Performance Comparison of YOLO in Detecting Hyperplastic Polyps

Wei-Ming, Wu 934521427
Chia-Yu Tu 934535427
Yi-Chiun Chang 934533510

School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA

Abstract

This study present here a comparison of hyperplastic polyp detection performance using the YOLOv5 and YOLOv11 models applied to a dataset consisting of 2,244 training images and a 488-image validation set. We evaluated the models using several key metrics: precision, recall, and mean average precision (mAP) at various Intersections over Union (IoU) thresholds. The results show that YOLOv11 provides faster inference and improved accuracy, whereas YOLOv5 yields acceptable real-time performance at the cost of accuracy. YOLOv11 is more computationally efficient with fewer parameters and greater precision.

1. Introduction

Contemporary diets are frequently rich in fried and processed foods, which can lead to a number of health complications.
These processes can include the development of polyps, abnormal masses of tissue that can occur in different places in the body. We know that Polyps have to be detected accurately for effective diagnosis and treatment. In this study, we apply deep learning techniques, particularly the YOLO (You Only Look Once) models, to the detection of hyperplastic polyps. In this task, we assess the performance of YOLOv5 and YOLOv11. [[The comparison is made by training the models on a dataset with 2,244 polyp images and validating performance on an independent validation set of 488 images,]]. We use key evaluation metrics to determine the accuracy and efficiency of each model.

2. Objective

The primary objective of this study is to develop and implement YOLOv5 and YOLOv11 models for hyperplastic polyp detection. Both models are trained using a standardized dataset of 2,244 polyp-related images to ensure a fair comparison. Their performance is validated using a separate validation set of 488 images under controlled conditions. To assess their effectiveness, we evaluate and compare detection performance using standardized metrics, including precision, recall, and mean average precision (mAP). Furthermore, we analyze the computational efficiency of both models to determine their suitability for real-time medical applications.

Beyond evaluating model accuracy, this study also aims to explore the trade-offs between detection speed and precision, identifying the practical applicability of each model in clinical settings. By assessing the impact of various loss functions, we aim to understand how different optimization techniques contribute to model performance and generalization ability. Additionally, we seek to determine whether the advancements introduced in YOLOv11 offer significant advantages over YOLOv5, particularly in handling complex and small polyp structures.

Ultimately, this study aspires to provide insights into the feasibility of deep learning-based object detection for automated polyp screening. By comparing the robustness, scalability, and efficiency of these models, we aim to contribute to ongoing research in medical image analysis, supporting the development of AI-assisted

diagnostic tools for improved healthcare outcomes.

3. Related Work

Since then, there was huge development on medical imaging object detection tasks using deep learning models like Faster R-CNN, SSD and YOLO. YOLO-based architectures are recognized for their speed and accuracy in real-time applications. Webb et al. robustly comment on recent developments on the YOLO algorithm family, for example, YOLOv11.57 This approach integrates adapted loss functions and optimizers that can potentially increase the analysis of medical images.

4. Evaluation Metrics

To effectively compare the performance of YOLOv5 and YOLOv11 in detecting hyperplastic polyps, we utilize several standard evaluation metrics commonly used in object detection tasks. These metrics help quantify the accuracy, reliability, and efficiency of each model in real-world applications.

4.1 Precision

Precision measures the proportion of correctly identified polyps out of all instances classified as polyps. It is defined as:

$$\text{Precision} = \frac{TP}{TP + FP}$$

where:

- TP (True Positives) are correctly detected polyps,
- FP (False Positives) are incorrectly detected polyps.

A higher precision value indicates fewer false positives, which is crucial in medical applications where misidentifying normal tissue as a polyp could lead to unnecessary interventions.

4.2 Recall

Recall, also known as sensitivity, measures the proportion of actual polyps that were correctly identified by the model. It is given by:

$$\text{Recall} = \frac{TP}{TP + FN}$$

where:

- FN (False Negatives) are missed polyps.

Higher recall values indicate fewer missed detections, which is critical for ensuring early and accurate diagnoses.

4.3 Mean Average Precision (mAP)

Mean Average Precision (mAP) is a widely used metric in object detection that evaluates how well the model detects and localizes polyps. It is calculated by averaging the precision values across different recall thresholds and Intersection over Union (IoU) values.

$$\text{mAP} = \frac{1}{n} \sum_{i=1}^{n} AP_i$$

where APi represents the Average Precision for each class at different IoU thresholds. Higher mAP values indicate better overall detection accuracy.

