International Journal of Applied Earth Observation and Geoinformation 145 (2025) 104999

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

High-resolution local climate zone mapping via deep mixed-scene decomposition of remote sensing imagery

Jiayi Li [a,b], Xinji Tian [b], Wenrui Wang [c,*], Lilin Tu [b], Yang Lu [d,e], Jie Jiang [f], Xin Huang [b]

[a] Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resource, Shenzhen 518034, China
[b] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[c] The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[d] School of Civil Engineering and Architecture, Wuhan Institute of Technology, Wuhan 430074, China
[e] Village Culture and Human Settlements Research Center, Wuhan Institute of Technology, Wuhan 430074, China
[f] School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

ARTICLE INFO

Keywords:
Local climate zone
Deep mixed-scene decomposition
High-resolution remote sensing images
OpenStreetMap
Pure images
Mixed scenarios

ABSTRACT

Under rapid urbanization, traditional single-class LCZ mapping methods fail to represent the coexistence of multiple land-cover types within the same block, resulting in blurred boundaries and reduced accuracy for urban heat-island modeling. To address this, LCZ mapping is reformulated as a mixed-scene unmixing task and tackled with a novel deep-learning framework, MSU-Net. Real street-block morphologies from OpenStreetMap are combined with 1 m Google Earth imagery to create multi-scale inputs that preserve both global block layouts and local detail. MSU-Net comprises a primary unmixing branch reinforced by two auxiliary guidance branches—one driven by purified-image semantic cues, the other by sparse local spatial reconstruction, and a Dual Cross-Attention Fusion (DCAF) module that integrates global–local and global–purified features under non-negativity and sum-to-one constraints. Two block-level datasets covering Wuhan and Shenzhen were created. MSU-Net outperforms existing methods on these datasets, boosting overall accuracy by 15–17 %, reducing mean absolute error by over 35 %, and cutting weighted-difference error by around 30 %. Transfer learning further confirms its robustness across cities with distinct morphologies.

1. Introduction

Under rapid urbanization, the urban heat island (UHI) effect, in which urban temperatures exceed those of surrounding rural areas, has intensified, posing serious threats to energy demand, human health, and exacerbating air pollution and heat-related risks (Liu et al., 2025). Accurate assessment of UHI is essential for designing effective mitigation strategies and also contributes to achieving the United Nations' Sustainable Development Goals. (Cao et al., 2023). To provide a scientific basis for UHI analysis, (Stewart and Oke, 2012) introduced the Local Climate Zone (LCZ) typology, establishing a standardized quantitative framework for UHI assessment and urban-form classification that has become the international benchmark. LCZs are defined as horizontally coherent zones ranging from several hundred meters to a few kilometers, characterized by consistent surface cover, structural attributes, material properties, and human activity levels (Feng and Liu, 2022). They comprise ten built types (LCZ 1–10) and seven natural types (LCZ A–G) (Baqa et al., 2025; Xu et al., 2021), effectively capturing urban and peri-urban heterogeneity (Fig. 1).

Remote sensing, with its wide-area coverage and consistent observations, has become the primary means for LCZ mapping (Vavassori et al., 2024; Zhu et al., 2024). By surveying recent publications on LCZ mapping, the current state of research is summarized in Table 1. Spatial resolutions of remote sensing imagery have advanced from the hectometer to the meter scale, thereby providing a richer and more detailed data source for LCZ mapping. As Table 1 shows, medium-resolution imagery remains popular in LCZ mapping due to its accessibility and broad applicability. Meanwhile, high-resolution imagery, when combined with deep-learning–based interpretation, offers markedly superior capability to resolve fine-scale spatial structures.

Although LCZ mapping methods have proven effective, they struggle to represent increasingly intricate urban forms. LCZs are defined as homogeneous zones with consistent structure and function, yet rapid urbanization produces areas where contrasting building heights,

* Corresponding author.
E-mail address: 2017302590121@whu.edu.cn (W. Wang).

https://doi.org/10.1016/j.jag.2025.104999
Received 7 August 2025; Received in revised form 11 November 2025; Accepted 28 November 2025
Available online 3 December 2025

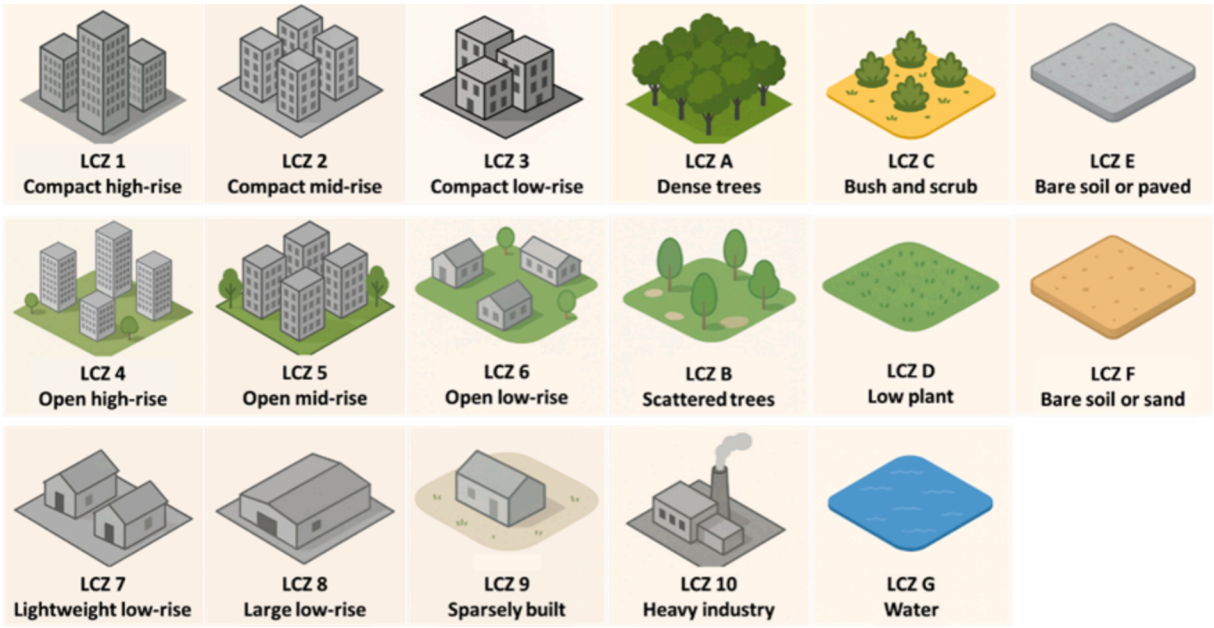


Fig. 1. Local Climate Zone (LCZ) Typology.

densities and land uses coexist within the same block(Li et al., 2024; Zhao et al., 2023). For example, commercial podiums (LCZ 2) often abut residential towers (LCZ 1) in mixed-use districts, while industrial blocks (LCZ 8), housing (LCZ 5) and green spaces intermingle along urban fringes (Fig. 2). Assigning a single LCZ class to such heterogeneous zones leads to confusion and blurred boundaries, reducing the accuracy of microclimate simulations and functional zoning(Anjos et al., 2020; Quan, 2019). Although introducing subclasses can address local classification issues, an overabundance of tailored categories undermines comparability across studies(Stewart and Oke, 2012). A more robust solution is to model each location as a mixture of standard LCZ types and recover their proportions through an "**unmixing/decomposition**" process. Recent advances in sub-meter remote sensing and deep-learning interpretation now make it feasible to identify and quantify these mixed scenarios by integrating structural and functional cues into a hybrid feature space(Adegun et al., 2023). This mixed-type LCZ mapping approach promises to capture urban surface complexity more faithfully and to enhance applications in climate analysis, green planning and fine-scale governance.

This paper addresses these challenges by introducing a high-resolution LCZ mapping approach based on deep mixed-scene decomposition. First, real street-block morphology from OpenStreetMap is combined with high-resolution imagery into multi-scale inputs that retain both block-level layouts and object-level details. LCZ mapping is then reformulated as a mixed-scene unmixing task and addressed using the proposed Mixed Scene Unmixing Network (MSU-Net), which integrates a main unmixing branch with two auxiliary branches: one for pure-class semantic prediction and the other for sparse local patch reconstruction. Finally, a Dual Cross-Attention Fusion (DCAF) module further enhances feature integration across semantic and spatial dimensions, enabling accurate decomposition of mixed LCZ scenes.

The paper is organized as follows. Section 2 introduces the study area and dataset. Section 3 outlines the proposed method and novel modules. Section 4 presents results, comparisons, and ablations. Section 5 discusses the advantages of high-resolution imagery and cross-city transferability.

2. Study areas and data

2.1. Study area

Wuhan (Fig. 3a), the largest city in central China, covers about 8,569 km^2 at the confluence of the Yangtze and Han Rivers. Approximately one-quarter of its area is occupied by water bodies, and together with flat terrain and a polycentric urban structure, the city exhibits diverse forms that are well-suited for mixed-scene decomposition. Shenzhen (Fig. 3b), one of China's first Special Economic Zones, spans 1,997 km^2 on the eastern Pearl River Delta. Although both cities experience a subtropical monsoon climate, Wuhan has distinct seasons, whereas Shenzhen is marked by abundant rainfall and high humidity.

Over the past two decades, both cities have undergone rapid densification and outward expansion. Wuhan's built-up area expanded from ~600 to 1,500 km^2, forming a complex polycentric structure, while Shenzhen developed high-density, vertical, and mixed-use urban forms under land constraints. Both cities exhibit strong structural and functional mixing: Wuhan's older districts combine street-level commerce, low-rise housing, and industrial uses, with transitional mixed zones at the urban fringes, whereas Shenzhen features integrated complexes of high-rise residences, commercial centers, and industrial zones. This rapid urbanization has intensified anthropogenic heat emissions, leading to pronounced urban thermal challenges and making both cities ideal testbeds for LCZ mapping.

2.2. Data

To accommodate the irregular shapes and spatial characteristics of actual LCZ blocks, OpenStreetMap (OSM) road network data were combined with high-resolution imagery (Fonte et al., 2019), with individual street blocks treated as the fundamental mapping unit. High-resolution remote sensing images in the RGB spectral bands, with a spatial resolution of 1 m, were acquired from the Google Earth platform (https://earth.google.com/) for the period 2019–2020, following Google's usage guidelines, for non-commercial scientific research. A total of 635 images covering Wuhan and Shenzhen, each sized 3 500 × 3 500 pixels, were used in this study.

We manually annotated high-resolution Google images by cutting them into block-level patches using the OSM road network, and assigned

Table 1
Comparison of LCZ interpretation methods at different scales.

	Pixel-level	Object-level	Scene-level
Primary Input Data	Medium-resolution imagery (10–30 m; e.g., Landsat/Sentinel-2)(Ching et al., 2018)	High-resolution imagery (e.g., Pleiades, SPOT-6)(La et al., 2020; Ma et al., 2021a)	High-resolution imagery with extended scene context(Bechtel et al., 2019; Manyanya et al., 2022)
Features	Primarily spectral features, fused with spatial/temporal/radar cues (Demuzere et al., 2019; Zhang et al., 2021)	Object-scale spatial pattens and spectral features(Dos Anjos et al., 2017; Ma et al., 2021b)	Multi-scale spatial features, often fused with temporal/spectral cues(Huang et al., 2021; Yu et al., 2022; Zhang et al., 2022)
Interpretation Unit	Regular grid cells (e.g., 100 m) (Shih et al., 2020; Verdonck et al., 2017)	Irregular object segment (Collins and Dronova, 2019; Ma et al., 2021b; Yoo et al., 2019)	Sliding-window grids with majority-voting (e.g., 100–640 m). (Kim et al., 2021; Rosentreter et al., 2020)
Method	Supervised classifiers (e.g., Random Forest, RF) on handcrafted features(Liang et al., 2021; Wang et al., 2021)	Image segment first, then classify the object (RF/CNN) (Liu et al., 2019)	Grid-based CNNs(Liu et al., 2023; Taubenböck et al., 2020)
Output	Pixel-level LCZ labels, aggregated to coarser grids (Anjos et al., 2020; Vandamme et al., 2019)	Object-level LCZ assignments, mosaicked spatially(Simanjuntak et al., 2019)	Scene-level LCZ maps per grid(Gawlikowski et al., 2020; Yoo et al., 2020)
Advantages	Well-established workflow; leverages open-source data (Aslam et al., 2021; Ching et al., 2018)	Better spatial segmentation; preserves object integrity(Zhou et al., 2021)	Captures high-level semantics; flexible scale(Feng et al., 2020; Zhao et al., 2020)
Limitations	Heavily spectral-driven; limited spatial context(Wang et al., 2024)	Computationally intensive; fails to extract semantically coherent objects across varied scales (e.g., LCZ8 vs. LCZ6) (Wang et al., 2024)	Sliding windows may sacrifice fine detail for efficiency(Liu and Shi, 2020; Wang et al., 2024)

decomposition-level labels based on visual interpretation (Fig. 2). The above annotation process was done manually by three experienced remote sensing image interpreters. In Wuhan, block areas range from 71.32 m^2 to 10,476,445.94 m^2, with a mean of 374,924.53 m^2, while in Shenzhen they range from 8,003.72 m^2 to 27,782,903.75 m^2, with a mean of 109,964.11 m^2. To mitigate class imbalance, data augmentation (i.e., rotation, flipping, scaling, and color jittering) was applied to rare classes in Shenzhen (LCZ1, LCZ3, LCZF, and LCZG). The class distributions in mixed and pure images for both cities are shown in Fig. 4. Wuhan includes 3,409 training and 1,041 testing blocks, while Shenzhen has 11,753 and 4,799, respectively.

3. Methodology

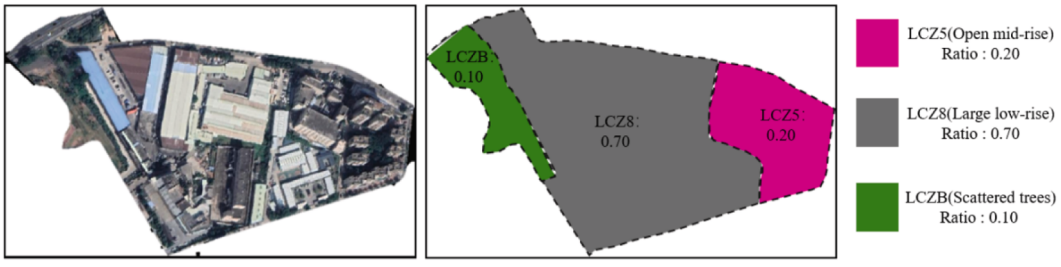
3.1. Overall structure

This study tiles imagery into block-level units using OSM-derived road networks and proposes MSU-Net (Fig. 5) to predict LCZ composition for each block. A multi-scale input strategy is used: globally, each irregular block was down-sampled to 256 × 256 pixels to preserve the overall urban layout (the "mixed patch" in Fig. 5); locally, high-resolution crops are cropped from this global image at 256 × 256 pixels (the "local patch" in Fig. 5). Patches consisting entirely of background are discarded, and each remaining local patch is paired with its corresponding global mixed image to form a global–local input set. This design mitigates the high-frequency information loss from resampling and enhances the model's ability to capture complex urban morphology.

