Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An Intelligent YOLO26-Based Early Warning Framework for Drone Detection in Critical Infrastructure Surveillance

Younis Arrabi

younis135@yahoo.com

Article

Keywords: Drone Detection, YOLO26, Early Alerting, Object Detection, Deep Learning, Computer Vision, Critical Infrastructure Security, Real-Time Surveillance

Posted Date: July 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10259678/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

An Intelligent YOLO26-Based Early Warning Framework for Drone Detection in Critical Infrastructure Surveillance

Younis Arrabi

PhD in Information Technology, College of Informatics,

Midocean University, Comoros.

Email:younis135@yahoo.com

Funding

The author declares that no funding was received for this study.

Abstract

Objective: The main objective of this research is to design and implement a real-time system for the detection and alert of drones using YOLO26 deep learning framework, to increase the level of surveillance of critical objects of interest such as sensitive infrastructures, restricted areas, telecoms towers, airports and industrial sites.

Methodology: The study shall follow an experimental-based approach and design a custom-made drone image and video database, that comprises of public available drone

datasets. For the simulation of a realistic CCTV like environment, realistic surveillance video sequences as well as synthetically generated video as well as real-life video of different scenarios and environments shall be included in the database. The drone images as well as videos of different drones in various scenarios like very far objects, small objects, blurred motion, low light, cloudy background and confused objects (like birds or actual aircraft) shall be included in the database. After training the YOLO26 model for the detection of drones in videos, the performance of the model shall be compared with several baseline models like YOLOv8, YOLO11, Faster R-CNN as well as typical CNN-based object detection models. For the evaluation of the performance of the model, the custom database shall be divided into three subsets, i.e. training set, validation set, and test set via stratified sampling. The various performance metrics, including mAP, precision, recall, F1-score, false alarm rate, detection delay, FPS, and early alert time shall be calculated and compared.

Expected Results: By implementing the YOLO26 model, we aim to get high accuracy for the small objects like the drones, and sufficient processing speed for real time processing on the edge as well as on the CCTV system. The results will also show whether YOLO26 can improve the early detection of small as well as distant drones compared to the previous versions of YOLO.

Recommendations: The proposed AI-based drone detection and early alerting system is recommended for deployment in critical infrastructure surveillance environments such as telecom towers, airports, oil and gas facilities, restricted zones, and industrial sites. The system should be integrated with existing CCTV or surveillance camera networks to provide real-time drone detection, automated alert generation, and early warning before drones enter sensitive areas. Future work should focus on expanding the dataset to include night-time footage, long-distance drones, harsh weather conditions, and drone-like objects such as birds and aircraft to reduce false alarms. Further optimization is also recommended for edge devices to ensure low-latency operation and practical field deployment.

Keywords: Drone Detection; YOLO26; Early Alerting; Object Detection; Deep Learning; Computer Vision; Critical Infrastructure Security; Real-Time Surveillance.

Introduction

Unmanned aerial vehicles (UAVs), commonly known as drones, have experienced rapid growth and can now be used for a wide array of applications, including photography, package delivery, and agricultural monitoring and inspection. However, with great technology comes great risk, and the unauthorized use of drones to conduct operations in restricted or sensitive areas, near airports, critical infrastructure, oil and gas sites, military installations, cell towers, and large public gatherings can pose serious security, safety, and business risks and allow for potential surveillance, smuggling, sabotage and service disruption among others. The use of drones around aerodromes is considered a safety concern and can occur by accident, by intent, in a criminal manner or even with terrorist motivation (European Union Aviation Safety Agency [EASA], 2021), and therefore the development of reliable, automated and real-time drone detection systems has become a pressing need for research and for operational use.

There are several types of detection systems for drones. The most common are radar systems for long distance detection, radio-frequency (RF) monitoring systems for identifying drones that communicate with their controllers, and acoustic sensors that can pick up the engine sound of a drone. There are also infrared and optical cameras, which can provide visual detection and also be used for classification. In addition, there are multi-sensor systems that use a combination of the above-mentioned sensors for better detection performance. Each type of sensing has its own performance limitations, and can also have a high number of false alarms, such as due to birds or other small flying objects. In this sense, active radar, passive radar, RF analysis, acoustic detection, image-based detection, and also fusion is used for counter-UAS detection. As mentioned before, each of them has its own set of deployment constraints, which makes difficult to choose the most suitable one for a given scenario, Besada et al. (2022) noted that counter-UAS detection can rely on active radar, passive radar, RF analysis, acoustic detection, image-based detection, or fusion, but each sensing approach involves different performance limitations, false-alarm behavior, and deployment constraints.

Computer vision-based approaches are also very promising as they can leverage on existing surveillance systems such as CCTV, PTZ and fixed cameras. As long as a computer vision model is trained and deployed on the video stream captured by these sensors, there is no need to install any additional sensors on the drones or their operators. Drone detection in video streams, however, is challenging as drones are typically very small objects, far away from the cameras, moving fast, appearing under motion blur, often appearing on top of clouds or complex background, and can easily be confused with other

small flying objects such as birds, model aircraft, etc., as well as with other moving objects such as people, vehicles, etc. The problem therefore is a small-object detection problem in real-time, which requires both high recall and low false-positive rates. False negatives could lead to delayed reaction while false positives could lead to decreased trust of the system by its operators.

Many deep learning models have been proposed to detect objects in images and videos. Recent years have seen a dramatic increase in the development of deep learning-based object detection models. Among these, the You Only Look Once (YOLO) family of object detection models treat object detection as a regression problem. In YOLO, objects are detected directly from full images in a single evaluation, thereby making the model highly suitable for a wide range of real-time object detection applications. The YOLO models have continued to evolve in recent years with significant improvements in terms of detection accuracy and deployment efficiency, and have been applied to numerous surveillances, industrial monitoring, and other safety-critical applications.

According to Jocher et al. (2026), YOLO26 is the newest generation of real-time object detection models. They can be deployed on a wide variety of hardware platforms and are designed to be extremely efficient. YOLO26 has a dual-head architecture which allows it to perform end-to-end object detection without requiring non-maximum suppression. This makes it especially suitable for object detection in video, such as in surveillance scenarios. The detection-head of YOLO26 does not use Distribution Focal Loss, which makes the detection-head of YOLO26 less complex. It uses Progressive Loss and STAL (Small Target Aware Label Assignment) to improve the learning of small objects.

