INTERNATIONAL JOURNAL OF ADVANCED COMPUTING AND INFORMATICS
VOLUME 2, ISSUE 1, 2026, pp. 53 – 61
ISSN: 3089-7483, DOI: https://doi.org/10.71129/ijaci.v2i1.pp53-61

REVIEW ARTICLE

YOLOv8 for Object Detection: A Comprehensive Review of Advances, Techniques, and Applications

Vivi Afifah [a*], **Surni Erniwati** [b]

[a]Department of Digital Business, Universitas Siber Indonesia, Jakarta, Indonesia
[b]Faculty of Information and Communication Technology, Universitas Teknologi Mataram, Mataram, Indonesia

ABSTRACT - This paper presents a review focusing on the most current advancements in object detection techniques using YOLOv8 and their applications across a range of fields, such as surveillance systems, autonomous driving, smart agriculture, industrial quality control, and medical image analysis. YOLOv8, launched by Ultralytics in early 2023, represents the newest evolution of the You Only Look Once framework., YOLOv8 demonstrates notable improvements in both detection accuracy and processing speed compared to earlier versions. This review explores key updates in YOLOv8's architecture, including enhanced backbone designs, efficient training processes, and optimized loss functions, alongside the incorporation of attention mechanisms and compact model structures. These findings from 40+ selected studies confirm YOLOv8's potential as a reliable and efficient solution for object detection tasks in both academic research and real-world applications. This study serves as a comprehensive reference for researchers and practitioners seeking to optimize aiming to enhance YOLOv8-based models across diverse practical applications.

Keywords: YOLOv8, Object Detection, Data Augmentation, Deep Learning, Computer Vision

1. Introduction

Object detection serves as a crucial pillar within the domain of computer vision, a highly active field of research with extensive applications spanning diverse sectors such as agriculture, remote sensing, intelligent traffic systems, and critical autonomous driving technologies, where comprehending the vehicle's environment through sensors is paramount [1]. Object detection is an essential and challenging problem in the field of computer vision that focuses on identifying and localizing instances of semantic objects of a certain class (such as people, cars, animals, or traffic signs) through digital image and video formats [1]. It plays a pivotal role in a wide range of practical applications, including intelligent surveillance systems, autonomous driving, smart farming, industrial automation, and medical diagnostics [2]. With the evolution of deep learning techniques, particularly convolutional neural networks (CNNs), the performance of object detection models has seen significant improvements in terms of both accuracy and speed [3]. However, one of the most persistent challenges in this domain is the detection of small objects, which are typically characterized by limited pixel coverage, low resolution, and high susceptibility to occlusion and background clutter. These issues are especially prominent in real-world scenarios such as remote sensing, medical imaging, and traffic surveillance, where critical small-scale targets like tumors, distant pedestrians, or traffic signs must be accurately detected [4]. Moreover, as larger objects tend to dominate feature representations in convolutional networks, small object features are often suppressed during multi-scale aggregation processes, resulting in lower recall and precision rates. Consequently, small object detection has become a focal research area that demands more adaptive architectures, improved feature calibration, and multi-scale learning strategies [5].

Deep learning has played a transformative role in advancing object detection, particularly through the development of models like the You Only Look Once (YOLO) family. These models harness convolutional neural networks (CNNs) to directly learn spatial hierarchies and semantic features from raw image data, enabling the simultaneous localization and classification of objects with remarkable speed and accuracy [6]. YOLO reframes object detection as a single regression problem, allowing it to predict bounding boxes and class probabilities from full images in one unified step [7]. Since the original YOLO release, successive versions YOLOv2 through YOLOv7 have integrated deeper backbones, enhanced feature pyramid structures, and optimized training strategies that embody the growing sophistication of deep learning architectures [8]. The latest iteration, YOLOv8, released by Ultralytics in 2023, represents a major leap forward by adopting a more powerful backbone and head design, anchor-free detection mechanisms, and improved loss functions, all of which reflect the maturation of deep learning methodologies in the field [9][10]. Furthermore, YOLOv8 integrates flexible design principles that improve generalization across datasets and application domains, making it highly adaptable for real-time deployment [11].

Recent studies have demonstrated the role of deep learning innovations in enhancing YOLOv8's capabilities, particularly in small object detection. For example, introduced SO-YOLOv8, which integrates attention mechanisms,

Received 15 January 2025, Accepted 06 July 2025
Available online 28 July 2025

*** Corresponding Author**
E-mail: vivi.afifah@cyber-univ.ac.id (Vivi Afifah)

multi-scale training, and data augmentation to address the inherent challenges of detecting small and occluded objects. Similarly, applied YOLOv8 in combination with the DeepSORT tracking algorithm to develop an automated, real-time monitoring system for coral reefs, achieving high precision and robust multi-object tracking in underwater environments. These examples highlight that the ongoing evolution of YOLOv8 is inseparable from advancements in deep learning, especially in addressing complex detection scenarios involving size, motion, or occlusion.

Despite its rapid uptake, there is currently no comprehensive study that systematically synthesizes the architecture, benchmark results, and practical deployments of YOLOv8 across various domains [12]. Existing research often isolates specific applications or provides limited comparative insights, making it challenging for practitioners to access an integrative understanding of the model's strengths, weaknesses, and optimization pathways [13]. Additionally, with the growing volume of research proposing modifications to YOLOv8 for specific detection tasks, the need for a structured, up-to-date synthesis is becoming increasingly critical.

