Journal of Real-Time Image Processing (2024) 21:52
https://doi.org/10.1007/s11554-024-01415-x

RESEARCH

Deeplab-YOLO: a method for detecting hot-spot defects in infrared image PV panels by combining segmentation and detection

Ye Lei[1] · Xiaoye Wang[1,2,3] · Aimin An[1,2,3] · Haijiao Guan[1]

Received: 24 August 2023 / Accepted: 5 January 2024 / Published online: 13 March 2024

Abstract
Aiming at the problem of difficult operation and maintenance of PV power plants in complex backgrounds and combined with image processing technology, a method for detecting hot spot defects in infrared image PV panels that combines segmentation and detection, Deeplab-YOLO, is proposed. In the PV panel segmentation stage, MobileNetV2 was introduced into the Deeplabv3+ model. Empty convolution in the atrous spatial pyramid pooling (ASPP) structure was improved, and established a relationship between layer-level features, the CBAM attention mechanism was combined, which achieved fast and accurate segment of PV panels and avoided false detection of hot spots. In the hot-spot recognition stage, a lightweight MobileNetV3 network was designed to replace the YOLO v5 backbone network, a small defect prediction head was added, and EIOU was used as a loss function, which improved the speed and accuracy of hot-spot detection and enhanced the performance of the YOLO v5 model. The experimental results show that the optimized Deeplabv3+ model and YOLO v5 model improve the accuracy of segmenting PV panels in images and identifying hot-spot defects by 2.61% and 0.7%, respectively, compared with the original model. This proposed method can accurately segment the PV panels and then identify different sizes of hot-spot defects on the PV panels.

Keywords Deep learning · Deeplabv3+ · YOLO v5 · Infrared image of PV panel · Defect detection

1 Introduction

In recent years, with the increasing environmental pollution and energy crisis, people have turned their attention to the development and utilization of new energy sources, and PV power generation occupies an important position in new energy sources [1], it is clean energy that uses solar energy to convert into electric energy. The PV panels of photovoltaic power plants are susceptible to shading by dust, bird droppings, leaves, etc. Long-term coverage on the surface of the PV panels will cause the internal circuit characteristics of the shaded part to change and become a load-consuming energy, resulting in hot-spot faults. In addition, PV panel modules will also cause their circuit to short-circuit due to internal defects to form a hot-spot effect [2]. These defects will not only affect the power generation efficiency of PV but also may have safety hazards. Therefore, the detection of defects in PV panels has become a crucial step in inspection work [3]. However, most centralized PV power plants are built in high mountains, deserts, and the sea, and the environment where PV panels are located may be complex or remote, which is not suitable for routine inspection.

At present, traditional inspection methods cannot meet the inspection requirements in terms of accuracy and speed. However, it is more common and accurate to identify and detect hot spots by capturing images and processing them. To automatically and accurately identify defects in images, researchers have proposed traditional machine

✉ Xiaoye Wang
ywangx06@semi.ac.cn

✉ Aimin An
anaimin@lut.edu.cn

Ye Lei
leiye0110@163.com

Haijiao Guan
1553698182@qq.com

1 College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China

2 Key Laboratory of Gansu Advanced Control for Industrial Processes, Lanzhou 730050, China

3 National Demonstration Center for Experimental Electrical and Control Engineering Education, Lanzhou University of Technology, Lanzhou 730050, China

vision algorithms based on manual feature extraction and classifiers [4–6]. Wang et al. [7] simulated foreign object occlusion experiments on PV panels under visible light and used improved YOLO v5 to detect PV panel occlusion, and finally simulated under Simulink to visualize and analyze the verification. Wang et al. [8] proposed an improved U-Net to segment defects such as cracks, black spots, and overwhelmed in solar cell images to address the shortage of defect segmentation. Zhang et al. [9] proposed an adaptive learning method based on YOLO v5 by introducing deformable convolutional blocks, increasing the number of prediction heads, and K-means clustering anchor box algorithm to detect defects in photovoltaic cells. However, the above methods are all in the PV panels are not encapsulated in the non-working state to collect images, it can not be applied to the inspection of photovoltaic power plants, and the detection of external factors such as fallen leaves, bird droppings accumulation of dust, etc. is incomplete. Rong et al. [10] proposed several algorithms to improve the YOLO series based on convolutional neural networks for the characteristics of PV string hot spots under IR imaging and made a test comparison. Su et al. [11–13] proposed a bag-of-words model based on unsupervised feature extraction to obtain global features for defect classification and improved YOLO v3 to capture defect information of different scale feature layers for small target defect detection under complex background interference.

The YOLO family is a typical one-stage target detection algorithm, always committed to improving the best speed and accuracy for real-time applications, and has been used in many computer vision applications. It uses anchor box to combine classification and target localization. Zhu et al. [14] replaced the original predictor head with a Transformer predictor's head based on YOLO v5, increased the number of predictor's heads, and also integrated the attentional convolution block (CBAM), which is more suitable for capturing small targets in UAV scenarios. Kim et al. [15] proposed an online copy-and-paste and hybrid data enhancement method to alleviate the class imbalance in the dataset during training and effectively improve the classification performance of the YOLO v5 detection model. Xiao et al. [16] combined Transformer with the backbone network of YOLO v5 and utilized Bi-FPN for multi-scale feature fusion of the model, achieving more effective extraction of feature information at different scales. Deep learning for semantic segmentation originated from fully convolutional networks (FCNs) [17], which improves the original CNN structure using up-convolution for upsampling and can be used without fully connected layers for dense prediction to achieve pixel-level classification. SegNet [18] shifts the maximum pooling index to the decoder, thus improving the segmentation resolution. Deeplab [19] mainly uses null convolution and proposes spatial pooling of pyramidal models in spatial dimensions using fully connected conditional random fields (CRF), starting with the Deeplabv2 [20, 21] algorithm, which uses an encoder and decoder architecture. Baheti et al. [22] used expanded Xception as a backbone network in the Deeplabv3+ structure and performed well for unstructured road recognition. Xi et al. [23] introduced an attention module on the original Deeplabv3+, which solved the problem of small and irregular target feature representation and improved segmentation performance. Guan et al. [24] proposed a combined semantic segmentation and target detection method for PV panel hot-spot detection. The PV panels are identified in the infrared images using improved YOLO v4, and the PV panels are extracted to segment the hot spots with improved Deeplabv3+. Wu et al. [25] proposed a lightweight pruning method for detecting underwater garbage in complex backgrounds based on YOLO v5. In the feature extraction part, MobileNetV3 network and CBAM attention mechanism were combined, and the improved network was pruned, achieving a relatively balanced effect in detection accuracy and speed. Hou et al. [26] proposed an object detection method M-YOLO for infrared images, and designed parallel feature fusion methods from top to bottom and bottom-up. It used MobileNetV2 network and two detection heads to achieve a lightweight model, which can detect infrared ground targets in complex environments. Other common semantic segmentation algorithms include U-Net and PSPNet.

