Estimation of urban-scale photovoltaic potential: A deep learning-based approach for constructing three-dimensional building models from optical remote sensing imagery

Longxu Yan [a,b], Rui Zhu [c,*], Mei-Po Kwan [d], Wei Luo [e,f], De Wang [a], Shangwu Zhang [a,b], Man Sing Wong [g], Linlin You [h], Bisheng Yang [i], Biyu Chen [i], Ling Feng [c]

[a] College of Architecture and Urban Planning, Tongji University, Shanghai 200092, China
[b] Key Laboratory of Spatial Intelligent Planning Technology, Ministry of Natural Resources, Shanghai 200092, China
[c] Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, Singapore 138632, Republic of Singapore
[d] Department of Geography and Resource Management, The Chinese University of Hong Kong, Hong Kong, China
[e] GeospatialX Lab, Geography Department, National University of Singapore, 117570, Singapore
[f] Saw Swee Hock School of Public Health, National University of Singapore, 117549, Singapore
[g] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China
[h] School of Intelligent Systems Engineering, Sun Yat-Sen University, Shenzhen 518107, China
[i] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

ARTICLE INFO

ABSTRACT

Building-integrated photovoltaics are increasingly used to build low-carbon buildings and promote energy transition. However, the absence of three-dimensional (3D) building models may hinder accurate estimation of photovoltaic (PV) potential on 3D urban surfaces. This study develops a detail-oriented deep learning approach, which for the first time constructs 3D buildings from high-resolution satellite images and estimates PV potential. Specifically, two convolutional neural networks, i.e., the Rooftop Segmentation Model and Height Prediction Model, were developed by advancing the basic DeepLabv3+ architecture and integrating dedicated layers, adaptive activation functions, and hybrid losses. Next, the two models were trained and tested on a self-made dataset targeted at Shanghai and an open datasets under standard data augmentation and transfer learning strategies. Then, morphological post-processing procedures were developed to cluster and regularize individual rooftops with estimated heights. Finally, PV potentials in typical areas were estimated and compared. Accuracy assessments suggest satisfactory rooftop segmentation and building height estimation. The absolute relative error between the PV potentials derived from the actual and predicted building models showed little difference, implying the reliability of the extracted buildings. The proposed model is novel and effective for constructing 3D building models that can facilitate PV penetration and urban studies in various fields.

1. Introduction

1.1. Background and motivation

Because of the development of solar photovoltaic (PV) technology, solar PV framing has been considered an emerging way of promoting sustainable energy transition (Wong et al., 2016). Unlike massive solar farming plants, which generate great amounts of electricity for a national grid, building-integrated photovoltaics (BiPV) show promising feasibility for developing low-carbon solar cities (Zhu et al., 2022a). This relies on appropriate site selection to ensure that BiPVs are placed at solar-abundant locations to build an efficient PV system, which requires an accurate estimation of spatio-temporal solar potential distribution on three-dimensional (3D) urban surfaces (Zhu et al., 2022b). To achieve this, it is crucial to obtain precise and up-to-update 3D building models to estimate shadow effects from surrounding buildings, which are mainly produced by advanced surveying technologies, such as Light Detection and Ranging (LiDAR) (Hu et al., 2018; Yang et al., 2016).

However, the precise 3D city models are usually confidential because of policy or security issues and unavailable in many countries due to limited budgets. Also, such data can quickly become outdated for fast-growing cities. Although building footprint polygons obtained from open-sourced maps, such as OpenStreetMap (2022), are accurate in many global cities, they usually do not contain the building height attribute. Since urban morphology significantly affects spatiotemporal solar distribution on 3D urban surfaces (Zhu et al., 2020), estimating PV potential without considering shadow effects from 3D buildings may cause unforeseeable uncertainties (Zhang et al., 2022a; Zhong et al., 2021). Therefore, the unavailability of 3D building models may have impeded the widespread deployment of PV in the development of solar cities. Thus, this study aimed to develop a simple and effective model for building accurate 3D building models, fulfilling large-scale estimation of PV potential.

1.2. Rooftop segmentation models

Over the past few decades, precise and high-resolution surveying data, such as 3D point clouds and Digital Surface Model (DSM) derived from LiDAR data, have been used as major data sources for PV potential estimation (Aslani & Seipel, 2022). For example, one study utilized airborne laser scanning (ALS) data to extract tilted rooftops and estimate solar potential (Jochem, Höfle, Rutzinger & Pfeifer, 2009), and another one recognized rooftop areas with an 87% success rate by using relatively low-resolution (2 m) DSM data (Gooding, Crook & Tomlin, 2015). These datasets are fundamental for accurately mapping the solar potentials of rooftops. However, some cities cannot obtain such data covering the entire region with affordable and frequent updates.

To address the limitation of missing data, some studies have proposed new methods of extracting rooftop areas from high-resolution remote sensing images, either from commercial satellite image providers such as WorldView (Norman, Shafri, Idrees, Mansor & Yusuf, 2020) or publicly available platforms such as Google Earth Engine (Huang, Mendis & Xu, 2019a; Zhang et al., 2022b; Zhong et al., 2021). For example, one study used deep learning-based image segmentation to extract rooftop areas in 90 Chinese cities with an overall accuracy of more than 97%, which was extracted from the 18-level Google Earth Satellite (GES) images with a resolution of 0.6 m (Zhang et al., 2022b). This implies that sub-meter resolution GES images allow for an accurate extraction of rooftop areas, which shows its advantages in terms of considerably low acquisition cost, spatially broad coverage, and relatively frequent updates.

Over the past few years, conventional edge detection algorithms have been used to extract rooftop areas from satellite imagery; for example, object-based image analysis was applied to Quickbird satellite images (Kabir, Endlicher & Jägermeyr, 2010). These methods usually use low- or mid-level features to classify building and non-building objects; this classification is significantly affected by threshold adjustments and determined empirical rules (Huang, Zhang, Xin, Sun & Zhang, 2019b). With continuous urbanization along with expanded urban areas, diversified rooftops, and complicated urban landscapes, setting up appropriate thresholds and using suitable empirical rules has become, thus impeding the wide utilization of these methods. Alternatively, with the advancements in artificial intelligence and remote sensing technologies, rooftop areas can be extracted from high-resolution remote sensing images using deep learning models. For instance, one study incorporated rooftop features into a Support Vector Machine to categorize rooftops into six classes (Mohajeri et al., 2018).

Recently, many studies extracted rooftop areas based on the image semantic segmentation model (Huang et al., 2019a; Li et al., 2021; Zhang et al., 2022b; Zhong et al., 2021a, 2021b). Qian et al. (2022a) extracted rooftop areas in 90 Chinese cities from GES images using a deep learning-based semantic segmentation network (i.e., Deep-Labv3+), which achieved an overall accuracy and F1-score of 97.95% and 83.11%, respectively. Further, Qian et al. (2022b) proposed a

network based on DeepLabv3+ to achieve the more challenging task of segmenting rooftop structure lines from satellite imagery, wihch needs a high capability of sensitively and accurately detecting linear-object features. The network (i) utilized transfer learning to fine-tune parameters and designed hybrid loss functions to alleviate class imbalance issues and obtain refined delineation, (ii) upgraded the backbone and integrated the Dural-Attention Mechanism with Atrous Spatial Pyramid Pooling in Encoder, and (iii) adapted PointRend in Decoder to refine the line segmentation. These studies enlighten us on the utilization of DeepLabv3+ to refine rooftop area segmentations.

1.3. Building height prediction

Previous studies developed many methods to estimate building heights in various spatial scales. The most widely used method is to retrieve large-scale building heights from single Synthetic Aperture Radar (SAR) data (Sun et al., 2022b) or LiDAR data (Lao et al., 2021). This method can build accurate and large-scale 3D city models; however, some cities cannot obtain such data, especially those experiencing rapid urbanization but confronting a limited budget. In contrast, some other studies have estimated building heights by detecting shadows from satellite imagery and calculated building heights based on complex geometric modeling (Liasis & Stavrou, 2016; Qi, Zhai & Dang, 2016). However, it faces challenges in certain urban areas with a high density of tall buildings because (i) buildings adjacent to each other will not create shadows, (ii) some low-rise buildings are always in shadow during the daytime, and (iii) separating shadow polygons affiliated to specific building instances is difficult. For instance, some shadows projected onto building façades severely impede height estimation. Although the above-mentioned problem can be solved by detecting and measuring the heights of building façades captured in street view images (Bshouty, Shafir & Dalyot, 2020; Yan & Huang, 2022), this is only effective for buildings along streets that have been captured by cameras and where up-to-date street view images are not readily available.

Another approach aggregated building heights based on grid cells with various spatial resolutions, and examples of this approach include using AW3D30 DSM to produce 30-m resolution building height maps of China (Huang et al., 2022); developing the Random Forest machine learning model for building height prediction based on Landsat and SAR images at 1-km resolution in China, Europe, and the United States (US) (Li, Koks, Taubenböck & Van Vliet, 2020a); and utilizing Sentinel-1 SAR data at a 500-m resolution that covers a total area greater than 500 km^2 in seven cities in the US (Li, Zhou, Gong, Seto & Clinton, 2020b). These building height datasets are important for large-scale analysis related to urban morphology; however, they are difficult to be used for research on micro-scale investigations.

To achieve micro-scale 3D building construction, one study utilized ZY-3 multi-view images to estimate building heights at a spatial resolution of 2.5 m (Cao et al., 2021). Specifically, a multi-spectral, multi-view, and multi-task deep neural network was developed, which is a multi-task learning framework with the fusion of ZY-3 multi-spectral and multi-view images that achieved competitive results in 42 Chinese cities compared to the Random Forest model. This novel study enlightens us on exploring the possibility of accurately estimating building heights from optical bands of single-view/orthographic satellite images by learning the relationships between building heights and features.