Despite the numerous implementations of the various approaches for the detection of drones there are a few open research questions that need to be addressed. The first open research question is how to enhance the small and distant object detection in various different scenarios. This includes scenarios under various weather and lighting conditions and different camera angles. As mentioned before many of the current approaches focus on the detection of objects and do not include the early alerting logic for the various different operational scenarios. Also, most of the approaches for the detection of objects do not consider the real-time aspect of the alerting process. This implies that the system needs to be able to generate an early warning in real-time and that the system needs to be able to notify the operator of the detected drone in real-time as well. Another important

open research question is how to compare the new approaches for real-time object detection with the more established approaches. An important aspect in the comparison of the approaches is the feasibility of the approach for real-world deployment. As mentioned before the system needs to process the video frames in real-time and the system needs to be able to run on various different edge-devices. The performance of the system also needs to be sufficient to process the video in real-time while maintaining a high number of correct detections and a low false alarm rate. In order to be able to compare the approaches for the detection of drones several metrics need to be used. First of all the standard metrics for object detection such as precision, recall, F1-score and mean average precision need to be used in order to compare the approaches for the detection of drones. In addition to the standard metrics that are used for the object detection also the latency, the false alarm rate, the number of frames that are required to detect the drone and the number of frames that are required to remove the alert from the system need to be taken into account. All of these metrics need to be sufficient in order to provide an adequate evaluation of the performance of the system.

This study will propose a novel framework based on a YOLO26 detector for real-time small and distant objects such as drones detection and provide early alerts. The study will set up a novel dataset for the small object such as a drone for detection. Also, it will utilize various data augmentation techniques to simulate the complex scenarios in surveillance video such as variable lighting, weather, angle of view, etc. The novel framework will be trained to classify images or video frames into objects of interest (drones) and background. The framework will be compared with various existing detectors that are suitable for real-time small object detection. The study will not only employ the metrics commonly used to evaluate object detection such as precision, recall, F1-score, and average precision (AP) for object of interest, but also other metrics such as detection latency, false positive per frame (FPFF), frame persistence, and number of early warnings generated. In summary, the expected contribution of this study is the development of an intelligent computer vision-based framework that is capable of implementing proactive drone detection and providing early warning for surveillance of critical infrastructure.

Problem Statement:

Traditional drone monitoring relies on radar, RF sensors, acoustic sensors, or manual CCTV observation. These approaches have limitations. Radar and RF systems can be

costly, acoustic detection may be affected by environmental noise, and manual CCTV monitoring is reactive and not scalable. Computer vision offers a practical alternative because many sites already have surveillance cameras, but detecting drones remains challenging due to their small size, high speed, background similarity, lighting variation, and similarity to birds or aircraft.

Research Objectives

1. To design and annotate a drone detection dataset covering multiple drone types, distances, backgrounds, and environmental conditions.
2. To train a YOLO26-based model for real-time drone detection in image and video streams.
3. To develop an early alerting logic that generates alerts based on detection confidence, object movement, distance estimation, and repeated frame confirmation.
4. To compare YOLO26 performance against YOLOv8 and other baseline models using accuracy and latency metrics.
5. To assess the feasibility of deploying the model on edge devices for critical infrastructure surveillance.

Literature Review

Drone Detection and Security Challenges

The increasing use of unmanned aerial vehicles (UAVs), commonly known as drones, has created significant opportunities in areas such as logistics, agriculture, inspection, mapping, emergency response, and infrastructure monitoring. However, drones also introduce serious security and operational risks when they are used in unauthorized or restricted areas. Critical infrastructure, airports, military facilities, oil and gas plants, telecom towers, and public-event locations may be exposed to surveillance, smuggling, service disruption, or physical threats caused by unauthorized drone activity. As a result, reliable drone detection has become an important research area within security surveillance and counter-UAS systems.

Drone detection is challenging because drones are often small, fast-moving, and visually similar to birds or other flying objects. In video surveillance, drones may appear as very small targets, especially when they are far from the camera. Their appearance can also be

affected by lighting conditions, weather, motion blur, camera angle, background complexity, and low image resolution. These conditions increase the probability of missed detections and false alarms. Therefore, an effective drone detection system must not only identify drones accurately but also operate in real time and generate early alerts before the drone reaches a sensitive or restricted zone.

Sensor-Based Drone Detection Methods

Several sensor-based technologies have been proposed for drone detection, including radar, radio-frequency (RF) detection, acoustic detection, infrared sensing, and multi-sensor fusion. Radar-based systems are widely used because they can detect flying objects over long distances and can operate during day and night conditions. Radar can also estimate object range and velocity, which makes it useful for tracking airborne objects. However, small drones usually have a low radar cross-section, fly at low altitude, and move at relatively low speed compared with conventional aircraft. These characteristics make drone detection difficult and may increase confusion with birds or other small objects (Besada et al., 2022).

RF-based systems detect communication signals between the drone and its controller. This approach can be effective when drones rely on active communication links. However, RF detection may be limited when drones operate autonomously, use encrypted communication, frequency hopping, or non-standard transmission methods. Acoustic detection uses the sound signature of drone motors and propellers. It is relatively low cost and passive, but its performance can be affected by wind, traffic, industrial noise, and distance. Infrared and thermal sensors can support night-time detection, but they require specialized equipment and may be affected by environmental temperature and weather conditions.

Multi-sensor fusion has been proposed to overcome the weaknesses of individual sensors. By combining radar, RF, acoustic, optical, and thermal sensors, detection systems can improve reliability and reduce false alarms. However, multi-sensor systems are usually more complex, expensive, and difficult to deploy at scale. For many organizations, especially those that already have CCTV or surveillance cameras, computer-vision-based drone detection provides a more practical and scalable approach.

Computer Vision-Based Drone Detection

Computer vision-based drone detection uses images or video frames to identify drones visually. This approach is attractive because many facilities already have existing surveillance cameras that can be reused for automated monitoring. Unlike RF or acoustic sensors, vision-based systems can provide visual confirmation, object localization, and

evidence through recorded footage. They can also be integrated with video management systems and security operation centers.

