Review

Not peer-reviewed version

YOLOv10 to Its Genesis: A Decadal and Comprehensive Review of The You Only Look Once Series

Ranjan Sapkota *, Rizwan Qureshi, Marco Flores-Calero, Chetan Badgujar, Upesh Nepal, Alwin Poulose, Peter Zeno, Uday Bhanu Prakash Vaddevolu, Hong Yan, Manoj Karkee

Posted Date: 24 June 2024

doi: 10.20944/preprints202406.1366.v2

Keywords: You Only Look Once; YOLO; YOLOv10 to YOLOv1; CNN; Deep learning; Object detection; Artificial intelligence; Computer vision; Healthcare; Autonomous Vehicles; Industrial manufacturing; Surveillance; Agriculture; YOLOv10; YOLOv9; YOLOv8

Article

YOLOv10 to Its Genesis: A Decadal and Comprehensive Review of The You Only Look Once Series

Ranjan Sapkota [1,*], Rizwan Qureshi [2], Marco Flores-Calero [3], Chetan Badgujar [4], Upesh Nepal [5], Alwin Poulose [6], Peter Zeno [7], Uday Bhanu Prakash Vaddevolu [8], Hong Yan [9] and Manoj Karkee [1]

1 Department of Biological Systems Engineering, Washington State University, United States
2 Center for Research in Computer Vision, The University of Central Florida, Orlando, USA
3 Department of Electrical, Electronics and Telecommunications, Universidad de las Fuerzas Armadas, Av. General Rumiñahui s/n, Sangolquí 171-5-231B, Ecuador
4 Biosystems Engineering and Soil Sciences, The University of Tennessee, Knoxville, TN 37996
5 Cooper Machine Company, Inc., Wadley, Georgia, 30477
6 School of Data Science, Indian Institute of Science Education and Research Thiruvananthapuram (IISER TVM), Vithura, Thiruvananthapuram 695551, Kerala, India
7 ZenoRobotics, LLC, Billings, MT 59106, USA
8 Agricultural and Biological Engineering, University of Florida, Gainesville, FL 32611
9 Department of Electrical Engineering, and Center for Intelligent Multidimensional Data Analysis, City University of Hong Kong, 999077, Hong Kong, China
* Correspondence: ranjan.sapkota@wsu.edu

Abstract: This review systematically examines the progression of the You Only Look Once (YOLO) object detection algorithms from YOLOv1 to the recently unveiled YOLOv10. Employing a reverse chronological analysis, this study examines the advancements introduced by YOLO algorithms, beginning with YOLOv10 and progressing through YOLOv9, YOLOv8, and subsequent versions to explore each version's contributions to enhancing speed, accuracy, and computational efficiency in real-time object detection. The study highlights the transformative impact of YOLO across five critical application areas: automotive safety, healthcare, industrial manufacturing, surveillance, and agriculture. By detailing the incremental technological advancements in subsequent YOLO versions, this review chronicles the evolution of YOLO, and discusses the challenges and limitations in each earlier versions. The evolution signifies a path towards integrating YOLO with multimodal, context-aware, and General Artificial Intelligence (AGI) systems for the next YOLO decade, promising significant implications for future developments in AI-driven applications.

Keywords: You Only Look Once; YOLO; YOLOv10 to YOLOv1; CNN; deep learning; object detection; real-time object detection; artificial intelligence; computer vision; healthcare; autonomous vehicles; industrial manufacturing; surveillance; agriculture

---

1. Introduction

Object detection is a critical component of computer vision, enabling systems to identify and locate objects within an image or video frame [1]. Real-time object detection has become integral to numerous applications requiring immediate analysis and interaction with dynamic environments [2–4]. For instance, real-time object detection is indispensable in autonomous vehicles and robotics [5], allowing the system to quickly recognize and track different objects such as vehicles, pedestrians, bicycles, and other obstacles, enhancing navigational safety and efficiency [6]. The utility of object recognition extends beyond vehicular applications, and is also pivotal in action recognition within video sequences, useful in digital surveillance, monitoring, sports analysis, and human-machine interaction [2,7]. These areas benefit from the capability to analyze and respond to situational dynamics in real-time, illustrating its broad applicability, acceptance, and impact. However, the problem of object detection involves several challenges:

 doi:10.20944/preprints202406.1366.v2

- Complexity of Real-World Scenes: Real-world environments are highly variable and unpredictable. Objects can appear in various orientations, scales, and lighting conditions, making it difficult for a detection algorithm to generalize and maintain accuracy [8].
- Occlusions and Clutter: Objects may be partially obscured by other objects, leading to incomplete information that must be accurately interpreted [9].
- Speed and Efficiency: Many applications, such as autonomous driving and real-time surveillance, require rapid processing of visual data to make timely decisions, demanding both high accuracy and low latency from detection algorithms [10].

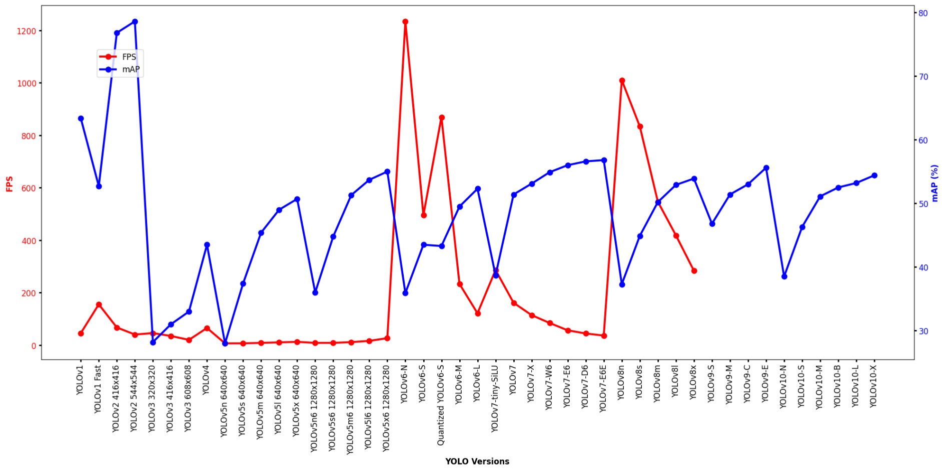


Figure 1. This visualization summarizes technical performance of YOLO models, comparing speed (FPS) and accuracy (mAP) from YOLOv1 to YOLOv10.

1.1. Traditional Approaches

Before the advent of deep learning, object detection relied on a combination of hand-crafted features and machine learning classifiers [11]. Some of the notable traditional methods include:

- **Correlation Filters:** Used to detect objects by correlating a filter with the image, often struggling with variations in object appearance [12].
- **Gabor Features:** Extracted texture features using Gabor filters, which are effective for texture representation but computationally intensive [13].
- **Histogram of Oriented Gradients (HOG):** Captures edge or gradient structures that characterize the shape of objects, typically combined with Support Vector Machines (SVM) for classification [14].
- **Local Binary Patterns (LBP):** Utilizes pixel intensity comparisons to form a binary pattern, used in texture classification and face recognition [15].
- **SVM and Multilayer Perceptrons (MLP):** Traditional classifiers used in conjunction with the aforementioned features to detect and classify objects [16]

While these methods can work well under controlled conditions, they often require careful manual feature selection and struggle to generalize across diverse real-world scenarios [17].

1.2. Emergence of Convolutional Neural Networks

The introduction of Convolutional Neural Networks (CNNs) revolutionized object detection [18] by automating feature extraction and enabling end-to-end learning. CNNs are particularly effective because:

- **Hierarchical Feature Learning:** CNNs learn to extract low-level features (e.g., edges, textures) in early layers and high-level features (e.g., object parts, shapes) in deeper layers, facilitating robust object representation [19].

 doi:10.20944/preprints202406.1366.v2

- **Spatial Invariance:** Convolutional layers enable CNNs to recognize objects regardless of their position within the image, enhancing detection robustness [20].
- **Scalability**: CNNs can be scaled to handle larger datasets and more complex models, improving performance on a wide range of tasks [21].

1.3. The R-CNN

Object detection presents a unique challenge for CNNs due to the variable number of objects in an image, which prevents the direct application of CNNs with fixed output layers [22]. While a sliding window-based brute force search could be used to select and classify regions [23], this approach is computationally prohibitive because it requires applying the CNN model to numerous region proposals of varying sizes and aspect ratios, making it inefficient for real-time applications.

In 2013, Ross Girshick et al. proposed the R-CNN (Region-based CNN) architecture to address these challenges. R-CNN uses the selective search algorithm to generate about 2000 region proposals, which are then processed by a CNN to extract features [24]. Fast R-CNN improved this process by integrating region proposal feature extraction and classification in a single pass [25]. Faster R-CNN further advanced the approach by introducing Region Proposal Networks (RPNs) for end-to-end training, eliminating the need for selective search [4].

1.4. You Only Look Once Approach

The "You Only Look Once" (YOLO) object detection algorithm was first introduced by Joseph Redmon et al., [26] in 2015, revolutionized real-time object detection by combining region proposal and classification into a single neural network, significantly reducing computation time. YOLO's unified architecture divides the image into a grid, predicting bounding boxes and class probabilities directly for each cell, enabling end-to-end learning [26]. YOLO is versatile, and its real-time detection capabilities have revolutionized agriculture [27] and healthcare [28], where accuracy and speed are paramount.

In agriculture, YOLO models detect and classify crops [29], pests, and diseases [30–32], facilitating precision agriculture techniques and automating farming operations to increase productivity and optimizing inputs. Additionally, in remote sensing, YOLO contributes to object recognition in satellite [33,34] and aerial imagery [35,36], which supports urban planning, land use mapping, and environmental monitoring. These capabilities demonstrate YOLO's contribution to critical global challenges such as urban development and environmental conservation. In healthcare, YOLO has been instrumental in assisting and improving diagnostic processes and treatment outcomes. The applications include, but are not limited to, cancer detection [37,38], skin segmentation [39], and pill identification [40,41] which showcase the model's ability to adapt to different needs, and essential tasks.

Surveillance and Security systems also leverage YOLO for real-time monitoring and rapid identification of suspicious activities [42,43]. By integrating these models into surveillance systems, security personnel can more effectively monitor and respond to potential threats, enhancing public safety [44]. Similarly, in the context of public health measures like social distancing and face mask detection during pandemics [45,46], YOLO models provided essential support in enforcing health regulations. In industrial applications, YOLO aids in surface inspection processes to detect defects and anomalies [47,48], ensuring quality control in manufacturing and production.

1.5. Motivation and Organization of the Study

Since "You Only Look Once" has been widely adopted in the field of computer vision, a search for this keyword in Google Scholar yields approximately 5,550,000 results as of June 9, 9:05 PM Pacific Daylight Time. The acronym "YOLO" further emphasizes its popularity, generating around 210,000 search results. Thousands of researchers have cited YOLO papers, highlighting its significant influence. This study aims to review the YOLO's decadal progress and its advancements over time, as visually summarized in the mind-map ,shown in Figure 2.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 June 2024 doi:10.20944/preprints202406.1366.v2

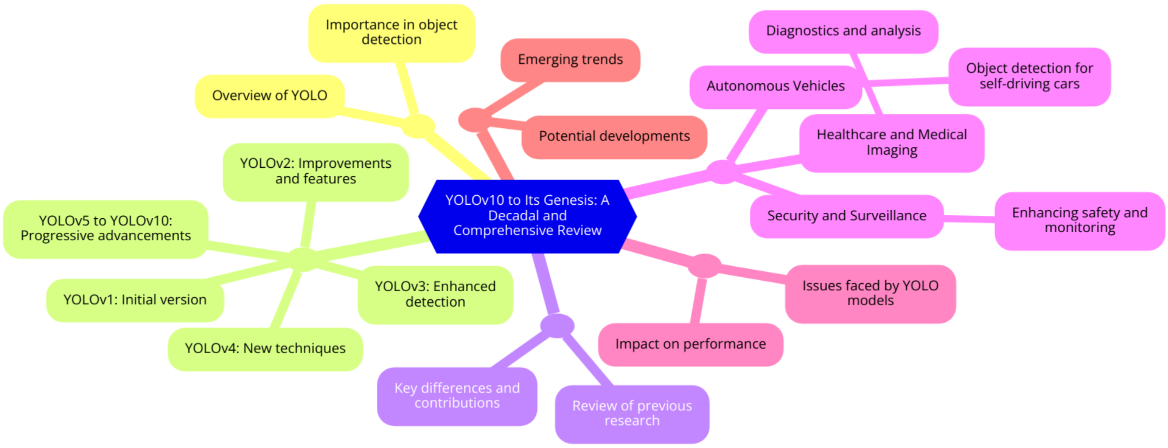


Figure 2. Block diagram showing organization of this review article: The structure includes YOLO Trajectory discussing the development path, Prior YOLO literature: Context and Distinctions providing background and differentiations, Review of YOLO Versions detailing each version, Applications highlighting various use cases, Challenges, Limitations and Future Directions addressing current issues and potential advancements, and Conclusion summarizing the findings. Each section systematically contributes to a comprehensive understanding of the YOLO framework's evolution and impact.

The comprehensive analysis begins with **YOLO Trajectory**, discussing the development path from YOLOv1 to YOLOv10. Next, **Prior YOLO literature: Context and Distinctions** provides background and differentiations among existing works. **Review of YOLO Versions** details the key features and improvements of each version. The **Applications** section highlights various use cases across different domains. Following this, **Challenges, Limitations and Future Directions** addresses current issues and potential advancements. Finally, the **Conclusion** summarizes the findings. Each section is further broken down into subsubsections: **YOLO Trajectory** includes **Significance of Latency and mAP Scores in YOLO** and **Single stage detection in YOLO**; **Prior YOLO literature: Context and Distinctions**; **Review of YOLO Versions** covers **YOLOv10, YOLOv9 and YOLOv8**, **YOLOv7, YOLOv6 and YOLOv5**, and **YOLOv4, YOLOv3, YOLOv2 and YOLOv1**; **Applications** discusses **Autonomous Vehicles**, **Healthcare and Medical Imaging**, **Security and Surveillance**, **Manufacturing**, and **Agriculture**; **Challenges, Limitations and Future Directions** explores **YOLO and the Artificial General Intelligence - AGI**, **Yolo on the Edge Devices**, and **Future Prospects**. This structured approach ensures a detailed and systematic review of the YOLO framework's evolution and impact.

2. YOLO Trajectory

YOLOv1 [26] was introduced in 2015 as a novel approach to object detection, offering good accuracy and speed by processing images in a single stage. The first YOLO version laid the foundation for real-time applications, setting a new standard for subsequent developments.

Figure 3 shows the timeline history of YOLO from its release version YOLOv1 to the upto date version YOLOv10. YOLOv2, or YOLO9000 [49,50], expanded on this foundation by improving the resolution at which the system operated and by being capable of detecting over 9000 object categories, thus enhancing its versatility and accuracy. YOLOv3 further advanced these capabilities by implementing multi-scale predictions and a deeper network architecture, which allowed better detection of smaller objects [51]. The series continued to evolve with YOLOv4 and YOLOv5, each introducing more refined techniques and optimizations to improve detection performance (i.e., accuracy and speed) even further [52–54]. YOLOv4 incorporated features like Cross-Stage Partial (CSP) connections and Mosaic data augmentation, while YOLOv5, developed by Ultralytics, brought significant improvements in terms of ease of use and performance, establishing itself as a popular choice in the computer vision community. Subsequent versions, YOLOv6 through YOLOv10, have continued to build on this success,

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 June 2024 doi:10.20944/preprints202406.1366.v2

focusing on enhancing model scalability, reducing computational demands, and improving real-time performance metrics. Each iteration of the YOLO series has set new benchmarks for object detection capabilities and significantly impacted various application areas, from autonomous driving and traffic monitoring to healthcare and industrial automation.

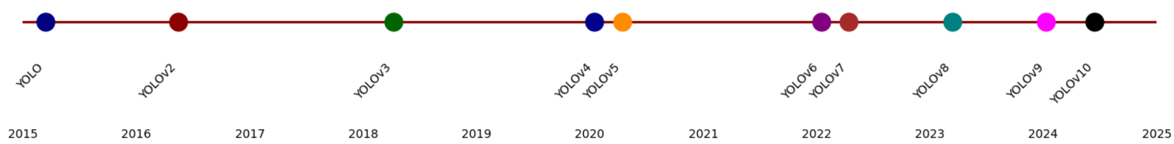


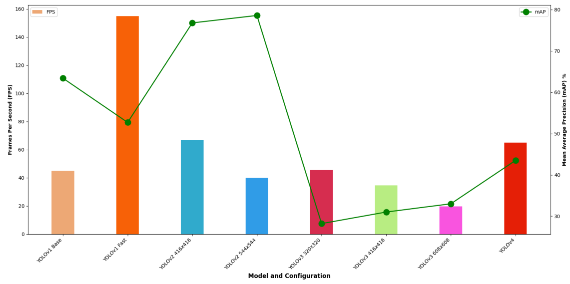
Figure 3. Timeline of YOLO versions from 2015 to 2024, illustrating the development progression from YOLOv1 to YOLOv10.

YOLOv10 [55], the latest iteration, introduces multiple model variants such as YOLOv10-N, YOLOv10-S, YOLOv10-M, YOLOv10-B, YOLOv10-L, and YOLOv10-X, achieving precision (AP) scores ranging from 38.5% to 54.4% on MS-COCO dataset [55]. Notably, YOLOv10-N and YOLOv10-S exhibit the lowest latencies at 1.84 ms and 2.49 ms, respectively, making them highly suitable for applications requiring low latency. These models outperform their predecessors, with YOLOv10-X achieving the highest mAP of 54.4% and a latency of 10.70 ms, reflecting a well-balanced enhancement in both accuracy and inference speed. According to Wang et al. [55], comparing YOLOv10 with YOLOv9 and YOLOv8 reveals a trend of incremental improvements. YOLOv9, featuring models like YOLOv9-N, YOLOv9-S, YOLOv9-M, YOLOv9-C, and YOLOv9-X, achieves mAP scores from 39.5% to 54.4% [56]. While the mAP scores are comparable to YOLOv10, the latency of YOLOv9 models is generally higher, particularly for YOLOv9-X, which matches YOLOv10-X in mAP but not in latency, indicating YOLOv10's superior efficiency [56]. YOLOv8 models, including YOLOv8-N, YOLOv8-S, YOLOv8-M, YOLOv8-L, and YOLOv8-X, show mAP scores ranging from 37.3% to 53.9% and latencies from 6.16 ms to 16.86 ms. Although YOLOv8 models perform well, they lag behind YOLOv10 and YOLOv9 in terms of both accuracy and latency, suggesting that the architectural refinements in YOLOv10 have effectively enhanced both detection performance and computational efficiency.

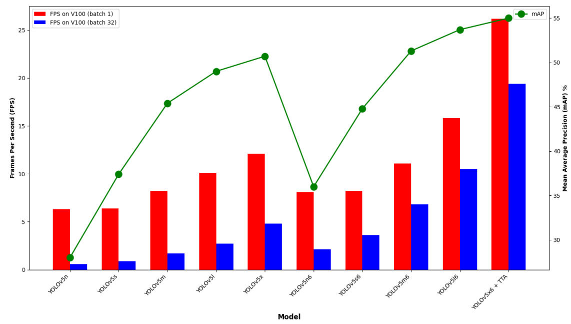
Further analysis of earlier YOLO versions, such as YOLOv7, YOLOv6, YOLOv5, YOLOv4, YOLOv3, YOLOv2, and YOLOv1, underscores the rapid advancements in this domain [57]. YOLOv7-tiny and YOLOv7 models achieve mAP scores of 56.4% and 51.2%, respectively, but with significantly higher latencies, indicating their focus on higher accuracy at the cost of speed [58]. YOLOv6 models (YOLOv6-N, YOLOv6-S, YOLOv6-M, YOLOv6-L) achieve mAP scores from 37.0% to 51.8% with moderate latencies. YOLOv5, a popular model, shows a competitive mAP of 50.7% and a latency of 140 ms [59]. Earlier versions like YOLOv4, YOLOv3, YOLOv2, and YOLOv1, with mAP scores of 43.5%, 57.9%, 76.8%, and 63.4% respectively [52], laid the groundwork for subsequent improvements, though they exhibit higher latencies compared to the latest versions.

The evolution of the YOLO as initially presented in scholarly articles, YOLOv1 through YOLOv4 were documented extensively in the literature [26,60–62]. These versions, showcased in Figure 4a, were fundamental in advancing object detection technologies, providing robust source code on GitHub and paving the way for further innovations. With the commercial landscape evolving, Ultralytics released YOLOv5 and YOLOv8, not through traditional academic channels but directly on GitHub, creating a pivotal shift in deployment and adaptation [63–65]. Subsequent versions, YOLOv6 and YOLOv7, marked a return to the academic realm, with detailed documentations and enhancements presented in [58,59]. Figure 4b shows the FPS and mAP comparision.

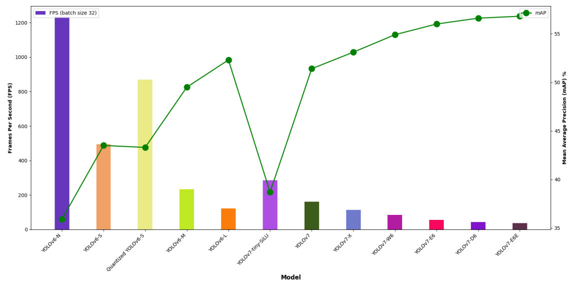
Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 June 2024 doi:10.20944/preprints202406.1366.v2



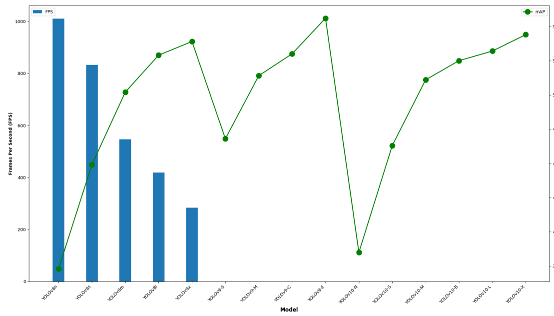
(a) Performance metrics for YOLOv1 to YOLOv4, illustrating advancements in object detection technology. Detailed in [26,60–62].



(b) Performance analysis of YOLOv5 by Ultralytics, highlighting significant improvements in speed and accuracy. Refer to [62–65].



(c) Comparative performance of YOLOv6 and YOLOv7, emphasizing enhanced detection capabilities and efficiency. Documented in [58,59].



(d) Analysis of YOLOv8, YOLOv9, and YOLOv10, showcasing continuous advancements in model precision and processing speed. Refer to [55,56,62–65]. YOLOv9 and YOLOv10 have no data in the paper to present FPS

Figure 4. Comprehensive analysis of the performance metrics for YOLO versions. Subfigure (a) covers YOLOv1 to YOLOv4, (b) details YOLOv5, (c) compares YOLOv6 and YOLOv7, and (d) showcases YOLOv8, YOLOv9, and YOLOv10.