1.4. solar pv estimation

Solar potential has been estimated by mainly focusing on four progressive perspectives: (i) estimating land-surface solar irradiation by developing physical models (Feng & Wang, 2021) and machine learning models (Liao, Zhu & Wong, 2022); (ii) developing physical models to estimate solar distribution on urban surfaces (Catita, Redweik, Pereira & Brito, 2014; Jakubiec & Reinhart, 2013; Zhu et al., 2020; Zhu, You, Santi, Wong & Ratti, 2019); (iii) investigating spatial and temporal
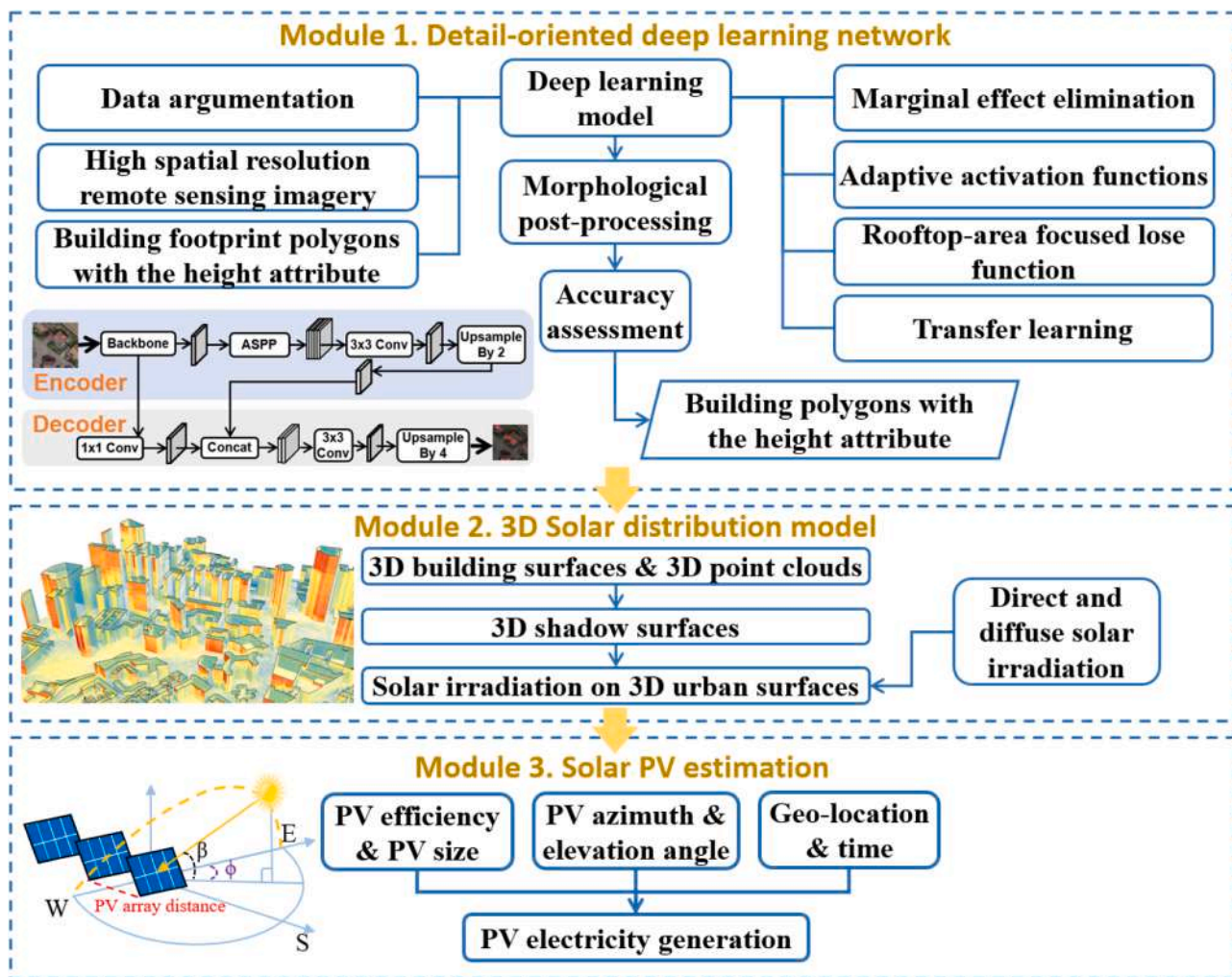
Fig. 1. A framework of 3D solar PV potential estimation based on the detail-oriented deep learning network.

distribution patterns of solar irradiation on urban surfaces (Catita et al., 2014; Jakubiec & Reinhart, 2013; Lindberg, Jonsson, Honjo & Wästberg, 2015; Lobaccaro, Carlucci, Croce, Paparella & Finocchiaro, 2017; Peronato, Rey & Andersen, 2018); and (iv) analysing and optimizing solar accessibility by designing new urban forms or hybrid systems (Bianchi, Branchini, Ferrari & Melino, 2014; Zhang et al., 2019; Zhu et al., 2019). To estimate solar distribution on 3D urban surfaces that supports complex building geometries (e.g. two buildings having concave rooftops are nested with each other) usually occurring in megacities and enable big data computation over a large area, a study developed an accurate solar estimation model at high spatiotemporal resolutions (Zhu et al., 2020). This suggested that urban solar potential is comprehensively affected by geo-location that determines time-dependent direct and diffuse solar irradiation, historical weather that affects land-surface solar irradiation qualified by land cover and urban morphology that creates shadow effects from surrounding buildings. Thus, the model developed by this study (Zhu et al., 2020) is appropriate for estimating PV potential on both actual and estimated building surfaces to evaluate the effectiveness of the proposed framework.

1.5. Contribution

In light of the limitations of previous methods, this study is innovative in that 3D urban surfaces were accurately, economically, easily, and rapidly constructed from two-dimensional (2D) satellite imagery. It

addresses the critical issue that the most updated 3D building dataset is unavailable in most cities. Our framework is feasible to construct 3D building datasets for global cities, which is particularly useful for a wide range of urban applications, such as estimating PV potential for urban solar farming.

The remaining sections of this paper are organized as follows. Section 2 develops a deep learning-based framework to segment building rooftop areas and predict building heights from satellite images. Section 3 introduces the model for PV estimation adapted for complex urban areas. Section 4 presents the study area, pre-processing, transfer learning, and the experiments and analyses for the rooftop area segmentation and building height prediction. Section 5 compares PV potential obtained from actual and estimated 3D urban surfaces. Finally, Section 6 presents the discussion and conclusion.

2. Deep learning-based 3D building model for PV potential estimation

2.1. Research framework

This study proposes a research framework for estimating solar PV potential on 3D urban surfaces in the following major steps (Fig. 1). First, a detail-oriented deep-learning network is proposed to accurately segment building rooftop areas and predict building heights from satellite imagery, resulting in the construction of 3D buildings. To achieve this, this study advances one of the state-of-the-art model architectures
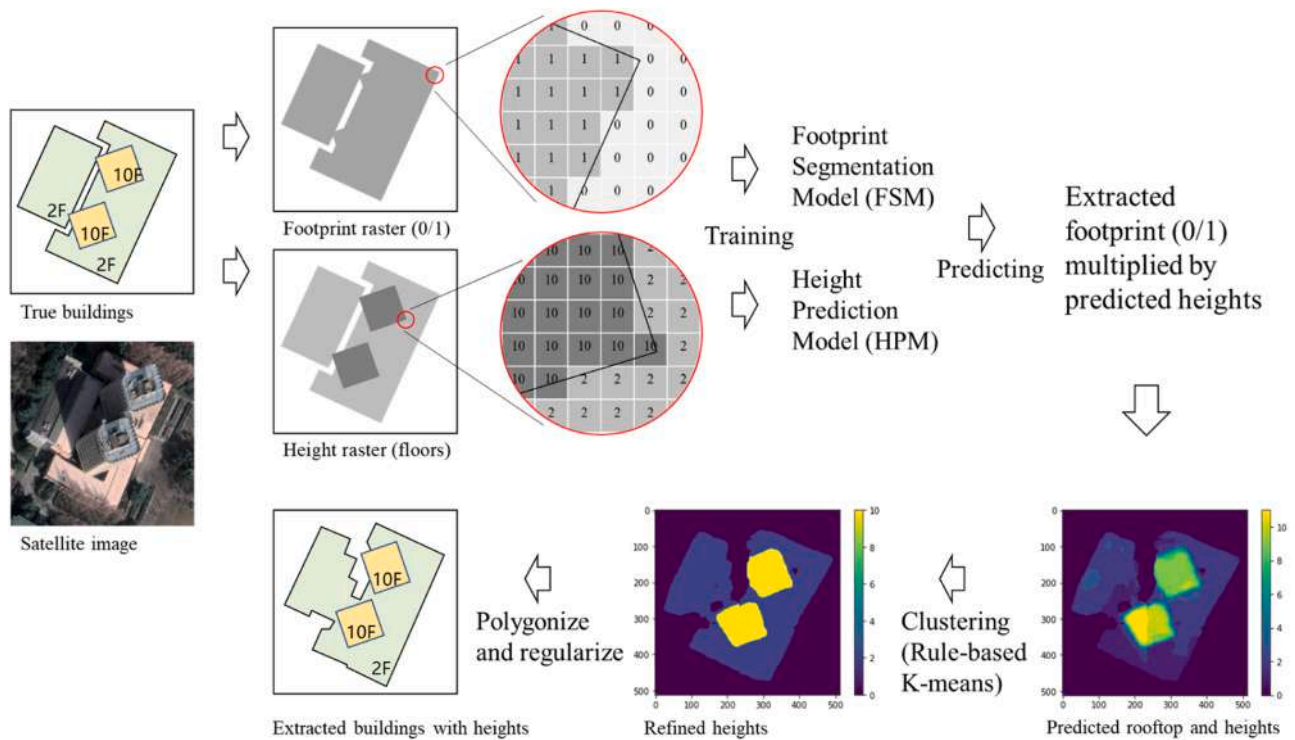
Fig. 2. Workflow of 3D building construction based on the detail-oriented deep learning network.

(i.e., the DeepLabv3+) by integrating adaptive activation functions, hybrid loss functions, transfer learning, morphological post-processing, etc. Second, a 3D solar distribution model is adapted into the framework, which estimates land surface solar irradiation (determined by direct and diffuse solar radiation) and quantifies the 3D shadow effects from surrounding buildings. Third, solar PV potential is modeled by incorporating crucial parameters, such as PV conversion efficiency, PV size, PV installation layouts, and the associated time and geo-location.

2.2. Overview of the detail-oriented deep learning network

To construct 3D buildings based on 2D satellite images, we first convert 2D rooftop polygons enriched with the height attribute to two types of raster images, one of which identified urban areas belonging to either rooftops or ground, while the other one presented the number of storeys in the buildings (Fig. 2). Next, two deep-learning networks, i.e., the Rooftop Segmentation Model (RSM) and Height Prediction Model (HPM), were developed and trained to perform rooftop segmentation and building height estimation, respectively. Furthermore, images of 3D buildings were obtained by multiplying binary rooftop images with building height images. Finally, rule-based K-means clustering was utilized to refine building heights, and boundary regularization was processed to construct regular 3D buildings.
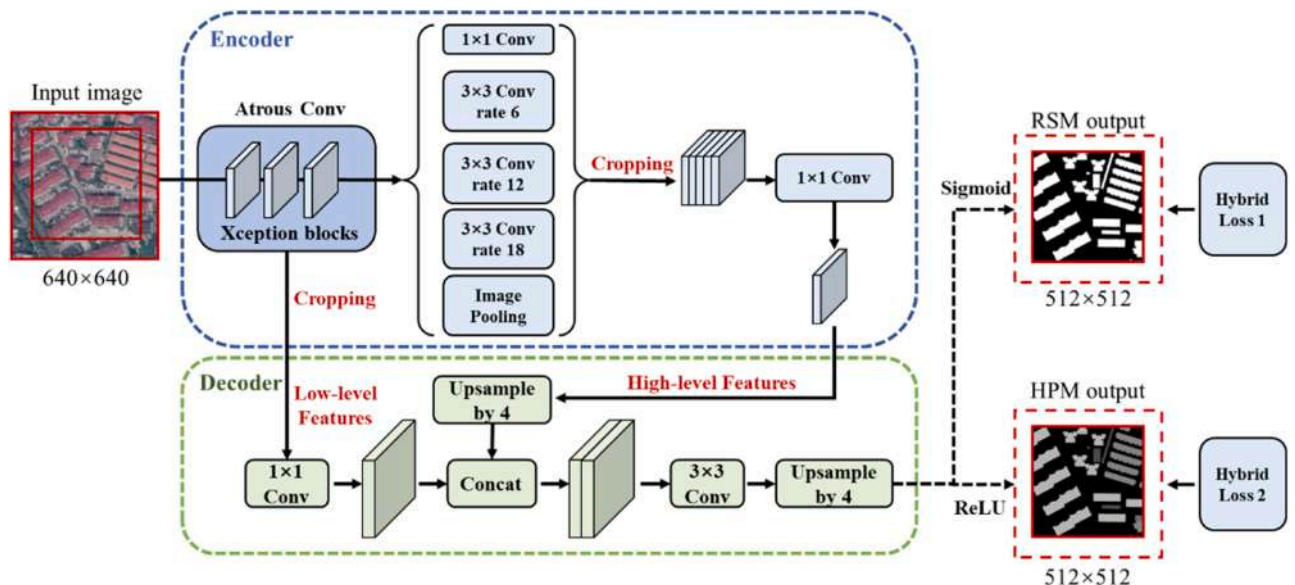

