Habitat International 167 (2026) 103663

Contents lists available at ScienceDirect

Habitat International

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

Urban vacant land identification and its distribution rules of China's urban system based on fast segment anything model

Xinyu Wang [a], Ying Long [a,b,*]

[a] School of Architecture, Tsinghua University, Beijing, 100084, China
[b] Hang Lung Center for Real Estate, Key Laboratory of Ecological Planning & Green Building, Ministry of Education, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Keywords:
Urban Vacant Land
Deep Learning
Segment Anything Model
Zipf's Law
Scaling Law

ABSTRACT

Urban Vacant Land (UVL) is both a resource and a challenge for sustainable urban development. However, large-scale UVL identification remains understudied. This study addresses this gap by proposing an innovative automated UVL identification method utilizing the cutting-edge Segment Anything Model (SAM), applied across all 2446 Natural Cities (NCs) in China. Our findings reveal several distribution patterns. First, the UVL ratio, which refers to the proportion of the UVL area in each NC, follows a log-normal distribution and remains independent of city size. Second, the UVL area adheres to Zipf's law and scaling law, where larger cities tend to have larger UVL areas. Third, we identify five distinct UVL spatial types based on intra-city distribution patterns. In large cities, UVL tends to cluster to form local types, while in smaller cities, they are more dispersed, forming central, peripheral, and scatter types. Forth, regional analysis reveals significant spatial heterogeneity in UVL types across China. Global and peripheral types require special attention, as they present unique challenges due to high UVL ratios and larger average UVL sizes. This study not only advances the methodological framework for UVL identification and providing a comprehensive UVL dataset for China, but also delivers actionable insights for sustainable urban development through the application of AI technology.

1. Introduction

The accelerated pace of urbanization and economic growth in recent decades has triggered unprecedented metropolitan expansion and construction booms (Liu et al., 2020). However, behind these achievements, other urban processes are emerging, including deurbanization, population loss, and urban shrinkage, which have caused inadequate utilization and land abandonment (He et al., 2017; Kim et al., 2018; Wang et al., 2024). Insufficient demands and utilization will bring urban vacant land (UVL), not only in shrinking cities but also in developing cities, resulting in widespread issues across the globe (Tu et al., 2024).

UVL typically refers to land within a city that remains unoccupied (Smith et al., 2017). Compared with other vacant land and wilderness, UVL is land within built-up areas or urban boundaries, where there are not many buildings (Bowman & Kim, 2016; López et al., 2021). In addition, in UVL areas, there is a lack of stable economic or social activities, including communications in public spaces, commercial activities, and other functional activities (Kim et al., 2015; Park et al., 2021). Newman and Kim (2017) named these nonproductive spaces with low regeneration potential as 'urban shrapnel'. A recent survey of major US cities found that most UVL is small (70.7 %), irregularly shaped (39.7 %), and poorly located or disconnected (41.4 %), posing uncertainty for urban regeneration (Bowman & Kim, 2016).

The cause of UVL is the mismatch between population demands and land use support. A decrease in population directly reduces demand in various sectors such as the economy and society, thus causing economic decline (Accordino & Johnson, 2000). The economic downturn leads to a reduction in job opportunities, causing significant outward migration of the population (Eraydin & Özatağan, 2021). When job opportunities spread out and wealthier people move to the suburbs, the already slow city real estate market is even worse (Kiviaho & Toivonen, 2023). The oversupply of land and housing leads to vacant buildings and empty urban spaces (Glaeser & Gyourko, 2005; Steinführer & Kabisch, 2008). For example, changes in the industrial division of labor due to economic globalization have resulted in the loss of 5 million industrial jobs in the United States between 2000 and 2016, leading to significant worker outflow in shrinking cities (Xie et al., 2018). This forces both cities and regions to experience a drain of population and capital, further intensifying the vacancy issue in the housing market, ultimately leading to a serious UVL problem (Martinez-Fernandez & Cristina, 2007). However,

* Corresponding author. School of Architecture, Tsinghua University, Beijing, 100084, China.
E-mail address: ylong@tsinghua.edu.cn (Y. Long).

https://doi.org/10.1016/j.habitatint.2025.103663
Received 23 March 2025; Received in revised form 1 November 2025; Accepted 2 November 2025

even when redevelopment demand exists, redevelopment does not always occur. According to detailed infill data from Germany, multi-family projects and low-density infill follow different decision-making logics, and many vacant parcels remain undeveloped despite having building rights (Ehrhardt et al., 2025).

The emergence of UVL has double-edged impacts on urban spaces and citizens (Zhang et al., 2023; Zhu et al., 2023). Extensive long-term vacant land can lead to widespread urban decline, due to factors including population decline (Rieniets, 2009), ineffective regeneration policies (Németh & Langhorst, 2014), economic recession (Ryan, 2012), and urban decentralization (Audirac, 2009). The surplus leads to a depressed real estate market, directly affecting economic development (Glaeser & Gyourko, 2005), especially for cities experiencing fiscal crises (Kim et al., 2018; Kremer et al., 2013; Németh & Langhorst, 2014). UVL and vacant properties provide hiding places for criminals, thereby increasing crime rates in surrounding areas (Branas et al., 2018). Furthermore, long-term vacancies can foster social isolation and a sense of helplessness among residents, negatively affecting community cohesion (Stern & Lester, 2021). Poor management of vacant land can cause environmental problems, such as garbage accumulation, which in turn affects the mental health of residents and their overall well-being (Garvin et al., 2013). While recent research highlights that activating vacant land can improve spatial efficiency and public value (Wang et al., 2025), scholars also caution that such interventions may lead to ecological trade-offs and uneven development (Qu et al., 2025).

However, some researchers view UVL as a resource for sustainable urban development, playing a positive role in the urban system (Pagano & Bowman, 2000). Related research highlights vacant lots as intentional landscapes that can provide stormwater volume retention benefits and reduce significant water flow into deteriorating wastewater systems (Kelleher et al., 2020). In Detroit, vacant land is linked to higher bird diversity, whereas roads and other urban features are associated with reduced bird occurrence and species richness (Dennison et al., 2025). In Benin City, UVL supports informal and community-driven temporary uses, highlighting the role of tactical urbanism as context-appropriate approaches in developing cities (Agheyisi, 2025). In addition, its existence offers numerous opportunities for urban development, including improving the ecological environment and social value of the city, attracting many designers and scholars for planning and design discussions (Kim et al., 2018).

Given the significant opportunities and uncertainties that UVL brings for future development, research on UVL is extremely important. Although existing studies have made important progress in identifying UVL, several limitations remain. First, several studies are confined to individual cities or small regions, lacking broader spatial generalizability, which limits the adaptability of their methods and hinders meaningful cross-city comparisons. For example, Smith et al. (2017) identified the distribution of UVL in Phoenix, Arizona, using remote sensing and cadastral data. Castro (2022) explored the spatial characteristics and land use status of UVL in Chicago. Snow et al. (2017) conducted an open-source and cloud-based UVL survey in Clayton, Pennsylvania. In China, Hu and Zhuang (2024) applied high-resolution remote sensing imagery and the SegFormer model to identify UVL in Hangzhou. Cheng et al. (2025) employed a similar architecture with multiscale feature modules, achieving high segmentation accuracy in UVL identification in Chengdu. Lin et al. (2025) apply explainable AI and geospatial similarity reasoning to identify underutilized urban land, offering an interpretable and data-efficient framework for spatial regeneration planning. While these studies represent meaningful progress in developing UVL identification methods and contribute to local urban planning, they still lack standardized classification and consistent evaluation indicators, such as accuracy and efficiency, that would address cross-city or national-level patterns.

Second, although several recent studies have attempted to conduct identification at a national scale, they remain limited in data consistency, coverage, or urban representativeness. Xu and Ehlers (2022) introduced multiple open data, including Open Street Map tags, Wikidata and social media open data to extract urban vacant land for Germany at a national scale. However, the heterogeneity of these data sources may introduce uncertainty and limit their applicability in other regions. Mao et al. (2022) developed an automatic framework for identifying UVL based on the DeepLab V3 model and applied it to 36 major metropolitan cities in China. Tu et al. (2024) extended the scope to 497 shrinking cities worldwide, combining deep learning techniques with manual labeling to examine the expansion of UVL. These studies have limitations in offering systematic insights into the broader urban context and in providing a unified understanding of UVL across diverse city types.

Third, these limitations have been paid to the spatial analysis of UVL and to a deeper understanding of its distribution patterns. Due to the inconsistency of datasets, it is challenging for researchers to develop a further understanding of how UVL is spatially organized. Most existing studies focus on UVL identification as a binary classification problem, aiming to determine whether an area is vacant and to delineate its precise boundary, while neglecting the broader spatial characteristics of UVL, such as clustering patterns or relationship with urban fabric.

To address these gaps, this research develops a deep learning-based model to identify UVL at a national scale and applies it to China. Recent advances in computer vision have enabled a more efficient segmentation model for urban studies. Streamlined segmentation models such as U-Net and DeepLabV3+ are effective in structured environments. However, they often struggle with irregular object boundaries and the detection of small or fragmented features (Chen et al., 2018). The Segment Anything Model (SAM), introduced by Kirillov et al. (2023), offers a more flexible and high-capacity segmentation architecture based on transformer modules and prompt-aware mask decoding. Several evaluations have shown that SAM achieved better performance in preserving fine-grained boundaries, generating high-quality object masks with minimal human supervision, and identifying small or irregular regions, particularly in high-resolution and complex urban imagery (Li et al., 2025; Ma et al., 2024; Osco et al., 2023). However, SAM requires substantial computing resources, making it impractical for deployment on standard hardware. The FastSAM, a lightweight model successfully reduces the requirement, by training on only 2 % of SAM's original dataset while achieving comparable performance, which makes it suitable for large-scale UVL identification using personal or resource-constrained hardware environments (Zhao et al., 2023).

Building on this identification method, this study aims to advance the understanding of UVL distribution at both national and city levels by analyzing indicators such as total UVL area, UVL ratio, and spatial distribution patterns. We believe these innovative methods and new empirical findings will provide both the public and decision-makers with a comprehensive understanding of UVL spatial boundaries, specific locations, and current conditions (Deng & Ma, 2015). Moreover, the findings will highlight the potential of UVL as valuable assets for urban regeneration, economic revitalization, and community empowerment (Li et al., 2018; López et al., 2021; Song et al., 2020).

2. Materials and methods

This study systematically identifies UVL and its spatial patterns in Chinese cities using advanced deep learning methods (Fig. 1). The research area encompasses the boundaries of 2,446 Natural Cities (NCs), which were reconstructed in 2020 with a minimum threshold of 5 km^2 based on relevant research (Ma & Long, 2020). The research framework is as follows. First, high-resolution satellite images of the research area were downloaded, cleaned, and processed into 55,670 raster patches (2,520 × 2,520 pixels). A total of 5,548 samples were meticulously annotated via the X-AnyLabeling platform to create a COCO-standard UVL dataset, which included both a training set and a test set. Second, the FastSAM model underwent 100 training epochs using the training set, with the F_2 score utilized to measure model performance based on the

