International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 15 Issue 3, March-2026

An Offboard Intelligent System for Real-Time UAV-Based Human Detection

Indranil Mukherjee[1], Pranali Gadade[2], Arya Kamble[3], Aishwarya Kadam[4], Archana Gore[5],

[1,2,3,4,5]Department of Artificial Intelligence and Machine Learning, Navsahyadri Education Society's Group of Institutions, Pune, India

ABSTRACT

Unmanned Aerial Vehicles (UAVs) are increasingly utilized for aerial surveillance, disaster response, and search-and-rescue operations. However, deploying computationally intensive deep learning models directly on lightweight UAV platforms is constrained by limited onboard processing capability, memory, and power resources. This paper proposes an offboard deep learning framework for real-time human detection in UAV-based surveillance systems. In the proposed architecture, an ESP32-CAM module mounted on the UAV functions as a video acquisition and IP-based streaming unit, transmitting real-time aerial video to a ground control station. Human detection is performed offboard using the YOLOv8s model, thereby reducing onboard computational burden while maintaining real-time performance.

To enhance detection robustness and generalization, two publicly available aerial datasets, Manipal UAV and SARD, were integrated to form a diversified training corpus. A scale-aware bounding box height filtering mechanism was introduced to remove high-altitude samples containing extremely small human annotations, improving detection reliability. The curated dataset was partitioned into training, validation, and testing subsets in an 80:10:10 ratio and converted into YOLO-compatible format. The YOLOv8s model was trained for 100 epochs on Google Colab using optimized hyperparameters.

Experimental results demonstrate stable convergence without significant overfitting, achieving a precision of 0.96, recall of 0.92, mAP@0.5 of 0.95, and mAP@0.5:0.95 of 0.65. These findings validate the effectiveness of the proposed offboard inference framework for reliable real-time UAV-based human detection.

Keywords: ***You Only Looks Once (YOLOv8), UAV Surveillance, Human Detection.***

1. INTRODUCTION

Unmanned Aerial Vehicles (UAVs) have observed rapid advancements over the past decade and are increasingly deployed in applications such as aerial surveillance, disaster response, environmental monitoring, border security, and search-and-rescue missions. Their ability to access hazardous or inaccessible terrains, combined with real-time aerial imaging capabilities, makes them highly valuable in time-critical operations [11]. In particular, automated human detection from aerial imagery plays a crucial role in search-and-rescue scenarios, disaster assessment, and crowd monitoring. Rapid identification of human presence can significantly reduce response time and improve mission efficiency [8].

Traditionally, UAV-based surveillance systems rely on manual monitoring of live video streams by ground control operators. However, continuous visual inspection of high-volume aerial footage is both cognitively demanding and prone to human error, especially under stressful or prolonged operational conditions [1]. To overcome these limitations, recent research has focused on integrating deep learning-based object detection algorithms into UAV systems to enable automated target identification [3].

Deep convolutional neural network (CNN) architectures, particularly single-stage detectors such as YOLO (You Only Look Once), have demonstrated remarkable performance in real-time object detection tasks. Models such as YOLOv5 and YOLOv7 have been successfully applied for detecting humans from drone imagery in search-and-rescue operations and thermal aerial videos [2]. The latest generation, YOLOv8, provides improved detection accuracy, optimized feature extraction, and efficient inference capabilities, outperforming earlier versions in several evaluation metrics.

Despite these advancements, deploying computationally intensive deep learning models directly on lightweight UAV platforms remains challenging. UAV-mounted embedded systems are typically constrained by limited processing power, memory capacity, and battery resources. Onboard inference increases energy consumption and thermal load, potentially reducing flight endurance and affecting system stability [11]. Existing UAV-based human detection systems primarily implement onboard processing using single-board computers such as Raspberry Pi or other edge devices, which introduces additional hardware complexity and may limit scalability [8].

To address these limitations, this paper proposes an offboard deep learning framework for real-time UAV-based human detection. In the proposed architecture, an ESP32-CAM module mounted on the UAV functions as a lightweight vision sensor and IP-based streaming device. Real-time aerial video is transmitted via Wi-Fi to a ground control laptop, where human detection is performed using the YOLOv8s model. By offloading computationally intensive inference tasks to a ground station, the proposed framework significantly reduces onboard computational burden while maintaining real-time detection performance.

