ICIC Express Letters
Part B: Applications
Volume **14**, Number **9**, September **2023**
 ISSN 2185-2766
pp. 917–925

FAST SAFETY HELMET WEARING DETECTION MODEL BASED ON YOLOV5

Yefei Zhang, Weicheng Xie*, Changquan Song, Xiaoming Yang and Hai Huang

School of Electrical Engineering and Electronic Information
Xihua University
No. 999, Jinzhou Road, Jinniu District, Chengdu 610039, P. R. China
{ 1344461017; 198746715; 1098435491 }@qq.com; 0120030022@xhu.mail.edu.cn
*Corresponding author: scxweicheng@mail.xhu.edu.cn

Received December 2022; accepted February 2023

Abstract. *The current safety helmet detection algorithm has poor real-time detection due to the existence of a large number of computing parameters in the ordinary convolution of the bottleneck module in the backbone network. In response to this problem, an object detection method of YOLOv5 is improved to optimize the efficiency of safety helmet wearing detection. Firstly, the C3 bottleneck module in the backbone network is replaced by the basic module of the lightweight network ShuffleNetV2. The channel mixing mode is adopted to decrease the computational complexity of the model. The experiment concluded that the improved detection model can achieve a recognition speed of 81.2 frames per second and a space of 9.8 MB without losing accuracy and reducing parameters, which improves the speed of detection.*
Keywords: Safety helmet detection, YOLOv5, ShuffleNetV2, Bottleneck module

1. **Introduction.** The safety of construction site has become one of the key issues that enterprises pay attention to and workers worry about. In the early stage, manual inspection was usually used to detect the wearing of the helmet. Since each construction link was scattered in different spaces, the problem of high personnel mobility led to low efficiency of manual supervision. Safety helmet detection mode is iterating and updating constantly. Nevertheless, the existing safety helmet detection model still has abundant problems in the case of complex construction scenes. For example, the detection accuracy is not high in the environment where the construction scene is changing. The complexity of the network model calculation leads to low detection efficiency and inability to prompt and early warning in time. At the same time, the safety helmet detection model has higher requirements on the hardware environment. In such a construction environment, many safety problems will occur. For example, workers do not wear safety helmets as required, which often leads to safety accidents. Recently, some scholars have made research on the real-time detection of safety helmet wearing [1], but the effect is not good; there are also corresponding researches on real-time detection of fire [2] and other safety issues. Consequently, it is particularly important to study a lightweight safety helmet wearing detection model in dangerous construction work scenarios.

Traditional algorithms are hard to identify targets and fail to meet real-time demands. Therefore, high-precision models have been proposed successively, such as Faster R-CNN [3] and Mask R-CNN [4]. Although the above algorithms have some advantages in detection accuracy, they ignore the defects of many parameters and large amount of calculation in the algorithm. Computational complexity is an important performance metric if the method is deployed on computer platform. Therefore, the end-to-end network model has

DOI: 10.24507/icicelb.14.09.917

been proposed by researchers, such as single shot multibox detector (SSD) [5], you only look once (YOLO) [6], and RetinaNet [7]. The detection efficiency has been improved.

Therefore, many researchers applied the YOLO algorithm to the detection of safety helmet wearing, and made innovations and improvements to the algorithm. For example, Chen et al. [8] introduced the function of Retinex image enhancement to improve the image quality of complex scenes. Then, the actual size of the safety helmet is clustered by K-means++ cluster analysis, but the calculation of the high-precision target box cannot locate the target in time. Ge et al. [9] integrated the high and low level features of the network and used adversarial training to improve the robustness of the model. However, feature fusion layer was added to the model resulting in a certain redundant structure. Zhao et al. [10] added the DenseBlock module and SE-Net [11] channel attention module to improve the model's recognition ability of small targets, but the densely connected layers would increase the computational parameters. Li et al. [12] proposed a layered positive sample selection mechanism to improve the fitting ability of the network model, but the post-processing algorithm adopted would make the prediction box repeatedly calculate in video detection, resulting in frame loss. All of the above improved algorithms have reduced the error-detection rate, but there will be some problems. There is a bottleneck in the computing power of the algorithm application platform. The network of the above improved model is too deep, the parameter calculation amount is large, and it is difficult to deploy the platform in the later stage. The recognition efficiency of the above model is still possible to improve. Therefore, the researchers optimized and modified the network model continuously, and proposed some lightweight network models successively such as MobileNetV2 [13], ShuffleNetV2 [14] and EfficientNetV2 [15], which optimized the recognition efficiency of the network model and reduced the size of the model.

