

Performance Analysis of YOLOv3, YOLOv4 and MobileNet SSD for Real Time Object Detection

Shahab Ul Islam^{*}, Giampaolo Ferraioli[†], Vito Pascazio[‡], Sergio Vitale[§], Muhammad Amin^{**}

Abstract

Real time identification of specific objects by machine learning approach has exhibited excellent outcomes; however, in the actual databases real-life images are prone to in-focus, unnecessary, blurred, noisy, shaky and jittery images. These problems create a great deal of wiggle room in success evaluation of real-time object recognition algorithms. To address these issues, object recognition mainly comprises applying the technique of You Only Look Once (YOLO). The unique feature of YOLO is the possibility of identifying all objects in the image. Other algorithms, unfortunately, are incapable of coping with the complexity of the whole images. Fast object detection is a key feature of YOLO classifiers since they are able to locate the best boxes and provide probabilities for each box in a very short time, surpassing the speed of other algorithms. YOLO localizes objects on pictures with its high level of precision. This study provides the real-time performance analysis of YOLOv3, YOLOv4 and MobileNet SSD for object detection. In a recent experiment, different object detection models were compared for their accuracy and speed. YOLOv4 boasted the highest precision (99.50%) and recall (98%) but took the longest to process an image (1.91 seconds). YOLOv3 offered a good balance, achieving a mean average precision (mAP) of 96.5% and recall of 94.6% with a faster inference time of 0.345 seconds. MobileNet SSD prioritized speed, delivering the quickest inference time (0.145 seconds) but with the lowest accuracy (precision: 91.4%, recall: 88.6%). This highlights the trade-off between accuracy and speed in object detection. YOLOv4 prioritizes accuracy, YOLOv3 offers a balance, and MobileNet SSD prioritizes speed.

Keywords: Object Detection; YOLOv3; YOLOv4; MobileNet SSD; Deep Learning.

Introduction

Humans can instantly recognize the features, locations, and interactions of images with just the naked eye. The visual systems of

^{*}Corresponding Author: Department of Engineering, Parthenope University of Naples, Naples 80138, Italy, shahabul.islam001@studenti.uniparthenope.it

[†]Department of Science and Technology, Parthenope University of Naples, Naples 80138, Italy, giampaolo.ferraioli@uniparthenope.it

[‡]Department of Engineering, Parthenope University of Naples, Naples 80138, Italy, vito.pascazio@uniparthenope.it

[§]Department of Engineering, Parthenope University of Naples, Naples 80138, Italy, sergio.vitale@uniparthenope.it

^{**}Department of Electronics, University of Peshawar, Peshawar 25120, Pakistan, mamin@uop.edu.pk

human are efficient and accurate, allowing to perform complex tasks like driving with minimal effort. The development of faster and more accurate object recognition systems are revolutionizing self-driving cars that won't need fancy sensors anymore. Improved object recognition transforms the assistive devices that gives real-time visual cues to users. Due to the advancements in object recognition system, robots become more adaptable and effective, leading to more open-ended designs (Nawaz et al., 2023). A classifier is used for detection of an object and test it at different points and scales in the test image.

The object detection through YOLOv3, YOLOv4-tiny and MobileNet SSD algorithms is incredibly efficient (Pang et al., 2020). It utilizes a single convolutional network to simultaneously predict the locations and categories of objects within an image. This approach stands in stark contrast to traditional detection systems that employ multiple classifiers, making YOLO significantly faster and more computationally efficient. OpenCV library has exciting and powerful techniques to perform the object detection and localization.

This trendy open-source library used for computer vision provides an easy interface for these algorithms and leverage their capabilities for the object detection applications as discussed in Diwan et al (2023). You Only Look Once Version 3 (YOLO) classifiers acknowledged for its speed and accuracy. The real time object detection dividing and predicting bounding boxes, classes, grid and probabilities for each grid cell. YOLOv3 further improves the architectural changes and enhancements by predecessors (Vijayakumar et al., 2024).

YOLOv4 is another version that achieves remarkable speed without compromising much on detection accuracy, making it an excellent choice for real-time object recognition on edge devices and platforms with limited computational power (Quach et al., 2023). 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 maintains competitive accuracy and achieves impressive inference speed, making it suitable for real-time applications on edge devices (Ragab et al., 2024).

On the other hand, MobileNet SSD (Single Shot MultiBox Detector) 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 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 (Tan et al., 2021).