2. Background Work

KalpaJayalath et. al. presents an autonomous drone-based system for real-time human detection to support search and rescue missions. The drone uses a TensorFlow neural network to identify humans in aerial images and automatically navigates toward detected locations. The system processes images onboard, sends selected frames for operator verification, and uses GPS data for autonomous flight control [1]. USMAN AZMAT et. al. proposes a human action recognition system for drone-recorded videos that addresses challenges like dynamic backgrounds, motion blur, and occlusions. The method extracts human silhouettes, skeleton keypoints, and motion features before classifying actions using a deep convolutional neural network. Experimental results on benchmark datasets show high recognition accuracy, achieving up to 95%[2]. F. H. Zaman et .al.proposed on developYOLOv5-based method for automated human detection in drone-assisted search and rescue operations. The model is trained on aerial videos captured in simulated rescue scenarios using different flight patterns and zoom levels. Results show improved detection performance and real-time processing capability, especially at higher zoom levels [3].G. Kucukayan et. al. This paper proposes YOLO-IHD, a new deep learning model for autonomous human detection in indoor environments using drones. The model is trained on a custom indoor aerial dataset and improves accuracy through optimized convolutional layers and attention mechanisms. Results demonstrate reliable real-time performance for applications such as disaster response and indoor search-and-rescue [4]

Hyun-Ki Jung et. al. represented a real-time drone detection system using a modified YOLO deep learning model optimized for detecting intruding drones in sensitive areas. It also introduces a semi-automatic dataset labeling method based on Kernelized Correlation Filters to accelerate training data preparation. Experimental results demonstrate the effectiveness of the proposed detector[6]. I. P. Saryet .al. focuses on human detection using UAV aerial images with deep learning models YOLOv5 and YOLOv8. The performance of both models is compared using precision, recall, and F1-score metrics. Results show that YOLOv8 achieves better overall detection performance than YOLOv5 [5]. Chenchen Jiang et. al. reviews the integration of YOLO algorithms with UAV technology, known as YOLO-Based UAV Technology (YBUT), for real-time multi-target detection and classification. It discusses the development, applications in fields such as engineering, transportation, and agriculture, and the benefits of combining drone and deep learning technologies. The study also highlights future research directions and potential applications of YBUT [7].XianxuZhai et. al. This paper proposes a UAV-based thermal infrared (TIR) object detectionframework using YOLO deep learning models for images and videos captured by FLIR cameras. The method addresses challenges such as complex scenes, low resolution, and limited datasets. Results show high detection performance with 88.69% mAP and real-time speed up to 50 FPS using the YOLOv5-smodel[9]. Chen Liu et. al. proposes an optimized YOLOv8-based method for tiny UAV detection in complex environments with small targets and varying lighting conditions. The model enhances small-object detection using a high-resolution detection head, improved feature extraction, and attention mechanisms. Experimental results show significant improvements in precision, recall, mAP, and reduced model size, making it suitable for real-world deployment[10]**.**

Sangle et. al. represented an autonomous drone system that detects humans in real-time during search and rescue missions using aerial video. A custom TensorFlow neural network identifies human presence, and the drone automatically moves toward the location for better verification. The system also sends selected frames and GPS data to the operator for further action[8].W. Guettala et. al. proposes a UAV-based thermal infrared human detection system using the YOLOv7 deep learning model. It detects humans in low-resolution thermal images and videos with 72.5% accuracy at IoU 0.5 and high speed (161 FPS). The system improves public safety by enabling human detection in complex environments using UAV thermal cameras[11]. Nilesh Parmanand Motwani et. al. improves UAV aerial object detection by enhancing the YOLOv8s model to handle small targets, scale variations, and complex backgrounds in drone images. An improved WIoUv3 loss function is introduced to reduce localization errors and increase detection accuracy. Experimental results show an improvement in mAP@0.5 to 40.7%[12].Summary of related work shown in Table.1.

Table. 1. Summary of related work

Authors / Year	Target Type	Dataset Type	Sensor / Radar / Camera Type	Signature Format / Features Used	Model Used	Result / Accuracy
Jayalath&Munasinghe (2021) *Drone-based Autonomous Human Identification for Search and Rescue*	Human (outdoor SAR)	Self-developed drone-captured dataset	RGB Camera + GPS	Image frames + location data	Custom TensorFlow CNN	Real-time detection on Raspberry Pi; Successful flight tests (no explicit accuracy % reported)
Kamaru Zaman et al. (2023) *Human Detection from Drone Using YOLOv5*	Human (Search & Rescue)	Real-world aerial dataset (drone footage)	UAV RGB Camera	Bounding box annotations	YOLOv5 (S, M, L)	High detection precision; capable of real-time detection
Indri PurwitaSary et al. (2023)*Performance Comparison of YOLOv5 and YOLOv8*	Human (Aerial images)	Aerial image dataset	UAV camera	Range FFT vectors + image features	YOLOv5 vs YOLOv8	YOLOv8 outperforms YOLOv5: +2.82% Precision, +0.98% F1-Score; YOLOv5 higher Recall by 0.54%
Kucukayan&Karacan (2024) *YOLO-IHD: Indoor Drone Human Detection*	Human (Indoor)	Self-created indoor UAV dataset	Indoor drone camera	Optimized convolutional features + attention	YOLO-IHD (modified YOLOv7-tiny)	mAP@0.5 improved by 42.51% (IHD dataset) and 33.05% (VisDrone) over YOLOv7-tiny
Sangle et al. (2020) *Smart Human Detection Drone for Rescue*	Human under debris	Sensor-based dataset	PIR Sensor + Ultrasonic + Wi-Fi	IR radiation patterns	Sensor-based detection (non-AI)	PIR detects humans up to 8 m; Successful geolocation tests using firebase
(Additional YOLO UAV Paper in Uploads, 2023) *fYOLOv5 for Search & Rescue*	Human	Real-world SAR dataset	UAV RGB Camera	Bounding box + enhanced frames	fYOLOv5 (Lightweight)	Reported improvement in detection speed and efficiency (exact values not included in snippet)

