Large-scale automatic identification of urban vacant land using semantic segmentation of high-resolution remote sensing images

Lingdong Mao [a], Zhe Zheng [a], Xiangfeng Meng [b], Yucheng Zhou [a], Pengju Zhao [a], Zhihan Yang [a], Ying Long [b,*]

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

HIGHLIGHTS

- Accuracy of the proposed framework is over 90 percent of professional auditors.
- Efficiency of the automatic framework is 15 times higher than the manual method.
- City stratification greatly improves robustness of large-scale identification.
- Urban vacant land identification results in 36 major Chinese cities are obtained.

ARTICLE INFO

ABSTRACT

Urban vacant land is a growing issue worldwide. However, most of the existing research on urban vacant land has focused on small-scale city areas, while few studies have focused on large-scale national areas. Large-scale identification of urban vacant land is hindered by the disadvantage of high cost and high variability when using the conventional manual identification method. Criteria inconsistency in cross-domain identification is also a major challenge. To address these problems, we propose a large-scale automatic identification framework of urban vacant land based on semantic segmentation of high-resolution remote sensing images and select 36 major cities in China as study areas. The framework utilizes deep learning techniques to realize automatic identification and introduces the city stratification method to address the challenge of identification criteria inconsistency. The results of the case study on 36 major Chinese cities indicate two major conclusions. First, the proposed framework of vacant land identification can achieve over 90 percent accuracy of the level of professional auditors with much higher result stability and approximately 15 times higher efficiency compared to the manual identification method. Second, the framework has strong robustness and can maintain high performance in various cities. With the above advantages, the proposed framework provides a practical approach to large-scale vacant land identification in various countries and regions worldwide, which is of great significance for the academic development of urban vacant land and future urban development.

1. Introduction

Urban vacant land (UVL) is a growing issue worldwide (Gobster, Hadavi, Rigolon, & Stewart, 2020; Li, Zhou, Bai, Pickett, & Han, 2018; Martinez-Fernandez, Audirac, Fol, & Cunningham-Sabot, 2012; Newman, Bowman, Jung Lee, & Kim, 2016). Both disordered urban expansion and urban shrinkage have led to the emergence of UVL (Bowman, 2004; Haase, Haase, Kabisch, Kabisch, & Rink, 2012; Kelleher, Golden, Burkholder, & Shuster, 2020). UVL is a waste of land resources, but it is a potential opportunity for urban development, as it is of great significance for improving the urban spatial structure and formulating urban development strategies (Branas et al., 2018; Heckert,

---

2015; Németh & Langhorst, 2014; Song et al., 2020).

Conventional methods to obtain UVL data include two types. The first approach is to send a questionnaire survey to the city planning department (Bowman & Pagano, 2000; Newman et al., 2016). Through this method, one can acquire authoritative statistical UVL data at the city scale. To investigate finer-scale spatial and temporal properties of UVL, many studies have adopted the second method, i.e., manual UVL identification based on visual interpretation of high-resolution remote sensing (HRRS) images. Li et al. (2018) manually delineated vacant land in 0.3 m resolution aerial photos of Shanghai city in 2000, 2005, and 2010 to systematically investigate the spatial–temporal distribution of vacant land at a fine scale. Li et al. (2019) conducted a case study of underutilized land in the rust belt city Changchun through visual interpretation of 1.0 m resolution satellite images in combination with field surveys. Similar UVL identification methods were adopted in case studies of Guangzhou city (Song et al., 2020) and Atibaia city (Sperandelli, Dupas, & Dias Pons, 2013). Despite the significance of these works, most of the existing studies based on the conventional manual identification method are focused on small-scale city areas. If applied to large-scale areas including many cities, the manual method has following two limitations: (1) the labour and time costs are too high, and (2) the identification results are usually highly variable due to the individual subjective differences in the identification criteria between different auditors.

Automatic UVL identification methods, which can reduce labour and time costs and control result variability, have been attracting attention recently. Shukla and Jain (2020) used an object-based image analysis method to automatically extract urban vacant parcels from unmanned aerial vehicle data and achieved overall high accuracy. However, the study focused on a small area and lacked the variety of UVL. Xu and Ehlers (2022) defined four typologies of vacant land and applied a rule-based data fusion framework integrating remote sensing images, geographical information system (GIS) layers, and citizen science data. While a large number of vacant sites were identified in 63 urban and rural districts in Germany, they found it difficult to systematically detect brownfield land.

With the rapid development of deep learning techniques in computer vision, basic remote sensing image segmentation tasks have been successfully addressed by representative semantic segmentation models, including FCN (Long, Shelhamer, & Darrell, 2015), U-Net (Ronneberger, Fischer, & Brox, 2015), SegNet (Badrinarayanan, Kendall, & Cipolla, 2017), and DeepLab (Chen, Papandreou, Kokkinos, Murphy, & Yuille, 2017; Chen, Papandreou, Schroff, & Adam, 2017; Chen, Zhu, Papandreou, Schroff, & Adam, 2018), which are potential tools to realize large-scale automatic UVL identification. To put it into practice, there are two main challenges: (1) developing the technical framework for accurate and efficient automatic UVL identification and (2) solving the problem of criteria inconsistency in large-scale cross-domain UVL identification.

To address these challenges, this research proposes a large-scale automatic identification framework of UVL based on semantic segmentation of HRRS images and city stratification. The main contributions of this paper are (1) realizing automatic UVL identification based on semantic segmentation of HRRS images, which can efficiently obtain accurate results, and (2) introducing city stratification for large-scale cross-domain UVL identification, which can effectively deal with the problem of identification criteria inconsistency. Results of a case study on 36 major Chinese cities prove that the proposed framework has good accuracy, high stability, high efficiency and strong robustness for large-scale UVL identification. The framework provides a practical approach to large-scale UVL identification in various countries and regions, which is of great significance for the advancement of UVL studies worldwide.

The remainder of this paper is structured as follows. In Section 2, the study area and data are introduced. Section 3 presents the methods of large-scale automatic identification of UVL, including city stratification, data labelling, model training, and prediction and postprocessing.

Section 4 shows the results, including framework performance evaluation, UVL identification results of 36 cities, and framework robustness tests and ablation study. The discussion and conclusions are shown in Section 5.

2. Study area and data

In this paper, 36 municipalities, provincial capitals, and sub-provincial cities in China (Fig. 1) were chosen to study the nationally large-scale automatic identification framework of UVL. These large cities are distributed in various provinces across the country, representing the urban development of their respective provinces and exhibiting diverse geographical features and urban forms.

The Google Earth HRRS images of the 36 cities were downloaded from BIGEMAP (www.bigemap.com), with three red–greenblue bands and a spatial resolution of approximately 0.3 m. We chose the images taken in the summer or autumn of 2019, which met the following requirements to maintain good data quality: (1) the surface vegetation of all the cities was in a growing state, and (2) the cloud interference was controlled at a very low level. For the urban boundary data, because the administrative divisions of Chinese cities are much larger than their actual urban development scope, we used 36 cities' central urbanized areas instead of the traditional administrative areas (generally the largest urban patch among all patches in a city). The adopted urbanized areas of Chinese cities were proposed by Ma and Long (2019), which are contiguous urban built-up areas at the community scale extracted from the superposition analysis of the community areas and urban construction lands.