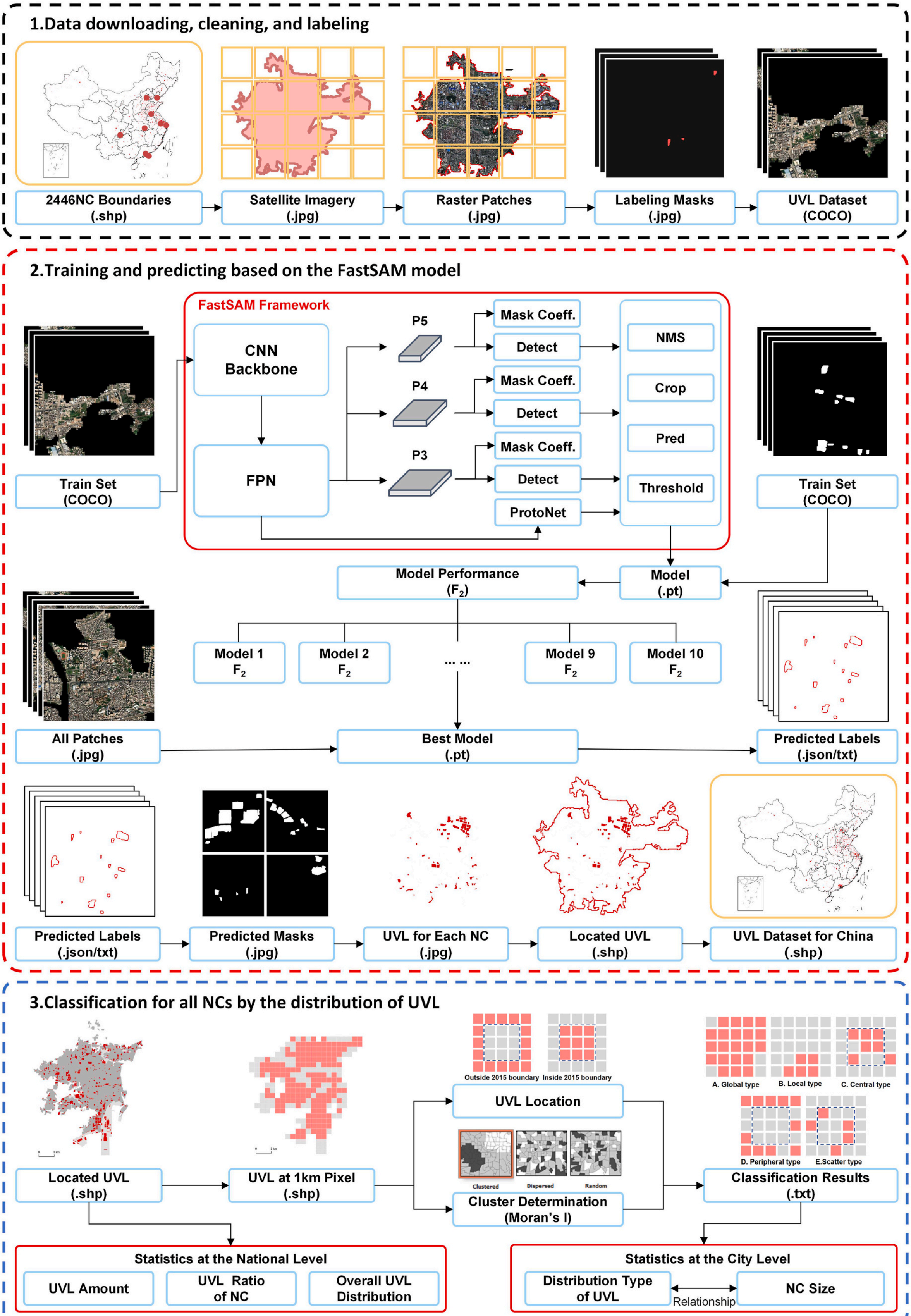

Fig. 1. The overall framework of this study.

test set. This training process was independently repeated 10 times, and the best-performing model was selected as the final output. The trained model was then deployed for full-coverage prediction, followed by necessary format transformation, merging, and georeferenced vector conversion using ArcGIS to generate the UVL dataset for the 2,446 NCs in China. A classification system was developed based on UVL location and cluster determination to support the analysis of UVL at both national and city levels. This included the amount, spatial characteristics, and the relationship between NC size and UVL distribution.

2.1. Research area

In China, researchers have different opinions regarding urban area boundaries. According to the definition of a city in China, the boundary of a city, or urban area defined by the National Bureau of Statistics of China (NBSC), is based on administrative units and utilized for population statistics. However, this boundary, known as administrative city boundaries, still contains a significant amount of rural area, rendering it unsuitable for urban studies, including UVL identification (Jin et al., 2017; Long, 2016; Wang & Long, 2024).

To solve this unsuitability, several conceptions, including natural city boundary, Chengqu (in Chinese), or functional urban areas (FUAs) are proposed (Jiang et al., 2020; Ma & Long, 2020; Wang & Long, 2023). From the perspective of built-up area, these concepts utilize Night Light Imagery, Land Use Cover datasets, or related geography data to identify the accurate boundary of built-up area and aggregate these boundaries into cities based on a certain standard. These boundaries are more accurate and detailed for urban studies and land analysis (Cao et al., 2024). This research introduces the concept of natural cities to obtain China's natural city boundaries (NC boundary) in 2020 (Jiang & Miao, 2015). Natural cities are defined as physically continuous urban agglomerations identified through built-up land continuity and population activity (Long et al., 2018). NCs emphasize that cities should be understood as physical entities shaped by land development and human activity, rather than purely administrative constructs (Jiang et al., 2015; Jiang & Miao, 2015). Such an approach ensures that our analysis focuses solely on urbanized spaces with functional coherence and spatial integrity (Ma & Long, 2020).

When identifying NC boundaries, a minimum threshold is needed to exclude small towns and fragmented rural settlements, because the distribution of natural cities in an urban system follows Zipf's Law, which is a long-tail pattern (Jiang et al., 2015; Long, 2016). In such distributions, a small number of large cities dominate in number, which may introduce noise. Therefore, a 5 km^2 minimum area threshold was adopted to retain only urban entities with sufficient spatial and functional coherence according to previous research (Wang & Long, 2023). We follow the method raised by Ma and Long (2020) to obtain 2446 NCs in China. The detailed method and identification results are demonstrated in the Supplementary Text.

2.2. Definition of UVL

The focus of different studies on the definition of UVL varies. Some scholars emphasize that vacant land reflects specific land-use information related to human activity (Xu & Ehlers, 2022), while others identify UVL directly from remote-sensing images (Mao et al., 2022; Hu & Zhuang, 2024).In this study, UVL is defined as land within urban areas that is vacant in both spatial form and functional use. From a spatial perspective, UVL refers to land with no or almost no buildings on the surface. From a functional perspective, UVL is characterized by the absence of a clear use, meaning that the land is not explicitly utilized or effectively developed. Such parcels generally lack productive, residential, or ecological service functions, and remain in an unused state. Unlike underutilized land, which emphasizes insufficient use, UVL highlights a complete lack of function rather than functional inefficiency.

Mao et al. (2022) manually annotated remote-sensing imagery and categorized UVL into six sub-types, including bare land without film and weeds and demolished structures. In this study, we adopted the criteria proposed by Mao et al. to determine whether a parcel qualifies as UVL. Specifically, UVL includes undeveloped bare land, idle grassland without clear recreational or ecological use, land parcels left vacant after demolition, abandoned construction waste land, and other parcels that have lost their original residential, industrial, or commercial functions. By contrast, active agricultural fields, water bodies, functioning green spaces (e.g., public parks), and construction sites under active development were excluded (Mao et al., 2022; Tu et al., 2024).

2.3. Data downloading, cleaning, and labeling

As discussed in the Introduction, the remote sensing characteristics of UVL are identified based on relevant research, which has provided several examples to determine whether a parcel of land can be categorized as UVL (Mao et al., 2022; Tu et al., 2024).

This study utilizes the Esri World Imagery Wayback dataset as the primary input for identifying UVL (https://livingatlas.arcgis.com/wayback). Esri World Imagery Wayback, a service provided by Esri, allows users to access and browse historical satellite imagery since 2014. With high-resolution imagery covering the globe and an easy-to-use interface, users can select images from specific dates for comparison and analysis of geographical and environmental changes.

Firstly, we downloaded 2023 satellite imagery with a precision of 17 level (about 1.19m), which totals 1080 GB in size from Esri World Imagery Wayback dataset. Secondly, to facilitate subsequent processing by deep learning models, the study utilized Python 3.7 and ArcGIS software (Arcpy) to segment all images into raster patches with a resolution of 2520 $\times$ 2520 pixels, as indicated in related research (Mao et al., 2022). This segmentation resulted in a total of 55,670 patches. Thirdly, ten percent of patches are selected as annotated samples for labeling and divided into train sets and test sets for training our deep learning model. On average, annotating a sample raster patch (2520*2520 pixels) takes approximately 1.2–1.5 min. In a typical day, annotators complete approximately 300–350 image patches, requiring 8–9 h of continuous work. For the overall annotated samples plan, which entails annotating a total of 5548 patches (about 10 % of the national dataset), completion is estimated to require approximately 14 days of annotation work. The platform for labeling is X-AnyLabeling (Wang, 2023).

To improve the accuracy and consistency of annotations, annotators received professional training before conducting the annotation. During the annotation process, regular audits and feedback were conducted by other 12 urban researchers. For each annotated patch, the study recorded the annotation time, the annotator information, and the annotation accuracy assessment. The labeling results (ground truth) of 6 samples are demonstrated in Fig. 2.

Finally, the study selected all patches containing UVL and created a UVL dataset in a YAML file based on the COCO dataset format (Lin et al., 2014). The COCO dataset format is a widely used data annotation format that supports various types of object detection and segmentation tasks. This dataset can serve as input for another deep learning model for training and future UVL, offering valuable data support. The COCO dataset of this research can be freely downloaded at Figshare: https://figshare.com/s/0146bb75bf5a5c025c57.

2.4. Training and predicting based on the FastSAM model

The FastSAM model is a simplified version of the Segment Anything Model (SAM), designed to significantly enhance processing speed while maintaining accuracy (Kirillov et al., 2023; Zhao et al., 2023). By redefining the task into simpler segmentation generation and prompt mechanisms, and using conventional Convolutional Neural Networks (CNN) with an instance segmentation branch, FastSAM can deliver performance comparable to SAM but with a 50-fold increase in speed on

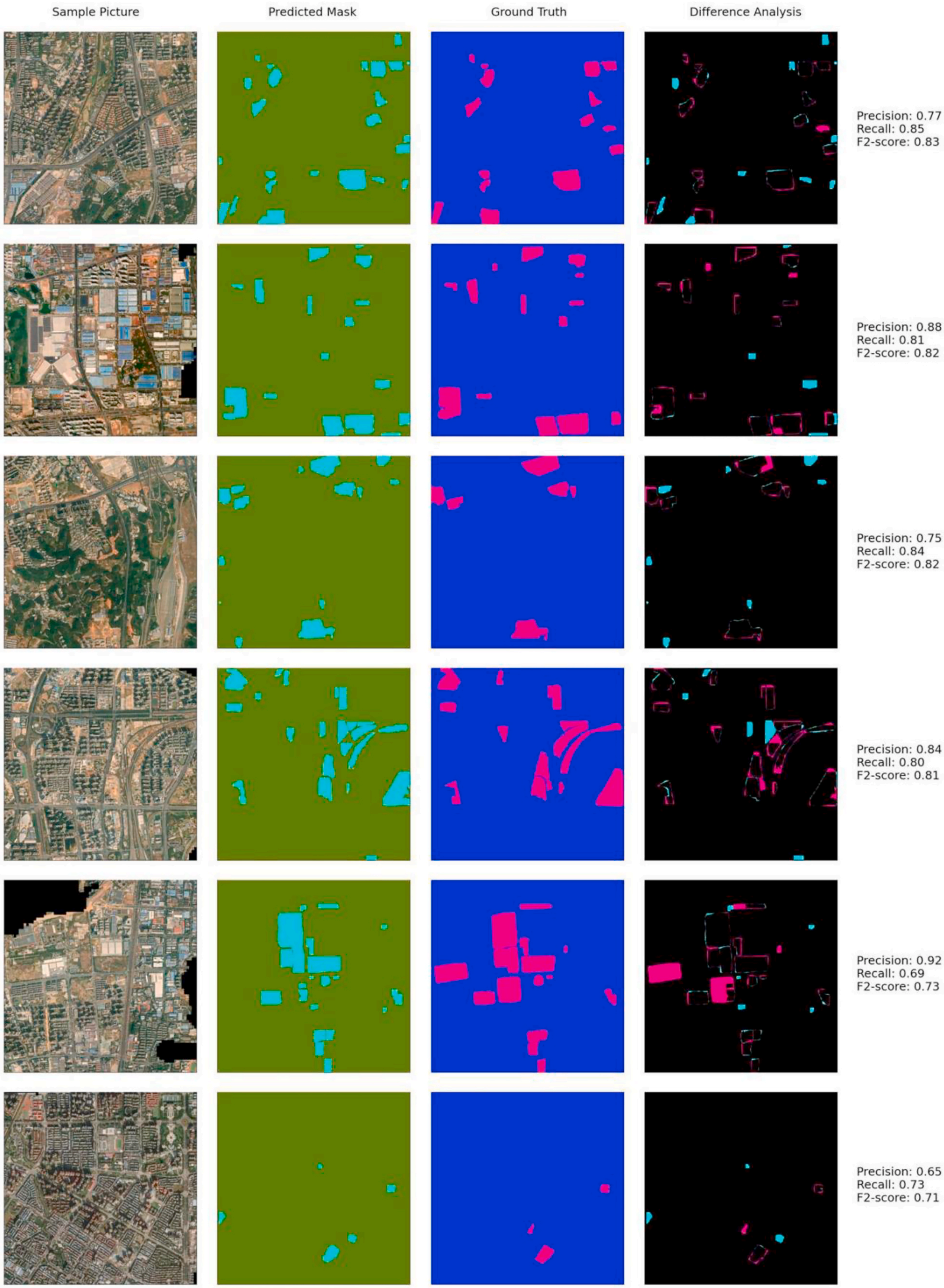

Fig. 2. The comparison between predicted masks and labeling results of 6 samples. Each sample's recall, precision and F_2 score are shown on the right. These samples indicate that the best trained model has reached satisfactory identification results.

the same hardware setup (Kirillov et al., 2023; Zhao et al., 2023). It can process and segment large volumes of images efficiently, making it suitable for real-time applications such as autonomous driving and video surveillance. These advantages offer an opportunity for land-related research to overcome boundary challenges. More information can be found at: https://github.com/CASIA-IVA-Lab/FastSAM.

