Advances and prospects on estimating solar photovoltaic installation capacity and potential based on satellite and aerial images

Advances and Prospects on Estimating Solar Photovoltaic Installation Capacity and Potential based on Satellite and Aerial Images

Hongzhi Mao[a], Xie Chen[a], Yongqiang Luo[a], Jie Deng[b], Zhiyong Tian[a,1], Jinghua Yu[a], Yimin Xiao[c], Jianhua Fan[d]

[a] *College of Environmental Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China*

[b] *School of Computing and Engineering, University of West London, United Kingdom*

[c] *School of Civil Engineering, Chongqing University, Chongqing 400045, China*

[d] *Department of Civil and Mechanical Engineering, Technical University of Denmark, Brovej 118, 2800 Kgs. Lyngby, Denmark*

Abstract

Solar photovoltaic (PV) system, as one kind of the most promising renewable energy technologies, plays a key role in reducing carbon emissions to achieve the targets of global net zero carbon. In the past few decades, PV installations have seen a rapid growth. Predicting the installed amount and the capacity of solar PV systems is therefore useful for formulating effective carbon reduction policies in the related area. In the present study, the methods of identifying PV installation based on satellite and aerial images have been reviewed. Suggestions have been put forward to optimize the identification process and to predict the potential of rooftop PV installation. The results show that the specific purposes of PV identification can be categorized as image classification, object detection and semantic segmentation. The available identification methods encompass pixel-based analysis method (PBIA), object-based analysis method (OBIA) and deep learning. Deep learning has a high accuracy in segmentation for all sizes of PV systems, with precision and recall of rooftop PV segmentation in the range of 41-98.9% and 54.5-95.8%, respectively. OBIA has the best accuracy in detecting centralized PV systems with relatively low-resolution multispectral images.

1 Corresponding author: Zhiyong Tian, tianzy0913@163.com; zhiyongtian@hust.edu.cn

Furthermore, a grading segmentation strategy for PV segmentation in the large region is presented, combining the three identification methods and the images with different resolutions. In addition, the potential of rooftop PV installation can be predicted by segmenting the available roof area in the images. After considering the shading effects, upper structure and other uses, the roof availability coefficient tends to be in the range of 0.25 to 0.46. It is also suggested to combine PV and roof segmentation to estimate the installation potential more accurately, in the context of rapid growth of the rooftop PV.

Highlights

- The sources and their characteristics of satellite and aerial images are analyzed
- The methods of PBIA, OBIA, and deep learning are classified and compared
- The applications of PV identification model are summarized
- Optimization of potential prediction of rooftop PV is discussed
- A grading strategy for PV segmentation with high efficiency, low cost and ensuring accuracy is proposed

Keywords: solar photovoltaic, identification, satellite and aerial images, deep learning

Word Count: 9163

1 Introduction

As of April 2022, 128 countries have committed to achieving carbon neutrality [1]. Nevertheless, it is a big challenge for every country to achieve carbon neutrality synchronously, especially for developing countries like China, which is the world's largest carbon emitter. The Chinese government has set goals of peaking in 2030 and achieving carbon neutrality by 2060 [2]. The proportion of electricity provided by non-fossil fuels in China must reach 77% by 2030 and 85% by 2050 in the whole energy structure, among which 17% of the electricity demand will be met by solar energy [3-5]. In this situation, it is not realistic to continue to use natural gas as the primary fuel in a mixed energy supply mode for carbon neutrality, in the meanwhile renewable energy must be vigorously developed as soon as possible [6].

Solar energy, as a kind of clean and renewable energy, plays an important role in the development of global

renewable energy applications. The technologies to harness solar energy embrace solar PV, solar thermal applications, and solar thermal energy storage [7, 8]. Among these technologies, it is reported that the global installed capacity of solar PV in 2020 is 127 GW, accounting for more than 49% of the total new renewable energy capacity [9]. Whilst China market has contributed to 48.2 GW, with a cumulative installed capacity of 253 GW, accounting for one third of the global installed capacity [10]. Nevertheless, it is estimated that 2,200-2,800 GW of solar capacity and 1,500-2,600 GW of wind capacity are still needed for China to achieve the Paris Agreement. On the other hand, the cost of solar PV production and installation is falling rapidly, with the globally weighted average cost of electricity being fallen by 77% for utility-scale PV between 2010 and 2018 [11]. Solar PV will continue to be the fastest-growing renewable energy application in the context [12].

Regarding the installation site of solar PV, farmland is the most common land type for the installation of centralized solar PV systems, followed by arid areas and grasslands [13]. On the other hand, electricity demand in cities is greater than in rural areas, while urban areas do not have a lot of land for centralized PV installation, resulting in a mismatch between PV power generation and energy demands. It is much more difficult to further develop large-scale centralized PV power plants in urban land areas. Rooftop PV technology has the advantage of saving land and currently accounts for 40% of the world's cumulative installed solar PV capacity [14]. It is estimated that rooftop PV can contribute 3.7-4.5% of China’s total power consumption in 2030 if 20% of the urban roof area is utilized [15]. In addition, rooftop PV can improve the living conditions of citizens who reside in backward areas not connected to power grids, by alleviating their energy accessibility problems. Therefore, it is expected that rooftop PV will grow massively in the soon future.

With the huge potential of a rooftop solar PV installation, it is of great significance to estimate the total installed solar PV capacity and power generation accurately. This will benefit policymakers and stakeholders. In the environmental sector, adequate energy statistics can provide a more complete picture of the country's progress

towards achieving carbon neutrality, allowing timely adjustments to national development goals. For government policymakers, understanding the accurate PV installation situation status leads to a better understanding of the implementation effect of incentive policies and how to adjust policies timely. For the electric power operation departments, the power grid working conditions can be better predicted with the photovoltaic layout and the future weather conditions, in order to prevent drastic voltage changes and power outage accidents. As to producers, production plans can be reasonably arranged in terms of targeted investment in publicity.

Currently, methods for estimating the number and capacity of installed PV systems include official registers, crowdsourced field surveys, behind-the-meter analysis, and identification in satellite and aerial images. In general, in order to grant installation permits or financial subsidies, government departments usually register PV information. However, most centralized PV systems are in government registries, and the registries for distributed PV systems are incomplete [16]. It is because their installation time was too early, or owners tried to evade permit fees, among other reasons. Although the capacity of a single distributed PV system is small, there are so many of them that they still account for 30% and 52% of the UK and Germany's total PV capacity, respectively [17, 18]. The crowdsourced field survey requires a lot of manpower, and the quality of data uploaded by different volunteers varies greatly [18]. It also requires high costs of data updates. The capacity of the centralized PV system can be estimated by analyzing the operation state of the electricity meter and the local solar irradiance. However, when the PV system is small, or the system is equipped with energy storage devices such as batteries, the prediction accuracy is low currently [19, 20].

In recent years, using satellite and aerial images to identify land cover types, road distribution and buildings has become a popular research means [21, 22]. Using satellite images to identify solar PV is confirmed as a feasible method. In 2018, Yu et al. [23] identified a total of (1.4702 ± 0.0007) million PV systems in the United States through aerial images, which exceeded the 1.02 million installations in

government registers at that time. Therefore, this method is considered the best method at present. In addition, after the satellite and aerial images are imported into the Geographic Information System (GIS), with a large number of geographic information such as terrain, location of roads and buildings, climate and economy data analyzed, the suitability of PV installation at each location could be calculated through the Analytic Hierarchy Process or multi-objective optimization [24-26]. It can guide the location of PV systems. The advantages and disadvantages of the above four methods for estimating the PV installation are summarized in Table 1.

Table 1 Comparison of methods for estimating PV installation

Methods	Advantages	Disadvantages
Official registers	It has complete information about almost all centralized PV systems. The information about the area, capacity, and tilt angle has high credibility.	The information about many distributed PV systems is incomplete. The information about the PV systems dismantled cannot be updated in time
Crowdsourced field surveys	It is suitable for small-area statistics. It can be statistics flexibly of various types of information of PV systems.	The quality of data uploaded by different volunteers is not uniform. A lot of manpower is required. The cost of updating data is high
Behind-the-meter analysis	The actual power of the PV system can be predicted directly. It could be used to guide power grid management.	The prediction error of distributed PV systems is large. A large number of actual meter data is difficult to obtain. The application of batteries affects prediction accuracy. It cannot be further used to estimate the PV installation potential.
Identification in satellite and aerial images	Most satellite and aerial images could be got for free. The number of PV systems can be counted fully. The information about location, area, and azimuth is accurate. It has a fast statistics speed and requires less manpower. The database is easy to update. The installation potential could be analyzed in combination with building identification.	A lot of computing resources are required. High-resolution aerial images covering the globe are hard to be obtained.

Recently, some review articles have mentioned the research status of PV identification using satellite and aerial images. Hoog et al. [27] summarized several classic studies and pointed out existing research challenges and opportunities. Cong et al. [28] and Gaviria et al. [29] stated the applications of artificial intelligence methods in the PV power grid and mentioned several classic studies on PV identification. However, different studies use satellite images and aerial images from different sources. The resolution, spectral band distribution, and download cost of these image data sources are different. At the same time, the

specific purposes of PV identification are also different, including image classification, object detection, and semantic segmentation. Object detection is suitable for counting the location and number of PV systems. Semantic segmentation can also be used to further estimate the area and capacity of the PV system. There are also a variety of methods of image identification, which involve machine learning methods and deep learning methods. At present, there is a lack of a comprehensive investigation of these studies, and the advantages, disadvantages, and applicable scenarios of various image data sources and identification methods have not been compared and analyzed in detail. This may lead to resource waste and insufficient identification accuracy in future research. Thus, it is necessary to scrutinize and analyze the existing methods in detail.

The present paper aims to review existing studies on PV and roof identification using satellite and aerial imagery from a holistic perspective. The available satellite and aerial image database sources are discussed in Section 2. Various models and methods are analyzed and compared in Section 3. In Section 4, the applications of the models in wide regions as well as the study of identifying roof area and its available area are discussed, which is of great significance for better exploiting the potential of roof PV. Finally, further research into identifying and counting the existing PV area as well as analyzing the exploitable potential of rooftop PV is discussed.

2 Image Collection and Model Metrics

2.1 Image Data Sources

Image data for solar PV installation identification can be obtained from satellite images or aerial images. Table 2 provides basic information about these image data sources. Commonly used satellite images come from Landsat-8, Sentinel, SPOT-6/7, and Gaofen-2, the first two of which are openly accessed by Google Earth Engine [30-34]. Satellite images generally contain multiple spectral bands with short revisit cycles but low resolution. It is worth noting that the satellite images need to be preprocessed to ensure quality before

being applied. Preprocessing includes geometric correction, radiation correction, and denoising. Among them, geometric correction is used to eliminate the errors caused by satellite position and imaging distortion of sensors. Radiation correction includes radiometric calibration, atmospheric correction, and topographic correction. Radiometric calibration is the radiometric correction for the sensors. Atmospheric correction is used to eliminate the influence of atmospheric absorption and scattering on radiative transfer. For the images of hilly and mountainous areas, topographic correction can eliminate the radiance error caused by the slope and the orientation. The cloud cover will affect the identification accuracy. Methods such as median filtering and homomorphic filtering could be used to eliminate the influence of a small number of thin clouds. If there are thick clouds in the image, a set of images with close time interval need to be cut and stitched to obtain a complete image without clouds [13]. Another method is to discard the local image where the clouds are located [35].

Higher resolution images can be obtained by performing aircraft aerial missions. Google Earth and Google Static Maps are the most used aerial image data sources in PV identification. They collect and provide global RGB images with a resolution of up to 0.15-0.3 m/px [36]. However, these images are charged. A thousand 640*640 pixels images require 1.6-2.0 USD [37]. When identifying the distribution of PV at the continent level or the global level, the charged fee for accessing high-resolution images in the full range is high. National Agriculture Imagery Program (NAIP)) provides 1 m/px image in the United States, including red, green, blue, and near infrared spectra [38]. Some studies show that this resolution is sufficient for building contour identification. Moreover, Spain, Germany, Switzerland, and other countries have their own high-resolution aerial images [39-41]. The update speed of the aerial image library is slower. Images of parts from the NAIP, for example, are updated every five years. In addition, Unmanned Aerial Vehicles (UAV) flying at low altitudes, allow their equipped cameras to shoot higher resolution images, with image resolution

being 0.1 m or even higher. However, due to the limited flight range and high time cost, UAV images are often used for PV identification on residential roofs in a small neighborhoods or damage detection of large centralized PV plants [42].

Table 2 Basic information about the image data sources for PV identification

Source	Coverage	Resolution	Description	Free or not
Landsat-8 [30]	Global	30 m/px or 15 m/px	11 spectral bands. The resolution is 30 m/px at the wavelength of 0.43 to 2.29 μm and 1.36 to 1.38 μm. The resolution is respectively 15 m and 100 m at the wavelength of 0.53-0.68 μm and 10.60-12.51 μm. It has a panchromatic band with a 15 m spatial resolution. The revisiting period is 16 days.	Yes
Sentinel-1 [31]	Global	10 m/px	C-band (5.407 GHz) SAR (Synthetic Aperture Radar) imagery. The VV and VH backscatter values can be used for PV at the acquisition mode of Interferometric Wide swath. The revisiting period is 6 days.	Yes
Sentinel-2 [32]	Global	10 m/px or lower	There are 13 spectral bands. The commonly used optical bands are visible light and near infrared with a total of 4 bands with a resolution of 10 m/px, and Red Edge and SWIR with a total of 6 bands with a resolution of 20 m/px. The revisiting period is 5 days.	Yes
SPOT [33]	Global	1.5 m/px or lower	Panchromatic images (0.455-0.745 μm) are supplied with a resolution of 1.5 m. Four spectral bands of images with a resolution of 6m, including red, green, blue, and near infrared (0.76-0.890) bands. The revisiting period is 26 days.	No
Google Earth & Earth Static Maps API [36]	Global	0.15 m/px to 15 m/px	Combines satellite and aerial imagery from multiple sources; much of the available imagery is between 1-3 years old.	High resolution images is yes
USGS [43]	The United States	0.15 m/px to 1 m/px	Ortho-rectified aerial imagery taken at different altitudes and different update cycles. Consolidation of multiple databases.	Yes
NAIP [38]	Continental US	1 m/px	Aerial orthophoto image. 4 spectral bands, including red, green, blue and near infrared. Images were taken during the agricultural growing season in the continental United States. The image is updated every five years.	No
Gaofen-2 [34]	Global	0.8 m/px-1 m/px or 4 m/px	Panchromatic spectral images with a resolution of 0.8-1 m/px and 4 spectral bands with a resolution of 4m are provided. The revisiting period is 64 days.	No
Digital Orthophotos (DOP) of Germany [39]	Germany	0.2 m/px	The product consists of georeferenced, differentially rectified aerial images of the surveying administrations of the German federal states. They are true-to-scale raster data of photographic images of the earth's surface. The product includes color images (RGB) as well as infrared images and color infrared (CIR) images.	No

PNOA: Plan Nacional de Ortofotografía Aérea [40]	Spain	0.5 m/px	The photograms were acquired in optimal meteorological conditions at a low flight altitude using calibrated photogrammetric cameras equipped with 3-band RGB sensors (8 bits per band). The imagery was orthorectified, radiometrically corrected, and has topographic corrections applied using ground points measured with accurate GPS systems	No
Swiss Federal Office of Topography [41]	Switzerland	0.25 m/px	Aerial photography of Switzerland. The application data are updated at the beginning of each quarter.	Yes

2.3 Training Dataset

A large database of images that have been annotated with PVs is needed in machine learning models or deep learning models as training data. The manual annotation of PV distribution on satellite images is labor-intensive and time-consuming. Therefore, the publicly available set of images with PV labels will be helpful for subsequent researchers.

Bradbury et al. [44] collected 601 satellite images of four California cities based on the USGS image set and manually annotated geospatial coordinates and border vertices for more than 19863 solar panels, which constitute a basic dataset for many subsequent studies. Jiang et al. [45] provided a dataset of images at 0.1 m/px, 0.3 m/px, and 0.8 m/px resolution for a portion of Jiangsu Province, China. These images are classified according to topographic distribution, with each containing a different area of PV, and PV mask images matching each image are also provided. The Amir dataset [46] consists of a set of aerial images collected from 12 countries and large PV plants located on six continents. Using OpenStreetMap as a platform and crowdsourced data from volunteers, Stowell et al. [18] have built more than 260,000 PV maps across the UK. Using Sentinel-2 satellite imagery at 10 m-60 m spatial resolution, Ortiz et al. [47] identified a total of 1076 large PV systems throughout India with available their latitude and longitude coordinates, installed areas, and polygon profiles. These datasets are publicly available and cover images from 0.1 m resolution to 10 m resolution, which can help to enrich the datasets of subsequent researchers, reduce costs and improve the accuracy of PV detection models. The above freely available data sets are described in Table 3.

Table 3 Basic information about public image database

Dataset	Location	Area	Resolution	Number	Annotation type
Bradbury et al [44]	4 cities of California	1352.25 km^2	0.3 m/px	601 images	Supply the corresponding PV mask images
Jiang et al [45]	Parts of Jiangsu Province, China	~1000 km^2	0.8 m/px	763 images	
		~218 km^2	0.3 m/px	2308 images	
		~0.4 km^2	0.1 m/px	645 images	
Costa et al [48]	24 centralized PV systems in Brazil	~722.5 km^2	10 m/px	144 images	
Amir dataset [46]	12 countries	/	/	3580 images	Supply orthographic images and the corresponding PV mask images
Stowell et al [18]	The whole of the U.K.	/	/	260,000 systems	Supply the latitude and longitude coordinates for each system. And supply the polygon shapes as geojson format
Ortiz et al [47]	The whole of India	/	/	1076 large PV systems	

3. Classification of PV Identification Models

The specific purposes of PV identification models can be categorized as image classification, object detection, semantic segmentation, and instance segmentation, as shown in Fig 1. Image classification classifies the images into positive samples and negative samples by judging whether the image contains PV or not. The distribution position and quantity of PV can be framed by object detection. Semantic segmentation can extract each pixel represented as PV in the image and then predict the area.

