Solar Energy 240 (2022) 90–103

Contents lists available at ScienceDirect

Solar Energy

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

Modified U-Net based photovoltaic array extraction from complex scene in aerial infrared thermal imagery

Yu Shen [a,b], Tao Fan [c], Guangzhi Lai [c], Zhixiong Na [c], Hu Liu [c], Ziyao Wang [a,b], Yiye Wang [a,b], Yiping Jiao [a,b], Xinyi Chen [a,b], Zhouwei Lou [a,b], Jinxia Zhang [a,b,*], Kanjian Zhang [a,b], Haikun Wei [a,b,*]

[a] Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Nanjing, 210096, PR China
[b] School of Automation, Southeast University, Nanjing, 210096, PR China
[c] State Grid Corporation of China, Beijing, 100031, PR China

ARTICLE INFO

Keywords:
PV array extraction
Complex scene
Unmanned aerial vehicle
Hotspot
U-Net
Infrared thermal imagery

ABSTRACT

Hotspot detection in infrared thermal (IRT) images obtained by unmanned aerial vehicles (UAVs) is essential for the reliability, safety and efficiency of photovoltaic (PV) plants. We observe that some objects in the complicated environment, which are outside the region of PV arrays, might significantly interfere with the performance of automatic hotspot detection. This work focuses on the problem of PV array extraction from complex scene, which has not been solved in previous literature. Existing studies have attempted image processing methods to extract PV arrays. However, these approaches rely on manually designed features and parameters, which are impractical for complex scene. In this paper, a deep learning model, namely modified U-Net, is proposed to solve the problem. The proposed model can learn features and parameters automatically. Base on the classical U-Net, several innovations are provided: (1) batch normalization layers are adopted to alleviate internal covariate shift problems; (2) "He initialization" is used to increase the robustness and speed up the convergence; and (3) "RMSprop" is adopted to update parameters adaptively. 1211 IRT images of a PV plant containing complex background are collected to verify the effectiveness of the proposed method. Extensive experiments are conducted to compare the modified U-Net with five deep learning models and three image processing methods. The results demonstrate that the proposed method performs better than existing methods. Cooperating with the modified U-Net for PV array extraction from complex scene, even a simple hotspot detection method can achieve the accuracy of 99.79% and the F1 score of 0.9548.

1. Introduction

In order to utilize solar energy, the generation and installation of photovoltaic (PV) modules has been growing rapidly in the last decade. During transport, installation and subsequent stages, defects of PV modules may occur (IEA, 2018). These defects will cause power loss and degradation of PV modules, and even create safety issues (Köntges et al., 2014). Therefore, it is necessary to inspect and maintain PV modules regularly to increase the reliability, safety and performance during their service life time. When a PV module is defective, one cell or group of cells will be warmer than other parts. The overheated area is called "hotspot" (IEC, 2016). Appearing frequently in PV plants, the hotspot phenomena may not only indicate power loss but also accelerate degradation of PV modules (Niazi et al., 2019; Dunderdale et al., 2020a). With the development of unmanned aerial vehicles (UAVs) in recent years (Gallardo-Saavedra et al., 2018), it is practical to implement the infrared thermal (IRT) cameras on board UAVs to collect images efficiently for hotspot detection (Bommes et al., 2021).

Since automatic hotspot detection in aerial IRT images are attractive for large-scale PV plants (Huerta Herraiz et al., 2020a; Su et al., 2021), some researchers focused on how to detect hotspots in each PV module. The study (Niazi et al., 2019) extracted the texture and histogram of gradient features, and then used a Naive Bayes classifier to detect hotspots. The study (Akram et al., 2020) detected hotspots by transfer learning. The deep convolutional neural network (CNN) was pre-trained from electroluminescence images dataset and then fine-tuned on infrared images dataset. The study (Dunderdale et al., 2020b) presented a deep learning and feature-based approach to identify different defect types according to the patterns of hotspots in PV modules. The study (Ali et al., 2022) divided the IRT images of PV modules

* Corresponding authors at: School of Automation, Southeast University, Nanjing, 210096, PR China.
E-mail addresses: jinxiazhang@seu.edu.cn (J. Zhang), hkwei@seu.edu.cn (H. Wei).

https://doi.org/10.1016/j.solener.2022.05.017
Received 16 March 2021; Received in revised form 20 March 2022; Accepted 8 May 2022
Available online 23 May 2022

Nomenclature

Acronyms

PV	Photovoltaic
UAV	Unmanned aerial vehicle
IRT	Infrared thermal
CNN	Convolutional neural network
DBSCAN	Density-based spatial clustering of applications with noise
ROI	Region of interest
ReLU	Rectified linear unit
MIoU	Mean intersection over union
FWIoU	Frequency weighted intersection over union
CIoU	Class-wise intersection over union

Symbols

p	Predicted distribution
q	Annotated distribution
L_k	Cross-entropy loss of pixel k
L	Total cross-entropy loss
C_{rc}^n	Mean intensity of the grid cell of the rth row and cth column in the nth PV module
S_m^n	Mean intensity of the mth sub-string in the nth PV module
M^n	Mean intensity of the nth PV module

Parameters

Th_1	Threshold of grid cell
Th_2	Threshold of sub-string
Th_3	Threshold of PV module

into non-overlapping regions, and computed color image descriptors of these regions as features. Then the machine learning algorithm was used to classify the PV module into three classes including normal, hotspot, and defective. However, these methods require PV modules to be separated before hotspot detection.

There are also some works focused on PV module separation. The study (Huerta Herraiz et al., 2020b) adopted the region-based CNNs to detect PV modules. However, roughly representing the PV modules by the rectangular bounding boxes is imprecise. The study (Xu et al., 2021) considered the boundaries of photovoltaic modules as straight lines. These boundaries were extracted by edge detection, Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm and linear fitting. However, in complex scene, various kinds of objects in the environment appear to be interferential colors or edges in IRT images. The boundaries of some objects outside the region of PV arrays could also be straight lines, which may interfere with the separation of PV modules. Therefore, a more effective and practical way is extracting PV arrays from the complex background before separating PV modules. Without appropriate extraction of PV arrays, the PV module separation and hotspot detection algorithms mentioned above cannot work.

Several image processing methods have been designed to extract PV arrays. The study (Dotenco et al., 2016) designed a data-driven approach including normalization, thresholding, orientation estimation and refinement to extract PV arrays. The study (Kim et al., 2016) eliminated background by thresholding, removed the remaining noise by dilatation, filled in the holes and applied erosion to select the region of PV arrays in IRT images. The studies (Arenella et al., 2017; Carletti et al., 2020) recognized the boundaries of PV arrays by Hough transformation or Canny algorithm. The study (Vidal de Oliveira et al., 2019) adopted Gaussian filter to decrease noise, Laplacian filter to detect boundaries and threshold filter to extract the PV arrays. The study (Montanez et al., 2020) improved the contrast of images by the Equalization of Adaptive Limited Contrast Histogram technique, reduced noise by Gaussian filter, and then selected the region of PV arrays by an adaptive thresholding algorithm. The study (Wang et al., 2021) proposed an approach including the guided filter-based image enhancement, line detection and background subtraction to extract PV arrays. However, these image processing methods rely on features, parameters or thresholds, which need to be manually designed. Finding the appropriate parameters or thresholds usually takes a lot of time and effort. To make matters worse, it is hard to represent the PV array as simple features due to projective deformation, lens distortion and noise. Consequently, these methods are impractical for complex scene.

Fortunately, the development of computer vision and machine learning provides powerful end-to-end deep learning models, which perform well for image segmentation on public datasets (Brostow et al., 2009; Zhang, 2020). These models simulate the human visual mechanisms to extract features, and have the ability to update parameters automatically by learning from annotated training samples. The increasing installation of PV modules and the development of UAVs make it possible to collect more and more images for creating datasets. Therefore, deep learning models become practical for the inspection and monitoring of PV systems. Among those deep learning models, U-Net is a popular network, which has advantages on datasets with fixed image semantics and structures (Zhang et al., 2019). PV arrays appear as approximate quadrilateral regions with surrounding edges in IRT images. These features can be considered as fixed semantics or structures, which are suitable to be extracted based on U-Net.

The performance of U-Net depends on its structure, as well as the initialization and optimization of parameters. Based on these observations, we propose a deep learning method, namely modified U-Net, to extract PV arrays from complex scene. The contributions of our work are summarized as follows:

(1) **This work focuses on the problem of PV array extraction from complex scene, which has not been solved in previous literature.** The proposed method can learn features and parameters automatically from the training dataset, which is practical for PV array extraction from the complicated background.

(2) **Several innovations are provided to improve the performance of U-Net:** Batch normalization layers are adopted to alleviate internal covariate shift problems; ‘‘He initialization’’ is used to increase the robustness and speed up the convergence; An adaptive optimization method, namely “RMSprop” is adopted to update parameters.

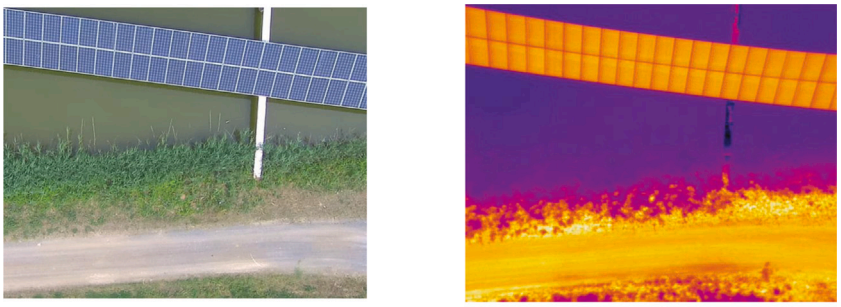
(3) Aiming at PV array extraction from complex scenes, 1211 IRT images containing complicated background are collected to verify the effectiveness of the proposed method. **The modified U-Net achieves state-of-the-art mean intersection over union (MIoU) of 99.12% for PV array extraction from complex scene** in aerial IRT images to the best of our knowledge.

