A Comparative Experimental Evaluation of YOLOv5, YOLOv7, and YOLOv8 for Object Detection

Airuddin Ahmad[1*] and Haw Ying Song[2]

[1*] Universiti Poly-Tech Malaysia
Jalan 6/91, Taman Shamelin Perkasa, 56100 Cheras, Kuala Lumpur, Malaysia
airuddin@uptm.edu.my

[2] Universiti Terbuka Malaysia
No 559, Jalan Langgar 05000 Alor Setar Kedah, Malaysia
easonhaw@oum.edu.my

Abstract. The purpose of this comparative study is to investigate the differences between the YOLOv5, YOLOv7, and YOLOv8 when detecting specific objects. Therefore, this study will focus on examining the precision performance, recall capabilities, IoU performance, and the balance between precision and recall of YOLOv5, YOLOv7, and YOLOv8 compared under systematic evaluation. This study was conducted by employing an experimental design. The experimental setup will be employed by using Google Colab as the platform, and the Google Colab Runtime type is Python 3 and T4 GPU. Additionally, the dataset is selected from the Roboflow Public Dataset, a total of 2781 images were selected. The results indicate that YOLOv5 is better performing in precision, recall, and the balance between precision and recall. Additionally, YOLOv8 is better in the value of Intersection over Union (IoU) performance. This study contributes insights that can have implications for future research, the education sector, institutions, technical optimization, and industrial development. Future researchers can use this study as a reference to delve into more diverse and in-depth studies on object detection algorithms.

Keywords: YOLOv5, YOLOv7, YOLOv8, Comparative Study

1 Introduction

Object detection is a crucial aspect of computer vision, focusing on localizing and classifying objects within images (Airuddin et al., 2024; Kundu, 2023). It is more complex than simple image classification because it handles multiple objects in a single image. According to Chen et al. (2024), object detection models utilize various techniques from computer vision, machine learning, and deep learning to identify patterns and features of different object classes. These models provide comprehensive annotations, including bounding boxes and labels for identified objects, such as traffic lights, vehicles, and people. Object detection is fundamental to AI vision techniques like image classification, retrieval, and co-segmentation, enabling advanced

N. A. Ishak et al. (eds.), *Proceedings of the International Conference on Cross- Disciplinary Academic Research 2025 - Track 1 Advances in Computing, Electronics, Engineering, and Mathematics (ICAR-T1 2025)*, Advances in Engineering Research 296,
https://doi.org/10.2991/978-94-6239-636-4_2

applications such as streamlined delivery services, security systems, and self-driving vehicles.

Despite the widespread use and adaptability of object detection, diverse theoretical perspectives exist among professionals regarding the effectiveness of different YOLO (You Only Look Once) models, with varying opinions on the superiority of YOLOv5, YOLOv7, or YOLOv8. Previous research often focused on older YOLO iterations, creating a gap in the systematic evaluation of the latest models (YOLOv5, YOLOv7, and YOLOv8). There is a continuous demand for improved precision, recall, and a delicate balance between these metrics in object detection algorithms. Furthermore, while Intersection over Union (IoU) is a critical metric for assessing prediction accuracy, fewer studies specifically examine IoU performance among YOLOv5, YOLOv7, and YOLOv8. There is also ambiguity regarding the trade-offs between accuracy, recall, and IoU performance across these models.

While numerous object detection models exist, this study focuses on YOLOv5, YOLOv7, and YOLOv8 because recent literature indicates a lack of systematic and consistent comparative evaluations among these latest YOLO versions. Concentrating on these models enables a focused investigation into performance trade-offs within a single, evolving object detection framework.

This study aims to conduct a comparative analysis of YOLOv5, YOLOv7, and YOLOv8 to address the identified research gaps. The specific objectives include investigating and comparing the precision performance, recall capabilities, IoU performance, and the balance between precision and recall of these three algorithms under systematic evaluation. Correspondingly, the research questions explore how these metrics compare among YOLOv5, YOLOv7, and YOLOv8.

This study formulated four specific hypotheses to guide its comparative analysis of YOLOv5, YOLOv7, and YOLOv8 object detection algorithms:

- **Precision Performance**

It was hypothesized that YOLOv5 would generally exhibit better precision performance when compared to YOLOv7 and YOLOv8 under systematic evaluation.

