the MDB-CD model significantly outperforms a baseline model trained only on imagery, achieving higher overall accuracy (OA) of 0.858 and Kappa coefficients of 0.818 across change detection. Specifically, the model generates an OA matrix of UFZ change detection transitions between 2017 and 2019, revealing 81 distinct transitions in UFZs. Notably, the integration of MP data proved instrumental in improving the model capturing subtle changes that RS imagery alone could not distinguish. An ablation study further highlights the significant accuracy improvements achieved by integrating RS imagery and MP data, emphasizing the value of a multimodal approach for detecting UFZ changes. This work highlights the value of incorporating social sensing data into urban change detection, offering a robust solution for dynamic urban planning and development.

1. Introduction

Change detection identifies differences in land-use distribution and composition over time, which is crucial for urban management and socioeconomic assessment (Viana et al., 2019). Detecting changes in urban functional zones (UFZs) can guide infrastructure development, optimizing land use and environmental protection. With the continuous development of world economy and the accelerated pace of urbanization, a large amount of land has been continuously developed and expropriated, resulting in significant changes in the original land cover in a relatively short period of time. The rapid development of urbanization and economy has brought a huge demand for understanding which UFZs have changed (Rui et al., 2025). Therefore, effectively identifying UFZs changes is essential for effective urban dynamic management and sustainable development.

Traditional UFZs change detection methods mostly rely on remote sensing (RS) imagery. For example, Liu et al. (2023) using RS images proposed an attention-based multiscale transformer network (AMTNet) that utilized a CNN-transformer structure to address the challenges of complex textures and seasonal variations. Basavaraju et al. (2022) using bi-temporal satellite imagery introduced an urban change detection network (UCDNet) to effectively tackle the challenge of edge preservation in urban change detection. Peng et al. (2021) based on high-resolution RS datasets developed a novel semisupervised convolutional network for change detection (SemiCDNet) to mitigate the need for large amounts of labeled data. RS data provides valuable physical information about UFZs changes, but these single-modal data often fail to comprehensively characterize the complex nature of urban functional changes. It has limitations in capturing the social and mobility aspects of urban spaces change (Rui et al., 2020).

Multimodal data fusion has shown great potential in various urban studies (Du et al., 2024). For instance, Dong et al. (2024) proposed a novel framework ChangeCLIP from image-text pairs to leverages robust semantic information for change detection. Su et al. (2024) combined satellite imagery with points of interest (POI) data to improve land use classification accuracy. Qiao et al. (2024) introduces an end-to-end deep learning-based multisource dynamic fusion network for UFZs identification on integrated POI, RS image and building footprint data. However, urban functional changes are gradual processes influenced by both spatial modifications and shifts in human behavior patterns. These studies primarily focus on static data, such as text, POI, and building footprint information, while overlooking the

* Corresponding author

influence of human mobility on dynamic transitions in urban functions (Fang et al., 2024).

To address these limitations, this study proposes a multimodal fusion framework to detect UFZs changes by integrating RS images and social sensing data. The framework learns deep features from images at the pixel scale while extracting mobility change patterns of UFZs through social sensing data at the human mobility scale.

2. Study Area and Datasets

2.1 Study Area

The study was conducted in Shenzhen, located in southern Guangdong Province, China. As the nation's first special economic zone, Shenzhen has undergone rapid development over the past three decades and now covers approximately 2,020.5 km^2 with a permanent resident population of 17.79 million. The city was chosen as the study area due to its characteristics as a high-density urban environment with complex functional zoning, making it an ideal case study for examining UFZs dynamics change. The study area is shown in Figure 1.