(4) **This work makes a great contribution to the researches and applications of automatic hotspot detection in aerial IRT images.** Cooperating with the proposed approach for PV array extraction from complex scene, even a simple hotspot detection method can achieve the accuracy of 99.79% and the F1 score of 0.9548.

The rest parts of this paper are organized as follows. Basic concepts and the overall framework are introduced in Section 2. In Section 3, the PV array extraction approach using the modified U-Net is described in detail. The hotspot detection method is described in Section 4. Subsequent to this, experiments and results are reported and discussed in Section 5. Finally, conclusions are drawn in Section 6.

2. Basic concept

This section introduces basic concepts including infrared thermal imaging, complex scene, PV array extraction, U-Net and hotspot detection, which are helpful to understand the following sections.



(a) visible image (b) IRT image

Fig. 1. Visible and IRT images of PV modules. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.1. Infrared thermal imaging

Any object emit thermal radiation if its temperature is higher than 0 K. The spectral radiant emittance from the perfect radiator named black body is revealed by Planck's law (Hudson, 2006). A blackbody with higher temperature will emit more radiation. Take advantage of this property, IRT imaging cameras typically measure the infrared radiation between 8 and 14 μm to estimate the temperature of objects (Köntges et al., 2014).

For sources that are not blackbodies, the emissivity is defined by the ratio of the radiation W' of the material surface to the radiation W of a blackbody at the same temperature (Hudson, 2006), as shown in (1).

$$\text{emissivity} = \frac{W'}{W} \tag{1}$$

The emissivity of solar cell cover glass is around 0.85 to 0.9 if the angle of view is within 60° to 90° (Köntges et al., 2014; Subedi et al., 2019). However, the emissivity of aluminum framework is less than 0.2 in general (Wen and Chai, 2011). It means that the aluminum frameworks, which appear at the boundaries of PV modules, will be shown as lower temperature by IRT imaging cameras. Fig. 1 shows the visible and IRT images of PV modules. In Fig. 1(a), the solar cells covered by glass are shown in deep blue regions and the aluminum frameworks are shown in thick white edges surrounding PV modules. In Fig. 1(b), the PV modules are shown in orange regions and the aluminum frameworks are shown in deep orange edges surrounding PV modules. Due to projective deformation, lens distortion and noise, the PV modules appear as approximate quadrilaterals in IRT images although they are actually rectangles. These characteristics should be noticed for PV array extraction.

2.2. Complex scene and PV array extraction

PV plants may be constructed in a variety of environments, including but not limited to mountains, fish ponds and roofs. The IRT images captured by UAVs will not only contain PV modules, but also other irrelevant objects. These irrelevant objects with different temperature or emissivity, will be shown by varying colors or brightness in IRT images. The temperature of some objects in the environment may be even higher than PV modules, which will interfere with the process of automatic inspection of PV modules in IRT images. This kind of environment is called "complex scene".

For intelligent inspection (such as hotspot detection) of PV modules in complex scene, a reasonable idea is to extract PV arrays in IRT images as region of interest (ROI) and ignore other irrelevant regions. The IRT image captured by the UAV from complex scene is shown in Fig. 2(a), and PV arrays extracted as ROI are shown in Fig. 2(b). It is obvious that the interference of other objects in complex scene can be avoid by PV array extraction, which is the focus of this paper.

2.3. U-Net

U-Net is an end-to-end model for image semantic segmentation, which can predict the class of objects at every pixel. The architecture of U-Net consists of a contracting path to extract features from low level to high level, and a expanding path to output finer pixel-wise predictions (Ronneberger et al., 2015). The parameters of U-Net can be learned by annotated training samples, making it possible to extract features from images automatically.

Compared with other end-to-end models, U-Net has advantages on datasets with fixed image semantics and structures. As mentioned in Section 2.1, PV arrays appear as approximate quadrilaterals with surrounding edges in IRT images. These features can be considered as fixed semantics or structures, which are suitable to be extracted based on U-Net. However, the original U-Net is designed for biomedical image segmentation, and it needs to be adjust to achieve desired performance when used for PV array extraction from the complex background in IRT images. The modified U-Net for PV array extraction will be described in Section 3.

2.4. Hotspots detection

The temperature difference between hotspot and the normal area can be multiples of 10 K (IEA, 2018). With appropriate measurement conditions, hotspots can be found in IRT images by different color or brightness. IRT image patterns observed in outdoor measurements are summarized in Köntges et al. (2014) and Tsanakas et al. (2016).

According to Section 1, there are three important issues for hotspot detection in aerial IRT images. The first one is PV array extraction from complex scene, the second one is PV module separation, and the third one is detecting overheated regions in each PV module. In this paper, we mainly focus on the extraction of PV arrays in IRT images collected by UAVs. Base on the concepts described in this section, we propose a novel approach for PV array extraction using the modified U-Net and then apply it for hotspot detection in aerial IRT images. The framework is shown in Fig. 3 and more details will be described in Sections 3 and 4.

3. PV array extraction

Inspired by the characteristics of U-Net and new insights in the field of machine learning, we design a modified U-Net for PV array extraction from the complex background in IRT images. The architecture of the modified U-Net is described in Section 3.1. Next the training and prediction process is described in Section 3.2. Then, the improvements of the proposed model compared with the original U-Net are summarized in Section 3.3.

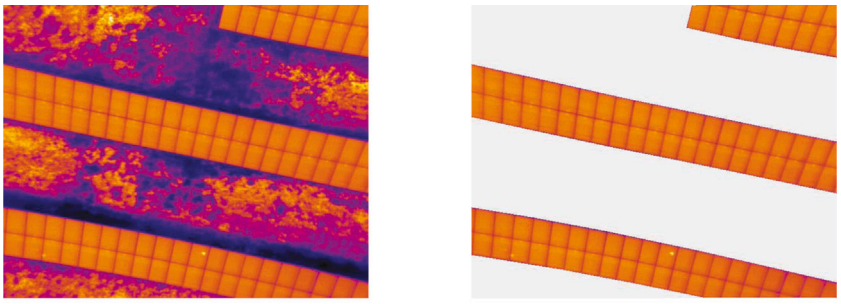


Fig. 2. Complex scene and ROI. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

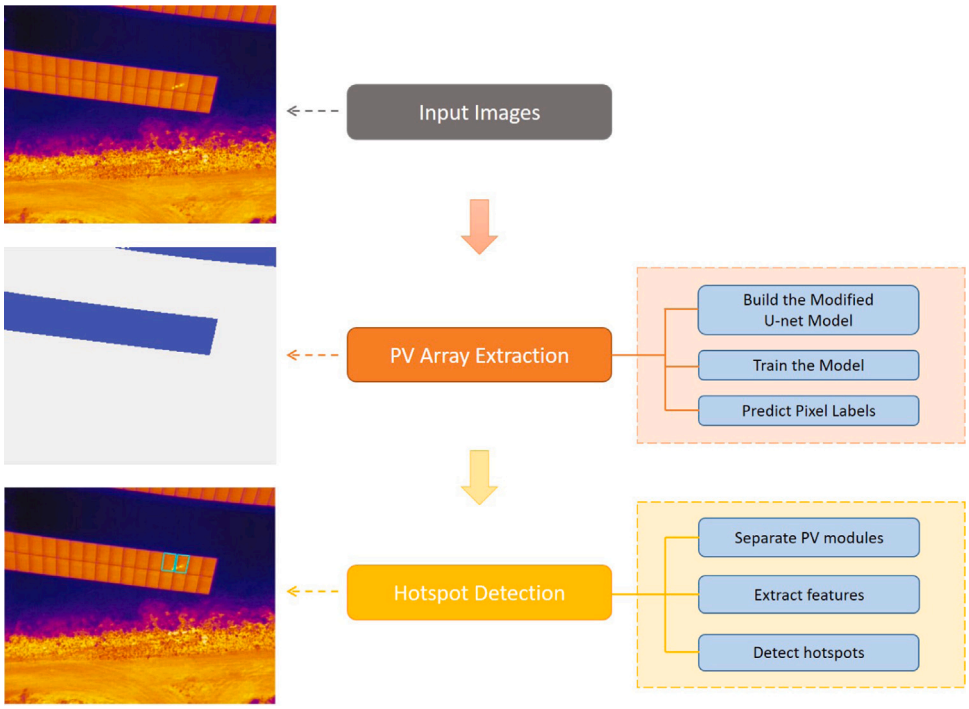


Fig. 3. The framework of PV array extraction and hotspot detection.

3.1. The architecture of the modified U-Net

The architecture of our modified U-Net is shown in Fig. 4. This model consists of a contracting path (left side in Fig. 4) to extract high-level features and a expansive path (right side in Fig. 4) to output pixel labels. At each level, feature maps are calculated by a convolutional layer with zero-padding, normalized by a batch normalization layer and then nonlinearized by a Rectified Linear Unit (ReLU) layer.

On the contracting path, the height and width of feature maps are both halved by a maxpooling layer from low level to high level. On the expansive path, the height and width of feature maps are both doubled by an up-sampling layer from high level to low level. Feature maps obtained by up-sampling and copied from the left side are concatenated together to provide more accurate output. The feature maps at level 2 on the right are converted to 2 channels, corresponding to 2 classes: PV array and background, respectively. The softmax function is used to activate the feature maps and produce probabilistic results. Pixel labels are predicted to be the class with the highest probability and then resized to 512 × 640, as output segmentation images. Different operation layers in Fig. 4 are described as follows:

Convolutional layers are essential for feature extraction. They compute feature maps corresponding to filters/kernels with learnable parameters. The size of filters is 3 × 3 in our model. We adopted zero-padding to keep the size of images unchanged. The operation of the convolutional layer is illustrated in Fig. 5. To simplify the visualization, we only show one channel of the input.

The kernel convolves with each channel of the input and generates the corresponding output matrix. Then the feature map is obtained by adding all these output matrices. It should be noticed that the parameters of the kernel are learnable and each kernel results in one feature map. In the feature map, the element with larger value indicates that its feature is more significant.

