

You Only Look Once (YOLO) Real-time Object Detection Algorithms: A Systematic Literature Review

Christine Bukola Asaju¹

chrisamaju0202@gmail.com

Florence Funke Abiola²

funke.abiola.940@gmail.com

Department of Computer Science, University of Ilorin, Kwara State,
Nigeria¹

Department of Computer Science, Federal Polytechnic, Idah Kogi State,
Nigeria²

ABSTRACT

Real-time object detection is essential in various computer vision applications, such as autonomous driving, surveillance, and robotics. Among deep learning-based object detection frameworks, the You Only Look Once (YOLO) algorithm family stands out for its impressive speed, accuracy, and efficiency. This paper provides a systematic review of YOLO-based object detection models, examining their development over the years, enhancements, and practical uses. It explores the evolution of YOLO from its initial version to the latest advancements, focusing on key architectural changes, training methodologies, and performance optimizations. Furthermore, YOLO is compared with other leading object detection models, which outline its strengths and limitations. The review also explores domain-specific applications of YOLO, including medical imaging, autonomous navigation, and industrial automation. By consolidating findings from recent research, this study sheds light on existing challenges and potential future directions in real-time object detection using YOLO.

Keywords: *YOLO; Object Detection; Deep Learning; Real-time Detection; Computer Vision; Systematic Review.*

INTRODUCTION

In the constantly developing science of computer vision, object detection is an important method that gives computers the remarkable ability to identify and locate a vast array of objects in visual input Challa (2023). From surveillance systems that keep an eye on public spaces Mirzaei, Nezamabadi-Pour, Raoof, and Derakhshani (2023) to autonomous vehicles

navigating congested streets Charroud et al. (2024), object detection has a wide range of innovative applications. These include face detection, face recognition, pedestrian counting, security systems, vehicle detection Kaur and Singh (2022), and more. Object detection goes beyond the limitations of basic image recognition by not only determining the presence of objects in an image but also specifying their exact locations within the visual field Kanjee (2023). This advanced capability is achieved through the implementation of sophisticated algorithms, with convolutional neural networks (CNNs) and deep learning techniques playing a central role. These algorithms are subjected to extensive training on large and diverse datasets, allowing them to learn and identify complex patterns, shapes, and distinguishing features that characterize individual objects. By leveraging deep learning methodologies, object detection systems can achieve high levels of accuracy and adaptability, making them essential for applications ranging from autonomous vehicles and surveillance to medical imaging and augmented reality.

Object detection algorithms are generally categorized into one-shot and two-shot approaches, depending on how many times an input image is processed by the network Chen et al. (2022). Two-shot algorithms were among the first to be developed, with notable examples including R-CNN (Region-based Convolutional Neural Networks) girshick2014rich, Fast R-CNN, and Faster R-CNN girshick2015fast. These algorithms typically involve an initial step to generate region proposals, followed by a second step that classifies and refines the detected objects. Building upon the foundation of Faster R-CNN, several optimization techniques have been introduced to enhance performance, including improvements in feature extraction networks, refinements to the Region Proposal Network (RPN), advancements in Region of Interest (ROI) classification, enhanced sample post-processing, and the use of larger mini-batches for training Chen et al. (2022). These developments have led to the emergence of more sophisticated and widely adopted algorithms, such as Region-based Fully Convolutional Networks (R-FCN) and Mask R-CNN, which further refine detection accuracy and segmentation capabilities Chen et al. (2022).

In two-shot object detection, the input image is processed twice to accurately determine both the presence and the precise location of objects. During the first pass, the algorithm generates a series of initial predictions or possible object positions. These preliminary detections are then refined in the second pass, where the model reassesses the proposed regions

and makes the final predictions with improved accuracy. Although this approach requires greater computational resources compared to single-shot object detection, it generally achieves higher precision, making it particularly suitable for applications that require detailed and reliable object localization Sharma and Syal (2023). Single-shot object detection predicts the existence and location of objects in an image by processing the input in a single pass. These methods are highly computationally efficient, since they analyze the entire image in one step, making them well suited for real-time object detection, especially in resource-constrained environments Sharma and Syal (2023). Among single-shot detection models, YOLO (You Only Look Once) is a seminal approach known for its balance between speed and accuracy. Following YOLO, SSD (Single Shot MultiBox Detector) emerged as another influential model, paving the way for further advancements such as Redefined Single Shot Multibox Detector (R-SSD), Deconvolutional Single Shot Detector (DSSD), Deeply Supervised Object Detector (DSOD), and Feature Space Singularity for Out-of-Distribution Detection (FSSD). Although there are several one-shot detection models, YOLO generally outperforms others in terms of speed and accuracy.

Given its computational efficiency and real-time capabilities, the objective of this study is to perform a systematic review of YOLO algorithms used for object detection between the years 2024 and 2025.