Fig. 5 illustrates the architecture of MSU-Net, which comprises a primary unmixing branch (center row) and two auxiliary branches (top and bottom rows). The primary branch receives a global block representation from a shared-weight encoder (in this study, a Swin Transformer) and incorporates local detail features from the upper auxiliary branch and endmember semantic features from the lower auxiliary branch via the Dual Cross-Attention Fusion (DCAF) module (Section 3.2). Finally, ReLU and Softmax enforce non-negativity and sum-to-one constraints for physically consistent decomposition.

Inspired by hyperspectral unmixing, where a few pure endmembers enhance decomposition (Hong et al., 2022), a limited set of pure LCZ blocks is introduced to reinforce the unmixing branch. A global Purified-image Semantic Guidance (PSG) branch is proposed, extracting semantic cues from pure block classifications to guide mixed-scene decomposition. This high level semantic guidance from scene-based classification bridges low-level features and high-level semantics (Wang et al., 2024). As shown in the top row of Fig. 5, this branch includes a shared-weight encoder followed by a full constraint module.

Furthermore, it has been noted that built-up classes LCZ1 to LCZ6 are often confused because their key local spatial cues are not adequately captured (Zonato et al., 2020). To address this, local spatial guidance



a. RGB visualization of a street block b. Manually annotation

Fig. 2. The "transition zone" in a single block with mixed industrial buildings (LCZ 8), residences (LCZ 5) and green spaces (LCZ B). The street-block boundaries from OSM were overlaid on high-resolution imagery to illustrate typical mixed scenarios. When the LCZ mixing ratios were manually annotated, the proportion of each pure LCZ class within the entire block was recorded as that class's mixing ratio (i.e., abundance); for example, LCZ B, LCZ 8 and LCZ 5 occupies 10%, 70%, and 20% of the block area, respectively. Image data © Google. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

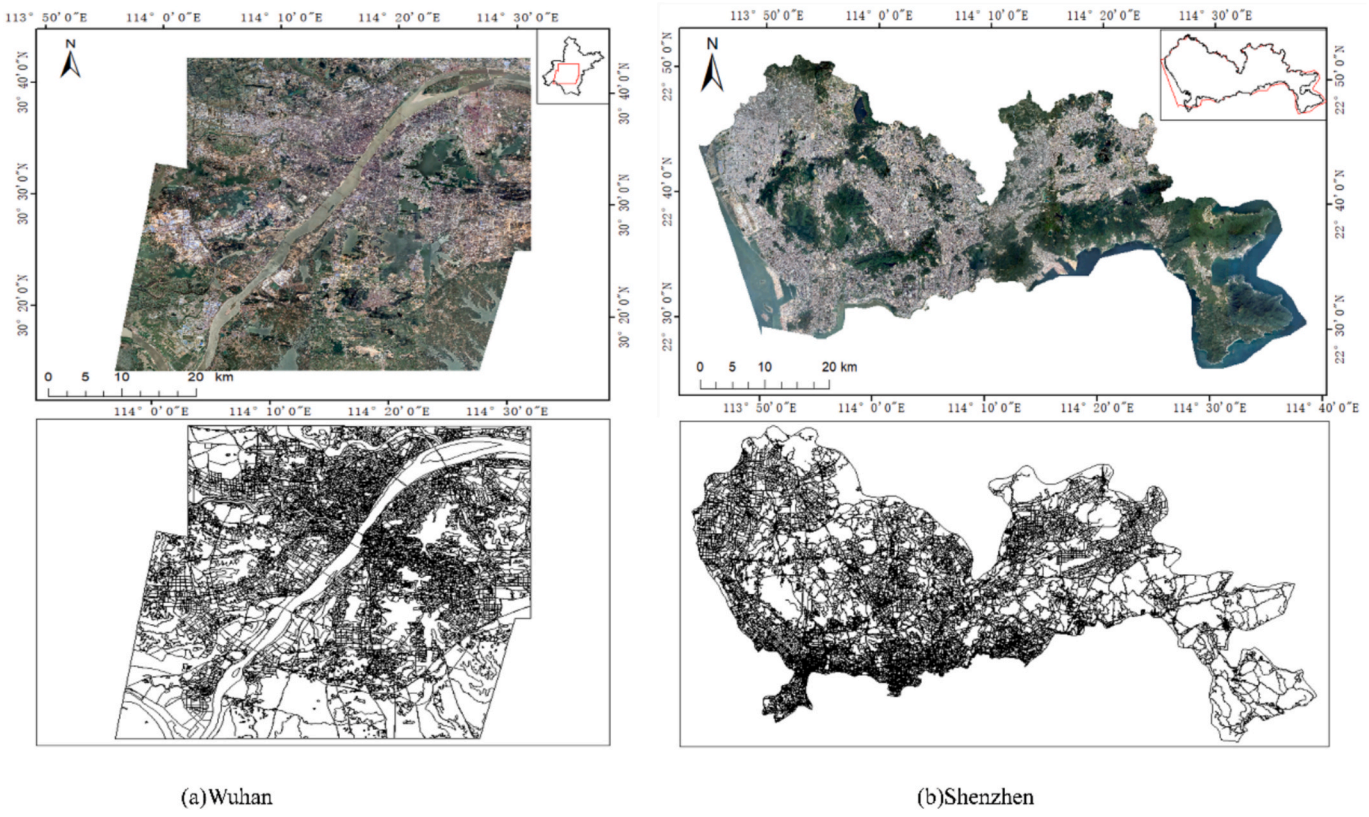
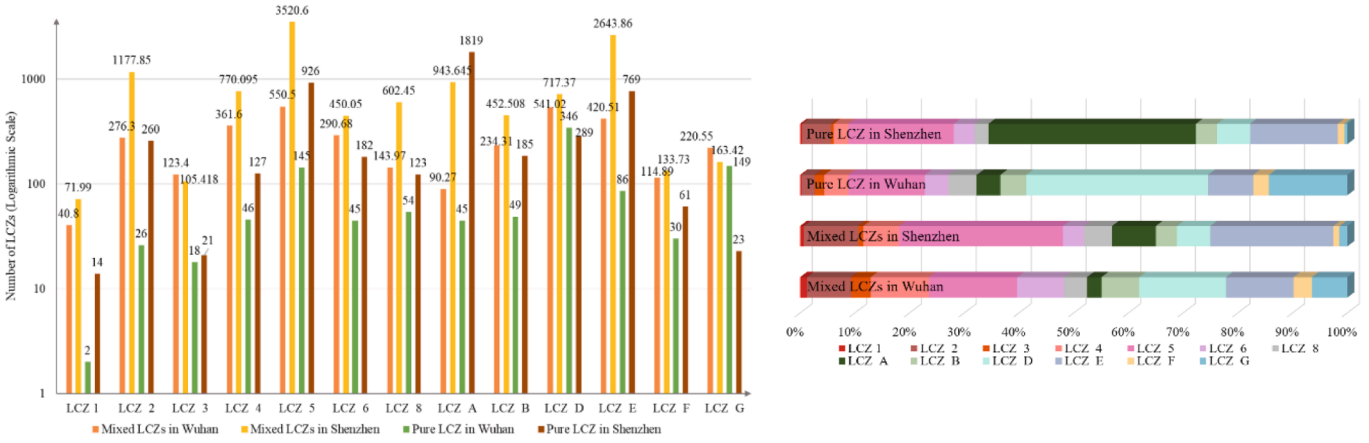


Fig. 3. The study areas and their corresponding OSM derived street scene boundaries. Image data © Google.



(a) Number of each LCZ category (b) Proportion of each LCZ category

Fig. 4. The distribution of each LCZ category in the study areas. For a mixed scene, the quantity refers to the ratio of a certain type of LCZ pixels in the entire block.

(LSG) is provided for LCZ scene unmixing by utilizing fine spatial details extracted from local image patches. An autoencoder-based branch was constructed for this purpose. Local patches are encoded by the shared-weight encoder, and a full-constraint module in combination with a pooling feature pyramid is employed to derive multi-scale spatial features. Because local decomposition results are highly sparse, an $L_{1/2}$ sparsity constraint (Xu et al., 2010) is incorporated to generate more effective local spatial guidance.

3.2. Dual Cross-Attention Fusion (DCAF)

The DCAF module integrates features along two complementary dimensions. The first dimension fuses global and local representations via a cross-attention mechanism to balance spatial context. Small blocks (smaller than 100 m × 100 m) provide rich local cues (e.g., textures) but lack the wider context needed to distinguish classes with similar textures. For example, LCZ3 and LCZ6 exhibit nearly identical local textures and can only be differentiated by their overall spatial arrangement. Large blocks (larger than 500 m × 500 m) preserve global layout but lose fine details when down-sampled. To reconcile this, query vectors from global features attend to local keys, aggregating local values into the global space. This enhances long-range dependency modeling and produces coherent embeddings that capture both contextual layout and fine-scale detail.

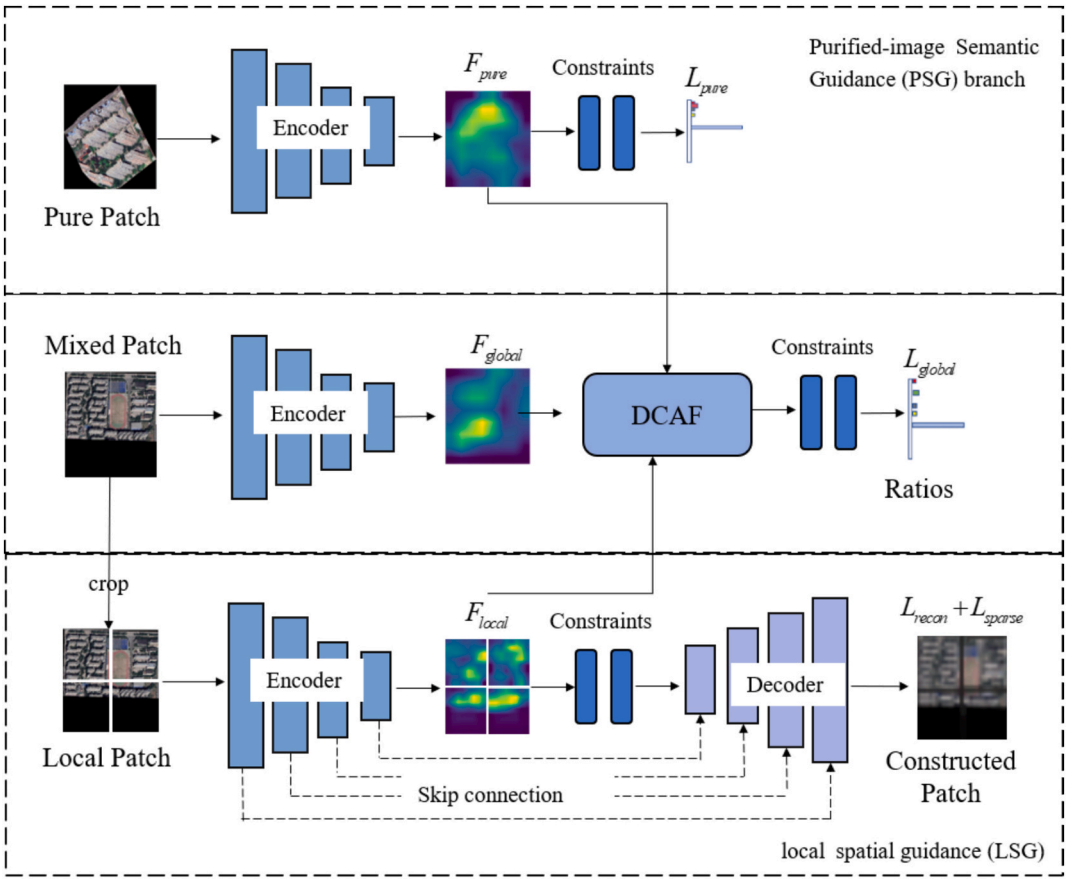


Fig. 5. The framework of MSU-Net. Global (down-sampled block image) and local (high-resolution crop) inputs are encoded in parallel. A primary unmixing branch then fuses the global feature with local detail cues and semantic endmember cues via the Dual Cross-Attention Fusion (DCAF) module. Finally, ReLU and SoftMax activations enforce non-negativity and sum-to-one constraints for proportional LCZ decomposition. The "ratio" refers to the proportion of a specific pure LCZ type within the entire block area. Image data © Google.

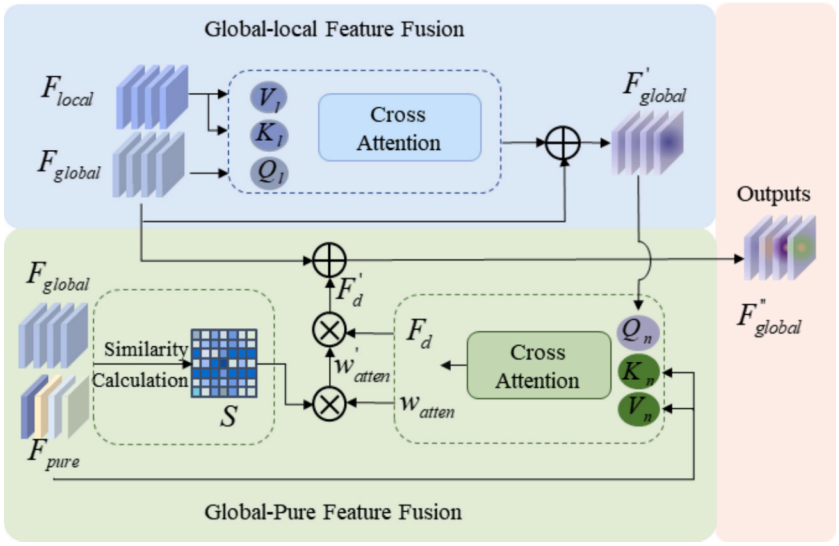


Fig. 6. Similarity-guided dual cross-attention fusion module (DCAF).