After the above analysis, the first-order target detection and semantic segmentation algorithm do have powerful real-time processing capability and low hardware requirements. This article proposes a Deeplab-YOLO hot-spot defect detection method that combines segmentation and detection with infrared images and based on the differences and features in the shape, size, and color of PV panels and hot spots. On the one hand, it can meet the accuracy of segmentation and enhance the edge features of the target. On the other hand, it can effectively identify different differential hot spots. The results showed that the improved model has significant performance improvements and is more suitable for detecting hot spots on PV panels, which provides technical support for the operation and maintenance of PV power plants.

The main contributions of this paper can be summarized as follows:

(1) We propose a rapid detection method for identifying hot-spot defects by combining image segmentation with object detection;
(2) We improve the Deeplabv3+ segmentation model to accurately segment the PV panels from the infrared image, solving the interference of heat-absorbing objects in the environment of the PV panels presenting "false hot spots" in the image when detecting hot-spot defects.

(3) And then improve the YOLO v5 detection mode to achieve a higher detection rate of hot spots, which solves the problem of large differences in the size of hot spots in images caused by different reasons and small hot spots missed detection.

The rest of this article is organized as follows. Section 2 describes the overall research process, model structure, and principles in detail. Section 3 presents the dataset sources, experimental setup and implementation details. Conclusions and future work are in Sect. 4.

2 PV panel segmentation and hot-spot detection

2.1 Overall research program

The method of this article focuses on two aspects: segmentation of PV panels and detection of hot spots. Different annotation software is used to create a dataset with PV panels and hot spots as the target, respectively, segment the panels using an improved Deeplabv3+ model to exclude bright spots caused by endothermic objects in the background, and then use a one-stage object detection algorithm YOLO v5 to identify hot spots on the PV panels. The overall flowchart is shown in Fig. 1.

2.2 A feature-enhanced segmentation model for PV panels

This experiment adopts a Deeplabv3+ network model optimized based on MobileNetV2 network. Using MobileNetV2 instead of Xception network as the backbone feature extraction network can greatly reduce the number of model parameters, achieve lightweight design of the model, and ensure that deep convolution completes feature extraction in high dimensions, improving model computational performance.

Second, the depth separable convolution is introduced into the Deeplabv3+-based ASPP structure to replace the empty convolution with the empty depth separable convolution. At the same time, the feature map output from the upper level of the ASPP expansive rate convolution layer is stitched with the original feature map channel and input to the next level of the expansive rate convolution layer to extract features to achieve different-sensory field fusion and intensively use multi-scale information.

Finally, the CBAM [27] (convolutional block attention module) attention mechanism is introduced in the Deeplabv3+ encoder module. And it is a lightweight attention module that was proposed in 2018, as its structure is illustrated in Fig. 2.

CBAM includes two sub-modules: the channel attention module (CAM) and the spatial attention module (SAM), which can filter the extracted features, making the information retained in the encoding stage more conducive to the accuracy of segmentation. After adding CBAM to the encoder of the Deeplabv3+ model, the network can learn the features of PV panels well, suppress redundant information and help extract the edge features of PV panels. The channel operation process is as follows:

$$\begin{aligned} M_C(F) &= \sigma(\mathrm{MLP}(\mathrm{AvgPool}(F)) + \mathrm{MLP}(\mathrm{MaxPool}(F))) \\ &= \sigma(W_1(W_0(F^c_{\mathrm{avg}})) + W_1(W_0(F^c_{\mathrm{max}}))) \end{aligned} \tag{1}$$

Input a feature map of $H \times W \times C$ size, pass through parallel max pooling layers and average pooling layers respectively,