- **Recall Capabilities**

The study hypothesized that YOLOv5 would generally demonstrate better recall capabilities than both YOLOv7 and YOLOv8 under systematic evaluation.

- **Intersection over Union (IoU) Performance**

It was hypothesized that YOLOv8 would generally show superior Intersection over Union (IoU) performance compared to YOLOv5 and YOLOv7 under systematic evaluation.

- **Balance Between Precision and Recall**

The final hypothesis posited that YOLOv8 would generally achieve a better balance between precision and recall when compared to YOLOv5 and YOLOv7 under systematic evaluation.

These hypotheses were designed to investigate the distinct performance characteristics of each YOLO model across key object detection metrics.

2 Literature Review

2.1 Precision Performance

The review reveals conflicting findings regarding the precision performance of the YOLO models. Gašparović et al. (2023) and Olorunshola et al. (2023) found YOLOv5 to exhibit the highest overall detection accuracy and superior precision compared to YOLOv7. Specifically, Olorunshola et al. (2023) reinforced YOLOv5's superiority over YOLOv7 in mAP@0.5 and mAP@0.5:0.95.

In contrast, He et al. (2023) identified YOLOv8 as the most adept model, achieving an 80% precision rate, with YOLOv7 and YOLOv5 following at 78.6% and 78.1% respectively. The paper notes that these differing conclusions stem from variations in experimental images and environments.

Research by Olorunshola et al. (2023) and Yadav et al. (2023) indicated that YOLOv7 demonstrated superior recall capabilities compared to YOLOv5 and YOLOv8. Olorunshola et al. (2023) reported YOLOv7 with a recall value of 0.564 against YOLOv5's 0.534. However, a study by Sary et al. (2023) focusing on human detection in aerial imagery found YOLOv5 to have superior recall performance (0.7594) compared to YOLOv8 (0.7540). The paper identifies a minor research gap due to insufficient evidence to conclusively assert YOLOv7's recall capabilities as superior across all contexts.

2.2 Intersection over Union (IoU) Performance

The Intersection over Union (IoU) ratio is highlighted as a crucial threshold for assessing the accuracy of object detection predictions by quantifying the overlap between predicted and ground truth bounding boxes.

Adegun et al. (2023) found YOLOv7 to have the highest mAP at 50-95% IoU (25%), surpassing YOLOv5 (18.4%) and YOLOv8 (17.5%). Conversely, Liu et al. (2023) and Sama et al. (2023) presented findings where YOLOv8 performed better in mAP 50:95, particularly in complex traffic scenarios and simulated environments. Sama et al. (2023) reported YOLOv8 at 87.8% mAP50-95, compared to YOLOv5 at 82.2% and YOLOv7 at 61.8%. The review notes the inconsistency in findings and emphasizes the need for comprehensive evaluation across a broader spectrum of IoU thresholds.

2.3 Balance Between Precision and Recall (F1-Score)

Achieving a harmonious balance between precision and recall is crucial for object detection algorithms.

Sary et al. (2023) and Rahman et al. (2023) suggested YOLOv8 exhibits superior equilibrium, with YOLOv8 outperforming YOLOv5 in F1-Score (0.7974 vs. 0.7876). Rahman et al. (2023) specifically noted YOLOv8's F1-score of 0.96, outperforming YOLOv5 (0.92) and YOLOv7 (0.89). However, Thakuria et al. (2023) found a modified YOLOv7 (YOLOv7_improved) to achieve an impressive F1-score of 96.9%, surpassing both YOLOv8 (95.9%) and YOLOv5 (95.8%) in their specific evaluations. The review concludes that the disparate outcomes underscore the necessity for a systematic assessment of this metric.

In conclusion, the existing variability and often contradictory findings in comparative studies of YOLOv5, YOLOv7, and YOLOv8 across key performance metrics. Although prior research has examined the performance of YOLOv5, YOLOv7, and YOLOv8, the lack of consistency in experimental design and evaluation criteria has resulted in fragmented conclusions. In particular, the trade-offs between precision, recall, and IoU performance remain insufficiently clarified. Therefore, a comprehensive and systematic evaluation within a single experimental framework is necessary to provide clearer and more reliable insights.