We utilize the UVL dataset to train our UVL identification model based on the FastSAM model. We utilized the FastSAM-X model and trained it on the UVL dataset for 100 epochs. This training process was repeated for 10 distinct models from which the one with the most exceptional performance was selected. The outcome of this election was a model that achieved the optimal F_2 score. The formula for the F_2 score specifically is:

$$F_2 = \frac{5precision \cdot recall}{4precision + recall}$$

Precision, which measures the model's accuracy, is the number of true positives divided by true positives and false positives. Recall, which measures the model's sensitivity, is true positives divided by true positives and false negatives.

In this study, we adopted the default parameter settings of the FastSAM model, and used a confidence threshold of 0.184 for predicting all patches. Under this configuration, the model achieved a best performance F_2 score of 0.76. Six samples of the predicted result based on the best model were shown in Fig. 2.

The prediction step was performed on a local GPU with CUDA 11.8 and PyTorch 2.0.1, and the entire prediction process took approximately 10 h to complete. Upon completion of the predictions, all predicted images were converted into a vector dataset using ArcGIS 10.8 and ArcPy. The key was to ensure that the generated vector data corresponded accurately with actual geographic locations. Through these steps, the image data was successfully transformed into precise and useable geographic information data.

Finally, we obtained the UVL shapefile of 2446 NCs with actual geographic locations. The research successfully generated a UVL dataset in Chinese cities, providing essential foundational information for urban planning and environmental management.

2.5. Classification for all NCs by the distribution of UVL

To examine the spatial distribution of UVL in each NC, we referenced related research on urban population distribution, which has provided comprehensive classification techniques and grid-scale urban analyses (Blanco et al., 2009; Jiang et al., 2020; Meng et al., 2021; Schetke & Haase, 2008). By employing a similar framework, this method categorized UVL into five primary patterns: global, local, central, peripheral, and scattered.

The global type refers to UVL being spread across most of the city's area. The local type describes UVL concentrated within one or more particular city clusters. UVL in the central type clusters in the old city, and the peripheral type appears near outer boundaries. The scatter type is distributed relatively evenly throughout.

Based on the characteristics of different patterns, the classification framework is arranged in Fig. 3. It is essential to recognize that UVL is confined to specific parts of the city (Pagano & Bowman, 2000), which poses challenges to defining the UVL distribution type and categorizing UVL into some types, such as the global type and local scatter type. Thus, this research utilizes a 1 km grid system to classify the distribution of UVL.

Specifically, grid cells where the proportion of UVL exceeds the average (4.78 %, which represents the percentage of the total UVL area relative to the total NC area according to the final identification result) are designated as 'UVL pixels', while others are classified as non-UVL pixels. The preliminary analysis based on 1 km grids indicated that UVL pixels constitute 29.8 % of the total grids, and 88.1 % of the UVL are covered by UVL pixels. Moreover, if we set a higher proportion for UVL to determine grids as UVL pixels, such as from 4.78 % to 10 %, it results in only 16.6 % of grids qualifying as UVL pixels. This restricts the classification process, particularly in smaller cities that have insufficient UVL pixels.

The classification is based on 1 km UVL pixels. As shown in Fig. 3, in the first step, cities with a UVL pixel ratio greater than 50 %, are classified as global type, indicating that UVL is distributed in most areas of the city. Those with a UVL pixel ratio of less than 50 % will be further classified. In the second step, the study analyzes the degree of clustering of UVL pixels in the remaining cities. If the distribution is clustered, meaning that UVL is concentrated in a certain area of the city, these cities are classified as local type distribution (clustered type). If the distribution is dispersed or random, further classification is needed. The degree of clustering is measured by Global Moran's Index (Moran's I), a widely used method for spatial autocorrelation analysis, as well as its p-value (Liu et al., 2024; Wang et al., 2023). Moran's I, with values

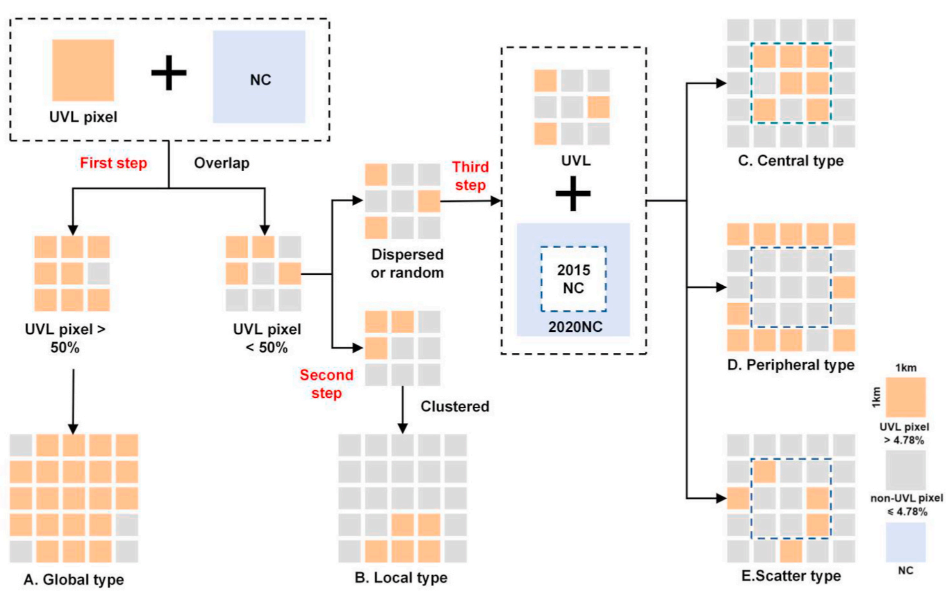

Fig. 3. The classification framework for UVL distribution. NC (Natural Cities): Urban areas identified based on related research. UVL pixel: A 1 km grid cell where the proportion of UVL exceeds the average threshold (4.78 %). Non-UVL pixel: A 1 km grid cell that is not identified as a UVL pixel.

between −1 and 1, indicates both the presence and significance of clustering patterns. Positive values suggest clustering, while negative values imply dispersion; a value closer to ±1 signifies a stronger spatial correlation. Moreover, for all the datasets, the p-value is used to assess the significance of spatial autocorrelation results. A small p-value (typically <0.05) indicates sufficient evidence to conclude that the observed spatial pattern. Thus, in this study, NCs with Moran's I > 0.2 and p-value <0.05 are identified as clustered, and other NCs, which means the UVL distribution pattern is dispersed (Moran's I < −0.2 and p-value <0.05) or random, are served as input for the next step. In the third step, using the 2015 urban boundary data, the study examines the locations of UVL. When a significant concentration of these pixels lies within the 2015 NC boundary, the city is classified as a central type. Conversely, if the concentration is on the periphery, it is considered a peripheral type. When clustering is balanced between the center and periphery, resulting in an indistinct distribution pattern relative to urban boundaries, the city is classified as scatter type. Fig. 4 illustrates representative cities for the five classification types, which gives us a new perspective to understand the UVL status in China.

3. Results

We introduce a method that utilizes the FastSAM model, a streamlined and faster version of the Segment Anything Model (SAM), to identify 204,769 UVL parcels across 2,446 cities (Kirillov et al., 2023; Zhao et al., 2023). The total area of UVL in Chinese cities is determined to be 4,442.1 square kilometers, representing 4.78 % of the total NC area. The results for each city, including its NC area, UVL area, UVL ratio, and UVL distribution type, are provided in Supplementary Data.

3.1. The total area of UVL at the national scale follows Zipf's law

Based on the test method of previous research, we conducted four tests for the UVL dataset to figure out whether the total area of UVL follows Zipf's law (Jiang & Miao, 2015): (1) for all NCs (2436 are available); (2) for NCs with UVL more than 0.1 square kilometers (2294 NCs); (3) for the top 90 % of NCs by UVL area (2193 NCs); (4) for NCs with UVL more than 0.3 square kilometers (1758 NCs). All results are illustrated in Fig. 5, demonstrating that the total area of UVL in each city follows Zipf's law, which posits that the total area of UVL in each city is inversely proportional to its rank in a list at the log scale.

It is important to note that, compared to other studies on Zipf's law in urban research, the fitting indicators for UVL are slightly lower compared with other indicators, like population, urban area, and GDP, particularly in the presence of some cities with lower UVL ratio (Bettencourt, 2013; Bettencourt et al., 2007; Jiang & Miao, 2015). As shown in Fig. 5c and d, when we reduce these cities with some UVL ratio, the R^2 value improves significantly, which confirms the validity of Zipf's law in UVL (Jiang et al., 2015). This phenomenon is because, compared to population and economic indicators in China, UVL is more susceptible to policy and is subject to unique circumstances (Liu et al., 2014). In China, the issues related to property rights have made the demolition of buildings extremely challenging, although the buildings are vacant (Gao et al., 2021). As a result, UVL, which should have emerged in these cities, has manifested itself in the form of vacant properties and vacant houses. China has a significant number of vacant buildings in recent years that have not been converted into UVL, commonly known as ghost cities (Williams et al., 2019). That is one of the reasons why some cities have lower UVL ratios as shown in Fig. 5.

3.2. UVL ratio at the national scale shows log normal distribution

The UVL ratio, defined as the proportion of UVL area relative to the total area of each city, serves as a key indicator for understanding spatial distribution patterns. Interestingly, when ranking all cities according to their UVL ratio (Fig. 6a), the study found a highly linear relationship between the logarithm of ranking and the UVL ratio ($R^2 = 0.992$). This indicates that the UVL follows a log-normal distribution. The R^2 value suggests that this relationship is significant and strong and almost all variations can be explained by this distribution. What's more, there is no relationship between NC area and UVL ratio. This is indicated by a Pearson correlation coefficient of 0.005, reflecting an almost negligible linear relationship, and a p-value of 0.79, which is well above the common significance threshold.

The presence of a log-normal distribution implies that the probability density function (PDF) of the logarithm of UVL ratio corresponds to a normal distribution as shown in Fig. 6b. The study plotted its PDF (the blue line in Fig. 6b and fitted it with a log-normal distribution curve (the red line in Fig. 6b). We can describe the UVL ratio using the mathematical formula of the normal distribution as follows:

$$Log(X) \sim N(\mu, \sigma^2)$$

where X represents the UVL ratio, μ is the scale parameter, and σ is the shape parameter.

This analysis highlights the distributional characteristics of the UVL ratio in an urban system. First, the UVL ratio exhibits a right-skewed distribution with a long tail on the right. This implies that its mode ($e^{\mu-\sigma^2}$), representing the peak of the distribution curve, is smaller than both its median (e^{μ}) and mean ($e^{\mu+\frac{\sigma^2}{2}}$). Among these, the median is a more accurate representation of the central tendency of the distribution, as 64.13 % of cities have a UVL ratio lower than the mean. The fitted curve's median is 3.59 %, while the observed median is 3.74 %. Second, similar to the normal distribution, the UVL ratio's 95 % Confidence Interval (95 % CI) ranges from 0.88 % to 15.06 %. Specifically, 66.89 % of NCs have a UVL ratio below 5 %, and 91.84 % are below 10 %. This suggests that cities with higher UVL ratios experience substantially greater development pressure compared to those with lower ratios, as their deviation from the average UVL ratio is much larger. Therefore, cities with high UVL ratios should be given adequate attention and prioritized for targeted management strategies.

3.3. UVL near periphery has larger area

The relationship between UVL, and the time of its emergence is another key focus of this study. Identification of the years when UVL appeared has rarely been discussed in existing research, making it difficult to analyze. To facilitate subsequent exploration, we adopted an analytical method commonly used in land use studies for further research: using a spatial-to-temporal approach to identify years of UVL (Tavares et al., 2012; Yin et al., 2018). This method involves delineating the city's boundaries during different periods under a unified standard. If a vacant lot is located between the boundaries of year t and year t+1, it is considered to have emerged between t and t+1. This approach is well suited to the rapid expansion of urban areas in China in the past 20 years, allowing a quick and straightforward determination of the years of large-scale UVL (Long et al., 2018). However, it is important to note that this method has limitations; it can only identify the years of UVL within areas of urban expansion, while UVL that arises within older urban districts, such as those from demolitions, cannot be identified. Based on this method, we have made many interesting discoveries.

