Safety Helmet Detection Based On YOLOV4-M

Si Gao, Yiting Ruan, Yue Wang, Weijie Xu, Mingfeng Zheng

(Hangzhou Hikvision Digital Technology Co., Ltd.)

Abstract

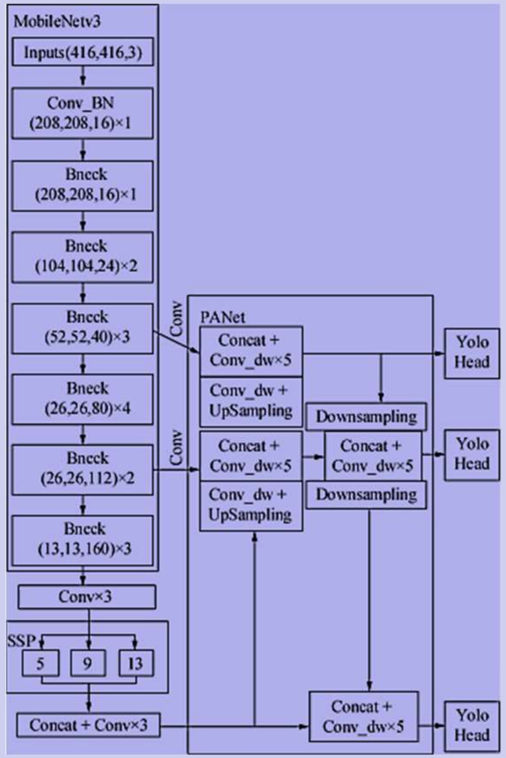
In the intelligent monitoring of the construction site, there is often the problem of not wearing a safety helmet. In order to effectively judge the wearing condition of the helmet, a detection method of YOLOV4-M is proposed. This method replaces the backbone network in YOLOV4 with MobileNetV3, uses depthwise separable convolution to reduce the amount of parameters of the backbone network, and uses H-swish to improve the performance of the model. Experiments show that the final detection accuracy rate is 98.2%, and the FPS reaches 40 frames, which can be well applied to the detection of helmets.

EXPERIMENT

(c) (d)

In Figure , avatars with helmets will be marked in green, and avatars without helmets will be marked in red. In Figure (a) and (b), it can be found that although the helmet is not worn on the head, it is still detected. At the same time, in this scene, only the detection frame is required to be marked, and the corresponding confidence information is not displayed. .Figure (c) and Figure (d) are another scene, from which it can be found that the method also has a better effect on some overlapping parts.

Methods



When detecting the helmet, first, the input image is subjected to a convolution with a step size of 2 and the width and height of the normalized compressed image are expanded and the number of channels is expanded, and then the Resblock module of the main feature extraction network of YOLOv4 is replaced by the bneck module. And build MobileNet-YOLO according to the network structure of MobileNetv3; at the same time, in the process of feature fusion, the ordinary convolution in the PANet network is replaced with a depthwise separable convolution.

Results

methods	MAP
YOLOV1-MobileNetV3	0.921
YOLOV2-MobileNetV3	0.930
YOLOV3-MobileNetV3	0.964
YOLOV4-M	0.982
FasterRCNN-MobileNetV3	0.980
SSD-MobileNetV3	0.931
YOLOV4	0.962

It can be found from the table that, agreeing to use MobileNetV3, the accuracy of Faster RCNN using the two-stage method is basically greater than that of YOLO and SSD of the one-stage method, but YOLOV4-M is slightly higher than Faster RCNN, which shows that the use of YOLOV4-M is effective, while also 2 percentage points higher than the improved YOLOV4. In order to further explore the effect of exploration in the model, we selected 317 samples with occlusion from the test set samples for testing.

Acknowledgment

We would like to thank the participants and companies who take part in this study.This work is supported by Research on muti-spectral spectral video intelligent inspection technology for substation equipment under Science and technology project of State Grid Corporation of China (5200-201919048A-0-0-00)

name	parameter
memory	24GB
information	Tesla K80
CUDA version	CUDA 11.0
Deep learning framework	Torch 1.4.0

This paper adopts the self-made helmet dataset. In the actual scene of the factory, a total of 1844 samples were collected. According to the ratio of 2:1, it is divided into training set and test set. LabelImg labeling software is used to calibrate the data set, and then a standard data set is established according to VOC format. The development platform used in this experiment is Ubuntu 16.0.

methods	MAP	FPS
YOLOV4	0.943	25
YOLOV4-M	0.940	40
YOLOV4-tiny	0.910	42

From the experimental results, we can find that the improved YOLOV4-M is only 0.3% lower in accuracy than YOLOV4, but the speed is increased by 60%. percentage point.

Conclusion

Aiming at the problem of helmet detection, this paper combines MobileNetV3 and YOLOV4 to propose a new model for helmet detection based on YOLOV4-M. After experiments, the method achieves a detection accuracy of 98.2% and has a detection speed of 40 FPS, which meets the testing requirements.