Journal of Remote Sensing
A SCIENCE PARTNER JOURNAL

RESEARCH ARTICLE

The Last Puzzle of Global Building Footprints—Mapping 280 Million Buildings in East Asia Based on VHR Images

Qian Shi[1,2], Jiajun Zhu[1*], Zhengyu Liu[1*], Haonan Guo[3], Song Gao[4], Mengxi Liu[1], Zihong Liu[5], and Xiaoping Liu[1,2]

[1]School of Geography and Planning, Sun Yat-sen University, Guangzhou 510006, China. [2]Guangdong Provincial Key Laboratory for Urbanization and Geo-simulation, Guangzhou 510275, China. [3]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430072, China. [4]Geospatial Data Science Lab, Department of Geography, University of Wisconsin–Madison, Madison, WI 53706, USA. [5]School of Geospatial Engineering and Science, Sun Yat-sen University, Zhuhai 528478, China.

*Address correspondence to: zhujj26@mail2.sysu.edu.cn (J.Z.); liuzhy236@mail2.sysu.edu.cn (Zhengyu L.)

Citation: Shi Q, Zhu J, Liu Z, Guo H, Gao S, Liu M, Liu Z, Liu X. The Last Puzzle of Global Building Footprints—Mapping 280 Million Buildings in East Asia Based on VHR Images. *J. Remote Sens.* 2024;4:Article 0138. https://doi.org/10.34133/remotesensing.0138

Submitted 27 October 2023
Accepted 17 March 2024
Published 9 May 2024

Building, as an integral aspect of human life, is vital in the domains of urban management and urban analysis. To facilitate large-scale urban planning applications, the acquisition of complete and reliable building data becomes imperative. There are a few publicly available products that provide a lot of building data, such as Microsoft and Open Street Map. However, in East Asia, due to the more complex distribution of buildings and the scarcity of auxiliary data, there is a lack of building data in these regions, hindering the large-scale application in East Asia. Some studies attempt to simulate large-scale building distribution information using incomplete local buildings footprints data through regression. However, the reliance on inaccurate buildings data introduces cumulative errors, rendering this simulation data highly unreliable, leading to limitations in achieving precise research in East Asian region. Therefore, we proposed a comprehensive large-scale buildings mapping framework in view of the complexity of buildings in East Asia, and conducted buildings footprints extraction in 2,897 cities across 5 countries in East Asia and yielded a substantial dataset of 281,093,433 buildings. The evaluation shows the validity of our building product, with an average overall accuracy of 89.63% and an F1 score of 82.55%. In addition, a comparison with existing products further shows the high quality and completeness of our building data. Finally, we conduct spatial analysis of our building data, revealing its value in supporting urban-related research. The data for this article can be downloaded from https://doi.org/10.5281/zenodo.8174931.

Introduction

As important carriers in human life, buildings play a tremendous role in fields such as urban sustainable development, building energy modeling, and urban planning [1–3]. To facilitate precise analysis and exploration of urban spatial structure across different disciplines, it is imperative to have access to high-quality, accurate, and comprehensive vector data of buildings [4]. Furthermore, the urbanization in the digital age has led to an escalating demand for dependable information of building rooftops [5–7]. Acquiring precise building rooftops information is vital for assessing the trends of urban and rural development and safeguarding the development of urban and rural ecosystems [8–10]. To provide urban planning suggestions in large-scale applications, meticulous analyses of spatial distribution and development trends of buildings are necessary. However, the execution of these analyses requires a substantial amount of reliable building data that is currently lacking. Consequently, generating a highly accurate large-scale building product holds immense significance for ongoing urban research [11–13].

Some efforts have been made to generate large-scale building data. Nevertheless, there are still limitations that hinder their applicability in East Asian countries. For instance, Microsoft has developed online building data encompassing over a billion buildings across numerous countries worldwide but lacks buildings in East Asia and the neighboring regions, resulting in spatial deficiencies. The OpenStreetMap (OSM) proposed the global building data online, covering a substantial portion of the existing human settlements in the world [14]. However, OSM building data exhibits the highest completeness in European and North American regions but drastically low completeness in Asia, particularly East Asian countries [15,16]. Furthermore, while the information of buildings and various man-made structures has been readily available for urban areas, a important number of rural buildings remain unmapped in OSM systems [17]. In summary, the quality of existing data in East Asia is inadequate, failing to meet the requirements of relevant fields [18]. Therefore, it is crucial to enhance the completeness of building footprints data in East Asian countries, especially in China. Although Zhang et al. [19] has provided building data

in selected 90 Chinese cities, the coverage area still remains limited and the complete distribution of buildings is lacking, making it impossible to conduct urban analyses on a national scale. The availability of refined building footprints data holds important potential for various research applications, including rooftop photovoltaic (RPV) potential evaluation, building-level function classification, and analysis of urban building forms. However, the incomplete buildings footprints data in East Asia limits the precision of large-scale applications and analysis in this region. For instance, referencing to Zhang et al. [20], the evaluation of large-scale RPV potential is hindered by insufficient detailed buildings data in certain areas, leading to inaccuracies in assessing specific buildings area and related information. They employ regression techniques to simulate and generate buildings area and other information as substitute data using a limited amount of error-prone buildings data. It is obvious that the presence of data errors impacts the accuracy of the final results. If more complete buildings footprints data can be used in these regions, more accurate large-scale analysis results can be obtained, and more valuable and accurate information can be provided to other studies. Therefore, further efforts are needed to generate more complete and accurate buildings data.

Based on the aforementioned analyses, it is evident that existing products primarily focus on European and American regions. Moreover, building completeness in East Asia is generally lower than in other regions. We identify 3 main factors contributing to this situation: (a) Training data in East Asia are scarce and of relatively poor quality, making it difficult to train a reliable model for large-scale mapping of buildings in East Asia. The Inria [21] and Massachusetts [22] datasets do not include East Asian data. The WHU [23], DeepGlobe [24], and Spacenet [25] datasets have limited coverage for these regions, and their image resolution and labeling quality are not sufficient to train a robust model and impede effective buildings extraction in East Asia. Given that existing building extraction models are predominantly supervised machine learning methods, discrepancies in training data significantly impact the mapping results in East Asian regions. Directly utilizing the model training by the current open dataset only allows for the extraction of a limited number of buildings in East Asia that possess clear boundaries, well-defined semantics, and standardized shapes. However, it fails to accurately extract the complex small-scale and diverse buildings that are characteristic of East Asia. Besides, the semantic information of these buildings is often lost, leading to significant errors. Accurate building extraction in East Asia still remains a huge challenge. As a result, in East Asian regions, the online building product is lacking and with subpar data quality. (b) The building appearances in European and American regions show great differences from those in East Asia. East Asian regions, particularly China, present diverse building appearances in terms of size, shape, and irregular distribution, making it challenging to accurately extract the features of buildings. Figure 1 illustrates the building appearances with regularity and distinct characteristics, along with noticeable gaps between buildings in European and American regions. Conversely, the East Asian regions showcase complex building layouts, substantial differences in appearance, size, and density, blurry gaps between buildings, and inconsistent

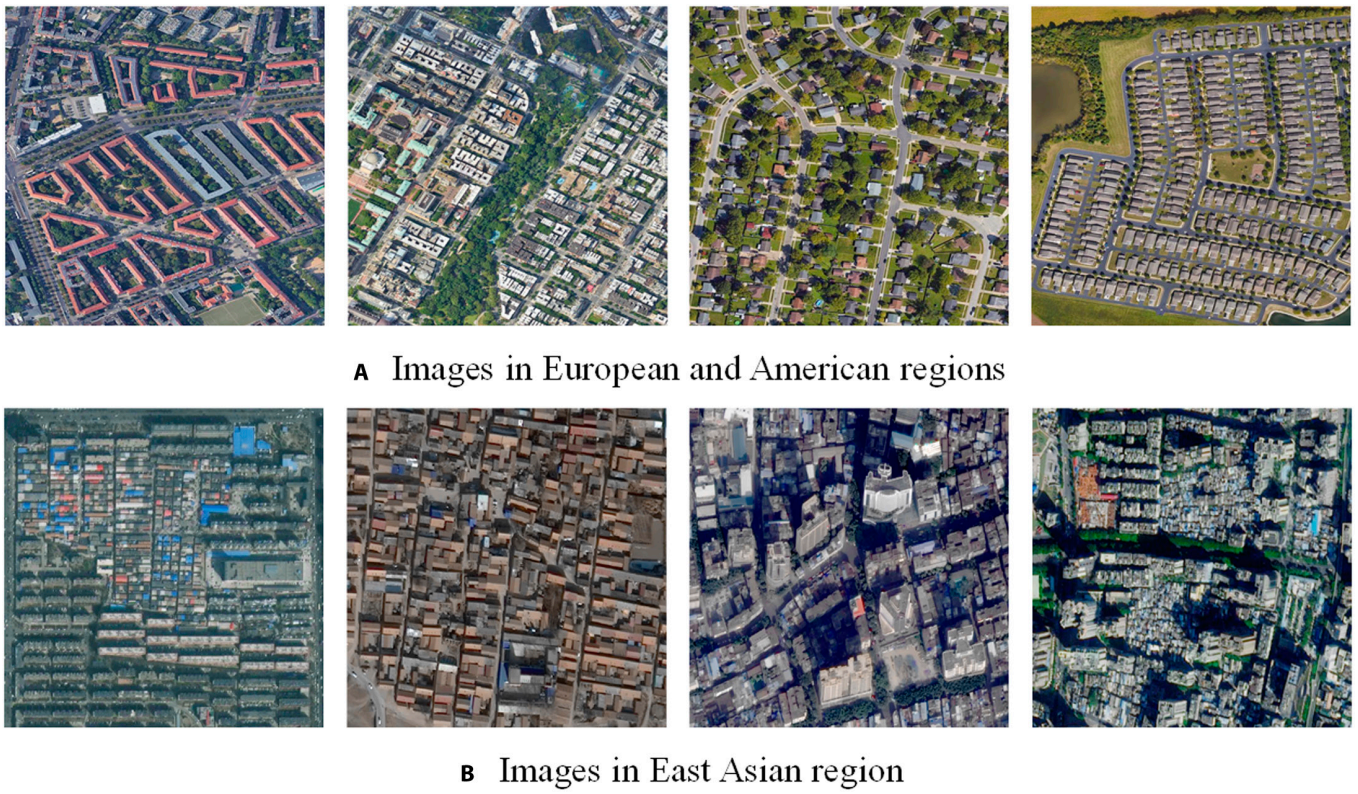

Fig. 1. Google Earth images with different distribution in different regions. (A) European and American regions. (B) East Asian regions. The images illustrate the building appearances with regularity and distinct characteristics, along with noticeable gaps between buildings in European and American regions. Conversely, the East Asian regions showcase complex building layouts, substantial differences in appearance, size, density, blurry gaps between buildings, and inconsistent characteristics. Images are from © Google Earth 2021.

characteristics. Additionally, the coexistence of high-rise and low-rise buildings further amplifies feature differences, significantly impacting model extraction capabilities. Furthermore, the presence of urban villages complicates precise building extraction. (c) The quality of open-access images in the East Asian regions, especially China, is inferior to that in European and American regions. Figure 2 demonstrates some Google Earth images at the same resolutions in different regions, highlighting the relatively poor quality of the images in East Asia, which poses challenges for accurate building feature extraction. The aforementioned disparities in building characteristics contribute to the difficulty of building extraction tasks in East Asian regions, resulting in a scarcity of publicly available large-scale building data. Hence, it is imperative to overcome the challenges associated with mapping buildings in East Asian regions and obtain complete vector data for buildings in East Asia to better support global research.