Rectified Linear Unit (ReLU) layers applies the activation function, as shown in (2), to each element in feature maps computed by convolutional layers, for non-linearity of the model.

$$R(x) = \max(0, x) \tag{2}$$

As shown in Fig. 6, ReLU activation function simulates the activation or inhibition mechanism of human neurons: If the value of an element in the feature map is larger than zero, the element is activated; else if the value of an element in the feature map is less than zero, the element is inhibited.

Batch normalization layers convert input features to normalized features with mean zero and variance one according to mini-batches of

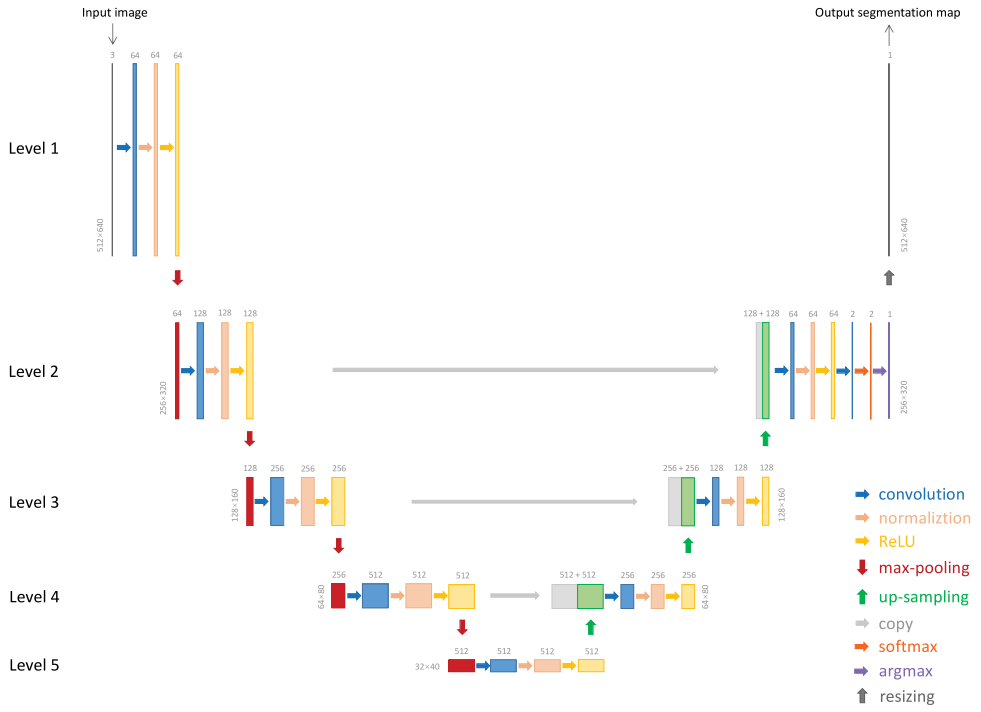


Fig. 4. The architecture of the modified U-Net. (The arrows represent different operations shown in the lower right corner. The boxes correspond to feature maps. The number of channels is provided on top of boxes. The height and width of feature maps keep invariant at each level, denoted at the left or right side of boxes.)

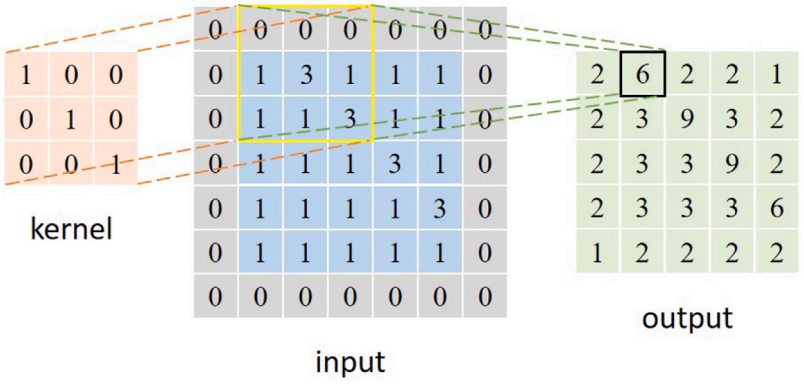


Fig. 5. The operation of the convolutional layer. (The size of the input is simplified as 5 × 5 for visualization. Each blue box denotes an element of the input. The gray boxes surrounding the input show zero-padding. The orange boxes denote parameters of the kernel, which can be considered as a matrix. And the numbers in the yellow bounding box in the input can be considered as another matrix. The dot product of these two matrices generates the value in the black bounding box in the output. It can be noticed that the location of the black bounding box in the output corresponds to the center of the yellow bounding box in the input.) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

training samples (Ioffe and Szegedy, 2015), as shown in (3),

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \epsilon}} \tag{3}$$

where x is the element in the feature map, $\hat{x}$ is the normalized element in the feature map, μ is the mean value of x over mini-batches, σ^2 is the variance of x over mini-batches and ϵ is a small constant for numerical stability.

It can be observed in Fig. 5 that during the operation of the convolutional layer, the output generated by each channel of the input depends on the parameters of the kernel. Consequently, the elements in the feature map are also influenced by the parameters of the corresponding kernel. During the training process, updating the parameters of the kernel would cause the change in the distribution of its corresponding feature maps over training samples, which is known as internal covariate shift (Ioffe and Szegedy, 2015).

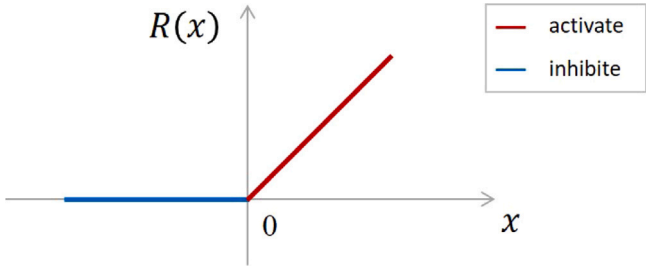


Fig. 6. ReLU activation function.

However, ReLU layers are expected to activate the elements with significant features, meanwhile, inhibite other elements. If the elements in feature maps over training samples have zero means and unit variances, the ReLU layers would work more effectively and the model

would perform better. Therefore, we insert batch normalization layers after convolutional layers and before ReLU layers. Batch normalization is a good way to reduce internal covariate shift and alleviate problems caused by bad initialization of parameters (Ioffe and Szegedy, 2015). It also benefits the training process.

Maxpooling layers are in charge of down-sampling. In each 2×2 receptive field, only the max element is retained, other 3 elements are discarded. There is no overlap between different receptive fields with the stride of 2. After maxpooling, the height and width of feature maps will be halved.

Up-sampling layers double the height and width of feature maps simply by repeating the elements.

Concatenate layers combine feature maps computed by up-sampling layers with copies of feature maps at the same level on the left side of the model. It can alleviate the information loss caused by preceding operation.

Softmax layer converts feature maps to probabilistic results, as shown in (4),

$$p = \frac{e^{f_i}}{\sum_j e^{f_j}} \tag{4}$$

where i, j represent different classes and f_i, f_j denote calculated values in the previous layer.

Argmax layer predicts pixel labels to be the one with the highest probability.

Resizing layer resizes pixel labels to 512×640 as output segmentation images.

3.2. Train the model and predict pixel labels

The model described in Section 3.1 has trainable parameters in convolutional layers. The weights of kernels or filters and biases need to be initialized before training the model. In the training stage, for input images, the modified U-Net provides corresponding output segmentation images with pixel labels. Then the loss function compares the difference between predicted pixel labels and annotated pixel labels. This loss can be propagated from latter layers to previous layers by partial derivative and the chain rule, namely "back propagation". Next, parameters are updated by the optimization method. With updated parameters, the model provides new predicted pixel labels. The loss function is calculated again. Parameters are optimized and updated once more. This loop will be executed for a specified number of epochs and stop. In the test stage, the modified U-Net provides predicted pixel labels according to test images. The training and prediction process is illustrated in Fig. 7.

Some essential steps are explained as follows:

Initialization: Before training the model, all the parameters need to be initialized. It is common practice to initialize CNNs by small values randomly drawn from Gaussian distribution with fixed variance. But for deep models, the highly non-linear property calls for more appropriate initialization method to accelerate the training process and avoid disconvergence. In our work, a robust initialization method considering the ReLU activation, called "He initialization", He et al. (2015) is adopted to randomly generate values from a uniform distribution $[-l, l]$ defined by (5),

$$l = \sqrt{6/\text{input\_num}} \tag{5}$$

where input_num represents the number of inputs in the weight tensor in each layer.

This method is designed for deep neural networks with Rectified activation units. It provides good initialization of parameters, which can speed up the convergence, increase the robustness of deep models and improve the accuracy as well (He et al., 2015).

Loss function: Since the softmax classifier is used in the model, we adopt the cross-entropy loss to evaluate the difference between predicted pixel labels and annotated pixel labels. The cross-entropy (Li, 2016) between two distributions: p and q is shown in (6),

$$H(q, p) = -\sum_x q(x) \log p(x) \tag{6}$$

where p represents the predicted distribution shown in (4) and q denotes the annotated distribution. More detailedly, if y_k is the class label of pixel k, $p = [0, \ldots, 1, \ldots, 0]$ with a single 1 at the y_kth position. As a result, the cross-entropy loss of pixel k is defined as (7).

$$L_k = -\log\left(\frac{e^{f_{y_k}}}{\sum_j e^{f_j}}\right) = -f_{y_k} + \log\sum_j e^{f_j} \tag{7}$$

The total cross-entropy loss is defined as (8).

$$L = \sum_k L_k \tag{8}$$

Optimization: The simplest optimization method is to change parameters with a fixed learning rate. But updating all the parameters equally and globally faces the trade-off between stability and efficiency. In deep neural networks, the appropriate learning rates may vary widely (Hinton, 2012). With a small learning rate, the training process will be time-consuming. With a large learning rate, the model may not converge. To solve this problem, various adaptive learning methods have been proposed. In our work, an effective and robust optimization method called "RMSprop" (Hinton, 2012) is adopted. It updates parameters adaptively using a global learning rate divided by the moving average of squared gradients (Li, 2016), as shown in (9) and (10),

$$ms(w, t) = \text{decay\_rate} \cdot ms(w, t-1) + \left(1 - \text{decay\_rate}\right) \cdot \left(\frac{\partial L}{\partial w}(t)\right)^2 \tag{9}$$

$$w = w - \text{learning\_rate} \cdot \frac{\partial L}{\partial w}(t) / (\sqrt{ms(w, t)} + \text{eps}) \tag{10}$$

where $\text{ms}(w, t)$ represents the moving average of the squared gradient for each weight; w is the weight to be updated; decay_rate is a hyperparameter set as 0.9; learning_rate is set as 0.001; eps is a small constant for numerical stability, set as 10^{-7}.

