Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 06 Issue 01
2023

Research Article
https://doi.org/10.56979/601/2023

Revolutionizing Real-Time Object Detection: YOLO and MobileNet SSD Integration

Daud Khan[1], Muhammad Waqas[1], Mohsin Tahir[1], Shahab Ul Islam[2], Muhammad Amin[1], Atif Ishtiaq[1], and Latif Jan[1]*

[1]Iqra National University, Peshawar, Pakistan.
[2]Parthenope University of Naples, Italy.
*Corresponding Author: Latif Jan. Email: latifjan@inu.edu.pk

Received: November 11, 2023 **Accepted:** November 29, 2023 **Published:** December 05, 2023

Abstract: Real-time object detection using machine learning techniques has improved algorithm performance, but issues like blurring, noise, and rotating jitter in real-world images impact detection methods. You Only Look Once (YOLO) is a faster and more accurate real-time object detection algorithm that can detect multiple objects in a single image, unlike other Convolutional Neural Network (CNN) based algorithms. This paper integrates YOLO (version 3) v3 and MobileNet Single Shot Detector (SSD), resulting in faster image detection and accurate localization. It also compares lighter versions of YOLOv3 and YOLOv4 in terms of accuracy.The integration of YOLOv3 and MobileNet SSD enables real-time object detection in various applications like augmented reality, robotics, surveillance systems, and autonomous vehicles. It enhances security, enables immediate responses to potential threats, and allows robots to perceive and interact with their environment. Finally, the work provides an insight into the performance, and capabilities of YOLOv3 and MobileNet SSD, leading to an informed decision-making process for integrating both algorithms in OpenCV.

Keywords: You Only Look Once, Real Time Object Detection, MobileNet Single Shot Detector, Convolutional Neural Network.

1. Introduction

The object detection through You Only Look Once version 3 (YOLOv3), and MobileNet Single Shot Detector (SSD) algorithms. YOLOv3 and other classifiers detect accurate detection using the OpenCV library in existing powerful techniques to perform object detection location. However, OpenCV library tends to open-source used for computer vision, providing and integrating the interface of these algorithms to leverage their capabilities for object detection applications [1]. YOLOv3 classifiers are acknowledged for their speed and accuracy. Real-time object detection divides and predicts bounding boxes, classes, grid, and probabilities for each grid cell. YOLOv3 also needs further improvement with architectural changes and enhancements by predecessors [4]. It is popular for its capability to detect objects across various aspects and scale ratios within a single pass. Real-time object detection implemented in YOLOv3 is a famous and powerful object detection classifier acknowledged for its speed and accuracy [5]. This algorithm is specifically designed for mobile and embedded devices with limited computational resources. MobileNet SSD provides a good balance between speed and accuracy, making it suitable for real-time object detection on resource-constrained devices. Object detection using MobileNet SSD (Single Shot MultiBox Detector) is a popular and efficient approach for performing real-time object detection tasks. MobileNet SSD combines the MobileNet architecture, a lightweight and efficient neural network, with the SSD framework to achieve accurate and fast object detection [7].YOLOv4-tiny, on the other hand, is a compact version of YOLOv4 designed specifically for scenarios where computational resources are limited. It utilizes advanced techniques such as CSPDarknet53 as the backbone network, which significantly reduces the model's size and computational requirements while maintaining competitive accuracy. YOLOv4-tiny achieves remarkable speed without compromising much on detection accuracy, making it an excellent choice for real-time object detection on edge devices and platforms with limited computational power [8].YOLOv4-tiny, a compact variant of YOLOv4, is specifically designed for real-time object detection on resource-constrained platforms. It incorporates advanced techniques such as CSPDarknet53 as the backbone network, reducing model size and computational requirements. Despite its compact nature, YOLOv4-tiny maintains competitive accuracy and achieves impressive inference speed, making it suitable for real-time applications on edge devices [9].

