Minimum Detectable Object Area in YOLOv11: Effects of Resolution and Model Scale

[1]Ketan Kanjiya, [2]Piyush Sonani, [3]Upendrasinh Zala
[1]Chief Research Officer, [2]Chief Technology Officer, [3]Chief Executive Officer
[1]Information Technology Department,
[1]Kshatrainfotech Pvt Ltd, Ahmedabad, India
[1]ketan@neuramonks.com, [2]piyush@neuramonks.com, [3]upen@neuramonks.com

***Abstract*—Detecting objects that occupy only a small fraction of an image remains a fundamental limitation of modern object detection systems. Although recent YOLO architectures employ multi-scale feature fusion to improve robustness across object sizes, the minimum object area required for reliable detection is rarely quantified explicitly. This paper presents a systematic, variant wise analysis of the detection limits of YOLOv11, focusing on how input image resolution, object scale, and model capacity jointly influence detectability. Experiments evaluate five object categories across image resolutions ranging from 2560×2560 to 40×40 pixels, progressively reducing object size while maintaining consistent evaluation conditions. The minimum detectable object area is defined as the smallest object consistently detected above a 70% confidence threshold and is reported as a percentage of total image area. Results show that input resolution strongly dominates detection feasibility, with minimum detectable area increasing sharply as resolution decreases. While larger YOLOv11 variants consistently outperform smaller models at moderate resolutions, all variants exhibit a resolution-dependent detection floor below which detection collapses regardless of model capacity. Additionally, object characteristics such as shape and texture significantly influence detection thresholds. These findings provide practical insights into the operational limits of YOLOv11 and inform model selection and resolution design for small object detection applications.**

***Index Terms*—Yolo, Object Detection, Minimum Object Area, Deep Learning, Detection Threshold Analysis**

I. Introduction

Object detection systems are increasingly deployed in environments where target objects occupy only a small fraction of the image, such as surveillance footage, aerial imagery, medical scans, and crowded public spaces [1-4]. Although modern deep learning based detectors have achieved impressive overall accuracy, detecting very small objects remains a persistent challenge. In such cases, limited pixel coverage constrains the availability of discriminative features, making reliable localization and classification difficult.

One-stage detectors, particularly the YOLO family, are widely adopted due to their real-time performance and unified detection pipeline. Foundational architectures such as YOLO, SSD, and Fast R-CNN established the standard pipeline consisting of a feature extraction backbone and a detection head [5-7]. Successive YOLO versions have introduced architectural improvements aimed at enhancing multi-scale feature representation and computational efficiency [8]. However, while these advances improve general detection capability, the minimum object area required for reliable detection is rarely quantified explicitly. In practice, understanding this lower bound is critical for system design, dataset curation, and deployment decisions, especially in applications involving distant or visually small targets.

The detectability of an object is strongly influenced by its spatial extent relative to the input image. Downsampling and resizing operations can cause small objects to vanish or lose structural coherence [9,10]. Prior studies have shown that resolution degradation and increased object distance jointly reduce detection performance, particularly for small objects [11]. Domain specific works, such as tiny vehicle detection [12], medical image analysis [13], and UAV based detection [14], further emphasize that small object sensitivity varies substantially across architectures and configurations. However, these studies typically focus on task specific enhancements rather than systematically identifying detection thresholds.

YOLO architectures employ multi-scale feature fusion through FPN and PAN structures to address object size variability. While these mechanisms improve robustness across scales, they do not guarantee detection below a certain object size, particularly when feature maps lack sufficient spatial resolution to represent extremely small regions. Moreover, different YOLOv11 variants vary in depth, parameter count, and receptive field structure, suggesting that their sensitivity to small object areas may differ significantly.

This paper investigates the detection limits of YOLOv11 by conducting a variant wise analysis of the minimum object area required for successful detection. Using a controlled experimental setup, objects are systematically evaluated across decreasing spatial areas while maintaining consistent training conditions and dataset composition. Performance is assessed using standard detection metrics to identify the point at which detection reliability degrades for each YOLOv11 variant. By explicitly quantifying minimum detectable object area, this study provides practical insights into the strengths and limitations of YOLOv11 variants and offers guidance for selecting appropriate models in small object dominated applications.

