Comparative Study of Deep Learning Algorithms BetweenYOLOv5 and Mobilenet-SSDv2 As Fast and Robust Outdoor Object Detection Solutions

Ryan Satria Wijaya*[1], Santonius Hasibuan[1], Anugerah Wibisana[1], Eko Rudia-wan Jamzuri[1], Mochamad Ari Bagus Nugroho[2]

[1]Department of Electrical Engineering, Politeknik Negeri Batam, Batam, Indonesia

[2]Department of Electrical Engineering, Politeknik Elektronika Negeri Surabaya, Surabaya, Indonesia

ryan@polibatam.ac.id

Abstract. Object detection is one of the most popular applications among young people, especially millennials and Generation Z. The use of object detection has become widespread in various aspects of daily life, such as facial recognition, traffic management, and autonomous vehicles. In the implementation of object detection, large and complex datasets are required. Thus, it is important to choose an efficient object detection algorithm that yields good results. This research compares the performance of YOLOv5 and MobileNet-SSDv2 using the same dataset, demonstrating that YOLOv5 outperforms MobileNet-SSDv2 in terms of speed and accuracy in object detection. The results indicate that YOLOv5 is capable of detecting objects more rapidly and accurately compared to MobileNet-SSDv2, especially under varying lighting conditions. Several factors affecting the performance of these algorithms include the complexity of the dataset used, the available processor speed, and the memory capacity that can be utilized.

Keywords: YOLOv5, MobileNet-SSDv2, Performance Comprasion, Object Detection, Deep Learning Algorithms

1 Introduction

This research was conducted due to several important aspects related to object detection using deep learning technology, which is currently a highly interesting topic in computer vision research. Deep learning algorithms such as YOLOv5 and MobileNet-SSDv2 have been widely used in various applications and have shown excellent results in detecting objects with high accuracy. However, the performance ofeach algorithm still needs to be further investigated to understand their effectiveness invarious real-world conditions. Although these algorithms are widely used, the specificchallenge this paper seeks to address is selecting the most efficient algorithm between YOLOv5 and

L. Lumombo et al. (eds.), *Proceedings of the 7th International Conference on Applied Engineering (ICAE 2024)*, Advances in Engineering Research 251,
https://doi.org/10.2991/978-94-6463-620-8_8

MobileNet-SSDv2 for outdoor object detection applications that require high speed, accuracy, and resource efficiency. The need for the right algorithm is crucial in applications such as autonomous vehicles, surveillance, and long-range objectdetection. Several previous studies have compared the performance of these two algorithms. For instance, research by Wang et al. (2021) shows that YOLOv5 outperforms MobileNet-SSDv2 in terms of speed and accuracy on the extensive COCO dataset. This result provides valuable insights for developers in choosing the right algorithm for applications that require fast detection.

On the other hand, research by Ma et al. (2021) shows that MobileNet-SSDv2 is highly effective for surface defect detection with high accuracy, especially on small objects and with additional image processing. This algorithm aids in surface inspection to efficiently identify defects in various materials. Moreover, research by Liu et al. (2021) demonstrates that YOLOv5 performs very well in safety helmet detection in high-frame-rate videos, which can be applied in construction safety to ensure proper helmet use by workers. They combined YOLOv5 with super-resolution reconstruction techniques to improve helmet detection accuracy in video. Another study by Phadtare et al. (2021) compared YOLOv3 with MobileNet-SSDv2 in an aerial surveillance system. The study shows that YOLOv3 performs better in detecting objects with low resolution on drones, providing key insights for the development of object detection systems for monitoring and security using drones.

Finally, research by Huang et al. (2021) shows that combining the MobileNet-SSDv2 model with YOLOv3 yields good results for vehicle detection at night, which is one of the challenges in object detection under low-light conditions. Although this paper mentions the use of annotated datasets, it is important to provide more details on the annotation process. The dataset annotation involves manually labeling objects in the images, which are then used to train the YOLOv5 and MobileNet-SSDv2 models. This labeling includes assigning labels to objects such as red and green balls in the images, which are used to evaluate the accuracy and performance of object detection in both models.