Fig. 6 illustrates how global and local features are fused by computing attention affinities between the global feature F_{global} and the local feature F_{local}. First, both features are flattened along their spatial dimensions. Each flattened feature is then projected into its own subspace: the global feature is linearly transformed by W_l^Q to produce the query vector Q_l, while the local feature is transformed by W_l^K and W_l^V to produce the key vector K_l and the value vector V_l, respectively. These projections are defined as follows:

$$Q_l = F_{global}W_l^Q, K_l = F_{local}W_l^K, V_l = F_{local}W_l^V \quad (1)$$

Table 2
Definition for every evaluation metric.

Type	Metrics	Formulation
Purified-scene	Over Accuracy (OA)	$\frac{\sum_{i=1}^{M}(TP_i+TN_i)}{\sum_{i=1}^{M}(TP_i+TN_i+FP_i+FN_i)}$
	Over Precision, (OP)	$\frac{\sum_{i=1}^{M}TP_i}{\sum_{i=1}^{M}(TP_i+FP_i)}$
Mixed-scene	Mean Absolute Error, (MAE)	$\frac{1}{M}\sum_{i=1}^{M}\lvert p_{true}^{(i)}-p_{pred}^{(i)}\rvert$
	Weighted Difference, (WD)	$\frac{1}{M}\sum_{i=1}^{M}p_{true}^{(i)}\lvert p_{true}^{(i)}-p_{pred}^{(i)}\rvert$
	Natural MAE, (MAE_N)	$MAE_N=\frac{1}{M_N}\sum_{i=1}^{M_N}\lvert p_{N,true}^{(i)}-p_{N,pred}^{(i)}\rvert$
	Built-up MAE, (MAE_B)	$MAE_B=\frac{1}{M_B}\sum_{i=1}^{M_B}\lvert p_{B,true}^{(i)}-p_{B,pred}^{(i)}\rvert$

Notes: M denotes the total number of LCZ mapping classes, M_N denotes the total number of natural classes (LCZ A–G); M_B denotes the total number of built-up classes (LCZ 1–10); TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively; p indicates the proportion.

Table 3
Prediction accuracies of MSU-Net in two study areas.

Study area	OA	OP	MAE	WD	MAE_B	MAE_N
Wuhan	0.835	0.852	0.0544	0.0187	0.0890	0.0049
Shenzhen	0.800	0.808	0.0655	0.0182	0.0733	0.0217

The similarity between Q_l and K_l is then calculated to generate a spatial correspondence map linking global and local features. Peaks in this correspondence map indicate critical alignment points between the two feature sets. Softmax-normalized similarity scores weight the value vectors V_l. This weighting performs a soft selection in feature space, amplifying strongly responding vectors and suppressing redundant information. To preserve the original global context, a residual connection is added to maintain the initial feature distribution. The resulting locally enhanced global feature F'_{global} is therefore given by:

$$F'_{global}=F_{global}+r\cdot Conv\left(\widetilde{V}\right),\widetilde{V}=Softmax\left(\frac{Q_lK_l^T}{\sqrt{d}}\right)\cdot V_l \tag{2}$$

where r dynamically adjusts the intensity of cross-layer information flow, and d scales channel dimensions to stabilize gradients. The incorporation of the purified LCZ feature F_{pure} as a semantic prior enhances the model's capacity to resolve fragmented semantic cues within mixed scenes. A semantic similarity matrix S is computed between the global feature F_{global} and the purified feature F_{pure} to quantify inter-class association strengths. This computation is given by:

$$S=Softmax\left(\frac{F_{pure}\cdot F_{global}}{\left\|F_{pure}\right\|_2\cdot\left\|F_{global}\right\|_2}\right) \tag{3}$$

Subsequently, the spatially enhanced global feature is fused with the high-purity semantic prior through a cross-attention mechanism to achieve complementary refinement. First, both the enhanced global feature F'_{global} and the purified feature F_{pure} are flattened along their spatial dimensions. Each flattened feature is then projected into its own subspace: the enhanced global feature is linearly transformed by W_p^Q to produce the query vector Q_p, while the purified feature is transformed by $W_p^K$and W_p^V to produce the key vector K_p and the value vector V_p respectively. These projections are defined as:

$$Q_p=F'_{global}W_p^Q,K_p=F_{pure}W_p^K,V_p=F_{pure}W_p^V \tag{4}$$

The similarity between Q_p and K_p is then calculated to produce a spatial correspondence map between the enhanced global feature and the purified semantic prior:

$$w_{atten}=Softmax((Q_pK_p^T)/\sqrt{d})) \tag{5}$$

where w_{atten} are the cross-attention weights. The resulting cross-attention weights, denoted w_{atten}, are refined by element-wise multiplication with the semantic similarity matrix S. This refinement ensures that only global responses with strong alignment to the purified semantics are retained. To preserve the original global context and maintain stable gradient flow, a residual connection is added to the initial global feature:

$$F''_{global}=F_{global}+(w_{atten}\times S)\times w_{atten}.V_p \tag{6}$$

3.3. Model training

3.3.1. Multi-task learning

MSU-Net is trained on three complementary sub-tasks: global mixed-scene decomposition, sparse-scene decomposition, and purified-image classification. Cross-entropy loss is employed for the global decomposition and the purified-image classification, while the local branch uses an $L_{1/2}$ sparsity-regularized reconstruction loss. Task weights are automatically balanced using the GradNorm strategy (Chen et al., 2018), which scales each task's gradients so that all tasks contribute equally to the overall gradient magnitude. Optimization is performed using the Adam algorithm with an initial learning rate of $1\times 1e^{-4}$, $\beta_1=0.9$, $\beta_2=0.999$, a weight-decay factor of $1\times 1e^{-2}$, and a batch size of 20. A learning-rate warmup is implemented via timm's CosineLRScheduler (Oba et al., 2025) to prevent early gradient explosion. The model was implemented in Python 3.7 using PyTorch 1.11 on a Windows 10 system, with computations accelerated on a CUDA 11.6 platform.

We further evaluated the efficiency of MSU-Net. Training one epoch on the Wuhan dataset took approximately 1,898 s, while training on the Shenzhen dataset took about 4,504 s. During inference, processing a single sample required approximately 0.30 s.

3.3.2. Accuracy assessment

MSU-Net's performance is evaluated from two complementary perspectives. In purified-scene evaluation, Overall Accuracy (OA) and Overall Precision (OP) follow standard LCZ classification protocols (Liu and Shi, 2020; Qiu et al., 2018). Each pure-class scene is processed by MSU-Net, and the class with the highest predicted proportion is compared to the reference label (Zhang and Du, 2015), with matches counted as correct. In mixed-scene decomposition evaluation, the model produced proportional shares for every LCZ class within each scene. Accuracy is measured by four metrics: Mean Absolute Error (MAE), defined as the average absolute difference between predicted and true class proportions; Weighted Difference (WD), calculated as the proportionally weighted average of these absolute errors to reduce the impact of outlier classes; Built-up MAE (MAE_B) and Natural MAE (MAE_N) measure detailed assessment of built-up and natural classes. All metrics are defined in Table 2.

4. Result

4.1. General result

4.1.1. LCZ mapping accuracy using MSU-Net

Table 3 shows that OA and OP exceed 0.8 in both Wuhan and Shenzhen, demonstrating MSU-Net's effectiveness in purified-scene classification. MAE remains below 0.07, indicating accurate LCZ block decomposition. MAE_B and MAE_N show that errors in built-up classes are slightly greater in Wuhan, whereas errors in natural classes are markedly higher in Shenzhen (0.0217 versus 0.0049). The differences in MAE_N between Shenzhen and Wuhan likely stem from urban morphology and landscape composition. Wuhan's extensive water bodies and vegetation create more homogeneous natural patches, whereas Shenzhen's dense, fragmented natural areas are harder to classify. Despite different overall MAE, the WD are nearly identical

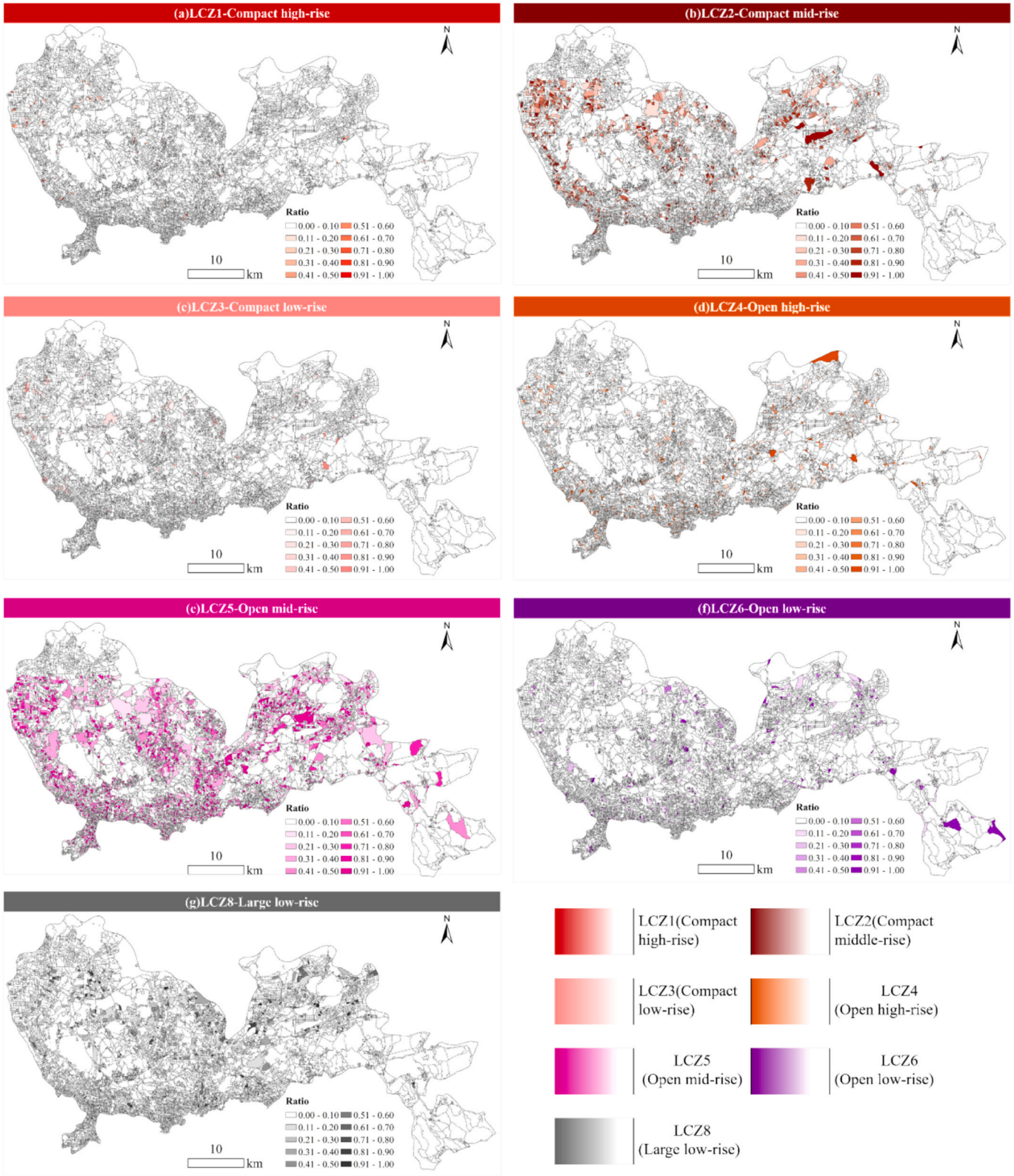


Fig. 7. Shenzhen's built-type LCZ classes as decomposed by MSU-Net. The "ratio" refers to the proportion of a specific pure LCZ type within the entire block area. A darker color indicates a higher ratio.

(Wuhan: 0.0182; Shenzhen: 0.0187). This finding indicates that predictive errors are more tightly clustered in Shenzhen, while larger deviations occur in specific LCZ categories in Wuhan, necessitating detailed per-class investigation.

4.1.2. *Visual LCZ mapping result by MSU-Net*

Figs. 7–10 illustrate LCZ maps of built-up and natural classes in Shenzhen and Wuhan. Color intensity indicates class proportion within each block, while the original block boundaries and geographic locations are retained. Detailed insets of two local regions in Shenzhen

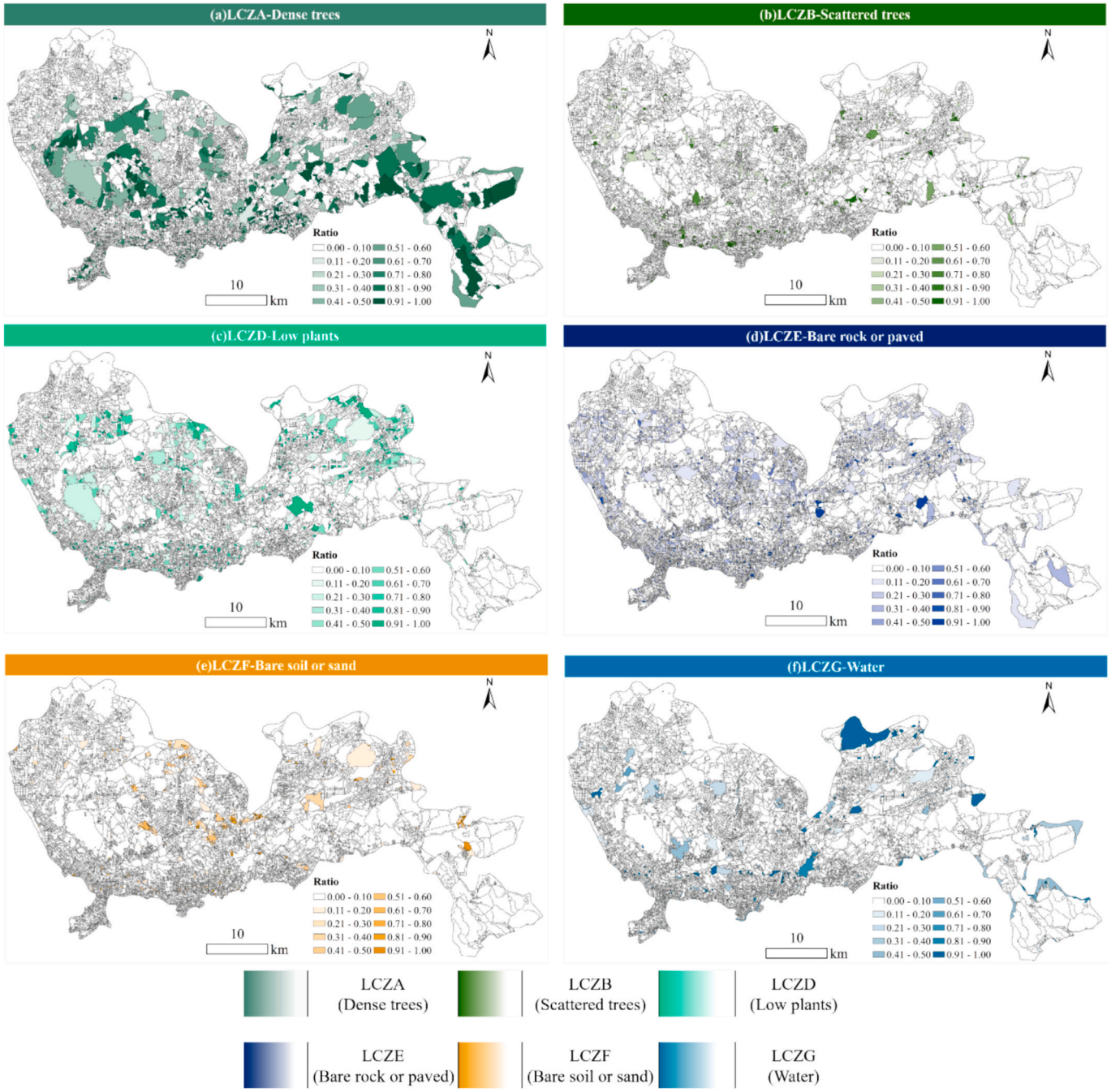


Fig. 8. Shenzhen's natural LCZ classes as decomposed by MSU-Net. The "ratio" refers to the proportion of a specific pure LCZ type within the entire block area. A darker color indicates a higher ratio.