To address this gap, this study presents a Systematic Literature Review (SLR) of YOLOv8-based object detection models, with a particular emphasis on the challenges and advancements in small object detection. The objectives of this review are fourfold. First, it aims to examine the architectural innovations introduced in YOLOv8 and evaluate how these improvements influence detection performance. Second, the study assesses YOLOv8's effectiveness across various benchmark datasets using standard evaluation metrics, including precision, recall, F1-score, and mean Average Precision (mAP). Third, it summarizes the modifications and enhancements proposed by other researchers to optimize YOLOv8, particularly for scenarios involving small or occluded objects. Finally, this review analyzes the practical implementation of YOLOv8 across diverse application domains, such as UAV aerial imaging, smart agriculture, road infrastructure monitoring, forest fire detection, and medical imaging, thereby offering a comprehensive view of its real-world applicability and adaptability [14].

The contributions of this review are as follows:

a. This study provides a systematic and up-to-date synthesis of recent research utilizing YOLOv8 for object detection, with a particular focus on small object scenarios.

b. It categorizes and compares architectural and training modifications proposed across studies, identifying key strategies for improving detection accuracy in challenging environments.

c. It presents a critical analysis of YOLOv8's performance across various datasets and domains, offering benchmarks that can inform future model design.

d. It identifies current research gaps and provides practical recommendations for advancing small object detection using deep learning, particularly within the YOLOv8 framework.

By offering a structured and comprehensive overview of YOLOv8's state-of-the-art applications and research directions, this review aims to support developers, researchers, and practitioners in designing and deploying more accurate and efficient object detection systems.

2. Related Works

YOLOv8, launched by Ultralytics in early 2023, represents the newest evolution of the You Only Look Once (YOLO) framework, marking a significant advancement in object detection. It demonstrates notable improvements in both detection accuracy and processing speed compared to earlier versions, making it a reliable and efficient solution for various object detection tasks. This iteration adopts a more powerful backbone and head design, incorporates anchor-free detection mechanisms, and utilizes improved loss functions, reflecting the maturation of deep learning methodologies in the field. YOLOv8's flexible design principles also enhance its generalization across datasets and application domains, making it highly adaptable for real-time deployment [2].

The architecture of YOLOv8 comprises four main components: Input, Backbone, Neck, and Head. The Input section handles image preprocessing tasks like data augmentation, adaptive anchor calculation, and adaptive grayscale padding [4]. The Backbone and Neck modules form the core of the network; YOLOv8 employs a modified CSPDarknet53 as its backbone, featuring the C2f module that improves upon YOLOv5's C3 module by combining high-level features with contextual information to enhance detection accuracy and gradient flow. Additionally, the Spatial Pyramid Pooling Fusion (SPPF) module integrates multiple max-pooling layers to merge feature maps, enabling the model to handle objects of various scales effectively. The Neck module fuses features extracted by the backbone across different resolutions, utilising a combination of Path Aggregation Network (PAN) and Feature Pyramid Network (FPN) to enhance feature integration and multi-scale representation [5]. The Head of YOLOv8 adopts a decoupled structure, separating object classification from localization tasks, each with its own loss function, and integrates the Task-Aligned Assigner for dynamic sample assignment, improving accuracy and robustness. Being an anchor-free model, it eliminates the need for predefined anchor boxes, simplifying training on diverse datasets, and is the first YOLO version to incorporate Soft-NMS, which reduces false positives and negatives by preserving overlapping target information [3].

YOLOv8 and its enhanced variants have found widespread applications across diverse fields. In autonomous driving and traffic systems, it has been used for object detection in adverse weather, brake light status recognition, pedestrian identification, and traffic sign detection. For UAV-based aerial imaging, it excels at detecting small-scale objects like other drones. In agriculture, applications include pest identification, disease detection, fruit recognition, and weed detection, contributing to precision farming [6]. Infrastructure inspection benefits from YOLOv8 in detecting road surface defects and structural cracks in dams and bridges. In the medical field, it aids in paediatric wrist fracture detection, elbow osteochondritis dissecans (OCD) identification, and tumour localization. Within manufacturing and industrial contexts, it's used for detecting defects in solar cells, flaws in 3D printing, bullet hole identification, and foreign objects in packaging. Lastly, in security and surveillance, YOLOv8 monitors smoking behaviour and performs multi-object tracking, such as fish movement analysis [7].

Despite its advancements, YOLOv8 faces several limitations. It often struggles with occlusion, scale variance, and abrupt changes in illumination, which can, with poor image resolution and varying annotation formats hindering performance evaluation [8]. degrade detection accuracy in complex environments. The quality and consistency of datasets also pose challenges. The inherent downsampling process can lead to the loss of crucial spatial information, particularly affecting small object detection. Distinguishing between visually similar objects also remains a challenge. Moreover, its relatively high computational requirements can make deployment difficult on edge devices or resource-constrained environments. Generalization across different object types, lighting conditions, and backgrounds can be limited, impacting its adaptability in diverse settings [9].

To address these limitations, researchers have proposed numerous modifications. For small object and occlusion detection, variants like YOLOv8-SnakeVision introduce Dynamic Snake Convolution and Context Aggregation Attention Mechanisms. DGW-YOLOv8 enhances the backbone with Deformable ConvNets v2 and global attention to reduce feature loss. UAV-YOLOv8 includes additional detection scales, and YOLOv8-MeY adds a larger detection layer. YOLO-SE incorporates a dedicated prediction head for small objects and a transformer-based detection head [10].