3. Methods

To achieve accurate and efficient UVL identification for all 36 large cities, we proposed a large-scale automatic identification framework, which consisted of four steps (Fig. 2): (1) city stratification, (2) data labelling, (3) model training, and (4) prediction and postprocessing. The goal and the contents of each step are introduced below.

3.1. City stratification

In large-scale areas including various cities, the characteristics of the vacant lands and non-vacant lands are usually divergent among different city categories, which leads to inconsistency in UVL identification criteria. To address the problem, we proposed a city stratification scheme, including (1) stratifying the cities into different categories and (2) selecting a representative city in each category to establish a model for UVL identification in the rest of the cities.

The principle of city stratification was to reach a balance between performance and cost. First and foremost, the features of the vacant lands and non-vacant lands in different categories should be stratified clearly. Second, the total cost of model establishment in all categories should be as low as possible, which requires that the number of categories should not be too large.

To apply city stratification, we investigated the features of vacant lands and non-vacant lands in the 36 cities. In this paper, urban vacant land refers to 'the lands completely vacant without any developed use, or the lands that are demolished, derelict, or under early-stage construction within the urban built-up area' (Li et al., 2018). To further specify and standardize the UVL definition, five professional auditors independently identified the UVL in the same test areas in ArcGIS10.6, cross-validated the results and discussed the differences. After three rounds of cross-validation tests (Appendix A), the auditors finally reached a consensus on the specific identification criteria of various types of vacant land and non-vacant land with similar features (see Fig. 3 for examples). According to the identification criteria, we made careful observation of the HRRS images of all 36 cities and found that major types of vacant land and non-vacant land in a city were related to
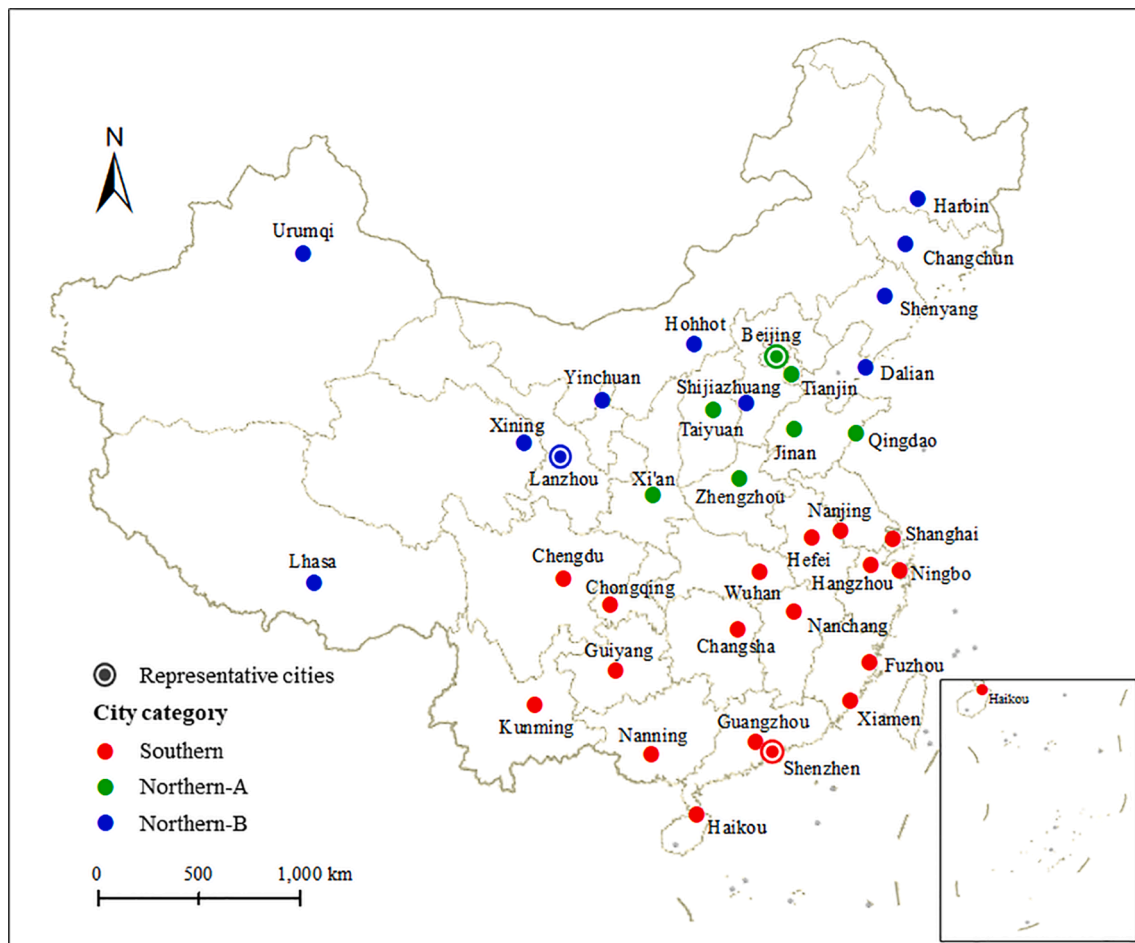
Fig. 1. Locations, categories and representative cities of 36 major cities in China.

vegetation and thus generally influenced by geographical location and climate. Following the principle of city stratification and taking into account the physico-geographical regionalization of China, we stratified the 36 cities into three categories: southern, northern-A, and northern-B, whose spatial distribution is shown in Fig. 1. Each category had different types of vacant land and non-vacant land, leading to different UVL identification criteria. The common and different aspects of the UVL identification criteria of the three city categories are summarized in Table 1.

Subsequently, the five auditors made a general assessment of the typicality of the vacant land and non-vacant land in every city according to the observation of the HRRS images and selected Shenzhen, Beijing, and Lanzhou as the representative cities from southern, northern-A, and northern-B categories, respectively. We trained one semantic segmentation model per representative city (three models in total) and used it to predict the UVL in other cities within the same city category. The models were trained with labelled data. The data labelling process is explained in Section 3.2 and the model training process is explained in Section 3.3.

3.2. Data labelling

The goal of data labelling was to obtain high-quality labelled data for model training. In each representative city, we chose training and test tiles (4 km × 4 km) as the label area. To guarantee a good model training effect, the training tiles should meet the requirements as follows: (1) covering a wide variety of urban vacant land and non-vacant land to enhance the robustness of the model, (2) maintaining a relatively high vacancy rate to reduce model bias caused by the data imbalance problem, and (3) ensuring a sufficient number of tiles to control the overfitting problem. The test tiles were not used for model training, but for robustness tests in Section 4.3, thus only needed to satisfy requirement (1) to fully test the model performance. According to the above requirements, the training and test tiles were selected through observation of the HRRS images of the whole representative cities by professional auditors. It should be mentioned that a smaller tile size was also applicable but more tiles would be needed to offer sufficient model training data. After the determination of the label area, five professional auditors finely delineated the UVL to obtain refined labels, as shown in Fig. 4.

After labelling, there were two steps for data preprocessing. First, in order to balance the overall and detailed features of vacant land in the images, export the training and test datasets from ArcGIS with an optimal spatial resolution 1.6 m rather than the original resolution (see Appendix B for explanation). Second, split the images and labels into small patches as the model input data (224 px × 224 px). A training tile was split into 121 training patches of size 256 px × 256 px with edges partially overlapped to fully utilize the data, and a test tile was split into 144 test patches of size 224 px × 224 px.