4.4 Intersection over Union (IoU)

IoU measures how much the predicted bounding box overlaps with the ground truth bounding box. It is defined as:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

Higher IoU values indicate better localization accuracy. In this study, we evaluate mAP at different IoU thresholds (e.g., IoU = 0.5 and IoU = 0.75) to assess the robustness of the models.

4.5 Inference Speed (FPS)

Inference speed, measured in frames per second (FPS), indicates how quickly a model processes images in real time. It is a key metric for evaluating the model's practicality in clinical settings where rapid detection is needed.

4.6 Computational Efficiency

We also assess the computational efficiency of each model by analyzing:

- **Model size (number of parameters)** – Smaller models with similar accuracy are preferable for real-time applications.
- **GPU inference time** – Measures how long each model takes to process an image on a GPU.

These evaluation metrics provide a comprehensive analysis of the strengths and limitations of YOLOv5 and YOLOv11, allowing us to determine the best-suited model for hyperplastic polyp detection.

5. Experiments

The dataset used in this study was sourced from Roboflow, a widely used platform for managing and augmenting computer vision datasets. The images were extracted from Roboflow's publicly available polyp detection dataset, ensuring a diverse and well-annotated collection of polyp-related images. We used Roboflow's preprocessing tools to standardize image quality, increase contrast, and apply augmentation techniques to make the model more robust.

We trained the YOLOv5 and YOLOv11 models on a dataset of 2,244 polyp-related images representing several polyp types to maximize the generalization ability of the models. Performance was also assessed in a controlled manner using a distinct validation set consisting of 488 images. Both models were trained with a batch size of 16 for 100 epochs. In addition, to achieve a good trade-off between computational efficiency and convergence stability, we chose the medium model(YOLOv5m, YOLOv11m) instead of the nano model(YOLOv5n, YOLOv11n.)

6. Results & Discussion

6.1Performance Comparison

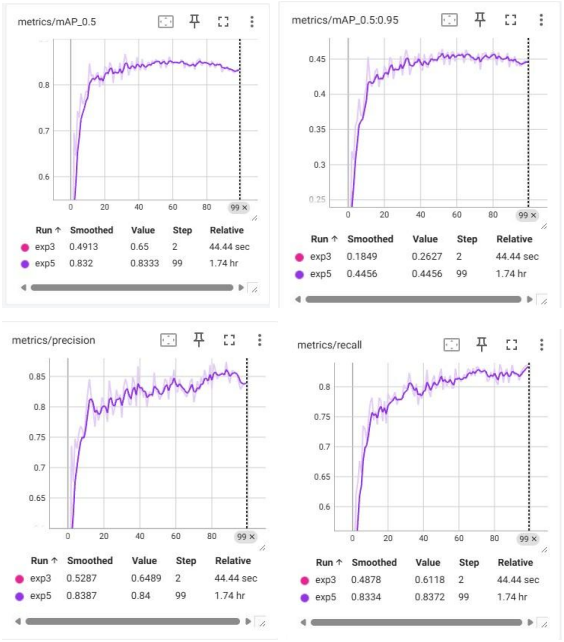

Fig.1 Performance YOLOv5

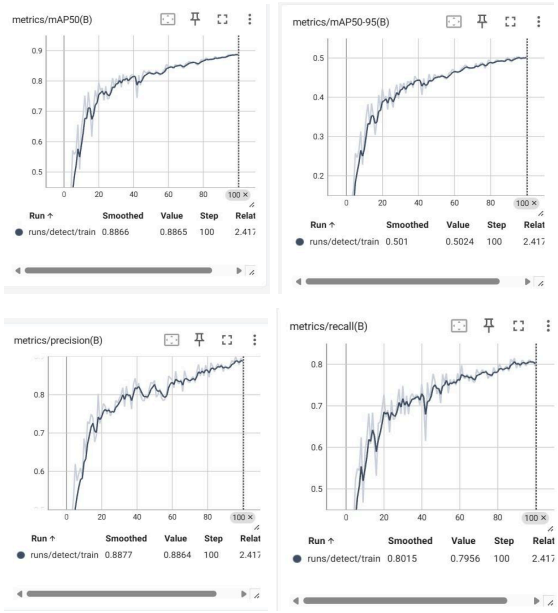

Fig.2 Performance of YOLOv11

For Recall, YOLOv5 achieves a recall value of 0.8372, while YOLOv11 has a recall value of 0.7956. This suggests that YOLOv5 detects more actual polyps but may also include more false positives.