This research study provides a performance analysis of YOLOv3, YOLOv4 and MobileNet SSD deep pre-trained networks in terms of accuracy and inference time. This analysis is helpful for research community to easily select required algorithm for implementation of real time detection on edge devices.

Literature Review

It is very important to acknowledge and build upon previous research in real time object detection to ensure progress in the field. Image processing, machine learning, and deep learning techniques play a significant role in enabling accurate object detection and localization. There is extant research which recommends some useful techniques for object detection. To better encompass the understanding from the extant literature on object detection, this study starts with a review of studies and techniques on real time object detection with high accuracy.

There are so many methods invented by researchers to address the object detection task; this task is fundamental in computer vision. Before presenting the proposed object detection technique, this work surveys prior work on the topic. Thus, the goal of such analysis is to understand how objects are currently being recognized with the help of the mentioned techniques. Real-time object detection is one of those problems that are somewhat easy to describe, yet rather difficult to solve in computer vision. It includes identification as well as categorization of objects that are present in real-time video streams or feeds from cameras. In recent years, two star players have emerged: These include YOLOv3 which is famous for one-shot-detection and MobileNet SSD which is a single-shot detector. Real-time object detection is another field in which both have made great progress.

Researchers have been comparing different object detection methods for a while now, especially for specific situations. The problem is, new algorithms are popping up all the time, and it's hard to keep up! There's a big gap in research that compares these latest advancements, like YOLOv3 and YOLOv4, to the older methods. This is especially true when it comes to figuring out which ones are best at spotting multiple objects in an image or video.

Salim et al. (2023) describe YOLO algorithm in the context of object detection and the usage of YOLO classifiers for prediction. The authors describe how the best methods derived from the CNN model and the YOLOv3 algorithm increase the speed and accuracy.

As for YOLO's newer generations, it has been demonstrated in the literature that they are indeed capable of object detection; nevertheless, there is a lack of studies on road object detection particularly utilizing these models. Existing works including Kumar et al. (2020) object detection proposal for surveillance cameras also have certain unexplained aspects. Their approaches are only partially described, and there are no clear guidelines for assessing the effectiveness of the results; that is why it is difficult to draw certain conclusions.

However, there's good news for the newer YOLO models. Studies comparing their performance in various object detection tasks, such as Rahman et al. (2020) research on insulator detection using YOLOv4 has shown promising results. Interestingly, in this particular case, YOLOv4 even outperformed YOLOv5. This suggests that the new YOLO models have significant potential. Nevertheless, further research specifically focused on road object detection is still necessary.

Zaidi et al. (2022) describes a system that acts like a super-powered party scanner. It uses a special technique to spot objects in real-time, like a friends at the party.

Khan et al. (2023) tackled a similar problem to improve locating objects in an image, instead of just saying "there's something in the image." Their approach involved a special technique called "structured output regression" which basically helps the computer predict the exact location and size of the object.

For self-driving cars to route safely, they must continuously identify and track their surroundings, including other vehicles, pedestrians, and obstacles. That's why real-time object detection is essential. Algorithms like YOLOv3 and MobileNet SSD excel in this area, providing the ability to detect objects in real-time. With this information, autonomous vehicles can make informed decisions, ensuring a smooth and safe journey (Masurekar et al., 2020).

Existing studies have made valuable contributions to the deep learning model for real time detection. This work builds on past research by evaluating the performance of YOLOv3, YOLOv4 and MobileNet SSD regarding accuracy and time and build an efficient system that can detect a real time object.

Methodology

This section provides an in-depth methodology of real time object detection. Figure 1 shows the different stages of the object detection process, such as data collection, image preprocessing, model training, model testing and evaluation.

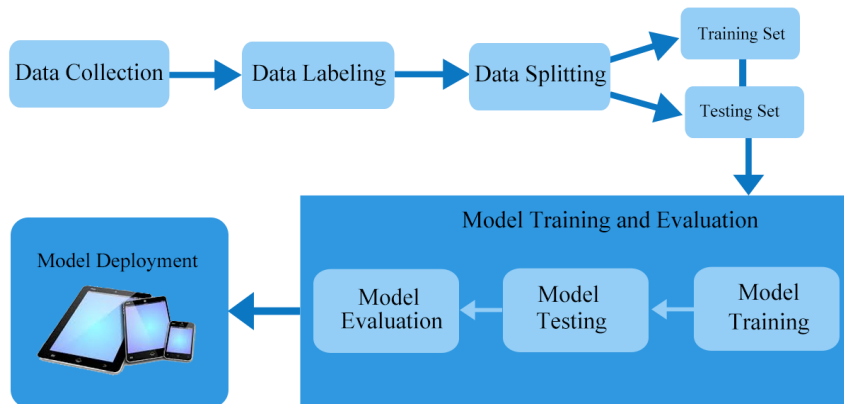


Figure 1: The Proposed Research Methodology.

