Journal of Mechanical Design and Testing 5(1), (2023), 35-43
DOI: http://dx.doi.org/10.22146/jmdt.85645
e-ISSN: 2685-8029, ISSN: 2716-1293

Recognition of Safety Helmet Wearing of Operator Local Ground Floor Unit 2 Suralaya PGU Based on Improved YOLOv3*

F. M. Ruaz*

PT. PLN Indonesia Power, Suralaya Power Generation Unit, Suralaya, Indonesia
*E-mail: firlan.ruaz@indonesiapower.co.id

Abstract
PT. PLN Indonesia Power Suralaya Power Generation Unit requires its operators on the local ground floor to wear safety helmets to prevent head injuries while performing their tasks. Observation of workers wearing safety helmet must be ensured by the company to prevent head injury by operators. The Culture Transformation Program of K3 PT PLN Indonesia Power brings a new observation method to detect safety helmet wearing of operator local ground floor using real-time detection. YOLOv3 is an application which can be used to real-time detection helmet wearing of workers using the darknet53 algorithm based on YOLOv1 and YOLOv2 by using data images that have been collected. Based on YOLOv3 model, the improved version of YOLOv3 is proposed to improve accuracy and speed detection of safety helmet wearing by combining multi-scale detection training. Different YOLOv3 versions will be used to compare results of helmet recognition. The improved YOLOv3 show results 96.63% mAP_{50} and 725,711 milliseconds better than the other versions of YOLO to detect helmet safety. The experimental result show that the improved YOLOv3 have satisfying with the detection speed and accuracy of safety helmet wearing detection by operators at PT PLN Indonesia Power Suralaya Power Generation Unit.

Keywords: *YOLO, safety helmet, deep learning, real-time detection, deep residual network.*

Abstrak
PT PLN Indonesia Power Unit Pembangkitan Suralaya mewajibkan operatornya di area local untuk memakai *safety helm* untuk mencegah cedera kepala saat melakukan tugas mereka. Pengamatan pekerja yang memakai helm pengaman harus dipastikan oleh perusahaan untuk mencegah cedera kepala yang disebabkan oleh aktivitas pekerjaan para operator. Program Transformasi Budaya K3 PT PLN Indonesia Power menghadirkan metode observasi baru untuk mendeteksi pemakaian *safety helm* oleh operator lantai dasar local dengan menggunakan deteksi *real-time*. YOLOv3 adalah aplikasi yang dapat digunakan untuk mendeteksi pemakaian *safety helm* oleh pekerja secara *real-time* menggunakan algoritma darknet53 berbasis YOLOv1 dan YOLOv2 dengan menggunakan data citra yang telah dikumpulkan. Berdasarkan model YOLOv3, versi YOLOv3 yang ditingkatkan diusulkan untuk meningkatkan akurasi dan kecepatan deteksi pemakaian helm pengaman dengan menggabungkan pelatihan-pelatihan deteksi multi-skala. Versi YOLOv3 yang berbeda akan digunakan untuk membandingkan hasil pengenalan *safety helm*. Versi YOLOv3 yang ditingkatkan menunjukkan hasil 96,63% mAP50 dan 725.711 milidetik lebih baik daripada YOLO versi lain untuk mendeteksi penggunaan *safety helm*. Hasil percobaan menunjukkan bahwa versi YOLOv3 yang ditingkatkan memuaskan hasil yang diperoleh dengan kecepatan deteksi dan akurasi deteksi pemakaian *safety helm* oleh operator-operator di Unit Pembangkitan Listrik PT PLN Indonesia Power Suralaya.

Kata kunci : YOLO, *safety helm*, *deep learning*, deteksi *real-time*, *deep residual network*.

*Best presenter dalam Science Technology and Management (STeM) MeetUp, 22 November 2022 di Yogyakarta.