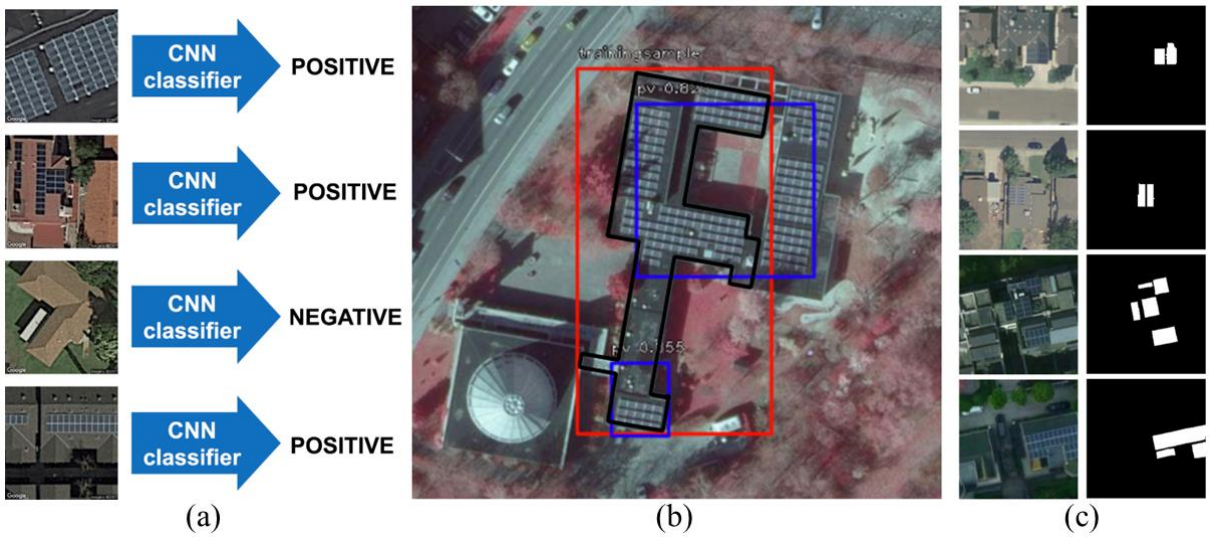

Fig 1 Different degrees of identification PV. (a) Image classification [23] (b) Object detection [49] (c) semantic segmentation [50]

In terms of the adopted methods, PV identification models can be divided into the conventional machine learning method and the deep learning method. The conventional machine learning methods can be further divided into Pixel-Based Image Analysis (PBIA) and Object-Based Image Analysis (OBIA).

3.1 Pixel-Based Image Analysis (PBIA)

PBIA is a method that extracts the features according to the spectral information of every PV pixel and classifies the unknown pixels by using the features. Most GIS software can provide some PBIA classifier toolkits, including maximum likelihood, iso cluster, principal components analysis, Support Vector Machine (SVM), and Random Forest. Some studies have used some of these methods [51]. Table 4 summarizes the basic information for studies using PBIA. The classification methods include statistical recognition and multi-classifier fusion.

Table 4 Basic information for studies using PBIA

Reference	Image Resolution	Location	PV scale	Type of the identification	Method	Result metrics
Ji et al [52]	1.2 m	Oldenburg,Germany	Rooftop PV	Object Detection	Mathematical statistic classification	overall accuracy is from 92.8% to 99.3%
Karoui et al [53]	0.84-1.6 m	a part of Toulouse, France	Rooftop PV	Object Detection	Multi part nonnegative matrix factorization	Normalized Mean Square Error of PV=23.73%
Malof et al [54]	0.3 m	the city of Fresno	Rooftop PV	Object Detection	Random Forest	precision=60%; recall=70%
Xia et al [35]	10 m	China	Centralized water PV	Semantic Segmentation	Random Forest	SPV: precision=96.9% recall=84.4%; FPV: precision=100%; recall=96.2%

Czirjak et al. [55] tested the spectrum distribution of polycrystalline silicon PV panels and found that their reflectance in the visible band (400-700nm) was low on the whole test spectral band ranging from 400 to 2500 nm, but increased sharply between 0.950-1.175 μm. They successfully detected PV in hyperspectral images using reference spectral distribution features. The results show that the PV quantity accuracy is high, but the shape and area are not accurate. Ji et al. [52] detected several hydrocarbon surface materials including PV in the image by the hydrocarbon index. Furthermore, by using 5 spectral indices, PV pixels can be screened out separately.

A single pixel may contain spectral information from several objects when the image resolution is low. Karoui et al. [53] designed an iterative multiplicative gradient-based algorithm to calculate the PV abundance of boundary pixels. Compared to the traditional nonnegative matrix factorization algorithm, it is more suitable for application in urban areas with uneven landscapes or irradiance.

The above three studies used the classification of statistical recognition. The classification of multi-classifier fusion refers to the multiple integrations of a single classifier. Random forest classifier is generally

used, has better accuracy, and can effectively prevent overfitting. Malof et al. [54] extracted 6 features of each pixel, including the mean and variance of the RGB channels of its window pixels. Each window is 3 by 3 pixels. The random forest method is used to assign a confidence value to each pixel point. Using random forest, Xia et al. [35] accurately distinguished centralized water PV, water surface and wetland by using the short infrared band of 1.610 μm, VH band information and Normalized Difference Vegetation Index. The segmentation results are shown in Fig 2.

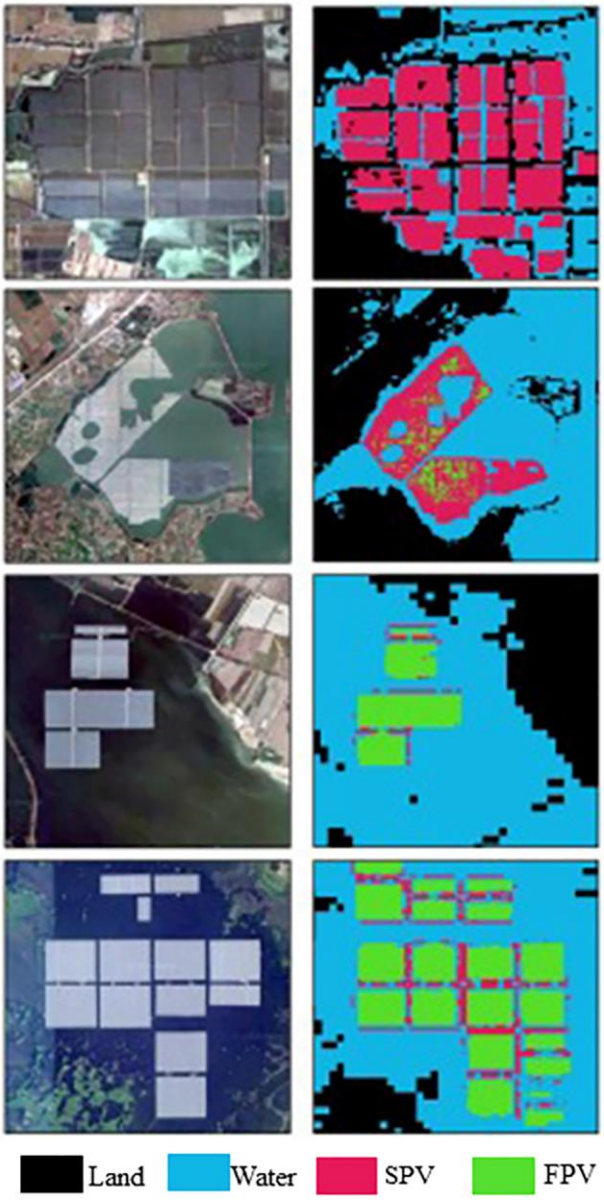

Fig 2 Results of segmentation water PV by PBIA [35]

3.2 Object- Based Image Analysis (OBIA)

Compared with PBIA, OBIA adds a step in the early stage, i.e. the generation of candidate regions. Firstly, the image is divided into several homogeneous objects. After extracting the typical feature of each object, different models are used to classify objects of unknown categories. There are two core problems with this method. One is the selection of the segmentation algorithm and optimal segmentation size, the other is to select appropriate features to establish the rule set. In addition to the spectral features of every single pixel, the geometric, texture features and other spatial features can be extracted. SVM and Random Forest are commonly used regional classifiers.

Malof et al. [56] extracted the mean RGB value and a multivariate normal probability distribution with mean and covariance parameters of each region in the image. Then, SVM was trained to classify the features. When detecting centralized PV systems in Ningxia, China, Zhang et al. [57] found that all accuracy and Kappa of detection increased by 0.87% and 3.4%, after 8 texture features were added, such as entropy, correlation and inverse difference moment, respectively, compared to PBIA. Plakman et al. [58] selected the mean, standard deviation, and median of each band in Sentinel-2 and Sentinel-1 as features to train a Random Forest classifier, as well as Short-Wave Infrared (SWIR), Red Edge, Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI). Also using Sentinel images, Vasku et al. [59] compared the performance of several classifiers. The results showed that the Random Forest and Classification and Regression Tree were the best. SVM was found to have more misjudgment after observing the image result.

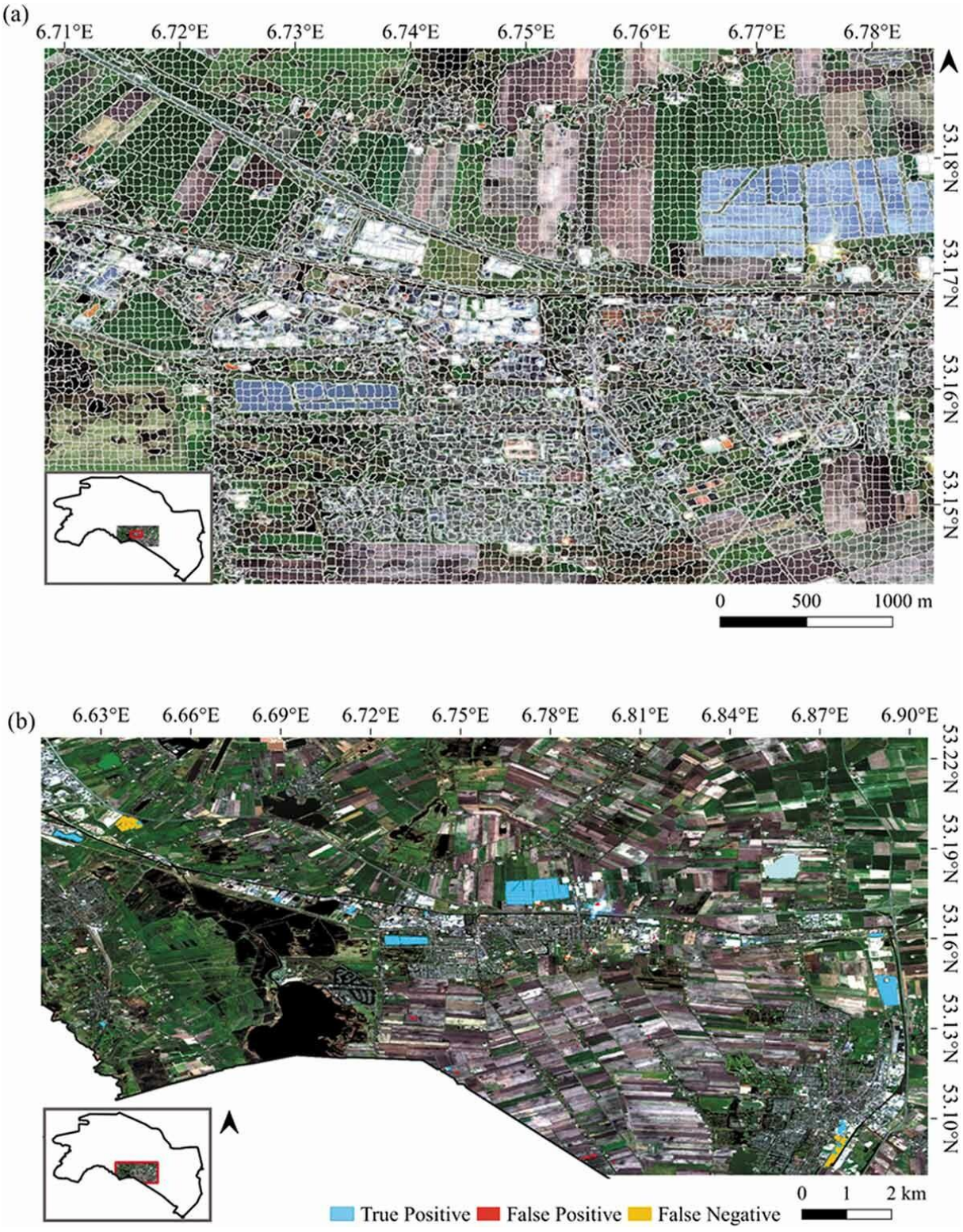

Fig 3 The results of two steps of OBIA. (a) several homogeneous objects generation (b) Detection PV [58]

In addition, the template matching technique is also used for PV identification. A small number of panels in the PV array can be used to create templates to identify the location and size of other panels in the same

array [60]. It is suitable for very-high-resolution images.

The basic information of above studies is shown in Table 5.

Table 5 Basic information for researches using OBIA

Reference	Image Resolution	Location	PV scale	Type of the identification	Method	Result metrics
Malof et al [56]	0.3 m	100 building images	Rooftop PV	Object Detection	SVM	Recall= 94%
Zhang et al [57]	30 m	Ninxia Province,China	Centralized PV (> 0.21 km^2)	Object Detection	Random Forest	precision=98.53%; recall=92.19%
Plakman et al [58]	10-20 m	Netherlands	Centralized PV (>1034 m^2)	Semantic Segmentation	Random Forest	For object: recall=85.86%; precision=92.39%; IoU=80.19%; F=99.97%
Vasku et al [59]	10 m	Denmark	Centralized PV (>12 km^2)	Semantic Segmentation	Classification and Regression Tree	precision=70.73%, recall=52.94%
					Random Forest	precision=71.60%, recall=54.90%
					SVM	precision=71.60%, recall=52.94%
					Naïve Bayes	precision=63.64%, recall=37.25%
Wang et al [60]	0.05 m	Drone photos	Centralized PV	Semantic Segmentation	region–line primitive association analysis and template matching	precision and recall measures are both higher than 0.99.

3.3 Methods Based on Deep Learning

In recent years, methods based on deep learning have become very popular in image identification. Among them, the convolutional neural network (CNN) is the most common method. The most basic CNN is a classified network. The most basic CNN network includes the convolution layer, pooling layer, activation function, and full connection layer. The convolution layer uses the convolution kernels to extract image information, and the values in the convolution kernel require to be trained. The activation function can make the process of information extraction non-linear. The pooling layer can integrate a certain range of information, enlarge the receptive field, and reduce the size of the image. The full connection layer is used to establish a linear relationship between the output data of the penultimate layer and the actual category. When the model is used for prediction, the output of the full connection layer is the probability predicted for

each category. Several common CNN classification networks have been proposed and improved, including AlexNet, VGG series, ResNet series, Inception series, etc. As the demand increases, object detection and semantic segmentation models are proposed. They have more complex architectures and often use the aforementioned classification networks as backbones [61].

3.3.1 Image Classification

The basic information of the investigations on image classification using CNN is summarized in Table 6. Malof et al. [62] firstly applied CNN to PV identification task. They created a CNN network containing 6 convolutional layers and 3 max-pool layers to filter false positive samples in the detection results of PBIA method in reference [54]. Thereafter, they created a CNN with VGG as the backbone [63]. They segmented satellite images into a large number of small size images, and used the CNN classifier to classify each small image. Using a larger low-resolution image set, Golovko et al. [64] trained a model containing 3 convolutional layers and 3 fully connected layers, achieving 89.1% precision of classification. Moraguez et al. [65] created a CNN containing 8 convolutional layers to analyze the influence of image resolution, and the results showed that the change of image resolution between 0.3 m and 0.6 m would not have an obvious influence on the accuracy of image classification.

Table 6 Basic information for researches of image classification using deep learning models

Reference	Image Resolution	Location	PV scale	Architecture	Backbone	Result metrics
Malof et al [62]	0.3 m	the city of Lemoore, CA	Rooftop PV	RF+CNN	/	precision=72%, recall=80%
Malof et al [63]	0.3 m	the city of Lemoore, CA	Rooftop PV	CNN with 3 VGG modules, 2 fully connected blocks	VGG16	The precision is about 95% when the recall is 80%
Golovko et al [64]	low-quality images	/	Rooftop PV	CNN with 3 convolutional layers	/	The precision = 89.1%; recall= 83.3%
Moraguez et al [65]	0.3 m	Riverside, CA and Redlands, CA	Rooftop PV	CNN with 8 convolutional layers	/	precision=91.1%; recall=82.9%
	0.4 m				/	precision=91.9%; recall=92.4%; F1=92.3%
	0.6 m				/	precision=91.3%; recall=90.0%
	1.25 m				/	precision=73.8%; recall=75.2%

3.3.2 Object Detection

Object detection can frame the PV in positive images. The trained model is used to screen and correct the position and size of the anchor frame, so as to achieve the best match to the object. Most object detection frameworks are either two-state networks or one-state networks. Generally, one-state networks include the Feature pyramid Networks (FPN), YOLO series, SSD, and RetianNet [66]. Two-state networks, including R-CNN, Faster RCNN, etc., are also known as candidate region-based methods, which can be applied to both object detection and semantic segmentation.

The features from the shallow convolutional layers lack sufficient semantic information but have a lot of location information. The features from deep convolution are low resolution but have rich semantic information. Therefore, feature fusion enables detection networks to obtain both high semantic information and precise location information. For rooftop PV, Yuan et al. [67] built a CNN consisting of 7 stages. each of which consists of a convolutional layer and optionally a max-pooling layer. Different from regular CNN, there is an integration stage upsampling feature maps from four stages and generating a feature stack. The

network uses early stages to capture edges and corners. Although it used feature fusion, the feature maps from each layer are simply connected, and the detection is still from the last layer of the backbone network.

