IARJSET

ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁐ Peer-reviewed & Refereed journal ⁐ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11550

The YOLO Odyssey: A Deep Dive into Versions 1-9: Introducing Versions of Algorithm, Exploring Applications, and Unveiling Limitations

Mrs. Mariyam E. Maniyar[1], Ms. Vaidehi S. Navarkar[2], Ms. Janhavi S. Shindikar[3]

Assistant Professor, MCA Department, K. K. Wagh Institute of Engineering Education and Research, Nashik, India[1]

Student, MCA Department, MCA, K. K. Wagh Institute of Engineering Education and Research, Nashik, India[2]

Student, MCA Department, MCA, K. K. Wagh Institute of Engineering Education and Research, Nashik, India[3]

Abstract: The You Only Look Once (YOLO) algorithm stands as a cornerstone in the realm of object detection, celebrated for its unparalleled accuracy and efficiency. In this research endeavour, we embark on a comprehensive exploration of the various iterations of the YOLO algorithm. Through meticulous comparative analysis, we unveil the evolutionary trajectory of each YOLO version, shedding light on the motivations behind their respective updates. Our investigation delves deep into the intricacies of target recognition and feature selection methodologies, underscoring the algorithm's continual refinement. Furthermore, this study offers valuable insights into the applications of YOLO in diverse domains, including the financial sector. By elucidating the nuances of YOLO and its counterparts, this paper enriches the discourse surrounding object detection literature.

Keywords: Machine learning, Object detection, YOLO algorithm, YOLO versions

I. INTRODUCTION

What is Object Detection?
Finding items in digital images is the goal of the computer vision task known as object detection. In this sense, it is an example of artificial intelligence—that is, teaching computers to see in the same way as humans do, namely by identifying and categorizing objects based on semantic categories.

YOLO: The field of computer vision has also benefited from the work of the YOLO model family. Since its introduction by Joseph Redmon in 2014, YOLO has been the subject of intense research and development, as well as several developer and researcher implementations and community adoption. A set of single-stage detection systems known as YOLO (You Only Look Once) is based on the open-source CNN framework Darknet. The YOLO architecture was first created in 2016 with speed as its top priority. The Ultralytics team's YOLOv5 and YOLOv8, which are updated globally, drive production object detection models.

IARJSET
ISSN (O) 2393-8021, ISSN (P) 2394-1588
International Advanced Research Journal in Science, Engineering and Technology
Impact Factor 8.066 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 11, Issue 5, May 2024
DOI: 10.17148/IARJSET.2024.11550

YOLO Version History:

A Timeline of YOLO versions

- **Yolo versions comparison table**

Yolo Version	Applications	Limitations
YOLO V1: YOLO v1, or "You Only Look Once" Version 1, marks a milestone in object detection methodologies by introducing a unified approach where a single neural network predicts bounding boxes and class probabilities directly from entire images in one pass	1. Object detection in real-time applications. 2. Autonomous vehicles for detecting pedestrians, vehicles, and traffic signs. 3. Surveillance systems for detecting and tracking objects. 4. Industrial automation 5. Sport Analysis	1. Struggles with small object detection due to its single-scale approach. 2. Lower accuracy compared to newer versions like YOLOv4 and YOLOv5. 3. Handling overlapping Objects 4. Speed and Efficiency
YOLO V2: In contrast to YOLO V1, which adopts a shallower neural network design and adds anchor boxes, YOLO V2 opens the door to improved localization accuracy over a range of object sizes and shapes.	1. Real-time object detection for self-driving cars. 2. Security and surveillance systems. 3. Industrial automation for detecting defects in manufacturing processes 4. Environmental Monitoring 5. Augmented Reality	1. Still limited in handling small objects and occlusions. 2. Requires significant computational resources for real-time performance. 3. Training Data Requirements 4. Limited Contextual Understanding
YOLO V3: YOLOv3, which builds on its predecessors, has an impressively well-designed architecture that makes use of cutting-edge methods like feature pyramid networks (FPN) and a unique prediction approach called "multi-scale prediction."	1. Real-time object detection for smart home systems (e.g., detecting intruders). 2. Traffic management for identifying vehicles and pedestrians. 3. Medical imaging for identifying and analysing anomalies.	1. May struggle with heavily occluded or densely packed scenes. 2. Limited accuracy for very small objects.

