YOLO Architectures and their Applications – A Review

Bhargavi Vijendra Sangam[1]; Swetha V.[2]; Ananya T.[3]

[1]Department of CSE (Cyber Security) APS College of Engineering Bengaluru, India
ORCID: 0009-0007-8376-6018

[2]Department of CSE (AIML) APS College of Engineering Bengaluru, India
ORCID: 0009-0008-6347-1033

[3]Department of CSE (Cyber Security) APS College of Engineering Bengaluru, India
ORCID: 0009-0005-3683-4175

Publication Date: 2026/06/13

Abstract: Object detection has become one of the most critical fields of study in computer vision due to its broad application range including surveillance, health care, autonomous driving, agriculture, robotics, and industrial automation. YOLO is considered one of the most efficient object detection frameworks due to its unique feature of being able to localize and classify objects in one feedforward network pass. Starting with its inception in 2016, YOLO architecture underwent multiple changes to improve accuracy, inference speed, and efficiency. This review paper provides an overview of how YOLO architectures have evolved from YOLO v1 to YOLO v8. The advancements, strengths, weaknesses, and enhancements of each iteration are discussed in detail. In addition, some of the applications of YOLO for object detection are analyzed. The paper concludes with potential future challenges and trends for real-time object detection.

Keywords: *YOLO, Object Detection, Deep Learning, Computer Vision, Real-Time Detection, CNN.*

How to Cite: Bhargavi Vijendra Sangam; Swetha V.; Ananya T. (2026) YOLO Architectures and their Applications – A Review. *International Journal of Innovative Science and Research Technology*, 11(6), 215-218.
https://doi.org/10.38124/ijisrt/26jun198

I. INTRODUCTION

Object detection plays a crucial role in computer vision that involves recognition and localization of the object of interest in the image/video frame. With advancements in artificial intelligence and deep learning technologies, computers have become capable of making sense of visual data. The use cases of the object detection algorithms include autonomous vehicles, intelligent surveillance systems, medical image processing, agriculture, and quality control processes in industries. Early models of object detection algorithms such as R-CNN, Fast R-CNN, and Faster R-CNN demonstrated good results regarding the accuracy of detection but were limited by their inability to make decisions in real time because of computational limitations and high computational complexity of algorithms. The YOLO (You Only Look Once) approach was created by Redmon et al. in 2016 to address the problem described above. This algorithm approached the object detection as a single-stage regression process which combined localization and classification at once. It ensured high speed of decision-making. Various versions of the YOLO algorithm appeared after its inception, each containing some innovations in design. The objective of this paper is to review the evolution of the YOLO algorithm from YOLO v1 to YOLO v8.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun198

II. EVOLUTION OF YOLO ARCHITECTURES

A. YOLOv1 (2016)

The appearance of YOLOv1 became a breakthrough in the area of object detection. Unlike other algorithms that utilized multi-staged pipelines and used region proposals, YOLOv1 introduced a revolutionary approach, where object detection became an instance of regression problems. The algorithm proposed a single network which made predictions for class scores and bounding box coordinates for the whole image at once [1]. In particular, an image in YOLOv1 was segmented into the grid of cells which were responsible for predicting coordinates and confidence scores for the objects with centers located in these regions. Thus, YOLOv1 managed to make predictions rather effectively and at the same time process images in real time. While the approach in question had many benefits, the lack of locality resulted in some issues, such as the failure of the model to detect small objects. Moreover, it could not provide proper detection for objects placed too close to each other. Still, YOLOv1 served as the basis for further development of YOLO and demonstrated the efficiency of such an approach to object detection using deep learning techniques.

B. YOLO9000/YOLOv2(2017)

In turn, the limitations found in the initial YOLO version led to the creation of the new one, called YOLO9000, and better known under the name of YOLOv2. Several innovations were used in YOLO9000, namely batch normalization, anchor boxes, high-resolution classifier, and multi-scale training. These features helped significantly improve performance of the network [2]. While batch normalization provided bet-ter training stability and speed, anchor boxes helped with the precise detection of various-sized objects. Finally, multi-scale training increased flexibility in terms of the usage of YOLO9000 at various resolutions. Overall, the new model provided a better recall rate and localization quality compared to its predecessor. An essential feature of YOLO9000 was its capability of joint training on datasets for object detection and image classification. Thanks to this, the algorithm managed to classify more than 9 thousand types of objects while keeping its high efficiency. Therefore, YOLO9000 can be considered a better generalized version of previous models in the series.

C. YOLOv3 (2018)

Several architectural changes were implemented in YOLOv3 in order to increase accuracy without reducing computational efficiency. One of the key changes was using a new backbone network called Darknet-53 that used residual connections to enhance feature extraction and enable deeper network training [3]. In addition, YOLOv3 included a multi-scale prediction method, wherein object detection was done across three levels of features. Therefore, this version could detect small, medium, and large objects, solving one of the major problems of the earlier versions, especially those associated with small target detection.