Feature pyramid network (FPN), a classic and effective feature fusion network, is used as a neck network by many object detection models. It includes two branches. The bottom-up branch is used to generate multi-scale features, and the top-down branch is used to transfer rich semantic information from the high level to the bottom level. The two highest layers are fused after passing through a convolution layer respectively to generate a new feature map. The new feature map is fused with the convolution result from the third highest layer to generate a second new feature map. Before each fusion, the smaller image should be enlarged by up-sampling. And so on, multiple fusion feature maps with different scales are generated, and they could be used to detect objects of different scales. Based on this idea, Ioannou et al. [68] created a CNN model consisting of three convolution layers and two up-sampling layers in the detection task of the centralized PV. In the upper sampling layer, feature maps of the same scale are fused to transfer information from the shallow layer to the deep layer, as shown in Fig 4. In addition, they found that blindly increasing the training time of epoch cannot significantly improve prediction performance, but lead to over-fitting. For the one-stage detection network, RetinaNet is essentially an FPN with the traditional cross-entropy loss being replaced by a new Focal Loss. Kleebauer et al. [49] firstly trained a RetinaNet with ResNet101 as the backbone. By performing the detection task in Hesse State, a second RetinaNet was retrained with the same structure and parameters using true positive bounding boxes as input. The second RetinaNet can be used for PV detection in other regions of Germany with high accuracy and robustness.

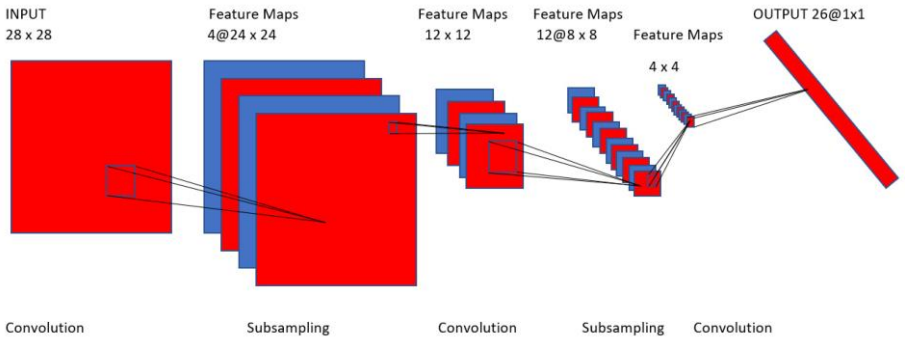

Fig 4 Network architecture for fusion features from up-sampling layers [68]

As for two-state networks, Golovko et al. [69] used Faster-RCNN to detect rooftop PV, where the model used FPN as a neck. However, objects of small size in the image, such as the top of the chimney, are also identified as small PV. Combined with the characteristics of two-state networks, it can be seen that the candidate region method adopts a large number of down-sampling coding methods and does not fully consider the global information, resulting in unsatisfactory detection effects for the distributed PV systems with a small area.

SolarFinder combines machine learning models with CNN [70]. It calculates eight features from the color, texture, and shape of each region on RGB images. Then SVM is used to identify the regions that may be rooftop PV, and CNN is used to classify these possible regions. Its performance is better than that of pure SVM and pure CNN. Therefore, combining OBIA with deep learning may be a way to further improve the performance of deep learning.

The basic information of the above studies is summarized in Table 7.

Table 7 Basic information for researches of object detection using deep learning models

Reference	Image Resolution	Location	PV scale	Architecture	Backbone	Result metrics
Yuan et al [67]	0.3 m	Washington, San Francisco and Boston	All kinds	CNN with 7 convolution layers	/	San Francisco: precision=85.5%; recall=87.3% Boston: precision=81.2%, recall=84%
Ioannou et al [68]	0.3-0.8 m	Greece	Centralized PV	CNN	Inception v3	precision=53%, recall=52%; F1-Score=52%
Kleebauer et al [49]	0.2 m	Hesse, Germany	Rooftop PV	RetinaNet+FPN	ResNet101	precision=92.77%, recall=84.47%
Golovko et al [69]	low-quality images	/	Rooftop PV	Faster-RCNN	ResNet50	precision=92.99%
Li et al [70]	low-quality images	13 geospatial regions in U.S.	Rooftop PV	OBIA-pure thresholding	/	Matthews Correlation Coefficient (MCC)=0.06
				PBIA-pure SVMs	/	MCC=0.25
				Pure CNNs	/	MCC=0.17
				SVM+CNN	/	MCC=0.31

3.3.3 Semantic Segmentation

Semantic segmentation can extract PV region at the pixel level. The area and capacity of PV can also be predicted by counting the number of pixels. The common semantic segmentation networks for PV can be divided into several types in terms of structure and training methods, as shown in Fig 5.

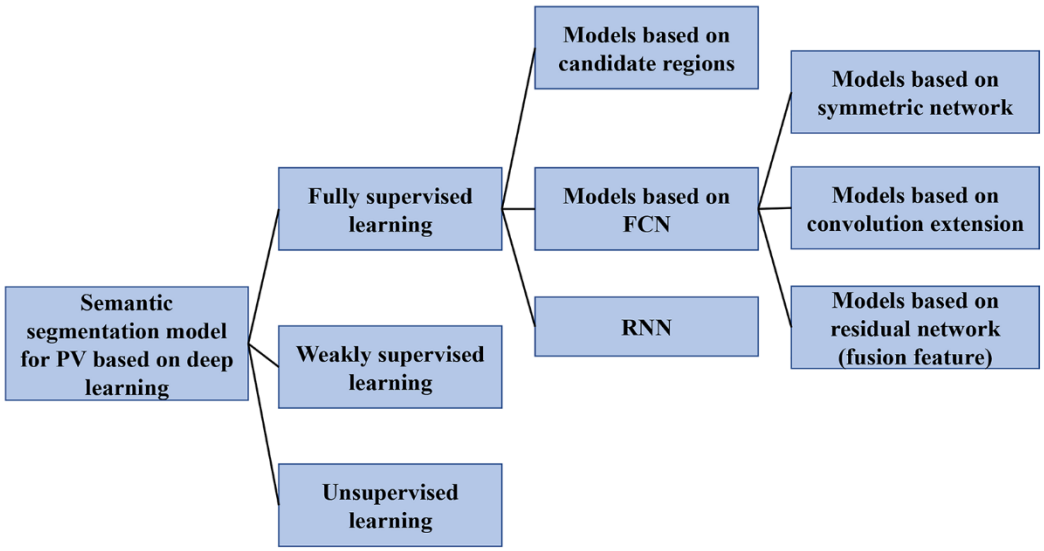

Fig 5 Classification of semantic segmentation models of deep learning for PV

3.3.3.1 Candidate Region-based Method

The candidate region-based object detection method mentioned in the above section is originally designed for detection, but in the process of development, it has the ability of semantic segmentation and even instance segmentation. In addition to the candidate region-based models mentioned in the previous section, Mask-RCNN is originally an instance segmentation network and has also been used to segment PV. The performance of segmentation PV using Mask R-CNN is shown in Table 8. For centralized PV systems, Sizkouhi et al. [71] established a Mask-RCNN with modified VGG16 as the backbone. All fully connected layers in VGG16 are removed, only two last layers are finetuned, and the other hidden layers are frozen. Schulz et al. [72] adapted the Mask-RCNN in combination with a multi-class labeling technique to build the DetEEktor. They chose ResNet101+FPN+RPN as a feature extractor. It can simultaneously identify roof-mounted PV systems, free-field PV systems, roof-mounted solar thermal systems, free-field solar thermal systems, biomass plants, and wind power plants. Liang et al. [73] added an overs-tile strategy and right-angle polygon fit algorithm to Mask-RCNN, and the performance was further improved.

Table 8 Basic information for segmentation researches using candidate region-based methods

Reference	Image Resolution	Location	PV scale	Architecture	Backbone	Result metrics
Sizkouhi et al [71]	/	12 countries	Centralized PV	Mask R-CNN	VGG16	accuracy=96.93%
Schulz et al [72]	/	BadBrambach and Chemnitz, Germany	All kinds	Mask RCNN	ResNet101 + FPN	Rooftop PV: precision=81%; recall=73%; F1-score =97%; IoU =77% Distributed PV: precision =97%; recall72%; F1-score83%; IoU=86%
Liang et al [73]	0.15 m	The United States	All kinds	Mask R-CNN	ResNet50+FPN	precision=95.5%; recall=92.1%; IoU=86.5%
				Mask R-CNN+ overlap-title	ResNet50+FPN	precision=95.7%; recall=93.5%; IoU=87.4%;
				Mask R-CNN+ overlap-title+ right-angle polygon fit algorithm	ResNet50+FPN	precision=96.2%; recall=95.5%; IoU=88.8%
				The method of DeepSolar	ResNet50+FPN	precision=93.1%; recall=88.5%

3.3.3.2 FCN-based Method

FCN is implemented based on the CNN by replacing the last fully connected layers with convolutional layers. Moreover, FCN generates a feature map with the same size as the original image through up-sampling. Compared with CNN, FCN can take images of any size as input, and can directly achieve end-to-end pixel-level prediction. To take advantage of multispectral images, Ishii et al. [74] first created an FCN with seven spectral input channels. The model is effective in detecting large PV systems in Japan. On the other hand, when predicting each pixel, multiple convolutional layers and pooling layers are applied to the original image for expanding the receptive field and extracting more information. However, the pooling layer will reduce the image size and resolution. After the prediction of the small--size feature map, the upsampling methods of the nearest neighbor or linear interpolation are applied to resize the predicted map to its original size. Such prediction results are fuzzy and smooth, and many details will be lost. Therefore, FCN is not suitable for the

segmentation of small objects. After that, there are more and more semantic segmentation models derived from FCN, to reduce the amount of detail lost by enlarging the receptive field.

The symmetric semantic segmentation models with encoder-decoder architecture are proposed. Encoders usually use CNN and down-sampling to reduce resolution and extract image feature mapping. The decoder converts low-resolution images and features into image segmentation and maps to pixel-level prediction. Common networks include DeconvNet, SegNet, U-Net, LinkNet, etc.

SegNet records the pooling index calculated in the maximum pooling layers of the encoder, and they are applied to upsample the feature map of the corresponding size in the decoder. This allows more accurate upsampling results. Edun et al. [75] adopted a modified VGG16 in the encoder to modify the input size and remove dense layers. Camilo et al. [76] changed the output of SegNet to 2-way Softmax. They form PV segmentation masks by classifying densely extracted image patches. The results showed that SegNet performed much better than VGG. When the Conditional Random Field, which is a postprocessing step, is combined with FCN to form CRFasRNN, the segmentation performance is still inferior to SegNet [22]. Other optimization methods were also tried, e.g., the signed Distance function (Dis) of boundaries. The Dis adopts a method of pixel labeling where the PV and non-PV areas of the image are represented by the distance between the pixels and the PV boundary points. The positive and negative numbers indicate the pixels inside and outside the PV area respectively. This labeling method can improve the imbalance of the number of pixels in different categories of samples. The results show that the precision of SegNet-Dis is 9.6% higher than that of SegNet-Bin, which adopts the Binary function as the labeling method. Furthermore, the accuracy of the model can be further improved by 1.2% after the fusion of near infrared spectra. House et al. [77] combined SegNet with the CNN classification network. The former is used to segment building roofs, and the latter can determine whether these roofs are PV-installed. This idea can be used to calculate the percentage

of area occupied by PV in the installed PV roof, which is important for estimating the PV potential of the roof.

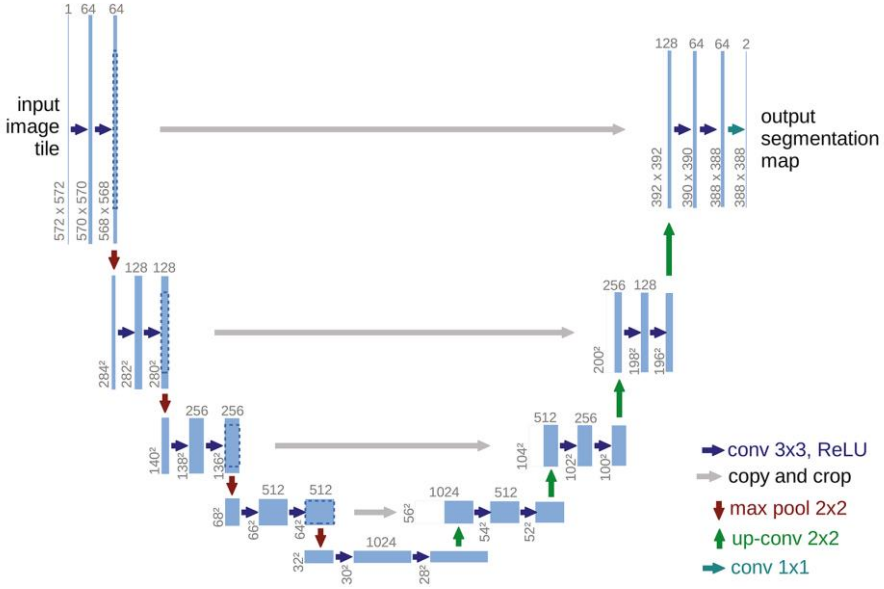

Fig 6 U-Net architecture [78]

The structure of the U-Net proposed firstly is shown in Fig 6. It includes the contracting path of up-sampling and the expanding path of down-sampling. Compared with SegNet, U-Net fuses the feature map to be upsampled with the feature map of the same depth and size in the encoder through skip layers. In this way, feature fusion is achieved, and its idea is similar to FPN in object detection. By the skip layer, the upsampled feature map contains more low-level semantic information. U-Net can achieve higher segmentation accuracy, of course, it also has more hyperparameters in the decoder than SegNet. By comparing precision, recall, and IoU in large-scale centralized PV power plants, U-Net is proved to be better than FCN [79]. U-Net is widely used to identify PV of all sizes in images of all resolutions, including large centralized systems and rooftop distributed systems [13, 80-82]. However, when identifying home systems with less than 100W capacity, the accuracy of the network will be significantly reduced. Jie et al. [50] introduced a Gated fusion module and an edge detection network to the U-Net to solve this problem. Some other optimization models based on U-

Net are proposed. The multitask collision-maximization Attention Network (EMANet) is an unsupervised clustering algorithm without convolution operation, which can efficiently capture the information of the most interesting region in the whole image. Hou et al. [83] combined U-Net with EMANet to improve the IoU of centralized PV systems segmentation. An adaptive cross learning driven U-Net method (Adaptive CrossNets) is developed to utilize cross learning, each generic U-Net in the community first updates the parameters individually on each epoch, and then learns the parameters from the optimal individual on a particular epoch [84]. The generic U-Net that performs best in the community is selected as the final model for CrossNet. Most U-Net segmentation models use ResNet as the backbone. Zech et al. [85] respectively selected ResNet18, ResNet34, ResNet50, and ResNet101 as the backbone to segment roof PV in Oldenburg of Germany. Jie et al. [86] also created a model for identifying large PV power stations, adding three modules between the encoder and decoder. The modules embrace Feature Refinement Residual Module (FRRM), Chained Dilation Attention Module (CDAM), and Global Channel Attention Module (GCAM). FRRM can refine the characteristics of each stage of the encoder, rather than directly connecting them to the corresponding decoder block. The CDAM expands the receiver range and selects useful features without reducing resolution. The basic information of the above studies is summarized in Table 9.

Table 9 Basic information for researches of semantic segmentation using FCN and FCN-based method

Reference	Image Resolution	Location	PV scale	Architecture	Backbone	Result metrics
Ishii et al [74]	30 m	Japan	Centralized PV (>5 MW)	FCN		For object: IoU>50%; precision>60%; recall>75%
Edun et al [75]	0.596 m	The U.S. 387 locations, 669 PV arrays	Centralized PV and distributed PV	encoder–decoder architecture	VGG16	accuracy=98.66%, dice coefficient=81.74%
Camilo et al [76]	0.3 m	the city of Fresno	Rooftop PV	CNN	VGG	When recall=80%, precision is about 50%
				SegNet	VGG	When recall=80%, precision is about 90%
Yang et al [22]	1 m	9 citys	Buildings	FCN-4s-Bin	VGG16	precision=31.0%
				FCN-8s-Bin		precision=34.5%
				FCN-4s-CRFasRNN-Bin		precision=29.8%
				FCN-8s-CRFasRNN-Bin		precision=35.4%
				SegNet-Bin		precision=74.1%
				SegNet-Dis		precision=83.7%
				SegNet-Bin-Fused		precision=74.2%
				SegNet-Dist-Fused		precision=84.9%
House et al [77]	/	An area of about 0.01 km2	Rooftop PV	SegNet +CNN	VGG16	Precision=98.9%; recall=95.8%
González et al [79]	/	12 countries	Centralized PV	FCN		recall=94.16%; IoU=87.47%; Dice coef=89.61%
				U-Net	/	recall=95.44%; IoU=90.42%; Dice coef=91.42%
Kruitwagen et al [13]	1.5 m&10 m	Global	Centralized PV (larger than 10000 m^2)	U-Net +ResNet50 / RNN U-Net is used for semantic segmentation. ResNet and RNN is image classifier for filtering false positives	/	For PV with an area greater than 10000W: recall=90%; IoU=90%; precision=98.6%
Wu et al [80]	0.5 m	17 cities around the world	Rooftop PV	U-Net: The coding layer of U-Net is replaced with a pre-trained Resnet50.	ResNet50	Count recall=91.90%; Area recall=96.25%
Ren et al [81]	0.03 m	Rwanda	solar home systems (<100 W)	U-Net	ResNet-50	For object: the maximum value of F1-Score is 79%. When the recall is 89%, the precision is 41%.
Castello et al [82]	0.25 m	Swiss	Rooftop PV	U-Net		IoU=64%; accuracy=94%; F1-Score=80%

Jie et al [50]	2 m	7 provinces of China	Centralized PV	LinkNet	ResNet34	mIoU=81.17%; mean precision=87.60%; mean recall=90.71%
				U-Net		mIoU=82.21%; mean precision=87.36%; mean recall=92.06%
				U-Net+FRRM+CDAM+GCAM		mIoU=87.39%; mean precision=93.76%; mean recall=92.18%
Hou et al [83]	/	China	Centralized PV	U-Net	ResNet101	mIoU=86.54%
				EmaNet		mIoU=93.79%
				U-Net and EmaNet		mIoU=93.94%
Zhuang et al [84]	0.3 m	Fresno, Oxnard, Stockton	Rooftop PV	U-Net with trasfer learning		mIoU=72.792%, Variance=1.286e-4
				U-Net without trasfer learning		mIoU=40.017%, Variance=1.191e-2
				CrossNets (a cross-learning driven U-Net method)		mIoU=74.268%, Variance=2.481e-5
				Adaptive CrossNets		mIoU=74.279%, Variance=1.458e-5
Zech et al [85]	0.2 m	Oldenburg, Germany	Rooftop PV	U-Net	ResNet18	IoU=65%; precision=83%, recall=76%; F1-Score=79%
					ResNet34	IoU=68%; precision=84%, recall=77%; F1-Score=80%
					ResNet50	IoU=69%; precision=84%, recall=79%; F1-Score=81%
					ResNet101	IoU=68%; precision=86%, recall=76%; F1-Score=81%
Jie et al [86]	0.3 m	Fresno, Stockton, Modesto	Rooftop PV	SegNet	Eff-b1	IoU=66.97%; precision=83.48%, recall=77.20%; F1-Score=80.22%
				LinkNet		IoU=69.23%; precision=83.60%, recall=80.11%; F1-Score=81.82%
				U-Net		IoU=70.28%; precision=83.83, recall=81.30%; F1-Score=82.54%
				FPN		IoU=71.11%; precision=84.79%, recall=81.50%; F1-Score=83.11%
				U-Net+GFM+EDN		IoU=73.60%; precision=86.17%, recall=83.45%; F1-Score=84.79%

LinkNet has a U-shaped structure as well, but with fewer parameters. It uses a residual block to replace the convolution layer and directly connects the encoder and the decoder to improve accuracy. In comparison, the advantages of SegNet and LinkNet lie in simple structure and fast calculation speed, but it is not as accurate as U-Net.