Early computer vision approaches relied on handcrafted features, background subtraction, motion detection, and classical machine learning classifiers. Although these methods were useful in controlled environments, they were often sensitive to illumination changes, camera movement, complex backgrounds, and object scale variation. The introduction of deep learning significantly improved object detection by allowing models to learn visual features automatically from large image datasets. Convolutional Neural Networks (CNNs), Faster R-CNN, Single Shot MultiBox Detector (SSD), and YOLO-based detectors have been widely applied to visual object detection tasks.

For drone detection, computer vision still faces several technical challenges. First, drones usually occupy a very small portion of the image, making feature extraction difficult. Second, drones may appear against complex backgrounds such as clouds, buildings, trees, mountains, or desert environments. Third, birds and distant aircraft may create false positives. Fourth, real-time operation is required because late detection may reduce the value of the warning. These challenges make drone detection a small-object, real-time, and low-false-alarm detection problem.

Deep Learning and YOLO-Based Object Detection

Deep learning object detectors are generally divided into two-stage and one-stage detectors. Two-stage detectors, such as Faster R-CNN, first generate candidate object regions and then classify them. These models can achieve high accuracy but may require more computational time, making them less suitable for real-time surveillance. One-stage detectors, such as SSD and YOLO, predict object classes and bounding boxes directly in a single pass, which makes them more suitable for real-time applications.

The YOLO family has become one of the most widely used real-time object detection approaches. Redmon et al. (2016) introduced YOLO as a unified detection model that frames object detection as a regression problem, predicting bounding boxes and class probabilities directly from full images. This design allows fast inference and makes YOLO suitable for real-time video analysis. Later YOLO versions improved detection accuracy, feature extraction, multi-scale detection, anchor-free prediction, and deployment flexibility.

YOLO models are particularly relevant to drone detection because the task requires fast processing of video streams. A drone detection system must identify small airborne targets quickly enough to generate useful alerts. However, conventional YOLO models may still struggle with very small drones, especially when the target is distant, partially

blurred, or similar to the background. Therefore, recent research has focused on improving small-object detection through enhanced feature pyramids, attention mechanisms, lightweight backbones, higher input resolution, and improved loss functions.

YOLO26 for Drone Detection

YOLO26 represents a recent advancement in real-time object detection. It was designed to improve accuracy, efficiency, and deployment readiness across different hardware platforms. YOLO26 introduces native non-maximum suppression-free end-to-end inference, a lighter detection head, Progressive Loss, the MuSGD optimizer, and Small Target Aware Label Assignment (STAL). These features are important for drone detection because drones often appear as small targets and require fast inference for early warning.

The Small Target Aware Label Assignment mechanism is especially relevant to drone detection because small drones may otherwise receive insufficient positive label assignment during training. Progressive Loss can improve training stability by shifting supervision toward the inference-time detection head. In addition, NMS-free inference may reduce post-processing latency and support faster real-time deployment. Therefore, YOLO26 provides a strong technical foundation for developing a drone detection and early alerting framework.

Early Alerting in Drone Detection Systems

Most drone detection studies focus on whether the drone is detected or not. However, for operational security, detection alone is not sufficient. A practical system should also generate early alerts based on detection confidence, repeated-frame confirmation, movement direction, and proximity to restricted zones. Early alerting is important because security teams need time to verify the threat, activate response procedures, and prevent the drone from entering sensitive areas.

An early alerting framework can classify alerts into different levels. For example, a single low-confidence detection may be treated as an observation, while repeated detection across several frames may trigger a warning. If the drone moves toward a restricted area, the system may escalate the warning to a high alert. If the drone enters the restricted zone, the system may generate a critical alert. This approach can reduce unnecessary false alarms while maintaining operational readiness.

Object tracking can further improve early alerting by confirming drone movement across frames. Tracking also helps distinguish real drones from temporary visual noise or

isolated false detections. Therefore, combining YOLO26 detection with frame persistence, object tracking, and rule-based alert escalation can provide a more operationally useful system than detection alone.

Related Work and Research Gaps

Previous studies have shown that drone detection can be performed using radar, RF, acoustic, infrared, visual, and multi-sensor methods. Sensor-based approaches provide useful capabilities but may suffer from cost, environmental limitations, false alarms, and deployment complexity. Vision-based approaches are more scalable when surveillance cameras already exist, but they face challenges related to small-object detection, background complexity, and real-time processing.

Recent YOLO-based studies have improved drone and UAV detection accuracy, especially through lightweight models and small-object optimization. However, several research gaps remain. First, many studies focus on detection accuracy but do not fully address early alerting and operational warning logic. Second, the detection of small and distant drones remains a major challenge in complex surveillance environments. Third, many models are evaluated using standard metrics such as mAP, precision, recall, and F1-score, but fewer studies evaluate latency, false alarm rate, frame persistence, and edge-device readiness. Fourth, there is still a need to evaluate newer architectures such as YOLO26 for drone detection, particularly because of its small-target-aware training and real-time deployment features.

This study addresses these gaps by proposing a YOLO26-based drone detection and early alerting framework. The study will evaluate detection performance using standard object detection metrics and operational metrics such as inference latency, frames per second, false alarm rate, and early alert generation. The contribution of this study is not limited to detecting drones but extends to developing a practical early warning framework suitable for critical infrastructure surveillance.

Table 1: Synthesis of Related Works on Drone Detection and Early Alerting Using Deep Learning

