Advanced Vision-Based Object Detection for Autonomous MAV Medicine Delivery

Subhramayee Swetapadma
School of Mechanical Science
Odisha University of Technology and Research
Bhubaneswar, India
subhramayee02@gmail.com

Rati Ranjan Dash
School of Mechanical Science
Odisha University of Technology and Research
Bhubaneswar, India
rrdash@outr.ac.in

***Abstract*— The increasing demand for efficient and autonomous healthcare logistics in remote and inaccessible regions has driven advancements in vision-based navigation for Micro Aerial Vehicles (MAVs). This research presents a fully vision-based object detection system for MAVs, eliminating the need for additional sensors such as LiDAR or GPS. The proposed system leverages the YOLOv8 model to enable real-time detection, target identification, and obstacle avoidance. A structured methodology, including dataset preparation, annotation, model training, and evaluation, ensures high accuracy and robust performance. The system achieves a mean Average Precision of 94.9% in multi-class detection and operates effectively in real-time environments. Experimental results demonstrate the feasibility of deploying a vision-only navigation framework for autonomous medicine delivery, reducing hardware complexity and operational costs. The proposed approach enhances scalability, making it suitable for broader applications in disaster relief, surveillance, and smart logistics.**

Keywords*— *Vision-based navigation, object detection, MAV, YOLOv8, autonomous medicine delivery.

I. Introduction

The delivery of essential healthcare services to remote and underserved regions remains a significant challenge due to infrastructural barriers and geographic obstacles [1]. Traditional logistics systems often struggle to provide timely and efficient healthcare access, particularly in areas with limited transportation networks. Micro Aerial Vehicles (MAVs) offer a promising solution by leveraging autonomous navigation and vision-based object detection to overcome these challenges [2], [3].

Vision-based navigation enables MAVs to detect objects, identify targets, and avoid obstacles without relying on additional sensors such as LiDAR or GPS. This eliminates hardware complexity and reduces operational costs, making autonomous MAVs a viable option for applications such as medicine delivery, disaster relief, and surveillance [4], [5]. The key to achieving efficient autonomous navigation lies in developing an accurate, real-time object detection framework tailored for MAVs operating in dynamic environments [6], [7].

This research focuses on designing and implementing a vision-only object detection system for MAVs using the YOLOv8 model. YOLOv8 is a state-of-the-art deep learning-based object detection framework that balances speed, accuracy, and computational efficiency, making it suitable for real-time applications [8], [9]. The system is trained using annotated datasets and optimised for robust performance in real-world conditions. Key objectives include enhancing detection accuracy, reducing computational overhead, and ensuring scalability for deployment in resource-constrained settings [10], [11].

The remainder of this paper is structured as follows: Section II presents a review of related work on vision-based object detection and MAV navigation. Section III outlines the methodology, including dataset preparation, model training, and evaluation metrics. Section IV discusses experimental results, including real-time detection performance and accuracy metrics and provides a detailed discussion of the findings, identifying challenges and areas for improvement. Finally, Section V concludes the paper with insights into future research directions and potential applications of the proposed framework.

II. Literature Review

A. Vision-Based Object Detection for MAV Navigation

The use of vision-based object detection in MAVs has gained significant attention due to its ability to navigate autonomously without relying on external sensors such as LiDAR or GPS. Traditional MAV navigation systems have primarily relied on sensor fusion, combining LiDAR, GPS, and IMUs for localization and obstacle detection [1]. However, these methods increase hardware complexity and cost, making them less suitable for resource-constrained applications.

Recent advancements in deep learning-based object detection models have enabled vision-only navigation, improving real-time target identification and obstacle avoidance. The YOLO (You Only Look Once) series has emerged as a leading framework due to its balance of speed and accuracy. Redmon et al. introduced YOLOv1 as a real-time object detection system that significantly reduced detection latency compared to region-based methods [2]. Over subsequent iterations, YOLOv8 has been optimised with anchor-free designs and decoupled detection heads, enhancing its performance in complex environments [3].

Some studies have demonstrated the feasibility of vision-based navigation for UAVs in healthcare logistics. Zou and Liu [4] reviewed vision-only systems for MAVs, emphasising their cost-effectiveness and adaptability. Similarly, Yu et al. [5] explored target tracking in UAV-based healthcare logistics using YOLO models, highlighting their effectiveness in identifying delivery points with high accuracy. Despite these advancements, challenges remain in optimising detection models for real-time deployment on MAVs with limited computational resources.