RMSProp algorithm adjusts the learning rate of each weight based on the magnitudes of its gradients, which has a beneficial effect for the optimization of parameters (Li, 2016).

3.3. The improvements of the proposed model

Compared with the original U-Net, the improvements of the proposed model are summarized as follows.

Batch normalization: In terms of the architecture of the model, batch normalization layers are inserted after convolutional layers and before ReLU layers. They convert input features to normalized features with mean zero and variance one. Batch normalization can alleviate problems caused by internal covariate shift and/or bad initialization. It can also accelerate the learning of parameters.

Initialization: In the beginning of training process, a robust initialization method designed for deep neural networks with Rectified activation units, called "He initialization", is adopted to randomly generate values from a uniform distribution. It provides good initialization of parameters, which can speed up the convergence, increase the robustness of deep models and improve the accuracy as well.

Optimization: As the core of the training process, an adaptive optimization method called "RMSprop" (Hinton, 2012) is adopted to update parameters using a global learning rate divided by the moving average of squared gradients. This optimization method can not only accelerate the training process, but also ensure the stability when the parameters of convolutional layers are updated.

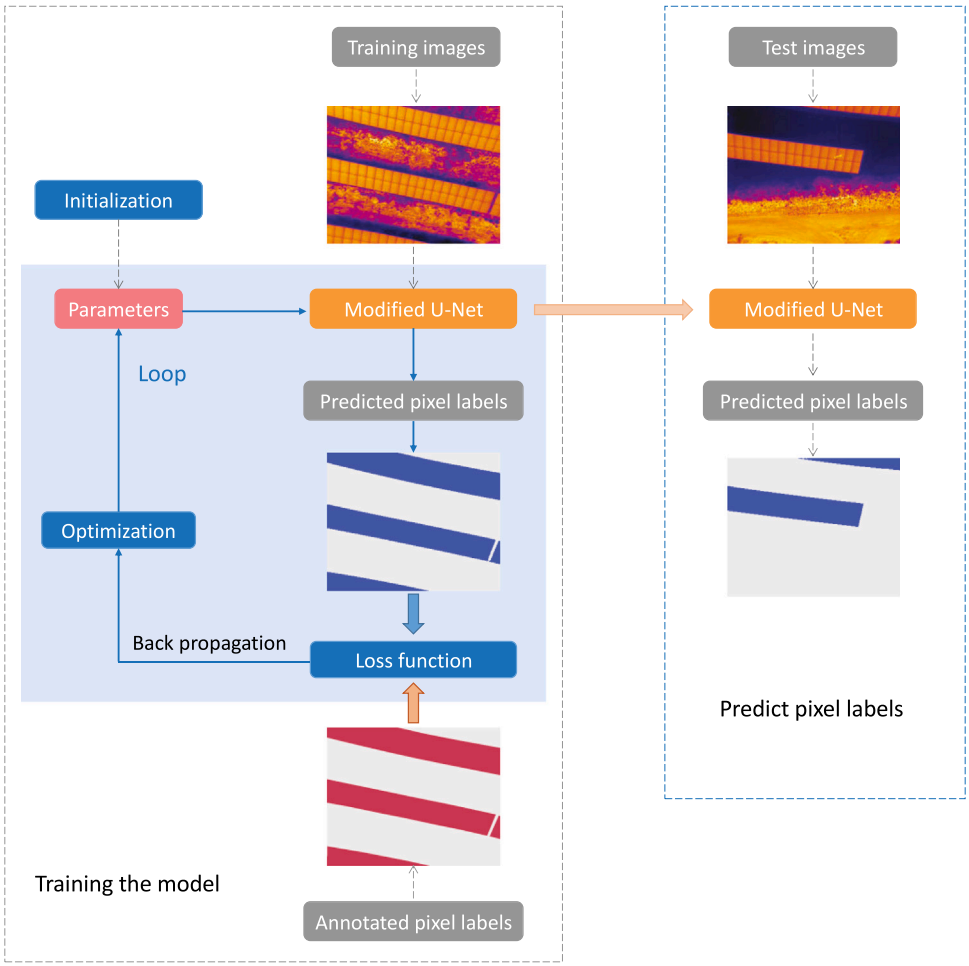


Fig. 7. Training and prediction.

4. Hotspot detection

In order to verify the validity of PV array extraction, a simple method for hotspot detection is designed in this section. Since we mainly focus on PV array extraction in this paper, the hotspot detection method is described in brief. Firstly, PV arrays are extracted by the proposed approach shown in Section 3. Secondly, PV modules are separated, as described in Section 4.1. Thirdly, hotspots are detected in each PV module, as described in Section 4.2.

4.1. PV module separation

For the IRT image shown in Fig. 8(a), its pixel labels are predicted in Fig. 8(b) using the proposed method described in Section 3. Subsequently, the PV modules are separated using the approach (Xu et al., 2021) including the following steps:

Step 1: PV array near the bounds of the image is ignored, as shown in Fig. 8(c).

Step 2: The boundary of the remaining PV array is detected by Canny algorithm and the longest line is found by Hough transformation as shown in Fig. 8(d).

Step 3: The thermal image is converted to the grayscale image, and the PV array is extracted according to Fig. 8(e).

Step 4: The PV array is rotated according to the direction of the longest line, and the intensity of the PV array is inverted as shown in Fig. 8(f).

Step 5: Peaks are found by horizontally scanning, as shown in Fig. 8(g).

Step 6: The points near the upper and lower bounds of each array are ignored and the remaining points are clustered using DBSCAN algorithm, as shown in Fig. 8(h).

Step 7: Linear fitting is adopted for the points of each cluster, and these lines intersect the boundary of each array, as shown in Fig. 8(i).

Step 8: The horizontal boundaries between two adjacent vertical boundaries are obtained in the same way, as shown in Fig. 8(j).

Step 9: The incomplete PV modules near the bound of the image are ignored, the masks of PV modules are shown in Fig. 8(k).

Step 10: The masks of PV modules in the IRT image can be obtained by rotating Fig. 8(k), as shown in Fig. 8(l).

4.2. Hotspot detection in each PV module

Each PV module is divided into grid cells, shown in Section 4.2.1. Then multi-level features are extracted, described in Section 4.2.2. Next, PV modules with overheated regions are detected, shown in Section 4.2.3.

4.2.1. Grid cell segmentation

Each PV module is divided into grid cells by the following steps.

Step 1: PV modules in IRT images can be represented as quadrilaterals. The vertexes of each PV module are found by the mask and the quadrangular boundaries are shown in Fig. 9. Module 1 is normal, while modules 2 and 3 present hotspots.

Step 2: Rectangular PV modules are recovered by removing perspective distortion (Hartley and Zisserman, 2004).

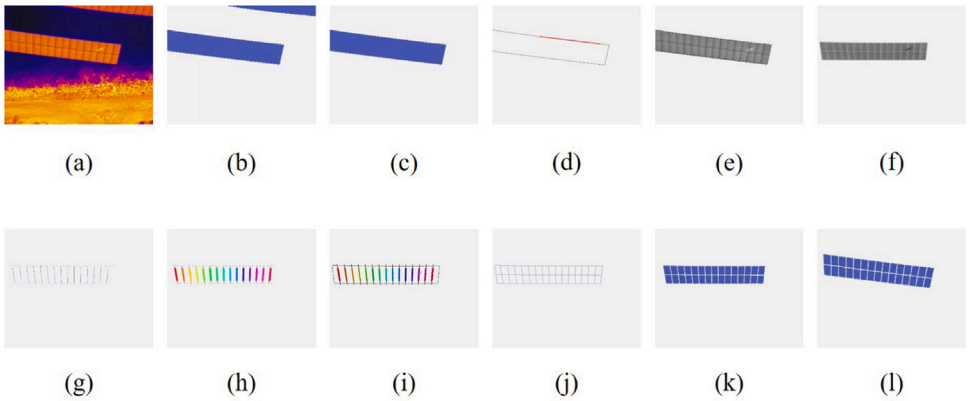


Fig. 8. Separation of PV modules.

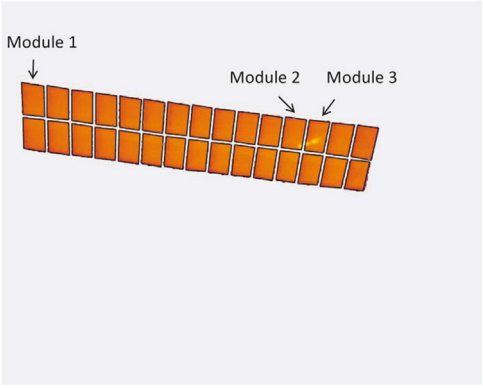


Fig. 9. Quadrangular boundaries of separate PV modules.

Step 3: Each rectified PV module is divided into 10 × 6 grid cells. The rectified PV modules 1, 2, 3 and their segmentation of grid cells are shown in Fig. 10.

4.2.2. Feature extraction

Inspired by the study (Dotenco et al., 2016; Kim et al., 2016), we extract multi-level features to detect hotspots. For each PV module, the following features are calculated.

Grid cell level: Mean intensity of grid cells in the nth PV module are represented as C_{ij}^{n}, $i = 1, 2, \ldots, 10$; $j = 1, 2, \ldots, 6$. These features are adopted to identify overheated cells.

