Construction worker hardhat-wearing detection based on an improved BiFPN

Chenyang Zhang, Zhiqiang Tian*, Jingyi Song, Yaoyue Zheng, Bo Xu

{zcyuniant, jingyisong, z1037268262, xuborj}@stu.xjtu.edu.cn, *zhiqiangtian@xjtu.edu.cn

Abstract

Work in the construction site is considered to be one of the occupations with the highest safety risk factor. Therefore, safety plays an important role in construction site. One of the most fundamental safety rules in construction site is to wear a hardhat. To strengthen the safety of the construction site, most of the current methods use multi-stage method for hardhat-wearing detection. These methods have limitations in terms of adaptability and generalizability. In this paper, we propose a one-stage object detection method based on convolutional neural network. We present a multi-scale strategy that selects the high-resolution feature maps of DarkNet-53 to effectively identify small-scale hardhats. In addition, we propose an improved weighted bi-directional feature pyramid network (BiFPN), which could fuse more semantic features from more scales. The proposed method can not only detect hardhat-wearing, but also identify the color of the hardhat. Experimental results show that the proposed method achieves a mAP of 87.04%, which outperforms several state-of-the-art methods on a public dataset.

Contributions

(1) Different from the high-cost sensor-based and multi-stage vision-based hardhat wearing detection, the proposed method is based on an end-to-end one-stage model, which is effective and robustness in a complex situation.

(2) We conduct the position regression at the high-resolution feature maps to boost the detection accuracy of the small objects. To fuse more semantic information of features, we propose an improved simple and effective weighted bi-directional feature pyramid network.

(3) Moreover, we adopt a multi-scale training strategy to generate different scales of inputs, which can further improve the performance and robustness of the model. The experimental results show that the proposed network is effective and robust.

Proposed Approach

We propose a one-stage object detection model. The model uses DarkNet-53 as a feature extractor, and sends the extracted features into an improved BiFPN for feature fusion. The fused features are used for prediction cross scale. Moreover, generalized IoU loss, focal loss, and cross-entropy loss are used as bounding box regression loss, confidence loss, and classification loss, respectively. An overview of the proposed method is shown in Fig. 1.

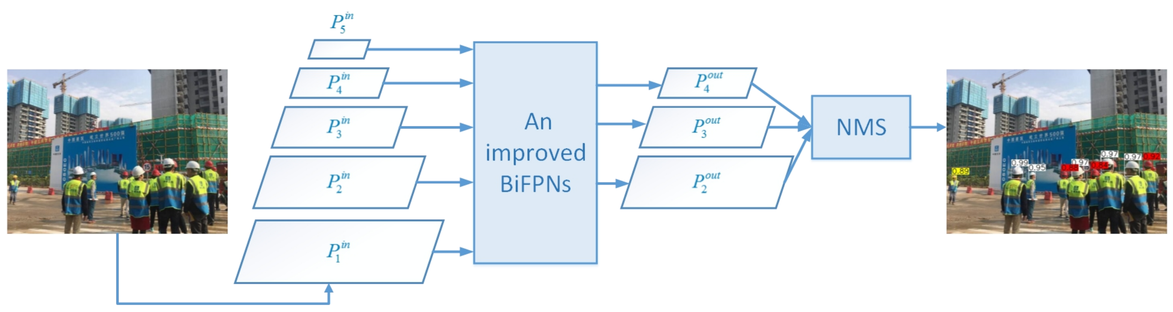

Fig.1. Architecture of one-stage hardhat wearing detection model.

A.Backbone:Darknet 53

Table I. Architecture of Darknet-53

	Layer Type	# of Filters	Filter Size	Output	P
	Convolutional	32	3×3	256×256	
	Convolutional	64	3×3/2	128×128	
1×	Convolutional	32	1×1		
	Convolutional	64	3×3		
	Residual			128×128	P_1^{in}
	Convolutional	128	3×3/2	64×64	
2×	Convolutional	64	1×1		
	Convolutional	128	3×3		
	Residual			64×64	P_2^{in}
	Convolutional	256	3×3/2	32×32	
8×	Convolutional	128	1×1		
	Convolutional	256	3×3		
	Residual			32×32	P_3^{in}
	Convolutional	512	3×3/2	16×16	
8×	Convolutional	256	1×1		
	Convolutional	512	3×3		
	Residual			16×16	P_4^{in}
	Convolutional	1024	3×3/2	8×8	
4×	Convolutional	512	1×1		
	Convolutional	1024	3×3		
	Residual			8×8	P_5^{in}
	Avgpool		Global		
	Connected		1000		
	Softmax				

