remote sensing

MDPI

Article

A Method for Segmentation and Identification of Urban Functional Areas with Generalization Ability Based on Scale-Aware Feature-Enhanced Attention Mask R-CNN

Chao Wang [1,*], Ziyang Chen [1], Shuzhe Huang [1], Pengfei Li [2] and Wei Wang [1]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430070, China; czyx_520@whu.edu.cn (Z.C.); szhuang@whu.edu.cn (S.H.); wangwei8091@163.com (W.W.)
[2] Yunnan Provincial Mapping Institute, Kunming 650034, China; qazlipengfei@126.com
* Correspondence: c.wang@whu.edu.cn

Highlights

What are the main findings?

- SFA-Mask R-CNN achieves over 91% functional zone classification accuracy across three cities by integrating CBAM, SA-FPN, and ASPP modules.
- Wuhan's transitional-stage model demonstrates the strongest cross-city transferability (85.37% in Chengdu), while the tested gradient reversal layer-based adversarial domain adaptation strategy fails to outperform direct transfer under the experimental conditions of this study.

What are the implications of the main findings?

- Cities at intermediate developmental stages provide the most generalizable training data, enabling efficient "train once, apply to multiple cities" large-scale mapping.
- Urban developmental stage is a key factor governing model transferability, offering new guidance for cross-city functional zone identification.

Abstract

Accurate identification of urban functional zones is critically important for urban planning and sustainable development. However, existing deep learning methods exhibit limited cross-city generalization capabilities, rendering them inadequate for large-scale automated mapping applications. To address this challenge, this study proposes a cross-city identification framework based on SFA-Mask R-CNN (Scale-Aware Feature-enhanced Attention Mask R-CNN), which achieves precise segmentation and classification of functional zones by integrating three complementary feature enhancement mechanisms: a Convolutional Block Attention Module (CBAM) for discriminative feature recalibration, a novel Scale-Aware FPN (SA-FPN) that dynamically adjusts feature pyramid layer weights according to the scale distribution of input imagery to improve cross-city transferability, and an ASPP (Atrous Spatial Pyramid Pooling) module for multi-scale contextual feature extraction, combined with multi-source data fusion. Using three cities along the Yangtze River Economic Belt—Chengdu, Wuhan, and Shanghai—as study areas, we systematically evaluated the model's cross-city generalization performance. The results show that the Wuhan model exhibits the strongest cross-city transferability, achieving 85.37% accuracy in Chengdu, suggesting that models trained on cities at transitional development stages possess superior generalizability. In contrast, the gradient reversal layer-based domain adaptation approach tested in this study failed to effectively enhance model performance.

Check for updates

Academic Editor: Victor Mesev

Received: 29 March 2026
Revised: 12 June 2026
Accepted: 26 July 2026
Published: 3 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

This study provides a practical technical pathway for "train once, apply to multiple cities" large-scale functional zone mapping and offers new perspectives for research on regional disparities in urban development.

Keywords: urban functional zone identification; Mask R-CNN; cross-city generalization; domain adaptation; multi-source data fusion

1. Introduction

Urban functional zones, as fundamental constituent units of urban spatial structure, reflect the spatial configuration and functional differentiation of urban land use. Accurate identification and analysis of these zones are critical for urban planning, optimization of resource allocation, and sustainable development. With the rapid advancement of China's urbanization process, cities at different developmental stages exhibit markedly distinct spatial patterns of functional zones. These variations pose new challenges for the automated identification of urban functional zones.

Early urban functional zone identification relied primarily on questionnaire surveys conducted among the public and experts, enabling delineation of functional zones based on public preferences [1]. With advances in geographic information technology, geometric algorithm-based methods using cadastral data gradually emerged. By combining Delaunay triangulation with the Graham algorithm, automated functional zone classification was implemented on NET and Geomedia platforms [2]. However, traditional approaches to urban functional zone identification largely depend on manual surveys and expert knowledge, resulting in low efficiency, strong subjectivity, and difficulty in updating, which renders them inadequate for the dynamic monitoring demands of rapidly urbanizing cities [3].

The introduction of remote sensing technology has provided new technical pathways for urban functional zone identification. Remote sensing imagery offers rich information on urban physical characteristics, and by analyzing images acquired at different time periods, dynamic monitoring of urban land use changes can be achieved [4]. During this stage, object-based image analysis became the mainstream approach; through multi-scale segmentation and hierarchical strategies, researchers were able to extract urban land-use information from high-resolution remote sensing imagery, detect urban land-use changes, and classify urban areas based on visually interpreted functional characteristics [5].

However, because remote sensing imagery primarily reflects physical characteristics such as urban texture and cannot fully capture urban social-semantic attributes, researchers have gradually incorporated urban static data (e.g., road networks and points of interest POIs) into functional zone classification frameworks. Early POI density-based scoring methods enabled quantitative analysis by calculating POI densities within grid cells, allowing intuitive descriptions of mixed functional zones [6]. Subsequently, the quadrat density method was introduced, using frequency density and type proportion as indicators for refined identification of single and mixed functional zones [7]. Building on this foundation, weighted scoring approaches established links between facility distribution and functional attributes by assigning different weights to POIs across categories and hierarchical levels, thereby determining functional zone types through composite density scores [8].

However, simple density statistics overlook spatial distribution characteristics. To address this limitation, methods incorporating kernel density estimation (KDE) were developed to account for POI spatial autocorrelation, effectively reducing point-data discretization and producing smoother, more continuous functional identification results [9].

Principal Component Analysis (PCA) further simplified multidimensional data complexity by reducing standardized block-level POI density data to extract composite indicators representing functional zones [10]. In addition, weighted frequency density ratio methods refined traditional approaches by using building-scale POI ratios within buffer zones to achieve automated classification of building functional types [11]. Collectively, these methods overcome the limitations of simple density statistics and provide more precise and diverse technical pathways for urban functional zone identification.

Furthermore, unsupervised identification methods based on clustering have been introduced into functional zone analysis, enabling automated delineation of functional zones by extracting features from urban data according to regional functional characteristics or the spatiotemporal patterns of human mobility. The K-means algorithm, owing to its computational simplicity, efficiency, and effective clustering performance, has become the most widely used method for urban functional zone identification. It has been extensively applied using multi-source data such as mobile phone data [12], taxi GPS data [13], check-in data [14], and POI data. However, K-means requires manual specification of the number of clusters, and its results are sensitive to the initialization of cluster centers.

To address these limitations, the fuzzy c-means clustering algorithm was introduced, enabling automatic discovery of functional patterns from the multidimensional feature space of POI data and successfully achieving automated delineation and functional interpretation of urban areas [15]. In addition, Gaussian Mixture Models (GMM) offer stronger descriptive capability than K-means by capturing complex data distributions. By integrating dynamic data such as GPS movement trajectories with unsupervised learning approaches, researchers have further improved the accuracy of functional zone identification [16]. Moreover, enhanced methods based on Dynamic Time Warping (DTW) distance k-medoids [17], multi-feature weighted decision algorithms [18], and related approaches have been successively proposed, continuously enriching clustering-based methodologies for urban functional zone identification.

In recent years, the rapid development of deep learning technology has brought new technical breakthroughs to urban functional zone identification. Convolutional Neural Networks (CNN), with their powerful image feature extraction capabilities, can automatically learn hierarchical semantic features from remote sensing imagery. End-to-end deep learning architectures have increasingly been applied to functional zone identification. By constructing multi-scale semantic feature extraction networks and directly learning the spatial distribution and texture patterns of functional zones from high-resolution remote sensing imagery, fine-grained functional zone delineation has been achieved, significantly outperforming traditional handcrafted feature-based methods [19].

To fully leverage the complementary advantages of multi-source data, multi-modal deep fusion methods have emerged as a major research focus. By separately constructing feature extraction branches for remote sensing imagery and rasterized POI data, and introducing attention mechanisms in the fusion layer to dynamically adjust the contribution weights of different modalities, these approaches successfully integrate physical morphological features with social functional semantics, achieving notable advances in high-precision functional zone identification, such as in Ningbo [20]. Hybrid modeling strategies further enrich the methodological framework. Through spatial unit construction based on Minimum Identification Units (MIU), combined with Deep Embedded Clustering (DEC) models, street-block-level functional zone segmentation has been achieved [21]. Alternatively, by proposing the SALT feature system—which integrates building morphology, population activity dynamics, and high-resolution imagery texture information—and combining it with ensemble learning frameworks, excellent classification performance has been demonstrated in single-city applications [22].

However, existing research still faces numerous challenges in cross-city applications. First, most deep learning models are trained on single-city datasets and therefore exhibit limited cross-city generalization capability. This limitation primarily arises from substantial differences in spatial characteristics among cities at different stages of development. Mature cities tend to feature high-density, small-scale, finely delineated functional zones with clear boundaries, whereas emerging cities are characterized by low-density, large-scale, and more extensive spatial layouts with pronounced transitional zones. When models trained on a single city are applied to cities with different development levels and spatial patterns, recognition accuracy declines markedly.

This performance degradation results from several compounding factors: (1) Scale mismatch: model parameters such as receptive field size and feature pyramid levels optimized for one city do not transfer effectively, leading to boundary omission in large-scale functional zones or over-segmentation of small-scale zones; (2) Feature overfitting: models tend to overfit city-specific spatial morphological features—such as building density, road network configuration, and greenspace distribution—while failing to capture more universal semantic associations relevant to functional zone identification; (3) Category imbalance: substantial differences in the frequency and mixing degree of functional zone types across cities introduce category bias or confusion when models are applied to new cities, particularly resulting in poor identification performance for functional zone types that are underrepresented in the training data.

Second, there is a lack of systematic comparative analysis targeting different urban development characteristics, making it difficult to uncover the underlying causes of model performance differences. Existing research primarily focuses on method optimization within single cities and lacks horizontal comparative experimental designs across cities at different developmental stages. As a result, it remains challenging to quantify the specific mechanisms through which urban characteristics—such as urbanization level, spatial morphological complexity, and the degree of functional zone mixing—influence model performance. This research gap prevents researchers from identifying which types of cities provide more representative training data with stronger generalization potential and fails to offer theoretical guidance for constructing universally applicable identification frameworks.

To address these issues, this study selected three representative cities along the Yangtze River Economic Belt—Chengdu, Wuhan, and Shanghai—as research cases, and proposed a cross-city urban functional zone identification framework based on SFA-Mask R-CNN (Scale-Aware Feature-enhanced Attention Mask R-CNN). Through targeted optimization of the classic Mask R-CNN architecture, including the introduction of the SA-FPN (Scale-Aware Feature Pyramid Network) to replace the standard FPN and dynamically adjust feature pyramid layer weights according to the scale distribution of input imagery, the introduction of the CBAM (Convolutional Block Attention Module) at the feature extraction stage for discriminative feature recalibration, and the integration of an ASPP (Atrous Spatial Pyramid Pooling) module for multi-scale contextual feature extraction—the challenges of identifying functional zones at different spatial scales were effectively addressed. SA-FPN enables the model to adapt its feature fusion to the scale characteristics of each target city without requiring city-level annotations. CBAM enables the model to adaptively focus on salient functional zone features through the cascading of channel and spatial attention. ASPP extracts features in parallel using multiple atrous convolutions with different dilation rates, enhancing the model's ability to perceive functional zone boundaries across scales. Combined with multi-source geographic data fusion, a comprehensive functional zone classification system was constructed. More importantly, by designing comparative experiments involving single-city modeling and domain-adaptive joint modeling, the cross-city generalization performance of the model was systematically evaluated, providing a scien-

tific basis for deeper understanding of functional zone identification patterns across cities at different stages of development.