Meanwhile, the FCN-based dilation convolution semantic segmentation model is also employed, such as Deeplab v3 and DeepLab v3+ [87]. They refer to the idea of the skip layer in U-Net and applied the Atrous Spatial Pyramid Pooling (ASPP) module for feature fusion. At the same time, they apply the dilated convolution layers, which reduces the use of pooling layers, expands the receptive field, and avoids the size reduction of the feature map. RefineNet and PSPNet belong to the FCN-based residual networks. Costa et al. [48] compared the performance of U-Net, DeepLab v3+, PSPNet, and FPN models on segmenting large PV systems, and matched different backbones. The results show that U-net presents better segmentation accuracy with the optimal backbone of Efficient-net-b7. While PSPNet has the shortest training period. Jiang et al. [45] compared the segmentation capabilities of U-Net, RefineNet, and DeepLab v3+ for various PV systems using images with three resolutions. The results show that DeepLab v3+ obtains the highest accuracy and precision, but the lowest recall on three datasets. This indicates that DeepLab v3+ tends to guarantee that the extracted PVs are accurate, while RefineNet and U-Net attempt to identify as many PVs as possible. Meanwhile, DeepLab v3+ is superior to U-Net for the identification of roof PV with high precision. A comparison of the above architectures is shown in Table 10.

Table 10 Comparison of researches using FCN-based method

Reference	Image Resolution	Location	PV scale	Architecture	Backbone	Result metrics
Costa et al [48]	10 m	Brazil	Centralized PV	U-Net	Eff-b7	Overall accuracy=98.08%; IoU=91.17%; F-Score=95.38%
					Eff-b0	Overall accuracy =98.05%; IoU=90.97%; F-Score=95.27%
					ResNet101	Overall accuracy =97.96%; IoU=90.58%; F-Score=95.06%
					ResNet50	Overall accuracy =97.98%; IoU=90.70%; F-Score=95.12%
				DeepLab v3+	Eff-b7	Overall accuracy =97.83%; IoU=89.98%; F-Score=94.73%
					Eff-b0	Overall accuracy =97.77%; IoU=89.82%; F-Score=96.64%
					ResNet101	Overall accuracy =97.46%; IoU=88.47%; F-Score=93.88%
					ResNet50	Overall accuracy =97.02%; IoU=86.63%; F-Score=92.84%
				PSPNet	Eff-b7	Overall accuracy =97.35%; IoU=88.03%; F-Score=93.64%
					Eff-b0	Overall accuracy =96.73%; IoU=85.43%; F-Score=92.14%
					ResNet101	Overall accuracy =97.06%; IoU=86.98%; F-Score=93.04%
					ResNet50	Overall accuracy =97.23%; IoU=87.60%; F-Score=93.39%
				FPN	Eff-b7	Overall accuracy =97.38%; IoU=87.99%; F-Score=93.61%
					Eff-b0	Overall accuracy =97.45%; IoU=88.21%; F-Score=93.73%
					ResNet101	Overall accuracy =97.58%; IoU=89.21%; F-Score=94.30%
					ResNet50	Overall accuracy =97.25%; IoU=87.74%; F-Score=93.47%
Jiang et al [45]	0.8 m	Parts of Jiangsu province, China	centralized PV and ground distributed PV	U-Net	/	IoU=77.6%; precision=87.1%; recall=86.4%; F1-Score=86.8%
				RefineNet	/	IoU=77.3%; precision=84.8%; recall=88.4%; F1-Score=86.6%
				DeepLab v3+	/	IoU=79.0%; precision=87.7%; recall=85.7%; F1-Score=79.0%
	0.3 m		All kinds	U-Net	/	IoU=85.8%; precision=89.7%; recall=93.5%; F1-Score=91.6%
				RefineNet	/	IoU=87.8%; precision=95.7%; recall=93.7%; F1-Score=94.7%
				DeepLab v3+	/	IoU=90.8%; precision=95.9%; recall=93.1%; F1-Score=94.5%
	0.1 m	Haian county, China	Rooftop PV	U-Net	/	IoU=78.7%; precision=78.7%; recall=90.0%; F1-Score=86.4%
				RefineNet	/	IoU=85.9%; precision=90.9%; recall=89.7%; F1-Score=90.3%
				DeepLab v3+	/	IoU=86.8%; precision=92.8%; recall=89.4%; F1-Score=91.1%

3.3.3.3 Weakly Supervised and semi-supervised Learning

The various types of deep learning methods mentioned above are all fully supervised learning methods, whose training sets require high-quality pixel-level labels. However, it is difficult to obtain and costly. Weakly supervision requires relatively coarse-grained labeling information with lower labeling cost. Therefore, it is easier to obtain more training sets, which enables PV identification over a wider range. The basic information of the weakly supervised learning models for PV is shown in Table 11.

CAM (Class Activation Mapping) is a method for generating weakly supervised labels. CAM replaces the fully connected layer of the classification network with a global average pooling (GAP) layer to obtain the value of each feature graph. Following a simple weighting calculation, a coarser and pixel-level region of different categories can be obtained. In some studies, the CAM results were directly used as the detection results. Imamoglu et al. [88] trained a CNN network with a feedback structure (FB-Net) for image classification and the model has two feedback paths. In addition, they input the features to the multi-channel pulse-coupled neural network (m-PCNN), to obtain the CAM of the centralized PV systems and detect the area growth over the years. Then, since more data sets were available, Kouyama et al. [89] adopted the simpler Grad-CAM with consistent accuracy and replaced the convolutional block in FB-Net with ResNet to improve the identification accuracy of the PV boundaries.

Yu et al. [23] established DeepSolar by employing the Greedy layer-wise training approach. They added a branch of image classification on a classification network to greedily extract features at a low-level hierarchy. Later, a new "convolutional layer-gap-linear classifier" structure was added at the end of the last convolutional layer on the branch to train the segmentation capability. The classification framework and the greedy layer-wise training are shown in Fig 7 The results show that CAM which can be used directly as segmentation results can be obtained via this method. Zhang et al. [90] thought that the results of CAM were rough and incomplete, and further took CAM generated from the fourth convolution layer in a classification model as the input, to train a pseudo supervised deep convolutional network with label correction strategy (PS-CNNLC) for semantic segmentation. The labels were updated by comparing the output with the input. When the output simultaneously satisfies three criteria, the iteration ends and the result is exported. They further developed optimization methods, such as using uncertainty estimation and forward correction to reduce the noise of pseudo-labels, or proposing a confidence-aware loss function to reduce the contribution

of low-quality pseudo-labels to the training segmentation networks [91, 92].

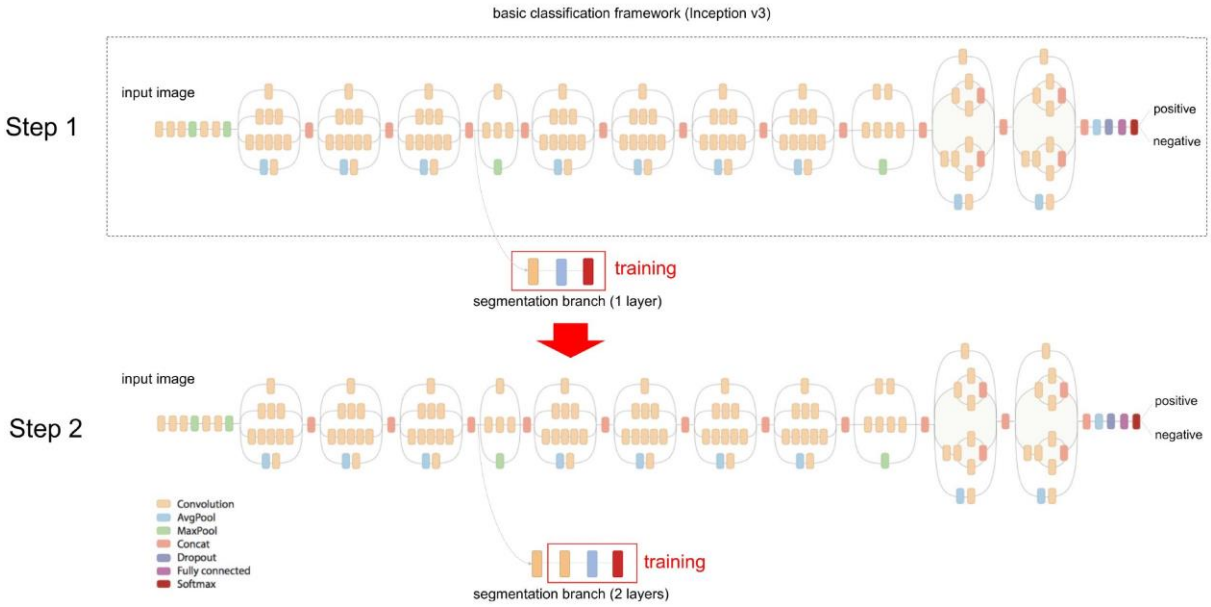

Fig 7 Classification framework and greedy layer-wise training for segmentation [23]

Similar to the weakly supervised learning, a semi-supervised learning method has been applied. It projects the point coordinates of PV with known geographic locations into satellite images, and then the pixels with similar spectral characteristics near the point are clustered into a PV system to generate the weakly labels. The false positive regions in the first result are labeled as negative (Hard Negative Mining, HNM) for a second training [47].

Table 11 Basic information of researches using weakly supervised and unsupervised learning models

Reference	Image Resolution	Location	PV scale	Type of the identification	Architecture	Backbone	Result metrics
Imamoglu et al [88]	30 m	Japan	Centralized PV (>5 MW)	Detection	Feedback CNN (FB-Net) +m-PCNN	/	For object: the precision is 83.81%, IoU is 0.56
Kouyama et al [89]	30 m	Japan	Centralized PV (> 5MW)	Detection	FB-Net+Grad CAM	/	/
Yu et al [23]	0.15 m	The United States	All kinds	Segmentation	Cassification: Inception v3, including 155 convolution layers Segmentation: the greedy layer-wise training	/	Image Cassification: In residential areas: precision=93.1%; recall= 88.5% in non-residential areas: precision=93.7%; recall=90.5% Semantic Segmentation mean relative error (MRE)=3.0% for residential areas; 2.1% for non-residential areas; Mean absolute percentage error (MAPE)=24.6%
Mayer et al [93]	0.1 m	Germany	All kinds	Segmentation	Cassification: Inception v3 Segmentation: DeepLab v3	Resnet-101	Image Cassification precision=93%; recall= 93% Semantic Segmentation Mean absolute percentage error (MAPE)=18.5%; mIoU=74.1%; mean relative error (MRE)=3.9%
Zhang et al [90]	0.3 m	Canberra, Australia	Rooftop PV	Semantic Segmentation	PS-CNNLC	/	precision=92.63%; recall=82.45%
					PS-CNN	/	precision=83.29%; recall=89.38%
					The method of DeepSolar	/	precision=67.67%; recall=77.93%
Zhang et al [91]	0.15-0.3 m	Canberra, Australia	Rooftop PV	Segmentation	effective uncertainty-aware forward correction method	/	precision=79.29%; recall=76.21%; F1=77.72%; IoU=63.56%
					PS-CNNLC	/	precision=65.68%; recall=83.15%; F1=73.39%; IoU=57.96%
					The method of DeepSolar	/	precision=65.17%; recall=60.23%; F1=62.60%; IoU=63.56%
Zhang et al [92]	0.15-0.4 m	Canberra, Australia	Rooftop PV	Segmentation	Grad-CAM	ResNet18	precision=88.31%, recall=54.50%, F1=67.40%, IoU=80.83%
						VGG16	precision=75.55, recall=72.34%, F1=73.91%, IoU=58.62%
					SP-RAN	ResNet18	precision=64.55%, recall=89.80%, F1=75.18%; IoU=60.23%
						VGG16	precision=73.13%, recall=84.02, F1=78.20%; IoU=64.21%
Ortiz et al [47]	10 m	India	Centralized PV	Segmentation	U-Net+ HNM	/	IoU=80.67%, precision=91.03%, recall=86.6%, recall for count is 94.4%

3.3.3.4 Statistic of Architecture and Backbone

The times of each architecture used in the above segmentation models are counted in Fig 8. The symmetric network based on FCN is the most frequently used, among which U-Net and the optimized models based on it are the most popular ones. Combined with the studies of architecture comparison, it shows that U-Net and DeepLab v3+ have no obvious difference in segmentation performance.

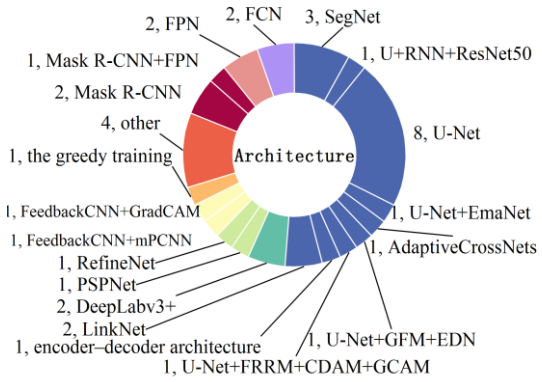

Fig 8 Number of times of all kinds of architectures for segment PV

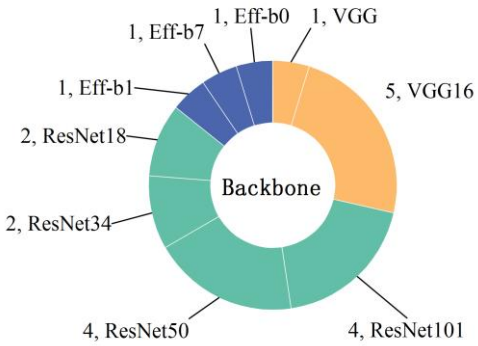

Fig 9 Number of times of all kinds of backbone for segment PV

Fig 9 shows the times of various backbone networks used for feature extraction, referring to the references that pointed out their backbone types. ResNet is the most used backbone, followed by VGG16. And it is found that the performance obtained by ResNet50 is often better than that obtained by ResNet101

[48]. In addition, Efficient-net-b7 is more advanced and deeper to Efficient-net-b0, ResNet50 and ResNet101, however, the maximum difference of accuracy among them is less than 1% [45]. It indicates that blindly increasing the depth of backbone or using more advanced backbone does not significantly improve the segmentation accuracy.

3.3.4 Statistic of Training Optimizer

Among the above references on PV identification using deep learning methods, there are 27 references that described the training optimizer in their model training. The statistical results are shown in Fig 10. It can be seen that the Adaptive Moment Estimation (Adam) optimizer and Stochastic Gradient Descent (SGD) optimizer are the two most commonly used optimizers. Among them, most semantic segmentation models with fully supervised learning use the Adam optimizer. The models for image classification and object detection prefer the SGD optimizer.

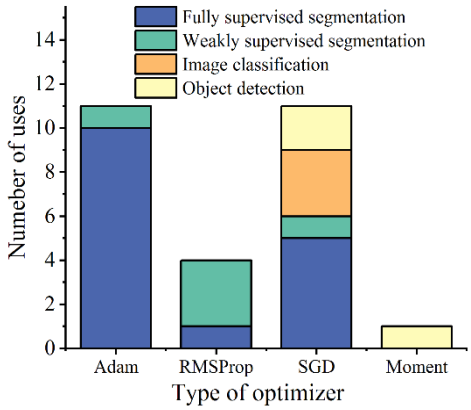

Fig 10 Number of times of all kinds of optimizers for PV identification

3.4 Image Augment

The PV region in satellite images is much smaller than the background. Image augment can increase the number of positive samples, so as to balance the sample proportion in the training set of deep learning to ensure identification performance. Meanwhile, image augment can also improve the robustness of OBIA and

deep learning models. Among the above 48 references using OBIA or deep learning, 22 references used image augment. Fig 11 statistically the times of the methods used. The most used image augment method is rotating the positive sample image, followed by flipping and adjusting the brightness range. Among them, random sampling is generally used for negative samples to reduce their proportion in the training set.