Moreover, logistic regression was used in this version for predicting objectness, which ensured more stable training results and higher accuracy. In terms of its performance, the YOLOv3 architecture found broad use in traffic monitoring systems, surveillance cameras, robots, and industrial automation systems because of its efficient balancing of speed and precision.

D. YOLOv4 (2020)

The creation of YOLOv4 had the aim of balancing both object detection accuracy and speed in order to become applicable to practical use in conventional hardware. Among other things, it employed the network CSPDarknet53 for feature extraction as the backbone network while increasing efficiency and reducing the number of calculations [4].

In order to boost object detection performance, it also included such innovations as Spatial Pyramid Pooling (SPP) and Path Aggregation Network (PANet). They allowed for a better integration and usage of features at multiple scales which significantly increased the accuracy of the localization of both large and small objects.

Moreover, YOLOv4 also used some sophisticated training techniques such as Mosaic data augmentation, Mish activation functions, DropBlock regularization, and Complete Intersection over Union (CIoU) loss function. All those elements helped achieve remarkable results in terms of accuracy, generalizability, and robustness, thus allowing it to be widely adopted in object detection.

E. YOLOv7 (2022)

One more advance in the development of YOLO framework happened in 2022 with the inclusion of a set of optimization techniques called trainable bag-of-freebies. Such an innovation helped to increase accuracy even further without adding to the cost of inference [5].

Apart from that, YOLOv7 included advanced feature aggregation techniques, more efficient model scaling algorithms, and label assignment methods which improved its accuracy further. Thus, YOLOv7 became one of the most accurate frameworks for real-time object detection and could success-fully operate in various situations.

It managed to achieve the best results in accuracy among the most common real-time object detectors when compared on benchmark datasets. Its good ratio of speed to accuracy allowed it to be used effectively not only in high-end hardware but also on edge devices which significantly widened its scope of applications.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jun198

F. YOLOv8 (2023)

Another improvement among many in the YOLO object detection family is YOLOv8 that uses the latest architectural advances in order to boost its performance. One of the key differences of this version is its reliance on an anchor-free detector which significantly simplified and improved the detection process [6].

The usage of advanced feature extraction techniques and training strategies helped achieve good robustness in difficult conditions such as those involving small objects, occlusions, difficult backgrounds, and different lighting conditions. It also performs well in other computer vision tasks such as instance segmentation, object classification, and object pose estimation. Due to its good performance in multiple vision tasks, YOLOv8 can be successfully applied in such areas as agriculture, autonomous systems, healthcare, industrial automation, and surveillance. With its high accuracy and performance and due to its ease of use, it quickly became one of the most popular object detection frameworks.

III. COMPARATIVE ANALYSIS OF YOLO VERSIONS

Since YOLOv1 was introduced in 2016, there have been significant advancements made in the area of object detection. Each model strives to solve the weaknesses of its predecessor by improving detection, efficiency, and real-time capabilities. Since there are increasingly greater demands for efficient object detection algorithms, this is the key point that researchers strive to achieve.

As compared to other models before, YOLOv1 relied on an entirely new technique of object detection via an all-in-one solution. While this algorithm proved to be better with real-time operations, it faced difficulties with small object detection and localization. These problems were solved by YOLO9000 with anchor boxes, batch normalization, and multi-scale training.

An innovation of YOLOv3 included a deeper network (Darknet-53) and multi-scale prediction. They improved the detection rate greatly. Moreover, a number of improvements was made in YOLOv4, such as applying a new architecture (CSPDarknet53), using Spatial Pyramid Pooling (SPP), and introducing Path Aggregation Network (PANet). Other innovations involve using Mosaic data augmentation, CIoU loss function, and other strategies that help train a model.

Finally, more recent YOLO versions concentrated on optimizing performance and efficiency. For instance, trainable bag-of-freebies techniques are used in YOLOv7 to enhance detection capabilities while decreasing inference cost. In turn, YOLOv8 uses anchor-free detection and increased capability for extracting features which allows for good generalization and robustness.

Main features, contributions, and drawbacks of various architectures are compared in Table I.