The main contributions of this study include: (1) Proposing the SFA-Mask R-CNN architecture, which integrates three complementary modules addressing distinct and non-redundant challenges in cross-city functional zone identification. While the individual building blocks—CBAM [23], ASPP [24], and feature pyramid networks [25]—have been introduced in prior work, their specific integration and the proposed SA-FPN constitute a targeted design for the cross-city scenario. Unlike existing adaptive pyramid architectures such as PANet and BiFPN, which focus on cross-level feature flow aggregation or fixed learned weighting, the proposed SA-FPN introduces a dynamic layer-level weighting mechanism that estimates FPN level importance from the scale distribution of each input image at inference time, without city-level annotations or manual reconfiguration—directly addressing the inter-city scale heterogeneity arising from differences in urban developmental stages. CBAM resolves what features to attend to through channel and spatial recalibration, and ASPP resolves how wide the receptive field should be through parallel atrous convolutions. This tri-level design is validated through systematic ablation experiments and cross-city generalization tests; (2) Revealing the influence of urban developmental stages on model generalization capability, demonstrating that the Wuhan model—representing a transitional stage characterized by both rapid development and structural optimization—exhibits the strongest cross-city adaptability, thereby providing a feasible technical pathway for "train once, apply to multiple cities" large-scale functional zone mapping; (3) Based on empirical analysis of three representative cities along the Yangtze River Economic Belt, quantitatively elucidating the mechanisms by which urban developmental stages affect functional zone identification accuracy, offering new technical approaches and analytical perspectives for research on regional disparities in urban development.

2. Study Area and Data

2.1. Study Area Selection

This study selected Chengdu, Wuhan, and Shanghai as research areas based on their typicality and representativeness of urban development along the Yangtze River Economic Belt (Figure 1). The Yangtze River Economic Belt spans eastern, central, and western China, covering 11 provinces and municipalities, and is one of the country's most economically dynamic regions. Urban development levels within this region exhibit pronounced gradient differences, providing ideal experimental conditions for examining the adaptability and generalization capability of cross-city urban functional zone identification methods.

The three study cities represent typical development models across the eastern, central, and western regions of the Yangtze River Economic Belt. Shanghai, as the leading city of the Yangtze River Economic Belt and China's economic center, exemplifies the developmental characteristics of mature eastern cities. Its built-up area covers approximately 1400 square kilometers, with a high level of urbanization, clearly delineated functional zones, and a relatively stable spatial structure, forming a multi-center development pattern with the Lujiazui Financial Center as the core. Wuhan, located in the middle reaches of the Yangtze River, is a key node city in central China and represents rapidly developing central cities. Its built-up area spans approximately 800 km^2 and is currently undergoing rapid urbanization and industrial upgrading. The distinctive geographical configuration of "two rivers and four banks" has shaped a relatively clear functional zone distribution, with zones exhibiting dynamic change. Chengdu, as the core city of western China, represents the development model of emerging western cities. Its built-up area covers approximately 900 km^2 and has experienced rapid urban expansion and functional zone restructuring in

recent years, characterized by fast-growing urban space and the coexistence of old and new functional zones.

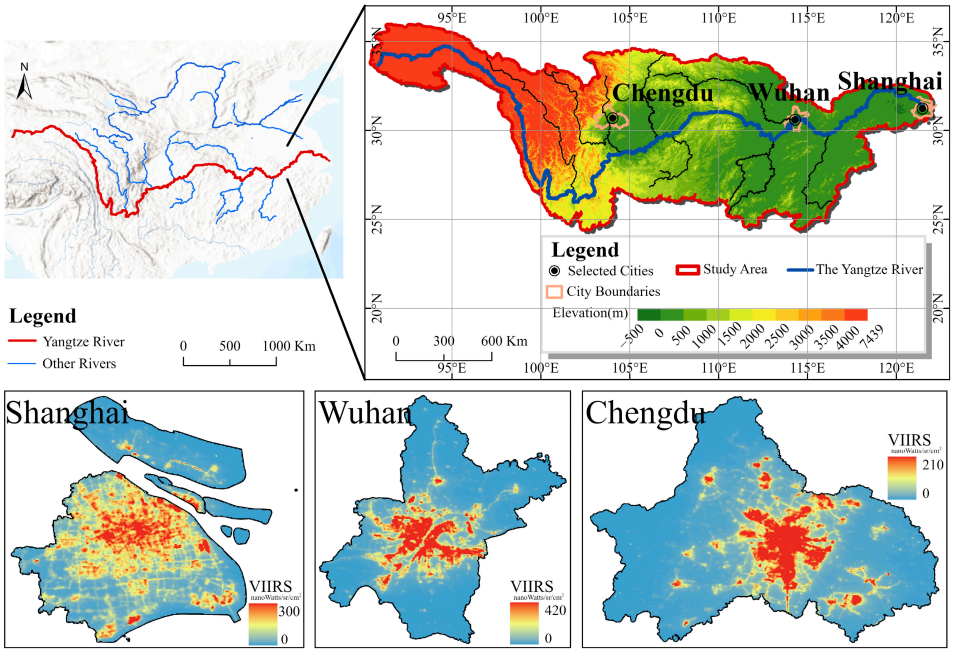


Figure 1. Overview of the study area.

The substantial differences among the three cities in urban scale, developmental stage, spatial morphology, and functional zone characteristics result in markedly distinct spatial patterns of urban functional zones. These differences not only reflect the regional gradient characteristics of urban development in China but also provide ideal experimental scenarios for validating the generalization capability of cross-city functional zone identification methods, thereby helping to elucidate the mechanisms through which urban developmental stages influence functional zone identification accuracy.

2.2. Data Sources and Acquisition

This study adopted a multi-source heterogeneous data fusion strategy to fully exploit the complementary advantages of different data sources. High-resolution remote sensing imagery was obtained from the Google Earth platform, acquiring RGB images of the three study cities in 2024 with a spatial resolution of 1 m. These images cover the main built-up areas of each city, with a total coverage exceeding 3000 km^2, providing a solid data foundation for precise instance segmentation of urban functional zones. Google Earth imagery offers advantages such as high image quality, frequent updates, and global coverage, enabling clear representation of urban elements including building morphology, road networks, and greenspace distribution.

Nighttime light data were derived from the NPP/VIIRS (Suomi National Polar-orbiting Partnership Visible Infrared Imaging Radiometer Suite) product, with a spatial resolution of 500 m and using monthly composite data for 2024 [26]. Nighttime light intensity effectively reflects the spatial distribution of human activity intensity and economic vitality, providing important socioeconomic information for functional zone classification. Compared with traditional DMSP/OLS data, NPP/VIIRS offers higher spatial and radiometric resolution, enabling more precise capture of intra-urban variations in activity intensity.

To address the spatial resolution mismatch between the 1 m remote sensing imagery and the 500 m VIIRS nighttime light data, multi-source feature extraction was conducted at the functional zone instance level rather than at the pixel level. For each functional zone polygon delineated by SFA-Mask R-CNN, zonal statistics (mean and standard de-

viation of VIIRS digital number values) were computed within the instance boundary, effectively reconciling the resolution disparity through instance-level aggregation. The annual mean composite of monthly VIIRS products for 2024 was adopted to minimize seasonal fluctuations and temporal noise. For POI data, kernel density estimation (KDE) with bandwidth determined by least-squares cross-validation was applied to generate continuous density surfaces for each of the 20 major Amap POI categories, from which mean density values within each instance boundary were extracted; point-in-polygon counting was additionally used to obtain absolute category counts and proportional composition per instance. All three data sources were targeted at the 2024 reference year, and the use of annual composites for nighttime light data ensured temporal consistency adequate for land-use classification purposes.

POI (Points of Interest) data were obtained from the Amap open platform, providing POI records for the three study cities in 2024 and covering more than 20 major categories, including commercial services, education and research, medical and health facilities, transportation infrastructure, and government institutions [27]. The POI data were subjected to refined classification and coordinate correction. After removing duplicate records and data points with low positional accuracy, we retained more than 500,000 valid POI records. POI data directly reflect actual land-use functions and service types within urban areas, providing a critical semantic basis for functional zone classification.

The data preprocessing workflow included geometric correction, projection transformation, and spatial registration. All datasets were standardized to the WGS84 coordinate system with a UTM projection to ensure spatial consistency across data sources. Remote sensing imagery underwent atmospheric correction and radiometric calibration, nighttime light data were processed to remove outliers and noise, and POI data were subjected to geocoding and attribute standardization. Establishing a unified spatial data framework provided a solid foundation for subsequent multi-source data fusion analysis.

3. Methodology

This study develops a cross-city urban functional zone identification framework based on deep learning and multi-source data fusion. The framework (Figure 2) employs the SFA-Mask R-CNN (Scale-Aware Feature-enhanced Attention Mask R-CNN) architecture to perform instance segmentation of functional zones from high-resolution remote sensing imagery, yielding precise boundary delineations. These segmentation results are subsequently integrated with nighttime light and POI data to construct a multidimensional feature system, within which a random forest classifier achieves semantic functional zone classification. Three comparative experimental settings—single-city independent modeling, cross-city direct transfer, and domain-adaptive joint modeling—were designed to systematically evaluate the model's adaptability and generalization across cities at different developmental stages.

3.1. SFA-Mask R-CNN Instance Segmentation

The proposed SFA-Mask R-CNN architecture builds upon the Mask R-CNN framework [28] and introduces three targeted enhancements to address the core challenges of urban functional zone identification: large-scale spatial variation and blurred inter-zone boundaries (Figure 3). It is important to note that SFA-Mask R-CNN is trained as a binary instance segmentation model: the network learns to distinguish foreground (urban functional zone instances) from background, without predicting any semantic category label such as residential, industrial, or commercial. The output of SFA-Mask R-CNN is therefore a set of instance-level polygonal masks, each delineating the precise spatial boundary of one detected functional zone unit, accompanied only by a binary foreground confidence score.

Semantic classification of these delineated instances—into categories such as residential, industrial, office, entertainment, and transportation—is subsequently performed by the random forest classifier described in Section 3.2, which operates on a multi-dimensional feature vector extracted for each instance. This two-stage design is intentionally adopted to separate spatial morphology recognition, at which deep learning excels, from semantic functional zone categorization, which requires heterogeneous multi-modal attributes (spectral indices, POI density, nighttime light intensity) more effectively handled by feature-based machine learning. It also enhances cross-city transferability, as the morphological segmentation model can be directly applied to new cities without re-annotating semantic categories. The backbone network employs ResNet50 [29] to extract hierarchical deep features through residual learning, and a CBAM (Convolutional Block Attention Module) [23] is embedded after each residual stage to adaptively recalibrate feature responses across both channel and spatial dimensions, suppressing irrelevant background interference. The central architectural innovation of this study is the proposed SA-FPN (Scale-Aware Feature Pyramid Network), which replaces the standard FPN [25] and constitutes the primary contribution of the SFA-Mask R-CNN framework.