Fig. 3. Detail-oriented deep learning network eliminating marginal effects.
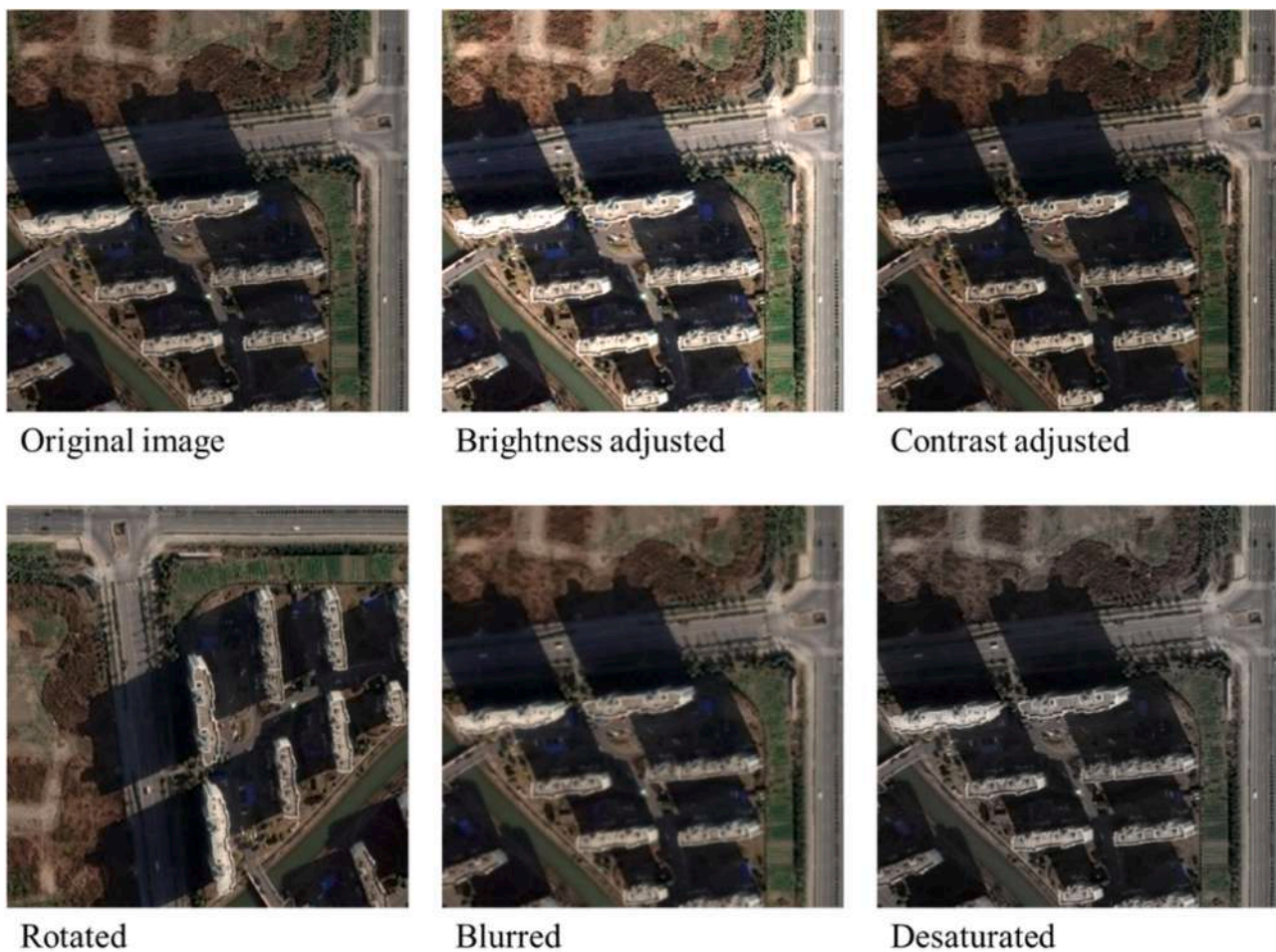
Fig. 4. One original image and five augmented images.

2.3. A refined deep learning network for constructing 3D buildings

2.3.1. Semantic image segmentation model

The RSM and HPM were developed using DeepLabv3+, a semantic segmentation model constructed by the Encoder-Decoder structure (Qian et al., 2022b). The encoding phase utilizes Deep Convolutional Neural Network and Atrous Spatial Pyramid Pooling (ASPP) to capture multi-scale features from the images, whereas the decoding phase up-samples the features obtained from the ASPP module by using bilinear interpolation to obtain an enlarged feature map. Next, the feature map concatenates with low-level features from the backbone for reconstructing the output at appropriate dimensions.

2.3.2. Deep learning network eliminating marginal effects

Building shadows are supposed to be crucial information for accurately segmenting rooftops and estimating building heights because a shadow is always associated with its rooftop and the building height and changes the brightness of the geo-objects in the shadow. However, shadows affiliated with buildings are often incomplete in discrete images, which may affect the accuracy of estimating building heights. To better account for shadow, an optimized training and estimation strategy has been proposed. Specifically, the input and output images are centralized at the same location with the input image having a larger extent. In this case, building shaodws in the marginal area of each image can also be considered during a series of convolution processes. To accomplish that, input images are extended outward for 64 pixels in each edge, resulting in an extension of the image size from 512×512 pixels to 640×640 pixels, which is named a patch (Fig. 3). Further, two cropping processes are adapted into the Encoder to align input and output. One is made during the skipping path from the Encoder to the Decoder, while the other is made right after the concatenation of ASPP. This ensures that the low- and high-level features have the same spatial extent.

In a hypothetical scenario, it is possible to detect rooftops and predict building heights based on a single deep learning model when the training datasets are pairs of remote sensing images and building footprint polygons with the height attribute. However, our preliminary study has confirmed that this approach will obtain unsatisfactory rooftop segmentation even though the accuracy of height estimation will be relatively high. Therefore, it is necessary to perform rooftop segmentation and height prediction as two independent and consecutive tasks. To achieve this, three crucial improvements are introduced: adaptive activation functions, hybrid loss functions to strengthen attention on rooftop areas, and transfer learning.

2.3.3. Adaptive activation functions

Adaptive activation functions are selected to predict building rooftop areas and heights. For RSM, the Sigmoid function is used to convert real values into a number between 0 and 1, which is interpreted as a probability (Zhou & Vosselman, 2012); this enables binary classification that indicates either non-building or building pixels when the probability is lesser or greater than 0.5, respectively. For HPM, the Rectified Linear Unit function is used to predict the number of storeys at each pixel (Zhang, Li, Li & Liu, 2018), which is a piecewise linear function where input (float values) is directly replicated as output (integer values) when it is positive. Otherwise, it will output zero.

2.3.4. Hybrid loss functions

To properly train RSM and HPM, two hybrid loss functions are constructed. For RSM (a binary segmentation task), the loss is designed by combining the commonly used Binary Cross Entropy (*BCE*) and the Dice coefficient loss (*Dice*) Eqs. (1)-(3). The Dice coefficient loss (Li, He, Li & Shen, 2022; Qian et al., 2022b), sometimes classified as one of the region-based losses, focuses more on the overlap between the true and estimated buildings, which is critical for boundary detection tasks. More specifically, the Hybrid Loss 1 (HL1) for RSM is:

$$HL1 = 1 - Dice + BCE \qquad (1)$$

$$Dice = \frac{2\sum_{i=1}^{N} y_i \widehat{y_i} + smooth}{\sum_{i=1}^{N} y_i^2 + \sum_{i=1}^{N} \widehat{y_i}^2 + smooth} \qquad (2)$$

$$BCE = -\frac{1}{N} \sum_{i=1}^{N} y_i \log(\widehat{y_i}) + (1 - y_i)\log(1 - \widehat{y_i}) \qquad (3)$$

where $\widehat{y_i} \in \{0, 1\}$ and $y_i \in [0, 1]$ are the truths and probabilities of a pixel being buildings, the *smooth* is simply a constant set to 1.0 to avoid division by zero.

For HPM, the objective is to estimate the number of storeys on each pixel, that is, the outputs are continuous numbers rather than probabilities. Thus, the constructed loss is essentially the mean squared error with more weights attached to the building pixels so that the model focuses more on height estimation rather than on building boundary detection Eqs. (4)-(6). The Hybrid Loss 2 (HL2) for HPM is:

$$HL2 = 0.9SE_{building} + 0.1SE_{non-building} \qquad (4)$$

$$SE_{building} = \frac{1}{N} \sum_{i=1}^{N} I(h_i > 0)(h_i - \widehat{h_i})^2 \qquad (5)$$

$$SE_{non-building} = \frac{1}{N} \sum_{i=1}^{N} I(h_i = 0)(h_i - \widehat{h_i})^2 \qquad (6)$$

where h_i and $\widehat{h_i}$ are the true and predicted building heights (number of storeys), respectively; $SE_{building}$ and $SE_{non-building}$ refer to the mean squared errors of the building and nonbuilding pixels, respectively; $I(\cdot)$ is the indicator function and {0.9, 0.1} are the weights attached to the building and non-building pixels, respectively.

2.3.5. Transfer learning

To improve prediction accuracy and training efficiency, a transfer learning strategy has been designed with four hierarchical training steps. First, the DeepLabv3+ model is pre-trained based on two datasets, Visual Object Classes Challenge (2022) and Cityscapes Dataset (2022), so that the model will obtain preliminary knowledge on geo-object segmentation. Next, the RSM is obtained by adjusting the input/output size and adding necessary layers to the pre-trained Deep-Labv3+ model; the RSM is trained on aerial images with labelled rooftops in several global cities, allowing it to quickly obtain a general ability to segment rooftops. Further, to segment rooftops in a city with a unique urban landscape and rooftop patterns, the RSM is trained on the footprint training dataset of the city. Moreover, the HPM is obtained by adjusting the activation function of the last layer of the trained RSM and re-trained on the training dataset of the city that contains building height information. The TensorFlow and keras framework is implemented in all training and testing procedures.