II. Dataset Preparation

In addition to real-world aerial imagery, a synthetic evaluation setup was designed to precisely measure the minimum detectable object size across different input resolutions and model configurations. This experiment does not rely on an existing dataset; instead, a controlled synthetic dataset was generated programmatically.

Five common object categories: man, pizza, kite, dog, and laptop were selected to represent diverse shapes, textures, and aspect ratios. Background images were created at fixed spatial resolutions of 2560×2560, 1280×1280, 640×640, 320×320, 160×160, 80×80, and 40×40 pixels. For each resolution, objects were rendered at multiple scales while preserving their original aspect ratios. This synthetic dataset provides a controlled environment for isolating the effects of object size and image resolution, independent of

background complexity or environmental variability. Representative examples of the object cutouts and an object placed on a uniform white background are shown in figure 1 for visual reference.

Figure 1: Representative examples illustrating object cutouts and an example object rendered on a uniform white background. Thin borders are added for visual separation from the page background.

III. Methodology

Figure 2 illustrates the YOLOv11 object detection architecture used in this study.

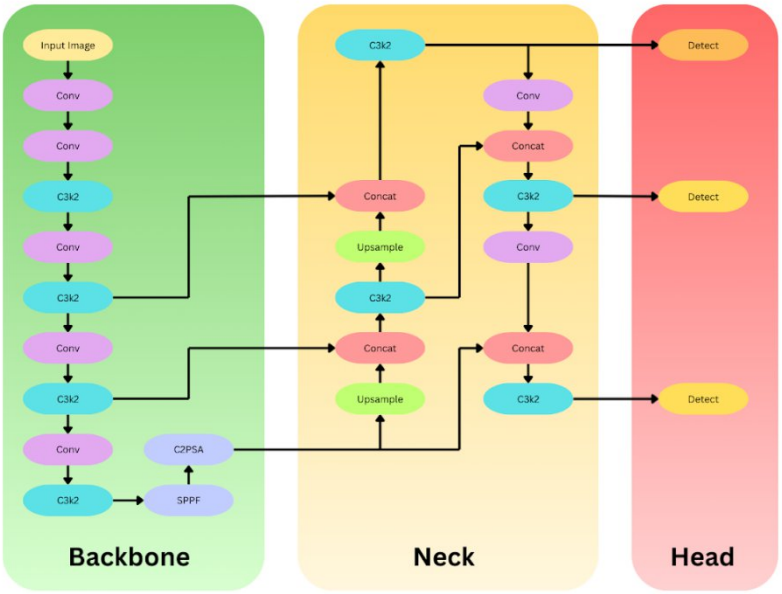

Figure 2: The architecture of the YOLOv11 object detection framework.

This study focuses on quantifying the minimum object area detectable by YOLOv11 variants under controlled conditions. Five object categories: man, pizza, kite, dog, and laptop were rendered onto uniform white background images at multiple spatial resolutions ranging from 2560×2560 to 40×40 pixels. For each background resolution and object category, object size was progressively reduced while preserving the original aspect ratio.

Detection was performed using YOLOv11 models under identical evaluation settings. The minimum detectable object area, measured in pixels as the rendered bounding box area of the object, was defined as the smallest object size that could be consistently detected with a confidence score of at least 70%. This criterion enables a precise assessment of detection limits as object scale decreases relative to image resolution.

This controlled experimental setup isolates the effects of object size, image resolution, and model architecture, eliminating confounding factors such as background clutter, occlusion, and illumination variability. As a result, the analysis provides a fine grained characterization of detector sensitivity to small objects across YOLOv11 variants.

All experiments were conducted on an Intel i5 CPU with 16 GB RAM, using CPU based inference to ensure consistent evaluation conditions across all model variants. By maintaining uniform preprocessing, confidence thresholds, and evaluation protocols, observed performance differences can be attributed primarily to variations in object area, input resolution, and YOLOv11 model capacity.

IV. Results and Discussions

This section presents the results of the minimum detectable object area analysis conducted across different YOLOv11 variants, image resolutions, and object categories. Five object categories man, pizza, laptop, kite, and dog were placed on uniform white background images of seven resolutions: 2560×2560, 1280×1280, 640×640, 320×320, 160×160, 80×80, and 40×40. These background resolutions are reported as column headers in Table 1.