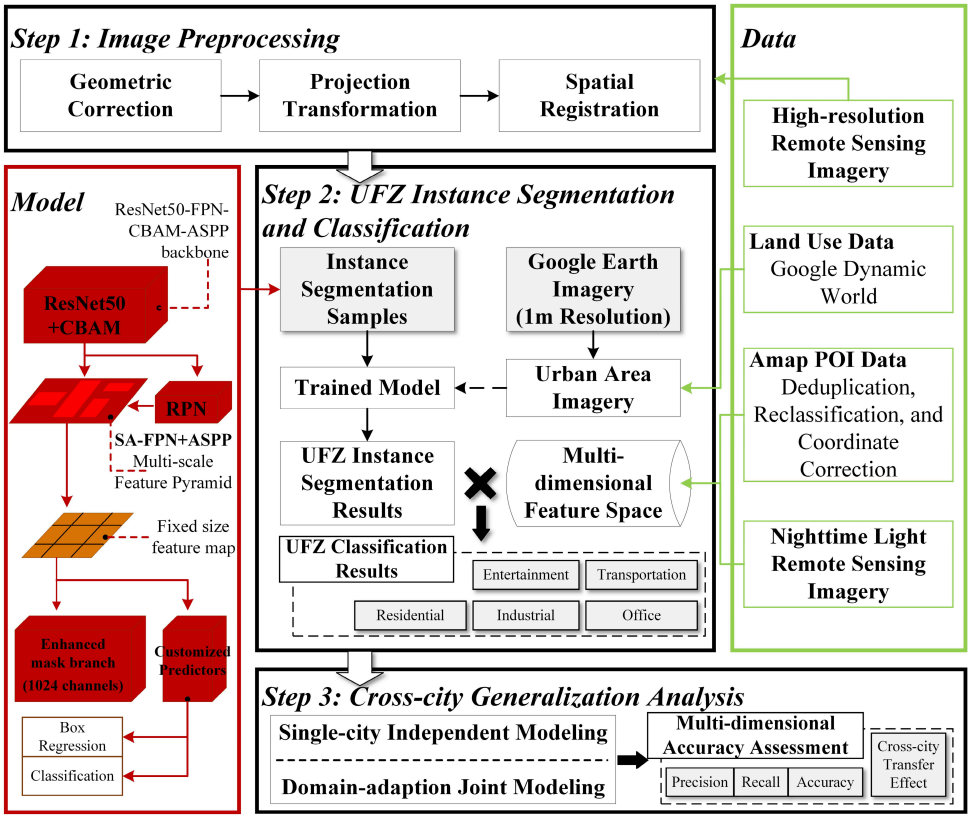


Figure 2. Technical framework.

Standard FPN constructs multi-scale feature pyramids through lateral connections and a top-down pathway, merging adjacent feature levels by direct element-wise addition. While effective for single-city scenarios, this fixed aggregation strategy assigns equal importance to all pyramid levels regardless of the scale distribution of functional zones in the input image. In cross-city applications, cities at different developmental stages exhibit substantially different functional zone scale distributions. Emerging cities such as Chengdu are dominated by large-scale, low-density functional zones, which require stronger contributions from low-resolution, semantically rich pyramid levels. Mature cities such as Shanghai contain dense, fine-grained zones that depend more on high-resolution

levels for boundary delineation. A fixed-weight FPN cannot adapt to these inter-city differences, limiting cross-city transferability.

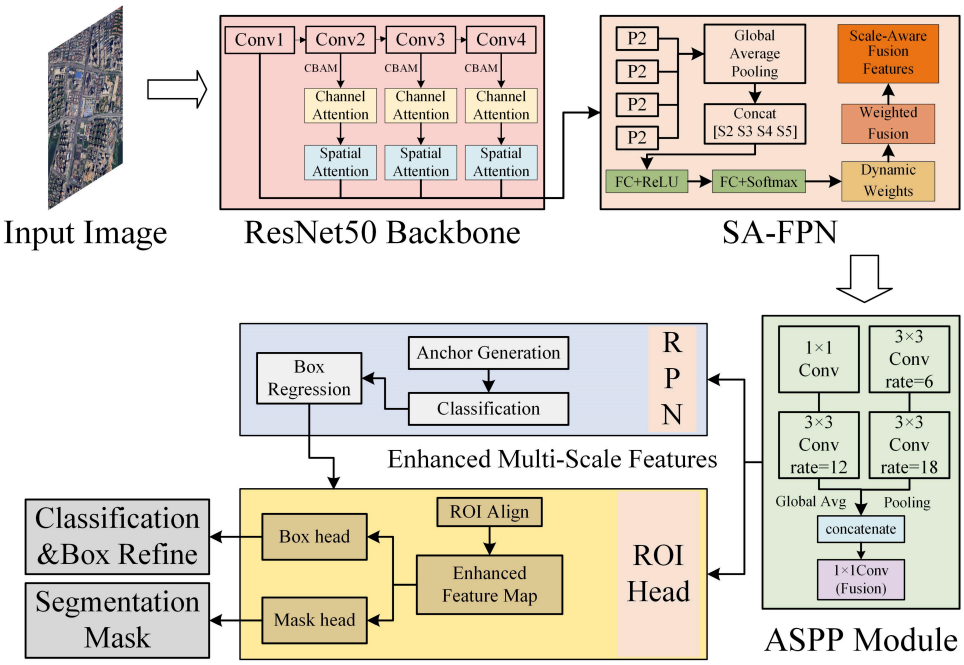


Figure 3. Model architecture.

The proposed SA-FPN addresses this limitation by introducing a dynamic layer-level weighting mechanism that estimates the relative importance of each FPN level from the scale characteristics of the input image. After the standard top-down pathway produces feature maps $\{F_2, F_3, F_4, F_5\}$ at pyramid levels $l \in \{2, 3, 4, 5\}$, global average pooling is applied to each level to produce a compact scale descriptor:

$$g_l = GAP(F_l), \quad g_l \epsilon \mathbb{R}^C \tag{1}$$

The level-wise descriptors are then concatenated into a unified scale representation:

$$g = \left[g_2 \bigoplus g_3 \bigoplus g_4 \bigoplus g_5 \right] \tag{2}$$

This representation is passed through a two-layer fully connected network with ReLU activation and Softmax normalization to produce a set of dynamic level weights $\omega = \{\omega_2, \omega_3, \omega_4, \omega_5\}$:

$$\omega = Softmax(W_2 \cdot ReLU(W_1 \cdot g)) \tag{3}$$

where W_1 and W_2 are learnable weight matrices. Each scalar weight ω_l reflects the estimated importance of pyramid level l given the scale distribution of the current input image. The SA-FPN output at each level is computed as:

$$P_l = \omega_l \cdot F_l \tag{4}$$

Let $C = 256$ denote the uniform channel dimension at all FPN levels. Global average pooling applied to $F_l \in \mathbb{R}^{\{H_l \times W_l \times C\}}$ produces $g_l \in \mathbb{R}^C$. These are concatenated into $g \in \mathbb{R}^{\{4C\}}$. The two-layer MLP maps g through $W_1 \in \mathbb{R}^{\{H \times 4C\}}$ ($H = 256$) with ReLU activation, yielding an intermediate hidden representation $z = ReLU(W_1 \cdot g) \in \mathbb{R}^H$, then through $W_2 \in \mathbb{R}^{\{4 \times H\}}$ followed by Softmax, yielding dynamic level weights $\omega \in \mathbb{R}^4$. Critically, these weights are computed independently for each input image at inference

time: while W_1 and W_2 are learned parameters, the output ω_l varies with each forward pass, allowing the model to automatically adapt its pyramid-level weighting to the scale distribution of the specific input image without any manual configuration.

Because the weights are derived directly from the input features at inference time, the model automatically increases the contribution of higher-resolution levels for inputs dominated by small-scale functional zones, and of lower-resolution levels for inputs containing large-scale zones. Critically, this mechanism requires no city-level annotations and no manual reconfiguration when applied to a new city, making SA-FPN inherently adaptive to inter-city scale differences. In the context of urban functional zone identification—where functional zones range from compact commercial street blocks to large industrial parks, and where scale distributions vary markedly across cities at different developmental stages—this layer-level dynamic weighting provides a structurally motivated advantage over fixed-weight feature aggregation.

Following SA-FPN, an ASPP (Atrous Spatial Pyramid Pooling) module [24] is appended to further enrich multi-scale contextual representations. The ASPP module adopts a parallel multi-branch architecture comprising a 1×1 convolution, three 3×3 atrous convolutions with dilation rates of 6, 12, and 18, and a global average pooling branch. The outputs are concatenated and fused via a 1×1 convolution, providing enriched contextual information for the downstream Region Proposal Network (RPN) and ROI heads. While SA-FPN governs adaptive weighting across pyramid levels according to input-image scale distribution, ASPP expands the effective receptive field within each level without increasing parameter count, enabling the model to simultaneously perceive context at multiple granularities. The synergy between these two modules—SA-FPN for input-adaptive cross-level contribution adjustment and ASPP for within-scale contextual expansion—equips SFA-Mask R-CNN with robust capability to handle the full spectrum of spatial scale variation encountered across cities at different developmental stages.

The detection and segmentation heads were additionally enhanced: the mask prediction branch employs an improved MaskRCNNPredictor with intermediate feature dimensions expanded to 1024 channels, and an adaptive detection threshold mechanism was implemented in the ROI head to dynamically adjust confidence thresholds according to functional zone density characteristics in each city. Model training adopted stochastic gradient descent with momentum (lr = 0.005, momentum = 0.9, weight decay = 0.0005). The composite loss function weighted classification, bounding box regression, and mask prediction losses at 1.0, 1.0, and 2.0, respectively, to emphasize segmentation boundary accuracy. Data augmentation including random horizontal flipping, random rotation ($\pm 15^{\circ}$), and color jittering was applied throughout training to improve generalization.

3.2. Multi-Source Data Fusion for Functional Zone Classification

The multi-source feature fusion strategy adopted in this study is a late fusion (decision-level fusion) approach. Features are independently extracted from each data source at the functional zone instance level and subsequently concatenated into a 42-dimensional feature vector for input to the random forest classifier. This design was deliberately chosen to accommodate the heterogeneity in spatial resolution, data modality, and semantic granularity among the three sources, avoiding the cross-scale misalignment that would arise with early fusion at the pixel or patch level. The complete feature vector comprises three groups: (1) 18 remote sensing features, including spectral indices (NDVI and NDBI mean and standard deviation), eight gray-level co-occurrence matrix (GLCM) texture descriptors (contrast, correlation, energy, homogeneity, entropy, variance, dissimilarity, and angular second moment), and four geometric shape descriptors (area, perimeter-to-area ratio, compactness, and elongation); (2) two nighttime light features (mean and standard

deviation of VIIRS digital number values within the instance boundary); and (3) 22 POI features (KDE-derived density statistics for each of the 20 major Amap POI categories, dominant category proportion, and Shannon entropy of the categorical POI composition). This feature set provides a three-dimensional representation spanning physical morphology, socioeconomic activity intensity, and functional semantic attributes. These three dimensions are conceptually non-equivalent and do not always co-vary uniformly—for example, residential and commercial zones may exhibit similar built-up morphology and nighttime activity intensity yet differ fundamentally in functional type, and POI records may reflect registered rather than actual usage patterns. The two-stage design of this framework partially addresses this heterogeneity by confining Stage 1 (SFA-Mask R-CNN) to physical-morphological boundary delineation and reserving multi-dimensional feature integration for Stage 2, thereby avoiding conflation of these dimensions in the spatial delineation process [30].

Functional zone classification was performed using the random forest algorithm, which exhibits stable performance when handling multidimensional heterogeneous features and offers strong capabilities for feature importance evaluation and resistance to overfitting [31,32]. Random forest operates by constructing multiple decision trees and applying a voting mechanism, effectively addressing scale differences and noise across diverse data sources. The model training process comprised three stages: feature importance analysis, parameter optimization, and cross-validation. Optimal values for key parameters, including the number of trees and maximum tree depth, were determined through grid search, with five-fold cross-validation used to assess the model's generalization performance.