2. Literature and Related Work

Joseph Redmon wrote the real-time object YOLO algorithm for detecting objects and determining predictions through YOLO classifiers [1]. Optimal techniques based on the CNN model and YOLOv3 algorithm, explaining efficiency and increasing reliability, were elaborated by Juan Du [2]. Other researchers focused on localizing and learning objects with a structured approach to producing output regression, as written by Matthew B. Blaschko. The researchers emphasized, in this paper, object localization and dealing with bounding box approaches for the localization of objects to overcome the limitations of sliding window techniques [3]. Real-time object detection is a challenging task in computer vision, including detecting and localizing objects of concern in real-time video streams or live camera feeds. In recent years, two famous classifiers, YOLOv3 (You Only Look Once) and MobileNet SSD (Single Shot Detector), have gained important attention and achieved remarkable and impressive performance in real-time object detection [4]. YOLOv3 is a state-of-the-art object detection algorithm targeting high accuracy and analyzing the speed of algorithms. It adopts a single-shot detection method, dividing the input image into a grid and predicting bounding boxes and class probabilities directly from each grid cell. YOLOv3 uses a deep neural network architecture with multiple detection layers to capture objects at different scales and achieve high detection accuracy. It can detect a wide range of object classes and handle complex scenes efficiently [5].MobileNet SSD, on the other hand, combines the MobileNet architecture and the SSD framework to achieve real-time object detection on resource-constrained devices. MobileNet is a lightweight neural network architecture designed for mobile and embedded devices, using depth-wise separable convolutions to reduce computational complexity while maintaining good accuracy. The SSD framework allows for multi-scale feature extraction and detection, enabling the detection of objects at various sizes and aspect ratios [6]. Both YOLOv3 and MobileNet SSD offer advantages in terms of real-time performance, accuracy, and versatility. They have been widely adopted in various applications, including surveillance systems, autonomous vehicles, robotics, and augmented reality [7]. In surveillance systems, real-time object detection using YOLOv3 or MobileNet SSD enables the detection and tracking of objects of interest, such as people, vehicles, or suspicious items, in live video streams. This enhances security and enables immediate responses to potential threats [8]. In autonomous vehicles, real-time object detection is crucial for identifying and tracking pedestrians, vehicles, and obstacles in the vehicle's surroundings. YOLOv3 and MobileNet SSD can provide real-time object detection capabilities, enabling autonomous vehicles to make informed decisions and ensure safe navigation [9]. In robotics, real-time object detection allows robots to perceive and interact with their environment. YOLOv3 and MobileNet SSD can be used to detect objects for object manipulation, scene understanding, or collaborative tasks between robots and humans [10].In augmented reality applications, real-time object detection using YOLOv3 or Mobile Net SSD enables the recognition and tracking of objects in real-world scenes. This allows for interactive and immersive experiences, where virtual objects can be placed and interacted with in real-time [11].YOLOv4 Tiny, being a more recent iteration, incorporates several advancements and demonstrates improved performance [12]. YOLOv4 Tiny generally achieves better accuracy than YOLOv3 MobileNet SSD while maintaining a small model size and real-time inference capabilities. It is particularly suitable for applications where computational resources are limited, such as embedded systems, drones, edge devices, vehicles and pedestrian's detections using YOLOv3 and YOLOv4 self-driving cars [13].

This paper aims to integrate the YOLOv3 with MobileNet SSD for real-time object detection, with a focus on improving algorithm performance in challenging real-world images affected by blurring, noise, and rotating jitter.

The objective is to overcome the limitations of sliding window techniques by emphasizing object localization and bounding box approaches for the accurate detection and localization of objects.

The paper also aims to analyze and predict the performance efficacy of classifiers in object detection, recommending the evaluation and benchmarking of algorithms on specific use cases to determine the most suitable option based on accuracy, speed, and deployment constraints.

2.1 Contribution of the paper

The paper revolutionizes real-time object detection by integrating YOLOv3 and MobileNet SSD, resulting in faster image detection compared to other algorithms.

The integration of YOLOv3 and MobileNet SSD enables the detection of multiple objects in a single image, unlike many other CNN-based algorithms that can only detect one object at a time.

The paper presents a comparison between lighter versions of YOLOv3 and YOLOv4, along with day and night time, in terms of accuracy.

