Application of YOLOv5 in Device Detection of Hydropower Station

Shouyuan Zhao, Chao Wen(✉), Yifeng Zhao, Liangliang Nie, Xiaoyu Zhang, Jialin Zou, and Yuxi Wu

CSG Power Generation Co., Ltd. Maintenance and Test Branch, Guangzhou 511400, China
xfliu0102@163.com

Abstract. Efficient detection and classification of devices are very important for effective maintenance in hydropower stations. However, the conventional manual approach suffers from low accuracy, efficiency, and reliability. Nowadays, object detection based on deep learning has achieved great development. In particular, the YOLO series models demonstrate significant advantages in terms of high accuracy and speed and robustness in complex image backgrounds, which have been widely used in numerous contexts for multi-object recognition tasks. Thus, in this paper, a YOLOv5-based algorithm is proposed to realize real-time multi-device recognition in hydropower stations. We labelled about 600 device photos of 27 categories collected from the site, based on which comparative experiments were carried out to evaluate the performance of YOLOv5 models of different configurations. The experimental results show that the mAP of YOLOv5m outperforms the others, the mAP of YOLOv5m can reach 95.3% and its precision can reach 94.1%. The study tests and reveals the effectiveness of YOLOv5 for device detection in hydropower stations. As such, the study on the one hand provides a useful asset management tool for the maintenance team, for another provides valuable empirical data for other studies that apply deep learning models in similar industrial situations.

Keywords: Mixed reality · Maintenance training · Power plants

1 Introduction

Energy is not only the basic guarantee for the survival of human society, but also an important material basis for economic and social development. Hydropower, as a clean and renewable energy, plays an irreplaceable role in the context of green and low-carbon development. In addition, traditional hydropower stations are inefficient in safety management, devices maintenance, technicians training, emergency drills and other aspects. Therefore, the intelligent and digital construction of hydropower station is particularly important. For example, the realization of intelligently detecting and monitoring devices, information exchange and sharing between hydropower station devices and system, etc., are the main direction of the future development of hydropower station. This paper mainly from the aspects of device detection to carry out research.

Intelligent hydropower station can give the most intuitive feelings to administrative staff and engineering technicians in its various tasks scenario. For example, in the scene

W. Cao et al. (Eds.): ISNEET 2022, LNEE 1017, pp. 43–52, 2023.
https://doi.org/10.1007/978-981-99-0553-9_6

of devices monitoring and maintenance work, we often need to operate the target device and synchronize the status and other information of the device to the cloud in real-time, so as to realize the intelligent operation and management of hydropower station. But before that, device detection is the prerequisite for the realization of the above intelligent and digital hydropower station. However, ordinary object detection in industrial scenes such as hydropower stations is still a thorny problem due to various devices, complex image background and susceptible to light and other factors. Furthermore, under the noisy working environment of high decibel for a long time, it is inevitable that the device detection efficiency of manual method is low, and we have to consider that sometimes the engineering technicians, especially the new ones who are not familiar with the work content, may also have the situation of detection errors. As summarized above, an efficient object detection algorithm is crucial for intelligent hydropower station. It provides intuitive and efficient solutions in improving efficiency of training technicians, work tasks and productivity, etc.