3.3. Cross-City Model Generalization Experimental Design

To systematically evaluate the adaptability and transferability of deep learning models across cities at different developmental stages, this study designed three sets of comparative experiments focusing on the cross-city application potential of single-city models (Figure 4). The experimental design first assesses the performance of single-city models tailored to specific urban characteristics, then evaluates their direct cross-city transfer capability, and finally introduces domain adaptation techniques to explore potential pathways for enhancing cross-city generalizability. Through comparative analysis of these three experimental settings, cross-city adaptability patterns in urban functional zone identification can be comprehensively revealed, providing scientific basis for constructing scalable automated identification systems.

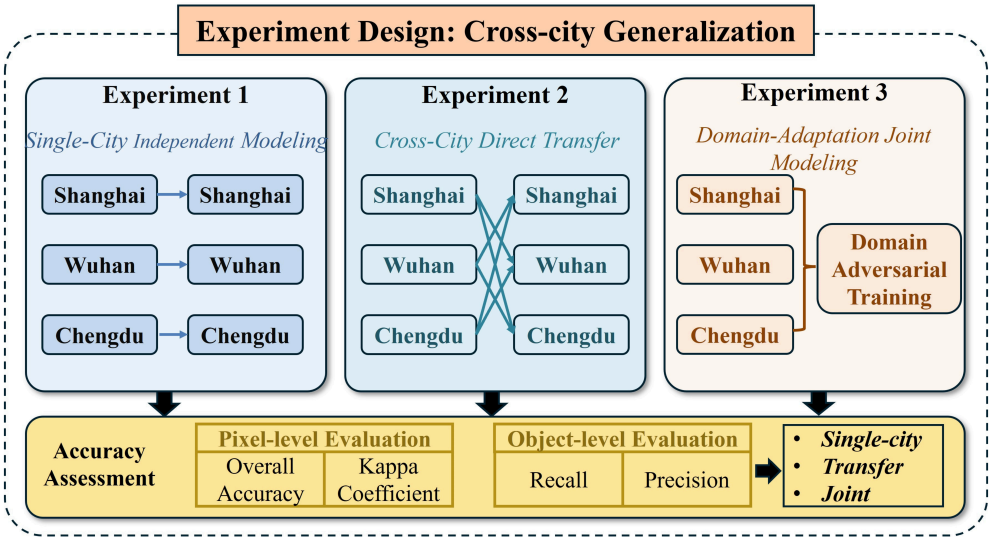


Figure 4. Experimental design for cross-city generalization.

3.3.1. Single-City Independent Modeling and Cross-City Direct Transfer Experiments

The single-city independent modeling experiment aimed to construct optimal functional zone identification models for each city, serving as baselines for evaluating cross-city performance. SFA-Mask R-CNN models were trained independently using data from Chengdu, Wuhan, and Shanghai, with each city providing approximately 150–200 annotated samples of 1024 × 1024 pixels. While the per-city sample count of 150–200 images may appear modest by conventional deep learning benchmarks, we note that each 1024 × 1024 pixel image at 1 m resolution covers approximately 1.05 km^2 of urban area and contains on average 80 annotated functional zone instances. Combined with ImageNet/COCO pre-trained backbone initialization, extensive data augmentation, and the spatially large and morphologically consistent nature of the segmentation targets, this training regime effectively mitigates overfitting and achieves high generalization, as evidenced by the strong single-city accuracy and competitive cross-city transfer performance reported in Section 4.

The training and test sets comprised 70% and 30% of the samples, respectively, with complete spatial separation ensured between them. During model training, targeted parameter adjustments and structural fine-tuning were conducted to accommodate the distinct functional zone characteristics of each city. Specifically, for Chengdu, which features larger-scale functional zones, the SA-FPN module's dynamic scale-aware weighting naturally adapts to the scale distribution of each input image, reducing the need for manual feature-level tuning. In addition, the RPN anchor scales were enlarged to better match the size distribution of Chengdu's larger functional zone instances. For Shanghai, characterized by high functional zone density, the IoU threshold for non-maximum suppression (NMS) was reduced to minimize missed detections in dense areas. For Wuhan, which exhibits diverse characteristics, standard configurations were maintained to balance identification of functional zones across different scales and densities. These city-specific optimization strategies fully exploited the potential of single-city data, thereby providing high-quality baseline models for subsequent cross-city comparisons.

The cross-city direct transfer experiment evaluated model performance when single-city models were applied directly to other cities without any adaptation. The optimal model trained for each city was deployed to the other two cities for functional zone segmentation and classification, yielding nine transfer scenarios (three source cities × three target cities). During transfer, all parameters and configurations of the source-city models were kept unchanged, and inference was performed directly on the target-city test sets. This experimental design enables a pure assessment of the generalizability of the learned feature representations, thereby revealing how urban development characteristics influence model transferability. This experiment constitutes the core of the study, as its results directly address the key scientific question of whether a model trained in one city can be effectively transferred to other cities, thus reducing annotation and training costs for large-scale functional zone mapping. By comparing performance across different transfer directions, the training data sources with the strongest cross-city adaptability can be identified, providing a basis for selecting technical strategies in practical applications.

3.3.2. Domain Adaptation Joint Modeling Experiment

The domain adaptation joint modeling experiment explored the potential for enhancing cross-city model generalizability through domain adaptation techniques [33]. This experiment integrated training data from the three cities and introduced a domain-adversarial training mechanism to reduce feature distribution discrepancies across cities. The domain adaptation framework comprises three core components: a feature extractor, a functional zone segmentor, and a domain discriminator. The feature extractor, built upon

the SFA-Mask R-CNN backbone, extracts deep feature representations of functional zones. The functional zone segmentor performs instance segmentation based on the extracted features. The domain discriminator attempts to identify the city of origin for each sample using its feature representations.

The training process adopted an adversarial learning strategy, optimizing model parameters by minimizing segmentation loss while maximizing domain confusion. Specifically, the feature extractor generates domain-invariant feature representations to "deceive" the domain discriminator while simultaneously optimizing functional zone segmentation performance, thereby preventing accurate identification of sample city origins. Meanwhile, the domain discriminator continuously improves its ability to distinguish samples from different cities. This adversarial interaction encourages the model to learn universal functional zone features shared across cities while suppressing city-specific bias features. The training loss function is designed as:

$$L_{total} = L_{seg} + \lambda L_{domain} \tag{5}$$

where L_{seg} is segmentation loss (including classification loss, bounding box regression loss, and mask loss), L_{domain} is domain adversarial loss (implemented based on gradient reversal layer), and λ is a balance coefficient, with optimal values selected from [0.1, 0.5, 1.0] through grid search.

During joint training, a data balancing strategy was adopted to ensure that samples from different cities were represented in similar proportions within each training batch (approximately 33% per city) through resampling, thereby avoiding bias toward cities with larger datasets. To further enhance the effectiveness of domain adaptation, city-level batch normalization mechanisms were introduced. These mechanisms maintain independent normalization statistics for each city, enabling adaptation to city-specific data distribution characteristics while preserving shared feature extraction capability. It should be noted that this experiment was designed to explore the application potential of domain adaptation techniques in cross-city urban functional zone identification, providing technical references for the development of more universally applicable cross-city identification methods in future research.

3.3.3. Experimental Result Evaluation

Experimental evaluation and comparative analysis employed a multi-level indicator system to comprehensively assess segmentation accuracy and classification performance [34]. Instance segmentation performance was evaluated using both pixel-level and object-level metrics. Pixel-level metrics included overall accuracy (OA) and the F1 score, with OA reflecting pixel-wise classification correctness and the F1 score representing the harmonic mean of precision and recall. Object-level metrics included recall and precision, which assess the model's ability to detect all true functional zones and the accuracy of predicted zones, respectively. Functional zone classification performance was evaluated using overall accuracy, the Kappa coefficient, and the F1 score, enabling analysis of identification accuracy and confusion among different functional zone types [35]. In addition, user's accuracy and producer's accuracy were calculated for each functional zone category to identify the model's strengths and weaknesses in recognizing specific functional zone types.

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

$$Overall\ Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

The comparative analysis was conducted across three dimensions: First, single-city model performance was compared to evaluate the effectiveness of customized optimization for different urban characteristics. Second, cross-city transfer performance was analyzed to identify the most representative source cities and the most challenging transfer directions. Third, the effectiveness of domain adaptation was evaluated by quantifying performance improvements achieved through domain-adversarial training relative to direct transfer and simple data integration. Through this systematic comparative framework, cross-city adaptability patterns of urban functional zone identification methods can be comprehensively revealed, providing a quantitative basis for selecting technical approaches in large-scale automated urban functional zone mapping.

The data partitioning strategy was critical for ensuring result reliability. For each city, datasets were divided into 70% training and 30% testing sets, with complete spatial separation enforced to avoid performance overestimation caused by spatial autocorrelation. In domain-adaptive joint modeling, training data from all cities were used for model optimization, while testing strictly relied on independent test sets for each city, ensuring objective evaluation. All experiments employed the same test sets for performance assessment, allowing direct comparison of results across different modeling strategies.

4. Results

4.1. Ablation Study

To validate the individual contribution of each proposed module, ablation experiments were conducted using the Wuhan dataset. Starting from the standard Mask R-CNN as the baseline, each enhancement module was incrementally added: the Convolutional Block Attention Module (CBAM), the Atrous Spatial Pyramid Pooling module (ASPP), and the Scale-Aware Feature Pyramid Network (SA-FPN). The fully integrated model incorporating all three modules is referred to as SFA-Mask R-CNN. All variants were trained and evaluated under identical experimental settings. Results are reported in Table 1 across four metrics: Overall Accuracy (OA), F1 score, Recall, and Precision.

Table 1. Ablation study results on the Wuhan dataset.

Model Variant	OA (%)	F1 (%)	Recall (%)	Precision (%)
Baseline	65.95	73.43	63.95	58.20
HTC	74.21	84.10	72.97	60.00
Mask2Former	76.23	85.14	81.08	65.21
w/CBAM	72.80	79.15	67.10	61.07
w/ASPP	75.65	79.82	71.80	68.90
w/SA-FPN	77.88	82.06	75.39	72.80
Proposed Model	85.03	89.00	85.54	86.18

The baseline Mask R-CNN achieved an OA of 65.95% and an F1 score of 73.43%, with relatively low Recall (63.95%) and Precision (58.20%), indicating limited capability in detecting and localizing functional zone instances. Adding CBAM improved OA to 72.80% and F1 to 79.15%, with moderate gains in both Recall and Precision. This improvement demonstrates that channel and spatial attention recalibration effectively suppresses irrelevant background information and enhances discriminative feature learning. Adding ASPP alone yielded an OA of 75.65% and F1 of 79.82%, with Recall increasing to 71.80% and

Precision to 68.90%. The multi-scale contextual representations provided by ASPP's parallel atrous convolutions improve the model's ability to perceive functional zones of varying spatial extents. Among the three modules evaluated individually, SA-FPN produced the largest performance gain, achieving an OA of 77.88%, F1 of 82.06%, Recall of 75.39%, and Precision of 72.80%. This result confirms that dynamically adjusting the contribution of each pyramid level according to the scale distribution of the input image is particularly beneficial for functional zone segmentation, where target objects span a wide range of spatial scales. By up-weighting the pyramid levels most relevant to the dominant functional zone scales in the input, SA-FPN enables more effective cross-scale feature aggregation than the fixed-weight fusion used in standard FPN.

The complete SFA-Mask R-CNN model, which integrates all three modules, achieved an OA of 85.03%, F1 of 89.00%, Recall of 85.54%, and Precision of 86.18%, substantially outperforming all partial configurations. Notably, the performance gain from the full model exceeds the sum of individual improvements, suggesting that the three modules are complementary rather than redundant. CBAM enhances feature discriminability, SA-FPN enables adaptive cross-scale fusion, and ASPP expands the effective receptive field—together forming a coherent and mutually reinforcing architecture. These results validate the design rationale of SFA-Mask R-CNN and justify the inclusion of all three modules.