We utilize the 2015 NC boundaries from Ma and Long (Long, 2016; Ma & Long, 2019), as well as the urban extension dataset from 1987 to 2017 (Gong et al., 2020), to summarize the average area of different eras. The two findings are shown as follows.

First, urban expansion from 2015 to 2023 has accounted for nearly half of UVL in China. It was found that 52.1 % of UVL is located within the 2015 NC boundaries, while 47.9 % is situated outside the 2015 NC boundaries. This indicates that the urban expansion from 2015 to 2023 has contributed to the addition of 98,000 lots of UVL, representing 49.6 % of the total vacant land area by 2023.

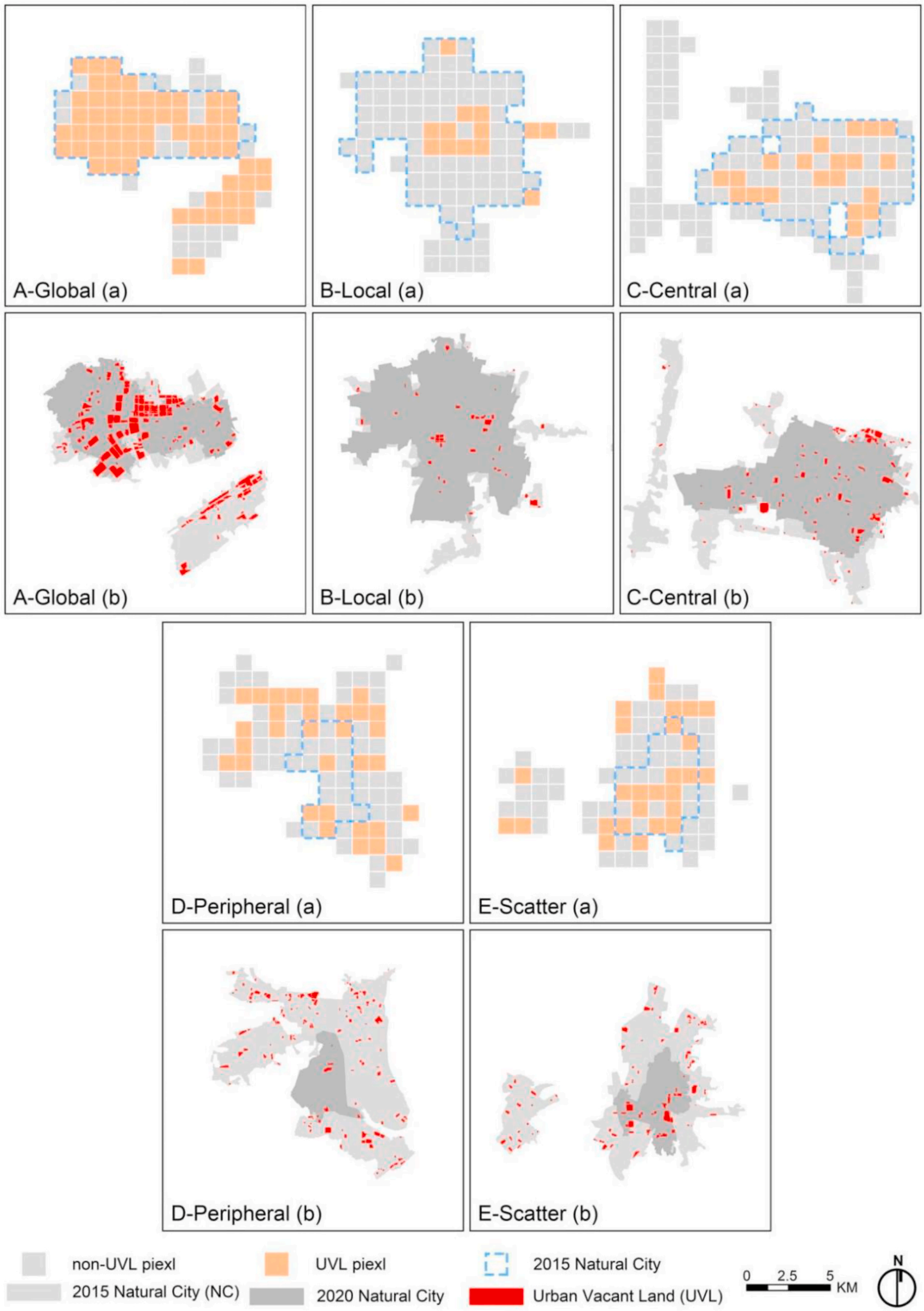

Fig. 4. Typical cities for each of the five classification types. A(a), B(a), C(a), D(a) and E(a) are classification results based on the classification framework. A(b), B(b), C(b), D(b) and E(b) are the original results of the FastSAM identification model.

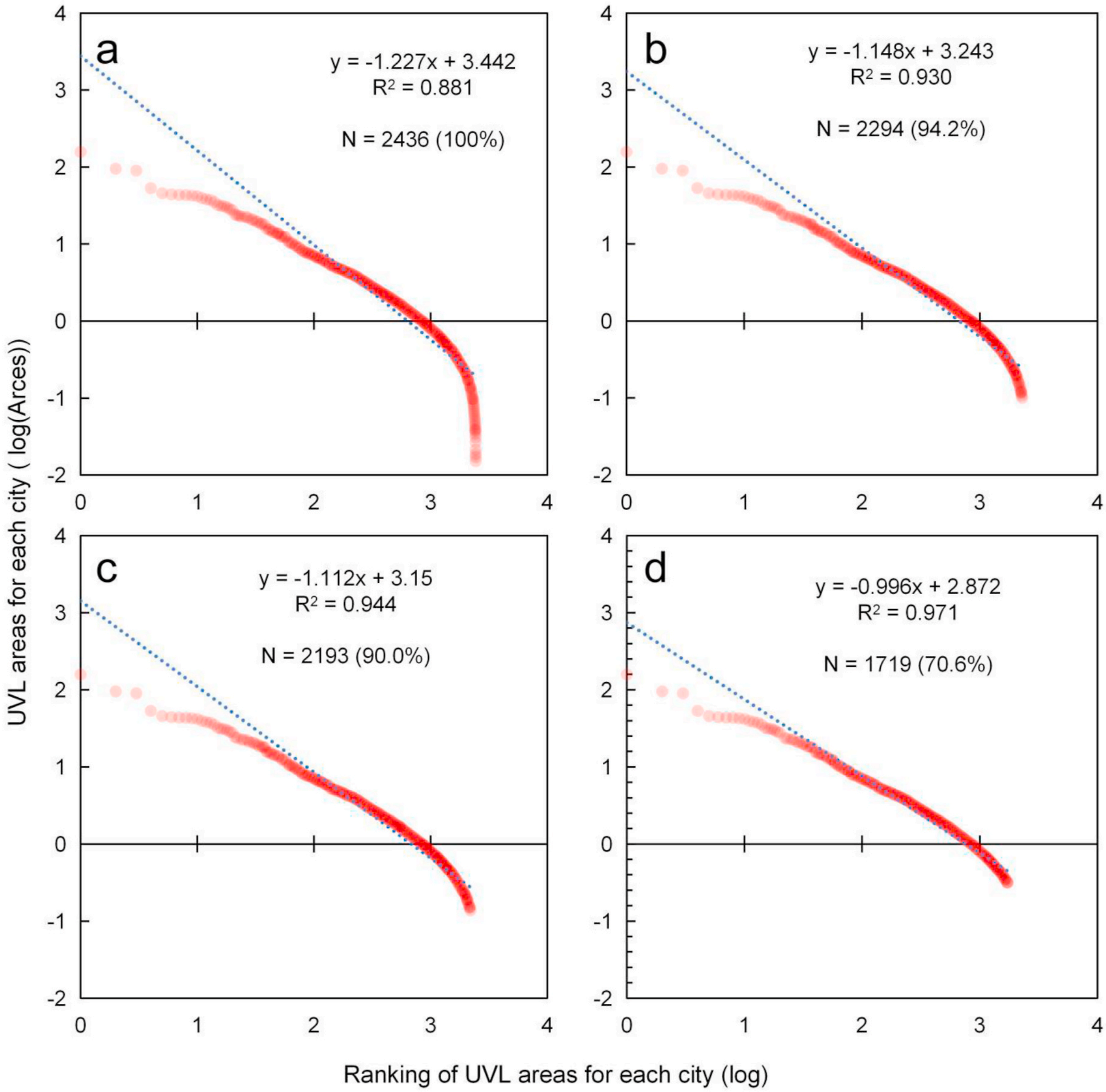

Fig. 5. The total UVL area of NCs and rank-size plots. Figure a: the rank-size plot for all 2436 NCs. Figure b: NCs with UVL more than 0.1 square kilometers (2294 NCs), which is the top 94.2 % NCs. Figure c: the top 90 % of NCs by UVL area (2193 NCs). Figure d: NCs with UVL more than 0.3 square kilometers (1758 NCs), which are the top 70.6 % NCs. Both the X and Y axes are on a log scale. The significance test confirmed a good fit ($p < 0.001$).

Second, the closer to the city's periphery, the larger the average size of the UVL. The average UVL size within the 2015 city boundaries is 2.1 ha, while the average size for the period 2015 to 2023 is 2.6 ha. Furthermore, when analyzing land changes from 1987 to 2017, it is observed that vacant lands located on the city boundaries from 1985 to 1986 had an average size of 1.4 ha. In contrast, those located on the 2016–2017 boundaries had an average size of approximately 2.6 ha.

3.4. *UVL spatial distribution patterns in NC are categorized into five patterns*

Table 1 and Fig. 7 present categorized results and related information about the five city types. Although there is no correlation between the UVL ratio and the NC area, the NC area and the UVL area adhere to a scaling law, with an overall fit of $R^2 = 0.72$ (Fig. 7a). Furthermore, Fig. 7a shows that when cities are categorized into five types, the UVL area and NC area within each type still follow the scaling law, demonstrating the strong robustness of this distribution pattern.

However, notable differences exist among these five types. As illustrated in Fig. 7b, smaller cities (e.g., cities with 5–10 km^2 NC area) are predominantly of the central and peripheral types (orange and light orange type). As the size of the city increases, the prevalence of these two types decreases significantly, while the local type becomes dominant. Based on the classification method (please see 2.5 Classification for all NCs by the distribution of UVL), the central, peripheral, and scatter types indicate that UVL is not clustered within the city. This suggests that in smaller cities, UVL tends to be dispersed in the urban boundary. As cities grow, UVL gradually concentrates and forms distinct clusters. Notably, in cities larger than 300 km^2, nearly all exhibit the local type, where UVL is clustered, leading to increased land-use efficiency.

Fig. 7c illustrates the characteristics of these five types, while Table 1 provides additional information. Local and global types have a higher average UVL area compared with other types for two reasons. For the local type, as mentioned earlier, large cities predominantly follow this type, which means the local type tends to have a higher average NC area and the higher NC area will lead to more UVL due to the scaling law as shown in Fig. 7a. For the global type, where UVL is distributed across most areas of the city, this category is characterized by the highest UVL

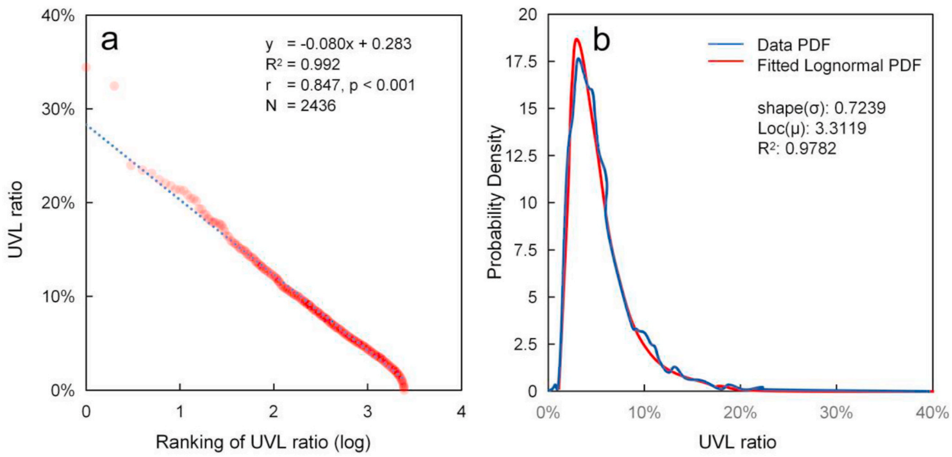

Fig. 6. Results of the distribution of UVL ratio. Figure a is the lognormal distribution of UVL ratio. Figure b is the Probability Density Function (PDF) of UVL Ratio. The X-axis represents the urban vacant land ratio, and the Y-axis represents its probability density. The blue line represents the probability density curve (PDF) of UVL ratio, while the red line depicts the standard lognormal distribution curve. The shape fitting parameter is 0.7138, the scale parameter is −3.2827, and the fitted R2 is 0.9864, indicating an excellent fit. The significance test confirmed a good fit ($p < 0.001$). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Results of UVL distribution patterns.