B. Real-Time Object Detection in UAVs

Real-time object detection is crucial for MAVs, as it ensures dynamic adaptability to environmental changes.

Single-shot detection frameworks such as SSD (Single Shot MultiBox Detector) [6] and EfficientDet [7] have been explored for UAV applications. SSD provides a tradeoff between speed and accuracy, while EfficientDet improves detection efficiency using neural architecture search.

Liu and Zhang [8] demonstrated real-time object detection in UAVs using YOLOv4 for agricultural monitoring. Their study confirmed YOLO's capability to maintain high detection accuracy while operating in real time. Similarly, Handa and Grabner [9] investigated UAVs for medical supply delivery, utilising object detection to identify landing zones and obstacles. These studies reinforce the importance of lightweight models that balance computational efficiency and detection performance.

C. *Challenges in Vision-Only Navigation*

Despite significant progress, vision-only navigation faces challenges such as occlusion handling, varying lighting conditions, and high computational demands. Zhao et al. [10] proposed a lightweight detection algorithm for UAV aerial imagery, addressing the issue of detecting small and complex objects in real-world scenarios. Their approach incorporated dynamic convolution techniques to improve feature extraction.

Furthermore, research by Sun and Li [11] explored optimisation strategies for lightweight object detection networks in UAVs, including model pruning and quantisation. Their findings suggest that reducing model complexity without sacrificing accuracy remains a key challenge in deploying vision-based systems on MAVs.

D. *Applications of Vision-Based Navigation*

The applications of vision-based navigation extend beyond healthcare logistics. UAVs equipped with vision-based detection have been utilised in disaster relief operations, surveillance, and precision agriculture. Wu and Yang [12] investigated UAV-based medicine delivery, demonstrating its potential to improve healthcare accessibility in remote areas. Additionally, Xu et al. [13] explored autonomous UAV navigation for smart logistics, highlighting the scalability of vision-based detection models in urban environments.

The literature emphasises the growing potential of vision-based MAV navigation in various domains. However, optimising object detection models for lightweight, real-time deployment remains a research priority. This study builds upon existing advancements by implementing a vision-only object detection system using YOLOv8, aiming to enhance real-time navigation performance while minimising hardware dependencies.

III. Methodology

The methodology for developing a vision-only object detection framework for MAVs using the YOLOv8 model is presented. The framework is designed to enable autonomous navigation without external sensors by incorporating dataset preparation, model training, real-time detection, and performance evaluation [1], [2].

A. *Dataset Preparation*

The dataset used for training consisted of images collected under diverse conditions to improve model robustness. Annotation was performed using the LabelImg tool, where bounding boxes were manually drawn around target objects [3]. To ensure consistency, all images were resized to 640×480 pixels. Data augmentation techniques, including flipping, rotation, and scaling, were applied to enhance generalisation and improve model performance in varying environments [4], [5].

B. *Model Training*

The YOLOv8 model was configured using data.yaml file, specifying dataset paths and class labels. The model was trained for 100 epochs with a batch size of 8 and an adaptive learning rate of 0.001 using the Stochastic Gradient Descent (SGD) optimiser [6]. The training process was conducted on an NVIDIA GPU to optimise computational efficiency, with loss reduction and accuracy monitored throughout the training phase [7], [8].

C. *Real-Time Object Detection*

For real-time detection, video frames were processed sequentially using OpenCV. The YOLOv8 model analysed each frame, generating bounding boxes and class labels. Detection confidence was controlled using a threshold of 0.2, while an Intersection over Union (IoU) threshold of 0.5 ensured accurate bounding box overlap [9]. The detected objects were visualised in real-time with annotated bounding boxes, and detection results were logged in CSV format for further analysis [10].

D. *Performance Evaluation*

The system's effectiveness was assessed using key performance metrics, including precision, recall, F1-score, and mean Average Precision (mAP) [11]. Precision and recall were used to evaluate detection accuracy, while the F1-score provided a balanced measure of model performance. The mAP50 and mAP50-95 metrics quantified the system's ability to detect objects across multiple IoU thresholds [12]. Confusion matrices were generated to identify misclassifications and analyse model reliability [13].

