© 2026 ADYPJIET, Volume 1(1), Online ISSN: 3049-1517, June 2026, pp. 67-73

Ajeenkya DY Patil Journal of Innovation in Engineering & Technology

Journal Homepage: https://www.adypsoe.in/adypjiet

Crowd and Traffic Analysis in Drone Imagery Using AI/ML

[1]Samruddhi Jagtap
[2]Atharv Kulkarni,
[3]Janvi Adam,
[4]Purva Shingare,
[5]Vinay Nagalkar
[1-4] UG students, Ajeenkya D Y Patil School of Engineering, Pune, INDIA
[5]Assistant Professor, Ajeenkya D Y Patil School of Engineering, Pune, INDIA

Article History:
Received: 01/04/2026
Revised: 16/04/2026
Accepted: 23/06/2026

Abstract:

Urbanization has significantly increased the complexity of crowd management and traffic control, necessitating the development of intelligent, automated surveillance systems. This paper proposes a comprehensive drone-based monitoring framework that leverages YOLOv8, a state-of-the-art deep learning algorithm, for that accurate detection and classification of pedestrians and vehicles. The model was trained on a curated dataset of 16,000 high-resolution aerial images collected from multiple urban scenarios, including busy intersections, public gatherings, and congested roadways. Data augmentation techniques such as rotation, flipping, scaling, and exposure adjustment were applied to enhance generalization and improve performance under varying lighting and perspective conditions. For deployment, the trained YOLOv8 model was ported to a Raspberry Pi 5 with 8 GB RAM and a connected Pi Camera v3 mounted on a quadcopter, enabling edge-based real- time inference at 18 frames per second without relying on cloud processing. The system achieved a mean average precision (mAP) of 85.3%, with class-wise precision of 92.8% for pedestrians and 95.2% for vehicles. An experimental evaluation across three urban locations, under diverse weather and lighting conditions, confirmed the robustness, low latency, and reliable operation of the system, even in high-density scenes. By combining on-edge AI processing with autonomous aerial mobility, the proposed framework improves response time, enhances data privacy, and offers scalable solutions for smart city applications. The study demonstrates the feasibility of deploying lightweight, high- accuracy object detection models on embedded platforms, marking a step toward autonomous, real-time aerial surveillance for urban safety management and intelligent traffic monitoring. Urbanization has increased the need for intelligent traffic and crowd monitoring systems. This study proposes a drone-based framework using YOLOv8 for accurate real-time detection of pedestrians and vehicles, deployed on a Raspberry Pi for edge processing. The system achieves high accuracy and low latency, making it suitable for scalable smart city surveillance applications.

Keywords:
Crowd Analysis, Traffic Monitoring, YOLOv8, Raspberry Pi 5, Drone Surveillance, Edge AI

1.0 Introduction

Urban centers worldwide are experiencing increasing challenges due to rising population density and vehicle congestion, which result in safety risks, traffic inefficiencies, and delays in emergency response. Managing crowds and vehicular flow in such complex environments has become a critical concern for city planners and public safety authorities [1,2]. Traditional closed-circuit television (CCTV) systems, while widely deployed, are limited by fixed viewpoints, restricted spatial coverage, and reliance on wired infrastructure, making them insufficient for large-scale, dynamic monitoring [3]. These systems also struggle to provide real-time analytics necessary for rapid decision- making during high-density events or emergencies. Recent developments in unmanned aerial vehicles (UAVs) and artificial intelligence (AI) offer promising alternatives. Drones equipped with AI-based vision systems provide flexible, high-mobility monitoring platforms capable of capturing wide-area aerial imagery in real time. Combined with advanced deep learning-based object detection algorithms such as YOLOv5, YOLOv8, and Faster R- CNN these systems can automatically detect, classify, and track pedestrians and vehicles with high accuracy [4]. YOLOv8, an improved variant of the “You Only Look Once” family, introduces a more efficient backbone, optimized detection layers, and anchor-free heads, achieving a balance between speed and precision suitable for edge deployment. The primary objective of this study is to design, implement, and evaluate a drone-based framework for real-time crowd and traffic analysis using YOLOv8 trained on a large-scale aerial dataset of 16,000 annotated images. This research aims to contribute to the development of autonomous, scalable surveillance systems capable of providing accurate crowd density estimation, efficient traffic flow assessment, and improved urban safety management. The proposed system seeks to enhance operational efficiency, reduce latency, and offer a practical solution for smart city applications, bridging the gap between conventional monitoring methods and intelligent aerial surveillance [5].