The technical analysis of these versions, as visualized from YOLOv1 to YOLOv10, highlights a progressive enhancement in both speed and accuracy. Performance metrics such as FPS and mAP were critically analyzed using Python and Matplotlib, illustrating the trade-offs inherent in each version's design. YOLOv6 through YOLOv10, documented in Figure 4c,d illustrate the continuous improvements, with later models optimizing computational efficiency and detection precision [55,56,58]. Each figure reflects the intricate balance between processing speed and accuracy, providing insights into the model's performance across various configurations and input resolutions. This ongoing development trajectory showcases the dynamic interplay between academic research and commercial applications, driving forward the capabilities of object detection systems in real-world scenarios.

2.1. Significance of Latency and mAP Scores in YOLO

Latency (L) and mAP are important metrics for describing the performance of object detection models like YOLO [66,67]. Latency measures the time taken by the model to process an image and produce predictions [67]. This includes all the steps required for the detection process, such as image preprocessing, model inference, and postprocessing, and is typically measured in milliseconds (ms). Lower latency is essential for real-time applications such as autonomous driving, surveillance, and robotics, where timely and accurate detection is crucial [68]. High latency can result in delays that are detrimental in these fast-paced environments, potentially compromising operational safety and

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 June 2024 doi:10.20944/preprints202406.1366.v2

effectiveness [69]. FPS (Frames per second) is another critical metric that complements latency by indicating how many images the model can process each second. Together, latency and FPS provide a comprehensive overview of a model's performance in real-time scenarios. Figure 4a illustrates the mAP and FPS rates, while Figure 4b illustrates the latency value of all 10 YOLO versions showcasing their evolution and effectiveness in real-time applications.

Likewise, mAP is a comprehensive metric used to evaluate the accuracy of object detection models [70]. It considers both precision and recall (Table 1), and it is calculated by taking the average precision (AP) across all classes and then averaging these AP scores [70,71]. It provides a balanced view of how well the model performs across different object categories and varying conditions within the dataset. Other metrics used for comprehensive evaluation of YOLO models [72,73] are detailed in Table 1.

Table 1. Summary of Performance Metrics Used in Model Evaluation.

No.	Performance Metric	Symbol	Equation	Description
1	Precision	P	$P = \frac{TP}{TP+FP}$	Ratio of true positive detections to the total predicted positives.
2	Recall	R	$R = \frac{TP}{TP+FN}$	Ratio of true positive detections to the total actual positives.
3	F1 Score	$F1$	$F1 = 2 \cdot \frac{P \cdot R}{P+R}$	Harmonic mean of precision and recall, balancing both metrics to provide a single performance measure for the model
4	Intersection over Union	IoU	$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$	Measures the overlap between the predicted and actual bounding boxes.
5	Frames Per Second	FPS	$\text{FPS} = \frac{1}{L}$	Number of images the model processes per second, inversely related to latency.
6	Non-Maximum Suppression	NMS	-	NMS is a post-processing step in YOLO to remove redundant bounding-boxes.

Here, True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) are the key performance evaluators where TP are instances where the model correctly identifies an object as present. TN occur when the model correctly predicts the absence of an object. FP arise when the model incorrectly identifies an object as present, and FN happen when the model fails to detect an object that is actually present. These metrics are crucial for assessing the accuracy and reliability of the YOLO object detection [70,71,73].

2.2. *Single Stage Detection in YOLO*

The Single Shot MultiBox Detector (SSD) [74] introduced in 2015 has revolutionized object detection by streamlining the process through a single-stage approach, significantly inspiring subsequent developments in YOLO models [74–76]. Unlike two-stage models like R-CNN, which rely on a region proposal step before actual object detection, SSD and by extension, YOLO variants, perform detection and classification in a single sweep across the image. This paradigm shift enhances the detection process by eliminating intermediate steps, thus facilitating faster and more efficient object detection suitable for real-time applications. The architecture of SSD, which YOLO models have adapted, utilizes multiple feature maps at different resolutions to detect objects of various sizes, employing a diverse array of anchor boxes at each feature map location to improve localization accuracy [77,78].

Figure 5 shows an example of a YOLO model that integrates SSD's architecture principles, specifically focusing on enhancing real-time detection capabilities through improved feature extraction using Multi-Headed Attention (MA) layers. These adaptations from SSD's methodology have enabled YOLO models such as YOLOv8, YOLOv9, and YOLOv10 to achieve significant improvements in processing speed and detection accuracy, making them highly effective for applications requiring rapid and reliable object detection [79,80]. The SSD-inspired single-shot mechanism directly classifies and localizes objects, reducing computational overhead and enabling the deployment of these models in resource-constrained environments such as mobile and edge devices. The continuous refinement of

 doi:10.20944/preprints202406.1366.v2

these techniques in YOLO models underscores an ongoing evolution aimed at balancing the demanding accuracy requirements with the need for speed in diverse real-world scenarios [76].

Table 2. Presenting the latency in milliseconds for various YOLO versions, highlighting the progression and improvements in speed across different iterations. The data reflects the enhancements made in real-time object detection capabilities, with references to the respective research papers for each version.

YOLO Version	Latency (ms)	Reference
YOLOv1 Base	22.22	[26]
YOLOv1 Fast	6.45	[26]
YOLOv2 (416x416)	14.93	[60]
YOLOv2 (544x544)	25	[60]
YOLOv3 (320x320)	22	[61]
YOLOv3 (416x416)	29	[61]
YOLOv3 (608x608)	51	[61]
YOLOv4	15.38	[62]
YOLOv5n (640x640)	158.73	[62–65]
YOLOv5s (640x640)	156.25	[62–65]
YOLOv5m (640x640)	121.95	[62,65]
YOLOv5l (640x640)	99.01	[62,65]
YOLOv5x (640x640)	82.64	[62,65]
YOLOv5n6 (1280x1280)	123.46	[62,65]
YOLOv5s6 (1280x1280)	121.95	[62,65]
YOLOv5m6 (1280x1280)	90.09	[62,63,65]
YOLOv5l6 (1280x1280)	63.29	[62,63,65]
YOLOv5x6 (1280x1280)	38.17	[62,65]
YOLOv6-N	0.81	[59]
YOLOv6-S	2.02	[59]
Quantized YOLOv6-S	1.15	[59]
YOLOv6-M	4.29	[59]
YOLOv6-L	8.26	[59]
YOLOv7-tiny-SiLU	3.50	[58]
YOLOv7	6.21	[58]
YOLOv7-X	8.77	[58]
YOLOv7-W6	11.90	[58]
YOLOv7-E6	17.86	[58]
YOLOv7-D6	22.73	[58]
YOLOv7-E6E	27.78	[58]
YOLOv8n	0.99	[63–65]
YOLOv8s	1.20	[63–65]
YOLOv8m	1.83	[63–65]
YOLOv8l	2.39	[63–65]
YOLOv8x	3.53	[63–65]
YOLOv9 and YOLOv10	Not Reported in the paper	[55,56]

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 June 2024 doi:10.20944/preprints202406.1366.v2

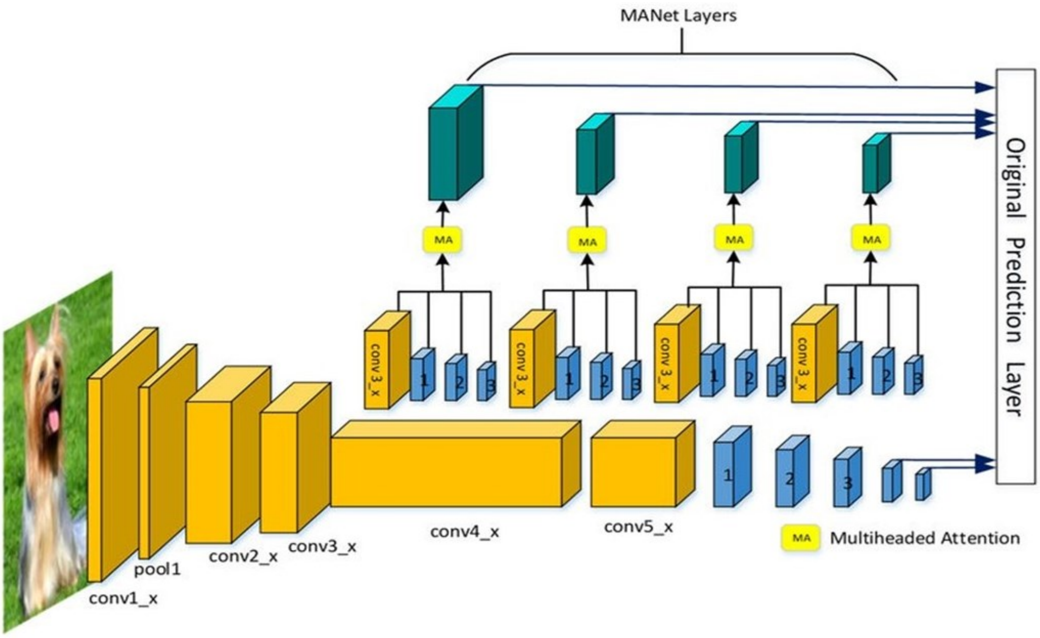


Figure 5. Enhanced YOLO model architecture incorporating SSD's single-stage detection approach with Multi-Headed Attention (MA) layers for superior real-time object detection performance [81].

3. Prior YOLO literature: Context and Distinctions

We collected the existing published literature on YOLO to document and critically analyze the past knowledge, including major highlights and limitations, which are briefly summarized and discussed here:

- "*A Review of YOLO Algorithm Developments*" by Peiyuan Jiang et al. [82] provided an insightful overview on YOLO algorithm development and its evolution through its versions. The authors analyze the fundamental aspects of YOLO's to object detection, comparing its various iterations to traditional CNNs. They emphasizes the ongoing improvements in YOLO, particularly in enhancing target recognition and feature extraction capabilities. It also discusses YOLO's application in specific fields like finance, highlighting its practical implications in feature extraction for image-based news analysis [82].
- "*A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023)*" by Ragab et al. [83] presented a systematic review of YOLO's application in the medical field, that analyzes how different variants, particularly YOLOv7 and YOLOv8, have been employed for various medical detection tasks. They highlight the algorithm's significant performance in lesion detection, skin lesion classification, and other critical areas, demonstrating YOLO's superiority over traditional methods in terms of accuracy and computational efficiency. Despite its successes, the review identifies challenges, such as the need for well-annotated datasets and addresses the high computational demands of YOLO implementations. The paper suggested directions for future research to optimize YOLO's application in medical object detection [83].
- "*A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS*" by Terven et al. [84] provides an extensive analysis of the evolutionary trajectory of the YOLO algorithm, detailing how each iteration has contributed to advancements in real-time object detection. Their review covers the significant architectural and training enhancements from YOLOv1 through YOLOv8 and introduces YOLO-NAS and YOLO with Transformers. This study serves as a valuable resource for understanding the progression in network architecture, which has progressively improved YOLO's efficacy in diverse applications such as robotics and autonomous driving.
- "*YOLOv1 to v8: Unveiling Each Variant–A Comprehensive Review of YOLO*" by Hussain [57], provided in-depth analyses on the internal components and architectural innovations of each YOLO

variant. It provided a deep dive into the structural details and incremental improvements that have marked the evolution of YOLO, presenting a well-structured analysis complete with performance benchmarks. This methodological approach not only highlights the capabilities of each variant but also discusses their practical impact across different domains, suggesting the potential for future enhancements like federated learning to improve privacy and model generalization [57].

- "*YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection*" by Muhammad Hussain [85] reviewed and showed rapid progression of the YOLO variants, focusing on their critical role in industrial applications, specifically for defect detection in manufacturing. Starting with YOLOv1 and extending through YOLOv8, the paper illustrates how each version has been optimized to meet the demanding needs of real-time, high-accuracy defect detection on constrained devices. Hussain's work not only examines the technical advancements within each YOLO iteration but also validates their practical efficacy through deployment scenarios in the manufacturing sector, emphasizing YOLO's alignment with industrial needs [85].

The existing literature shows a significant lack of comprehensive reviews incorporating the latest YOLO releases, specifically YOLOv9 and YOLOv10. As we mark a decade of the YOLO algorithm's evolution, it is crucial to systematically document and critically analyze the newer models to provide well-documented, synthesized, up-to-date insights and comparative analyses across a broader range of applications to the broad research and technical community. This state-of-the-art review paper aims to bridge this gap by exploring the advancements and capabilities of YOLOv9 and YOLOv10, offering a detailed perspective on their impact and potential within the ever-evolving landscape of object detection technologies.

In this review paper, we adopt a unique reverse-chronological approach to analyze the progression of YOLO, beginning with the most recent versions and moving backward. The analysis is divided into three distinct subsections. The first section covers the latest iterations, YOLOv10, YOLOv9, and YOLOv8, where we delve into the architecture and advancements that define the forefront of object detection technology. This approach not only shows the most cutting-edge developments but also sets the stage for understanding the incremental improvements that have been realized over time. The second section reviews YOLOv7, YOLOv6, and YOLOv5, tracing further back in the series to highlight the evolutionary steps that have contributed to the enhancements observed in the later versions. We analyze each model's technical and scientific aspects to provide a comprehensive view of the progress within these iterations. The third section addresses the earlier YOLO versions, offering a complete historical perspective that enriches the reader's understanding of the foundational technologies and the methodologies, refined through successive updates. Additionally, we discuss the application of the YOLO models in reverse order across five critical real-world domains: autonomous vehicles, healthcare and medical image analysis, security and surveillance, manufacturing industry, and agriculture. For each application, we present a detailed examination and a corresponding tabular data in reverse chronological order, showcasing how YOLO technologies have been adapted and implemented to meet specific industry needs and challenges. This reverse review strategy not only emphasizes the state-of-the-art but also provides a narrative of technological evolution, illustrating how each iteration builds upon the last to push the boundaries of what's possible in object detection. By understanding where YOLO technology stands today and how it got there, readers gain a comprehensive view of its capabilities and potential future directions. This methodical unpacking of the YOLO series not only highlights technological advancements but also offers insights into the broader implications and utility of these models in practical scenarios, setting the groundwork for anticipating future innovations in object detection technology.

 doi:10.20944/preprints202406.1366.v2

4. Review of YOLO Versions

This section reviews YOLO series models, starting from the advanced and latest version, YOLOv10, and progressively tracing back to the foundational YOLOv1. By first highlighting the most recent technological advancements, this approach enables immediate insights into the state-of-the-art capabilities of object detection. Subsequently, the narrative is reversed, exploring how earlier models laid the groundwork for these innovations.

4.1. YOLOv10, YOLOv9 and YOLOv8

YOLOv10 [55], developed at Tsinghua University, China, represents a breakthrough in the YOLO series for real-time object detection, achieving unprecedented performance. This version eliminates the need for non-maximum suppression (NMS) [86], a traditional bottleneck in earlier models, thereby drastically reducing latency. YOLOv10 introduces a dual assignment strategy in its training protocol, which optimizes detection accuracy without sacrificing speed with the help of one-to-many and one-to-one label assignments, ensuring robust detection with lower latency [87]. The architecture of YOLOv10 includes several innovative components that enhance both computational efficiency and detection performance. Among these are lightweight classification heads [88] that reduce computational demands, spatial-channel decoupled downsampling to minimize information loss during feature reduction [89], and rank-guided block design that optimizes parameter use [90]. These architectural advancements ensure that YOLOv10 operates synergistically across various scales—from YOLOv10-N (Nano) to YOLOv10-X (Extra Large), making it adaptable to diverse computational constraints and operational requirements [55]. According to wang et al. [55], performance evaluations on benchmark datasets like MS-COCO [91] demonstrate that YOLOv10 not only surpasses its predecessors—YOLOv9 and YOLOv8—in both accuracy and efficiency but also sets new industry standards. For instance, YOLOv10-S significantly outperforms comparable models with an improved mAP and lower latency. This version also incorporates holistic efficiency-accuracy driven design, large-kernel convolutions, and partial self-attention modules, which collectively improve the trade-off between computational cost and detection capability. The architecture diagrams of YOLOv10, YOLOv9, and YOLOv8 are summarized in Figures 6, 7, and 8, respectively.

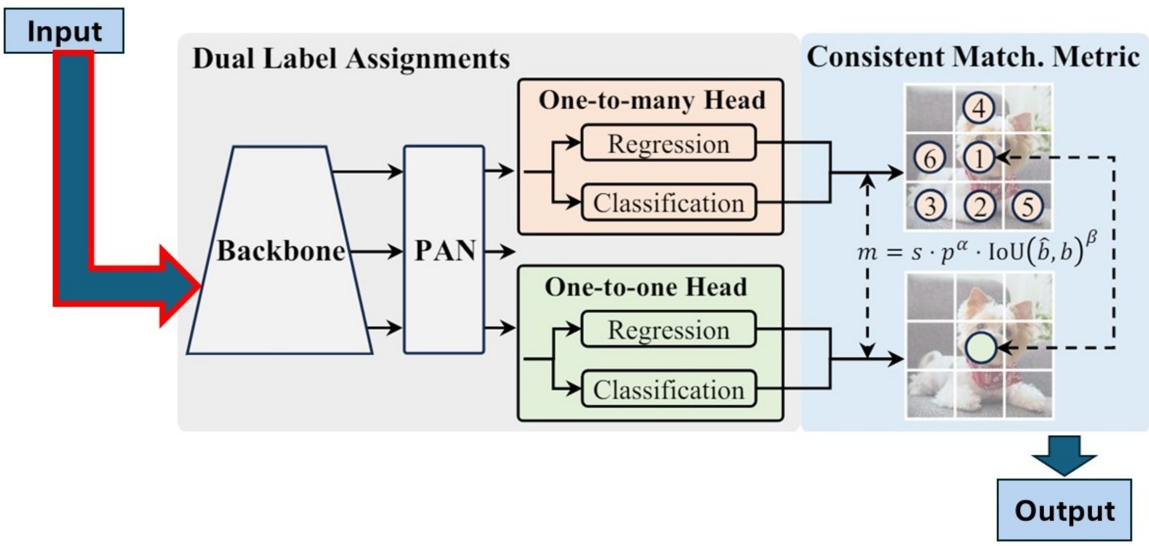


Figure 6. YOLOv10 architecture, which employs a dual label assignment strategy to improve detection accuracy. A backbone processes the input image, while PAN (Path Aggregation Network) enhances feature representation. Employed heads are (1) one-to-many head for regression and classification tasks, and (2) one-to-one head for precise localization [55].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 June 2024 doi:10.20944/preprints202406.1366.v2

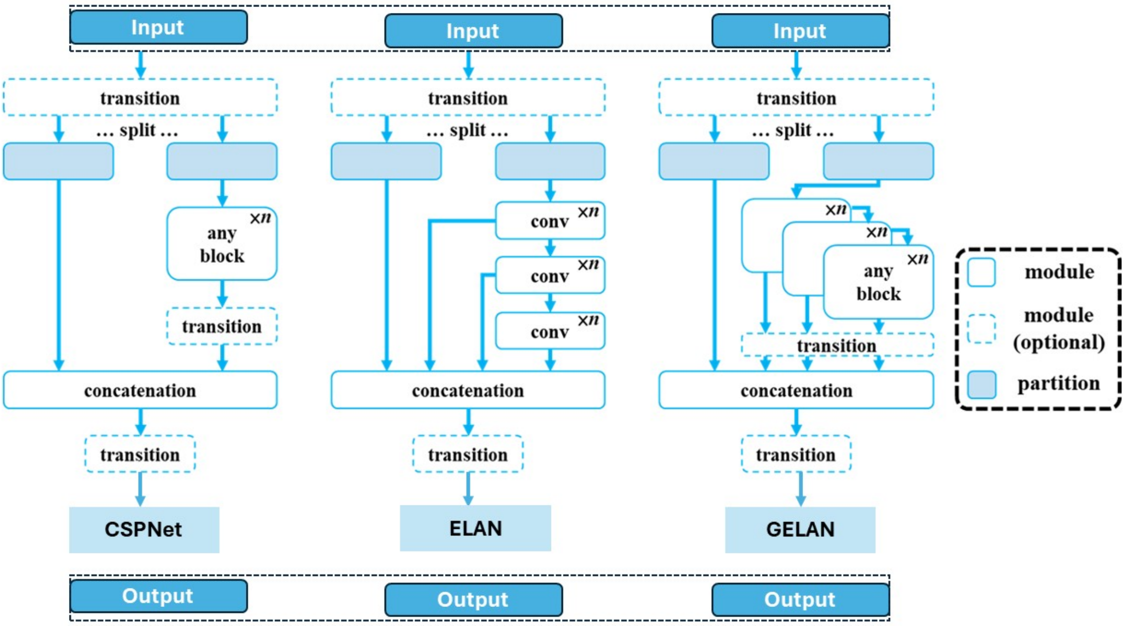


Figure 7. YOLOv9 architecture [56] with CSPNet, ELAN, and GELAN modules. CSPNet enhances gradient flow and reduces computational load through feature map partitioning. ELAN focuses on linear aggregation of features for improved learning efficiency, while GELAN generalizes this approach to combine features from multiple depths and pathways, providing greater flexibility and accuracy in feature extraction.

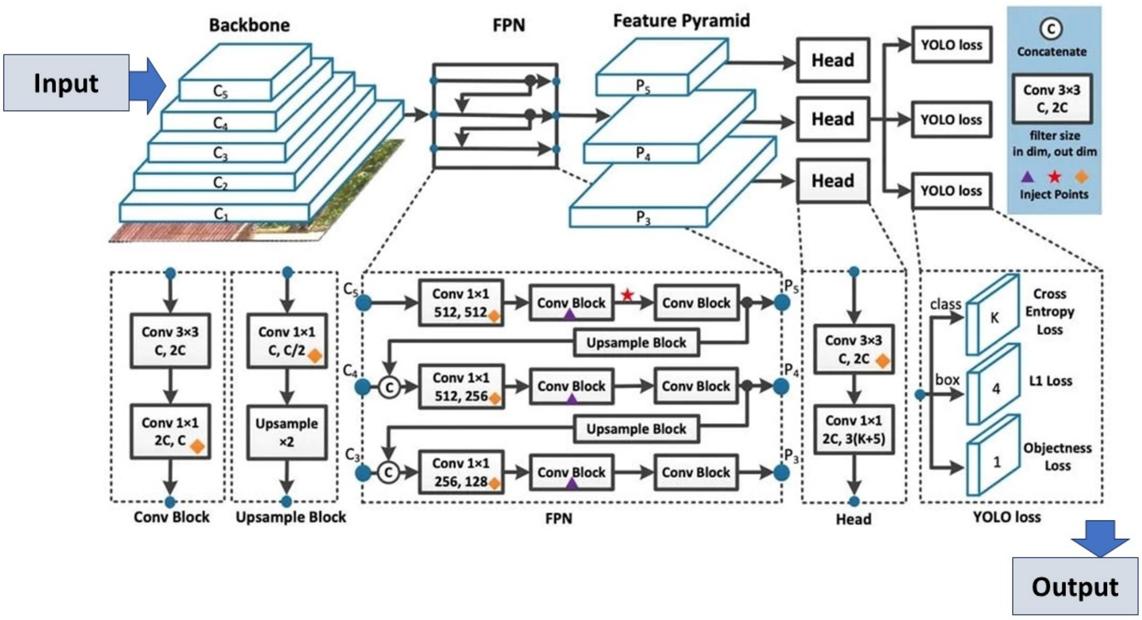


Figure 8. YOLOv8 architecture [92]: showcasing the key components and their connections. The backbone network processes the input image through multiple convolutional layers (C1 to C5), extracting hierarchical features. These features are then passed through the Feature Pyramid Network (FPN) to create a feature pyramid (P3, P4, P5), which enhances detection at different scales. The network heads perform final predictions, incorporating convolutional blocks and upsample blocks to refine features.