2.4. Data augmentation

Satellite imagery is usually organized as a set of spatially contiguous tiles with the same size if the image data size is significantly large. To effectively construct a training dataset, several data augmentation methods are proposed in this study (Fig. 4). First, a patch (640 × 640 pixels) is selected from a tile (1600 × 1600 pixels) at a random location so that the training data is no longer at fixed blocks. Second, each patch has a 5% probability of being adjusted on (i) contrast using Gamma Correction (γ ranging between ±5) and brightness (ranging between ±10%), and (ii) saturation with 0–50% random change, where 0 means completely black and white. Third, each patch has a 10% probability of being adjusted by (i) random rotation at 90°, 180°, or 270°, and (ii) performing the Gaussian blur function with a 3 × 3 or 5 × 5 convolution size.

2.5. Morphological post-processing

Rooftop polygons converted from estimated rooftop pixels may show irregular serration along their boundaries, and some polygons with significantly small areas or low heights (e.g., lower than one meter) are probably noises. Height estimations may also fluctuate around the true values. For example, the real building in Fig. 2 comprises two parts (2 and 10 storeys), but the prediction demonstrates random fluctuations between 2 and 10. To solve this problem, this study proposes morphological post-processing to refine building geometries. First, a building height raster is generated by multiplying the binary output of the RSM and the corresponding height prediction from the HPM. Second, the building heights are further refined using the rule-based K-means clustering algorithm such that the building instances can be elaborately separated despite the unavailability of instance segregation models. In addition, final building polygons with refined heights are generated using the polygonisation and regularization tools in ArcGIS Pro. Specifically, the rule-based K-means clustering algorithm is as follows:

1) Extract the pixels of a group of adjacent buildings with rounded heights denoted as {hi}.
2) Get the unique height values of {hi} denoted as {Hm}.
3) Calculate the proportion of pixel numbers (equivalent to footprint area) and the proportion of the summation of the pixel numbers (equivalent to the floor area) of each unique height Hm, denoted as $\{P_m^{footprint}\}$ and $\{P_m^{floor}\}$, respectively.
4) Count the number of unique height values that have their $P_m^{footprint}$ or P_m^{floor} exceeding Ω, denoted as Kn. After several tests, we found that Ω=20% was sufficient to identify significant height values and omit other fluctuating heights produced by the model.
5) Run K-means clustering algorithm on {hi} with K set to Kn.
6) Assign the cluster centres' height values to all the pixels in the same cluster and then get the refined heights.

2.6. Accuracy assessment

In this study, we constructed a confusion matrix [TP, TN, FP, FN] that presents the numbers of true-positive, true-negative, false-positive, and false-negative pixels, respectively. Specifically, TP means that the pixels belong to the footprint area in the real world and the prediction is correct, and FN means that the pixels belong to the footprint area but the prediction is wrong. In contrast, FP denotes that the pixels are not in the rooftop area and the prediction is wrong, while TN denotes that the pixels are not in the rooftop area and the prediction is correct. Next, the performance of the proposed model was systematically evaluated using a set of indices, *mIoU* (Eq. (7)), *precision* (Eq. (8)), *recall* (Eq. (9)), *accuracy* (Eq. (10)) and *F1-score* (Eq. (11)), which are commonly used in classification and image segregation problems. Meanwhile, the mean absolute error (*MAE*) and the mean percentage error (*MPE*) are calculated (Eqs. (12)-(13)) to measure the accuracy of the models in regression-like situations.

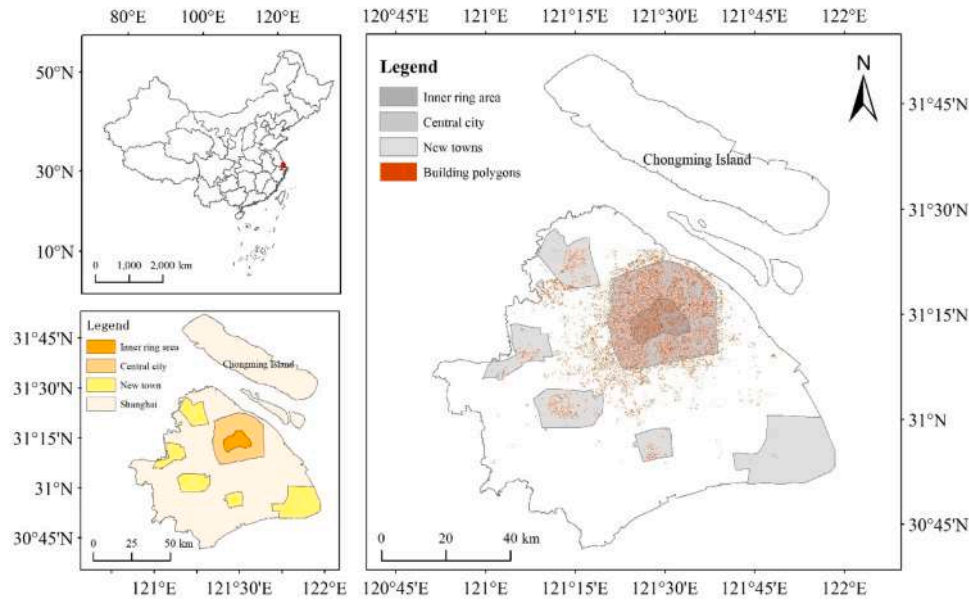$$mIoU = \frac{TP}{FP + FN + TP} \qquad (7)$$

Fig. 5. The study area in Shanghai, located in the eastern part of China.

$$precision = \frac{TP}{TP + FP} \tag{8}$$

$$recall = \frac{TP}{TP + FN} \tag{9}$$

$$accuracy = \frac{TP}{TP + FP} \tag{10}$$

$$F1 - score = \frac{2 \times precision \times recall}{precision + recall} \tag{11}$$

$$MAE = \frac{1}{N} \sum_{N} |h_i - \widehat{h_i}| \tag{12}$$

$$MPE = \frac{\sum Area_{predict} - \sum Area_{true}}{n \sum Area_{true}} \tag{13}$$

3. Estimation of PV potential on 3D urban surfaces

3.1. Land surface solar irradiation estimation

It is noteworthy that cities on the same latitude can have distinctly different climates and varying solar potential over time. To accurately quantify land surface solar irradiation associated with a specific time and location, cloud cover under real weather conditions needs to be considered, as it conclusively determines solar radiation affected by the atmosphere (Zhu et al., 2019). Thus, the average atmospheric transmittivity (*t*) and diffuse proportion (*d*) are calculated based on Eqs. (14)-(15) (Huang, Rich, Crabtree, Potter & Fu, 2008), using the statistical data on the proportions of sunny, partly cloudy and cloudy days, which are presented by p_{sunny}, $p_{partlycoludy}$, and p_{cloudy}, respectively.

$$t = 0.70 \times p_{sunny} + 0.50 \times p_{partlycloudy} + 0.30 \times p_{cloudy} \tag{14}$$

$$d = 0.20 \times p_{sunny} + 0.45 \times p_{partlycloudy} + 0.70 \times p_{cloudy} \tag{15}$$

3.2. Estimation of solar potential on 3D urban surfaces

A well-developed model has been used to estimate solar irradiation on 3D urban surfaces (Zhu et al., 2020). The model utilizes 2D footprints of a building having the height attribute to build a set of 3D polygons

denoted by **P**, which consists of horizontal rooftops **R**, vertical façades **F** and the ground **G**. On this basis, each polygon $p \in \textbf{P}$ can be discretized by spatial contiguous and homogenous grid cells such that a complete set of 3D centroids denoted by **O** can be obtained from the grid cells and used to present urban surfaces as 3D point clouds. In more detail, $\forall o \in \textbf{O}$, o contains several crucial elements, including the point ID i, the 3D co-ordinate c, and the accumulated solar irradiation u during a period, i.e. $o = \langle i, c, u \rangle$. Then, solar irradiation can be modeled as a set of parallel 3D vectors (**E**) that passes through the atmosphere and approaches **O** at a given time and location. Meanwhile, a set of 3D solar vectors will intersect with the rooftop vertexes, leading to the creation of initial shadow surfaces modelled by 3D parallelograms.

Further, these shadow surfaces will be modified and recorded by **S** based on the following situations. First, shadow surfaces will be removed if façades are entirely in shadow, facing the sunlight or adjacent to a shorter one. Second, shadow surfaces will be reshaped if rooftop edges are partly in shadow or shadow surfaces intersect with other rooftops or façades. Ultimately, $\forall o \in \textbf{O}$, u will be 0 if point clouds are in shadow and calculated based on the energy conservation law if they are in sunlight. This model was developed in a relational database management system (PostgreSQL) with a series of SQL functions to support Big Data geo-computation, which utilized several database accelerating technologies, such as Spatial Index, Parallel Query, Views, and Temporary Tables.