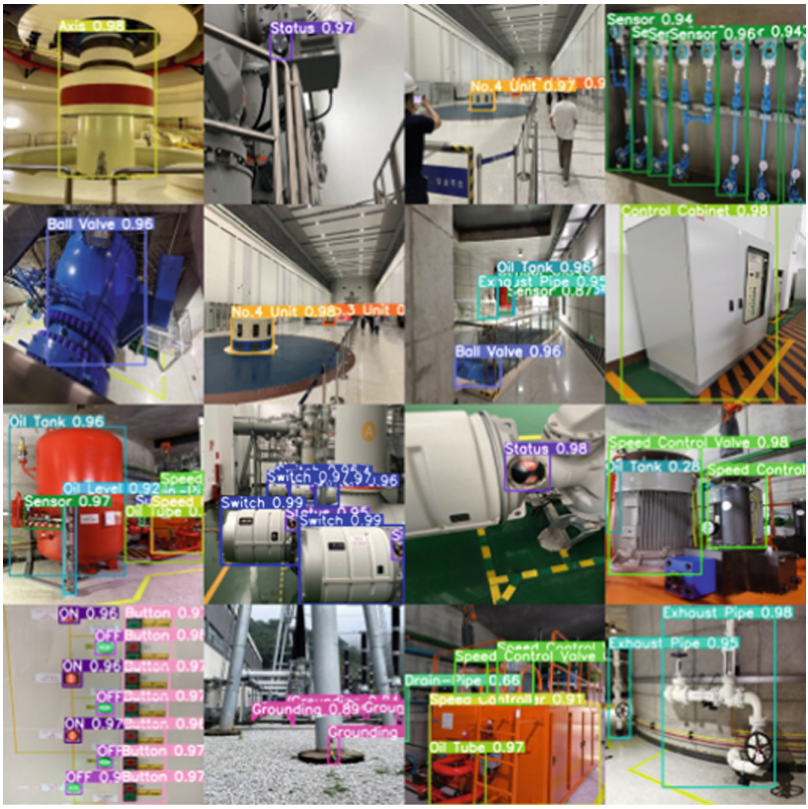

Fig. 1. We trained YOLOv5 with the photos of industrial scene, and then carried out real-time device detection of the hydropower station.

In recent years, tasks of computer vision based on deep learning method have been widely approved by industrial community. The goal of object detection is to find out all the objects of interest in the image (in our application, the objects that interested refer to the devices in the hydropower station), and accurately predict their positions and categories, which is one of the core problems in the field of computer vision. As one of the most powerful algorithms in the object detection, YOLOv5 has been widely used in the industrial community. In summary, we tested YOLOv5s, YOLOv5m and YOLOv5l in the image of the maintenance scenario of hydropower station. The experimental results show that YOLOv5s has low accuracy but can carry out real-time detection, while YOLOv5l has high accuracy, however, due to the complex model structure and image background, real-time detection cannot be achieved because of the large amount of calculation in forward reasoning, while the accuracy and real-time detection speed of YOLOv5m can meet the requirements of industrial scenes. And it can achieve ideal results in some small object detection tasks as shown in Fig. 1.

2 Related Work

In this part, we will introduce object detection models that have been applied in various fields of industry and our method to detect devices of hydropower station.

2.1 Object Detection

In recent years, researchers have put forward many solutions to these problems. As stated by Zheng et al. [1], the concept of object detection is a large scope, which usually also includes behavior detection [2], face detection [3], vehicle detection [4], fruit detection [5] and so on. Object detection algorithms based on deep learning are mainly classified into two types, one is two-stage framework based on region proposal and the other is one-stage framework based on classification and regression.

Based on Two-Stage Framework: It usually consists of region proposal generation, using CNN to extract image features, classification and regression of bounding boxes. The R-CNN [6] can be said is the pioneer of the use of deep learning method for object detection, it is not like the traditional method using sliding window to determine all possible region, instead, a series of regions proposals that may be objects are extracted in advance, and then features are extracted on these region proposals using deep training network, and the features are sent into the SVM classifier of each class to judge whether it belongs to this class or not. Finally, the position of bounding box is adjusted by using regressor. The Fast R-CNN [7] proposed by Girshick inputs the images to be processed and the RoI into the deep convolutional neural network to obtain the feature map. In the feature map, feature vectors are extracted through the full connection layer and sent to two branches, one is the classification of objects and the other is the regression of bounding boxes. In other word, multi-task learning mechanism is adopted in this model, its evaluation performance on 16.5k dataset reached nearly 70% mAP. Based on the above two models, the Faster R-CNN [8] proposed by He et al. abandoned the Selective Search method of RCNN and directly used Region Proposal Network to generate region

proposals, and then makes separate predictions for each of its region proposals. Its mAP on the VOC2007 dataset can reach 73.2%. But it is obvious that it can be costly in terms of time because of the multiple runs of detection and classification.

