Research and application of small target detection method

Zhenheng Wang¹, Huajun Wang¹*, Jun Wan¹, Yu Ma¹, Cong Huang¹, Yao Yang¹
¹ Chengdu University of Technology, Chengdu, Sichuan, 610000, China
*Corresponding author’s e-mail: wangzhenheng@stu.cdut.edu.cn

Abstract: With the advent of the big data era, more and more data are used in image detection. In order to solve the problem of small target detection in traditional image detection algorithms, this paper proposes an improved target detection algorithm based on YOLO-v5. The algorithm proposes to connect two models in series, using the former model for target pre-positioning, and the latter model for target detection. This method trains two different model parameters, tests them separately, and finally connects them for target detection. The test results show that the improved algorithm improves the ability in small target detection, and at the same time has a speed advantage.

1. Introduction
With the rapid development of artificial intelligence technology, artificial intelligence technology is more and more used in industrial production. Facing the increasing demand, more and more production safety problems also appear. Especially in production occasions, some workers' illegal operation may produce some unexpected potential safety hazards. Workers are five times more likely to be injured at work than other workers. Therefore, in order to solve this problem, the combination of artificial intelligence technology and on-site construction image can scan the potential safety hazards of workers, such as workers who do not wear protective equipment, or whether there are operation violations, and send the image into the reasoning model and make judgment. Putting artificial intelligence technology into use can save human resources, monitor workers' operation at any time and judge whether the operation meets the specifications, which greatly reduces the potential safety hazards caused by workers' non-standard operation. Therefore, the development of artificial intelligence technology can promote the production and development of other industries at the same time, and bring us continuous supplement and improvement.

2. Related work
This paper mainly introduces the basic principle of image processing through deep learning, and applies it to detect the smoking problem of oil workers. Because the cigarette account for a very small proportion in the whole picture, and the cigarette photographed by the camera are blurred, it is very difficult for workers to detect smoking. Because the previous image detection algorithms often have the problems of difficult detection and large error in the detection of small objects. In this paper, an image detection algorithm based on yolo-v5 is proposed. In this method, two network models are connected in series to obtain a new network model structure, which solves the problem that small objects are difficult to locate and detect, and improves the prediction accuracy.
3. The method

3.1. Basic principles of deep learning

Deep learning is a branch of machine learning method, which is based on artificial neural network with representational learning. Learning can be supervised, semi supervised or unsupervised. Deep learning architectures include deep neural networks\(^1\), deep belief networks\(^2\), deep reinforcement learning\(^3\), recursive neural networks and convolutional neural networks. These networks have been widely used in computer vision, speech recognition, natural language processing, machine translation, bioinformatics, drug design, medical image analysis, material inspection and board game programs. In these practical applications, the results obtained through continuous machine learning can be equal to experts, and even surpass experts in some aspects. In engineering construction, the monitoring of workers' code of conduct is often a difficult problem, which often leads to the consequence of weak supervision due to insufficient manpower. Therefore, the method of deep learning image processing can effectively improve the supervision and ensure the life safety of workers.

3.2. Principle of yolov5 algorithm

YOLO (you only look once) is a high-performance general target detection model. Yolov1\(^4\) creatively completed the two tasks of classification and target positioning using the first-order structure. The subsequent improvement in speed and accuracy of yolov2\(^5\) and yolov3\(^6\) further accelerated the landing of object detection in the industry, Yolov4\(^7\) can complete training on an ordinary GPU (1080ti). From yolov1 to now, yolov5 series has been developed to yolov5. Compared with yolov4, yolov5 is more flexible. To some extent, yolov5 model is the SOTA (state of the art) in all known Yolo implementations. It provides four versions, namely yolov5s, yolov5m, yolov5l and yolov5x. The size and accuracy of the four models increase in turn, distinguished by the number of bottlenecks, Channel and layer control factors similar to efficientnet\(^8\) are adopted to realize version change, and models of appropriate size can be selected according to application scenarios. This paper mainly realizes model compression and acceleration to make it easier to be applied to embedded devices with limited resources. Therefore, the selection range of benchmark model is set in yolov5s series.

After absorbing the advantages of the previous version and other networks, YOLOv5 has changed the previous YOLO target detection algorithm's faster detection speed but low accuracy. The YOLOv5 target detection algorithm has improved detection accuracy and real-time performance, satisfying Video image real-time detection needs, and the structure is more compact. Its network model is divided into 4 parts, namely Input (input), Backbone (backbone network), Neck (multi-scale feature fusion module) and Prediction (prediction). Its network structure is shown as in Figure 1.