Real-time object recognition came into existence through Joseph Redmon's YOLO framework. He introduced a grid-based method for concurrently predicting bounding boxes and class probabilities Jiang, Zhao, Li, and Jia (2020). YOLO is especially well-suited for applications that require real-time performance because of its creative design, which makes detection extremely effective. YOLO is especially well-suited for applications that require real-time performance because of its innovative design, which makes detection extremely effective. The YOLO architecture has continuously evolved since its conception, with each iteration from YOLOv1 to the most recent YOLOv12, introducing improvements in efficiency, speed, and accuracy. Important developments include new loss functions designed to maximize performance and the addition of anchor boxes, which enhance the model's detection capabilities for objects of various sizes and forms. Every YOLO iteration also has optimized backbones, which help with quicker processing and improved detection capabilities. The YOLO framework's development between 2015 and 2023 demonstrates the iterative improvements achieved in response to new object detection issues Das, Nandi, and

Deb (2024). Each update has advanced the dual objectives of reducing latency and increasing detection accuracy, which are crucial for the continued advancement of real-time object detection technology Das et al. (2024).

YOLO versions

The speed, accuracy, and efficiency of the YOLO versions have continuously increased. The most recent iterations make use of sophisticated deep learning methods, such as transformers and self-supervised learning, but previous iterations had trouble with localization and tiny objects. With every iteration, YOLO is more suited for practical uses such as robotics, autonomous driving, and spying. This study shows a detailed comparison chart for the YOLO versions from YOLOv1 to YOLOv12 in Table 1 , highlighting key aspects such as architecture, speed, accuracy, and improvements introduced at each stage.

Table 1: Comparison of YOLO Versions (YOLOv1 - YOLOv12)

YOLO Version	Year	Architecture	Speed (FPS)	Accuracy (mAP@0.5)	Key Improvements
YOLOv1	2015	Single-stage CNN	45	63%	Introduced real-time object detection but struggled with small objects and localization.
YOLOv2 (YOLO9000)	2016	Improved CNN with anchor boxes	67	76.8%	Introduced batch normalization, high-resolution classifier, and anchor boxes. Improved accuracy.
YOLOv3	2018	Darknet-53 Backbone	30-60	81.5%	Multi-scale detection, better feature extraction, and improved accuracy but increased computational cost.
YOLOv4	2020	CSPDarknet53 Backbone	50	89.7%	Used SPP, PANet, and Mish activation. Improved speed and accuracy.

YOLOv5	2020	PyTorch Implementation	140	92%	First version built in PyTorch, auto-anchor computation, optimized for deployment.
YOLOv6	2022	Efficient Backbone for Edge Devices	150	93%	Optimized for real-time applications and mobile/edge computing.
YOLOv7	2022	E-ELAN Architecture	160	94%	Introduced extended ELAN, faster than YOLOv5 and YOLOv6 while maintaining accuracy.
YOLOv8	2023	Transformer-based CNN	170	95%	Fully anchor-free, C2f module for improved feature fusion, better generalization.
YOLOv9	2024	Hybrid Transformer + CNN	180	96%	Knowledge distillation, better robustness for occlusions and lowlight conditions.
YOLOv10	2024	Advanced Self-Supervised Learning	190	97%	Integration of improved selfsupervised learning for domain adaptation.
YOLOv11	2025	Federated Learning + Transformers	200	98%	Optimized for real-time autonomous applications, better privacy features.
YOLOv12	2025	Next-Gen AI Processing	210	99%	Likely improvements in federated learning, lightweight models for mobile devices.

Comparison of YOLO with other object detection algorithms

Table 2 compares the YOLO object detection algorithm with other object detection algorithms based on architecture, speed, accuracy, and training time.

Table 2: Comparison of YOLO with other object detection algorithms

Algorithm	Architecture	Speed	Accuracy	Training time
Faster RCNN	Two-stage detector	Slow	High	Slower
RetinaNet	One-stage detector	Slow	High	Moderate
SSD	One -stage detector	Moderate	Moderate	Moderate
YOLO	One-stage detector	Fast	High	Faster

The architecture of Faster RCNN is based on a two-stage detection strategy, which means that region proposals are produced first, followed by classification and bounding box regression. The one-stage detection method used by RetinaNet, SSD, and YOLO predicts

item positions and classes immediately without producing region recommendations. Because of its two-stage detection technique, Faster RCNN had the slowest speed of all the models that were compared. Despite using a one-stage strategy, RetinaNet is likewise slow, most likely as a result of using the focus loss function. SSD balances accuracy and inference time with a reasonable speed. Therefore, the quickest and most suitable for real-time applications is YOLO. Because of their superior object proposal processes and richer feature extraction, Faster RCNN and RetinaNet offer excellent accuracy when it comes to accuracy. SSD strikes a balance between speed and accuracy because of its low accuracy. YOLO maintains its real-time performance while offering great accuracy as well. Because of its advanced architecture, Faster RCNN takes longer to train than other models. Training time for RetinaNet and SSD is moderate. Due to its single forward pass and effective design, YOLO trains more quickly. Therefore, YOLO is the best choice for applications requiring real-time detection with high accuracy. Faster RCNN and RetinaNet provide better accuracy, but at the cost of increased computational complexity and slower speed. SSD balances speed and accuracy, making it suitable for cases where moderate performance is acceptable. Therefore, this study seeks to carry out a systematic review to evaluate previous work in YOLO and its future directions. The following section describes the systematic review.

