ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

# OBJECT DETECTION IN DEEP LEARNING USING YOLO - A SURVEY

**[1]Dr. W. Rose Varuna, [2]K.V. Sri Lakshmanan, [3]R. Preamkumar and [4]S. Vigneshkumar**

[1]Assistant Professor, Department of Information Technology, Bharathiar University, Coimbatore-641046

[2,3,4]M.sc Information Technology, Department of Information Technology, Bharathiar University, Coimbatore-641046

[1]rosevaruna@buc.edu.in, [2]srilakshmanankv@gmail.com, [3]preamkumar0304@gmail.com and [4]vigneshkumar02082002@gmail.com

## ABSTRACT

*This research paper thoroughly explores the YOLO (You Only Look Once) family of algorithms, a popular and efficient single-stage object detection framework. The review covers the evolution of YOLO from its beginning to the latest versions, analyzing the performance of each iteration. It emphasizes the diverse applications of YOLO, particularly in real-time object detection on embedded systems. The paper also discusses recent advancements in compression algorithms to optimize YOLO models and addresses the challenge of reducing model size for deployment on resource-constrained devices. Lastly, it outlines potential research directions for the YOLO family, including new architectural designs and training strategies. Overall, the paper serves as a valuable reference for researchers exploring YOLO and its role in object detection.*

*Keyword: Deep Learning, YOLO, Object detection, Real-time detection, Embedded system.*

## 1. INTRODCTION

The main task in computer vision is object detection, which involves identifying the things of interest in input images and correctly classifying each object that is found. Because of it have many uses and the latest developments in technology, object detection has become increasingly popular in the field of computer vision in recent years.

The system is a one-stage detector directly detecting targets without candidate boxes. One-stage detectors like YOLOv5 are efficient, though slightly less accurate than two-stage detectors. The article summarizes YOLO versions 1 to 8 for researchers' needs. It covers YOLO family architectures, image size, average precision (AP), Frames Per Second (FPS), and parameters. The survey also discusses model compression and future insights.

Deep learning is a form of machine learning that uses neural networks trained on large datasets to perform tasks without explicit programming. The "deep" in deep learning signifies the use of multiple layers in these networks to understand intricate patterns from input data. These networks consist of connected nodes in layers, with input and output layers managing data and predictions. During training, the network adjusts its parameters based on prediction errors. Deep learning excels in tasks such as image and speech recognition, Natural Language Processing, and pattern recognition, utilizing models like Convolutional Neural Networks (CNNs) for images and Recurrent Neural Networks (RNNs) for sequences.

The YOLO (You Only Look Once) algorithm in deep learning is a popular approach for object detection. Unlike traditional methods, YOLO divides an image into a grid and simultaneously predicts multiple bounding boxes and class probabilities for each grid cell in a single pass through the neural network. This allows YOLO to efficiently detect and classify multiple objects. The algorithm comprises a backbone neural network, often a convolutional neural network (CNN), and a detection layer that predicts bounding boxes and class probabilities. YOLO's real-time object detection capabilities have found applications in areas such as autonomous vehicles, surveillance, and image analysis. Various versions, from YOLOv1 to YOLOv8, have been developed, each bringing improvements in terms of accuracy and speed.

## 2. LITERATURE REVIEW

**Joseph Redmon et.al** 2015.[1] Object detection systems, like YOLO (You Only Look Once), aim to identify objects in images efficiently. YOLOv1 treats object detection as a single regression problem, predicting both bounding boxes and class probabilities in one go. Other approaches like Fast R-CNN and Faster R-CNN improve efficiency by using region of interest pooling and a Region Proposal Network (RPN) for end-to-end training. Selective Search and the Deformable Parts Model (DPM) offer alternative methods for generating region proposals. Hyper Net suggests using a hyper-network for proposal generation. Various studies compare the speed and accuracy of these methods, highlighting trade-offs. A comprehensive review discusses the strengths and weaknesses of deep learning-based object detection, including YOLO and its counterparts.

**Sakshi Gupta et al**.[2020],[2].Research on YOLOv2 focuses on enhancements for faster and more accurate object detection, such as "YOLO9000: Better, Faster, Stronger." Darknet framework, detailed in the "Darknet: Open Source Neural Networks in C" paper, supports YOLO implementations. Studies like "Speed/accuracy trade-offs for modern convolutional object detectors" by Huang et al. compare detection models, aiding in understanding their performance. Practical resources like the YOLOv2 GitHub repository and NVIDIA's CUDA Toolkit Documentation facilitate code implementation and GPU acceleration.