B. Mutil-scale feature fusion:An improved BiFPN

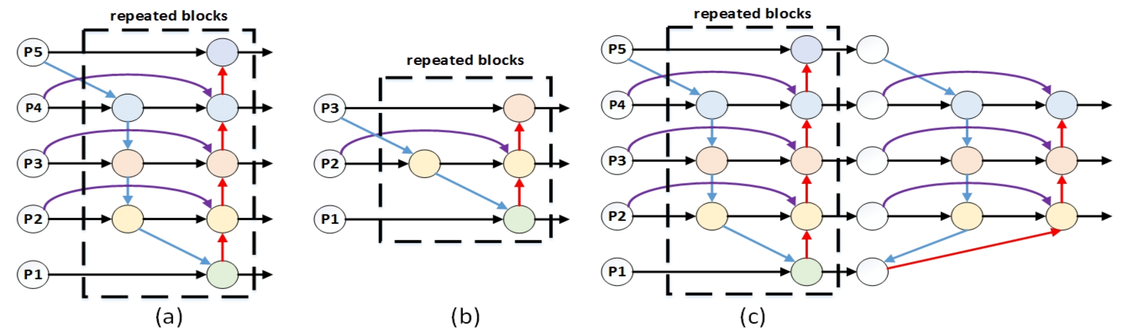

Fig.2. Feature pyramid network design-(a) is the original BiFPN, which has 5 levels of input and output; (b) is 3 levels of input and output BiFPN; (c) is an improved BiFPN, both in the n-1 layer 5 levels of input and output, but the n-th layer is 5 levels of input and 3 levels of output.

C.Predictions Across Scales

There are three sizes of the bounding boxes, which are used in the proposed method. Each size predicts 3 bounding boxes, so the resulting tensor is N×N×[3×(4+1+5)], which contains 4 bounding box offsets, an objectness prediction, and 5 class predictions, representing blue, white, yellow, red, and not wearing a hardhat, respectively. The vectors used for prediction are obtained from the improved BiFPN.

D.Loss

The loss function is composed of three categories, which are bounding box regression loss, L_{bbox}, confidence loss, L_{conf} and classification loss, L_{class}. The loss function is defined as follows:

$$L = L_{bbox} + L_{conf} + L_{class} .$$

E.Implementation Detail

We used NMS, fine-tune, K-means anchors and multi-scale training to implement our model.

Results

Table 2. Detection results with respect to GDUT-HWD test.

methods	Input size	Backbone	mAP (%)	AP(%)					Avg. precision, area (%)		
				Blue	White	Yellow	Red	None	small	medium	large
Faster R-CNN	300×500	VGG16	65.67	70.80	68.03	69.57	60.95	59.01	37.86	80.75	86.67
FPN	384×640	ResNet-50	73.26	77.61	75.49	76.86	69.67	66.66	50.43	84.83	90.32
SSD 512	512×512	VGG16	83.27	86.15	85.46	88.16	80.57	76.02	66.13	88.36	90.96
SSD -RPA 512	512×512	VGG16	83.89	86.35	86.05	89.17	80.10	77.76	67.05	87.88	90.86
YOLO v3	512×512	Darknet-53	83.69	84.94	85.82	86.81	82.68	78.22	69.55	87.84	89.57
Ours	320×320	Darknet-53	79.88	81.89	81.92	83.41	78.69	73.39	61.89	85.53	89.47
	512×512	Darknet-53	87.04	87.85	88.68	89.30	87.20	82.17	74.39	87.92	87.88

Fig. 3. Detection results on GDUT-HWD test with our model. Each color corresponds to an object category. The blue, white, yellow, red box represents the detected hardhat is in blue, white, yellow, and red, respectively. And the number in the box indicates the confidence score. The green box represents that there is no hardhat.

Conclusion

We found that the robustness based on the deep learning method was better than that based on the traditional visual method because of the complexity of the scene on the construction site. Therefore, We proposed a new method based on deep learning for hardhat-wearing detection, which could reduce the risk of reducing accidents caused by not wearing a hardhat. We did relevant experiments on GDUT-HWD and prove that our model is the most advanced performance.

References

[1]Jixiu Wu, Nian Cai, Wenjie Chen, Huiheng Wang, Guotian Wang, "Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset," Automation in Construction, vol. 106, pp. 102894, 2019.

[2]J. Redmon, A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, 2018.

[3]Tan, M., Pang, R., Le, Q.V., "Efficientdet: Scalable and efficient object detection," arXiv preprint arXiv:1911.09070, 2019.