To further validate the superiority of SFA-Mask R-CNN over existing state-of-the-art instance segmentation models, we additionally compared our method against HTC (Hybrid Task Cascade) and Mask2Former under identical training/testing conditions on the Wuhan dataset. As shown in Table 1, SFA-Mask R-CNN outperforms both methods across all metrics, exceeding HTC by 10.82% in OA and Mask2Former by 8.80% in OA. The particularly large margins in Precision (+26.18% over HTC, +20.97% over Mask2Former) indicate that general-purpose segmentation frameworks struggle with the boundary localization demands of urban functional zone identification, a limitation that the proposed SA-FPN's dynamic scale-aware weighting mechanism specifically addresses.

4.2. SFA-Mask R-CNN Segmentation Performance Evaluation

This study independently constructed SFA-Mask R-CNN instance segmentation models using training data from Shanghai, Wuhan, and Chengdu and evaluated their performance on the corresponding city-specific test sets. Both pixel-level metrics—overall accuracy (OA) and F1 score—and object-level metrics—recall and precision—were employed for comprehensive performance assessment (Figure 5). Pixel-level metrics reflect the accuracy of pixel-wise classification, whereas object-level indicators evaluate the model's capability to detect and localize complete functional zone instances.

From the pixel-level perspective, the Chengdu model achieved the best performance on its local test set, with an overall accuracy of 85.36% and an F1 score of 89.91%. The Wuhan model showed comparable performance, with an overall accuracy of 85.03% and an F1 score of 89.00%. In contrast, the Shanghai model performed relatively weaker on its local test set, with an overall accuracy of 83.51% and an F1 score of 85.55%, ranking lowest among the three cities. These results indicate that the functional zone characteristics of Chengdu and Wuhan exhibit strong learnability for deep learning models in pixel-level classification tasks. In contrast, certain characteristics of Shanghai's functional zones increase the complexity of pixel-level classification.

Object-level indicators exhibited patterns that were not entirely consistent with pixel-level results. The Shanghai model achieved the highest recall at 89.81%, substantially exceeding that of Chengdu (77.53%) and Wuhan (85.54%), indicating a strong capability to detect complete functional zone instances and effectively avoid omissions. Although the Chengdu model attained the highest F1 score, its relatively low recall suggests limitations in

identifying all functional zone instances, with some being only partially detected or missed. The Wuhan model demonstrated more balanced performance across object-level metrics, with recall of 85.54% and precision of 86.18%, reflecting stable segmentation capability. In terms of precision, the Wuhan model (86.18%) and the Shanghai model (87.35%) both outperformed the Chengdu model (81.88%), indicating that the Wuhan and Shanghai models achieved more accurate boundary localization with fewer false detections.

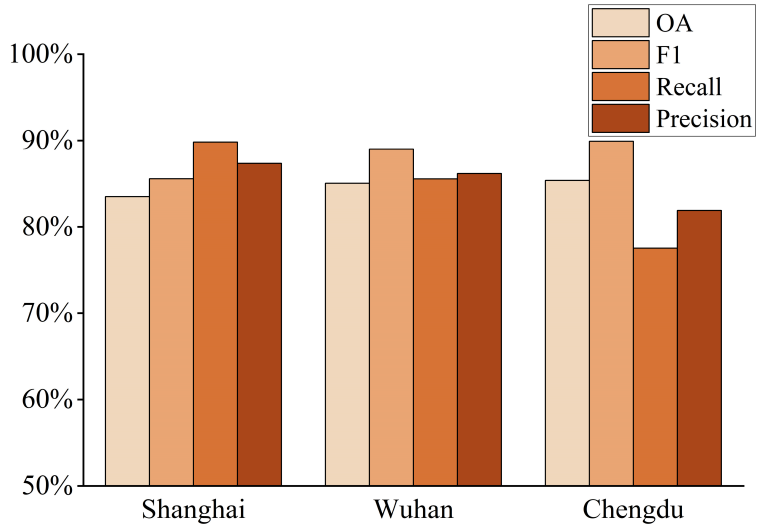


Figure 5. Accuracy of single-city independent models.

The performance differences among the local models of the three cities provide an important foundation for subsequent analysis of cross-city generalizability. The pixel-level accuracy advantages of the Chengdu and Wuhan models, together with the Shanghai model's strong object-level recall performance, reflect the distinct influences of city-specific functional zone characteristics on deep learning model learning processes. These differences are manifested not only in overall performance levels but also in the trade-offs between pixel-level and object-level tasks, thereby establishing a data-driven basis for understanding model transfer performance across cities (Figure 6).

4.3. Cross-City Model Generalization Performance Comparison

4.3.1. Single-City Model Cross-City Transfer Performance

The single-city independent modeling cross-city application experiment systematically evaluated model generalization capability through nine cross-city scenarios (three source cities × three target cities). The results show that models trained on different cities exhibit pronounced performance differences and clear asymmetries when applied to other cities (Figure 7).

The Wuhan model demonstrated the strongest generalization capability in cross-city transfer. When applied to Chengdu, the Wuhan model achieved an overall accuracy of 85.37% and an F1 score of 89.44%. This performance is comparable to the Wuhan model's local results (OA = 85.03%, F1 = 89.00%) and even slightly exceeds the performance of the Chengdu model on its own test data (OA = 85.36%, F1 = 89.91%). At the object level, the Wuhan model achieved a recall of 87.53% and a precision of 91.70% in Chengdu, indicating a well-balanced detection performance. When transferred to Shanghai, the Wuhan model achieved an overall accuracy of 80.43% and an F1 score of 85.42%. Although these values are lower than those obtained in Chengdu, they remain relatively high. Notably, the Wuhan model's object-level precision in Shanghai reached 91.43%, comparable to that of the locally trained Shanghai local model, demonstrating strong transferability in functional zone boundary localization accuracy.

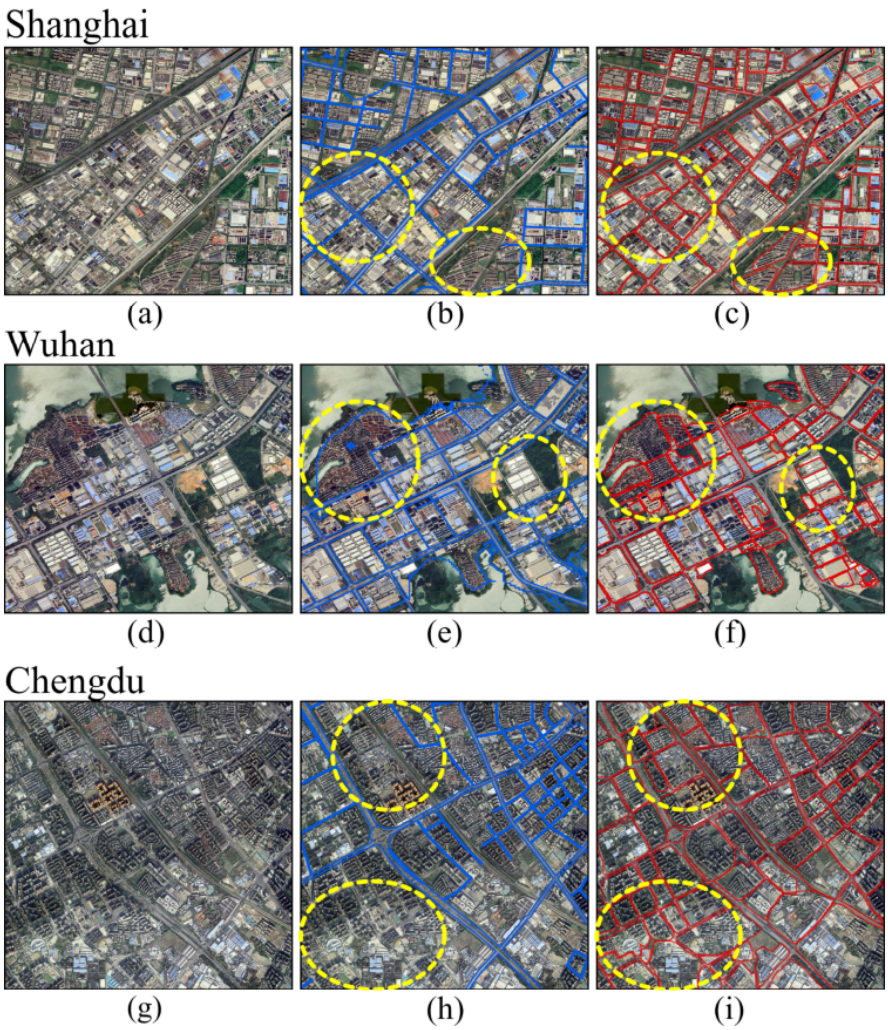


Figure 6. Instance segmentation results for each city ((**a**,**d**,**g**) are true-color images; (**b**,**e**,**h**) are EULUC-China segmentation results; (**c**,**f**,**i**) are instance segmentation results of this study).

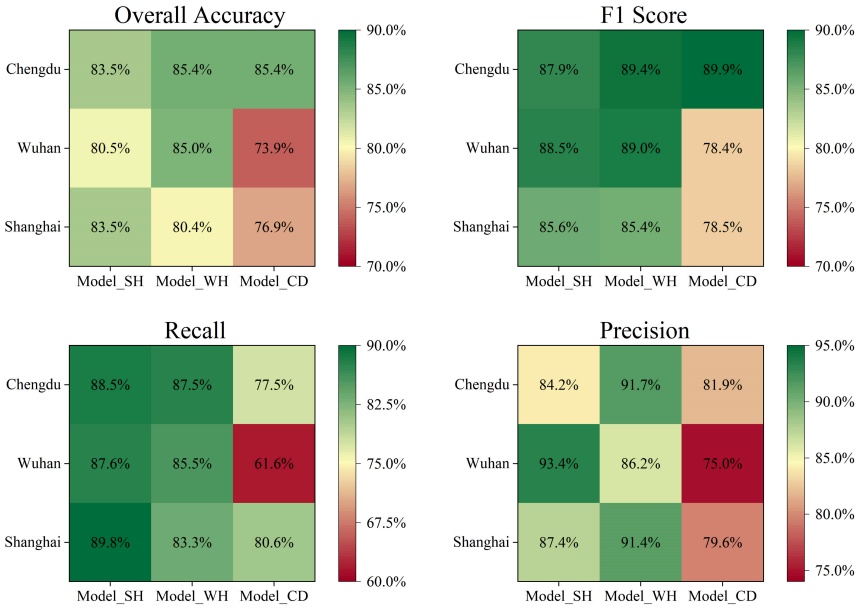


Figure 7. Comparison of cross-city transfer performance of single-city models (Model_SH represents the single-city model generated using Shanghai data; Model_WH represents the single-city model generated using Wuhan data; Model_CD represents the single-city model generated using Chengdu data).

The Shanghai model also demonstrated strong cross-city adaptability. When applied to Wuhan, the Shanghai model achieved overall accuracy of 80.53%, an F1 score of 88.52%, and a recall of 87.57%. When transferred to Chengdu, it attained an overall accuracy of 83.46% and an F1 score of 87.88%, indicating stable performance. Notably, bidirectional transfer between the Shanghai and Wuhan models exhibited nearly symmetric performance characteristics (Figure 8). The Wuhan model achieved an overall accuracy of 80.43% in Shanghai, while the Shanghai model achieved 80.53% in Wuhan, indicating minimal performance difference. This symmetry suggests that functional zones of Wuhan and Shanghai may share a degree of similarity or complementarity in their spatial characteristics.