Year	Authors	Algorithm / Method	Dataset	Key Metrics	Classes	Gaps Identified
2016	Redmon et al.	YOLO real-time object detection	COCO / general object detection dataset	Real-time detection speed and competitive accuracy	Multiple object classes	General object detection model; not specifically optimized for drone

						detection or early warning
2022	Besada et al.	Counter-UAS sensor review: radar, RF, acoustic, image-based, and fusion methods	Review-based study	Detection capability, sensor limitations, false alarm behavior	UAV / non-UAV	Focused on sensor technologies; limited experimental evaluation of real-time vision-based early alerting
2023	Seidaliyeva et al.	Drone detection and classification techniques review	Review-based study	Comparative analysis of detection methods	UAV types and non-UAV objects	Review study only; no new YOLO-based experimental model or operational alerting framework
2024	Liu et al.	Vision-based drone detection in complex environments	Drone image/video dataset	Precision, recall, mAP	Drone / background	Detection focus; limited emphasis on early warning logic and real-time operational deployment
2024	Zhao et al.	Lightweight UAV small-target detection model	UAV small-object dataset	mAP, FPS, model size	UAV / background	Improved small-object detection, but limited early alerting and critical infrastructure application
2026	Jocher et al.	YOLO26 real-time end-to-end vision model	General object detection benchmarks	mAP, latency, efficiency	Multiple object classes	Advanced architecture, but requires domain-specific validation for drone detection and early alerting
Present Study	Proposed Research	YOLO26-based drone detection and early alerting framework	Custom drone surveillance dataset	mAP, precision, recall, F1-score, FPS, latency, false alarm rate, early alert time	Drone, bird, aircraft, helicopter, background object	Addresses prior gaps by combining YOLO26, small-object drone detection, real-time inference,

						and operational early alerting

Methodology

This study uses a research design of experimental research. The independent variable is the detection architecture (YOLO26 vs. CNN baseline) while mAP, Precision, Recall, and F1-Score for an held-out test set are the dependent variables. All the experimental parameters are fully specified below for reproducibility.

YOLO26

Ultralytics YOLO26 is a unified family of real-time computer vision models introduced in the Ultralytics YOLO26 paper. It supports multiple vision tasks, including object detection, instance segmentation, pose estimation, image classification, and oriented object detection. The model introduces native end-to-end inference, a lightweight detection head, an enhanced training strategy, and task-specific heads designed to improve accuracy, efficiency, and deployment flexibility (Ultralytics, 2026).

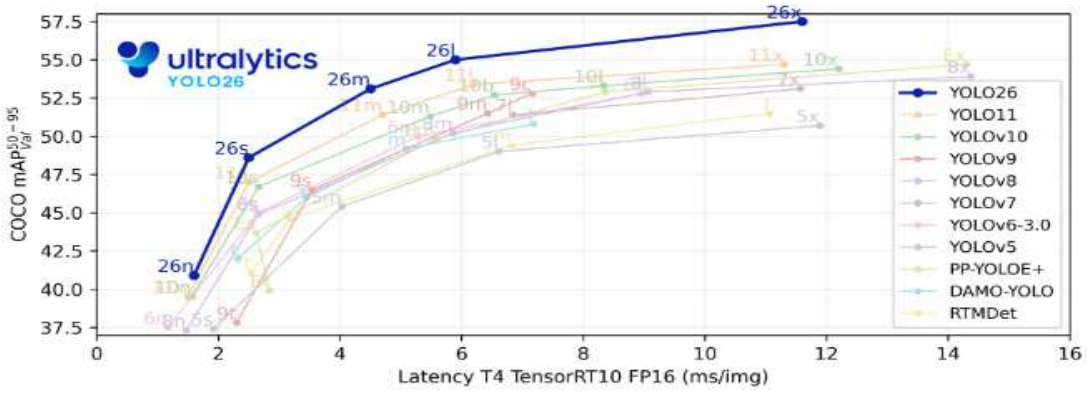


Figure 1:YOLO26 performance compared with older versions.

Dataset Construction and Annotation

A custom dataset was constructed from publicly available drone image repositories and surveillance video frames to ensure diversity in drone types, background conditions, lighting variations, object scales, and viewing angles. The dataset included images captured under different environmental conditions, such as clear sky, cloudy background, urban areas, open fields, industrial sites, and long-distance surveillance

views. Since drones often appear as small and fast-moving objects in video streams, special attention was given to collecting images with different drone sizes, altitudes, orientations, and partial occlusion conditions.

The initial collection contained approximately 3200 annotated drone-related instances belonging to five classes: drone, bird, aircraft, helicopter, and background flying object. These classes were selected to improve the model's ability to distinguish drones from visually similar airborne objects and reduce false alarms during real-time surveillance. All images were manually annotated using axis-aligned bounding boxes with Roboflow, which enforced a consistent YOLO-format label schema consisting of class index, normalized center-x, normalized center-y, width, and height.

Figures 2 and 3 show representative examples of the annotation process. Figure 2 presents a single-class annotation in which a drone is precisely bounded within the image frame. Figure 3 illustrates a multi-class annotation scenario where a drone and a visually similar airborne object, such as a bird or aircraft, are annotated separately. This type of annotation is critical for real-world drone detection because surveillance systems must accurately identify unauthorized drones while minimizing false alarms caused by birds, aircraft, helicopters, or other background objects.



Figure 2: Single-class drone annotation

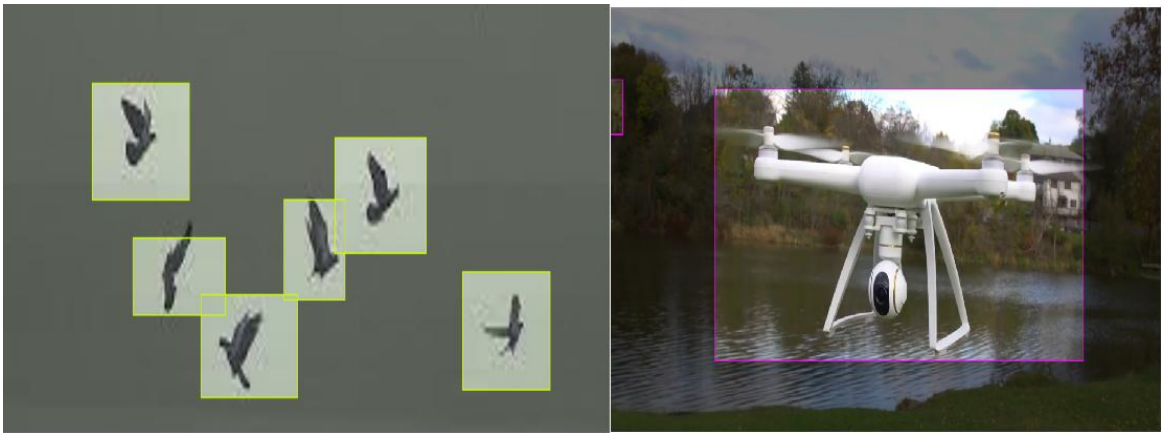


Figure 3: Multi-class drone vs bird annotation

The dataset was divided into training, validation, and testing subsets using a 70%, 20%, and 10% split, respectively. Stratified splitting was applied to maintain a similar class distribution across the three subsets and ensure that each class, including drone, bird, aircraft, helicopter, and background flying object, was proportionally represented.