2 Method

In this study, we adopted two object detection methods that have proven to be effec-tive, namely YOLOv5 and MobileNet-SSDv2. This method has commonly utilized in a wide array of fields and has its own advantages in object detection.

2.1 YOLOv5

YOLOv5 is the latest innovation in the object detection algorithm which commonly utilized in a wide array of fields. Wu et al. (2021) developed a new technique by combining a local convolution neural network (CNN) with the YOLOv5 object detection algorithm to improve identifying small objects in remote imagery [6]. This approach improves YOLOv5's ability to recognize small objects in remote imagery, so that detection results become more accurate and efficient. In addition, Zhu et al. (2021) also improved the performance of YOLOv5 by developing the TPH-YOLOv5 which is

equipped with a Transformer Prediction Head, especially Intended for object recognition in drone shooting conditions [7]. This approach enhances accuracy and efficiency of object recognition. using YOLOv5 in drone images. In agriculture, Chen et al. (2021) used enhanced YOLOv5 to model plant disease recognition [8]. This allows the system to quickly and accurately detect and identify diseases in crops, whichin turn allows for timely control measures to protect crops and increase crop yields.

Furthermore, Yao et al. (2021) developed a real time kiwi defect identification algorithm using YOLOv5 as a framework [9]. This algorithm allows the system to automatically detect defects in kiwi fruit in real time. This helps in increasing the efficiency and accuracy of the kiwifruit sorting process, as well as improving the quality of the final product. In the area of image recognition, Yang et al. (2021) developed a YOLOv5-based facial mask recognition system [10]. In the context of the COVID-19 pandemic, this system can help detect whether someone is wearing a mask or not in real-time. This is useful for monitoring compliance with mask use policies and helping to maintain public health. Lastly, several recent studies have also reported improved performance and further enhancements to YOLOv5. For example, Mahendrakar et al. (2023) conducted a performance analysis of YOLOv5 and Faster R-CNN for self-guided navigation around uncooperative objects [12]. They concluded that YOLOv5 provided better results in this context.

Furthermore, Ren et al. (2023) developed an enhanced Real-time object detection using the YOLOv5 framework in vehicle camera shooting conditions [13]. Their research results show an increase in performance compared to the previous version. Rafi et al. (2023) also performed a performance analysis of the YOLO model for vehicle recognition in the South Asian region [14]. They revealed that the YOLOv5 model gave satisfactory results in this context. In addition, Benjumea et al. (2023) proposed YOLO-Z, a method to improve YOLOv5 forsmall object detection in autonomous vehicles [15]. Their experimental results show improved performance in this regard. As such, YOLOv5 has become a cornerstone of innovation in various fields, including remote sensing, agriculture, image recognition, and robotics. Continuing improvement and development in YOLOv5 help improve object detection performance and provides a more reliable solution. The training process was conducted with a batch size of 2 and a total of 1000 epochs, with a learningrate adjusted to achieve optimal results.

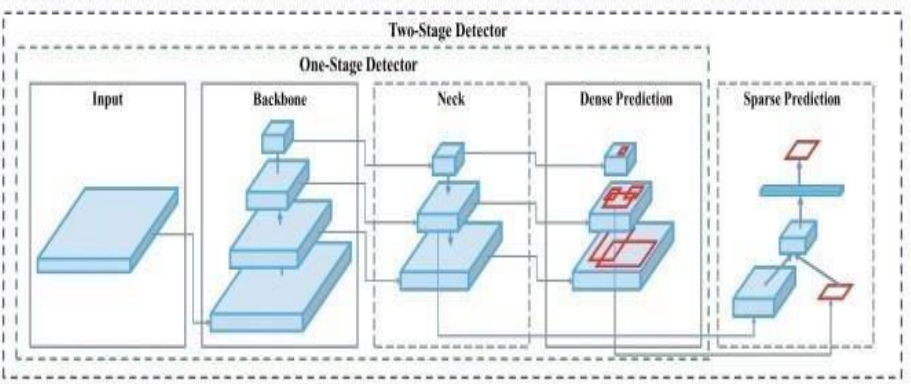

Fig. 1. Network architecture.