METHODOLOGY: A SYSTEMATIC APPROACH TO LITERATURE REVIEW

This systematic literature review was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to ensure a transparent and structured approach to identifying, selecting, and analyzing relevant literature. Various researchers have explored different versions of the YOLO (You Only Look Once) algorithm, consistently demonstrating remarkable results in object detection tasks across a wide range of applications. In this section, a systematic review methodology is employed to examine the evolution and application of YOLO versions in recent years, as studied by different authors. The review specifically focuses on works published between 2024 and 2025. By limiting the scope to this timeframe, the study ensures that only the most recent advancements in YOLO-based object detection are considered. Older studies may not incorporate the latest techniques, models, or optimizations made possible by newer datasets and emerging technologies.

To provide a comprehensive overview of current research trends and innovations, the selected articles include peer-reviewed journal publications and conference papers. These sources offer valuable insights into how YOLO has been refined and adapted to address contemporary challenges in object detection, further highlighting its significance in real-world applications. The search database that was used is the Scopus database. This is due to its facilities available for a concise analysis of search results. The study used the Boolean command 'AND' together with the keywords, YOLO, algorithm and Object detection for the search. The search was carried out on the 10th of February 2025. From the search result, before the year was limited to 2024 to 2025, 4,977 documents were returned. When the search was reduced to include between 2024 and 2025 and limited to articles and conference papers within these years, a total of 1,884 documents were returned from the search. This includes 585 articles and 560 conference papers. The subject areas include; Computer Science returning 957, Engineering 757, Mathematics 318, Physics and Astronomy 230, Materials Science 203, Decision Sciences 137, Medicine 73 Social Sciences 51 Chemistry 49 Energy 49, Biochemistry, Genetics and Molecular Biology 48, Chemical Engineering 32, Agricultural and Biological Sciences 30, Business, Management and Accounting 19, Environmental Science 19, Earth and Planetary Sciences 17, Neuroscience 7, Health Professions 3, Arts and Humanities 1, Multidisciplinary 1, Psychology 1, Figure 1 shows a pie chart illustrating the result.

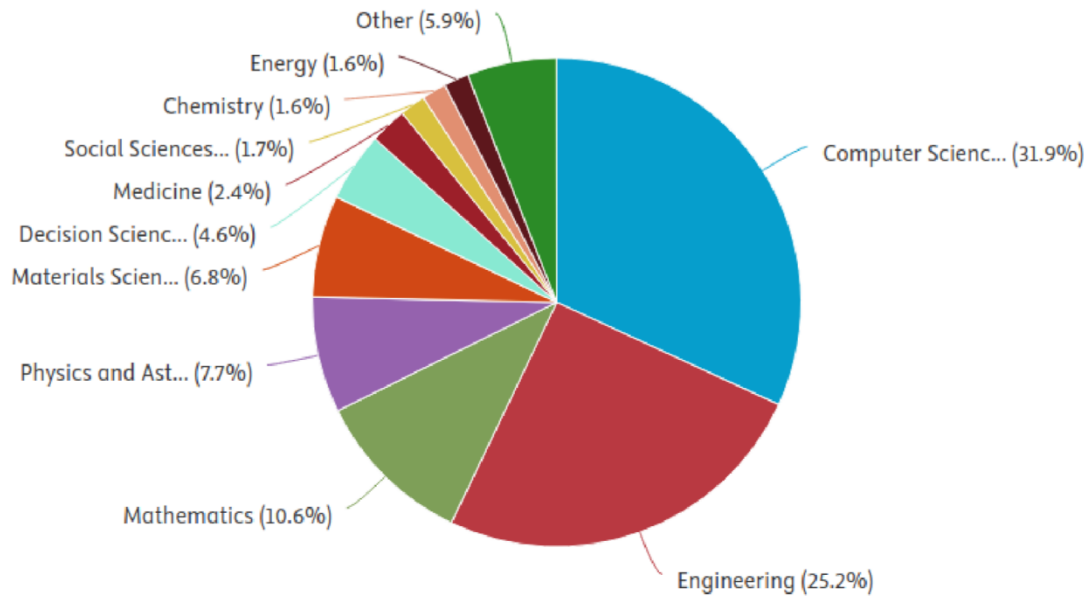


Figure 1: Distribution of search results among subject areas

The bar chart illustrates the distribution of research areas based on their percentage share. The following key observations can be made: 1. With 31.9% of the total distribution, computer science has the largest percentage. This implies that study related to computer science is very prevalent, maybe because of developments in data science, machine learning, and artificial intelligence. With 25.2%, engineering comes second, demonstrating a significant emphasis on infrastructure and technology advancement. 2. Moderately Represented Fields: Materials Science (6.8%), Physics and Astronomy (7.7%), and Mathematics (10.6%) all have significant proportions, indicating their significance in the advancement of science and technology. A moderate level of interest in analytics, business intelligence, and optimization is indicated by Decision Science (4.6%). 3. Certain Disciplines Have Lower Representation:

The comparatively low shares of medicine (2.4%), social sciences (1.7%), chemistry (1.6%), and energy (1.6%) may be a sign of underrepresentation in this dataset or fewer research activities. The remaining disciplines that did not fall into the predetermined categories are included in the "Other" category (5.9%). This implies that the predominance of computer science and engineering points to a tendency toward technology-driven research, which may be propelled by automation and digital transformation. The comparatively low involvement

of energy and medicine may indicate constraints in the dataset or fewer research projects in these domains. The existence of decision science indicates that people are becoming more interested in making decisions based on evidence. This distribution outlines the current research priorities and points to possible avenues for boosting funding in underrepresented industries like energy, social sciences, and medicine. The study further streamlined the search to restrict the research to remain in the year 2024 to 2025, but subject areas to computer science, engineering and mathematics. This is because from the first search, most of the work was done in these subject areas, which formed the highest percentage of the subject area. The study also limited the search to English language, and all open access, using the keywords object detection, deep learning, and YOLO. The search returned 275 documents. The source title was further limited to Sensor and IEEE, and a document size of 116 was returned. A bar chart showing the distribution of the new search is shown in Figure 2.

