Construction worker hardhat-wearing detection based on an improved BiFPN

Chenyang Zhang, Zhiqiang Tian, Jingyi Song, Yaoyue Zheng, Bo Xu
School of Software Engineering, Xi'an Jiaotong University, Xi'an, China

INTRODUCTION

The construction industry plays an important role in national economy. However, for construction, which is very important to the country, its employees have the most dangerous occupations in the world. Therefore, the construction safety management regulations attach great importance to the correct use of personal protective equipment (PPE), especially hardhat. The wearing of hardhat can protect construction workers from head injury. But in order to check the wearing of hardhat, it will cost a lot of manpower and time. Our work can assist in checking the wearing of hardhat and reduce the cost of manpower and time.

PROBLEM

Currently, the automatic hardhat-wearing detection methods can mainly fall into two categories: **sensor-based methods** and **vision-based methods**.

Though the sensor-based methods can provide a promising result, the high cost is the main cause that it cannot be adopted widely. For this reason, vision-based solutions have become the overwhelming majority.

PROBLEM

It is extremely challenging for the **vision-based methods** to achieve impressive results. There are two main challenges:

1) the complexity of the environment, including the overexposure, excessive dimness, excessive background noise, and so on;

2) the uncertainty of personnel, the heavy occlusion of the crowd, and the size diversity of personnel.

PROPOSED APPROACH

1. Detection Model

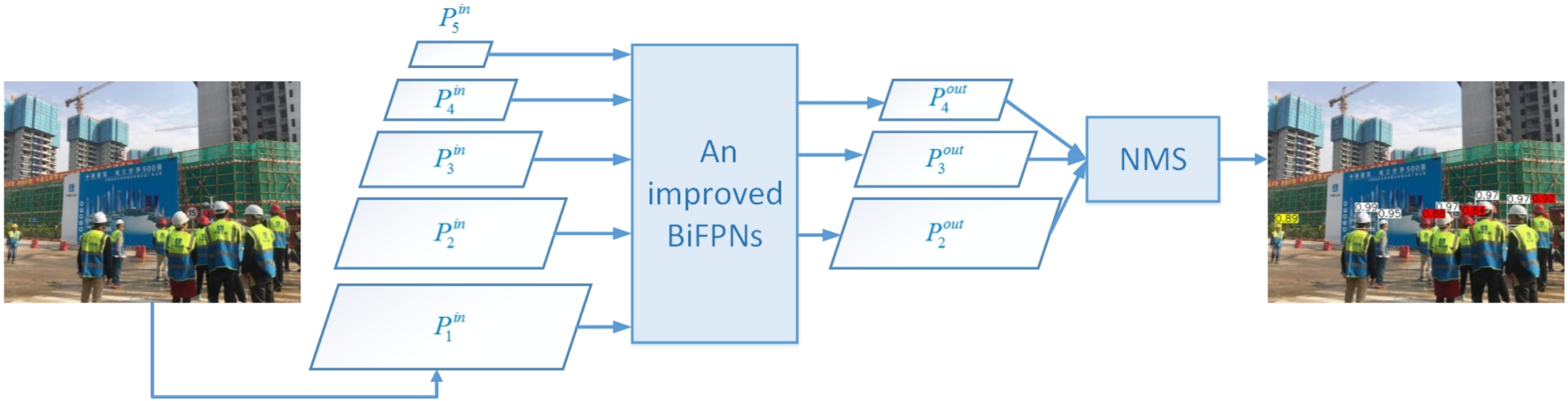

Fig.1 Architecture of one-stage hardhat wearing detection model.