Data augmentation techniques like Mosaic and MixUp, random brightness/contrast jittering, rotation, scale augmentations, CutMix, and elastic deformation have been widely adopted to bolster YOLOv8's robustness against occlusion, lighting variability, and object scale [11]. However, these traditional methods still have limitations, such as performance degradation under heavy occlusion or sensitivity to extreme illumination shifts [13].

Further research directions include developing lightweight GAN architectures to reduce computational demands and exploring self-supervised and weakly supervised training to alleviate the need for paired data. Adaptive training strategies that dynamically adjust to varying environmental conditions could bolster model robustness, and investigating transfer learning and domain adaptation techniques will be crucial for bridging the gap between synthetic and real data distributions [16]. For broader adoption and meaningful progress, establishing standardized evaluation protocols and unified benchmarking frameworks is imperative, as the current lack of consistency hinders fair comparisons across different GAN-enhanced detection models [17]. This standardization, particularly for evaluating fidelity, impact on accuracy, inference latency, and resource efficiency across various hardware configurations, will guide future development efforts and ensure the realisation of reliable, scalable, and high-performance GAN-powered object detection systems [18].

3. Methods

Deep learning has emerged as a transformative approach in the domain of object detection, garnering substantial attention from the research community. This growing interest has led to a proliferation of survey studies that examine various methodological, architectural, and application-oriented aspects of object detection. While these reviews have significantly advanced our understanding of object detection techniques, they often generalize across multiple models and fail to provide a focused analysis of specific architectures.In particular, the distinctive role and technical advancements introduced by YOLOv8a recent evolution in the YOLO family have not been comprehensively addressed in existing literature. To bridge this gap, the present section conducts a focused literature review, aiming to position this study within the broader research context and to underscore its unique contributions. A comparative summary of prior surveys and their scope is presented in **Table 1** to further clarify the novelty and relevance of this work.

Table 1. Summary of relevant surveys

Reference	Year	Primary focus	YOLOv8 Coverage	Main Contribution
[1]	2023	UAV Small-Object Detection	Extensive	Lightweight head and UAV-specific augmentations for small targets
[2]	2023	Multi-object pedestrian tracking	Moderate	Integration with OC-SORT for improved trajectory stability
[3]	2025	Real-time video surveillance	Moderate	Added transformer-attention module for cluttered/low-light scenes
[4]	2023	Classroom Behavior Recognition	Limited	Custom anchors and head fine-tuning for pose-based detection
[5]	2023	Leaf-Infection Region Segmentation	Moderate	Polygon-box refinement layer for high-precision masks

Several recent surveys have concentrated on advancing one-stage detection frameworks, particularly YOLOv8, by targeting domain-specific challenges and architectural refinements. The scheme by Wang et al. [1] and Xiao and Feng [2] are exemplified this trend by augmenting the core YOLOv8 architecture for improved small-object detection and multi-object pedestrian tracking, respectively. Wang et al. applied a lightweight detection head and UAV-specific data augmentations to boost mAP on aerial imagery, demonstrating extensive coverage of YOLOv8's neck and head modules. In parallel, Xiao and Feng integrated YOLOv8's detection outputs with the OC-SORT tracker an approach characterized by moderate YOLOv8 coverage that yielded more stable real-time trajectories under crowded conditions. Nimma et al. [3] further extended YOLOv8's capabilities into real-time video surveillance by embedding a transformer-attention module into the feature-fusion pipeline, striking a balance between detection accuracy and inference speed in low-light and cluttered scenarios.

Domain specific applications have also driven targeted YOLOv8 adaptations. Chen et al. [4] tackled classroom behavior recognition by fine tuning YOLOv8's anchors and detection head to recognize student poses, a use case with limited coverage of only the head module but resulting in an 8% accuracy uplift at over 25 FPS. Meanwhile, Zhu et al. [5] addressed segmentation of pest infected leaf regions by layering a polygon-box refinement network atop YOLOv8's segmentation branch, achieving mask precision up to 92%. These studies highlight how moderate to limited coverage of YOLOv8's backbone, neck, or head modules can nonetheless deliver significant gains when customized for particular environments ranging from indoor behavior analysis to precision agriculture.

Despite this progress, there remains a notable gap in comprehensive evaluations that compare the trade-offs among different levels of YOLOv8 architectural coverage backbone pruning, neck fusion strategies, attention modules and their impacts on both detection accuracy and computational efficiency across diverse application domains. Existing surveys tend to focus either on a single module's enhancement or on a narrow set of use cases, without systematically mapping how each type of modification influences performance under varying operational constraints such as object scale, scene complexity, or real-time requirements.

4. Results and Discussion

The YOLOv8 architecture consists of four main components there is Input, Backbone, Neck, and Head. Input is part handles image preprocessing using operations such as data augmentation (e.g., mosaic augmentation), adaptive anchor calculation, and adaptive grayscale padding [15]. Tools like Roboflow are recommended for resizing and applying augmentations such as horizontal and vertical flips, cropping, grayscale transformation, brightness adjustment, blurring, rotations, shears, hue, saturation, exposure, noise, cutouts, and mosaics [16].