Type	Global	Local	Central	Peripheral	Scatter	Overall
Zipf's law fitting of NC area (R^2)	0.96	0.95	0.96	0.99	0.97	0.97
Average NC area (km^2)	30.37	71.15	19.95	15.49	27.07	40.28
Average UVL area (km^2)	3.26	3.09	0.61	0.65	0.90	1.92
Average number of UVL	105.17	146.12	40.59	37.24	57.89	89.26
UVL ratio (%)	10.75	4.34	3.04	4.17	3.83	4.78
Total NC area (km^2)	8381.11	60689.85	8297.62	7386.41	6767.61	91522.59
Total UVL area (km^2)	900.69	2635.42	252.31	307.97	259.02	4355.41
Total NC number	276	853	416	477	250	2272

ratio compared to other types. These reasons lead to the higher average UVL area in global and local types.

Furthermore, a comparison between the central and peripheral types, which are predominantly observed in small and medium-sized cities, provides valuable insights. The average number of UVL in the peripheral type is 37.24, which is smaller than that in the central type (Table 1). However, despite having fewer UVL, the peripheral type demonstrates a larger total UVL area compared to the central type. This observation supports the findings in 3.3 UVL near periphery has larger area, indicating that UVL near the periphery tends to be larger than those located closer to the city center.

To evaluate the robustness of the five-type UVL classification, we conducted a sensitivity analysis using an alternative urban boundary dataset, the Global Urban Boundary (GUB) dataset (Li et al., 2020). Specifically, we replaced the NC boundaries with the GUB boundaries and repeated the UVL distribution classification process under the same grid-based and clustering criteria (as described in 2.5 Classification for all NCs by the distribution of UVL).

Following the procedure illustrated in Fig. 3, the classification of global and local types remained unaffected because their identification is based on the degree of clustering and the overall UVL ratio without involving boundary. In contrast, central, peripheral, and scatter types (a total of 1,143 cities) are defined in Step 3, where urban boundaries play a crucial role. Under the GUB boundaries, the number of central type cities increased to 579, peripheral to 381, and scatter to 183. Among these cities, 775 (67.8 %) retained the same classification under both boundary definitions, while 368 cities showed a classification mismatch. Of these mismatches, 48.37 % involved the peripheral type and 38.04 % involved the scatter type. Most of these transitions occurred between peripheral and scatter categories, indicating that their classification is relatively sensitive to the urban boundary definition. In contrast, the central type exhibited the highest stability, suggesting that its identification is less affected by boundary changes.

These findings suggest that, although the overall five-type classification framework is not entirely immune to boundary effects, the robustness is still influenced by urban boundary. Therefore, for applications involving sensitive urban boundary contexts, such as cross-city comparisons or regional policy planning, caution should be exercised when interpreting peripheral and scatter classifications.

3.5. Five UVL spatial distribution patterns across regions in China

This study further analyzes the spatial patterns of the five UVL types and examines their distribution across different regions. The regional analysis is based on the official administrative division of China, as defined by the National Bureau of Statistics. Specifically, the 6 regions include North China, Northeast China, East China, Central-South China, Southwest China, and Northwest China, as detailed in Table 2. The spatial distribution of different UVL types shows spatial heterogeneity, as shown in Fig. 8. Fig. 9 demonstrates the log-log transformed relationship between UVL area and NC area across five UVL types and six regions in China. These five types exhibit distinct spatial heterogeneity. For example, the Local type is primarily concentrated in North China, East China, and Northwest China, while the Peripheral type is mainly distributed in East and Northwest China. A comparison of intra-regional UVL distribution patterns reveals notable differences across regions as shown in Fig. 9. In North China, UVL is primarily composed of Local and Central types. In Northwest and East China, Local and Peripheral types

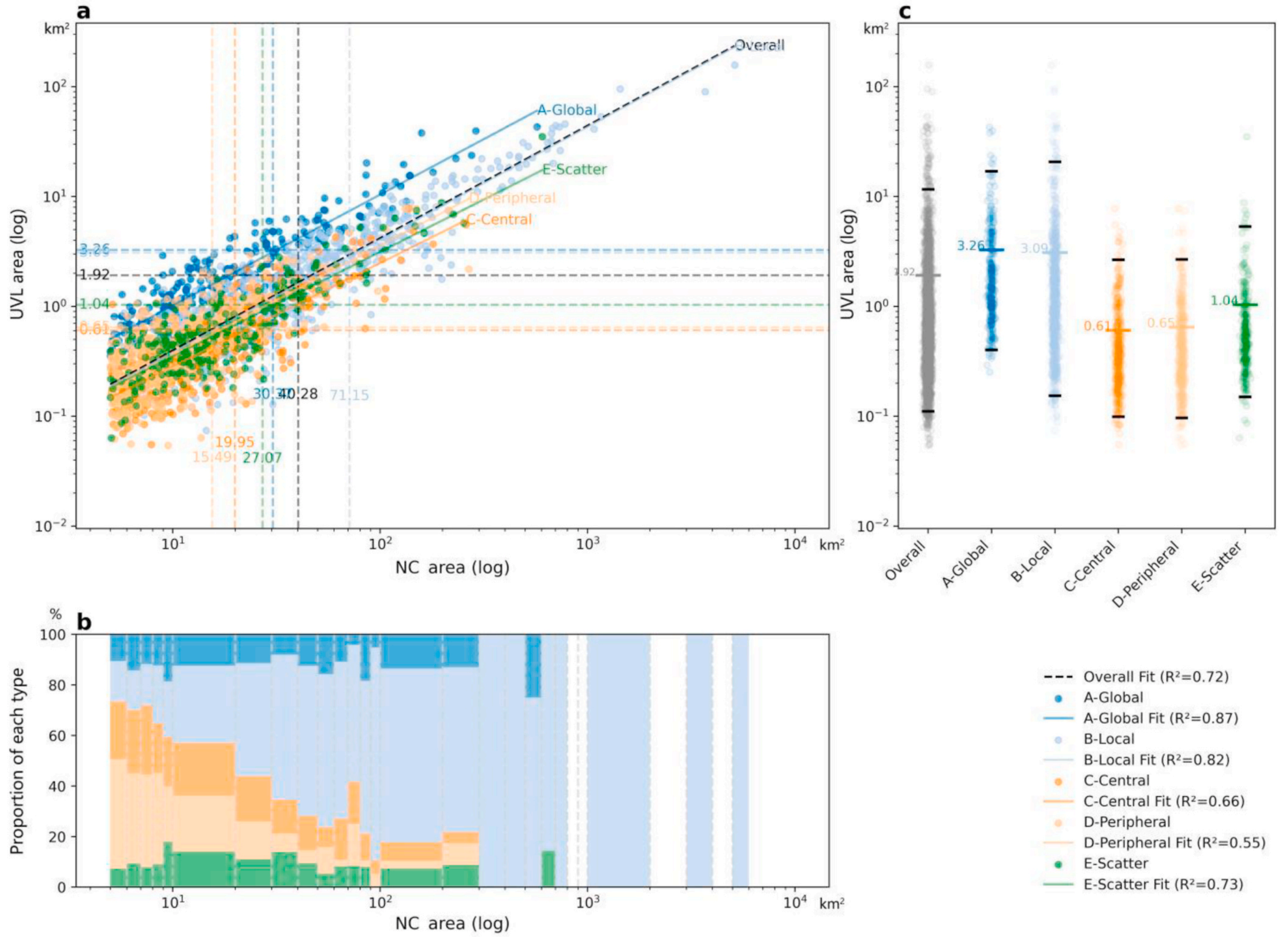

Fig. 7. The distribution results of the five types at the national scale. As shown in the legend, each type has a specific color. In Figure a, each plot represents a city (NC), which has two attributes: the total area (NC area, X-axis) and the UVL area (Y-axis). Figure a shows a scaling relationship between the UVL area and the NC area (both log-transformed) across cities. In Figure b, the X-axis is NC area, the same as in Figure a, while the Y-axis indicates the proportion of each of the five types within certain ranges. Figure b depicts how the proportion of each UVL type varies with NC area size. Medium and large cities are predominantly characterized by local types, while small cities are primarily central and peripheral types. In Figure c, the X-axis shows the five distribution types, and the Y-axis is the same as in Figure a, representing the log UVL area. Figure c illustrates the average and 95 % Confidence Interval of the five types. Cities following global and local types tend to have higher UVL areas, while cities following central and peripheral types exhibit lower UVL areas than the average. The significance tests for all five types and the overall model confirm a good fit ($p < 0.001$). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Commonly used official administrative division in China.

Region	Provinces/Municipalities Included
North China	Beijing, Tianjin, Hebei, Shanxi, Inner Mongolia
Northeast China	Liaoning, Jilin, Heilongjiang
East China	Shanghai, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shandong
Central-South China	Henan, Hubei, Hunan, Guangdong, Guangxi, Hainan
Southwest China	Chongqing, Sichuan, Guizhou, Yunnan, Tibet
Northwest China	Shaanxi, Gansu, Qinghai, Ningxia, Xinjiang

are more prevalent. In Northeast China, the dominant types are Central and Peripheral, indicating that UVL is mainly concentrated within urban cores and along the periphery, likely due to industrial decline and urban shrinkage. In contrast, Central-South China and Southwest China are dominated by the Local type, suggesting a more clustered distribution of urban vacant land.

4. Discussion

This paper presents a method for identifying UVL using deep learning models and provides access to both the models and their training data. The approach was applied to all cities throughout China, generating a national-scale dataset on UVL. This dataset offers a critical resource for researchers aiming to analyze the current conditions, distribution patterns, and characteristics of UVL.

Compared with existing research that relied on manual classification (Castro, 2022; Smith et al., 2017) and city-specific semantic segmentation (Hu & Zhuang, 2024; Mao et al., 2022), this study introduces an efficient segmentation model to enable consistent, high-efficiency UVL identification across 2446 cities. While previous national-scale datasets, such as Tu et al. (2024) focused on shrinking cities using hybrid manual and deep learning workflows, our approach emphasizes full automation with standardized annotation and quality control, improving replicability and scalability for all China's cities. Through the analysis of this dataset, the study enhances the understanding of UVL within urban systems, including the distribution of UVL areas, UVL ratios, and the

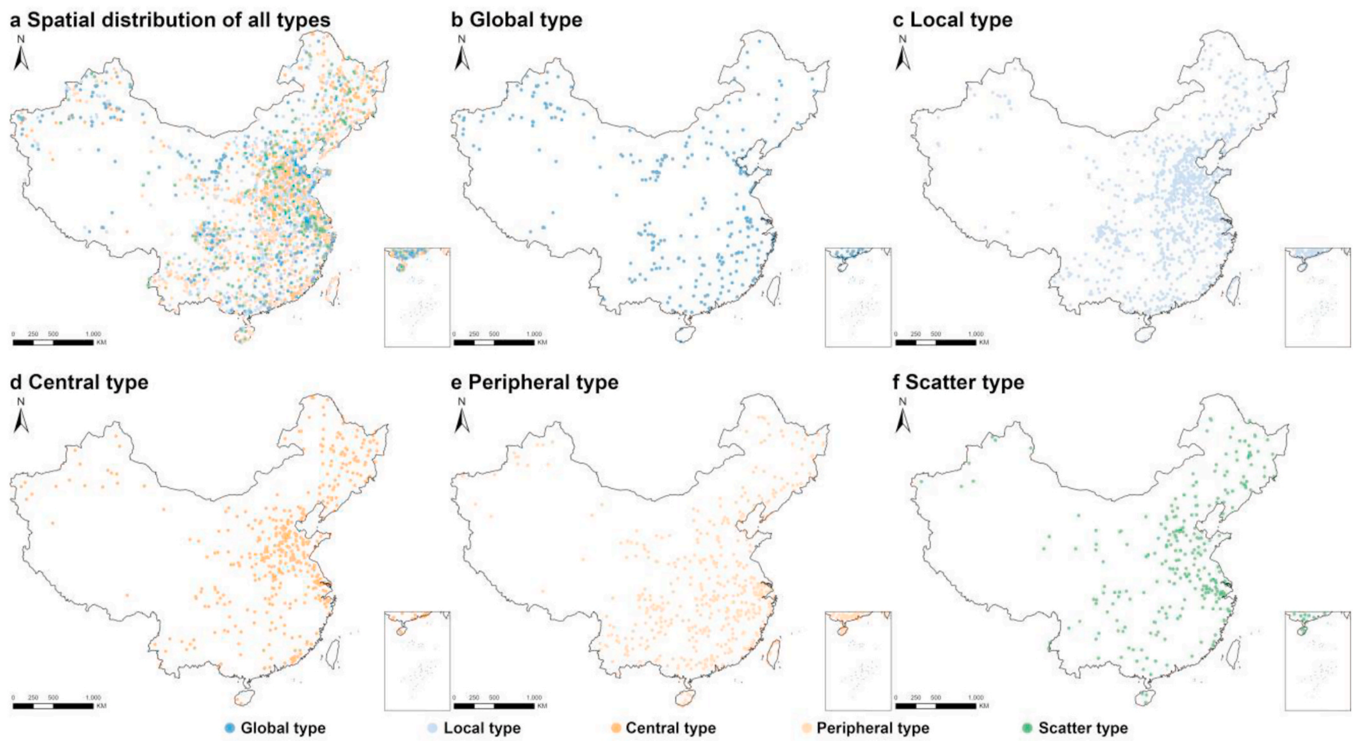

Fig. 8. Spatial distribution results of the five types at the national scale. These figures are based on the standard map provided by the Standard Map Service of the Ministry of Natural Resources of China (Map No. GS(2016)1570), with no modifications to the national boundaries. Figure a shows the composite distribution of all five UVL types, while figure b–f illustrate the spatial distribution of each specific type: (b) global type, (c) local type, (d) central type, (e) peripheral type, and (f) scatter type.

spatial patterns of UVL in each NC. The detailed contributions and discussions are as follows.