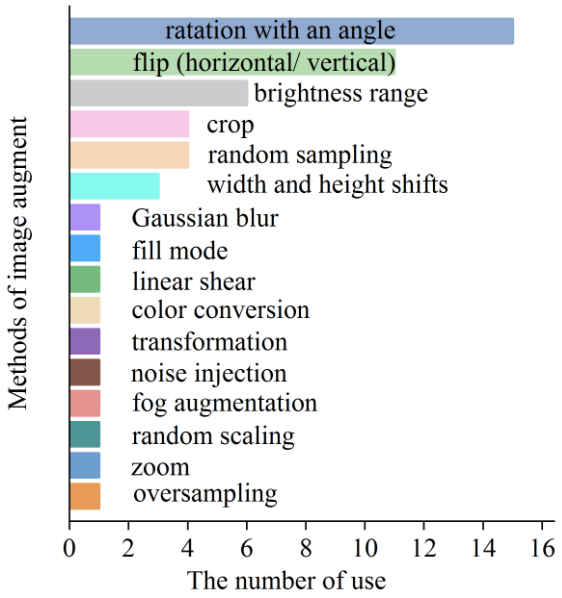

Fig 11 Number of times of all kinds of methods for image augment

3.5 Comparison of Different Methods contrast

The above identification models are classified according to application scenarios, i.e., PV scale, image resolution and identification purpose level. Table 12 summarizes the range of performance metrics for each method. Only precision, recall and IoU (most used). are summarized here, because the performance metrics used in each study are different. The application scenarios and performance of each method are roughly compared, even though the image datasets used in each study are different.

Table 12 Performance comparison of four methods under different identification tasks

Identify rooftop PV in images with resolution less than 2 m/px				
	PBIA	OBIA	Fully supervised deep learning	Weakly supervised deep learning
Image classification	/	/	precision=72-95%; recall=75.2-92.4%	/
Object detection	precision=60%; recall=70%	recall=94% in 100 buildings images.	precision=81.2-92.99%; recall=84-87.3%	/
Semantic segmentation	/	/	precision=41-98.9%; recall=73-95.8%	precision=64.55-92.63%; recall=54.5-89.38%
Identify centralized PV in images with resolution less than 2 m/px				
Semantic segmentation	/	Segment other panels within the same array: precision and recall measures are both higher than 0.99.	precision=53-98.6%; IoU=77.3-90%; recall=52-90%	/
Identify centralized PV in images with a resolution greater than or equal to 10 m/px				
Object detection	/	precision=92.39-98.53%; recall=85.86-92.19% IoU=80.19%	precision=60-83.81%; recall>75%; IoU=50-56%	/
Semantic segmentation	For water PV: precision=96.9-100% recall=84.4-96.2%	precision=63.64-71.6%; recall=37.25-54.9%	IoU=85.43-91.17%; F1=83-96.64%; overall accuracy=96.73-98.08%	precision=91.03%; recall=86.6%; IoU=80.67%

Combining Table 4 and Table 5, for the rooftop PV detection task in high-resolution images, when OBIA is applied in a small area, the recall can reach 94%. Since more texture features can be obtained, the detection performance of OBIA is higher than that of PBIA. However, PBIA and OBIA are only suitable for detection in a small region, and the performance in RGB images is not as good as in multispectral images. The deep learning method is suitable for city-scale detection and can maintain high precision, ranging from 81.2% to 92.99%. Deep learning is generally used for the tasks of rooftop PV segmentation with high-resolution images. The range of the precision and the recall of fully supervised learning methods are 41-98.9% and 73-95.8%, respectively. Compared to the weakly supervised learning method, the fully supervised learning method has a higher upper limit of performance. When the model structure is too simple, the segmentation

accuracy may be very low. Weakly supervised learning is suitable for segmentation in a larger geographical region. The well-trained weakly supervised learning model has a maximum precision of 92.63% with a recall of 82.45%, making it better than most fully supervised learning models. However, the performances from the same model but with different training datasets differ greatly. It indicates that the size of image datasets and label quality significantly affect the model performance.

Fully supervised learning methods can be used to segment centralized PV in high-resolution images. However, compared with models using low-resolution images, its advantages are not obvious. For example, the maximum deviation in IoU is only 1.17%. As for OBIA, which is only used to identify other PV panels in the same array where the template PV is located, can be applied for fault detection of the PV array.

Low-resolution images no higher than 10 m/px are generally satellite images from Landsat or Sentinel with information from multispectral bands. OBIA can outperform deep learning in centralized PV detection through the proper use of multispectral information and texture features. It can achieve the highest precision of 98.53% with a recall of 92.19%, which is significantly higher than those of deep learning. However, the deep learning method is still better than OBIA in the segmentation task, and the weakly supervised learning is inferior to the fully supervised learning. In addition, PBIA can achieve excellent segmentation performance for water concentrated PV due to the simple background and the obvious spectral differences between ground and targets.

The advantages, disadvantages, and application scenarios of these methods have been presented in Table 13.

Table 13 The advantages, disadvantages, and application scenarios of identification methods

Model approach		advantages	disadvantages	Applicable scenario
PBIA		The calculation speed is fast and the true positive of the result is high.	Without utilizing the spatial features between pixels, it is easy to cause incomplete shapes when detecting or dividing PV. Easy to produce the salt-and-pepper effect.	It is suitable for object detection in a small area, or when there is multispectral information and the PV spectrum is obviously different from the surrounding background. It has an excellent performance in segmenting water centralized PV.
OBIA		Fast calculation speed, better use of spectral features, and context information. It can obtain good performance using multispectral images. And the true positive of the result is high.	The selection of typical features is diverse, so it is difficult to formulate simple, accurate, and robust rule sets.	Detect the rooftop PV in small areas. Detect centralized PV using multispectral images. PV panel fault detection in a centralized PV.
Deep learning method	Total supervision learning	It has the highest identification accuracy. The true negative rate of identification results is high.	The physical meaning of the model is not clear. The model performance can only be optimized by repeated parameter tuning, which has the characteristics of experience and blindness. A large amount of accurately labeled training set data is needed.	It is suitable for the detection and segmentation of rooftop PV in a relatively large area and the detection and segmentation of centralized PV in a large area.
	Weakly supervised learning	It does not need lots of pixel-level annotated training set data. The training set can be constructed through border-level annotation, graffiti-level annotation, and image-level annotation.	The model training is complicated and the accuracy is lower than that of the fully supervised method The physical meaning of the model is not clear	It is suitable for the detection and segmentation of PV systems of various sizes at the national or international level.

4 Application of PV Identification

4.1 Determining the Azimuth and Tilt Angle of PV panels

The azimuth and tilt angle of PV panels will affect the accuracy of power generation estimation. The PV area in 2D images needs to be converted to the actual area with the tilt angle. The azimuth can be judged according to the rotation angles of the polygons. Edun et al. [75] used the Candy edge detection method and Hough Transform to determine the direction of the longest edge of PV to estimate its azimuth. The azimuth could also be estimated by identifying the ridge line in images [94, 95].

According to a known tilt angle distribution law of a certain region, the tilt angle of PV in the vicinity can be assigned at random [95]. Killinger et al. [96] calculated the tilt angle of many PV systems worldwide, which accounted for 1.7% of global installed capacity in the year. The results showed that the average tilt angle was between 16.1° (Australia) and 35.6° (Belgium). Combined with 3D image information, Mayer et al. [87] used aerial orthographic images with 0.1 m/px to segment PV and assumed that every 6 m^2 of PV area could have a capacity of 1 kilowatt peak (kWp). Compared to the capacity information from the official registry, the total capacity obtained is 22.70% to 31.73% lower without considering the tilt correction for the rooftop PV with a capacity of less than 30 kWp. This difference could be reduced to between 8.8% and 19.49% when all the tilt angles are equal to the local optimum tilt angle. Assuming that the tilt angle is equal to the optimal angle in that region, the tilt angle will not lead to capacity estimation deviation for the rooftop PVs with a capacity of more than 30 kWp.

4.2 Predicting the PV at national or global level

The fundamental goal of PV identification is to estimate PV area and capacity. Most studies ended up evaluating the performance of the model through test sets from the same locations as the training set. However, few studies extended their model to identify the PV in a wider region and to predict their area and capacity.

Hou et al. [83] used SolarNet to identify and map 439 large-scale centralized photovoltaic power stations in China, covering a total area of nearly 2,000 km^2. Kruitwagen et al. [13] counted PV output greater than 10 kW in the world and collected a total of 68,661 systems in 131 countries. As for the distributed PV and rooftop PV, Yuan et al. [67] performed PV segmentation on satellite maps of San Francisco and Boston, covering an area of 108 km^2 respectively, and projected the PV segmentation results onto OSM maps. The location of the PV distribution and the size of the coverage can be seen. DeepSloar identified approximately 1.4702 million PV systems in urban areas and all areas in the United States with nightlight intensity greater

than 128 [23].

So et al. [97] proposed the estimation relationship between PV area and capacity

$$c = \gamma\alpha + \gamma_0 \quad (1)$$

where γ is a proportionality constant, indicating the capacity per unit of surface area. γ_0 is a bias parameter.

For the solution, So et al proposed two regression methods: (i) a global estimate of capacity per unit area; (ii)a unique estimate of capacity per unit area for each solar array. The second method assumes that the PV capacity is related to the brightness or intensity of the image, and the area of the PV plate is also related to the power generation efficiency. Analysis results show that the second method can improve the estimation accuracy by about 9%. Based on this method, Malof et al. [98] used SolarMapper to estimate the PV capacity of 168 municipal regions in Connecticut. Their γ calculated by the first method was 0.88, and the second method was 0.91.

4.3 Predicting rooftop PV potential

Rooftop PV potential can be divided into four classes: physical potential, geographical potential, technical potential, and economic potential, as shown in Fig 12. Physical potential refers to the total amount of solar energy a region can receive. The data can be obtained from weather station statistics or from some sunshine simulation software. The geographic potential is the roof area available for PV installation. The roof geometry, tilt angle, azimuth, the shading effects of other buildings and trees, and the upper structures located on rooftops should be considered. Technical potential refers to the maximum power generation that can be generated, considering the conversion efficiency of PV modules and system performance efficiency. The former is about 19.8%, from the Tracking the Sun [16]. The latter is generally affected by panel fouling, component loss, ambient temperature and other conditions, and is generally 0.75-0.9 [99]. Economic potential is the profitability of the rooftop PV system after considering installation and operation costs,

lifetime, interest rate and other economic parameters and government policy. Among them, the geographical potential is the focus of this paper.

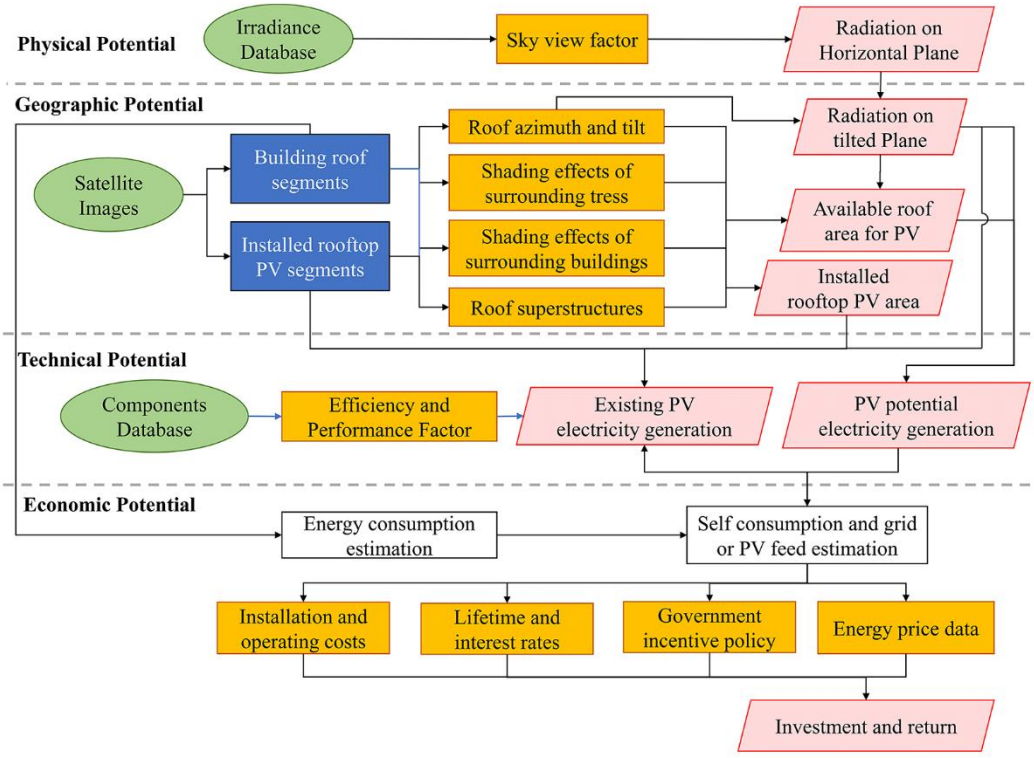

Fig 12 The relationship and influencing factors of four kinds of rooftop PV potentials

Some early studies suggested a rough positive correlation between roof area and local population density [100, 101]. Joshi et al. [14] established the relationship between population density, road length, the built-up area identified by satellite images and roof area to predict global PV potential. However, most of the building footprint data in the training set were from North America, Europe, and Africa. There are doubts about the accuracy of the model for China, India, and other regions.

Segmenting roofs in satellite images, a more accurate method becomes more likely to be used in a large region as computer technology improves. Table 14 shows some studies on roof segmentation. Since building contours differ from roofs in non-orthographic images, it does not include studies on building segmentation. It shows that U-Net is the best segmentation model with a maximum precision of 91.9%. In addition, FPN

and SegNet also have good performance. At the same time, Zhong et al. [102] found that compared to the rural, the urban area is smaller but the number of buildings is larger and the background is more complex. Increasing the proportion of urban images in the training dataset can significantly improve the model performance.

Table 14 Basic information of studies for roof segmentation

Reference	Location	Method	Backbone	Performance
Song et al [103]	a sub-district of Beijing	OBIA	/	area error=8.24%
Qin et al [104]	7 cities, China	FCN	VGG16	overall accuracy=94.67%, mIoU=85%
		FCN+CRF	VGG16	overall accuracy=94.69%, mIoU=83%
Huang et al [105]	Wuhan, China	U-Net	ResNet	Compared with GIS: when the shading factor is 0, the error of PV potential is 9.23%. when the shading factor is 0.2, the error of PV potential is 0.1%.
Lee et al [106]	6 cities, the U.S.	FPN	ResNet-101	precision=91.1%
		U-Net	ResNet-102	precision=91.9%
		Mask-RCNN	ResNet-103	precision=86.3%
Sampath et al [107]	Bangalore, India	U-Net	/	Accuracy=94.73%，Dice coefficient=78.24%
		SegNet	/	Accuracy=94.07%，Dice coefficient=69.88%
		FCN	/	Accuracy=93.45%，Dice coefficient=61.41%
Streltsov et al [108]	Gainesville, Florida, and San Diego, California	Modified U-Net	ResNet-152	Building recall for Gainesville=84%; Building recall for San Diego=88%; residential precision=99%; commercial precision=74%
Zhong et al [102]	Nanjing, China	DeepLab v3	/	Non-strategic sampling method: accuracy=90%，precision=80%，recall=71%，F1=80%.
				Strategic sampling method: accuracy=92%，precision=82%，recall=79%，F1=81%.

The study by Huang et al shows that assuming 100% utilization of roof areas for PV will lead to a potential error of 9.23% [105]. Roof availability area is influenced by various factors. Sun et al. [109] trained

a model based on U-Net that could identify roof orientation to filter out roofs facing north. Qian et al. [110] propose a network model with the encoder-decoder structure that can identify the boundary lines of the roof structure. Several methods have been proposed for laying fixed-size PV panels in roof images, considering in detail the PV installation angle and maintenance distance [94, 109, 111]. This method can improve the estimation accuracy but is very complicated. For residential in villages and towns in northern China, roofs are almost not shadowed by surrounding buildings, and the ratio of PV available area to total roof area in the whole village is about 0.58 [109]. However, this study ignored the area of solar thermal systems, which are common in this region. If the roof availability coefficient can be determined, the geographical potential of rooftop PV could be estimated by simple mathematical methods. Table 15 lists some methods of calculating the coefficient and the influencing factors considered. 3D building information of urban blocks is imported into GIS software to simulate solar radiation at a typical time, and to obtain the available roof area in most studies. The roof availability coefficient obtained by this method is between 0.5 and 0.975. However, it only considers the shading effect between buildings, which results in a serious overestimation of geographical potential. Khan et al. [112] found that 20-25% of the roof area was used for other use including upper structures through field measurement. Considering the azimuth, shading effect and other uses, the coefficient is reduced to 0.25-0.46 through human inspection and segmentation from satellite images as well as the field survey. Also, the attenuation of low-rise buildings from surrounding trees is approximately 15-30% [113].

Table 15 Roof availability coefficient methods based on different methods

Reference	Location	Building type assessed	The factors considered	Method	Roof availability coefficient
Izquierdo et al [114]	Spain	All in urban	Shading between buildings, other roof uses	Human inspection from satellite images	0.34
Bergamasco et al [115]	Turin, Italy	Residential	Shading between buildings, other roof uses, azimuth	Segmentation from satellite images	0.431
		Industrial			0.460
Bergamasco et al [116]	Piedmont , Italy	Residential		Theoretical analysis	0.065
		Industrial			0.304
Lobaccaro et al [117]	Lugano-Paradiso, Switzerland	Residential	Shading between buildings	Irradiation simulation for 3D model	0.7
		Industrial			0.9
Ko et al [118]	Taiwan, China	All			0.5
Ren et al [119]	A block of Hong Kong, China	Urban			0.643
Xu et al [120]	Wuhan, China	Industrial buildings, middle and high-rise residential buildings			0.901
		Commercial and office buildings			0.946
		Low-rise residential buildings			0.975
Singh et al [121]	Mumbai, India	All	Shading between buildings, other roof uses		0.28-0.4
Mishra et al [99]	Uttarakhand, India			Field survey	0.25
Khan et al [112]	Karachi, Pakistan	Residential and commercial buildings	other roof uses	Field survey	0.75-0.8
Sun et al [109]	North China	Residential in village and town	Azimuth, maintenance distance	Simulated installation in satellite images, field survey	Ratio of installed PV area to roof area: Flat roof: 0.82 N-S pitched roof: 0.47 E-W pitched roof: 0.83

In addition, many roofs do not fully utilize the available area for rooftop PV installation in practice. Combined with manual inspection, Mainzer et al. [95] found that only 30% to 80% of the roof available area

was installed with PV systems in Freiburg, Germany. De Vries et al. [94] compared the installation situation of 215 roofs obtained by the model and the actual. There are 66 and 45 roofs were using two-thirds and one-third of their available area for PV installation, respectively. The former is due to ignoring some roof obstacles, while the main reason for the latter is that the owners are unwilling to install it. The erosion of rooftop PV geographic potential caused by this situation requires further investigation and summary. At the same time, installing PV on available roofs as much as possible is a matter of economy and policy.