3.3. Model training

In this paper, we adopted the DeepLabv3 (Chen, Papandreou, Schroff, & Adam, 2017) model, which is one of the best methods for image segmentation (Chen et al., 2019). The model acquires a large receptive field based on the encoder ResNet (He, Zhang, Ren, & Sun, 2016) pretrained in ImageNet (Deng et al., 2009) and atrous spatial pyramid pooling consisting of parallel convolution with different expansion rates.
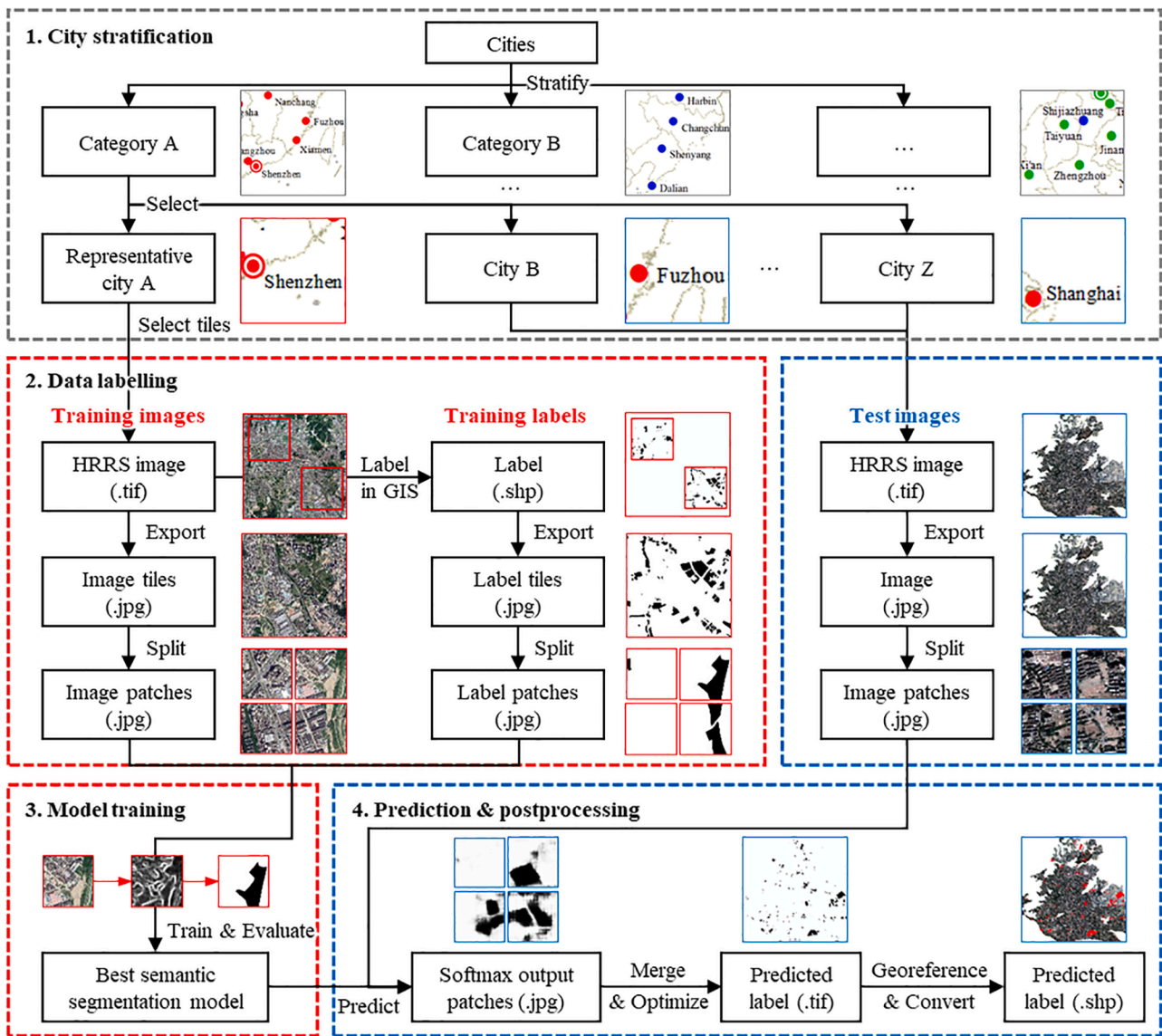
Fig. 2. Large-scale automatic identification framework of urban vacant land. The red boxes relate to model establishment, and the blue boxes relate to prediction.

The goal of the model was to discover as much true UVL as possible while maintaining a relatively high accuracy to reduce the cost of the necessary manual refinements on the results. In this paper, the precision (Equation (1)) indicated the proportion of the true UVL to all the UVL identified by the model, representing the model's accuracy, whereas the recall (Equation (2)) indicated the proportion of the true identified UVL to all the UVL in the given samples, reflecting the model's capability of discovering the true UVL. Considering the goal, both metrics were important, but the recall was more important. Therefore, we introduced the F_β-score (Equation (3)), a metric that combines precision and recall, and uses the β parameter to control their relative weights (Duque, Patino, & Betancourt, 2017). A value of $\beta = 1$ gives equal weights to the precision and the recall, whereas a value of $\beta = 2$ prioritizes the recall. Therefore, we chose F2-score as the final model evaluation metric.

$$precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (1)$$

$$recall = \frac{TruePositives}{TruePositives + FalseNegatives} \quad (2)$$

$$F_\beta = (1 + \beta^2) \cdot \frac{precision \cdot recall}{(\beta^2 \cdot precision) + recall} \quad (3)$$

In the model training process, the training patches obtained in Section 3.2 were divided into the training set and the validation set. The training set contained 75 percent of the patches while the validation set contained the rest. We used the training set to train the neural network parameters and the validation set to evaluate the model validation F2-score. We trained a series of models with different backbones and training techniques (see Appendices C and D) and selected the best models for final prediction.

3.4. Prediction and postprocessing

The goal of prediction and postprocessing was to acquire high-quality UVL identification results in the shapefile format. To realize the goal, we used softmax outputs of the semantic segmentation model and proposed the hybrid prediction and the edge optimization method to optimize the results.

Without result optimization, the model outputs in this paper should be pixel-level binary classes (UVL and non-UVL). However, some problems arose when we directly output the UVL prediction results,

Fig. 3. Examples of vacant land and non-vacant land with similar features. (a) is bare land without film and weeds. (b) is land with demolished structures. (c) is the derelict hardened ground with garbage. (d) is land under early-stage construction. (e) is land with weeds. (f) is bare land covered with green film. (j) is land under late-stage construction. (l) is green farmland. (m) is brown woodland.

which are discussed in the following paragraphs. To further optimize the results, we used softmax layer outputs of the semantic segmentation model, i.e. pixel-level softmax scores (0 ~ 1) instead of binary values (0 or 1). This method helped preserve the output information, which was fundamental for further optimization.