Figure 4 presents a sample mosaic of the training batch, illustrating the diversity of drone appearances under different surveillance conditions. The sample includes variations in drone size, altitude, orientation, distance from the camera, background type, and lighting conditions. These examples demonstrate the complexity of real-world drone detection, where drones may appear as small objects against sky, urban, desert, industrial, or cloudy backgrounds.

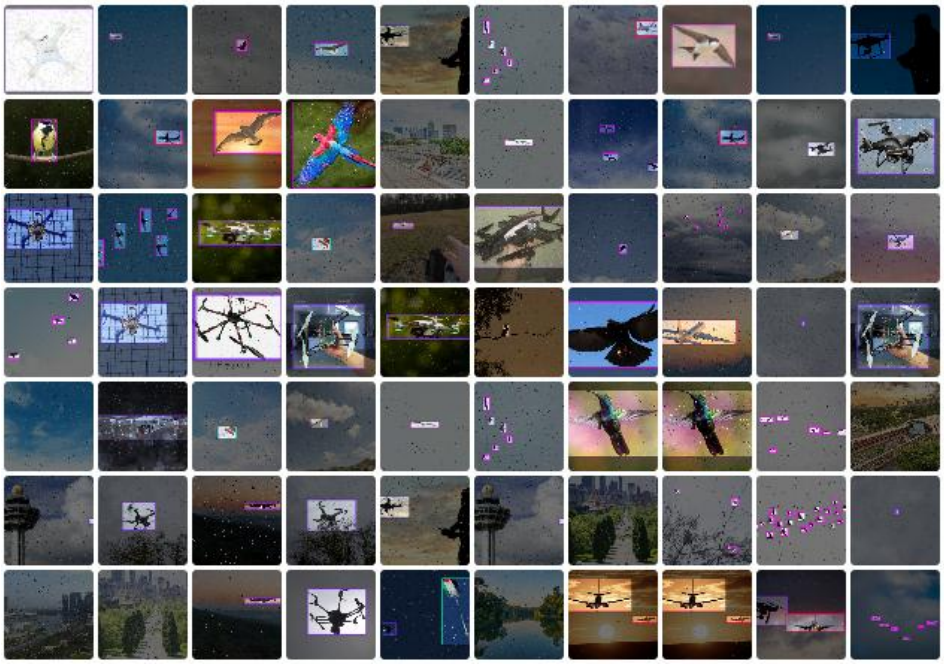


Figure 4: Training batch sample

Figure 5 shows a validation batch with corresponding YOLO26 prediction results overlaid on the images. High confidence values for drone detections indicate that the model has learned relevant visual features and generalized well to unseen validation images. The presence of drone-like objects, such as birds, aircraft, and helicopters, in the validation set is important for evaluating the model's ability to distinguish unauthorized drones from visually similar airborne objects and reduce false alarms during early alerting.



Figure 5: Validation batch with predictions

Data Augmentation Pipeline

To enhance the model's generalization capability and overcome the limited size of the base dataset, a systematic data augmentation pipeline was applied to the training set only. The original dataset of approximately 3200 annotated drone-related instances was expanded to around 10000 images through multiple augmentation techniques. These augmentations were implemented online during training rather than as a fixed offline set of pre-generated images. This approach ensured that each training epoch exposed the model to a statistically varied version of the input images, improving robustness and reducing the risk of overfitting.

Noise injection, as illustrated in Figure 6, was applied to simulate Gaussian sensor noise commonly observed in low-cost CCTV and surveillance cameras, especially under low-light or long-distance monitoring conditions. This augmentation is particularly important for drone detection because drones often appear as small objects in the frame, and image noise can obscure their shape, edges, and rotor structure. By training the model with noisy image samples, YOLO26 learns to extract reliable drone-related visual features

even under degraded video quality, poor visibility, and challenging surveillance conditions.



Figure 6: Noise Augmentation - Gaussian noise is added over the image

Greyscale conversion, shown in Figure 7, was applied to simulate monochrome surveillance cameras and black-and-white night-vision systems commonly used in security monitoring. This augmentation helps improve drone detection across different camera types and lighting conditions.

Training YOLO26 on greyscale images encourages the model to focus on shape, edges, rotor structure, contours, and texture rather than color alone. This improves detection performance in low-light, cloudy, distant, or degraded scenes and enhances the model's ability to generalize across real-world surveillance environments.

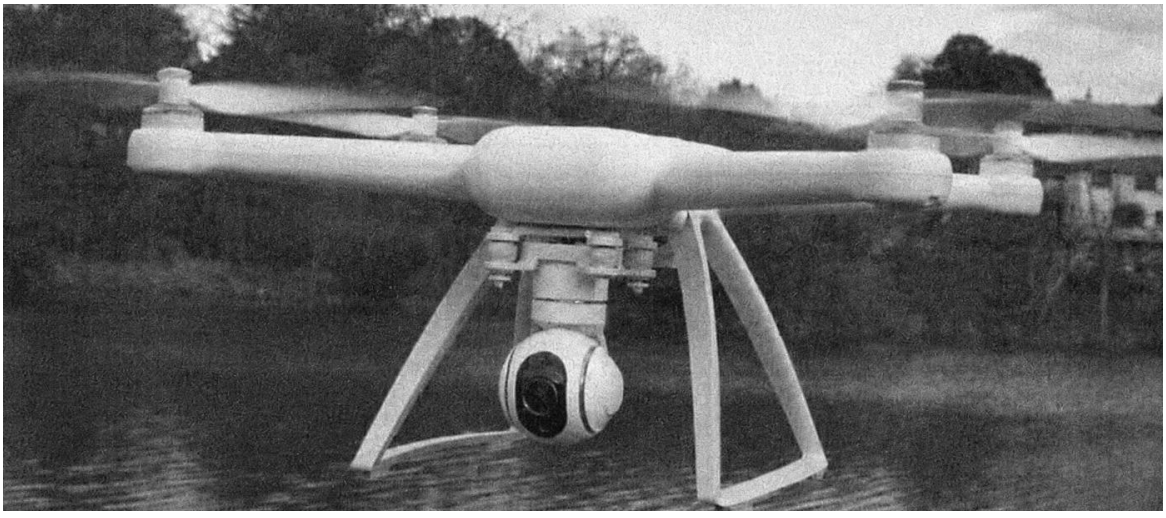


Figure 7. Greyscale Augmentation – The drone image is converted into monochrome format to simulate black-and-white surveillance cameras and night-vision

video. This augmentation helps the YOLO26 model learn drone-related shape and texture features independently of color information.