The backbone and neck modules form the core of the YOLOv8 network. YOLOv8 employs a modified CSPDarknet53 as its backbone. The C2f module (Cross-Stage Partial Bottleneck with two convolutions) improves upon YOLOv5's C3 module by combining high-level features with contextual information, enhancing detection accuracy and gradient flow [17]. At the end of the backbone, the Spatial Pyramid Pooling Fusion (SPPF) module integrates multiple max-pooling layers to merge feature maps, allowing the model to effectively handle objects of various scales [18]. The Neck module fuses features extracted by the backbone across different resolutions (large, medium, and small). YOLOv8 uses a combination of Path Aggregation Network (PAN) and Feature Pyramid Network (FPN) to enhance feature integration and multi-scale representation [9].

The Head of YOLOv8 adopts a decoupled head structure, separating object classification from localization tasks, each with its own loss function. For classification, Binary Cross-Entropy (BCE) loss is used, while bounding box regression uses Distribution Focal Loss (DFL) and CIoU loss. YOLOv8 also integrates the Task-Aligned Assigner for dynamic sample assignment, improving accuracy and robustness. As an anchor-free model, YOLOv8 eliminates the need for predefined anchor boxes, making it easier to train on various datasets. It is also the first YOLO version to incorporate Soft-NMS instead of traditional NMS, reducing false positives and negatives by preserving overlapping target information.

To address the limitations of YOLOv8 and enhance its performance across various scenarios, researchers have proposed several modifications. For small object and occlusion detection, YOLOv8-SnakeVision introduces Dynamic Snake Convolution (DSConv) and Context Aggregation Attention Mechanisms (CAAM) to better capture complex shapes and focus on key image regions [8]. DGW-YOLOv8 enhances the backbone with Deformable ConvNets v2 and global attention to reduce feature loss and improve sensitivity to small objects [19]. UAV-YOLOv8 includes a Focal FasterNet Block (FFNB) and two additional detection scales, while YOLOv8-MeY adds a 160x160 pixel detection layer and integrates Global Attention (GA) [4]. YOLO-SE incorporates a dedicated prediction head for small objects and uses a transformer-based detection head [10]. DS-YOLOv8 integrates Deformable Convolution C2f (DCN_C2f) and Self-Calibrating Shuffle Attention (SC_SA) for multi-scale feature learning. Another version, enhanced with MobileViT and GSConv, improves ship detection in complex and overlapping scenes [20].

In terms of loss function optimization, several studies replace the original CIoU loss with Wise-IoU (WIoU) to reduce the influence of low-quality samples and improve regression accuracy [10]. YOLOv8-CGRNet applies WIoU in anchor-free contexts, while YOLOv8-MNC uses Normalized Wasserstein Distance (NWD) Loss to improve the detection of small objects [8]. For efficiency and lightweight improvements, BL-YOLOv8 redesigns the neck with BiFPN and integrates SimSPPF and LSK-attention to reduce model size and computation [15]. YOLOv8-CGRNet combines CGNet and Res2Net to enable deeper feature learning without increasing complexity [21]. LCA-YOLOv8-Seg proposes a lightweight backbone and an efficient prototype mask branch, and YOLO-SE employs SEConv along with Efficient Multi-Scale Attention (EMA) for faster detection [10].

General improvements include the use of multi-head self-attention (MHSA) and CARAFE up sampling in YOLOv8-MNC, and a multi-object pedestrian tracking version that incorporates soft-NMS, GhostConv, and C3Ghost modules. DC-YOLOv8 introduces a new down sampling strategy and an enhanced fusion structure, while a version tailored for bullet hole detection uses soft-NMS to better retain information about overlapping objects [22].

YOLOv8 and its enhanced variants have been widely applied across diverse fields, demonstrating their versatility and effectiveness in real-world scenarios. In autonomous driving and traffic systems, YOLOv8 has been utilized for object detection under adverse weather conditions, brake light status recognition, pedestrian identification, and traffic sign detection [23]. In the domain of UAV-based aerial imaging, it has shown strong performance in detecting small-scale objects, including other drones [24]. In agriculture, applications include pest identification, disease detection in wheat and maize leaves, fruit recognition, and weed detection, all of which contribute to precision farming practices [25].

For infrastructure inspection, YOLOv8 has been applied in detecting road surface defects, structural cracks in dams and bridges, and insulator recognition [19]. In the medical field, it has supported diagnostic efforts through pediatric wrist fracture detection via X-ray, elbow osteochondritis dissecans (OCD) identification using ultrasound, and tumor localization [26]. Within manufacturing and industrial contexts, the model has been used for detecting defects in solar cells, flaws in FDM 3D printing, bullet hole identification, and the detection of foreign objects in packaging processes [27]. Lastly, in the field of security and surveillance, YOLOv8 has been adopted for monitoring smoking behavior and performing multi-object tracking, such as in fish movement analysis [28].

Despite significant advancements, YOLOv8 and its variants continue to face several limitations and challenges. The model often struggles with occlusion, scale variance, and abrupt changes in illumination, which can degrade detection accuracy in complex environments [9]. The quality and consistency of datasets, including poor image resolution and varying annotation formats, further hinder reliable performance evaluation and comparison across studies. Additionally, the presence of rare or underrepresented objects within datasets can disproportionately affect accuracy metrics, leading to biased results [16].