4.1. Enhancing UVL identification through unified AI methods

AI models for building segmentation in land-related tasks face substantial challenges, primarily due to complex definitions and ambiguous boundaries (Wang & Li, 2021; Xue et al., 2024). Most prior studies adopt either U-Net or DeepLab variants (Mao et al., 2022) and are limited by their sensitivity to small-scale features or irregular shapes. FastSAM, by contrast, leverages prompt-aware mask decoding and transformer backbones to produce finer boundaries, which are essential for UVL's fragmented spatial morphology. In addition, unlike Xu and Ehlers (2022), who utilized heterogeneous open data with inconsistent semantic meaning, we built a unified, COCO-standard UVL training set with over 5,000 high-resolution samples, ensuring model accuracy, interpretability, and potential reuse in related land classification tasks. Against this backdrop, this research advances solutions to these challenges and offers an open dataset for applying and enhancing the SAM model in land-related studies, thereby creating a valuable resource for cross-disciplinary communities and decision-makers.

To address challenges in boundary delineation, this study developed an urban boundary framework to ensure that the UVL identification dataset has sufficient resolution to support parcel scale analysis. Leveraging this framework, the SAM model was applied to produce UVL with more continuous and relatively accurate boundaries, which are essential in urban land segmentation studies. Small UVL was filtered out to enhance dataset clarity and relevance. The results from this study indicate that the SAM model can effectively generate smoother, more precise boundaries, which represent a meaningful advancement for applying AI model in urban land segmentation research.

Moreover, the transferability of the FastSAM model to other land type identification tasks, such as agricultural land, green spaces, or brownfields, deserves further discussion and investigation. These land types often exhibit complex spatial patterns and fuzzy boundaries, similar to UVL segmentation. Given the model's ability to produce smoother, accurate boundaries in the UVL identification context, it is reasonable to expect that, with proper fine-tuning and domain-specific labeling, FastSAM could be effectively extended to these other domains. Such generalizability would enhance the model's utilization for broader land use analysis and cross-context applications, in support of urban sustainability, agricultural monitoring, and related land policy. The model, training code, labeling dataset, and UVL distribution datasets are available on Figshare, providing valuable resources for advancing future research.

4.2. Distribution characteristics of UVL at city scale

While previous studies contributed valuable insights into UVL in selected cities or regions, they lacked the methodological consistency to support comparative statistical analysis across diverse urban systems (Mao et al., 2022; Tu et al., 2024; Xu & Ehlers, 2022). The present study, by employing a unified data annotation protocol, a single deep learning model (FastSAM), and a standardized definition of Natural Cities, enables a national-scale analysis of UVL distributions with comparable metrics such as Zipf's law fitting, UVL ratio distributions, and typological classification. This level of consistency is crucial for identifying robust statistical regularities across cities with differing sizes, geography, or development stages.

Furthermore, previous work seldom examined whether UVL indicators followed specific statistical laws, nor did they provide empirical thresholds (such as UVL ratio percentiles or confidence intervals) to guide comparative urban analysis or policy prioritization. Our findings bridge this gap and offer foundational benchmarks. We demonstrate that the UVL area follows Zipf's law and exhibits a strong correlation with NC area on the log scale, indicating that larger cities are likely to have more

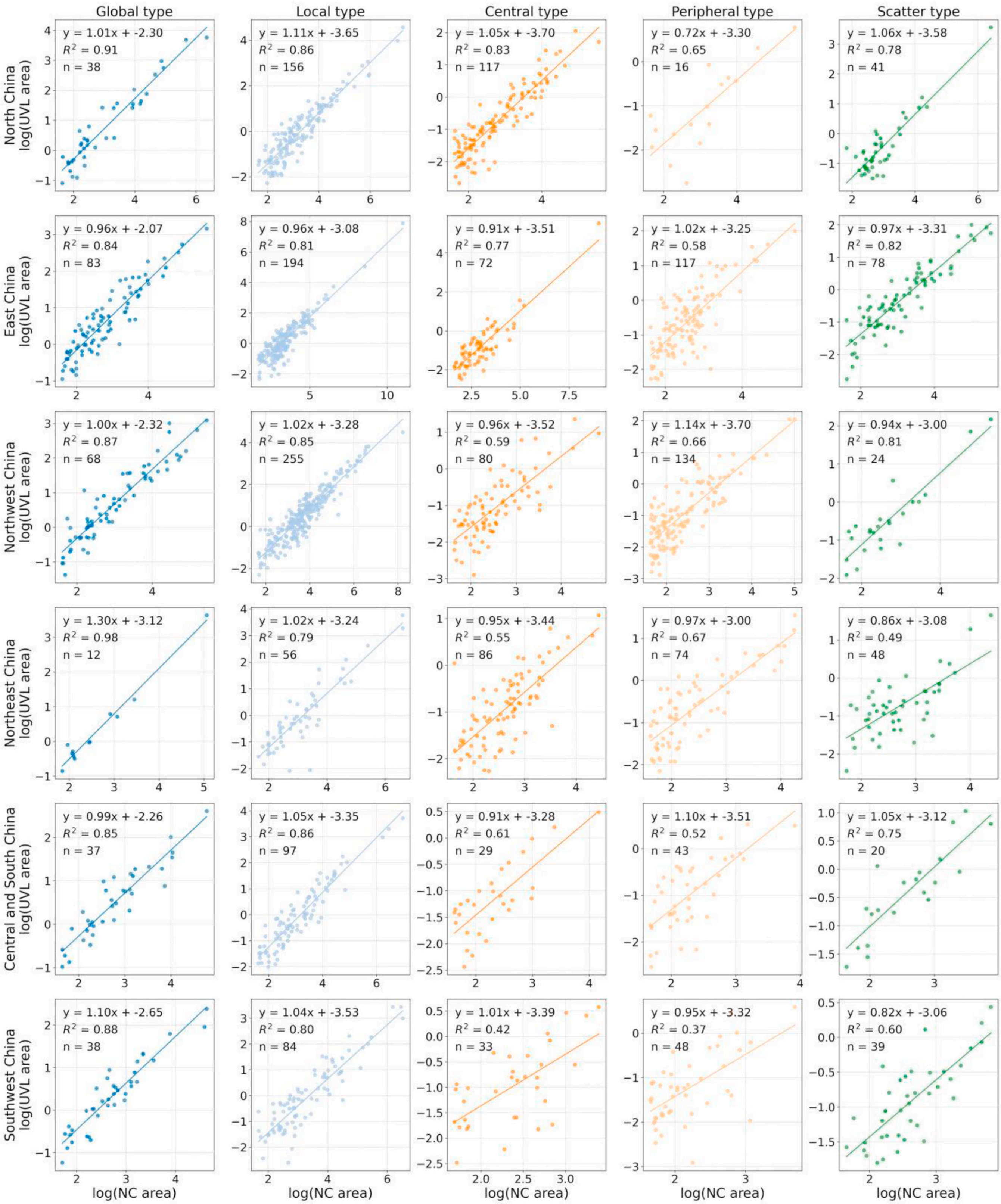

Fig. 9. Scaling law relationship between UVL area and NC area across five types and regions in China. This figure presents scatter plots of log-transformed UVL area versus UVL density across five UVL spatial types (Global, Local, Central, Peripheral, Scatter) and six major regions of China (North China, East China, Northwest China, Northeast China, Central and South China, Southwest China). Each subplot shows the linear regression line (in log-log space), with the regression equation, coefficient of determination (R^2), and sample size (n) annotated.

UVL than smaller ones. This phenomenon is driven by the increase in scale rather than the UVL ratio, as the UVL ratio is independent of NC area.

The UVL ratio exhibits a strong log-normal distribution, characterized by a long tail, indicating that some cities have significantly higher UVL ratios than the average. These cities require adequate attention to address the development challenges associated with their elevated UVL ratios. This distribution also indicates that the median is a more appropriate measure of central tendency for the UVL ratio than the mean.

4.3. Spatial distribution characteristics of UVL for each NC and its policy suggestions

Nearly 50 % of UVL in China are located outside 2015 built-up areas, with their average size increasing as they approach the urban periphery. This pattern is predominantly due to urban expansion, where new areas are developed around the peripheries of existing urban centers, leading to an increase in the size and amount of vacant land.

Based on these categorization results, several conclusions can be drawn. First, UVL in large cities tends to cluster together, forming distinct UVL clusters, whereas UVL in small and medium cities is more dispersed. Consequently, the local type is the predominant pattern in large cities, while the central, peripheral, and scatter types are more commonly found in small and medium cities. Second, the global and local types exhibit larger UVL areas, due to the widespread distribution of UVL. Third, the global and peripheral types require particular attention when addressing UVL-related challenges. For the global type, cities typically have a UVL ratio exceeding 10 %, placing them in the top 10 % according to the log normal distribution of UVL ratios. For the peripheral type, while these cities have lower total UVL areas and UVL ratios, the average size of individual UVL is larger, presenting unique challenges compared to other city types.

To support effective governance, urban strategies should be tailored to the distinct characteristics of each UVL type. In global type cities, where UVL is extensive and spans multiple development zones, long-term strategic planning should prioritize its integration through coordinated land use and multifunctional reuse. Peripheral type UVL, typically large and located at the urban edge, requires improved infrastructure and phased development to enhance accessibility and avoid leapfrog growth. Local type UVL, found within built-up areas, offers strong potential for targeted renewal, including temporary uses and community-based planning. Central type UVL near historical or core areas, though limited in size, holds strategic importance and should be integrated into broader redevelopment programs. Scattered type UVL, due to its fragmentation, demands flexible zoning tools and adaptive reuse strategies to promote gradual reintegration into the urban fabric.

4.4. Comparison with related research

To further validate our identification results, we compared them with previous UVL studies. A representative example is the study by Pagano and Bowman (2000), which collected data on UVL in approximately 70 U.S. cities through manual survey methods, relying primarily on field interviews and questionnaire responses from urban planners and government officials. The dataset included key indicators such as total urban area, total UVL area, and UVL ratio.

As shown in Supplementary Text, our visualization of the results indicates that the distribution of UVL area across U.S. cities follows Zipf's law, with the regression achieving a coefficient of determination of 0.652. When the smallest cities in the sample (the bottom 10 %) were excluded, the model fit improved markedly ($R^2 = 0.844$), further confirming the universality and robustness of this relationship. Their data also revealed that the correlation between urban area and UVL ratio was very weak ($R^2 < 0.03$), an observation that is consistent with the findings of our study.

