Multi-Scale Object Detection using YOLOv8

Hanaanee Hana[1,✉]

[1]Department of Computer Science & Engineering, Faculty of Engineering, University of Moratuwa, Sri Lanka

Aerial object detection plays a crucial role in applications such as air traffic monitoring, defense systems, and disaster management. This paper explores multi-scale, anchor-free detection of aerial objects—specifically airplanes—using the YOLOv8 framework on the OpenImages V6 dataset. We incorporate multi-scale training strategies such as tiling, progressive resizing, and fine-tuning on aerial-related classes to enhance detection accuracy across different altitudes and object sizes. By leveraging YOLOv8's anchor-free architecture, our approach demonstrates improved localization of small and large-scale flying objects while maintaining computational efficiency. Experimental results are expected to indicate significant improvements in mean Average Precision (mAP), validating the effectiveness of the proposed multi-scale anchor-free training pipeline.

computer vision | multi-scale object detection | anchor-free detection | open images | YOLOv8 | airplane

Introduction

Aerial object detection remains a challenging area in computer vision due to scale variations, occlusions, and environmental noise. YOLOv8, being an anchor-free detector, eliminates the need for manually defined anchor boxes and provides better generalization across object sizes and shapes. While previous works have explored general-purpose detection or human detection on multiple datasets, research on multi-scale aerial detection remains limited especially when it comes to Open Images v6 using YOLOv8. In this study, we focus on improving airplane detection using YOLOv8 with multi-scale and tiling strategies. The selected classes (*Airplane, Helicopter, Drone, Bird, Rocket*) reflect a diverse aerial domain, making this an ideal benchmark to evaluate anchor-free detection performance. Our contributions include a two-stage multi-scale training pipeline and a tiling-based augmentation framework optimized for aerial imagery.

Methodology

We adopt a two-stage training process that combines multi-scale augmentation with fine-tuning for aerial objects:

Data Preparation:

- A subset of OpenImages V6 containing all 5 classes along with other background classes was filtered to retain aerial-related categories: *Airplane*, *Helicopter*, *Drone*, *Bird*, and *Rocket*.
- Each main class was limited to a maximum of 500 instances for balanced training.
- Images were optionally tiled into 640×640 patches with 20% overlap to improve detection of small-scale flying objects.
- The dataset was split into 80%, 10%, and 10% for training, validation, and testing respectively.

Training:

- **Stage 1:** Train a generalized YOLOv8 model on all 5 main classes with multi-scale augmentation to build robust feature representations.
- **Stage 2:** Fine-tune the model exclusively on the aerial classes using progressive resizing (from 640px to 1280px) to improve scale adaptability.
- Hyperparameters used include: learning rate of 0.01 for Stage 1 and 0.002 for Stage 2, momentum of 0.937, and weight decay of 0.0001.

Anchor-Free Detection:

- YOLOv8's anchor-free head was leveraged to predict bounding boxes directly using object center points and distributional focal loss (DFL).
- This reduces the complexity associated with anchor box tuning and improves detection consistency across varying scales.

Evaluation:

- Validation and test sets were used to compute standard COCO metrics: Precision, Recall, mAP@0.5, and mAP@0.5–0.95.
- Test-Time Augmentation (TTA) was applied to assess generalization under varying image scales.
- Dataset balance and class frequency were analyzed to understand bias across aerial categories.
- **Hyperparameter Optimization:**
  To identify optimal training parameters, a lightweight hyperparameter sweep was conducted using YOLOv8's built-in training function. The search space included variations in initial learning rate (`lr0`) and weight decay values. Each configuration was trained for 12 epochs on the tiled dataset and evaluated on the validation set to measure preliminary performance. The tested configurations are summarized in Table 1. The optimal configuration (`lr0 = 0.005`, `weight_decay = 0.0001`) demonstrated stable convergence and improved validation mAP, which was subsequently used for the final Stage 2 fine-tuning.

Table 1. Hyperparameter Sweep Configurations and Results

Run	lr0	Weight Decay	Precision	mAP@0.5
1	0.010	0.0001	0.816	**0.616**
2	0.005	0.0001	**0.817**	**0.616**
3	0.010	0.0100	0.799	0.604

Experimental Results

Dataset Statistics:

The aerial subset exhibited moderate class imbalance, with the *Airplane* and *Helicopter* classes having the highest instance counts. The tiling procedure significantly increased the number of effective training samples, especially improving the representation of small-scale objects.