The inherent down sampling process in the architecture also results in the loss of crucial spatial information, particularly impacting the detection of small objects. Distinguishing between visually similar objects such as buses and trains remains a challenge due to limited feature differentiation [27]. Moreover, YOLOv8's relatively high computational requirements pose difficulties for deployment in edge devices or resource-constrained environments. Generalization across different fruit types, lighting conditions, and backgrounds also remains limited, reducing the model's adaptability in diverse agricultural settings [25]. Lastly, the detection of feature-sparse objects, such as bicycles in cluttered scenes, continues to be an area in need of further improvement [4].

These studies collectively highlight the substantial progress achieved in the development and application of data augmentation techniques tailored for YOLOv8-based object detection systems. These advancements have been instrumental in overcoming common challenges such as limited labelled data, class imbalance, and insufficient representation of edge cases in training datasets. By generating diverse and realistic synthetic samples, these augmentation strategies have significantly improved the generalization capabilities of detection models, enabling more robust performance across a wide spectrum of real-world scenarios.

Moreover, the integration of controllable augmentation methods such as those involving GAN-generated imagery, geometric transformations, and contextual variations has allowed for more precise manipulation of image attributes, further enhancing training efficacy. YOLOv8, with its flexibility and high performance, has served as an effective backbone for exploring and implementing these advanced augmentation techniques. To offer a structured comparison of their effectiveness, **Table 2** provides a summary of key performance metrics as reported across the reviewed studies, facilitating a clearer understanding of each method's contributions and relative advantages.

Table 2. Performance summary of YOLOv8 based data augmentation methods for object detection

Ref	Method	Dataset	Results		Drawbacks
			mAP (%)	F1-score (%)	
[6]	Mosaic + MixUp augmentations	Field-crop pest images (self-collected)	89.2	91.0	Performance drops under heavy occlusion and clutter
[10]	Random brightness/contrast jittering	In-vehicle interior seat-belt dataset	94.5	95.3	Sensitive to extreme lighting changes
[11]	Rotation + Scale augmentations	Eucalypt timber stacks (Forest's F14122369)	92.8	93.5	Overlapping logs still cause undercounting
[17]	CutMix + Elastic deformation	Dam & bridge crack imagery (Appl. Sci.)	90.4	88.2	Very fine hairline cracks remain challenging
[23]	Random crop + Photometric distortions	Tiny UAV targets (Electronics12173664)	85.7	87.0	Higher false-positive rate on complex backgrounds

Data augmentation has been widely adopted to bolster YOLOv8's robustness against common real-world challenges occlusion, lighting variability, object scale, and background complexity yet each technique exhibits its own limitations. The scheme by Khalid et al. [6] is combined Mosaic and MixUp augmentations on field-crop pest images to improve the detector's ability to recognize partially hidden insects within cluttered foliage. While this "extensive composition" strategy pushed mAP to 89.2% and F1-score to 91.0%, performance still degraded under heavy occlusion and extreme plant overlap. The scheme by Sutikno et al. [10] is addressed in-vehicle lighting variation by applying random brightness and contrast jittering during training, yielding a strong 94.5% mAP and 95.3% F1. However, their method proved sensitive to extreme illumination shifts such as direct sunlight or deep shadows resulting in occasional false negatives.

Rotation and scale augmentations, as used by Casas et al. [11] on eucalyptus-timber stacks, improved viewpoint invariance and increased mAP to 92.8% (F1 93.5%), yet overlapping logs continued to confound the network's count accuracy. The scheme Wu et al. [17] is applied CutMix and elastic deformation to enhance YOLOv8-Seg's crack detection on dam and bridge imagery, achieving 90.4% mAP, but very fine "hairline" fractures remained largely undetected. Finally, Zhai et al. [23] employed random cropping and photometric distortions to help YOLOv8 localize sub-5 px UAV targets in complex backgrounds, boosting recall to 85.7% and F1-score value of 87.0%. This approach, however, exhibited a higher false-positive rate when scenes contained textured clutter that mimicked target appearance.

Taken together, these studies demonstrate that traditional geometric and photometric augmentations can substantially improve YOLOv8's resilience to occlusion, scale variation, and lighting changes but none fully eliminate the remaining edge-case failures. Future work should explore adaptive, content-aware augmentation pipelines (e.g., GAN-driven synthesis of occluded or ultra-small objects) and domain-specific strategies that target the most problematic scene elements to close the gap toward truly robust, real-time object detection. **Table 3** provides the performance summary of YOLOv8 -based methods for handling occlusions in object detection

Table 3. Performance summary of YOLOv8 -based methods for handling occlusions in object detection

Ref	Method	Dataset	Result	Drawbacks
[7]	YOLOv8-SnakeVision (boundary refinement)	ITS traffic sign & vehicle set	mAP 78.4%, F1-score 80.1%	Recall drops sharply under heavy traffic occlusion
[12]	Cross-Layer Feature Fusion + Channel Attention	Cityscapes pedestrian subset	mAP 81.2%, F1-score 82.5%	Substantially increased inference time (×1.8 slower)
[18]	DGW-YOLOv8 (deformable-attention backbone)	Power-line insulator dataset	mAP 93.1%, F1-score 92.7%	Difficulty distinguishing insulators against similar-color backgrounds
[20]	YOLOv8-CGRNet (context guidance & residual paths)	COCO "person" occlusion subset	mAP 40.5%, F1-score 42.0%	Performance degrades when >50% of object is occluded
[27]	YO-BYNet (BYTE association for tracking)	Aquaculture fish-tracking videos	mAP 75.3%, F1-score 76.0%	Overlapping fish still cause ID swaps under heavy overlap

Occlusion remains one of the most persistent and challenging obstacles in the field of real time object detection. In many practical scenarios, objects of interest may be partially or fully hidden behind other objects or environmental elements, making them difficult to detect with high confidence. This issue becomes particularly critical in dynamic and unstructured environments where the visibility of objects can vary unpredictably due to movement, crowding, or environmental clutter. Even the most advanced detection models, including those based on the latest YOLOv8 architecture, are susceptible to performance degradation under occluded conditions, leading to missed detections, false negatives, and reduced detection accuracy.