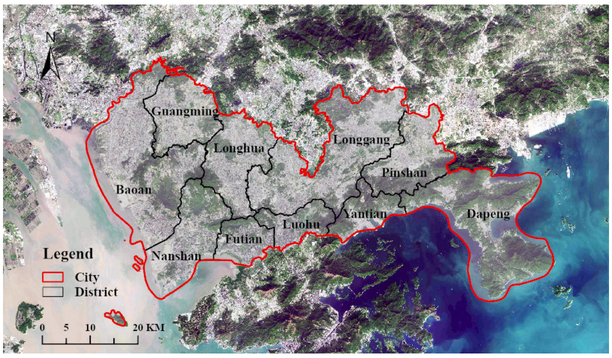

Figure 1. Map of the study area

2.2 Datasets

The study utilized high-resolution RS imagery from the Gaofen-2 satellite for 2017 and 2019, with a spatial resolution of 0.9 meter. These images were labeled and classified according to the second national land survey ground truth data into nine categories: Water, Agriculture, Greenland, Commercial, Industry, Residential, Public, Education, and Transportation. The mobile positioning (MP) data provided by China Unicom is used to capture human mobility in Shenzhen. This dataset captures over 163 thousand records of mobile phone users each day for the years 2017 and 2019. The data were sampled at one-hour intervals, ensuring a fine-grained temporal resolution. Each record comprises key attributes, including an anonymized user ID, geographic coordinates, and a timestamp. It's worth noting that personal information is not recorded to protect private privacy.

3. Methods

The study proposed a novel multimodal dual-branch change detection (MDB-CD) deep learning framework, for detecting UFZs changes by combining RS imagery and social sensing data. The framework consists of a dual-branch architecture that processes both the visual and social feature of urban areas. The workflow begins with two temporal branch network and each network integrates multimodal branches. RS-branch using RS imagery provides detailed spatial information and extract visual semantic features that represent the physical changes about urban transformations. MP-branch using mobile positioning (MP) data - a typical social sensing data - captures temporal mobility patterns in human mobility, offering insights into functional shifts in urban spaces. These two complementary branches are then fused to provide a comprehensive characterization of UFZs changes. The framework enables a more nuanced understanding of urban functional transitions by combining physical spatial modifications with dynamic human mobility patterns. This integrated approach aims to overcome the limitations of traditional single-modal methods and static data.

3.1 Spatial Grid Generation

To analyze UFZs changes, we conduct a grid-based approach using a 500-meter spatial resolution. This grid size is selected to balance the granularity of analysis with the characteristics of both RS and mobility data. The RS images are first clipped according to these 500-meter grid cells, establishing the basic spatial units for our analysis. For consistency in data integration, we process the MP data using the same spatial grid structure. Within each grid cell, we generate statistical summaries of human mobility patterns. The processed MP data is stored in a 7-day, 24-hour, 6-week format, allowing for a comprehensive understanding of mobility patterns within each grid. Then the MP data records stored in NPY format for efficient data handling and subsequent analysis. Through this process, we obtained two datasets including 6824 labeled images and mobility records of each grid in 2017 and 2019. All of these data categorized into one of 9 scene categories for both time periods.

3.2 Multimodal Dual-Branch Fusion

3.2.1 RS-branch： The study uses the SE-ResNeXt101 structure in RS-branch, which excels at extracting detailed visual features from RS images (He et al., 2016; Hu et al., 2018). The SE-ResNeXt101 combines the powerful split-transform-merge strategy of ResNeXt with Squeeze-and-Excitation (SE) blocks, enabling the model to capture multi-scale spatial patterns and contextual information. The SE blocks adaptively recalibrate channel-wise feature responses, enhancing the model's ability to focus on the most relevant visual changes in the RS imagery. This architecture is particularly effective in detecting changes in urban environments, where subtle variations in land cover can be crucial for understanding urban dynamics.

Initially, the RS images pass through the convolutional layers of the pre-trained network, which extract low-level features like edges, textures, and shapes. As the network deepens, more complex features are captured, such as building outlines, road networks, and vegetation patterns. To tailor the network to the specific task of RS image analysis, we replace the original 1000-class classification layer with a custom fully connected layer. This layer reduces the dimensionality of the extracted features to 256. To ensure a consistent input size for the fully connected layer, Adaptive Average Pooling is applied to downsample the feature maps, making them compatible with the network's final layers.