![YOLOv5 Structure](image.png)

**Figure 1.** Network structure diagram of YOLOv5
YOLOv5 also has an iterative update, which has been updated to v5.0, where v3.0, v4.0, and v5.0 have their own advantages. The structure of YOLOv5-P5 in v4.0 and v5.0 is the same, and v5.0 newly proposes YOLOv5- Compared with the previous version, the P6 model has added 1 output layer, a total of 4 output layers P3, P4, P5, and P6. The corresponding down-sampling amplitudes are 8, 16, 32, and 64 respectively. The newly added inspection layer is conducive to detection large objects can obtain higher accuracy through higher resolution training. In addition, the YOLOv5-P5 model uses the newly supported SiLU() activation function in PyTorch1.7 to replace the LeakyReLU() and Hardswish() activation functions used in the previous version, so that only the SiLU activation function is used anywhere in the network, and Some Conv modules of BottleneckCSP in the previous version have been deleted. The previous version of BottleneckCSP and the improved BottleneckCSP changes in YOLOv5-P5 are shown in Figure 2. The improved BottleneckCSP is called the C3 module.

![Figure 2 BottleneckCSP and C3 module](image)

4. Experiment and analysis
This experiment is carried out under Windows system. NVIDIA GeForce GTX 2060 graphics card, Intel Core i7-10875h 8 core 16 thread CPU processor and 32g memory are adopted. The deep learning framework is pytorch-yolov5.

4.1. Data preprocessing
The data set used in this experiment is smoking pictures taken from cameras in different scenes. Firstly, the data set is divided into training set, test set and verification set, and then the data is marked, and the hands and heads of all people in the figure are marked.

The training steps of yolo-v5 network are as follows: 1) convert the labeled JSON label into TXT format label.2) The obtained TXT format label is divided into two parts. A part of the pictures and labels are integrated as the training set of the network model. In the second part, the remaining files are sorted together as the verification set of the model training.3) Finally, put the sorted data into the project and write the path into the yaml file to start training the network model.

4.2. Model training and results.
In this paper, the yolo-v5 framework is used to construct the prediction model, and the approximate position of cigarette is determined by prepositioning. Then, a combinatorial model is formed by using the idea of integrated series. In general, the first mock exam accuracy of a series structure is better than that of a single model. Finally, the training data are used to detect the smoking behavior in the test set, and show the results of smoking detection.

4.2.1. Model training. Figure 3 and Figure 4 show the training results of the head and hand detection model and the cigarette detection model.
4.2.2. Detection results. The smoking behavior detection results obtained by combining the two models are shown in Figure 5.

Figure 3 Results of head and hand detection model

Figure 4 Results of cigarette detection model

Figure 5 Results of cigarette detection model
As can be seen from Figure 5, although the cigarette is very small in the figure, the position of the cigarette can still be detected because of the prepositioning operation.

5. Conclusion
In the paper, we first introduce the research background of artificial intelligence technology in image processing, points out that image processing technology can solve industrial problems, then analyzes the basic framework and network structure of yolo-v5, and applies it to cigarette detection. This paper proposes a cascade structure of YOLO-v5 models, which can be used to detect small targets in images. Through the analysis of the experimental results: when the mAP of the head and hand detection model reaches more than 90% and the mAP of the cigarette end detection model reaches more than 80%, the cigarette can be successfully detected by the combined models. The experimental results show that the model series structure proposed in this paper has better detection results and obvious advantages in small target detection. The application of improved YOLO-v5 model to cigarette detection proves that image detection can be applied to the detection of small targets. The follow-up work will solve the problem of how to improve the target detection accuracy of the algorithm, verify and improve the model in practical applications.

References
[1] LIU Z, GAO J, YANG G, et al. Localization and classification of paddy field pests using a saliency map and deep convolutional neural network[J]. Scientific Reports, 2016, 6: 20410.
[2] Gao N, Gao L, Gao Q L, et al. An intrusion detection model based on deep belief networks // 2014 2nd International Conference on Advanced Cloud and Big Data. Huangshan, China: IEEE, 2014: 247-252.
[3] LI H, WEI T, REN A, et al. Deep reinforcement learning: framework, applications, and embedded implementations[C]. IEEE International Conference on Computer-Aided Design, 2017: 847-854.
[4] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 779-788.
[5] Redmon J, Farhadi A. YOL09000: better, faster, stronger[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017: 7263-7271.
[6] Redmon J, Farhadi A. YOLOv3: An incremental improvement[J/OL]. arXiv: 1804.02767[2018-4-9]. https://arxiv.org/abs/1804.02767
[7] Bochkovskiy A, Wang C-Y, Liao H-Y M. Yolov4: Optimal speed and accuracy of object detection[J/OL]. arXiv: 2004.10934[2020-4-23]. https://arxiv.org/abs/2004.10934.
[8] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]. International Conference on Machine Learning, 2019: 6105-6114.