In view of the above problems of real time of safety helmet wearing detection, the YOLOv5 network model is improved on the premise of reducing the calculation parameters and maintaining the accuracy of the model. Firstly, a lightweight feature extraction unit is constructed, and the bottleneck module (C3) is replaced by the basic module of ShuffleNetV2 to reduce the parameters of the network model and speed up the target feature extraction. Only a weight model with a size of 9.8 MB can make the mean average precision (mAP) recognized by the safety helmet reach 92.26%.

The specific structure of each chapter of the thesis is as follows.

1) The first chapter elaborates on the research background, current research status, and existing problems of helmet wearing detection.
2) Chapter 2 introduces the basic model of the algorithm.
3) Chapter 3 introduces the lightweight network model and makes improvements to the network, and provides a detailed description.
4) Chapter 4 introduces the experimental environment and evaluation indicators.
5) Chapter 5 presents the experimental results and compares and analyzes them.
6) Chapter 6 provides a summary of the entire article and prospects for the issues that need to be addressed.

2. **Algorithm Principle.** Considering that excellent general-purpose one-stage detection algorithms have been continuously proposed in recent years, the detection accuracy has gradually caught up with the two-stage detection accuracy. The one-stage network can be easily trained end-to-end, and is more widely used in industry and has a certain industrial deployment foundation. The YOLOv5 algorithm is the fifth-generation improved version of the YOLO algorithm series, which has a wide range of applications in the field of target detection. The network achieved the best trade-off between speed and accuracy at the time, and it is also one of the preferred algorithms for object detection in the industry. Therefore, the paper uses the YOLOv5s network with the smallest width and depth as the base model for improvement. We can see this in Figure 1 that the YOLOv5