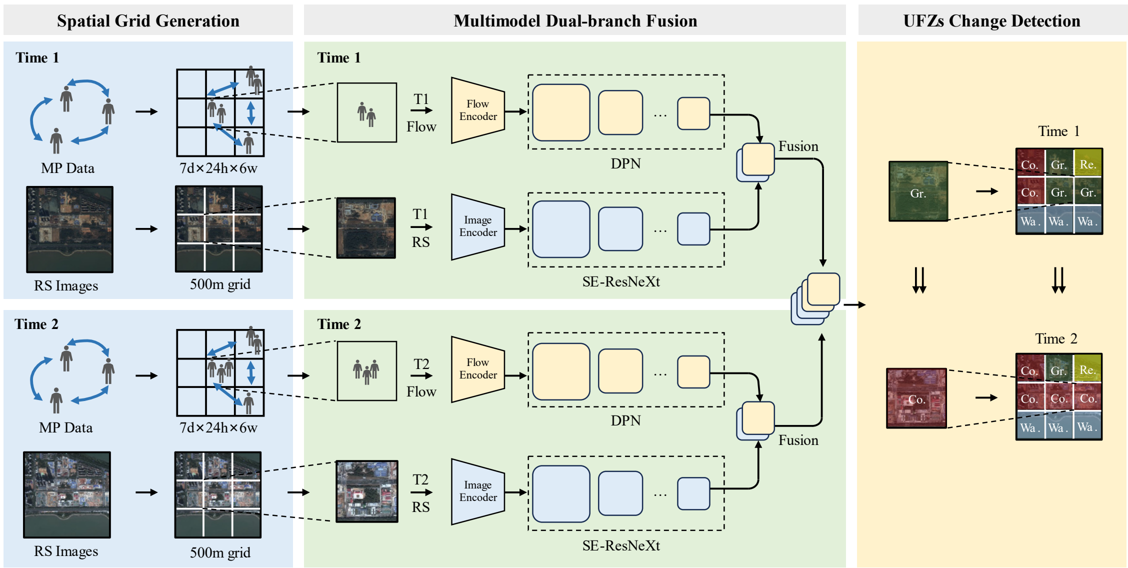

Figure 2. Workflow of the proposed multimodal framework.

3.2.2 MP-Branch: The study uses a simplified DPN-26 (Dual Path Network) structure in MP-branch, chosen for its ability to capture both temporal and spatial patterns in mobility data efficiently (Chen et al., 2017). The DPN-26 network leverages residual connections and dense connectivity, allowing for effective feature reuse and reduced computational costs. This branch processes aggregated mobility statistics from each grid cell, extracting patterns that reflect human activity and mobility. These mobility patterns are essential for detecting changes in how people interact with urban spaces, helping to identify shifts in functionality and land-use patterns in response to evolving human activities.

The raw mobility data first traverses a series of convolutional layers that transform initial spatial-temporal traces into increasingly abstract representations. The network's innovative Bottleneck blocks, characterized by dense connectivity and residual connections, progressively distill complex mobility patterns across four hierarchical layers. Each convolutional block systematically reduces spatial dimensions while expanding feature depth, capturing nuanced aspects of human mobility and interaction within urban spaces. The initial convolutional layers extract basic mobility characteristics, such as mobility frequency and density, while deeper layers synthesize more complex behavioral patterns. Ultimately, global average pooling condenses the extracted features into a compact 64-dimensional representation, effectively capturing the essence of human mobility dynamics.

3.2.3 Multimodal and Multitemporal Fusion: RS-branch and MP-branch independently processes its respective input data—RS images and mobility data—capturing the learned features through the output feature maps. The fusion layer integrates multi-modal features from both 2017 and 2019, combining information from the RS-branch and MP-branch for each time period. This dual-temporal fusion enables the model to compare and contrast urban changes over time, with adaptive weighting applied to prioritize the most relevant data. By analyzing data from these two distinct timestamps, the layer enables comprehensive change detection through cross-modal and cross-temporal integration. In areas with significant physical changes detected by RS images, the model gives higher importance to RS features, while in regions where mobility patterns shift but structural changes are minimal, it emphasizes MP features. This adaptive weighting ensures that the most relevant information from both data sources and time periods is effectively utilized for change detection.