PROPOSED APPROACH

2. Multi-scale feature fusion:An improved BiFPN

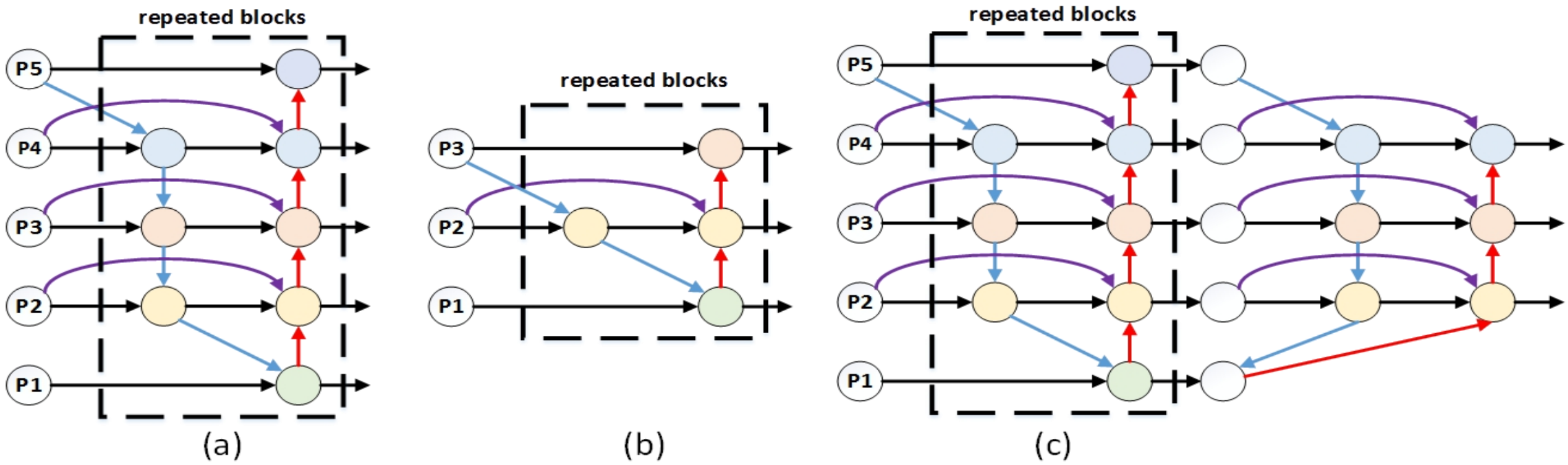

Fig. 2 Feature pyramid network design-(a) is the original BiFPN, which has 5 levels of input and output; (b) is 3 levels of input and output BiFPN; (c) is an improved BiFPN, both in the n-1 layer 5 levels of input and output, but the n-th layer is 5 levels of input and 3 levels of output.

PROPOSED APPROACH

3. Loss

The loss of the detection is composed of three parts, namely bounding box regression loss, $Loss_{bbox}$, confidence loss, $Loss_{conf}$ and classification loss, $Loss_{class}$.

$$L = Loss_{bbox} + Loss_{conf} + Loss_{class} \tag{1}$$

Among them, the **confidence loss** uses focal loss, which mainly solves the problem of a serious imbalance in the ratio of positive and negative samples. **Classification loss** uses cross entropy loss, which is the most common loss function for classification problems.

PROPOSED APPROACH

3. Loss

The bounding box regression loss uses Generalized IoU Loss.

$$GIoU = IoU - \frac{|C/(A \cup B)|}{|C|} \tag{2}$$

In its expression, C is the smallest bounding box surrounding A and B, and IoU is the intersection ratio of A and B. GIoU can reflect the distance between two boxes, and its value range is [-1,1]. When the two boxes overlap, the value of GIoU is 1; when the two boxes do not overlap and the distance is infinite, the value of GIoU is -1.

DATASET

Table 1. The number of instances in the GDUT-HWD dataset.

Label	Numbers		
	Trainval	Test	Total
Blue	1251	1361	2612
White	1813	1968	3781
Yellow	1936	1962	3898
Red	2148	2083	4231
None	2171	2200	4371
Small	4237	4713	8950
Medium	4098	3826	7924
Large	984	1035	2019

RESULT

Fig.3 Detection results on GDUT-HWD test with our model.

RESULT

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

Methods	Input size	Backbone	mAP	AP				
				Blue	White	Yellow	Red	None
Faster R-CNN	300×500	VGG16	65.67	70.80	68.03	69.57	60.95	59.01
FPN	384×640	ResNet-50	73.26	77.61	75.49	76.86	69.67	66.66
SSD 512	512×512	VGG16	83.27	86.15	85.46	88.16	80.57	76.02
SSD-RPA 512	512×512	VGG16	83.89	86.35	86.05	89.17	80.10	77.76
YOLO v3	512×512	Darknet-53	83.69	84.94	85.82	86.81	82.68	78.22
Ours	320×320	Darknet-53	79.88	81.89	81.92	83.41	78.69	73.39
	512×512	Darknet-53	**87.04**	**87.85**	**88.68**	**89.30**	**87.20**	**82.17**

RESULT

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

Methods	Input size	Backbone	Small	Medium	Large	Avg
Faster R-CNN	300×500	VGG16	37.86	80.75	86.67	68.43
FPN	384×640	ResNet-50	50.43	84.83	90.32	75.19
SSD 512	512×512	VGG16	66.13	**88.36**	**90.96**	81.82
SSD -RPA 512	512×512	VGG16	67.05	87.88	90.86	81.93
YOLO v3	512×512	Darknet-53	69.55	87.84	89.57	82.32
Ours	320×320	Darknet-53	61.89	85.53	89.47	78.96
	512×512	Darknet-53	**74.39**	87.92	87.88	**83.40**

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. M.-W. Park, N. Elsafty, Z. Zhu, “Hardhat-wearing detection for enhancing on-site safety of construction workers,” Journal of Construction Engineering, vol. 2015, pp. 1-8, 2015.
2. B.E. Mneymneh, M. Abbas, H. Khoury, “Vision-based framework for intelligent monitoring of hardhat wearing on construction sites,” Journal of Computing in Civil Engineering, vol. 33, no. 2, pp. 04018066, 2019.
3. 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.
4. J. Redmon, S. K Divvala, R. Girshick, A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” in CVPR, 2016.
5. J. Redmon, A. Farhadi, “YOLO9000: Better, Faster, Stronger,” in CVPR, 2017.
6. J. Redmon, A. Farhadi, “YOLOv3: An Incremental Improvement,” arXiv preprint arXiv:1804.02767, 2018.
7. Tan, M., Pang, R., Le, Q.V., “Efficientdet: Scalable and efficient object detection,” arXiv preprint arXiv:1911.09070, 2019.

Thank you!