5 Discussion

5.1 Selection of Image Data Sources and Identification Models

The satellite image with a resolution of less than 10 m/px can be downloaded free of charge, and have the advantages of multispectral information, short revisit periods and fast update. It can be used to identify centralized PV power stations and estimate their installation dates. The 0.3 m/px is the threshold resolution of the image segmentation task for rooftop PV [122]. High-resolution aerial images of a few countries are openly accessed with short update cycles. Although Google Earth/Static Maps offer high-resolution images that cover the globe, they are taken on inconsistent dates, and their update cycle is longer. Moreover, the images are not free of charge. To reduce the cost, multispectral images from Sentinel and Landsat-8 can be used to identify and filter out the centralized PV power stations as well as the large region of vegetation or water.

The barrier to improving the PV identification model is the large number of high-quality PV image datasets. Existing public PV image datasets can cover various sizes of PV, but the overall number is still not abundant.

By and large, the aim of PV identification mainly includes image classification, object detection and semantic segmentation. There are three common methods, i.e., PBIA, OBIA and deep learning. The

calculation speed of PBIA and OBIA is fast, and it can obtain excellent identification performance when using multispectral images for the centralized PV. OBIA has very excellent detection performance in identifying centralized PV even with low spectral images. Deep learning has high accuracy in PV segmentation of all sizes, but its model convergence is difficult, and the prediction speed is slow, with a large number of samples needed in training. Combined with the selection of images at different resolutions, this study proposes a strategy of grading identification for PV, which is used to segment all kinds of PV in a large region, such as national and global, as shown in Fig 13. Firstly, the PBIA and the 10 m/px and 20 m/px multispectral images from Sentinel can be used to identify the centralized water PV in the water region obtained by historical water surface data. Secondly, OBIA and the multispectral images from Sentinel can be used to detect centralized PV on land. It can be used to filter out large areas of water and vegetation through NDVI and NDWI, thus reducing the area needed to identify in subsequent tasks. The recall of objects should be kept high by adjusting the probability threshold. Then the images around the positive detection objects are cropped out and input into a deep learning segmentation model. This significantly reduces the identification region of the deep learning segmentation model. Finally, the deep learning method is used to segment rooftop PV and distributed PV systems in the remaining region in aerial images with the resolution of not less than 0.3 m/px. Such a grading identification strategy can improve the efficiency of the overall identification task, reducing the cost of image acquisition and ensuring accuracy.

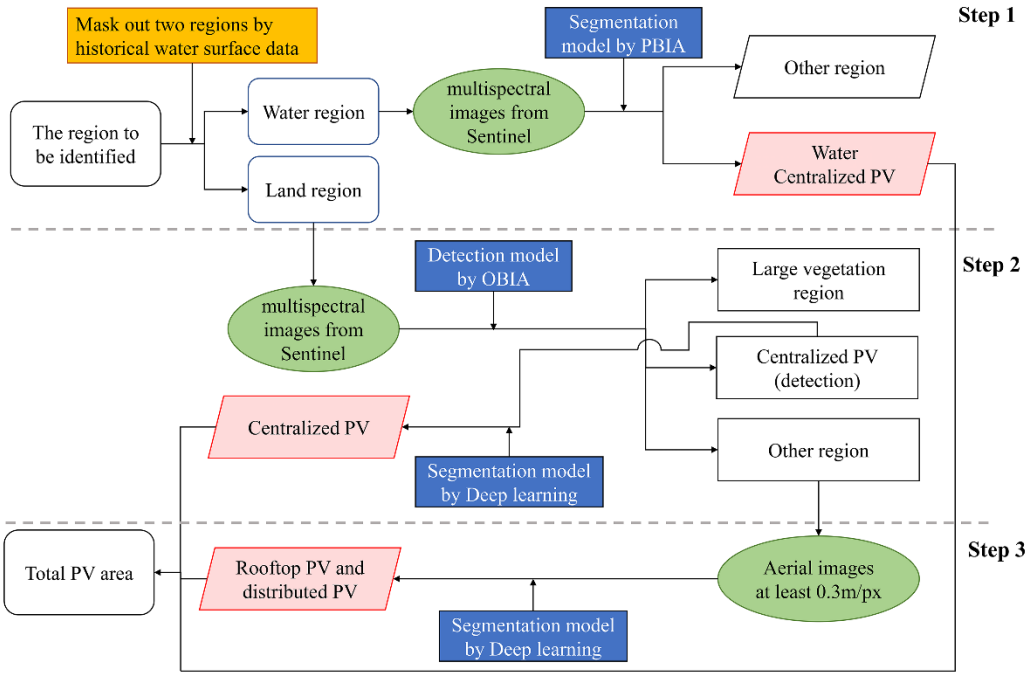

Fig 13 Strategy of grading identification PV

A case study is applied to illustrate the strategy of detection first detection using OBIA and then segmentation using deep learning for centralized PV. In Jiangsu datasets [45], the dataset with a resolution of 0.3m /pixel includes five centralized PV image sub-datasets with different backgrounds and one rooftop PV image sub-dataset. In this case, the first five image sub-datasets are selected as the first image dataset, which includes 2072 images, each with a size of 1024×1024. The ratio of the training dataset to the test dataset is 8:2. On this basis, each image is divided into four 512×512 images, and the images with PV pixels are selected and grouped into the second image dataset, as shown in Fig 14. It includes 5551 images in total. This process approximately represents the process of using the detection model to screen the candidate regions for further segmentation.

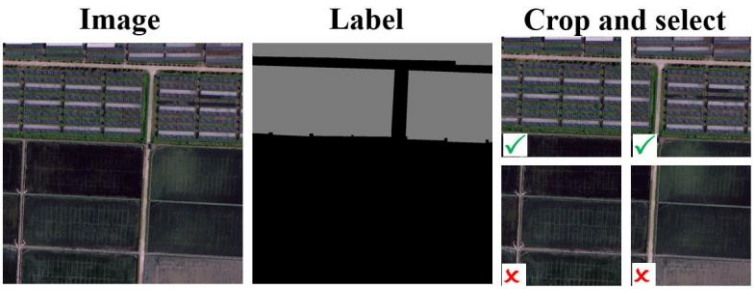

Fig 14 Case Study's Generation of Image Datasets

The two models are trained on two datasets respectively. They all apply DeepLab v3 as the architecture, ResNet-101 as the backbone, and Adam as the training optimizer, with a learning rate of 0.001. After training 50 epochs, the results of the optimal epochs represent the model performance. For the model using the complete dataset, the precision, recall, and F1-score for pixels for PV are 96.61%, 96.80%, and 96.71%, respectively. For the second model, the precision, recall, and F1-score for pixels for PV are 96.71%, 97.17% and 96.94%, respectively. As Fig 15(a) and (b), it can be seen that the continuity of the positive regions of the first model is less, and there are more false negative regions inside them. In addition, it may predict a large area of background areas as positive, as shown in Fig 15(c). The second model has better continuity and clearer boundary. In this case, manual image cropping is assumed to be a detection process. Referring to the above, for centralized PV systems, detection using OBIA and multispectral images can obtain better results than that using depth learning for the whole region. Therefore, this case has a certain reference value.

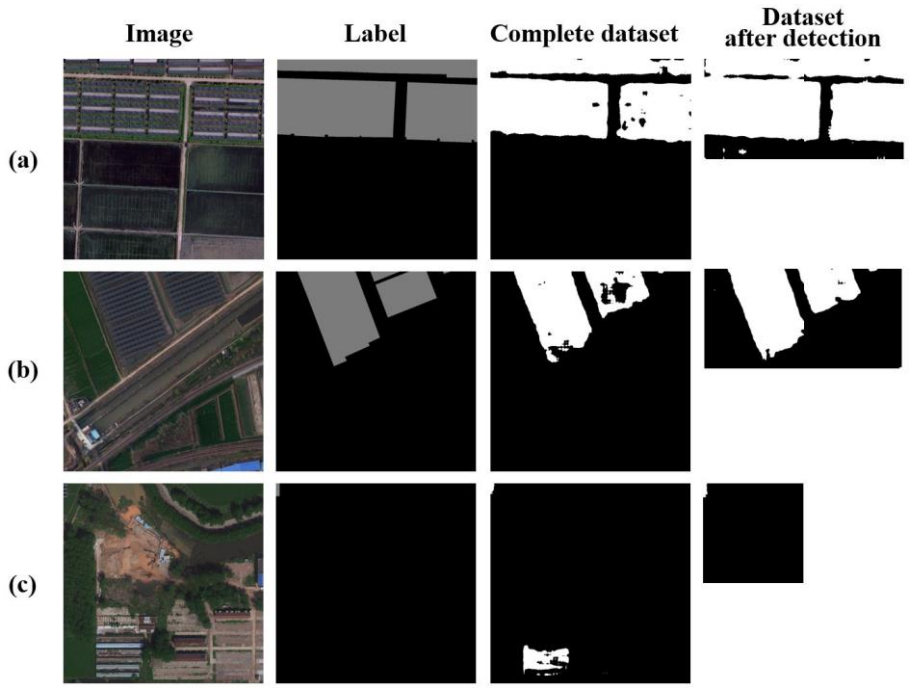

Fig 15 Segmentation Results of Two Identification Strategies

5.2 Optimization of Identification Models

Identification models for segmentation should still maintain high performance in the regions out of training datasets. For example, DeepSolar was trained using image datasets from the U.S, and when it was applied to Germany, its recall for PV number was reduced to 40.7% [93]. To improve the applicability of the models, the training dataset should include satellite images of various scenarios, including different landforms (e.g., cities, farmland, deserts, grasslands, etc.), as well as considering the diversity of architectural styles of different cities. When a model is applied to a similar new scenario, the performance of the model can be significantly improved only by adding a small amount of satellite images from the new scenario to the training dataset for further training. Studies have shown that the ratio of new satellite images from the new similar scenario to the original training datasets can be less than 10% [93, 123]. Some objects with similar colors or shapes are also easy to be identified as PV, such as solar thermal systems, glass skylights, swimming pools and roads. It results in an increased false positive rate while reducing precision and accuracy. The model can be optimized by adding images of these confusable objects or taking false positive results to

the training dataset for further training with negative labels [72, 124]. In addition, removing the identification results that coincide with the location of roads, or retaining the identification results only within the polygon contour of roofs can be used [67, 70]. It is found that PV showed significantly different color characteristics in sunlit situations and shade-facing situations [122]. However, almost all PV panels exhibit highly similar color histogram patterns. It suggests that the different color features are only caused by brightness relative to lighting conditions. Therefore, changing the color pattern in positive samples can be used as an image augment method, and the color pattern can also be used to filter out the false positives in the results.

5.3 Estimation of Capacity and Installation Potential

The tilt angle correction should be performed to estimate the actual PV area and build the relationship between the area and the capacity. The tilt angle could be based on the known tilt angle distribution law in the vicinity if it exists, or is assumed to be equal to the local optimum tilt angle.

For predicting the rooftop PV potential, simulating the placement of fixed-size PV panels on each roof is a piece of detailed and accurate work, but it is too complicated to be not suitable for a wide application. Summarizing the roof availability coefficient is still the most feasible method. The solar shadow simulation of buildings in a block with 3D information only helps to estimate the effect of shading on this coefficient. The coefficient reduction resulting from the roof's upper structure and other uses should also be considered.

5.4 Outlooks

(1) As for non-orthographic aerial images, such as from Google Maps, the image distortion and the area estimation error caused by the unknown shooting height and the unknown orthographic point location are worthy of a further study.

(2) RNN, Temporal Cluster Matching technology, and PCNN have been used to predict the installation time of the centralized PV systems [13, 47, 88]. It is worth applying the identification to monitor the

installation growth rate. Identifying the dynamic growth process of rooftop PV can provide implications for the government to make more effective policies.

(3) With the proliferation of rooftop PV, it is important to periodically estimate the rooftop PV coverage ratio and the existing geographic potential in a city. And the installed systems should be considered in future studies of geographic potential. Furthermore, it is necessary to conduct statistical analysis on the roof availability coefficient after PV installation, as well as on the proportion of PV installed area to total roof area. The latter can be used directly to improve the prediction accuracy of the potential of rooftop PV installation.

6. Conclusions

This paper has reviewed the image sources available and the existing public image datasets, followed by classification and analysis of the existing PV identification models. The application of predicting PV area and capacity in large regions and of predicting the geographical potential of rooftop PV has been discussed. The following conclusions and outlooks can be drawn:

(1) The series of Sentinel can provide free, global multispectral imagery at 10 m/px with a short revisit period. It can be used to identify centralized PV systems and to infer their installation dates. Aerial images with a resolution of no less than 0.3 m/px are required for rooftop PV identification. Google Earth/Static Maps and other platforms can provide images of some countries meeting the required resolution. It is noted that these images may not be openly accessed and could be taken in different years, making it difficult for predicting yearly PV installation capacity.

(2) The PV identification models can be divided into image classification, object detection and semantic segmentation subject to different purposes. Especially, the PV area can be predicted by the models of semantic segmentation. The PV identification models are commonly established by three main methods,

including PBIA, OBIA and deep learning. The first two methods have a fast identification speed, but both need to establish feature sets beforehand and their performance depends on the spectral information. OBIA can be superior to the deep learning method in detecting centralized PV. The deep learning method has the best performance in segmenting all sizes of PV systems including rooftop PV.

(3) In the semantic segmentation model, the most widely used model with fully supervised learning is U-Net and its optimized architecture. The U-Net and the DeepLab v3+ are the models with the best accuracy, and there is no obvious performance difference between them. For backbone networks, the ResNet series and Efficient-Net-b series are commonly used. By far Efficient-net-b7 has the best performance, followed by ResNet-101 which has a simpler structure and is more commonly used. And the complexity and training efficiency of models also need to be concerned. Blindly increasing the complexity of the model does not necessarily improve the accuracy. To improve the robustness of the models, rotation, flip, and changing brightness are the most commonly used image augment methods. Considering the susceptibility of PV color characteristics in images to light, changing brightness should be preferred.

(4) A grading segmentation strategy for all sizes of PV is presented. Firstly, PBIA and multispectral images from Sentinel are used to segment concentrated water PV and filter out the large regions of water. Secondly, OBIA is used to detect centralized PV in the remaining area of the image and maintain a high recall rate by adjusting the threshold. The regions around positive detection results are cropped into the candidate images for segmentation by the models based on the deep learning method. Finally, the deep learning model and the aerial images with a resolution of at least 0.3 m/px are used to segment roof PV and distributed PV in the remaining region. A simplified case study has been applied to illustrate that detection first and then segmentation can improve the segmentation accuracy of the centralized PV.

(5) The available area on the roof is the key parameter to predict the geographic potential of solar rooftop

PV installation capacity. After considering roof azimuth, shadow effect between buildings and other uses of the roof, the roof availability coefficient is in the range of 0.25 to 0.46. The actual roof available coefficients for various functional buildings need to be investigated and summarized widely in the future.

Declaration of Competing Interest

None

CRediT author statement

Hongzhi Mao: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Writing - original draft. **Xie Chen**: Resources, Writing - review & editing. **Yongqiang Luo** and **Jie Deng**: Writing - review & editing. **Zhiyong Tian**: Project administration, Supervision, Funding acquisition, Writing - review & editing. **Jinghua Yu**, **Yimin Xiao** and **Jianhua Fan**: Writing - review & editing.

Acknowledgement

All the authors wish to thank the financial support from National Key R&D Program of China (No.2021YFE0113500), the Fundamental Research Funds for the Central Universities, China (Grant Number: 2020kfyXJJS097), and Research Project of the Ministry of Housing and Urban-Rural Development of China "Research and Demonstration of Optimal Configuration of Energy Storage System in Nearly Zero Energy Communities" (K20210466).