Table 1 Yolo Comparison Table

Version	Year	Contributions	Drawbacks
YOLOv1	2016	First implementation of a real-time single-stage object detector	Ineffective on small and closely localized targets
YOLO9000	2017	Application of anchor boxes, batch normalization, and multi-scale training	Localization errors in complicated environments
YOLOv3	2018	Applying Darknet-53 backbone network and residual connections	Greater computational costs compared to YOLOv2
YOLOv4	2020	Use of CSPDarknet53, SPP, and PANet; im-plementation of mo-saic data augmenta-tion, CIoU loss, etc.	Complex architectural and training procedures
YOLOv7	2022	Using trainable bag-of-freebies and opti-mized scaling methods	Dependence on tuning of hyperparameters
YOLOv8	2023	Employing anchor-free detection and feature extraction, and detection of multiple targets	Lack of benchmarking results due to recent development

In this study, YOLOv8 model (You Only Look Once, version 8) is used as the main computer vision model for automatic, real-time pest detection. YOLOv8 takes live videos or still images acquired from a camera set up in the agricultural environment, and recognizes and classifies different pest types such as aphids, beetles, grasshoppers, and sawflies. [9] [10] Based on recognizing different pests and their count in the real-time environment, YOLOv8 helps determine when to deploy different defense mechanisms through Raspberry Pi 4B, including acoustic buzzer, LEDs, or even relay-controlled pesticides based on the pest level detected by the model. Through an efficient image recognition method and IoT-enabled hardware deployment, it becomes possible to implement an efficient pest management plan to control pests and avoid pesticide overuse.

V. CHALLENGES AND LIMITATIONS

Despite some significant advances in this regard, some problems still remain concerning YOLO models. Object detection within cluttered scenes remains a challenging problem. Problems of object occlusion, variations in lighting conditions, and object crowding may lead to incorrect predictions. More-over, deployment of such models on edge devices demands additional optimization considering their computation requirements.

It can be inferred from the comparison performed that with every new release, YOLO fixes some issues of its predecessors. After providing a foundation of architecture in its initial stage, YOLO addressed such issues as localization, extraction of features, detection of small objects, and problems in training. At the moment, modern architectures such as YOLOv7 and YOLOv8 offer a good balance between speed and detection performance.

IV. APPLICATIONS OF YOLO

- *Object Detection Applications*

YOLO algorithms have found extensive use in real-time object detection applications [7], [8] because of the tradeoff achieved by YOLO algorithms in terms of speed and accuracy. Applications in the realm of intelligent transportation include detection of vehicles, monitoring of the traffic, and pedestrian detection. With the capability of real-time processing of video streams, the system can come to quick conclusions in critical situations. A number of researchers have shown that advanced YOLO variants like YOLOv4, YOLOv7, and YOLOv8 are capable of providing high accuracy and speeds. YOLO-based object detection is extensively used in surveillance and security applications. Surveillance systems incorporate YOLO detectors for purposes of detecting crowds, monitoring intrusion, and abnormal behavior recognition. Over time, YOLO algorithms have greatly advanced in their capacity to detect tiny and partly hidden objects.

V. FUTURE RESEARCH DIRECTIONS

Future directions of research related to object detection with the use of YOLO architecture might include construction of energy-efficient architectures, usage of transformer models, application of self-supervised learning approaches, fusion of information provided by multiple sensors, and techniques in the sphere of explainable artificial intelligence.

VI. CONCLUSION

The YOLO framework has been an essential part of making progress in real-time object detection. From the YOLOv1 to the latest YOLO9000, YOLOv3, YOLOv4, YOLOv7, and YOLOv8 versions, many improvements have been made in terms of speed, accuracy, and efficiency. YOLO has proven its performance in various fields such as health care, surveil-lance, agriculture, robotics, and industry automation. With the progress being made in the field of research, YOLO will continue to play a pivotal role in computer vision applications.

REFERENCES

[1]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[2]. J. Redmon and A. Farhadi, “Yolo9000: Better, faster, stronger,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 7263–7271.

[3]. ——, “Yolov3: An incremental improvement,” *arXiv preprint arXiv:1804.02767*, 2018.

[4]. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “Yolov4: Op-timal speed and accuracy of object detection,” *arXiv preprint arXiv:2004.10934*, 2020.

[5]. C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, “Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 7464–7475.

[6]. D. Reis, J. Hong, J. Kupec, and A. Daoudi, “Real-time flying object detection with yolov8,” *arXiv preprint arXiv:2305.09972*, 2024.

[7]. S. fSathish and A. Sangeetha, “Real time object detection using yolov8,” *International Journal of Scientific Research & Engineering Trends*, vol. 11, no. 6, Nov-Dec 2025.

[8]. D. Reis, J. Kupec, J. Hong, and A. Daoudi, “Real-time flying object detection with yolov8,” *arXiv preprint arXiv:2305.09972*, 2024.

[9]. M. Pavan, N. SaiKiran, and B. Sangam, “Bio-inspired tri modal pest defense system with predictive analytics,” *KS School of Engineering and Management*, 2025.

[10]. K. Neha, A. Arya, and R. Singh, “Wheat disease detection using yolo and drone-captured images,” *Journal of Graphic Era University*, vol. 13, no. 02, pp. 411–438, 2025. [Online]. Available: https://doi.org/10.13052/jgeu0975-1416.1327