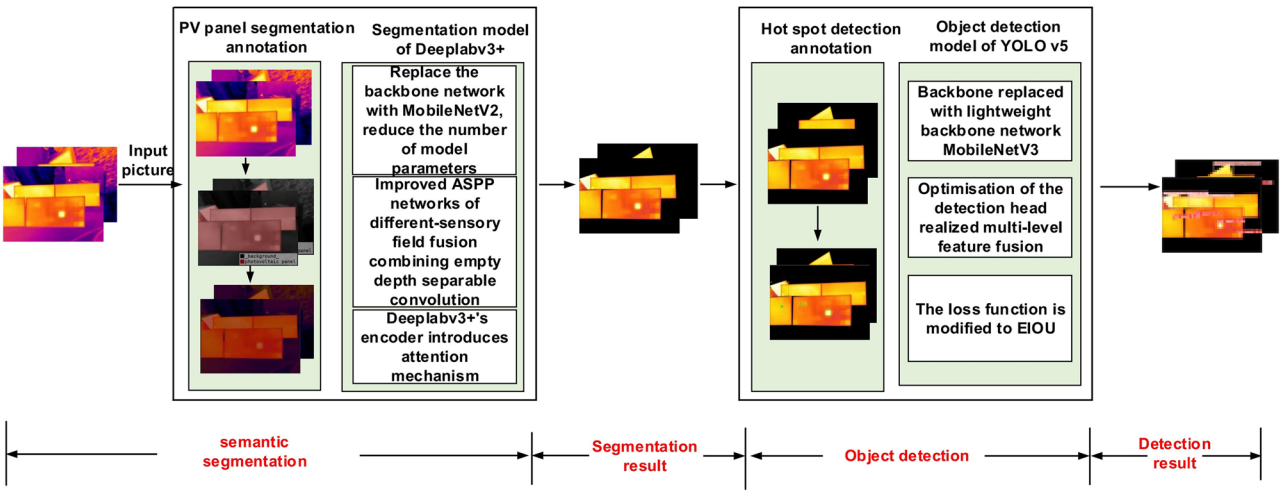


Fig. 1 PV segmentation and hot-spot detection process

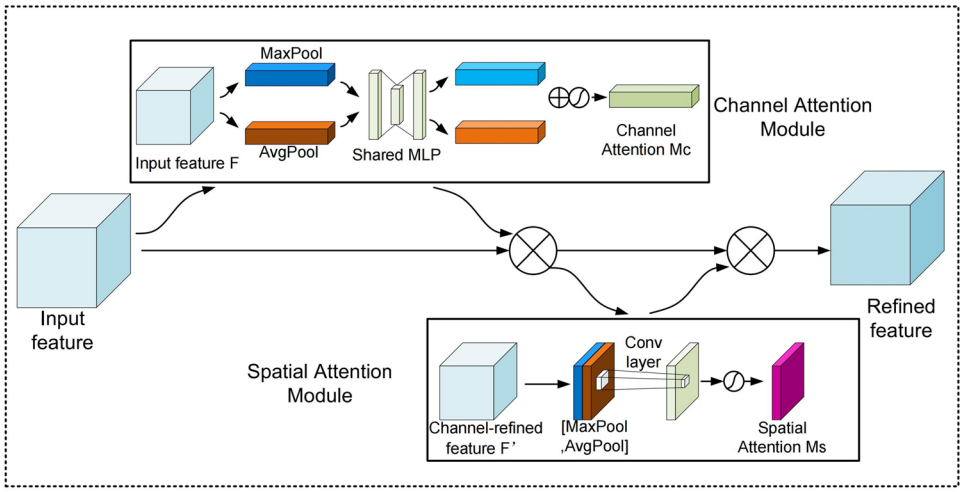


Fig. 2 Structure of CBAM

and transform the feature map into two $1 \times 1 \times C$ channel features. Then, send the two channel features into a two-layer neural network composed of multi-layer perceptrons and hidden layers. The number of neurons in the first layer is C/r, and the number of neurons in the second layer is C. The activation function is ReLU. After adding the output elements of the two features by product, the required channel attention map M_C can be obtained through the sigmoid activation function. And the spatial operation process is as follows:

$$\begin{aligned} M_S(F) &= \sigma(f^{7\times7}([\mathrm{AugPool}(F);\mathrm{MaxPool}(F)])) \\ &= \sigma(f^{7\times7}([F^S_{\mathrm{avg}};F^S_{\mathrm{max}}])) \end{aligned} \tag{2}$$

The channels of input F' are compressed dimensionally using average pooling and max pooling to aggregate channel information, the two obtained channel features of size $H \times W \times 1$ are fused by a convolutional layer of size 7×7, and finally, the Sigmoid activation function is used to multiply the weight coefficients with the input feature F' to obtain the spatial attention map M_S.

2.3 A multi-scale fusion model for hot-spot detection

The YOLO v5 algorithm was introduced in 2020 by the Ultralytics LLC team, and it comes in four different models: YOLO v5s, YOLO v5m, YOLO v5l, and YOLO v5x. Among them, YOLO v5s has the fastest inference speed and precise accuracy. Its architecture mainly consists of three parts: backbone, neck, and head. After inputting an image, backbone aggregates the image features at different image granularities to form image features. Then, neck stitches the extracted image features and transmits them to the prediction layer, and the head predicts the image features to generate bounding boxes and prediction categories.

In this experiment, the lightweight backbone network MobileNetV3 was used to replace the original backbone network CSPDarknet53 of YOLO v5 [28]. MobileNetV3 is composed of deep separable convolution and inverse residual structure, and is dimensionally increased by 6 times through point-by-point convolution, allowing deep convolution to extract hot-spot features in higher dimensional spaces, achieving the effect of feature reuse. The overall architecture of Deeplab-YOLO model is shown in Fig. 3.

Second, in order to solve the problem of uneven size of hot spots in images and incomplete recognition of hot spots. The hot spots caused by obstructions such as leaves and bird droppings are very small, while the heat spots caused by internal damage to PV panels are relatively large. The original YOLO v5 does not achieve accurate recognition of small targets. In response to the above issues, an improved YOLO v5 introduced a decoupling head for detecting small defects. It can be more sensitive to feature extraction of small hot spots on PV panels. Multi-level feature fusion is realized.

The Loss function of the model reflects the difference between the predicted value and the true value, especially the IOU that determines the regression loss in the Loss function. so in the improved YOLO v5, we use EIOU [29] as the regression loss function, which minimizes the difference between the width and height of the prediction box and the real box, making the rate of convergence of the model faster. Optimized the sample imbalance problem in bounding box regression tasks.

The EIOU consists of the overlap loss of the prediction box and the real box (L_{EIoU}), the real box of the prediction box and the real box (L_{dis}), and the width and height loss of the prediction box and the real box (L_{asp}):