3 Methodology

This study employs a quantitative research approach, specifically an experimental study design, to conduct a comparative analysis of YOLOv5, YOLOv7, and YOLOv8 object detection models. The primary goal is to examine their precision performance, recall capabilities, and Intersection over Union (IoU) distinctions, as well as the balance between precision and recall. Experimental research is considered the most suitable method for systematically investigating and elucidating causal relationships between applied variables and their effects.

3.1 Research Instruments and Environment

The experimental setup for this study utilizes Google Colab as the primary platform for analysing the precision, recall, mAP 50, and mAP 50-95 of the YOLO models (Kundu et al., 2023). Google Collaboratory, or "Colab," is a cloud-based Jupyter notebook environment that facilitates machine learning and AI experimentation through Python code execution. The specific runtime environment used is Python 3 and a T4 GPU. The experiments were conducted on an HP Pavilion 15 au102-tx, accessing Google Colab via Google Chrome Version 120.0.6099.129.

3.2 Data Collection

For object detection, a dataset of 2781 images were sourced from the Roboflow Public Dataset. Roboflow Public Datasets are known for providing comprehensive details and sufficient images for experimental research. The images selected specifically feature cats in various indoor and outdoor settings, at different distances, to thoroughly test the models' animal detection capabilities. Using a single-class object dataset allows the comparative analysis to focus on model performance differences without the confounding effects introduced by multi-class imbalance or class-specific complexity. The dataset underwent preprocessing, including Auto-Orient and resizing to a standardized 416 x 416 (width x height) resolution, which aligns with the specifications for YOLOv5, YOLOv7, and YOLOv8. The 2781 cat images were divided into three subsets: 87% (2433 images) for training, 8% (232 images) for validation, and 4% (116 images) for testing. All YOLO models were trained for 20 epochs using default hyperparameter settings recommended by their respective official implementations, including learning rate, batch size, and optimizer configuration, to ensure a fair and consistent comparison across models.

3.3 Data Analysis

To effectively analyze the findings, the study employed Mean Average Precision (mAP) and F1-score as key metrics. mAP assesses model efficacy in object detection and information retrieval, deriving its evaluation from sub-metrics like Confusion Matrix, IoU, Recall, and Precision (Kanstrén et al., 2020). The F1-Score is a composite metric that integrates both precision and recall, calculated as their harmonic mean. IoU is crucial for computing F1 or mAP, quantifying the spatial overlap between predicted and ground truth bounding boxes. The study utilized five types of information metrics: Precision, Recall, mAP 50, mAP 50-95, and F1 Score, to evaluate the optimal performance of each YOLO model across these criteria.

In essence, the methodology outlines a systematic experimental approach using established tools and datasets to rigorously compare the performance of YOLOv5, YOLOv7, and YOLOv8 on specific object detection tasks, focusing on key performance indicators.

4 Findings

The performance of YOLOv5, YOLOv7, and YOLOv8 on Google Colab within the context of this study. From the focus on precision, recall capabilities, and Intersection over Union (IoU), the inquiry seeks to offer a comprehensive understanding of these algorithms' efficacy in object detection.

This study also delves into the delicate equilibrium between precision and recall within the contextual domains of these 3 YOLO models. Therefore, this study will contribute valuable insights into the strengths and limitations of YOLOv5, YOLOv7, and YOLOv8, enhancing the discourse on these advanced models. Following is an example for Table 1.