3.3 Evaluation Metric

To evaluate the performance of MDB-CD model, we used five metrics: overall accuracy (OA), recall, F1 score (F1), and kappa coefficient (KAPPA). OA measures the precision of the model across all change categories (Congalton and Green, 2008). Recall quantifies the model's ability to identify true positives within each category (Powers, 2020). F1 averages precision and recall for each category (Rijsbergen, 1979). KAPPA assesses the consistency between the classification results and random classification (Cohen, 1960). Their specific formulas are as follows.

$$\mathrm{OA} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{1}$$

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

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

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

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

where TP = True Positives (correctly classified as positive)
TN = True Negatives (correctly classified as negative)
FP = False Positives (incorrectly classified as positive)
FN = False Negatives (incorrectly classified as negative)
p_o = Overall classification accuracy
$a_1, a_2, \ldots a_i$ = True samples in category i
$b_1, b_2, \ldots b_i$ = Predicted samples in category i
n = Total number of samples

4. Result and Analysis

This study presents the results of the MDB-CD model for detecting UFZ changes, including the OA matrix for 81 transition pairs from 2017 to 2019. It also shows three typical scenes in Shenzhen to demonstrate the model's performance. Ablation experiments highlight the importance of data fusion and attention mechanisms for accurately classifying UFZ changes in complex urban environments.

4.1 UFZs Transition Result Based on MDB-CD Model

The MDB-CD model achieves an overall test accuracy of 0.858 and a Kappa value of 0.818, demonstrating its ability to effectively capture nuanced features from both the physical and social dimensions of urban environments. Figure 3 shows the OA matrix of the 81 types of UFZs change detection transition pair from 2017 to 2019.

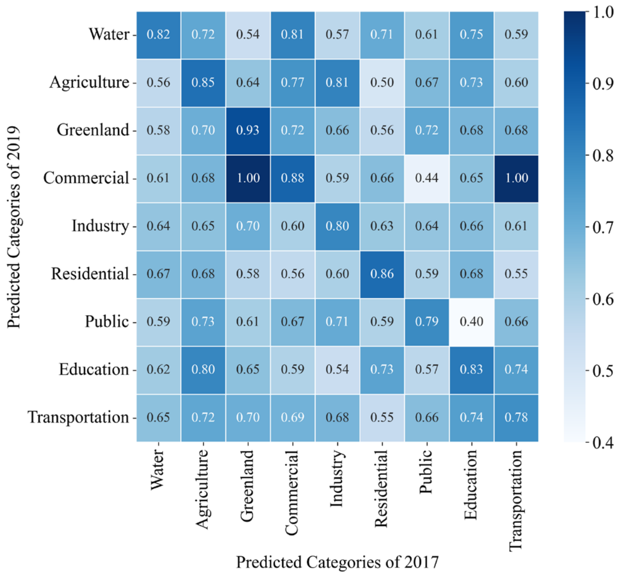

Figure 3. The overall accuracy (OA) matrix of UFZs change detection transition pair from 2017 to 2019

Some UFZs transitions exhibit relatively high OA, particularly along the diagonal of the matrix, indicating changes within the same category. These values remain stable, reflecting minimal misclassification within each category. For example, the OA of Greenland to Greenland is exceptionally high at 0.93, suggesting that Greenland has remained largely stable over the two-year period, with few errors in classification. This stability in certain categories reflects the persistence of UFZs types like Greenland, which experience fewer alterations in urban environments.

The transitions between the Education, Public, and Commercial categories show relatively low OA values. Specifically, the OA from Education to Public and Commercial is 0.4 and 0.65, from Public to Commercial and Education is 0.44 and 0.57, and from Commercial to Education is 0.59. These low values suggest challenges in accurately classifying transitions between these categories. One reason could be the small sample size of these UFZs types, which limits the model’s ability to learn their distinctive features. Another reason is the similarity in spatial characteristics of Education, Public, and Commercial scenes, making it difficult to distinguish them in high-density urban areas. The overlap in features, combined with the dynamic nature of urban development, complicates accurate classification and transition detection.

Notably, the transitions from Greenland to Commercial and from Transportation to Commercial both have an OA of 1, indicating perfect classification and suggesting that the MDB-CD model was able to accurately detect and classify these changes. This accuracy underscores the significant role of MP data in the MDB-CD model’s ability to distinguish between different UFZs categories, particularly in dynamic urban environments. The inclusion of MP data helps capture human mobility features, which are essential for identifying subtle shifts between urban landscapes and built-up areas. In cases where traditional remote sensing (RS) data alone may struggle to differentiate between such categories—due to similarities in spatial characteristics or ambiguous UFZs types—MP data provides crucial contextual information that allows the model to resolve these ambiguities.