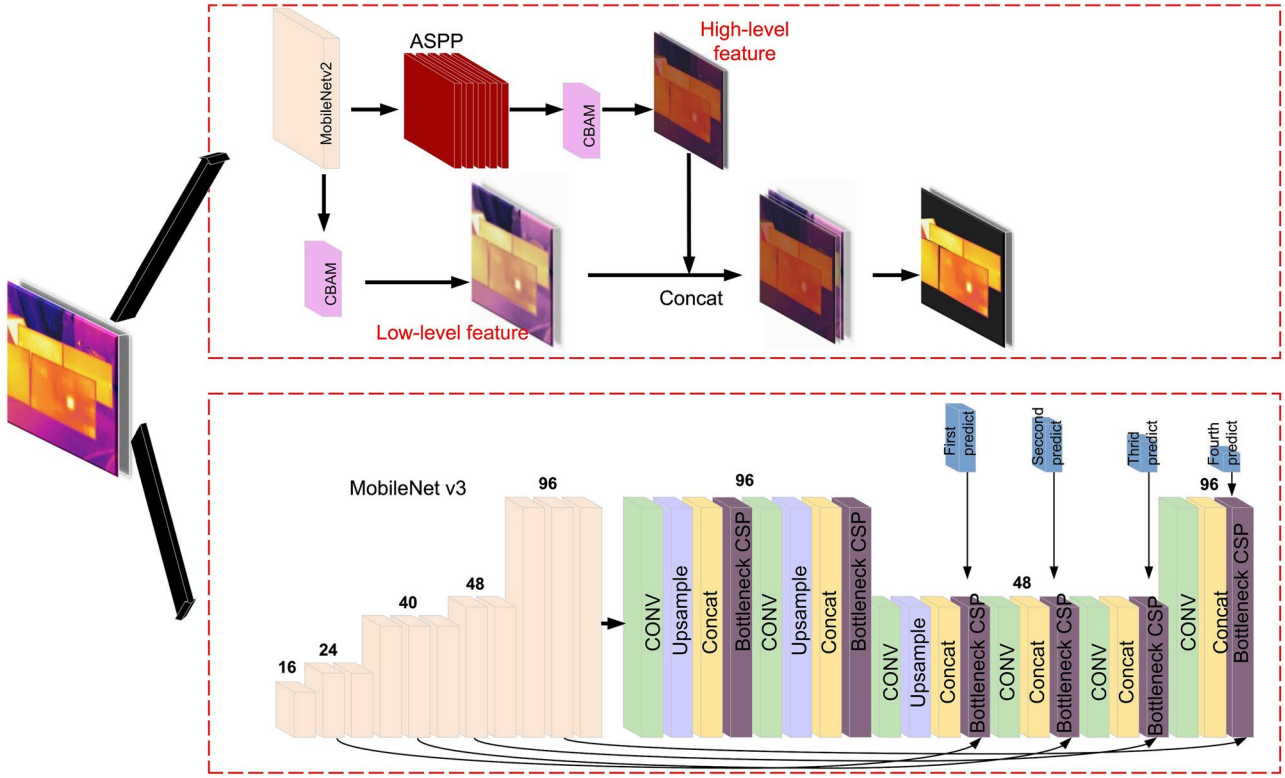


Fig. 3 Structure of Deep-YOLO

$$L_{\mathrm{EIoU}} = L_{\mathrm{IoU}} + L_{\mathrm{dis}} + L_{\mathrm{asp}} = 1 - \mathrm{IoU} + \frac{\rho^2(b, b^{\mathrm{gt}})}{c^2} + \frac{\rho^2(w, w^{\mathrm{gt}})}{C_w^2} + \frac{\rho^2(h, h^{\mathrm{gt}})}{C_h^2} \quad (3)$$

where b and b_{gt} are the coordinates of the center points of the prediction box and the real box, respectively, $\rho^2(b, b_{\mathrm{gt}})$ denote the Euclidean distance between them, and C_w^2, C_h^2 are the width and height of the minimum outer rectangle of the prediction box and the real box, respectively. Similarly, $\rho^2(w, w_{\mathrm{gt}})$ denotes the difference between the width of the prediction box and the real box, and $\rho^2(h, h_{\mathrm{gt}})$ denotes the difference between the height of the prediction box and the real box.

3 Experimental verification and result analysis

3.1 Dataset and experimental settings

The experimental data in this article were captured by a UAV equipped with an infrared camera. A total of 672 defective infrared images of PV panels were collected, and a total of 482 images containing PV panels and hot spots were selected for network training and testing. To address the risk of insufficient network training sample datasets and overfitting of the network. The experiment performed various image transformations on the obtained images. Including mirror flipping, left–right flipping, multiply the pixel value of the image by any value between 0.8 and 1.2 to increase the brightness or change its color of the images, sharpen the image, add Gaussian noise to the pixel points at random ± 5, perform affine transformation on some images, image scaling, rotation, shear transformation, fill the images. Expanding them to 3374 images with a resolution of 640 × 480.

In the segmentation stage of the PV panels, the training and validation sets were divided according to 9:1, and the initial learning rate was set to 0.0005. In the freezing and unfreezing stages, 50 trainings are conducted, respectively, 100 in total. Take the average of every 5 training sessions as the result of one session. In the hot-spot recognition stage, the training set, validation set and test set are divided according to 6:2:2 and trained 200 times. The expanded image is shown in Fig. 4.

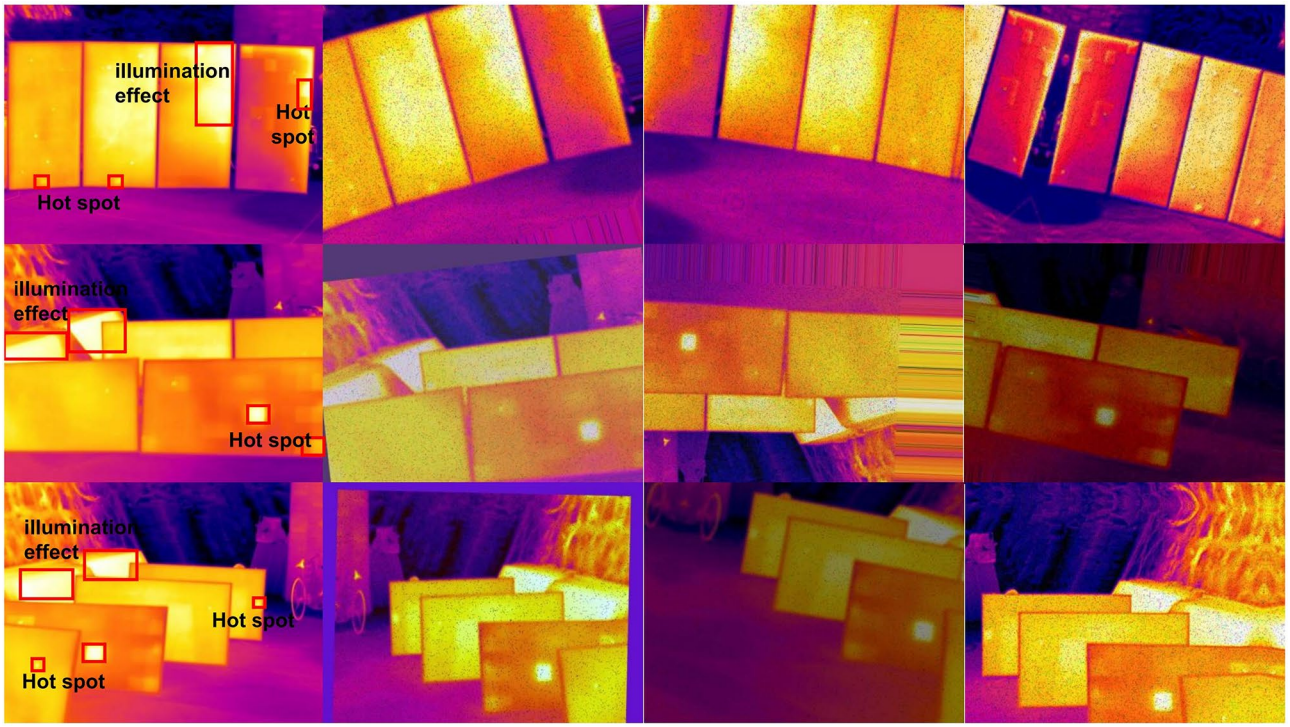


Fig. 4 Image enhancement

3.2 Results and analysis of PV panel segmentation

The mean pixel accuracy (MPA) and mean intersection over union (MIOU) are often used as evaluation indicators in segmentation models:

