Benchmarking YOLOv8, YOLOv9, and YOLOv10 for Detecting Tiny and Small Objects in Complex Environments

George Albert Bitwire
Graduate School of Electronic and Electrical Engineering, Kyungpook National University
Daegu, Republic of Korea
bitwire@knu.ac.kr

Dong Seog Han*
Graduate School of Electronic and Electrical Engineering, Kyungpook National University
Daegu, Republic of Korea
dshan@knu.ac.kr

***Abstract*—In recent years, object detection algorithms have evolved significantly, with the you only look once(YOLO) series standing out for their real-time performance and accuracy. This paper presents a comprehensive benchmarking of the latest iterations—YOLOv8, YOLOv9, and YOLOv10—specifically focusing on their capabilities in detecting tiny and small objects within complex environments. These scenarios, characterized by high object density, cluttered backgrounds, and varying lighting conditions, pose significant challenges for detection algorithms. Through extensive experiments on the VisDrone-DET datasets, we evaluate the performance of these models in terms of precision, recall, and processing speed. The findings indicate that YOLOv10 demonstrates superior accuracy, making it a viable option for resource-constrained applications. This study serves as a valuable resource for researchers and practitioners seeking to deploy object detection models in complex real-world environments, particularly where detecting tiny and small objects is critical.**

***Index Terms*—YOLOv8, YOLOv9, YOLOv10, Deep Learning, Object Detection, VisDrone-DET.**

I. Introduction

The detection of tiny and small objects in complex environments has become a significant challenge in the field of computer vision, particularly in applications requiring high precision and reliability. Object detection models like the you only look once (YOLO) series have gained a widespread popularity for their ability to perform real-time detection with impressive accuracy. However, the performance of these models can vary considerably when applied to scenarios involving tiny and small objects or where factors such as object scale, occlusion, and background clutter play a critical role.

In recent years, the YOLO architecture has undergone significant advancements, with the introduction of YOLOv8 [1], YOLOv9 [2], and YOLOv10 [3]. Each iteration has introduced improvements and optimizations all aimed at enhancing accuracy and speed [4]. Despite these advancements, there remains a lack of comprehensive evaluation focused specifically on the detection of tiny and small objects in complex environments.

This paper seeks to address this gap by benchmarking the performance of YOLOv8, YOLOv9, and YOLOv10 in detection of tiny and small objects under challenging condition. We evaluate these models across a variety of complex environments that simulate real-would scenarios, utilizing a challenging dataset VisDrone-DET [5]. This dataset is specifically designed to test detection performance in settings with cluttered backgrounds, varying lighting conditions, and significant occlusions, providing a rigorous benchmark for accessing the capabilities of YOLOv8, YOLOv9, and YOLOv10 in detecting tiny and small objects under real-world conditions.

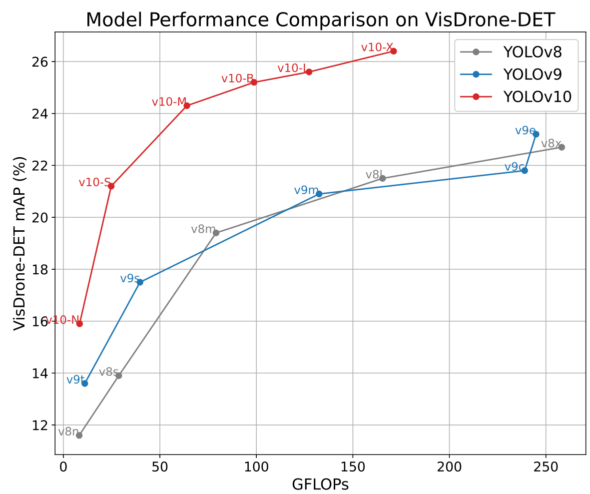

Fig. 1. Shows mAP vs GFLOPs for the model variations YOLOv8, YOLOv9 and YOLOv10.

II. Method

In this section, we delve into the YOLO models and evaluation metrics used.