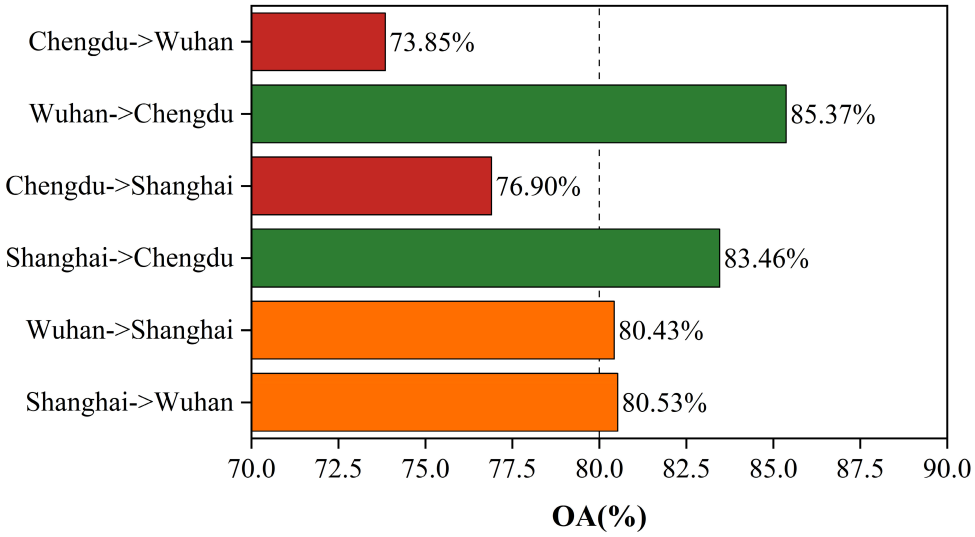


Figure 8. Overall accuracy of single-city models in different transfer directions. (The dashed line represents 80%).

The Chengdu model exhibited notably poor cross-city transfer performance, with the most pronounced performance degradation across all transfer scenarios. When applied to Wuhan, overall accuracy declined to 73.85% and the F1 score to 78.37%, representing a substantial drop relative to the Chengdu model's local performance. More critically, the Chengdu model's object-level recall in Wuhan was only 61.64%, the lowest value among all test scenarios, indicating severe functional zone omission, with many instances failing to be effectively detected. When applied to Shanghai, the Chengdu model's overall accuracy increased slightly to 76.90% and the F1 score to 78.47%; however, overall performance remained weak. Although recall improved to 80.56% and precision to 79.60%—both higher than in Wuhan—these values still lagged considerably behind the cross-city transfer performance of the other models.

From the directional perspective of cross-city transfer, the experimental results exhibit clear asymmetries. Transfer between Wuhan and Chengdu shows strong unidirectional effectiveness: the Wuhan model achieved an overall accuracy of 85.37% when transferred to Chengdu, whereas the Chengdu model achieved only 73.85% when transferred to Wuhan, a difference of 11.52%. A similar asymmetry is observed between Shanghai and Chengdu, with the Shanghai model achieving an overall accuracy of 83.46% when transferred to Chengdu, compared with 76.90% for the Chengdu model transferred to Shanghai, a difference of 6.56%. In contrast, bidirectional transfer between Wuhan and Shanghai exhibited much greater symmetry, with an overall accuracy difference of only 0.10%, representing the most balanced transfer performance among all cross-city scenarios.

4.3.2. Domain Adaptation Joint Model Performance

The domain adaptation joint modeling strategy integrated training data from the three cities with the aim of enhancing cross-city generalization by increasing training sample diversity. During data preparation, a balancing strategy was applied to ensure that samples from each city contributed similar proportions to the training set (approximately 33% per city) through resampling, thereby preventing bias toward cities with larger datasets. However, the experimental results indicate that the joint model did not achieve the expected performance gains (Figure 9). Specifically, it failed to outperform either the best single-city models or the direct cross-city transfer models when evaluated across all three cities.

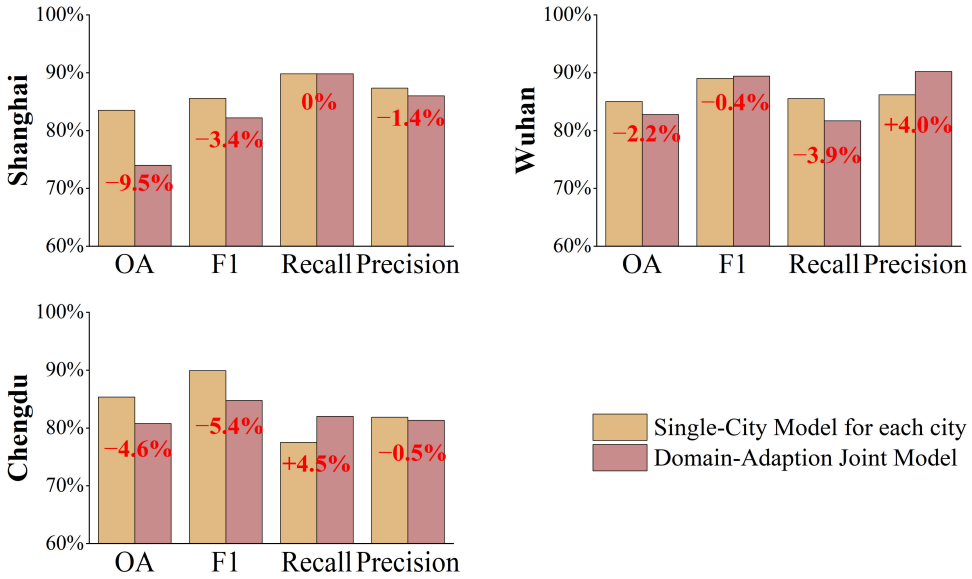


Figure 9. Accuracy comparison between single-city independent modeling and domain adaptation joint modeling.

The joint model achieved an overall accuracy of 82.79% and an F1 score of 89.42% on the Wuhan test set. While this performance exceeded that of most cross-city transfer scenarios, it remained inferior to the locally trained Wuhan model (OA = 85.03%). At the object level, the joint model achieved a recall of 81.69% and a precision of 90.20%. Although precision remained high, the reduced recall compared with the Wuhan local model's 85.54% indicates that the joint model compromised its ability to detect complete functional zone instances.

On the Chengdu test set, the joint model achieved an overall accuracy of 80.78% and an F1 score of 84.77%, which were clearly lower than those of the Chengdu local model (OA = 85.36%) and substantially lower than the Wuhan model's transfer performance in Chengdu (OA = 85.37%). Object-level recall and precision were 82.02% and 81.34%, respectively, with both metrics failing to match the performance of either the Chengdu local model or the best cross-city transfer model. These results indicate that the joint training strategy had a negative effect in Chengdu. Despite increased data diversity, the approach did not improve the model's ability to learn the specific functional zone characteristics of Chengdu.

The joint model performed worst in Shanghai among the three cities, achieving an overall accuracy of only 74.01% and an F1 score of 82.19%. Compared with the Shanghai local model (OA = 83.51%), performance declined by 9.50 percentage points, representing the largest reduction relative to single-city optimal models. Although object-level recall remained high at 89.81%, comparable to the Shanghai local model, overall segmentation quality was substantially inferior to that of single-city models. Precision reached 85.99%,

which is within an acceptable range; however, the pronounced decline in overall performance indicates that the joint model was ineffective in capturing the complex characteristics of Shanghai's functional zones.

Comparisons between the joint model and single-city local models reveal performance declines across all cities. In Wuhan, the joint model showed a decrease of 2.24% relative to the local model; in Chengdu, the decline was 4.58%; and in Shanghai, the reduction was largest at 9.50%. This overall degradation indicates that simple domain adaptation-based joint training strategies failed to effectively enhance model generalization and may even impair performance in individual cities due to training data heterogeneity. Further comparison with the optimal cross-city transfer models (e.g., the Wuhan model applied to Chengdu and Shanghai) shows that the joint model also failed to demonstrate any advantage. These results suggest that, under the current experimental settings, carefully selecting representative source cities for model training can yield better cross-city application performance than simply merging data from multiple cities.

4.4. Functional Zone Classification and Spatial Pattern Analysis

Based on the functional zone instance boundaries obtained from SFA-Mask R-CNN, this study integrated remote sensing imagery features, nighttime light data, and POI data, and applied a random forest algorithm to achieve intelligent functional zone classification for the three cities. Classification accuracy evaluation results indicate that Wuhan achieved the highest classification performance, with an overall accuracy of 93.42% and a Kappa coefficient of 0.895. Shanghai achieved an overall accuracy of 91.97% with a Kappa coefficient of 0.860, while Chengdu achieved an overall accuracy of 91.21% with a Kappa coefficient of 0.851. All three cities exhibited high classification accuracy; however, the observed differences reflect the significant influence of urban developmental stages and spatial patterns on functional zone identification performance.

The spatial distribution of functional zones in the three cities exhibits pronounced regional differences (Figure 10). Shanghai demonstrates a typical mature metropolitan multi-center network pattern, forming an interconnected system of specialized functional centers such as Lujiazui, Zhangjiang, and Hongqiao. Residential areas display a concentric distribution pattern, while industrial functional zones are primarily concentrated in suburban areas beyond the outer ring. Wuhan's functional zone distribution presents a prominent riverine axial development model. The Yangtze River and Han River naturally divide the city into three major sections, each forming relatively independent yet interconnected functional clusters, fully reflecting the strong influence of the "two rivers and four banks" geographical setting on urban functional layout. Chengdu's functional zone distribution exhibits a clear radial expansion pattern centered on Tianfu Square, with functional zones expanding outward along major transportation corridors to form a "one center, multiple axes" spatial structure. Rapid urban spatial expansion has resulted in extensive urban-rural transition zones, characterized by mixed residential, industrial, and agricultural land-use features.

Compared with the EULUC-China dataset, which reports a Level I overall accuracy of 61.2% across Chinese cities (approximately 45% in Chengdu, 75% in Shanghai, and 80% in Wuhan), the proposed framework achieved substantially higher classification accuracy in all three study cities. Specifically, the improvements over EULUC Level I accuracy were +46.21 percentage points in Chengdu (91.21% vs. ~45%), +16.97 percentage points in Shanghai (91.97% vs. ~75%), and +13.42 percentage points in Wuhan (93.42% vs. ~80%). Notably, the most pronounced improvement was observed in Chengdu, a city in a rapid expansion stage where EULUC's performance was relatively weak, suggesting that the

proposed framework's SA-FPN and multi-source data fusion strategy are particularly effective for emerging western cities with large-scale, heterogeneous functional zones.