4. Empirical investigation

4.1. Study area

Shanghai, a commercial center in Asia, is one of the biggest cities across the globe with a population of more than 25 million and an electricity consumption of more than 1500 GWh in 2021 (Shanghai Bureau of Statistics, 2022). Since Shanghai has a long history and has developed at an extraordinarily rapid rate over the past 40 years, it is polycentric and has a complex urban landscape with various building types, a mixture of modern skyscrapers and old villages (Shen & Wu, 2012; Chen et al., 2022; Yan, Wang, Zhang & Ratti, 2021). In an attempt to cope with the conflicts between energy transition and economic development, the demand for solar energy in Shanghai has increased (Chen et al., 2022), making it an ideal study area for constructing 3D urban surfaces and estimating PV potential. The study covered almost
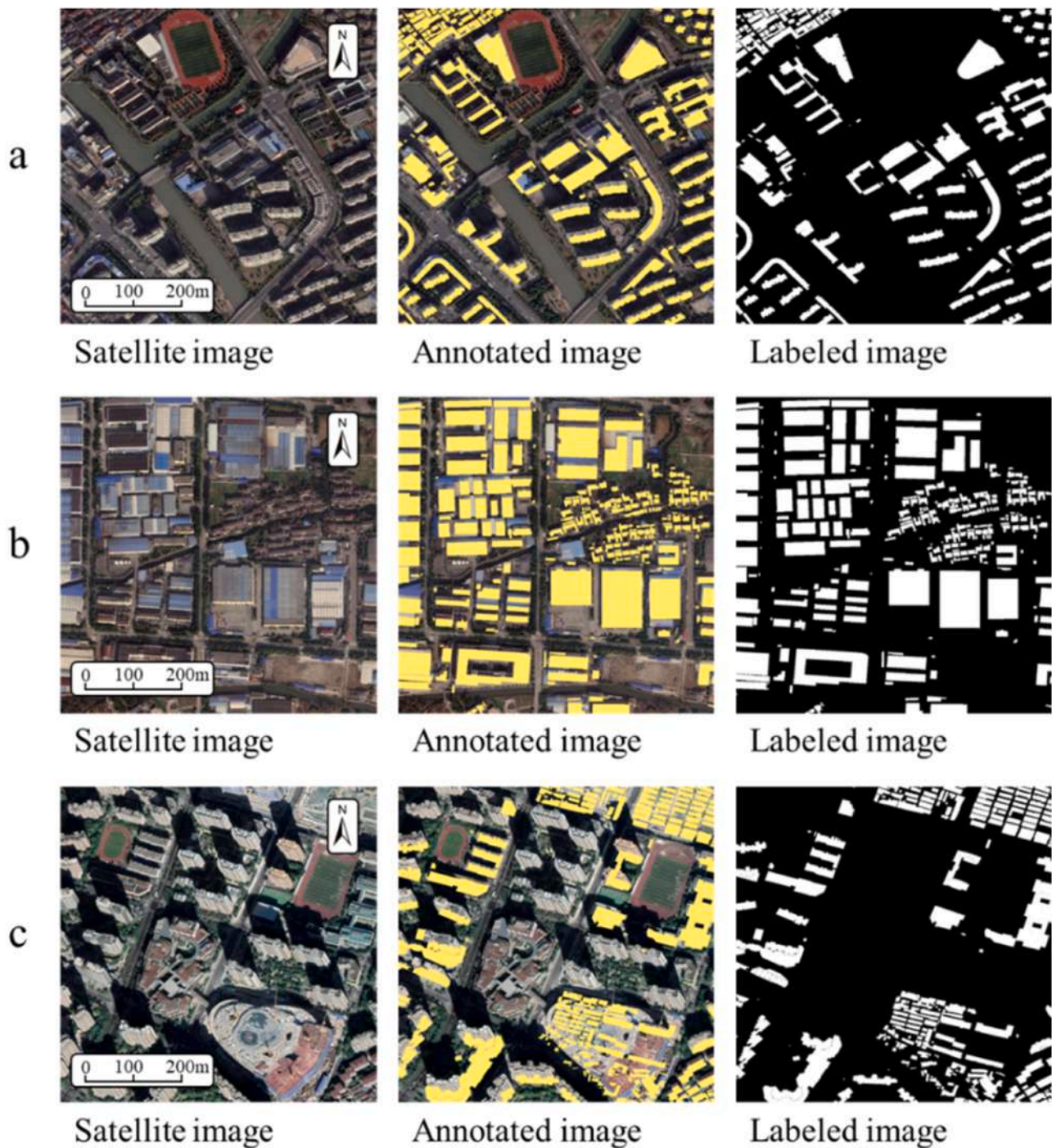
Fig. 6. Examples of labelled images in Shanghai. (a) A selected tile with high-quality labels in the central area. (b) A selected tile with high-quality labels in the suburban area. (c) An abandoned tile with low-quality labels.

the whole administrative area of Shanghai (Fig. 5) except for Chongming Island (5400 km^2), where the building density is relatively low.

4.2. Datasets

Optical remote sensing images covering the whole of Shanghai were obtained from Google Earth via Python scripts, which were acquired in the historical maps of 2014 and 2019 with three tunnels (Red, Green, and Blue) at a spatial resolution of 0.3–0.6 m. The reasons for using Google Earth satellite images are (1) they have sufficient pixel resolution for the building extraction task, (2) these images are pre-screened and merged in a way that most of them are cloudless, and (3) they cover almost all cities around the world so that our trained models may have the maximum range of application. The corresponding building footprint polygons in 2015 with height information (i.e. the number of storeys) were obtained from the Baidu Map, which has been used in numerous studies (Chen et al., 2020; Wang et al., 2021; Yao et al., 2017). The Baidu building polygons generally cover the whole central city, important new towns, and some of the suburban areas of Shanghai, providing a comprehensive representation of various building types and densities. To build a robust training model, an additional training dataset was obtained from the Inria Aerial ImageLabeling Dataset

Table 1

Summary of the experimental datasets.

Datasets	No. of Tiles	Area (km²)	Resolution (m)	Height information
Inria	5640	405	0.3	No
Shanghai (2014)	8950	740	0.2–0.6	Yes
Shanghai (2019)	7350	610	0.2–0.6	Yes
Total	21,940	1755	–	–

(2022), which provides satellite images at 0.3 m resolution with labelled building rooftops in three cities (i.e., Austin, Chicago, and Kitsap County) in the US and two (i.e. Vienna, and Western Tyrol) in Austria. The daily cloud cover data between 2009 and 2022 in Shanghai were obtained from World Weather Online (2022).

4.3. Data pre-processing

Since high-resolution images lead to huge data size, we divided the images and organized a set of tiles each containing 1600×1600 pixels. To obtain high-quality training and testing datasets in Shanghai, building polygons used as labels were geo-referenced onto the two-year satellite images. Since the accuracy and the update frequency of building polygons vary dramatically across areas, tiles with high-quality labels were manually screened and maintained for training and validation, while the rest were abandoned when they got outdated or misaligned with building footprint labels (Fig. 6). The OpenCV modules in Python were utilized in the above-mentioned procedures. Eventually, 16,300 well-matched tiles were selected for the experiment, covering about 1350 km² in Shanghai (Table 1). Together with the Inria dataset, we have 21,940 tiles with high-resolution satellite images covering about 1755 km² of four city areas worldwide. Since the number of tiles was sufficient, the selected tiles were organized into three categories, taking 95% and 5% for training and testing, respectively.

4.4. Transfer learning

As described in Section 2.3.5, the proposed model was first trained based on the Visual Object Classes Challenge and Cityscapes Dataset to obtain a general knowledge of geo-object segmentation. Next, the model selected a patch from each of the tiles in the Inria dataset at a random location for further training to gain specific capability on rooftop area segmentation. Lastly, the model was retrained by applying the same method used for the Shanghai dataset to accurately segment rooftops in Shanghai.

4.5. Rooftop segmentation accuracy

The evaluation based on Shanghai satellite images in 2014 achieved the best result, with *mIoU, recall, precision, accuracy*, and *F1-score* equalling 0.75, 0.84, 0.87, 0.94 and 0.84, respectively (Table 2), indicating that most of the building or non-building pixels are both correctly classified. Meanwhile, the rooftop area's MPE was limited to 1.8%, and the Pearson correlation was significantly high with $R = 0.97$ ($p<0.001$). These values indicate a high accuracy of the RSM segmentation. In contrast, the accuracy on Shanghai satellite images in 2019 is the lowest

regarding all metrics. This is probably because the satellite images in 2014 best matched the ground truth, i.e., the Baidu rooftop polygons acquired in 2015. Although the satellite images in 2019 that are not well matched with the building polygons are manually abandoned, the screening process still introduces some errors. Nevertheless, more than 80% of the building or non-building pixels in 2019 are still correctly classified, which is a satisfactory result. Additionally, despite that the trained model was particularly developed for Shanghai, it sufficed for Inria rooftop segmentations, achieving a *mIoU* of 0.75 and outperforming 76.8% of similar studies that utilized the Inria dataset (Leaderboard, 2022). Compared with similar large-scale building extraction studies, e.g., the one conducted by Microsoft in the U.S. and Canada, which obtained approximate IoUs of 0.86 and 0.76 and pixel-wised recall of 92% and 72% (Microsoft, 2023a, 2023b), our model can still be qualified for application.

4.6. Height estimation accuracy

Table 3 shows that the MAE of building height was 0.79 and 0.87 for the Shanghai testing dataset in 2014 and 2019, respectively, and the Pearson correlations are significantly high, implying that even on the testing dataset the estimated number of stroeys varied from the actual number within just one floor, and indicating an accurate prediction of building height. The MPE of building floor area was relatively large at 7.5% and 8.1% for the testing dataset. The reason is that the building floor area is calculated by multiplying the predicted rooftop/footprint area and the floor number so that the MPE inherits both the error of RSM and HPM. Meanwhile, this may also indicate significant challenges in building height prediction using solely satellite images: a small error in height prediction may lead to larger mistakes in building floor statistics.

4.7. Geo-visualized comparison between actual and estimated buildings

Quantitative accuracy assessments demonstrated that estimated and actual buildings could accurately match each other. To provide a direct visual impression, the two datasets are shown in Fig. 7, further validating two important characteristics. First, the predicted rooftop locations and shapes accurately match the actual ones, except for some predicted buildings that are missing in ground truth, which is due to the lack of timely updates on Baidu map. Second, the predicted building heights mostly match the actual ones quite well. However, 3D view also reveals that the height error of high-rise buildings is more significant than lower ones. The reason may be that training samples of high-rise buildings are not seemingly sufficient, making the model not that smart in estimating the height of tall blocks. Nevertheless, visual

Table 3

Accuracy assessment of the HPM.

Height prediction model	Pixel-wised statistics Height MAE	Building floor area statistics	
		MPE	Pearson correlation ($p<0.001$)
Testing dataset (2014)	0.79	7.5%	0.98
Testing dataset (2019)	0.87	8.1%	0.97

Table 2

Accuracy assessment of the RSM.

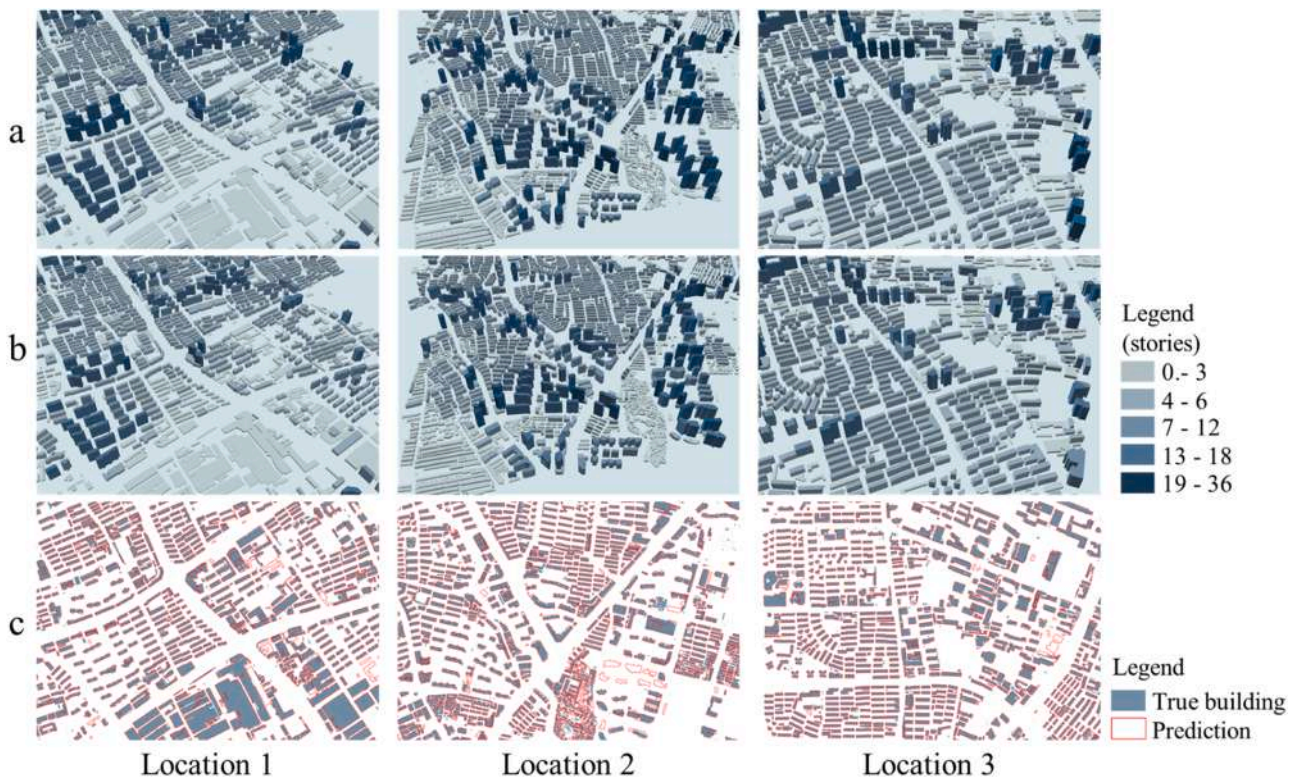| Rooftop segmentation model	Pixel-wised statistics					Rooftop area statistics	
	mIoU	Recall	precision	accuracy	F1-score	MPE	Pearson correlation($p<0.001$)
Testing dataset (Inria)	0.75	0.87	0.85	0.92	0.84	3.2%	0.95
Testing dataset (2014)	0.75	0.84	0.87	0.94	0.84	1.8%	0.97
Testing dataset (2019)	0.70	0.80	0.84	0.92	0.80	4.2%	0.95