**Omkar Masurekar et al** 2020[3]. Object detection, vital in computer vision, aids tasks like self-driving cars and assisting the visually impaired. YOLO (You Only Look Once) stands out for real-time object identification, employing object localization. It categorizes into classification-based (e.g., CNN, RNN) and regression-based (e.g., YOLO), predicting classes and bounding boxes in a single pass for speed, despite potential localization errors. Challenges include real-world dataset testing with issues like blurred images, tackled by proposed YOLOv3 applications in security, traffic monitoring, and aiding the visually impaired through audio feedback, initially focusing on detecting three specific objects.

**Podakanti Satyajith Chary**. 2023[4] The review focuses on a real-time object detection system using YOLOv4, integrated into Google Colab for collaborative GPU-supported development. It draws configurations from the official repository, incorporates webcam functionality, and shares open-source code for transparency. Training data cover diverse scenarios, with the model undergoing transfer learning and fine-tuning. Performance evaluation includes standard metrics and real-time speed measurement, addressing challenges through hyperparameter tuning, finding applications in surveillance, autonomous vehicles, and human-computer interaction, proving effective and accurate in dynamic scenarios, advancing computer vision.

**Manikandan et.al** 2023. [5] The literature discusses how deep learning, particularly the YOLO algorithms, enhance accuracy and speed in real-time object detection. YOLO's grid-based approach and unique loss function improve accuracy by addressing challenges like overlapping objects. Ongoing research focuses on deep learning advancements and integrating object detection with other tasks. YOLOv5, noted for its efficiency with a hybrid network, performs well in tasks like autonomous driving, emphasizing ethical data use and responsible deployment for significant speed and accuracy improvements.

**Jonathan Atrey et.al** 2022.[6] This encompasses recent studies in face mask detection systems, highlighting the utilization of deep learning techniques such as Convolutional Neural Networks (CNNs) and YOLOv6 architectures. Key themes include real-time detection, dataset diversity, and practical applications during the COVID-19 pandemic. Works emphasize the importance of robust models trained on diverse datasets and provide comprehensive overviews of methodologies, challenges, and future directions in this research domain. These studies collectively contribute to advancing face mask detection technology for public safety and health monitoring.

**Chien-Yao Wang et.al** 2022.[7] The research focuses on improving real-time object detection using the Microsoft COCO dataset. It explores different versions of the YOLO series, like YOLOv7, designed for various hardware platforms and service needs. The study optimizes models for efficiency and discusses the impact of

activation functions. It evaluates performance using standard metrics and compares the proposed models with others in the field. Overall, the research aims to enhance real-time object detection while considering practical deployment factors.

**Dillon Reis et.al** 2023. [8] The text discusses the challenges of object detection, particularly in distinguishing between similar classes like F-14 and F-18 fighter jets. It explores the architecture of YOLOv8, emphasizing its use of activation maps to understand feature extraction at different stages of the model. Experimental results showcase the model's ability to detect small objects and classify various flying objects accurately, even in challenging conditions. Transfer learning is employed to refine the model, demonstrating improved performance in detecting and classifying small, pixelated objects and distinguishing between different types of flying objects. Overall, the text highlights the effectiveness of YOLOv8 and transfer learning in addressing object detection challenges.

## 3. METHODOLOGY

***Yolo Version 1***

The methodology for utilizing YOLO (You Only Look Once) for object detection begins with the collection and preprocessing of a diverse dataset containing various objects of interest, along with annotations detailing bounding boxes and class labels. Subsequently, the YOLO architecture is implemented and trained using this dataset, fine-tuning its parameters through optimization techniques such as Stochastic Gradient Descent (SGD). Evaluation of the trained model involves assessing its performance metrics, particularly mean Average Precision (mAP), and comparing it with other state-of-the-art methods to understand its relative strengths and weaknesses. To enhance the model's accuracy and speed, optimization strategies are employed, including fine-tuning hyperparameters and exploring different model variations. Finally, deployment of the trained YOLO model in real-world applications entails integration into existing systems and continuous monitoring and updating to ensure reliability and adaptability to evolving scenarios and requirements. This iterative process ensures that the YOLO model effectively addresses the need for fast and accurate object detection across diverse use cases and environments.[9]