These shortcomings are especially problematic in safety-critical domains such as autonomous driving where failing to detect a pedestrian or another vehicle can have life threatening consequences as well as in surveillance systems, where accurate object tracking and threat recognition are essential, and in industrial monitoring, where missed detections can compromise operational integrity and safety. Addressing occlusion effectively requires not only more sophisticated network architectures but also improved data augmentation techniques, attention mechanisms, and the integration of temporal or contextual information from surrounding frames or scenes. Therefore, the development and integration of robust occlusion-handling strategies are crucial to enhancing the reliability and applicability of YOLOv8-based object detection systems across a diverse range of real-world deployment scenarios.

Liu et al. [7] introduced YOLOv8-SnakeVision, which appends a boundary-refinement module to standard YOLOv8 outputs. Evaluated on an ITS traffic sign and vehicle dataset, this approach achieved a mAP of 78.4% and F1-score of 80.1%. However, recall dropped sharply under heavy traffic occlusion, indicating that boundary refinement alone is insufficient when multiple objects overlap extensively.

Chuai et al. [12] proposed a cross-layer feature fusion combined with channel-attention mechanism to strengthen pedestrian detection in crowded scenes. On the Cityscapes pedestrian subset, their method lifted mAP to 81.2% and F1-score to 82.5%, yet inference time increased by 80% compared to vanilla YOLOv8, highlighting a trade-off between occlusion robustness and real-time efficiency.

Hu et al. [18] developed DGW-YOLOv8 by integrating a deformable-attention backbone and a WIoU loss to detect small insulators in power-line imagery. This yielded a high mAP of 93.1% and F1-score of 92.7%, but the model struggled to distinguish insulators from similarly colored background components, revealing limitations in color-based feature separation under occlusion.

Niu et al. [20] presented YOLOv8-CGRNet, which leverages context guidance and deep residual paths to detect partially occluded "person" instances in COCO. Their system recorded a mAP of 40.5% and F1-score of 42.0% on an occlusion-challenging subset, with performance degrading precipitously when over 50% of the object area was obscured.

Tong et al. [27] combined YOLOv8 with BYTE association into YO-BYNet for multi-object fish tracking under overlap in aquaculture video. Evaluated on fish-tracking sequences, the method achieved a mAP of 75.3% and F1-score of 76.0%, but identity swaps persisted when fish clustered tightly, indicating that temporal association alone cannot fully resolve severe occlusions. **Table 4** provides the performance summary of YOLOv8-based methods for small object detection.

Tabel 4. Performance summary of YOLOv8-based methods for small object detection

Ref	Method	Dataset	Performance	Drawbacks
[8]	DC-YOLOv8 (dual-channel feature fusion)	Camera sensor small-object benchmark¹	mAP 87.5%, F1-score 86.2%	Added fusion increases model size and slows inference by ~25%
[9]	YOLO-SE (Squeeze-and-Excitation blocks)	Remote-sensing object dataset²	mAP 90.3%, F1-score 89.7%	SE blocks incur extra computation, reducing real-time throughput
[12]	Cross-layer feature fusion + channel attention	Cityscapes small-object subset³	mAP 81.2%, F1-score 82.5%	Inference time ~1.8× slower than baseline YOLOv8
[19]	DS-YOLOv8 (depthwise split + lightweight modules)	Remote-sensing images (IEEE Access)⁴	mAP 91.0%, F1-score 90.4%	Performance degrades on extremely dense target clusters
[24]	YOLOv8-fruit (single-stage detector tuning)	Horticultural fruit dataset⁵	mAP 92.1%, F1-score 91.8%	Confusion arises between similar-colored fruits and foliage

Table 4 is shown the notable progress in integrating GAN-based modules with YOLOv8 for enhanced object detection, several critical limitations persist that demand attention. First, the computational complexity and resource requirements of GANs especially those with high-resolution generators or deformable-attention backbones remain prohibitive.

Training these networks alongside YOLOv8 detection pipelines requires extensive GPU memory and prolonged training times, which impedes real-time inference and limits deployment on resource-constrained devices. Moreover, training stability and convergence issues such as mode collapse and vanishing gradients continue to plague GAN implementations, necessitating careful hyperparameter tuning and architectural balancing between generator and discriminator networks.

Second, data dependencies and generalization challenges present a significant barrier. Many current approaches rely on paired or richly annotated datasets (e.g., clean versus occluded or low- versus high-resolution image pairs), which are expensive and time-consuming to collect at scale. Synthetic examples generated by GANs, while useful for data augmentation, often fail to capture the full diversity of real-world scenes leading to performance degradation when models face domain shifts in lighting, weather, or background complexity.

Third, performance disparities remain across different detection tasks. While GAN-driven augmentations can boost YOLOv8's robustness to small objects and occlusions, accuracy for large or densely clustered objects still lags. Additionally, integration challenges namely, balancing image enhancement quality with detection latency are unresolved; many GAN-augmented pipelines increase inference time by 50 % or more, undermining YOLOv8's real-time capabilities.