YOLOv3 and MobileNet SSD can be used for object manipulation, scene understanding, collaborative tasks between robots and humans, and augmented reality applications.

The paper provides a comprehensive understanding of the underlying concepts, methodologies, and software lifecycle associated with implementing YOLOv3 and MobileNet SSD in OpenCV.

The performance, and capabilities of YOLOv3 and MobileNet SSD are explored, leading to an informed decision-making process for integrating both algorithms in OpenCV.

3. Implementation of the YOLO Algorithm

The You Only Look Once (YOLO) algorithm is renowned for its impressive high accuracy and real-time object detection capabilities. The YOLO classifier employs forward propagation based on the network to make accurate predictions. The non-max suppression, named in Andrew Ng's video explanation, refines YOLO's recognized diagrams. YOLOv3, a real-time object detection classifier, analyzes objects in images or video frames and can be implemented in OpenCV using pre-trained models or custom datasets. YOLOv3 in OpenCV utilizes deep neural networks, employing a single-pass approach for direct bounding box and class probability predictions. Known for its speed and accuracy, YOLOv3 finds applications in surveillance, autonomous driving, and general object recognition.

3.1 YOLOv3 Architecture

YOLOv3, a deep learning model, facilitates real-time object detection in images and videos. It employs convolutional layers to extract features, predicting bounding boxes and class probabilities for each object in an image. The architecture includes a feature pyramid network and a prediction module for detecting objects at different scales, emphasizing speed and accuracy for applications like self-driving cars, surveillance, and robotics.

3.2 YOLOv3 in OpenCV

In OpenCV, YOLOv3 utilizes deep neural networks with a single-pass approach, ensuring high-accuracy object detection. Its applications span surveillance, autonomous driving, and general object recognition, showcasing its versatility and effectiveness.

Figure 1. Object Detection usingle single object

3.3 MobileNet SSD in OpenCV

MobileNet SSD seamlessly integrates the MobileNet architecture with the Single Shot MultiBox Detector (SSD) framework, finely tuned for efficient real-time object detection. OpenCV provides pre-trained MobileNet SSD models, pretrained on datasets like COCO or VOC, facilitating object detection tasks without requiring extensive training. Recognized for its efficiency, MobileNet SSD in OpenCV is especially well-suited for resource-constrained devices like mobile phones or embedded systems [14-15]. It delivers swift inference times while maintaining commendable detection accuracy, making it ideal for real-time applications. In OpenCV, you can harness the library's functionalities to seamlessly incorporate YOLOv3 MobileNet SSD and YOLOv4 Tiny for object detection tasks. This encompasses loading pre-trained models, conducting inference on images or videos, and visualizing detected objects through bounding boxes and labels. Both YOLOv3 MobileNet SSD and YOLOv4 Tiny in OpenCV provide robust capabilities for object detection, and the choice between them hinges on specific requirements such as accuracy, speed, and deployment constraints [16-17]. Evaluating and benchmarking these algorithms on your particular use case is recommended to determine the most suitable option.

Table 1. Ensuring the Test case is working correctly

No.	Test case/Test script	Attribute and value	Expected result	Result
1.	Start the project and run the project	No errors generated	Successfully run the project main page display	Pass
2.	Take the real-time object detection images from data set	Click on the upload button	Successfully image uploaded	Pass forward phase
3.	Startup of real-time object detection of various images	Image should high quality	Detect the accurate detection location	Pass

Whereas in the table it shows how the algorithm shows the performance accurately.

Table 2. Ensuring algorithm show the performance accurately.

No.	Test case/Test script	Attribute and value	Expected result	Result
1.	Take the high-quality pixels images from the given dataset	Identifying the quality features of the images	Successfully detect the real-time image location	Pass
2.	Check the algorithms accuracy	Compare the accuracy rate through confusion matrix	Successfully algorithm identify the objects	Pass

Now to compare the YOLOv3, SSD, YOLOv4 detection algorithms performances in table 3.

Table 3. Ensuring each algorithms accuracy rate and its working performance on dataset images using real-time object detection.