***Yolo Version 2***

YOLOv2, or YOLO9000, improves on YOLOv1's unified object detection by introducing anchor boxes for flexible bounding box predictions and using multiple scales to better detect objects of various sizes. It employs the Darknet-19 architecture, a network with 19 convolutional layers, and combines object detection and classification training on diverse datasets. YOLOv2 introduces the idea of YOLO9000, aiming to recognize over 9000 object categories. With real-time processing capabilities, YOLOv2 enhances accuracy and expands object detection abilities by refining predictions and incorporating hierarchical classification, making it effective for applications requiring swift and comprehensive object recognition.[10]

***Yolo Version 3***

YOLOv3, the third version of the You Only Look Once (YOLO) algorithm, improves object detection by using a Feature Pyramid Network and the Darknet-53 architecture for better feature extraction. It predicts bounding boxes at three different scales, making it effective for spotting objects of various sizes. YOLOv3 maintains real-time processing capabilities, predicting class probabilities and object ness scores using anchor boxes. With three YOLO "heads" for multi-scale predictions and Non-Maximum Suppression to refine results, YOLOv3 excels in detecting objects across categories, having been trained on the COCO dataset with insights from YOLO9000. This makes YOLOv3 accurate and suitable for real-time applications.

***Yolo Version 4***

The YOLOv4 algorithm is specifically designed to perform real-time object detection using a single comprehensive pass through the neural network. Its primary goal is to efficiently identify and locate objects within live webcam feeds. The algorithm achieves this by configuring the YOLOv4 model, loading and preprocessing diverse datasets, and training the model using transfer learning for better adaptability. Additionally,

it seamlessly integrates with webcams in the Google Co lab environment and evaluates performance metrics, aiming to strike a balance between accuracy and speed. The ultimate purpose is to create a versatile and efficient system applicable in various domains such as surveillance, robotics, and human-computer interaction.

***Yolo Version 5***

Implementing YOLOv5 for object detection on the COCO dataset involves several key steps. First, preprocess the data by splitting it into training, validation, and testing sets, and apply augmentation techniques to increase diversity. Then, select YOLOv5 as the model architecture and fine-tune it on the training data while monitoring performance metrics like loss and mAP. Evaluate the trained model on the validation set, optimizing hyperparameters and exploring techniques like knowledge distillation for further improvement. Address ethical considerations such as dataset bias and fairness, and consider future directions such as enhancing object recognition in complex environments and extending the COCO dataset. By following this methodology, YOLOv5's effectiveness in real-time object detection and tracking can be maximized while ensuring responsible and impactful deployment.

***Yolo Version 6***

The methodology entails first collecting and preparing a diverse dataset of ten thousand images categorized into 'mask', 'without a mask', and 'incorrectly worn mask' classes, ensuring variations in angles, lighting, distances, backgrounds, and human attributes. Subsequently, the YOLOv4, YOLOv5, and YOLOv6 object detection algorithms are trained and evaluated using this dataset, with performance metrics analyzed for comparison. Following this, YOLOv6 is implemented in a real-world scenario, such as a local convenience store, to detect mask-wearing individuals in real-time and assess its impact on public behaviour. Validation of the model's performance using additional images is conducted, alongside discussion of future research directions. Ultimately, a comprehensive report is prepared to document findings and insights for stakeholders and policymakers in the field of face mask detection and public health safety.

***Yolo Version 7***

YOLOv7, the latest iteration in the YOLO series of object detection systems, represents a significant advancement over its predecessors by incorporating novel architectural enhancements and training methodologies. Building upon the original YOLO framework's revolutionary approach of single-shot detection, YOLOv7 introduces innovations like improved backbone architectures and optimization strategies aimed at enhancing both accuracy and efficiency. With its adaptability to various deployment scenarios and techniques such as model scaling, YOLOv7 offers state-of-the-art performance in real-time object detection applications, making it well-suited for diverse environments ranging from edge devices to cloud servers.

***Yolo Version 8***

The methodology involves recognizing the malicious use of drones and addressing limitations in detection technologies by developing a real-time flying object detection model. This model, based on YOLOv8 architecture with CSPDarknet53 backbone, is trained initially on diverse flying object datasets, then refined using transfer learning to simulate real-world conditions. Training involves annotation, splitting data, and evaluating performance across various object sizes and backgrounds. The refined model undergoes extensive testing and evaluation, including confusion matrix analysis and activation map interpretation, to ensure its ability to detect small, camouflaged, and distant flying objects accurately. Finally, the refined model is deployed for real-world implementation, with documentation provided for future research and collaboration.