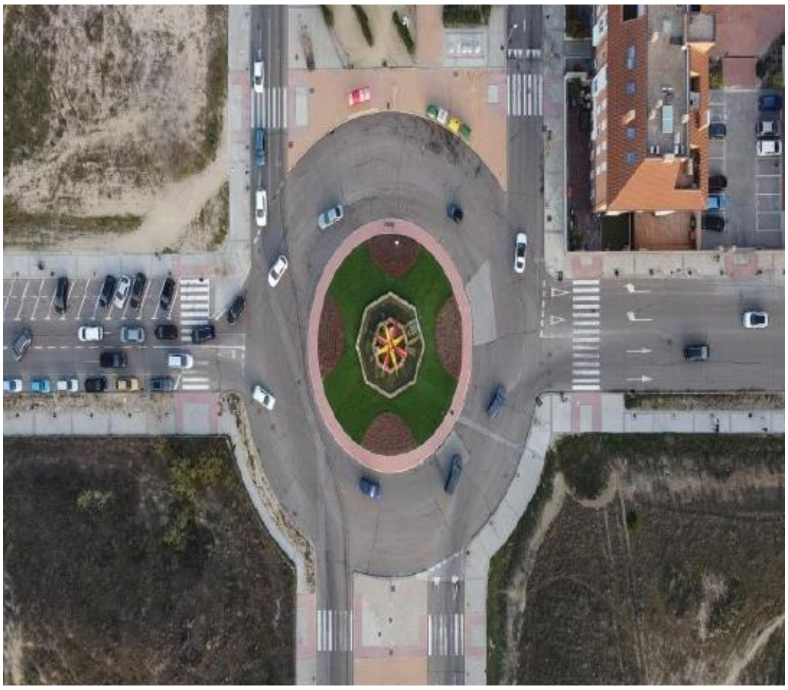

Figure 1. Aerial Image

2.0 Materials and Methods:

2.1 Dataset Preparation

The foundation of any deep learning-based surveillance system lies in the quality and diversity of its dataset. For this study, a comprehensive aerial dataset of 16,000 high-resolution images was collected from various urban environments, including dense road networks, multi-lane intersections, pedestrian-heavy zones, public gatherings, and event locations. The images were captured at different times of day, including morning, afternoon, and evening, under diverse weather conditions such as sunny, cloudy, and partially rainy scenarios. This variety ensured that the trained model could generalize across real-world conditions, including variations in illumination, shadow patterns, occlusions, and perspective distortions that are common in drone-captured imagery. Each image was meticulously annotated to mark pedestrians, vehicles, and other objects of interest, ensuring high-quality ground truth data for supervised training [6]. The annotation process followed strict guidelines, with multiple reviewers cross- verifying labels to minimize human error. To further enhance the diversity of the dataset and improve model robustness, data augmentation techniques were extensively applied. These included rotation at random angles, horizontal and vertical flipping, scaling to simulate varying altitudes, brightness and contrast adjustments to account for different lighting conditions, Gaussian noise addition, exposure correction, and random cropping to simulate partial occlusions. Such augmentations were critical to ensuring that the YOLOv8 model could accurately detect small and densely packed targets, such as pedestrians in crowded streets or vehicles in traffic jams, under varying real- world conditions [7]. The dataset was split into training, validation, and testing subsets in a ratio of 70:15:15, resulting in 11,200 images for training, 2,400 images for validation, and 2,400

© 2026 ADYPJIET, Volume 1(1), Online ISSN: 3049-1517, June 2026, pp. 67-73

images for testing. This split was designed to balance the learning process while allowing proper evaluation of the model's generalization performance on unseen data.