(Figs. 11–12) and Wuhan (Figs. 13–14) are provided, where color saturation encodes class proportion and distinct hues represent LCZ categories.

In both cities, the mixed-scene decomposition reveals characteristic spatial patterns of built-up classes. In Shenzhen, LCZ 1–2 (Figs. 7a and 7b) cluster in core functional zones like central business district and technology park while LCZ 3(Fig. 7a) occur as scattered patches (0.5–0.8 proportion) in old districts and urban villages. Overall, a concentric development pattern emerges with a high-rise core surrounded by progressively lower-rise zones, confirming the ring-layered expansion typical of high-density cities. In Wuhan, a multi-core structure appears. LCZ 1 (Fig. 9a) predominates in core business districts and along major riverside development axes such as the "Two Rivers, Four Banks" corridor, where its share often exceeds 0.7. LCZ 3 (Fig. 9c) across the downtown and historic preservation zones; areas with proportions of 0.8 to 1.0 correspond closely to officially protected sites. LCZ 1–3 (Fig. 9a–c) reaches 63 % in the high-density interval of 0.5 to 0.8, which is 22 % higher than LCZ 4–6 (Fig. 9d–f), characterizing Wuhan's infill-oriented spatial form. LCZ 8 (Fig. 9h, 0.6–0.9 in industrial parks and 0.2 to 0.4 in logistics areas) display a bimodal distribution, indicating MSU-Net's ability to distinguish between production-oriented and storage-oriented building typologies.

Natural classes show pronounced spatial distinctions. In Shenzhen, LCZ A (Fig. 8a) and LCZ B (Fig. 8b) cluster in hilly suburbs. LCZ D (Fig. 8c) and LCZ F (Fig. 8e) occur along coastal zones and developing districts, indicating construction sites or awaiting development. LCZ E (Fig. 8d) aligns with eastern volcanic belts, while LCZ G (Fig. 8f) form high-proportion corridors (0.41–1.00) along rivers and reservoirs.

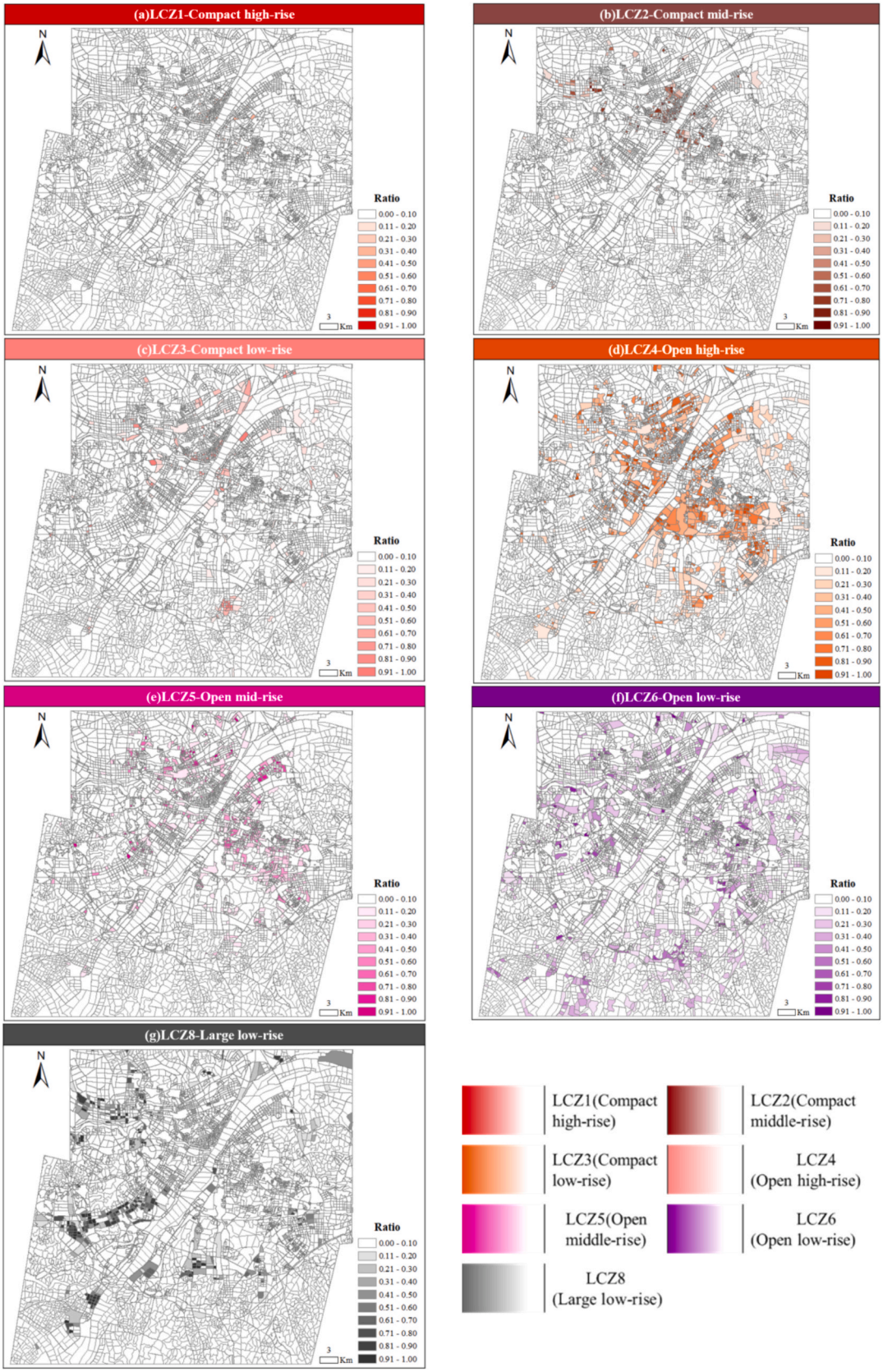


Fig. 9. Wuhan's built-type LCZ classes as decomposed by MSU-Net. The "ratio" refers to the proportion of a specific pure LCZ type within the entire block area. A darker color indicates a higher ratio.

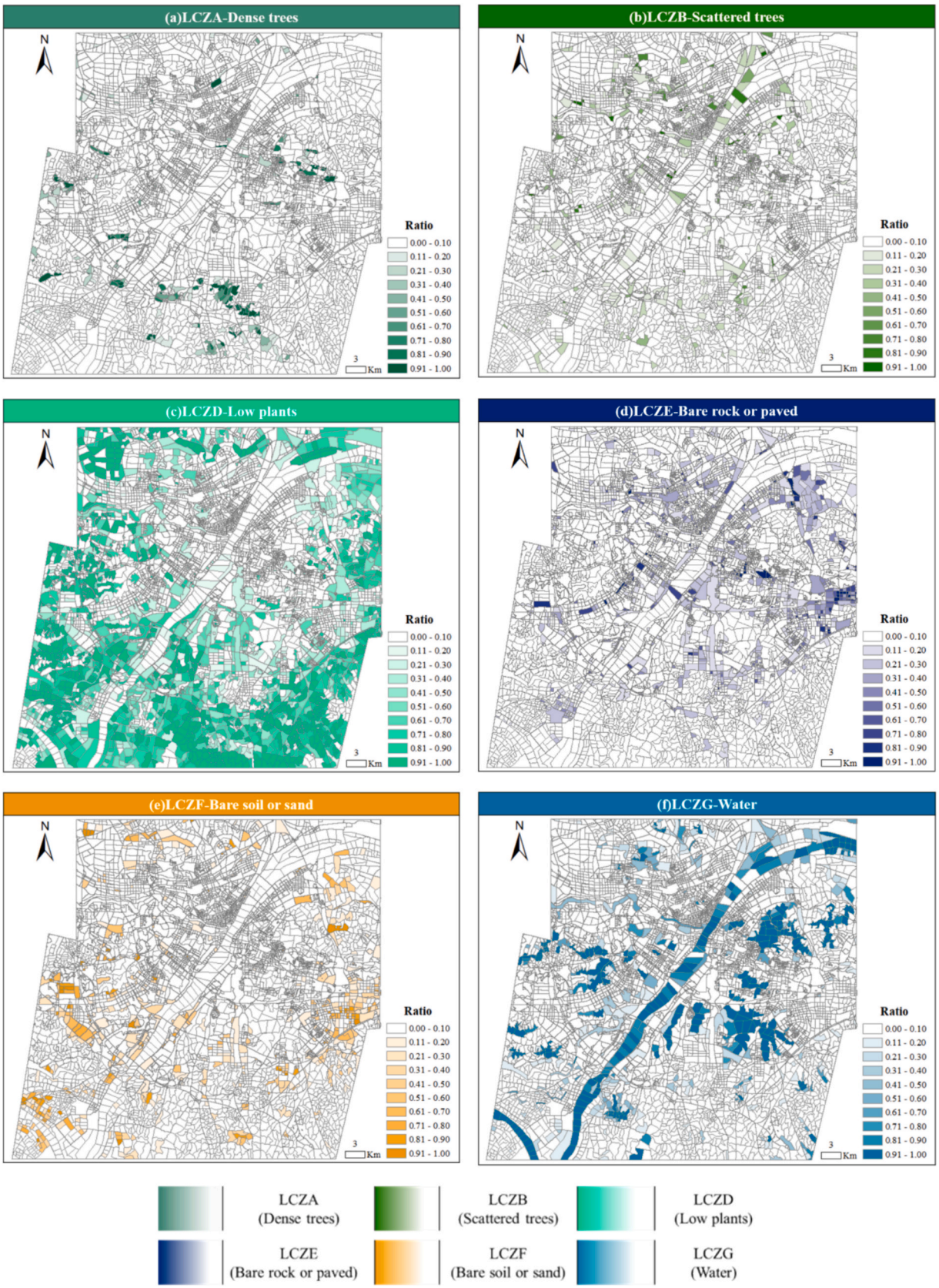


Fig. 10. Wuhan's natural LCZ classes as decomposed by MSU-Net. The "ratio" refers to the proportion of a specific pure LCZ type within the entire block area. A darker color indicates a higher ratio.

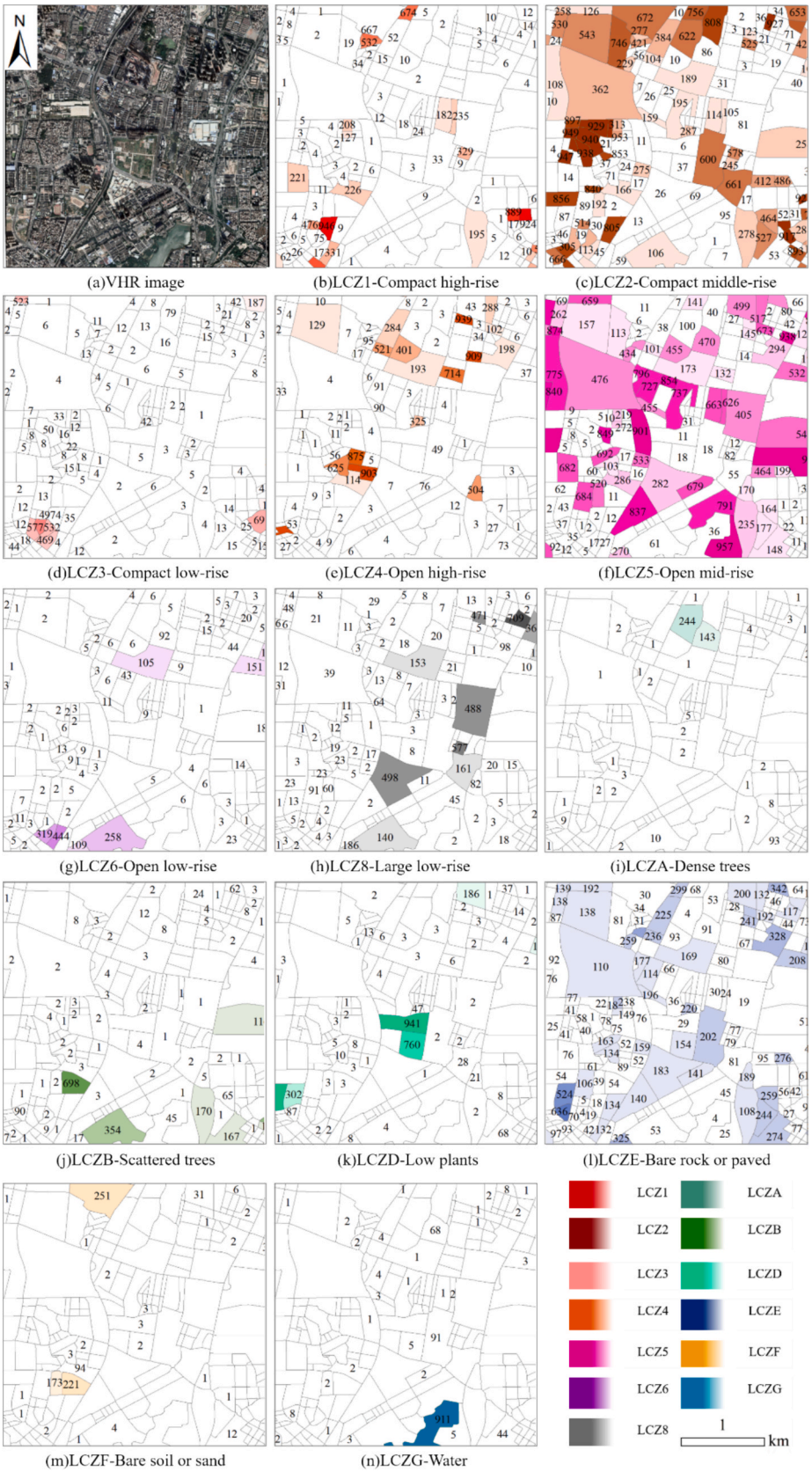


Fig. 11. Visualization of unmixing results for region 1 in Shenzhen. Numeric labels are shown to three decimal places, with values below 0.001 omitted. All values are multiplied by 1000 for integer presentation. Image data © Google.