E. *Tools and Hardware*

The implementation utilised Ultralytics for model training, OpenCV for video processing, and Matplotlib for performance visualisation [14]. The model was trained on an NVIDIA GPU, while real-time detection was performed on an Intel CPU to ensure efficiency in deployment [15].

This methodology provides a structured approach to implementing a vision-only object detection framework for MAVs. By leveraging YOLOv8, the system achieves high accuracy, real-time detection, and computational efficiency, making it suitable for autonomous navigation applications. The next topic describes experimental results and validation of the proposed system.

IV. Results And Discussion

The experimental results and performance analysis of the vision-only object detection framework for MAVs are discussed. The evaluation focuses on detection accuracy, real-time processing efficiency, and system reliability under different environmental conditions. Performance metrics, including precision, recall, F1-score, and mean Average Precision (mAP), are used to assess the model's effectiveness [1], [2]. Additionally, confusion matrices, evaluation curves, and real-time detection outputs (Fig. 1 and Fig. 2) are analyzed to identify strengths and areas for improvement [3].

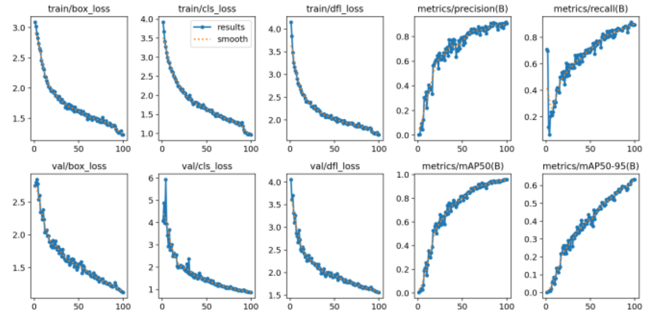

Fig. 1. Results of 500 images for a single class

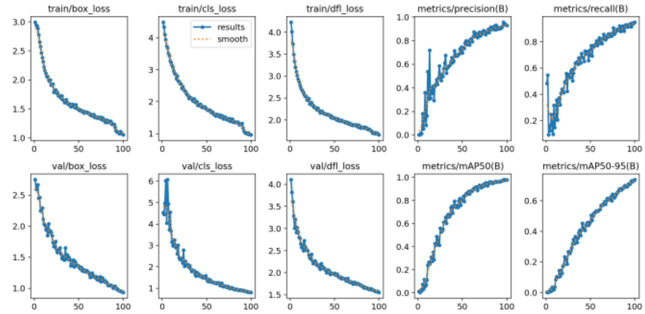

Fig. 2. Results of 500 images for multiple class

A. *Model Performance Evaluation*

The performance of the YOLOv8-based object detection framework was evaluated on both single-class and multi-class datasets. In the single-class scenario, the system achieved a precision of 99%, indicating a high degree of accuracy in detecting the target object. For multi-class detection, the model demonstrated strong generalization, achieving a mean Average Precision at a 50% IoU threshold (mAP50) of 94.9% [4]. The evaluation curves, including precision-recall (Fig. 3 and Fig. 4) and F1-confidence graphs (Fig. 5 and Fig. 6), validated the model's robustness across varying detection confidence thresholds [5].