Table 2. Validation Metrics (Without Test-Time Augmentation)

Metric	Precision	Recall	mAP@0.5	mAP@0.5–0.95
Value	0.742	0.632	0.653	0.407

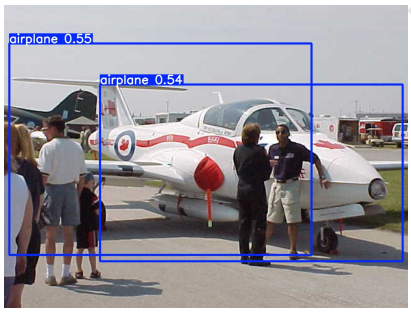
Table 3. Validation Metrics (With Test-Time Augmentation)

Metric	Precision	Recall	mAP@0.5	mAP@0.5–0.95
Value	**0.830**	0.634	**0.681**	**0.465**

Testing Results and Visual Predictions:

Qualitative analysis was performed on the test set to evaluate real-world detection performance. Sample predictions are illustrated in Fig. 1, showcasing the model's ability to localize diverse aerial targets across varying scales and backgrounds.

(a) Sample 1 – Multi-object aerial scene with varied scales.



(e) Sample 2 – Mixed aerial objects within complex landscape.

Fig. 1. Sample test predictions demonstrating accurate localization and classification across aerial scenes.

Observations:

- **Quantitative:** Test-Time Augmentation (TTA) improved both precision and mAP values, indicating enhanced robustness against viewpoint and scale variations.
- **Qualitative:** Visual inspection confirms consistent bounding box localization across densely populated aerial scenes and small object detections such as drones.
- **Training Strategy:** Progressive resizing led to smoother convergence and improved large-object recognition, while tiling provided a significant boost for small-object detection.
- **Architecture:** The anchor-free YOLOv8 architecture effectively generalized to multi-scale features without relying on manual anchor tuning.

Conclusion

This study demonstrates a robust anchor-free, multi-scale detection pipeline using YOLOv8 for aerial objects on the OpenImages V6 dataset. The combination of tiling, progressive resizing, and fine-tuning across aerial-related classes significantly improves model accuracy across varying object scales. The findings indicate the potential of anchor-free models for real-world aerial applications such as surveillance, traffic monitoring, and disaster response. Future work will explore cross-dataset generalization and lightweight deployment for edge-based drone systems.

hyperref

You can access the project notebook here: Project Notebook Link

References

1. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real Time Object Detection," in *Proc. CVPR*, 2016.
2. J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger," in *Proc. CVPR*, 2017.
3. A. Kuznetsova et al., "The Open Images Dataset V4: Unified image classification, object detection, and visual relationship detection at scale," *arXiv preprint arXiv:1811.00982*, 2018.
4. J. Lin et al., "Feature Pyramid Networks for Object Detection," in *Proc. CVPR*, 2017.
5. J. Terven and D. Cordova-Esparza, "A Comprehensive Review of YOLO Architectures in Computer Vision," *arXiv preprint arXiv:2304.00501*, 2023.
6. Z. Wang et al., "MFF-YOLO: An Improved YOLO Algorithm Based on Multi-Scale Feature Fusion," *TST*, vol. 28, 2025.
7. L. Wang et al., "A multi-scale small object detection algorithm SMA-YOLO for UAV imagery," *Sci. Rep.*, vol. 15, 2025.
8. S. Li et al., "PD-YOLO: A novel weed detection method based on multi-scale feature fusion," *Front. Plant Sci.*, vol. 16, 2025.
9. Z. Zhang et al., "An anchor-free object detector based on soften optimized bi-directional FPN," in *Proc. CVPR*, 2022.
10. D. Reis et al., "Real-Time Flying Object Detection with YOLOv8," *arXiv preprint arXiv:2305.09972*, 2024.
11. H. Tan et al., "EfficientDet: Scalable and Efficient Object Detection," in *Proc. CVPR*, 2020.
12. X. Li et al., "Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection," *arXiv preprint arXiv:2006.04388*, 2020.
13. X. Zhuang et al., "Multiscale semantic enhancement network for object detection," *Sci. Rep.*, vol. 13, no. 34277, 2023.
14. K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," *IEEE TPAMI*, vol. 37, 2014.
15. L. Liu et al., "Path Aggregation Network for Instance Segmentation," in *Proc. CVPR*, 2018.