Table 1. The Overall Performance of YOLOv5, YOLOv7 and YOLOv8

	YOLOv5	**YOLOv7**	**YOLOv8**
Precision	0.996	0.605	0.97
Recall	0.987	0.736	0.973
mAP 50-95	0.714	0.208	0.744
F1 Score	0.99	0.66	0.97

The findings of this research are:

- **Precision Performance**: YOLOv5 demonstrated the highest precision performance (precision = 0.996, mAP 50 = 0.995), followed by YOLOv8 (precision = 0.97, mAP 50 = 0.984). YOLOv7 exhibited the lowest precision performance (precision = 0.605, mAP 50 = 0.527).
- **Recall Capabilities**: YOLOv5 showed the best recall capabilities (recall = 0.987), followed by YOLOv8 (recall = 0.973). YOLOv7 had the lowest recall capabilities (recall = 0.736).
- **Intersection over Union (IoU) Performance**: YOLOv8 achieved the highest IoU performance (mAP 50:95 = 0.744), followed by YOLOv5 (mAP 50:95 = 0.714). YOLOv7 had the lowest IoU performance (mAP 50:95 = 0.208).
- **Balance Between Precision and Recall**: YOLOv5 performed the best in balancing precision and recall (F1 score = 0.99), followed by YOLOv8 (F1 score = 0.97). YOLOv7 ranked the lowest in this aspect (F1 score = 0.66).

The superior precision and recall achieved by YOLOv5 may be attributed to its relatively stable and mature architecture, which has been extensively optimized for general object detection tasks. Its balanced network depth and feature aggregation mechanisms likely contribute to more accurate bounding box predictions and reduced false positives in single-object detection scenarios. YOLOv8's higher IoU performance can be explained by architectural refinements and improved loss functions that enhance bounding box regression accuracy. These improvements enable tighter alignment between predicted and ground truth bounding boxes, particularly in well-defined object structures such as cats. The comparatively lower performance of YOLOv7 across several metrics may be influenced by its more complex architectural design, which, while powerful in certain contexts, may require longer training durations or more diverse datasets to fully realize its performance potential. The observed trade-off between precision–recall performance and IoU scores suggests that models optimized for accurate object classification may not always produce the most precise bounding box localization, highlighting inherent design trade-offs among different YOLO versions. The use of a single-class dataset with consistent object characteristics likely favored models that perform well under controlled conditions, which may partially explain YOLOv5's strong precision and recall results compared to models designed for broader, multi-class generalization. These findings indicate that performance differences among YOLOv5, YOLOv7, and YOLOv8 are not solely determined by model version, but are also shaped by architectural design choices, training configurations, and dataset characteristics.

5 Conclusion and Future Study

This study conducted a comparative analysis of YOLOv5, YOLOv7, and YOLOv8 object detection algorithms, confirming distinct differences among them. The research validated the effectiveness of YOLOv5 in precision, recall, and the balance between these two metrics. It also observed that YOLOv8 generally exhibited superior Intersection over Union (IoU) performance compared to YOLOv5 and YOLOv7 under systematic evaluation.

5.1 Performance Metrics Breakdown

- **Precision Performance:** YOLOv5 demonstrated the best precision performance (0.996 precision, 0.995 mAP 50), followed by YOLOv8 (0.97 precision, 0.984 mAP 50). YOLOv7 exhibited the least favourable precision performance (0.605 precision, 0.527 mAP 50). This finding aligns with previous studies by Gašparović et al. (2023) and Olorunshola et al. (2023).

- **Recall Capabilities:** YOLOv5 showed the most robust recall capabilities (0.987 recall), positioning it at the forefront. YOLOv8 secured the second position (0.973 recall), while YOLOv7 demonstrated comparatively diminished efficacy (0.736 recall). This supports the hypothesis that YOLOv5 generally has better recall capabilities.

- **Intersection over Union (IoU) Performance:** YOLOv8 attained the highest mean Average Precision (mAP) at the 50-95 index (0.744), indicating superior IoU performance. YOLOv5 followed as the second-best (0.714 mAP 50-95), and YOLOv7 lagged significantly (0.208 mAP 50-95). This confirms the hypothesis that YOLOv8 generally has better IoU performance.

- **Balance Between Precision and Recall (F1-Score):** YOLOv5 was identified as the optimal performer in the balance between precision and recall, achieving an F1-score of 0.99. YOLOv8 followed closely with an F1-score of 0.97, while YOLOv7 showed the least favorable performance with an F1-score of 0.66. Contrary to some prior literature suggesting YOLOv8's superiority in this aspect, this study found YOLOv5 to maintain the best balance. Consequently, the initial hypothesis that YOLOv8 would have a better balance was rejected.