For each background resolution and model variant, the object size was progressively reduced until the detector failed to produce a valid detection. The smallest object size that remained detectable was recorded. Results are reported as the minimum detectable object area expressed as a percentage of the total image area.

Table 1: Minimum detectable object area (% of image area) across YOLOv11 variants and image resolutions.

Model	Object	2560	1280	640	320	160	80	40
yolov11n	man	0.13	0.14	0.11	0.15	0.41	6.30	no detection
yolov11n	laptop	3.45	3.43	4.93	13.97	40.20	no detection	no detection
yolov11n	kite	0.06	0.06	0.12	0.21	0.61	9.34	no detection
yolov11n	pizza	1.91	2.92	2.74	2.20	3.09	11.48	no detection
yolov11n	dog	0.27	0.70	0.61	0.66	1.72	no detection	no detection
yolov11s	man	0.10	0.08	0.10	0.15	0.41	1.64	33.75
yolov11s	laptop	0.72	1.21	1.50	8.20	15.84	no detection	no detection
yolov11s	kite	0.19	0.20	0.22	0.30	0.43	11.78	no detection
yolov11s	pizza	4.05	4.57	4.22	7.37	10.89	13.66	no detection
yolov11s	dog	0.62	0.46	0.61	0.61	1.56	10.56	no detection
yolov11m	man	0.05	0.05	0.05	0.15	0.41	7.22	47.25
yolov11m	laptop	0.56	1.31	1.84	4.79	13.31	no detection	no detection
yolov11m	kite	0.04	0.04	0.07	0.43	0.43	6.88	65.88
yolov11m	pizza	1.56	3.99	5.18	9.88	9.60	11.16	44.63
yolov11m	dog	0.60	0.54	0.47	0.47	0.77	17.02	18.06
yolov11l	man	0.03	0.06	0.05	0.15	0.21	1.64	28.88
yolov11l	laptop	0.26	0.86	1.50	4.14	25.78	no detection	no detection
yolov11l	kite	0.04	0.11	0.49	1.37	4.12	2.44	no detection
yolov11l	pizza	3.28	4.27	3.53	4.69	6.18	12.38	44.63
yolov11l	dog	0.13	0.23	0.15	0.66	1.72	16.00	no detection
yolo11x	man	0.03	0.04	0.05	0.10	0.21	1.64	12.38
yolo11x	laptop	0.39	0.79	0.80	1.58	2.75	no detection	no detection
yolo11x	kite	0.14	0.16	0.31	1.03	0.43	2.44	6.88
yolo11x	pizza	1.91	1.76	1.58	3.37	3.09	4.81	no detection
yolo11x	dog	0.14	0.13	0.13	0.25	0.56	4.52	no detection