2. METHODOLOGY

3.1 System Overview

The proposed system implements an offboard deep learning architecture for real-time UAV-based human detection. Unlike conventional onboard processing approaches, computationally intensive inference operations are shifted from the UAV platform to a ground control station to minimize onboard resource consumption. A custom-built UAV equipped with an ESP32-CAM module captures real-time aerial video frames during flight. The ESP32-CAM functions as a lightweight image acquisition and IP-based streaming unit, transmitting video frames wirelessly to a ground station via Wi-Fi communication. No deep learning inference is executed onboard, thereby significantly reducing power consumption, thermal stress, and computational overhead on the UAV. At the ground station, incoming frames are processed sequentially using the trained YOLOv8s object detection model. The model performs human detection and generates bounding box predictions for identified individuals in each frame. By offloading inference to a GPU-enabled ground system, the architecture enables stable real-time detection performance while enhancing UAV flight endurance and operational stability. This offboard processing framework is particularly suitable for surveillance, disaster response, and search-and-rescue scenarios where reliable real-time human detection is critical.

3.2 Dataset Preparation and Fusion

In this study, two publicly available aerial human detection datasets, ManipalUAV and SARD, were consolidated to construct a unified training corpus. During the merging process, duplicate filename conflicts were programmatically resolved by appending incremental suffixes to repeated filenames, thereby preventing data overwriting and preserveng dataset integrity.Representative samples from the curated UAV-based human detection dataset are illustrated in Fig. 1.

(a)

(b)

(c)

(d)

Fig. 1.Representative aerial human detection samples from the consolidated dataset: (a) open-area campus environment, (b) challenging illumination conditions, (c) search-and-rescue operational scenario, and (d) unrestricted forest region with dense background complexity.

Following dataset consolidation and preprocessing, a total of 7358 images were retained in the final curated dataset.All annotation files adhered to the standard YOLO object detection format, where each object instance is represented as:
(class_id,x_c, y_c, w, h)

Here, x_c and y_c denote the normalized center coordinates of the bounding box, while w and h represent normalized width and height relative to the image dimensions. A structured YAML configuration file was subsequently generated to define the training, validation, and testing image directories, along with the single target class label (“human”).

To mitigate the adverse effects of extreme scale variation in aerial imagery, a scale-aware bounding box filtering mechanism was introduced. Let h_{pixel} denote the bounding box height in pixel units and H_{image} represent the image height. The normalized bounding box height is computed as:

International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 15 Issue 3, March-2026

$$h_n = \frac{h_{pixel}}{H_{image}}$$

$$h_n \geq \tau$$

An altitude-proxy threshold τ was defined such that:

where $\tau = 0.08$ Bounding boxes with normalized height below this threshold were excluded from the dataset. This filtering strategy reduces noise introduced by extremely small object instances, enhances feature representation quality, and stabilizes gradient learning during training.Finally, the refined dataset was partitioned into training (80%), validation (10%), and testing (10%) subsets to ensure balanced optimization and unbiased performance evaluation.

3.3 Model Architecture and Training

The proposed human detection system is implemented using the YOLOv8s (You Only Look Once, Version 8 – small variant) object detection architecture. YOLOv8 is a single-stage detection framework designed to achieve a balance between detection accuracy and real-time inference speed. In this study, the small (s) variant was selected to ensure efficient processing of aerial video streams while maintaining high detection performance.

The YOLOv8 architecture is composed of three primary components: the backbone, neck, and detection head. The backbone is responsible for hierarchical feature extraction from input images and consists of convolutional layers integrated with C2f modules, which enhance feature propagation and gradient flow. The neck employs a multi-scale feature aggregation mechanism to combine low-level spatial features with high-level semantic features, thereby improving detection of human targets at varying scales. The detection head performs bounding box regression and classification in a decoupled manner, enabling precise localization and improved classification confidence.