Baseline Architecture: Convolutional Neural Network

A standard Convolutional Neural Network (CNN) architecture was used as a baseline model to benchmark the performance of the proposed YOLO26-based drone detection system. The CNN architecture consisted of convolutional layers for feature extraction, max-pooling layers for reducing spatial dimensions, and fully connected layers for final classification. CNN-based deep learning became widely recognized after demonstrating strong image classification performance on the ImageNet dataset (Krizhevsky et al., 2012).

For drone object detection, the CNN baseline followed a two-stage detection pipeline. In the first stage, candidate regions likely to contain flying objects were generated using region proposal methods. In the second stage, these candidate regions were classified by the CNN as drone, bird, aircraft, helicopter, or background object. This approach is conceptually similar to the R-CNN detection paradigm, where object localization and classification are performed through sequential region proposal and CNN-based classification stages (Girshick et al., 2014).

The CNN baseline was selected because it represents one of the foundational deep learning approaches used in earlier computer vision and object detection studies. However, due to its two-stage processing structure, CNN-based detection is generally slower than one-stage detectors such as YOLO. This limitation is particularly important in drone surveillance, where real-time detection and early alerting are required. Therefore, comparing YOLO26 with a CNN baseline provides an evidence-based assessment of whether YOLO26 offers superior accuracy, lower latency, and better suitability for real-time drone detection in surveillance video streams.

Training Configuration

The YOLO26 model was trained using Google Colaboratory with an NVIDIA T4 GPU with 16 GB VRAM, providing the parallel processing capability required for efficient deep learning model training. Training was conducted for 50 epochs with an initial learning rate of 0.01 and cosine annealing scheduling to support stable convergence during optimization. The batch size was set to 16, depending on GPU memory availability and image resolution.

Pre-trained COCO weights were used for model initialization through transfer learning. This approach helps accelerate model convergence and improves generalization when training on a domain-specific drone detection dataset, particularly when the available dataset size is limited (Wang et al., 2021). Early stopping with a patience value of 10 was applied to reduce the risk of overfitting and ensure that training stopped when validation performance no longer improved.

Since drone detection is a small-object detection problem, the training configuration also emphasized image-scale variation, augmentation, and validation monitoring. These settings were important to improve YOLO26 performance in detecting drones under different surveillance conditions, including long-distance views, low-resolution frames, motion blur, variable lighting, and complex backgrounds.

Evaluation Metrics

Model performance was evaluated using four complementary object detection metrics: Mean Average Precision (mAP), Precision, Recall, and F1-Score. These metrics were selected to provide a comprehensive assessment of YOLO26 performance in detecting drones and distinguishing them from visually similar airborne objects such as birds, aircraft, and helicopters.

Mean Average Precision (mAP) was used as the main detection performance metric. It measures the area under the Precision–Recall curve across all target classes and Intersection over Union (IoU) thresholds. In this study, mAP provides an overall indication of how accurately the model detects and localizes drone-related objects in surveillance images and video frames.

Precision was calculated as: $\text{Precision} = TP / (TP + FP)$

Recall was calculated as: $\text{Recall} = TP / (TP + FN)$

The F1-Score was calculated as the harmonic mean of Precision and Recall:

$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$

The F1-Score provides a balanced measure of detection performance, particularly when both false alarms and missed drone detections are operationally important. These metrics were calculated for each class individually and also as aggregate values across all classes to evaluate the overall effectiveness of the YOLO26-based drone detection model.

Results and Discussion:

The mAP learning curve for the YOLO26 model across 50 training epochs is presented in Figure 08. The model initially achieved an mAP of 85.5% at epoch 1. During the early training phase, a temporary decline to approximately 62% was observed at epoch 3, which can be attributed to the interaction between the learning rate schedule, augmentation pipeline, and the model's adjustment from pre-trained COCO weights to the drone detection domain. After this short adaptation phase, the model rapidly improved, exceeding 95% mAP by epoch 5 and reaching 98.42% at epoch 50, with a peak mAP of 98.91% observed at epoch 37.

This "dip-then-rise" pattern is commonly observed in transfer learning scenarios, especially when a model trained on general object categories is fine-tuned on a specialized dataset with different visual characteristics. In this study, drone objects are often small, distant, and visually similar to birds, aircraft, or background noise, which requires the model to adapt its learned feature representation to the new domain. The temporary decline therefore reflects the initial cost of domain adaptation before the network stabilizes and learns drone-specific features.

The absence of divergence between early and later epochs indicates stable convergence and suggests that the augmentation strategy, transfer learning initialization, and early-stopping regularization helped prevent overfitting. Overall, the learning curve demonstrates that YOLO26 successfully learned robust drone detection features and maintained high performance throughout the training process.