Object Detection Method

This method divides one image into an $S \times S$ grid. For each grid cell, B bounding boxes with confidence are predicted. Figure 1 shows the object detection model in which the model is trained on the COCO dataset which is a large-scale dataset specifically designed for tasks related to computer vision, particularly in object detection, image segmentation, and image captioning. The data is split into training and testing set. After training and testing the model is implemented on the edge device.

YOLOv3 Network Architecture

YOLOv3 is one of such advantages that operates in real time and is incredibly accurate in object recognition and tagging. In OpenCV one can use YOLOv3 models that are pre-trained and stored in the library's listing. A model can be trained by feeding it with a custom datasets that researchers selected and prepared according to their needs. OpenCV also provides an API for using already trained YOLOv3 models and creating object recognition functions, either image or video streams. YOLOv3 with OpenCV comes to the fore with deep neural networks as it may be operated by a one-pass method that does not regress but predicts the bounding boxes and class probabilities directly. It can perceive objects ranging from tiny to hours in exactitude and also finds its use in areas covering surveillance, driverless cars, and object detection simply. YOLOv3 is an Object tracker and identifies object segments in the images or video frames (Gunawan et al., 2022). It stresses the recognition of the whole image at one go before bounding boxes and class possibilities are predicted straight away. YOLOv3 employs deep neural network architecture that consists of varieties of layers on top of each other in order to detect and analyze objects properly at different scales (Gai et al., 2023).

YOLOv3 tessellates the initial image into a grid and supplies a bounding box estimate that includes class probabilities for every grid cell. The performance of YOLOv3 in particulate and complex background detection is high. Particle detection is a challenge for object detecting models. Therefore YOLOv3 is the best solution. It detects a wide variety of different objects and is most preferred for such jobs as surveillance, autonomous driving, object recognition and many more (Ji et al., 2023).

YOLO V4 Network Architecture

The YOLOv4-tiny is a model of an object detection that is in the state of art and has been the favorite of several experts due to its high speed and accuracy (Jiang et al., 2020). YOLOv4 is a small-scale version, tailored to realize the objectives of YOLOv4 for resource-limited scenes, and edge devices with minimal computing power. YOLOv4-tiny is an architectural heir from the previously released YOLOv4, thereby using advanced techniques and designing to outperform its ancestors. The designed model is a mix of deep convolutional neural networks and anchor-based object detection approach that will locate and highlight the objects with a high level of accuracy in just seconds. YOLOv4-tiny offers a wise trade-off between accuracy and speed that is appropriate for scenarios that demand real-time detection of objects on devices that are incumbent on limited resources, such as drones, smart-phones, and embedded systems (Nepal et al., 2022). Its small size and the possibility of easy incorporation into different types of equipment such as security systems, robots and vehicles of the future make it a more versatile device.

MobileNet SSD in Architecture

MobileNet SSD implementation combines MobileNet architecture and SSD framework to provide real-time detection of objects with no need of setting thresholds and anchors. Open CV has a bunch of pre-trained MobileNet SSD models which have been trained on large volumes of data like VOC or COCO (Feroz et al., 2022). It encompasses loading pre-trained models, visualizing detected objects through bounding boxes and labels and showcasing captured images or videos. YOLOv3 MobileNet SSD, and yolov4 tiny on the OpenCV platform offer good and useful mechanisms for object detection (Younis et al., 2020). The conundrum lies in which one to choose for applications that have conflicting requirements such as accuracy, speed, and implement on constraints. The evaluation and inadequacy in terms of your particular case will help in determining the best option of algorithm.

Results and Discussion

The section explains the expertness and possibility of algorithm to adapt to different situations by examining different metrics. Accuracy and relevance are two indicators employed to determine the effectiveness of the algorithm. Alongside the accuracy scoring, we have evaluated the algorithm's inference time for its efficiency.

The inference time is the total time it takes the algorithm to detect objects in an image. All the algorithms were used to detect different object and their accuracy were then compared based on the above discussed metrics. Figure 2 shows chair detection using YOLOv3 with the accuracy of 0.98.



Figure 2: Chair Detection Using Yolo V3.



Figure 3: Bike Detection Using Yolo V4.