4.2 UFZs Change Detection Based on MDB-CD Model

To demonstrate the results of the MDB-CD model, we present three scenes in Shenzhen in Figure 4(a). The yellow boxes mark the positions of the detected scene image scenes in Shenzhen. These locations represent typical examples of newly developed or transformed UFZs.

Figure 4.1 illustrates the transition of a public facility which was under construction in 2017 and completed before 2019. The MDB-CD model successfully detected the entire building within its grid. This location corresponds to the Shenzhen World Exhibition and Convention Center. However, one grid, which included part of the convention halls, was misclassified as a Public. This misclassification highlights the influence of human mobility patterns captured by MP data, which obscured the classification features presented in the RS images.

Figure 4.2 captures the development of a commercial center featuring the CITIC Financial Center and China Merchants Bank Headquarters, occupying two grid cells in the lower right corner. The change detection accurately tracked their construction progress from 2017 to 2019, successfully identifying the UFZs classification changes associated with landmark buildings. The combination of RS and MP data effectively captures category changes in the presence of significant building transformations.

Figure 4.3 depicts the development of a logistics hub in Shenzhen. Specifically, it refers to the Pinghu National Logistics Hub, located in Longgang District, which was among the first 23 national logistics hubs designated in China in 2019. This result highlights the critical role of MP data in identifying land-use changes. The RS imagery alone might suggest this area belongs to categories such as Greenland or Agricultural. However, the inclusion of MP data during the early construction phase allowed the MDB-CD model to correctly classify it as Transportation.
In conclusion, the MDB-CD model has demonstrated robust detection capabilities across various types of land-use changes, showing significant improvement over traditional RS-only methods. However, the results indicate that the model still faces some limitations in mixed-use areas with complex human mobility, suggesting the need for integration with additional data sources for further refinement.

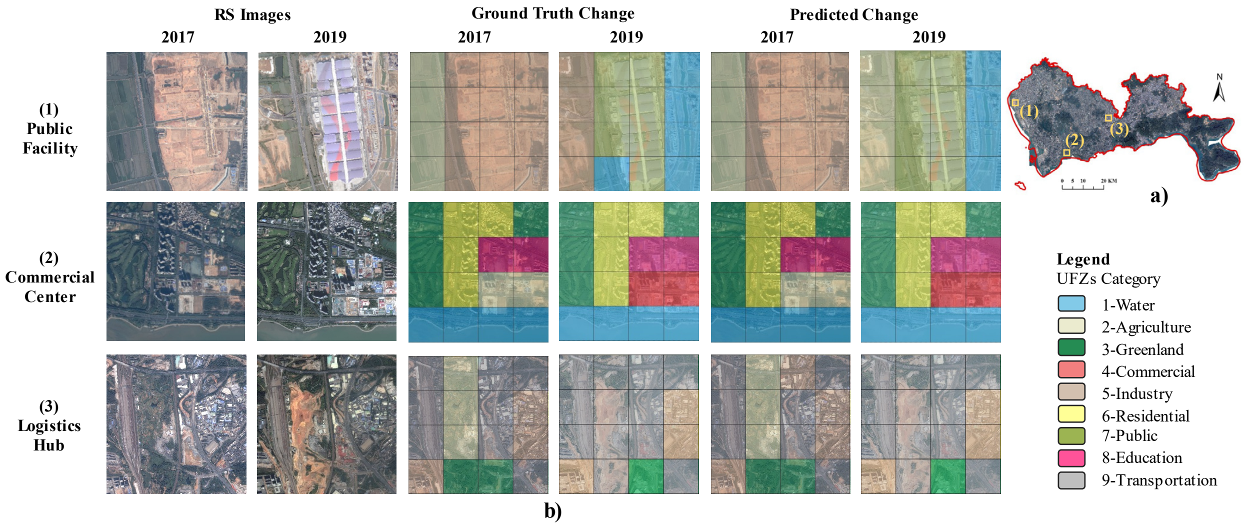

Figure 4. Representative urban scenes. a) locations of scenes; b) RS Images, Ground Truth Change and Predicted Change of UFZs in 2017 and 2019