1. INTRODUCTION

One of the Culture Transformation Programs of K3 PLN Indonesia Power requires workers to wear protective equipment, including safety helmets. Ground floor's operators at PT PLN Indonesia Power Suralaya Power Generation Unit (PGU) must wear helmets, and their compliance can be monitored through manual or computer-based methods. While manual monitoring can be costly in terms of labours, computer-based monitoring provides real-time detection and requires fewer workers.

Industry 4.0 has revolutionized technology and manufacturing, with machine learning being a key area of development. With the help of GPU-based parallel computing, machine learning with deep neural networks can now be trained faster, enabling new applications such as object detection. There are several methods for object detection using deep learning, including R-CNN, SPP-net, Fast R-CNN, Faster R-CNN, and Mask R-CNN. YOLO (You Only Look Once) is an object detection algorithm that has been successfully developed into YOLOv2 and YOLOv3, with the latter achieving good detection accuracy and speed in various fields, including construction and manufacturing. However, there are still challenges in detecting objects in complex environments such as the ground floor area of the PLN Indonesia Power Suralaya Power Generation Unit. This paper proposes using image data from the writer's camera phone and web image crawler to create a helmet detection dataset. The YOLOv3 model is used with K-means clustering analysis to obtain more accurate dimensional frame information, and multi-scale images are used to train the model on images of different resolutions. After multiple rounds of training, the optimal parameters are obtained, resulting in improved detection rates and accuracy.

Deep learning is one of subjects of machine learning. The object detection of deep learning is divided into two categories. One is the combination of prediction box and convolutional neural network (CNN) classification. A typical CNN is structured with multiple layers: an input layer, a convolutional layer, an active layer, a pooling layer, a fully connected layer and finally, an output layer. The input layer is used to initialize the input image data and make all the available dimensions zero-centered. This layer is also responsible for normalizing the scale of all input data to a range within 0 and 1, which would help in accelerating the speed of converging. This normalization is also helpful in reducing redundancy by whitening the data. Principal Component Analysis (PCA) is done to degrade and decorate the available dimensions of the extracted data while focusing on key dimensions (Krizhevsky, 2012). A region with convolutional neural network feature (R-CNN) was proposed. It includes the extract region, the proposal computer CNN feature, the support vector machine and the bounding boxes regression (Girshick, 2015). But the speed and accuracy of R-CNN is not so good and it requires a fixed-size input image. The method of spatial pyramid pooling (SPP-net) generates a fixed-length representation regardless of image size/scale (He K, 2016). However, the detection speed of SPP-net cannot satisfy requirement of real-time object detection. Then the fast region-based convolutional neural network (Fast R-CNN) was proposed (Girshick, 2015). The Fast R-CNN employs innovations to improve the training and testing speed object detection while it also increases detection accuracy, but the Fast R-CNN is not the really end-to-end. With the faster region-based convolutional neural network (Faster R-CNN) was proposed (Redmon, 2015). The Faster R-CNN also proposes the region proposal network (RPN), and it realized the end-to-end. As an extension of the Faster R-CNN, an approach was proposed which can efficiently detect objects in an image while mask for each instance (Mask R-CNN). Mask R-CNN proposed a segmentation instance (He K, 2016). However, all above these algorithms have difficulties in the real-time object detection. In the field of object detection, the method of using deep learning has become mainstream. Redmon J proposed the YOLO (You Only Look Once) object detection algorithm (Redmon, 2015). Which have been successfully developed into YOLOv2 and YOLOv3. YOLOv2 focused on small object detection, increased the mean accuracy of mAP (mean average precision) by 2%. The latest YOLOv3 further strengthened multi-label classification and network architecture, taking into account both accuracy and detection speed, which has good detection effect in construction and other fields. The latest YOLOv3 development is to further strengthened multi-label classification and network architecture, taking into account both accuracy and detection speed, which has good detection effect in construction and other manufacturer fields. But there are still some deficiencies in detection accuracy in plant environment like ground floor area PLN Indonesia Power Suralaya Power Generation Unit.