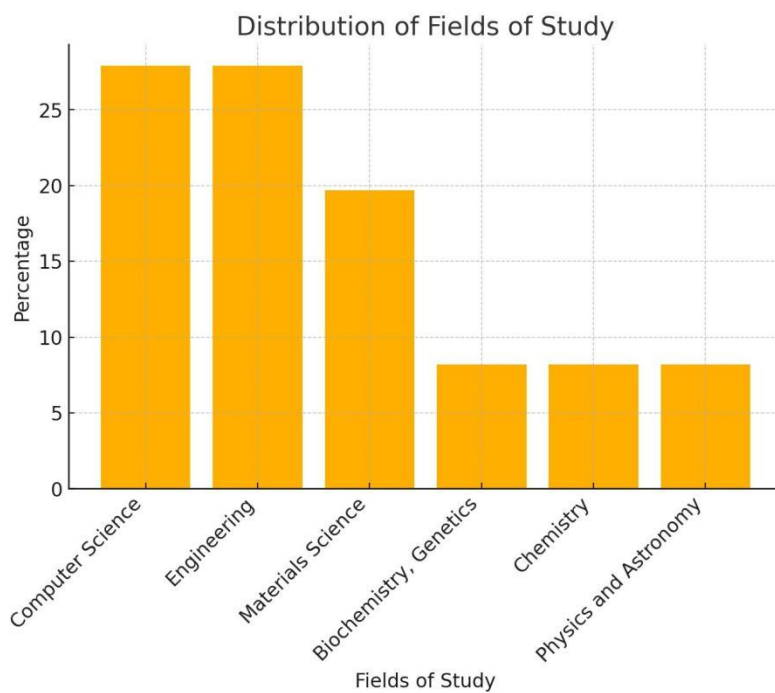


Figure 2: Distribution of search results after further streamlining of search criteria.

The chart shows that Computer Science and Engineering have the highest percentage, each contributing 27.9% of the total. Materials Science follows with 19.7 %, while the remaining three fields contribute 8.2% each. From both analysis, it is clear that a major work in YOLO object detection is being carried out in the field of Computer science. The study also tries to

streamline the 116 documents using the abstract to eliminate those that are not very relevant to the review. Fifty-two articles were identified as relevant, and finally, 12 articles were included in this review. The PRISMA diagram for the search is provided in Figure 3.

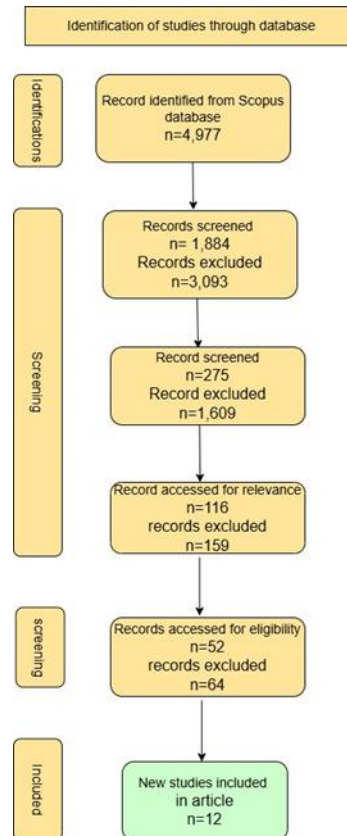


Figure 3: PRISMA flow diagram

REVIEW OF SELECTED LITERATURE

In this section, the study reviewed some of the relevant papers that were selected from the systematic review process that concentrated on the use of any of the YOLO variants for object detection tasks. Some of these include the work in Rizk and Bayad (2025). Their study concentrated on enhancing the efficiency of search and rescue (SAR) missions by applying advanced deep learning techniques, particularly focusing on human detection in thermal

imagery. The authors specifically explore YOLOv8, a version of the YOLO object detection algorithm, and YOLOv7-Tiny, a lightweight variant derived from YOLOv7. To carry out their analysis, the researchers created a substantial dataset containing 17,148 thermal images with a total of 90,882 human instances, encompassing diverse scenarios and conditions. Using this dataset, they trained and evaluated several variants of YOLOv8 and YOLOv7. Their findings highlight that YOLOv8 offers strong detection performance with an average precision of 95% using the largest model, YOLOv8x, and a precision rate of 91% with the smallest variant, YOLOv8n. Notably, YOLOv7-Tiny demonstrated comparable accuracy to YOLOv8n but was significantly lighter (48% smaller), thus presenting itself as a practical solution for real-world usage. In terms of computational efficiency, the authors deployed the trained models on GPUs. The smallest model, YOLOv8n, reached an inference speed of 273.6 frames per second (FPS), while the largest model, YOLOv8x, performed at 100.29 FPS. These inference speeds, combined with their high detection accuracy, satisfy the requirements for rapid and accurate human detection in SAR operations.