The YOLOv10 series delineates a broad array of models, each tailored to specific performance needs within real-time object detection frameworks. Starting with YOLOv10-N (Nano), it demonstrates a rapid detection capability with a mAP of 38.5% at an exceptionally reduced latency to 1.84 ms, making it highly suitable for scenarios demanding quick responses. Progressing through the series, YOLOv10-S (Small) and YOLOv10-M (Medium) offer progressively higher mAP values of 46.3% and 51.1% at latencies of 2.49 ms and 4.74 ms, respectively, providing a balanced performance for versatile applications. The larger variants, YOLOv10-B (Balanced) and YOLOv10-L (Large), cater to environments requiring detailed detections, with mAPs of 52.5% and 53.2% and latencies of 5.74 ms and 7.28 ms respectively. The largest model, YOLOv10-X (Extra Large), excels with the highest mAP of 54.4% at a latency of 10.70 ms, designed for complex detection tasks where precision is paramount. These configurations underscore YOLOv10's adaptability across a spectrum of operational requirements.

Reflecting on YOLO's evolution, starting from YOLOv1, which set the benchmark with an mAP of 63.4% and a latency of 45 ms, to the latest YOLOv10, significant technological strides have been evident. YOLOv10's predecessors, YOLOv9 and YOLOv8, display comparable mAP scores to YOLOv10 but with marginally higher latency, indicating the incremental enhancements YOLOv10 brings to the table. Specifically, YOLOv9 and YOLOv8 models, such as YOLOv9-N and YOLOv8-N, showcase mAPs of 39.5% and 37.3%, respectively, at latency indicative of their generational improvements. Meanwhile, the higher end of these series, YOLOv9-X, and YOLOv8-X, achieve mAPs of 54.4% and 53.9%, respectively, with YOLOv10 outperforming them in efficiency. The YOLO series, from YOLOv1 through YOLOv8, YOLOv9, and now YOLOv10, has continually advanced the frontier of real-time object detection, enhancing both the speed and accuracy of detections, and thus broadening the scope for practical applications in sectors like autonomous driving, surveillance, and real-time video analytics.

The paper on YOLOv9 [56] marks a significant advancement in real-time object detection by addressing the efficiency and accuracy challenges associated with earlier versions, particularly through the mitigation of information loss in deep neural processing. It introduces the innovative Programmable Gradient Information (PGI) and the Generalized Efficient Layer Aggregation Network (GELAN) architecture. These enhancements focus on preserving crucial information across the network, ensuring robust and reliable gradients that prevent data degradation, which is common in deep neural networks [93]. Compared to its successor, YOLOv10, YOLOv9 sets a foundational stage by addressing the information bottleneck problem that typically hinders deep learning models. While YOLOv9's PGI strategically maintains data integrity throughout the processing layers, YOLOv10 builds upon this foundation by completely eliminating the need for NMS and further optimizing model architecture for reduced latency and enhanced computational efficiency. YOLOv10 also introduces dual assignment strategies for NMS-free training, significantly enhancing the system's response time without compromising accuracy, which reflects a direct evolution from the groundwork laid by YOLOv9's innovations [94]. Furthermore, YOLOv9's GELAN architecture represents a pivotal improvement in network design, offering a flexible and efficient structure that effectively integrates multi-scale features. While GELAN contributes significantly to YOLOv9's performance, YOLOv10 extends these architectural improvements to achieve even greater efficiency and adaptability [95]. It reduces computational overhead and increases the model's applicability to various real-time scenarios, showcasing an advanced level of refinement that leverages and enhances the capabilities introduced by YOLOv9.

YOLOv8 was released in January 2023 by Ultralytics, marking a significant progression in the YOLO series with an introduction of multiple scaled versions designed to cater to a wide range of applications [63,96]. These versions included YOLOv8n (nano), YOLOv8s (small), YOLOv8m (medium), YOLOv8l (large), and YOLOv8x (extra-large), each optimized for specific performance and computational needs. This flexibility made YOLOv8 highly versatile, supporting a multitude of vision tasks such as object detection, segmentation, pose estimation, tracking, and classification,

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 June 2024 doi:10.20944/preprints202406.1366.v2

significantly broadening its application scope in real-world scenarios [96]. The architecture of YOLOv8 underwent substantial refinements to enhance its detection capabilities. It retained a similar backbone to YOLOv5 but introduced modifications in the CSP Layer, now evolved into the C2f module—a cross-stage partial bottleneck with dual convolutions that effectively combine high-level features with contextual information to bolster detection accuracy. YOLOv8 transitioned to an anchor-free model with a decoupled head, allowing independent processing of objectness, classification, and regression tasks which, in turn, improved overall model accuracy [97]. The output layer employed a sigmoid activation function for objectness scores and softmax for class probabilities, enhancing the precision of bounding box predictions. YOLOv8 also integrated advanced loss functions like CIoU [98] and Distribution Focal Loss (DFL) [99] for bounding-box optimization and binary cross-entropy for classification, which proved particularly effective in enhancing detection performance for smaller objects. YOLOv8's architecture, demonstrated in detailed diagrams, features the modified CSPDarknet53 backbone with the innovative C2f module, augmented by a spatial pyramid pooling fast (SPPF) layer that accelerates computation by pooling features into a fixed-size map. This model also introduced a semantic segmentation variant, YOLOv8-Seg, which utilized the backbone and C2f module, followed by two segmentation heads designed to predict semantic segmentation masks efficiently. This segmentation model achieved state-of-the-art results on various benchmarks while maintaining high speed and accuracy, evident in its performance on the MS COCO dataset where YOLOv8x reached an AP of 53.9% at 640 pixels image size—surpassing the 50.7% AP of YOLOv5—with a remarkable speed of 280 FPS on an NVIDIA A100 using TensorRT. As we progress backward through the YOLO series, from YOLOv10 to YOLOv8 and soon to YOLOv7, these architectural and functional advancements highlight the series' evolutionary trajectory in optimizing real-time object detection networks.

YOLOv8 was released in January 2023 by Ultralytics, marking a significant progression in the YOLO series with an introduction of multiple scaled versions designed to cater to a wide range of applications [63,96]. These versions included YOLOv8n (nano), YOLOv8s (small), YOLOv8m (medium), YOLOv8l (large), and YOLOv8x (extra-large), each optimized for specific performance and computational needs. This flexibility made YOLOv8 highly versatile, supporting a multitude of vision tasks such as object detection, segmentation, pose estimation, tracking, and classification, significantly broadening its application scope in real-world scenarios [96]. The architecture of YOLOv8 underwent substantial refinements to enhance its detection capabilities. It retained a similar backbone to YOLOv5 but introduced modifications in the CSP Layer, now evolved into the C2f module—a cross-stage partial bottleneck with dual convolutions that effectively combine high-level features with contextual information to bolster detection accuracy. YOLOv8 transitioned to an anchor-free model with a decoupled head, allowing independent processing of objectness, classification, and regression tasks which, in turn, improved overall model accuracy [97]. The output layer employed a sigmoid activation function for objectness scores and softmax for class probabilities, enhancing the precision of bounding box predictions. YOLOv8 also integrated advanced loss functions like CIoU [98] and Distribution Focal Loss (DFL) [99] for bounding-box optimization and binary cross-entropy for classification, which proved particularly effective in enhancing detection performance for smaller objects. This model also introduced a semantic segmentation variant, YOLOv8-Seg [100], which utilized the backbone and C2f module, followed by two segmentation heads designed to predict semantic segmentation masks efficiently. This segmentation model achieved state-of-the-art results on various benchmarks while maintaining high speed and accuracy, evident in its performance on the MS COCO dataset where YOLOv8x reached an AP of 53.9% at 640 pixels image size—surpassing the 50.7% AP of YOLOv5—with a remarkable speed of 280 FPS on an NVIDIA A100 using TensorRT. As we progress backward through the YOLO series, from YOLOv10 to YOLOv8 and soon to YOLOv7, these architectural and functional advancements highlight the series' evolutionary trajectory in optimizing real-time object detection networks.

 doi:10.20944/preprints202406.1366.v2

4.2. YOLOv7, YOLOv6 and YOLOv5

The YOLOv7 model introduces enhancements in object detection tailored for drone-captured scenarios, particularly through the Transformer Prediction Head (TPH-YOLOv5) variant [101], which emphasizes improvements in handling scale variations and densely packed objects [58]. By incorporating TPH and the Convolutional Block Attention Module (CBAM) [102], YOLOv7 substantially boosts its capacity to focus on relevant regions in cluttered environments. These features particularly enhance the model's ability to detect objects across varied scales, an essential trait for drone applications where altitude changes affect object size perception drastically. The model integrates sophisticated strategies like multi-scale testing [103] and a self-trained classifier, which refines its performance on challenging categories by specifically addressing common issues in drone imagery such as motion blur and occlusion. These adaptations have shown notable improvements, with YOLOv7 achieving competitive results in drone-specific datasets and challenges [104]. The model's adaptability and robustness in such specialized conditions demonstrate its potential beyond conventional settings, catering effectively to next-generation applications like urban surveillance and wildlife monitoring.

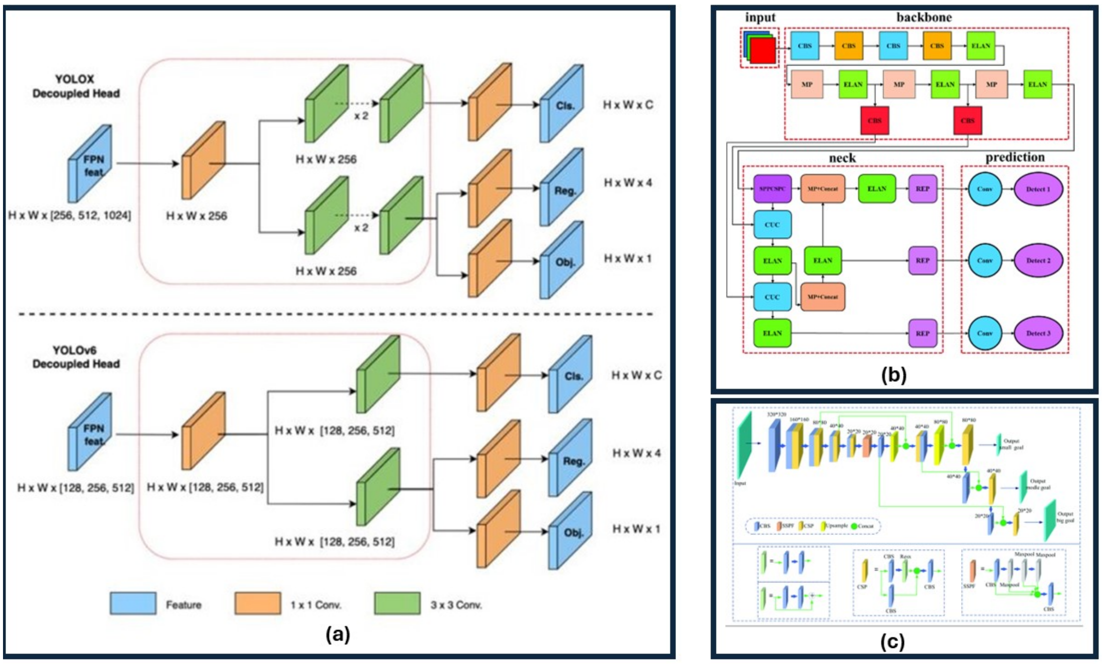


Figure 9. Comparative architectures of YOLOv5 [105], YOLOv6 [106], and YOLOv7 [107]. (a) Decoupled head structures for YOLOv5 and YOLOv6, showing feature extraction from the Feature Pyramid Network (FPN) and subsequent classification (Cls.), regression (Reg.), and objectness (Obj.) predictions. (b) Detailed backbone, neck, and prediction modules of YOLOv7, highlighting ELAN and other components. (c) Overall pipeline of YOLOv5, including backbone, detection heads, and feature extraction blocks, showcasing the architectural advancements across versions.

YOLOv6 emerges as a robust solution in industrial applications by delivering a finely balanced trade-off between speed and accuracy, crucial for deployment across various hardware platforms [59]. It iterates on previous versions by incorporating cutting-edge network designs, training strategies, and quantization techniques to enhance its efficiency and performance significantly. This model has been optimized for diverse operational requirements with its scalable architecture, ranging from YOLOv6-N to YOLOv6-X, each offering different levels of performance to suit specific computational budgets [108]. Significant innovations in YOLOv6 include the use of advanced label assignment techniques and loss functions that refine the model's predictive accuracy and operational efficiency. By leveraging state-of-the-art advancements in machine learning, YOLOv6 not only excels in traditional object detection metrics but also sets new standards in throughput and latency, making it exceptionally suitable for real-time applications in industrial and commercial domains.

The subsequent versions of YOLO, namely YOLOv6 and YOLOv7 each introduce innovative features that build on the foundation set by YOLOv5. YOLOv6, released in October 2021, introduced lightweight nano models optimized for mobile and CPU environments, alongside a more effective backbone for improved small object detection. YOLOv7 further advanced this development by incorporating a new backbone network, PANet [109], enhancing feature aggregation and representation, and introducing the CIOU loss function for better object scaling and aspect ratio handling. YOLO-v6 significantly shifts the architecture to an anchor-free design, incorporating a self-attention mechanism to better capture long-range dependencies and employing adaptive training techniques to optimize performance during training [110]. These versions collectively push the boundaries of object detection performance, emphasizing speed, accuracy, and adaptability across a range of deployment scenarios.

YOLOv5 has significantly contributed to the YOLO series evolution, focusing on user-friendliness and performance enhancements [64,65]. Its introduction by Ultralytics brought a streamlined, accessible framework that lowered the barriers to implementing high-speed object detection across various platforms. YOLOv5's architecture incorporates a series of optimizations including improved backbone, neck, and head designs which collectively enhance its detection capabilities. The model supports multiple size variants, facilitating a broad range of applications from mobile devices to cloud-based systems [64]. YOLOv5's adaptability is further evidenced by its continuous updates and community-driven enhancements, which ensure it remains at the forefront of object detection technologies. This version stands out for its balance of speed, accuracy, and utility, making it a preferred choice for developers and researchers looking to deploy state-of-the-art detection systems efficiently.

YOLOv5 marks a significant evolution in the YOLO series, focusing on production-ready deployments with streamlined architecture for real-world applications. This version emphasizes reducing the model's complexity by refining its layers and components, enhancing its inference speed without sacrificing detection accuracy. The backbone and feature extraction layers were optimized to accelerate processing, and the network's architecture was simplified to facilitate faster data throughput. Importantly, YOLO v5 enhances its deployment flexibility, catering to edge devices with limited computational resources through model modularity and efficient activations. These architectural refinements ensure YOLO v5 operates effectively in diverse environments, from high-resource servers to mobile devices, making it a versatile tool in the arsenal of object detection technologies.

4.3. YOLOv4, YOLOv3, YOLOv2 and YOLOv1

The introduction of YOLOv4 [62] in 2020 marked the latest in these developments, employing CSPDarknet-53 [111]as its backbone. This modified version of Darknet-53 uses Cross-Stage Partial connections to reduce computational demands while enhancing learning capacity. YOLOv4 incorporates innovative features such as Mish activation [112], replacing traditional ReLU to maintain smooth gradients, and utilizes new data augmentation techniques like Mosaic and CutMix [113]. Additionally, it introduces advanced regularization methods including DropBlock regularization [114] and Class Label Smoothing to prevent overfitting [115], alongside optimization strategies termed BoF (Bag of Freebies) [116] and BoS (Bag of Specials) that enhance training and inference efficiency. Following the success of YOLOv4, YOLOv3 was introduced in 2018, which utilized the Darknet-53 architecture with influences from residual learning. This version was trained initially on ImageNet, helping it to effectively detect objects across various sizes due to its multi-scale detection capabilities within the architecture.

YOLOv3 [61] improved detection accuracy, especially for small objects, through its use of three different scales for detection, thereby capturing essential features at various resolutions. Earlier, YOLOv2 and the original YOLO (YOLOv1) laid the groundwork for these advancements [117]. Released in 2016, YOLOv2 introduced a new 30-layer architecture with anchor boxes from Faster R-CNN and batch normalization [118] to speed up convergence and enhance model performance. YOLOv1, debuting in

2015 by Joseph Redmon, revolutionized object detection with its single-shot mechanism that predicted bounding boxes and class probabilities in one network pass, utilizing a simpler Darknet-19 architecture. This initial approach significantly accelerated the detection process, establishing the foundational techniques that would be refined in later versions of the YOLO series.

The introduction of YOLOv4 [62] in 2020 marked the latest in these developments, employing CSPDarknet-53 [111]as its backbone. This modified version of Darknet-53 uses Cross-Stage Partial connections to reduce computational demands while enhancing learning capacity. YOLOv4 incorporates innovative features such as Mish activation [112], replacing traditional ReLU to maintain smooth gradients, and utilizes new data augmentation techniques like Mosaic and CutMix [113]. Additionally, it introduces advanced regularization methods including DropBlock regularization [114] and Class Label Smoothing to prevent overfitting, alongside optimization strategies termed BoF (Bag of Freebies) and BoS (Bag of Specials) that enhance training and inference efficiency. Following the success of YOLOv4, YOLOv3 was introduced in 2018, which utilized the Darknet-53 architecture with influences from residual learning. This version was trained initially on ImageNet, helping it to effectively detect objects across various sizes due to its multi-scale detection capabilities within the architecture. YOLOv3 improved detection accuracy, especially for small objects, through its use of three different scales for detection, thereby capturing essential features at various resolutions. Earlier, YOLOv2 and the original YOLO (YOLOv1) laid the groundwork for these advancements. Released in 2016, YOLOv2 introduced a new 30-layer architecture with anchor boxes from Faster R-CNN and batch normalization to speed up convergence and enhance model performance. YOLOv1, debuting in 2015 by Joseph Redmon, revolutionized object detection with its single-shot mechanism that predicted bounding boxes and class probabilities in one network pass, utilizing a simpler Darknet-19 architecture. This initial approach significantly accelerated the detection process, establishing the foundational techniques that would be refined in later versions of the YOLO series. YOLOv4 and YOLOv3, showcasing their advanced architectures and features, are illustrated in Figure 10a and b, respectively, while YOLOv2 and YOLOv1 are depicted in Figure 11a and b, showcasing the foundational developments in the series.

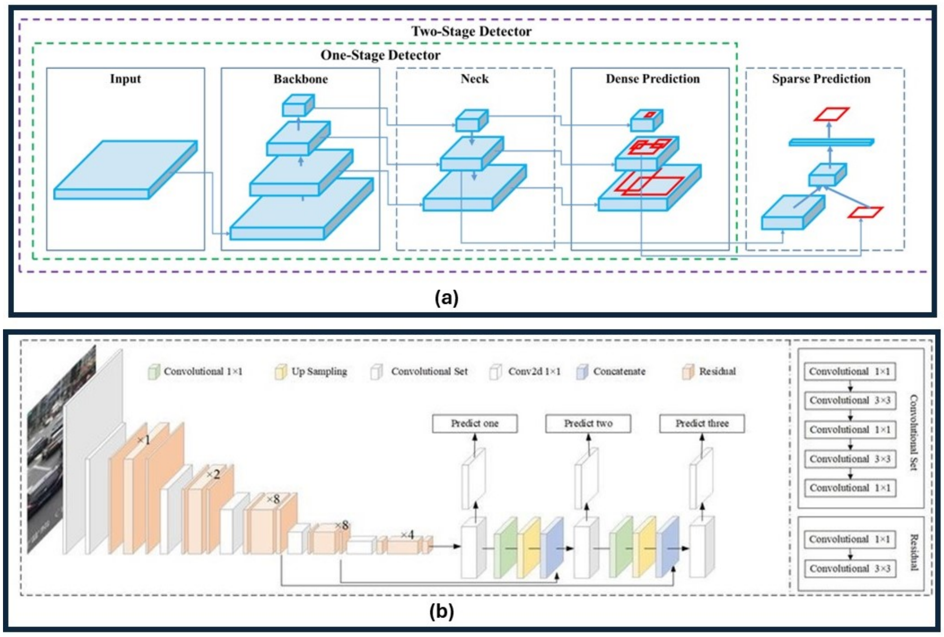


Figure 10. Comparison of YOLOv4 [62] and YOLOv3 [61] architectures. (a) YOLOv4 architecture showing a two-stage detector with backbone, neck, dense prediction, and sparse prediction modules. (b) YOLOv3 architecture featuring convolutional and upsampling layers leading to multi-scale predictions. This highlights the structural advancements in object detection between the two versions

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 June 2024 doi:10.20944/preprints202406.1366.v2

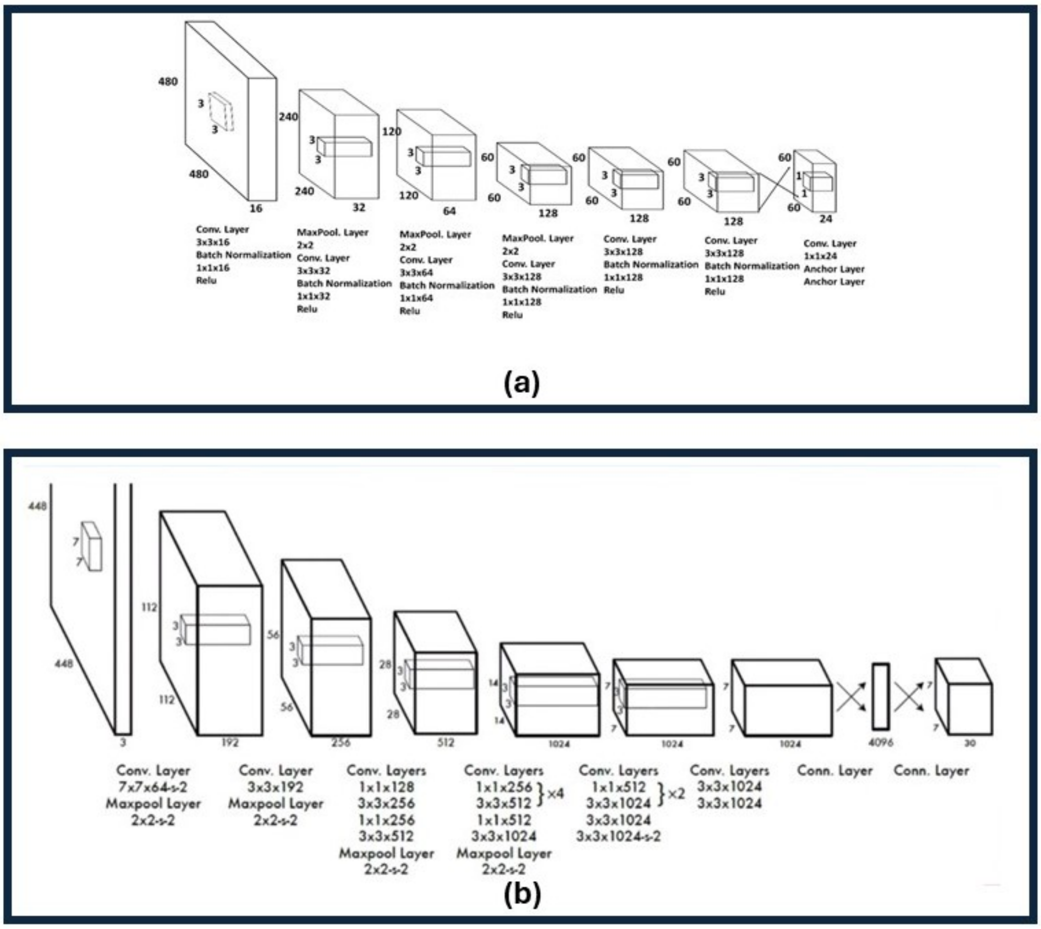


Figure 11. Comparison of YOLOv1 [117] and YOLOv2 [60] architectures. (a) YOLOv1 architecture, showing the sequence of convolutional layers, max-pooling layers, and fully connected layers used for object detection. This model performs feature extraction and prediction in a single unified step, aiming for real-time performance. (b) YOLOv2 architecture, illustrating improvements such as the use of batch normalization, higher resolution input, and anchor boxes.