In the proposed framework, the YOLOv8s model was trained on the merged and scale-filtered aerial human dataset using an input resolution of 640 × 640 pixels. The training process was conducted for 100 epochs with GPU acceleration on Google Colab. The loss function is composed of three components: bounding box regression loss, classification loss, and Distribution Focal Loss (DFL), which together optimize localization precision and classification reliability.By integrating YOLOv8s within the offboard processing architecture, the system achieves reliable human detection performance while preserving UAV energy efficiency and operational stability.
The training parameter configuration is shown in Table 2.

Table. 2. Training parameter configuration

Parameter	Discription	Values
Model	Object detection architecture used	YOLOv8s
Dataset	Consolidated aerial human dataset (ManipalUAV + SARD)	7358 images
Number of Classes	Total detection categories	1 (Human)
Image Size	Input resolution during training	640× 640 pixels
Epochs	Total number of training iterations	100
Optimizer	Optimization algorithm	SGD
Initial Learning Rate	Starting learning rate	0.01
Batch Size	Images Processed Per batch	4
Device	Computational hardware	T4GPU (device=0)
Dataset Split	Train / Validation / Test ratio	80% /10% / 10%
Training Platform	Computational environment	GoogleColab (GPU)

The evaluation metrics indicate strong detection capability. Precision increases throughout training and stabilizes at approximately 0.96, indicating a low false-positive rate. Recall reaches approximately 0.92, showing that most human instances are successfully detected. The mAP@0.5 converges close to 0.95, confirming high detection accuracy at standard IoU threshold. The mAP@0.5:0.95 stabilizes around 0.65, demonstrating reliable performance across stricter IoU thresholds.Shown in Fig. 2, the classification loss, bounding box loss, and distribution focal loss of the model training and validation sets gradually decrease with the increase of training times, indicating that the model is continuously studying and has a certain extent of generalization capability, which can obtain more accurate features.