Sub-string level: Mean intensity of sub-strings in the nth PV module are denoted as S_{k}^{n}, $k = 1, 2, 3$. These features are useful to detect overheated sub-strings.

Module level: Mean intensity of the nth PV module is represented as M^{n}, $n = 1, 2, \ldots, N$, where N is the number of PV modules in the IRT image. These features are adopted to identify overheated modules.

4.2.3. Hotspot detection

Based on the features extracted above, we design the hotspot detection algorithm including the following steps.

Step 1: Set $n = 1$.

Step 2: Iterate through each grid cell of the nth PV module, if there exist i and j that satisfy $C_{ij}^{n} - M^{n} > Th_1$ (Th_1 is the threshold, set as 25), this cell is regarded as hotspot.

Step 3: If the kth sub-string of the nth PV module satisfies $S_{k}^{n} - M^{n} > Th_2$ (Th_2 is the threshold, set as 10), this sub-string is considered to be overheated.

Step 4: If the nth PV module satisfies $M^{n} - \overline{M} > Th_3$ ($\overline{M}$ is the mean intensity of the total N modules in the IRT image, and Th_3 is the threshold, set as 15), this module is regarded as overheated.

Step 5: If $n < N$, let $n = n + 1$ and skip to step 2; else stop.

Table 1
Dataset information — samples.

	Positive	Negative	Total
Number	675	536	1211

Table 2
Dataset information — PV modules.

	Normal	Hotspot	Total
Number	8918	210	9128

Table 3
Software and version.

Software	Python	Conda	Tensorflow	Keras	Opencv-python
Version	3.6.10	4.8.2	2.2.0	2.4.3	4.2.0

5. Experiments and results

5.1. Experimental preparation

5.1.1. Dataset

IRT images are captured using the thermographic camera FLIR Vue Pro (640, 19 mm) with the unmanned aerial vehicle from a PV plant in Jiangsu, China. These images are obtained when the sky is clear without clouds. The solar irradiance are higher than 800 W/m^2. We collect 1211 IRT images in total, including 675 positive samples and 536 negative samples. PV modules are present in positive samples and absent in negative samples, as shown in Fig. 11(a) and (b) respectively.

All these samples are split into three folds for cross validation. In subsequent experiments, one fold is regarded as the test set and the other two are used as training set. The dataset information about samples is listed in Table 1. In addition, we counted PV modules appeared in these samples. There are 9128 complete modules in total, including 8918 normal modules and 210 modules with hotspots. This information is listed in Table 2.

5.1.2. Software

Python is used as the programming language, which is installed and managed by Conda. In addition, two open-source frameworks for machine learning, namely TensorFlow and Keras are adopted. The open-source computer vision library known as OpenCV-python is also utilized. The versions of related software are listed in Table 3.

5.2. PV array extraction

Firstly, the ground truth of pixels is labeled by the open annotation tool called LabelMe. All the pixels are classified into two classes: PV

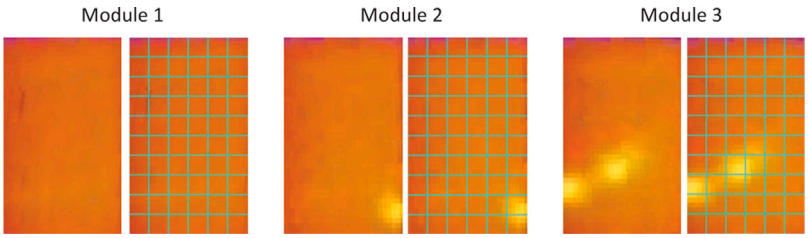
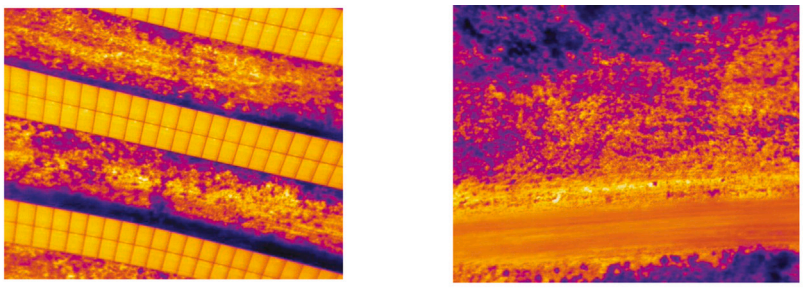
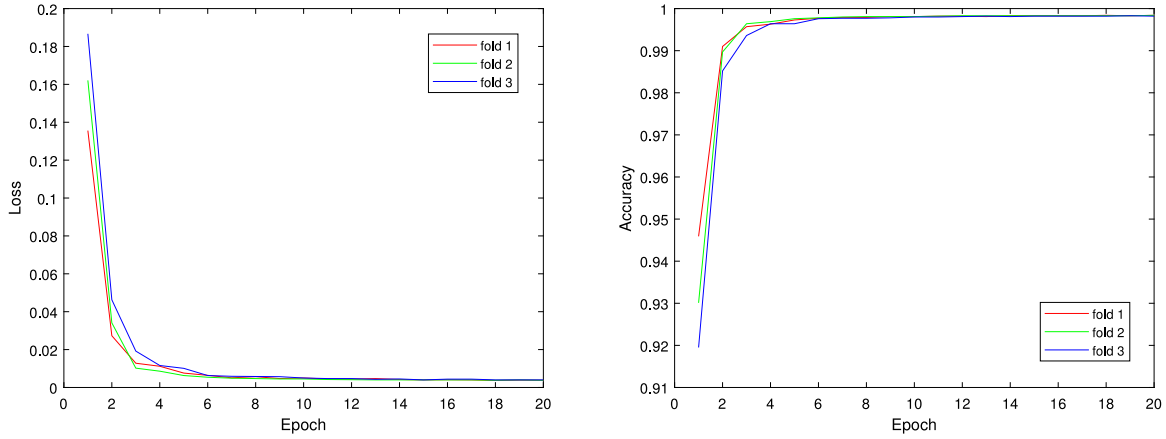


Fig. 10. Rectified PV modules and the segmentation of grid cells.



(a) positive sample (with PV modules) (b) negative sample (without PV modules)

Fig. 11. The positive sample and the negative sample.



(a) training loss (b) accuracy

Fig. 12. The training loss and accuracy with epoches for three folds.

array and background. Next, we use our modified U-Net for PV array extraction from complex scene.

The training loss and accuracy with epoches of our modified U-Net are shown in Fig. 12. The training loss/accuracy curves are smooth and gradually convergent, which implies that our modified U-Net is robust for PV array extraction.

We use mean intersection over union (MIoU), frequency weighted intersection over union (FWIoU) and class-wise intersection over union (CIoU) in this section for performance evaluation of PV array extraction.

Mean intersection over union (MIoU) is the most commonly used evaluation for semantic segmentation, defined as the average intersection over union between ground truth pixel labels and predicted pixel labels (Long et al., 2015), as shown in (11),

$$\mathrm{MIoU} = \frac{1}{k+1}\sum_{i=0}^{k}\frac{P_{ii}}{\sum_{j=0}^{k}P_{ij}+\sum_{j=0}^{k}P_{ji}-P_{ii}} \tag{11}$$

where k denotes the number of classes; P_{ij} represents the number of pixels with ground truth class i predicted as class j.

Frequency weighted intersection over union (FWIoU) is calculated according to the frequency, namely the number of pixels for each class (Long et al., 2015), as shown in (12).

$$\mathrm{FWIoU} = \frac{1}{\sum_{i=0}^{k}\sum_{j=0}^{k}P_{ij}}\cdot\sum_{i=0}^{k}\frac{P_{ii}\cdot\sum_{j=0}^{k}P_{ij}}{\sum_{j=0}^{k}P_{ij}+\sum_{j=0}^{k}P_{ji}-P_{ii}} \tag{12}$$

Class-wise intersection over union (CIoU) is defined to evaluate the performance for each class (Schneider et al., 2017), as shown in (13).

$$\mathrm{CIoU}(i) = \frac{P_{ii}}{\sum_{j=0}^{k}P_{ij}+\sum_{j=0}^{k}P_{ji}-P_{ii}} \tag{13}$$

5.2.1. Comparison with existing deep learning models

Five existing deep learning models for semantic segmentation are adopted for comparison, namely FCN (Long et al., 2015; Shelhamer et al., 2017), SegNet (Badrinarayanan et al., 2017), U-Net (Ronneberger et al., 2015), Res U-Net (Zhang et al., 2019) and VGG U-Net (Sovetkin