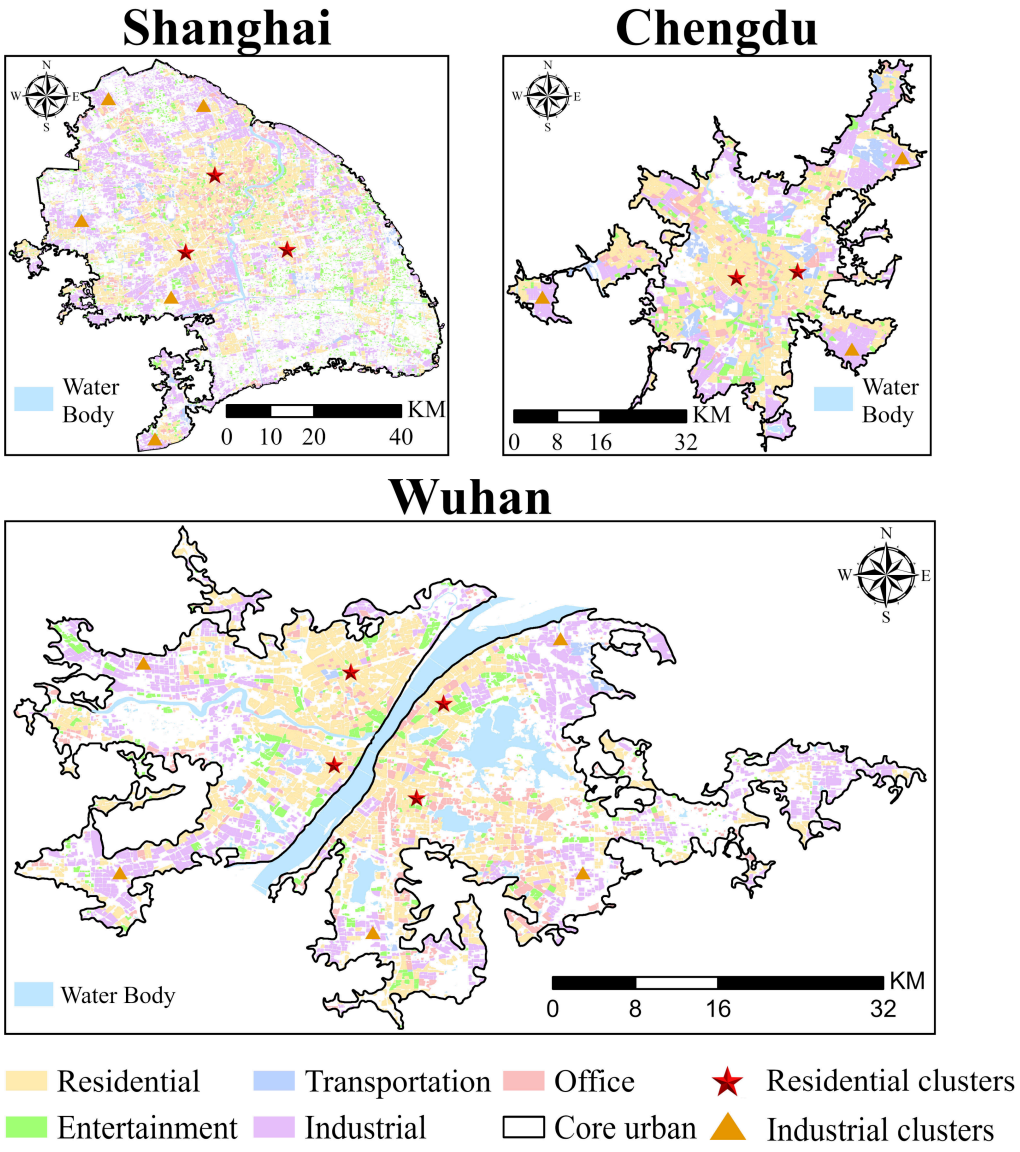


Figure 10. Functional zone classification results of core urban areas in the three cities (core urban area data sourced from [36]).

5. Discussion

5.1. Coupling Mechanism Between Urban Developmental Stages and Model Generalization Capability

Cross-city model generalizability experiments revealed the profound impact of urban developmental stages on deep learning model performance. This influence affects not only local recognition accuracy but, more importantly, determines cross-city adaptability. The results show that the Wuhan model exhibits the strongest cross-city transfer capability, indicating the feasibility of using a single model to achieve large-scale, multi-city automated functional zone mapping.

The Wuhan model's strong generalization capability stems from the unique representativeness of its training data [37]. As a core city in the middle reaches of the Yangtze River, Wuhan is in a transitional stage where rapid development and structural optimization coexist, with functional zone characteristics combining the standardization of mature eastern cities and the dynamism of emerging western cities. This characteristic allows Wuhan's

functional zones to encompass a wide range of spatial morphology types, including highly concentrated core commercial areas alongside large-scale industrial parks and dispersed residential areas, as well as traditionally well-defined functional zones alongside emerging zones with blurred boundaries. Experimental results validate this advantage: the Wuhan model achieved an overall accuracy of 85.37% when applied to Chengdu, slightly outperforming the Chengdu local model (85.36%), and maintained an overall accuracy of 80.43% when applied to Shanghai, which is essentially symmetric with the Shanghai-to-Wuhan transfer performance (80.53%).

More importantly, the internal spatial heterogeneity resulting from Wuhan's "two rivers and four banks" geographical configuration plays a key role in enhancing generalization capability. Wuhan's intermediate position on both scale (mean zone area: 0.059 km^2) and density (16.96 instances/km^2) metrics, combined with the highest within-city CV (2.11) among the three cities, quantitatively grounds this representativeness argument. This characteristic—where a single city encompasses multiple development states—endows the Wuhan model with bidirectional transferability, allowing downward compatibility with emerging cities and upward adaptability to mature cities.

In contrast, the Chengdu and Shanghai models exhibit clear stage-specific limitations. To operationalize these stage designations, we computed three morphological characterization metrics from the instance segmentation outputs: mean functional zone instance area, spatial density (instances per km^2), and within-city coefficient of variation of instance area. Chengdu exhibits the largest mean zone area (0.072 km^2) at the lowest spatial density (13.88 instances/km^2), quantitatively confirming its large-scale, low-density morphological profile; these characteristics learned during its rapid expansion stage show strong regional specificity. When applied to other cities, performance declined markedly, with overall accuracy dropping to 73.85% in Wuhan and object-level recall decreasing to as low as 61.64%. The Shanghai model, meanwhile, faces challenges associated with mature urban complexity: functional zone boundaries often exhibit gradual transitions, substantial internal heterogeneity, and prevalent vertical functional stacking. These characteristics result in relatively lower pixel-level accuracy for the Shanghai local model (83.51%) but excellent object-level recall (89.81%), reflecting the stability and completeness of functional zone spatial patterns in mature cities.

From a regional development perspective, the functional zone pattern differences among the three cities vividly reflect the gradient development characteristics of the Yangtze River Economic Belt. Shanghai's high degree of specialization and intensification, Wuhan's cluster-based diversification, and Chengdu's rapid expansion together constitute the comprehensive spectrum of China's urbanization process. The strong generalizability demonstrated by the Wuhan model in this study indicates that selecting cities in intermediate developmental states that encompass transitional features as training data sources enables models to capture a broader range of developmental stages, thereby achieving an efficient technical pathway of "train once, apply to multiple cities" [38].

The underperformance of domain adaptation can be attributed to three compounding mechanisms. First, the scale differences among the three cities are structural and geometric rather than merely statistical: Chengdu is dominated by large-scale, low-density functional zones, while Shanghai contains dense, fine-grained zones at substantially smaller scales. Domain-adversarial training can reduce statistical feature distribution discrepancies across domains, but cannot resolve geometric scale gaps that operate at the level of spatial structure and object size. Second, jointly training on cities with such contrasting scale characteristics introduces conflicting learning objectives—Chengdu's large-scale zones require broad contextual and low-frequency spatial representations, whereas Shanghai's fine-grained zones demand high-frequency, boundary-sensitive features. Forcing a single

domain-invariant representation under these conflicting requirements produces a suboptimal feature space for both domains, a classic manifestation of negative transfer. Third, the domain discriminator operates on globally pooled backbone features without distinguishing between functional zone categories, resulting in coarse-grained alignment that cannot address category-specific distribution differences across cities (e.g., industrial zones in Chengdu exhibit fundamentally different morphological signatures from those in Shanghai). It should be noted that a grid search over $\lambda \in \{0.1, 0.5, 1.0\}$ was performed and no setting yielded improvement over direct transfer, suggesting that the failure is attributable primarily to structural scale heterogeneity rather than to implementation-specific factors. Future work should explore category-conditional domain adversarial training, where domain alignment is enforced separately for each functional zone type, as well as scale-aware adaptation strategies, to better address these structural inter-city differences.

5.2. Method Advantages and Technical Merits

This study adopted a two-stage strategy of "instance segmentation—intelligent classification," which offers significant technical advantages over end-to-end deep learning approaches. This decoupled design allows deep learning to focus on its strength in spatial morphology recognition—accurately extracting functional zone boundaries through the SFA-Mask R-CNN architecture—while traditional machine learning is responsible for semantic classification based on heterogeneous multi-source data. This clear division of labor avoids feature conflicts that can arise when end-to-end methods simultaneously address morphological identification and semantic understanding, enabling each stage to achieve optimal performance.

The improvements to the SFA-Mask R-CNN architecture constitute the core innovation of this study. Three complementary modules—CBAM, SA-FPN, and ASPP—were introduced to address the key technical challenges of spatial scale variation and blurred boundaries in urban functional zone identification. CBAM enables adaptive focus on salient functional zone features through the cascading of channel and spatial attention, suppressing interference from irrelevant background information. SA-FPN replaces the standard FPN and introduces a dynamic layer-level weighting mechanism that estimates the importance of each pyramid level from the scale characteristics of the input image. By adaptively adjusting the contribution of each FPN level at inference time, the model naturally emphasizes the pyramid levels most relevant to the functional zone scales present in the target city, without requiring any city-level annotations or manual reconfiguration. This design is particularly important for cross-city generalization, as it allows the model to accommodate the different scale distributions of functional zones across cities at different developmental stages. The ASPP module employs parallel atrous convolutions with different dilation rates to expand the receptive field without increasing parameter count, allowing the model to simultaneously perceive functional zones at multiple scales, from small commercial street blocks to large industrial parks. Experimental results demonstrate that these enhancements enabled the model to achieve pixel-level accuracy exceeding 83% in single-city scenarios while maintaining recognition accuracy above 80% in cross-city transfer scenarios.

The multi-source data fusion strategy further enhanced classification performance. The two-stage framework fully exploits the complementary strengths of different data sources by flexibly integrating physical morphological features from remote sensing imagery, functional semantic attributes from POI data, and activity intensity information from nighttime lights in the second stage, thereby achieving a three-dimensional feature representation encompassing "physical—social—functional" dimensions. The random forest algorithm effectively handles scale differences and noise inherent in heterogeneous

data, enabling robust functional zone classification with overall accuracy exceeding 91%, which is substantially higher than results obtained using single data sources. Benchmarking against EULUC-China (Level I overall accuracy: 61.2%), the proposed framework achieved improvements of approximately +13, +17, and +46 percentage points in Wuhan, Shanghai, and Chengdu, respectively, further confirming the effectiveness of the instance segmentation paradigm and multi-source fusion strategy over existing large-scale mapping products.

In addition, the two-stage strategy demonstrates stronger transferability in cross-city applications. Because the spatial morphological characteristics of functional zone boundaries are relatively stable across cities, trained segmentation models can be directly transferred to target cities, requiring only adaptation of the classification model to local characteristics. This substantially reduces annotation costs and computational overhead for large-scale cross-city functional zone mapping.

5.3. Technical Limitations and Future Prospects

This study's methodology still has technical limitations that warrant improvement. First, the approach exhibits a strong dependence on high-quality annotated data. Training SFA-Mask R-CNN requires precise annotations of functional zone boundaries, with each city needing approximately 150–200 annotated samples. The annotation process is not only time-consuming but also requires substantial domain expertise. Although the Wuhan model demonstrated strong cross-city transfer capability, cities with unique geographical environments or distinctive development patterns may still require supplementary local samples for model fine-tuning.

Second, the improved network architecture increases computational complexity. The introduction of CBAM, SA-FPN, and ASPP modules raises the number of model parameters and computational cost, resulting in high processing demands when applied to large-scale urban areas. Recognition accuracy in highly mixed functional zones also remains limited. This limitation is particularly evident in mature cities such as Shanghai, where vertical functional stacking modes—such as ground-floor commercial, mid-floor office, and high-floor residential uses—are prevalent. Remote sensing data primarily capture two-dimensional surface information, certain biases in functional zone classification are unavoidable.