Figure 2. Training Data I

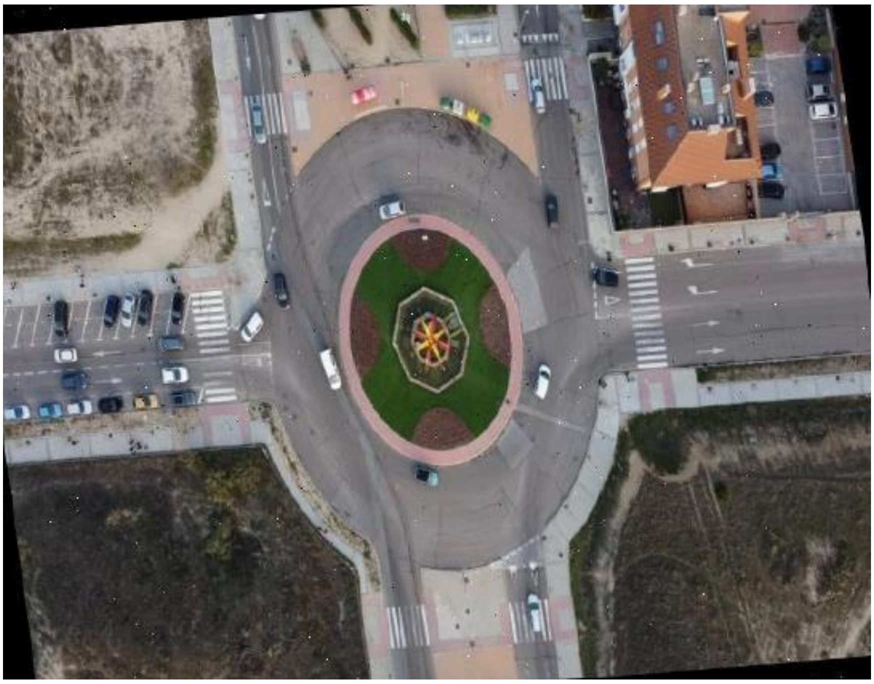

Figure 3. Training Data II

2.2 Model Architecture

© 2026 ADYPJIET, Volume 1(1), Online ISSN: 3049-1517, June 2026, pp. 67-73

The YOLOv8 small variant was chosen for its ability to provide a strong balance between computational efficiency and high detection accuracy. YOLOv8 represents an evolution of the YOLO (You Only Look Once) family, incorporating a more advanced backbone network, optimized detection heads, and anchor-free mechanisms for better localization of objects of varying sizes. The model was trained using transfer learning, leveraging pre-trained weights on large-scale object detection datasets to accelerate convergence and enhance performance with limited aerial imagery. The training was conducted on an NVIDIA RTX 4080 GPU, providing the computational power required for processing high-resolution images efficiently. The model underwent 200 epochs of fine-tuning with a batch size of 16 and an initial learning rate of 0.001, employing learning rate schedulers to optimize convergence. Advanced augmentation strategies such as Mosaic augmentation were applied to combine multiple images into a single training sample, increasing contextual diversity and improving small-object detection. The Intersection over Union (IoU) loss function was used to optimize bounding box predictions, and anchor-free detection heads improved the localization of objects in high-density scenes where traditional anchor-based methods often fail. Additional regularization techniques were applied to prevent overfitting and ensure that the trained model could maintain high performance on unseen data.

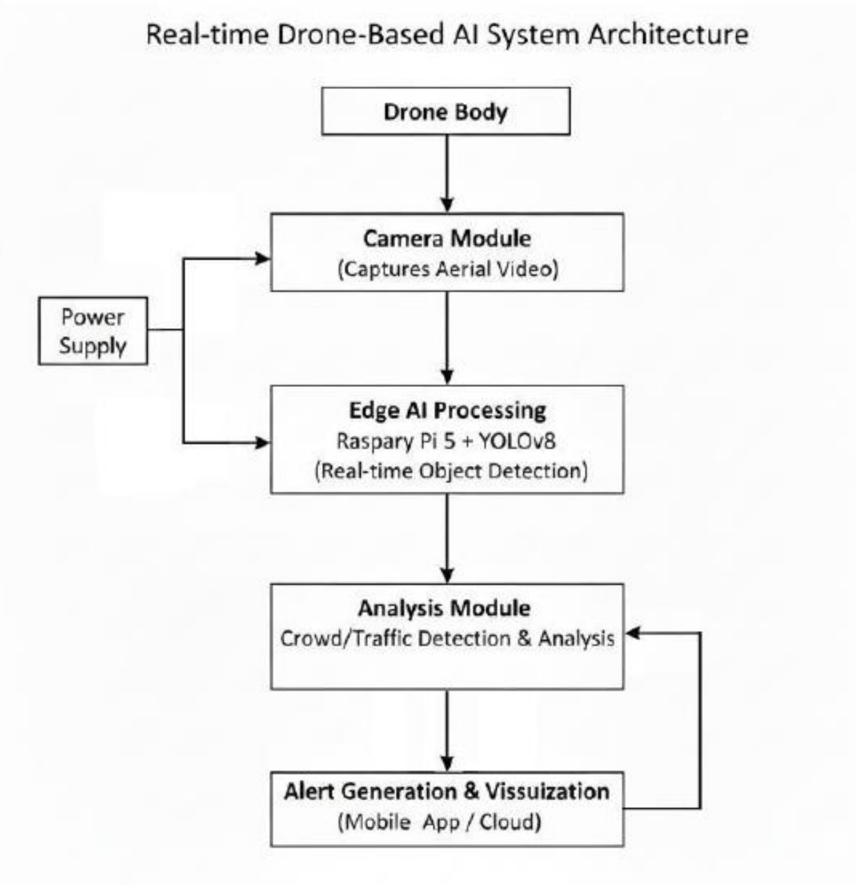

Figure 4. Flowchart

2.3 Hardware Implementation

Deployment of deep learning models in real-time aerial surveillance requires careful consideration of hardware constraints. To enable edge inference, the trained YOLOv8 model was deployed on a Raspberry Pi 5 with 8 GB RAM, paired with a Pi Camera v3 mounted on a quadcopter drone. This setup was selected to eliminate dependency on cloud servers, reduce latency, and protect sensitive urban surveillance data.