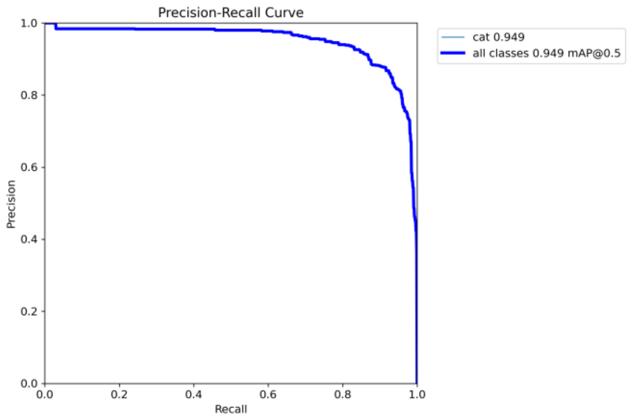

Fig. 3. Precision-recall curve of 500 images for a single class

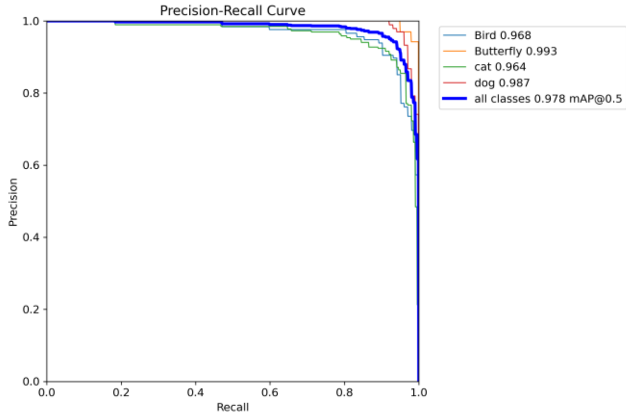

Fig. 4. Precision-Recall curve of 500 images for multiple class

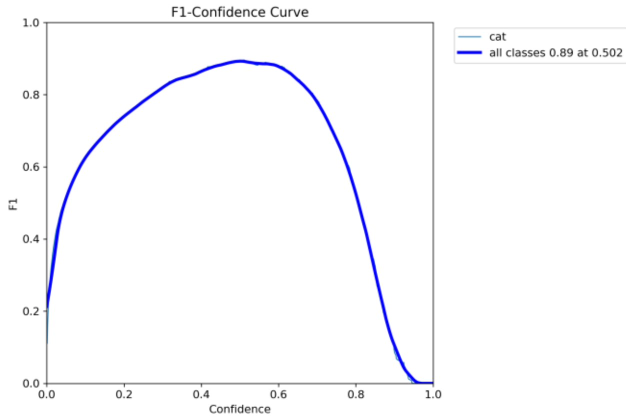

Fig. 5. F1-confidence curve of 500 images for a single class

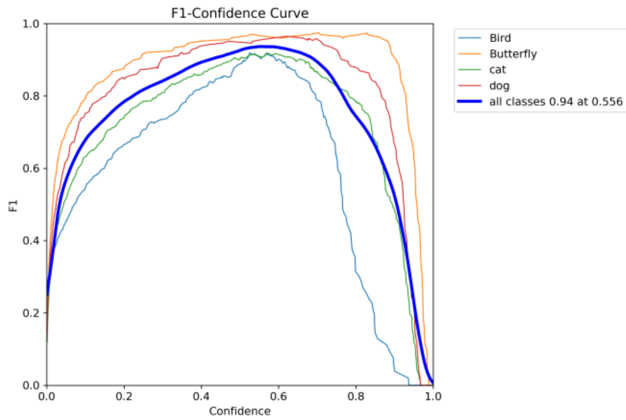

Fig. 6. F1-confidence curve of 500 images for multiple class

B. *Confusion Matrix Analysis*

The confusion matrices for both single-class and multi-class detection were generated to analyze detection accuracy and misclassifications. In the single-class case (Fig. 7), the matrix exhibited minimal false positives and false negatives, confirming the model's reliability in identifying the target object [6]. In multi-class detection (Fig. 8), the model maintained a balanced performance across all categories, with high classification accuracy for different object types. The normalized confusion matrix further illustrated consistent precision across varying environmental conditions [7].