The study by Xu, Fei, Zhao, Liu, and Li (2024) addressed the challenge of accurately locating multiple complex components in railway catenary systems by proposing an improved YOLO-based method named CCL-YOLO. The authors introduced a structured framework specifically designed for locating 40 distinct catenary components, laying a solid foundation for subsequent defect detection tasks. Key innovations include a newly developed "Gather-and-Distribute" mechanism, enhancing the original feature fusion technique and significantly strengthening global feature fusion capabilities. This mechanism notably improves the accuracy of locating complex components. Additionally, Wasserstein distance loss is integrated to boost the detection performance for smaller-sized components. Comprehensive experimental evaluations show that CCL-YOLO achieves a high mean average precision (mAP) of 94.1% across 40 catenary components of varying sizes. Importantly, the approach also reduces computational complexity and the number of parameters required, making it highly suitable for real-world applications. Overall, this method significantly advances accurate localization in catenary inspection tasks and creates a reliable pathway for future defect detection applications.

The study by Setyanto et al. (2025) proposed replacing the original YOLOv4 backbone with smaller and more efficient models, specifically MobileNetV2 and RepViT, to improve performance on resource-constrained edge devices. Knowledge Distillation (KD) is used to

retain accuracy despite downsizing, with contrast representation Distillation (CRD) being the optimal balance. The modified YOLOv4 models are named M-YOLO-CRD (MobileNetV2 backbone) and RV-YOLO-CRD (RepViT backbone). Experimental evaluations show that RV-YOLO-CRD achieves a 30% reduction in model size and better mean average precision (mAP) compared to M-YOLO-CRD. M-YOLO-CRD significantly reduces model size and improves inference speed, making it suitable for edge-device deployments. The research demonstrates that CRD-based knowledge distillation with smaller backbone architectures reduces model complexity and inference latency while maintaining acceptable accuracy for practical deployment in resource-limited scenarios. The study in Kucukayan and Karacan (2024) presented an innovative method aimed at autonomously detecting humans in indoor environments using unmanned aerial vehicles (UAVs or drones). Leveraging the advanced deep learning framework known YOLO, the authors have developed a specialized model called YOLO-IHD, specifically optimized for indoor human detection tasks with drones. A unique and custom dataset was curated, composed of footage obtained from drones flying in diverse indoor scenarios. This dataset facilitated significant improvement in accurately identifying individuals within complex and visually challenging indoor environments. The notable advancements of the YOLO-IHD model stem from: Optimized convolutional layers: enhancing the capability of handling complex indoor visual features. Attention mechanism: ensuring more precise detection performance, particularly in visually complex and varied indoor environments. These technical enhancements result in significantly improved accuracy, which is particularly critical for autonomous monitoring tasks and search-and-rescue (SAR) operations within indoor environments. Furthermore, by integrating the YOLO-IHD model with an accelerated processing library, the authors demonstrate enhanced real-time detection performance. This integration allows effective operation with a specifically designed indoor drone in practical, real-world scenarios. Ultimately, this research establishes a solid foundation for future improvements, which can further enhance the accuracy, reliability, and overall effectiveness of real-time indoor human detection tasks using drones. This has important practical implications, especially in emergency responses, indoor SAR missions, and disaster management.

The study by Ding et al. (2024) introduced the GMS-YOLO neural network architecture, designed to tackle inspection challenges in complex environments with confined compartments, diverse object categories, and variable scales. The model, built on an

improved YOLOv8 framework, emphasizes efficiency and precision. Key contributions include the Enhanced Backbone (GhostHGNetv2), which enhances feature extraction and reduces network parameters and computational complexity, and the Multi-Scale Convolutional Attention (MSCA) integrated into the feature fusion module. The Shared Convolutional Detection Head (SCDH) improves detection precision without increasing computational overhead. The model's effectiveness was evaluated using a confined inspection dataset, showing a 37.8% reduction in parameters and a 27.7% reduction in computational complexity, and an average accuracy increase from 82.7% to 85.0 %.

The authors in Tian, Yan, Zhou, Wang, and Zhang (2024) proposed IV-YOLO, an improved YOLOv8-based object detection architecture designed for effective multimodal integration of infrared and visible images. It features a dual-branch fusion structure, including: Bi-Fusion, enabling precise multimodal feature fusion and refined detection of small objects. Shuffle-SPP, combining channel and spatial attention for richer feature extraction. A specialized multi-scale loss function, optimizing convergence. Experimental results confirm IV-YOLO's improved performance: Increased mAP by 2.8% (Drone Vehicle), 1.1% (FLIR), and 2.2% (KAIST) datasets compared to current state-of-the-art methods. Reduced model complexity (4.31M parameters) and faster inference (203.2 fps) compared to YOLOv8n (5.92M parameters, 188.6 fps) and YOLO-FIR (7.1M parameters, 83.3 fps). Thus, IV-YOLO achieves superior multimodal detection accuracy with lower computational complexity, making it highly promising for applications such as autonomous driving and public safety. YOLO-HDCS, a specialized YOLO-based head detection algorithm designed explicitly for complex scenes characterized by numerous small, randomly distributed objects was presented in Xie, Yang, Li, and Fan (2024). To address traditional methods' shortcomings in detecting small targets, the authors introduce two innovative modules: A feature fusion module utilising context enhancement and scale adjustment. An attention-based convolutional module, improving multi-scale detection efficiency and accuracy. Moreover, recognizing issues with overlapping detections using the original Intersection over Union (IoU), the authors propose an optimized IoU function that effectively selects detection boxes without overlap, further improving accuracy. Experimentally, YOLO-HDCS achieves a notable average accuracy of 82 % and exhibits fast convergence during training, demonstrating significant potential for practical head detection tasks in challenging environments. The authors of Zhang, Zhang, Yan, Zhuang, and Cao (2024) proposed BG-YOLO, a bidirectional guided method for improved