Fig. 7. Comparison of actual buildings and post-processed building predictions. (a) Actual buildings in 3D view. (b) Predicted buildings in 3D view. (c) Boundaries comparison in 2D view.
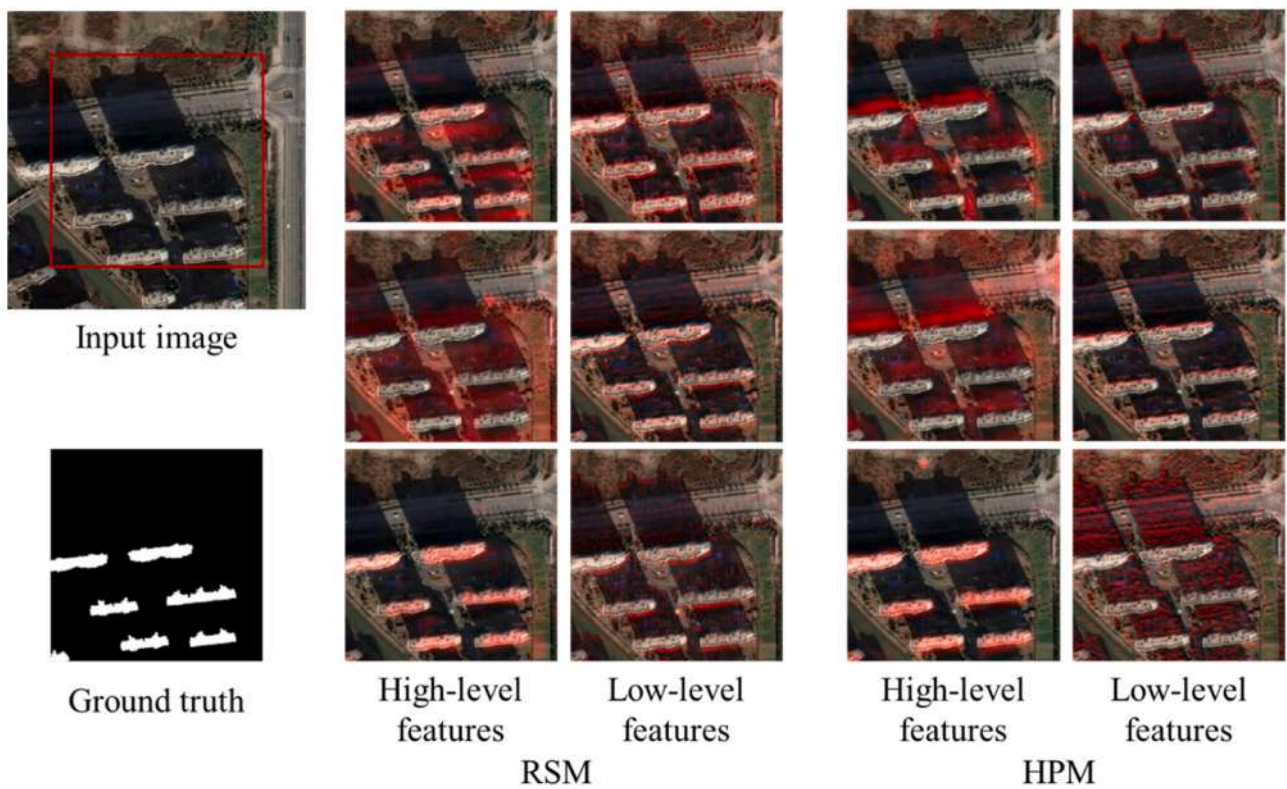

Fig. 8. Intermediate outputs of high- and low-level features of the advanced RSM and HPM.
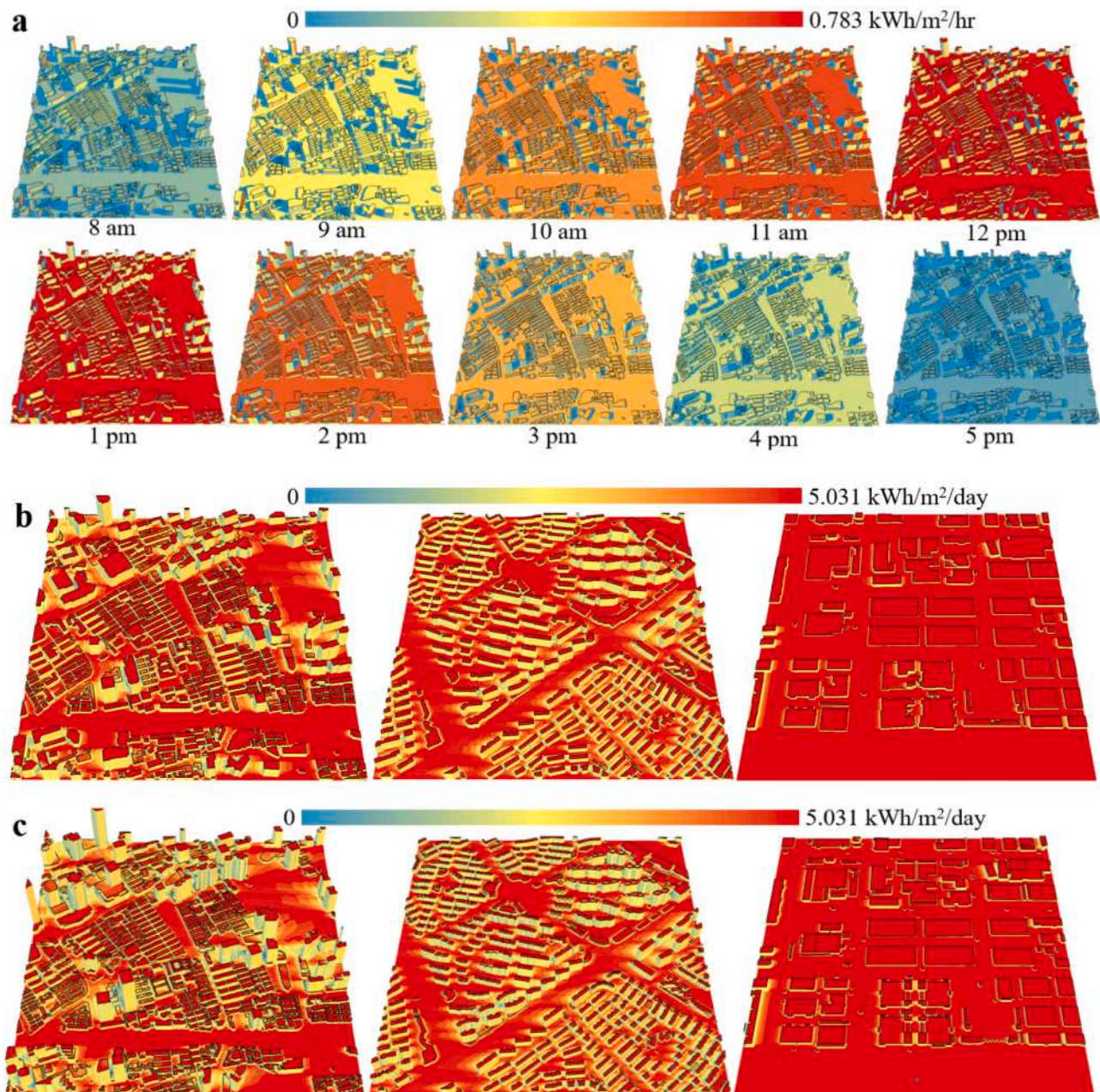
Fig. 9. Spatial distribution of solar potential on the 15th of September 2022. (a) Hourly distribution based on the predicted 3D urban surfaces. (b) Accumulated solar potential based on the predicted 3D urban surfaces in the business (left), residential (middle) and industrial (right) areas. (c) Accumulated solar potential based on the surveyed 3D urban surfaces in the business (left), residential (middle) and industrial (right) areas.

comparisons generally indicates that the model is effective in various urban morphologies where building densities are low or high, rooftop areas are small or large, and building heights are short or tall.

4.8. Interpretability of the RSM and HPM

Motivated by the impressive performance of the detail-oriented deep learning network that has been well-packaged as a black box, we further explored the intermediate outputs of the model, as conducted in many studies (e.g., Sun et al., 2022a; Zhang & Zhu, 2018). To accomplish this, both high- and low-level features, as shown in Fig. 3, were extracted from the network, highlighted in red and overlapped with input images (Fig. 8). We found that high-level features of the RSM successfully learned to identify abstract objects, including the green space between buildings, rooftops, and the open space in each street block, while high-level features of HPM focused more on the side profiles of buildings and the shadows. In contrast, the low-level features of both the RSM and HPM focused particularly on basic textures and veins, especially of rooftops, impervious surfaces and shadows. This implies that our refined network, which coped with the shadow effect, can effectively improve the height estimation. Our analysis also revealed that land use and land cover associated with specific textures and hues are important geo-features for 3D building construction.

5. Spatiotemporal distribution of 3D solar PV potential

Fig. 9(a) shows the hourly solar distribution in the business district based on the estimated 3D urban surfaces on the 15th of September 2022; this figure successfully models spatiotemporal variation of solar distribution with shadow effects made by surrounding buildings. To
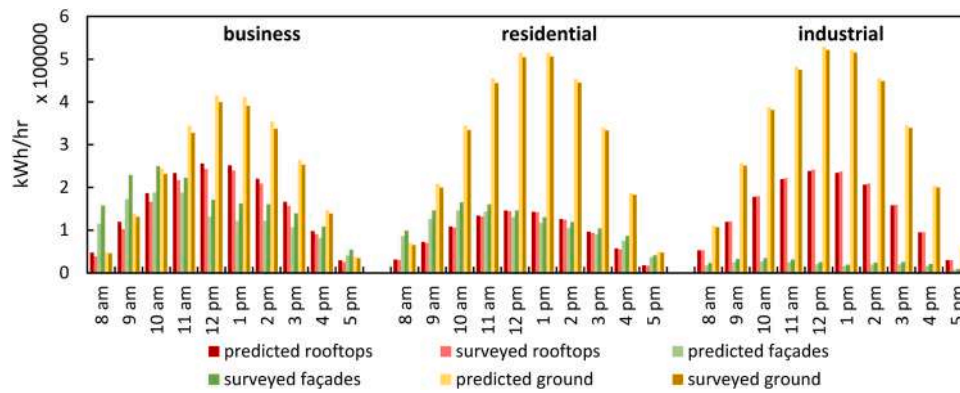
Fig. 10. Statistics of hourly solar potential based on the predicted and surveyed 3D urban surfaces in three districts (i.e., business, residential, and industrial) on the 15th of September 2022.