## 4. RESULT AND ANALYSIS

**Table 1:** Comparison of Existing Methods

| YOLO Version | Description | Result |
|---|---|---|
| YOLOv1 | Divides image into a grid for bounding box predictions, uses a unified | Real-time processing, pre-trained on ImageNet, fine- |

| | | |
|---|---|---|
| | approach for object detection | tuned |
| YOLOv2 | Introduces anchor boxes for flexible bounding box predictions, employs Darknet-19 architecture | Enhanced accuracy, real-time processing, hierarchical classification |
| YOLOv3 | Utilizes Feature Pyramid Network and Darknet-53 architecture, predicts bounding boxes at three scales | Detects objects of various sizes, real-time processing |
| YOLOv4 | Specifically designed for real-time object detection, employs transfer learning for better adaptability | Single comprehensive pass through the network, balance between accuracy and speed |
| YOLOv5 | Focuses on small target detection, adjusts backbone's down-sampling process and feature fusion | Streamlined algorithm for efficiency, improved detection of small objects |
| YOLOv6 | Anchor-free detection for industrial applications, incorporates Efficient Rep backbone for higher parallelism | Efficient object detection, self-distillation for faster detection |
| YOLOv7 | Outperforms other detectors in speed and accuracy, introduces Extended efficient layer aggregation network | Object detection in dynamic scenarios, retail monitoring |
| YOLOv8 | Employs CSPDarknet-53 backbone for lightweight design, utilizes Decoupled-Head for classification separation | Gradient flow optimization, anchor-free methodology |

## 5. CONCLUSION

The survey paper offers a comprehensive exploration of the YOLO (You Only Look Once) family of algorithms, elucidating their evolution and performance in object detection within the realm of deep learning. It meticulously analyses the key features and advancements spanning from YOLOv1 to YOLOv8, shedding light on their adeptness in facilitating real-time detection on embedded systems and their wide-ranging applications across diverse domains.

Furthermore, by delving into research studies and practical implementations, the paper underscores the pivotal role of YOLO in pivotal tasks such as surveillance, autonomous vehicles, and healthcare. It underscores YOLO's prowess in addressing intricate challenges associated with object detection, thus highlighting its relevance and impact in various real-world scenarios.

## 6. FUTURE ENHANCEMENT

As future enhancement in YOLO algorithms aims to enhance efficiency through advanced compression techniques and model optimization, while tailoring models for specific domains like healthcare and agriculture. Additionally, there's a focus on developing lightweight YOLO models for edge device deployment, addressing limitations in computational resources. Improving robustness against environmental variations and occlusions is crucial, alongside efforts to enhance interpretability and explainability for better trust in model decisions.

## REFERENCES

[1] Joseph Redmon, and Santosh Divvala. *You Only Look Once: Unified, Real-Time Object Detection*. *CVF*, 2016. pp. 779-788.

[2] P. Viola and M. J. Jones. Robust real-time face detection. International journal of computer vision, 57(2):137–154, 2004.

[3] Sakshi Gupta, and Dr. T. Uma Devi. *YOLOv2 Based Real Time Object Detection*. vol. 8, *IJCST*, 2020. pp. 26-30

[4] Swetha M S, et al. “Object Detection and Classification in Globally Inclusive Images Using Yolo” proceedings of the International Journal of Advance Research in Computer Science and Management Studies (IJARCM) in Dec 2018

[5] Omkar Masurekar, and Omkar Jadhav. *Real Time Object Detection Using YOLOv3*. vol. 7, *IRJET*, 2020. pp. 3764-3768.

[6] Satyajith Chary, Podakanti . "Real Time Object Detection Using YOLOv4." *Ijraset*, vol. 11, 2023, pp. 1375-1379, https://doi.org/10.22214/ijraset.2023.57602.

[7] Mani Kandan, et al. *OBJECT DETECTION USING YOLO V5*. *ECB*, 2023. pp. 6266-6233.

[8] "Real-world Application of Face Mask Detection System Using YOLOv6." *ResearchGate*, 23 Feb. 2023, www.researchgate.net/publication/368849613.

[9] Chien-Yao Wang, and Alexey Bochkovskiy. "YOLOv7: Trainable Bag-of-freebies Sets New State-of-the-art for Real-time Object Detectors." *ResearchGate*, 10 Jul. 2022, www.researchgate.net/publication/361807900.

[10] Dillon Reis, and Jordan Kupec. Real-Time Flying Object Detection with YOLOv8. ArXiv, 2023.pp. 1-10.