Multi-source data fusion processes also face numerous challenges. Significant differences exist in spatial and temporal resolution, update frequency, and data quality among the various data sources. POI data suffer from issues such as location drift and inconsistent category annotation. The 500 m spatial resolution of nighttime light data limits its ability to represent fine-scale functional zones. These data quality issues, along with uncertainty propagation during the fusion process, may affect the final classification accuracy.

The application effectiveness of domain adaptation techniques fell short of expectations, revealing deeper challenges in cross-city functional zone identification [39]. The high heterogeneity of urban functional zone features—such as Chengdu's large-scale patterns versus Shanghai's small-scale patterns, and Wuhan's cluster-based structure versus Shanghai's networked structure—imposes greater demands on domain adaptation methods. Future research should explore a broader range of adaptation strategies, such as target-city fine-tuning with limited labeled samples and MMD-based feature alignment as an alternative to adversarial training, to better address structural differences among cities and enhance methodological universality in large-scale, multi-city applications.

6. Conclusions

To address key challenges in cross-city urban functional zone identification—including scale differences, boundary ambiguity, and limited generalization capability—this study

proposes the SFA-Mask R-CNN (Scale-Aware Feature-enhanced Attention Mask R-CNN) architecture and constructs a cross-city identification framework integrated with multi-source data fusion. Using three representative cities along the Yangtze River Economic Belt—Chengdu, Wuhan, and Shanghai—as empirical cases, we systematically evaluated the adaptability of deep learning methods across cities at different developmental stages. The SFA-Mask R-CNN architecture effectively overcomes the technical challenges of large spatial scale variation and blurred functional zone boundaries by incorporating three modules: the CBAM attention mechanism for discriminative feature recalibration, the SA-FPN for scale-aware cross-scale feature fusion, and the ASPP module for multi-scale contextual feature extraction. Local models for the three cities achieved high segmentation accuracy on their respective test sets, with pixel-level overall accuracy ranging from 83% to 85% and F1 scores between 85% and 90%, thereby validating the effectiveness of the improved architecture. Furthermore, the two-stage strategy of "instance segmentation—intelligent classification" fully exploits the strengths of deep learning in spatial morphology extraction and multi-source data fusion in semantic classification, resulting in functional zone classification accuracy exceeding 91%.

Cross-city model generalizability experiments revealed the significant impact of urban developmental stages on model adaptability. The Wuhan model demonstrated the strongest cross-city transfer capability, achieving an overall accuracy of 85.37% when applied to Chengdu and maintaining 80.43% accuracy when applied to Shanghai. This indicates that cities in transitional stages—characterized by the coexistence of rapid development and optimization adjustment—possess functional zone characteristics that span a broader spectrum of types. Consequently, models trained on such cities exhibit both downward compatibility with emerging cities and upward adaptability to mature cities. In contrast, the Chengdu model showed markedly weaker cross-city transfer performance, with overall accuracy declining to 73–77% when applied to other cities, reflecting the stage-specific and regional limitations associated with large-scale, low-density characteristics during rapid expansion phases. Domain adaptation joint modeling experiments further revealed that simple multi-city data integration strategies failed to effectively enhance model generalization, suggesting that more sophisticated adaptation techniques are required. The spatial pattern differences in functional zones among the three cities strongly reflect the gradient development characteristics of the Yangtze River Economic Belt. These differences not only affect functional zone identification accuracy but, more importantly, substantially influence model transferability.

Author Contributions: Conceptualization, C.W. and Z.C.; methodology, Z.C.; software, Z.C.; validation, C.W. and S.H.; formal analysis, W.W.; investigation, P.L.; resources, C.W.; data curation, Z.C.; writing—original draft preparation, Z.C.; writing—review and editing, C.W. and S.H.; visualization, Z.C.; supervision, C.W.; project administration, C.W.; funding acquisition, C.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China [grant number 2024YFC3811004], the Natural Science Foundation of Hubei Province [grant number 2024AFB551] and the Major Science and Technology Special Projects and Key R & D Projects in Yunnan Province in 2024 (No. 202403ZC380001). The numerical calculations in this paper were performed on the supercomputing system at the Supercomputing Center of Wuhan University.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: During the preparation of this manuscript/study, the authors used ChatGPT (OpenAI, GPT-5) for the purpose of text polishing and language refinement. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ning, H.; Luoping, Z.; Qiming, L.; Peier, W. The Research on Approach and Method of Sustainable Development-basedUrban Functional Zoning and Application. *Urban Stud.* **2009**, *16*, 63–70. [CrossRef]
2. Ji-rong, G.; Xian-wei, C.; Hai-long, Y. Spatial clustering algorithm on urban function oriented zone. *Sci. Surv. Mapp.* **2011**, *36*, 65–67+64. [CrossRef]
3. Yu, M.; Xu, H.; Zhou, F.; Xu, S.; Yin, H. A Deep-Learning-Based Multimodal Data Fusion Framework for Urban Region Function Recognition. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 468. [CrossRef]
4. Huang, C.; Deng, L.; Fang, C. Satellite Remote Sensing and Dynamic Variation of Ur ban Land Use Changes—Taking Chengdu City as an Example. *Geo-Inf. Sci.* **2007**, *9*, 118–123. [CrossRef]
5. Kuang, W.; Zhang, S.; Liu, J.; Shao, Q. Methodology for classifying and detecting intra-urbanland use change:A case study of Changchun city during the last 100 years. *J. Remote Sens.* **2010**, *14*, 345–355. [CrossRef]
6. Chi, J.; Jiao, L.; Dong, T.; Gu, Y.; Ma, Y. Quantitative Identification and Visualization of Urban Fuctional Area Based on POI Data. *J. Geomat.* **2016**, *41*, 68–73. [CrossRef]
7. Luo, S.; Liu, Y.; Gao, S.; Wang, P. Quantitative identification of urban functional areas based on spatial grid. *Bull. Surv. Mapp.* **2020**, *1*, 214–217. [CrossRef]
8. Kang, Y.; Wang, Y.; Xia, Z.; Chi, J.; Jiao, L.; Wei, Z. Identification and Classification of Wuhan Urban Districts Based on POI. *J. Geomat.* **2018**, *43*, 81–85. [CrossRef]
9. Wang, J.; Ye, Y.; Fang, F. A Study of Urban Functional Zoning Based on Kernel Density Estimation and Fusion Data. *Geogr. Geo-Inf. Sci.* **2019**, *35*, 66–71.
10. Wang, Q.; Dai, Z. The analysis of urban spatial structurebased on POI data and principalcomponent analysis. *Territ. Nat. Resour. Study* **2018**, *6*, 12–16. [CrossRef]
11. Cao, Y.; Liu, J.; Wang, Y.; Wang, L.; Wu, W.; Su, F. A Study on the Method for Functional Classification of Urban Buildings by Using POI Data. *J. Geo-Inf. Sci.* **2020**, *22*, 1339–1348. [CrossRef]
12. Soto, V.; Frías-Martínez, E. Automated land use identification using cell-phone records. In Proceedings of the 3rd ACM International Workshop on MobiArch, Bethesda, MD, USA, 28 June 2011; pp. 17–22.
13. Cai, L.; Hui, F.; Ye, M.; Kang, K.; Zhao, X. Semi-Supervised Urban Land Using Classification MethodBased on Uncertainty Sampling. *J. Jilin Univ. (Inf. Sci. Ed.)* **2016**, *34*, 550–555. [CrossRef]
14. Wang, Y.; Wang, T.; Tsou, M.-H.; Li, H.; Jiang, W.; Guo, F. Mapping Dynamic Urban Land Use Patterns with Crowdsourced Geo-Tagged Social Media (Sina-Weibo) and Commercial Points of Interest Collections in Beijing, China. *Sustainability* **2016**, *8*, 1202. [CrossRef]
15. Yunliang, J.; Moxuan, D.; Jing, F.; Shaowen, G.; Yong, L.; Xinqiang, M. Research on identifying urban regions of different functions based on POI data. *J. Zhejiang Norm. Univ. (Nat. Sci.)* **2017**, *40*, 398–405. [CrossRef]
16. Song, X.; Pu, Y.; Liu, D.; Feng, Y. Mining urban functional areas using pedestrians' movement trajectories. *Acta Geod. Cartogr. Sin.* **2015**, *44*, 82–88.
17. Chen, Y.; Liu, X.; Li, X.; Liu, X.; Yao, Y.; Hu, G.; Xu, X.; Pei, F. Delineating urban functional areas with building-level social media data: A dynamic time warping (DTW) distance based k-medoids method. *Landsc. Urban Plan.* **2017**, *160*, 48–60. [CrossRef]
18. Jiang, G.; Hu, F.; Shi, L. Urban functional area identification based on call detail record data. *J. Comput. Appl.* **2016**, *36*, 2046–2050. [CrossRef]
19. Zhou, W.; Ming, D.; Lv, X.; Zhou, K.; Bao, H.; Hong, Z. SO–CNN based urban functional zone fine division with VHR remote sensing image. *Remote Sens. Environ.* **2020**, *236*, 111458. [CrossRef]
20. Xie, L.; Feng, X.; Zhang, C.; Dong, Y.; Huang, J.; Liu, K. Identification of Urban Functional Areas Based on the Multimodal Deep Learning Fusion of High-Resolution Remote Sensing Images and Social Perception Data. *Buildings* **2022**, *12*, 556. [CrossRef]
21. Lin, A.; Huang, B.; Wu, H.; Luo, W. An MIU-based deep embedded clustering model for urban functional zoning from remote sensing images and VGI data. *Int. J. Appl. Earth Obs. Geoinf.* **2024**, *128*, 103689. [CrossRef]
22. Wu, H.; Luo, W.; Lin, A.; Hao, F.; Olteanu-Raimond, A.-M.; Liu, L.; Li, Y. SALT: A multifeature ensemble learning framework for mapping urban functional zones from VGI data and VHR images. *Comput. Environ. Urban Syst.* **2023**, *100*, 101921. [CrossRef]
23. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
24. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef] [PubMed]
25. Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.

26. Li, X.; Zhou, Y.; Zhao, M.; Zhao, X. A harmonized global nighttime light dataset 1992–2018. *Sci. Data* **2020**, *7*, 168. [CrossRef] [PubMed]
27. Yao, Y.; Li, X.; Liu, X.; Liu, P.; Liang, Z.; Zhang, J.; Mai, K. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 825–848. [CrossRef]
28. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
29. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
30. Zhang, X.; Du, S.; Wang, Q. Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data. *ISPRS J. Photogramm. Remote Sens.* **2017**, *132*, 170–184. [CrossRef]
31. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
32. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
33. Ganin, Y.; Lempitsky, V. Unsupervised domain adaptation by backpropagation. In Proceedings of the International Conference on Machine Learning, Lille, France, 7–9 July 2015; pp. 1180–1189.
34. Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 740–755.
35. Congalton, R.G.; Green, K. *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*; CRC Press: Boca Raton, FL, USA, 2019.
36. Li, X.; Gong, P.; Zhou, Y.; Wang, J.; Bai, Y.; Chen, B.; Hu, T.; Xiao, Y.; Xu, B.; Yang, J. Mapping global urban boundaries from the global artificial impervious area (GAIA) data. *Environ. Res. Lett.* **2020**, *15*, 094044. [CrossRef]
37. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2009**, *22*, 1345–1359. [CrossRef]
38. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.-S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [CrossRef]
39. Tzeng, E.; Hoffman, J.; Saenko, K.; Darrell, T. Adversarial discriminative domain adaptation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7167–7176.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.