Table 4
Accuracy statistics of the three partitions in three urban districts.

Index	Area	Rooftop	Façade	Ground
R^2	Business	0.9974	0.9812	0.9999
	Residential	0.9998	0.9983	0.9999
	Industrial	0.9999	0.9951	0.9999
Relative error (%)	Business	9.3590	31.3194	4.8354
	Residential	2.7115	13.7284	3.0238
	Industrial	0.7821	26.7684	1.7134

Table 5
PV potential on the 15th of September 2022.

Area	Building datasets	Rooftop (MWh)	Façade (MWh)	Ground (MWh)	Absolute relative error (%)
Business	Predicted	283.42	222.87	422.76	2.9493
	Measured	261.99	291.07	403.40	
Residential	Predicted	164.27	185.72	552.29	0.9553
	Measured	160.92	210.89	539.07	
Industrial	Predicted	269.59	34.08	590.15	0.3047
	Measured	272.05	43.21	581.29	

investigate the similarity of solar distribution on 3D urban surfaces constructed by deep learning-based estimation (Fig. 9(b)) and LiDAR scanning (Fig. 9(c)), we compared the accumulated solar distributions in three districts (i.e. business, residential and industrial) on the same day which had distinctly different urban morphology. Visually, solar distributions computed from the two urban-surface sources are highly similar. Statistics of hourly solar irradiation on three urban partitions (i.e., rooftops, façades and ground) are presented in Fig. 10 and suggest that there is only an insignificant difference based on two difference data sources, i.e., predicted versus surveyed urban surfaces. Specifically, R^2 was larger than 0.99 for all three partitions (Table 4), implying a robust regression between hourly solar irradiations from the two sets of urban surfaces. Meanwhile, the absolute relative errors of estimated solar potential on rooftops and ground were less than 10% (Table 4), suggesting a satisfactory estimation accuracy on horizontal surfaces. However, the errors were relatively large for solar estimation on façades, especially in the business district which had a high density of tall buildings. We also found that the errors mainly came from the early morning (between 8 am and 10 am) when solar irradiation has a low elevation angle. This means that the errors in building heights and differences in façade orientations may affect façade solar estimation considerably.

Finally, assuming that the PV conversion efficiency is 22% and the performance ratio is 80% (Zhu et al., 2022c), the PV potentials based on the predicted and surveyed building datasets are presented in Table 5, which shows that the PV potentials based on estimated and surveyed 3D

building surfaces are extraordinarily similar on rooftops, façades, and ground. For the total PV potential on all surfaces, the absolute relative errors were only 2.95%, 0.95%, and 0.30% for the business, residential and industrial areas, respectively. These results suggest that the PV potential estimations using the constructed 3D urban surfaces are effective in the scope of large-scale analysis.

6. Discussion and conclusion

This study proposes a novel deep learning-based approach for constructing 3D buildings and estimating solar PV potential. The two hierarchical models for segmenting rooftop areas (with an MPE between 1.8% and 4.2%) and predicting building heights (with an MPE between 7.5% and 8.1%) with morphological post-processing have achieved satisfactory accuracy based on several testing datasets. The models also achieved a highly accurate estimation of PV potential on rooftops and ground; however, the accuracy was relatively low on façades. Although the total PV potential on the estimated 3D urban surfaces was highly accurate in all three large areas, the future study can conduct accuracy assessments on individual buildings by developing an adaptive spatio-temporal statistical model.

This study is significant in three aspects. First, we developed a deep learning-based approach to accurately constructing *all* 3D building surfaces in a densely urban area from optical bands of satellite imagery. It is innovative in integrating two semantic segmentation networks by eliminating the marginal effects, revising the outputs from possibility distributions to real numbers, and proposing hybrid loss functions, which successfully constructed 3D buildings purely from 2D satellite imagery. Second, this study provides an in-depth understanding of the deep learning network built by the Decoder and Encoder structure. The results suggest the effectiveness of spatial and channel attention that the high-level features can identify abstract objects in the RSM and pay special attention to the side profiles of buildings and shadows in the HPM, and the low-level features particularly focus on the basic textures and veins. Third, this study proposes an alternative approach for constructing 3D building models when the current 3D city models built by surveying technologies become outdated or unavailable. In this aspect, our approach is particularly useful for easy, fast, and economical 3D-city data generation that can be used for a variety of applications, such as urban planning and PV installation optimization. Thus, our approach can facilitate the development of many fast-growing countries where frequent LiDAR data updates are unavailable.

The dates of satellite images and building polygons used for validation and testing are not entirely consistent with each other, adversely affecting accuracy assessment. For example, visual examination confirmed that the developed network successfully segmented rooftop areas from satellite imagery; however, some of them were not included in the 3D building dataset. This implies that the actual accuracy for both

rooftop areas and building heights should be higher than the current statistics. Meanwhile, the results show that relative errors of estimated building heights are larger than those of segmented rooftop areas. Nevertheless, R^2 is very high for all three urban partitions, indicating a robust regression between the predicted and measured building surfaces. Therefore, it is possible to integrate the regression models into transfer learning for urban areas having similar morphological characteristics, which can considerably improve building height estimation.

Notably, morphological post-processing can be further improved. The Regularize Building Footprint tool provided by ArcGIS Pro-was used to regularize the polygons converted from the predicted buildings. However, the tool focuses on eliminating undesirable geometric artefacts, while many building properties, such as the angles which usually stay within a certain range and edges which are usually perpendicular to each other and generally maintaining a certain length, are not taken into consideration. Therefore, we speculate that morphological post-processing can be improved using well-tuned regularization algorithms that combine several manually defined rules.

This study used the same solar irradiation model and input parameters based on the estimated and surveyed building data for solar distribution evaluation. This means that solar distribution differences are created only because the same solar irradiation arrives at different building surfaces (e.g., different heights and façade orientations). To create large variations of solar conditions, hourly comparisons were made when solar irradiations changed significantly in azimuths (from 85.81° to 266.10°) and elevation angles (from 3.86° to 61.71°) from 8 am to 5 pm on 15th September 2022. This allows us to confidently evaluate the robustness of the constructed building model. Nevertheless, the evaluation can be more robust if computing more days in a year, such as considering solstices and equinoxes or the characteristic declination days (Brito, e Silva & Freitas, 2021), which helps to systematic estimate and evaluate annual solar PV potential while minimize the total computation.

The assumption that all rooftops are flat causes a certain deviation from reality. Motivated by the satisfactory results of this study, future work can develop a similar deep-learning network to estimate the slope and aspect of all rooftops. Since rooftop surfaces in different slopes (flat or steep) and aspects (facing the sun or shadow) present various brightness and textures, their high- and low-level features can be captured and learned by deep learning networks. This can be achieved when the datasets from real observations, such as 3D point clouds obtained from LiDAR scanning, are available to calculate slopes and aspects for training, validation, and testing. Further, rooftop structure lines (e.g. boundaries and valley lines) have been segmented from high-resolution satellite imagery (Qian et al., 2022b), which means that rooftop surfaces can be determined and used as input features to facilitate rooftop slope and aspect prediction.

In conclusion, this study is innovative in presenting a model that easily, rapidly, economically and accurately constructs 3D buildings from 2D satellite imagery and estimates PV potential on continuous urban surfaces. The proposed framework is technically effective for integrating deep learning-based rooftop area segmentation and building height estimation with transfer learning, shadow effect elimination and morphological post-processing. The proposed framework can be generalized and is deliverable in other cities; this is particularly useful for a variety of applications related to the urban built environment.

Declaration of Competing Interest

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

Data availability

Data will be made available on request.

References

Aslani, M., & Seipel, S. (2022). Automatic identification of utilizable rooftop areas in digital surface models for photovoltaics potential assessment. *Applied Energy, 306*, Article 118033.

Bianchi, M., Branchini, L., Ferrari, C., & Melino, F. (2014). Optimal sizing of grid-independent hybrid photovoltaic–battery power systems for household sector. *Applied Energy, 136*, 805–816.

Brito, M. C., e Silva, R. A., & Freitas, S. (2021). Characteristic Declination—A Useful Concept for Accelerating 3D Solar Potential Calculations. *Energy Technology, 9*(3), Article 2000943.

Bshouty, E., Shafir, A., & Dalyot, S. (2020). Towards the generation of 3D OpenStreetMap building models from single contributed photographs. *Computers, Environment and Urban Systems, 79*, Article 101421.

Cao, Y., & Huang, X. (2021). A deep learning method for building height estimation using high-resolution multi-view imagery over urban areas: A case study of 42 Chinese cities. *Remote Sensing of Environment, 264*, Article 112590.

Catita, C., Redweik, P., Pereira, J., & Brito, M. C. (2014). Extending solar potential analysis in buildings to vertical facades. *Computers & Geosciences, 66*, 1–12.

Chen, W., Zhou, Y., Wu, Q., Chen, G., Huang, X., & Yu, B. (2020). Urban building type mapping using geospatial data: A case study of Beijing, China. *Remote Sensing, 12* (17), 2805.

Chen, Z., Yu, B., Li, Y., Wu, Q., Wu, B., Huang, Y., et al. (2022). Assessing the potential and utilization of solar energy at the building-scale in Shanghai. *Sustainable Cities and Society, 82*, Article 103917.

Cityscapes Dataset (2022). Semantic Understanding of Urban Street Scenes. Accessed 19 September 2022 https://www.cityscapes-dataset.com/.

Feng, F., & Wang, K. C. (2021). Merging ground-based sunshine duration observations with satellite cloud and aerosol retrievals to produce high-resolution long-term surface solar radiation over China. *Earth System Science Data, 13*, 907–922.

Gooding, J., Crook, R., & Tomlin, A. S. (2015). Modelling of roof geometries from low-resolution LiDAR data for city-scale solar energy applications using a neighbouring buildings method. *Applied Energy, 148*, 93–104.

Hu, P., Yang, B., Dong, Z., Yuan, P., Huang, R., Fan, H., et al. (2018). Towards Reconstructing 3D Buildings from ALS Data Based on Gestalt Laws. *Remote Sensing, 10*, 1127.

Huang, H., Chen, P., Xu, X., Liu, C., Wang, J., Liu, C., et al. (2022). Estimating building height in China from ALOS AW3D30. *ISPRS Journal of Photogrammetry and Remote Sensing, 185*, 146–157.