5. Applications

YOLO has many reacl-time practical applications, such as, autonomous vehicles for obstacle detection and traffic sign recognition, enhancing safety and navigation [35]. Additionally, YOLO is employed in surveillance for intrusion detection and anomaly identification, and in healthcare for detecting anomalies in medical images, aiding in accurate and efficient diagnostics [119].

5.1. Autonomous Vehicles

Each YOLO version has been pivotal in advancing the capabilities of autonomous vehicles by providing highly efficient and accurate real-time detection systems. Each iteration of YOLO has brought improvements that enhance the vehicle's ability to perceive its environment quickly and accurately, which is critical for safe navigation and decision-making [120]. Starting with YOLOv1 [26], the YOLO algorithm revolutionized the approach by performing detection tasks directly from full images in a single network pass, allowing for the detection of objects at a remarkable speed [121]. This initial model was pivotal, setting a high standard for real-time object detection and establishing a framework that future versions would build upon. Subsequent iterations, including YOLOv2 and YOLOv3, continued to refine this approach by introducing concepts such as real-time multi-scale processing and improved anchor box adjustments, which enhanced the accuracy and robustness of the detections. These versions were particularly adept at handling the variable scales of objects seen

 doi:10.20944/preprints202406.1366.v2

in driving environments—from nearby pedestrians to distant road signs—making them invaluable for autonomous driving applications. YOLOv4 and later versions further pushed the boundaries by integrating advanced neural network techniques and optimizations that improved detection accuracy while maintaining the high-speed processing necessary for real-time applications [122,123]. These advancements in YOLO technology have not only bolstered the capabilities of autonomous vehicles in terms of environmental perception and decision-making but have also significantly contributed to advancements in automotive safety and operational reliability [124].

Ye et al. (2022) developed an end-to-end adaptive neural network control for autonomous vehicles that predicts steering angles using YOLOv5, enhancing vehicle navigation precision [125]. Mostafa et al. (2022) compared the effectiveness of YOLOv5, YOLOX, and Faster R-CNN in detecting occluded objects for autonomous vehicles, improving detection reliability [126]. Jia et al. (2023) proposed an enhanced YOLOv5 detector for autonomous driving, which offers increased speed and accuracy [127]. Chen et al. (2023) utilized an improved YOLOv5-OBB algorithm for autonomous parking space detection in electric vehicles, enhancing operational efficiency [128]. Liu and Yan (2022) customized YOLOv7 for vehicle-related distance estimation, providing essential metrics for safe navigation [129]. Mehla et al. (2023) evaluated YOLOv8 against EfficientDet in autonomous maritime vehicles, highlighting the superior detection capabilities of YOLOv8 [130]. Patel et al. (2024) enhanced traffic sign detection using YOLOv8, promoting safer driving environments [131].

YOLOv8 and YOLOv9 are at the forefront of transforming the landscape of autonomous vehicle technologies, playing a pivotal role in enhancing the operational safety and efficiency of self-driving cars. These models have excelled in real-time object detection, a crucial aspect of autonomous driving, especially under the challenging and variable conditions typical in real-world traffic environments. For instance, in the Robotaxi-Full Scale Autonomous Vehicle Competition, YOLOv8 was specifically adapted to recognize and interpret traffic signs, providing real-time alerts that are essential for safe driving [132]. Moreover, YOLOv8-QSD, an enhanced version, addresses the need for detecting smaller objects such as traffic signs and signals, demonstrating its utility with a notable accuracy rate and efficiency in processing, making it ideal for high-speed driving scenarios [133].

Further advancements with YOLOv8 have led to significant improvements in object detection in adverse weather conditions, an area of particular concern for autonomous driving. The application of transfer learning techniques using datasets from diverse weather conditions has markedly increased the detection performance of YOLOv8, ensuring reliable recognition of crucial road elements like pedestrians and obstacles under challenging weather scenarios [134]. Additionally, the development of YOLOv8 for specific tasks such as brake light status detection illustrates the algorithm's flexibility and its potential in enhancing interpretability and safety for autonomous vehicles [135]. These innovations underscore the critical role of YOLOv8 and YOLOv9 in pushing the boundaries of what is possible in the autonomous vehicle industry, highlighting their impact in meeting the rigorous demands for safety and reliability in self-driving technologies [136]. Table 3 illustrates different applications of YOLO in the autonomous vehicle industry, presented in reverse chronological order from the most recent versions to the older ones.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 June 2024 doi:10.20944/preprints202406.1366.v2

Table 3. Studies on YOLO applications in autonomous vehicles, focusing on object detection and real-time performance improvements for enhanced safety.

Title of Paper	Description of Work	Purpose and YOLO Usage	Version	Ref. and Year
"Transforming Aircraft Detection Through LEO Satellite Imagery and YOLOv9 for Improved Aviation Safety"	Utilizes YOLOv9 with LEO satellite imagery for enhanced detection of aircraft in wide-area airport environments.	Aims to improve airport security and aviation safety by integrating advanced YOLO-based object detection with satellite imagery.	YOLOv9	[137], 2024
"YOLOv8-QSD: An Improved Small Object Detection Algorithm for Autonomous Vehicles Based on YOLOv8"	Developed an anchor-free, BiFPN-enhanced YOLOv8 model for better small object detection in driving scenarios.	Enhances detection of small objects for autonomous vehicles with reduced computational demands, tested on SODA-A dataset.	YOLOv8-QSD	[133], 2024
"Object Detection in Dense and Mixed Traffic for Autonomous Vehicles With Modified Yolo"	Adapted YOLOv7 with deformable layers and softNMS for object detection in heavy Indonesian traffic.	Enhances detection and classification of objects around autonomous vehicles using a modified YOLOv7, tested on a novel Indonesian traffic dataset.	YOLOv7-MOD	[138], 2023
"Local Regression Based Real-Time Traffic Sign Detection using YOLOv6"	Utilized YOLOv6 with a Logistic Regression classifier for enhanced traffic sign detection.	Improves real-time traffic sign identification using YOLOv6 optimized for embedded systems and smartphones. Tested on ITSDBD.	YOLOv6	[139], 2022
"Small-object detection based on YOLOv5 in autonomous driving systems"	Investigated and refined YOLOv5 for improved detection of small objects like traffic signs and traffic lights.	Enhances small object detection through architectural adjustments to YOLOv5, tested on BDD100K, TT100K, and DTLD datasets.	YOLOv5	[140], 2023
"Deep convolutional neural network for enhancing traffic sign recognition developed on Yolo V4"	Analyzed YOLO V4 and YOLO V4-tiny with SPP for better feature extraction in traffic sign recognition.	Compares the enhancement of traffic sign recognition performance by integrating SPP into YOLO V4 backbones.	YOLOv4	[141], 2022
"The improvement in obstacle detection in autonomous vehicles using YOLO non-maximum suppression fuzzy algorithm"	Employed a hybrid of fuzzy logic and NMS in YOLO for better obstacle detection in autonomous driving.	Enhances obstacle detection accuracy and speed using a modified YOLO algorithm.	YOLOv3	[142], 2021
"Object Tracking for Autonomous Vehicle Using YOLOV3"	Evaluated YOLOv3 for object tracking in autonomous vehicles.	Two models were provided, one trained using only the online COCO dataset, and the other trained with additional images from various locations at Universiti Malaysia Pahang (UMP).	YOLOv3	[143], 2022

 doi:10.20944/preprints202406.1366.v2

5.2. Healthcare and Medical Imaging

YOLO has marked a significant technological advancement, especially with the introduction of newer versions such as YOLOv7 and YOLOv8 [144–146]. The recent iterations of YOLO, particularly YOLOv7, YOLOv8, and YOLOv9, could significantly enhance medical diagnostics by offering advanced computational efficiency and improved feature extraction capabilities, making them suitable for real-time medical imaging applications. Such capabilities are crucial in urgent care scenarios, where swift diagnosis can be pivotal. For instance, YOLOv8's sophisticated algorithms excel in accurately delineating complex biological structures, vital for identifying pathologies in conditions like vascular diseases or tumors. Similarly, YOLOv9's rapid processing power enables immediate analysis of medical images, essential in emergency medical responses where timely intervention is critical. These versions have the potential to revolutionize healthcare by facilitating early detection of diseases and supporting continuous patient monitoring, transforming the traditional approach of healthcare diagnostics into one that integrates accurate, swift diagnostics seamlessly with routine medical examinations. Unlike the traditional methods which depend heavily on manual annotation and are prone to errors and subjectivity, YOLO algorithms automate the detection and localization of medical anomalies such as tumors, lesions, and other pathological markers across various imaging modalities. This automation is driven by YOLO's unique architecture that efficiently predicts multiple bounding boxes and class probabilities in a single analysis, enhancing diagnostic accuracy and reducing the potential for human error.

In the field of medical imaging and diagnostics, the adoption of the YOLO object detection algorithm has showcased promising improvements in accuracy and efficiency, particularly with its latest versions like YOLOv5, YOLOv6, YOLOv7, and YOLOv8. For instance, Luo et al. (2021) leveraged YOLOv5 in conjunction with ResNet50 to enhance chest abnormality detection, demonstrating the algorithm's proficiency in identifying subtle medical conditions [147]. Similarly, Wu et al. (2022) developed Me-YOLO, an adapted version of YOLOv5, to improve the detection of medical personal protective equipment, highlighting the model's adaptability to varied medical use cases [148]. Moreover, advancements like the CSFF-YOLOv5 by Zhao et al. (2024) introduced modifications for better feature fusion, significantly boosting the detection accuracy in femoral neck fracture cases [149]. This specificity is further explored by Goel and Patel (2024), who enhanced YOLOv6 for lung cancer detection using an advanced PSO optimizer, underscoring the potential of YOLO algorithms in facilitating early disease diagnosis and treatment [150]. Additionally, the extension of YOLOv6 by Norkobil Saydirasulovich et al. (2023) for improved fire detection in smart city environments exemplifies the algorithm's versatility beyond traditional medical applications, proving its efficacy in diverse environmental conditions [151]. Each of these developments not only enhances specific medical diagnostic processes but also paves the way for integrating these advanced object detection systems into broader healthcare applications, as illustrated by the innovative uses of YOLOv7 and YOLOv8 in detecting whole body bone fractures and enhancing hospital efficiency [152,153]. These studies collectively demonstrate the significant advancements brought by YOLO in the healthcare sector, ensuring more precise, efficient, and versatile diagnostic solutions.

Recent versions such as YOLOv7, YOLOv8 and YOLOv9 have been effectively demonstrated across a variety of healthcare applications. Razaghi et al. (2024) utilized YOLOv8 for the innovative diagnosis of dental diseases, highlighting its precision in identifying dental pathologies [154]. Similarly, Pham and Le (2024) leveraged YOLOv8 for the detection and classification of ovarian tumors from ultrasound images, showcasing the model's adaptability to different medical imaging modalities [155]. Krishnamurthy et al. (2023) applied custom YOLO architectures to enhance object detection capabilities during endoscopic surgeries, illustrating the potential of YOLO in surgical settings [156]. Furthermore, Palanivel et al. (2023) discussed the application of YOLOv8 in cancer diagnosis through medical imaging, further cementing YOLO's role in critical healthcare applications [157].

Continuing with advancements, Karaköse et al. (2024) introduced CSFF-YOLOv5, an improved YOLO model for femoral neck fracture detection, utilizing advanced feature fusion techniques [158].

Inui et al. (2023) demonstrated YOLOv8's effectiveness in detecting elbow osteochondritis dissecans in ultrasound images, which supports its use in orthopedic diagnostics [146]. Bhojane et al. (2023) employed YOLOv8 for detecting liver lesions from MRI and CT images, underscoring the algorithm's capability across various imaging technologies [159]. Additionally, Zhang et al. (2023) developed an improved detection model for microaneurysms using YOLOv8, which illustrates continuous enhancements in YOLO's application to highly specific medical tasks [94].

Table 4 illustrates the different uses of YOLO versions in security and survelliance:

5.3. Security and Surveilliance

In the ever-evolving field of security systems, YOLO's application extends to detecting unauthorized entries and identifying potential threats swiftly, thereby bolstering security measures [160]. Recent YOLO models such as YOLOv6 build on this by improving detection accuracy through deeper network layers that process images with greater precision [161]. Meanwhile, YOLOv7 offers advanced customization options that allow security systems to be finely tuned to specific surveillance needs, enhancing the adaptability and effectiveness of threat detection [161,162]. These YOLO versions support high-resolution video feeds, ensuring that security personnel can engage with real-time data to make informed decisions quickly. Further advancements in surveillance systems are embodied by YOLOv8 and YOLOv9, which introduce significant innovations in deep learning for security applications [163,164]. YOLOv8's architecture is designed to handle complex environments where traditional surveillance systems may fail, such as varying lighting and weather conditions. This version's robust performance in diverse scenarios enhances its utility in comprehensive security strategies. On the other hand, YOLOv9 pushes the boundaries of speed and accuracy, providing unparalleled real-time analysis and detection capabilities. Its deployment in surveillance systems ensures that even the subtlest anomalies are detected, reducing the likelihood of security breaches. The integration of recent versions of YOLO such as YOLOv8 and YOLOv9 into security frameworks not only streamlines operations but also ensures a proactive approach to threat management, keeping public and private spaces safer across the globe [165,166].

The application of YOLO models in surveillance and security systems highlights their pivotal role in enhancing real-time response and precision. Majeed et al. [160] investigated the effectiveness of a YOLOv5-based security system within a real-time environment, underscoring its capability to significantly improve operational efficiency in dynamic settings. Similarly, Affes et al. [161] conducted a comparative study across YOLOv5, YOLOv6, YOLOv7, and YOLOv8, focusing on their performance in intelligent video surveillance systems. Their analysis demonstrated the incremental improvements in detection accuracy and processing speed, crucial for real-time security applications. Further advancing the field, Cao and Ma [162] utilized a refined YOLOv7 model to enhance campus security through improved target detection capabilities, highlighting the model's precision in identifying potential threats in densely populated environments. Chatterjee et al. [163] introduced a YOLOv8-based intrusion detection system specifically tailored for physical security and surveillance, which significantly contributes to safeguarding assets and individuals by detecting unauthorized entries or activities effectively. Additionally, Sandhya and Kashyap [164] employed YOLOv8 for real-time object-removal tampering localization in surveillance videos, a crucial technology for maintaining the integrity of video evidence and ensuring the reliability of surveillance feeds. Together, these studies showcase the robustness of YOLO architectures in addressing diverse and complex security challenges, providing substantial improvements in both the efficacy and efficiency of surveillance operations.

Recent studies have significantly leveraged advanced YOLO models to enhance surveillance and security across various domains. Bakirci and Bayraktar [165] discussed optimizing ground surveillance for aircraft monitoring using YOLOv9, highlighting its efficacy in real-time security applications. Similarly, Chakraborty et al. [167] explored a multi-model approach for violence detection, incorporating YOLOv8 to improve public safety through automated surveillance. These advancements indicate a shift towards reliable and efficient security systems for complex scenarios.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 June 2024 doi:10.20944/preprints202406.1366.v2

Table 4. Studies on YOLO applications in healthcare and medicine, emphasizing object detection for diagnostic imaging and real-time medical analysis

Title of Paper	Description of Work	Purpose and YOLO Usage	Version	Ref. and Year
"Efficient Skin Lesion Detection using YOLOv9 Network"	Utilized YOLOv9 for advanced skin lesion detection, leveraging deep learning to enhance diagnostic accuracy and speed.	Developed improved skin lesion identification using YOLOv9, showcasing significant advances in detection performance.	YOLOv9	[145], 2023
"Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm"	Employed YOLOv8 with data augmentation on the GRAZPEDWRI-DX dataset for detecting fractures in pediatric wrist X-ray images.	Enhanced fracture detection in pediatric wrist trauma using YOLOv8, achieving superior mAP compared to previous versions. Designed an app for surgical use.	YOLOv8	[145], 2023
"Chapter 4 - Medical image analysis of masses in mammography using deep learning model for early diagnosis of cancer tissues"	Utilizes YOLOv7 to detect and diagnose cancerous tissues in mammogram images, leveraging advancements in deep learning for early cancer detection.	Aims to enhance early detection of breast cancer using YOLOv7, improving diagnostic accuracy with deep learning integration. Performance measured by Precision, Recall, and F1-score.	YOLOv7	[168], 2024
"Improving YOLOv6 using advanced PSO optimizer for weight selection in lung cancer detection and classification"	Enhanced YOLOv6 with Particle Swarm Optimization for weight optimization in lung cancer detection from CT scans.	Utilized advanced PSO to optimize YOLOv6 for higher accuracy in detecting lung cancer, significantly outperforming previous methods on the LUNA 16 Dataset.	YOLOv6	[150], 2024
"One-Stage methods of computer cision object detection to classify carious lesions from smartphone imaging"	Utilized YOLO v5, YOLO v5X, and YOLO v5M to detect and classify carious lesions from smartphone images.	Aimed to automate caries detection with enhanced accuracy using YOLO. mAP, P, and R metrics validated performance.	YOLOv5, YOLOv5X, YOLOv5M	[169], 2023
"An Improved Method of Polyp Detection Using Custom YOLOv4-Tiny"	Customized YOLOv4-tiny with Inception-ResNet-A block for enhanced detection of polyps in wireless endoscopic images.	Developed to imporve the detection performance of polyp detection using a modified YOLOv4-tiny. Demonstrated significant perforamnce improvement.	YOLOv4-Tiny	[170], 2022
"Detection of dental caries in oral photographs taken by mobile phones based on the YOLOv3 algorithm"	Utilized YOLOv3 for detecting dental caries from mobile phone images, employing image augmentation and enhancement for improved accuracy.	Enhanced detection and diagnosis of dental caries using YOLOv3, with evaluation of diagnostic precision, recall, and F1-score across different datasets.	YOLOv3	[171], 2021
"Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network"	Employed YOLOv2 for automatic detection and diagnosis of thyroid nodules in ultrasound images, enhancing diagnostic precision.	Compared AI performance with radiologists using YOLOv2, showing improved accuracy and specificity in thyroid nodule diagnosis. ROC curve analysis confirms effectiveness.	YOLOv2	[172], 2019
"Real-Time Facial Features Detection from Low Resolution Thermal Images with Deep Classification Models"	Developed a method to localize facial features from low-resolution thermal images by modifying existing deep classification networks for real-time detection.	Demonstrates how spatial information can be restored and utilized from classification models for facial feature detection, significantly reducing dataset preparation time while maintaining high precision.	Custom Deep Classification Model and YOLO	[173], 2018

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 June 2024 doi:10.20944/preprints202406.1366.v2

Chen et al. [174] delve into the application of an enhanced YOLOv8 model for large-scale security and low-altitude drone-based law enforcement, demonstrating its potential in managing security risks effectively. Further, Pashayev et al. [175] utilize YOLO8 for intelligent face recognition in smart cameras, contributing to the development of smarter, more responsive surveillance technologies. Additionally, Kaç et al. [176] investigate image-based security techniques for critical water infrastructure surveillance, employing YOLO models to ensure robust monitoring. Lastly, Gao et al. [177] introduce an improved YOLOv8s network model for contraband detection in X-ray images, underscoring the versatility and precision of YOLO models in enhancing contraband security measures.

Recent advancements in surveillance technologies have leveraged the YOLO's capabilities, particularly in managing crowd dynamics and detecting critical events. Antony et al. [178] explored the use of YOLOv8 alongside ByteTrack for crowd management, emphasizing the system's efficiency in improving surveillance and public safety. This integration marks a significant step towards enhancing real-time monitoring capabilities during large public gatherings. Concurrently, Zhang [179] utilized a YOLO model to detect fire and smoke in IoT surveillance systems, showcasing the model's ability to respond swiftly to emergency situations, thus bolstering safety protocols within environments.

In security, Khin et al. [180] conducted a comparative study of YOLOv8 with other models like RetinaNet and EfficientDet for gun detection, emphasizing YOLOv8's superior accuracy in detecting firearms within a custom dataset. It underlines the critical role of precise object detection to prevent potential threats. Additionally, Nkuzo et al. [181] provided a comprehensive analysis of the YOLOv7 in detecting car safety belts in real-time, illustrating its importance in enforcing road safety measures. Moreover, Chang et al. [182] developed an improved YOLOv7, equipped with feature fusion and attention mechanisms, tailored for detecting safety gear violations in high-risk environments like construction, to enhance workplace safety standards. Table 5 presents the various YOLO usage in security and surveillance.

5.4. Manufacturing

In the landscape of industrial manufacturing, the deployment of YOLO algorithms significantly enhances the capability of automated optical inspection (AOI) systems. Each iteration of the YOLO family, from YOLOv2 to YOLOv5, and beyond into the latest versions like YOLOv6 and YOLOv7, brings forward substantial improvements in detecting defects across various manufacturing domains [85,183–185]. The high accuracy and real-time processing capabilities of YOLOv6 and YOLOv7, for instance, allow for immediate identification of production flaws, crucial for maintaining workflow efficiency on fast-paced production lines [186,187]. Advancing into the domain of smart manufacturing, YOLO algorithms are pivotal in revolutionizing quality control mechanisms [188,189]. The continuous evolution from YOLOv5 to YOLOV6, YOLOv7, YOLOv8,and upto 10th version of YOLO exemplifies the adaptation of deep learning to meet the stringent quality demands of modern manufacturing processes. These algorithms reduce the need for labor-intensive manual inspections, thereby minimizing the margin for human error and enhancing the overall speed of quality assessments [85,183–185,188,189].

For instance, [190] pioneered YOLO-IMF, an enhanced version of YOLOv8 tailored for precise surface defect detection in industrial settings, exemplifying the algorithm's efficacy in real-time environments. This refinement aims to cater to the high demands for accuracy in manufacturing sectors where defects can significantly impact quality and safety. Continuing this trend, [191] introduced Yolo-SD, which utilizes simulated feature fusion for few-shot learning, enhancing YOLOv8's capability in detecting industrial defects under varied conditions. Similarly, [192] extended YOLOv8's utility in monitoring 3D printing processes by optimizing hyperparameters to detect faults more accurately, reflecting a targeted approach to maintaining production integrity. [193] adapted YOLOv8 to inspect cylindrical parts, a critical aspect of quality control in specialized manufacturing. Lastly, [194] leveraged a conditioned version of YOLOv8, named Cond-YOLOv8-seg, to assess the uniformity of industrially produced materials, showcasing the model's versatility across different manufacturing scenarios. These innovations underscore the pivotal role of YOLO algorithms in driving forward the capabilities of industrial inspection systems, highlighting their impact on enhancing operational efficiency and product quality.