The average UVL ratio in Chinese cities is 4.78 %, which is significantly lower than the 15.4 % reported for U.S. cities. In China, 66.89 % of NCs have a UVL ratio below 5 %, while 91.84 % fall below 10 %. However, this does not imply that the UVL issue in China is less significant than that in the United States. Compared to the United States, Chinese cities generally exhibit higher population densities and stronger land development controls. For example, urban planning in China tends to enforce compact development strategies. In addition, population projections under the Shared Socioeconomic Pathways indicate that China is undergoing rapid demographic transformation, with substantial internal migration and population aging, which affect urban land use differently than in the United States (Chen et al., 2020; Tatem, 2017; Wang et al., 2022). Moreover, the presence of "ghost cities" and widespread housing vacancies (Jin et al., 2017) poses unique challenges. In particular, China faces substantial difficulties in converting its large stock of vacant housing into UVL due to demolition barriers (Gao & Ryan, 2021). Empirical evidence from localized practices, supported by publicly available government data, further indicates that UVL has become a focal point of attention for both governmental authorities and academic researchers (Li et al., 2018; Zhang et al., 2023).

5. Conclusion

This study presents a robust framework based on a segmentation model for the identification and analysis of UVL at the national scale across China. By leveraging an advanced FastSAM model, we successfully generated a comprehensive UVL dataset covering 2,446 NCs. This openly accessible dataset provides a valuable foundation for urban planning researchers, practitioners, and policymakers.

The results reveal that UVL exhibits distinct spatial and statistical characteristics. Specifically, the total UVL area follows Zipf's law, while the UVL ratio conforms to a log-normal distribution with a long tail, reflecting significant inter-city variation.

This study presents a framework based on a segment model for the identification and analysis of UVL at a national scale across China. By employing advanced deep learning techniques, including the FastSAM model, we successfully generated a UVL dataset that covers 2,446 Natural Cities (NCs), providing a useful and open dataset for urban planning researchers and policymakers.

The results reveal that UVL exhibits unique spatial and statistical properties. The total UVL area follows Zipf's law, while the UVL ratio demonstrates a log-normal distribution with a long tail, indicating substantial variability among cities. Furthermore, cities were classified into five spatial types, global, local, central, peripheral, and scattered, based on the distribution patterns of UVL. Larger cities tend to form clustered UVL zones (local type), whereas smaller cities display more fragmented or peripheral patterns.

This research makes meaningful contributions in three key dimensions. First, from a data openness perspective, this study provides a suite of openly accessible resources, including national-scale UVL maps, deep learning-based segmentation models, training datasets, and manually labeled samples. These resources are shared via Figshare and GitHub, ensuring transparency and reproducibility.

Second, in terms of algorithm universality, the FastSAM-based segmentation approach proves to be highly adaptable to the complex and often ambiguous boundaries. Moreover, this model has the potential to be transferred to other land-related tasks, such as detecting agricultural land, identifying informal settlements, or mapping green infrastructure. This generalizability makes it a versatile tool for both urban researchers and land management professionals in various fields.

Third, regarding policy operability, the study proposes a typology-driven framework and tries to align UVL spatial patterns with differentiated management strategies. By categorizing cities into five UVL types, the framework allows policymakers to move beyond one-size-fits-all approaches and design interventions that are spatially and contextually appropriate. For instance, in global type cities with widespread

UVL, policies should focus on regional-scale coordination, while peripheral type cities may benefit more by flexible and temporary land use policies.

Future work could address several limitations identified in this study. First, future research should focus on improving the accuracy of UVL identification by enhancing model precision and optimizing computational strategies. Due to current GPU constraints, the segmentation process relies on reduced-resolution imagery, which may limit the model's ability to detect fine-grained urban land features. Second, the integration of long-term time-series data on UVL represents a valuable direction for future research, helping to uncover temporal dynamics and developmental trajectories. These data can support a deeper investigation into the mechanisms driving the formation and persistence of UVL. The dataset and identification framework developed in this study provide a solid foundation for pursuing these inquiries.

CRediT authorship contribution statement

Xinyu Wang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation. **Ying Long:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Funding

This research was supported by the by the National Natural Science Foundation of China (No. 52178044, 62394331 and 62394335).

Declaration of competing interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.habitatint.2025.103663.

Data availability

All relevant data, including Python tools, labeled datasets, models, and UVL results, are available for download from Figshare or GitHub.●

UVL identification results: https://figshare.com/s/f57bb8825ab157637d24. Python tools: https://github.com/wangxy20/UVL_identification_FastSAM or https://figshare.com/s/93c66411d8ee337dab15.●

Model for UVL identification: https://figshare.com/s/a204a7aeba8e172e2f4d.●

Label Dataset for training: https://figshare.com/s/3a05a0d41fcd1cbf12b9.

References

Accordino, J., & Johnson, G. T. (2000). Addressing the vacant and abandoned property problem. *Journal of Urban Affairs, 22*(3), 301–315. https://doi.org/10.1111/0735-2166.00058

Agheyisi, J. E. (2025). Temporary use of urban vacant spaces: A pro-poor land use strategy in the global south. *Journal of Urban Affairs, 47*(3), 978–1002. https://doi.org/10.1080/07352166.2023.2213830

Audirac, I. (2009). Urban shrinkage amid fast metropolitan growth: Two faces of contemporary urbanism. Retrieved from https://www.researchgate.net/publication/228815324_Urban_Shrinkage_Amid_Fast_Metropolitan_Growth_Two_Faces_of_Contemporary_Urbanism. (Accessed 3 January 2025).

Bettencourt, L. M. A. (2013). The origins of scaling in cities. *Science, 340*, 1438–1441. https://doi.org/10.1126/science.1235823

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences, 104*, 7301–7306. https://doi.org/10.1073/pnas.0610172104

Blanco, H., Alberti, M., Olshansky, R., Chang, S., Wheeler, S. M., Randolph, J., London, J. B., Hollander, J. B., Pallagst, K. M., Schwarz, T., Popper, F. J., Parnell, S., Pieterse, E., & Watson, V. (2009). Shaken, shrinking, hot, impoverished and informal: Emerging research agendas in planning. *Progress in Planning, 72*, 195–250. https://doi.org/10.1016/j.progress.2009.09.001

Bowman, R. J., & Kim, B. (2016). A current inventory of vacant urban land in America. *Journal of Urban Design, 21*(3), 302–319. https://doi.org/10.1080/13574809.2016.1167589

Branas, C. C., South, E., Kondo, M. C., Hohl, B. C., Bourgois, P., Wiebe, D. J., & MacDonald, J. M. (2018). Citywide cluster randomized trial to restore blighted vacant land and its effects on violence, crime, and fear. *Proceedings of the National Academy of Sciences, 115*(12), 2946–2951. https://doi.org/10.1073/pnas.1718503115

Cao, N., Wen, Y., Wang, J., Gao, X., & Sun, P. (2024). Identifying shrinking cities from a physical city perspective and influencing factors: A case study of the Chengdu-Chongqing economic circle in China. *Applied Geography, 168*, Article 103314. https://doi.org/10.1016/j.apgeog.2024.103314

Castro, A. (2022). *Vacant lots in Chicago: A typology of vacant Lot uses and their local context*. Northeastern Illinois University. Retrieved from https://proxy.library.upenn.edu/login?url=https://www.proquest.com/dissertations-theses/vacant-lots-chicago-typology-lot-uses-their-local/docview/2668173199/se-2?accountid=14707. (Accessed 3 July 2025).

Chen, Y., Li, X., Huang, K., Luo, M., & Gao, M. (2020). High-resolution gridded population projections for China under the shared socioeconomic pathways. *Earth's Future, 8*, Article e2020EF001491. https://doi.org/10.1029/2020ef001491

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 801–818).

Cheng, X., Yang, J., Li, B., Zhao, B., Pan, D., Shen, Z., … Liu, M. (2025). Enhancing urban land utilization through segformer: A vacant land analysis in Chengdu. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18*, 6070–6085. https://doi.org/10.1109/JSTARS.2025.3538920

Cheng, X., Yang, J., Li, B., Zhao, B., Pan, D., Shen, Z., Zhu, Q., & Liu, M. (2025). Enhancing urban land utilization through segformer: A vacant land analysis in chengdu. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18*, 6070–6085. https://doi.org/10.1109/JSTARS.2025.3538920

Deng, C., & Ma, J. (2015). Viewing urban decay from the sky: A multi-scale analysis of residential vacancy in a shrinking U.S. city. *Landscape and Urban Planning, 141*, 88–99. https://doi.org/10.1016/j.landurbplan.2015.05.002

Dennison, C. J., Pearson, A. L., Hanson, J. O., Brown, C. D., & Buxton, R. T. (2025). The role of vacant lots in promoting avian species diversity and occupancy in a post-industrial city. *Landscape and Urban Planning, 265*, Article 105518. https://doi.org/10.1016/j.landurbplan.2025.105518

Ehrhardt, D., Behnisch, M., Michaeli, M., & Jehling, M. (2025). Understanding incremental densification–determinants of residential infill on vacant lots. *Landscape and Urban Planning, 260*, Article 105375. https://doi.org/10.1016/j.landurbplan.2025.105375

Eraydin, A., & Özatağan, G. (2021). Pathways to a resilient future: A review of policy agendas and governance practices in shrinking cities. *Cities, 115*, Article 103226. https://doi.org/10.1016/j.cities.2021.103226

Gao, S., & Ryan, B. D. (2021). Implementation challenges of state-led redevelopment in shrinking cities: Case study of shantytown redevelopment in Yichun, Northeast China. *Journal of Urban Planning and Development, 147*, Article 05021001. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000661

Garvin, E., Branas, C., Keddem, S., Sellman, J., & Cannuscio, C. (2013). More than just an eyesore: Local insights and solutions on vacant land and urban health. *Journal of Urban Health, 90*(3), 412–426. https://doi.org/10.1007/s11524-012-9782-7

Glaeser, E. L., & Gyourko, J. (2005). Urban decline and durable housing. *Journal of Political Economy, 113*(2), 345–375. https://doi.org/10.1086/427465

Gong, P., Li, X., Wang, J., Bai, Y., Chen, B., Hu, T., Liu, X., Xu, B., Yang, J., & Zhang, W. (2020). Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. *Remote Sensing of Environment, 236*, Article 111510. https://doi.org/10.1016/j.rse.2019.111510

He, S. Y., Lee, J., Zhou, T., & Wu, D. (2017). Shrinking cities and resource-based economy: The economic restructuring in China's mining cities. *Cities, 60*, 75–83. https://doi.org/10.1016/j.cities.2016.07.009

Hu, X., & Zhuang, S. (2024). Large-scale spatial–temporal identification of urban vacant land and informal green spaces using semantic segmentation. *Remote Sensing, 16*(2), 216. https://doi.org/10.3390/rs16020216

Jiang, B., & Miao, Y. (2015). The evolution of natural cities from the perspective of location-based social media. *The Professional Geographer, 67*, 295–306. https://doi.org/10.1080/00330124.2014.968886

Jiang, B., Yin, J., & Liu, Q. (2015). Zipf's law for all the natural cities around the world. *International Journal of Geographical Information Science, 29*, 498–522. https://doi.org/10.1080/13658816.2014.988715

Jiang, Z., Zhai, W., Meng, X., & Long, Y. (2020). Identifying shrinking cities with NPP-VIIRS nightlight data in China. *Journal of Urban Planning and Development, 146*, Article 04020034. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000598

Jin, X., Long, Y., Sun, W., Lu, Y., Yang, X., & Tang, J. (2017). Evaluating cities' vitality and identifying ghost cities in China with emerging geographical data. *Cities, 63*, 98–109. https://doi.org/10.1016/j.cities.2017.01.002

Kelleher, C., Golden, H. E., Burkholder, S., & Shuster, W. (2020). Urban vacant lands impart hydrological benefits across city landscapes. *Nature Communications, 11*, 1–11. https://doi.org/10.1038/s41467-020-15338-1

Kim, G., Miller, P. A., & Nowak, D. J. (2015). Assessing urban vacant land ecosystem services: Urban vacant land as green infrastructure in the City of Roanoke, Virginia. *Urban Forestry and Urban Greening, 14*(3), 519–526. https://doi.org/10.1016/j.ufug.2015.05.003

Kim, G., Miller, P. A., & Nowak, D. J. (2018). Urban vacant land typology: A tool for managing urban vacant land. *Sustainable Cities and Society, 36*, 144–156. https://doi.org/10.1016/j.scs.2017.09.014

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W. Y., Dollár, P., & Girshick, R. (2023). Segment anything. *Proceedings of the IEEE International Conference on Computer Vision, 3992*, 3992–4003. https://doi.org/10.1109/ICCV51070.2023.00371

Kiviaho, A., & Toivonen, S. (2023). Forces impacting the real estate market environment in shrinking cities: Possible drivers of future development. *European Planning Studies, 31*(2), 189–211. https://doi.org/10.1080/09654313.2022.2121604