$$\mathrm{MPA} = \frac{1}{n+1}\sum_{i=0}^{n}\frac{R_{ii}}{\sum_{j=0}^{n}R_{ij}} \tag{4}$$

$$\mathrm{MIOU} = \frac{1}{\mathrm{n}+1}\sum_{i=0}^{n}\frac{R_{ii}}{\sum_{j=0}^{n}R_{ij}+\sum_{j=0}^{n}R_{ji}-R_{\mathrm{ii}}}. \tag{5}$$

Suppose there are $n+1$ classes (n target classes and one background class), R_{ii} denotes the number of correctly classified pixels, R_{ij} denotes the number of pixels belonging to class i but assigned to class j, and R_{ji} denotes the number of pixels belonging to class j but assigned to class i.

To verify the superiority of the improved Deeplabv3+ model, the effects of different semantic segmentation models trained with the same training parameters and dataset were compared, respectively. From Table 1, it can be seen that the MPA and IOU of the four models are relatively high, but the amount of U-Net network models is large, and the MPA index of PSPNet is relatively small. Therefore, Deeplabv3+ was chosen as the segmentation model. After improvement, the MPA and MIOU of Deeplabv3+ reached 95.01% and 97.62%, respectively, which were 2.61% and 12.94% higher than the original Deeplabv3+.

The comparison of training losses is shown in Fig. 5. Overall, the training loss values of the four models tend to stabilize with the increase in iteration times. The U-Net and improved Deeplabv3+ models have better fitting results in the later stage. Although U-Net has slightly better stationarity, the improved Deeplabv3+ model has stronger convergence.

Table 1 Performance comparison results between different methods

Models	MPA (%)	MIOU (%)	GFLOPS (G)	Parameters (M)
Deeplabv3+	92.40	84.68	53.026	5.818
Improved Deeplabv3+	**95.01**	**97.62**	**93.4**	**2.28**
U-Net	93.06	96.25	452.31	24.892
PSPNet	91.49	78.62	6.034	2.377

Bold values indicate the data that ranks within the top one in terms of performance

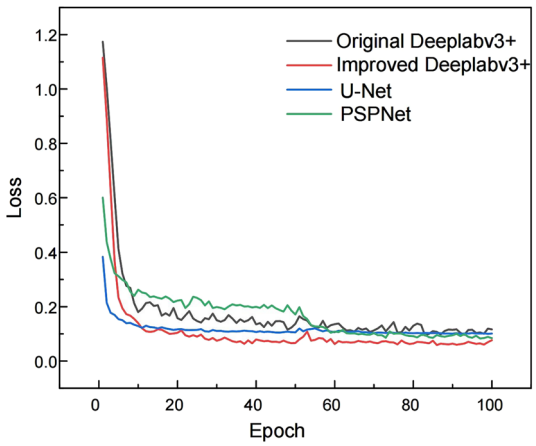


Fig. 5 The comparison curve of train loss

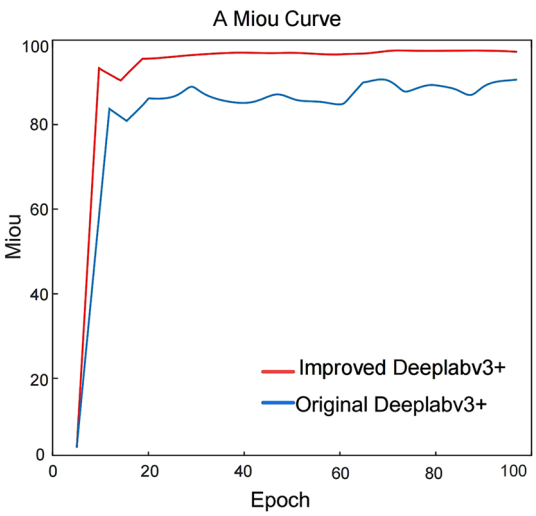


Fig. 6 Average cross-parallel ratio comparison curve

Figure 6 shows the average intersection and merger ratio curves of the model. It can be seen that the improved Deeplabv3+ curve is higher and smoother than the original Deeplabv3+ curve. This means that the improved network has a high overlap between the real and predicted box for each class segmentation, and the segmentation is more accurate, which can segment the PV panels from the complex background more accurately.

Randomly select 4 infrared images for testing. From Fig. 7, it can be seen that the improved Deeplabv3+ segmentation can better preserve the edge information of the PV panel, and the recognition is clearer and more accurate. However, other segmentation models blur the detailed features of the PV panel edges, resulting in wrong segmentation.

3.3 Results and analysis of PV panel hot-spot detection

Precision (P), recall (R), average detection precision (AP), FPS, and model size are commonly used to measure the feasibility of the model for practical applications when detecting hot spots. FPS is the number of frames per second for detecting images, and model size is determined by the number of model parameters. Each indicator is defined as shown in Eqs. (6)–(9):

$$\text{precision} = \frac{\text{tp}}{\text{tp} + \text{fp}} \tag{6}$$

$$\text{recall} = \frac{\text{tp}}{\text{tp} + \text{fn}} \tag{7}$$

$$\text{AP} = \int_0^{\text{recall}} \text{precision} d_{\text{recall}} \tag{8}$$

$$\text{FPS} = \frac{s}{t} \text{FPS} = \frac{s}{t} \tag{9}$$

where tp is the number of actual positive samples that are predicted to be positive, fp is the number of actual positive samples that are predicted to be negative, fn is the number of actual negative samples that are predicted to be positive, s is the number of images processed, and t is the amount of time required to process the images. AP is the area of the P/R curve.