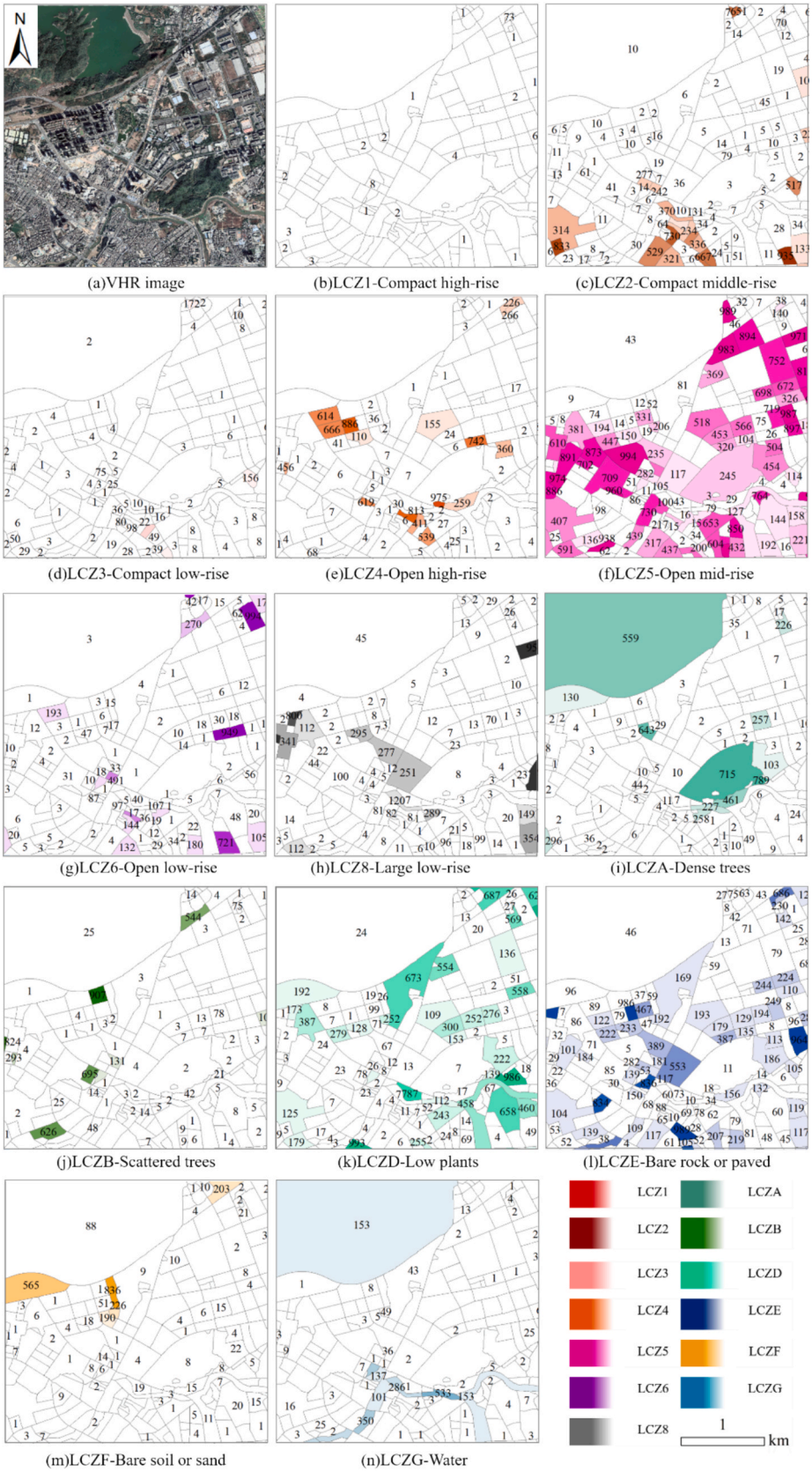


Fig. 12. Visualization of unmixing results for region 2 in Shenzhen. Numeric labels are shown to three decimal places, with values below 0.001 omitted. All values are multiplied by 1000 for integer presentation. Image data © Google.

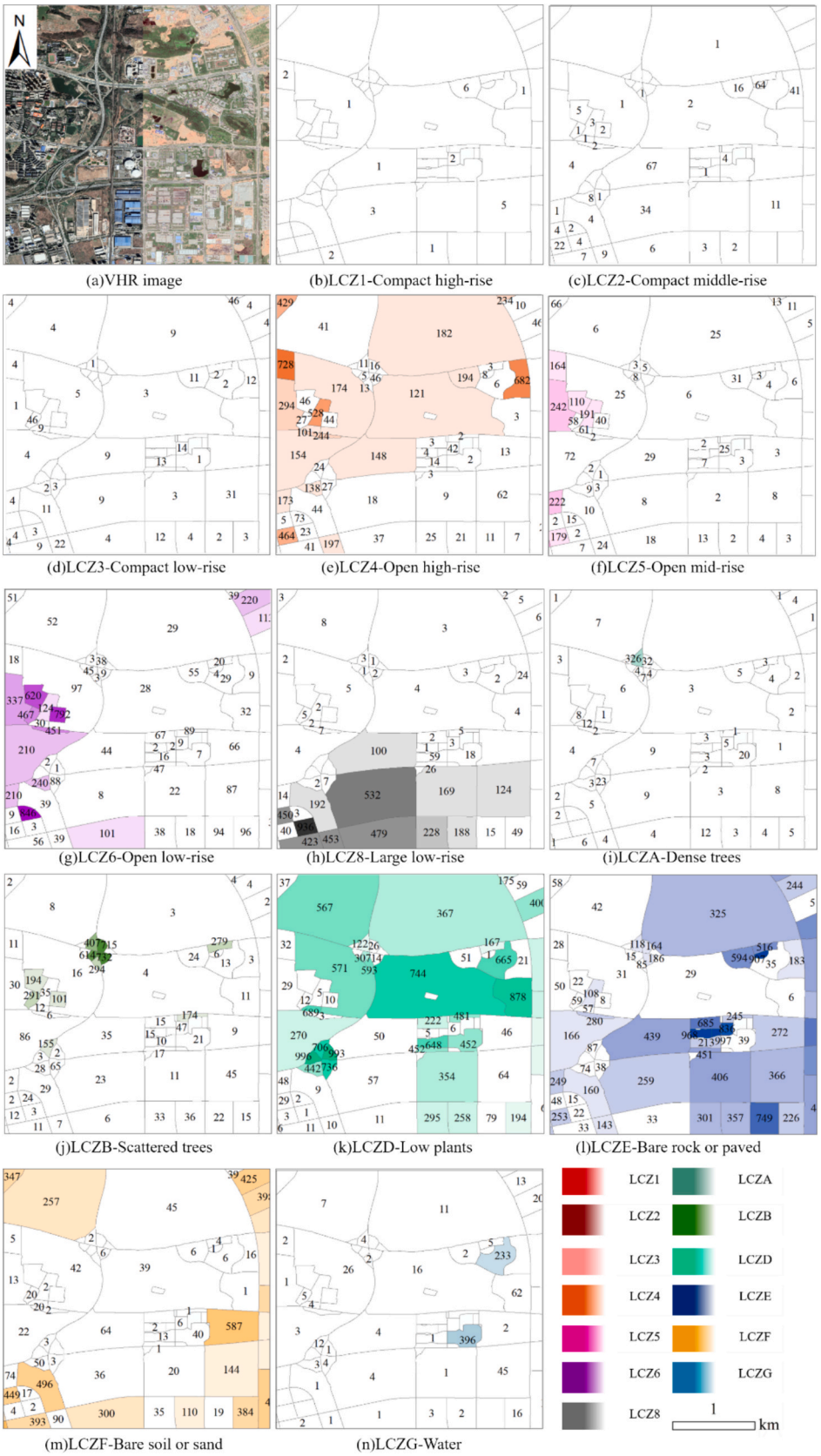


Fig. 13. Visualization of unmixing results for region 1 in Wuhan. Numeric labels are shown to three decimal places, with values below 0.001 omitted. All values are multiplied by 1000 for integer presentation. Image data © Google.

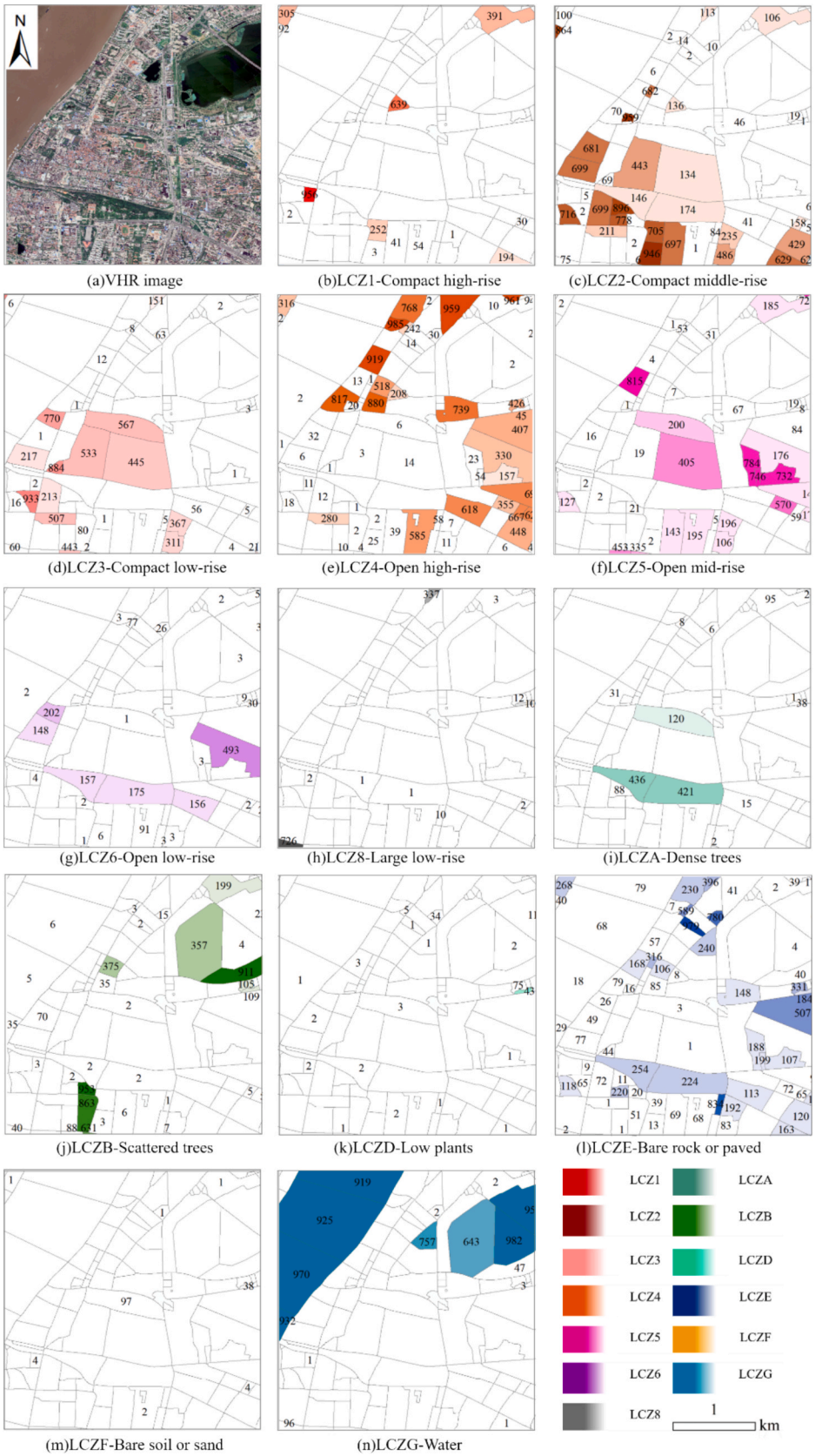


Fig. 14. Visualization of unmixing results for region 2 in Wuhan. Numeric labels are shown to three decimal places, with values below 0.001 omitted. All values are multiplied by 1000 for integer presentation. Image data © Google.

Table 4
Comparative evaluation of LCZ mapping methods on Shenzhen.

Method	OA	OP	MAE	WD	MAE_B	MAE_N
WUDAPT	0.683	0.685	0.1042	0.0435	0.1434	0.0652
LCZ-Net	0.775	0.771	0.0821	0.0265	0.0896	0.0219
Sen2lcz-Net	0.763	0.751	0.0882	0.0278	0.0894	0.0412
LCZ-CNN	0.756	0.762	0.0845	0.0314	0.0935	0.0476
MCFUNet-LCZ	0.723	0.751	0.0700	0.0206	0.0657	0.0751
MSU-Net	0.800	0.808	0.0655	0.0182	0.0733	0.0217

Table 5
Comparative evaluation of LCZ mapping methods on Wuhan.

Method	OA	OP	MAE	WD	MAEB	MAEN
WUDAPT	0.692	0.685	0.1042	0.0422	0.1526	0.0569
LCZ-Net	0.798	0.796	0.0746	0.0271	0.1032	0.0265
LCZ-CNN	0.758	0.761	0.0810	0.0336	0.1245	0.0357
Sen2lcz-Net	0.789	0.782	0.0782	0.0282	0.1112	0.0208
MCFUNet-LCZ	0.766	0.751	0.0656	0.0191	0.0584	0.0741
MSU-Net	0.835	0.852	0.0544	0.0187	0.0890	0.0049

In Wuhan, LCZ A (Fig. 10a) concentrate in the east. LCZ B (Fig. 10b) form continuous ecological transition belts along the urban-fringe and Third Ring Road. LCZ D (Fig. 10c) is more prevalent in the west and south. LCZ E (Fig. 10d) matches 91 % of highest-value hilly exposures (value over 0.9). LCZ G (Fig. 10f) align into high-proportion belts (0.8–1.0) along major waterways, with surrounding lakes exhibiting gradual decay, illustrating the organizing influence of water systems on urban form.