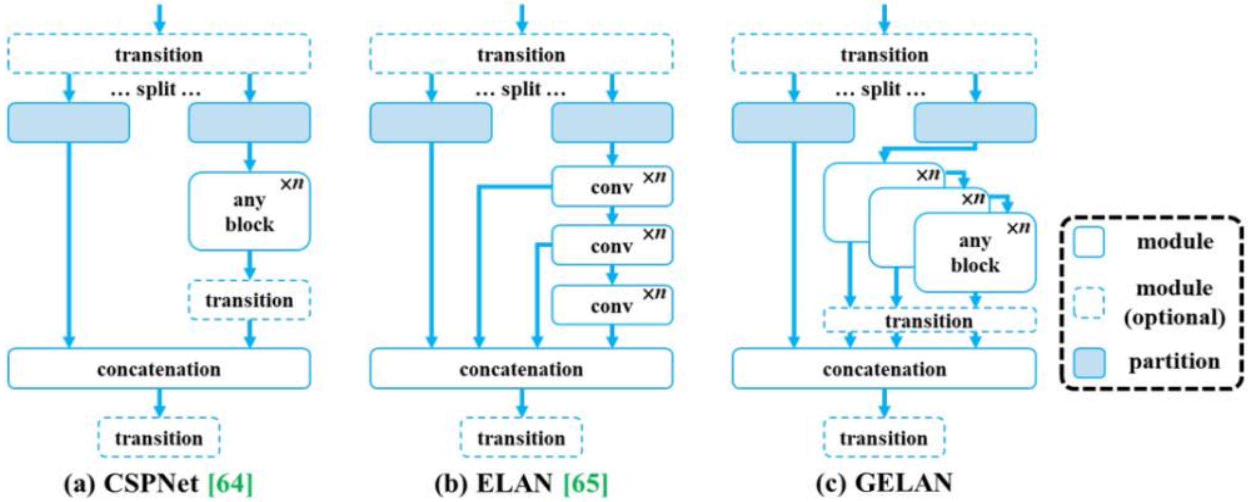
YOLO V4: Beyond mere iteration, YOLOv4 embodies a transformative fusion of cutting-edge techniques and architectural enhancements, poised to redefine the boundaries of accuracy, speed, and versatility in real-world applications. By seamlessly integrating state-of-the-art advancements in network architecture, optimization strategies, and training methodologies.	1. Advanced surveillance systems with improved accuracy. 2. Robotics for object detection and manipulation. 3. Environmental monitoring for detecting wildlife and anomalies.	1. Requires powerful hardware for real-time processing. 2. Still challenges with small object detection in cluttered scenes.
YOLOv5: In 2020, Glen Jocher, the CEO and founder of Ultralytics, launched YOLOv5, a few months after YOLOv4. It was created in Pytorch rather than Darknet and takes advantage of many of the enhancements covered in the YOLOv4 section. YOLOv5 includes the AutoAnchor Ultralytics algorithm.	1. Mobile and edge devices for real-time object detection. 2. Retail for inventory management and customer tracking. 3. Sports analytics for player tracking and activity recognition.	1. Limited to single-scale detection, affecting small object detection. 2. May require optimization for specific deployment scenarios.
YOLOv6: In September 2022, Meituan Vision AI Department published YOLOv6 to ArXiv. The network architecture consists of a PAN topology neck, an effective decoupled head using a hybrid-channel approach, and an efficient backbone with RepVGG or CSPStackRep blocks.	1. Improved real-time object detection for mobile devices. 2. Agricultural applications for crop monitoring and pest detection. 3. Wildlife conservation for tracking endangered species.	1. Limited to single-scale detection, similar to YOLOv5. 2. Performance may vary based on the dataset and deployment conditions.
YOLOv7: In July 2022, YOLOv7 was released on ArXiv by the same authors as YOLOv4 and YOLO R. It was faster and more accurate than any known object detector at the time, achieving 5 to 160 frames per second. It was learned without pre-trained backbones using only the MS COCO dataset, like YOLOv4.	1. High-speed object detection for robotics and drones. 2. Retail for automated checkout systems. 3. Smart cities for traffic management and public safety.	1. Computational Resource Demand 2. Complexity in Customization 3. Real-time Performance vs. Accuracy Trade-off 4. Ethical and Privacy Concerns
YOLOv8: YOLOv8 was released in January 2023 by Ultralytics, the company that developed YOLOv5. Five scaled variants were made available by YOLOv8: YOLOv8n (nano), YOLOv8s (small), YOLOv8m (medium), YOLOv8l (large), and YOLOv8x (extra big). Numerous vision tasks, including object identification, segmentation, pose estimation, tracking, and classification, are supported by YOLOv8.	1. Cutting-edge real-time object detection for various industries. 2. Augmented reality for object recognition and interaction. 3. Environmental monitoring for disaster response and wildlife conservation.	1. Computational Resources 2. Data Bias and Overfitting 3. Complexity in Customization

DOI: 10.17148/IARJSET.2024.11550

YOLOv9: In February 2024, Chien-Yao Wang, I-Hua Yeh, and Hong-Yuan Mark Liao released YOLOv9, the most recent iteration of the programme. Designed to outperform all transformer- and convolution-based techniques, it is an enhanced real-time object identification model.