No.	Test case/Test script	Attribute and value	Expected result	Result
1.	Check the data feature engineering techniques implemented	Remove the missing, unnecessary values form the dataset	The successfully implemented data cleaning process	Pass

2.	Detect the accuracy rate of each algorithm on various images	Calculate the accuracy rate	Apply confusion matrix techniques	Pass

Table 4. To provide the accurate results desired to the test case.

No.	Test case/Test script	Attribute and value	Expected result	Result
1.	Check the object detection abilities	Detect accurate location of the detection	Successfully detect the region of image recognition	Pass
2.	Prepare the test data	Dataset should include images & videos	Successfully added various images for the detection	Pass

Finally, the Integration testing is a software testing technique that focuses on testing the interactions and communication between different components or modules of a software system The goal of integration testing is to verify that the integrated components work together as intended, ensuring that the system functions correctly as a whole.

Table 5. To Identify the Errors

No.	Test case/Test script	Attribute and value	Expected result	Result
1.	Check the errors during the project execution time	Type of errors	Identifying the errors successfully	Pass
2.	Data Flow and Dependences	Verifies the correct data flow of data	It ensures that the data is transmitted, transformed, and received accurately across different components	Pass
3.	Compatibility and system behavior	Assesses the overall system behavior, performance and compatibility	It ensures that the integrated system meets the specified requirements operates efficiently	Pass

Figure 2. Object detection of two objects

YOLOv4 tiny is a state-of-the-art object detection model that has gained significant attention for its exceptional speed and accuracy. It is a compact version of YOLOv3, designed specifically for resource-constrained environments and edge devices with limited computational power. YOLOv3 builds upon the success of its predecessors, incorporating advanced techniques and architectural improvements to achieve remarkable performance. The model leverages a combination of deep convolutional neural networks and anchor-based object detection, enabling it to detect and localize objects with impressive precision in real time. YOLOv3 strikes a fine balance between speed and accuracy, making it ideal for applications that require real-time object detection on devices with limited resources, such as drones, smartphones, and embedded systems. Its compact size and efficient architecture have opened up new possibilities for deploying object detection systems in a wide range of scenarios, including surveillance, robotics, and autonomous vehicles.

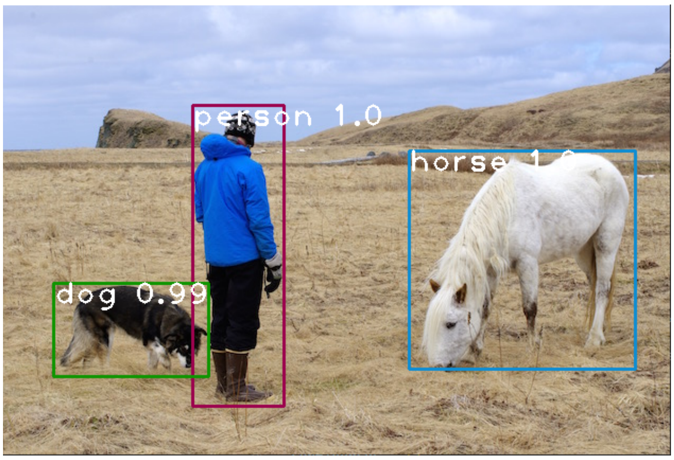

Figure 3. Multiple Object Detections

The YOLO algorithm setup the detection network is an end-to-end, one-step structure which was developed. The large number of labeled images target the bounding and detected the localize area of the images. Bounding images prediction and probabilities categorical obtained the relatively fast box (BB). The fast box connected probability become categorical prediction directly obtained. The confusion matrix is the measurement of the mean, average and precision show the performance through various algorithms and methods performing. Sometime variations, scales and other issues determined through the confusion matrix adapts multiple convolution layers for multi-scale object detection.