As highlighted in the local visualization, the model achieves high accuracy in capturing diverse building morphologies and natural features across both cities. The unmixing results faithfully represent spatial overlaps and transitions among LCZ types, providing a detailed basis for fine-scale urban morphology analysis. Nevertheless, a few failure modes remain: in mixed building blocks the model can confuse high-, mid-, and low-rise classes (Fig. 11); in areas where low vegetation borders water, small water patches may be missed (Fig. 12); along low-vegetation–bare-soil interfaces, the estimated fractions can be biased (Fig. 13); and in areas where compact and open buildings coexist, the model occasionally blends the two classes (Fig. 14).

In summary, Shenzhen follows a concentric pattern of mid- to high-rise aggregation surrounded by a low-rise gradient, whereas Wuhan exhibits a multi-core concentric pattern in which built form and natural land cover coexist. The mixed-scene decomposition network accurately unmixes complex scenarios such as interleaved LCZ 3 and LCZ D (Figs. 7c and 9c), providing precise, fine-scale data for high-resolution urban climate research.

4.2. Comparing with current LCZ methods

Given the absence of methods specifically designed for large-scale decomposing mixed LCZ blocks, five representative LCZ mapping algorithms are selected as benchmarks: WUDAPT (Ching et al., 2018), LCZ-Net(Liu and Shi, 2020), LCZ-CNN(Huang et al., 2021), Sen2LCZ-Net (Qiu et al., 2020), and MCFUNet-LCZ (Ji et al., 2023). WUDAPT is a pixel-level approach, averaging spectral features within blocks and classifying them via random forest. LCZ-Net adopts a thirteen-layer convolutional backbone with residual blocks (He et al., 2016) and squeeze-and-excitation modules (Hu et al., 2020), using global average pooling with channel weighting to reinforce spatial detail. Sen2LCZ-Net extracts multi-level features through four convolutional layers, fuses them in a dense layer, and outputs class scores via a fully connected layer. LCZ-CNN replaces its final fully connected layer with global average pooling after three convolutional stages to produce class probabilities. MCFUNet-LCZ proposed a multiscale feature fusion architecture composed of three modules, namely dual-pooling fusion, coding-decoding fusion, and global probability fusion, to enhance feature representation and improve LCZ classification accuracy.

To ensure fairness, all methods use the same OSM-derived block-level dataset. In WUDAPT, block-wise mean spectral features are classified via random forest. Other deep-learning methods directly input block images for mixed-scene decomposition. As no dedicated unmixing methods currently exist for LCZ mapping, existing classification models are repurposed as decomposition backbones, with SoftMax outputs interpreted as class proportions.

As shown in Table 4, performance in Shenzhen varies significantly. WUDAPT performs worst (OA = 0.683, MAE = 0.1042, MAE_B = 0.1434), indicating RF's limitations for mixed-scene decomposition in high-density urban blocks. Three deep-learning methods (LCZ-Net, Sen2LCZ-Net, LCZ-CNN) improve OA by average 11.7 % and reduce MAE_B by average 37.8 %, highlighting their superior spatial modeling. MSU-Net achieves the best results in the core metrics (OA = 0.800, MAE = 0.0655, WD = 0.0182), and reduces MAE_N by 66.7 % relative to WUDAPT (0.0217), validating its strength in natural class decomposition. For built-up classes, MSU-Net also outperforms all other methods.

Table 5 shows similar trends in Wuhan, with some regional variation. WUDAPT's built-up MAE further increases to 0.1526, confirming that traditional classifiers struggle with the irregular building layouts of older urban districts. In Wuhan, MSU-Net attains an OA of 0.835, which is 4.4 % higher than its Shenzhen result, and MAE_N (MAE_N = 0.0049) achieves a highly accurate decomposition of natural LCZ types.

Mixed-scene decomposition results for Wuhan (Figs. 15 and 16) show that MSU-Net demonstrates superior performance.

MSU-Net clearly outlines high-proportion built-up classes (LCZ 2–6) with preserved block structure and boundaries, while enhancing natural class (LCZ D and E) discrimination and ensuring smooth transitions without overestimation or misclassification. In contrast, WUDAPT yields coarse and low-detail results. LCZ-Net and LCZ-CNN perform moderately in mid-proportion regions but under-respond in high-proportion zones. Sen2LCZ-Net detects some high-proportion regions but introduces many false positives. Overall, MSU-Net outperforms all methods in spatial accuracy, class fidelity, and mixed-scene representation, proving most effective for mapping urban surface-cover proportions.

4.3. Ablation studies

An ablation study on the Wuhan dataset evaluates the contributions of the Purified-image Semantic Guidance (PSG) branch, Local Spatial Guidance (LSG) branch, and Dual Cross-Attention Fusion (DCAF) module (Table 6). Adding the PSG branch alone raises OA from 0.803 to 0.825, showing global priors aid fragmented detection. Including the LSG branch further reduces MAE by 17.38 % and built-up MAE by 15.96 %, indicating local spatial features improve boundary delineation. Integrating DCAF lowers the weighted difference to 0.0187, i.e., a 7.88 % drop versus simple concatenation of all three branches, confirming that the dual attention mechanism strengthens spatial coherence and semantic alignment. Together, these modules improve mixed-scene unmixing via global-semantic guidance, local refinement, and focused attention.

To further illustrate the effectiveness of the DCAF module, Fig. 17 compares the CAM responses and classification results with and without DCAF. Fig. 17 shows that, without the DCAF module, the model generates highly similar CAM responses for LCZ2, LCZ3, LCZ5, and LCZ6, revealing limited discriminative capability among these classes. After incorporating the DCAF module, the activation maps exhibit greater differentiation among LCZ categories. The results clearly demonstrate that incorporating DCAF significantly improves the model's ability to decompose mixed scenes and to distinguish between different LCZ categories.

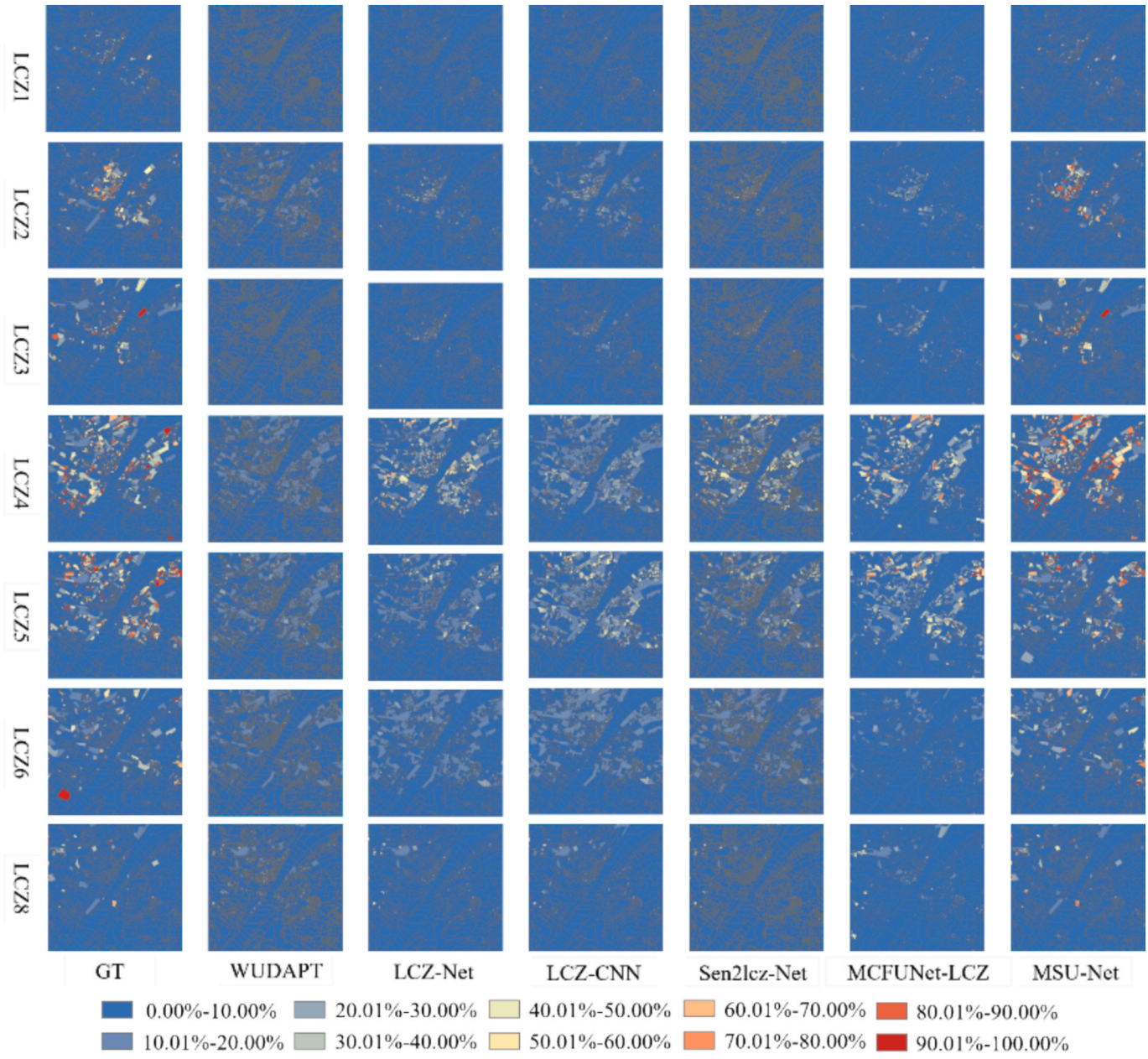


Fig. 15. Comparative visualization of built-type LCZ decomposition in Wuhan using various methods.

5. Discussion

5.1. Impact of high-resolution and multi-spectral imagery on LCZ mapping

Previous studies typically use medium- and low-resolution imagery (e.g., Landsat 30 m, Sentinel-2 10 m) for LCZ mapping, as LCZ zones span several hundred meters to a few kilometers. For example, WUDAPT resamples 30 m Landsat data to 100 m and performs pixel-wise classification. In this study, the impact of both spatial resolution and spectral richness is assessed via mixed-scene decomposition experiments on the Wuhan dataset. Five resolutions and spectral configurations are tested: 30 m (Landsat), 10 m with 3 bands (Sentinel-2 RGB), 10 m with 10 bands (the 10 m and 20 m Sentinel-2 bands), 5 m (resampled Google Earth), and 1 m (this study's source data). Results (Table 7) indicate that:

1. When resolution is increased from 30 m to 1 m within the WUDAPT framework, OA rises from 0.546 to 0.683 (+25.1 %) and MAE falls from 0.1254 to 0.1042 (–16.9 %). MAE_B decreases by 48 %, from 0.2759 to 0.1434, demonstrating that high-resolution imagery enables more precise delineation of complex building outlines
2. With 1 m inputs, MSU-Net yields an OA of 0.800, i.e., a further 11.3 % gain over 30 m inputs, and reduces MAE_N from 0.0677 to 0.0217 (–68 %), showing that deep learning fully exploits the textural richness of high-resolution data to enhance LCZ mapping accuracy
3. For both WUDAPT and our model, including 10 Sentinel-2 bands notably improves accuracy, as richer spectral information enhances the model's ability to estimate building fractions more precisely

5.2. Visualization of LCZ decomposition in urban blocks

Fig. 18 presents three representative examples. The first example depicts a block where compact and open buildings are interwoven. The model successfully distinguishes between these two urban forms,

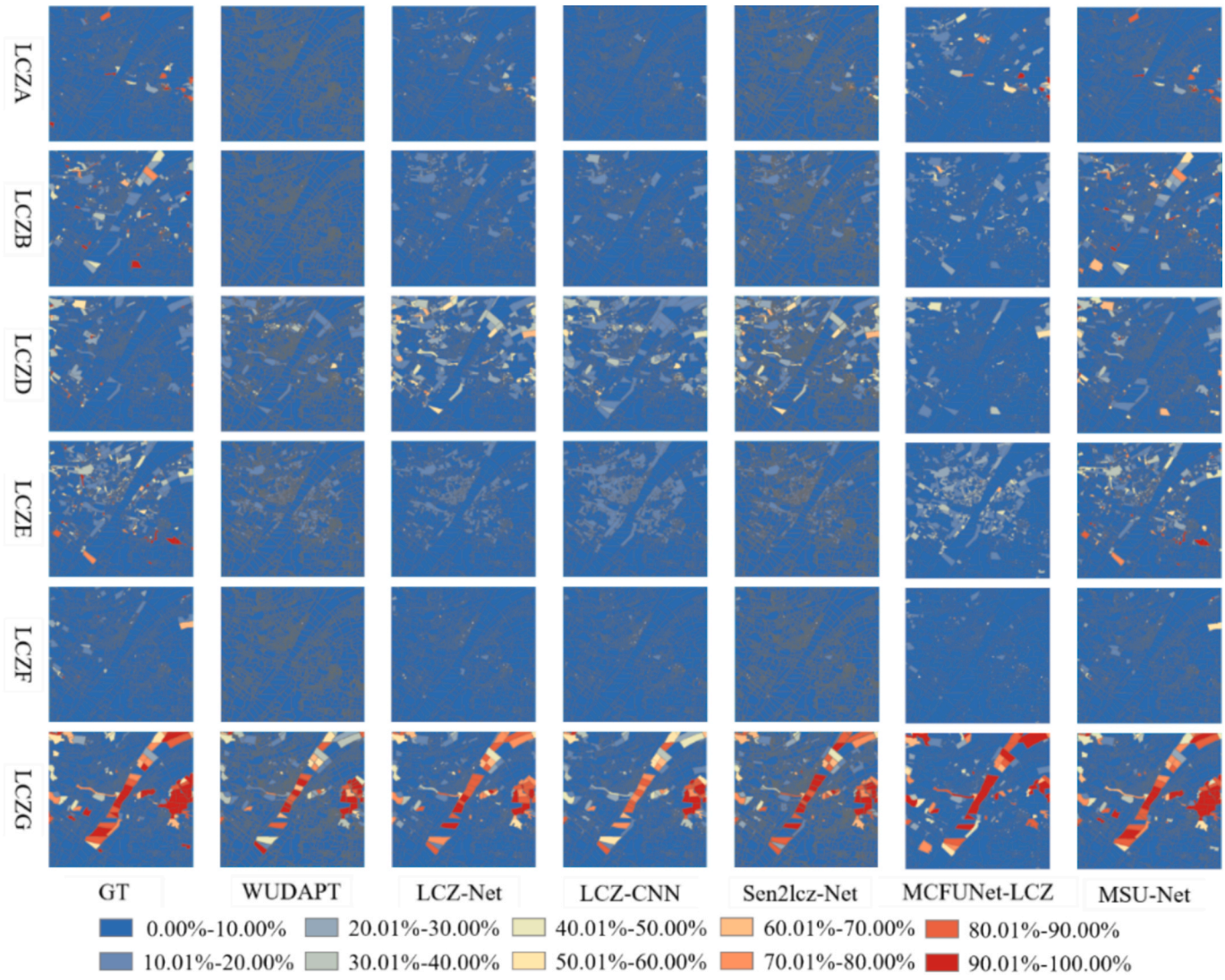


Fig. 16. Comparative visualization of natural LCZ decomposition in Wuhan using various methods.