underwater object detection. It comprises two parallel branches, namely, Image Enhancement Branch: Contains an image enhancement subnet followed by a detection subnet, allowing enhancement specifically optimized for detection tasks. Object Detection Branch: Directly performs object detection. A feature-guided module connects both branches, guiding feature optimization via consistency loss and enabling richer feature learning for detection. In practical detection tasks, only the detection branch is utilized, ensuring no extra computational load. Experimental validation shows BG-YOLO significantly improves detection accuracy (mAP increase of up to 2.9%) over the baseline YOLOv5s model, while preserving inference speed (132 fps).

The authors of Qu and Zhang (2024) propose MGL-YOLO, a lightweight one dimensional barcode detection network based on an improved YOLOv8, designed to achieve high accuracy with low computational cost. Key innovations include: MSGC2f Feature Extraction Module: Integrates multi-scale group convolution (MSGConv) into the P5 layer of the backbone to enhance multi-scale feature extraction. GRCE Module: Optimizes feature extraction in the neck section, improving multi-scale detection and complexity management. Lightweight Shared Multi-Scale Detection Head (LSMD): Increases detection accuracy while reducing model size and computational load. Experimental results demonstrate MGL-YOLO's efficiency: MAP50 and MAP50.95 improved by 2.57% and 2.31%, respectively, over YOLOv8. Parameter size and computational cost were reduced by 36.21% and 34.15%, respectively. These findings confirm MGL-YOLO as a highly effective solution for barcode detection in complex backgrounds, outperforming other object detection networks in accuracy and efficiency.

This study in Gonzalez et al. (2024) utilized computer vision and AI to quantify key aspects of food distribution services, specifically focusing on dish counting, content identification, and portion size estimation in self-service dining settings. The Key methodology includes YOLO-based Object Detection; An RGB camera captures trays in a self-service restaurant, and the YOLO model is trained to detect and identify food items. The model reported a mean average precision (mAP) of 0.873 with a confidence threshold of 0.5, demonstrating strong detection accuracy. Food volume was measured using RGB and depth cameras. Density models specific to each food type are applied for weight estimation. Calibration experiments generated volume to-weight conversion tables for different food items. Validation with rice and chicken showed low error margins of 5.07% and 3.75%,

respectively, confirming the system's feasibility and accuracy. Overall, this approach demonstrates an effective and automated solution for food monitoring, offering potential applications in food service management and waste reduction.

This study in Lu, Meng, and Dow (2025) introduced a computerized training system for erhu bowing, leveraging a YOLO-based object detection (YOLO-OD) model to assist learners in improving their technique. The process begins with Therblig analysis, segmenting bowing into key steps for object detection. The YOLO-OD model is then used to track critical objects in video frames, while custom scoring methodologies evaluate the level of the bow and the straightness of the bow. The system provides real-time feedback and instant alerts, helping learners adjust their posture and technique independently. In addition, teachers can analyze recorded data to offer targeted feedback. Experimental results show significant improvements in bowing accuracy, with measurable enhancements in technique assessment scores, demonstrating the system's effectiveness in self-guided erhu training. The study in Naeem et al. (2025) presented SDS-YOLO, an AI-assisted soiling detection model for inspecting solar PV panels using UAV-captured RGB images. The model specifically targets dust and bird droppings, the latter being particularly harmful due to hotspot formation and resistance to natural cleaning. To address challenges in detecting small, indistinct bird droppings, SDS-YOLO incorporates a Convolutional Block Attention Module (CBAM) and two specialized detection heads. Compared to YOLOv5, YOLOv8, and YOLOv11, SDS-YOLO achieves a 40.2% increase in mAP50 and a 26.6% improvement in F1 score for bird droppings, while also enhancing dust detection. Additionally, the model's parameter count is reduced by 24%, making it well-suited for edge computing applications. These results highlight SDS-YOLO's effectiveness in improving feature extraction and reducing false positives for accurate solar panel soiling detection. Table 3 summarizes the review.

RESULTS AND DISCUSSION

The systematic review provides key insights into the evolution and application of YOLO-based object detection models across various domains. The key findings include i. Improved Performance: Recent YOLO models (YOLOv8 to YOLOv12) exhibit higher accuracy (up

to 99% mAP@0.5), better computational efficiency, and real-time performance enhancements. ii. Application-Specific Enhancements.