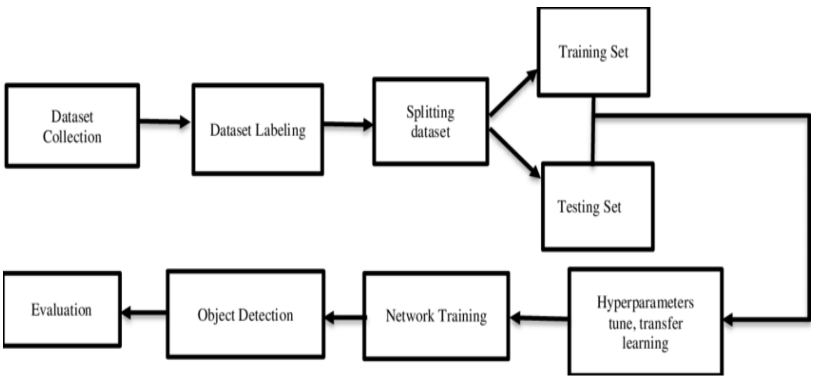

Figure 4. Real Time Object Detection Flow Diagram

4. Performance Analysis of YOLOv3

To verifying the efficiency and reliabilities of YOLOv3 detection classifiers in real-time detection for real images of objects with a small feature of dataset, this paper collected from another experiment with the same sample dataset and the SSD algorithm.

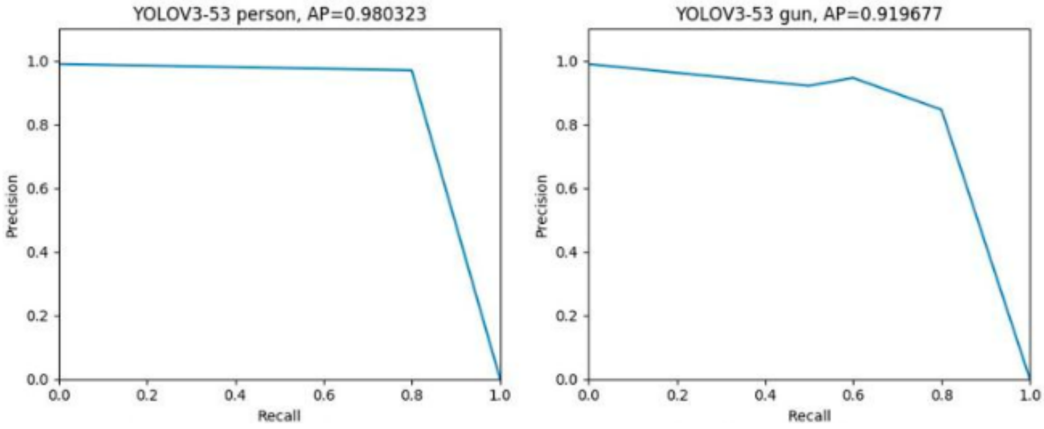

Figure 5. Object Detection Performance with 53 persons.

The figure shown average and high accuracy rates in terms of precision vs recall compared with given algorithms performance.

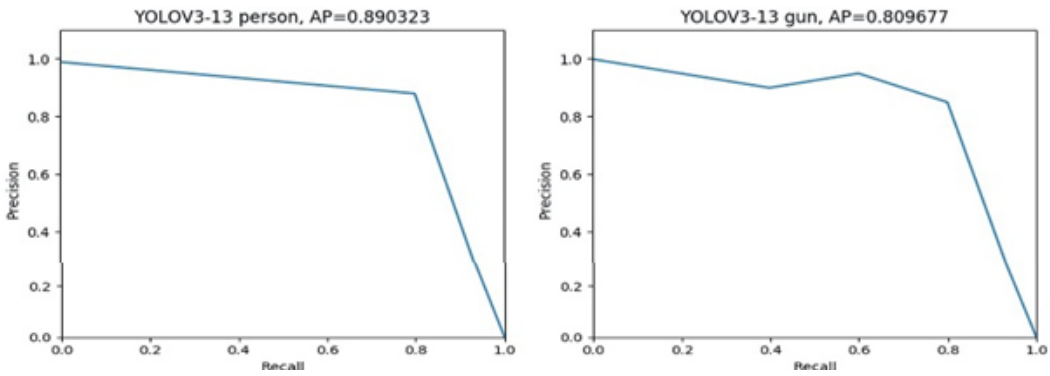

Figure 6: Object Detection Performance with 13 persons.