This paper uses the image data collected from the writer's camera phone by capture the operator local ground floor activities and the pictures obtained by the web image crawler to create a helmet detection data set. Firstly, based on the YOLOv3 model, performing K-means clustering analysis on the dimensions of the target frame obtains more accurate dimensional frame information, so that the model can obtain more edge information of the target object. Then using multi-scale images during the training process training enables the model to adapt images of different resolutions. Finally, train multiple times obtain the optimal parameters during model training. Experiments show that the YOLOv3 on COCO data set algorithm can improve the detection rate while ensuring the detection accuracy. The prediction of this method is based on the multi-scale feature. YOLOv3 combines the feature pyramid (Dollar, 2014). And the single shot multi boxes detector (Liu W, 2016), and the residual network

is also added in the YOLOv3 (He K, 2016). However, the network structure of YOLOv3 still can be optimized (Mazzia, 2020), but all this training module size is too big. Next, the lightweight algorithm of the Tiny YOLOv3 was proposed in (Zhang, 2019). The module size of this algorithm is smaller than the others and it can be applied to the real-time object detection. But it still loses part of the detection accuracy. The implementation of the algorithm is divided into several steps: region proposals extraction, CNN (Convolutional Neural Networks) feature computation and bounding-box regression. However, the region proposals need to be pre-fetched, which will take up a lot of disk space. At the same time, each region proposal needs to be calculated by the CNN network, and a large number of overlapping regions will bring waste of computing resources. Due to these shortcomings, Fast-RCNN proposed a solution for building ROI pooling layers and multi-task loss layers. Faster-RCNN used a method of adding additional RPN branch networks to integrate region proposals extraction into deep networks. These solutions have improved the speed of the R-CNN algorithms, but it is still difficult to meet the engineering requirements in real-time video. The regression-based target detection network is more advantageous in detection speed. The SSD network achieves a good balance of efficiency and effect due to the combination of regression and multi-scale features. After continuous iterative improvement of YOLO, at 320×320 YOLOv3 runs in 22 milliseconds at 28.2 mAP, as accurate as SSD but three times faster (Girshick, 2015). In terms of speed and accuracy, it is more appropriate to use YOLOv3 as the target detection network in engineering. The improved YOLOv3 model transforms object detection into a regression problem. YOLOv3 no longer produces candidate regions and directly produces the location and category of the object. The improved YOLOv3 has a faster detection speed and realizes real-time detection and use GPU power to accelerate speed detection. However, the detection accuracy is poor, especially for small objects. This improved YOLOv3 version more optimized feature extraction and fusion, also the improved YOLOv3 improves the mean average precision (mAP) of the self-made dataset, the improved YOLOv3 has more accuracy in the real-time detection. The experimental results show that we can get better detection results under the same detection time with using the improved YOLOv3.

2. SAFETY HELMET DETECTION METHOD

2.1. YOLOv3 Introduction

YOLO version 3(YOLOv3) uses Darknet-53 network structure based on YOLOv1 and YOLOv2 to extract features through 53 convolution layers and 5 maximum pooling layers, adding batch normalization and dropout removal operations to prevent over-fitting. YOLOv3 pre-detection system uses classifiers to perform detection tasks many times, and applies the model to multiple positions and proportions of images. Those areas with higher scores are regarded as detection results. Although YOLOv3 enhances the accuracy of small object detection, the recognition accuracy of safety helmet (mAP-50 between 50 and 60) cannot meet the commercial requirements under complex ground floor area of PLN Indonesia Power Suralaya PGU and various motion postures.

YOLOv3 uses regression problem to compute the target score. YOLOv3 use Darknet-53 as the backbone network and uses three scale predictions. YOLOv3 run neural network on a new image at test time to predict detections. The algorithm divides the input image into S×S grids. If the centre point of the object's ground truth falls within a certain grid, the grid is responsible for detecting the object. Each grid outputs B prediction bounding boxes, including position information of the bounding box (centre point coordinates x, y, width w, height h), and prediction confidence.