In recent years, compared with traditional methods, the ever-improving deep learning methods perform better in building extraction [26–28], making it possible for large-scale buildings mapping. For instance, Girard et al. [29] proposed a frame field learning network with a direction learning strategy that produces high-quality building footprints results. Jiang et al. [30] developed a boundary-enhanced network with a boundary postprocessing module to enhance the extraction of building boundaries. Guo et al. [31] introduced a coarse-to-fine boundary refinement network for more accurate building contours. Liu et al. [32] incorporated vector line learning to focus on building boundaries. Lin et al. [33] proposed a novel buildings edge-aware refined network (BEARNet) by learning edge prior for building extraction from high-resolution remote sensing images. Chen et al. [34] proposed a contour-guided and local structure-aware encoder–decoder network (CGSANet) by edge information learning. Yu et al. [35] proposed an edge fine-tune module as postprocessing method for building extraction. Jung et al. [36] proposed a method to elaborate edges of buildings detected in remote sensed images to enhance the boundaries of segmentation masks by using information obtained from holistically nested edge detection network. Although these methods have made progress in building extraction, the following problems may still exist when applying to large-scale mapping: (a) Insufficient generalization ability: Using a single model to extract buildings of the entire East Asian region leads to poor results due to variations in building appearance, size, color, and density. (b) Inaccurate building boundary representation: Although these methods enhance the extraction results of building boundaries, they still differ from the geometric appearance of real buildings, requiring further improvements for more precise building data. Besides, these methods just consider the multiscale information of building boundaries by a simple supervision or a multitask learning strategy without deeply fusing the boundaries information and semantic feature. It will ignore the interaction between the multiscale boundaries and the semantic learning, affecting the final results. (c) High computational cost: The models with deep structures and multiple branches often perform well and can obtain more accurate results. However, due to the large model parameters and high computational costs, these models are always with low efficiency, significantly impacting large-scale buildings mapping. To address these challenges and achieve large-scale building mapping in East Asian regions, we propose a comprehensive

Fig. 2. Google Earth images with different resolution in different regions. (A) European and American regions. (B) East Asian regions. The images show the relatively poor quality of the images in East Asia. The visual sharpness and spatial resolution of some images in East Asia look inconsistent, which may be obtained by resampling. These images are not helpful for precise building extraction. Images are from © Google Earth 2021.

large-scale mapping (CLSM) framework with 3 strategies: (a) Strong generalization ability: We propose a region-based adaptive fine-tuning strategy. Firstly, a pretraining model trained by our own dataset is used for initial inference in a subset of East Asian regions. The predicted results with high evaluation scores are then selectively utilized to generate pseudo labels, which are employed to fine-tune the models in specific regions and optimize the mapping results. (b) Stable boundary optimization ability: We introduce a boundary guidance strategy and a regularization module. The boundary information is directly weighted to enhance the semantic features of buildings and improve the extraction of building boundaries. Additionally, a generative adversarial learning network (GAN) [37] is employed to enhance the regularity of building boundaries [38–40]. Finally, the building footprints are further constrained by predefined shape features, resulting in more accurate geometric representation [41] (c) High model efficiency: We adopt model distillation [42] to transfer knowledge from an ensemble model to a lighter-weight model. This approach significantly improves efficiency while maintaining a high building extraction ability. By implementing these strategies, we have successfully generated high-quality and complete building footprints data with better coverage in East Asian regions. This fills the existing gaps in public products.

We utilized high-resolution Google Earth images with a resolution of 0.5 m to extract buildings in East Asia, including China, Japan, South Korea, North Korea, and Mongolia, and generated high-quality building footprints data. Our building dataset encompasses 2,897 cities across 5 countries, totaling 281,093,433 buildings (the details for each countries can be shown in Table 1). To validate our proposed building data, we conducted a comprehensive evaluation encompassing quantitative assessment, visual inspection, and comparison with existing products. Analysis of the results demonstrates a high level of consistency between our product and the ground truth buildings of the corresponding images, with predicted building shapes closely resembling the actual geometric shapes. Furthermore, we conducted manual annotations and accuracy evaluation in multiple sample areas, achieving an overall accuracy exceeding 87%. Our results outperform the existing products proposed by OSM and Zhang et al. [19], by displaying superior shape representation and completeness. Consequently, our proposed products satisfactorily meet the current demands of complete building data in East Asia, bridging the gaps in publicly available building footprints data and offering valuable support for urban analysis and other related research [43–45]. Utilizing our proposed products, we conducted identification and analysis of the central business districts (CBDs) through kernel density analysis of the building data, which provides valuable insights for urban analysis and other pertinent studies [46,47].

Table. 1. Statistic of our buildings data in East Asian countries

Location	Count	Size
China	248,856,847	68.90 GB
Japan	26,849,566	4.84 GB
North Korea	972,452	156 MB
South Korea	3,793,567	653 MB
Mongolia	621,001	255 MB
Total	281,093,433	74.75 GB

The contributions of this paper can be summarized as follows:

1. We generate a high-quality and comprehensive buildings data in East Asian regions, filling the gaps in the existing public buildings data. Our buildings data can also serves as valuable data for relevant studies in various domains.
2. A comprehensive deep-learning-based large-scale mapping framework (CLSM) in view of the complexity of buildings in East Asia is designed, setting a reference for research in related fields.
3. Our buildings data undergoes rigorous evaluation from multiple perspectives. We utilized our buildings data to analyze the CBDs of 2 select cities in China. The analysis results confirm the utility and effectiveness of our buildings data in facilitating urban-related studies.

Methods

Study area

The study area of this paper encompasses East Asia, specifically China, Japan, South Korea, North Korea, and Mongolia. The availability of publicly accessible building products in these regions is extremely limited, with existing data exhibiting incompleteness and low accuracy, thereby hindering related research applications. East Asia stands out as one of the most developed regions in Asia, experiencing rapid urbanization. Acquisition of large-scale building data in East Asia would significantly contribute to urban management and spatial planning studies in the region. East Asia region is situated along the coastal region of East Asia, characterized by a considerably high level of overall development, particularly in Japan, South Korea, and China. Consequently, there is an urgent demand for comprehensive building data within East Asia.

Data

Goole Earth imagery

Google Earth, an open-source platform, offers global remote sensing images at various spatial resolutions. To obtain high-quality building footprints data in the East Asian regions, it is necessary to extract detailed features from high-resolution remote sensing images. Consequently, we selected the most recent level 18 images available from 2020 to 2022 in the East Asian regions, which possess a spatial resolution of 0.5 m. The overall size of these images exceeds 100 TB. Due to the computer hardware performance limitations, we utilized a sliding window approach with a size of 512×512 to extract buildings. The final results were then merged to derive the desired vector data.

Global urban boundaries

The delineation of global urban areas in 7 years (i.e., 1990, 1995, 2000, 2005, 2010, 2015, and 2018) is established by global urban boundaries (GUB) [48], which relies on the 30-m global artificial impervious area data [49]. Since the images acquired for this study span from 2020 to 2022, the GUB data from 2018 is selected as a reference to construct the training dataset. Given

the presence of diverse land use types within the East Asian imagery, we utilized the GUB data to classify urban and non-urban areas into distinct sampling regions for positive and negative samples. This approach was adopted to enhance the scientific rigor of the training dataset.

Local climate zones

The local climate zone product, as introduced in [50], is developed through the utilization of Sentinel-1 and Sentinel-2 data alongside deep learning techniques. This local climate zone classification system incorporates various building categories characterized by distinct heights and densities. In the process of data preparation, these categories were employed to ensure the acquisition of diverse training samples with varying appearances, aiming to enhance the accuracy of large-scale mapping.

A CLSM framework

In this paper, we present a comprehensive framework for large-scale mapping, as illustrated in Fig. 3. Our focus lies specifically on generating high-quality building footprints data in the East Asian regions. The workflow encompasses 3 key steps: (a) data preparation, (b) model training, and (c) large-scale mapping. The data preparation stage holds crucial significance since contemporary building extraction techniques heavily rely on adequate training data [51,52]. Furthermore, the scarcity and complexity of building samples in East Asian regions, particularly in China, contribute to the deficiency and poor quality of existing public products. This challenges the accurate learning of essential features. To address this issue, we placed special emphasis on capturing the complexities of buildings in East Asian regions, particularly China, during the data preparation phase. This allows us to construct a diverse training dataset that encompasses various building types. Furthermore, during the model training phase, we employed specific optimization modules to enhance the performance of building extraction. Lastly, in the large-scale mapping stage, we adopted some strategies including region-based adaptive fine-tuning, to generate a refined and high-quality building footprints data in the East Asian regions.

Boundary-enhanced network (BE-Net)

In order to achieve precise and high-quality mapping outcomes, we propose a boundary-enhanced network (BE-Net) for the building extraction task. The architectural representation of BE-Net can be observed in Fig. 4. Our design incorporates the Attention U-Net [53] as the baseline model, augmented with global attention, boundary enhancement, and regularization modules. The Attention U-Net has been proven effective in

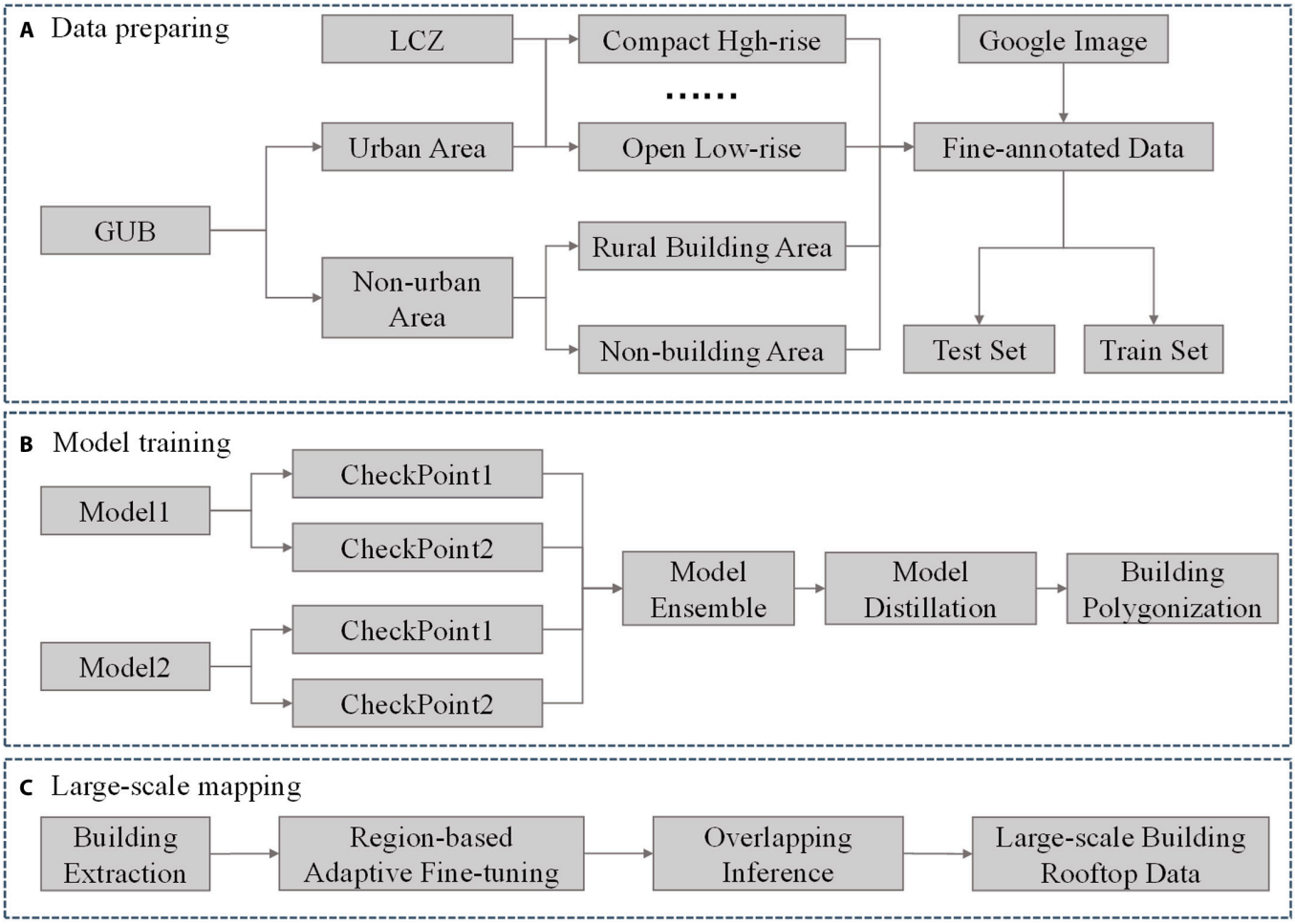

Fig. 3. Diagram of the CLSM to generate vector data of buildings in East Asia. It encompasses 3 key steps: (A) data preparation, (B) model training, and (C) large-scale mapping.