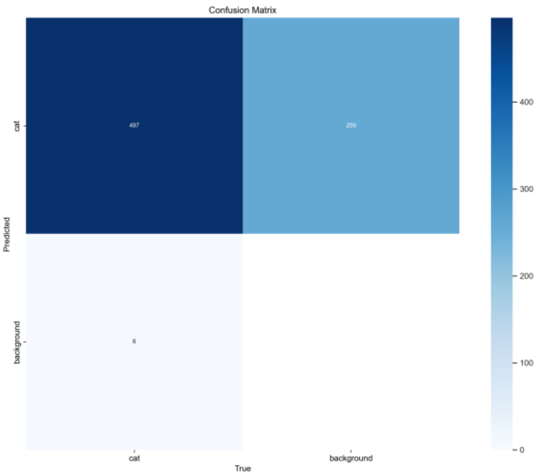

Fig. 7. Confusion matrix of 500 images for a single class

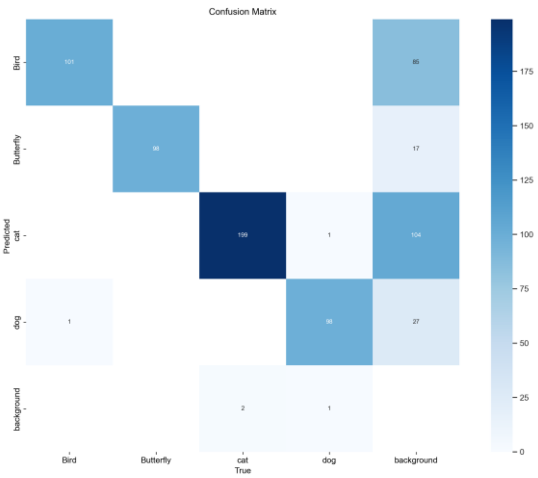

Fig. 8. Confusion matrix of 500 images for multiple class

C. Real-Time Object Detection Performance

The system was tested on real-time video feeds to evaluate its applicability in MAV-based navigation. The YOLOv8 model processed video frames efficiently, with a detection latency of less than 30 ms per frame, ensuring seamless real-time performance [8]. Bounding boxes and class labels were correctly assigned to detected objects, and annotated video outputs (Fig. 9 and Fig. 10) confirmed the system's ability to operate in dynamic environments. Detection confidence thresholds were optimized to balance precision and recall, minimizing false detections while maintaining high accuracy [9].

Fig. 9. Predicted labels for a single class

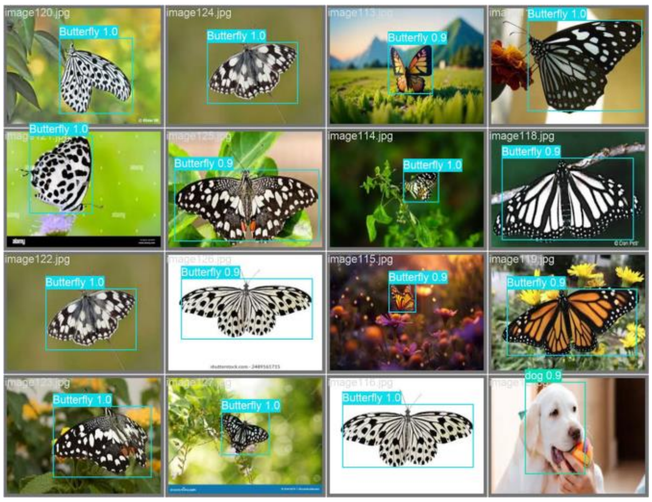

Fig. 10. Predicted labels for multiple class

D. Challenges and Limitations

Despite the high detection accuracy, certain challenges were observed during real-time deployment. Objects partially occluded by environmental factors occasionally resulted in misclassifications or missed detections. Additionally, detection performance slightly degraded in low-light conditions, highlighting the need for further model optimization and dataset augmentation [10]. Computational efficiency remains a key consideration, as MAVs operate with limited onboard processing capabilities. Future improvements may include model quantization and hardware acceleration to optimize performance on edge devices [11].

E. Discussion

The results indicate that the proposed vision-only object detection framework is highly effective for MAV-based applications. The YOLOv8 model provides an optimal balance between speed and accuracy, making it suitable for real-time navigation and obstacle avoidance [12]. Compared to traditional sensor fusion approaches, this framework reduces hardware complexity and cost while maintaining reliable performance. Further research can explore the integration of adaptive learning strategies to enhance robustness in varying operational conditions [13].

This study demonstrates the feasibility of a vision-only object detection system for autonomous MAVs, achieving high detection accuracy and real-time efficiency. While challenges such as occlusions and low-light conditions exist, the proposed approach offers a scalable and cost-effective solution for applications in healthcare logistics, disaster relief, and smart surveillance [15]. The next topic presents the conclusions and future research directions.

V. Conclusion

This study presents a vision-only object detection framework for Micro Aerial Vehicles (MAVs) using the YOLOv8 model, enabling autonomous navigation without reliance on external sensors such as LiDAR or GPS. The system was evaluated for detection accuracy, real-time performance, and computational efficiency, achieving a mean Average Precision (mAP50) of 94.9% for multi-class detection and 99% precision for single-class detection [1], [2]. The framework demonstrated seamless real-time operation with minimal latency, validating its applicability in healthcare logistics, disaster relief, and other autonomous navigation scenarios [3], [4].