YOLOv3 draws on the anchor box idea of Faster R-CNN (Girshick, 2015). YOLOv3 abandon the manually selected anchor box and run k-means clustering on the dimensions of bounding boxes to get good priors. YOLOv3 uses this method to obtain 9 cluster centers, which can better cover the characteristics of the ground truth of the train set. YOLO also adopted a multi-scale prediction method similar to the FPN (Feature Pyramid Networks) network (Girshick, 2015). Because of predictions on multiple scale feature maps, YOLOv3 has acquired image features at different scales and greatly improved the detection of small targets. Combining the anchor box and multi-scale prediction idea, YOLOv3 first assigns several anchor boxes to each scale feature map according to the length and width of the anchor boxes. Calculate the IoU (Intersection-over-Union) of anchor boxes to each ground truth and assign the ground truth to the feature map of the anchor box closest to its shape. When performing bounding box regression training, back propagation will cause the predicted bounding box to approach the ground truth.

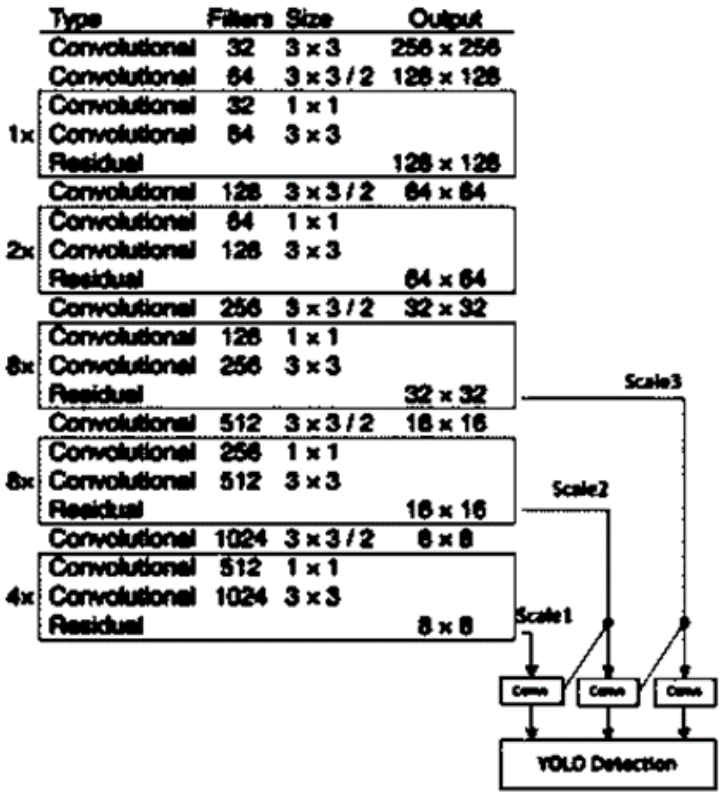

Figure 1. YOLOv3 network structure

2.2. Principal Improvement Version of YOLOv3

This paper discusses the three primary disadvantages of YOLOv3, which are as follows:

1. YOLOv3 has a large network, which results in many model parameter sizes and more physical memory. Due to this reason, it requires high-performance hardware equipment to achieve excellent performance. However, it is difficult to realize real-time detection in mobile devices or cheap devices.
2. The size of input images for YOLOv3 is fixed, and if we normalize the size of images, it can cause image distortion, which can affect the detection effort.
3. Compared to the two-stage object detection algorithm of the RCNN series, YOLOv3 has a poorer ability to recognize the positions of objects and has a low recall rate for multi-target detection (Ren S, 2015).