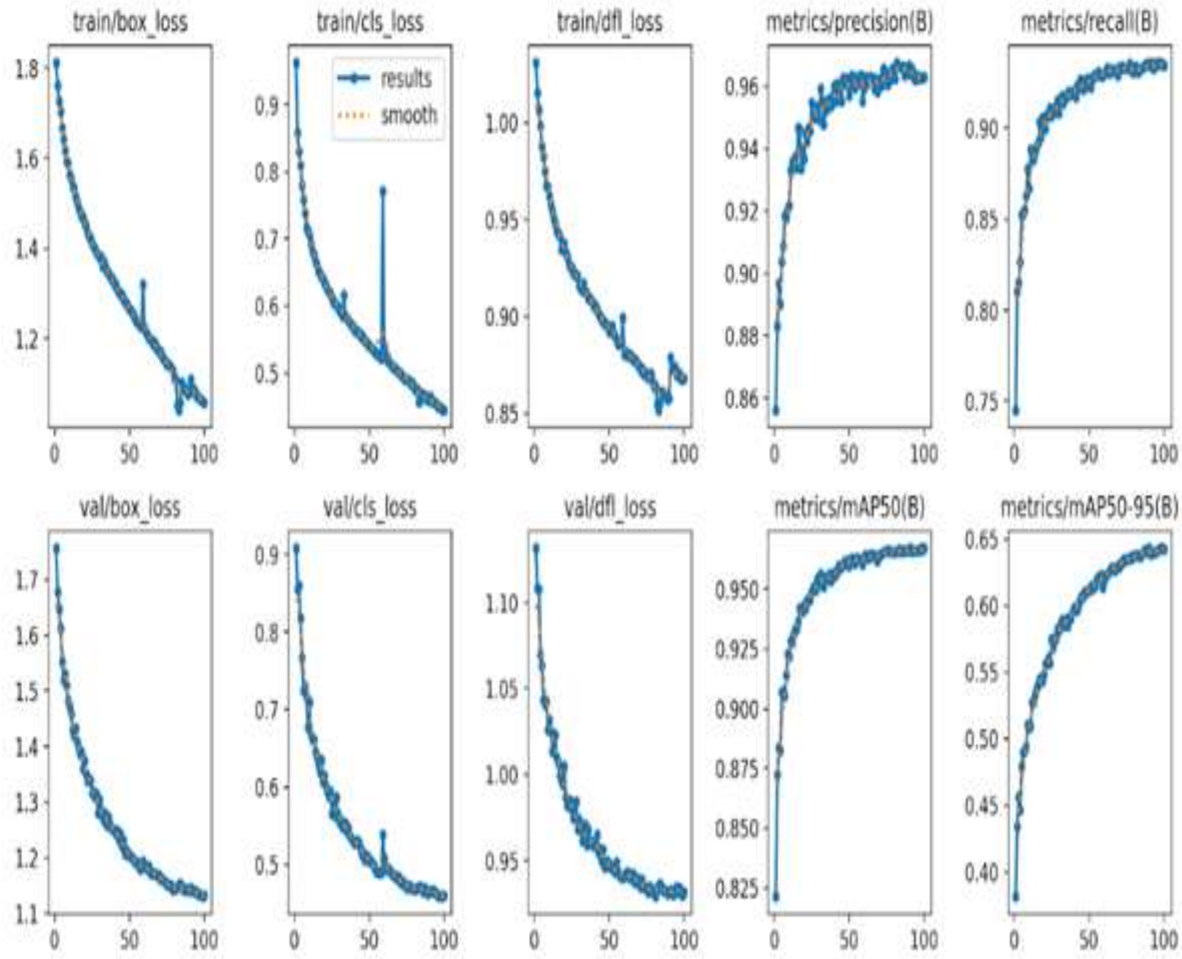

Fig .2. Training indicator chart

Overall, the training convergence behavior, high precision (≈0.96), strong recall (≈0.92), mAP@0.5 (≈0.95), and dominant confusion matrix diagonal collectively confirm that the proposed UAV-based human detection framework achieves stable optimization and reliable real-time detection performance.

3.4 Hardware Setup

The proposed system is implemented using a custom-built quadcopter UAV configured for offboard human detection. The platform consists of a quadcopter frame integrated with four brushless DC motors and Electronic Speed Controllers (ESCs) to ensure stable thrust generation and controlled maneuverability. A KK2.1.5 flight controller is employed for real-time attitude stabilization by regulating roll, pitch, and yaw dynamics. Manual control and emergency override are enabled through a radio transmitter–receiver pair.

The UAV is powered by a lithium-polymer (Li-Po) battery, while a DC–DC buck converter provides regulated voltage to onboard electronic components. An ESP32-CAM module is mounted on the UAV for real-time aerial video acquisition and Wi-Fi-based IP streaming to a ground station, where human detection is performed offboard to minimize onboard computational load and power consumption.Actual drone image shown in Fig.3

Fig .3. Actual drone image

The UAV operates within a direct line-of-sight (LOS) communication range for stable control. When the ESP32-CAM module is integrated, the effective Wi-Fi transmission range is approximately 20–30 meters under normal outdoor conditions. The achievable flight altitude depends on the type of Wi-Fi connectivity used for video streaming. When connected through a mobile hotspot, stable video transmission is maintained up to approximately 10 meters (≈30 feet). However, when connected via a dedicated Wi-Fi router, the operational altitude increases to approximately 30–40 meters (≈100–130 feet). The average flight endurance of the UAV is approximately 8-10 minutes per charge, depending on payload weight, environmental conditions, and battery health.

4. RESULTS AND DISCUSSION

The numerical evaluation indicates high detection reliability with strong precision–recall balance. The achieved mAP scores demonstrate effective bounding box localization across varying IoU thresholds. These results confirm that the proposed offboard detection framework maintains stable inference performance under aerial imaging conditions.

4.1 Confusion Matrix Analysis

The normalized confusion matrix shows strong diagonal dominance. Approximately 95% of true human instances are correctly classified as human, while only about 5% are misclassified. The background class also shows high prediction confidence, indicating minimal confusion between human and background categories. This confirms that the trained model effectively distinguishes human targets from aerial background scenes.confusion matrixshown in Fig. 4.