Figure 3 shows the accuracy of the YOLOv4 which is 0.99, while Figure 4 shows the image of clock detected using MobileNet SSD with the accuracy of 0.94.

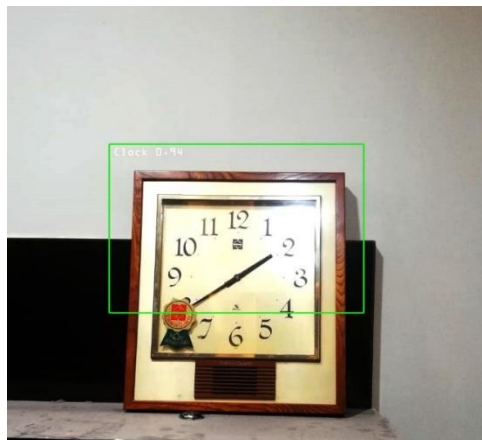











Figure 4: Clock Detection Using MobileNet SSD.

Efficiency Analysis

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. The table 1 shows the performance through precision, recall, and accuracy carvers of YOLOv3, YOLOv4 tiny and MobileNet SSD algorithms. The table shown average and high accuracy rate compared with given algorithms performance. With the help of algorithms, we can easily analyze and predicted the performance efficacy. With object detection test YOLOv3 and YOLOv4 both gives good accuracy but yolo takes more time in detecting object than MobileNet as shown in Table 1. The classifiers extracting meaningful features from the images to improve the accuracy through confusion matrix approaches.

Figure 5 and figure 6 shows the performance of YOLOv3, YOLOv4 and MobileNet SSD under both daytime and nighttime conditions. To improve processing speed, further optimizations are needed through compressed design. YOLOv3L, on the other hand, exhibits daytime performance of 96.5% and night-time performance of 92.5%. YOLOv4 demonstrates remarkable accuracy, achieving 98.5% during the day and 94.2% at night. MobileNet SSD also demonstrates a remarkable accuracy, achieving 86.4% during the day and 83.7% at night. The overall performance range for both YOLO models is illustrated in the figure. However, there appears to be a discrepancy between RGB images performing better during daytime and thermal images performing better at night.

Table 1 Performance of Algorithms with Single Object

Input Image	Yolo V3	Yolo V4	MobileNet SSD
a			
Precision%	96.50	98.50	86.44
Recall %	94.98	97.98	84.90
Inference Time	0.345 Sec	1.80 Sec	0.145Sec
Input Image	Yolo V3	Yolo V4	MobileNet SSD
b			
Precision%	96.30	99.50	91.44
Recall %	94.68	98.98	89.90
Inference Time	0.245 Sec	1.91 Sec	0.165Sec
Input Image	Yolo V3	Yolo V4	MobileNet SSD
C			
Precision%	96.50	99.00	88.64
Recall %	94.98	98.98	86.70
Inference Time	0.485 Sec	0.876 Sec	0.195Sec

MobileNet SSD prioritizes real-time performance on devices with limited resources by using a lightweight MobileNet architecture and the SSD framework for multi-scale object detection. YOLOv3 strikes a balance between high accuracy and processing speed. It employs a single-shot detection approach with deep neural networks for multi-scale object detection, making it efficient for complex scenes. YOLOv4 focuses on improving YOLOv3's performance while maintaining its real-time capabilities. It introduces features like "bag of freebies" (data augmentation techniques) and improved feature aggregation. Comparing with other work both MobileNet SSD and YOLO prioritize speed over absolute accuracy compared to some object detection algorithms that utilize complex architectures (e.g., Faster R-CNN). However, they still achieve good accuracy while being faster for real-time applications. Compared to research exploring object detection for a broad range of categories, MobileNet SSD and YOLOv3/YOLOv4 have shown promise in detecting object in real time environment. MobileNet SSD excels in this area due to its lightweight design.