Based on the city stratification defined in Section 3.1, UVL prediction using the representative city's model within the same city category led to good results in all 36 cities except Chengdu, which was considered to be an atypical city. Labelling data and training models specially for Chengdu would cost too much. Through further observation on HRRS images, we found that Chengdu's UVL features were not so typical and were mixed between the southern and northern-B city categories to some extent. Therefore, we used hybrid prediction by the southern and northern-B models, which improved the identification results in Chengdu without increasing the cost. The detailed effects are demonstrated in Section 4.3 and Appendix G. The process of hybrid prediction in atypical cities was as follows: (1) select two or more city categories' representative city models according to the UVL features of the atypical city, (2) predict UVL using the selected models one by one to obtain the corresponding softmax output maps, and (3) perform a weighted average on the multiple softmax score maps to obtain the final hybrid

score map for further optimization. In Chengdu, the optimal combination of weights on the southern and northern-B model output maps was (0.5, 0.5).

To acquire the final result in shapefile format, the softmax score map needed to be transformed into a black and white image. This was realized by directly truncating the 0 ~ 1 softmax scores at 0.5, but the result was found to have spots on the edges (Fig. 5c), which caused difficulty in its conversion to shapefile format. To obtain clear edges, we introduced an edge optimization method whose effect is shown in Fig. 5d. The method consisted of the following two steps. (1) Smoothing. Blur the softmax score map to eliminate the spots and reduce noise while still keeping the general features clear. (2) Truncation. Set a threshold around 0.5. Transform the smoothed score map to a black and white image by setting the scores below the threshold to 0 and the others to 1. In this paper, we used the 'cv2.blur()' function in the OpenCV-Python library (Mordvintsev and Abid, 2017). Through visual observation, we chose the blur parameter 'kernel size' as (20, 20) and the truncation threshold as 0.6 to obtain the best edge optimization effect.

The complete process of this part was as follows (Fig. 2): (1) obtain softmax output patches from the model, (2) merge the patches as a whole raster (.tif), (3) optimize the result using the proposed

Table 1

Common and different aspects of the vacant land identification criteria of the three categories of cities.

City category	Vacant land	Non-vacant land with similar features
All in common	(a) Bare land without film and weeds (b) Land with demolished structures (c) Derelict hardened ground with garbage (d) Land under early-stage construction	(g) Park or golf course (h) Parking lot (i) Hardened ground in good condition (j) Land under late-stage construction
Southern	(e) Land with weeds	(k) Green woodland (l) Green farmland
Northern-A	(f) Bare land covered with green film	(m) Brown woodland (n) Brown farmland (o) Woodland covered with green film
Northern-B	(e) Land with weeds (f) Bare land covered with green film	(k, m) Green or brown woodland (l, n) Green or brown farmland (p) Grassland in good condition

optimization methods including hybrid prediction (if needed) and edge optimization, (4) georeference the raster to its real location in ArcGIS using georeferencing tools, (5) convert the raster to shapefile format using ArcGIS 'raster to polygon' tool, and (6) calculate the area of polygons and delete small-area polygons which were considered to be noise.

4. Results

4.1. Framework performance evaluation

The best models of the three city categories were selected after a series of parameter optimization (see Appendices B, C, and D for detailed comparisons). The final best validation F2-scores of the southern, northern-A, and northern-B models were 0.852, 0.888, and 0.817, respectively. The average best validation F2-score reached a satisfactory value of 0.852.

To further evaluate the comprehensive performance of the framework, we compared the accuracy and efficiency of the conventional manual identification method and those of the proposed automatic identification framework by the following two quantitative evaluation indicators.

The accuracy indicator was the intersection over union (IoU) value between the identification results and the gold standard. The gold standard was the identification results decided by multiple professional auditors together. The IoU is the most popular evaluation metric for the similarity between two image segmentation results (Isola, Zhu, Zhou, & Efros, 2017; Wang et al., 2018; Zhang, Isola, & Efros, 2016), defined by Equation (4).

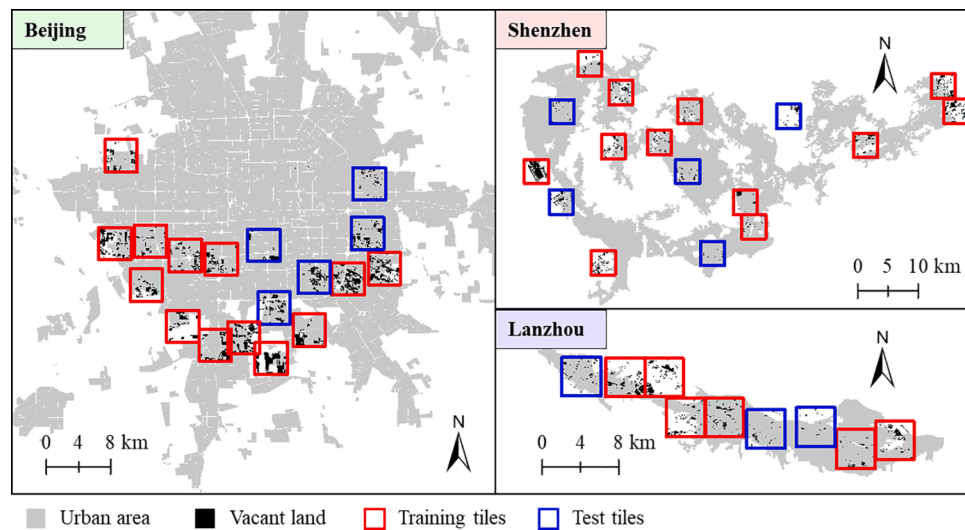$$IoU(A, B) = area(A \cap B)/area(A \cup B) \tag{4}$$

Fig. 4. Labels and tiles of the representative cities. The vacancy rates in the training tiles of Beijing, Shenzhen, and Lanzhou are 17.9%, 11.8%, and 7.7%, respectively.
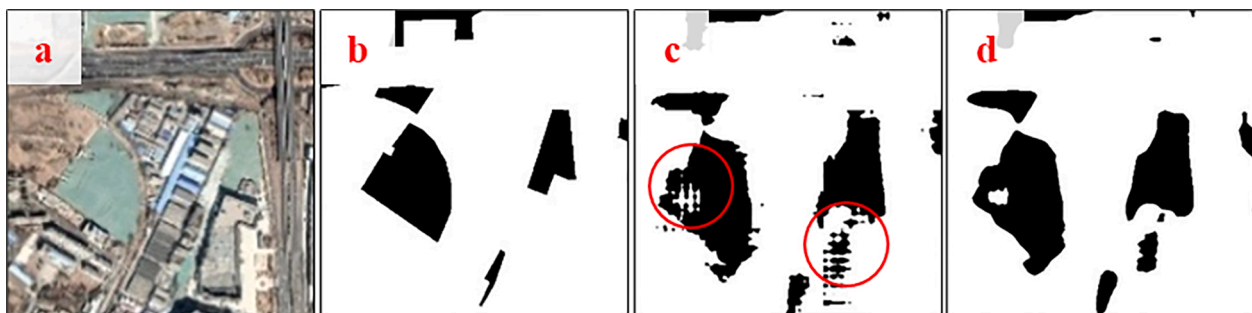

Fig. 5. Comparison between the optimized and non-optimized results. (a) is the image. (b) is the ground truth. (c) is the non-optimized result with spots (in red circles) on the edges. (d) is the optimized result with clear edges.