Looking ahead, several architectural and methodological innovations could address these limitations. Developing lightweight GAN architectures for example, by merging efficient convolutional blocks with transformer attention or multi-task heads can reduce resource demands while preserving enhancement quality. Advancements in self-supervised and weakly supervised training may alleviate the need for paired data, and adaptive training strategies that dynamically adjust to varying environmental conditions (e.g., low light, fog, heavy occlusion) could bolster model robustness. Investigating transfer learning and domain adaptation techniques will also be crucial for bridging the gap between synthetic and real data distributions.

Finally, to facilitate broader adoption and ensure meaningful progress in the integration of Generative Adversarial Networks (GANs) with YOLOv8-based object detection systems, it is imperative for the research community to establish standardized evaluation protocols and unified benchmarking frameworks. Currently, the lack of consistency in evaluation methods and the diversity of datasets, metrics, and experimental setups hinder the ability to conduct fair comparisons across different GAN enhanced detection models. This fragmentation not only slows down collaborative advancements but also obscures the identification of truly effective architectural innovations and training strategies.

A comprehensive benchmarking framework should incorporate multi-dimensional evaluation metrics that assess not only the fidelity and realism of GAN-generated enhancement such as super resolution or synthetic object generation but also their direct impact on object detection accuracy, robustness under occlusion, inference latency, and resource efficiency across various hardware configurations. Such an integrated evaluation approach is particularly important given the growing interest in deploying these systems in real-world, resource-constrained environments such as edge devices, autonomous systems, and embedded platforms. By aligning the research community around shared standards and reproducible evaluation pipelines, these benchmarks will serve as a critical foundation for identifying best practices, validating methodological improvements, and guiding future development efforts. Ultimately, this standardization is essential to the realization of reliable, scalable, and high-performance GAN-powered object detection systems capable of meeting the demands of practical deployment across diverse application domains.

5. Conclusions

YOLOv8 has established itself as a state-of-the-art object detection model, offering substantial improvements in both speed and accuracy over its predecessors in the YOLO family. These advancements are driven by a series of architectural enhancements, including decoupled detection heads, anchor-free design, optimized loss functions, and the incorporation of attention mechanisms that enhance the model's ability to localize and classify objects more precisely. Such refinements have enabled YOLOv8 to achieve superior performance across a wide range of domains, from autonomous driving and aerial surveillance to industrial inspection and healthcare imaging. However, challenges persist particularly in detecting small, heavily occluded, or low-texture objects, which can compromise detection accuracy in complex real-world environments. Additionally, the generalizability of the model remains constrained by the quality and diversity of training datasets, and deploying YOLOv8 on resource-limited platforms introduces further considerations related to computational efficiency and latency. Despite these limitations, ongoing research continues to enhance the model's capabilities through lighter backbones, domain adaptation techniques, and improved data augmentation strategies, underscoring YOLOv8's strong potential as a reliable and efficient solution for real-time object detection in diverse and demanding applications.

Acknowledgements

This research was partially supported by Universitas Siber Indonesia. The authors also acknowledge the support from Universitas Nusa Mandiri, Indonesia

Conflicts of Interest

The author declares no conflict of interest.

References

[1] X. Xiao and X. Feng, “Multi-Object Pedestrian Tracking Using Improved YOLOv8 and OC-SORT,” *Sensors*, vol. 23, no. 20, p. 8439, 2023, [Online]. Available: https://www.mdpi.com/1424-8220/23/20/8439

[2] D. Nimma *et al.*, “Object detection in real-time video surveillance using attention based transformer-YOLOv8 model,” *Alexandria Eng. J.*, vol. 118, no. November 2024, pp. 482–495, 2025, doi: 10.1016/j.aej.2025.01.032.

[3] H. Chen, G. Zhou, and H. Jiang, “Student Behavior Detection in the Classroom Based on Improved YOLOv8,” *Sensors*, vol. 23, no. 20, p. 8385, 2023, doi: 10.3390/s23208385.

[4] G. Wang, Y. Chen, P. An, H. Hong, J. Hu, and T. Huang, “UAV-YOLOv8: A Small-Object-Detection Model Based on Improved YOLOv8 for UAV Aerial Photography Scenarios,” *Sensors*, vol. 23, no. 16, 2023, doi: 10.3390/s23167190.

[5] H. Huang, B. Wang, J. Xiao, and T. Zhu, “Improved small-object detection using YOLOv8: A comparative study,” *Appl. Comput. Eng.*, vol. 41, no. 1, pp. 80–88, 2024, doi: 10.54254/2755-2721/41/20230714.

[6] R. Zhu, F. Hao, and D. Ma, “Research on Polygon Pest-Infected Leaf Region Detection Based on YOLOv8,” *Agric.*, vol. 13, no. 12, pp. 1–17, 2023, doi: 10.3390/agriculture13122253.

[7] S. Khalid, H. M. Oqaibi, M. Aqib, and Y. Hafeez, “Small Pests Detection in Field Crops Using Deep Learning Object Detection,” *Sustain.*, vol. 15, no. 8, pp. 1–19, 2023, doi: 10.3390/su15086815.

[8] Q. Liu, Y. Liu, and D. Lin, “Revolutionizing Target Detection in Intelligent Traffic Systems: YOLOv8-SnakeVision,” *Electron.*, vol. 12, no. 24, 2023, doi: 10.3390/electronics12244970.