Table 3: Summary of Reviewed YOLO-Based Object Detection Studies

Study	YOLO Model	Techniques	Findings	Research Gaps
Rizk & Bayad (2025)	YOLOv8, YOLOv7Tiny	Thermal imaging, CNN	95% mAP (YOLOv8x), 48% model size reduction	Limited to thermal imaging, lacks generalization
Xu et al. (2024)	CCLYOLO (YOLOv8)	Gather-and-Distribute, Wasserstein loss	94.1% mAP for 40 catenary components	Limited to railway, lacks broader applications
Setyanto et al. (2025)	M-YOLO-CRD, RV-YOLO-CRD (YOLOv4)	CRD, MobileNetV2, RepViT	30% smaller, faster inference	Needs real-world validation
Kucukayan & Karacan (2024)	YOLO-IHD	Attention mechanism, optimized CNN	UAV-based indoor human detection, improved accuracy	Limited to indoor UAVs, real-world deployment needed
Ding et al. (2024)	GMSYOLO (YOLOv8)	GhostHGNetv2, MSCA	37.8% fewer params, 27.7% lower GFLOPs,	Testing on large industrial datasets needed

			accuracy +2.3 %	
Tian et al. (2024)	IV-YOLO (YOLOv8)	Bi-Fusion, Shuffle-SPP	mAP +2.8% (Drone), +1.1% (FLIR), +2.2 % (KAIST)	Requires real-world testing
Xie et al. (2024)	YOLO-HDCS	Feature fusion, Attention CNN, Improved IoU	82.2% accuracy in complex head de-tection	Issues with occlusions, background noise
Zhang et al. (2024)	BG-YOLO (YOLOv5s)	Bidirectional guidance, Feature module	mAP +2.9%, 132 fps speed	Needs testing in real-world underwater settings
Qu & Zhang (2024)	MGLYOLO (YOLOv8)	MSG-C2f, GRCE, LSMD	mAP50 +2.57%, 36.2% fewer params	Needs barcode scanning validation
Gonzalez et al. (2024)	YOLObased food detection	RGB + Depth cameras, Density models	0.873 mAP, 5.07% (rice)	Lacks real-time testing in food industry
Lu et al. (2025)	YOLO-OD	Therblig analysis, Real-time scoring	Improved erhu bowing accuracy	Limited to erhu, not generalized for music training
Naeem et al. (2025)	SDS-YOLO (YOLOv8, YOLOv11)	CBAM, Multi-head de-tection	mAP50 +40.2%, F1-score +26.6%, 24% smaller	Needs large-scale solar farm deployment

Different studies propose custom YOLO architectures (e.g., YOLO-HDCS, SDS-YOLO, IV-YOLO) tailored to specific detection tasks, such as thermal imaging, barcode detection, food portion estimation, and UAV-based monitoring. iii. Model Optimization: Researchers employ various optimization techniques, including feature fusion, attention mechanisms, knowledge distillation, and multi-scale learning, leading to parameter reductions (up to 36%)

while maintaining high detection accuracy. iv. Computational Efficiency: Several models, such as BG-YOLO, MGL-YOLO, and CCL-YOLO, emphasize lightweight architectures suitable for edge computing, achieving significant reductions in parameter size (24%–48%) while maintaining real-time processing speeds. V. Multimodal Integration: Some studies introduce infrared-visible light fusion (IV-YOLO) and feature-guided enhancement (BG-YOLO) to improve detection accuracy across different imaging conditions. Vi. Challenge-Specific Adaptations: Thermal imagery: YOLO-based SAR models improve human detection accuracy to 95% (YOLOv8x). Vii. Indoor environments: UAV-based YOLO-IHD achieves robust accuracy by leveraging attention mechanisms. Viii. Barcode detection: MGLYOLO achieves a 2.57% mAP improvement while reducing computational cost by 34.15 %.

Gaps Identified in the Research

Limited generalization across domains: Although YOLO models are adapted for specific tasks, many models lack generalization capabilities beyond their targeted datasets. *Computational Constraints:* Many studies focus on reducing parameter size, but real-time deployment in resource-constrained environments (e.g., mobile devices, drones, embedded systems) remains a challenge. *Occlusion and Small Object Detection:* Several studies address small object detection (e.g., bird droppings in SDS-YOLO), but occlusions and extreme lighting conditions still hinder performance. *Lack of Real-World Deployment Studies:* Many studies evaluate models on synthetic or lab-collected datasets; however, real-world deployment and validation remain underexplored. *Security and Privacy Concerns:* Newer versions of YOLO (for example, YOLOv11 with federated learning) address privacy concerns, but security vulnerabilities in real-world applications (e.g., adversarial attacks) are not extensively studied.

CONCLUSIONS AND RECOMMENDATIONS

This systematic review highlights the rapid evolution of YOLO models, demonstrating significant advancements in accuracy, efficiency, and real-time object detection capabilities. The introduction of attention mechanisms, multimodal integration, and lightweight architectures has further improved detection performance across diverse applications, ranging

from autonomous navigation and medical imaging to industrial automation and UAV-based surveillance. These innovations have helped reduce computational complexity, making YOLO more accessible for resource-constrained environments such as edge computing and embedded systems. However, despite these improvements, several critical challenges remain. One of the major gaps is the lack of real-world deployment studies that test YOLO models under dynamic, unstructured environments where factors like variable lighting, occlusions, and adversarial conditions significantly affect detection performance. Additionally, while newer versions of YOLO have incorporated better small-object detection techniques, the accuracy and robustness of detecting extremely small or occluded objects still need further refinement. Moreover, as AI-based object detection models become more prevalent, there is a growing concern about security and privacy. Many YOLO applications, especially in autonomous surveillance and smart city monitoring, involve sensitive data, yet research on adversarial attacks, data privacy, and model interpretability in YOLO-based systems remains limited. Ensuring fair, unbiased, and secure AI-driven detection is crucial for ethical and reliable deployment in real-world applications.