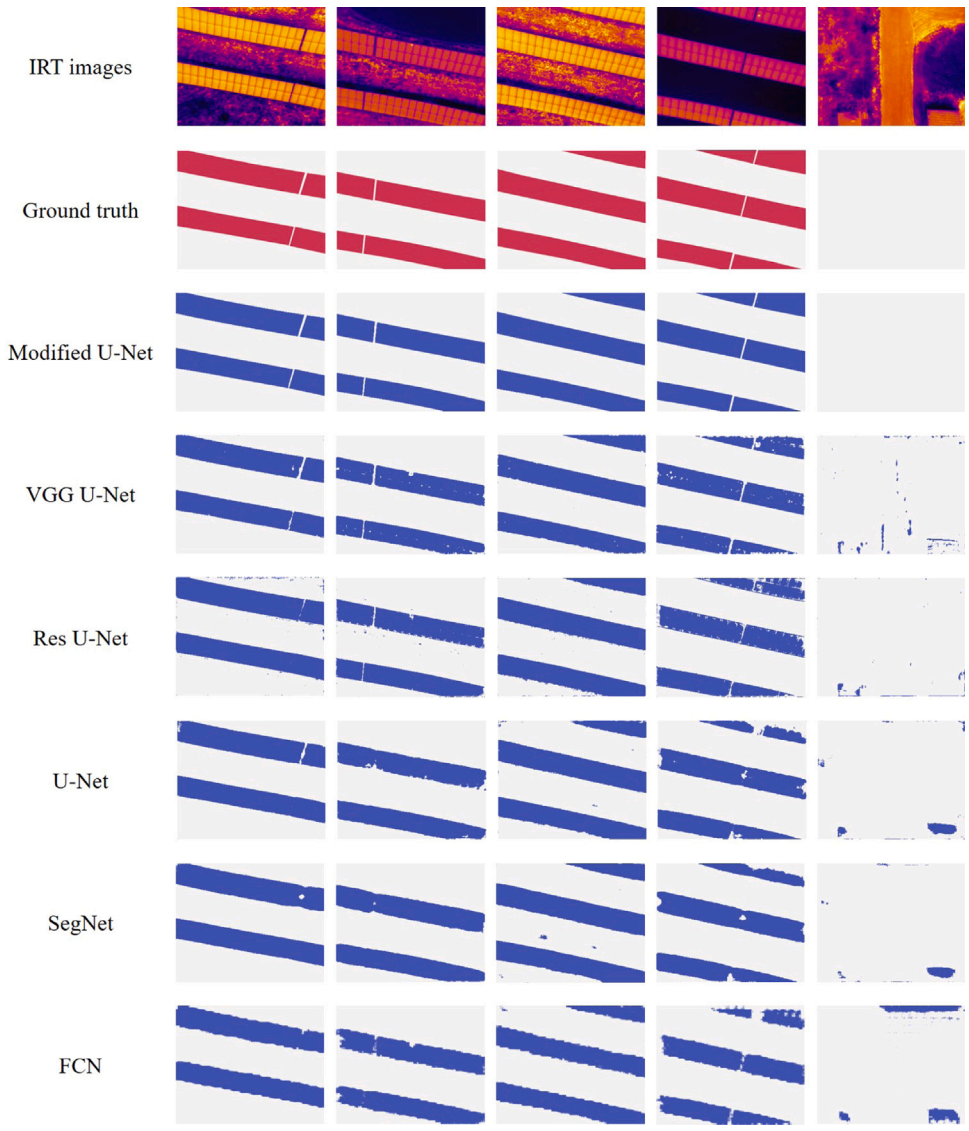


Fig. 13. PV array extraction from complex scene by deep learning models.

et al., 2021). All the samples are split into three folds for cross validation. One fold is regarded as the test set and the other two are used as training set.

The performance evaluation is shown in Table 4. Among these deep learning methods, U-Net works better than FCN and SegNet. Meanwhile, Res U-Net and VGG U-Net outperform the original U-Net owing to transform learning. Our modified U-Net achieves the highest MIoU, FWIoU and CIoUs for PV array extraction from complex scene in aerial infrared thermal imagery.

The results of PV array extraction by different deep learning methods are shown in Fig. 13. Each column represents a sample. Positive samples are shown in the left four columns and the negative sample is shown in the rightmost column. The infrared thermal images are shown in the first row and the ground truth is shown in the second row. PV arrays extracted by different methods are shown from the third to eighth rows. Although existing deep learning models can learn to extract features automatically from different kinds of complex environment, the predicted pixel labels of these models present some obvious mistakes.

The mainly weaknesses of these existing deep leaning models are summarized as follows: The PV arrays are extracted incompletely, especially for the results of FCN, SegNet and U-Net in the fourth column in Fig. 13; The narrow interval between two adjacent PV arrays is prone to be misrecognized as the region of PV array; Internal holes appear in the extracted PV array, especially for the results of VGG U-Net and Res U-Net; Some region in the background would be misidentified as the PV array, shown in the rightmost column in Fig. 13.

Compared with these deep learning models, the predicted pixel labels of the modified U-Net are consistent with the ground truth. The results verify that the proposed model works well and performs better than these existing deep learning models.

5.2.2. *Comparison with image processing methods*

The proposed model is also compared with three image processing methods (Kim et al., 2016; Montanez et al., 2020; Wang et al., 2021). The performance evaluation of these image processing methods are shown in Table 5. The average MIoU, FWIoU, CIoU for background and CIoU for PV array of the modified U-Net are 0.9912, 0.9960, 0.9977 and 0.9847 respectively, as shown in Table 4. Compared with Table 5, it can be found that the modified U-Net significantly outperforms these image processing methods.

The results of PV array extraction by the proposed model and the image processing methods (Kim et al., 2016; Montanez et al., 2020; Wang et al., 2021) are compared in Fig. 14. Each column represents a sample. Positive samples are shown in the left four columns and the negative sample is shown in the rightmost column. The infrared

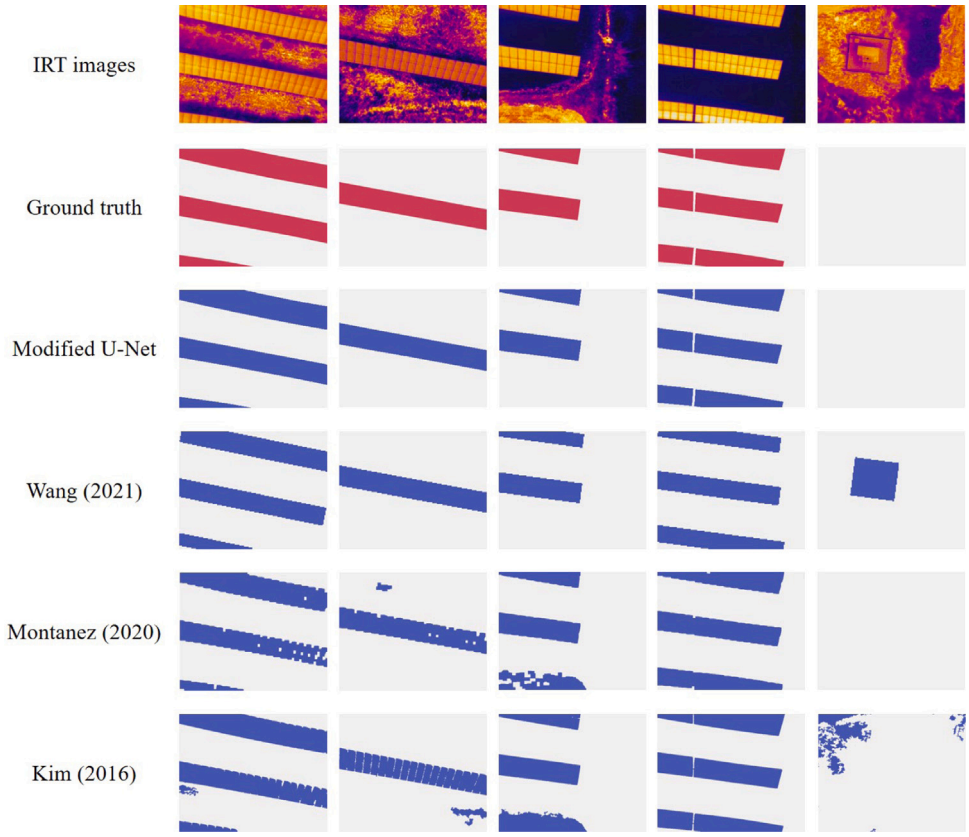
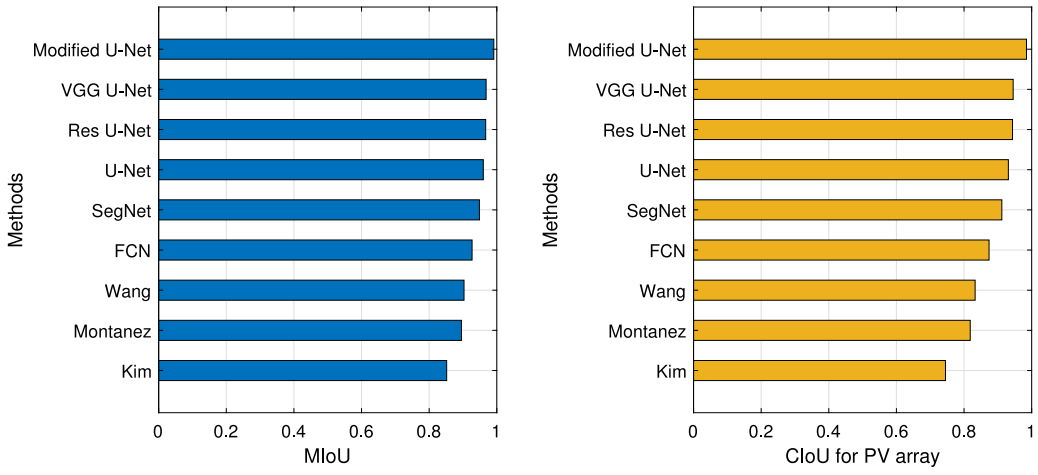


Fig. 14. PV array extraction from complex scene by the modified U-Net and image processing methods.



(a) bar graph of MIoU (b) bar graph of CIoU for PV array

Fig. 15. Comparison of MIoU and CIoU for PV array.

thermal images are shown in the first row and the ground truth is shown in the second row. PV arrays extracted by different methods are shown from the third to sixth rows.

The method (Kim et al., 2016) eliminates the background by thresholding. However, it is difficult to find the fixed threshold in complex scenes. Consequently, some regions in the background with high intensity would be misidentified as the PV array while some region in the PV array with low intensity would be misclassified as the background. The method (Montanez et al., 2020) uses an adaptive thresholding algorithm, performing better than the fixed threshold. However, there is still some misidentification between the PV array and the background. The method (Wang et al., 2021) represents the PV array as rectangles, which can avoid the distraction of complex background in most cases. However, the rectangular object in the background would be misidentified as the PV array. Besides, the PV array is not a standard rectangle in the IRT image due to projective deformation and lens distortion. As a result, there would be some errors at the edges of PV arrays.

Compared with these image processing methods, the proposed modified U-Net can learn to update parameters automatically from training samples. The predicted pixel labels of the modified U-Net are consistent

Table 4
Performance of deep learning models.