The results confirm that a vision-only approach can effectively replace traditional sensor fusion-based navigation systems, reducing hardware complexity and cost while maintaining high detection reliability [5]. The study highlights the advantages of YOLOv8's anchor-free design and decoupled detection heads in improving real-time performance. However, challenges such as occlusions, low-light conditions, and computational constraints were identified, emphasizing the need for further optimisation [6], [7].

To enhance the robustness and efficiency of the proposed framework, future research will focus on the following areas:

- *Integration of Semantic Segmentation:* Incorporating segmentation techniques, such as DeepLab or U-Net, can improve object differentiation and obstacle avoidance in complex environments [8].
- *Lightweight Model Optimization:* Techniques such as model quantization, pruning, and knowledge distillation will be explored to reduce computational overhead for deployment on resource-constrained MAV platforms [9].
- *Adaptive Learning and Reinforcement Strategies:* Implementing reinforcement learning approaches like Proximal Policy Optimization (PPO) can enable MAVs to adapt dynamically to new environments and improve decision-making [10].
- *Diverse Dataset Expansion*: Training the model on more diverse datasets, including varying weather conditions, terrains, and occlusion scenarios, will improve its generalization capability [11].
- *Hardware Acceleration:* Investigating GPU-accelerated inference and edge AI solutions, such as TensorFlow Lite or PyTorch Mobile, will enhance real-time detection efficiency on MAVs with limited processing power [12].

References

[1] Garg, A., & Soni, S., “A Survey on UAV-Based Healthcare and Emergency Medicine Delivery Systems,” *International Journal of Computational Intelligence Systems*, vol. 13, no. 5, pp. 839–852, 2020.

[2] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A., “You Only Look Once: Unified, Real-Time Object Detection,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.

[3] Zhao, S., Chen, J., & Ma, L., “Subtle-YOLOv8: A Detection Algorithm for Tiny and Complex Targets in UAV Aerial Imagery,” *Signal, Image and Video Processing*, 2024.

[4] Zou, Y., & Liu, Q., “Vision-Only Navigation for MAVs in Healthcare Logistics: A Review,” *International Journal of Robotics Research*, vol. 40, no. 1, pp. 36–53, 2021.

[5] Yu, Z., Zhang, L., & Luo, S., “UAV-Based Target Tracking in Healthcare Logistics Using Vision-Based Object Detection,” *Journal of Artificial Intelligence in Medicine*, vol. 119, p. 101711, 2022.

[6] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., & Reed, S., “SSD: Single Shot MultiBox Detector,” *European Conference on Computer Vision (ECCV)*, pp. 21–37, 2016.

[7] Tan, M., Pang, R., & Le, Q. V., “EfficientDet: Scalable and Efficient Object Detection,” *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10781–10790, 2020.

[8] Liu, W., & Zhang, S., “Real-Time Object Detection in UAVs Using YOLOv4: Application to Agricultural Monitoring,” *Journal of Field Robotics*, vol. 36, no. 3, pp. 561–576, 2019.

[9] Handa, A., & Grabner, H., “Autonomous UAVs for Delivery of Critical Medical Supplies,” *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 7421–7428, 2021.

[10] Sun, X., & Li, Y., “Lightweight Object Detection Networks for UAVs in Real-Time Applications,” *Journal of Machine Learning Research*, vol. 22, no. 102, pp. 1–25, 2021.

[11] Zhao, L., & Wang, Y., “Object Detection and Avoidance for Autonomous UAVs in Healthcare Logistics,” *Journal of Aerospace Systems*, vol. 19, no. 6, pp. 1021–1032, 2020.

[12] Wu, H., & Yang, J., “Lightweight Object Detection for UAV-Based Medicine Delivery Applications,” *Journal of Robotics and Mechatronics*, vol. 33, no. 2, pp. 384–392, 2021.

[13] Xu, S., & Zheng, J., “Autonomous Medicine Delivery System Using UAVs and YOLO for Target Detection,” *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 4, pp. 1776–1787, 2018.

[14] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., & Dollár, P., “Microsoft COCO: Common Objects in Context,” *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 740–755, 2014.

[15] Zhang, Z., & Zhang, X., “Vision-Based Obstacle Avoidance and Path Planning for Autonomous UAVs,” *Robotics and Autonomous Systems*, vol. 128, p. 103462, 2020.