YOLOv9 is available in four variants (v9-S, v9-M, v9-C, and v9-E), arranged according to the number of parameters. To enhance precision, it presents the Generalised Efficient Layer Aggregation Network (GELAN) and programmable gradient information (PGI). While GELAN optimises lightweight models with gradient route planning, PGI guards against data loss and guarantees precise gradient updates. Currently, object detection is the only computer vision task that YOLOv9 supports.

II. ARCHITECTURE



The Generalised Efficient Layer Aggregation Network (GELAN) and Programmable Gradient Information (PGI) are combined in YOLOv9 to produce a novel design that greatly enhances gradient flow and information retention. The ability of YOLOv9 to precisely understand an object's structure from a brief warm-up during training is depicted in the above figure. Building upon the foundation of YOLOv7, YOLOv9 adds an additional PGI and GELAN layer.

Applications:

1. Logistics and distribution: Using object detection, product inventory levels can be estimated to guarantee adequate stock levels and to gather data on customer behaviour.
2. Autonomous vehicles: To assist self-driving cars in safely navigating the road, autonomous vehicles can make use of YOLOv9 object detection.
3. People counting: Shop owners and mall operators can train the model to identify foot traffic in real time, estimate wait times, and other features.
4. Sports analytics: By tracking players' actions on a pitch, analysts can utilise the model to obtain pertinent information about the success of their team.

Key Takeaways:

Utilising auxiliary reversible branches, Programmable Gradient Information (PGI) is a method for producing gradients that are more dependable. Every model size, from small to huge, can be used effectively.

1. Generalised Efficient Layer Aggregation Network (GELAN): YOLOv9's foundation, GELAN is a new architecture that offers depth-wise parametrization for resource efficiency, increases flexibility by enabling interchangeable computational blocks, and guarantees consistent performance in a range of configurations with varying block types and depths for scalable object detection.
2. Better Performance with Less Complexity: YOLOv9's novel architecture reduces the complexity and processing needs of the model while increasing object detection speed and accuracy.
3. Flexibility in Various Model Sizes: YOLOv9 is flexible, providing five model variations (YOLOv9-n, YOLOv9-s, YOLOv9-m, YOLOv9-c, and YOLOv9-e) to meet a variety of needs, from compact to more complex, high-performance applications.

DOI: 10.17148/IARJSET.2024.11550

III. CONCLUSION

The evolution of the You Only Look Once (YOLO) algorithm from its inception in 2014 to the most recent iteration, YOLOv9, represents a significant journey in the realm of object detection. This comprehensive exploration of YOLO versions 1 through 9 has illuminated not only the advancements made in accuracy, speed, and versatility but also the persistent challenges faced in real-world applications.

Throughout this investigation, it became evident that each iteration of the YOLO algorithm was driven by a pursuit of enhanced performance and efficiency. From the pioneering YOLOv1, which introduced the concept of single-stage detection, to the groundbreaking YOLOv9, which leverages Generalised Efficient Layer Aggregation Network (GELAN) and Programmable Gradient Information (PGI) for improved precision and computational efficiency, the YOLO family has continually pushed the boundaries of what is possible in object detection.

Applications of YOLO algorithms span a wide range of domains, including autonomous vehicles, surveillance systems, medical imaging, retail, sports analytics, and environmental monitoring. These applications underscore the versatility and utility of YOLO across various industries and scenarios.

However, despite the remarkable progress made with each iteration, challenges persist. Issues such as small object detection, occlusion handling, computational resource demands, and real-time performance versus accuracy trade-offs remain areas of active research and development.

In conclusion, the YOLO Odyssey represents not just a series of algorithmic updates but a testament to the collaborative efforts of researchers, developers, and practitioners in advancing the field of object detection. As we look towards the future, it is clear that YOLO will continue to play a pivotal role in shaping the landscape of computer vision, driving innovation, and empowering applications that enhance our everyday lives.

REFERENCES

[1]. https://ieeexplore.ieee.org/document/9371695
[2]. https://ieeexplore.ieee.org/document/10404710
[3]. https://ieeexplore.ieee.org/document/10073281
[4]. https://towardsdatascience.com/evolution-of-yolo-yolo-version-1-afb8af302bd2
[5]. https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b
[6]. https://www.mdpi.com/2079-9292/11/15/2348
[7]. https://learnopencv.com/yolov9-advancing-the-yolo-legacy/#:~:text=Generalized%20Efficient%20Layer%20Aggregation%20Network%20(GELAN)%3A%20GELAN%2C%20the,performance%20across%20various%20configurations%20with
[8]. https://docs.ultralytics.com/yolov5/
[9]. https://arxiv.org/abs/2209.02976