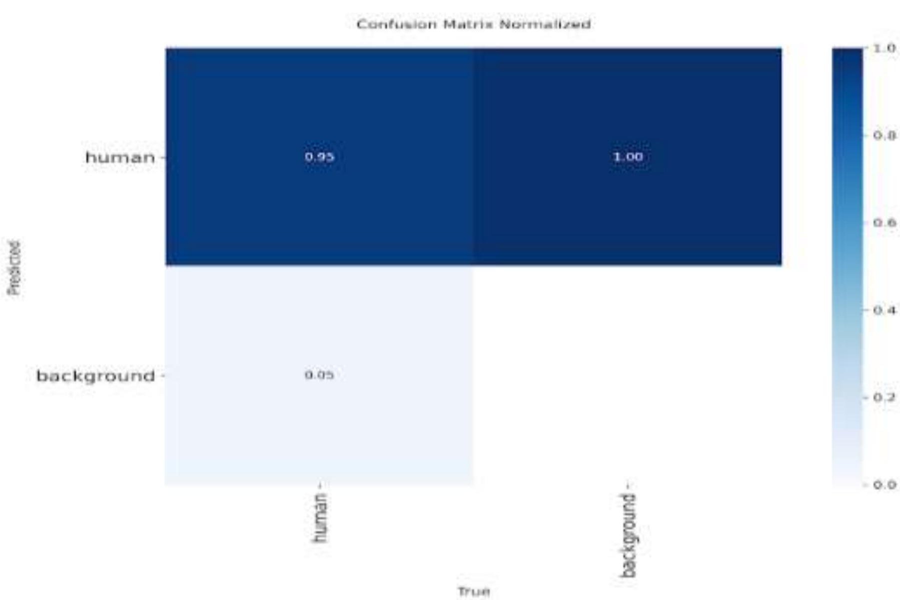

Fig .4. Confusion matrix

4.2 Precision-Recall Curve

The Precision–Recall curve demonstrates a near-rectangular shape concentrated toward the upper-right region, which indicates high precision maintained across a wide range of recall values. The Average Precision (AP) for the human class is approximately 0.967is shown in Fig. 5, confirming robust detection consistency across confidence thresholds.

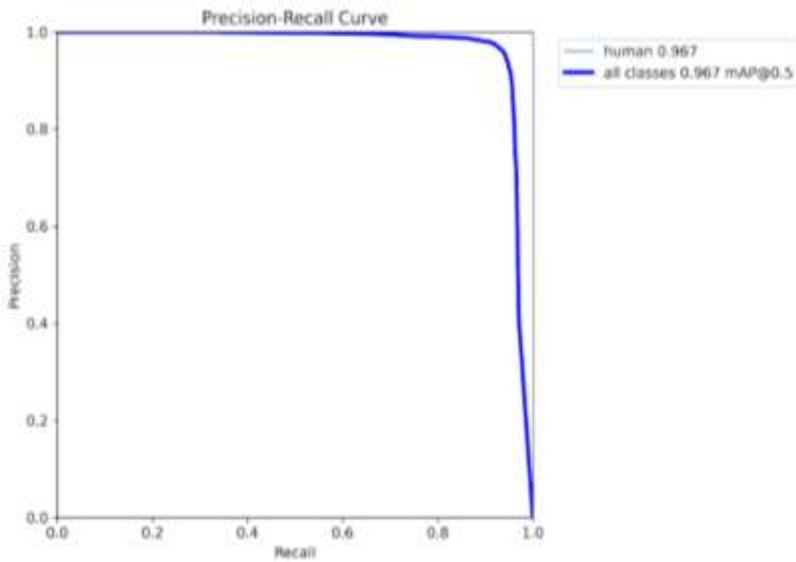

Fig .5. P-R curve

4.3 Integrated Performance Dashboard

An integrated performance dashboard was developed to monitor and manage real-time human detection events generated by the proposed UAV-based detection system. The dashboard provides a centralized interface that displays the processed live video feed along with detection outputs from the trained YOLOv8 model. Whenever a human is detected, a bounding box is drawn around the identified person in the video stream, and an alert mechanism is automatically activated. A beep sound is triggered at 5-second intervals to notify the operator of the detected human presence. Simultaneously, the system records the detection event by storing the corresponding date and time, alert level, and snapshot image of the detected frame as visual evidence.

The dashboard includes an Alert History module that maintains a structured log of all detection events. Each entry contains the detection timestamp, human presence status, alert level, and a reference to the stored snapshot image. This logging mechanism ensures traceability and enables post-event analysis. Additionally, all detected frames are automatically saved in a dedicated Snapshots section, where images are timestamped and linked to their respective alert records for documentation and reporting purposes.Dashboard for real-time detection system shown in Fig .6.

(a)

Alerts

Date	Time	Humans	Snapshot
2026-01-18	20:13:29	1	View
2026-01-18	20:13:29	1	View
2026-01-18	20:13:29	1	View
2026-01-18	20:13:29	1	View
2026-01-18	20:13:30	1	View
2026-01-18	20:13:30	1	View
2026-01-18	20:13:30	1	View
2026-01-18	20:13:31	1	View
2026-01-18	20:13:31	1	View
2026-01-18	20:13:31	1	View
2026-01-18	20:13:32	1	View
2026-01-18	20:13:32	1	View

Fig .6. Dashboard for Real-time detection

A map interface is also integrated into the dashboard to display the UAV's location during detection events, improving situational awareness and spatial tracking capabilities. By combining real-time monitoring, alert generation, event logging, visual evidence storage, and location mapping, the integrated performance dashboard enhances system usability and provides a comprehensive platform for intelligent UAV-based human surveillance and reporting.

4.4 Real-Time Detection Results

The proposed system was tested under real-time conditions using live video input. The trained YOLOv8 model successfully detected multiple human instances within a single frame, as shown in Fig. 5. Detected individuals are marked with bounding boxes and labeled accordingly, demonstrating effective multi-object detection capability.