Table 5. Studies on YOLO usage in security and surveillance, for real-time threat detection and enhanced monitoring to improved safety measures

Title of Paper	Description of Work	Purpose and YOLO Usage	Version	Ref. and Year
"YOLOv9-Enabled Vehicle Detection for Urban Security and Forensics Applications"	Implements YOLOv9 for aerial vehicle detection via UAVs, enhancing urban security and forensic capabilities.	Focus on utilizing YOLOv9 for real-time vehicle monitoring, facilitating efficient law enforcement and forensic analysis in urban settings.	YOLOv9	[166], 2024
"SC-YOLOv8: A Security Check Model for the Inspection of Prohibited Items in X-ray Images"	Developed a custom YOLOv8 model for X-ray image analysis to detect prohibited items. Enhanced model accuracy using a novel backbone structure and data augmentation.	Aimed to improve security screening effectiveness and reduce error rates in detecting prohibited items. Showcases an innovative use of YOLOv8 in security applications.	YOLOv8	[195], 2023
"Detection of Prohibited Items Based upon X-ray Images and Improved YOLOv7"	Improved YOLOv7 with spatial attention for contraband detection in X-ray images. Implemented large kernel attention mechanisms to improve texture and feature extraction to boost accuracy.	Aims to automate security inspections and enhance public safety by improving prohibited item detection with modified YOLOv7 . Demonstrates YOLOv7's adaptability in security systems.	YOLOv7	[196], 2022
"Suspicious Activity Trigger System using YOLOv6 Convolutional Neural Network"	Implements YOLOv6 to detect and classify suspicious activities in CCTV footage, enhancing home surveillance systems. Utilizes deep learning to automatically trigger alerts, improving response times and security effectiveness.	Aims to reduce property theft by integrating YOLOv6 into home security systems to auto-detect suspicious behavior and alert users. Demonstrates YOLOv6's effectiveness in real-world security applications.	YOLOv6	[197], 2023
"Real-time Object Detection for Substation Security Early-warning with Deep Neural Network based on YOLO-V5"	Utilizes YOLO-v5 to enhance substation security by detecting multiple threats like fire, unauthorized entry, and vehicle misplacement in real-time. Combines deep learning with video surveillance to reduce the need for extra hardware.	Designed to improve substation security management without costly additional equipment by detecting various security threats simultaneously using YOLO-v5. Demonstrates the application of YOLO-v5 in critical infrastructure protection.	YOLOv5	[198], 2022
"Fighting against terrorism: A real-time CCTV autonomous weapons detection based on improved YOLO v4"	Imporved YOLOv4 with SCSP-ResNet backbone and F-PaNet module for detecting weapons in CCTV footage, integrating synthetic and real-world data to enhance detection.	Aims to bolster security and counter-terrorism efforts by accurately identifying weapons in CCTV using an advanced YOLOv4 architecture, demonstrating significant performance improvements.	YOLOv4	[199], 2023
"Automatic tracking of objects using improvised Yolov3 algorithm and alarm human activities in case of anomalies"	Utilizes an enhanced YOLOv3 model to automatically track objects and alert for anomalies in live video feeds, comparing performance with CNNs and decision trees.	Designed to enhance surveillance systems by detecting and alerting on anomalies like bag stealing and lock-breaking, demonstrating rapid processing and high detection accuracy.	YOLOv3	[200], 2022
"Multi-Object Detection using Enhanced YOLOv2 and LuNet Algorithms in Surveillance Videos"	Employs a novel YOLOv2-LuNet combination for efficient multi-object tracking in video surveillance, enhancing feature extraction and object detection accuracy.	Designed to improve real-time surveillance by enabling robust multi-object tracking in challenging conditions. Highlights the effectiveness of combined YOLOv2 and LuNet approach.	YOLOv2	[201], 2024
"From Silence to Propagation: Understanding the Relationship between 'Stop Snitchin' and 'YOLO'"	Examines the cultural shift from 'Stop Snitchin" to 'YOLO' in urban hip-hop culture, highlighting the role of social media in promoting individualism and exceptionalism.	Aims to explore how social media influences criminal behavior and public perception, applying cultural criminology to assess changes in social interactions and deviance.	N/A	[202], 2015

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 June 2024 doi:10.20944/preprints202406.1366.v2

Additionally [203] introduced DCS-YOLOv8, a variant optimized for detecting steel surface defects, demonstrating its effectiveness in addressing the complexities of steel manufacturing. This adaptation ensures that even minor imperfections are identified, crucial for maintaining the structural integrity of steel products.Likewise, [204] further refined YOLOv8 to develop BL-YOLOv8, focusing on road defect detection. This model enhances the safety and maintenance of transportation infrastructure by enabling more accurate and real-time detection of road surface anomalies. Similarly, [205] presented a "Hardware-Friendly" YOLOv8 model designed for foreign object identification on belt conveyors, crucial for preventing equipment damage in materials handling. This version of YOLOv8 is tailored to perform well on the limited computational resources typical of industrial hardware systems. Finally, [206] employed an improved YOLOv8 algorithm for the detection of defects in automotive adhesives, a critical quality control measure for ensuring vehicle safety and durability. These applications of YOLOv8 exemplify its adaptability and precision in industrial settings, where high accuracy and efficiency are paramount for operational success and safety compliance.

The recent advancements in YOLOv7 have paved the way for significant improvements in industrial inspection and monitoring systems. Wu et al. (2023) developed an enhanced YOLOv7 model specifically tailored for detecting objects in complex industrial equipment scenarios, highlighting its application in real-world settings [207]. Similarly, Kim et al. (2022) implemented YOLOv7 in a real-time inspection system that leverages Moire patterns to detect defects in highly reflective injection molding products, demonstrating the algorithm's capability in manufacturing quality control [208]. Further, Chen et al. (2023) explored the defect detection capabilities of YOLOv7 for automotive running lights, contributing to safer automotive systems through precise quality assurance techniques [209].

Hussain et al. (2022) applied domain feature mapping with YOLOv7 to automate inspections of pallet racking in storage facilities, enhancing safety and efficiency in logistics operations [210]. Zhu et al. (2023) extended YOLOv7's utility to the identification and classification of surface defects in belt grinding processes, aiding in maintaining the integrity of manufacturing workflows [211]. Lastly, Zhang et al. (2024) innovated with YOLO-RDP, a lightweight version of YOLOv7, optimized for detecting steel defects in real-time, showcasing the adaptability of YOLOv7 to resource-constrained environments and promoting sustainable manufacturing practices [179]. Table 6 illustrates the different use of YOLO versions in the field of industrial manufacturing:

5.5. Agriculture

In agricultural environments, advanced object detection techniques such as YOLOv5 [212,213], YOLOv6 [214,215], YOLOv7 [216,217], and YOLOv8 [218–222] have proven to be instrumental in transforming traditional farming into precision agriculture[223]. YOLOv5, for example, has been adept at weed detection [218,224], enabling farmers to apply herbicides more effectively and economically by precisely identifying and localizing weed species amidst crops. This level of precision not only conserves resources but also mitigates the adverse environmental impact of excessive chemical usage. Furthermore, YOLOv6, YOLOv7 and YOLOv8 has enhanced capabilities in broader agricultural applications such as monitoring and analyzing crop health and growth patterns, significantly improving yield predictions and crop management strategies [225–227].

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 June 2024 doi:10.20944/preprints202406.1366.v2

Table 6. Studies on YOLO applications in the manufacturing industry, focusing on real-time defect detection and process optimization for improved efficiency

Title of Paper	Description of Work	Purpose and YOLO Usage	Version	Ref. and Year
"YOLO-IMF: An Improved YOLOv8 Algorithm for Surface Defect Detection in Industrial Manufacturing Field"	Proposes an enhanced YOLOv8, YOLO-IMF, for surface defect detection on aluminum plates. Replaces CIOU with EIOU loss function to better handle small and irregularly shaped targets, achieving significant improvements in precision.	Demonstrates YOLOv8's extended applicability in industrial settings by enhancing accuracy and defect detection capabilities.	YOLOv8	[190], 2023
"YOLOv7-SiamFF: Industrial Defect Detection Algorithm Based on Improved YOLOv7"	Introduces YOLOv7-SiamFF, an advanced defect detection framework employing YOLOv7 with Siamese network enhancements for superior defect identification and background noise suppression.	Enhances industrial defect detection by integrating attention mechanisms and feature fusion modules, achieving higher accuracy in pinpointing defect locations.	YOLOv7	[185], 2024
"A Novel Finetuned YOLOv6 Transfer Learning Model for Real-Time Object Detection"	Enhances real-time object detection by integrating a transfer learning approach with a pruned and finetuned YOLOv6 model, significantly boosting detection accuracy and speed.	Focuses on improving YOLOv6 for efficient object detection in embedded systems, using advanced pruning techniques for reduced model size without sacrificing performance.	YOLOv6	[228], 2023
"Real-time Tool Detection in Smart Manufacturing Using YOLOv5"	Utilizes YOLOv5 for advanced real-time tool detection in manufacturing environments, optimizing object detection capabilities for precise tool localization.	Aims to enhance smart manufacturing by leveraging YOLOv5 for accurate and real-time detection of various tools, contributing significantly to Industry 4.0 initiatives.	YOLOv5	[229], 2023
"Efficient Automobile Assembly State Monitoring System Based on Channel-Pruned YOLOv4"	Implements a channel-pruned YOLOv4 algorithm to optimize monitoring in automobile assembly, enhancing detection speed without compromising accuracy.	Designed to streamline assembly monitoring in industrial environments, showcasing YOLOv4's utility in enhancing operational efficiency and deployment readiness.	YOLOv4	[230], 2024
"YOLO V3 + VGG16-based Automatic Operations Monitoring in Manufacturing Workshop"	Utilizes a combined YOLO V3 and VGG16 framework to recognize and monitor industrial operations accurately for Industry 4.0 manufacturing workshops.	Aims to enhance production efficiency and quality by automating action analysis and process monitoring using advanced YOLO V3 and VGG16 technologies.	YOLO V3, VGG16	[231], 2022
"Improvements of Detection Accuracy by YOLOv2 with Data Set Augmentation"	Employs YOLOv2 with an innovative data set augmentation method to enhance the detection accuracy and confidence in identifying defective areas in industrial products.	Seeks to optimize defect detection and visualization on production lines, demonstrating YOLOv2's effectiveness with limited data augmentation options.	YOLOv2	[232], 2023

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 June 2024 doi:10.20944/preprints202406.1366.v2

The recent introduction of YOLOv7 and YOLOv8 has further pushed the boundaries of agricultural innovation. YOLOv7 [233–235] and YOLOv8 [100,222,236] has been specifically refined to detect small pests and subtle disease symptoms on crops, which are often overlooked by human inspectors. Its enhanced deep learning framework allows for integrating complex image recognition tasks that facilitate early detection, thereby preventing widespread crop damage. On the other hand, YOLOv8 has made significant strides in fruit detection tasks. Its application in orchards for detecting fruits such as apples [237,237] supports optimal harvesting by determining the right stage of fruit maturity. This maximizes the harvest quality and ensures that the fruits are picked at their nutritional peak, thereby enhancing their market value. The application of these advanced YOLO models YOLOv5, YOLOv6, YOLOv7, and YOLOv8 represents a leap towards a more sustainable and efficient agricultural sector.

Recent studies have demonstrated the efficacy of YOLO-based models in enhancing various aspects of agricultural automation and efficiency. Junos et al. (2021) optimized a YOLO-based object detection model to improve crop harvesting systems, showcasing the potential to boost yield and reduce labor costs [238]. Zhao et al. (2024) extended this application to real-time object detection combined with robotic manipulation, further aligning agricultural practices with advanced automation technologies [239]. Chen et al. (2021) developed an apple detection method using a tailored YOLOv4 algorithm, specifically designed to support harvesting robots operating in complex environments, which significantly enhances the precision and efficiency of fruit picking [240].

Further contributions include work by Nergiz (2023), who utilized YOLOv7 to enhance strawberry harvesting efficiency, providing practical solutions for small to medium-sized enterprises in the agricultural [241]. Wang et al. (2024) focused on planning harvesting operations in large strawberry fields using a deep learning-based image processing method, demonstrating the scalability of YOLO for larger agricultural operations [242]. Lastly, Zhang et al. (2023) introduced DCF-YOLOv8, an improved algorithm for agricultural pests and diseases detection by aggregating low-level features, which helps in early detection and management of crop health [222]. These studies collectively illustrate the transformative impact of YOLO-based models in modernizing agricultural practices, ensuring higher productivity and sustainability.

In orchard automation, the YOLO object detection models have been specifically pivotal in enhancing the accuracy and efficiency of fruit detection [243–245], flower identification [246–248], and automated harvesting processes [238,249,250]. These models adeptly identify and classify fruits at various stages of ripeness, detect flowers with high precision, and facilitate efficient harvesting operations. The development of YOLO models, has introduced significant improvements that cater specifically to the challenges of agricultural environments. For instance, YOLOv5's introduction of multi-scale predictions improved the detection of small and clustered objects like flowers and young fruits, which are critical during the early stages of crop yield management [251]. As the models advanced, YOLOv7 and YOLOv8 incorporated better segmentation techniques, which enhanced the differentiation between fruit types and maturity stages, critical for targeted harvesting [252,253].

Moreover, recent iteration, YOLOv9 have leveraged advanced algorithms with spatial pyramid pooling and attention mechanisms, which have refined the detection capabilities in plant disease detection [254]. [254] performed a comparative study on different important versions of YOLO (v5, v8 and v9) on a real-world dataset for tomato plant disease detection and suggested that YOLOv9 outperforms YOLOv5 and YOLOv8.

Table 7 illustrates the different use of YOLO versions in the field of Agriculture:

Table 7. Studies on YOLO usage in agriculture, emphasizing automated crop monitoring, pest detection, and yield estimation for enhanced productivity.

Title of Paper	Description of Work	Purpose and YOLO Usage	Version	Ref. and Year
"Automating Tomato Ripeness Classification and Counting with YOLOv9"	Implements YOLOv9 to automate and enhance the accuracy of classifying and counting ripe tomatoes, replacing labor-intensive visual inspections.	Aims to streamline tomato ripeness monitoring and counting, to enhance agricultural productivity and quality. Utilizes YOLOv9 for high accuracy in detection.	YOLOv9	[255], 2024
"A Lightweight YOLOv8 Tomato Detection Algorithm Combining Feature Enhancement and Attention"	Enhances YOLOv8 for tomato detection in agriculture using depthwise separable convolution and dual-path attention gate modules. Optimizes real-time detection for robotic tomato picking.	Aims to advance agricultural automation by boosting YOLOv8's efficiency and accuracy in tomato harvesting. Demonstrates improved performance over earlier YOLO versions.	YOLOv8	[90], 2023
"An Attention Mechanism-Improved YOLOv7 Object Detection Algorithm for Hemp Duck Count Estimation"	Implements CBAM-YOLOv7 to enhance feature extraction capabilities within YOLOv7 for precise hemp duck counting in agriculture, outperforming SE-YOLOv7 and ECA-YOLOv7 in precision and mAP.	Enhances livestock management by automating duck count with advanced object detection, reducing labor and improving accuracy.	YOLOv7	[216], 2022
"Detecting Crops and Weeds in Fields Using YOLOv6 and Faster R-CNN Object Detection Models"	Utilizes YOLOv6 and Faster R-CNN to detect crops and weeds for precise management.	Aims to boost agricultural productivity and environmental sustainability by improving accuracy in weed detection using YOLOv6.	YOLOv6	[214], 2023
"An improved YOLOv5-based vegetable disease detection method"	Enhances YOLOv5 for precise detection of vegetable diseases by upgrading CSP, FPN, and NMS modules to handle complex environmental interference.	Aims to improve food security by boosting the accuracy and speed of disease detection in vegetables using an improved YOLOv5 algorithm.	YOLOv5	[256], 2022
"Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments"	Implements a channel pruned YOLOv4 model to enhance efficiency and accuracy in detecting apple flowers, supporting the development of flower thinning robots.	Aims to optimize apple flower detection in orchards by applying channel pruning to YOLOv4, significantly reducing model size and improving processing speed while maintaining high accuracy.	YOLOv4	[246], 2020
"Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model"	Enhances YOLOv3-tiny with additional convolutional kernels for improved kiwifruits detection in orchards, in occlusions and varying lighting conditions.	Focus on increasing the efficiency of kiwifruit detection in dynamic orchard environmentswith a modified YOLOv3-tiny, demonstrating high performance.	YOLOv3-tiny	[257], 2021
"A Detection Method for Tomato Fruit Common Physiological Diseases Based on YOLOv2"	Implements YOLOv2 to detect and identify healthy and diseased tomato, using advanced image processing and data augmentation to enhance detection accuracy.	Aims to boost tomato yield and quality control through efficient detection of physiological diseases, demonstrating the effectiveness of YOLOv2 in agriculture.	YOLOv2	[258], 2019
"A Vision-Based Counting and Recognition System for Flying Insects in Intelligent Agriculture"	Utilizes YOLO for initial detection and counting, and SVM for fine classification of flying insects, for efficient in pest control.	Demonstrates a robust, efficient system for insect monitoring, greatly enhancing accuracy and speed in pest management.	YOLO, SVM	[259], 2018

 doi:10.20944/preprints202406.1366.v2

6. Challenges, Limitations and Future Directions

YOLOv10:

- As the latest version in the YOLO series, YOLOv10 has not yet seen widespread adoption in published research. Its release promises cutting-edge improvements in object detection capabilities, but the lack of extensive testing and real-world application data makes it difficult to ascertain its full potential and limitations.
- Preliminary evaluations suggest that while YOLOv10 might offer advancements in speed and accuracy, integrating it into existing systems could present challenges due to compatibility and computational demands. Potential users may hesitate to adopt this version until more comprehensive studies and benchmarks are available, which articulate its advantages over previous models.
- The expectation with YOLOv10, much like its predecessors, is that it will drive further research in object detection technologies. Its eventual widespread implementation could pave the way for addressing complex detection scenarios with higher accuracy, particularly in dynamic environments. However, as with any new technology, the adaptation phase will be crucial in understanding its practical limitations and operational challenges.

YOLOv9:

- Despite YOLOv9's enhancements in detection capabilities, it has only been featured in a handful of studies, which limits a comprehensive understanding of its performance across diverse applications. This lack of extensive validation may deter organizations from adopting it until more empirical evidence and comparative analyses establish its efficacy and efficiency over earlier versions.
- While YOLOv9 improves upon the speed and accuracy of its predecessors, it may still struggle with detecting small or overlapping objects in cluttered scenes. This is a recurring challenge in high-density environments like crowded urban areas or complex natural scenes, where precise detection is critical for applications such as autonomous driving and wildlife monitoring.
- Future developments for YOLOv9 could focus on enhancing its robustness in adverse conditions, such as varying weather, lighting, or occlusions. Integrating more adaptive and context-aware mechanisms could help in mitigating false positives and improving the reliability of the system under different operational conditions. The implementation of advanced training techniques such as federated learning could also be explored to enhance its adaptability and learning efficiency from decentralized data sources.

YOLOv8:

- YOLOv8 has shown significant improvements in object detection tasks, particularly in real-time applications. However, it continues to face challenges in terms of computational efficiency and resource consumption when deployed on lower-end hardware [260]. This can limit its applicability in resource-constrained environments where deploying advanced hardware solutions is not feasible [132].
- The future direction for YOLOv8 could involve optimizing its architectural design to reduce computational load without compromising detection accuracy. Enhancing its scalability to efficiently process images of varying resolutions and conditions can broaden its application scope. Moreover, incorporating adaptive scaling and context-aware training methods could potentially address the detection challenges in complex scenes, making it more robust against diverse operational challenges.

YOLOv7:

- Although YOLOv7 introduces significant improvements in detection accuracy and speed, its adoption across varied real-world applications reveals a persistent challenge in handling highly

 doi:10.20944/preprints202406.1366.v2

dynamic scenes. For instance, in environments with rapid motion or in scenarios involving occlusions, YOLOv7 can still experience drops in performance. The algorithm's ability to generalize across different types of blur and motion artifacts remains an area for further research and enhancement.

- The complexity of YOLOv7's architecture, while beneficial for accuracy, imposes a substantial computational burden. This makes it less ideal for deployment on edge devices or platforms with limited processing capabilities, where maintaining a balance between speed and power efficiency is crucial [161,261]. Efforts to streamline the model for such applications without significant loss of performance are necessary.
- Looking forward, there is significant potential in expanding YOLOv7's capabilities through the integration of semi-supervised or unsupervised learning paradigms. This would enable the model to leverage unlabeled data effectively, a common challenge in the real-world where annotated datasets are often scarce or expensive to produce. Additionally, enhancing the model's resilience to adversarial attacks and variability in data quality could further solidify its utility in security-sensitive applications like surveillance and fraud detection.

YOLOv6:

- One of the notable challenges with YOLOv6 is its handling of scale variability within images, which can affect its efficacy in environments where objects appear at diverse distances from the camera. While YOLOv6 shows improved accuracy and speed over its predecessors, it sometimes struggles with small or partially occluded objects, which are common in crowded scenes or complex industrial environments [151,262]. This limitation can be critical in applications such as automated surveillance or advanced manufacturing monitoring.
- YOLOv6, while efficient, still requires considerable computational resources when compared to other models optimized for edge devices. Its deployment in resource-constrained environments such as mobile or embedded systems often requires a trade-off between detection performance and operational efficiency. Further optimizations and model pruning are necessary to achieve the best of both worlds—real-time performance with reduced computational demands.
- Future enhancements for YOLOv6 could focus on incorporating more advanced feature extraction techniques that improve its robustness to variations in object appearance and environmental conditions. Additionally, integrating more adaptive and context-aware learning mechanisms could help overcome some of the challenges related to background clutter and similar adversities. Enhancing the model's capacity to learn from a limited number of training samples, through techniques such as few-shot learning or transfer learning, could address the scarcity of labeled training data in specialized applications.

YOLOv5:

- YOLOv5 has made significant strides in improving detection speed and accuracy, but it faces challenges in consistently detecting small objects due to its spatial resolution constraints. This is particularly evident in fields like medical imaging or satellite image analysis, where precision is crucial for identifying fine details. Techniques such as spatial pyramid pooling or enhanced up-sampling may be needed to increase the receptive field and improve the detection of smaller objects without compromising the model's efficiency [120,263,264].
- While YOLOv5 offers faster training and inference times compared to previous versions, its deployment on edge devices is limited by high memory and processing requirements [127,265]. Although optimized models like YOLOv5s provide a solution, they sometimes do so at the cost of detection accuracy. Optimizing network architecture through neural architecture search (NAS) could potentially offer a more balanced solution, enhancing both performance and efficiency for real-time object detection applications.
- The adaptability of YOLOv5 to varied environmental conditions and different types of data distribution remains an area for development. Future research could focus on enhancing the

 doi:10.20944/preprints202406.1366.v2

robustness of YOLOv5 through advanced data augmentation techniques and domain adaptation strategies. This would enable the model to maintain high accuracy levels across diverse application settings, from urban surveillance to complex natural environments, effectively handling variations in lighting, weather, and seasonal changes.