Figure 1 shows the YOLOv5 Network Architecture from three main parts, namely:

a. Backbone

Backbone is the part of the network that is responsible for extracting features from images. In YOLOv5, Efficient Net is used as the backbone for feature extraction. Efficient Net is combined with several convolution layers and pooling layers to produce better features.

b. Necks

Necks is part of the network that connects between the backbone with the head. In YOLOv5, neck uses Spatial Pyramid Pooling which functions to extract more detailed features at each image scale.

c. Head

Head is the part of the network that is responsible for predicting objects in images. In YOLOv5, head uses multiple convolution layers and nonlinear layers to produce an output in tensor form with the same size for each grid in the image. Each output tensor will be processed to determine the location, size, and class of objects in the image.

2.2 Mobilenet-SSDv2

Mobilenet-SSDv2 is an object detection model developed by combining the Mobilenet-SSDv2 model and Single Shot Detector algorithm (SSD). This model offers advantages in terms of light size and efficient use of repair resources. For example, in research by Ratna et al (2019), the Mobilenet-SSDv2 is used as a trained model in building object recognition applications applying the TensorFlow Object Detection API [16]. In addition, the Mobilenet-SSDv2 has also been adopted in various other studies. For example, a study by Sudibyo et al. (2021) regarding a based-on edge computing automatic vehicle classification system uses the Mobilenet-SSDv2 model as one of its components [17]. This model is also used in Tiny Deeply Supervised Object Detection, object detection model designed for limited resource use [18]. Li, Y et al. (2018) Apart from object detection applications, the Mobilenet-SSDv2 is also used in various other fields. For example, in research by Carballo et al. (2018), this model is applied in the use of machine learning for tracking solar panels [19].

Furthermore et al. (2020) uses the Mobilenet-SSDv2 to detect cracks in the road surface using deep learning and object detection [20]. Thus, the Mobilenet-SSDv2 has proven to be effective and widely usedin various studies for object detection applications and related fields such as vehicle classification systems, use of limited resources, solar tracking panels, and crack analy- sis on road surfaces. The training process for MobileNet-SSDv2 also uses a batch size of 2 and a total of 1000 epochs, with appropriate adjustments to the learning rate."

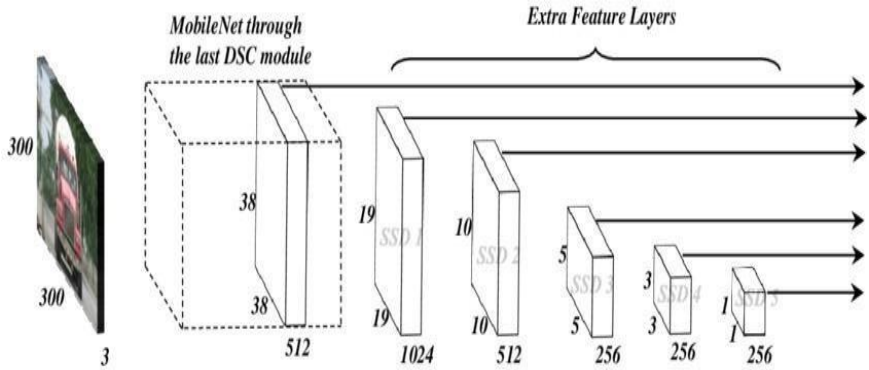

Fig. 2. Network Architecture Mobilenet-SSDv2

Figure 2 shows the Mobilenet-SSDv2 Network Architecture from two main parts, namely:

a. Mobilenet-SSDv2 Base Network

The final layer in the MobileNet-SSDv2 architecture is responsible for predicting object labels and bounding boxes within the image. The MobileNetv2 architecture is composed of multiple interconnected layers,commonly known as additional feature layers. These layers aim to capture more abstract and complex features in images that are useful for object detection.

b. MultiBox Head

MultiBox Head Is the last layer of the MobileNet-SSDv2 architecture. Thislayer is used to make predictions of object labels and bounding boxes within the image. Multi Box Head consists within a convolutional layer, depth wise separable convolution layer, and an activation layer.