With the help of algorithms, we can easily analyze and predicted the performance efficacy. The classifiers extracting meaningful features from the images to improve the accuracy through confusion matrix approaches.

YOLOv3 in OpenCV offer powerful capabilities for object detection, and the choice between them depends on specific requirements such as accuracy, speed, and deployment constraints. It is recommended to evaluate and benchmark the algorithms on your specific use case to determine the most suitable option.

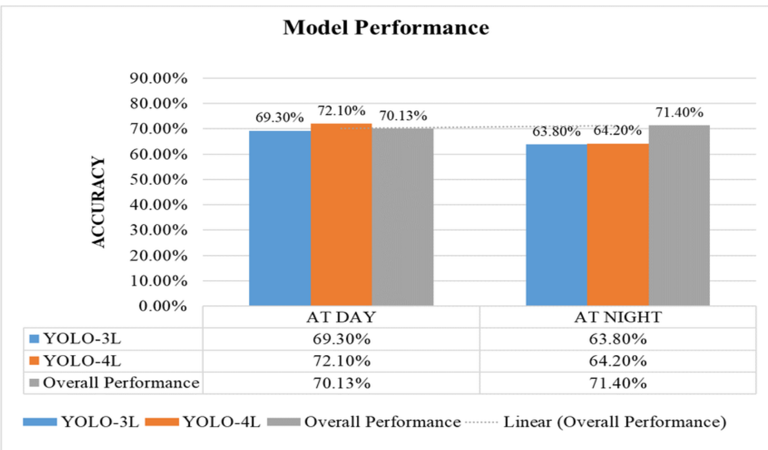

Figure 7. Comparing the performance of YOLO3L and YOLO4L across all executions based on processing time.

Figure 7 illustrates the performance range of YOLO-3L and YOLO-4L, showcasing their effectiveness during both daytime and nighttime scenarios. To enhance processing time, there is a need for improvements through compressed design. Notably, YOLO-4L excels with an accuracy of 72.10% during the day and 64.20% at night. Likewise, for YOLO-3L its day time performance is 69.30% will night time performance is 63.80% respectivel. The overall range of performance for both the YOLO models are recorded and depicted in the figure 7. Nevertheless, there seem to be some conflicts in which RGB images perform better during daytime while thermal images perform well during night time.

5. Conclusion & Future Work

In conclusion, the utilization of YOLOv3 and MobileNet SSD for real-time object detection in OpenCV proves to be a robust and prevalent approach, striking a balance between accuracy and speed for diverse use cases. Our comparative performance considered accuracy, speed, and suitability for different applications, leading to a well-informed solution that integrates both YOLOv3 and MobileNet SSD in OpenCV, showcasing their potential across various scenarios. Looking ahead, performance optimization remains a focal point, even for efficient algorithms like YOLOv3 and MobileNet SSD. Exploring avenues such as hardware acceleration, utilizing GPUs, or specialized hardware like Tensor Processing Units (TPUs), could further enhance processing speed. Additionally, the integration of object tracking techniques promises to elevate the capabilities of real-time object detection by providing continuous tracking across frames. Further optimization of the processing time of YOLOv3 and MobileNet SSD algorithms is recommended, particularly through compressed design, to improve their efficiency and reliability for real-time object detection. Evaluating and benchmarking the algorithms on specific use cases is recommended to determine the most suitable option based on requirements such as accuracy, speed, and deployment constraints. Exploring the performance and capabilities of YOLOv5, a lighter version of YOLOv4, could be a potential future direction for object detection, considering its potential improvements in accuracy and speed. Investigating the performance of YOLOv3 and MobileNet SSD algorithms on different types of images, such as RGB images during daytime and thermal images during nighttime, could provide insights into their adaptability to different environmental conditions. Considering the ethical implications of using real-time object detection algorithms in surveillance systems, autonomous vehicles, and robotics is an important area for future research.