In this paper, ablation experiments are used to validate the optimisation model. Specifically, the backbone network (MobileNetv3) is replaced for YOLO v5, a small target detection layer (smallhead) is introduced, the loss function (EIOU) is replaced, and the detection effect is compared under the same dataset and the same number of training iterations.

As can be seen from Table 2, the optimized model has the smallest number of model parameters and the minimum average network inference time per image. The optimized model achieved 94.2% accuracy in hot-spot recognition, which is 0.7% higher than the original YOLO v5 model. However, the addition of a small target detection layer model YOLO v5 smallhead which showed a certain decrease. This indicates that the infrared features of smaller hot-spot defects are not obvious compared to PV panels, so there are certain limitations on the detection of small targets. In terms of the recall rate of hot-spot defects, the YOLO v5 smallhead model has increased by 3.2% compared to the original YOLO v5 model, and compared to the optimized model improves by 3.5%. Overall, the optimized model can better meet the characteristics of hot-spot defects of PV panels in infrared images.

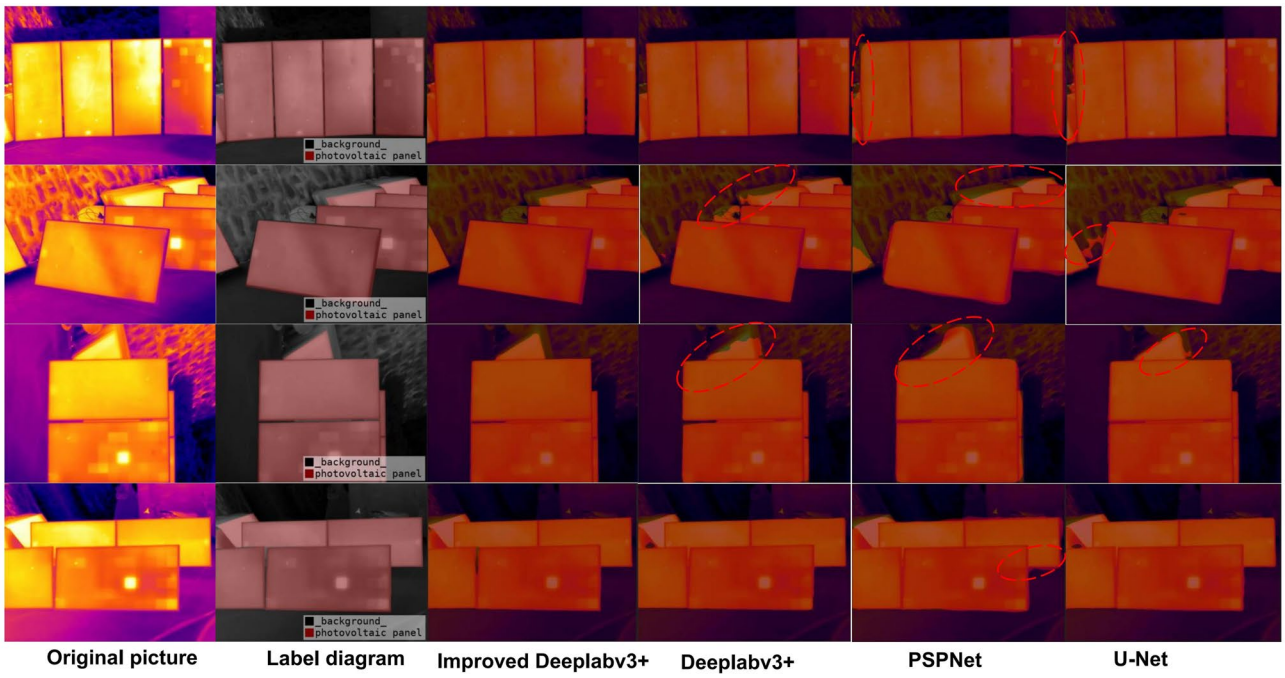


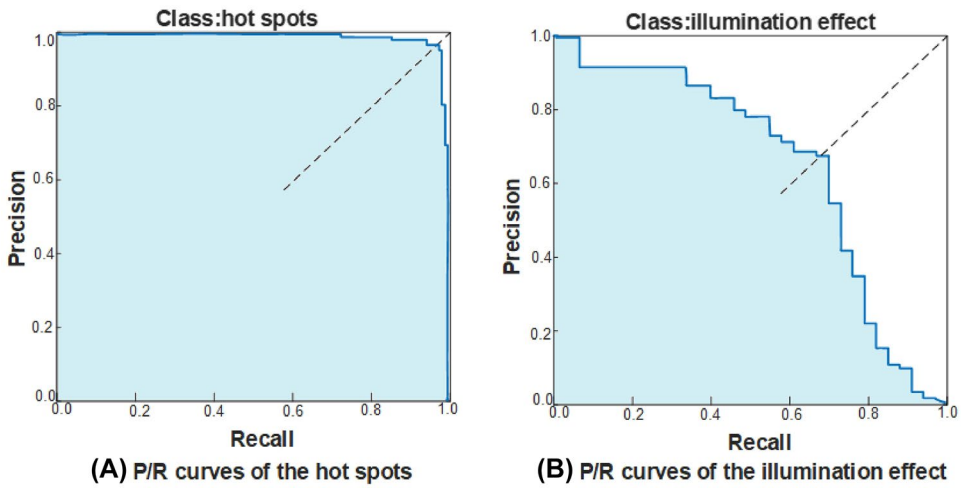
Fig. 7 The comparison of PV segmentation

Table 2 Statistical results of ablation experiments

MobileNetv3	Smallhead	EIOU	*P* (%)	*R* (%)	Parameters (M)	Inference time (ms)
			93.5	95.9	7.02	147.06
√			**96.1**	94.3	1.4	78.125
	√		78.5	99.1	22.5	101.01
		√	82	**1**	7.01	147.06
√	√	√	94.2	99.4	**0.83**	**55.87**

√ indicates the selected methods and bold values indicate the data that ranks within the top one in terms of performance

Fig. 8 P/R curve of optimizing model



The *P*/*R* curve of the optimized model for detecting hot spots is shown in Fig. 8. The horizontal and vertical axes represent recall and precision, respectively, and the mAP value of each type of hot spot is the area surrounded by the curve and the two axes. The point where the recall rate is equal to the precision is the balance point. From (a), it can be seen that since the hot spot has very prominent color characteristics in the infrared imaging image, the area enclosed by its mAP is close to 1, and the detection effect is better. However, the false hot spot in (b) is relatively poorly detected because it is affected by the light angle and shooting angle, and the hot-spot area is very similar and has blurred edges.