YOLOv4, YOLOv3, YOLOv2 and YOLOv1:

- The advancements in YOLOv4 brought significant improvements in speed and accuracy, but the model's performance remains inconsistent across various datasets, especially in class imbalance and rare object recognition. Its computational demand limits its practical deployment on low-power devices. Efforts to enhance model compression and environmental adaptability could further broaden its utility in real-world applications.
- YOLOv3 improved upon the balance of speed and accuracy, yet it struggles with small object detection due to its grid limitation. Its computational efficiency poses challenges for deployment in resource-constrained environments, prompting research towards optimization techniques to improve efficiency without sacrificing performance. Additionally, enhancing the model's robustness to environmental variations could improve its reliability for applications like autonomous driving and urban surveillance.
- Despite the incremental improvements introduced in YOLOv2, it faces challenges in detecting small objects, balancing speed with accuracy, and maintaining relevance with the advent of more capable successors. This version's reliance on a fixed grid system hampers its ability to perform in high-precision detection tasks. Future developments may shift towards adapting YOLOv2's core strengths in new architectures that enhance its spatial resolution and dynamic scaling capabilities.

For the versions of YOLO under YOLOv5, their use may decrease and discontinue in the future as newer versions are replacing the older YOLO versions in overall performance and efficiency.

- The potential for YOLOv4, YOLOv3, and YOLOv2 in future research involves exploring adaptive mechanisms that can tailor learning rates and augment data to better handle diverse operational scenarios. Integrating these models with newer technologies like model pruning and feature fusion may address existing inefficiencies and extend their applicability to a wider range of applications.
- YOLOv1 was revolutionary for its time, introducing real-time object detection by processing the entire image at once as a single regression problem. However, it faces significant challenges in dealing with small objects due to each grid cell predicting only two boxes and the probabilities for the classes. This structure often leads to poor performance on groups of small objects that are close together, such as flocks of birds or traffic scenes with multiple vehicles at a distance. Improvements in subsequent models focus on increasing the number of predictions per grid and incorporating finer-grained feature maps to enhance small object detection.
- Another limitation of YOLOv1 is the spatial constraints of its bounding boxes. Since each cell in the grid can only predict two boxes and has limited context about its neighboring cells, the precision in localizing objects, especially those with complex or irregular shapes, is often compromised. This challenge is particularly evident in medical imaging and satellite image analysis, where the exact contours of the objects are crucial. Advances in convolutional neural network designs and cross-layer feature integration in later versions seek to address these drawbacks.
- Despite the foundational advancements introduced by YOLOv1, its direct application has waned over the years, superseded by more robust iterations like YOLOv2 and YOLOv3. These later versions build upon the core principles of YOLOv1 but offer refined mechanisms for handling varied object sizes and aspect ratios. Future research directions are less likely to focus on YOLOv1 itself but may explore its integration into hybrid models or specialized adaptations that can leverage its speed for real-time applications where latency is critical, albeit with compensations in detection accuracy and granularity.

 doi:10.20944/preprints202406.1366.v2

- Future iterations could focus on dynamic grid systems, lighter network architectures, and advanced scaling features to tackle the challenges of small object detection and computational limitations. These improvements could enhance their deployment in emerging areas such as edge computing, where real-time processing and low power consumption are crucial.
- As newer models like YOLOv8 and YOLOv9 continue to evolve, the foundational aspects of YOLOv4, YOLOv3, and YOLOv2 can still offer valuable insights for developing hybrid models or specialized applications. Research may increasingly focus on leveraging these older versions for their speed attributes while compensating for their detection limitations through composite and hybrid modeling approaches.

Over the past decade, the series of You Only Look Once (YOLO) models have significantly impacted various sectors, demonstrating the powerful capabilities of deep learning in real-world applications. As a pioneering object detection algorithm, YOLO has facilitated rapid advancements across diverse fields by offering high-speed, real-time detection with commendable accuracy. One of the most notable applications has been in public safety and surveillance, where YOLO models have improved the efficacy of monitoring systems, enhancing the detection of suspicious activities and ensuring public safety more efficiently. In the realm of automotive technology, YOLO has been integral in developing advanced driver-assistance systems (ADAS), contributing to object detection that supports collision avoidance systems and pedestrian safety. Furthermore, YOLO has transformed the healthcare sector by accelerating medical image analysis, enabling quicker and more accurate detection of pathologies which is critical for diagnostics and treatment planning. In industrial settings, YOLO has optimized quality control processes by identifying defects in manufacturing lines in real-time, thereby reducing waste and increasing production efficiency. Additionally, in the retail sector, YOLO has supported inventory management through automated checkouts and stock monitoring, enhancing customer experience and operational efficiency.

6.1. YOLO and the Artificial General Intelligence - AGI

While YOLO is a specialized AI focused on object detection, its success underscores a critical component of Artificial General Intelligence (AGI) [266] the ability to process and interpret visual data. In the quest for AGI, integrating such advanced perception systems is crucial. An AGI system would need to combine YOLO-like real-time object detection with other cognitive abilities, such as natural language understanding and reasoning, to perform a wide range of tasks. For example, an AGI-powered robot could use YOLO for visual recognition to navigate and interact with its environment while simultaneously understanding and responding to verbal instructions, demonstrating a level of versatility and general intelligence akin to human capabilities.

6.1.1. YOLO as the “Neural Network That Can Do”

This generation of neural networks has amazed us with its advanced vision and language capabilities, pushing the boundaries of what AI can perceive and interpret. The next wave of neural networks, however, will be defined by their ability to not only understand but also to act and execute tasks in real-time. YOLO is poised to be a key player in this transformation. Its unparalleled speed and accuracy in object detection make it the ideal candidate for applications requiring immediate action, such as autonomous driving, robotics, and real-time surveillance. As we move towards a future where AI not only sees and speaks but also performs complex tasks autonomously, YOLO’s role will be instrumental in bridging the gap between perception and action. One such project is the "BEHAVIOUR", which is a human-centered simulation benchmark to evaluate embodied AI solutions [267] at the Stanford University [268]

6.2. YOLO on the Edge Devices

The deployment of YOLO on edge devices unlocks several promising avenues for future research and development. One potential direction involves enhancing the algorithm’s efficiency and accuracy

 doi:10.20944/preprints202406.1366.v2

for even more constrained environments, such as ultra-low-power Microcontrollers and embedded systems. This can be achieved through further optimization techniques, including model pruning, quantization, and the development of specialized hardware accelerators. Additionally, integrating YOLO with advanced communication protocols and edge computing frameworks could facilitate more seamless collaboration between edge devices and centralized cloud services, enhancing the overall system performance and scalability. Exploring the integration of YOLO with other AI-driven functionalities, such as anomaly detection and predictive analytics, may unlock new applications in areas like healthcare, smart cities, and industrial automation. As edge computing continues to evolve, the adaptation of YOLO to support federated learning paradigms could ensure the data privacy while enabling continuous learning and improvement of object detection models. These future directions will not only expand the capabilities of YOLO but also contribute significantly to the advancement of intelligent edge computing systems.

6.3. Future Prospects

In considering future challenges, it is envisioned that YOLO variants will continue to improve performance on small object targets, especially as they penetrate into more specialized areas such as precision manufacturing. Attention mechanisms could be integrated to enhance small object detection, while the use of vision transformers could further improve YOLO's performance on capturing global contextual dependencies. This trajectory suggests a necessity for enhancements in lightweight architectures that balance high accuracy with stringent FPS requirements. As YOLO progresses, meeting the demands of niche applications will likely drive further innovation in architectural design and optimization, ensuring its continued relevance in domains with stringent requirements for precision and efficiency. For instance, the integration of voice commands into surveillance systems, facilitated by models like ChatGPT, could transform security mechanisms, making them more interactive and responsive. In the healthcare sector, the incorporation of medical imaging with historical patient data and live symptom descriptions could significantly improve the personalization and accuracy of medical responses.

Looking further, YOLO's potential to adapt to such multimodal advancements will be instrumental in pioneering the next wave of intelligent applications. From autonomous vehicles that interpret both road signs and pedestrian gestures to smart homes that react to visual cues and voice instructions, the integration of YOLO with a broader spectrum of data types and deeper contextual understanding heralds a groundbreaking epoch in artificial intelligence. This transformative phase promises to significantly improve the interactivity and cognitive capabilities of Machine Vision systems, marking a pivotal shift in visual process automation.

6.4. Challenges in Statistical Metrics for Evaluation

Threat: Relying on a single statistical summary metric to measure YOLO detection capability may not fully reflect the performance of systems across various YOLO applications, necessitating the use of several metrics.

Mitigation: Despite this limitation, our main premise is that the selected metrics enable us to compare various YOLO systems and adequately assess their overall effectiveness. Recognizing the inherent limitations of statistical summaries is crucial when conducting a comprehensive evaluation of detection systems across different applications. Therefore, we aim to improve the clarity and reliability of our review by openly acknowledging these potential threats to construct validity. This approach provides a more nuanced understanding of the limitations associated with various aspects of YOLO techniques for object detection in diverse domains.

7. Conclusion

In this comprehensive review, we explored the evolution of the YOLO models from the most recent YOLOv10 to the inaugural YOLOv1. This retrospective analysis covered a decade of advancements,

 doi:10.20944/preprints202406.1366.v2

highlighting the pivotal improvements in each version and their respective impacts across five critical application areas: public safety, automotive technology, healthcare, industrial manufacturing, and retail. Our review outlined the significant enhancements in detection speed, accuracy, and computational efficiency that each iteration brought, while also addressing the specific challenges and limitations faced by earlier versions. Furthermore, we have identified gaps in the current capabilities of YOLO models and proposed potential directions for future research. Predicting the trajectory of YOLO's development, we anticipate a shift towards multimodal data processing, leveraging advancements in large language models and natural language processing to enhance object detection systems. This fusion is expected to broaden the utility of YOLO models, enabling more sophisticated, context-aware applications that could revolutionize interaction between AI systems and their environments. Thus, this review not only serves as a detailed chronicle of YOLO's evolution but also sets a prospective blueprint for its integration into the next generation of technological innovations.

Author Contributions: Ranjan Sapkota: principal conceptualizer, idea generation, research design, formal analysis, original draft preparation, manuscript writing, and editing. **Rizwan Qureshi, Marco Flores-Calero, Chetan Badgujar, Upesh Nepal, Alwin Poulose, Peter Zeno, Uday Bhanu Prakash Vaddevolu, Hong Yan, Manoj Karkee:** methodology refinement, critical revisions, manuscript review, and editing.

References

1. Liu, L.; Ouyang, W.; Wang, X.; Fieguth, P.; Chen, J.; Liu, X.; Pietikäinen, M. Deep learning for generic object detection: A survey. *International journal of computer vision* **2020**, *128*, 261–318.
2. Fernandez, R.A.S.; Sanchez-Lopez, J.L.; Sampedro, C.; Bavle, H.; Molina, M.; Campoy, P. Natural user interfaces for human-drone multi-modal interaction. 2016 International Conference on Unmanned Aircraft Systems (ICUAS). IEEE, 2016, pp. 1013–1022.
3. Wang, R.J.; Li, X.; Ling, C.X. Pelee: A real-time object detection system on mobile devices. *Advances in neural information processing systems* **2018**, *31*.
4. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* **2015**, *28*.
5. Flippo, D.; Gunturu, S.; Baldwin, C.; Badgujar, C. Tree Trunk Detection of Eastern Red Cedar in Rangeland Environment with Deep Learning Technique. *Croatian journal of forest engineering* **2023**, *44*, 357–368. doi:10.5552/crojfe.2023.2012.
6. Guerrero-Ibáñez, J.; Zeadally, S.; Contreras-Castillo, J. Sensor technologies for intelligent transportation systems. *Sensors* **2018**, *18*, 1212.
7. Hussain, R.; Zeadally, S. Autonomous cars: Research results, issues, and future challenges. *IEEE Communications Surveys & Tutorials* **2018**, *21*, 1275–1313.
8. Kaushal, M.; Khehra, B.S.; Sharma, A. Soft Computing based object detection and tracking approaches: State-of-the-Art survey. *Applied Soft Computing* **2018**, *70*, 423–464.
9. Khan, S.M.; Shah, M. Tracking multiple occluding people by localizing on multiple scene planes. *IEEE transactions on pattern analysis and machine intelligence* **2008**, *31*, 505–519.
10. Gupta, A.; Anpalagan, A.; Guan, L.; Khwaja, A.S. Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues. *Array* **2021**, *10*, 100057.
11. Zou, Z.; Chen, K.; Shi, Z.; Guo, Y.; Ye, J. Object detection in 20 years: A survey. *Proceedings of the IEEE* **2023**, *111*, 257–276.
12. Liu, S.; Liu, D.; Srivastava, G.; Połap, D.; Woźniak, M. Overview and methods of correlation filter algorithms in object tracking. *Complex & Intelligent Systems* **2021**, *7*, 1895–1917.
13. Hu, X.d.; Wang, X.q.; Meng, F.j.; Hua, X.; Yan, Y.j.; Li, Y.y.; Huang, J.; Jiang, X.l. Gabor-CNN for object detection based on small samples. *Defence Technology* **2020**, *16*, 1116–1129.
14. Surasak, T.; Takahiro, I.; Cheng, C.h.; Wang, C.e.; Sheng, P.y. Histogram of oriented gradients for human detection in video. 2018 5th International conference on business and industrial research (ICBIR). IEEE, 2018, pp. 172–176.
15. Karis, M.S.; Razif, N.R.A.; Ali, N.M.; Rosli, M.A.; Aras, M.S.M.; Ghazaly, M.M. Local Binary Pattern (LBP) with application to variant object detection: A survey and method. 2016 IEEE 12th international colloquium on signal processing & its applications (CSPA). IEEE, 2016, pp. 221–226.

16. Chiu, H.J.; Li, T.H.S.; Kuo, P.H. Breast cancer–detection system using PCA, multilayer perceptron, transfer learning, and support vector machine. *IEEE Access* **2020**, *8*, 204309–204324.
17. Xiang, Y.; Mottaghi, R.; Savarese, S. Beyond pascal: A benchmark for 3d object detection in the wild. IEEE winter conference on applications of computer vision. IEEE, 2014, pp. 75–82.
18. Xie, X.; Cheng, G.; Wang, J.; Yao, X.; Han, J. Oriented R-CNN for object detection. Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 3520–3529.
19. Li, X.; Song, D.; Dong, Y. Hierarchical feature fusion network for salient object detection. *IEEE Transactions on Image Processing* **2020**, *29*, 9165–9175.
20. Crawford, E.; Pineau, J. Spatially invariant unsupervised object detection with convolutional neural networks. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, Vol. 33, pp. 3412–3420.
21. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 10781–10790.
22. Zhiqiang, W.; Jun, L. A review of object detection based on convolutional neural network. 2017 36th Chinese control conference (CCC). IEEE, 2017, pp. 11104–11109.
23. Teutsch, M.; Kruger, W. Robust and fast detection of moving vehicles in aerial videos using sliding windows. Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2015, pp. 26–34.
24. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J.; Mercan, E. R-CNN for object detection. IEEE Conference, 2014.
25. Girshick, R. Fast r-cnn. Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440–1448.
26. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 779–788.
27. Badgujar, C.M.; Poulose, A.; Gan, H. Agricultural Object Detection with You Look Only Once (YOLO) Algorithm: A Bibliometric and Systematic Literature Review. *arXiv preprint arXiv:2401.10379* **2024**.
28. Patel, G.S.; Desai, A.A.; Kamble, Y.Y.; Pujari, G.V.; Chougule, P.A.; Jujare, V.A. Identification and separation of medicine through Robot using YOLO and CNN Algorithms for Healthcare. 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI). IEEE, 2023, Vol. 1, pp. 1–5.
29. Ajayi, O.G.; Ashi, J.; Guda, B. Performance evaluation of YOLO v5 model for automatic crop and weed classification on UAV images. *Smart Agricultural Technology* **2023**, *5*, 100231.
30. Morbekar, A.; Parihar, A.; Jadhav, R. Crop disease detection using YOLO. 2020 international conference for emerging technology (INCET). IEEE, 2020, pp. 1–5.
31. Li, D.; Ahmed, F.; Wu, N.; Sethi, A.I. Yolo-JD: A Deep Learning Network for jute diseases and pests detection from images. *Plants* **2022**, *11*, 937.
32. Cheeti, S.; Kumar, G.S.; Priyanka, J.S.; Firdous, G.; Ranjeeva, P.R. Pest detection and classification using YOLO AND CNN. *Annals of the Romanian Society for Cell Biology* **2021**, pp. 15295–15300.
33. Pham, M.T.; Courtrai, L.; Friguet, C.; Lefèvre, S.; Baussard, A. YOLO-Fine: One-stage detector of small objects under various backgrounds in remote sensing images. *Remote Sensing* **2020**, *12*, 2501.
34. Cheng, L.; Li, J.; Duan, P.; Wang, M. A small attentional YOLO model for landslide detection from satellite remote sensing images. *Landslides* **2021**, *18*, 2751–2765.
35. Chen, C.; Zheng, Z.; Xu, T.; Guo, S.; Feng, S.; Yao, W.; Lan, Y. Yolo-based uav technology: A review of the research and its applications. *Drones* **2023**, *7*, 190.
36. Luo, X.; Wu, Y.; Zhao, L. YOLOD: A target detection method for UAV aerial imagery. *Remote Sensing* **2022**, *14*, 3240.
37. Prinzi, F.; Insalaco, M.; Orlando, A.; Gaglio, S.; Vitabile, S. A YOLO-based model for breast cancer detection in mammograms. *Cognitive Computation* **2024**, *16*, 107–120.
38. Aly, G.H.; Marey, M.; El-Sayed, S.A.; Tolba, M.F. YOLO based breast masses detection and classification in full-field digital mammograms. *Computer methods and programs in biomedicine* **2021**, *200*, 105823.
39. Ünver, H.M.; Ayan, E. Skin lesion segmentation in dermoscopic images with combination of YOLO and grabcut algorithm. *Diagnostics* **2019**, *9*, 72.
40. Tan, L.; Huangfu, T.; Wu, L.; Chen, W. Comparison of RetinaNet, SSD, and YOLO v3 for real-time pill identification. *BMC medical informatics and decision making* **2021**, *21*, 1–11.

41. Suksawatchon, U.; Srikamdee, S.; Suksawatchon, J.; Werapan, W. Shape Recognition Using Unconstrained Pill Images Based on Deep Convolution Network. 2022 6th International Conference on Information Technology (InCIT). IEEE, 2022, pp. 309–313.
42. Arroyo, M.A.; Ziad, M.T.I.; Kobayashi, H.; Yang, J.; Sethumadhavan, S. YOLO: frequently resetting cyber-physical systems for security. Autonomous Systems: Sensors, Processing, and Security for Vehicles and Infrastructure 2019. SPIE, 2019, Vol. 11009, pp. 166–183.
43. Bordoloi, N.; Talukdar, A.K.; Sarma, K.K. Suspicious activity detection from videos using yolov3. 2020 IEEE 17th India Council International Conference (INDICON). IEEE, 2020, pp. 1–5.
44. Gorave, A.; Misra, S.; Padir, O.; Patil, A.; Ladole, K. Suspicious activity detection using live video analysis. Proceeding of International Conference on Computational Science and Applications: ICCSA 2019. Springer, 2020, pp. 203–214.
45. Kolpe, R.; Ghogare, S.; Jawale, M.; William, P.; Pawar, A. Identification of face mask and social distancing using YOLO algorithm based on machine learning approach. 2022 6th International conference on intelligent computing and control systems (ICICCS). IEEE, 2022, pp. 1399–1403.
46. Bashir, S.; Qureshi, R.; Shah, A.; Fan, X.; Alam, T. YOLOv5-M: A deep neural network for medical object detection in real-time. 2023 IEEE Symposium on Industrial Electronics & Applications (ISIEA). IEEE, 2023, pp. 1–6.
47. Pham, D.L.; Chang, T.W.; others. A YOLO-based real-time packaging defect detection system. *Procedia Computer Science* **2023**, *217*, 886–894.
48. Klarák, J.; Andok, R.; Malík, P.; Kuric, I.; Ritomskỳ, M.; Klačková, I.; Tsai, H.Y. From Anomaly Detection to Defect Classification. *Sensors* **2024**, *24*, 429.
49. Li, R.; Yang, J. Improved YOLOv2 object detection model. 2018 6th international conference on multimedia computing and systems (ICMCS). IEEE, 2018, pp. 1–6.
50. Nakahara, H.; Yonekawa, H.; Fujii, T.; Sato, S. A lightweight YOLOv2: A binarized CNN with a parallel support vector regression for an FPGA. Proceedings of the 2018 ACM/SIGDA International Symposium on field-programmable gate arrays, 2018, pp. 31–40.
51. Kim, K.J.; Kim, P.K.; Chung, Y.S.; Choi, D.H. Performance enhancement of YOLOv3 by adding prediction layers with spatial pyramid pooling for vehicle detection. 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS). IEEE, 2018, pp. 1–6.
52. Nepal, U.; Eslamiat, H. Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs. *Sensors* **2022**, *22*, 464.
53. Sozzi, M.; Cantalamessa, S.; Cogato, A.; Kayad, A.; Marinello, F. Automatic bunch detection in white grape varieties using YOLOv3, YOLOv4, and YOLOv5 deep learning algorithms. *Agronomy* **2022**, *12*, 319.
54. Mohod, N.; Agrawal, P.; Madaan, V. YOLOv4 vs YOLOv5: Object detection on surveillance videos. International Conference on Advanced Network Technologies and Intelligent Computing. Springer, 2022, pp. 654–665.
55. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. YOLOv10: Real-Time End-to-End Object Detection. *arXiv preprint arXiv:2405.14458* **2024**.
56. Wang, C.Y.; Yeh, I.H.; Liao, H.Y.M. YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *arXiv preprint arXiv:2402.13616* **2024**.
57. Hussain, M. YOLOv1 to v8: Unveiling Each Variant–A Comprehensive Review of YOLO. *IEEE Access* **2024**, *12*, 42816–42833.
58. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 7464–7475.
59. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; others. YOLOv6: A single-stage object detection framework for industrial applications. *arXiv preprint arXiv:2209.02976* **2022**.
60. Redmon, J.; Farhadi, A. YOLO9000: better, faster, stronger. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7263–7271.
61. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* **2018**.
62. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934* **2020**.
63. Ultralytics. Home — docs.ultralytics.com. https://docs.ultralytics.com/. [Accessed 28-05-2024].