5.2 Limitations and Recommendations

Limitations of the study include its exclusive focus on YOLOv5, YOLOv7, and YOLOv8, without comparing them to a broader range of object detection algorithms like Faster-CNN and SSD. The evaluation was also confined to cat objects, leaving performance on other object types unexplored. Furthermore, the study's assessment was limited by a short testing duration of 20 epochs, with implications of longer epochs remaining unknown. Recommendations for future research include incorporating a wider array of object detection algorithms, diversifying the types of objects used for testing, and extending the duration of epochs for more robust validation.

Acknowledgement

The authors would like to express their sincere appreciation to Universiti Poly-Tech Malaysia (UPTM) and Universiti Terbuka Malaysia (OUM) for their continuous academic and technical support throughout the completion of this research. Special thanks are extended to the developers and maintainers of the Roboflow Public Dataset and the Google Colab platform, whose resources and tools greatly facilitated the experimental work conducted in this study.

The authors are also grateful to colleagues and reviewers who provided valuable feedback and constructive suggestions that helped improve the quality and clarity of this paper. Their insights have significantly contributed to strengthening the research design, analysis, and presentation of findings.

Finally, the authors acknowledge the broader research community in computer vision and deep learning, whose ongoing innovations in object detection frameworks such as YOLOv5, YOLOv7, and YOLOv8 continue to inspire and advance scientific exploration in this field.

References

Adegun, A. A., Fonou, D. J.V., Viriri, S., & Odindi, J. (2023). State-of-the-Art Deep Learning Methods for Objects Detection in Remote Sensing Satellite Images. Sensors 2023, 23, 5849.

Airuddin, A., Othman, N., & Westergren, A. (2024). Utilizing Classroom Application of Facial Recognition Techniques through Internet of Things (IOT) Devices. International Journal, 5(1), 124-131.

Chen, W., Luo, J., Zhang, F., & Tian, Z. (2024). A review of object detection: Datasets, performance evaluation, architecture, applications and current trends. Multimedia Tools and Applications, 83(24), 65603-65661.

Gašparović, B., Mauša, G., Rukavina, J., & Lerga, J. (2023). Evaluating YOLOV5, YOLOV6, YOLOV7, and YOLOV8 in Underwater Environment: Is There Real Improvement. 2023 8th International Conference on Smart and Sustainable Technologies (SpliTech), Split/Bol, Croatia, 2023, pp. 1-4.

He, C. X., & Saha, P. (2023). Investigating YOLO Models Towards Outdoor Obstacle Detection For Visually Impaired People. ArXiv (Cornell University).

Kanstrén, T. (2020). A look at precision, recall, and F1-score. Towards Data Science, 12(10).

Kundu, R. (2023). YOLO: Algorithm for Object Detection Explained. V7 Labs. Retrieved from https://www. v7labs. com/blog/yolo-object-detection.

Olorunshola, O. E., Irhebhude, M. E., & Evwiekpaefe, A. E. (2023). A Comparative Study of YOLOv5 and YOLOv7 Object Detection Algorithms. Journal of Computing and Social Informatics, 2(1), 1–12.

Rahman, S., Rony, J. H., Uddin, J., & Samad, M. A. (2023). Real-Time Obstacle Detection with YOLOv8 in a WSN Using UAV Aerial Photography. J. Imaging 2023, 9, 216.

Sama, A. K., & Sharma, A. (2023). Simulated uav dataset for object detection. In ITM Web of Conferences (Vol. 54, p. 02006). EDP Sciences.

Sary, I. P., Andromeda, S., & Armin, E. U. (2023). Performance Comparison of YOLOv5 and YOLOv8 Architectures in Human Detection using Aerial Images. Ultima Computing: Jurnal Sistem Komputer, 15(1), 8–13.

Yadav, D., Upadhyay, P., Fatima, H., Jain, S., & Avinash, K. S. (2023). American Sign Language Detection using YOLOv5 and YOLOv8. Research Square (Research Square).

Thakuria, A., & Erkinbaev, C. (2023). Improving the network architecture of YOLOv7 to achieve real-time grading of canola based on kernel health. Smart Agricultural Technology, 5, 100300.