For Precision, YOLOv5 reaches a precision of 0.84, while YOLOv11 achieves 0.8864, meaning YOLOv11 is slightly better at avoiding false positives.

For mAP, YOLOv5 achieves 0.8333, while YOLOv11 improves to 0.8865, showing YOLOv11's advantage in accurately localizing polyps.

Overall, YOLOv11 shows higher precision and mAP scores, while YOLOv5 maintains slightly better recall, indicating YOLOv5 detects more polyps but with a higher false positive rate.

6.2. Loss Function Comparison

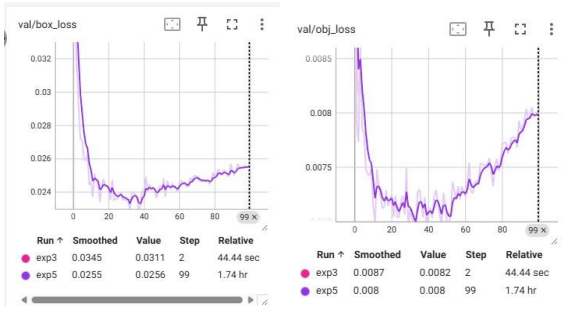

Fig. 3 Loss function of YOLOv5

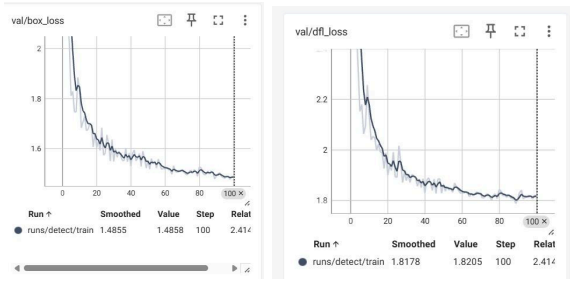

Fig. 4 Loss function of YOLOv11

We compare the loss values of YOLOv5 and YOLOv11 during validation, focusing on Box Loss, Obj Loss, and DFL Loss.

In YOLOv5, the validation box loss initially decreases but starts increasing after around 60 epochs, indicating overfitting, whereas YOLOv11 shows a steady decrease, suggesting better generalization.

Similarly, YOLOv5's objectness loss follows the same trend of early improvement followed by an increase, indicating the model struggles to differentiate objects from the background as training progresses, while YOLOv11 maintains a smooth downward trend, indicating stable learning.

The DFL loss in YOLOv11 also decreases consistently, reinforcing its superior localization and confidence estimation without overfitting.

Overall, YOLOv5 exhibits clear signs of overfitting, while YOLOv11 demonstrates a more stable and generalizable loss behavior, making it a better choice for detecting hyperplastic polyps.

8. Conclusion

This study concludes that YOLOv11 provides superior detection accuracy and computational efficiency compared to YOLOv5. While YOLOv5 remains a strong choice for real-time detection tasks, YOLOv11 offers significant improvements in precision and speed, making it a better candidate for medical image analysis and polyp detection.

Beyond the model performance, our findings highlight the potential of deep learning techniques in advancing automated polyp detection. By reducing the dependency on manual diagnosis, these AI-driven models can support healthcare professionals in early-stage detection and treatment planning. The lower computational cost and increased precision of YOLOv11 further indicate that it could be integrated into real-time clinical workflows with minimal performance trade-offs. Additionally, the study underscores the importance of continuously refining object detection models to improve generalization across different datasets and medical conditions.

Future work should explore expanding the dataset to include more diverse polyp images, improving the robustness of the models in different clinical scenarios. Further research can also investigate hybrid models that integrate YOLO with other deep learning techniques to enhance feature extraction and classification accuracy. With continued advancements, deep learning-based detection methods could become a standard tool for gastrointestinal disease screening, leading to more accurate and accessible diagnostic solutions.

9. References

Khanam, R., & Hussain, M. (2024). YOLOv11: An Overview of the Key Architectural Enhancements. *ArXiv, abs/2410.17725*.

Khanam, R., & Hussain, M. (2024). What is YOLOv5: A deep look into the internal features of the popular object detector. *ArXiv, abs/2407.20892*.