The five Beijing test tiles shown in Fig. 4 were used for accuracy evaluation (Appendix E). In each tile, one auditor was selected from the five professional auditors in turn to give the manual identification result, while the rest four auditors decided the gold standard together. The automatic identification results were predicted using the best Beijing (northern-A) model. For both manual and automatic methods, the IoU value between the identification result and the gold standard were calculated. As a result, the average IoU of the automatic framework reached 63.8%, over 90% that of the manual method (69.0%), which indicated that the accuracy of the automatic framework was very close

Southern category	Northern-A category	Northern-B category
Changsha	Beijing	Changchun
Chengdu	Jinan	Dalian
Chongqing	Qingdao	Harbin
Fuzhou	Taiyuan	Hohhot

Urban area ■ Vacant land
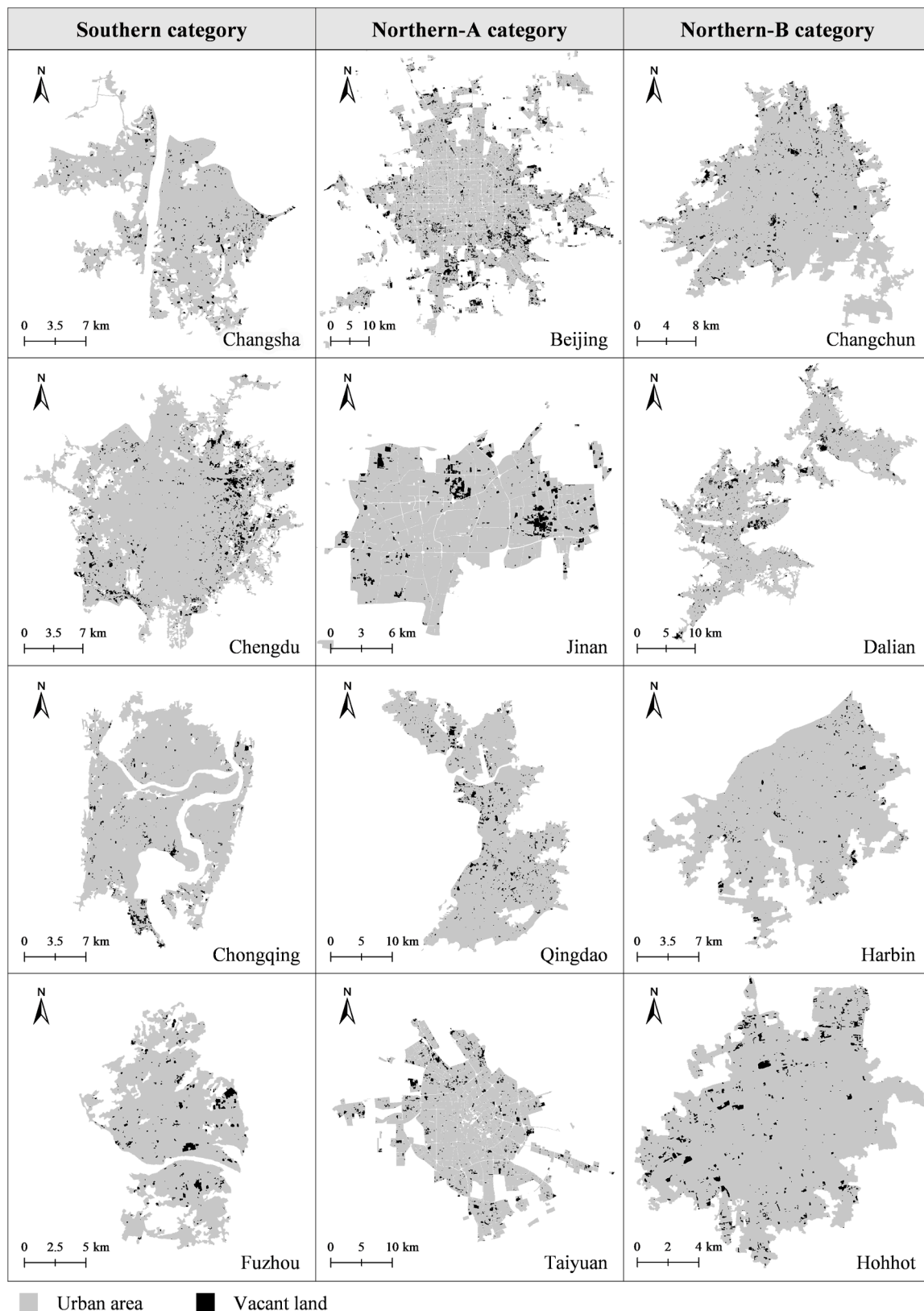
Fig. 6. Identification results of urban vacant land of the 36 cities in part (4 cities in each category).

to the level of professional auditors. Furthermore, the automatic framework achieved a much smaller standard deviation of 5.1% compared to the manual method (16.9%), which indicated that the automatic framework achieved high stability and reliability and effectively avoided high variance caused by the individual subjective difference of the auditors.

The efficiency indicator was the time required to identify the UVL per 100 km^2 urban area. For the manual method, the efficiency was simply the labelling efficiency, defined by Equation (5). For the automatic framework, the total identification time was the sum of the manual labelling time, the model training time, and the prediction and postprocessing time, and the efficiency was defined by Equation (6).

$$efficiency_{manual} = time_{label}/area_{label} \qquad (5)$$

$$efficiency_{automatic} = \left(time_{label} + time_{train} + time_{predict}\right)/area_{total} \qquad (6)$$

In this case, the times of labelling, training, and prediction were 64 h, 6 h, and 4 h, respectively. The labelling area and the total area were 704 km^2 and 12,711 km^2, respectively. According to Equations (5) and (6), the efficiency of the manual method was 9 h/100 km^2, while the efficiency of the automatic framework was 0.6 h/100 km^2, 15 times higher than the former.

Overall, compared to the conventional manual method, the proposed automatic identification framework achieved good accuracy, high stability, and high efficiency.

4.2. Identification results of 36 cities

The 36 cities' UVL identification results and the corresponding key statistical data are partly shown in Fig. 6 and Table 2, respectively, including various kinds of information, such as the vacancy rate, the spatial distribution, and other statistical properties of UVL on all scales.

According to the data in Table 2, the UVL rates of the 36 major cities except for Nanchang were all less than 10%, and the average UVL rate of the 36 cities was 4.9%, which was close to the UVL rate of New York City (5.08%) and much less than the average UVL rate of 65 typical shrinking cities in the United States (11.48%) (Song et al., 2020). This reflected that urban shrinkage in the 36 Chinese cities was generally unapparent. The data also revealed an interesting discovery: cities with larger urban areas tended to have a higher UVL rate, although the trend was not that obvious. This might reveal a possible phenomenon in which larger cities are more prone to disordered urban expansion and urban shrinkage.

From Fig. 6, three patterns of the spatial distribution of UVL were extracted: (1) the UVL is scattered in the whole urban area with a generally even distribution of the UVL density, e.g., Changsha and Harbin, (2) the UVL density increases gradually from the centre to the periphery of the urban area, e.g., Beijing and Chengdu, and (3) the UVL is concentrated in several parts of the urban area, e.g., Jinan. These patterns can be used to analyse the patterns of urban sprawl and shrinkage of each city in further studies.