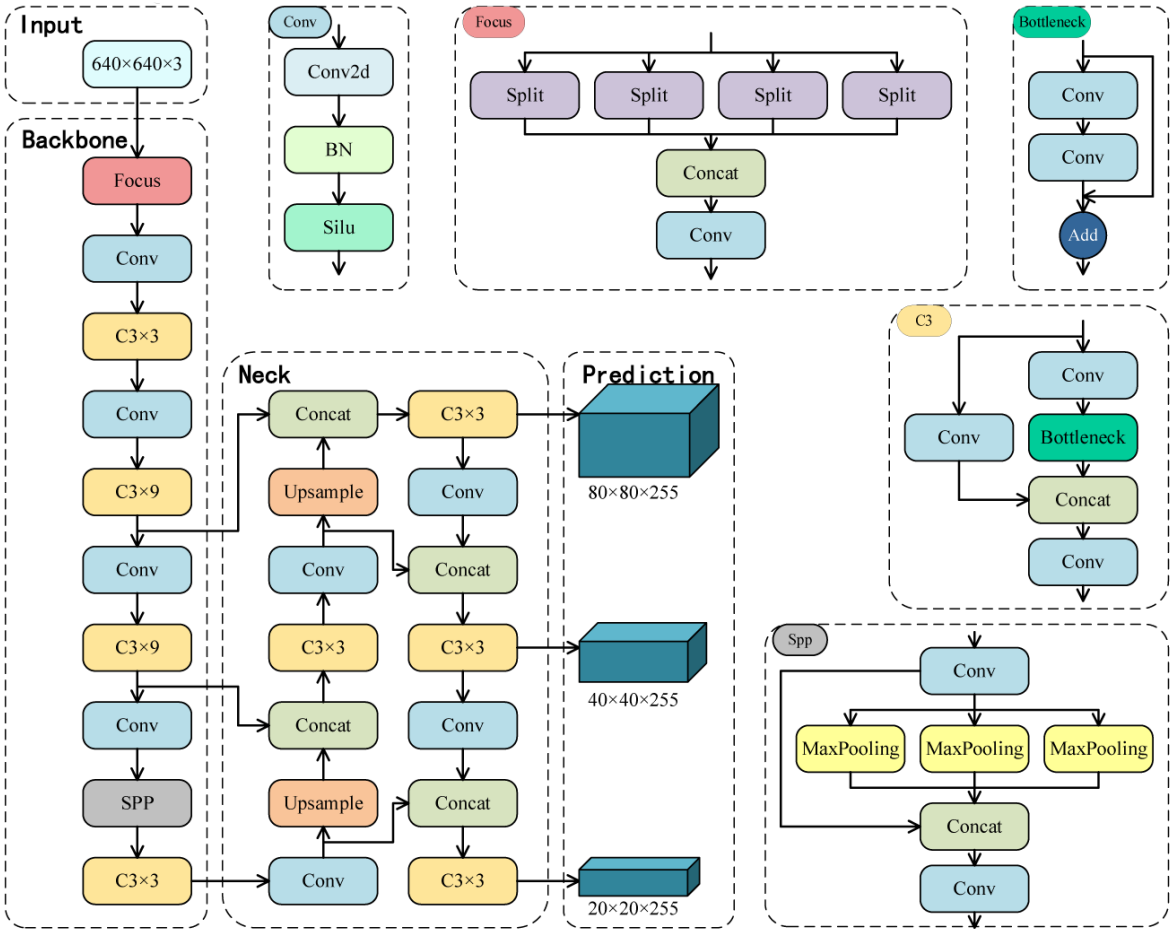

FIGURE 1. YOLOv5 network structure diagram

network model is constituted of four parts: Input, Backbone, Neck, and Prediction. The main detection process is as follows: the backbone network consists of modules such as the slice module (Focus), convolution module (Conv), bottleneck module (C3) and spatial pyramid pooling (SPP), which is the feature extraction process. Neck network adopts feature fusion mode from top down [16] and from bottom up [17] to express multi-scale features. In the end, the obtained feature maps are passed into the prediction layer. In order to remove the redundant prediction box, it is solved by non-maximum suppression (NMS). This process outputs the predicted target category and confidence score, and returns the position coordinate information of the target box.

3. **Introduction of Lightweight Network Modules.** ShuffleNetV2 is a lightweight network model optimized based on the mobile algorithm platform. Compared with the first version, the basic module of ShuffleNetV2 adopts the operation of random Channel split [18]. This process splits the input channel into two parts randomly. One group passes back directly and maintains its own mapping, and the other group takes computing operations backwards. The computing ways in this part is constituted by two convolution layers and one depthwise separable convolution. The two convolution layers are used 1×1 ordinary convolution operation rather than using group convolution. The amount of input and output channels is kept consistent to speed up the calculation of the network. After processing, the effect of element-wise addition is avoided by cascading the two groups of calculated output channels. Finally, a random channel mixing Channel shuffle operation is performed on the output feature map, so that the information between channels affects each other. The application of Channel split and Channel shuffle in the module reduces

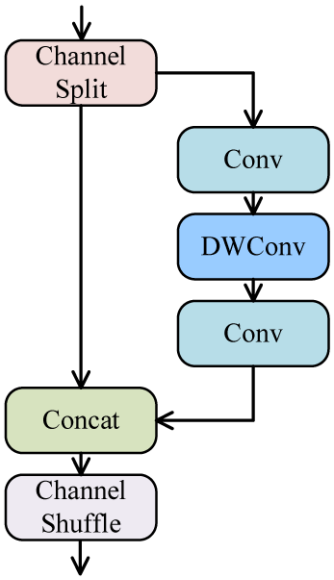

FIGURE 2. Basic unit of ShuffleNetV2

the size of the model, achieving a good effect on the improvement of detection efficiency. Figure 2 shows the structure of this part.

There are four bottleneck C3 modules in the backbone network of YOLOv5, which contain multiple convolutional layers. Although the convolution operation is the basic operation of feature extraction, the convolution kernel contains a large number of parameters, which caused the number of parameters increase dramatically. As a result, the C3 bottleneck module should be optimized. The C3 module in the original backbone network is removed and replaced by the ShuffleNetV2 basic module named ShuffleBlock. Thus, the number of parameters in the module can be reduced to solve the boundary effect in the channel and the insufficiency of deep feature extraction in the image. In order to make up for the limitation of ShuffleBlock, three ShuffleBlock modules are used for feature extraction in layer 2 and layer 9 of the backbone network. In the middle layers 4 and 6, seven ShuffleBlock modules are used for feature extraction.

The safety helmet detection algorithm not only needs to accurately identify the head target wearing the safety helmet and not wearing the safety helmet in the complex construction environment, but also needs to simplify the model calculation as much as possible to facilitate the deployment of the algorithm on the application platform. Therefore, the backbone network of the YOLOv5 network model is optimized and improved, and the number of network parameters and the size of model are reduced on the premise of constant accuracy. The final structure is shown in Figure 3.