The edge implementation required optimization of the YOLOv8 model using PyTorch Lite, allowing efficient execution on embedded hardware without compromising detection accuracy. OpenCV was used for real-time video frame capture, preprocessing, and resizing, ensuring smooth input to the neural network. To handle the drone's movement and vibrations, additional preprocessing such as frame stabilization and motion compensation was implemented, which improved bounding box accuracy and reduced detection jitter. Power management and computational efficiency were also considered, as drones operate on limited battery resources. The Raspberry Pi 5's low-power consumption and YOLOv8's lightweight architecture ensured that the system could sustain extended flight times while performing real-time inference at a rate of 18 frames per second (FPS).

In addition, the system was designed with scalability and adaptability in mind, allowing integration with other smart city infrastructures such as traffic management systems and emergency response networks. The modular architecture enables easy upgrades to more advanced models or hardware components as technology evolves. Furthermore, the use of on-device processing ensures reliable performance even in areas with limited network connectivity, making the framework suitable for deployment in both densely populated urban centers and remote locations.

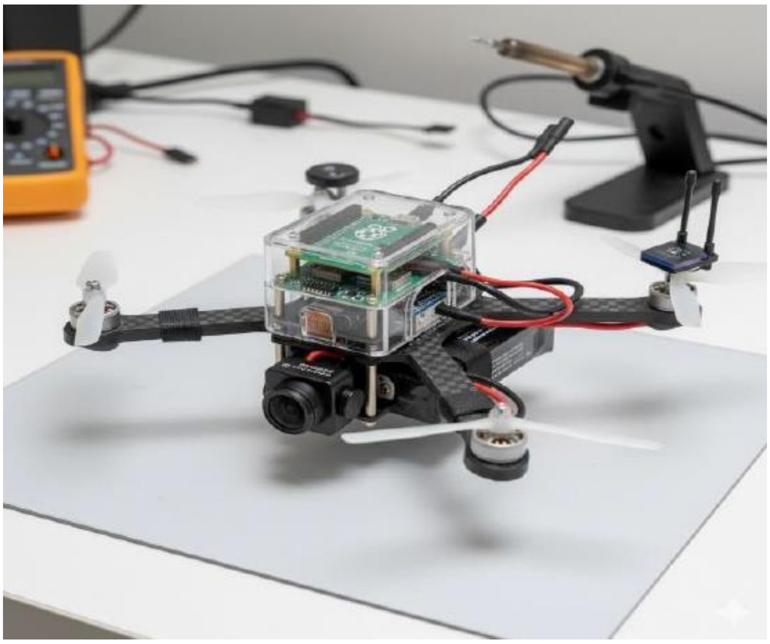

Figure 5 Hardware Image

References:

1. Umair Jilani, Muhammad Asif, Muhammad Yousuf Irfan Zia, Munaf Rashid, Sarmad Shams & Pablo Otero ,A systematic Review on Urban Road Traffic Congestion, Wireless Personal Communication, 140, Pages 81-109, 2025.

2. Theocharidis, T., Kavallieratou, E., Underwater communication technologies: A review. Telecommunication Systems, 88(2), 54,2025.

3. Samirsinh P Parmar, Swati H Chauhan, Legal and Ethical Guidelines in CCTV Surveillance: Addressing Key Issues in Educational Institutions, International Journal of Human Rights and Constitutional Studies 6(2):14-28,2024.

4. Fabio Chagas, Neno Ruseno, Aurilla Aurelie Bechina, Artificial Intelligence Approaches for UAV Deconfliction: A Comparative Review and Framework Proposal, Automation, 6(40),54, 2025.

5. Vivi Afifah, Surni Erniwati, YOLOv8 for Object Detection: A Comprehensive Review of Advances,Techniques, and Applications, 2(1) (2026): International Journal of Advanced Computing and Informatics.

6. Y. Akbari, Noor Almaadeed, Somaya Al-ma'adeed, Omar Elharrouss, Applications, databases and open computer vision research from drone videos and images: a survey, Artificial Intelligence Review 54(11),2021.

7. Wulan Sapitri, Yesi Novaria Kunang, Ilman Zuhri Yadi, Mahmud Mahmud, The Impact of Data Augmentation Techniques on the Recognition of Script Images in Deep Learning Models, Jurnal Online Informatika, 8(2):169-176.