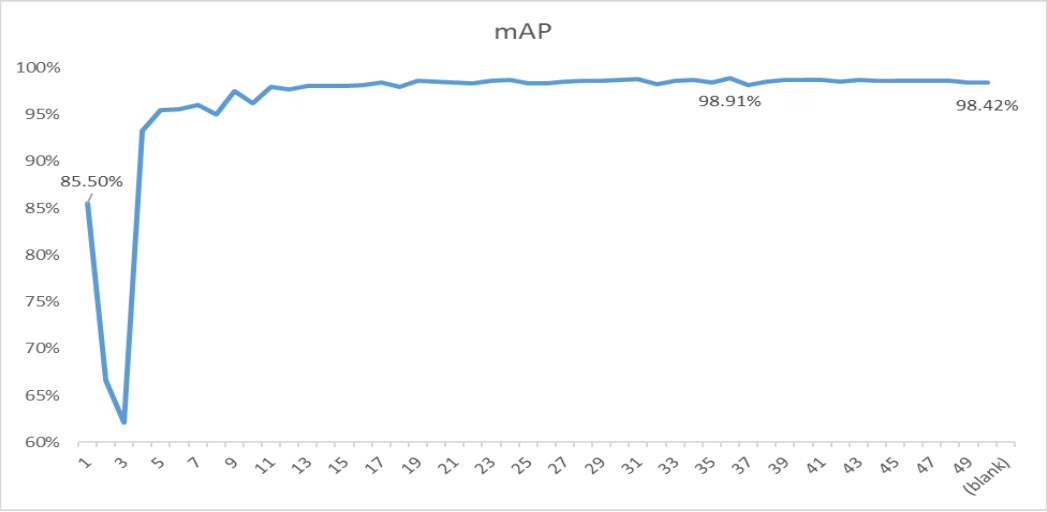


Figure 08. mAP Learning Curve for YOLO26 Drone Detection Over 50 Epochs – The curve shows an initial domain-adaptation dip during the early epochs, followed by rapid improvement and stable

convergence. The final mAP reached 98.42%, with a peak value of 98.91% at epoch 37, indicating strong detection performance and effective generalization.

The Precision learning curve is shown in Figure 09. Precision started at approximately 0.77 during the first epoch, then temporarily decreased to around 0.67 at epoch 3, following a similar pattern to the mAP curve. This early decline can be explained by the model's initial adjustment from general COCO pre-trained features to the drone detection dataset, where drones appear as small, distant, and visually challenging objects.

After this short adaptation stage, precision increased rapidly and stabilized above 0.97 from approximately epoch 7 onward. The smooth plateau observed between epochs 10 and 50 indicates that the YOLO26 model consistently produced highly accurate positive detections. In practical terms, most of the bounding boxes predicted by the model correctly contained drones or relevant airborne objects rather than background regions incorrectly classified as drones.

This high precision is especially important for drone detection and early alerting because false positives, such as birds, aircraft, clouds, or background noise being misclassified as drones, may create unnecessary alarms and reduce operator confidence in the system. Therefore, the stabilized precision curve confirms that the proposed YOLO26 model is effective in minimizing false drone alerts while maintaining reliable detection performance.

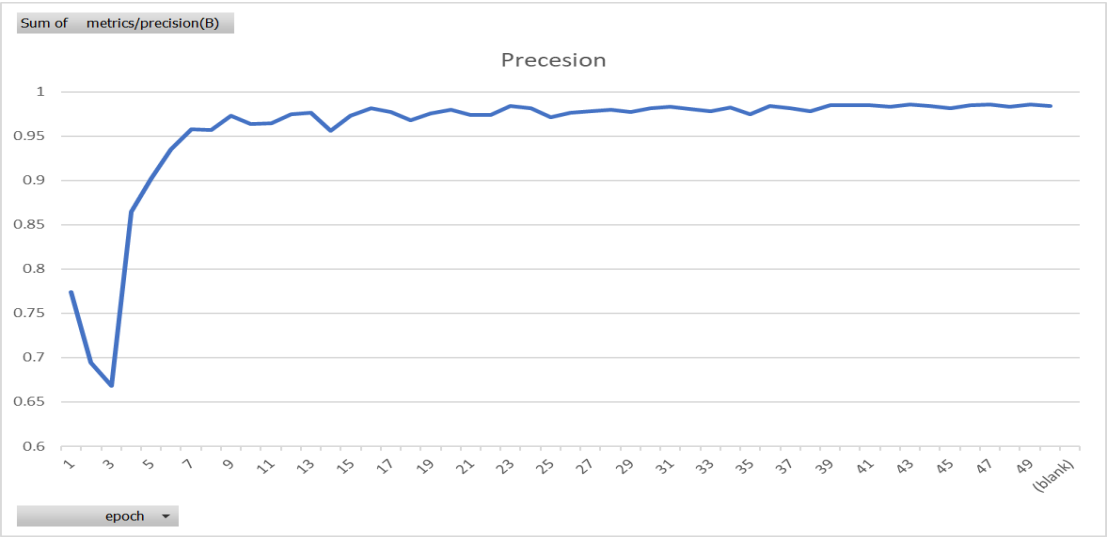


Figure 09. Precision Learning Curve for YOLO26 Drone Detection Over 50 Epochs – Precision initially declined during early domain adaptation, then rapidly improved and stabilized above 0.97 from epoch 7 onward, indicating a low false-positive rate and reliable drone detection performance.

The Recall learning curve is presented in Figure 10. Recall started at approximately 0.85 during the first epoch, then temporarily declined to around 0.64 at epoch 3, following the same early domain-adaptation pattern observed in the mAP and precision curves. After this initial adjustment phase, recall improved gradually, reaching approximately 0.95 by epoch 15 and stabilizing at around 0.983 by epoch 50.

The slower convergence of recall compared with precision is expected in drone detection tasks. During the early training stages, the model first learns to make confident detections for clearer and easier examples, such as large or close-range drones against simple backgrounds. As training progresses, the model begins to recognize more difficult cases, including small distant drones, partially blurred drones, drones against cloudy or complex backgrounds, and drone-like objects such as birds or aircraft.

The final recall value of approximately 98.2% indicates that the YOLO26 model missed fewer than 2 out of every 100 actual drone-related instances in the test set. This is a strong result for a multi-class airborne object detection task, especially where drones may appear as small objects under challenging surveillance conditions. High recall is particularly important for early alerting because missed detections may delay warning generation and reduce the effectiveness of the security response.