A. YOLO Models

YOLOv10 introduces Non-Maximum Suppression(NMS)-free training, utilizing consistent dual assignments to eliminate

Fig. 2. Comparison of Detection Results Across Models: Ground Truth vs. YOLOv8, YOLOv9, and YOLOv10 on VisDrone Dataset.

the need for NMS, significantly reducing inference latency. Additionally, the model exhibits a holistic model design, optimizing various components such as lightweight classification heads, spatial-channel decoupled down-sampling, and rank-guided block design, balancing efficiency with accuracy. Enhanced model capabilities are achieved through the incorporation of large-kernel convolutions and partial self-attention modules, boosting performance without incurring significant computational costs [3].

YOLOv9 improving on its predecessors by optimizing the backbone network for better feature extraction, refining the neck and head for enhanced multi-scale detection, and introducing a more effective loss function to better handle complex scenes. These upgrades result in improved accuracy, especially for smaller objects, while maintaining the model's signature high inference speed, making it suitable for real-time applications [2].

YOLOv8 introduces a more efficient backbone architecture, which enhances feature extraction and improves detection accuracy, particularly for small objects. YOLOv8 also refines the model's neck and head components, enabling better multi-scale feature fusion and more precise object localization and classification. Additionally, the model includes updates to its training process, such as an improved loss function and data augmentation techniques, which contribute to faster convergence and better generalization.

B. Model Evaluation

Metrics such as precision, recall, average precision(AP), mean average precision(mAP), and accuracy were used to evaluate the performance of the YOLO models on the VisDrone-DET. Precision is the ratio of true positive cases to the total number of cases classified as positive within the entire sample. Recall measures the proportion of actual positive cases that are accurately detected among all the true positive cases. AP measures the area under the precision-recall curve. This curve plots the precision (true positive rate) against the recall (sensitivity) as the detection threshold varies.

III. Discussion

From the result presented in Figs. 1 and 2, YOLOv10 consistently outperforms both YOLOv8 and YOLOv9 across all computational levels. YOLOv10 achieves the highest mAP, reaching approximately 26%, indicating its superior accuracy in object detection task on the VisDrone-DET dataset. The graph also reveals that as the GFLOPS increase, YOLOv10 maintains a steeper upward trend in mAP compared to other models, reflecting its efficiency in leveraging computational resources. In contrast, YOLOv8 and YOLOv9 exhibit more gradual improvements in mAP, with YOLOv8 peaking around 22% and YOLOv9 around 23%. This suggests that YOLOv10 is not only more accurate but also more effective in utilizing higher computational power, making it robust and scalable option for high-performance object detection.

IV. Conclusion

In conclusion YOLOv10 offers superior performance in terms of mAP across its varying model complexities. The significant gains in mAP particularly at higher GFLOPS levels, highlights its efficiency and effectiveness in utilizing computational resources to achieve a higher accuracy. Further experiments are recommended on more complex datasets such as the small object detection dataset(SODA), which includes SODA-D and SODA-A, focusing on driving and aerial scenarios respectively.

V. Acknowledgment

This research was supported by the Mobilus, Korea, under the KNU(Kyunpook National University) support program(202416240000)

References

[1] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics yolov8," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[2] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "Yolov9: Learning what you want to learn using programmable gradient information," 2024. [Online]. Available: https://arxiv.org/abs/2402.13616

[3] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han, and G. Ding, "Yolov10: Real-time end-to-end object detection," 2024. [Online]. Available: https://arxiv.org/abs/2405.14458

[4] Y. Quan, P. Wang, Y. Wang, and X. Jin, "Gui-based yolov8 license plate detection system design," in *2023 5th International Conference on Control and Robotics (ICCR)*, 2023, pp. 156–161.

[5] P. Zhu, L. Wen, D. Du, X. Bian, H. Fan, Q. Hu, and H. Ling, "Detection and tracking meet drones challenge," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 11, pp. 7380–7399, 2021.