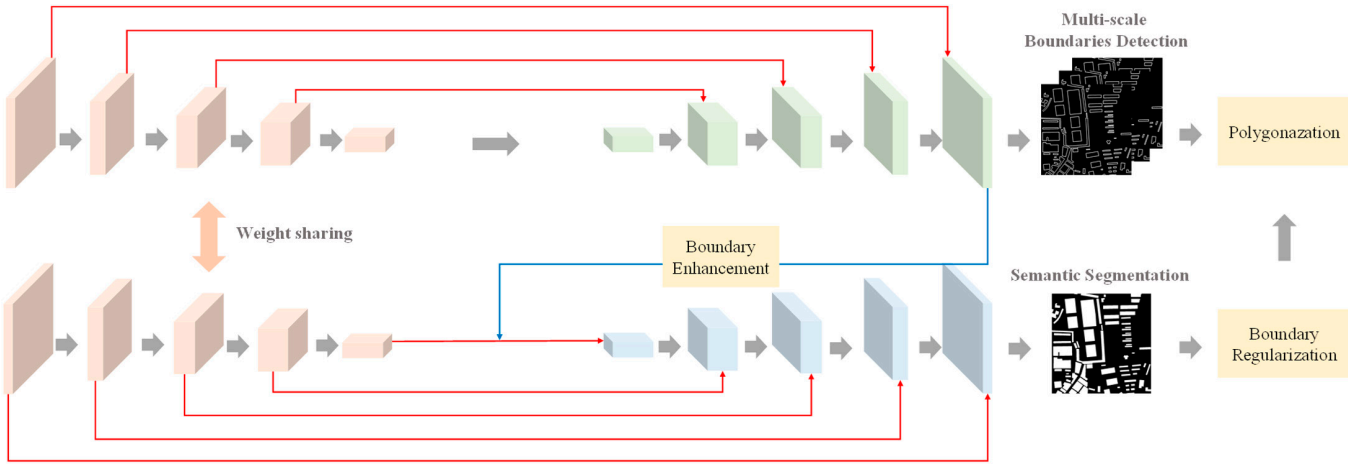

Fig. 4. Flowchart of the proposed BE-Net for building extraction. It includes 2 branches: (a) multiscale boundaries detection branch and (b) semantic segmentation branch. We designed the boundary enhancement module and boundary regularization module to obtain better results that is close to the real buildings.

semantic segmentation tasks within remote sensing domains [54]. Leveraging its structure as the foundational model greatly benefits the building extraction task. Based on Attention U-Net, we propose a boundary enhancement module and regularization module to further improve the performance of the network. The boundary enhancement module is designed for enriching the multiscale boundary information in the building extraction task. We designed a new branch for Attention U-Net to directly detect the multiscale boundaries of buildings in order to obtain more detailed features. We also added a global attention mechanism in this module to offer consideration for global dependencies among buildings, reducing the likelihood of omission. In the semantic branch of our network, we introduced a regularization module to enhance the accuracy of predicted building boundaries. We employ adversarial learning techniques and utilize a polygonization module to enforce shape constraints on buildings, ultimately yielding more precise building data.

Boundary enhancement module

The East Asian regions, notably China, exhibit complicated building layouts, significant variations in appearance, size, and density, as well as blurry gaps between buildings, and inconsistent characteristics. Additionally, the coexistence of high-rise and low-rise buildings further exacerbates these differences, greatly affecting the capabilities of model extraction. Therefore, to better capture more comprehensive boundary details for higher performance of buildings extraction in East Asia, we employed a multiscale building boundary learning strategy in this module. This strategy involves dividing the boundary labels into 3 levels and extracting boundary information at varying scales through 3 distinct convolution layers. Utilizing the multiscale boundary information learned by the boundary detection branch, we integrated it with the global attention module to augment the boundary feature of buildings and encourage a heightened focus on boundary pixels. As illustrated in Fig. 5, the different scales of boundary features were computed independently. They were then divided into 4 branches, which underwent dot product operations with the features within the attention module. Subsequently, a series of convolution layers and down-sample layers were employed to strengthen these features

We added an attention mechanism in the boundary enhancement module to further improve the performance of our proposed model. The architecture of the attention mechanism is derived from the GC Block of the Global Context Network [55]. The global attention mechanism improves upon the Non-Local Block [56] by reducing parameters while enhancing attention module effectiveness. Additionally, it incorporates the channel attention mechanism structure from the SE Block [57], resulting in a remarkable performance. Therefore, we adopted the GC Block as the fundamental structure for our global attention mechanism, depicted in Fig. 6. The mechanism operates as follows: First, the input feature undergoes dimension transformation to calculate similarity, resulting in an attention map generated by a softmax classifier. Next, weights are assigned to corresponding query points through a multiplication operation. Subsequently, the related feature is obtained via narrow convolution. Finally, the original feature and the obtained feature are combined through an addition operation, yielding the final feature representation.

After acquiring the enhanced feature through the GC Block, we proceed to reinforce it using the multiscale boundaries information obtained from the boundaries enhanced module. This process, as depicted in Fig. 5, can be described as follows:

$$Y = (1 + B) \circ X \tag{1}$$

where X represents the features obtained by the GC Block; Y represents the enhanced features; B represents the down-sampling boundary information; and $\circ$ symbolizes the dot product operation. To prevent excessive suppression of nonboundary pixel features and maintain accuracy in subsequent building extraction, a value of 1 is added to B. As shown in Fig. 6, the boundaries information is divided into 4 branches. Consequently, this process is iterated 4 times, targeting different scales to enhance the feature and improve the effectiveness of building boundary extraction.

Boundary regularization module

The regularization module achieves high-quality building shapes through adversarial learning, as illustrated in Fig. 7. This module bears a resemblance to GAN, comprising a generator and a discriminator. During training, both the generator and the

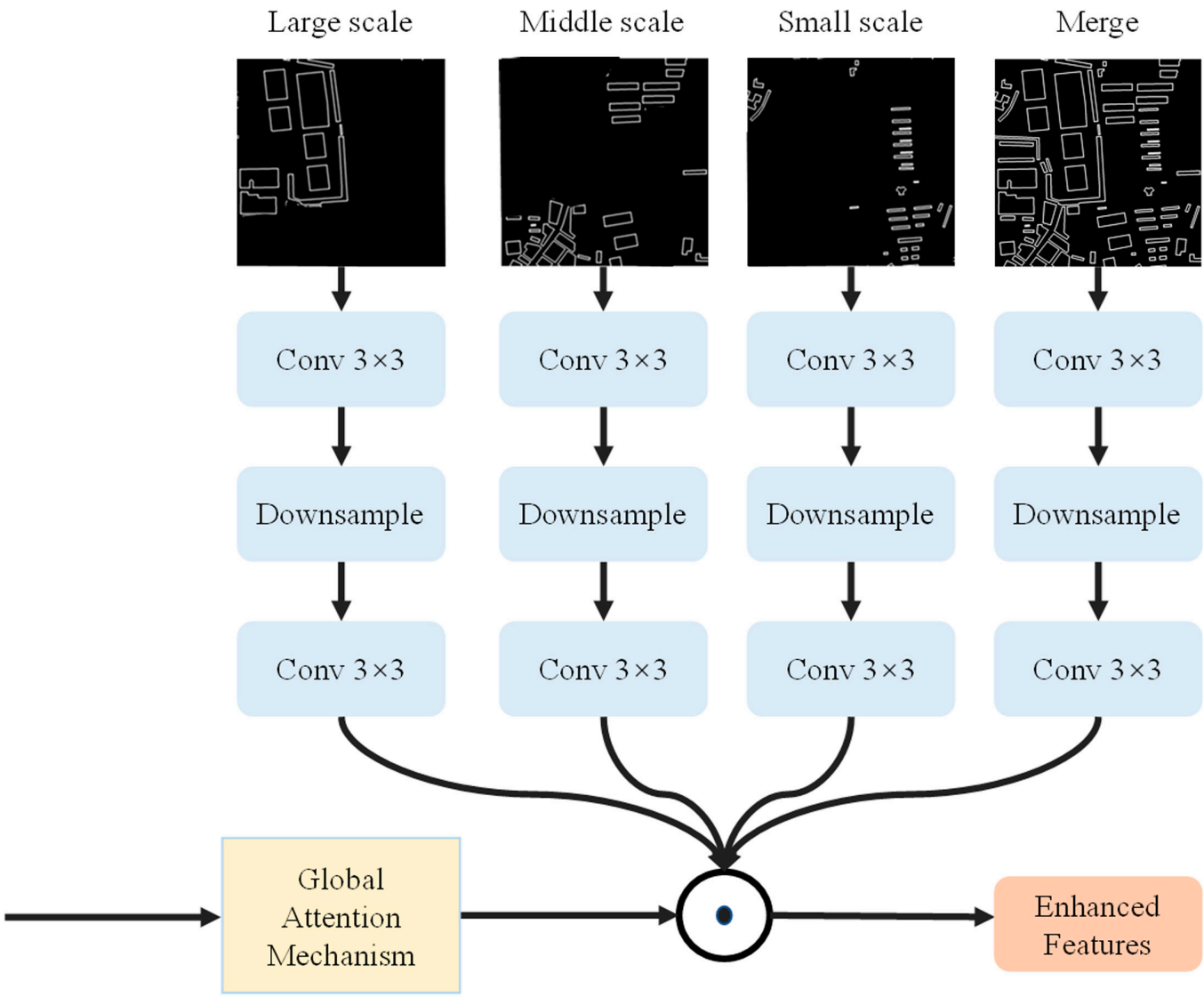

Fig. 5. Architecture of the boundary enhancement module. We utilize multiscale boundary information learned by the boundary detection branch to enhance the features obtained from the global attention mechanism.

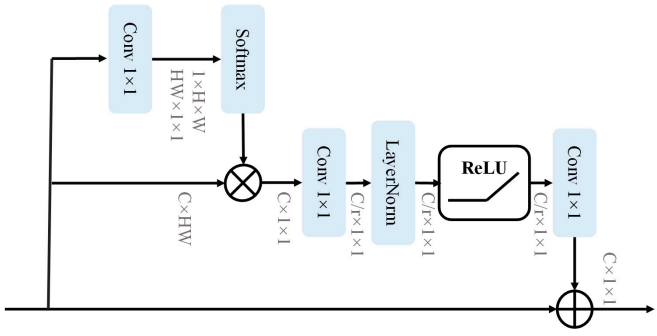

Fig. 6. Architecture of the global attention mechanism.

discriminator engage in adversarial learning. However, during inference, only the generator is utilized for optimizing building shapes. The shape learning process relies on a simple Res-Net [58] structure to produce optimized predictions. Due to the distinct data formats between real labels (with values [0, 1]) and prediction results (probability maps), the discriminator can easily distinguish them. To address this, we employed 2 separate encoders and a weight-sharing decoder to reconstruct the input data, reducing the disparity between the 2 formats. A simple convolutional neural network structure served as the discriminator, extracting features from the input result to discern whether it originates from the reconstructed mask or the regularized output. Through adversarial learning, a balance was achieved between the generator and discriminator, resulting in high-quality building shapes within the model prediction.

Polygonization module

Following the acquisition of relatively regular building results using our model, corner information for each building is derived based on its boundary and semantic information. By connecting this corner information, a closed polygon building footprints is generated. Despite the regularization module's ability to achieve satisfactory results, the vectorization process still yields numerous corners and relatively complex shapes. Given

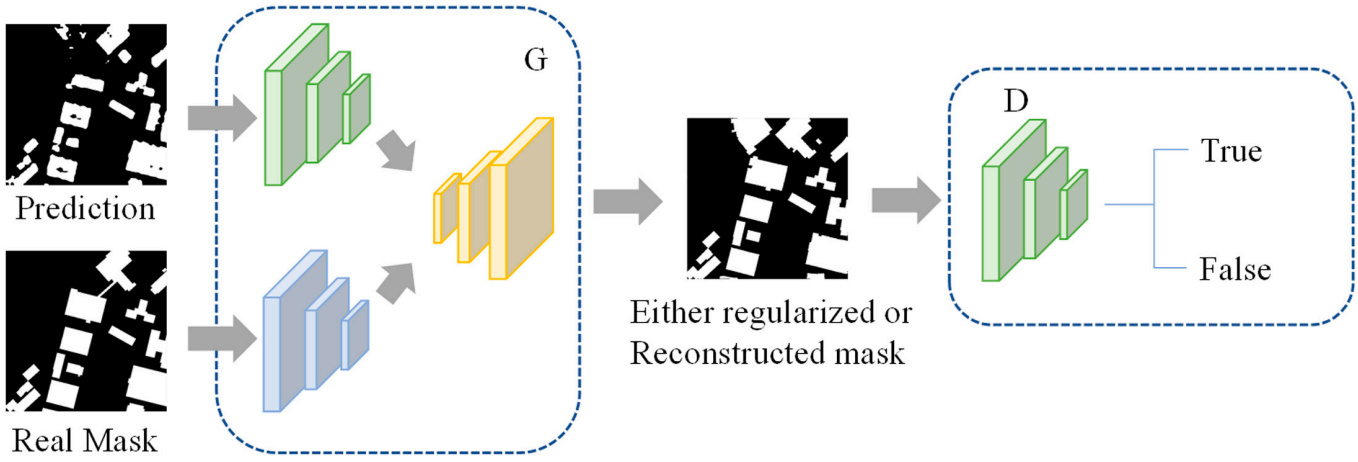

Fig. 7. Architecture of the boundary regularization module. It can further enhance the boundary information of buildings by adversarial learning and obtain the high-quality results that resemble the shape of real buildings.

that building shapes typically exhibit highly regular features, such as long straight lines and right angles, we employ the post-processing method proposed in [41]. This method serves to further simplify the result, enabling the extraction of building footprints data with more accurate geometric shapes.

Loss function

Our method comprises 2 main stages: feature extraction training and adversarial training. In the first stage, BE-Net was trained on a labeled dataset to achieve accurate prediction results. Subsequently, the generator and discriminator underwent adversarial learning to refine the predictions with improved spatial details. The aforementioned training processes contribute to the overall effectiveness and refinement of our method. The detailed information about the optimization process is as follows.