3 Results and Discussion

This part describes the findings of the analysis as well as explains a comprehensive discussion. Results can be presented in the form of pictures and tables that are made easy for readers to understand. In the discussion there are several important things needed to detect objects

3.1 Alternative Research Methodologies

Dataset. Dataset utilized in this study compares the performance of the YOLOv5 and Mobilenet-SSDv2 algorithms. It consists of 2671 Red Ball and Green Ball images, specially selected for YOLOv5 and Mobilenet-SSDv2, along with thecorresponding object labels. These images serve as input for each object detection algorithm. Performance assessment measures are employed to compare the efficacy of the two algorithms. The dataset plays a crucial role inensuring the accuracy and reliability of the performance comparison results between YOLOv5 and MobileNet-SSDv2.

Labeling the Dataset. After collecting the dataset, the next phase requires data annotations being prepared. Annotation refers to the process of assigning labels or categories todata in a data set, facilitating the training and testing of deep learning models.

Model Training. Following the dataset preparation, the subsequent step entails creating deeplearning models for Yolov5 and Mobilenet-SSDv2. These models are then trained using the prepared data. The model training process involves multiple iterations, including inputting data into the model, examining the resulting out- put, evaluating model performance, and optimizing parameters to enhance themodel's effectiveness.

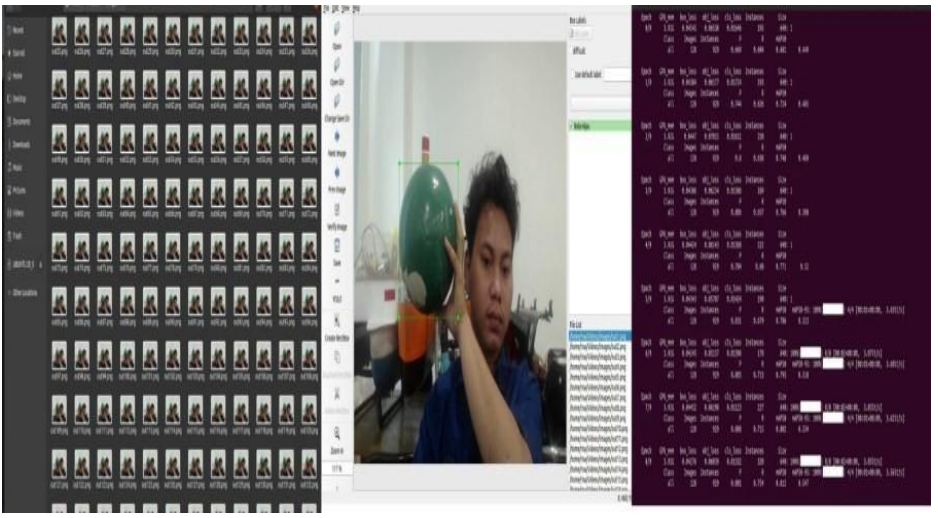

Fig 3. Dataset, label, train.

3.2 Experimental Results

Detection Results. The results of the object detection above show that Figure 4 which is on the left isa training data set containing green and red ball images to train the algorithm. Figure 4which is in the middle shows the performance of the YOLOv5 algorithm in identifyinggreen and red balls in the visible image. Meanwhile, Figure 4 which is on the right shows the performance of the Mobilenet-SSDv2 algorithm in recognizing the same object. These three images are useful for evaluating the performance and precision of object recognition algorithms and assisting in the selection of algorithms that suit the needs and goals of the application.

Experiment Results Mileage. Based on the experimental results obtained from the two algorithms in Table 1, it can be concluded that in the range of 1.5 meters to 4.5 meters, both the YOLOv5 and Mobilenet-SSDv2 algorithms successfully detect objects perfectly. The second model is capable of recognizing and labeling objects with high accuracy. However, at a range of 5.5 meters to 6.5 meters, the performance of the YOLOv5 and Mobilenet- SSDv2 algorithms has decreased in detecting objects

perfectly. In fact, at 7 meters andabove, the two algorithms are not able to detect objects properly.

Fig. 4. Dataset, YOLOv5, MOBILENET-SSDv2.

Table 1. Miliage comprasion.