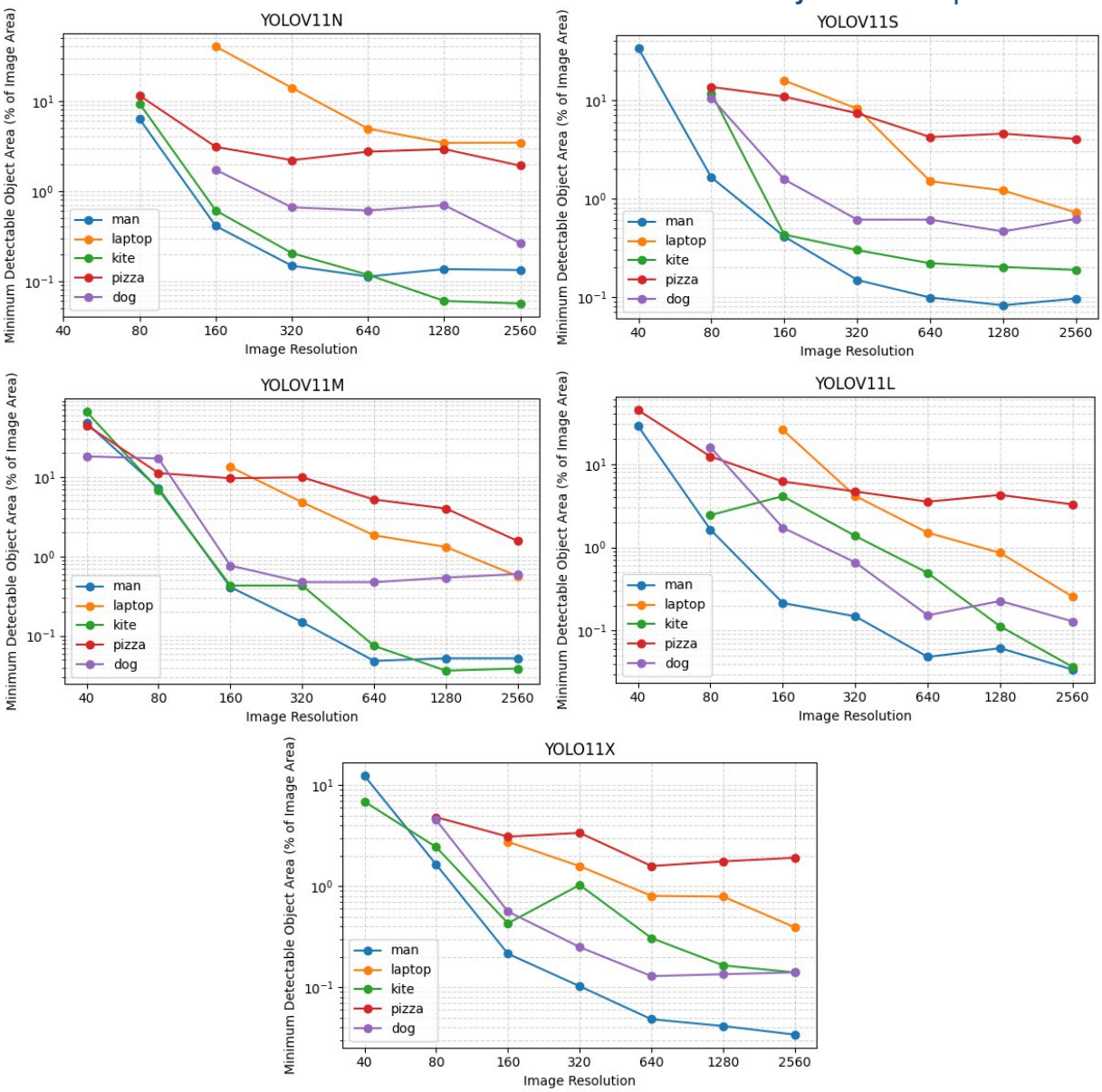

Figure 3: Minimum detectable object area (% of image) as a function of input resolution for five object categories (man, dog, laptop, pizza, kite). Each subplot corresponds to one YOLOv11 variant (n, s, m, l, x).

Resolution Dominates Minimum Detectable Object Area

Table 1 shows that for all YOLOv11 variants, the minimum detectable object area as a percentage of the image increases sharply as image resolution decreases. At high resolutions from 2560 to 640, most objects are detected at 1% area or less. At 320, thresholds rise to approximately 0.1% to 10%, at 160 many objects require more than 10%, and at 80 to 40 detection often fails. This indicates that detection feasibility is primarily driven by input resolution due to feature map downsampling and grid cell limitations. This trend is also illustrated in Figure 3, which shows minimum detectable area versus image resolution for all object categories across YOLOv11 variants.