According to the description above, the overall loss can be defined as:

$$L = L_F + L_R \tag{2}$$

where L_F and L_R represent the loss of the feature learning part and the regularization module, respectively. According to the abovementioned model description, L_F can be decomposed as:

$$L_F = L_S + \sum L_B^i \tag{3}$$

where L_S represents the loss of the semantic segmentation branch, and L_R^i stands for the loss of the boundary extraction branch at specific scale. Both of these losses are combined with the cross entropy loss and dice loss, which can be denoted as:

$$L_{CE} = \frac{1}{N}\sum_i -\left[y_i \log(p_i) + (1-y_i)\log(1-p_i)\right] \tag{4}$$

$$L_{dice} = 1 - \frac{2\left|\hat{Y}\right| \cap |Y|}{\left|\hat{Y}\right| + |Y|} \tag{5}$$

where y_i is the label of i, p_i is the corresponding predicted result. In Eq. 5, Y is the label and $\hat{Y}$ is the corresponding prediction result.

According to the description of the regularization module, L_R can be divided into 2 parts: (a) the reconstruction loss of the generator; and (b) the adversarial loss of the GAN. Therefore, based on [38], we can construct the L_R loss function. The detailed information can be found in [38]

Large-scale buildings mapping

To enhance the accuracy of large-scale mapping, some strategies should be employed to aid the inference process. Within our proposed framework, CLSM, specific strategies are designed to address various inference problems and facilitate improved prediction accuracy.

Model ensemble

In large-scale mapping applications, it is crucial to ensure model stability. To achieve this, we have incorporated an ensemble strategy, building upon the approach presented in [19] with certain modifications. In [19], the authors utilized multiple checkpoints of the same model to obtain ensemble results. In our approach, we enhanced model stability by training 2 models with distinct parameter settings, creating a multiview prediction network. This additional ensemble approach further improves the overall stability of the model.

Model distillation

The utilization of a multimodel and self-model ensemble strategy results in substantial computational costs, posing limitations in practical applications, particularly for large-scale mapping tasks. To overcome this issue, it is imperative to enhance the efficiency of building extraction while maintaining performance. In addressing this challenge, we employed model distillation to transfer knowledge from an ensemble model to a more lightweight counterpart. As depicted in Fig. 8, our approach involves 2 steps. Firstly, the prediction outputs of the ensemble model are utilized to generate pseudo labels, which serve as the learning target for the lightweight model. Subsequently, through backpropagation, the lightweight model is trained to approximate the performance of the ensemble model. This allows for efficient knowledge transfer and improves the effectiveness of building extraction while reducing computational overhead.

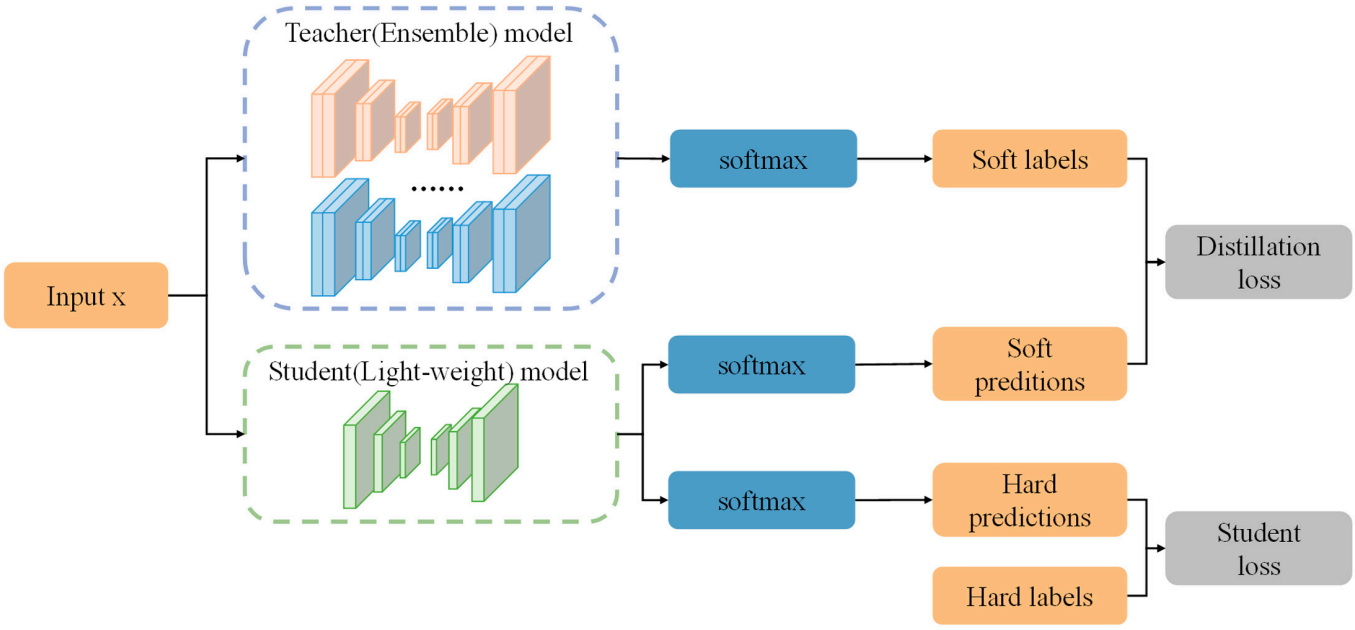

Fig. 8. Structure of the model distillation used in this paper. We utilize hard labels and the results obtained from the ensemble model to train a lightweight model, in order to improve the efficiency in the large-scale mapping process.

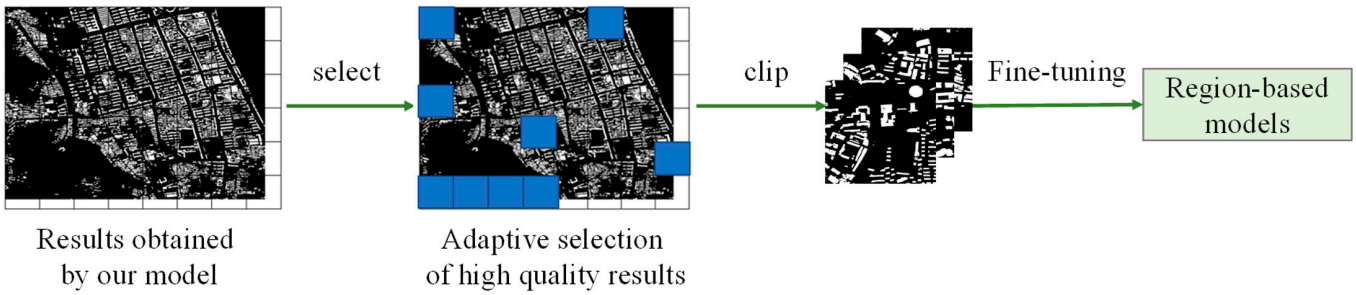

Fig. 9. Flowchart of the region-based adaptive fine-tuning. It can adaptively select the stable prediction results as the pseudo labels to fine-tuning the pretraining model in order to obtain more excellent results in specific regions.

Region-based adaptive fine-tuning

In large-scale mapping applications, the generalization ability of models plays a vital role in achieving effective transferability. To further enhance the transferability of our model, we proposed a region-based adaptive fine-tuning strategy to optimize its generalization capabilities. As illustrated in Fig. 9, our approach involves several steps. Firstly, each region undergoes a coarse extraction process through predetection, generating probability maps for individual regions. Subsequently, these probability maps are divided into blocks, and an adaptive evaluation metric is employed to quantitatively assess the quality of the prediction results. It is evident that accurate and stable prediction results should yield probabilities close to 1 for buildings and close to 0 for backgrounds. Conversely, unstable and ambiguous prediction results tend to yield probabilities close to 0.5. Leveraging this characteristic, we selected relatively stable prediction results to generate pseudo labels, which were then used for fine-tuning the model. This process yielded a submodel tailored to the specific region, enabling its effective utilization in large-scale mapping tasks.

Overlapping inference

In large-scale mapping applications, image clipping into patches is often necessary due to computer hardware limitations. However, this procedure can introduce significant discontinuities in the final merged results. To mitigate this issue, we adopted the concept of slide window prediction, as illustrated in Fig. 10. By defining windows with a certain overlapping rate (represented by the orange boxes), we calculated the results of the overlapping regions (the green boxes) using a mean operation to enhance their continuity. The nonoverlapping regions (indicated by the blue boxes) were retained without further processing. This strategy greatly reduces the visual artifacts caused by splicing. Moreover, by avoiding the clipping and merging processes altogether, the efficiency of our model is further improved, resulting in faster and more streamlined operations.

Results

In this section, a comprehensive evaluation of our proposed building data is presented, including quantitative evaluation, visual evaluation, and comparison with existing products.

Quantitative evaluation

We conducted fine-scale mapping of buildings in the East Asia region and will publicly release the corresponding building products. For the Chinese region, where existing products are scarce, we performed model inference on buildings across the

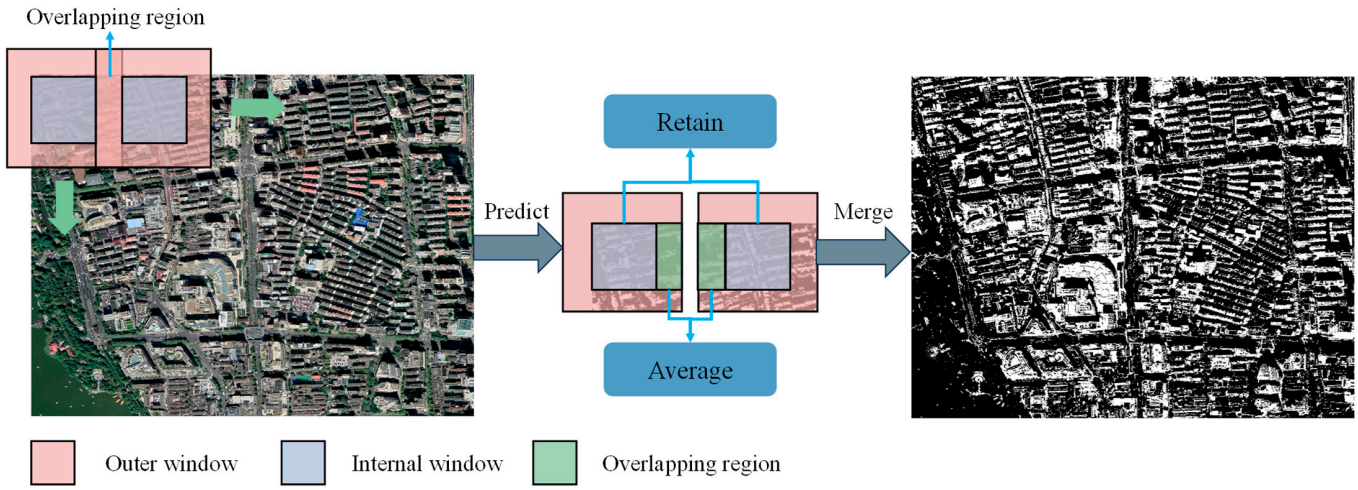

Fig. 10. Illustration of overlapping inference. By defining windows with a certain overlapping rate (represented by the orange boxes), we calculated the results of the overlapping regions (the green boxes) using a mean operation to enhance the continuity of the results.

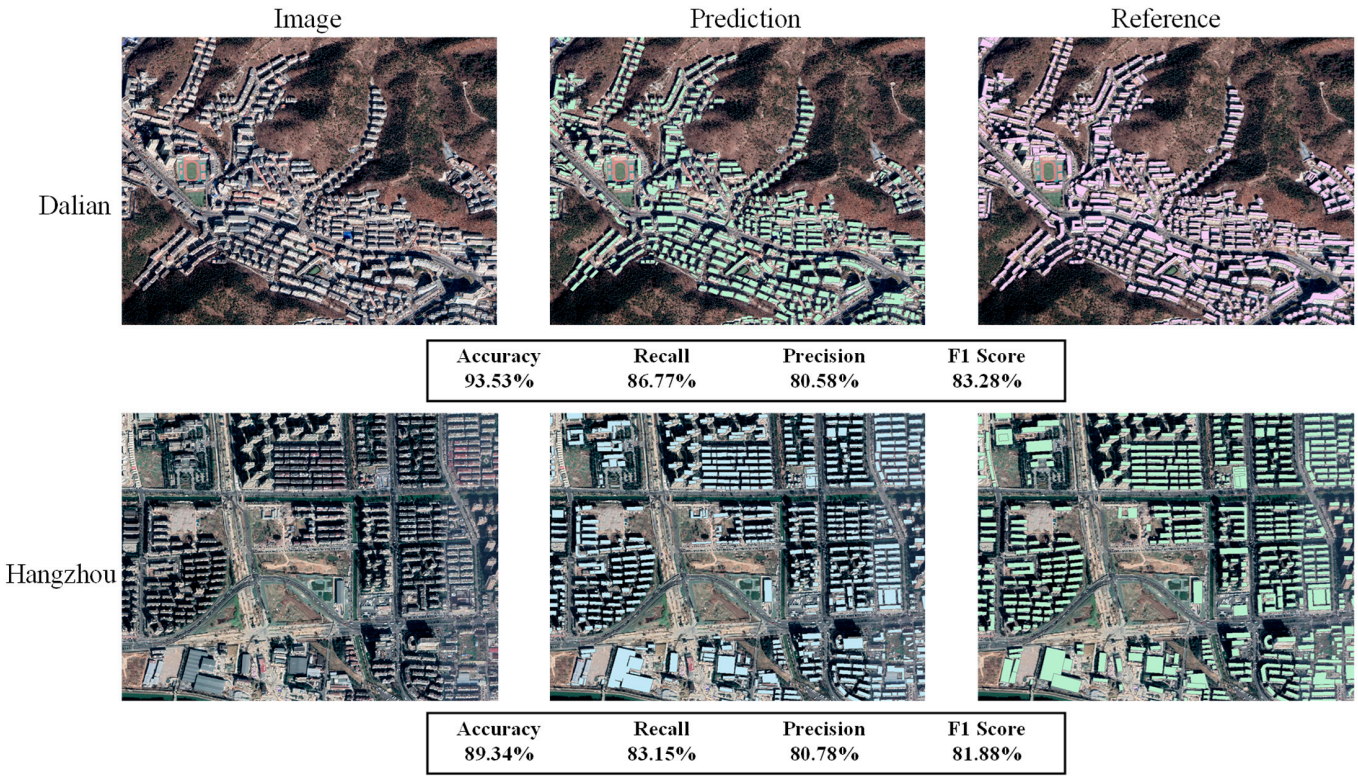

Fig. 11. The evaluation results in Dalian and Hangzhou. Images are from © Google Earth 2021.

entirety of China, generating a complete and high-quality building dataset. In Japan, South Korea, North Korea, and Mongolia, we optimized and supplemented the OSM data to obtain higher-quality building data in these regions. Throughout the large-scale mapping process, the data volume and workload of China were the largest, with no auxiliary data available for building mapping. Therefore, we primarily conducted a quantitative evaluation in China.