Despite these disadvantages, YOLOv3 is an excellent one-stage object detection algorithm that offers remarkable detection speed and accuracy and is widely used in the industry. It adopts the backbone network and multiscale feature extraction network. Darknet53 is the backbone network used by YOLOv3, which has better feature extraction abilities than Darknet19 and is better than lightweight networks such as MobileNet. The multiscale feature extraction network outputs feature maps of three different scales, which is suitable for object detection with different sizes and particularly improves the detection ability for small objects. In addition, the multiscale feature extraction network uses the idea of FPN and integrates the feature information of different sizes to effectively improve the detection effect. The residual network employed by Darknet53 can be represented by Eqs. (1)-(3) (Girshick, 2015)

$$X_1 = \sigma\{\beta(W_1, X)\} \tag{1}$$

$$X_2 = \sigma\{\beta(W_2, X_1)\} \tag{2}$$

$$X_3 = X + X_2 \tag{3}$$

where X represents an input feature, (W_1, X) represents an input feature undergoing a convolution with a weight of W_1, and the size of the convolution kernel of W_1 is 1×1. β represents batch normalization, and σ represents nonlinear ReLU activation. (W_2, X_1) represents an input feature undergoing a convolution with a weight of W2, the size of the convolution kernel of W_2 is 3×3, X_2 represents a backbone output feature of the residual

structure, and X3 represents a final output feature of the residual network. Darknet53 first performs a traditional 3 × 3 convolution on the input features and then stacks five residual blocks. The residual network number of each residual block is 1, 2, 8, 8, and 4. Residual blocks are connected through the convolution of downsampling. The outputs of the residual blocks (the 3rd, 4th and 5th) are taken as the input of the multiscale feature extraction network. The sizes of the convolution used in the multiscale feature extraction network including upsampling are 1 × 1 and 3 × 3. Finally, the outputs include feature maps whose scales are 13 × 13, 26 × 26 and 52 × 52.

Although YOLOv3 enhances the accuracy of small object detection, the recognition accuracy of safety helmet (mAP-50 between 50 and 60) cannot meet the commercial requirements under complex ground floor area of PT PLN Indonesia Power Suralaya PGU and various motion postures, it is improved through the following three aspects.

1. Predict and obtain a priori box of safety helmet by clustering algorithm, and classify scale features.
2. Optimizing accuracy by combining deep residual network with multi-scale detection training.
3. Optimize the selection of target box by adjusting the weight of loss function.

The improved YOLOv3 proposed by using the residual network and multi-scale fusion to optimize the network structure of the original YOLO (Ma, 2019).

3. EXPERIMENTAL RESULTS

3.1. Experiment environment

The application of YOLOv3 running on Microsoft Windows 11, Python v 3.10.7, Open CV v 4.6.0, Microsoft Visual Studio 2017, NVIDIA CUDA v 11.01, and NVIDIA CUDNN v 8.0.5 software to enhance the algorithm's performance. Additionally, this paper utilized hardware components, including an AMD Ryzen 3 1200 Quad Core processor, NVIDIA RTX 3050 graphics card, 16GB of RAM, and a 512GB SSD, to maximize system's processing speed and efficiency.

3.2. Outline Steps of YOLOv3

In this paper, there are five outline steps of YOLOv3 to detect the wearing of safety helmet:

a) Bounding Box Prediction
b) Class Prediction
c) Prediciton Across Scales
d) K-means Clustering
e) Feature extraction by darknet classification network

3.3. Differences of Recognition of Safety Helmet by Using Different YOLO

The application of YOLOv3 in this paper used four different YOLOv3 version based on iteration have been processed during training helmet dataset.

Figure 2. Recognition of Safety Helmet using YOLOv3

Figure 3. Recognition of Safety Helmet using The Improved YOLOv3

Figure 4. Recognition of Safety Helmet using Tiny YOLOv3

Figure 5. Recognition of Safety Helmet using Tiny PRN YOLOv3

3.4. Experimental Analysis

3.4.1 Comparison of experimental results

3.4.1.1 Detection by Using Different YOLOv3 Version

Firstly, YOLOv3 helmet detection used to compare four different version YOLOv3 version to detect safety helmet wearing of operator local ground floor at PT PLN Indonesia Power PGU Suralaya by collecting photo images of workers at PT PLN Indonesia Power Suralaya PGU environment. The test accuracy and detection results of different YOLOv3 version are show in Figure 6 and in Figure 8 by using image from figure 6.