4.3. Framework robustness tests and ablation study

Robustness is very important in large-scale automatic UVL identification. To test the robustness of the proposed framework, we carried out experiments of self-prediction, inter-category cross-prediction, intra-category prediction, and hybrid prediction on the best models of the three city categories. We also trained a hybrid model using the training tiles of all three representative cities to perform an ablation study on stratified prediction. The results (Appendices F, G, and H) proved that the framework based on city stratification had strong robustness for large-scale UVL identification and that using stratified prediction by the three models performed better than using one hybrid model.

In the self-prediction experiment, the model of each category was used to identify UVL on the test images of the training city. The results (see red boxes in Appendix F) showed that all the models were capable of

Table 2
Statistical data of the vacant land identification results in the 36 cities.

Category	City	Urban area (km^2)	UVL area (km^2)	UVL rate (%)
Southern	Changsha	260.8	8.6	3.3
	Chengdu	472.8	24.6	5.2
	Chongqing	290.6	7.6	2.6
	Fuzhou	124.3	5.2	4.2
	Guangzhou	335.3	19.4	5.8
	Guiyang	78.9	0.8	1.0
	Haikou	94.4	3.0	3.2
	Hangzhou	209.5	10.7	5.1
	Hefei	233.4	13.8	5.9
	Kunming	195.0	11.5	5.9
	Nanchang	197.8	24.7	12.5
	Nanjing	656.2	43.3	6.6
	Nanning	239.9	7.7	3.2
	Ningbo	145.1	6.4	4.4
	Shanghai	830.4	63.1	7.6
	Shenzhen	844.7	42.2	5.0
	Wuhan	451.3	19.4	4.3
	Xiamen	79.6	1.9	2.4
Northern-A	Beijing	1694.9	159.3	9.4
	Jinan	292.5	16.7	5.7
	Qingdao	497.4	20.9	4.2
	Taiyuan	461.9	29.6	6.4
	Tianjin	556.6	28.4	5.1
	Xi'an	313.5	12.2	3.9
	Zhengzhou	340.6	18.7	5.5
Northern-B	Changchun	272.6	18.0	6.6
	Dalian	454.2	26.8	5.9
	Harbin	359.7	6.5	1.8
	Hohhot	174.0	6.8	3.9
	Lanzhou	147.5	6.8	4.6
	Lhasa	38.1	0.7	1.9
	Shenyang	595.9	19.1	3.2
	Shijiazhuang	372.0	23.4	6.3
	Urumqi	247.4	13.4	5.4
	Xining	57.4	2.0	3.5
	Yinchuan	94.9	3.1	3.3

identifying the majority of the UVL and accurately delineating their shapes, which proved that all three models were well trained.

In the inter-category cross-prediction experiment, the model of each category was used to identify UVL on the test images of the representative cities of the other two categories. As expected, the cross-prediction results were poor. The average IoU between the results and the ground truth was 39.6%. The northern-A model was not capable of identifying approximately half of the UVL in the other two categories, while the southern and northern-B models over-identified the UVL in the northern-A category. The cross-prediction results of the southern and northern-B categories seemed to be acceptable on a large scale, but the details were not satisfactory. This experiment reflected that there were significant inter-category differences between the three categories of cities, which proved the necessity of city stratification.

In the intra-category prediction experiment, one city other than the representative city was selected randomly (except Chengdu) from each category to offer test images for the same category's model to identify UVL. The results (Appendix G) showed that all three models were capable of identifying the majority of the UVL and extracting quite accurate shapes within the same city category, which proved the effectiveness of city stratification and reflected the good generalization ability of the three models in their categories.

The hybrid prediction experiment using models of southern and northern-B categories was carried out on the atypical city Chengdu. As shown in Appendix H, the southern model over-identified some buildings as UVL, while the northern-B model under-identified land with weeds. By using hybrid prediction, the results took advantage of the two models and showed accurate identification of land with weeds and no over-identification of buildings. This experiment indicated that even dealing with the atypical city where intra-category prediction performed

poorly, the proposed framework with hybrid prediction still had good robustness. Although the hybrid prediction method was only used in Chengdu in this case study, the method is of great importance improving robustness of the framework in large-scale areas including a lot of cities.

To perform an ablation study on the effect of stratified prediction by the three representative city models, we trained one hybrid model using all the training tiles of the three representative cities for comparison. The best validation F2-score of the hybrid model was 0.850, close to the average F2-score 0.852 of the three models. Results showed that the hybrid model also had good robustness, as it was capable of identifying the majority of the UVL and extracting quite accurate shapes in cities of all three categories (Appendix G) and even in the atypical city Chengdu (Appendix H). This was reasonable because the hybrid model was trained with various types of UVL in all three city categories. Despite its robustness, its accuracy on the details was a bit lower than that of the framework with stratified prediction using the three models. In Appendix G, the hybrid model under-identified land with weeds in Shanghai and over-identified green farmland in Dalian while the framework with stratified prediction did not. The results of the hybrid model in Appendices F, G, and H also had more noise compared to the framework. Therefore, the framework using stratified prediction by the three models performed better than using one hybrid model.

5. Conclusions and discussion

In this paper, we proposed a large-scale automatic identification UVL framework based on semantic segmentation of HRRS images and city stratification, applied the framework in a case study of 36 major Chinese cities, and obtained accurate and efficient UVL identification results.

The large-scale automatic identification framework has performance advantages in comparison with the conventional manual identification method. First, using the automatic framework, the labour and time costs are greatly reduced, and efficiency can be increased by approximately 15 times. Second, the results of the automatic framework are highly stable and reliable, and their variance is much lower than that of the manual method. Finally, the automatic framework can reach 90 percent accuracy of the level of professional auditors.

The proposed framework based on city stratification has strong robustness and can perform well in various cities. City stratification is effective and necessary to achieve accurate large-scale UVL identification, as it can distinguish the inter-category differences and reduce the intra-category differences so that each city category's model can obtain good identification results within the category. For atypical cities where the intra-category model does not perform well, using hybrid prediction by multiple models can further improve the identification results. The framework using stratified prediction by multiple models performs better than using one hybrid model.

Several problems remain to be solved in future work, including (1) city stratification by automatic methods such as clustering instead of the manual method adopted in this paper, (2) using automatic methods to exclude low-quality images in the training dataset instead of manual examination, (3) using HRRS images with more bands to include more UVL features, and (4) further verifying the reliability of the hybrid prediction with more atypical cities.

6. Supplementary Material

Supplementary data to this article is available online at https://data.mendeley.com/datasets/3c8myvygjj.

Appendix A. Cross-validation tests on vacant land identification criteria.

Cross-validation results of three rounds of manual identification by five professional auditors in the process of establishing consensus on vacant land identification criteria. To include various kinds of vacant land and non-vacant land in southern and northern cities in China, two tiles of size 4 km × 4 km, one in Shenzhen (southern city) and one in Beijing (northern city), were selected as test areas in each round. The IoU between the identification results of auditor 1 and the others are shown. The average IoU increases from 44.8% to 66.8%, which indicates that the five auditors established an acceptable consensus in Round 3.

Auditor ID	1 & 2	1 & 3	1 & 4	1 & 5	Average	Standard deviation
Round 1 IoU (%)	52.5	52.3	34.1	40.2	44.8	7.9
Round 2 IoU (%)	68.5	66.4	53.1	59.8	62.0	6.0
Round 3 IoU (%)	73.1	70.9	59.4	63.7	66.8	5.5

Appendix B. Parameter analysis of data spatial resolution.