Distance	Yolov5			Mobilenet-SSDv2		
	Perfectly detected	Imperfectly detected	Not detected	Perfectly detected	Imperfectly detected	Not detected
1.5 (M)	√			√		
2.5 (M)	√			√		
3.5 (M)	√			√		
4.5 (M)	√			√		
5.5 (M)		√			√	
6.5 (M)		√			√	
>7 (M)			√			√

Figure 5 shows the results of object detection by calculating the distance from 1.5 meters to 6.5 meters using the YOLOv5 and Mobilenet-SSDv2 algorithms. In Table 1 on the left, there are detection results using the YOLOv5 algorithm, while on the right, there are detection results using the Mobilenet-SSDv2 algorithm.

Fig. 5. Object detection distance of YOLOv5 and MOBILENET-SSDv2 algorithms.

3.3 Confusion Matrix

Confusion matrix is a method for evaluating the performance of an algorithm or classification. In deep learning algorithms such as YOLOv5 and MobileNet-SSDv2, the fusion matrix is used to evaluate performance models in object detection tasks. In practice, the Confusion Matrix is used to calculate various evaluation metrics such as accuracy (accuracy), precision (precision), sensitivity (recall), and f1 scores to measure the performance of measuring objects in the model used.

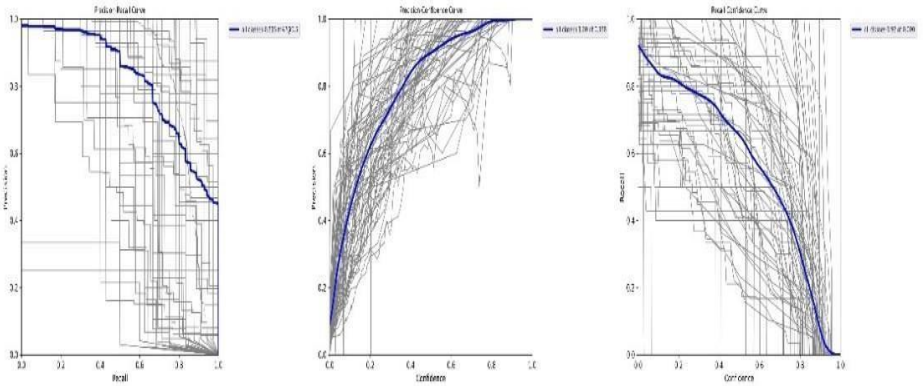

Fig. 6. Confusion Matrix YOLOv5.

i Precision - Confidence Curve: Shows the relationship between precision and confidence in the YOLOv5 model. In the context of the confusion matrix, precision refers to the model's ability to provide correct positive results

precisely, while trust refers to the level of confidence the model has in the results given.

ii Precision - Recall Curve: Shows the relationship between precision and recallin the YOLOv5 model. Recall in the confusion matrix refers to the model's ability to correctly identify all positive samples.

iii Recall - Confidence Curve: Shows the relationship between recall rates and confidence levels in the YOLOv5 model. In the context of the confusion matrix, recall refers to the model's ability to identify all positive samples present.

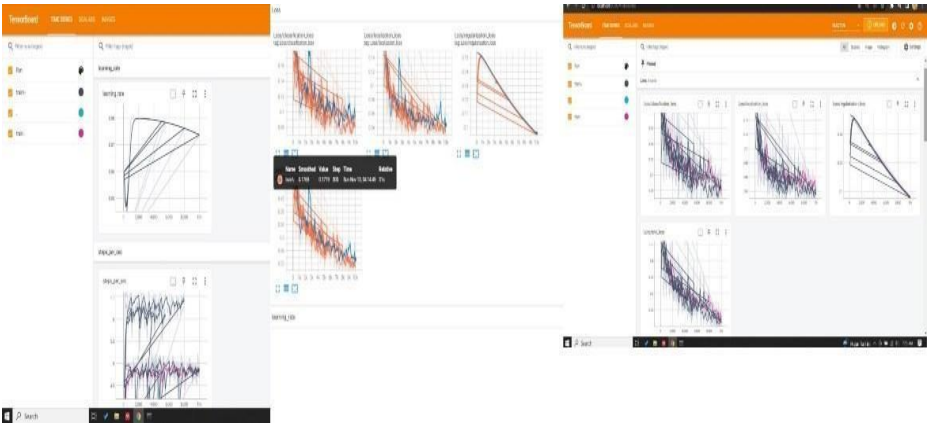

Fig. 7. Confusion Matrix MOBILENET-SSDV2.