References

[1] Tracker, N. Z. *Net Zero Numbers*. 2022. https://zerotracker.net/ (accessed 2022 8 May).
[2] UN. Progress Towards The Sustainable Development Goals. Council EaS: 2020; p 70520.
[3] Mallapaty, S. How China could be carbon neutral by mid-century. *Nature* **2020**, *586* (7830), 482-484. DOI: https://link.gale.com/apps/doc/A639031249/HRCA?
[4] He, G.; Lin, J.; Sifuentes, F.; Liu, X.; Abhyankar, N.; Phadke, A. Rapid cost decrease of renewables and storage accelerates the decarbonization of China's power system. *Nature Communications* **2020**, *11* (1), 2486. DOI: https://doi.org/10.1038/s41467-020-16184-x.
[5] He, J.; Li, Z.; Zhang, X.; Wang, H.; Dong, W.; Chang, S.; et al. Comprehensive report on China's Long-Term Low-Carbon Development Strategies and Pathways. *Chinese Journal of Population, Resources and Environment* **2020**, *18* (4), 263-295. DOI: https://doi.org/10.1016/j.cjpre.2021.04.004.
[6] De La Peña, L.; Guo, R.; Cao, X.; Ni, X.; Zhang, W. Accelerating the energy transition to achieve carbon neutrality. *Resources, Conservation and Recycling* **2022**, *177*, 105957. DOI: https://doi.org/10.1016/j.resconrec.2021.105957.
[7] Wang, F.; Harindintwali, J. D.; Yuan, Z.; Wang, M.; Wang, F.; Li, S.; et al. Technologies and perspectives for

achieving carbon neutrality. *The Innovation* **2021**, *2* (4), 100180. DOI: https://doi.org/10.1016/j.xinn.2021.100180.
[8] Luo, Y.; Cheng, N.; Zhang, S.; Tian, Z.; Xu, G.; Yang, X.; et al. Comprehensive energy, economic, environmental assessment of a building integrated photovoltaic-thermoelectric system with battery storage for net zero energy building. *Building Simulation* **2022**. DOI: https://doi.org/10.1007/s12273-022-0904-1.
[9] IRENA. *Renewable capacity statistics 2021*. 2021. https://www.irena.org/publications/2021/March/Renewable-Capacity-Statistics-2021 (accessed 2022 6 May).
[10] Zhang, L.; Chen, C.; Wang, Q.; Zhou, D. The impact of feed-in tariff reduction and renewable portfolio standard on the development of distributed photovoltaic generation in China. *Energy* **2021**, *232*, 120933. DOI: https://doi.org/10.1016/j.energy.2021.120933.
[11] Liu, Z.; Deng, Z.; He, G.; Wang, H.; Zhang, X.; Lin, J.; et al. Challenges and opportunities for carbon neutrality in China. *Nature Reviews Earth & Environment* **2022**, *3* (2), 141-155. DOI: https://doi.org/10.1038/s43017-021-00244-x.
[12] Bartie, N. J.; Cobos-Becerra, Y. L.; Fröhling, M.; Schlatmann, R.; Reuter, M. A. The resources, exergetic and environmental footprint of the silicon photovoltaic circular economy: Assessment and opportunities. *Resources, Conservation and Recycling* **2021**, *169*, 105516. DOI: https://doi.org/10.1016/j.resconrec.2021.105516.
[13] Kruitwagen, L.; Story, K. T.; Friedrich, J.; Byers, L.; Skillman, S.; Hepburn, C. A global inventory of photovoltaic solar energy generating units. *Nature* **2021**, *598* (7882), 604-610. DOI: https://doi.org/10.1038/s41586-021-03957-7.
[14] Joshi, S.; Mittal, S.; Holloway, P.; Shukla, P. R.; Ó Gallachóir, B.; Glynn, J. High resolution global spatiotemporal assessment of rooftop solar photovoltaics potential for renewable electricity generation. *Nature Communications* **2021**, *12* (1), 5738. DOI: https://doi.org/10.1038/s41467-021-25720-2.
[15] Wang, Y.; He, J.; Chen, W. Distributed solar photovoltaic development potential and a roadmap at the city level in China. *Renewable and Sustainable Energy Reviews* **2021**, *141*, 110772. DOI: https://doi.org/10.1016/j.rser.2021.110772.
[16] Lab, B. *Tracking the Sun*. 2021. https://emp.lbl.gov/tracking-the-sun (accessed 2022 5 May).
[17] Schwabeneder, D.; Dallinger, B.; Moisl, F. Existing and Future PV Prosumer Concepts. *PVP4Grid Project* **2018**. DOI: https://www.pvp4grid.eu/wp-content/uploads/2018/12/D2.1_Existing-future-prosumer-concepts_PVP4G-1.pdf.
[18] Stowell, D.; Kelly, J.; Tanner, D.; Taylor, J.; Jones, E.; Geddes, J.; et al. A harmonised, high-coverage, open dataset of solar photovoltaic installations in the UK. *Scientific Data* **2020**, *7* (1), 394. DOI: https://doi.org/10.1038/s41597-020-00739-0.
[19] Zhang, X.; Grijalva, S. A Data-Driven Approach for Detection and Estimation of Residential PV Installations. *IEEE Transactions on Smart Grid* **2016**, *7* (5), 2477-2485. DOI: https://doi.org/10.1109/TSG.2016.2555906.
[20] Erdener, B. C.; Feng, C.; Doubleday, K.; Florita, A.; Hodge, B.-M. A review of behind-the-meter solar forecasting. *Renewable & Sustainable Energy Reviews* **2022**, *160* (160), 112224. DOI: https://doi.org/10.1016/j.rser.2022.112224.
[21] Lian, R.; Wang, W.; Mustafa, N.; Huang, L. Road Extraction Methods in High-Resolution Remote Sensing Images: A Comprehensive Review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2020**, *13*, 5489-5507. DOI: https://doi.org/10.1109/JSTARS.2020.3023549.
[22] Yang, H. L.; Yuan, J.; Lunga, D.; Laverdiere, M.; Rose, A.; Bhaduri, B. Building Extraction at Scale Using Convolutional Neural Network: Mapping of the United States. *IEEE Journal of Selected Topics in Applied Earth*

Observations and Remote Sensing **2018**, *11* (8), 2600-2614. DOI: https://doi.org/10.1109/JSTARS.2018.2835377.

[23] Jiafan, Y.; Zhecheng, W.; Majumdar, A.; Rajagopal, R. DeepSolar: A Machine Learning Framework to Efficiently Construct a Solar Deployment Database in the United States. *Joule* **2018**, *2* (12), 2605-2617. DOI: https://doi.org/10.1016/j.joule.2018.11.021.

[24] Fakhraian, E.; Forment, M. A.; Dalmau, F. V.; Nameni, A.; Guerrero, M. J. C. Determination of the urban rooftop photovoltaic potential: A state of the art. *Energy Reports* **2021**, *7*, 176-185. DOI: https://doi.org/10.1016/j.egyr.2021.06.031.

[25] Sun, L.; Jiang, Y.; Guo, Q.; Ji, L.; Xie, Y.; Qiao, Q.; et al. A GIS-based multi-criteria decision making method for the potential assessment and suitable sites selection of PV and CSP plants. *Resources Conservation and Recycling* **2021**, *168*, 105306. DOI: https://doi.org/10.1016/j.resconrec.2020.105306.

[26] Alhammad, A.; Sun, Q.; Tao, Y. Optimal Solar Plant Site Identification Using GIS and Remote Sensing: Framework and Case Study. *Energies* **2022**, *15* (1), 312. DOI: https://doi.org/10.3390/en15010312.

[27] de Hoog, J.; Maetschke, S.; Ilfrich, P.; Kolluri, R. R. Using Satellite and Aerial Imagery for Identification of Solar PV. In *Proceedings of the Eleventh ACM International Conference on Future Energy Systems*, Virtual Event, Australia, 2020; Association for Computing Machinery: pp 308-313.DOI: https://doi.org/10.1145/3396851.3397681.

[28] Feng, C.; Liu, Y.; Zhang, J. A taxonomical review on recent artificial intelligence applications to PV integration into power grids. *International Journal of Electrical Power & Energy Systems* **2021**, *132*, 107176. DOI: https://doi.org/10.1016/j.ijepes.2021.107176.

[29] Gaviria, J. F.; Narváez, G.; Guillen, C.; Giraldo, L. F.; Bressan, M. Machine learning in photovoltaic systems: A review. *Renewable Energy* **2022**, *196*, 298-318. DOI: https://doi.org/10.1016/j.renene.2022.06.105.

[30] NASA/USGS. *Landsat*. 2022. https://landsat.gsfc.nasa.gov (accessed 2022 5 May).

[31] Agency, T. E. S. *Sentinel-1*. 2022. https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1 (accessed 2022 5 May).

[32] Agency, T. E. S. *Sentinel-2*. 2022. https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-2 (accessed 2022 5 May).

[33] Intelligence, A. *Spot 6/7*. 2022. https://www.intelligence-airbusds.com/en/8693-spot-67 (accessed 2022 5 May).

[34] Application, C. C. F. R. S. D. a. *Land Observation Satellite Data Service Platform*. 2014. http://www.cresda.com/CN/Satellite/3128.shtml (accessed 2022 5 May).

[35] Xia, Z.; Li, Y.; Guo, X.; Chen, R. High-resolution mapping of water photovoltaic development in China through satellite imagery. *International Journal of Applied Earth Observation and Geoinformation* **2022**, *107*, 102707. DOI: https://doi.org/10.1016/j.jag.2022.102707.

[36] Earth, G. 2022. https://www.google.com/earth/ (accessed 2022 5 May).

[37] Google. *Maps Static API Usage and Billing*. 2022. https://developers.google.com/maps/documentation/maps-static/usage-and-billing (accessed 2022 5 May).

[38] Agency, U. S. D. o. A. F. S. *National Agriculture Imagery Program (NAIP)*. 2022. https://naip-usdaonline.hub.arcgis.com/ (accessed 2022 5 May).

[39] BKG, G.-D. *Digitale Orthophotos Bodenauflösung 20 cm (DOP20)*. 2022. https://gdz.bkg.bund.de/index.php/default/digitale-geodaten/digitale-orthophotos/digitale-orthophotos-bodenauflosung-20-cm-dop20.html (accessed 2022 5 May).

[40] Nacional, I. G. *Plan Nacional de Ortofotografía Aérea*. 2022. https://pnoa.ign.es/presentacion-y-objetivo (accessed 2022 5 May).

[41] terrain, S. F. O. o. *Information system for aerial photographs / LUBIS viewer*. 2022. https://www.swisstopo.admin.ch/en/maps-data-online/maps-geodata-online/lubis.html (accessed 2022 5 May).
[42] Shihavuddin, A. S. M.; Rashid, M. R. A.; Maruf, M. H.; Hasan, M. A.; Haq, M. A. u.; Ashique, R. H.; et al. Image based surface damage detection of renewable energy installations using a unified deep learning approach. *Energy Reports* **2021**, *7*, 4566-4576. DOI: https://doi.org/10.1016/j.egyr.2021.07.045.
[43] Survey, U. G. *The National Map*. 2022. https://www.usgs.gov/programs/national-geospatial-program/national-map (accessed 2022 5 May).
[44] Bradbury, K.; Saboo, R.; Johnson, T. L.; Malof, J. M.; Devarajan, A.; Zhang, W.; et al. Data Descriptor: Distributed solar photovoltaic array location and extent dataset for remote sensing object identification. *Scientific Data* **2016**, *3*. DOI: https://doi.org/10.1038/sdata.2016.106.
[45] Jiang, H.; Yao, L.; Lu, N.; Qin, J.; Liu, T.; Liu, Y. J.; et al. Multi-resolution dataset for photovoltaic panel segmentation from satellite and aerial imagery. *Earth System Science Data* **2021**, *13* (11), 5389-5401. DOI: https://doi.org/10.5194/essd-13-5389-2021.
[46] Sizkouhi, A. M.; Aghaei, M.; Esmailifar, S. M. *Aerial Imagery of PV Plants for Boundary Detection*. IEEE Dataport, 2020. https://dx.doi.org/10.21227/g2bb-ms79 (accessed 2022 5 May).
[47] Ortiz, A.; Negandhi, D.; Mysorekar, S.; Kiesecker, J.; Nagaraju, S. K.; Robinson, C.; et al. An Artificial Intelligence Dataset for Solar Energy Locations in India. *arXiv* **2022**, *2202.01340[cs.LG]*. DOI: https://doi.org/10.48550/arXiv.2202.01340.
[48] Coelho Vieira da Costa, M. V.; Ferreira de Carvalho, O. L.; Orlandi, A. G.; Hirata, I.; de Albuquerque, A. O.; Vilarinho e Silva, F.; et al. Remote Sensing for Monitoring Photovoltaic Solar Plants in Brazil Using Deep Semantic Segmentation. *Energies* **2021**, *14* (10), 2960. DOI: https://doi.org/10.3390/en14102960.
[49] Kleebauer, M.; Horst, D.; Reudenbach, C. Semi-Automatic Generation of Training Samples for Detecting Renewable Energy Plants in High-Resolution Aerial Images. *Remote Sensing* **2021**, *13* (23), 4793. DOI: https://doi.org/10.3390/rs13234793.
[50] Jie, Y.; Ji, X.; Yue, A.; Chen, J.; Deng, Y.; Chen, J.; et al. Combined Multi-Layer Feature Fusion and Edge Detection Method for Distributed Photovoltaic Power Station Identification. *Energies* **2020**, *13* (24), 6742. DOI: https://doi.org/10.3390/en13246742.
[51] ArcGIS. *An overview of the Segmentation and Classification toolset*. 2022. https://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/an-overview-of-the-segmentation-and-classification-tools.htm (accessed 2022 20 October).
[52] Ji, C.; Bachmann, M.; Esch, T.; Feilhauer, H.; Heiden, U.; Heldens, W.; et al. Solar photovoltaic module detection using laboratory and airborne imaging spectroscopy data. *Remote Sensing of Environment* **2021**, *266*, 112692. DOI: https://doi.org/10.1016/j.rse.2021.112692.
[53] Karoui, M. S.; Benhalouche, F. Z.; Deville, Y.; Djerriri, K.; Briottet, X.; Houet, T.; et al. Partial Linear NMF-Based Unmixing Methods for Detection and Area Estimation of Photovoltaic Panels in Urban Hyperspectral Remote Sensing Data. *Remote Sensing* **2019**, *11* (18), 2164. DOI: https://doi.org/10.3390/rs11182164.
[54] Malof, J. M.; Bradbury, K.; Collins, L. M.; Newell, R. G. Automatic detection of solar photovoltaic arrays in high resolution aerial imagery. *Applied Energy* **2016**, *183*, 229-240. DOI: https://doi.org/10.1016/j.apenergy.2016.08.191.
[55] Czirjak, D. Detecting photovoltaic solar panels using hyperspectral imagery and estimating solar power production. *Journal of Applied Remote Sensing* **2017**, *11* (2), 026007. DOI: https://doi.org/10.1117/1.Jrs.11.026007.
[56] Malof, J. M.; Hou, R.; Collins, L. M.; Bradbury, K.; Newell, R. Automatic Solar Photovoltaic Panel

Detection in Satellite Imagery. In *4th International Conference on Renewable Energy Research and Applications (ICRERA)*, Palermo, ITALY, Nov 22-25, 2015; IEEE: 2015; pp 1428-1431.DOI: https://doi.org/10.1109/ICRERA.2015.7418643.

[57] Zhang, X.; Zeraatpisheh, M.; Rahman, M. M.; Wang, S.; Xu, M. Texture Is Important in Improving the Accuracy of Mapping Photovoltaic Power Plants: A Case Study of Ningxia Autonomous Region, China. *Remote Sensing* **2021**, *13* (19), 3909. DOI: https://doi.org/10.3390/rs13193909.

[58] Plakman, V.; Rosier, J.; van Vliet, J. Solar park detection from publicly available satellite imagery. *GIScience & Remote Sensing* **2022**, *59* (1), 461-480. DOI: https://doi.org/10.1080/15481603.2022.2036056.

[59] Vasku, M. An exploration of automatic detection of large-scale solar plants: application ofmachine learning-based image classification in Google Earth Engine. Aalborg University Copenhagen, Copenhagen, 2019.DOI: https://projekter.aau.dk/projekter/files/306324521/Vasku_Marek_master_thesis_AAUCPH_GIS_2019.pdf.

[60] Wang, M.; Cui, Q.; Sun, Y.; Wang, Q. Photovoltaic panel extraction from very high-resolution aerial imagery using region–line primitive association analysis and template matching. *ISPRS Journal of Photogrammetry and Remote Sensing* **2018**, *141*, 100-111. DOI: https://doi.org/10.1016/j.isprsjprs.2018.04.010.

[61] Garcia-Garcia, A.; Orts-Escolano, S.; Oprea, S.; Villena-Martinez, V.; Garcia-Rodriguez, J. A review on deep learning techniques applied to semantic segmentation. *arXiv* **2017**, 1704.06857 [cs.CV]. DOI: https://doi.org/10.48550/arXiv.1704.06857.

[62] Malof, J. M.; Collins, L. M.; Bradbury, K.; Newell, R. G. A deep convolutional neural network and a random forest classifier for solar photovoltaic array detection in aerial imagery. In *2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*, Birmingham, ENGLAND, 20-23 Nov. 2016, 2016; 2016; pp 650-654.DOI: https://doi.org/10.1109/ICRERA.2016.7884415.

[63] Malof, J. M.; Collins, L. M.; Bradbury, K. A Deep Convolutional Neural Network with Pre-traning, for Solar Photovoltaic Array Detection in Aerial Imagery. In *2017 IEEE International Geoscience and Remote Sensing Symposium*, Fort Worth, TX, Jul 23-28, 2017; IEEE: 2017; pp 874-877.DOI: https://doi.org/10.1109/IGARSS.2017.8127092.

[64] Golovko, V.; Bezobrazov, S.; Kroshchanka, A.; Sachenko, A.; Komar, M.; Karachka, A. Convolutional neural network based solar photovoltaic panel detection in satellite photos. In *9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems - Technology and Applications (IDAACS)*, Bucharest, ROMANIA, 2017

Sep 21-23, 2017; Vol. 1, pp 14-19.DOI: https://doi.org/10.1109/IDAACS.2017.8094501.

[65] Moraguez, M.; Trujillo, A.; de Weck, O.; Siddiqi, A. Convolutional Neural Network for Detection of Residential Photovoltaic Systems in Satellite Imagery. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Electr Network, 2020

Sep 26-Oct 02, 2020; IEEE: 2020; pp 1600-1603.DOI: https://doi.org/10.1109/igarss39084.2020.9324245.

[66] Li, K.; Wan, G.; Cheng, G.; Meng, L.; Han, J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing* **2020**, *159*, 296-307. DOI: https://doi.org/10.1016/j.isprsjprs.2019.11.023.

[67] Yuan, J.; Yang, H.-H. L.; Omitaomu, O. A.; Bhaduri, B. L. Large-scale solar panel mapping from aerial images using deep convolutional networks. In *2016 IEEE International Conference on Big Data (Big Data)*, Washington, DC, 5-8 Dec. 2016, 2016; IEEE: 2016; pp 2703-2708.DOI: https://doi.org/10.1109/BigData.2016.7840915.

[68] Ioannou, K.; Myronidis, D. Automatic Detection of Photovoltaic Farms Using Satellite Imagery and Convolutional Neural Networks. *Sustainability* **2021**, *13* (9), 5323. DOI: https://doi.org/10.3390/su13095323.

[69] Golovko, V.; Kroshchanka, A.; Bezobrazov, S.; Sachenko, A.; Komar, M.; Novosad, O. Development of

Solar Panels Detector. In *2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T)*, Kharkiv, Ukraine, 9-12 Oct. 2018, 2018; IEEE: pp 761-764.DOI: https://doi.org/10.1109/INFOCOMMST.2018.8632132.

[70] Li, Q.; Feng, Y.; Leng, Y.; Chen, D. SolarFinder: Automatic Detection of Solar Photovoltaic Arrays. In *2020 19th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, Sydney, NSW, Australia, 21-24 April 2020, 2020; IEEE: pp 193-204.DOI: https://doi.org/10.1109/IPSN48710.2020.00024.