4. **Experiment Setting.** Kaggle is a data science competition platform. The safety helmet data set is open to the public. It has a total of 8,465 pictures of workers wearing safety helmet, including indoor, outdoor and other construction scenes. At the same time, the head states of the workers in the dataset are divided into two types of targets for labeling: wearing safety helmet and not wearing safety helmet.

The experiment builds a deep learning platform on the Windows10 64-bit operating system. The main configuration is Intel(R) Core (TM) i9-10900X 3.70GHz CPU. The graphics card model is NVIDIA GeForce RTX 3090 and the video memory is 24 GB. The memory size is 64 GB. PyCharm is adopted as the integrated development environment, and the deep learning framework is written in Python language. The experiment adopts the stochastic gradient descent method for training.

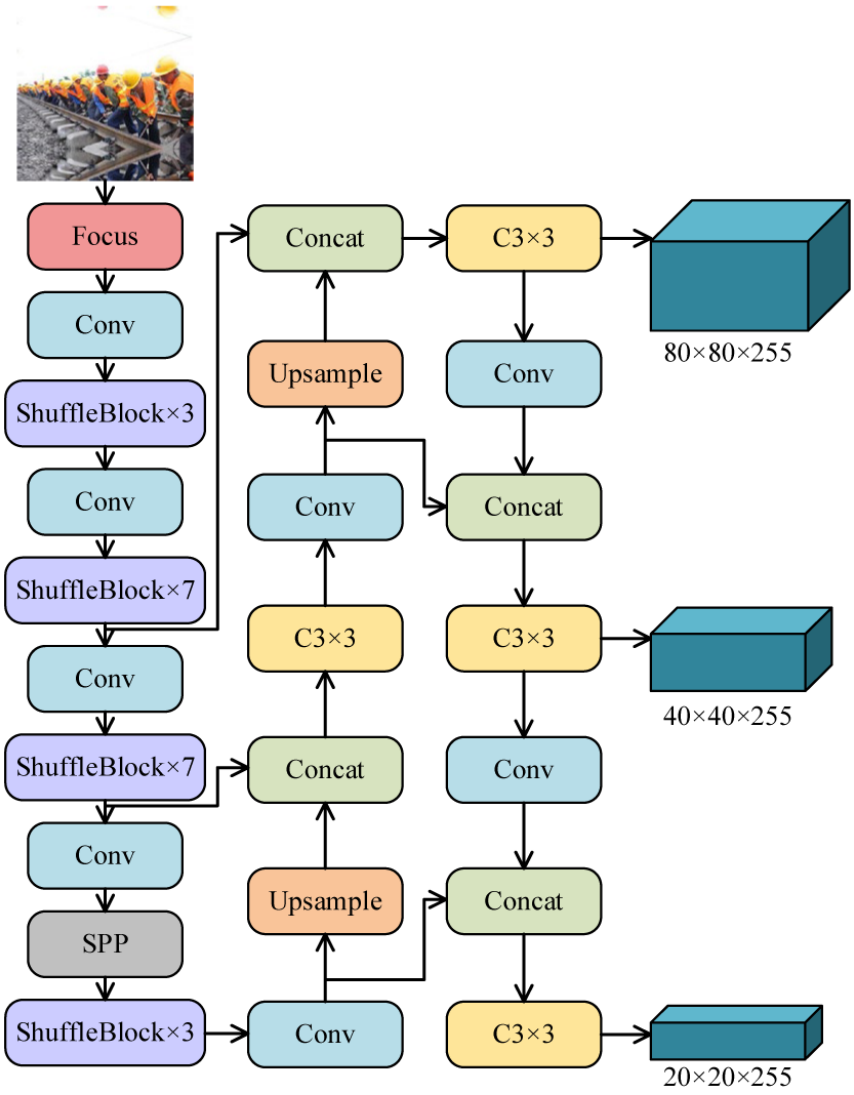

FIGURE 3. Schematic diagram of the improved YOLOv5 algorithm

In order to ensure that the model accuracy is not lost, while speeding up the model detection speed and reducing the weight of the model, taking this as the standard, mAP is used as the evaluation index of the model accuracy. The specific expression is as follows:

$$P = \frac{TP}{FP + TP} \tag{1}$$

$$R = \frac{TP}{FN + TP} \tag{2}$$

$$AP = \int_0^1 P(r)dr \tag{3}$$

$$mAP = \frac{1}{c}\sum_{j=1}^{c} AP_j \tag{4}$$

P is the precision rate. R is the recall rate. TP is the number of true positive targets, FP is the number of false positive targets, and FN is the number of false negative targets. In Formula (3) AP represents the average precision rate, which can combine the precision rate and the recall rate to analyze. In Formula (4) mAP represents the average value of each target category AP, which can reflect the recognition ability of the model. Finally, FPS is used to assess the recognition speed of the algorithm, that is the number of pictures

processed per second. Ws (weight size) is used as a measure of the weight file size after model training.