	Model	MIoU	FWIoU	CIoU	
				Background	PV array
Fold 1	FCN	0.9257	0.9659	0.9798	0.8715
	SegNet	0.9479	0.9760	0.9856	0.9103
	U-Net	0.9580	0.9808	0.9887	0.9272
	Res U-Net	0.9653	0.9842	0.9906	0.9400
	VGG U-Net	0.9705	0.9867	0.9922	0.9487
	Modified U-Net	**0.9914**	**0.9961**	**0.9978**	**0.9850**
Fold 2	FCN	0.9303	0.9671	0.9806	0.8801
	SegNet	0.9505	0.9764	0.9859	0.9151
	U-Net	0.9634	0.9828	0.9898	0.9371
	Res U-Net	0.9696	0.9856	0.9915	0.9477
	VGG U-Net	0.9668	0.9845	0.9910	0.9426
	Modified U-Net	**0.9910**	**0.9958**	**0.9976**	**0.9844**
Fold 3	FCN	0.9246	0.9645	0.9788	0.8703
	SegNet	0.9475	0.9752	0.9851	0.9099
	U-Net	0.9588	0.9807	0.9886	0.9290
	Res U-Net	0.9667	0.9845	0.9908	0.9427
	VGG U-Net	0.9680	0.9852	0.9914	0.9446
	Modified U-Net	**0.9912**	**0.9959**	**0.9977**	**0.9847**
Average	FCN	0.9269	0.9659	0.9797	0.8740
	SegNet	0.9486	0.9759	0.9855	0.9118
	U-Net	0.9601	0.9814	0.9890	0.9311
	Res U-Net	0.9672	0.9848	0.9910	0.9434
	VGG U-Net	0.9684	0.9855	0.9915	0.9453
	Modified U-Net	**0.9912**	**0.9960**	**0.9977**	**0.9847**

Table 5
Performance of image processing methods.

Method	MIoU	FWIoU	CIoU	
			Background	PV array
Kim et al. (2016),	0.8518	0.9305	0.9585	0.7451
Montanez et al. (2020)	0.8955	0.9524	0.9727	0.8183
Wang et al. (2021)	0.9031	0.9549	0.9733	0.8328

with the ground truth, which indicates that the proposed model works well and performs better than these image processing methods.

5.2.3. Discussion

MIoU is the most commonly used evaluation for semantic segmentation, and CIoU for PV array is closely related to the performance of PV array extraction from complex scene in IRT images. Therefore, these two indicators of different methods are visualized and compared in Fig. 15. Obviously the proposed modified U-Net achieves the highest MIoU and CIoU for PV array.

The image processing methods (Kim et al., 2016; Montanez et al., 2020; Wang et al., 2021) rely on manual designed features, parameters or thresholds, which are unable to learn from the large available IRT images obtained by UAVs. The accuracy of these methods is limited compared with deep learning models.

Among the deep learning models, U-Net performs better than SegNet and FCN because its concatenate layers alleviate the information loss. VGG U-Net and Res U-Net improve the performance of U-Net by transfer learning to obtain better initialization of parameters. However, they replace the base model of U-Net with VGG-Net and ResNet respectively, which have more layers and parameters. As a result, training these models requires more annotated samples and more time. The proposed model modifies the U-Net in other ways. The modified U-Net adopts batch normalization layers to alleviate internal covariate shift problems, "He initialization" to increase the robustness and "RMSprop" to update parameters adaptively. The experimental results show that the modified U-Net satisfactorily solves the problem of PV array extraction from complex scene, which outperforms the existing methods.

Table 6
Results of hotspot detection.

True positive	True negative	False positive	False negative
201	8908	10	9

Table 7
Performance evaluation of hotspot detection.

Accuracy	Precision	Recall	F1 score
99.79%	0.9526	0.9571	0.9548

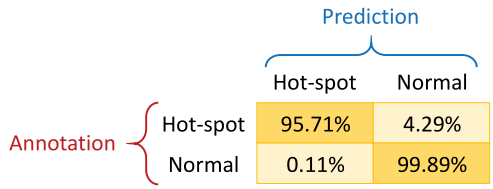


Fig. 16. Confusion matrices for hotspot detection.

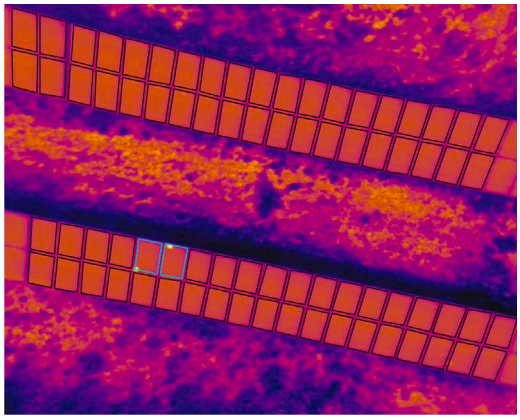


Fig. 17. Detect PV modules with hotspots in complex scene.

5.3. Hotspot detection

Accuracy, Precision, Recall, F1 score and confusion matrix are used in this section for performance evaluation of hotspot detection. They are defined as (14) to (17),

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{N} \times 100\% \tag{14}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{15}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{16}$$

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{17}$$

where True Positive denotes the number of hotspot samples predicted correctly; True Negative represents the number of normal samples predicted correctly; False Positive denotes the number of normal samples mis-classified as hotspot; False Negative represents the number of hotspot samples mis-classified as normal; N is the total number of samples.

The results of hotspot detection are shown in Table 6. True Positive and True Negative are much larger than False Positive and False Negative, which means that most of the PV modules are classified correctly.

The performance evaluation is shown in Table 7. The accuracy is 99.79% and the F1 score is 0.9548, indicating that the simple hotspot

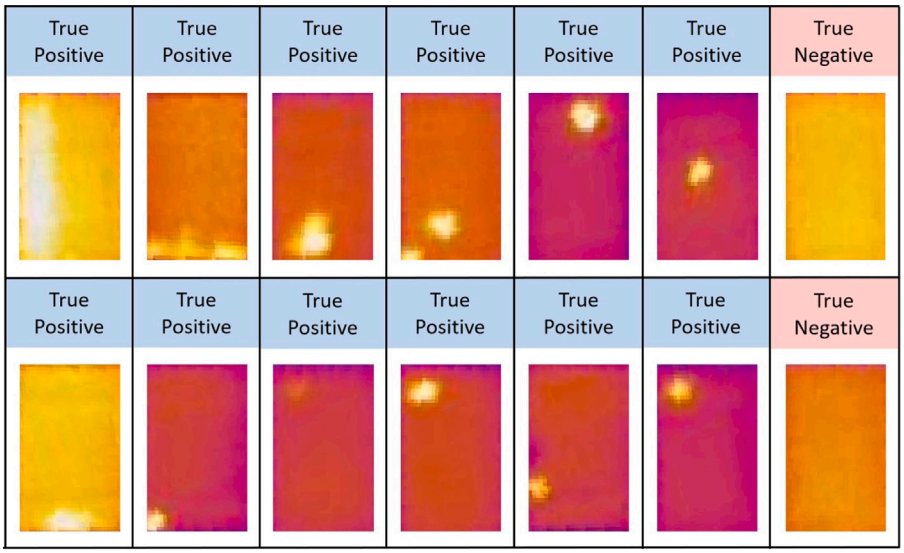


Fig. 18. Correct predictions of PV modules.

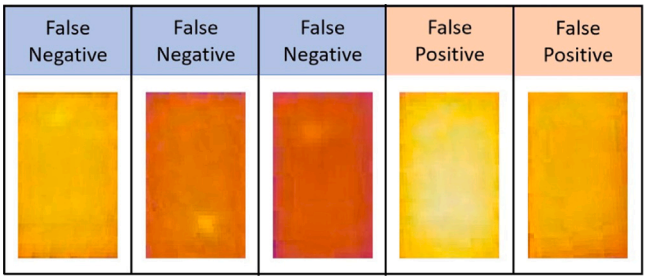


Fig. 19. Incorrect predictions of PV modules.

detection method performs well, cooperating with PV array extraction from the complex scene.

The confusion matrices are shown in Fig. 16. For normal PV modules, 99.89% of them are classified correctly and only 0.11% are mis-classified as hotspot. For PV modules with hotspots, 95.71% of them can be detected correctly while 4.29% are mis-classified as normal.

Hotspot detection in IRT images collected by the UAV are shown in Fig. 17. With accurate extraction of PV arrays from complex scene by the modified U-Net, PV modules are separated. Subsequently, they are classified as normal and hotspot, shown with black and cyan bounding boxes respectively.

Some correct predictions of PV modules are presented in Fig. 18. Different patterns of hotspots can be detected correctly, as the left 6 columns in Fig. 18. Most normal modules are predicted as normal correctly, as shown in Fig. 16 and the rightmost column in Fig. 18.

Some incorrect predictions of PV modules are shown in Fig. 19. The temperature of hotspot might be multiples of 10 K higher than the normal region in the PV module. However, the temperature in a normal PV module may also differ by several degrees, even under appropriate conditions. As a result, a few mild hotspots are failed to be detected, as shown in the left 3 columns in Fig. 19. And a few normal PV modules with higher temperature difference are predicted incorrectly as hotspot, shown in the right 2 columns in Fig. 19.

6. Conclusion

This work focuses on the problem of PV array extraction from complex scene, which has not been solved in previous literature. The proposed deep learning model, namely modified U-Net, can learn features and parameters automatically by the large available IRT images. Several innovations are provided to improve the performance of U-Net: (1) batch normalization layers is adopted to alleviate internal covariate shift problems; (2) "He initialization" is used to increase the robustness and speed up the convergence; and (3) "RMSprop" is adopted to update parameters adaptively. The extensive experiments are conducted to compare the modified U-Net with five deep learning models and three image processing methods. The results demonstrate that the proposed method performs better than existing methods. This work makes a great contribution to the researches and applications of automatic hotspot detection in aerial IRT images. Cooperating with the proposed approach for PV array extraction from complex scene, even a simple hotspot detection method can correctly classify most of the PV modules into normal or hotspot. However, a few mild hotspots are failed to be detected. In future research work, the hotspot detection algorithm shall be replaced with more advanced methods in computer vision and pattern recognition. In addition, the dataset would be expanded to validate the generalization ability of the proposed method. More IRT images might be collected from different PV stations, at various irradiance levels. The proposed modified U-Net could also be used in visible images collected by UAVs to extract PV arrays, which is useful for intelligent monitoring and inspection of PV plants.