Huang, J., Zhang, X., Xin, Q., Sun, Y., & Zhang, P. (2019b). Automatic building extraction from high resolution aerial images and LiDAR data using gated residual refinement network. *ISPRS Journal of Photogrammetry and Remote Sensing, 151*, 91–105.

Huang, S., Rich, P. M., Crabtree, R. L., Potter, C. S., & Fu, P. (2008). Modeling monthly near-surface air temperature from solar radiation and lapse rate: Application over complex terrain in Yellowstone national park. *Physical Geography, 29*, 158–178.

Huang, Z., Mendis, T., & Xu, S. (2019a). Urban solar utilization potential mapping via deep learning technology: A case study of Wuhan, China. *Applied Energy, 250*, 283–291.

Inria Aerial Image Labeling Dataset (2022). The automatic pixelwise labeling of aerial imagery. Accessed 19 September 2022 https://project.inria.fr/aerialimagelabeling.

Jakubiec, J. A., & Reinhart, C. F. (2013). A method for predicting city-wide electricity gains from photovoltaic panels based on LiDAR and GIS data combined with hourly Daysim simulations. *Solar Energy, 93*, 127–143.

Jochem, A., Höfle, B., Rutzinger, M., & Pfeifer, N. (2009). Automatic roof plane detection and analysis in airborne lidar point clouds for solar potential assessment. *Sensors, 9*, 5241–5262.

Kabir, M. H., Endlicher, W., & Jägermeyr, J. (2010). Calculation of bright rooftops for solar PV applications in Dhaka Megacity, Bangladesh. *Renewable Energy, 35*, 1760–1764.

Lao, J., Wang, C., Zhu, X., Xi, X., Nie, S., Wang, J., et al. (2021). Retrieving building height in urban areas using ICESat-2 photon-counting LiDAR data. *International Journal of Applied Earth Observation and Geoinformation, 104*, Article 102596.

Leaderboard (2022). Accuracy comparison obtained from different networks. Accessed 19 September 2022 https://project.inria.fr/aerialimagelabeling/leaderboard/.

Li, M., Koks, E., Taubenböck, H., & Van Vliet, J. (2020a). Continental-scale mapping and analysis of 3D building structure. *Remote Sensing of Environment, 245*, Article 111859.

Li, P., Zhang, H., Guo, Z., Lyu, S., Chen, J., Li, W., et al. (2021). Understanding rooftop PV panel semantic segmentation of satellite and aerial images for better using machine learning. *Advances in Applied Energy, 4*, Article 100057.

Li, X., He, M., Li, H., & Shen, H. (2022). A Combined Loss-Based Multiscale Fully Convolutional Network for High-Resolution Remote Sensing Image Change Detection. *IEEE Geoscience and Remote Sensing Letters, 19*, Article 8017505.

Li, X., Zhou, Y., Gong, P., Seto, K. C., & Clinton, N. (2020b). Developing a method to estimate building height from Sentinel-1 data. *Remote Sensing of Environment, 240*, Article 111705.

Liao, X., Zhu, R., & Wong, M. S. (2022). Simplified estimation modeling of land surface solar irradiation: A comparative study in Australia and China. *Sustainable Energy Technologies and Assessments, 52*, Article 102323.

Liasis, G., & Stavrou, S. (2016). Satellite images analysis for shadow detection and building height estimation. *ISPRS Journal of Photogrammetry and Remote Sensing, 119*, 437–450.

Lindberg, F., Jonsson, P., Honjo, T., & Wästberg, D. (2015). Solar energy on building envelopes –3D modelling in a 2D environment. *Solar Energy, 115*, 369–378.

Lobaccaro, G., Carlucci, S., Croce, S., Paparella, R., & Finocchiaro, L. (2017). Boosting solar accessibility and potential of urban districts in the Nordic climate: A case study in Trondheim. *Solar Energy, 149*, 347–369.

Microsoft (2023a). Open building footprints datasets in Canada. Assessed 9 February 2023 https://github.com/microsoft/CanadianBuildingFootprints.

Microsoft (2023b). Open building footprints datasets in the United Sates. Assessed 9 February 2023 https://github.com/microsoft/USBuildingFootprints.

Mohajeri, N., Assouline, D., Guiboud, B., Bill, A., Gudmundsson, A., & Scartezzini, J.-. L. (2018). A city-scale roof shape classification using machine learning for solar energy applications. *Renewable Energy, 121*, 81–93.

Norman, M., Shafri, H. Z. M., Idrees, M. O., Mansor, S., & Yusuf, B. (2020). Spatio-statistical optimization of image segmentation process for building footprint extraction using very high-resolution WorldView 3 satellite data. *Geocarto International, 35*(10), 1124–1147.

OpenStreetMap (2022). OpenStreetMap is a map of the world, created by people like you and free to use under an open license. Accessed 19 September 2022 https://www.openstreetmap.org/.

Peronato, G., Rey, E., & Andersen, M. (2018). 3D model discretization in assessing urban solar potential: The effect of grid spacing on predicted solar irradiation. *Solar Energy, 176*, 334–349.

Qi, F., Zhai, J. Z., & Dang, G. (2016). Building height estimation using. *Google Earth. Energy and Buildings, 118*, 123–132.

Qian, Z., Chen, M., Yang, Y., Zhong, T., Zhang, F., Zhu, R., et al. (2022a). Vectorized dataset of roadside noise barriers in China. *Earth System Science Data, 14*, 4057–4076.

Qian, Z., Chen, M., Zhong, T., Zhang, F., Zhu, R., Zhang, Z., et al. (2022b). Deep Roof Refiner: A detail-oriented deep learning network for refined delineation of roof structure lines using satellite imagery. *International Journal of Applied Earth Observation and Geoinformation, 107*, Article 102680.

Shanghai Bureau of Statistics (2022). Energy Consumption in Main Years. Assessed 19 September 2022 https://tjj.sh.gov.cn/tjnj/nj21.htm?d1=2021tjnjen/E0501.htm.

Shen, J., & Wu, F. (2012). Restless urban landscapes in China: A case study of three projects in Shanghai. *Journal of Urban Affairs, 34*(3), 255–277.

Sun, M., Zhang, F., Duarte, F., & Ratti, C. (2022a). Understanding architecture age and style through deep learning. *Cities (London, England), 128*, Article 103787.

Sun, Y., Mou, L., Wang, Y., Montazeri, S., & Zhu, X. X. (2022b). Large-scale building height retrieval from single SAR imagery based on bounding box regression networks. *ISPRS Journal of Photogrammetry and Remote Sensing, 184*, 79–95.

Visual Object Classes Challenge (2022). PASCAL Visual Object Classes Challenge 2012. Accessed 19 September 2022 http://host.robots.ox.ac.uk/pascal/VOC/voc2012/.

Wang, C., Wei, S., Du, S., Zhuang, D., Li, Y., Shi, X., et al. (2021). A systematic method to develop three dimensional geometry models of buildings for urban building energy modeling. *Sustainable Cities and Society, 71*, Article 102998.

Wong, M. S., Zhu, R., Liu, Z., Lu, L., Peng, J., Tang, Z., et al. (2016). Estimation of Hong Kong's solar energy potential using GIS and remote sensing technologies. *Renewable Energy, 99*, 325–335.

World Weather Online (2022). Weather API and Data for Businesses and Developers. Accessed 19 September 2022 https://www.worldweatheronline.com/.

Yan, L., Wang, D., Zhang, S., & Ratti, C. (2021). Understanding urban centers in Shanghai with big data: Local and non-local function perspectives. *Cities (London, England), 113*, Article 103156.

Yan, Y., & Huang, B. (2022). Estimation of building height using a single street view image via deep neural networks. *ISPRS Journal of Photogrammetry and Remote Sensing, 192*, 83–98.

Yang, B., Huang, R., Li, J., Tian, M., Dai, W., & Zhong, R. (2016). Automated Reconstruction of Building LoDs from Airborne LiDAR Point Clouds Using an Improved Morphological Scale Space. *Remote Sensing, 9*, 14.

Yao, Y., Liu, X., Li, X., Zhang, J., Liang, Z., Mai, K., et al. (2017). Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data. *International Journal of Geographical Information Science, 31*(6), 1220–1244.

Zhang, J., Xua, L., Shabunko, V., Tay, S. E. R., Sun, H., Lau, S. S. Y., et al. (2019). Impact of urban block typology on building solar potential and energy use efficiency in tropical high-density city. *Applied Energy, 240*, 513–533.

Zhang, K., Qian, Z., Chen, M., Yang, Y., Zhong, T., Zhu, R., et al. (2022a). Using street view images to identify road noise barriers with ensemble classification model and geospatial analysis. *Sustainable Cities and Society, 78*, Article 103598.

Zhang, L., Li, Z., Li, A., & Liu, F. (2018). Large-scale urban point cloud labeling and reconstruction. *ISPRS Journal of Photogrammetry and Remote Sensing, 138*, 86–100.

Zhang, Q. S., & Zhu, S. C. (2018). Visual interpretability for deep learning: A survey. *Frontiers of Information Technology & Electronic Engineering, 19*(1), 27–39.

Zhang, Z., Qian, Z., Zhong, T., Chen, M., Zhang, K., Yang, Y., et al. (2022b). Vectorized rooftop area data for 90 cities in China. *Scientific Data, 9*, 66.

Zhong, T., Zhang, K., Chen, M., Wang, Y., Zhu, R., Zhang, Z., et al. (2021b). Assessment of solar photovoltaic potentials on urban noise barriers using street-view imagery. *Renewable Energy, 168*, 181–194.

Zhong, T., Zhang, Z., Chen, M., Zhang, K., Zhou, Z., Zhu, R., et al. (2021a). A city-scale estimation of rooftop solar photovoltaic potential based on deep learning. *Applied Energy, 298*, Article 117132.

Zhou, L., & Vosselman, G. (2012). Mapping curbstones in airborne and mobile laser scanning data. *International Journal of Applied Earth Observation and Geoinformation, 18*, 293–304.

Zhu, R., Cheng, C., Santi, P., Chen, M., Zhang, X., Mazzarello, M., et al. (2022b). Optimization of photovoltaic provision in a three-dimensional city using real-time electricity demand. *Applied Energy, 316*, Article 119042.

Zhu, R., Kondor, D., Cheng, C., Zhang, X., Santi, P., Wong, M. S., et al. (2022a). Solar photovoltaic generation for charging shared electric scooters. *Applied Energy, 313*, Article 118728.

Zhu, R., Wong, M. S., Kwan, M. P., Min, C., Santi, P., & Ratti, C. (2022c). An economically feasible optimization of photovoltaic provision using real electricity demand: A case study in New York City. *Sustainable Cities & Society, 78*, Article 103614.

Zhu, R., Wong, M. S., You, L., Santi, P., Nichol, J., Ho, H. C., et al. (2020). The effect of urban morphology on the solar capacity of three-dimensional cities. *Renewable Energy, 153*, 1111–1126.

Zhu, R., You, L., Santi, P., Wong, M. S., & Ratti, C. (2019). Solar accessibility in developing cities: A case study in Kowloon East, Hong Kong. *Sustainable Cities and Society, 51*, Article 101738.