5. **Analysis of Experimental Results.** The original YOLOv5s model and the improved model were used to conduct comparative experiments. The experimental results were divided into three situations: crowded scenario, complex environment scenario, and multi-scale target scenario. As shown in Figures 4-6, the obtained results are distinguished by marking the target name and confidence score above the regression prediction box, and displaying them in different colors. From the experimental results, it can be concluded that the improved network model can detect two types of targets with safety helmets and heads without safety helmets correctly in different construction environments, but the original YOLOv5s model will have false and missed recognition. At the same time, the improved algorithm's prediction confidence score of the target is mostly higher than the original YOLOv5s model, and the prediction box can accurately lock the object we want to identify. Experimental result demonstrates that the improved algorithm has better detection effect on the target in more complex construction scenes. It has better feature extraction capacity and can accurately locate the target.

(a) YOLOv5s (b) Ours

FIGURE 4. Crowded scenario

In order to verify the application effect of the ShuffleNetV2 basic module on the YOLOv5 network, the experiments were carried out to analyze the ablation experiments with the original YOLOv5s. Firstly, the basic YOLOv5s model is trained, then the C3 module is replaced by ShuffleBlock module on the basic model and the same setting is trained. Using the test set to predict multiple versions of the model, the IoU threshold is set to 0.5, and the results are shown in Table 1. It can be concluded from Table 1 that while the functional modules are continuously replaced and optimized, the speed of the detection model is gradually improved, reaching 81.2 frames per second, the number of parameters is reduced to 4.903×10^6, and the weight size is further compressed to 9.8 MB.

TABLE 1. Ablation comparison experiment

Model	mAP@0.5/%	FPS/(frame·s^{-1})	Params/10^6	Weight size/MB
YOLOv5s	92.23	63.1	7.057	14.1
YOLOv5s+ShuffleBlock	92.26	81.2	4.903	9.8

(a) YOLOv5s (b) Ours

FIGURE 5. Complex environment scenario

(a) YOLOv5s (b) Ours

FIGURE 6. Multi-scale target scenario

The mean accuracy of the improved network model remains essentially unchanged compared with the original YOLOv5s, the detection speed is increased from 63.1 frames per second to 81.2 frames per second, and the weight file size of the model is compressed from 14.1 MB to 9.8 MB, which optimizes the detection speed and model size effectively, which increases the operating efficiency of the improved network model greatly under the PC terminal.

The improved model is compared with the current lightweight target detection model with better performance for experimental analysis. PeleeNet [19] and Tiny SSD [20] are trained separately on the premise of keeping the training set unchanged, as shown in Table 2. Comparing the improved model with the two trained models in terms of accuracy, the mean accuracy of the improved model on the safety helmet target is 5.51 and 13.64 percentage points higher than that of PeleeNet and Tiny SSD respectively, which promotes the feature extraction capacity and accurate target location regression. From the perspective of real-time detection, due to the lightweight design of the backbone network in the improved model, the computational complexity of the network is reduced, and the inference speed is higher than that of PeleeNet and Tiny SSD, respectively. That is an improvement of 14.8 fps and 27.7 fps on both models. In terms of network size, PeleeNet

Table 2. Comparison with different safety helmet detection algorithms

Algorithm	AP/%(IOU = 0.5)		mAP/%	FPS/(frame·s^{-1})	Weight size /MB
	Wearing safety helmet	Not wearing safety helmet			
PeleeNet	89.85	83.65	86.75	66.4	23.7
Tiny SSD	80.73	76.51	78.62	53.5	5.3
YOLOv5s	96.19	88.28	92.23	63.1	14.1
Ours	97.09	88.30	92.26	81.2	9.8

is 23.7 MB, while Tiny SSD is 5.3 MB. The weight size of the improved model is 9.8 MB, which is 4.5 MB larger than Tiny SSD model. However, the accuracy is basically unchanged, keeping a superior detection accuracy. The final analysis shows that the improved lightweight YOLOv5 network model has obvious advantages over PeleeNet and Tiny SSD in terms of average precision, detection speed and network storage space.