Based on One-Stage Framework: Generally, image pixels are directly mapped to bounding box coordinates and classification probability, and all bounding boxes can be predicted through the detector at once. This can greatly reduce the time cost, so it can meet the real-time requirements of portable mobile devices. Common detectors based on one-stage such as YOLO series and SSD models, are based on classification and regression frameworks. Before this, R-CNN series algorithm occupied half of the field of object detection. Then Redmon et al. [9] proposed a new object detection model named YOLO, its mAP achieved 75.0% on the VOC2007 test set and it is much less affected by complex image background than Fast R-CNN. He et al. [10] proposed RetinaNet which is also based on one-stage object detection algorithm, it uses FPN to get multi-scale feature map on the basis of feedforward ResNet and two subtasks of classification anchor and regression of positive anchor. Its accuracy is also close to Faster R-CNN based on two-stage but faster. Soon, Redmon et al. [11] introduced YOLO9000 (YOLOv2) using multi-scale training method that can detect more than 9000 object classes. At 67 FPS, it can achieved 76.8% mAP on the VOC2007. In 2018, Redmon [12] made some updates to YOLO, mainly in network structure, it uses DarkNet-53 as feature extraction and FPN for multi-scale feature maps and switch to logistic regression as its classification method, YOLOv3 takes much less time than RetinaNet to process an image. Bochkovskiy et al. [13] introduced some practical techniques based on traditional YOLO, which are superior in speed and precision to the fastest and most accurate models that previously proposed. Not long after YOLOv4 appeared, YOLOv5 was proposed by Ultralytics et al. YOLOv5 has made further improve ments on the basis of YOLOv4 and YOLOv3, mainly including the input layer, Backbone, Neck, and Head output layer, so that its speed and accuracy have been greatly improved. In addition, YOLOv5 can reach up to 55.8% mAP in the COCO dataset, and it has been widely used in various object detection scenarious in the industry, which is also why we choose YOLOv5 as the device detection model of hydropower station.

2.2 Device Detection

The realization of intelligent inspection and maintenance in the industrial scene attracts a large number of engineering personnel. For example, Li et al. [14] used Fast R-CNN for pedestrain detection, and their method was better than other methods in INRIA data set. Its mAP reached 61.61% on KITTI data set (Easy). Jiang et al. [15] used Faster R-CNN for face detection, and compared with R-CNN and Fast R-CNN in performance, in which the detection rate of both R-CNN and Fast R-CNN was around 90%, while the detection rate of Faster R-CNN reached 96%. Warsi et al. [16] regard YOLOv3 as the object detection algorithm of the gun detection system. In countries where guns are legal, the use of surveillance cameras for detection and early detection can help prevent such violent acts. Zhou et al. [17] applied YOLOv5 to the helmet detection in the factory to further standardize workers' work and reduce the occurrence of factory