4.3 Model Ablation Analysis

To verify the effectiveness of the proposed MDB-CD model in fusing two data sources, we conducted ablation experiments with single data sources. Table 1 presents the performance results for multiple models. The ResNet-based model utilizing RS imagery demonstrated substantially superior performance, achieving a test OA of 0.806 and a Kappa coefficient of 0.746. Notably, the DPN model using MP data alone exhibited lower performance, with a test OA of 0.545 and a Kappa coefficient of 0.412. Although this type of data offers valuable insights into human mobility and population distribution, the raw MP data alone is insufficient to capture the intricate changes in UFZs due to its limited spatial and contextual resolution. Highly populated areas such as Residential, Commercial, or Public zones often exhibit similar patterns in human mobility, which can complicate the identification of distinct urban functional changes. This underscores the importance of combining RS imagery and MP data to leverage their complementary strengths in accurately detecting urban functional changes.

Model	RS	MP	OA	Kappa
1. ResNet	√		0.806	0.746
2. DPN		√	0.545	0.412
3. ResNet &DPN	√	√	0.821	0.774
4. **MDB-CD** (SE-ResNet &DPN)	√	√	**0.858**	**0.818**

Table 1. Performance of all models on the dataset. Models 1–2: Single data; Models 3: Two-data fusion; Models 4: MDB-CD (proposed method)

Compared to classification based on single data, data fusion further significantly improves accuracy. Model 3, combining RS images and MP data without attention mechanisms, already surpasses single-data models, demonstrating the complementary nature of these data sources. Model 4 (MDB-CD), which integrates SE-ResNet and DPN with attention mechanisms, achieves the best results, improving the OA by 6.5% over the RS-only model and by 57.5% over the MP-only model. This highlights the model's capability to effectively weigh and extract features from heterogeneous data. Ablation experiments further validate the importance of each component in the MDB-CD model. The attention mechanisms, in particular, allow the model to dynamically focus on the most informative aspects of the data, prioritizing relevant features that are crucial for accurately detecting urban functional zone changes. As a result, the model not only improves the precision of predictions but also enhances its robustness in dealing with the complexities of urban environments.

In conclusion, the ablation experiments confirm the effectiveness of the MDB-CD model in integrating multi-source data for UFZ change detection. By leveraging the complementary strengths of RS images and MP data, and incorporating attention mechanisms to enhance feature extraction, the MDB-CD model significantly improves change detection performance. These results demonstrate the potential of fusing multimodal data for complex urban analysis tasks.

5. Conclusion

This study presents a novel multimodal dual-branch deep learning framework, MDB-CD, for change detection in UFZs by integrating RS imagery and social sensing data. The MDB-CD framework enhances the model's ability to detect changes in urban environments by combining physical spatial modifications and dynamic social behavior patterns. The dual-branch architecture processes the data through two distinct pathways. The RS-branch focuses on RS imagery to extract semantic features related to urban landscapes, such as building density, land cover changes, and infrastructure development. Meanwhile, the MP-branch analyzes MP data to capture temporal variations in human activity, reflecting shifts in social behavior and urban functionality. The MDB-CD framework is able to detect and understand both physical changes and social dynamics by fusing these two complementary modalities, offering a more comprehensive view of UFZs transitions over time. The model shows high OA of 0.858 and a Kappa coefficient of 0.818 across urban change detection. Specifically, the model generates an OA matrix of 81 distinct UFZs change detection transitions from

2017 to 2019. Notably, the integration of MP data proved instrumental in improving the model capturing subtle changes that RS imagery alone could not distinguish. Also, the study conducts an ablation analysis to highlights the significant accuracy improvements achieved by integrating RS imagery and MP data, emphasizing the value of a multimodal approach for detecting UFZs changes.

The proposed method demonstrates significant improvements in detecting UFZs changes, but several limitations remain. One limitation is the inherent imbalance in the data samples, which may affect the model's robustness. Exploring transfer learning and few-shot learning approaches could help mitigate the challenges of data sample imbalance. Also, the current approach of integrating RS imagery with MP data offers a promising foundation, but there is significant potential to expand the research scope by incorporating additional data sources. Future work will focus on explore the integration of high-resolution street-level imagery, infrastructure network data and other urban multiple data to create a more comprehensive understanding of UFZs dynamics.