6. **Conclusions.** A lightweight YOLOv5 target detection network is improved to solve the problem of a large number of computational parameters and high processing capacity of equipment in deep neural network model for safety helmet wearing detection. Firstly, without losing accuracy, ShuffleNetV2's basic module is used to perform lightweight operations on the network to simplify calculation parameters. The network is optimized to drop the inference time on the CPU and increase the recognition efficiency of the safety helmet target.

At present, the experiment only improves the detection speed, parameter quantity, and model size, and the detection accuracy will be considered later. In the future work, the problem of easy missed detection of small-scale targets under wide field of view will be studied. The network is optimized to find a better network model for the detection of worker helmet objects.

REFERENCES

[1] M. Sadiq, S. Masood and O. Pal, FD-YOLOv5: A fuzzy image enhancement based robust object detection model for safety helmet detection, *International Journal of Fuzzy Systems*, vol.24, pp.2600-2616, 2022.

[2] S. Kumar, H. Gupta, D. Yadav et al., YOLOv4 algorithm for the real-time detection of fire and personal protective equipments at construction sites, *Multimedia Tools and Applications*, vol.81, pp.22163-22183, 2022.

[3] S. Ren, K. He, R. Girshick et al., Faster R-CNN: Towards real-time object detection with region proposal networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.39, no.6, pp.1137-1149, 2017.

[4] K. He, G. Gkioxari, P. Dollár et al., Mask R-CNN, *Proc. of the IEEE International Conference on Computer Vision*, pp.2961-2969, 2017.

[5] W. Liu, D. Anguelov, D. Erhan et al., SSD: Single shot multibox detector, *European Conference on Computer Vision*, pp.21-37, 2016.

[6] J. Redmon, S. Divvala, R. Girshick et al., You only look once: Unified, real-time object detection, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.779-788, 2016.

[7] T.-Y. Lin, P. Goyal, R. Girshick et al., Focal loss for dense object detection, *Proc. of the IEEE International Conference on Computer Vision*, pp.2980-2988, 2017.

[8] S. Chen, W. Tang, T. Ji et al., Detection of safety helmet wearing based on improved Faster R-CNN, *2020 International Joint Conference on Neural Networks (IJCNN)*, pp.1-7, 2020.

[9] Q. Ge, Z. Zhang, L. Yuan et al., Fusion of environmental features and improved safety helmet wearing detection in YOLOv4, *Chinese Journal of Image and Graphics*, vol.26, no.12, pp.2904-2917, 2021.

[10] R. Zhao, H. Liu, P. Liu et al., Safety helmet detection algorithm based on improved YOLOv5s, *Journal of Beijing University of Aeronautics and Astronautics*, pp.1-16, 2021.

[11] J. Hu, L. Shen and G. Sun, Squeeze-and-excitation networks, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.7132-7141, 2018.

[12] Z. Li, W. Xie, L. Zhang et al., Toward efficient safety helmet detection based on YOLOv5 with hierarchical positive sample selection and box density filtering, *IEEE Transactions on Instrumentation and Measurement*, vol.71, pp.1-14, 2022.

[13] M. Sandler, A. Howard, M. Zhu et al., MobileNetV2: Inverted residuals and linear bottlenecks, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.4510-4520, 2018.

[14] N. Ma, X. Zhang, H.-T. Zheng et al., ShuffleNetV2: Practical guidelines for efficient CNN architecture design, *European Conference on Computer Vision*, pp.122-138, 2018.

[15] M. Tan and Q. Le, EfficientNetV2: Smaller models and faster training, *International Conference on Machine Learning*, pp.10096-10106, 2021.

[16] T.-Y. Lin, P. Dollár, R. Girshick et al., Feature pyramid networks for object detection, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.2117-2125, 2017.

[17] S. Liu, L. Qi, H. Qin et al., Path aggregation network for instance segmentation, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.8759-8768, 2018.

[18] X. Zhang, X. Zhou, M. Lin et al., ShuffleNet: An extremely efficient convolutional neural network for mobile devices, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.6848-6856, 2018.

[19] R. J. Wang, X. Li and C. X. Ling, Pelee: A real-time object detection system on mobile devices, *Proc. of the 32nd International Conference on Neural Information Processing Systems*, pp.1967-1976, 2018.

[20] A. Womg, M. J. Shafiee, F. Li et al., Tiny SSD: A tiny single-shot detection deep convolutional neural network for real-time embedded object detection, *2018 15th Conference on Computer and Robot Vision (CRV)*, pp.95-101, 2018.