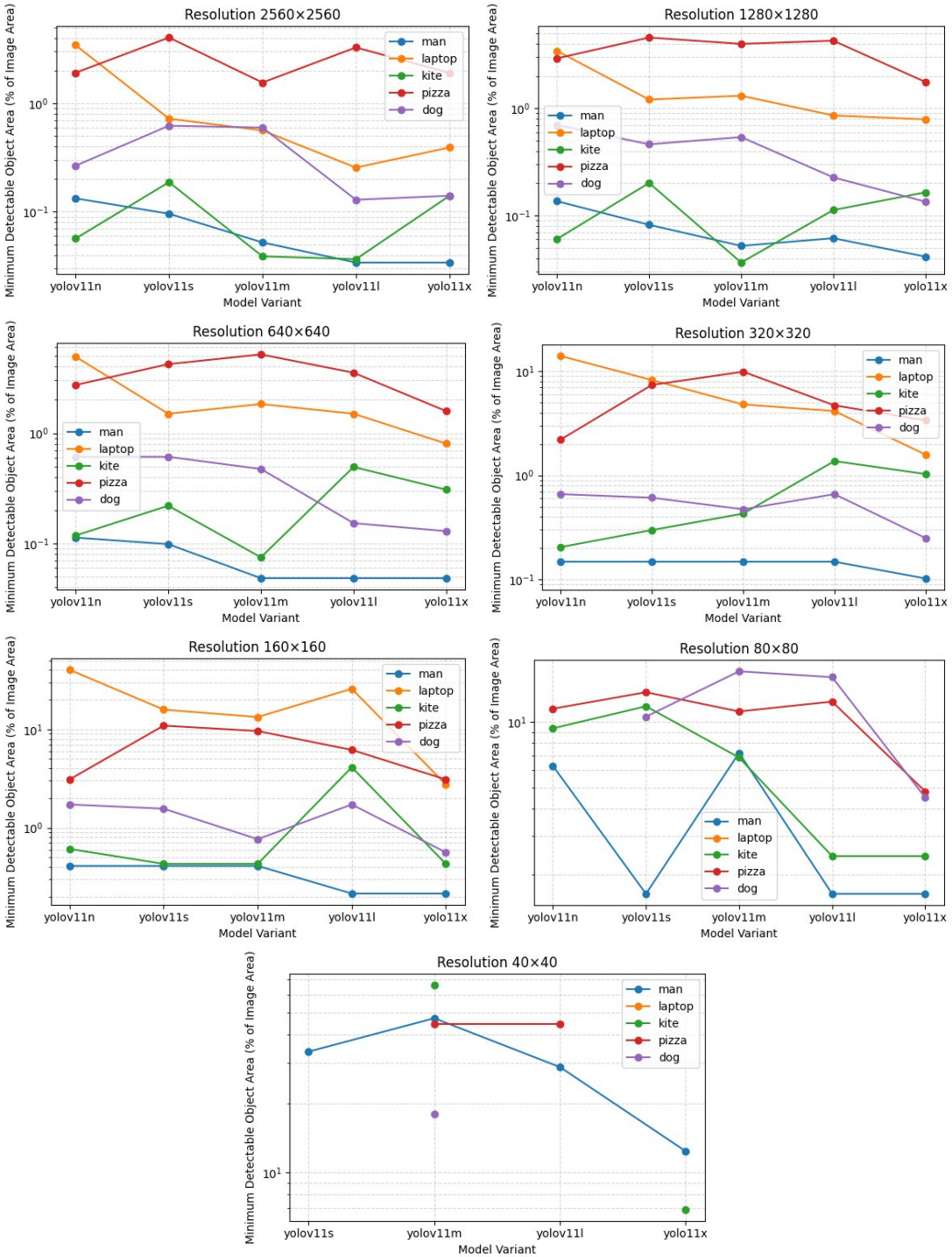

Figure 4: Minimum detectable object area (% of image) across YOLOv11 variants for seven resolutions (2560-40). Each subplot corresponds to a fixed resolution.

For the same resolution and object, larger YOLOv11 models such as m, l, and x generally detect smaller objects than lighter variants n and s. For example, the kite object at 80×80 pixels is detected at 9.34% by YOLOv11n, 11.78% by YOLOv11s, 6.88% by YOLOv11m, and 2.44% by both YOLOv11l and YOLOv11x, as shown in Table 1. However, at an extremely low resolution of 40×40 pixels, all models fail for most objects. This relationship is visualized in Figure 4, which shows minimum detectable area versus model variant across all resolutions.

Resolution Dependent Detection Floor

Across models, detection fails when objects occupy less than approximately 5% to 10% of the image at 80×80 pixels or less than approximately 1% to 2% at 320×320 pixels, as reported in Table 1. This defines a practical detection floor independent of model size.

Object Shape and Texture Influence Detectability

Objects with rich edges and texture such as man and dog are detectable at smaller areas than rigid or flat objects such as laptop and pizza. For instance, at 160×160 pixels using YOLOv11n, the minimum detectable area is 0.41% for man, 1.72% for dog, 3.09% for pizza, and 40.20% for laptop, as summarized in Table 1.

Detection Degrades Non Linearly

The increase in minimum detectable area with decreasing resolution is highly non linear. For example, for the kite object using YOLOv11n, the minimum detectable area is 0.06% at 2560, 0.12% at 640, 0.21% at 320, 0.61% at 160, 9.34% at 80, and detection fails at 40, as shown in Table 1 and Figure 4. Detection collapses abruptly once spatial information falls below a critical granularity.