i Learning_rate graph: Shows the change in learning rate during model training. Learning rate in the context of the confusion matrix is not directly related to the evaluation of the confusion matrix itself, but can affect the model trainingprocess and the accuracy of the resulting results.

ii Steps_per_sec Graph: Shows the model's training speed in steps per second. This is not directly related to the confusion matrix, but gives an idea of the speed of model training and its performance in processing data.

iii Loss/Classification_loss Graph: Shows the change in the value of the loss (loss) of the loss function related to classification. Loss in the context of the confusion matrix can provide an indication of the extent to which the model has errors in classifying data.

3.4 Object Detection Results in Various Conditions

In dark conditions, the YOLOv5 algorithm also succeeded in detecting objects, although not perfectly, while the Mobilenet-SSDv2 algorithm failed to detect objects at all, indicating that YOLOv5 is more reliable in dealing with situations with very low lighting.

Table 2. Comparison of conditions.

Condition	Yolov5			Mobilenet-SSDv2		
	Perfectly detected	Imperfectly detected	Not detected	Perfectly detected	Imperfectly detected	Not detected
Bright	√			√		
Dim	√				√	
Dark		√				√

Fig. 8. Condition comparison between YOLOv5 and MOBILENET-SSDv2 algo-rithms.

Figure 8 shows a comparison of bright, dim, and dark conditions using the YOLOv5 and Mobilenet-SSDv2 algorithms. In figure 7 there are 2 images combined into one. The top image is an image of the Yolov5 algorithm, and the image below is the Mobilenet-SSDv2 algorithm. In the picture, three lighting conditions can be seen, namely bright, dim, and dark conditions.

i In bright conditions, both the YOLOv5 and Mobilenet-SSDv2 algorithms successfully detect objects perfectly, showing good performance of both algorithms in that situation.
ii In dim conditions, the YOLOv5 algorithm was able to detect objects perfectly, while the Mobilenet-SSDv2 algorithm only managed to detect objects imperfectly, indicating that YOLOv5 has an advantage in more challenging lighting conditions.

4 Conclusion

From the above research results, it can be concluded that YOLOv5 and Mobilenet-SSDv2 are two deep learning algorithms that are very useful for implementing real-time computer vision applications. YOLOv5 excels in object detection speed and accuracy, making it suitable for use in applications requiring rapid response, such as autonomous vehicles or security monitoring systems. Meanwhile, Mobilenet-SSDv2 has a higher ability to detect small objects or in low lighting.

References

1 Wang, J., Sun, Y., Liu, Z., & Sarma, K. M. (2021). A comparison of YOLOv5 and MobileNet-SSDv2 on COCO dataset. arXiv preprintarXiv:2103.16808.
2 Liu, J., Wang, J., & Lu, Y. (2021). Safety helmet detection based on YOLOv5 driven by super- resolution reconstruction. Multimedia Tools and Applications, 80(26), 39389-39409.
3 Ma, M., Li, G., Wei, J., Li, Z., Li, C., & Li, J. (2021). Research on a SurfaceDefect Detection Algorithm Based on MobileNet-SSD. Journal of Physics: Conference Series, 1822(1), 012029. doi: 10.1088/1742- 6596/1822/1/012029
4 M Phadtare, V Choudhari, R Pedram. (2021). Comparison between YOLO and SSD mobile net for object detection in a surveillance drone. International Journal of Scientific Research in Engineering and Management, 5(3), 506-512.

5 Huang, S., Wu, H., Zhang, Y., & Liu, Q. (2021). M-YOLO: A Nighttime Vehicle Detection Method Combining Mobilenet v2 and YOLO v3. Journal of Physics: Conference Series, 1883(1), 012094. doi: 10.1088/1742- 6596/1883/1/012094.