64. Ultralytics. Comprehensive Guide to Ultralytics YOLOv5 — docs.ultralytics.com. https://docs.ultralytics.com/yolov5/. [Accessed 28-05-2024].
65. GitHub - ultralytics/yolov5: YOLOv5 in PyTorch > ONNX > CoreML > TFLite — github.com. https://github.com/ultralytics/yolov5. [Accessed 28-05-2024].
66. Tang, S.; Yuan, Y. Object detection based on convolutional neural network. International Conference-IEEE–2016, 2015.
67. Mao, H.; Yang, X.; Dally, W.J. A delay metric for video object detection: What average precision fails to tell. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 573–582.
68. Chen, B.; Ghiasi, G.; Liu, H.; Lin, T.Y.; Kalenichenko, D.; Adam, H.; Le, Q.V. Mnasfpn: Learning latency-aware pyramid architecture for object detection on mobile devices. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 13607–13616.
69. Pestana, D.; Miranda, P.R.; Lopes, J.D.; Duarte, R.P.; Véstias, M.P.; Neto, H.C.; De Sousa, J.T. A full featured configurable accelerator for object detection with YOLO. *IEEE Access* **2021**, *9*, 75864–75877.
70. Zhou, P.; Ni, B.; Geng, C.; Hu, J.; Xu, Y. Scale-transferrable object detection. proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 528–537.
71. Hall, D.; Dayoub, F.; Skinner, J.; Zhang, H.; Miller, D.; Corke, P.; Carneiro, G.; Angelova, A.; Sünderhauf, N. Probabilistic object detection: Definition and evaluation. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020, pp. 1031–1040.
72. Goutte, C.; Gaussier, E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. European conference on information retrieval. Springer, 2005, pp. 345–359.
73. Liang, Z.; Zhang, Z.; Zhang, M.; Zhao, X.; Pu, S. Rangeioudet: Range image based real-time 3d object detector optimized by intersection over union. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 7140–7149.
74. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer, 2016, pp. 21–37.
75. Fu, C.Y.; Liu, W.; Ranga, A.; Tyagi, A.; Berg, A.C. Dssd: Deconvolutional single shot detector. *arXiv preprint arXiv:1701.06659* **2017**.
76. Zhang, S.; Wen, L.; Bian, X.; Lei, Z.; Li, S.Z. Single-shot refinement neural network for object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 4203–4212.
77. Cui, L.; Ma, R.; Lv, P.; Jiang, X.; Gao, Z.; Zhou, B.; Xu, M. MDSSD: multi-scale deconvolutional single shot detector for small objects. *arXiv preprint arXiv:1805.07009* **2018**.
78. Lin, T.; Zhao, X.; Shou, Z. Single shot temporal action detection. Proceedings of the 25th ACM international conference on Multimedia, 2017, pp. 988–996.
79. Tang, X.; Du, D.K.; He, Z.; Liu, J. Pyramidbox: A context-assisted single shot face detector. Proceedings of the European conference on computer vision (ECCV), 2018, pp. 797–813.
80. Li, Z.; Yang, L.; Zhou, F. FSSD: feature fusion single shot multibox detector. *arXiv preprint arXiv:1712.00960* **2017**.
81. Jiang, J.; Xu, H.; Zhang, S.; Fang, Y. Object detection algorithm based on multiheaded attention. *Applied Sciences* **2019**, *9*, 1829.
82. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A Review of Yolo algorithm developments. *Procedia computer science* **2022**, *199*, 1066–1073.
83. Ragab, M.G.; Abdulkader, S.J.; Muneer, A.; Alqushaibi, A.; Sumiea, E.H.; Qureshi, R.; Al-Selwi, S.M.; Alhussian, H. A Comprehensive Systematic Review of YOLO for Medical Object Detection (2018 to 2023). *IEEE Access* **2024**.
84. Terven, J.; Córdova-Esparza, D.M.; Romero-González, J.A. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine Learning and Knowledge Extraction* **2023**, *5*, 1680–1716.
85. Hussain, M. YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. *Machines* **2023**, *11*, 677.
86. Rothe, R.; Guillaumin, M.; Van Gool, L. Non-maximum suppression for object detection by passing messages between windows. Computer Vision–ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part I 12. Springer, 2015, pp. 290–306.

87. Li, S.; Li, M.; Li, R.; He, C.; Zhang, L. One-to-few label assignment for end-to-end dense detection. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 7350–7359.
88. Bhagat, S.; Kokare, M.; Haswani, V.; Hambarde, P.; Kamble, R. WheatNet-lite: A novel light weight network for wheat head detection. Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 1332–1341.
89. Hu, Y.; Tan, W.; Meng, F.; Liang, Y. A Decoupled Spatial-Channel Inverted Bottleneck For Image Compression. 2023 IEEE International Conference on Image Processing (ICIP). IEEE, 2023, pp. 1740–1744.
90. Yang, G.; Wang, J.; Nie, Z.; Yang, H.; Yu, S. A lightweight YOLOv8 tomato detection algorithm combining feature enhancement and attention. *Agronomy* **2023**, *13*, 1824.
91. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer, 2014, pp. 740–755.
92. Jocher, G.; others. YOLOv8: A Comprehensive Improvement of the YOLO Object Detection Series. https://docs.ultralytics.com/yolov8/, 2022. Accessed: 2024-06-05.
93. Tishby, N.; Zaslavsky, N. Deep learning and the information bottleneck principle. 2015 ieee information theory workshop (itw). IEEE, 2015, pp. 1–5.
94. Zhang, B.; Li, J.; Bai, Y.; Jiang, Q.; Yan, B.; Wang, Z. An Improved Microaneurysm Detection Model Based on SwinIR and YOLOv8. *Bioengineering* **2023**, *10*, 1405.
95. Chien, C.T.; Ju, R.Y.; Chou, K.Y.; Chiang, J.S. YOLOv9 for Fracture Detection in Pediatric Wrist Trauma X-ray Images. *arXiv preprint arXiv:2403.11249* **2024**.
96. YOLOv8 Object Detection Model: What is, How to Use — roboflow.com. https://roboflow.com/model/yolov8. [Accessed 28-05-2024].
97. Ultralytics. Ultralytics YOLOv8 Solutions: Quick Walkthrough — ultralytics.medium.com. https://ultralytics.medium.com/ultralytics-yolov8-solutions-quick-walkthrough-b802fd6da5d7. [Accessed 28-05-2024].
98. Du, S.; Zhang, B.; Zhang, P.; Xiang, P. An improved bounding box regression loss function based on CIOU loss for multi-scale object detection. 2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML). IEEE, 2021, pp. 92–98.
99. Xu, S.; Wang, X.; Lv, W.; Chang, Q.; Cui, C.; Deng, K.; Wang, G.; Dang, Q.; Wei, S.; Du, Y.; others. PP-YOLOE: An evolved version of YOLO. *arXiv preprint arXiv:2203.16250* **2022**.
100. Yue, X.; Qi, K.; Na, X.; Zhang, Y.; Liu, Y.; Liu, C. Improved YOLOv8-Seg Network for Instance Segmentation of Healthy and Diseased Tomato Plants in the Growth Stage. *Agriculture* **2023**, *13*, 1643.
101. Zhu, X.; Lyu, S.; Wang, X.; Zhao, Q. TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 2778–2788.
102. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3–19.
103. Hnewa, M.; Radha, H. Integrated multiscale domain adaptive yolo. *IEEE Transactions on Image Processing* **2023**, *32*, 1857–1867.
104. Bai, Z.; Pei, X.; Qiao, Z.; Wu, G.; Bai, Y. Improved YOLOv7 Target Detection Algorithm Based on UAV Aerial Photography. *Drones* **2024**, *8*, 104.
105. Jocher, G.; Team, U. YOLOv5. https://github.com/ultralytics/yolov5, 2020. Accessed: 2024-06-05.
106. Team, M. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. https://github.com/meituan/YOLOv6, 2022. Accessed: 2024-06-05.
107. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696* **2022**. Accessed: 2024-06-05.
108. Sirisha, U.; Praveen, S.P.; Srinivasu, P.N.; Barsocchi, P.; Bhoi, A.K. Statistical analysis of design aspects of various YOLO-based deep learning models for object detection. *International Journal of Computational Intelligence Systems* **2023**, *16*, 126.
109. Wang, K.; Liew, J.H.; Zou, Y.; Zhou, D.; Feng, J. Panet: Few-shot image semantic segmentation with prototype alignment. proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 9197–9206.
110. Zhang, Z.; Lu, X.; Cao, G.; Yang, Y.; Jiao, L.; Liu, F. ViT-YOLO: Transformer-based YOLO for object detection. Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 2799–2808.

 doi:10.20944/preprints202406.1366.v2

111. Mahasin, M.; Dewi, I.A. Comparison of cspdarknet53, cspresnext-50, and efficientnet-b0 backbones on yolo v4 as object detector. *International Journal of Engineering, Science and Information Technology* **2022**, *2*, 64–72.
112. Misra, D. Mish: A self regularized non-monotonic activation function. *arXiv preprint arXiv:1908.08681* **2019**.
113. Yun, S.; Han, D.; Oh, S.J.; Chun, S.; Choe, J.; Yoo, Y. Cutmix: Regularization strategy to train strong classifiers with localizable features. Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 6023–6032.
114. Ghiasi, G.; Lin, T.Y.; Le, Q.V. Dropblock: A regularization method for convolutional networks. *Advances in neural information processing systems* **2018**, *31*.
115. Müller, R.; Kornblith, S.; Hinton, G.E. When does label smoothing help? *Advances in neural information processing systems* **2019**, *32*.
116. Zhang, Z.; He, T.; Zhang, H.; Zhang, Z.; Xie, J.; Li, M. Bag of freebies for training object detection neural networks. *arXiv preprint arXiv:1902.04103* **2019**.
117. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv preprint arXiv:1506.02640* **2016**. Accessed: 2024-06-05.
118. Santurkar, S.; Tsipras, D.; Ilyas, A.; Madry, A. How does batch normalization help optimization? *Advances in neural information processing systems* **2018**, *31*.
119. Vijayakumar, A.; Vairavasundaram, S. Yolo-based object detection models: A review and its applications. *Multimedia Tools and Applications* **2024**, pp. 1–40.
120. Benjumea, A.; Teeti, I.; Cuzzolin, F.; Bradley, A. YOLO-Z: Improving small object detection in YOLOv5 for autonomous vehicles. *arXiv preprint arXiv:2112.11798* **2021**.
121. Sarda, A.; Dixit, S.; Bhan, A. Object detection for autonomous driving using yolo [you only look once] algorithm. 2021 Third international conference on intelligent communication technologies and virtual mobile networks (ICICV). IEEE, 2021, pp. 1370–1374.
122. Cai, Y.; Luan, T.; Gao, H.; Wang, H.; Chen, L.; Li, Y.; Sotelo, M.A.; Li, Z. YOLOv4-5D: An effective and efficient object detector for autonomous driving. *IEEE Transactions on Instrumentation and Measurement* **2021**, *70*, 1–13.
123. Zhao, J.; Hao, S.; Dai, C.; Zhang, H.; Zhao, L.; Ji, Z.; Ganchev, I. Improved vision-based vehicle detection and classification by optimized YOLOv4. *IEEE Access* **2022**, *10*, 8590–8603.
124. Woo, J.; Baek, J.H.; Jo, S.H.; Kim, S.Y.; Jeong, J.H. A study on object detection performance of YOLOv4 for autonomous driving of tram. *Sensors* **2022**, *22*, 9026.
125. Ye, C.; Wang, Y.; Wang, Y.; Tie, M. Steering angle prediction YOLOv5-based end-to-end adaptive neural network control for autonomous vehicles. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering* **2022**, *236*, 1991–2011.
126. Mostafa, T.; Chowdhury, S.J.; Rhaman, M.K.; Alam, M.G.R. Occluded object detection for autonomous vehicles employing YOLOv5, YOLOX and Faster R-CNN. 2022 IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). IEEE, 2022, pp. 0405–0410.
127. Jia, X.; Tong, Y.; Qiao, H.; Li, M.; Tong, J.; Liang, B. Fast and accurate object detector for autonomous driving based on improved YOLOv5. *Scientific reports* **2023**, *13*, 9711.
128. Chen, Z.; Wang, X.; Zhang, W.; Yao, G.; Li, D.; Zeng, L. Autonomous Parking Space Detection for Electric Vehicles Based on Improved YOLOV5-OBB Algorithm. *World Electric Vehicle Journal* **2023**, *14*, 276.
129. Liu, X.; Yan, W.Q. Vehicle-related distance estimation using customized YOLOv7. International Conference on Image and Vision Computing New Zealand. Springer, 2022, pp. 91–103.
130. Mehla, N.; Ishita.; Talukdar, R.; Sharma, D.K. Object Detection in Autonomous Maritime Vehicles: Comparison Between YOLO V8 and EfficientDet. International Conference on Data Science and Network Engineering. Springer, 2023, pp. 125–141.
131. Patel, P.; Vekariya, V.; Shah, J.; Vala, B. Detection of traffic sign based on YOLOv8. AIP Conference Proceedings. AIP Publishing, 2024, Vol. 3107.
132. Soylu, E.; Soylu, T. A performance comparison of YOLOv8 models for traffic sign detection in the Robotaxi-full scale autonomous vehicle competition. *Multimedia Tools and Applications* **2024**, *83*, 25005–25035.
133. Wang, H.; Liu, C.; Cai, Y.; Chen, L.; Li, Y. YOLOv8-QSD: An Improved Small Object Detection Algorithm for Autonomous Vehicles Based on YOLOv8. *IEEE Transactions on Instrumentation and Measurement* **2024**.
134. Kumar, D.; Muhammad, N. Object detection in adverse weather for autonomous driving through data merging and YOLOv8. *Sensors* **2023**, *23*, 8471.

135. Oh, G.; Lim, S. One-Stage Brake Light Status Detection Based on YOLOv8. *Sensors* **2023**, *23*, 7436.
136. Afdhal, A.; Saddami, K.; Sugiarto, S.; Fuadi, Z.; Nasaruddin, N. Real-Time Object Detection Performance of YOLOv8 Models for Self-Driving Cars in a Mixed Traffic Environment. 2023 2nd International Conference on Computer System, Information Technology, and Electrical Engineering (COSITE). IEEE, 2023, pp. 260–265.
137. Bakirci, M.; Bayraktar, I. Transforming Aircraft Detection Through LEO Satellite Imagery and YOLOv9 for Improved Aviation Safety. 2024 26th International Conference on Digital Signal Processing and its Applications (DSPA). IEEE, 2024, pp. 1–6.
138. Wibowo, A.; Trilaksono, B.R.; Hidayat, E.M.I.; Munir, R. Object Detection in Dense and Mixed Traffic for Autonomous Vehicles With Modified Yolo. *IEEE Access* **2023**, *11*, 134866–134877.
139. Kaur, R.; Singh, J. Local Regression Based Real-Time Traffic Sign Detection using YOLOv6. 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N). IEEE, 2022, pp. 522–526.
140. Mahaur, B.; Mishra, K. Small-object detection based on YOLOv5 in autonomous driving systems. *Pattern Recognition Letters* **2023**, *168*, 115–122.
141. Dewi, C.; Chen, R.C.; Jiang, X.; Yu, H. Deep convolutional neural network for enhancing traffic sign recognition developed on Yolo V4. *Multimedia Tools and Applications* **2022**, *81*, 37821–37845.
142. Zaghari, N.; Fathy, M.; Jameii, S.M.; Shahverdy, M. The improvement in obstacle detection in autonomous vehicles using YOLO non-maximum suppression fuzzy algorithm. *The Journal of Supercomputing* **2021**, *77*, 13421–13446.
143. Hung, W.C.W.; Zakaria, M.A.; Ishak, M.; Heerwan, P. Object Tracking for Autonomous Vehicle Using YOLO V3. In *Enabling Industry 4.0 through Advances in Mechatronics: Selected Articles from iM3F 2021, Malaysia*; Springer, 2022; pp. 265–273.
144. Pandey, S.; Chen, K.F.; Dam, E.B. Comprehensive multimodal segmentation in medical imaging: Combining yolov8 with sam and hq-sam models. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 2592–2598.
145. Ju, R.Y.; Cai, W. Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm. *Scientific Reports* **2023**, *13*, 20077.
146. Inui, A.; Mifune, Y.; Nishimoto, H.; Mukohara, S.; Fukuda, S.; Kato, T.; Furukawa, T.; Tanaka, S.; Kusunose, M.; Takigami, S.; others. Detection of elbow OCD in the ultrasound image by artificial intelligence using YOLOv8. *Applied Sciences* **2023**, *13*, 7623.
147. Luo, Y.; Zhang, Y.; Sun, X.; Dai, H.; Chen, X.; others. Intelligent solutions in chest abnormality detection based on YOLOv5 and ResNet50. *Journal of healthcare engineering* **2021**, *2021*.
148. Wu, B.; Pang, C.; Zeng, X.; Hu, X. Me-yolo: Improved yolov5 for detecting medical personal protective equipment. *Applied Sciences* **2022**, *12*, 11978.
149. Zhao, X.; Wang, Q.; Zhang, M.; Wei, Z.; Ku, R.; Zhang, Z.; Yu, Y.; Zhang, B.; Liu, Y.; Wang, C. CSFF-YOLOv5: Improved YOLOv5 based on channel split and feature fusion in femoral neck fracture detection. *Internet of Things* **2024**, *26*, 101190.
150. Goel, L.; Patel, P. Improving YOLOv6 using advanced PSO optimizer for weight selection in lung cancer detection and classification. *Multimedia Tools and Applications* **2024**, pp. 1–34.
151. Norkobil Saydirasulovich, S.; Abdusalomov, A.; Jamil, M.K.; Nasimov, R.; Kozhamzharova, D.; Cho, Y.I. A YOLOv6-based improved fire detection approach for smart city environments. *Sensors* **2023**, *23*, 3161.
152. Zou, J.; Arshad, M.R. Detection of whole body bone fractures based on improved YOLOv7. *Biomedical Signal Processing and Control* **2024**, *91*, 105995.
153. Salinas-Medina, A.; Neme, A. Enhancing Hospital Efficiency Through Web-Deployed Object Detection: A YOLOv8-Based Approach for Automating Healthcare Operations. 2023 Mexican International Conference on Computer Science (ENC). IEEE, 2023, pp. 1–6.
154. Razaghi, M.; Komleh, H.E.; Dehghani, F.; Shahidi, Z. Innovative Diagnosis of Dental Diseases Using YOLO V8 Deep Learning Model. 2024 13th Iranian/3rd International Machine Vision and Image Processing Conference (MVIP). IEEE, 2024, pp. 1–5.
155. Pham, T.L.; Le, V.H. Ovarian Tumors Detection and Classification from Ultrasound Images Based on YOLOv8. *Journal of Advances in Information Technology* **2024**, *15*.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 24 June 2024 doi:10.20944/preprints202406.1366.v2

156. Krishnamurthy, V.; Balasubramanian, S.; Kanmani, R.S.; Srividhya, S.; Deepika, J.; Nimeshika, G.N. Endoscopic Surgical Operation and Object Detection Using Custom Architecture Models. International Conference on Human-Centric Smart Computing. Springer, 2023, pp. 637–654.
157. Palanivel, N.; Deivanai, S.; Sindhuja, B.; others. The Art of YOLOv8 Algorithm in Cancer Diagnosis using Medical Imaging. 2023 International Conference on System, Computation, Automation and Networking (ICSCAN). IEEE, 2023, pp. 1–6.
158. Karaköse, M.; Yetış, H.; Çeçen, M. A New Approach for Effective Medical Deepfake Detection in Medical Images. *IEEE Access* **2024**.
159. Bhojane, R.; Chourasia, S.; Laddha, S.V.; Ochawar, R.S. Liver Lesion Detection from MR T1 In-Phase and Out-Phase Fused Images and CT Images Using YOLOv8. International Conference on Data Science and Applications. Springer, 2023, pp. 121–135.
160. Majeed, F.; Khan, F.Z.; Nazir, M.; Iqbal, Z.; Alhaisoni, M.; Tariq, U.; Khan, M.A.; Kadry, S. Investigating the efficiency of deep learning based security system in a real-time environment using YOLOv5. *Sustainable Energy Technologies and Assessments* **2022**, *53*, 102603.
161. AFFES, N.; KTARI, J.; BEN AMOR, N.; FRIKHA, T.; HAMAM, H. Comparison of YOLOV5, YOLOV6, YOLOV7 and YOLOV8 for Intelligent Video Surveillance. *Journal of Information Assurance & Security* **2023**, *18*.
162. Cao, F.; Ma, S. Enhanced Campus Security Target Detection Using a Refined YOLOv7 Approach. *Traitement du Signal* **2023**, *40*.
163. Chatterjee, N.; Singh, A.V.; Agarwal, R. You Only Look Once (YOLOv8) Based Intrusion Detection System for Physical Security and Surveillance. 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). IEEE, 2024, pp. 1–5.
164. Sandhya.; Kashyap, A. Real-time object-removal tampering localization in surveillance videos by employing YOLO-V8. *Journal of Forensic Sciences* **2024**.
165. Bakirci, M.; Bayraktar, I. Boosting aircraft monitoring and security through ground surveillance optimization with YOLOv9. 2024 12th International Symposium on Digital Forensics and Security (ISDFS). IEEE, 2024, pp. 1–6.
166. Bakirci, M.; Bayraktar, I. YOLOv9-Enabled Vehicle Detection for Urban Security and Forensics Applications. 2024 12th International Symposium on Digital Forensics and Security (ISDFS). IEEE, 2024, pp. 1–6.
167. Chakraborty, S.; Zahir, S.; Orchi, N.T.; Hafiz, M.F.B.; Shamsuddoha, A.; Dipto, S.M. Violence Detection: A Multi-Model Approach Towards Automated Video Surveillance and Public Safety. 2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS). IEEE, 2024, pp. 1–6.
168. Julia, R.; Prince, S.; Bini, D. Medical image analysis of masses in mammography using deep learning model for early diagnosis of cancer tissues. In *Computational Intelligence and Modelling Techniques for Disease Detection in Mammogram Images*; Elsevier, 2024; pp. 75–89.
169. Salahin, S.S.; Ullaa, M.S.; Ahmed, S.; Mohammed, N.; Farook, T.H.; Dudley, J. One-stage methods of computer vision object detection to classify carious lesions from smartphone imaging. *Oral* **2023**, *3*, 176–190.
170. Doniyorjon, M.; Madinakhon, R.; Shakhnoza, M.; Cho, Y.I. An improved method of polyp detection using custom YOLOv4-tiny. *Applied Sciences* **2022**, *12*, 10856.
171. Ding, B.; Zhang, Z.; Liang, Y.; Wang, W.; Hao, S.; Meng, Z.; Guan, L.; Hu, Y.; Guo, B.; Zhao, R.; others. Detection of dental caries in oral photographs taken by mobile phones based on the YOLOv3 algorithm. *Annals of Translational Medicine* **2021**, *9*.
172. Wang, L.; Yang, S.; Yang, S.; Zhao, C.; Tian, G.; Gao, Y.; Chen, Y.; Lu, Y. Automatic thyroid nodule recognition and diagnosis in ultrasound imaging with the YOLOv2 neural network. *World journal of surgical oncology* **2019**, *17*, 1–9.
173. Kwaśniewska, A.; Rumiński, J.; Czuszyński, K.; Szankin, M. Real-time facial features detection from low resolution thermal images with deep classification models. *Journal of Medical Imaging and Health Informatics* **2018**, *8*, 979–987.
174. Chen, G.; Du, W.; Xu, T.; Wang, S.; Qi, X.; Wang, Y. Investigating Enhanced YOLOv8 Model Applications for Large-Scale Security Risk Management and Drone-Based Low-Altitude Law Enforcement. *Highlights in Science, Engineering and Technology* **2024**, *98*, 390–396.
175. Pashayev, F.; Babayeva, L.; Isgandarova, Z.; Kalejahi, B.K. Face Recognition in Smart Cameras by Yolo8. *KHAZAR JOURNAL OF SCIENCE AND TECHNOLOGY (KJSAT)* **2023**, p. 67.