Table 6
Ablation experiments on Wuhan.

	OA	OP	MAE	WD	MAE_B	MAE_N
Baseline	0.803	0.812	0.0725	0.0254	0.1103	0.0146
+ PSG	0.825	0.835	0.0721	0.0232	0.1101	0.0090
+ PSG + LSG	0.831	0.848	0.0599	0.0203	0.0927	0.0067
+ PSG + LSG + DCAF	0.835	0.852	0.0544	0.0187	0.0890	0.0049

although slight deviations remain in the estimated proportions. The second example shows an area combining high-rise, mid-rise, and low-rise buildings. The model accurately identifies the dominant building categories, while minor inaccuracies persist in their fractional decomposition. The third example corresponds to a scene where impervious surfaces are mixed with LCZB. The model demonstrates strong discriminative capability between these two classes, though a very small fraction of LCZB is misclassified as LCZD.

5.3. Generalization performance analysis

Given regional variation in urban morphology, transfer learning is essential for adapting LCZ models to new cities. To evaluate MSU-Net's generalization, Shenzhen is used as the source domain and Wuhan as the target. Two baseline scenarios are defined: direct inference on Wuhan (Table 8, first row), direct inference on Shenzhen (Table 8, second row). The experimental scenario involves fine-tuning the Shenzhen-trained model on limited Wuhan samples before inference (Table 8, last row).

Applying the Shenzhen-trained model to Wuhan directly yields only slight accuracy drops, with MAE_N within 0.01 and MAE_B differing by 0.03–0.05. Fine-tuning on Wuhan samples boosts OA to 0.831, reduces MAE_B by 38.3 % relative to direct Shenzhen inference, and lowers WD by 45.1 %. These findings confirm MSU-Net's strong cross-city robustness, particularly for natural land-cover identification.

6. Conclusion

This study improves LCZ mapping under urban heat-island conditions by overcoming the classification accuracy and cross-city transferability limitations of traditional methods. An innovative solution has been proposed that integrates high-resolution remote sensing imagery with deep learning techniques. For the first time, the concept of "mixed-scene decomposition" has been introduced, in which the irregular block morphology from OSM data and the rich spatial detail of imagery are combined to construct an application-oriented, high-resolution LCZ mapping pipeline.

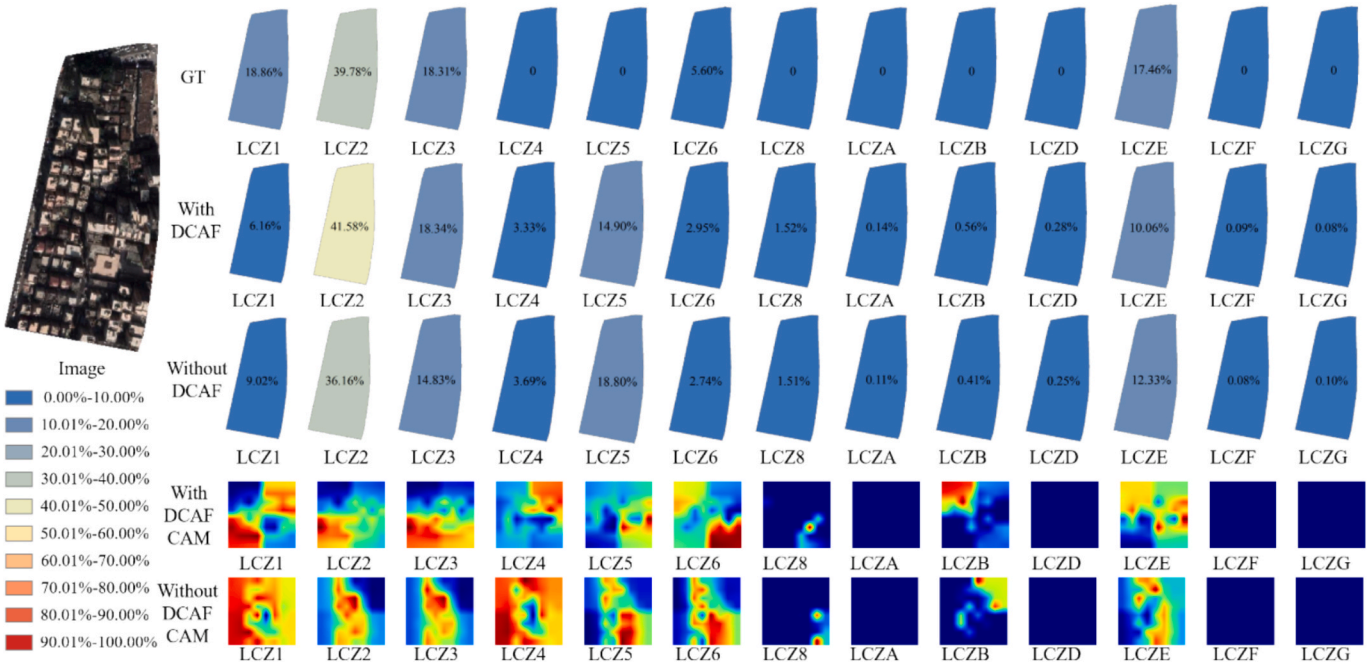


Fig. 17. Comparison of model results and corresponding CAM visualizations with and without the DCAF module.

Table 7
Comparison between different image sources using WUDAPT and MSU-Net on Wuhan.

Method	Spatial resolution (image source)	OA	OP	MAE	WD	MAE_B	MAE_N
WUDAPT	30 m(Landsat)	0.546	0.578	0.1254	0.0692	0.2759	0.0989
	10 m(Sentinel-2)	0.637	0.632	0.1187	0.0578	0.2165	0.0848
	10 m(Sentinel-2 10bands)	0.736	0.737	0.0743	0.0218	0.0812	0.0662
	5 m(Google earth)	0.677	0.681	0.1064	0.0576	0.1823	0.0712
	1 m (Google earth)	0.683	0.685	0.1042	0.0435	0.1434	0.0652
MSU-Net	30 m(Landsat)	0.719	0.723	0.0942	0.0359	0.1753	0.0677
	10 m(Sentinel-2)	0.755	0.764	0.0903	0.0345	0.1376	0.0629
	10 m(Sentinel-2 10bands)	0.806	0.793	0.0714	0.0216	0.0639	0.0778
	5 m (Google earth)	0.788	0.774	0.0899	0.0255	0.1123	0.0486
	1 m (Google earth)	0.800	0.808	0.0655	0.0182	0.0733	0.0217

To improve recognition in heterogeneous urban blocks, MSU-Net introduces a dual semantic–spatial guidance mechanism that enhances unmixing precision and stability. The DCAF module enables effective interaction and reinforcement between global–local and mixed–purified feature, boosting model expressiveness in complex scenes.

To validate the method, the first annotated dataset for mixed LCZ unmixing was constructed, and systematic experiments in Shenzhen and Wuhan confirmed the superiority of the proposed approach in accuracy, adaptability, and decomposition capability. Additional tests across 30 m to 1 m imagery and across cities with distinct spatial layouts confirmed its robustness to resolution and urban form. These results highlight its strong generalization and practical value, offering more interpretable, application-ready LCZ data to support urban climate modeling, green planning, and fine-scale governance.

However, this study has several limitations. First, in highly mixed and complex blocks, dense building distributions and shadows from tall buildings may interfere with the model, leading to potential confusion in building type identification. In future work, we plan to introduce multi-temporal imagery to reduce shadow effects and improve robustness under changing illumination. Second, the study relies on OSM data to provide realistic, irregular LCZ block morphologies. However, as OSM is crowdsourced and lacks standardized data quality control, some accuracy loss may occur. To overcome this, we intend to employ automatic road-network semantic segmentation to generate block boundaries, and integrate these boundary cues into a multi-task end-to-end network alongside classification, ensuring that boundary generation and LCZ classification mutually assist each other. Despite these limitations, the study provides a valuable framework for LCZ mapping and lays the groundwork for future improvements.

CRediT authorship contribution statement

Jiayi Li: Writing – original draft, Validation, Investigation. **Xinji Tian:** Writing – review & editing, Visualization, Validation, Data curation. **Wenrui Wang:** Writing – review & editing, Methodology, Conceptualization. **Lilin Tu:** Writing – review & editing, Validation, Supervision. **Yang Lu:** Visualization, Validation. **Jie Jiang:** Writing – review & editing, Supervision, Resources. **Xin Huang:** Supervision, Resources, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

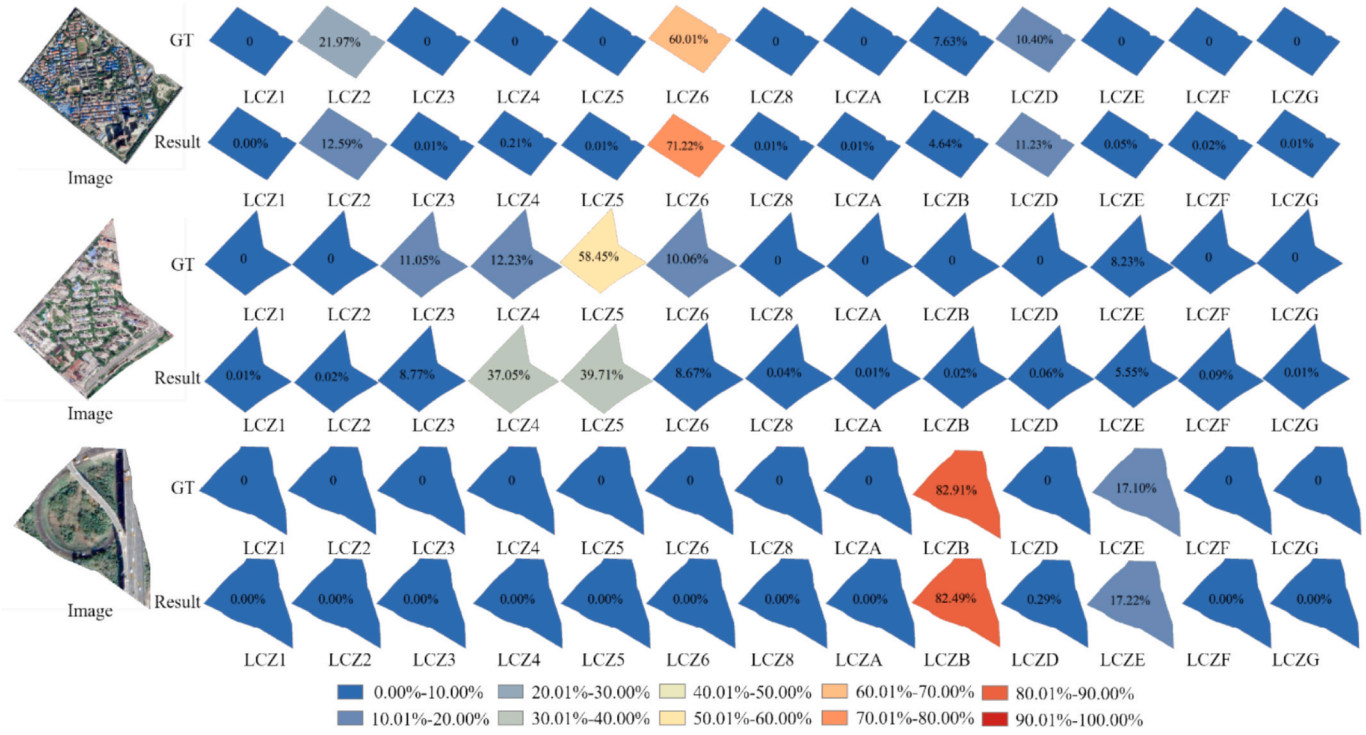


Fig. 18. The visual results of the MSU-Net.

Table 8
Results of transfer learning on datasets from different cities.

Source	Target	Transfer setting	OA	OP	MAE	WD	MAE_B	MAE_N
Wuhan	Wuhan	Direct prediction	0.835	0.852	0.0544	0.0187	0.0890	0.0049
Shenzhen	Wuhan	Direct prediction	0.828	0.829	0.0672	0.0213	0.0957	0.0079
		target fine-tuning	0.831	0.832	0.0550	0.0117	0.0591	0.0068

Acknowledgments

This work was supported in part by the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resource (under Grant KF-2023-08-08), and in part by the National Natural Science Foundation of China (under Grant 42471391 and 42271328), and in part by National Key Research and Development Program of China (under Grant 2024YFF1306102), the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources (under Grant KF-2023-08-08).

Data availability

The source code used in this study are openly available at https://github.com/wangqiaoqiaoqiao/LCZ-mapping. The dataset can be accessed from https://doi.org/10.6084/m9.figshare.30291571.v1.

References

Adegun, A.A., Viriri, S., Tapamo, J.R., 2023. Review of deep learning methods for remote sensing satellite images classification: experimental survey and comparative analysis. J. Big Data. 10. https://doi.org/10.1186/s40537-023-00772-x.

Anjos, M., Targino, A.C., Krecl, P., Oukawa, G.Y., Braga, R.F., 2020. Analysis of the urban heat island under different synoptic patterns using local climate zones. Build. Environ. 185. https://doi.org/10.1016/j.buildenv.2020.107268.

Aslam, A., Rana, I.A., Bhatti, S.S., 2021. The spatiotemporal dynamics of urbanisation and local climate: a case study of Islamabad, Pakistan. Environ. Impact Assess. Rev. 91, 106666. https://doi.org/10.1016/j.eiar.2021.106666.

Baqa, M.F., Lu, L., Guo, H., Song, X., Alavipanah, S.K., Nawaz-ul-Huda, S., Li, Q., Chen, F., 2025. Investigating heat-related health risks related to local climate zones using SDGSAT-1 high-resolution thermal infrared imagery in an arid megacity. Int. J. Appl. Earth Obs. Geoinf. 136, 104334. https://doi.org/10.1016/j.jag.2024.104334.