Firstly, to assess the accuracy of our product, we selected validation sampling areas in several Chinese cities and manually annotated reference building vector data based on reference images. Accuracy results were calculated for different cities to evaluate the performance across diverse regions. Figure 11 illustrates the accuracy evaluation results of Dalian and Hangzhou, demonstrating highly consistent predicted results with the reference data. Evaluation metrics reached a high level of performance. Likewise, Fig. 12 presents the accuracy evaluation results of Zhuhai and Shanghai, exhibiting an overall accuracy exceeding 87% and close proximity between the predicted and annotated vector data. On average, the recall and precision rates reached 84.52% and 81.06%, respectively, proving the model's ability to accurately identify buildings in various regions. The

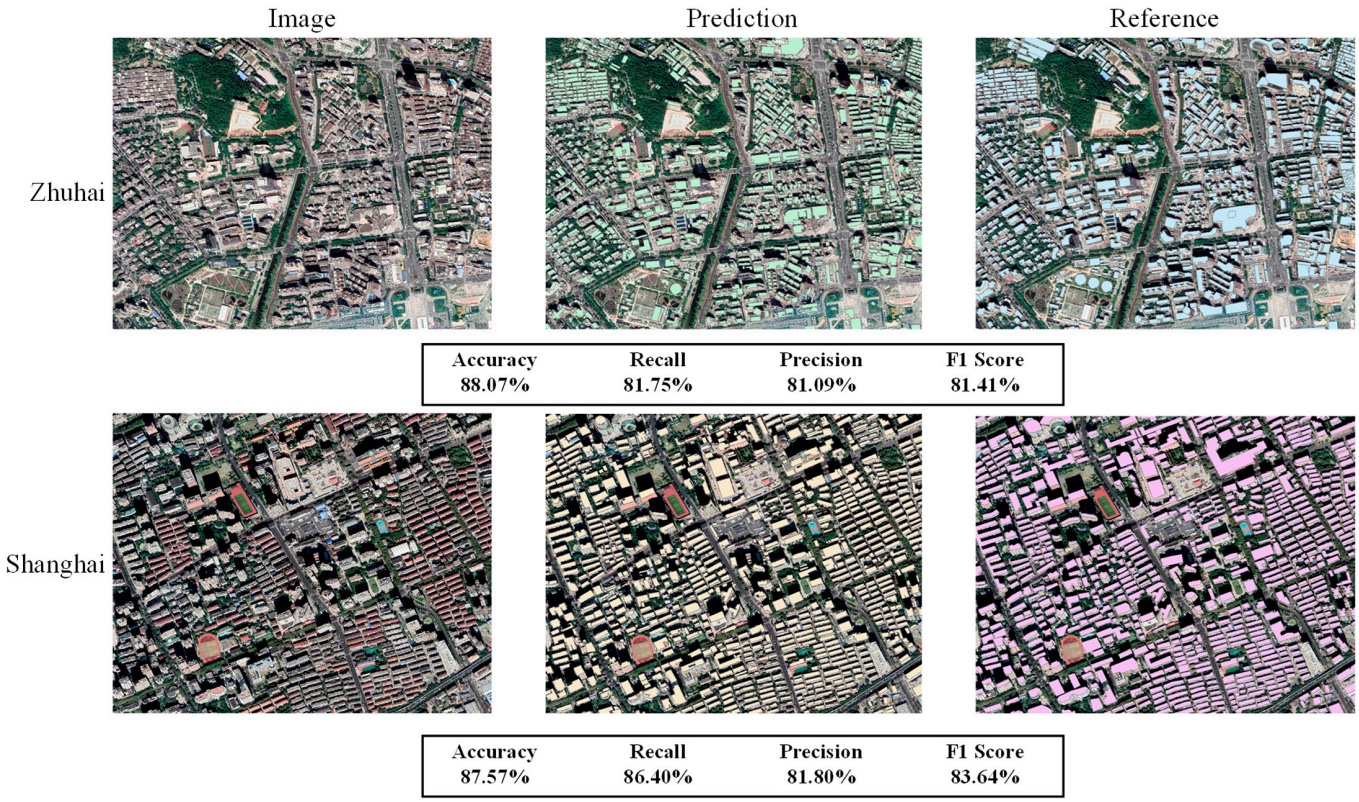

Fig. 12. The evaluation results in Zhuhai and Shanghai. Images are from © Google Earth 2021.

evaluation results depicted in Figs. 11 and 12 highlight the model's stability across diverse regions, consistently achieving multiple metrics over 80%. This capability enables detailed and reliable mapping of large-scale regions, further substantiating the high quality of our product.

Visual evaluation

Our proposed building data have been subjected to postprocessing techniques such as regularization, resulting in geometric shapes that closely resemble real buildings. Hence, this section focuses on presenting visual results to further exemplify the high quality of our product through a visual evaluation of different regions and various types of buildings.

To visually showcase the detailed prediction results, we selected several representative cities in China for visualization, as demonstrated in Fig. 13. Beijing, Shanghai, Guangzhou, Shenzhen, Chengdu, Chongqing, Nanjing, and Suzhou were chosen as these cities encompass diverse characteristics. Overall, the model achieved remarkable visualization outcomes in each city, accurately aligning with the imagery while portraying buildings with regular geometric shapes. However, varied building characteristics were observed across different cities, resulting in distinct visual effects. In cities like Guangzhou and Shenzhen, located in Guangdong province with numerous urban villages, buildings tend to be small and densely clustered. Conversely, Chongqing, situated in a mountainous region, exhibits a notable difference in the spatial distribution of buildings compared to other cities. Other cities also display variations in building appearance, yet the overall results remain complete with exceptional visualization quality. Figure 14 presents detailed results in different zoom-in areas, showcasing the buildings' boundaries depicted by green lines with precise geometric shapes. It further signifies the exceptional quality of our product.

The proposed building product demonstrates exceptional detail and facilitates the accurate identification of buildings in densely populated areas. This attribute holds particular significance for building mapping in China, especially within urban villages. Consequently, we have meticulously selected and presented results pertaining to several urban villages, as depicted in Fig. 15. We intersected the urban village areas with our buildings data, thereby preserving and exhibiting the buildings exclusively within these areas while temporarily concealing data pertaining to external buildings. It is worth noting that the urban village areas utilized in this study were derived from our own research and on-site surveys. Upon examining the imagery showcased in Fig. 15, we observed that buildings within urban villages exhibit dense packing and a chaotic layout, primarily comprising small-scale structures. Extracting buildings within such areas presents a significant challenge, primarily due to the poor quality or even absence of data in existing products. Despite this, our product yields diverse and plentiful building information specifically within urban villages, aligning remarkably well with the corresponding imagery. Notably, it excels in identifying individual small buildings characterized by regular geometric shapes, underscoring the robust performance of our model. This further highlights the superiority of our product and provides invaluable data support for research applications related to urban village analysis and other pertinent studies.

We have also included the corresponding results for visual analysis in regions such as Japan, South Korea, North Korea, and Mongolia, as illustrated in Fig. 16. Notably, noticeable variations

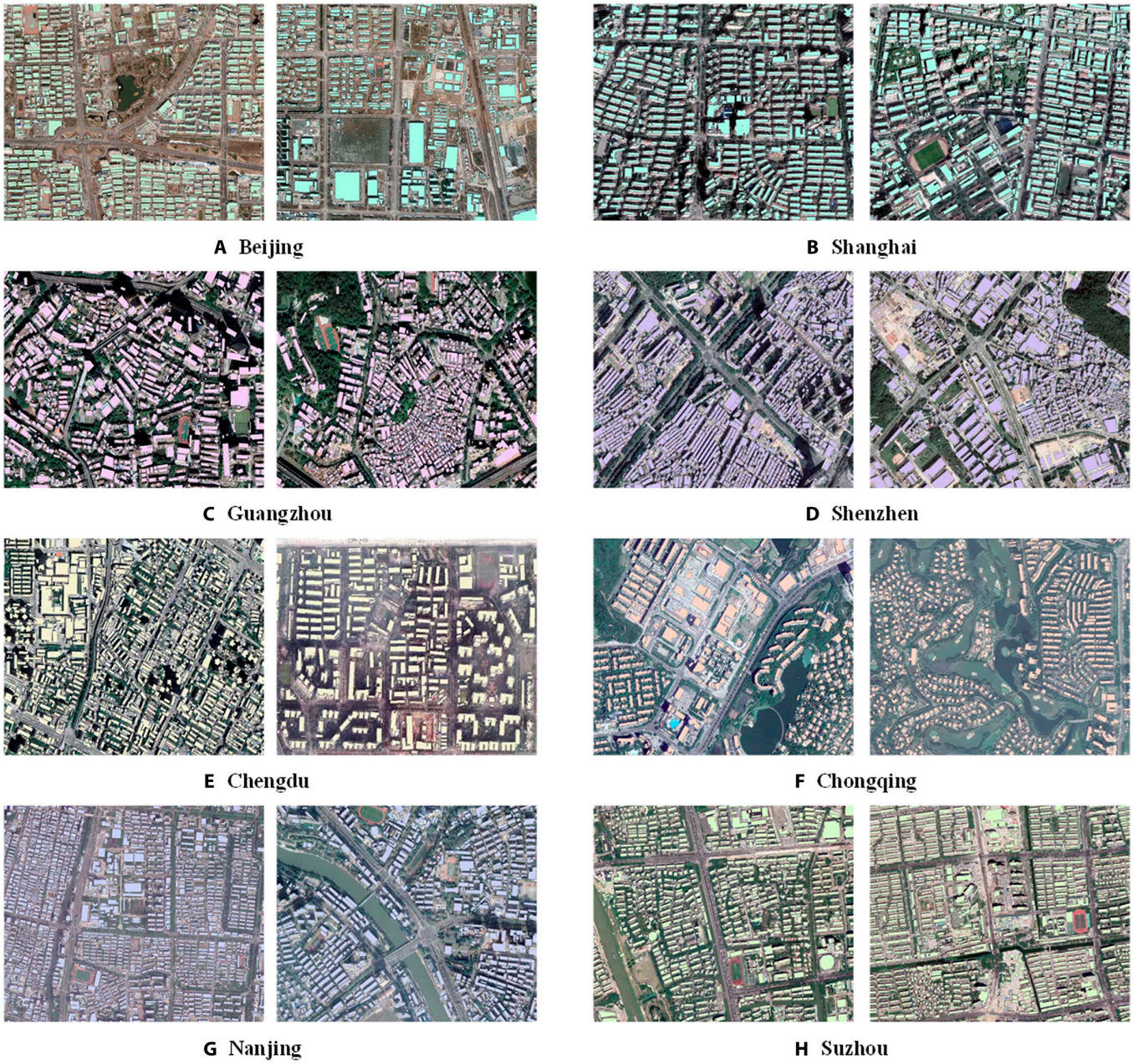

Fig. 13. The visual results in different cities. (A) Beijing. (B) Shanghai. (C) Guangzhou. (D) Shenzhen. (E) Chengdu. (F) Chongqing. (G) Nanjing. (H) Suzhou. Images are from © Google Earth 2021.

exist in the building characteristics across these different countries. In the case of North Korea and Mongolia, the number of buildings is comparatively lower, and their distribution appears more scattered. Our results demonstrate the accurate identification of buildings within these 2 countries. Conversely, South Korea and Japan exhibit a higher density of buildings, with a more orderly distribution, thereby yielding superior visual effects in the results obtained. Furthermore, our model showcases the ability to distinguish individual buildings within densely populated areas, capturing more intricate building data. This further validates the exceptional performance of our product.