Declaration of competing interest

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

Acknowledgments

This work was supported by the National Key Research and Development Program of China (Grant No. 2018YFB1500800), the National Natural Science Foundation of China (Grant No. 61773118, Grant No. 61703100, Grant No. 61973083), Science and Technology Project of State Grid Corporation of China (Intelligent operation and maintenance technology of distributed photovoltaic system SGTJDK00DYJS2000148).

References

Akram, M.W., Li, G., Jin, Y., Chen, X., Zhu, C., Ahmad, A., 2020. Automatic detection of photovoltaic module defects in infrared images with isolated and develop-model transfer deep learning. Sol. Energy 198 (November 2019), 175–186. http://dx.doi.org/10.1016/j.solener.2020.01.055.

Ali, M.U., Saleem, S., Masood, H., Kallu, K.D., Masud, M., Alvi, M.J., Zafar, A., 2022. Early hotspot detection in photovoltaic modules using color image descriptors: An infrared thermography study. Int. J. Energy Res. 46 (2), 774–785. http://dx.doi.org/10.1002/er.7201.

Arenella, A., Greco, A., Saggese, A., Vento, M., 2017. Real time fault detection in photovoltaic cells by cameras on drones. In: International Conference Image Analysis and Recognition. pp. 617–625. http://dx.doi.org/10.1007/978-3-319-59876-5_68.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. SegNet: A Deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (12), 2481–2495. http://dx.doi.org/10.1109/TPAMI.2016.2644615, arXiv:1511.00561.

Bommes, L., Pickel, T., Buerhop-Lutz, C., Hauch, J., Brabec, C., Peters, I.M., 2021. Computer vision tool for detection, mapping, and fault classification of photovoltaics modules in aerial IR videos. Prog. Photovolt., Res. Appl. 29 (12), 1236–1251. http://dx.doi.org/10.1002/pip.3448, arXiv:2106.07314.

Brostow, G.J., Fauqueur, J., Cipolla, R., 2009. Semantic object classes in video: A high-definition ground truth database. Pattern Recognit. Lett. 30 (2), 88–97. http://dx.doi.org/10.1016/j.patrec.2008.04.005.

Carletti, V., Greco, A., Saggese, A., Vento, M., 2020. An intelligent flying system for automatic detection of faults in photovoltaic plants. J. Ambient Intell. Humaniz. Comput. 11 (5), 2027–2040. http://dx.doi.org/10.1007/s12652-019-01212-6.

Dotenco, S., Dalsass, M., Winkler, L., Wurzner, T., Brabec, C., Maier, A., Gallwitz, F., 2016. Automatic detection and analysis of photovoltaic modules in aerial infrared imagery. In: IEEE Winter Conference on Applications of Computer Vision, WACV 2016. http://dx.doi.org/10.1109/WACV.2016.7477658.

Dunderdale, C., Brettenny, W., Clohessy, C., van Dyk, E.E., 2020a. Photovoltaic defect classification through thermal infrared imaging using a machine learning approach. Prog. Photovolt., Res. Appl. 28 (3), 177–188. http://dx.doi.org/10.1002/pip.3191.

Dunderdale, C., Brettenny, W., Clohessy, C., van Dyk, E.E., 2020b. Photovoltaic defect classification through thermal infrared imaging using a machine learning approach. Prog. Photovolt., Res. Appl. 28 (3), 177–188. http://dx.doi.org/10.1002/pip.3191.

Gallardo-Saavedra, S., Hernández-Callejo, L., Duque-Perez, O., 2018. Technological review of the instrumentation used in aerial thermographic inspection of photovoltaic plants. Renew. Sustain. Energy Rev. 93, 566–579. http://dx.doi.org/10.1016/j.rser.2018.05.027.

Hartley, R., Zisserman, A., 2004. Multiple view geometry in computer vision. http://dx.doi.org/10.1017/cbo9780511811685.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2380–7504. http://dx.doi.org/10.1109/ICCV.2015.123, arXiv:1502.01852.

Hinton, G., 2012. Neural Networks for Machine Learning. Technical Report.

Hudson, R.D., 2006. Infrared System Engineering. Wiley.

Huerta Herraiz, A., Pliego Marugán, A., García Márquez, F.P., 2020a. Photovoltaic plant condition monitoring using thermal images analysis by convolutional neural network-based structure. Renew. Energy 153, 334–348. http://dx.doi.org/10.1016/j.renene.2020.01.148.

Huerta Herraiz, A., Pliego Marugán, A., García Márquez, F.P., 2020b. Photovoltaic plant condition monitoring using thermal images analysis by convolutional neural network-based structure. Renew. Energy 153, 334–348. http://dx.doi.org/10.1016/j.renene.2020.01.148.

IEA, 2018. World Energy Outlook 2018: Highlights. Technical Report.

IEC, 2016. Terrestrial Photovoltaic (PV) Modules – Design Qualification and Type Approval – Part 2: Test Procedures. Technical Report.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: 32nd International Conference on Machine Learning, ICML 2015. arXiv:1502.03167.

Kim, D., Youn, J., Kim, C., 2016. Automatic photovoltaic panel area extraction from UAV thermal infrared images. J. Korean Soc. Surv. Geodesy Photogramm. Cartogr. 34 (6), 559–568. http://dx.doi.org/10.7848/ksgpc.2016.34.6.559.

Köntges, M., Kurtz, S., Packard, C., Jahn, U., Berger, K., Kato, K., Friesen, T., Liu, H., Van Iseghem, M., 2014. Review of Failures of Photovoltaic Modules. Technical Report.

Li, F.-F., 2016. CS231n: Convolutional Neural Networks for Visual Recognition. Technical Report, Stanford University.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. http://dx.doi.org/10.1109/CVPR.2015.7298965.

Montanez, L.E., Valentin-Coronado, L.M., Moctezuma, D., Flores, G., 2020. Photovoltaic module segmentation and thermal analysis tool from thermal images. In: 2020 IEEE International Autumn Meeting on Power, Electronics and Computing, ROPEC 2020. http://dx.doi.org/10.1109/ROPEC50909.2020.9258760, arXiv:2010.07356.

Niazi, K.A.K., Akhtar, W., Khan, H.A., Yang, Y., Athar, S., 2019. Hotspot diagnosis for solar photovoltaic modules using a naive Bayes classifier. Sol. Energy 34–43. http://dx.doi.org/10.1016/j.solener.2019.07.063.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. http://dx.doi.org/10.1007/978-3-319-24574-4_28, arXiv:1505.04597.

Schneider, L., Jasch, M., Fröhlich, B., Weber, T., Franke, U., Pollefeys, M., Rätsch, M., 2017. Multimodal neural networks: RGB-d for semantic segmentation and object detection. In: Scandinavian Conference on Image Analysis. pp. 98–109. http://dx.doi.org/10.1007/978-3-319-59126-1_9.

Shelhamer, E., Long, J., Darrell, T., 2017. Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (4), 640–651. http://dx.doi.org/10.1109/TPAMI.2016.2572683, arXiv:1411.4038.

Sovetkin, E., Achterberg, E.J., Weber, T., Pieters, B.E., 2021. Encoder-decoder semantic segmentation models for electroluminescence images of thin-film photovoltaic modules. IEEE J. Photovolt. 11 (2), 444–452. http://dx.doi.org/10.1109/JPHOTOV.2020.3041240, arXiv:2010.07556.

Su, Y., Tao, F., Jin, J., Zhang, C., 2021. Automated Overheated Region Object detection of photovoltaic module with thermography image. IEEE J. Photovolt. 11 (2), 535–544. http://dx.doi.org/10.1109/JPHOTOV.2020.3045680.

Subedi, I., Silverman, T.J., Deceglie, M.G., Podraza, N.J., 2019. Emissivity of solar cell cover glass calculated from infrared reflectance measurements. Sol. Energy Mater. Sol. Cells 190, 98–102. http://dx.doi.org/10.1016/j.solmat.2018.09.027.

Tsanakas, J.A., Ha, L., Buerhop, C., 2016. Faults and infrared thermographic diagnosis in operating c-si photovoltaic modules: A review of research and future challenges. Renew. Sustain. Energy Rev. 62, 695–709. http://dx.doi.org/10.1016/j.rser.2016.04.079.

Vidal de Oliveira, A.K., Mohammedreza, A., Rüther, R., 2019. Automatic fault detection of photovoltaic array by convolutional neural. In: 36th European Photovoltaic Solar Energy Conference and Exhibition. pp. 1302–1307.

Wang, N., Sun, Z.L., Zeng, Z., Lam, K.M., 2021. Effective segmentation approach for solar photovoltaic panels in uneven illuminated color infrared images. IEEE J. Photovolt. 11 (2), 478–484. http://dx.doi.org/10.1109/JPHOTOV.2020.3041189.

Wen, C.D., Chai, T.Y., 2011. Experimental investigation of emissivity of aluminum alloys and application of multispectral radiation thermometry. Appl. Therm. Eng. 31 (14–15), 2414–2421. http://dx.doi.org/10.1016/j.applthermaleng.2011.04.005.

Xu, Z., Shen, Y., Zhang, K., Wei, H., 2021. A segmentation method for PV modules in infrared thermography images. In: 13th IEEE PES Asia-Pacific Power and Energy Engineering Conference.

Zhang, B.S.P., 2020. The 2nd diabetic retinopathy – grading and image quality estimation challenge. URL: https://isbi.deepdr.org/.

Zhang, H., Hong, X., Zhou, S., Wang, Q., 2019. Infrared image segmentation for photovoltaic panels based on res-unet. In: Pattern Recognition and Computer Vision. In: LNCS, vol. 11857, Springer International Publishing, pp. 611–622. http://dx.doi.org/10.1007/978-3-030-31654-9_52.