Upon detection, the system automatically logs the event in the terminal and stores a timestamped snapshot of the frame for evidence and future analysis. The results confirm stable detection performance in dynamic scenes with multiple targets, validating the practical applicability of the proposed UAV-based surveillance system. Trained image and external image as shown in Fig .7 and Fig .8.

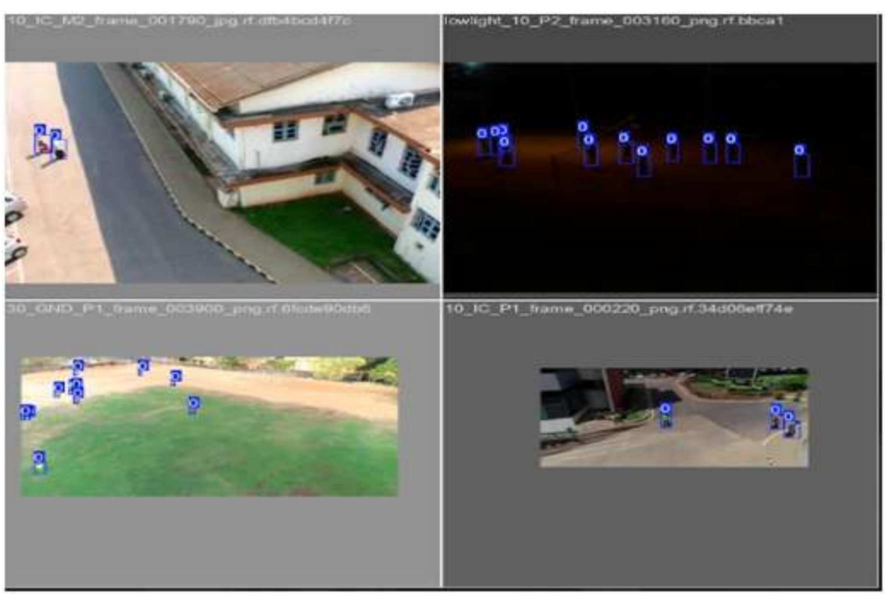

Fig. 7.Trained images detection result

(a)

(b)

Fig .8. External images detection result

5. CONCLUSION

In this work, a UAV-based human detection system was successfully developed and implemented using a trained YOLOv8 model for real-time surveillance. The system is capable of detecting multiple human targets from live video streams and generating instant alerts with bounding box visualization. Detection events are automatically logged with timestamps, and snapshots are stored for documentation and post-event analysis.

The integration of real-time monitoring, alert mechanisms, snapshot storage, and location tracking improves situational awareness and operational efficiency. Experimental results demonstrate reliable detection performance under practical conditions, validating the effectiveness of the proposed system for surveillance and security applications. Future improvements include integrating a GPS module for real-time location tracking and geo-tagging of detection events. The system range can be extended by enhancing communication modules. Additionally, increasing battery capacity will improve flight time for longer surveillance operations.

REFERENCES

[1]. K. Jayalath and S. R. Munasinghe, "Drone-based Autonomous Human Identification for Search and Rescue Missions in Real-time," in *Proc. IEEE Int. Conf. on Intelligent Computing and Applications (ICIAfS)*, Aug. 2021, pp. 1–6, doi: 10.1109/ICIAfS52090.2021.9606048.

[2]. U. Azmat, S. S. Alotaibi, M. Abdelhaq, N. Alsufyani, M. Shorfuzzaman, A. Jalal, and J. Park,"Aerial Insights: Deep Learning-Based Human Action Recognition in Drone Imagery," IEEE Access, vol. 11, 2023, doi: 10.1109/ACCESS.2023.3302353.

[3]. F. H. Zaman, N. T. Thamrin, N. Thari, and A. H. Hasmi, "Human Detection from Drone Using You Only Look Once (YOLOv5) for Search and Rescue Operation," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, May 2023.

[4]. G. Kucukayan and H. Karacan, "YOLO-IHD: Improved Real-Time Human Detection System for Indoor Drones," *Sensors*, vol. 23, 2023.

[5]. I. P. Sary, S. Andromeda, and E. U. Armin, "Performance Comparison of YOLOv5 and YOLOv8 Architectures in Human Detection Using Aerial Images," *Ultima Computing : Jurnal SistemKomputer*, vol. 15, no. 1, pp. 8–13, Jun. 2023, doi: 10.31937/sk.v15i1.3204.

[6]. H.-K. Jung and G.-S. Choi,"Improved YOLOv5: Efficient Object Detection Using Drone Images under Various Conditions," Applied Sciences, vol. 12, no. 7, p. 7255, 2022.