Comparison with existing products

To effectively highlight the superior performance of our product, we conducted a comparative analysis with several existing products. Given the disparities among the compared products, separate comparisons were undertaken for China and Japan.

The description of the 2 existing products is as follows: (a) The availability of OSM building data on a global scale provides researchers with valuable information. Nevertheless, the quality of this data in the East Asian regions is considered to be comparatively lower than that of other countries. In order to showcase the superior quality of our building data specifically in East Asia, we conducted a comparative analysis with OSM product. (b) The building data produced in [19] contains 90 cities in China by using deep learning methods and high-resolution remote sensing imagery. By conducting a comparison with this dataset within the same regions, we aimed to evaluate the quality of our proposed building data.

Fig. 14. The visual results in zoom-in areas. Images are from © Google Earth 2021.

In China, the quality of OSM data is notably inadequate, leading to incomplete building results. In order to address this limitation, we included the building footprints data proposed in [19] encompassing 90 cities in China for comparison purposes. Since our study encompasses the entirety of China, while other products only cover specific regions, we selected common areas of coverage for comparison, as illustrated in Fig. 17. Evidently, our product exhibits the highest level of building completeness, aligning remarkably well with the corresponding imagery, particularly when contrasted with OSM data. Moreover, in terms of visual effects, our results outperform the others, boasting accurate geometric shapes that closely resemble real buildings. Conversely, the results of Zhang et al. are comparatively less refined. Furthermore, the results of Zhang et al. are chaotic and lack the ability to distinguish individual buildings in certain densely populated areas comprising small buildings. In contrast, our results accurately capture diverse building characteristics and successfully differentiate individual buildings. In conclusion, our product surpasses existing alternatives in China with respect to both data quality and visual effects. Importantly, we have provided comprehensive building data for the entire nation, which is a significant advancement compared to the deficient offerings of the existing products.

Figure 18 showcases a comparative analysis conducted in Japan, with blue areas representing the results obtained from our proposed product and yellow areas denoting the outcomes of OSM products. Notably, our product has significantly enhanced and supplemented the publicly available OSM data, resulting in a more precise and comprehensive dataset. Specifically, the original OSM results displayed notable omissions in certain regions, whereas our product rectifies these deficiencies. Moreover, the extracted building shapes derived from our model exhibit a remarkable alignment with the geometric characteristics of actual buildings, thereby showcasing exceptional visualization effects. Figure 18 also demonstrates the stability of our model in accurately mapping buildings of diverse forms and sizes. The selected imagery depicts a high building density, distinct roof color variations, and a variety of building sizes and characteristics. Our predicted results closely correspond to buildings in imagery, providing further substantiation of the outstanding performance delivered by our product.

Discussion

Description and analysis of statistical information

To further elucidate the value of our product and showcase its exceptional performance, kernel density analysis was conducted to approximate the distribution of urban centers. By taking China as an illustrative case, we specifically selected 2 well-developed

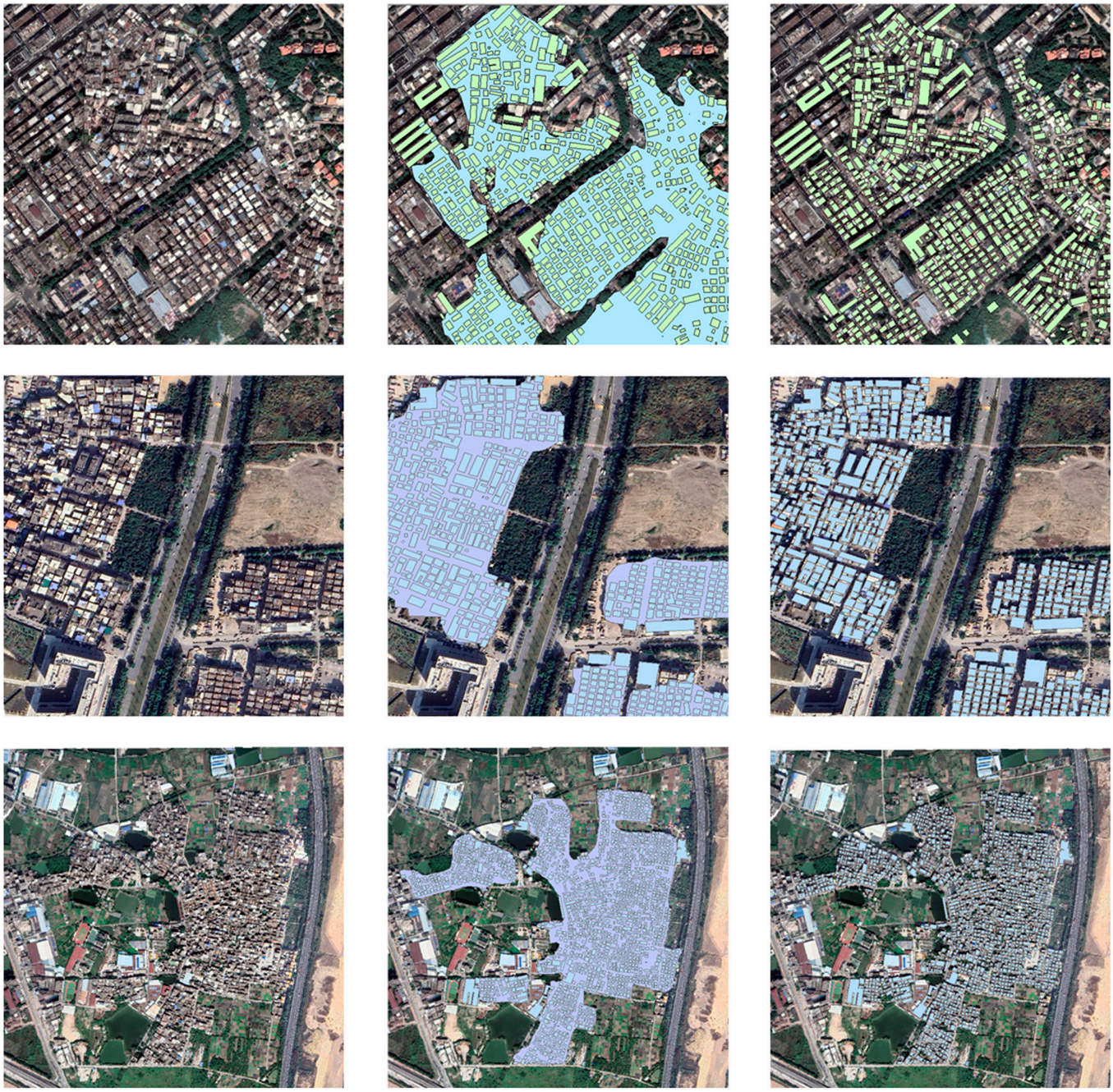

Fig. 15. The visual results in urban villages. (A) Images. (B) Results in urban village areas. (C) Results in urban village areas (the region of urban village is hidden). We intersected the urban village areas with our buildings data, thereby preserving and exhibiting the buildings exclusively within these areas while temporarily concealing data pertaining to external buildings. Images are from © Google Earth 2021.

cities, namely Guangzhou and Shanghai, and performed kernel density analysis on their respective building datasets, as displayed in Fig. 19. The results indicate the identification of hotspots within the delineated red box regions for both cities, which can serve as rough representations of the urban centers. Subsequent investigation reveals that these identified regions correspond to Zhujiang New Town in Guangzhou and Lujiazui in Shanghai, which are recognized as the CBDs of the respective cities. This correlation further substantiates the capability of our building product in facilitating urban center analysis and providing essential insights into the spatial layout of urban areas.

To further validate the accuracy of the building data presented in this paper, we conducted cross-verification and analysis by comparing it with population data. Following the methodology proposed in [14], we calculated the correlation using Landscan population data. To illustrate this, we selected 2 cities in China for analysis. The approach involved dividing the study area into grid cells and categorizing them into 4 types based on the buildings data in this paper and the Landscan population data.

1. Type I: With low building density and low population density

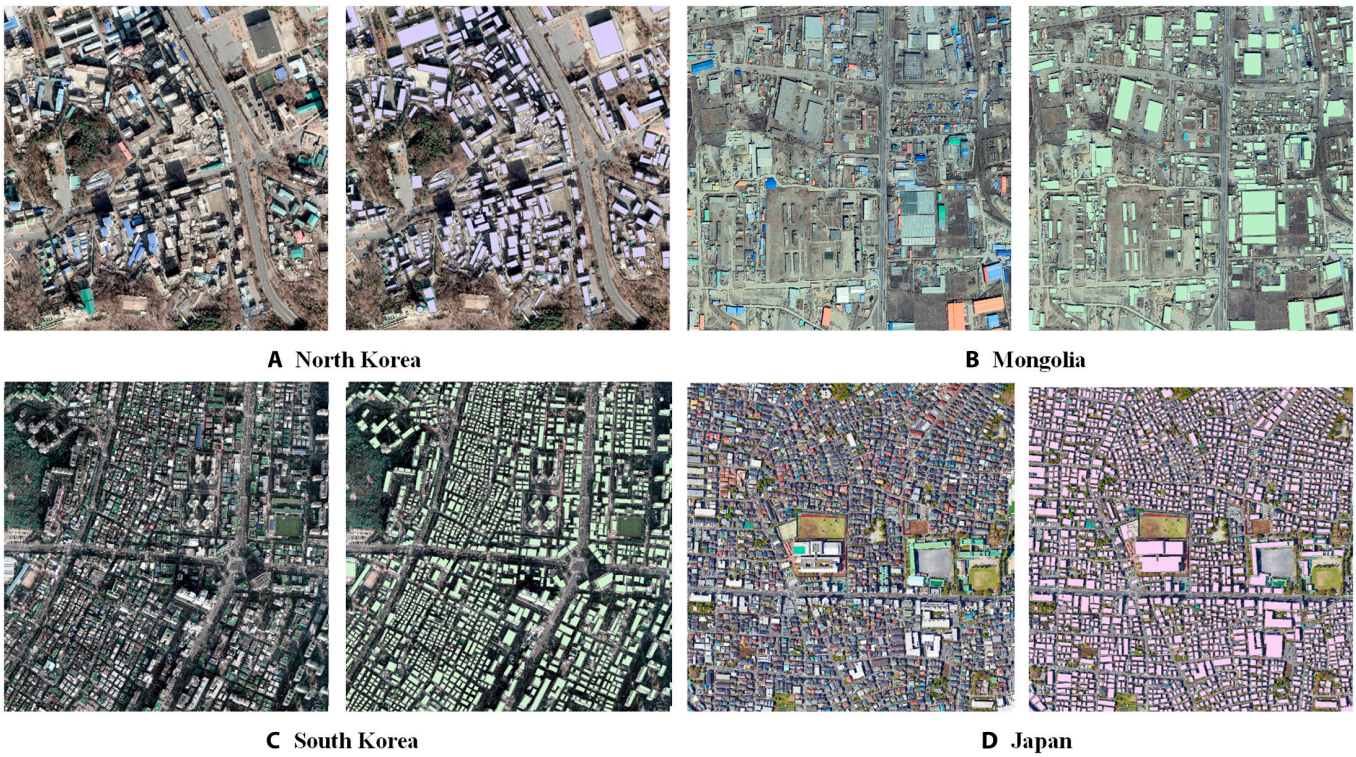

Fig. 16. The visual results in different cities. (A) North Korea. (B) Mongolia. (C) South Korea. (D) Japan. Images are from © Google Earth 2021.

2. Type II: With high building density and low population density
3. Type III: With low building density and high population density
4. Type IV: With high building density and high population density

Then, the relationship between buildings and population is calculated according to Eq. 6 to further verify the accuracy of building data:

$$C_{estimate} = \frac{N_{TypeIII} + N_{TypeIV}}{N_{TypeII} + N_{TypeIII} + N_{TypeIV}} \times 100\% \quad (6)$$

According to the calculated results, Shenzhen and Yinchuan were selected as exemplars, representing regions with well-developed population and economy and regions with average population and economy, respectively. The calculated results, as depicted in Fig. 20, demonstrate a strong correlation between the building data presented in this paper and the Landscan population data for both city types. This finding further validates the accuracy of our data.

We also conducted data analysis to determine the building coverage rates in various cities across China. Table 2 presents the top 10 cities with the highest building coverage rates in China. It is evident that cities characterized by a denser distribution of buildings and higher levels of development tend to exhibit higher building coverage rates. Notably, Dongguan gets the highest building coverage rate, reaching an impressive 13%. Additionally, the table highlights the prevalence of higher building coverage rates in cities within Guangdong Province, indicating significant building development in this region, likely influenced by its large population. The information provided by our product holds promise for facilitating research in urban scale analysis and offers valuable applications in this domain.