176. Kaç, S.B.; Eken, S.; Balta, D.D.; Balta, M.; İskefiyeli, M.; Özçelik, İ. Image-based security techniques for water critical infrastructure surveillance. *Applied Soft Computing* **2024**, p. 111730.
177. Gao, Q.; Deng, H.; Zhang, G. A Contraband Detection Scheme in X-ray Security Images Based on Improved YOLOv8s Network Model. *Sensors* **2024**, *24*, 1158.
178. Antony, J.C.; Chowdary, C.L.S.; Murali, E.; Mayan, A.; others. Advancing Crowd Management through Innovative Surveillance using YOLOv8 and ByteTrack. 2024 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET). IEEE, 2024, pp. 1–6.
179. Zhang, D. A Yolo-based Approach for Fire and Smoke Detection in IoT Surveillance Systems. *International Journal of Advanced Computer Science & Applications* **2024**, *15*.
180. Khin, P.P.; Htaik, N.M. Gun Detection: A Comparative Study of RetinaNet, EfficientDet and YOLOv8 on Custom Dataset. 2024 IEEE Conference on Computer Applications (ICCA). IEEE, 2024, pp. 1–7.
181. Nkuzo, L.; Sibiya, M.; Markus, E.D. A Comprehensive Analysis of Real-Time Car Safety Belt Detection Using the YOLOv7 Algorithm. *Algorithms* **2023**, *16*, 400.
182. Chang, R.; Zhang, B.; Zhu, Q.; Zhao, S.; Yan, K.; Yang, Y.; others. FFA-YOLOv7: Improved YOLOv7 Based on Feature Fusion and Attention Mechanism for Wearing Violation Detection in Substation Construction Safety. *Journal of Electrical and Computer Engineering* **2023**, *2023*.
183. Ahmad, H.M.; Rahimi, A. Deep learning methods for object detection in smart manufacturing: A survey. *Journal of Manufacturing Systems* **2022**, *64*, 181–196.
184. Pendse, R.; Rajput, H.; Saraf, S.; Sarwate, A.; Jadhav, J.; others. Defect Detection in Manufacturing using YOLOv7. *IJRAR-International Journal of Research and Analytical Reviews (IJRAR)* **2023**, *10*, 179–185.
185. Yi, F.; Zhang, H.; Yang, J.; He, L.; Mohamed, A.S.A.; Gao, S. YOLOv7-SiamFF: Industrial defect detection algorithm based on improved YOLOv7. *Computers and Electrical Engineering* **2024**, *114*, 109090.
186. Wang, H.; Xu, X.; Liu, Y.; Lu, D.; Liang, B.; Tang, Y. Real-time defect detection for metal components: a fusion of enhanced Canny–Devernay and YOLOv6 algorithms. *Applied Sciences* **2023**, *13*, 6898.
187. Ludwika, A.S.; Rifai, A.P. Deep learning for detection of proper utilization and adequacy of personal protective equipment in manufacturing teaching laboratories. *Safety* **2024**, *10*, 26.
188. Beak, S.; Han, Y.H.; Moon, Y.; Lee, J.; Jeong, J. YOLOv7-Based Anomaly Detection Using Intensity and NG Types in Labeling in Cosmetic Manufacturing Processes. *Processes* **2023**, *11*, 2266.
189. Zhao, H.; Wang, X.; Sun, J.; Wang, Y.; Chen, Z.; Wang, J.; Xu, X. Artificial intelligence powered real-time quality monitoring for additive manufacturing in construction. *Construction and Building Materials* **2024**, *429*, 135894.
190. Liu, Z.; Ye, K. YOLO-IMF: an improved YOLOv8 algorithm for surface defect detection in industrial manufacturing field. International Conference on Metaverse. Springer, 2023, pp. 15–28.
191. Wen, Y.; Wang, L. Yolo-sd: simulated feature fusion for few-shot industrial defect detection based on YOLOv8 and stable diffusion. *International Journal of Machine Learning and Cybernetics* **2024**, pp. 1–13.
192. Karna, N.; Putra, M.A.P.; Rachmawati, S.M.; Abisado, M.; Sampedro, G.A. Towards Accurate Fused Deposition Modeling 3D Printer Fault Detection using Improved YOLOv8 with Hyperparameter Optimization. *IEEE Access* **2023**.
193. Li, W.; Solihin, M.I.; Nugroho, H.A. RCA: YOLOv8-Based Surface Defects Detection on the Inner Wall of Cylindrical High-Precision Parts. *Arabian Journal for Science and Engineering* **2024**, pp. 1–19.
194. Hu, Y.; Wang, J.; Wang, X.; Sun, Y.; Yu, H.; Zhang, J. Real-time evaluation of the blending uniformity of industrially produced gravelly soil based on Cond-YOLOv8-seg. *Journal of Industrial Information Integration* **2024**, *39*, 100603.
195. Han, L.; Ma, C.; Liu, Y.; Jia, J.; Sun, J. SC-YOLOv8: A Security Check Model for the Inspection of Prohibited Items in X-ray Images. *Electronics* **2023**, *12*, 4208.
196. Yuan, J.; Zhang, N.; Xie, Y.; Gao, X. Detection of Prohibited Items Based upon X-ray Images and Improved YOLOv7. Journal of Physics: Conference Series; 3rd International Conference on Advanced Materials and Intelligent Manufacturing (ICAMIM 2022), , 2022; Vol. 2390, p. 012114. doi:10.1088/1742-6596/2390/1/012114.
197. Awang, S.; Rokei, M.Q.R.; Sulaiman, J. Suspicious Activity Trigger System using YOLOv6 Convolutional Neural Network. 2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). IEEE, 2023, pp. 527–532.

198. Xiao, Y.; Chang, A.; Wang, Y.; Huang, Y.; Yu, J.; Huo, L. Real-time object detection for substation security early-warning with deep neural network based on YOLO-V5. 2022 IEEE IAS Global Conference on Emerging Technologies (GlobConET). IEEE, 2022, pp. 45–50.
199. Wang, G.; Ding, H.; Duan, M.; Pu, Y.; Yang, Z.; Li, H. Fighting against terrorism: A real-time CCTV autonomous weapons detection based on improved YOLO v4. *Digital Signal Processing* **2023**, *132*, 103790.
200. Kashika, P.; Venkatapur, R.B. Automatic tracking of objects using improvised Yolov3 algorithm and alarm human activities in case of anomalies. *International Journal of Information Technology* **2022**, *14*, 2885–2891.
201. Mohandoss, T.; Rangaraj, J. Multi-Object Detection using Enhanced YOLOv2 and LuNet Algorithms in Surveillance Videos. *e-Prime-Advances in Electrical Engineering, Electronics and Energy* **2024**, *8*, 100535.
202. Smiley, C.J. From silence to propagation: Understanding the relationship between "Stop Snitchin" and "YOLO". *Deviant Behavior* **2015**, *36*, 1–16.
203. Yang, S.; Zhang, Z.; Wang, B.; Wu, J. DCS-YOLOv8: An Improved Steel Surface Defect Detection Algorithm Based on YOLOv8. Proceedings of the 2024 7th International Conference on Image and Graphics Processing, 2024, pp. 39–46.
204. Wang, X.; Gao, H.; Jia, Z.; Li, Z. BL-YOLOv8: An improved road defect detection model based on YOLOv8. *Sensors* **2023**, *23*, 8361.
205. Luo, B.; Kou, Z.; Han, C.; Wu, J. A "Hardware-Friendly" Foreign Object Identification Method for Belt Conveyors Based on Improved YOLOv8. *Applied Sciences* **2023**, *13*, 11464.
206. Wang, C.; Sun, Q.; Dong, X.; Chen, J. Automotive adhesive defect detection based on improved YOLOv8. *Signal, Image and Video Processing* **2024**, pp. 1–13.
207. Wu, Q.; Kuang, X.; Tang, X.; Guo, D.; Luo, Z. Industrial equipment object detection based on improved YOLOv7. International Conference on Computer, Artificial Intelligence, and Control Engineering (CAICE 2023). SPIE, 2023, Vol. 12645, pp. 600–608.
208. Kim, O.; Han, Y.; Jeong, J. Real-time Inspection System Based on Moire Pattern and YOLOv7 for Coated High-reflective Injection Molding Product. *WSEAS Transactions on Computer Research* **2022**, *10*, 120–125.
209. Chen, J.; Bai, S.; Wan, G.; Li, Y. Research on YOLOv7-based defect detection method for automotive running lights. *Systems Science & Control Engineering* **2023**, *11*, 2185916.
210. Hussain, M.; Al-Aqrabi, H.; Munawar, M.; Hill, R.; Alsboui, T. Domain feature mapping with YOLOv7 for automated edge-based pallet racking inspections. *Sensors* **2022**, *22*, 6927.
211. Zhu, B.; Xiao, G.; Zhang, Y.; Gao, H. Multi-classification recognition and quantitative characterization of surface defects in belt grinding based on YOLOv7. *Measurement* **2023**, *216*, 112937.
212. Wang, Y.; Yang, L.; Chen, H.; Hussain, A.; Ma, C.; Al-gabri, M. Mushroom-YOLO: A deep learning algorithm for mushroom growth recognition based on improved YOLOv5 in agriculture 4.0. 2022 IEEE 20th International Conference on Industrial Informatics (INDIN). IEEE, 2022, pp. 239–244.
213. Badgujar, C.M.; Armstrong, P.R.; Gerken, A.R.; Pordesimo, L.O.; Campbell, J.F. Real-time stored product insect detection and identification using deep learning: System integration and extensibility to mobile platforms. *Journal of Stored Products Research* **2023**, *104*, 102196. doi:10.1016/j.jspr.2023.102196.
214. Bhat, S.; Shenoy, K.A.; Jain, M.R.; Manasvi, K. Detecting Crops and Weeds in Fields Using YOLOv6 and Faster R-CNN Object Detection Models. 2023 International Conference on Recent Advances in Information Technology for Sustainable Development (ICRAIS). IEEE, 2023, pp. 43–48.
215. Bist, R.B.; Subedi, S.; Yang, X.; Chai, L. A novel YOLOv6 object detector for monitoring piling behavior of cage-free laying hens. *AgriEngineering* **2023**, *5*, 905–923.
216. Jiang, K.; Xie, T.; Yan, R.; Wen, X.; Li, D.; Jiang, H.; Jiang, N.; Feng, L.; Duan, X.; Wang, J. An attention mechanism-improved YOLOv7 object detection algorithm for hemp duck count estimation. *Agriculture* **2022**, *12*, 1659.
217. Kumar, P.; Kumar, N. Drone-based apple detection: Finding the depth of apples using YOLOv7 architecture with multi-head attention mechanism. *Smart Agricultural Technology* **2023**, *5*, 100311.
218. Chen, G.; Hou, Y.; Cui, T.; Li, H.; Shangguan, F.; Cao, L. YOLOv8-CML: A lightweight target detection method for Color-changing melon ripening in intelligent agriculture. *ResearchSquare* **2023**.
219. Sapkota, R.; Ahmed, D.; Karkee, M. Synthetic Meets Authentic: Leveraging Text-to-Image Generated Datasets for Apple Detection in Orchard Environments. *Available at SSRN 4770722* **2024**.
220. Ahmed, D.; Sapkota, R.; Churuvija, M.; Karkee, M. Machine vision-based crop-load estimation using yolov8. *arXiv preprint arXiv:2304.13282* **2023**.

 doi:10.20944/preprints202406.1366.v2

221. Sapkota, R.; Ahmed, D.; Karkee, M. Comparing YOLOv8 and Mask RCNN for object segmentation in complex orchard environments. *arXiv preprint arXiv:2312.07935* **2023**.
222. Zhang, L.; Ding, G.; Li, C.; Li, D. DCF-Yolov8: An Improved Algorithm for Aggregating Low-Level Features to Detect Agricultural Pests and Diseases. *Agronomy* **2023**, *13*, 2012.
223. Badgujar, C.M.; Poulose, A.; Gan, H. Agricultural object detection with You Only Look Once (YOLO) Algorithm: A bibliometric and systematic literature review. *Computers and Electronics in Agriculture* **2024**, *223*, 109090. doi:10.1016/j.compag.2024.109090.
224. Junior, L.C.M.; Ulson, J.A.C. Real time weed detection using computer vision and deep learning. 2021 14th IEEE International Conference on Industry Applications (INDUSCON). IEEE, 2021, pp. 1131–1137.
225. Yu, X.; Yin, D.; Xu, H.; Pinto Espinosa, F.; Schmidhalter, U.; Nie, C.; Bai, Y.; Sankaran, S.; Ming, B.; Cui, N.; others. Maize tassel number and tasseling stage monitoring based on near-ground and UAV RGB images by improved YoloV8. *Precision Agriculture* **2024**, pp. 1–39.
226. Khalid, M.; Sarfraz, M.S.; Iqbal, U.; Aftab, M.U.; Niedbała, G.; Rauf, H.T. Real-time plant health detection using deep convolutional neural networks. *Agriculture* **2023**, *13*, 510.
227. Gallo, I.; Rehman, A.U.; Dehkordi, R.H.; Landro, N.; La Grassa, R.; Boschetti, M. Deep object detection of crop weeds: Performance of YOLOv7 on a real case dataset from UAV images. *Remote Sensing* **2023**, *15*, 539.
228. Gupta, C.; Gill, N.S.; Gulia, P.; Chatterjee, J.M. A novel finetuned YOLOv6 transfer learning model for real-time object detection. *Journal of Real-Time Image Processing* **2023**, *20*, 42.
229. Zendehdel, N.; Chen, H.; Leu, M.C. Real-time tool detection in smart manufacturing using You-Only-Look-Once (YOLO) v5. *Manufacturing Letters* **2023**, *35*, 1052–1059.
230. Jiang, D.; Wang, H.; Lu, Y. An efficient automobile assembly state monitoring system based on channel-pruned YOLOv4 algorithm. *International Journal of Computer Integrated Manufacturing* **2024**, *37*, 372–382.
231. Yan, J.; Wang, Z. YOLO V3+ VGG16-based automatic operations monitoring and analysis in a manufacturing workshop under Industry 4.0. *Journal of Manufacturing Systems* **2022**, *63*, 134–142.
232. Arima, K.; Nagata, F.; Shimizu, T.; Otsuka, A.; Kato, H.; Watanabe, K.; Habib, M.K. Improvements of detection accuracy and its confidence of defective areas by YOLOv2 using a data set augmentation method. *Artificial Life and Robotics* **2023**, *28*, 625–631.
233. Jia, L.; Wang, T.; Chen, Y.; Zang, Y.; Li, X.; Shi, H.; Gao, L. MobileNet-CA-YOLO: An improved YOLOv7 based on the MobileNetV3 and attention mechanism for Rice pests and diseases detection. *Agriculture* **2023**, *13*, 1285.
234. Vaidya, S.; Kavthekar, S.; Joshi, A. Leveraging YOLOv7 for plant disease detection. 2023 4th international conference on innovative trends in information technology (ICITIIT). IEEE, 2023, pp. 1–6.
235. Umar, M.; Altaf, S.; Ahmad, S.; Mahmoud, H.; Mohamed, A.S.N.; Ayub, R. Precision Agriculture Through Deep Learning: Tomato Plant Multiple Diseases Recognition with CNN and Improved YOLOv7. *IEEE Access* **2024**.
236. Zayani, H.M.; Ammar, I.; Ghodhbani, R.; Maqbool, A.; Saidani, T.; Slimane, J.B.; Kachoukh, A.; Kouki, M.; Kallel, M.; Alsuwaylimi, A.A.; others. Deep Learning for Tomato Disease Detection with YOLOv8. *Engineering, Technology & Applied Science Research* **2024**, *14*, 13584–13591.
237. Ma, B.; Hua, Z.; Wen, Y.; Deng, H.; Zhao, Y.; Pu, L.; Song, H. Using an improved lightweight YOLOv8 model for real-time detection of multi-stage apple fruit in complex orchard environments. *Artificial Intelligence in Agriculture* **2024**.
238. Junos, M.H.; Mohd Khairuddin, A.S.; Thannirmalai, S.; Dahari, M. An optimized YOLO-based object detection model for crop harvesting system. *IET Image Processing* **2021**, *15*, 2112–2125.
239. Zhao, H.; Tang, Z.; Li, Z.; Dong, Y.; Si, Y.; Lu, M.; Panoutsos, G. Real-time object detection and robotic manipulation for agriculture using a YOLO-based learning approach. *arXiv preprint arXiv:2401.15785* **2024**.
240. Chen, W.; Zhang, J.; Guo, B.; Wei, Q.; Zhu, Z. An apple detection method based on des-YOLO v4 algorithm for harvesting robots in complex environment. *Mathematical Problems in Engineering* **2021**, *2021*, 1–12.
241. NERGİZ, M. Enhancing Strawberry Harvesting Efficiency through Yolo-v7 Object Detection Assessment. *Turkish Journal of Science and Technology* **2023**, *18*, 519–533.
242. Wang, C.; Han, Q.; Li, C.; Li, J.; Kong, D.; Wang, F.; Zou, X. Assisting the Planning of Harvesting Plans for Large Strawberry Fields through Image-Processing Method Based on Deep Learning. *Agriculture* **2024**, *14*, 560.

243. Chen, W.; Lu, S.; Liu, B.; Chen, M.; Li, G.; Qian, T. CitrusYOLO: a algorithm for citrus detection under orchard environment based on YOLOV4. *Multimedia Tools and Applications* **2022**, *81*, 31363–31389.
244. Mirhaji, H.; Soleymani, M.; Asakereh, A.; Mehdizadeh, S.A. Fruit detection and load estimation of an orange orchard using the YOLO models through simple approaches in different imaging and illumination conditions. *Computers and Electronics in Agriculture* **2021**, *191*, 106533.
245. Sapkota, R.; Ahmed, D.; Churuvija, M.; Karkee, M. Immature Green Apple Detection and Sizing in Commercial Orchards using YOLOv8 and Shape Fitting Techniques. *IEEE Access* **2024**, *12*, 43436–43452.
246. Wu, D.; Lv, S.; Jiang, M.; Song, H. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. *Computers and Electronics in Agriculture* **2020**, *178*, 105742.
247. Wang, J.; Gao, Z.; Zhang, Y.; Zhou, J.; Wu, J.; Li, P. Real-time detection and location of potted flowers based on a ZED camera and a YOLO V4-tiny deep learning algorithm. *Horticulturae* **2021**, *8*, 21.
248. Khanal, S.R.; Sapkota, R.; Ahmed, D.; Bhattarai, U.; Karkee, M. Machine Vision System for Early-stage Apple Flowers and Flower Clusters Detection for Precision Thinning and Pollination. *IFAC-PapersOnLine* **2023**, *56*, 8914–8919.
249. Xiao, F.; Wang, H.; Xu, Y.; Zhang, R. Fruit detection and recognition based on deep learning for automatic harvesting: an overview and review. *Agronomy* **2023**, *13*, 1625.
250. Yijing, W.; Yi, Y.; Xue-fen, W.; Jian, C.; Xinyun, L. Fig fruit recognition method based on YOLO v4 deep learning. 2021 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON). IEEE, 2021, pp. 303–306.
251. Zhang, Y.; Li, L.; Chun, C.; Wen, Y.; Xu, G. Multi-scale feature adaptive fusion model for real-time detection in complex citrus orchard environments. *Computers and Electronics in Agriculture* **2024**, *219*, 108836.
252. Zhou, J.; Zhang, Y.; Wang, J. A dragon fruit picking detection method based on YOLOv7 and PSP-Ellipse. *Sensors* **2023**, *23*, 3803.
253. Xiuyan, G.; ZHANG, Y. Detection of Fruit using YOLOv8-based Single Stage Detectors. *International Journal of Advanced Computer Science & Applications* **2023**, *14*.
254. Boudaa, B.; Abada, K.; Aichouche, W.A.; Belakermi, A.N. Advancing Plant Diseases Detection with Pre-trained YOLO Models. 2024 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS). IEEE, 2024, pp. 1–6.
255. Vo, H.T.; Mui, K.C.; Thien, N.N.; Tien, P.P. Automating Tomato Ripeness Classification and Counting with YOLOv9. *International Journal of Advanced Computer Science & Applications* **2024**, *15*.
256. Li, J.; Qiao, Y.; Liu, S.; Zhang, J.; Yang, Z.; Wang, M. An improved YOLOv5-based vegetable disease detection method. *Computers and Electronics in Agriculture* **2022**, *202*, 107345.
257. Fu, L.; Feng, Y.; Wu, J.; Liu, Z.; Gao, F.; Majeed, Y.; Al-Mallahi, A.; Zhang, Q.; Li, R.; Cui, Y. Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model. *Precision Agriculture* **2021**, *22*, 754–776.
258. Zhao, J.; Qu, J. A detection method for tomato fruit common physiological diseases based on YOLOv2. 2019 10th international conference on Information Technology in Medicine and Education (ITME). IEEE, 2019, pp. 559–563.
259. Zhong, Y.; Gao, J.; Lei, Q.; Zhou, Y. A vision-based counting and recognition system for flying insects in intelligent agriculture. *Sensors* **2018**, *18*, 1489.
260. Ye, R.; Gao, Q.; Qian, Y.; Sun, J.; Li, T. Improved Yolov8 and Sahi Model for the Collaborative Detection of Small Targets at the Micro Scale: A Case Study of Pest Detection in Tea. *Agronomy* **2024**, *14*, 1034.
261. Olorunshola, O.E.; Irhebhude, M.E.; Evwiekpaefe, A.E. A comparative study of YOLOv5 and YOLOv7 object detection algorithms. *Journal of Computing and Social Informatics* **2023**, *2*, 1–12.
262. Li, N.; Wang, M.; Yang, G.; Li, B.; Yuan, B.; Xu, S. DENS-YOLOv6: A small object detection model for garbage detection on water surface. *Multimedia Tools and Applications* **2023**, pp. 1–21.
263. Jung, H.K.; Choi, G.S. Improved yolov5: Efficient object detection using drone images under various conditions. *Applied Sciences* **2022**, *12*, 7255.
264. Wang, A.; Peng, T.; Cao, H.; Xu, Y.; Wei, X.; Cui, B. TIA-YOLOv5: An improved YOLOv5 network for real-time detection of crop and weed in the field. *Frontiers in Plant Science* **2022**, *13*, 1091655.
265. Wu, T.H.; Wang, T.W.; Liu, Y.Q. Real-time vehicle and distance detection based on improved yolo v5 network. 2021 3rd World Symposium on Artificial Intelligence (WSAI). IEEE, 2021, pp. 24–28.

 doi:10.20944/preprints202406.1366.v2

266. Pande, A.K.; Brantley, P.; Tanveer, M.H.; Voicu, R.C. From AI to AGI-The Evolution of Real-Time Systems with GPT Integration. SoutheastCon 2024. IEEE, 2024, pp. 699–707.
267. Manzotti, R. Embodied AI beyond embodied cognition and enactivism. *Philosophies* **2019**, *4*, 39.
268. Li, C.; Zhang, R.; Wong, J.; Gokmen, C.; Srivastava, S.; Martín-Martín, R.; Wang, C.; Levine, G.; Ai, W.; Martinez, B.; others. Behavior-1k: A human-centered, embodied ai benchmark with 1,000 everyday activities and realistic simulation. *arXiv preprint arXiv:2403.09227* **2024**.