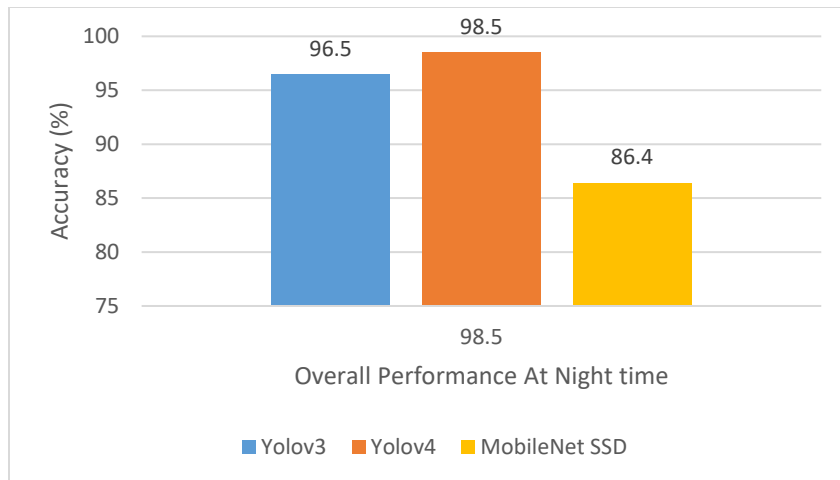


Figure 5: Overall Performance during Day

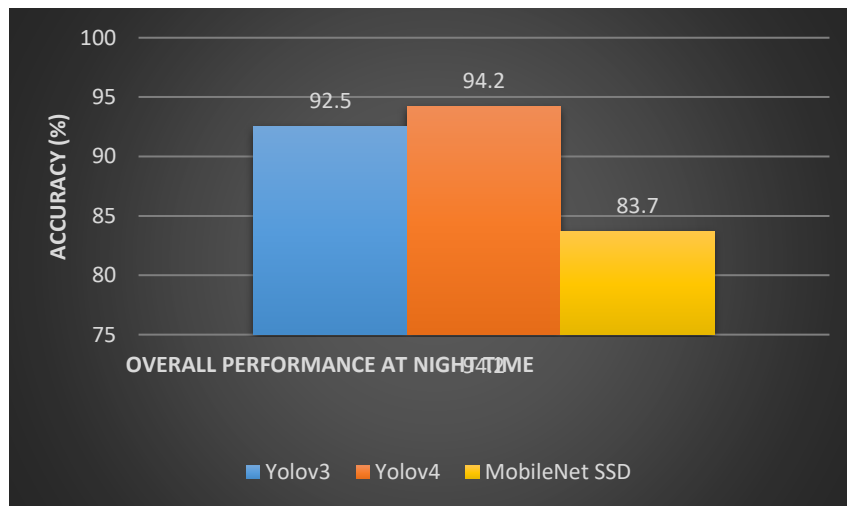


Figure 6: Overall Performance during Day

While YOLOv3 and YOLOv4 Tiny shows promise, existing research on YOLO for object detection in video, like (Nepal et al., 2022) work, lacks clarity in methodology and evaluation metrics, making it difficult to draw definitive conclusions (Feroz et al., 2022). More research is needed in this area. There's a lack of research comparing the latest YOLO models (YOLOv4) to their predecessors, particularly for detecting multiple objects in various environments (Feroz et al., 2022). MobileNet SSD, YOLOv3, and YOLOv4 represent valuable advancements in real-time object detection. They offer different strengths depending on the

application's needs (speed vs. accuracy, resource constraints). Further research is crucial, particularly for YOLO in real time object detection and comparing new YOLO generations to their predecessors.

Conclusion and Future Work

In this study, we implemented and proposed a solution for object detection using YOLOv3, YOLOv4 and MobileNet SSD classifiers. The comparative study shows that the YOLO classifiers have shown the best performance among other algorithms for object detection. An algorithm for the detection of items in pictures and videos can be applied to obtain a solution for multiple problems in different domains. For instance, this autonomous system can be implemented to automatically generate vehicle traffic lanes in order to ensure safety, detect security threats or spot visually evoking people with audio feedback. In this paper, we presented stimulation of multiple objects model that can detect in real time. The YOLOv4 variant can both achieve the top accuracy and work the way that is quick despite the fact of being effective. The reason for this is that the network relies on a forward pass and can figure out the objects of multiple scales simultaneously. The study used an insufficient data set which is the reason behind the unreliable results. A bigger data size will ensure the more precision in the activity detection systems. The study method was limited in that different types of detection methods in literature were not used. With different kind of algorithms being used it will aid in building the robustness of the detection systems. The best way to make sure that passive millimeter wave (PMMW) security requirements are met in the future would be to test them on more training databases of another type.