Kremer, P., Hamstead, Z. A., & McPhearson, T. (2013). A social–ecological assessment of vacant lots in New York City. *Landscape and Urban Planning, 120*, 218–233. https://doi.org/10.1016/j.landurbplan.2013.05.003

Li, X., Gong, P., Zhou, Y., Wang, J., Bai, Y., Chen, B., … Zhu, Z. (2020). Mapping global urban boundaries from the global artificial impervious area (GAIA) data. *Environmental Research Letters, 15*(9), Article 094044. https://doi.org/10.1088/1748-9326/ab9be3

Li, W., Wang, H., Wang, X., & Tu, T. (2025). Automatic identification of potential renewal areas in urban residential districts using remote sensing data and GeoAI. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18*, 8523–8535. https://doi.org/10.1109/JSTARS.2025.3549618

Li, W., Zhou, W., Bai, Y., Pickett, S. T. A., & Han, L. (2018). The smart growth of Chinese cities: Opportunities offered by vacant land. *Land Degradation & Development, 29*, 3512–3520. https://doi.org/10.1002/ldr.3125

Lin, C., Huang, Y., Liu, Y., Li, G., Zhou, Z., Zhong, Y., Wang, H., & Li, J. (2025). Identifying underutilized land by eXplainable artificial intelligence and geographic similarity ensemble model with limited samples. *Habitat International, 163*, Article 103503. https://doi.org/10.1016/j.habitatint.2025.103503

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common objects in context. *Lecture Notes in Computer Science, 8693*, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

Liu, Y., Fang, F., & Li, Y. (2014). Key issues of land use in China and implications for policy making. *Land Use Policy, 40*, 6–12. https://doi.org/10.1016/j.landusepol.2013.03.013

Liu, C., Hu, S., Wu, S., Song, J., & Li, H. (2024). County-level land use carbon emissions in China: Spatiotemporal patterns and impact factors. *Sustainable Cities and Society, 105*, Article 105304. https://doi.org/10.1016/j.scs.2024.105304

Liu, X., Huang, Y., Xu, X., Li, X., Li, X., Ciais, P., Lin, P., Gong, K., Ziegler, A. D., Chen, A., Gong, P., Chen, J., Hu, G., Chen, Y., Wang, S., Wu, Q., Huang, K., Estes, L., & Zeng, Z. (2020). High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. *Nature Sustainability, 3*(6), 564–570. https://doi.org/10.1038/s41893-020-0521-x

Long, Y. (2016). Redefining Chinese city system with emerging new data. *Applied Geography, 75*, 36–48. https://doi.org/10.1016/j.apgeog.2016.08.002

Long, Y., Zhai, W., Shen, Y., & Ye, X. (2018). Understanding uneven urban expansion with natural cities using open data. *Landscape and Urban Planning, 177*, 281–293. https://doi.org/10.1016/j.landurbplan.2017.05.008

López, J. I. M., Kim, G., Lei, Y., Newman, G., & Suppakittpaisarn, P. (2021). An assessment method and typology for the regeneration of vacant land in Quito, Ecuador. *Urban Forestry and Urban Greening, 62*, Article 127130. https://doi.org/10.1016/j.ufug.2021.127130

Ma, S., & Long, Y. (2019). Identifying spatial cities in China at the community scale. *Journal of Urban and Regional Planning, 11*, 37–50.

Ma, S., & Long, Y. (2020). Functional urban area delineations of cities on the Chinese mainland using massive Didi ride-hailing records. *Cities, 97*, Article 102532. https://doi.org/10.1016/j.cities.2019.102532

Ma, X., Wu, Q., Zhao, X., Zhang, X., Pun, M. O., & Huang, B. (2024). Sam-assisted remote sensing imagery semantic segmentation with object and boundary constraints. *IEEE Transactions on Geoscience and Remote Sensing, 62*, 1–18. https://doi.org/10.1109/TGRS.2024.3411398

Mao, L., Zheng, Z., Meng, X., Zhou, Y., Zhao, P., Yang, Z., & Long, Y. (2022). Large-scale automatic identification of urban vacant land using semantic segmentation of high-resolution remote sensing images. *Landscape and Urban Planning, 222*, Article 104384. https://doi.org/10.1016/j.landurbplan.2022.104384

Martinez-Fernandez, C., & Cristina, T. W. (2007). Shrinking cities: Urban management and environmental legacy. *Landschaftsentwicklung und Umweltforschung*.

Meng, X., Ma, S., Xiang, W., Kan, C., Wu, K., & Long, Y. (2021). Classification of shrinking cities in China using Baidu big data. *Acta Geographica Sinica, 76*, 2477–2488. https://doi.org/10.11821/DLXB202110010

Németh, J., & Langhorst, J. (2014). Rethinking urban transformation: Temporary uses for vacant land. *Cities, 40*, 143–150. https://doi.org/10.1016/j.cities.2013.04.007

Newman, G., & Kim, B. (2017). Urban shrapnel: Spatial distribution of non-productive space. *Landscape Research, 42*(6), 699–715. https://doi.org/10.1080/01426397.2017.1363877

Osco, L. P., Wu, Q., De Lemos, E. L., Gonçalves, W. N., Ramos, A. P. M., Li, J., & Junior, J. M. (2023). The segment anything model (sam) for remote sensing applications: From zero to one shot. *International Journal of Applied Earth Observation and Geoinformation, 124*, Article 103540. https://doi.org/10.1016/j.jag.2023.103540

Pagano, M. A., & Bowman, A. O. M. (2000). *Vacant land in cities: An urban resource*.

Park, Y., Newman, G. D., Lee, J. E., & Lee, S. (2021). Identifying and comparing vacant housing determinants across South Korean cities. *Applied Geography, 136*, Article 102566. https://doi.org/10.1016/j.apgeog.2021.102566

Qu, Y., Zhao, W., Zhan, L., Cui, Y., & Li, Y. (2025). Unlocking rural resources for revitalization: Examining the process and impact of compensated use of rural residential land through a social-ecological system framework. *Habitat International, 166*, Article 103581. https://doi.org/10.1016/j.habitatint.2025.103581

Rieniets, T. (2009). Shrinking cities: Causes and effects of urban population losses in the twentieth century. *Nature and Culture, 4*(3), 231–254. https://doi.org/10.3167/nc.2009.040302

Ryan, S. P. (2012). The costs of environmental regulation in a concentrated industry. *Econometrica, 80*(3), 1019–1061. https://doi.org/10.3982/ecta6750

Schetke, S., & Haase, D. (2008). Multi-criteria assessment of socio-environmental aspects in shrinking cities: Experiences from eastern Germany. *Environmental Impact Assessment Review, 28*, 483–503. https://doi.org/10.1016/j.eiar.2007.09.004

Smith, J. P., Li, X., & Turner, B. L. (2017). Lots for greening: Identification of metropolitan vacant land and its potential use for cooling and agriculture in Phoenix, AZ, USA. *Applied Geography, 85*, 139–151. https://doi.org/10.1016/j.apgeog.2017.06.005

Snow, J., Mullholland, T., Johnson, D., Ford, G., & DelCotto, L. (2017). Vacant land assessment report: Investigation and strategies to reduce blight in Clairton. Retrieved from https://groundedpgh.org/wp-content/uploads/2017/05/ClairtonVLAR-FINAL.pdf. (Accessed 3 July 2025).

Song, X., Wen, M., Shen, Y., Feng, Q., Xiang, J., Zhang, W., Zhao, G., & Wu, Z. (2020). Urban vacant land in growing urbanization: An international review. *Journal of Geographical Sciences, 30*, 669–687. https://doi.org/10.1007/s11442-020-1749-0

Steinführer, S. A., & Kabisch, A. M. (2008). *Demographic change and its impact on housing. Final report for the EUROCITIES network, Brussels and Leipzig*.

Stern, M., & Lester, T. W. (2021). Does local ownership of vacant land reduce crime? *Journal of the American Planning Association, 87*(1), 73–84. https://doi.org/10.1080/01944363.2020.1792334

Tatem, A. J. (2017). WorldPop, open data for spatial demography. *Scientific Data, 4*. https://doi.org/10.1038/sdata.2017.4

Tavares, A. O., Pato, R. L., & Magalhães, M. C. (2012). Spatial and temporal land use change and occupation over the last half century in a peri-urban area. *Applied Geography, 34*, 432–444. https://doi.org/10.1016/j.apgeog.2012.01.009

Tu, T., Wang, X., & Long, Y. (2024). Spatiotemporal changes of urban vacant land and its distribution patterns in shrinking cities on the globe. *Science of The Total Environment, 947*, Article 174424. https://doi.org/10.1016/j.scitotenv.2024.174424

Wang, W. (2023). *Advanced auto labeling solution with added features*. GitHub Repository. https://github.com/CVHub520/X-AnyLabeling.

Wang, S., & Li, W. (2021). GeoAI in terrain analysis: Enabling multi-source deep learning and data fusion for natural feature detection. *Computers, Environment and Urban Systems, 90*, Article 101715. https://doi.org/10.1016/j.compenvurbsys.2021.101715

Wang, X., & Long, Y. (2023). Future shrinking cities on the globe: A projection map for 2020–2100 based on global gridded population dataset. *Environment and Planning B: Urban Analytics and City Science, 50*, 1994–1997. https://doi.org/10.1177/23998083231186153

Wang, X., & Long, Y. (2024). Characteristics of "resident population in physical urban area" in identifying shrinking cities: An examination based on literature review and multi-source data. *Geographical Research, 43*(7), 1929–1939. https://doi.org/10.11821/dlyj020240054, 2024.

Wang, X., Meng, X., & Long, Y. (2022). Projecting 1 km-grid population distributions from 2020 to 2100 globally under shared socioeconomic pathways. *Scientific Data, 9*, 1–13. https://doi.org/10.1038/s41597-022-01675-x

Wang, X., Meng, X., Wang, C., Yang, L., Zhang, Y., & Long, Y. (2024). Identification of inefficient spaces in resource-depleted cities: A case study of Hegang City. *Resources Science, 46*, 1119–1130. https://doi.org/10.18402/resci.2024.06.06

Wang, Z., Wang, C., Dou, H., Cheng, G., Zhang, J., Lei, X., & Huang, X. (2025). A strategy of building a beautiful and harmonious countryside: Reuse of idle rural residential land based on symbiosis theory. *Habitat International, 155*, Article 103238. https://doi.org/10.1016/j.habitatint.2024.103238

Wang, J., Wang, J., & Zhang, J. (2023). Spatial distribution characteristics of natural ecological resilience in China. *Journal of Environmental Management, 342*, Article 118133. https://doi.org/10.1016/j.jenvman.2023.118133

Williams, S., Xu, W., Tan, S. B., & Foster, M. J. (2019). Ghost cities of China: Identifying urban vacancy through social media data. *Cities, 94*, 275–285. https://doi.org/10.1016/j.cities.2019.05.006

Xie, Y., Gong, H., Lan, H., & Zeng, S. (2018). Examining shrinking city of Detroit in the context of socio-spatial inequalities. *Landscape and Urban Planning, 177*, 350–361. https://doi.org/10.1016/j.landurbplan.2018.03.002

Xu, S., & Ehlers, M. (2022). Automatic detection of urban vacant land: An open-source approach for sustainable cities. *Computers, Environment and Urban Systems, 91*, Article 101729. https://doi.org/10.1016/j.compenvurbsys.2021.101729

Xue, B., Cheng, H., Yang, Q., Wang, Y., & He, X. (2024). Adapting segment anything model to aerial land cover classification with low-rank adaptation. *IEEE Geoscience and Remote Sensing Letters, 21*, 1–5. https://doi.org/10.1109/LGRS.2024.3357777

Yin, H., Prishchepov, A. V., Kuemmerle, T., Bleyhl, B., Buchner, J., & Radeloff, V. C. (2018). Mapping agricultural land abandonment from spatial and temporal segmentation of Landsat time series. *Remote Sensing of Environment, 210*, 12–24. https://doi.org/10.1016/j.rse.2018.02.050

Zhang, W., Li, Y., & Zheng, C. (2023). The distribution characteristics and driving mechanism of vacant land in Chengdu, China: A perspective of urban shrinkage and expansion. *Land Use Policy, 132*, Article 106812. https://doi.org/10.1016/j.landusepol.2023.106812

Zhao, X., Ding, W., An, Y., Du, Y., Yu, T., Li, M., Tang, M., & Wang, J. (2023). *Fast segment anything*.

Zhu, R., Newman, G., Li, D., Song, Y., & Zou, L. (2023). Associations between vacant urban lands and public health outcomes in growing and shrinking cities. *Urban Forestry and Urban Greening, 89*, Article 128127. https://doi.org/10.1016/j.ufug.2023.128127