Comparative experiments were also conducted with a range of mainstream methods, all of which had the same experimental configuration and dataset, and the results are shown in Table 3.

From Table 3, it can be seen that our proposed method has achieved optimal results in all the parameters. Enable the model to meet the detection accuracy while ensuring detection speed. Overall, the method proposed in this article is more suitable for detecting hot-spot defects in infrared images.

Tested with segmented images, the detection results of the original YOLO v5 model can be seen in Fig. 9(C), which clearly shows missed detections and low detection confidence. While in Fig. 9(D), all PV hot spots can be detected and classified, and smaller hot spots can also be detected with higher confidence levels. In comparison, the proposed Deeplab-YOLO method has more accurate detection results, can capture more hot-spot features, and has excellent generalization performance.

Table 3 Comparative experiment

Models	*P* (%)	*R* (%)	Parameters (M)	map50 (%)
YOLO v5s	93.5	95.9	7.02	98.1
YOLO v5m	71.1	98.1	20.97	97.8
YOLO v5l	83.1	98.3	46.27	98.4
YOLO v5x	84.6	98.3	86.38	98.6
YOLO v3	81.6	98.1	61.67	97.9
YOLO v3-spp	78.8	98.8	62.71	97.9
Proposed	**94.2**	**99.4**	**0.83**	**99.2**

Bold values indicate the data that ranks within the top one in terms of performance

4 Conclusions

This paper investigates the detection of hot-spot defects on PV panels under complex background in infrared images, and proposes a Deeplab-YOLO hot-spot defect detection

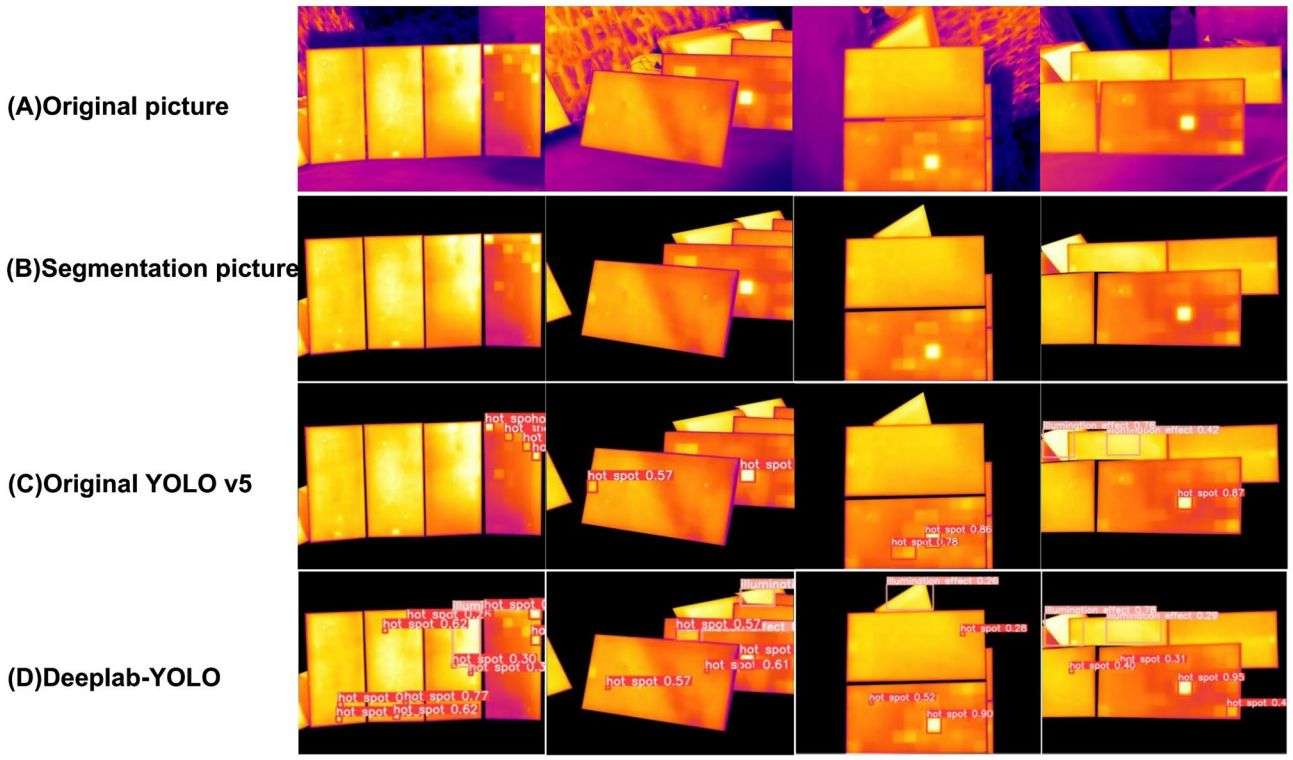


Fig. 9 PV panel hot-spot identification results graph

method. This method can effectively extract the PV panels and exclude the interference brought by the outside world when detecting hot spots, and at the same time, it can adapt to the variation of different sizes of hot spots and accurately detect hot spots with large morphological differences. Finally, compared with other methods, the detection speed is improved while the detection accuracy is also improved. The experimental results show that the proposed method has a high accuracy in detecting hot spots and performs best in all aspects. How to evaluate small sample datasets more accurately without losing speed will be considered in the next research.

Author contributions Lei Ye and Wang Xiaoye completed the main manuscript documents, Guan Haijiao provided ideas, and An Aimin provided guidance. All authors have reviewed the manuscript.

Funding This work is supported by the National Natural Science Foundation of China (61563032, 61963025), the Gansu Provincial Science and Technology Program (22YF7GA164, 22CX8GA131) and the Red Willow Outstanding Young Talent Program of Lanzhou University of Technology (Grant Nos. 062001).

Data availability Data will be made available on reasonable request.

Declarations

Conflict of interest The authors have no competing interests to declare.