References

- Diwan, T., Anirudh, G., & Tembhurne, J. V. (2023). Object detection using YOLO: Challenges, architectural successors, datasets and applications. *multimedia Tools and Applications*, 82(6), 9243-9275.
- Feroz, M. A., Sultana, M., Hasan, M. R., Sarker, A., Chakraborty, P., & Choudhury, T. (2022). 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.
- Gai, R., Chen, N., & Yuan, H. (2023). A detection algorithm for cherry fruits based on the improved YOLO-v4 model. *Neural Computing and Applications*, 35(19), 13895-13906.
- Gunawan, C. R., Nurdin, N., & Fajriana, F. (2022). Design of A RealTime Object Detection Prototype System with YOLOv3 (You Only

- Look Once). *International Journal of Engineering, Science and Information Technology*, 2(3), 96-99.
- Ji, S. J., Ling, Q. H., & Han, F. (2023). An improved algorithm for small object detection based on YOLO v4 and multi-scale contextual information. *Computers and Electrical Engineering*, 105, 108490.
- Jiang, Z., Zhao, L., Li, S., & Jia, Y. (2020). Real-time object detection method based on improved YOLOv4-tiny. *arXiv preprint arXiv:2011.04244*.
- Khan, D., Waqas, M., Tahir, M., Islam, S. U., Amin, M., Ishtiaq, A., & Jan, L. (2023). Revolutionizing Real-Time Object Detection: YOLO and MobileNet SSD Integration. *Journal of Computing & Biomedical Informatics*, 6(01), 41-49.
- Kumar, C., & Punitha, R. (2020, August). Yolov3 and yolov4: Multiple object detection for surveillance applications. In *2020 Third international conference on smart systems and inventive technology (ICSSIT)* (pp. 1316-1321). IEEE.
- Masurekar, O., Jadhav, O., Kulkarni, P., & Patil, S. (2020). Real time object detection using YOLOv3. *International Research Journal of Engineering and Technology (IRJET)*, 7(03), 3764-3768.
- Muwardi, R., Permana, J. M. R., Gao, H., & Yunita, M. (2023). Human Object Detection for Real-Time Camera using Mobilenet-SSD. *Journal of Integrated and Advanced Engineering (JIAE)*, 3(2), 141-150.
- Nawaz, M., Khalil, M., & Shehzad, M. K. (2023). MIYOLO: Modification of Improved YOLO-v3. *IETE Journal of Research*, 69(11), 8036-8044.
- Nepal, U., & Eslamiat, H. (2022). Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs. *Sensors*, 22(2), 464.
- Pang, L., Liu, H., Chen, Y., & Miao, J. (2020). Real-time concealed object detection from passive millimeter wave images based on the YOLOv3 algorithm. *Sensors*, 20(6), 1678.
- Quach, L. D., Quoc, K. N., Quynh, A. N., & Ngoc, H. T. (2023). Evaluating the effectiveness of YOLO models in different sized object detection and feature-based classification of small objects. *Journal of Advances in Information Technology*, 14(5), 907-917.
- Rahman, E. U., Zhang, Y., Ahmad, S., Ahmad, H. I., & Jobaer, S. (2021). Autonomous vision-based primary distribution systems porcelain insulators inspection using UAVs. *Sensors*, 21(3), 974.
- Ragab, M. G., Abdulkader, S. J., Muneer, A., Alqushaibi, A., Sumiea, E. H., Qureshi, R., ... & Alhussian, H. (2024). A Comprehensive

- Systematic Review of YOLO for Medical Object Detection (2018 to 2023). IEEE Access.
- Salim, R., Wulandari, M., & Calvinus, Y. (2023, December). Weapon detection using SSD MobileNet V2 and SSD resnet 50. In AIP Conference Proceedings (Vol. 2680, No. 1). AIP Publishing.
- Sindhvani, N., Verma, S., Bajaj, T., & Anand, R. (2021). 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(1), 131-146.
- Srithar, S., Priyadharsini, M., Sharmila, F. M., & Rajan, R. (2021, May). 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.
- Tan, L., Huangfu, T., Wu, L., & Chen, W. (2021). Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification. *BMC medical informatics and decision making*, 21, 1-11.
- Vijayakumar, A., & Vairavasundaram, S. (2024). Yolo-based object detection models: A review and its applications. *Multimedia Tools and Applications*, 1-40.
- Younis, A., Shixin, L., Jn, S., & Hai, Z. (2020, January). Real-time object detection using pre-trained deep learning models MobileNet-SSD. In *Proceedings of 2020 6th International Conference on Computing and Data Engineering* (pp. 44-48)
- Zaidi, S. S. A., Ansari, M. S., Aslam, A., Kanwal, N., Asghar, M., & Lee, B. (2022). A survey of modern deep learning based object detection models. *Digital Signal Processing*, 126, 103514.