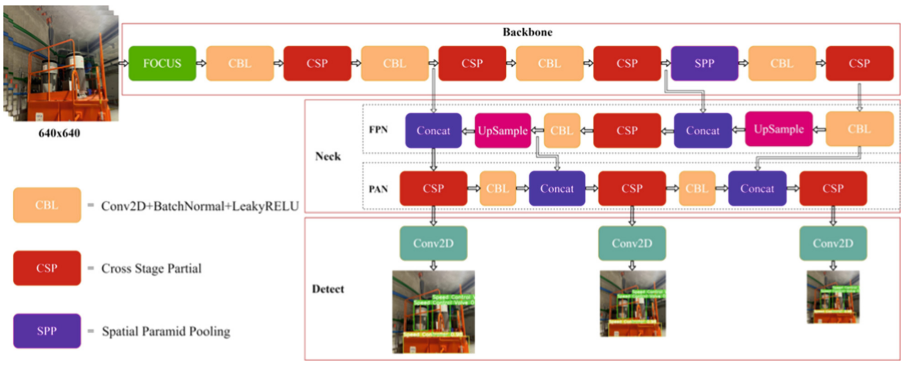

Fig. 2. The YOLOv5 network structure

safety accidents. Luo et al. [18] used YOLOv5-based framework to detect aircraft in remote sensing images, and simultaneously tested Faster R-CNN and YOLOv5. The result showed that Faster R-CNN could obtain 74.86% mAP at 8.93FPS, while YOLOv5 could achieve 81.51% mAP at 41.92FPS. This once again illustrates the shortcoming of the two-stage framework. Feng et al. [19] used YOLOv5 framework to solve the problem of mask recognition and successfully deployed it in malls and other public places for mask detection, which is of great significance in the context of COVID-19 pandemic. And Liu et al. [20] used YOLOv5 for real-time detection of railway signal lights to reduce the risk of railway traffic accidents. In this article, we use YOLOv5 for real-time electrical equipment detection in the scene such as inspection, maintenance, and the model will be deployed to mobile intelligent devices which has the function of augmented reality technology, through the combination of object detection and virtual simulation technology, it can better experience the intuitive feeling brought by human-computer interaction, thus not only can help engineering technicians with understanding of hydropower devices knowledge, but also improve the efficiency of inspection.

Besides, through reviewing a large number of literature, we find that there are few researchers who apply the algorithm mentioned above to the scene of electric devices detection in hydropower stations, which will be an innovative application. We hope that our application can fill this gap and also help similar studies in future.

3 Method

This section proposed the application of YOLOv5 in the detection of hydropower station devices. We introduce the network structure of YOLOv5, then we describe the sources of our data sets and their evaluation indicators in detail.

3.1 YOLOv5 Structure

Redmon, the originator of YOLO, withdrew from Computer Vision field after releasing YOLOv3. Then Bochkovskiy put forward YOLOv4, while YOLOv4 was still hot, Ultralytics et al. shortly launched YOLOv5 based on the previous versions, whose speed and

performance were greatly improved. In the input side, we cropped the collected device photos to the size of 640 × 640, and performed a pre-processing operation on them with data enhancement.

In Backbone, we input images to the backbone that consist of Focus and CSP [21] for feature extraction. The Focus part is to slice the input images of 640 × 640 × 3 and to splice tensor to get the feature map of 320 × 320 × 12, and then obtain the feature map of 320 × 320 × 48 through 48 convolution kernels, which is not included in the previous version. CSP is to reduce the computation and memory usage of the network and improve the reasoning ability and accuracy of the model.

In addition, feature maps of different scales are input into the Neck structure that consist of FPN and PAN [22] network for feature fusion. FPN samples high-level semantic features from the top-down, while PAN samples high-positioning features from the bottom-up, so as to strengthen the feature fusion ability of network.

Finally, the multi-scale feature maps are predicted and output in the Detect section, and the YOLOv5 network structure is shown in Fig. 2.

3.2 The Introduction of Dataset

Dataset are very important for model training. Our dataset is based on 600 device photos collected at the site of the hydropower station, including a total of 27 categories. The tool labelimg was used to annotate them, and the annotated data was converted from XML format to TXT format as required by YOLO. Finally, 85% of the device photos were randomly selected as training samples. The number of times that each devices are labeled is presented in Fig. 3.

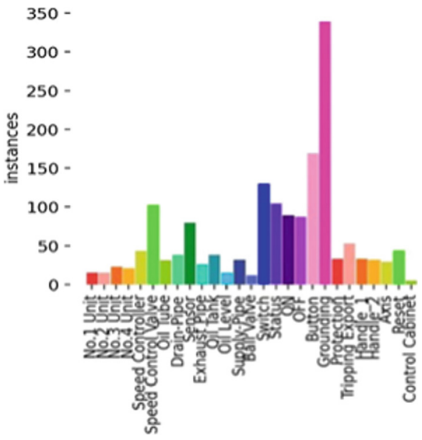

Fig. 3. The annotating times of device.

3.3 Evaluation Indicator

Unlike single-object detection, precision can not be directly used in multi-object detection due to the possibility of multiple devices in each image. In multi-target detection,

precision, mAP and recall are commonly used in multi-object detection for evaluating models. Where precision is defined as follows:

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{1}$$

rather than accuracy, it means the probability of being annotated true labels in the predicted objects, where TP refers to the number of objects that correctly detected, FP is the number of objects that incorrectly detected. And recall can be defined as follows:

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{2}$$

it means the probability of true samples predicted among all true samples, where FN refers to the number of objects in object detection under the condition of leakage. Finally, mAP index is used to evaluate detection precision in object detection, and its calculation formula is as follows:

$$mAP = \frac{\sum_{i=1}^{n} A_i P}{n} \tag{3}$$

it means the average of the average precision of each device, where n is the number of device categories and AP represents the average precision.