The application and significance of our data

The availability of buildings footprints data plays a crucial role in various studies by providing comprehensive and detailed buildings information, thereby enhancing the accuracy of related analyses. Our supplementary buildings footprints data specifically focuses on East Asia and serves as a pivotal resource for research in this region and beyond. It facilitates large-scale analyses and enables to obtain more comprehensive and detailed outcomes. For example, the direct utilization of buildings footprints data allows for the evaluation and analysis of RPVs, contributing to improved urban planning and energy management [20]. RPVs are crucial in achieving energy transition and climate goals, especially in cities with high building density and substantial energy consumption. Estimating RPV carbon mitigation potential at the city level of an entire large country is challenging given difficulties in assessing rooftop area. By providing our data directly to these relevant studies, we enable more accurate assessments of photovoltaic potential and carbon emission analysis, offering valuable insights for national planning. Moreover, buildings footprints data can be employed to predict future building development through geographic simulation, allowing for proactive urban planning and management. Analyzing building trends on a larger scale using our East Asia buildings footprints data enables the formulation of more precise plans. Additionally, our provided buildings footprints data is instrumental in the building-level function classification in East Asia [59]. Understanding the actual building functions is essential for many urban applications, such as city management, urban planning, and optimization of transportation systems. Leveraging our buildings footprints data, these related studies

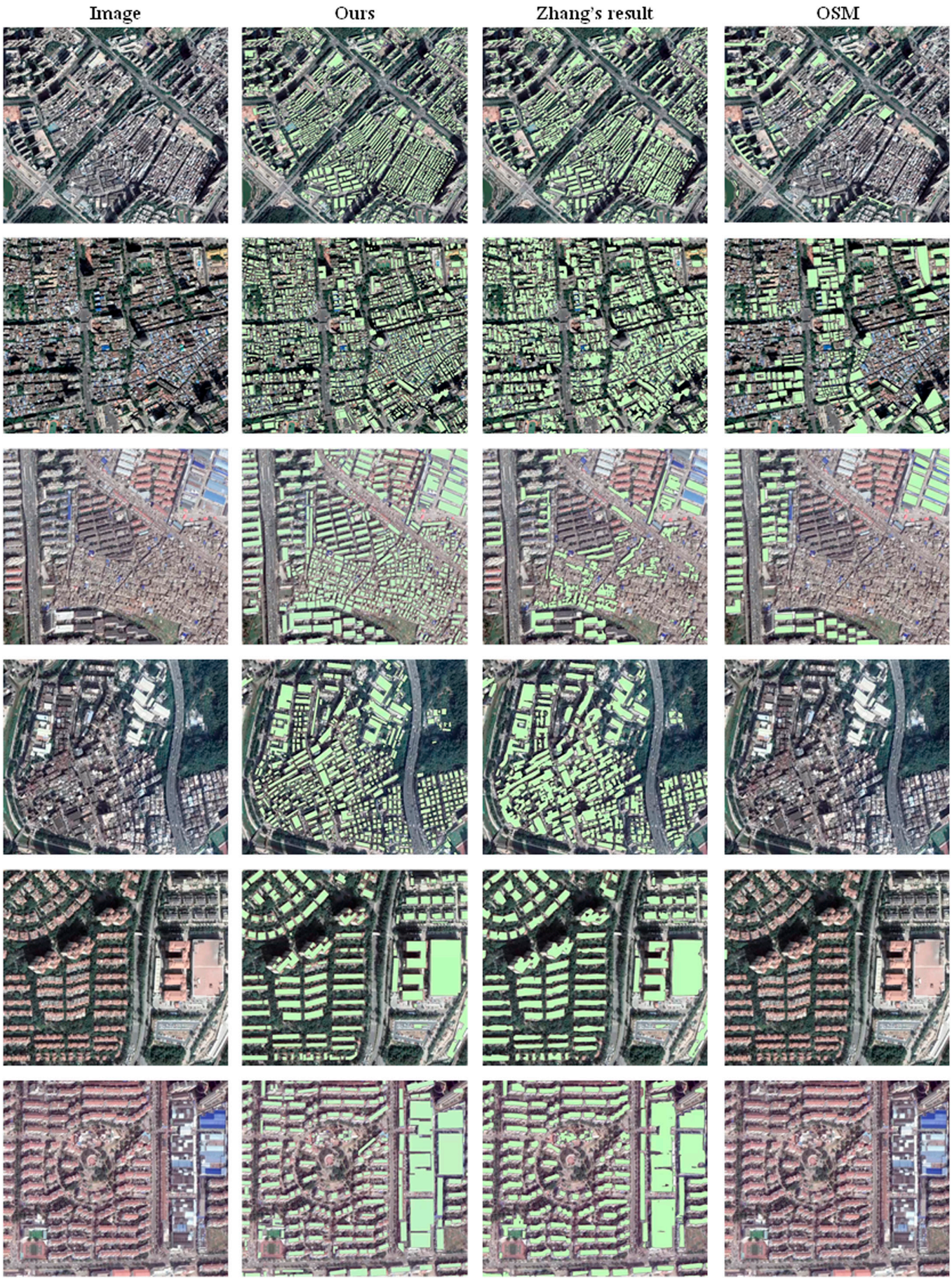

Fig. 17. The comparison results with different products in China. Images are from © Google Earth 2021.

Fig. 18. The comparison results with OSM in Japan. The blue blocks denote the results generated by our method, whereas the yellow ones represent the results derived from OSM. Images are from © Google Earth 2021.

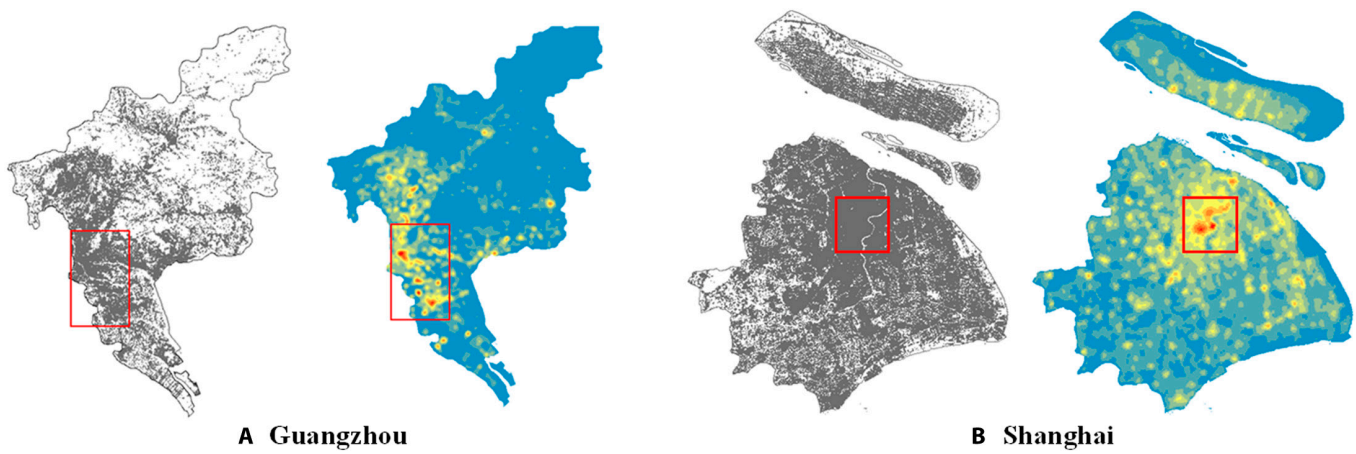

Fig. 19. The kernel density analysis results of different cities in China. (A) Guangzhou. (B) Shanghai.

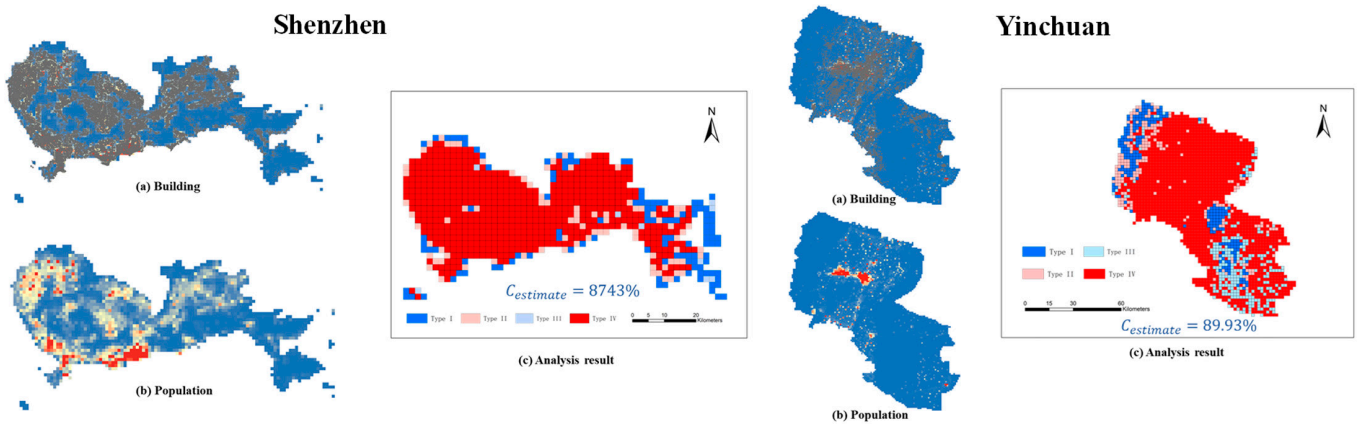

Fig. 20. The cross-validation results with Landscan population data. (A to C) We choose Shenzhen and Yinchuan as examples, representing the areas with more developed population and economy and the areas with ordinary population and economy respectively.

Table. 2. Statistical results of our buildings data in China. The results show the top 10 cities with the highest building coverage rates. The forward slash in the province column denotes a municipality city.

ID	City	Province	City area (km^2)	Building area (km^2)	Building coverage rate (%)
1	Dongguan	Guangdong	2,422.02	315.16	13.01
2	Macau	/	33.9	3.84	11.32
3	Shenzhen	Guangdong	1,920.12	201.83	10.51
4	Zhengzhou	Henan	7,523.89	736.54	9.79
5	Foshan	Guangdong	3,883.37	363.49	9.36
6	Shanghai	/	7,049.01	643.84	9.13
7	Jiaxing	Zhejiang	4,062.56	367.09	9.04
8	Zhongshan	Guangdong	1,755.81	151.32	8.62
9	Langfang	Hebei	6,429.87	533.55	8.30
10	Changzhou	Jiangsu	4,458.50	326.63	7.33

can undertake comprehensive, large-scale analyses, facilitating macro-level management of cities. In summary, the supplement of buildings footprints data in East Asia empowers related research to conduct extensive investigations in this region, which, as one of the world's most significant economic zones, contributes greatly to global research.

Limitations and future work

Limitations such as cloud cover and inadequate temporal resolution can impact the imagery used in deep-learning-based building extraction, leading to discontinuities and impaired quality. As a result, obtaining precise building products for specific periods remains challenging. The current findings are based on a particular timeframe, and there is room for improvement in terms of accuracy. Furthermore, the significant variability in imagery and building characteristics across different regions makes it arduous to achieve highly generalized results using a single model. While this study has enhanced the generalization capability of our model through strategies like adaptive fine-tuning, limitations still exist in current deep learning methods. The semisupervised fine-tuning approach has introduced considerable uncertainties, making it challenging to guarantee substantial improvements across all regions. For future work, our focus will be on developing novel and reliable methods for large-scale mapping by harnessing advancements in deep learning technology and leveraging the availability of high-quality remote sensing data. This includes exploring more stable and versatile approaches, such as incorporating universal big models like Segment Anything [60]

Building data plays a critical role in various urban-related studies. However, single-temporal building products may not fully satisfy the requirements of certain research applications. As the need for long-term building data is expected to rise in the future, our forthcoming work will involve integrating temporal inference methods like change detection and sequential interpolation with extensive time series data. This integration aims to analyze the construction years of buildings comprehensively. By leveraging the outcomes of this analysis, we will generate long-term building products that offer more precise, detailed, and superior-quality data for urban research.

In addition, the functions and classifications of buildings are crucial considerations in numerous relevant studies. However, the building products presented in this paper lack information pertaining to building categories. To further explore the potential application of building data, we will conduct a comprehensive analysis by integrating diverse sources and modes of data in future work. By employing cutting-edge methods, we aim to ascertain the functions of buildings, thereby facilitating additional research in this domain.