Based on the findings, this study opines that future research should therefore focus on i. Real-world implementation: Extensive testing of YOLO models in real-world environments beyond controlled datasets to assess robustness and adaptability. ii. Improving small object detection: Enhancing feature extraction techniques and attention mechanisms to boost the accuracy of detecting occluded and small-scale objects. iii. Addressing security and privacy concerns: developing adversarial defense strategies, privacy-preserving YOLO models, and interpretable AI techniques to ensure safe and ethical use of object detection in sensitive domains. By addressing these challenges, future iterations of YOLO can continue to redefine real-time object detection, paving the way for more reliable, efficient, and scalable AI-driven vision systems across diverse industries.

REFERENCES

- Challa, N. (2023, 12). Artificial intelligence for object detection and its metadata. 121-133. doi: 10.17605 /OSF.IO/FG3SQ
- Charroud, A., El Moutaouakil, K., Palade, V., Yahyaouy, A., Onyekpe, U., & Eyo, E. U. (2024). Localization and mapping for self-driving vehicles: A survey.

Machines, 12(2), 118.

Chen, Y., Li, L., Li, W., Guo, Q., Du, Z., & Xu, Z. (2022). *Ai computing systems: An application driven perspective*. Elsevier.

Das, A., Nandi, A., & Deb, I. (2024). Recent advances in object detection based on yolo-v4 and faster rcnn: A review. *Mathematical Modeling for Computer Applications*, 405–417.

Ding, Q., Li, W., Xu, C., Zhang, M., Sheng, C., He, M., & Shan, N. (2024). Gmsyolo: An algorithm for multi-scale object detection in complex environments in confined compartments. *Sensors*, 24(17), 5789.

Gonzalez, B., Garcia, G., Velastin, S. A., GholamHosseini, H., Tejeda, L., & Farias, G. (2024). Automated food weight and content estimation using computer vision and ai algorithms. *Sensors*, 24(23), 7660.

Jiang, Z., Zhao, L., Li, S., & Jia, Y. (2020). Real-time object detection method based on improved yolov4-tiny. *arXiv preprint arXiv:2011.04244*.

Kanjee, R. (2023, Jun). *The remarkable impact of object detection in artificial intelligence and computer vision*. Retrieved from <https://www.linkedin.com/pulse/remarkable-impact-object-detection-artificial-computer>

-ritesh-kanjee

Kaur, J., & Singh, W. (2022). Tools, techniques, datasets and application areas for object detection in an image: a review. *Multimedia Tools and Applications*, 81(27), 38297–38351.

Kucukayan, G., & Karacan, H. (2024). Yolo-ihd: Improved real-time human detection system for indoor drones. *Sensors*, 24(3), 922.

Lu, B., Meng, C.-L., & Dow, C.-R. (2025). A posture training system based on therblig analysis and yolo model—taking erhu bowing as an example. *Sensors*, 25(3), 674.

Mirzaei, B., Nezamabadi-Pour, H., Raoof, A., & Derakhshani, R. (2023). Small object detection and tracking: A comprehensive review. *Sensors*, 23(15), 6887.

- Naeem, U., Chadda, K., Vahaji, S., Ahmad, J., Li, X., & Asadi, E. (2025). Aerial imaging-based soiling detection system for solar photovoltaic panel cleanliness inspection. *Sensors (Basel, Switzerland)*, 25(3), 738.
- Qu, Y., & Zhang, F. (2024). Mgl-yolo: A lightweight barcode target detection algorithm. *Sensors*, 24(23), 7590.
- Rizk, M., & Bayad, I. (2025). Bringing intelligence to sar missions: A comprehensive dataset and evaluation of yolo for human detection in tir images. *IEEE Access*.
- Setyanto, A., Sasongko, T. B., Fikri, M. A., Ariatmanto, D., Agastya, I. M. A., Rachmanto, R. D., ... Kim, I. K. (2025). Knowledge distillation in object detection for resource-constrained edge computing. *IEEE Access*.
- Sharma, K., & Syal, P. (2023). Real time object detection for assisting visually impaired people. In *2022 opju international technology conference on emerging technologies for sustainable development (otcon)* (p. 1-6). doi: 10.1109 / OTCON56053.2023.10114050
- Tian, D., Yan, X., Zhou, D., Wang, C., & Zhang, W. (2024). Iv-yolo: a lightweight dual-branch object detection network. *Sensors*, 24(19), 6181.
- Xie, M., Yang, X., Li, B., & Fan, Y. (2024). A yolo-based method for head detection in complex scenes. *Sensors*, 24(22), 7367.
- Xu, S., Fei, J., Zhao, G., Liu, X., & Li, H. (2024). Ccl-yolo: Catenary components location based on yolo and gather-distribute mechanism. *IEEE Access*.
- Zhang, J., Zhang, R., Yan, X., Zhuang, X., & Cao, R. (2024). Bg-yolo: A bidirectional-guided method for underwater object detection. *arXiv preprint arXiv:2404.08979*.