Performance Analysis of YOLOv8: A Robust Approach for Aerial Object Detection

Fussy Mentari Dirgantara, Ramdhan Nugraha, *Soo Young Shin
Department of IT Convergence Engineering,
Kumoh National Institute of Technology (KIT), Gumi, South Korea
Email: {dirgantara, ramdhan, *wdragon} @kumoh.ac.kr

Abstract

This research investigated the object detection capabilities of the YOLOv8 model on the VisDrone2021 dataset. We trained the model on aerial video sequences encompassing diverse weather conditions and complex scenarios. Our analysis focused on evaluating precision, recall, and mean Average Precision (mAP) scores, potentially incorporating an ablation study to assess hyperparameter tuning and training parameter impacts. We may also benchmark against other model and discuss generalizability to real-world applications, acknowledging limitations and future research directions. This comprehensive evaluation will provide valuable insights into the effectiveness of the YOLOv8 model for aerial object detection tasks.

Ⅰ. Introduction

The burgeoning field of aerial object detection presents a multifaceted challenge, demanding both accuracy and robustness in the face of diverse and dynamic environments. Aerial video sequences, encompassing a multitude of object categories within intricate weather conditions and ever-shifting viewpoints, necessitate the development of sophisticated detection models capable of navigating these complexities.

Enter the YOLOv8 model, a cutting-edge architecture renowned for its swift and precise object detection capabilities. This research leverages the image detection system and analyzes the YOLOv8 on the VisDrone2021 dataset, a meticulously curated benchmark specifically designed for aerial object detection tasks [1].

Through careful analysis, this research delves into the YOLOv8n's performance on VisDrone2021. We analyze its ability to accurately localize and identify objects, employing established metrics such as precision, recall, and the coveted mean Average Precision (mAP). This comprehensive assessment aims to illuminate not only the strengths and limitations of the YOLOv8n model but also to contribute to the ongoing pursuit of robust and generalizable object detection solutions within the aerial domain.

Ⅱ. Method

The approach of our research consists of three distinct parts. The initial step entails the selection of a dataset. In the second phase, we utilize the YOLOv8 model to train on a dataset of 400 video clips, which are composed of 265,228 frames, as well as 10,209 static images [2]. Lastly, in the third section, we evaluate the effectiveness of the training model.

1. Dataset Selection

The first part of the research was data selection, to enhance the use of aerial object detection tasks, we need to use a specific dataset that is captured by various drone-mounted cameras. Choosing the right dataset for training and evaluating the YOLOv8 model for aerial object detection was important. We opted for the VisDrone2021 dataset, a meticulously curated aerial video collection specifically designed to mimic the complexities of real-world scenarios.

VisDrone2021 offers a rich tapestry of diverse scenarios [3]. Its sequences capture object categories like vehicles and pedestrians, amidst an ever-shifting kaleidoscope of weather conditions, lighting variations, and camera viewpoints. This variety ensures the model isn't simply collected on idealized situations, but also real-world aerial footage.

Ultimately, choosing VisDrone2021 as our training ground was a strategic decision, equipping the YOLOv8 model with the tools and experience it needs to excel in the demands of aerial object detection.

2. Training YOLOv8

The second part of this research focuses on training the YOLOv8 model using the chosen dataset. YOLOv8 has various models, from YOLOVv8n, s, m, l, and x. In this research, we focused on the YOLOv8n (stands for YOLOv8 nano) model which would be suitable for edge computing systems [4].

Hyperparameters—like batch size, learning rate, and optimizer—are carefully tuned to create the optimal learning environment. Data augmentation techniques, such as mosaic, enrich the dataset with variations, ensuring the model doesn't become overly reliant on specific data patterns.

During the training phase, the model engages in prediction and refinement. It receives image or video frames, extracting features through its backbone architecture and generating bounding boxes and class probabilities for detected objects. These predictions are compared to the ground truth labels, and any discrepancies trigger a loss calculation. This loss is then backpropagated through the network, guiding

adjustments to weights and biases, and fine-tuning the model's visual expertise.

Key metrics like loss, precision, recall, and mAP are monitored, providing insights into its progress and potential areas for improvement. Validation sets offer a crucial checkpoint, ensuring the model generalizes well to unseen data and doesn't succumb to overfitting. Visualization techniques allow researchers to observe its predictions firsthand, identifying any visual biases or errors.

3. Analyze the performance of the trained model

The third and final part of the study is to analyze the performance of the training model. YOLOv8n model training for 20 epochs and completed in 2.454 hours. The metrics result from this model can be seen in Figure 1 which shows that the model is slowly improving over the training. The final mAP50 value is 0.274, the mAP50-95 value is 0.157, the precision is 0.37, and the recall is 0.285.

The comparison of the train and validation loss is shown in Figure 2 which shows that all of the loss values are decreasing. The difference between the final predicted and true boxes loss of train and validation values are 1.471 and 1.431. The accuracy of the classes in each detection value is 1.177 for the train and 1.103 for validation. The object loss value of the train is 0.933 and for the validation is 0.935.

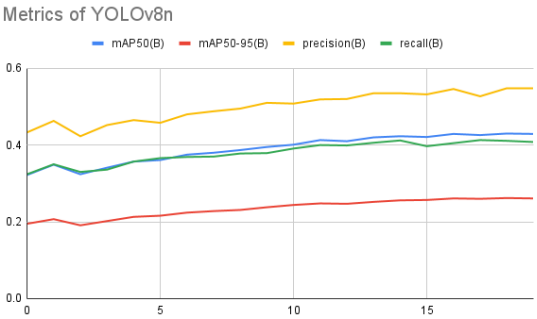

Fig 1. Performance Metrics

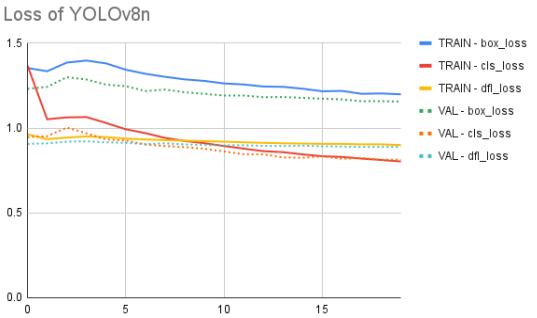

Fig 2. Loss Result of Trained Model

Fig 3. Comparison of the examples of the detection effect between YOLOv8n result (left) and YOLOv5nu result (right)

Figure 3 depicts the comparative analysis of the outcomes of YOLOv8n and YOLOv5nu. The outcome highlights the need to carefully choose the suitable model according to the specific demands of the aerial item detection task. The YOLOv8n model has been enhanced to effectively identify and locate smaller objects within the frame compared to the YOLOv5n model.

Ⅲ. Conclusion

In this study, we conducted an analysis of the YOLOv8 model. Especially in the YOLOv8n model, we find that the aerial image that is trained has better performance to detect smaller objects that are captured by the camera and will be suitable for real-life scenarios that need detailed information on the ground.

ACKNOWLEDGMENT

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2023-RS-2023-00259061) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation & Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education"(2018R1A6A1A03024003).

REFERENCES

[1] Wang, X., Yao, F., Li, A., Xu, Z., Ding, L., Yang, X., ... & Wang, S. (2023). DroneNet: Rescue Drone-View Object Detection. Drones, 7(7), 441.

[2] http://aiskyeye.com/.

[3] Sineglazov, V., & Kalmykov, V. (2021). Image processing from unmanned aerial vehicle using modified YOLO detector.

[4] https://github.com/ultralytics/yolov8.