References

1. Basit, B.A., Jung, J.-W.: Recent developments and future research recommendations of control strategies for wind and solar pv energy systems. Energy Rep. **8**, 14318–14346 (2022)
2. Lukas, B., Mathis, H., Claudia, B.-L., Tobias, P., Jens, H., Christoph, B., Andreas, M., Ian, M.P.: Anomaly detection in ir images of pv modules using supervised contrastive learning. Progr. Photovolt. Res. Appl. **30**(6), 597–614 (2022)
3. Bolun, D., Yang, R., He, Y., Wang, F., Huang, S.: Nondestructive inspection, testing and evaluation for si-based, thin film and multi-junction solar cells: an overview. Renew. Sustain. Energy Rev. **78**, 1117–1151 (2017)
4. Vijithananda, S.M., Jayatilake, M.L., Hewavithana, B., Gonçalves, T., Rato, L.M., Weerakoon, B.S., Kalupahana, T.D., Silva, A.D., Dissanayake, K.D.: Feature extraction from mri adc images for brain tumor classification using machine learning techniques. Biomed. Eng. Online **21**(1), 52 (2022)
5. Firuzi, K., Vakilian, M., Phung, B.T., Blackburn, T.R.: Partial discharges pattern recognition of transformer defect model by lbp & hog features. IEEE Trans. Power Deliv. **34**(2), 542–550 (2018)
6. Luo, Q., Sun, Y., Li, P., Oluyomi, S., Tian, L., He, Y.: Generalized completed local binary patterns for time-efficient steel surface defect classification. IEEE Trans. Instrum. Meas. **68**(3), 667–679 (2018)
7. Wang, J., Zhao, B., Yao, X.: Pv abnormal shading detection based on convolutional neural network. In: 2020 Chinese Control And Decision Conference (CCDC), pp. 1580–1583. IEEE (2020)
8. Guo, S., Wang, Z., Lou, Y., Li, X., Lin, H.: Detection method of photovoltaic panel defect based on improved mask r-cnn. J. Internet Technol. **23**(2), 397–406 (2022)
9. Zhang, M., Yin, L.: Solar cell surface defect detection based on improved yolo v5. IEEE Access **10**, 80804–80815 (2022)
10. Ying, Y., Qi, Y., Rong, L., Men, H., Joo, Y.H.: Anchor points based accurate fault locating in large-scale photovoltaic plants via aerial infrared videos. IEEE J. Photovolt. **12**(1), 437–443 (2021)
11. Binyi, S., Chen, H., Zhu, Y., Liu, W., Liu, K.: Classification of manufacturing defects in multicrystalline solar cells with novel feature descriptor. IEEE Trans. Instrum. Meas. **68**(12), 4675–4688 (2019)
12. Binyi, S., Chen, H., Chen, P., Bian, G., Liu, K., Liu, W.: Deep learning-based solar-cell manufacturing defect detection with complementary attention network. IEEE Trans. Ind. Inform. **17**(6), 4084–4095 (2020)
13. Binyi, S., Chen, H., Liu, K., Liu, W.: Rcag-net: residual channelwise attention gate network for hot spot defect detection of photovoltaic farms. IEEE Trans. Instrum. Meas. **70**, 1–14 (2021)
14. Zhao, Q., Liu, B., Lyu, S., Wang, C., Zhang, H.: Tph-yolov5++: boosting object detection on drone-captured scenarios with cross-layer asymmetric transformer. Remote Sens. **15**(6), 1687 (2023)
15. Kim, J.-H., Kim, N., Park, Y.W., Won, C.S.: Object detection and classification based on yolo-v5 with improved maritime dataset. J. Mar. Sci. Eng. **10**(3), 377 (2022)
16. Xiao, D., Xie, F.T., Gao, Y., Li, Z.N., Xie, H.F.: A detection method of spangle defects on zinc-coated steel surfaces based on improved YOLO-v5. Int J Adv Manuf Technol. **128**, 937–951 (2023)
17. Sun, W., Wang, R.: Fully convolutional networks for semantic segmentation of very high resolution remotely sensed images combined with dsm. IEEE Geosci. Remote Sens. Lett. **15**(3), 474–478 (2018)
18. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)
19. Niu, Z., Liu, W., Zhao, J., Jiang, G.: Deeplab-based spatial feature extraction for hyperspectral image classification. IEEE Geosci. Remote Sens. Lett. **16**(2), 251–255 (2018)
20. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 834–848 (2017)
21. Gurita, A., Mocanu, I.G.: Image segmentation using encoder-decoder with deformable convolutions. Sensors **21**(5), 1570 (2021)
22. Baheti, B., Innani, S., Gajre, S., Talbar, S.: Semantic scene segmentation in unstructured environment with modified deeplabv3+. Pattern Recogn. Lett. **138**, 223–229 (2020)
23. Xi, D., Qin, Y., Wang, Z.: Attention deeplabv3 model and its application into gear pitting measurement. J. Intell. Fuzzy Syst. **42**(4), 3107–3120 (2022)
24. Kuanqi, G., Yutong, L., Yuwei, Z., Lielie, Q., Nannan, Z., Yingli, C.: Photovoltaic hot spot detection of aerial infrared image based on deep learning. Electron. Meas. Technol. **45**(22), 7 (2022)
25. ChunMing, W., YiQian, S., TiaoJun, W., YaLi, L.: Underwater trash detection algorithm based on improved yolov5s. J. Real-Time Image Proc. **19**, 911–920 (2022)
26. Zhiqiang, H., Ying, S., Hao, G., Juanjuan, L., Sugang, M., Jiulun, F.: M-yolo: an object detector based on global context information for infrared images. J. Real-Time Image Process. **19**, 1009–1022 (2022)
27. Chen, B., Zhang, Z., Liu, N., Tan, Y., Liu, X., Chen, T.: Spatiotemporal convolutional neural network with convolutional block attention module for micro-expression recognition. Information **11**(8), 380 (2020)
28. Zhang, L., Cheng, J., Liu, J., Liu, T., Xiang, D., Yi, S.: Unsupervised ship detection in sar images using superpixels and cspnet. IEEE Geosci. Remote Sens. Lett. **20**, 1–5 (2023)
29. Zhang, Y.-F., Ren, W., Zhang, Z., Jia, Z., Wang, L., Tan, T.: Focal and efficient iou loss for accurate bounding box regression. Neurocomputing **506**, 146–157 (2022)