4 Experiment

Here we are going to introduce the experimental details of this paper, including the experimental environment, experimental data, result and analysis, then compare three YOLOv5 models of different weights and show the final results, and all experiments are running under Python3.7 and Windows10 environment.

4.1 Experimental Platform

In this paper, our experiment was run under the configuration that shown in the following table (Table 1):

Table 1. Experimental environment

Category	Parameters
CPU	Intel(R) Core(TM) i9-10920X CPU @ 3.50GHz
RAM	128G
GPU	Nvidia GeForce RTX 3090 24G
CUDA	11.3
Cudnn	8.4.1

4.2 Experimental Data

Due to the high pixel value of photos collected on site, if directly input them into the model that would bring huge computational overhead and poor real-time performance, that is undesirable and unnecessary. Therefore, we cropped the data sets to the size of 640 × 640 required by the model in batches. Then we divided randomly the data sets that are annotated into training set and test set, and the division ratio was 8.5:1.5. The annotating times of each device is shown in Fig. 3 and then the three different models of YOLOv5 will be trained and verified. Before the training, we set the training epoch to 300, and other hyperparameters such as learning rate are shown in the following table (Table 2):

Table 2. Network hyperparameters

Parameters	Values
Initial learning rate	0.00001
Optimizer	0.001
Epochs	300
IoU training threshold	0.7
Batch size	64

4.3 Result and Analysis

In order to obtain a model with better detection efficiency for power device in real scenes, we trained and tested YOLOv5s, YOLOv5m and YOLOv5l in exactly the same environment. The precision and classification loss of the above three models are shown in Fig. 4.

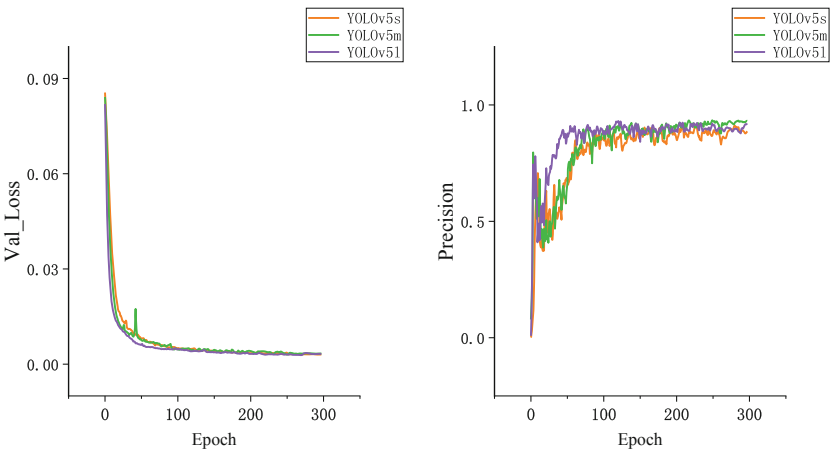

Fig. 4. The evaluation results of three models.

To facilitate comparison, the performance indicators of the three models are shown in the following table (Table 3):

Table 3. The performance of three different models

Model	mAP	Precision	FPS
YOLOv5s	0.925	92.3%	95
YOLOv5m	0.953	94.1%	64
YOLOv5l	0.952	93.7%	27

As shown in Fig. 4, the horizontal coordinate is the epoch of training, and the vertical coordinate is the validation class loss and precision. With the epoch approaching 300, the validation loss of three models were converging roughly the same, but from the perspective of precision, although YOLOv5m is slightly worse than YOLOv5l in the early epoch, its later performance has steadily improved and even surpassed YOLOv5l. Compared with YOLOv5m and YOLOv5l from the above table, FPS (Frame Per Second) refers to the number of images that can be displayed per second. Due to the complexity of the network, YOLOv5l has a large number of parameters to be calculated by the model, so its FPS is only 27. It is not effective in real-time detection of actual industrial scenes. Therefore, based on the above indicators, it can be concluded that YOLOv5m has high precision and fast speed in actual detection scenes, which can meet the requirements of industrial real-time detection.