Image	Ground truth	0.3 m SR	1.6m SR	5.0m SR
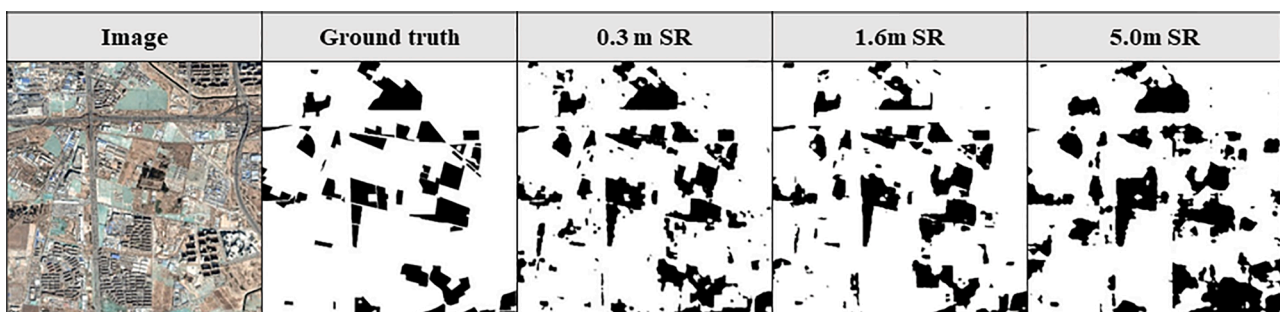
Results of different data spatial resolutions (SRs) compared to the

ground truth. The result of 0.3 m SR is accurate in shape but full of noise, the result of 1.6 m SR is accurate with less noise, and the result of 5.0 m SR is less accurate. This implies that with high data SR, the model lacks a large view of images and focuses too much on details; with proper SR, the model is capable of learning necessary details with a large view and identify vacant land based on the combined features of the vacant land and its surrounding environment; with low SR, the model lacks enough details to identify vacant land accurately. Therefore, 1.6 m is a proper choice of SR.

Appendix C. Parameter analysis of model backbones.
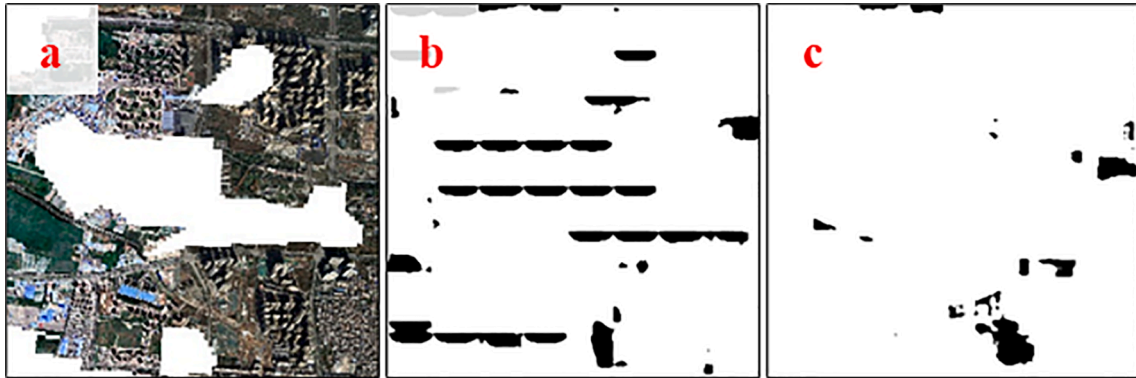
Identification results of vacant lands in Xi'an city using the northern-A model with different backbones. (a) is the satellite image, with the white part representing the area outside the city boundary. (b) is the result with strange textures using the backbone ResNet152, which indicates the severe overfitting problem and poor generalization ability of the model. (c) is the result of using the backbone ResNet18, which shows good generalization ability.

Appendix D. Parameter analysis of key training techniques.

Ablation study: the influence of the following key training techniques on model performance. (1) Learning rate decay, which can promote the model training when the model loss stops decreasing. (2) Weighted loss function, which can reduce model bias caused by the class imbalance problem in vacant land identification. (3) Data augmentation, which can enrich the training dataset and enhance the models' generalization ability. The results show that all the key training techniques have significant effects on the improvement of model performance.

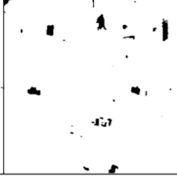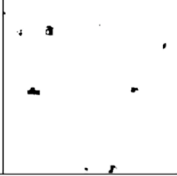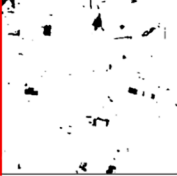| Training techniques	Best validation F2-score
Baseline	0.859
+ Learning rate decay	0.867 (+0.008)
+ Weighted loss function	0.875 (+0.008)
+ Data augmentation	0.888 (+0.013)
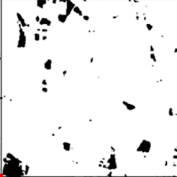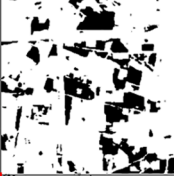
Note. The values in parentheses indicate the increment compared to the step above.

Appendix E. Results of accuracy evaluation.

Accuracy evaluation results in five Beijing test tiles. In each tile, one auditor was selected from five professional auditors in turn to give the manual identification result, while the rest four auditors decided the gold standard together. The automatic identification results were predicted using the best Beijing (northern-A) model. For both manual and automatic methods, the IoU between the identification result and the gold standard in each tile are shown. The automatic method achieved an average IoU of 63.8%, over 90% that of the manual method (69.0%), with a much smaller standard deviation.
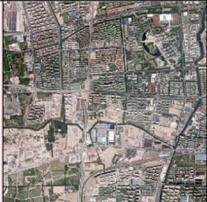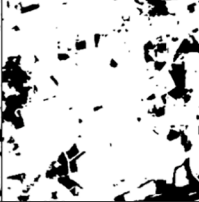
Tile ID	1	2	3	4	5	Average	Standard deviation
Manual IoU (%)	70.7	66.0	39.4	90.2	78.8	69.0	16.9
Automatic IoU (%)	65.5	65.4	55.1	70.8	62.4	63.8	5.1

Appendix F. Results of self-prediction and inter-category cross-prediction.

Self-prediction (in red boxes) and inter-category cross-prediction results of the models of three city categories and the hybrid model prediction results compared to the ground truth. The self-prediction results are accurate, which indicates that the models are well trained. The inter-category cross-prediction results are poor, which shows significant inter-category differences and proves the necessity of city stratification. The hybrid model prediction results are also quite accurate, because the hybrid model was trained with labelled data of all three representative cities. However, the hybrid model prediction results have more noise compared to the self-prediction results.

Appendix G. Results of intra-category prediction.

Intra-category prediction results of the models of three city categories and the hybrid model prediction results compared to the ground truth. The intra-category results are quite accurate, which proves the effectiveness of city stratification and reflects the good generalization ability of the three

models. The hybrid model prediction results are generally good, but the details are not as accurate as the intra-category results, such as under-identification of land with weeds in Shanghai and over-identification of green farmland in Dalian.

Appendix H. Results of hybrid prediction.