Acknowledgments

We thank H. Li for editing.
Funding: This study was supported in part by the National Key R&D Program of China under Grant 2022YFB3903402, in part by the National Natural Science Foundation of China under Grant 42222106, in part by the National Natural Science Foundation of China under Grant 61976234, and in part by the Fundamental Research Funds for the Central Universities, Sun Yat-sen University under Grant 22lgqb12.
Author contributions: Q.S.: Conceptualization, resources, investigation, funding acquisition and writing—review and editing. J.Z.: Methodology, validation, and writing—original draft preparation. Zhengyu L.: Data curation, formal analysis, and visualization. H.G.: Writing—review and editing. S.G.: Writing—review and editing. M.L.: Writing—review and editing. Zihong L.: Writing—review and editing. X.L.: Writing—review and editing.
Competing interests: The authors declare that they have no competing interests. Q.S. is an editorial board member for *Journal of Remote Sensing* and was not involved in the editorial review, or the decision to publish, this article. All authors declare that there are no other competing interests.

Data Availability

The buildings data generated in this paper can be downloaded at https://doi.org/10.5281/zenodo.8174931.

References

1. Johari F, Shadram F, Widén J. Urban building energy modeling from geo-referenced energy performance certificate data. Development, calibration, and validation. *Sustain Cities Soc*. 2023;96:Article 104664.
2. Nouvel R, Zirak M, Coors V, Eicker U. The influence of data quality on urban heating demand modeling using 3D city models. *Comput Environ Urban Syst*. 2017;64:68–80.
3. Chen Y, Tang L, Yang X, Bilal M, Li Q. Object-based multi-modal convolution neural networks for building extraction using panchromatic and multispectral imagery. *Neurocomputing*. 2020;386:136–146.
4. Zhao W, Bo Y, Chen J, Tiede D, Blaschke T, Emery WJ. Exploring semantic elements for urban scene recognition: Deep integration of high-resolution imagery and OpenStreetMap (OSM). *ISPRS J Photogramm Remote Sens*. 2019;151:237–250.
5. Biljecki F, Ohori KA, Ledoux H, Peters R, Stoter J. Population estimation using a 3D city model: A multi-scale country-wide study in the Netherlandsm. *PLoS One*. 2016;11(6):Article e0156808.
6. Acuto M, Parnell S, Seto KC. Building a global urban science. *Nat Sustain*. 2018;1(1):2–4.
7. Hu Q, Zhen L, Mao Y, Zhou X, Zhou G. Automated building extraction using satellite remote sensing imagery. *Autom Constr*. 2021;123:Article 103509.
8. Assouline D, Mohajeri N, Scartezzini JL. Quantifying rooftop photovoltaic solar energy potential: A machine learning approach. *Sol Energy*. 2017;141:278–296.
9. Assouline D, Mohajeri N, Scartezzini JL. Large-scale rooftop solar photovoltaic technical potential estimation using random forests. *Appl Energy*. 2018;217:189–211.
10. Dehwah AHA, Asif M. Assessment of net energy contribution to buildings by rooftop photovoltaic systems in hot-humid climates. *Renew Energy*. 2019;131:1288–1299.
11. Kazmi H, Fu C, Miller C. Ten questions concerning data-driven modelling and forecasting of operational energy demand at building and urban scale. *Build Environ*. 2023;239:Article 110407.
12. Yang Y, Gu Q, Wei H, Liu H, Wang W, Wei S. Transforming and validating urban microclimate data with multi-sourced microclimate datasets for building energy modelling at urban scale. *Energy Build*. 2023;295:113318.
13. Xu X, Ou J, Liu P, Liu X, Zhang H. Investigating the impacts of three-dimensional spatial structures on CO_2 emissions at the urban scale. *Sci Total Environ*. 2021;762:Article 143096.
14. Zhou Q, Zhang Y, Chang K, Brovelli MA. Assessing OSM building completeness for almost 13,000 cities globally. *Int J Digit Earth*. 2022;15(1):2400–2421.
15. Zhang Y, Zhou Q, Brovelli MA, Li W. Assessing OSM building completeness using population data. *Int J Geogr Inf Sci*. 2022;36(7):1443–1466.
16. Zhou Q. Exploring the relationship between density and completeness of urban building data in OpenStreetMap for quality estimation. *Int J Geogr Inf Sci*. 2018;32(2):257–281.
17. Vargas-Muñoz JE, Lobry S, Falcão AX, Tuia D. Correcting rural building annotations in OpenStreetMap using convolutional neural networks. *ISPRS J Photogramm Remote Sens*. 2019;147:283–293.
18. Esch T, Brzoska E, Dech S, Leutner B, Palacios-Lopez D, Metz-Marconcini A, Marconcini M, Roth A, Zeidler J. World settlement footprint 3D-A first three-dimensional survey of the global building stock. *Remote Sens Environ*. 2022;270:Article 112877.
19. Zhang Z, Qian Z, Zhong T, Chen M, Zhang K, Yang Y, Zhu R, Zhang F, Zhang H, Zhou F, et al. Vectorized rooftop area data for 90 cities in China. *Sci Data*. 2022;9(1):66.
20. Zhang Z, Chen M, Zhong T, Zhu R, Qian Z, Zhang F, Yang Y, Zhang K, Santi P, Wang K, et al. Carbon mitigation potential afforded by rooftop photovoltaic in China. *Nat Commun*. 2023;14(1):2347.
21. Maggiori E, Tarabalka Y, Charpiat G, Alliez P. Semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. Paper presented at: IEEE International Geoscience and Remote Sensing Symposium (IGARSS); 2017 Jul 23–28; Fort Worth, TX.
22. Mnih V. Machine learning for aerial image labeling [thesis]. [Toronto (Canada)]: University of Toronto; 2013.
23. Ji S, Wei S, Lu M. Fully convolutional networks for multisource building extraction from an open aerial and satellite imagery data set. *IEEE Trans Geosci Remote Sens*. 2018;57(1):574–586.
24. Demir I, Koperski K, Lindenbaum D, Pang G, Huang J, Basu S, Hughes F, Tuia D, Raskar RD. A challenge to parse the earth through satellite images. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); 2018 Jun 18–22; Salt Lake City, UT.
25. Bradbury K, Brigman B, Collins L, Johnson T, Lin S, Newell R, Park S, Suresh S, Wiesner S, Xi Y. Aerial imagery object identification dataset for building and road detection, and building height estimation. figshare. 2016.
26. Guo H, Shi Q, Marinoni A, Du B, Zhang L. Deep building footprint update network: A semi-supervised method for updating existing building footprint from bi-temporal remote sensing images. *Remote Sens Environ*. 2021;264:Article 112589.
27. Wei S, Zhang T, Ji S. A concentric loop convolutional neural network for manual delineation-level building boundary segmentation from remote-sensing images. *IEEE Trans Geosci Remote Sens*. 2021;60:3126704.
28. Zhou D, Wang G, He G, Yin R, Long T, Zhang Z, Chen S, Luo B. A large-scale mapping scheme for urban building from Gaofen-2 images using deep learning and hierarchical approach. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2021;14:11530–11545.
29. Girard N, Smirnov D, Solomon J, Tarabalka Y. Polygonal building extraction by frame field learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2021 Jun 20–25; Nashville, TN.
30. Jiang X, Zhang X, Xin Q, Xi X, Zhang P. Arbitrary-shaped building boundary-aware detection with pixel aggregation network. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2020;14:2699–2710.
31. Guo H, Du B, Zhang L, Su X. A coarse-to-fine boundary refinement network for building footprint extraction from remote sensing imagery. *ISPRS J Photogramm Remote Sens*. 2022;183:240–252.
32. Liu Z, Shi Q, Ou J. LCS: A collaborative optimization framework of vector extraction and semantic segmentation for building extraction. *IEEE Trans Geosci Remote Sens*. 2022;60:3215852.
33. Lin H, Hao M, Luo W, Yu H, Zheng N. BEARNet: A novel buildings edge-aware refined network for building extraction from high-resolution remote sensing images. *IEEE Geosci Remote Sens Lett*. 2023;20:3272353.

34. Chen S, Shi W, Zhou M, Zhang M, Xuan Z. CGSANet: A contour-guided and local structure-aware encoder–decoder network for accurate building extraction from very high-resolution remote sensing imagery. *IEEE J Sel Top Appl Earth Obs Remote Sens*. 2021;15:1526–1542.
35. Yu T, Wan H, Tang P, Sheng L. Building footprint extraction model based on deep supervision and post-processing technology. Paper presented at: 2022 3rd International Conference on Geology, Mapping and Remote Sensing (ICGMRS); 2022 Apr 22–24; Zhousan, China.
36. Jung H, Choi H-S, Kang M. Boundary enhancement semantic segmentation for building extraction from remote sensed image. *IEEE Trans Geosci Remote Sens*. 2021;60:3108781.
37. Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative adversarial networks: An overview. *IEEE Signal Process Mag*. 2018;35(1):53–65.
38. Ding L, Tang H, Liu Y, Shi Y, Zhu XX, Bruzzone L. Adversarial shape learning for building extraction in VHR remote sensing images. *IEEE Trans Image Process*. 2021;31:678–690.
39. Zorzi S, Fraundorfer F. Regularization of building boundaries in satellite images using adversarial and regularized losses. Paper presented at: IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium; 2019 Jul 28–02 Aug; Yokohama, Japan.
40. Zorzi S, Bittner K, Fraundorfer F. Machine-learned regularization and polygonization of building segmentation masks. Paper presented at: 2020 25th International Conference on Pattern Recognition (ICPR); 2021 Jan 10–15; Milan, Italy.
41. Gribov A. Optimal compression of a polyline while aligning to preferred directions. Paper presented at: 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW); 2019 Sep 22–25; Sydney, NSW, Australia.
42. Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. arXiv. 2015. https://doi.org/10.48550/arXiv.1503.02531
43. Diakite AA, Zlatanova S. Automatic geo-referencing of BIM in GIS environments using building footprints. *Comput Environ Urban Syst*. 2020;80:Article 101453.
44. Durst N, Sullivan E, Huang H, Park H. Building footprint-derived landscape metrics for the identification of informal subdivisions and manufactured home communities: A pilot application in Hidalgo County, Texas. *Land Use Policy*. 2021;101:Article 105158.
45. Huang X, Wang C. Estimates of exposure to the 100-year floods in the conterminous United States using national building footprints. *Int J Disaster Risk Reduct*. 2020;50:Article 101731.
46. Yu W, Ai T, Shao S. The analysis and delimitation of central business district using network kernel density estimation. *J Transp Geogr*. 2015;45:32–47.
47. Yang Z, Chen Y, Guo G, Zheng Z, Wu Z. Using nighttime light data to identify the structure of polycentric cities and evaluate urban centers. *Sci Total Environ*. 2021;780:Article 146586.
48. Li X, Gong P, Zhou Y, Wang J, Bai Y, Chen B, Hu T, Xiao Y, Xu B, Yang J, et al. Mapping global urban boundaries from the global artificial impervious area (GAIA) data. *Environ Res Lett*. 2020;15(9):Article 094044.
49. Gong P, Li X, Wang J, Bai Y, Chen B, Hu T, Liu X, Xu B, Yang J, Zhang W, et al. Annual maps of global artificial impervious area (GAIA) between 1985 and 2018. *Remote Sens Environ*. 2020;236:Article 111510.
50. Zhu XX, Qiu C, Hu J, Shi Y, Wang Y, Schmitt M, Taubenböck H. The urban morphology on our planet–global perspectives from space. *Remote Sens Environ*. 2022;269:Article 112794.
51. Zhang H, Liao Y, Yang H, Yang G, Zhang L. A local–global dual-stream network for building extraction from very-high-resolution remote sensing images. *IEEE Trans Neural Netw Learn Syst*. 2020;33(3):1269–1283.
52. Guo H, Shi Q, Du B, Zhang L, Wang D, Ding H. Scene-driven multitask parallel attention network for building extraction in high-resolution remote sensing images. *IEEE Trans Geosci Remote Sens*. 2020;59(5):4287–4306.
53. Oktay O, Schlemper J, Le Folgoc L, Lee M, Heinrich M, Misawa M, Mori K, McDonagh S, Hammerla NY, Kainz B, et al. Attention u-net: Learning where to look for the pancreas. arXiv. 2018. https://doi.org/10.48550/arXiv.1804.03999
54. Li C, Liu Y, Yin H, Li Y, Guo Q, Zhang L, Du P. Attention residual U-Net for building segmentation in aerial images. Paper presented at: IEEE International Geoscience and Remote Sensing Symposium (IGARSS); 2021 Jul 11–16; Brussels, Belgium.
55. Ni J, Wu J, Tong J, Chen Z, Zhao J. GC-Net: Global context network for medical image segmentation. *Comput Methods Prog Biomed*. 2020;190:Article 105121.
56. Wang X, Girshick R, Gupta A, He K. Non-local neural networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018;7794–7803.
57. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Paper presented at: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2018 Jun 18–23; Salt Lake City, UT.
58. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Paper presented at: 2016 IEEE Conference Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27–30; Las Vegas, NV.
59. Zhang X, Liu X, Chen K, Guan F, Luo M, Huang H. Inferring building function: A novel geo-aware neural network supporting building-level function classification. *Sustain Cities Soc*. 2023;89:Article 104349.
60. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo W-Y, et al. Segment anything. arXiv. 2023. https://doi.org/10.48550/arXiv.2304.02643