5 Conclusion

We use object detection model based on deep learning named YOLOv5 to detect and track power devices in novel industrial scenarios. Under the complex background of real industrial environment and the negative influence of light, YOLOv5 still shows its excellent detection ability. First of all, we cut the photos collected on site, annotate them and transform them into the formats that required, and do a pre-processing operation with data enhancement, which can enhance the ability of model feature extraction. Secondly, we use three different YOLOv5 models to train the collected data sets. Finally, we verified their detection effectiveness in a real industrial environment, with YOLOv5m demonstrating its superior detection capability in this novel industrial application.

Acknowledgments. This work was supported by the Technology Project of CSG POWER GENERAT-ION Co., Ltd (No. 022200KK52200001).

References

1. Zhao, Z.Q., Zheng, P., Xu, S., et al.: Object detection with deep learning: a review. IEEE Trans. Neural Netw. Learn. Syst. **30**(11), 3212–3232 (2019)

2. Shahverdy, M., Fathy, M., Berangi, R., et al.: Driver behavior detection and classification using deep convolutional neural networks. Expert Syst. Appl. **149**, 113240 (2020)
3. Jiang, H., Learned-Miller, E.: Face detection with the faster R-CNN. In: IEEE International Conference on Automatic Face and Gesture Recognition, pp. 650–657. IEEE (2017)
4. Sang, J., Wu, Z., Guo, P., et al.: An improved YOLOv2 for vehicle detection. Sensors **18**(12), 4272 (2018)
5. Wan, S., Goudos, S.: Faster R-CNN for multi-class fruit detection using a robotic vision system. Comput. Netw. **168**, 107036 (2020)
6. Girshick, R., Donahue, J., Darrell, T., et al.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)
7. Girshick, R.: Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1440–1448 (2015)
8. Ren, S., He, K., Girshick, R., et al.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. **39**(6), 1137–1149 (2017)
9. Redmon, J., Divvala, S., Girshick, R., et al.: You only look once: Unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779–788 (2016)
10. Lin, T.Y., Goyal, P., Girshick, R., et al.: Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2980–2988 (2017)
11. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7263–7271 (2017)
12. Redmon, J., Farhadi, A.: Yolov3: an incremental improvement. arXiv preprint arXiv:1804.02767 (2018)
13. Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M.: Yolov4: optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 (2020)
14. Li, J., Liang, X., Shen, S.M., et al.: Scale-aware fast R-CNN for pedestrian detection. IEEE Trans. Multimedia **20**(4), 985–996 (2017)
15. Jiang, H., Learned-Miller, E.: Face detection with the faster R-CNN. In: 2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017), pp. 650–657. IEEE (2017)
16. Warsi, A., Abdullah, M., Husen, M.N., et al.: Gun detection system using YOLOv3. In: 2019 IEEE International Conference on Smart Instrumentation, Measurement and Application (ICSIMA), pp. 1–4. IEEE (2019)
17. Zhou, F., Zhao, H., Nie, Z.: Safety helmet detection based on YOLOv5. In: 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA), pp. 6–11. IEEE (2021)
18. Luo, S., Yu, J., Xi, Y., et al.: Aircraft target detection in remote sensing images based on improved YOLOv5. IEEE Access **10**, 5184–5192 (2022)
19. Yang, G., Feng, W., Jin, J., et al.: Face mask recognition system with YOLOV5 based on image recognition. In: 2020 IEEE 6th International Conference on Computer and Communications (ICCC), pp. 1398–1404. IEEE (2020)
20. Liu, W., Wang, Z., Zhou, B., et al.: Real-time signal light detection based on yolov5 for railway. In: IOP Conference Series: Earth and Environmental Science, vol. 769, no. 4, p. 042069. IOP Publishing (2021)
21. Wang, C.Y., Liao, H.Y.M., Wu, Y.H., et al.: CSPNet: a new backbone that can enhance learning capability of CNN. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 390–391 (2020)
22. Liu, S., Qi, L., Qin, H., et al.: Path aggregation network for instance segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8759–8768 (2018)