Figure 6. Operator PT Indonesia Power PGU Suralaya Waiting The Bus To Go Back To House

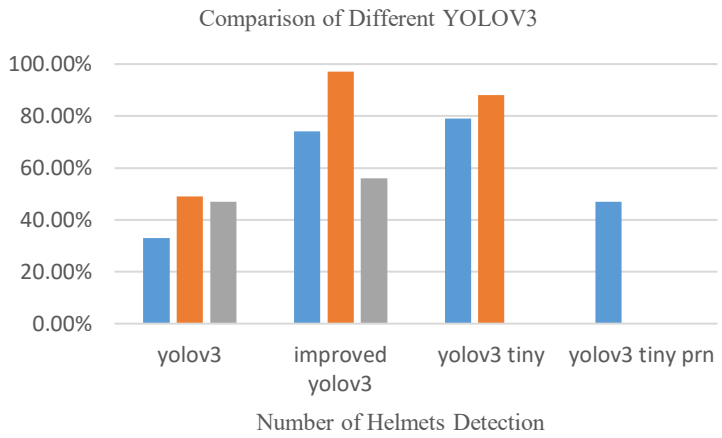

Figure 7. Comparison of Helmet Detection Using Different YOLOv3 Version

Figure 8. Detection results of Improved YOLOv3

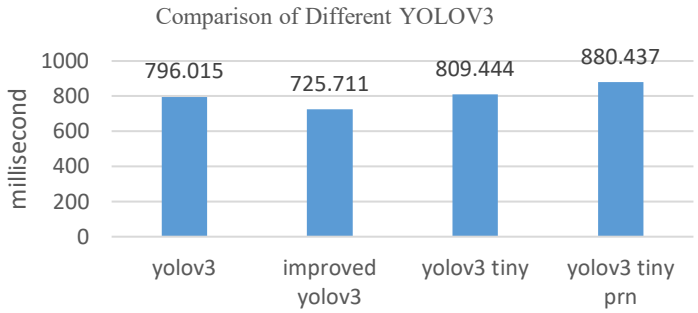

Figure 9. Detection Speed Comparison of Helmet Detection Using Different YOLOv3 Version

From the above results, it can be seen that the recognition accuracy of detection image from figure 7 varies greatly by using improved YOLOv3 version. The improved YOLOv3 version can detect three workers using safety helmet by average accuracy 75,67% within 725.711 ms, YOLOv3 can detect two worker using safety helmet by average accuracy 43,00% within 796.015 ms, Tiny YOLOv3 have better average accuracy by 83,50% but it just detects two workers using safety helmet and slower than Improved YOLOv3, detection speed within 809.444 ms, Tiny-PRN YOLOv3 detect one worker using safety helmet by average accuracy 47% within 880,437 ms.

3.4.1.2 Comparison of mAP (Mean Average Precision) at Different Version YOLOv3

Table 1. Comparison of mAP@IoU=50 and mAP@IoU=75

Darknet	YOLOv3	Improved YOLOv3	Tiny YOLOv3	Tiny-PRN YOLOv3
Total BFLOPS	65.312	65.312	9.673	3.406
AVG-Output	516922	516922	577025	275530
Detections Count	3371	1297	1767	3159
No Helmet AP@0.50	91.31%	99.00%	95.99%	84.63%
Helmet AP@0.50	69.16%	94.25%	87.21%	78.89%
No Helmet AP@0.75	12.46%	57.58%	33.70%	14.94%
Helmet AP@0.75	4.36%	39.27%	15.33%	10.69%
mAP @0.50	80.24%	96.63%	91.60%	81.76%
mAP @0.75	80.24%	48.43%	24.52%	12.81%

Based on the above verification analysis of different version YOLOv3, aiming at the detection requirements in PLN Indonesia Power Suralaya PGU environment to using comercial, this paper compares the effect of the improved algorithm for single object detection and multi-object detection on the conclusion that the overall accuracy of Improved YOLOv3 is meet the satisfactory based on detection speed and detection accuracy.

