Yolov5-based channel pruning is used for real-time detection of construction workers’ safety helmets and anti-slip shoes in informationalized construction sites

Xuehui SHEN1, Baojiang LI1,*, Xichao WANG1, Guochu CHEN1, Junquan LU2, Bin LIN3, Jianming ZHENG4

1School of Electrical Engineering, Shanghai Dianji University, Shanghai, 201306, China
2China Tobacco Hebei Industrial Co., Ltd, Shijiazhuang 050000, China
3Zhangjiakou Cigarette Factory Co., Ltd. Zhangjiakou 075000, China
4Hebei Tuoying Intelligent Technology Co., Ltd. Baoding, Hebei 071000, China

*Corresponding author’s e-mail: libj@sdju.edu.cn

Abstract: In view of the problems that the application of target detection model in edge equipment occupied too much memory and the operation was slow, a lightweight model based on YoloV5 is proposed to detect the wearing condition of safety helmet and anti-slip shoes. The channel pruning is carried out on the basis of the original model, and the channel with low weight is deleted to reduce the parameters of the model and improve the detection speed. The experimental results show that the recognition accuracy of the network after channel pruning is 93.2%, which is basically the same as that before channel pruning. Meanwhile, the number of parameters is reduced by 63.9%, and the detection speed is 123.45fps. It provides technical support for the development of embedded detection equipment.

1. Introduction
The production safety of enterprises is crucial, and every year there will be tragedies caused by the failure to wear the corresponding safety protective equipment [1]. In the cylinder driving operation of a tobacco factory in China, workers are usually required to wear safety helmets and anti-slip shoes before the next construction operation. Therefore, workers' wearing, including safety helmets and anti-slip shoes, need to be detected before the construction. At present, the wear detection before construction mainly relies on manual detection, and people need to be at the construction site all the time, which consumes a lot of human resources, which requires a machine that can replace manual detection to ensure the safety of construction personnel all the time. The traditional detection based on large servers requires a large amount of investment in the early stage. Meanwhile, the field is not good for the deployment of large equipment, so it relies on edge equipment for detection. The traditional large models run slowly in embedded devices and cannot meet the real-time requirements, so the model needs to be deleted.

Yolo has certain advantages in some target detection fields [2-4], but there is still room for YOLO to be optimized for use on less powerful devices, such as edge computing devices[5-6].
2. Data and methods

2.1. Data
The data was collected from a cigarette factory, the video was shot by IphoneXR, and then the video was captured as a picture. The fuzzy and incomplete pictures were deleted. A total of 11567 pictures were selected. The specific distribution of the data set is shown in Table 1, and the wearing example is shown in Figure 1.

| Data types                  | The serial number | Categories                                      | Number of training sets |
|-----------------------------|-------------------|-------------------------------------------------|-------------------------|
| Construction personnel      | 0                 | Without safety helmets and without anti-slip shoes | 3050                    |
| wear conditions             | 1                 | Without safety helmets and with anti-slip shoes  | 3093                    |
|                             | 2                 | With safety helmets and without anti-slip shoes  | 2791                    |
|                             | 3                 | With safety helmets and with anti-slip shoes     | 2633                    |
| Total                       |                   |                                                 | 11567                   |

Figure 1. Safety helmets and anti-slip shoes are worn, from left to right are with safety helmets and with anti-slip shoes, without safety helmets and with anti-slip shoes, with safety helmets and with anti-slip shoes, with safety helmets and without anti-slip shoes.

2.2. Using channel pruning-based Yolov5 deep learning algorithm for the real-time and accurate detections of safety helmets and anti-slip shoes

2.2.1. Channel-pruning based on Yolov5s
The essence of the channel pruning algorithm is to eliminate the connection points with low weight by identifying the channel of the network, and disconnect the input and output connected to the connection points [7]. Compared with the layer pruning algorithm, this algorithm reduces the number of parameters to be stored and has lower hardware requirements, so it is easy to be deployed on embedded devices, mobile terminals and other small computing platforms[8-9]. Therefore, in this study, a balance was made between flexibility and implementation cost. A channel pruning algorithm was used to prune the Yolov5s safety helmets and anti-slip shoes detection network after training.
In a Convolution-Bn-Activation module, the Bn layer can achieve channel scaling. The specific operation of the BN layer has two parts: as shown in Eqs. (1)-(2). After batch-normalization, a linear transformation will be carried out by evaluating the weight of the coefficient $\gamma$. When the weight of the coefficient $\gamma$ is very small, the corresponding activation ($Z_{out}$) will be correspondingly small. These outputs with small responses can be pruned out, thus achieving channel pruning at the BN layer.

$$Z = \frac{Z_{in} - \mu_B}{\sqrt{\sigma_B^2 + \theta}}$$  

(1)

$$Z_{out} = \gamma \hat{z} + \beta$$

(2)

Where $\mu_B$ and $\sigma_B$ are the mean and standard deviation values of input activations over $B$, $\gamma$ and $\beta$ are trainable affine transformation parameters (scale and shift) which provides the possibility of linearly transforming normalized activations back to any scales.