V. Conclusion

This study systematically analyzed the minimum detectable object area of YOLOv11 across multiple model variants and input resolutions, providing a clear characterization of detection limits for small objects. Results demonstrate that input image resolution is the dominant factor governing detectability, with minimum detectable object area increasing sharply as resolution decreases. While larger YOLOv11 variants consistently detect smaller objects than lightweight models at moderate resolutions, all variants exhibit a resolution-dependent detection floor beyond which increasing model capacity yields limited benefit. At very low resolutions, detection collapses regardless of model scale, indicating fundamental constraints imposed by spatial downsampling and feature map granularity. The analysis also reveals that object-specific characteristics, particularly shape and texture, significantly influence detection thresholds, with articulated and textured objects remaining detectable at smaller areas than rigid or low-texture objects. By explicitly quantifying detection limits, this work offers practical guidance for selecting appropriate YOLOv11 variants and input resolutions in applications dominated by small objects. These findings can inform dataset design, deployment strategies, and resolution trade-offs in real-world detection systems.

References

[1] A. F. Japar, H. R. Ramli, N. M. H. Norsahperi, and W. Z. W. Hasan, “Oil palm loose fruit detection using YOLOv4 for an autonomous mobile robot collector,” IEEE Access, vol. 12, pp. 138582–138593, 2024.

[2] S. Zhang, Y. Wu, C. Men, and X. Li, “Tiny YOLO optimization oriented bus passenger object detection,” Chinese J. Electron., vol. 29, no. 1, pp. 132–138, 2019.

[3] A. Priadana, D.-L. Nguyen, X.-T. Vo, J. Choi, R. Ashraf, and K. Jo, “HFD-YOLO: Improved YOLO network using efficient attention modules for real-time one-stage human fall detection,” IEEE Access, vol. 13, pp. 41248–41258, 2025.

[4] Q. Zhang, K. Ahmed, M. I. Khan, H. Wang, and Y. Qu, “YOLO-FCE: A feature and clustering enhanced object detection model for species classification,” Pattern Recognit., vol. 171, p. 112218, 2026.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, June 27–30, 2016, pp. 779–788.

[6] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “SSD: Single shot multibox detector,” in Proc. Eur. Conf. Comput. Vis. (ECCV), Amsterdam, The Netherlands, Oct. 8–16, 2016, pp. 21–37.

[7] R. Girshick, “Fast R-CNN,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Santiago, Chile, Dec. 7–13, 2015, pp. 1440–1448.

[8] A. A. Murat and M. S. Kiran, “A comprehensive review on YOLO versions for object detection,” Eng. Sci. Technol., Int. J., vol. 70, p. 102161, 2025.

[9] Y. Meng, J. Zhan, K. Li, et al., “A rapid and precise algorithm for maize leaf disease detection based on YOLO MSM,” Sci. Rep., vol. 15, p. 6016, 2025.

[10] B. Peng and T.-K. Kim, “YOLO-HF: Early detection of home fires using YOLO,” IEEE Access, vol. 13, pp. 79451–79466, 2025.

[11] Y. Hao, H. Pei, Y. Lyu, Z. Yuan, J.-R. Rizzo, Y. Wang, and Y. Fang, “Understanding the impact of image quality and distance of objects to object detection performance,” in Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS), Detroit, MI, USA, Oct. 1–5, 2023, pp. 11436–11442.

[12] D. Padilla Carrasco, H. A. Rashwan, M. Á. García, and D. Puig, “T-YOLO: Tiny vehicle detection based on YOLO and multi-scale convolutional neural networks,” IEEE Access, vol. 11, pp. 22430–22440, 2023.

[13] B. Lufyagila, B. Mgawe, and A. Sam, “Fine-tuned YOLO-based deep learning model for detecting malaria parasites and leukocytes in thick smear images: A Tanzanian case study,” Mach. Learn. Appl., vol. 21, p. 100687, 2025.

[14] X. Meng, F. Yuan, and D. Zhang, “Improved model MASW YOLO for small target detection in UAV images based on YOLOv8,” Sci. Rep., vol. 15, p. 25027, 2025.