4. CONCLUSION

This paper presents a new and improved method for detecting whether or not workers are wearing safety helmets at the PT PLN Indonesia Power Suralaya PGU Unit 2 in real-time. The method is based on the YOLOv3 detection model and uses a combination of deep residual network technology and multi-scale convolution feature to increase the accuracy of the detection. In addition, the method uses multi-scale detection training and adjusts the weight of the loss function while maintaining a high detection rate. The result is a significant improvement in the detection rate and accuracy, with a 76% increase in the accuracy of detecting a single object in the mine environment, from 0.43 to 0.76. The accuracy speed is also improved, decreasing from 796 ms to 725 ms, an 8.9% decrease. This new method meets the commercial safety requirements for real-time monitoring of safety helmet wearing in the working environment (ground floor area) of PT PLN Indonesia Power PGU Suralaya Unit 2.

5. ACKNOWLEDGMENT

We sincerely appreciate the support received from the knowledge team of PT PLN Indonesia Power and colleague from PT PLN Indonesia Power PGU Suralaya

6. REFERENCES

Dollar, P., Appel, R., & Belongie, S. (2014). Fast feature pyramids for object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(08), 1532–1545

Girshick, R. (2015) Fast R-CNN[C]//IEEE International Conference on Computer Vision. IEEE, 2015:1440-1448.

He K , Zhang X , Ren S , et al. (2016). Deep Residual Learning for Image Recognition[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society: Page 770- 778.

Krizhevsky, A., Sutskever, I., Hinton, G.E., (2012). ImageNet Classification With Deep Convolutional Neural Networks[C]//International Conference on Neural Information Processing Systems. Curran Associates Inc, 1097-1105.

Liu, W., Anguelov, D., & Erhan, D. (2016). SSD: Single shot multibox detector. Computer Vision - ECCV 2016, PT I. Lecture Notes in Computer Science, 9905, 21–37.

Ma, Q.,Zhu, B,Zhang, H.W.,Zhang, Y., Jiang, Y.C. (2019) Low-altitude UAV Detection and Recognition Method Based on Optimized YOLOv3.J. Laser & Optoelectronics Progress. Vol 56, No 20.

Mazzia, V., Khaliq, A., & Salvetti, F. (2020). Real-time apple detection system using embedded systems with hardware accelerators: An edge AI application. IEEE ACCESS, 08, 9102–9114.

Redmon, J., Divvala, S., Girshick, R., et al. (2015) You Only Look Once: Unified, Real-Time Object Detection. J. Page 779-788

Redmon, J., Farhadi, A. (2018). YOLOv3: An Incremental Improvement. J. IEEE Conference on ComputerVision and Pattern Recognition. IEEE, 89-95.

Redmon, J., Farhadi. A. (2017). YOLO9000: Better, Faster,Stronger[C]//IEEE Conference on Computer Visionand Pattern Recognition. IEEE. Page 6517-6525.

Ren S , He K , Girshick R , et al. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence. Vol 39, No 6:1137-1149

Ren, S., He, K., Girshick, R., et al. (2015). Faster R-CNN: Towards Real-Time Object Detection With Region Proposal Networks[C]//International Conference on Neural Information Processing Systems. MIT Press, 91-99.

Zhang, Y., Shen, Y. L., & Zhang, J. (2019). An improved Tiny YOLOv3 pedestrian detection algorithm. OPTIK, 183, 17–23.