Figure 2. Schematic diagram of channel pruning algorithm

The channel pruning of Yolov5 consists of the following steps:

1. Sparse training. L1 regularization is used to train the network sparsity on the BN layer of the safety helmets and anti-slip shoes detection network, and the parameters with less weight in the model are found out to make the network structure sparse.

2. Channel pruning. A network with a smaller width can be obtained by cutting off the input and output relations related to the weighted channel at a certain pruning rate. The network will automatically identify and remove the unimportant channels, which will hardly affect the generalization ability of the network.

3. Fine tuning of the model after channel pruning. After channel pruning, there may be a large loss of precision, so the precision of the model can be restored by fine-tuning the model.

During the whole process, the objective function is shown in Eq. (3)

$$L = \sum_{(x,y)} l(f(x, W), y) + \lambda \sum_{\gamma \in \Gamma} g(\gamma)$$

(3)

Where the first term was the training loss of the network; the second term is the L1 regular constraint term of the $\gamma$ coefficient of the BN layer; $x$ and $y$ are the input and output of the training, respectively; $W$ is the training parameter in the network; $\lambda$ is the penalty factor.

Parameter Settings at each stage of channel pruning are shown in Table 2.

| Parameter  | Value  |
|------------|--------|
| Sparse training | Batch-size | 26 |
| Sparse rate  | 0.001   |
Table 1. Channel pruning and parameter tuning

| Parameter       | Value   |
|-----------------|---------|
| Epochs          | 30      |
| Pruning rate    | 0.5     |
| Batch-size      | 26      |
| Epochs          | 100     |

Figure 3 shows the sparsity training process. As the sparsity training progresses, $\gamma$ gradually approaches 0, but does not reach 0, indicating that $\gamma$ has gradually become sparse. Starting from the 14th epoch, the change in $\gamma$ is not so obvious, indicating that the sparsity training can be stopped. An attempt was made to train 300 epochs to make $\gamma$ closer to 0, but there was no difference compared with the 14th epoch.

Figure 3 shows the BN layer $\gamma$ coefficient during sparse training.

After sparse training and channel pruning, the channel number distribution is shown in Figure 4. A total of 4496 channels are prunned, and the prune ratio is 46.7%. As you can see from Figure 4, the number of channels in most layers is significantly reduced. Table 3 shows the changes in the number and accuracy of parameters before and after channel pruning. Before and after channel pruning, the model size was reduced by nearly 63.9%, but the accuracy remained basically unchanged. Summing up the above two characteristics, it is shown that channel pruning is effective.
3. Results and Analysis

3.1. Evaluation of model performance

In order to evaluate the performance of the model, 3400 images are tested. In this study, four indexes of accuracy precision (P), recall (R), mean Average Precision (mAP) and detection speed (FPS) were selected for evaluation. Their calculation formulas are shown in Eqs. (4) - (6).

\[
\text{precision} = \frac{TP}{TP + FP} \times 100\% \tag{4}
\]

\[
\text{recall} = \frac{TP}{TP + FN} \times 100\% \tag{5}
\]

\[
mAP = \frac{\sum_{c=1}^{c} AP(c)}{c} \times 100\% \tag{6}
\]

Where TP, FP and FN are the numbers of true positive cases, false positive cases and false negative cases, respectively; and c is the number of detection categories. There are four wearing situations in this study, so c=4.

A total of 3403 images were selected for the test, including the same number of different wearables. Test results as shown in Table 4, indicate that channel construction Yolov5 is of high precision, detection
speed also has more than 100 frames at the same time, the in the embedded devices, small system is
good, at the same time for the application in embedded small platform provides technical reference.

Table 4. Detection results of safety helmets and anti-slip shoes by the proposed method

| Evaluation index | Precision (%) | Recall(%) | mAP(%) | Detection speed (fps) |
|------------------|--------------|----------|--------|----------------------|
| Test results     | 99.2         | 90       | 93.2   | 123.45               |

3.2. Comparison of different object detection algorithms
At present, in addition to the Yolo target detection network, there are other target detection networks
such as Faster-RCNN[10], RetinaNet[11],Yolov4-tiny[12] and test results are compared with them. As
shown in Table 5, compared with other algorithms, the mAP value of Yolov5s after channel pruning
decreased by less than 1%, but the model of network model decreased by at least 4.5 times. Both Faster-
RCNN and RetinaNet need to reach several hundred MB and obviously cannot run on embedded devices
and small mobile devices and cannot meet the real-time requirements. Yolov4-tiny can basically meet
the real-time requirements, but the mAP is 0.06% lower than that of Yolov5s after channel pruning. As
a result, Yolov5s for removing channels is up to 37 times faster than other networks without
compromising accuracy. The results show that channel pruning of Yolov5s can effectively detect
whether the operator is wearing a helmet and anti-slip shoes.