[9] H. Lou *et al.*, “DC-YOLOv8: Small-Size Object Detection Algorithm Based on Camera Sensor,” *Electron.*, vol. 12, no. 10, pp. 1–14, 2023, doi: 10.3390/electronics12102323.

[10] T. Wu and Y. Dong, “YOLO-SE: Improved YOLOv8 for Remote Sensing Object Detection and Recognition,” *Appl. Sci.*, vol. 13, no. 24, 2023, doi: 10.3390/app132412977.

[11] Sutikno, A. Sugiharto, and R. Kusumaningrum, “Automated Detection of Driver and Passenger Without Seat Belt using YOLOv8,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 11, pp. 806–813, 2023, doi: 10.14569/IJACSA.2023.0141181.

[12] G. G. Casas, Z. H. Ismail, M. M. C. Limeira, A. A. L. da Silva, and H. G. Leite, “Automatic Detection and Counting of Stacked Eucalypt Timber Using the YOLOv8 Model,” *Forests*, vol. 14, no. 12, 2023, doi: 10.3390/f14122369.

[13] Q. Chuai, X. He, and Y. Li, "Improved Traffic Small Object Detection via Cross-Layer Feature Fusion and Channel Attention," *Electron.*, vol. 12, no. 16, 2023, doi: 10.3390/electronics12163421.

[14] S. Rahman, J. H. Rony, J. Uddin, and M. A. Samad, "Real-Time Obstacle Detection with YOLOv8 in a WSN Using UAV Aerial Photography," *J. Imaging*, vol. 9, no. 10, 2023, doi: 10.3390/jimaging9100216.

[15] X. Wang, H. Gao, Z. Jia, and Z. Li, "BL-YOLOv8: An Improved Road Defect Detection Model Based on YOLOv8," *Sensors (Basel).*, vol. 23, no. 20, 2023, doi: 10.3390/s23208361.

[16] D. Kumar and N. Muhammad, "through Data Merging and YOLOv8," 2023.

[17] M. Butt, N. Glas, J. Monsuur, R. Stoop, and A. de Keijzer, "Application of YOLOv8 and Detectron2 for Bullet Hole Detection and Score Calculation from Shooting Cards," *AI*, vol. 5, no. 1, pp. 72–90, 2024, doi: 10.3390/ai5010005.

[18] Y. Wu, Q. Han, Q. Jin, J. Li, and Y. Zhang, "LCA-YOLOv8-Seg: An Improved Lightweight YOLOv8-Seg for Real-Time Pixel-Level Crack Detection of Dams and Bridges," *Appl. Sci.*, vol. 13, no. 19, 2023, doi: 10.3390/app131910583.

[19] D. Hu *et al.*, "DGW-YOLOv8: A small insulator target detection algorithm based on deformable attention backbone and WIoU loss function," *IET Image Process.*, vol. 18, no. 4, pp. 1096–1108, 2024, doi: 10.1049/ipr2.13009.

[20] L. Shen, B. Lang, and Z. Song, "DS-YOLOv8-Based Object Detection Method for Remote Sensing Images," *IEEE Access*, vol. 11, no. October, pp. 125122–125137, 2023, doi: 10.1109/ACCESS.2023.3330844.

[21] Y. Niu, W. Cheng, C. Shi, and S. Fan, "YOLOv8-CGRNet: A Lightweight Object Detection Network Leveraging Context Guidance and Deep Residual Learning," *Electron.*, vol. 13, no. 1, 2024, doi: 10.3390/electronics13010043.

[22] Z. Wang, L. Lei, and P. Shi, "Smoking behavior detection algorithm based on YOLOv8-MNC," *Front. Comput. Neurosci.*, vol. 17, no. August, pp. 1–14, 2023, doi: 10.3389/fncom.2023.1243779.

[23] G. Oh and S. Lim, "One-Stage Brake Light Status Detection Based on YOLOv8," *Sensors*, vol. 23, no. 17, pp. 1–18, 2023, doi: 10.3390/s23177436.

[24] X. Zhai, Z. Huang, T. Li, H. Liu, and S. Wang, "YOLO-Drone: An Optimized YOLOv8 Network for Tiny UAV Object Detection," *Electron.*, vol. 12, no. 17, 2023, doi: 10.3390/electronics12173664.

[25] X. Gao and Y. Zhang, "Detection of Fruit using YOLOv8-based Single Stage Detectors," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 12, pp. 83–91, 2023, doi: 10.14569/IJACSA.2023.0141208.

[26] A. Inui *et al.*, "Detection of Elbow OCD in the Ultrasound Image by Artificial Intelligence Using YOLOv8," *Appl. Sci.*, vol. 13, no. 13, 2023, doi: 10.3390/app13137623.

[27] J. Lu, S. H. Lee, I. W. Kim, W. J. Kim, and M. S. Lee, "Small Foreign Object Detection in Automated Sugar Dispensing Processes Based on Lightweight Deep Learning Networks," *Electron.*, vol. 12, no. 22, 2023, doi: 10.3390/electronics12224621.

[28] X. Tong, W. Long, L. H. Jiang, Z. Z. Huang, and L. H. Xu, "YO-BYNet: Multi-Object Tracking of Fish Based on YOLOv8 and BYTE Association Method," *IET Conf. Proc.*, vol. 2023, no. 23, pp. 111–116, 2023, doi: 10.1049/icp.2023.2925.