Bechtel, B., Demuzere, M., Mills, G., Zhan, W., Sismanidis, P., Small, C., Voogt;, J., 2019. SUHI analysis using local climate Zones—A comparison of 50 cities. Urban Clim. 28, 100451. https://doi.org/10.1016/j.uclim.2019.01.005.

Cao, R., Liao, C., Li, Q., Tu, W., Zhu, R., Luo, N., Qiu, G., Shi, W., 2023. Integrating satellite and street-level images for local climate zone mapping. Int. J. Appl. Earth Obs. Geoinf. 119, 103323. https://doi.org/10.1016/j.jag.2023.103323.

Chen, Z., Badrinarayanan, V., Lee, C.Y., Rabinovich, A., 2018. GradNorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In: Proc. 35th Int. Conf. Mach. Learn, pp. 1240–1251.

Ching, J., Mills, G., Bechtel, B., See, L., Feddema, J., Wang, X., Ren, C., Brousse, O., Martilli, A., Neophytou, M., 2018. WUDAPT: an urban weather, climate, and environmental modeling infrastructure for the anthropocene. Bull. Am. Meteorol. Soc. 99, 1907–1924.

Collins, J., Dronova, I., 2019. Urban landscape change analysis using local climate zones and object-based classification in the salt lake metro region, Utah, USA. Remote Sens. 11, 1–27. https://doi.org/10.3390/RS11131615.

Demuzere, M., Bechtel, B., Mills, G., 2019. Global transferability of local climate zone models. Urban Clim. 27, 46–63. https://doi.org/10.1016/j.uclim.2018.11.001.

Dos Anjos, C.S., Lacerda, M.G., do Livramento Andrade, L., Salles, R.N., 2017. Classification of urban environments using feature extraction and random forest. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, pp. 1205–1208. https://doi.org/10.1109/IGARSS.2017.8127174.

Feng, P., Lin, Y., He, G., Guan, J., Wang, J., Shi, H., 2020. A dynamic end-to-end fusion filter for local climate zone classification using SAR and multi-spectrum remote sensing data. In: IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 4231–4234. https://doi.org/10.1109/IGARSS39084.2020.9324427.

Feng, W., Liu, J., 2022. A literature survey of local climate zone classification: status, application, and prospect. Buildings. https://doi.org/10.3390/buildings12101693.

Fonte, C.C., Lopes, P., See, L., Bechtel, B., 2019. Using OpenStreetMap (OSM) to enhance the classification of local climate zones in the framework of WUDAPT. Urban Clim. 28, 100456. https://doi.org/10.1016/j.uclim.2019.100456.

Gawlikowski, J., Schmitt, M., Kruspe, A., Zhu, X.X., 2020. On the fusion strategies of Sentinel-1 and Sentinel-2 data for local climate zone classification. In: IGARSS

2020–2020 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 2081–2084. https://doi.org/10.1109/IGARSS39084.2020.9324234.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.
Hong, D., Gao, L., Yao, J., Yokoya, N., Chanussot, J., Heiden, U., Zhang, B., 2022. Endmember-guided unmixing network (EGU-Net): a general deep learning framework for self-supervised hyperspectral unmixing. IEEE Trans. Neural Netw. Learn. Syst. 33, 6518–6531. https://doi.org/10.1109/TNNLS.2021.3082289.
Hu, J., Shen, L., Albanie, S., Sun, G., Wu, E., 2020. Squeeze-and-excitation networks. IEEE Trans. Pattern Anal. Mach. Intell. 42, 2011–2023. https://doi.org/10.1109/TPAMI.2019.2913372.
Huang, X., Liu, A., Li, J., 2021. Mapping and analyzing the local climate zones in China's 32 major cities using Landsat imagery based on a novel convolutional neural network. Geo-Spat. Inf. Sci. 24, 528–557. https://doi.org/10.1080/10095020.2021.1892459.
Ji, W., Chen, Y., Li, K., Dai, X., 2023. Multicascaded feature fusion-based deep learning network for local climate zone classification based on the So2Sat LCZ42 benchmark dataset. IEEE J Sel. Top. Appl. Earth Obs. Remote Sens. 16, 449–467. https://doi.org/10.1109/JSTARS.2022.3226524.
Kim, M., Jeong, D., Kim, Y., 2021. Local climate zone classification using a multi-scale, multi-level attention network. ISPRS J. Photogramm. Remote Sens. 181, 345–366. https://doi.org/10.1016/j.isprsjprs.2021.09.015.
La, Y., Bagan, H., Yamagata, Y., 2020. Urban land cover mapping under the local climate Zone scheme using Sentinel-2 and PALSAR-2 data. Urban Clim. 33, 100661. https://doi.org/10.1016/j.uclim.2020.100661.
Li, J., Chen, Y., Zhao, D., Zhai, J., 2024. The impact of built environment on mixed land use: evidence from Xi'an. Land 13. https://doi.org/10.3390/land13122214.
Liang, T., He, J., Chen, L., Yao, Z., Zhang, L., Che, H., Gong, S., 2021. Simulation of the influence of a fine-scale urban underlying surface on the urban heat island effect in Beijing. Atmos. Res. 262, 105786. https://doi.org/10.1016/j.atmosres.2021.105786.
Liu, L., Lin, M., Du, Z., Liu, J., Chen, G., Du, J., 2023. Developing a CNN-based, block-scale oriented local climate Zone mapping approach: a case study in Guangzhou. Build. Environ. 240, 110414. https://doi.org/10.1016/j.buildenv.2023.110414.
Liu, S., Qi, Z., Li, X., Yeh, A.G.O., 2019. Integration of convolutional neural networks and object-based post-classification refinement for land use and land cover mapping with optical and sar data. Remote Sens. 11, 1–25. https://doi.org/10.3390/rs11060690.
Liu, S., Shi, Q., 2020. Local climate zone mapping as remote sensing scene classification using deep learning: a case study of metropolitan China. ISPRS J. Photogramm. Remote Sens. 164, 229–242. https://doi.org/10.1016/j.isprsjprs.2020.04.008.
Liu, W., Zhang, L., He, T., Zhang, M., 2025. Evaluation of urban heat island effects based on fine-resolution mapping of subpixel impervious surface dynamics over four cities in China. Int. J. Appl. Earth Obs. Geoinf. 142, 104740. https://doi.org/10.1016/j.jag.2025.104740.
Ma, L., Yang, Z., Zhou, L., Lu, H., Yin, G., 2021a. Local climate zones mapping using object-based image analysis and validation of its effectiveness through urban surface temperature analysis in China. Build. Environ. 206, 108348. https://doi.org/10.1016/j.buildenv.2021.108348.
Ma, L., Zhu, X., Qiu, C., Blaschke, T., Li, M., 2021b. Advances of local climate zone mapping and its practice using object-based image analysis. Atmos. 12, 1146.
Manyanya, T., Teerlinck, J., Somers, B., Verbist, B., Nethengwe, N., 2022. Sentinel-based adaptation of the local climate zones framework to a South African context. Remote Sens. 14, 1–23. https://doi.org/10.3390/rs14153594.
Oba, Y., Tezuka, T., Hasegawa, H., 2025. Prediction of Iterative Solvers' Convergence Using Pretraining by Natural Images. In: Proc. Int. Conf. Big Data Anal. Knowl. Discov. Springer, pp. 341–348.
Qiu, C., Schmitt, M., Mou, L., Ghamisi, P., Zhu, X.X., 2018. Feature importance analysis for local climate zone classification using a residual convolutional neural network with multi-source datasets. Remote Sens. 10, 1–14. https://doi.org/10.3390/rs10101572.
Qiu, C., Tong, X., Schmitt, M., Bechtel, B., Zhu, X.X., 2020. Multilevel feature fusion-based CNN for local climate zone classification from sentinel-2 images: benchmark results on the So2Sat LCZ42 dataset. IEEE J Sel. Top. Appl. Earth Obs. Remote Sens. 13, 2793–2806. https://doi.org/10.1109/JSTARS.2020.2995711.
Quan, J.L., 2019. Enhanced geographic information system-based mapping of local climate zones in Beijing, China. Sci. China Technol. Sci. 62, 2243–2260. https://doi.org/10.1007/s11431-018-9417-6.
Rosentreter, J., Hagensieker, R., Waske, B., 2020. Towards large-scale mapping of local climate zones using multitemporal Sentinel 2 data and convolutional neural networks. Remote Sens. Environ. 237, 111472. https://doi.org/10.1016/j.rse.2019.111472.
Shih, W.Y., Ahmad, S., Chen, Y.C., Lin, T.P., Mabon, L., 2020. Spatial relationship between land development pattern and intra-urban thermal variations in Taipei. Sustain. Cities Soc. 62, 102415. https://doi.org/10.1016/j.scs.2020.102415.
Simanjuntak, R.M., Kuffer, M., Reckien, D., 2019. Object-based image analysis to map local climate zones: the case of Bandung, Indonesia. Appl. Geogr. 106, 108–121. https://doi.org/10.1016/j.apgeog.2019.04.001.
Stewart, I.D., Oke, T.R., 2012. Local climate zones for urban temperature studies. Bull. Am. Meteorol. Soc. 93, 1879–1900. https://doi.org/10.1175/BAMS-D-11-00019.1.
Taubenböck, H., Debray, H., Qiu, C., Schmitt, M., Wang, Y., Zhu, X.X., 2020. Seven city types representing morphologic configurations of cities across the globe. Cities 105, 102814. https://doi.org/10.1016/j.cities.2020.102814.
Vandamme, S., Demuzere, M., Verdonck, M.L., Zhang, Z., Van Coillie, F., 2019. Revealing Kunming's (China) historical urban planning policies through local climate zones. Remote Sens. 11, 1–20. https://doi.org/10.3390/rs11141731.
Vavassori, A., Oxoli, D., Venuti, G., Brovelli, M.A., Siciliani de Cumis, M., Sacco, P., Tapete, D., 2024. A combined remote sensing and GIS-based method for local climate Zone mapping using PRISMA and Sentinel-2 imagery. Int. J. Appl. Earth Obs. Geoinf. 131, 103944. https://doi.org/10.1016/j.jag.2024.103944.
Verdonck, M.L., Okujeni, A., van der Linden, S., Demuzere, M., De Wulf, R., Van Coillie, F., 2017. Influence of neighbourhood information on 'local climate Zone' mapping in heterogeneous cities. Int. J. Appl. Earth Obs. Geoinf. 62, 102–113. https://doi.org/10.1016/j.jag.2017.05.017.
Wang, R., Wang, M., Ren, C., Chen, G., Mills, G., Ching, J., 2024. Mapping local climate zones and its applications at the global scale: a systematic review of the last decade of progress and trend. Urban Clim. 57, 102129. https://doi.org/10.1016/j.uclim.2024.102129.
Wang, Y., Ni, Z., Hu, M., Chen, S., Xia, B., 2021. A practical approach of urban green infrastructure planning to mitigate urban overheating: a case study of Guangzhou. J. Clean. Prod. 287, 124995. https://doi.org/10.1016/j.jclepro.2020.124995.
Xu, C., Hystad, P., Chen, R., Van Den Hoek, J., Hutchinson, R.A., Hankey, S., Kennedy, R., 2021. Application of training data affects success in broad-scale local climate zone mapping. Int. J. Appl. Earth Obs. Geoinf. 103, 102482. https://doi.org/10.1016/j.jag.2021.102482.
Xu, Z., Zhang, H., Wang, Y., Chang, X., Liang, Y., 2010. L 1/2 regularization. Sci. China Inf. Sci. 53, 1159–1169.
Yoo, C., Han, D., Im, J., Bechtel, B., 2019. Comparison between convolutional neural networks and random forest for local climate zone classification in mega urban areas using Landsat images. ISPRS J. Photogramm. Remote Sens. 157, 155–170. https://doi.org/10.1016/j.isprsjprs.2019.09.009.
Yoo, C., Lee, Y., Cho, D., Im, J., Han, D., 2020. Improving local climate zone classification using incomplete building data and sentinel 2 images based on convolutional neural networks. Remote Sens. 12, 1–22. https://doi.org/10.3390/rs12213552.
Yu, Y., Li, J., Yuan, Q., Shi, Q., Shen, H., Zhang, L., 2022. Coupling dual graph convolution network and residual network for local climate zone mapping. IEEE J Sel. Top. Appl. Earth Obs. Remote Sens. 15, 1221–1234. https://doi.org/10.1109/JSTARS.2021.3132394.
Zhang, R., Wang, Y., Hu, J., Yang, W., Chen, J., Zhu, X., 2022. SAR4LCZ-net: a complex-valued convolutional neural network for local climate zones classification using Gaofen-3 quad-pol SAR data. IEEE Trans. Geosci. Remote Sens. 60, 1–16. https://doi.org/10.1109/TGRS.2021.3137911.
Zhang, X., Du, S., 2015. A Linear Dirichlet Mixture Model for decomposing scenes: Application to analyzing urban functional zonings. Remote Sens. Environ. 169, 37–49. https://doi.org/10.1016/j.rse.2015.07.017.
Zhang, Y., Li, D., Liu, L., Liang, Z., Shen, J., Wei, F., Li, S., 2021. Spatiotemporal characteristics of the surface urban heat island and its driving factors based on local climate zones and population in beijing, china. Atmos. 12. https://doi.org/10.3390/atmos12101271.
Zhao, N., Zhong, Y., Ma, A., 2020. Mapping local climate zones with circled similarity propagation based domain adaptation. In: IGARSS 2020–2020 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 1377–1380. https://doi.org/10.1109/IGARSS39084.2020.9323191.
Zhao, X., Xia, N., Li, M.C., 2023. 3-D multi-aspect mix degree index: a method for measuring land use mix at street block level. Comput. Environ. Urban Syst. 104, 102005. https://doi.org/10.1016/j.compenvurbsys.2023.102005.
Zhou, Y., Wei, T., Zhu, X., Collin, M., 2021. A parcel-based deep-learning classification to map local climate zones from Sentinel-2 images. IEEE J Sel. Top. Appl. Earth Obs. Remote Sens. 14, 4194–4204. https://doi.org/10.1109/JSTARS.2021.3071577.
Zhu, Q., Ran, L., Zhang, Y., Guan, Q., 2024. Integrating geographic knowledge into deep learning for spatiotemporal local climate zone mapping derived thermal environment exploration across Chinese climate zones. ISPRS J. Photogramm. Remote Sens. 217, 53–75. https://doi.org/10.1016/j.isprsjprs.2024.08.004.
Zonato, A., Martilli, A., Di Sabatino, S., Zardi, D., Giovannini, L., 2020. Evaluating the performance of a novel WUDAPT averaging technique to define urban morphology with mesoscale models. Urban Clim 31, 100584. https://doi.org/10.1016/j.uclim.2020.100584.