Table 5. Detection results of safety helmets and anti-slip shoes by different object detection algorithms

| Algorithms       | mAP (%) | Model-size (MB) | Detection speed (fps) |
|------------------|---------|-----------------|----------------------|
| Faster-RCNN      | 93.27   | 330.3           | 3.33                 |
| RetinaNet        | 93.29   | 442.9           | 2.37                 |
| Yolov4-Tiny      | 92.94   | 23.5            | 65.12                |
| Ours             | 93.2    | 5.2             | 123.45               |

4. Conclusions
In this study, the channel pruning method is used to reduce the hardware requirements for model
operation and reduce the cost for large-scale application. The principle of channel pruning is to judge
the weight of the connection points in the network model of the neural network through the weight
coefficient. The connection points with small weight coefficient are disconnected from the
Corresponding input and output, and these connection points have little influence on the final accuracy
of the model and will not affect the final detection result. After channel pruning, a total of 4496 channels
with channel weight coefficient less than 0.5 are deleted, reducing the channel of Yolov5s by 63.9%.
After constant parameter fine-tuning, the operating speed reaches 123.45fps and mAP reaches 93.2%,
which provides technical support for the development of embedded monitoring robot. Compared with
the Faster-RCNN, RetinaNet and Yolov4-tiny algorithms, the model size is reduced by 73 times and the
mAP is reduced by less than 1%. The running speed is 1.89 times faster than yolov4-Tiny, which meets
the real-time and accuracy requirements. In future research, the efficiency of the channel pruning
algorithm will be further studied to improve the accuracy of detection.

Acknowledgements:
The authors would like to thank China Tobacco Hebei Industrial Co., Ltd for providing people’s wearing
data, and pay special thanks to Qiongqiong LI, Zhaoran WANG et al. as their work provides useful
information to this paper.
References
[1] Hoła, A.; Sawicki, M.; Szóstak, M. Methodology of Classifying the Causes of Occupational Accidents Involving Construction Scaffolding Using Pareto-Lorenz Analysis. Appl. Sci. 2018, 8, 48. https://doi.org/10.3390/app8010048
[2] Li M, Zhang Z, Lei L, et al. Agricultural Greenhouses Detection in High-Resolution Satellite Images Based on Convolutional Neural Networks: Comparison of Faster R-CNN, Yolo v3 and SSD[J]. Sensors, 2020, 20(17).
[3] Morera, Á.; Sánchez, Á.; Moreno, A.B.; Sappa, Á.D.; Vélez, J.F. SSD vs. Yolo for Detection of Outdoor Urban Advertising Panels under Multiple Variabilities. Sensors 2020, 20, 4587. https://doi.org/10.3390/s20164587.
[4] Morera N, N Sánchez, Moreno A B , et al. SSD vs. Yolo for Detection of Outdoor Urban Advertising Panels under Multiple Variabilities[J]. Sensors (Basel, Switzerland), 20(16).
[5] Zhang X, He Y, Jian S . Channel Pruning for Accelerating Very Deep Neural Networks[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2017.
[6] Rueda F M, Grzeszick R, Fink G A. Neuron Pruning for Compressing Deep Networks using Maxout Architectures[J]. Springer, Cham, 2017.
[7] P. Molchanov, A. Mallya, S. Tyree, I. Frosio and J. Kautz, "Importance Estimation for Neural Network Pruning," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 11256-11264, doi: 10.1109/CVPR.2019.01152.
[8] Z. Liu, J. Li, Z. Shen, G. Huang, S. Yan and C. Zhang, "Learning Efficient Convolutional Networks through Network Slimming," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2755-2763, doi: 10.1109/ICCV.2017.298.
[9] Kong, Y.H., Zhu, C.C., Che, L.L., and D. Wu, et al. Flower recognition in complex background Computers and Electronics in Agriculture model pruning based on MobileNets. Sci. Technol. Eng. 2018, 178(2020)10574211, 18 (19), 84–88. https://doi.org/10.3969/j.issn.1671-1815.2018.19.013.
[10] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6):1137-1149.
[11] Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP (99):2999-3007.
[12] Bochkovskiy A, Wang C Y, Liao H. Yolov4: Optimal Speed and Accuracy of Object Detection[J]. 2020.