[71] Sizkouhi, A. M. M.; Aghaei, M.; Esmailifar, S. M.; Mohammadi, M. R.; Grimaccia, F. Automatic Boundary Extraction of Large-Scale Photovoltaic Plants Using a Fully Convolutional Network on Aerial Imagery. *IEEE Journal of Photovoltaics* **2020**, *10* (4), 1061-1067. DOI: https://doi.org/10.1109/JPHOTOV.2020.2992339.

[72] Schulz, M.; Boughattas, B.; Wendel, F. DetEEktor: Mask R-CNN based neural network for energy plant identification on aerial photographs. *Energy and AI* **2021**, *5*, 100069. DOI: https://doi.org/10.1016/j.egyai.2021.100069.

[73] Liang, S.; Qi, F.; Ding, Y.; Cao, R.; Yang, Q.; Yan, W. Mask R-CNN based segmentation method for satellite imagery of photovoltaics generation systems. In *39th Chinese Control Conference (CCC)*, Shenyang, China, 2020

Jul 27-29, 2020; IEEE: 2020; pp 5343-5348.DOI: https://doi.org/10.23919/CCC50068.2020.9189474.

[74] Ishii, T.; Simo-Serra, E.; Iizuka, S.; Mochizuki, Y.; Sugimoto, A.; Ishikawa, H.; et al. Detection by classification of buildings in multispectral satellite imagery. In *23rd International Conference on Pattern Recognition (ICPR)*, Cancun, Mexico, 4-8 Dec. 2016, 2016; IEEE: pp 3344-3349.DOI: https://doi.org/10.1109/ICPR.2016.7900150.

[75] Edun, A. S.; Perry, K.; Harley, J. B.; Deline, C. Unsupervised azimuth estimation of solar arrays in low-resolution satellite imagery through semantic segmentation and Hough transform. *Applied Energy* **2021**, *298*, 117273. DOI: https://doi.org/10.1016/j.apenergy.2021.117273.

[76] Camilo, J.; Wang, R.; Collins, L. M.; Bradbury, K.; Malof, J. M. Application of a semantic segmentation convolutional neural network for accurate automatic detection and mapping of solar photovoltaic arrays in aerial imagery. *arXiv* **2018**, 1801.04018 [cs.CV]. DOI: https://doi.org/10.48550/arXiv.1801.04018.

[77] House, D.; Lech, M.; Stolar, M. Using Deep Learning to Identify Potential Roof Spaces for Solar Panels. In *2018 12th International Conference on Signal Processing and Communication Systems (ICSPCS)*, Cairns, QLD, Australia, 17-19 Dec. 2018, 2018; IEEE: pp 1-6.DOI: https://doi.org/10.1109/ICSPCS.2018.8631725.

[78] Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Munich, Germany, 2015

Oct 05-09, 2015; Springer, Cham: 2015; Vol. 9351, pp 234-241.DOI: https://doi.org/10.1007/978-3-319-24574-4_28.

[79] Pérez-González, A.; Jaramillo-Duque, Á.; Cano-Quintero, J. B. Automatic Boundary Extraction for Photovoltaic Plants Using the Deep Learning U-Net Model. *Applied Sciences-Basel* **2021**, *11* (14), 6524. DOI: https://doi.org/10.3390/app11146524.

[80] Wu, A. N.; Biljecki, F. Roofpedia: Automatic mapping of green and solar roofs for an open roofscape registry and evaluation of urban sustainability. *Landscape and Urban Planning* **2021**, *214*, 104167. DOI: https://doi.org/10.1016/j.landurbplan.2021.104167.

[81] Ren, S.; Malof, J.; Fetter, T. R.; Beach, R.; Rineer, J.; Bradbury, K. Utilizing geospatial data for assessing energy security: Mapping small solar home systems using unmanned aerial vehicles and deep learning. *arXiv* **2022**, 2201.05548 [eess.IV]. DOI: https://doi.org/10.48550/arXiv.2201.05548.

[82] Castello, R.; Roquette, S.; Esguerra, M.; Guerra, A.; Scartezzini, J.-L. Deep learning in the built

environment: automatic detection of rooftop solar panels using Convolutional Neural Networks. *Journal of Physics: Conference Series* **2019**, *1343* (1), 012034. DOI: https://doi.org/10.1088/1742-6596/1343/1/012034.
[83] Hou, X.; Wang, B.; Hu, W.; Yin, L.; Wu, H. SolarNet: A Deep Learning Framework to Map Solar Power Plants In China From Satellite Imagery. *arXiv* **2019**, 1912.03685[cs.CV]. DOI: https://doi.org/10.48550/arXiv.1912.03685.
[84] Zhuang, L.; Zhang, Z.; Wang, L. The automatic segmentation of residential solar panels based on satellite images: A cross learning driven U-Net method. *Applied Soft Computing* **2020**, *92*, 106283. DOI: https://doi.org/10.1016/j.asoc.2020.106283.
[85] Zech, M.; Ranalli, J. Predicting PV Areas in Aerial Images with Deep Learning. In *47th IEEE Photovoltaic Specialists Conference (PVSC)*, Electr Network, 2020
Jun 15-Aug 21, 2020; IEEE: 2020; pp 767-774.DOI: https://doi.org/10.1109/PVSC45281.2020.9300636.
[86] Jie, Y.; Yue, A.; Liu, S.; Huang, Q.; Chen, J.; Meng, Y.; et al. Photovoltaic power station identification using refined encoder-decoder network with channel attention and chained residual dilated convolutions. *Journal of Applied Remote Sensing* **2020**, *14* (1), 016506. DOI: https://doi.org/10.1117/1.JRS.14.016506.
[87] Mayer, K.; Rausch, B.; Arlt, M.-L.; Gust, G.; Wang, Z.; Neumann, D.; et al. 3D-PV-Locator: Large-scale detection of rooftop-mounted photovoltaic systems in 3D. *Applied Energy* **2022**, *310*, 118469. DOI: https://doi.org/10.1016/j.apenergy.2021.118469.
[88] Imamoglu, N.; Kimura, M.; Miyamoto, H.; Fujita, A.; Nakamura, R. Solar Power Plant Detection on Multi-Spectral Satellite Imagery using Weakly-Supervised CNN with Feedback Features and m-PCNN Fusion. *arXiv* **2017**, 1704.06410 [cs.CV]. DOI: https://doi.org/10.48550/arXiv.1704.06410.
[89] Kouyama, T.; Imamoglu, N.; Imai, M.; Nakamura, R. Verifying Rapid Increasing of Mega-Solar PV Power Plants in Japan by Applying a CNN-Based Classification Method to Satellite Images. In *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, Waikoloa, HI, USA, 26 Sept.-2 Oct. 2020, 2020; IEEE: pp 4104-4107.DOI: https://doi.org/10.1109/IGARSS39084.2020.9324469.
[90] Zhang, J.; Jia, X.; Hu, J. Pseudo Supervised Solar Panel Mapping based on Deep Convolutional Networks with Label Correction Strategy in Aerial Images. In *2020 Digital Image Computing: Techniques and Applications (DICTA)*, Melbourne, Australia, 29 Nov.-2 Dec. 2020, 2020; IEEE: pp 1-8.DOI: https://doi.org/10.1109/DICTA51227.2020.9363379.
[91] Zhang, J.; Jia, X.; Hu, J. Uncertainty-Aware Forward Correction for Weakly Supervised Solar Panel Mapping From High-Resolution Aerial Images. *IEEE Geoscience and Remote Sensing Letters* **2021**, *19*, 1-5. DOI: https://doi.org/10.1109/LGRS.2021.3132935.
[92] Zhang, J.; Jia, X.; Hu, J. SP-RAN: Self-Paced Residual Aggregated Network for Solar Panel Mapping in Weakly Labeled Aerial Images. *IEEE Transactions on Geoscience and Remote Sensing* **2021**, *60*, 1-15. DOI: https://doi.org/10.1109/TGRS.2021.3123268.
[93] Mayer, K.; Wang, Z.; Arlt, M.-L.; Neumann, D.; Rajagopal, R. DeepSolar for Germany: A deep learning framework for PV system mapping from aerial imagery. In *2020 International Conference on Smart Energy Systems and Technologies (SEST)*, Istanbul, Turkey, 7-9 Sept. 2020, 2020; IEEE: pp 1-6.DOI: https://doi.org/10.1109/SEST48500.2020.9203258.
[94] de Vries, T. N. C.; Bronkhorst, J.; Vermeer, M.; Donker, J. C. B.; Briels, S. A.; Ziar, H.; et al. A quick-scan method to assess photovoltaic rooftop potential based on aerial imagery and LiDAR. *Sol Energy* **2020**, *209*, 96-107. DOI: https://doi.org/10.1016/j.solener.2020.07.035.
[95] Mainzer, K.; Killinger, S.; McKenna, R.; Fichtner, W. Assessment of rooftop photovoltaic potentials at the urban level using publicly available geodata and image recognition techniques. *Sol Energy* **2017**, *155*, 561-573. DOI: https://doi.org/10.1016/j.solener.2017.06.065.

[96] Killinger, S.; Lingfors, D.; Saint-Drenan, Y.-M.; Moraitis, P.; van Sark, W.; Taylor, J.; et al. On the search for representative characteristics of PV systems: Data collection and analysis of PV system azimuth, tilt, capacity, yield and shading. *Sol Energy* **2018**, *173*, 1087-1106. DOI: https://doi.org/10.1016/j.solener.2018.08.051.

[97] So, B.; Nezin, C.; Kaimal, V.; Keene, S.; Collins, L.; Bradbury, K.; et al. Estimating the electricity generation capacity of solar photovoltaic arrays using only color aerial imagery. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Fort Worth, TX, 23-28 July 2017, 2017; IEEE: 2017; pp 1603-1606.DOI: https://doi.org/10.1109/IGARSS.2017.8127279.

[98] Malof, J. M.; Li, B.; Huang, B.; Bradbury, K.; Stretslov, A. J. A. Mapping solar array location, size, and capacity using deep learning and overhead imagery. *arXiv* **2019**, *1902.10895[cs.CV]*. DOI: https://doi.org/10.48550/arXiv.1902.10895.

[99] Mishra, T.; Rabha, A.; Kumar, U.; Arunachalam, K.; Sridhar, V. Assessment of solar power potential in a hill state of India using remote sensing and Geographic Information System. *Remote Sensing Applications: Society and Environment* **2020**, *19*, 100370. DOI: https://doi.org/10.1016/j.rsase.2020.100370.

[100] Wiginton, L. K.; Nguyen, H. T.; Pearce, J. M. Quantifying rooftop solar photovoltaic potential for regional renewable energy policy. *Computers, Environment and Urban Systems* **2010**, *34* (4), 345-357. DOI: https://doi.org/10.1016/j.compenvurbsys.2010.01.001.

[101] Bódis, K.; Kougias, I.; Jäger-Waldau, A.; Taylor, N.; Szabó, S. A high-resolution geospatial assessment of the rooftop solar photovoltaic potential in the European Union. *Renewable and Sustainable Energy Reviews* **2019**, *114*, 109309. DOI: https://doi.org/10.1016/j.rser.2019.109309.

[102] Zhong, T.; Zhang, Z.; Chen, M.; Zhang, K.; Zhou, Z.; Zhu, R.; et al. A city-scale estimation of rooftop solar photovoltaic potential based on deep learning. *Applied Energy* **2021**, *298*, 117132. DOI: https://doi.org/10.1016/j.apenergy.2021.117132.

[103] Song, X.; Huang, Y.; Zhao, C.; Liu, Y.; Lu, Y.; Chang, Y.; et al. An Approach for Estimating Solar Photovoltaic Potential Based on Rooftop Retrieval from Remote Sensing Images. *Energies* **2018**, *11* (11), 3172. DOI: https://doi.org/10.3390/en11113172.

[104] Qin, Y.; Wu, Y.; Li, B.; Gao, S.; Liu, M.; Zhan, Y. Semantic Segmentation of Building Roof in Dense Urban Environment with Deep Convolutional Neural Network: A Case Study Using GF2 VHR Imagery in China. *Sensors* **2019**, *19* (5), 1164. DOI: https://doi.org/10.3390/s19051164.

[105] Huang, Z.; Mendis, T.; Xu, S. Urban solar utilization potential mapping via deep learning technology: A case study of Wuhan, China. *Applied Energy* **2019**, *250*, 283-291. DOI: https://doi.org/10.1016/j.apenergy.2019.04.113.

[106] Lee, S.; Iyengar, S.; Feng, M.; Shenoy, P.; Maji, S. DeepRoof: A Data-driven Approach For Solar Potential Estimation Using Rooftop Imagery. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, AK, USA, 2019; Association for Computing Machinery: pp 2105–2113.DOI: https://doi.org/10.1145/3292500.3330741.

[107] Sampath, A.; Bijapur, P.; Karanam, A.; Umadevi, V.; Parathodiyil, M. Estimation of rooftop solar energy generation using Satellite Image Segmentation. In *2019 IEEE 9th International Conference on Advanced Computing (IACC)*, Tiruchirappalli, India, 13-14 Dec. 2019, 2019; IEEE: pp 38-44.DOI: https://doi.org/10.1109/IACC48062.2019.8971578.

[108] Streltsov, A.; Malof, J. M.; Huang, B.; Bradbury, K. Estimating residential building energy consumption using overhead imagery. *Applied Energy* **2020**, *280*, 116018. DOI: https://doi.org/10.1016/j.apenergy.2020.116018.

[109] Sun, T.; Shan, M.; Rong, X.; Yang, X. Estimating the spatial distribution of solar photovoltaic power

generation potential on different types of rural rooftops using a deep learning network applied to satellite images. *Applied Energy* **2022**, *315*, 119025. DOI: https://doi.org/10.1016/j.apenergy.2022.119025.
[110] Qian, Z.; Chen, M.; Zhong, T.; Zhang, F.; Zhu, R.; Zhang, Z.; et al. 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* **2022**, *107*, 102680. DOI: https://doi.org/10.1016/j.jag.2022.102680.
[111] Walch, A.; Castello, R.; Mohajeri, N.; Scartezzini, J.-L. Big data mining for the estimation of hourly rooftop photovoltaic potential and its uncertainty. *Applied Energy* **2020**, *262*, 114404. DOI: https://doi.org/10.1016/j.apenergy.2019.114404.
[112] Khan, J.; Arsalan, M. H. Estimation of rooftop solar photovoltaic potential using geo-spatial techniques: A perspective from planned neighborhood of Karachi – Pakistan. *Renewable Energy* **2016**, *90*, 188-203. DOI: https://doi.org/10.1016/j.renene.2015.12.058.
[113] Krapf, S.; Kemmerzell, N.; Khawaja Haseeb Uddin, S.; Hack Vázquez, M.; Netzler, F.; Lienkamp, M. Towards Scalable Economic Photovoltaic Potential Analysis Using Aerial Images and Deep Learning. *Energies* **2021**, *14* (13), 3800. DOI: https://doi.org/10.3390/en14133800.
[114] Izquierdo, S.; Rodrigues, M.; Fueyo, N. A method for estimating the geographical distribution of the available roof surface area for large-scale photovoltaic energy-potential evaluations. *Sol Energy* **2008**, *82* (10), 929-939. DOI: https://doi.org/10.1016/j.solener.2008.03.007.
[115] Bergamasco, L.; Asinari, P. Scalable methodology for the photovoltaic solar energy potential assessment based on available roof surface area: Further improvements by ortho-image analysis and application to Turin (Italy). *Sol Energy* **2011**, *85* (11), 2741-2756. DOI: https://doi.org/10.1016/j.solener.2011.08.010.
[116] Bergamasco, L.; Asinari, P. Scalable methodology for the photovoltaic solar energy potential assessment based on available roof surface area: Application to Piedmont Region (Italy). *Sol Energy* **2011**, *85* (5), 1041-1055. DOI: https://doi.org/10.1016/j.solener.2011.02.022.
[117] Lobaccaro, G.; Frontini, F. Solar Energy in Urban Environment: How Urban Densification Affects Existing Buildings. *Energy Procedia* **2014**, *48*, 1559-1569. DOI: https://doi.org/10.1016/j.egypro.2014.02.176.
[118] Ko, L.; Wang, J.-C.; Chen, C.-Y.; Tsai, H.-Y. Evaluation of the development potential of rooftop solar photovoltaic in Taiwan. *Renewable Energy* **2015**, *76*, 582-595. DOI: https://doi.org/10.1016/j.renene.2014.11.077.
[119] Ren, H.; Xu, C.; Ma, Z.; Sun, Y. A novel 3D-geographic information system and deep learning integrated approach for high-accuracy building rooftop solar energy potential characterization of high-density cities. *Applied Energy* **2022**, *306*, 117985. DOI: https://doi.org/10.1016/j.apenergy.2021.117985.
[120] Xu, S.; Li, Z.; Zhang, C.; Huang, Z.; Tian, J.; Luo, Y.; et al. A method of calculating urban-scale solar potential by evaluating and quantifying the relationship between urban block typology and occlusion coefficient: A case study of Wuhan in Central China. *Sustainable Cities and Society* **2021**, *64*, 102451. DOI: https://doi.org/10.1016/j.scs.2020.102451.
[121] Singh, R.; Banerjee, R. Estimation of rooftop solar photovoltaic potential of a city. *Sol Energy* **2015**, *115*, 589-602. DOI: https://doi.org/10.1016/j.solener.2015.03.016.
[122] Li, P.; Zhang, H.; Guo, Z.; Lyu, S.; Chen, J.; Li, W.; et al. Understanding Rooftop PV Panel Semantic Segmentation of Satellite and Aerial Images for Better Using Machine Learning. *Advances in Applied Energy* **2021**, *4*, 100057. DOI: https://doi.org/10.1016/j.adapen.2021.100057.
[123] Wang, R.; Camilo, J.; Collins, L. M.; Bradbury, K.; Malof, J. M. The poor generalization of deep convolutional networks to aerial imagery from new geographic locations: an empirical study with solar array detection. In *2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, Washington, DC, USA, 10-12

Oct. 2017, 2017; IEEE: pp 1-8.DOI: https://doi.org/10.1109/AIPR.2017.8457965.
[124] Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks ? In *28th Conference on Neural Information Processing Systems (NIPS)*, Montreal, CANADA, Dec 08-13, 2014; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K., Eds.; 2014; Vol. 27.