Test image	Ground truth	Hybrid prediction	Hybrid model	Southern model	Northern-B model
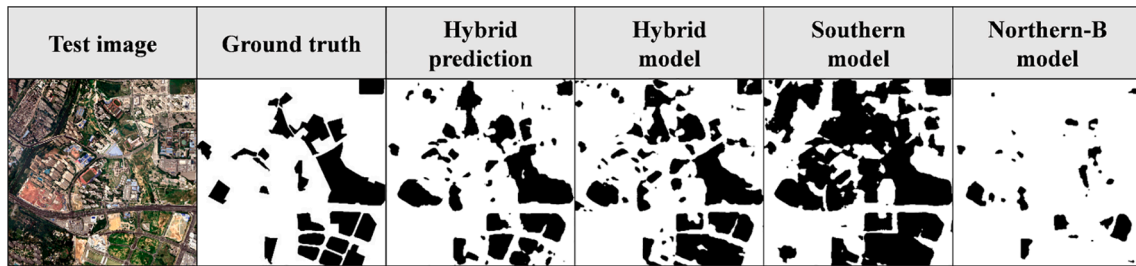
The identification results in Chengdu using the southern model, the northern-B model, the hybrid model, and the hybrid prediction by the southern and northern-B models compared to the ground truth. The southern model over-identifies some buildings, while the northern-B model under-identifies the land with weeds. The hybrid prediction result shows accurate identification of land with weeds and no over-identification of buildings. The hybrid model performs almost as well as the hybrid prediction method but with more noise. The results indicate that the framework with hybrid prediction and the hybrid model both have good robustness dealing with atypical cities.

References

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(12), 2481–2495. https://doi.org/10.1109/TPAMI.2016.2644615

Bowman, A. O. M. (2004). *Terra incognita: Vacant land and urban strategies.* Washington, D.C.: Georgetown University Press.

Bowman, A. O. M., & Pagano, M. A. (2000). Transforming America's cities: Policies and conditions of vacant land. *Urban Affairs Review, 35*(4), 559–581. https://doi.org/10.1177/10780870022184534

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. https://doi.org/10.1073/pnas.1718503115

Chen, G., Li, C., Wei, W., Jing, W., Woźniak, M., Blažauskas, T., & Damaševičius, R. (2019). Fully convolutional neural network with augmented atrous spatial pyramid pool and fully connected fusion path for high resolution remote sensing image segmentation. *Applied Sciences, 9*(9), 1816. https://doi.org/10.3390/app9091816

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 40*(4), 834–848. https://doi.org/10.1109/TPAMI.2017.2699184

Zhang, R., Isola, P., & Efros, A. A. (2016, October). *Colorful image colorization. Paper presented at the European Conference on Computer Vision.*

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018, September). *Encoder-decoder with atrous separable convolution for semantic image segmentation. Paper presented at the Proceedings of the European Conference on Computer Vision (ECCV).*

Deng, J., Dong, W., Socher, R., Li, L. J., Kai, L., & Li, F.-F. (2009). *ImageNet: A large-scale hierarchical image database.* In *Paper presented at the 2009 IEEE Conference on Computer Vision and Pattern Recognition.* https://doi.org/10.1109/CVPR.2009.5206848

Duque, J. C., Patino, J. E., & Betancourt, A. (2017). Exploring the potential of machine learning for automatic slum identification from VHR imagery. *Remote Sensing, 9*(9), 895. https://doi.org/10.3390/rs9090895

Gobster, P. H., Hadavi, S., Rigolon, A., & Stewart, W. P. (2020). Measuring landscape change, lot by lot: Greening activity in response to a vacant land reuse program. *Landscape and Urban Planning, 196*, Article 103729. https://doi.org/10.1016/j.landurbplan.2019.103729

Haase, D., Haase, A., Kabisch, N., Kabisch, S., & Rink, D. (2012). Actors and factors in land-use simulation: The challenge of urban shrinkage. *Environmental Modelling & Software, 35*, 92–103. https://doi.org/10.1016/j.envsoft.2012.02.012

He, K., Zhang, X., Ren, S., & Sun, J. (2016, June). *Deep residual learning for image recognition. Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

Heckert, M. (2015). A spatial difference-in-differences approach to studying the effect of greening vacant land on property values. *Cityscape, 17*(1), 51–60.

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017, July). *Image-to-image translation with conditional adversarial networks. Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

Kelleher, C., Golden, H. E., Burkholder, S., & Shuster, W. (2020). Urban vacant lands impart hydrological benefits across city landscapes. *Nature Communications, 11*(1), 1563. https://doi.org/10.1038/s41467-020-15376-9

Li, W., Wang, D., Li, H., Wang, J., Zhu, Y., & Yang, Y. (2019). Quantifying the spatial arrangement of underutilized land in a rapidly urbanized rust belt city: The case of Changchun city. *Land Use Policy, 83*, 113–123. https://doi.org/10.1016/j.landusepol.2019.01.032

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*(10), 3512–3520. https://doi.org/10.1002/ldr.3125

Chen, L.C., Papandreou, G., Schroff, F., & Adam, H. (2017). Rethinking atrous convolution for semantic image segmentation. *arXiv e-prints*, arXiv:1706.05587.

Long, J., Shelhamer, E., & Darrell, T. (2015, June). *Fully convolutional networks for semantic segmentation.* Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Ma, S., & Long, Y. (2019). Identifying spatial cities in China at the community scale. *Journal of Urban and Regional Planning, 11*(1), 37–50. In Chinese.

Martinez-Fernandez, C., Audirac, I., Fol, S., & Cunningham-Sabot, E. (2012). Shrinking cities: Urban challenges of globalization. *International Journal of Urban and Regional Research, 36*(2), 213–225. https://doi.org/10.1111/j.1468-2427.2011.01092.x

Mordvintsev, A., & Abid, K. (2017). Opencv-Python tutorials documentation. Retrieved from https://storage.googleapis.com/kaggle-forum-message-attachments/904222/16192/OpenCV-Python%20Tutorials-2017.pdf.

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. D., Bowman, A. O. M., Jung Lee, R., & 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

Ronneberger, O., Fischer, P., & Brox, T. (2015, October). *U-Net: Convolutional networks for biomedical image segmentation.* Paper presented at the Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Cham.

Shukla, A., & Jain, K. (2020). Automatic extraction of urban land information from unmanned aerial vehicle (UAV) data. *Earth Science Informatics, 13*(4), 1225–1236. https://doi.org/10.1007/s12145-020-00498-x

Song, X., Wen, M., Shen, Y., Feng, Q., Xiang, J., Zhang, W., … Wu, Z. (2020). Urban vacant land in growing urbanization: An international review. *Journal of Geographical Sciences, 30*(4), 669–687. https://doi.org/10.1007/s11442-020-1749-0

Sperandelli, D. I., Dupas, F. A., & Dias Pons, N. A. (2013). Dynamics of urban sprawl, vacant land, and green spaces on the metropolitan fringe of São Paulo, Brazil. *Journal of Urban Planning and Development, 139*(4), 274–279. https://doi.org/10.1061/(ASCE)UP.1943-5444.0000154

Wang, T. C., Liu, M. Y., Zhu, J. Y., Tao, A., Kautz, J., & Catanzaro, B. (2018, June). *High-resolution image synthesis and semantic manipulation with conditional GANs. Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

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