References

1. Mao, Qi-Chao, Hong-Mei Sun, Yan-Bo Liu, and Rui-Sheng Jia. "Mini-YOLOv3: real-time object detector for embedded applications." Ieee Access 7 (2019): 133529-133538.
2. Masurekar, Omkar, Omkar Jadhav, Prateek Kulkarni, and Shubham Patil. "Real time object detection using YOLOv3." International Research Journal of Engineering and Technology (IRJET) 7, no. 03 (2020): 3764-3768.
3. Gai, Wendong, Yakun Liu, Jing Zhang, and Gang Jing. "An improved Tiny YOLOv3 for real-time object detection." Systems Science & Control Engineering 9, no. 1 (2021): 314-321.
4. Zhang, Xiuling, Xiaopeng Dong, Qijun Wei, and Kaixuan Zhou. "Real-time object detection algorithm based on improved YOLOv3." Journal of electronic imaging 28, no. 5 (2019): 053022-053022.
5. Srithar, S., M. Priyadharsini, F. Margret Sharmila, and R. Rajan. "Yolov3 Supervised machine learning framework for real-time object detection and localization." In Journal of Physics: Conference Series, vol. 1916, no. 1, p. 012032. IOP Publishing, 2021.
6. Gunawan, Chichi Rizka, Nurdin Nurdin, and Fajriana Fajriana. "Design of A Real-Time Object Detection Prototype System with YOLOv3 (You Only Look Once)." International Journal of Engineering, Science and Information Technology 2, no. 3 (2022): 96-99.
7. Gong, Hua, Hui Li, Ke Xu, and Yong Zhang. "Object detection based on improved YOLOv3-tiny." In 2019 Chinese automation congress (CAC), pp. 3240-3245. IEEE, 2019.
8. Pang, Lei, Hui Liu, Yang Chen, and Jungang Miao. "Real-time concealed object detection from passive millimeter wave images based on the YOLOv3 algorithm." Sensors 20, no. 6 (2020): 1678.
9. Tan, Lu, Tianran Huangfu, Liyao Wu, and Wenying Chen. "Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification." BMC medical informatics and decision making 21 (2021): 1-11.
10. Chen, Zhihao, Redouane Khemmar, Benoit Decoux, Amphani Atahouet, and Jean-Yves Ertaud. "Real time object detection, tracking, and distance and motion estimation based on deep learning: Application to smart mobility." In 2019 Eighth International Conference on Emerging Security Technologies (EST), pp. 1-6. IEEE, 2019.
11. Shill, Apu, and Md Asifur Rahman. "Plant disease detection based on YOLOv3 and YOLOv4." In 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), pp. 1-6. IEEE, 2021.
12. Nepal, Upesh, and Hossein Eslamiat. "Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs." Sensors 22, no. 2 (2022): 464.
13. Khan, Asjad M. "Vehicle and pedestrian detection using YOLOv3 and YOLOv4 for self-driving cars." PhD diss., California State University San Marcos, 2021.
14. Feroz, Md Alamin, Marjia Sultana, Md Rakib Hasan, Aditi Sarker, Partha Chakraborty, and Tanupriya Choudhury. "Object detection and classification from a real-time video using SSD and YOLO models." In Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2021, pp. 37-47. Springer Singapore, 2022.
15. Sindhwani, Nidhi, Shekhar Verma, Tushar Bajaj, and Rohit Anand. "Comparative analysis of intelligent driving and safety assistance systems using YOLO and SSD model of deep learning." International Journal of Information System Modeling and Design (IJISMD) 12, no. 1 (2021): 131-146.
16. Heda, Lokesh, and Parul Sahare. "Performance Evaluation of Yolov3, Yolov4 and Yolov5 for Real-Time Human Detection." 2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS), 2023. https://doi.org/10.1109/pcems58491.2023.10136081.
17. Sozzi, Marco, Silvia Cantalamessa, Alessia Cogato, Ahmed Kayad, and Francesco Marinello. "Automatic Bunch Detection in White Grape Varieties Using Yolov3, Yolov4, and Yolov5 Deep Learning Algorithms." Agronomy 12, no. 2 (2022): 319. https://doi.org/10.3390/agronomy12020319.