Acknowledgment

This study is supported and funded by National Science Foundation of China (42311530335), 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), and the Key Project of Shenzhen Commission of Science and Technology (No.JCYJ20220818 100200001).

References

Basavaraju, K.S., Sravya, N., Lal, S., Nalini, J., Reddy, C.S., Dell’Acqua, F., 2022. UCDNet: A deep learning model for urban change detection from bi-temporal multispectral sentinel-2 satellite images. *IEEE Trans. Geosci. Remote Sens.* 60, 1–10. doi.org/10.1109/TGRS.2022.3161337

Cao, R., Tu, W., Li, Qing, Liu, J., Zhu, J., Zhang, Q., Li, Qingquan, Qiu, G., 2020. Deep learning-based remote and social sensing data fusion for urban region function recognition. *ISRPS J. Photogramm. Remote Sens.* 163, 82–97. doi.org/10.1016/j.isprsjprs.2020.02.014

Cao, R., Tu, W., Chen, D., Zeng, W., 2025. Mapping urban villages in China: Progress and challenges. *Computers, Environment and Urban Systems.*

Chen, Y., Li, J., Xiao, H., Jin, X., Yan, S., Feng, J., 2017. Dual path networks, in: *Advances in Neural Information Processing Systems.* Curran Associates, Inc.

Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*. 20, 37–46. doi.org/10.1177/001316446002000104

Congalton, R.G., Green, K., 2008. Assessing the accuracy of remotely sensed data: Principles and practices, second edition, 2nd ed. CRC Press, Boca Raton. doi.org/10.1201/9781420055139

Dong, S., Wang, L., Du, B., Meng, X., 2024. ChangeCLIP: Remote sensing change detection with multimodal vision-language representation learning. *ISPRS Journal of Photogrammetry and Remote Sensing*. 208, 53–69. doi.org/10.1016/j.isprsjprs.2024.01.004

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

Fang, B., Li, M., Huang, Z., Yue, Y., Tu, W., Guo, R. (2024). Revealing multi-scale spatial synergy of mega-city region from a human mobility perspective. *Geo-Spatial Information Science*, 1–16. doi.org/10.1080/10095020.2024.2379060

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. doi.org/10.1109/CVPR.2016.90

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks. Presented at *the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141.

Liu, Wei, Lin, Y., Liu, Weijia, Yu, Y., Li, J., 2023. An attention-based multiscale transformer network for remote sensing image change detection. *Isprs J. Photogramm. Remote Sens.* 202, 599–609. doi.org/10.1016/j.isprsjprs.2023.07.001

Peng, D., Bruzzone, L., Zhang, Y., Guan, H., Ding, H., Huang, X., 2021. SemiCDNet: A semisupervised convolutional neural network for change detection in high resolution remote-sensing images. IEEE Trans. *Geosci. Remote Sens.* 59, 5891–5906. doi.org/10.1109/TGRS.2020.3011913

Powers, D.M.W., 2020. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. doi.org/10.48550/arXiv.2010.16061

Qiao, H., Jiang, H., Yang, G., Jing, F., Sun, W., Lu, C., Meng, X., 2024. A multisource dynamic fusion network for urban functional zone identification on remote sensing, POI, and building footprint. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.* 17, 10583–10599. doi.org/10.1109/JSTARS.2024.3404094

Rijsbergen, C.J.V., 1979. Information retrieval. Butterworths.

Su, C., Hu, X., Meng, Q., Zhang, L., Shi, W., Zhao, M., 2024. A multimodal fusion framework for urban scene understanding and functional identification using geospatial data. Int. J. *Appl. Earth Obs. Geoinf.* 127, 103696. doi.org/10.1016/j.jag.2024.103696

Viana, C.M., Oliveira, S., Oliveira, S.C., Rocha, J., 2019. 29 - land use/land cover change detection and urban sprawl analysis, in: Pourghasemi, H.R., Gokceoglu, C. (Eds.), Spatial Modeling in GIS and R for Earth and Environmental Sciences. Elsevier, pp. 621–651. doi.org/10.1016/B978-0-12-815226-3.00029-6