[7]. C. Jiang, H. Ren, X. Ye, J. Zhu, H. Zeng, Y. Nan, M. Sun, X. Ren, and H. Huo, "Object Detection from UAV Thermal Infrared Images and Videos Using YOLO Models," International Journal of Applied Earth Observation and Geoinformation, vol. 112, p. 102912, 2022.

[8]. A. Sangle, H. Bhanji, R. Vade, and S. S. Pansambal, "A Smart Human Detection Drone for Rescue," *SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology*, vol. 12, no. 1, pp. 16–19, 2020.

[9]. X. Zhai, Z. Huang, T. Li, H. Liu, and S. Wang, "YOLO-Drone: An Optimized YOLOv8 Network for Tiny UAV Object Detection," Sensors, vol. 24, no. 3, 2024, Art. no. 851.

[10]. C. Liu, F. Meng, Z. Zhu, and L. Zhou, "Object Detection of UAV Aerial Image Based on YOLOv8," Journal of Artificial Intelligence and Systems, vol. 5, no. 3, 2023.

[11]. W. Guettala, A. Sayah, L. Kahloul, and A. Tibramine, "Real-Time Human Detection by Unmanned Aerial Vehicles," *in Proc. IEEE Int. Conf. on Robotics and Automation (ICRA)*, Paris, France, pp. 123–128, 2023.

[12]. N. P. Motwani and S. Soumya, "Human Activities Detection Using Deep Learning Technique—YOLOv8," ITM Web of Conferences, vol. 56, Art. no. 03003, 2023, doi: 10.1051/itmconf/20235603003.

[13]. A. A. B. Abdelnabi and G. Rabadi, "Human Detection From Unmanned Aerial Vehicles' Images for Search and Rescue Missions: A State-of-the-Art Review," *IEEE Access*, vol. 12, pp. 21437–21461, 2024, doi: 10.1109/ACCESS.2024.3479988.

[14]. D. P. Simões, H. C. de Oliveira, and D. R. Pereira, "Unicamp-UAV: An open dataset for human detection in UAV imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 231, pp. 119–136, Jan. 2026, doi: 10.1016/j.isprsjprs.2025.10.026.

[15]. J. R. Terven and D. M. Córdova-Esparza, "A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction*, vol. 6, no. 1, pp. 1–27, 2024.

[16]. D. L., P. K. Megharaj, P. P., N. Kiran, A. R. K. P., and G. S., "State-of-the-art object detection: An overview of YOLO variants and their performance," in *Proc. 2023 4th Int. Conf. on Smart Electronics and Communication (ICOSEC)*, IEEE, 2023, pp. 1–6.

[17]. W. Yi and H. Guan, "BSF-YOLOv8: A Small Object Detection Algorithm for UAV Aerial Photography Based on YOLOv8," *IAENG International Journal of Computer Science*, vol. 51, no. 1, pp. 1–10, 2024.

[18]. M. Wu, W. Xie, X. Shi, P. Shao, and Z. Shi, "Real-Time Drone Detection Using Deep Learning Approach," in Proc. Int. Conf. Communication, Control, Computing and Electronics Engineering, 2018, doi: 10.1109/ICCOINS.2018.8510564.

[19]. M. N. Sulaiman, M. R. M. Razif, C. A. Z. C. Hassan, N. H. M. Mustapha, S. W. S. Azhar, and M. N. Azman,"Wireless Surveillance with Human Detection Using Artificial Intelligence and Drone," Journal of Telecommunication, Electronic and Computer Engineering, vol. 17, no. 3, pp. 51–58, 2025.

[20]. H. Chang and Z. Wang, "UAV Vehicle Detection System Based on YOLOv8," Journal of Physics: Conference Series, vol. 2872, 2024, Art. no. 012019.

[21]. W. Rahmaniar and A. Hernawan,"Real-Time Human Detection Using Deep Learning on Embedded Platforms: A Review," Journal of Robotics and Control, vol. 2, no. 6, Nov. 2021, doi: 10.18196/jrc.26123.

[22]. Z. Bao, "The UAV Target Detection Algorithm Based on Improved YOLOv8," in Proceedings of the International Conference on Image Processing, Machine Learning and Pattern Recognition (IPMLP), Guangzhou, China, Sept. 2024, pp. 264–269, doi: 10.1145/3700906.3700949.

[23]. C. Jiang, H. Ren, X. Ye, J. Zhu, H. Zeng, Y. Nan, M. Sun, X. Ren, and H. Huo, "Object Detection from UAV Thermal Infrared Images and Videos Using YOLO

[24]. M. E. Sahinet al.,"Deep Learning-Based Real-Time Drone Detection Using YOLO," Drones, vol. 7, no. 3, p. 190, 2023, doi: 10.3390/drones7030190.