6 Wu, W., Liu, H., Li, L., Long, Y., Wang, X., Wang, Z., Li, J., & Chang, Y. (2021). Application of local fully Convolutional Neural Network combinedwith YOLO v5 algorithm in small target detection of remote sensing image. Journal of Ambient Intelligence and Humanized Computing, 12(11), 11599- 11609. doi: 10.1007/s12652-021-03283-7.

7 Zhu, X., Lyu, S., Wang, X., & Zhao, Q. (2021). TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios. IEEE Transactions on Geoscience and Remote Sensing. doi: 10.1109/TGRS.2021.3111818.

8 Chen, Z., Wu, R., Lin, Y., Li, C., Chen, S., Yuan, Z., Chen, S., & Zou, X. (2021). Plant Disease Recognition Model Based on Improved YOLOv5. IEEE Access, 9, 102633-102644. doi: 10.1109/ACCESS.2021.3097340.

9 Yao, J., Qi, J., Zhang, J., Shao, H., Yang, J., & Li, X. (2021). A Real-Time Detection Algorithm for Kiwifruit Defects Based on YOLOv5. IEEE Access, 9, 122461-122473. doi: 10.1109/ACCESS.2021.3108317.

10 Yang, G., Feng, W., Jin, J., Lei, Q., Li, X., Gui, G., & Wang, W. (2021). Face Mask Recognition System with YOLOV5 Based on Image Recognition. Journal of Physics: Conference Series, 1859(1), 012181. doi: 10.1088/1742-6596/1859/1/012181.

11 Yan, B., Fan, P., Lei, X., Liu, Z., & Yang, F. (2022). A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5. Journal of Physics: Conference Series, 1967(1), 012029. doi: 10.1088/1742-6596/1967/1/012029.

12 Ratna Aningtiyas, P., Sumin, A., & Wirawan, S. (2019). Pembuatan AplikasiDeteksi Objek Menggunakan TensorFlow Object Detection API dengan Memanfaatkan SSD MobileNet V2 Sebagai Model Pra-Terlatih. JurnalTeknik Informatika dan Sistem Informasi, 5(1), 57-65.

13 Mahendrakar, T., Ekblad, A., Fischer, N., White, R. T., Wilde, M., Kish, B.,& Silver, I. (2023). Performance Study of YOLOv5 and Faster R-CNN for Autonomous Navigation around Non-Cooperative Targets. Florida Institute of Technology. Unpublished manuscript.

14 Ren, Z., Zhang, H., & Li, Z. (2023). Improved YOLOv5 Network for Real-Time Object Detection in Vehicle-Mounted Camera Capture Scenarios. Sensors, 23(10), 4589.

15 Rafi, M. M., Chakma, S., Mahmud, A., Rozario, R. X., Munna, R. U., Abedin, M. A., Wohra, R. H. J., Mahmud, K. R., & Paul, B. (2022). Performance Analysis of Deep Learning YOLO Models for South Asian Regional Vehicle Recognition. Department of Computer Science and Engineering, University of Liberal Arts Bangladesh, Dhaka, Bangladesh.

16 Benjumea, A., Teeti, I., Cuzzolin, F., & Bradley, A. (2021). YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles. Email addresses.

17 Sudibyo, R. W., Mahmudah, H., Hadi, M. Z. S., & Sa'adah, N. (2021). Edge Computing-Based Automated Vehicle Classification System Using the MobileNet V2 Model. IEEE Access, 9, 42226-42239.
18 Li, Y., Lin, W., & Li, J. (2018). Tiny-DSOD: Lightweight Object Detection for Resource-Restricted Usages. arXiv preprint arXiv:1807.11013.
19 Carballo, J. A., Bonilla, J., Berenguel, M., Fernández-Caballero, A., & Baeyens, E. (2018). Machine learning for solar trackers. Renewable and Sustainable Energy Reviews, 94, 38-50.
20 Hassan, S. I., O'sullivan, D., McKeever, S., McGowan, R., Feighan, K., & Power, D. (2020). Detecting Patches on Road Pavement Images Acquired with 3D Laser Sensors using Object Detection and Deep Learning. Journal of Computing in Civil Engineering, 34(6), 04020063. doi:10.1061/(asce)cp.1943-5487.0000968.

.