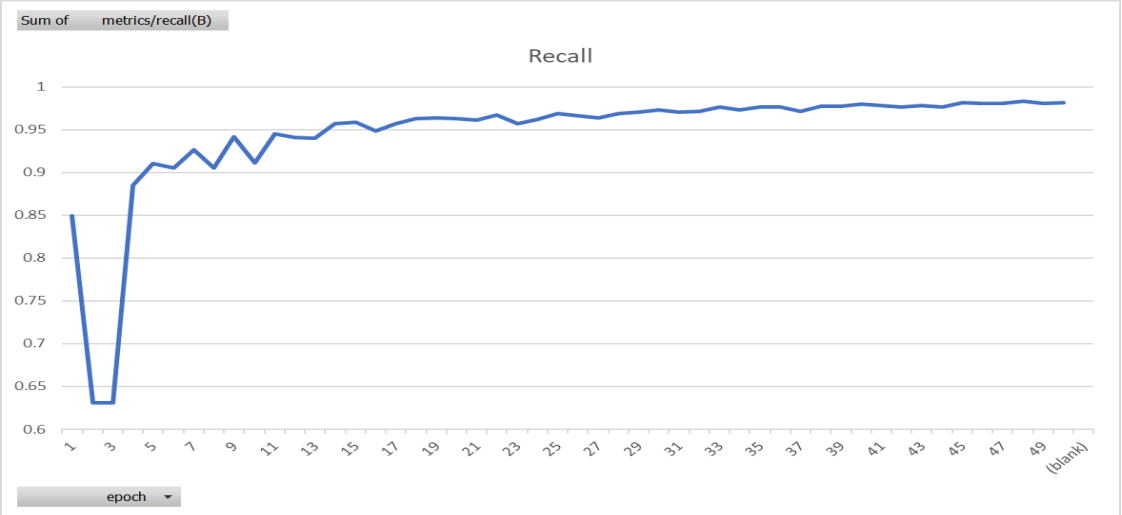


Figure 10. Recall Learning Curve for YOLO26 Drone Detection Over 50 Epochs – Recall showed an initial decline during early domain adaptation, then gradually improved and stabilized at approximately 0.983 by epoch 50, indicating strong capability in detecting true drone instances and minimizing missed detections.

Comprehensive Training Dashboard

The complete training and validation results for the YOLO26 model over 50 epochs are shown in Figure 11. The full training dashboard includes loss curves and performance metrics that provide a comprehensive view of the model's learning behavior throughout the training process.

The training loss panels, including box loss, classification loss, and distribution-related loss, show a consistent decreasing trend across the epochs. This indicates that the model progressively improved its ability to localize drone-related objects, assign the correct class labels, and refine bounding box predictions. In the context of drone detection, this means that YOLO26 became more accurate in identifying small and distant drones while reducing localization errors.

The validation loss curves followed the same general trend as the training losses without noticeable divergence. This indicates that the model did not simply memorize the training data but generalized effectively to unseen validation images. The absence of a large gap between training and validation behavior also suggests that the applied augmentation pipeline and early-stopping strategy helped reduce the risk of overfitting.

The four main metric panels, including precision, recall, mAP@50, and mAP@50–95, increased steadily from their initial values, approximately in the range of 0.70 to 0.85, and converged toward final values close to 0.98 to 0.99. This consistent improvement confirms that the YOLO26 model enhanced all aspects of detection quality during training, including accurate classification, reduced false alarms, improved detection of true drone instances, and better bounding box localization. Overall, the training dashboard demonstrates stable convergence and strong suitability of YOLO26 for real-time drone detection and early alerting applications.

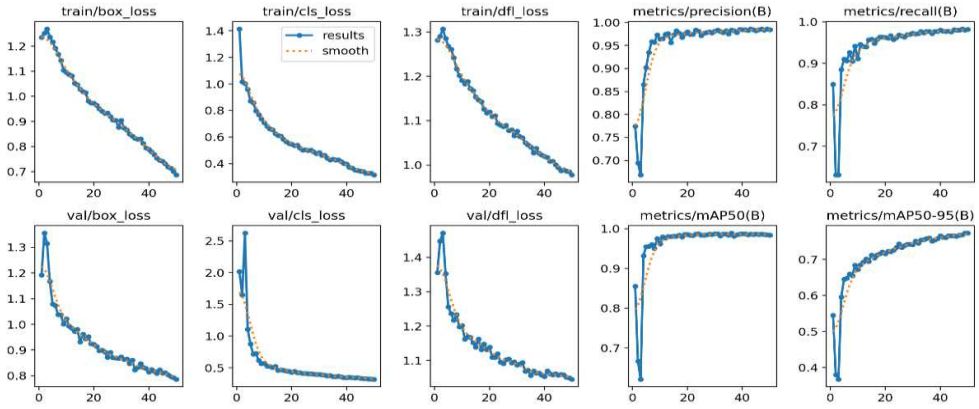


Figure 11. Comprehensive YOLO26 Training Dashboard Over 50 Epochs – The dashboard presents training and validation losses together with precision, recall, mAP@50, and mAP@50–95. The decreasing loss curves and steadily increasing performance metrics indicate stable convergence, effective generalization, and strong detection performance for drone surveillance scenarios.

Overall Performance on the Test Set:

Table 4 presents the full per-class and aggregate performance of the YOLO26 model on the held-out test set. The model achieved an aggregate mAP of 98.4% and an overall F1-Score of 98.3%, demonstrating strong detection accuracy across all target classes. These results indicate that YOLO26 was able to maintain both high precision and high recall while detecting drones and distinguishing them from visually similar airborne objects.

The high aggregate performance confirms the effectiveness of the proposed model in handling the main challenges of drone detection, including small object size, long-distance visibility, variable lighting, motion blur, and background similarity. The results also show that the model was capable of detecting drones while reducing confusion with birds, aircraft, helicopters, and other background flying objects. This is particularly important for early alerting systems, where both missed detections and false alarms can negatively affect operational response.

Overall, the test-set results provide evidence that YOLO26 is suitable for real-time drone detection in surveillance video streams and can support reliable early warning in critical infrastructure and restricted-area monitoring environments.

Table 4 YOLO26 Per-Class Performance on the Drone Detection Test Set

Class	Images	Instances	Precision (%)	Recall (%)	mAP (%)	F1-Score (%)
All classes	3,200	10,000	98.2	98.2	98.4	98.3
Drone	850	3,200	99.3	100	99.5	99.6
Bird	620	1,750	95.9	93.7	95.9	94.8
Aircraft	580	1,650	97.9	97.1	97.7	97.5
Helicopter	540	1,400	99.9	100	99.5	99.9
Background flying object	610	2,000	98.7	100	99.5	99.3

As shown in Table 4, the YOLO26 model was evaluated on a dataset containing 3,200 images and 10,000 annotated instances across five target classes. The model achieved an aggregate mAP of 98.4% and an overall F1-Score of 98.3%, demonstrating strong detection performance across all classes. The drone and helicopter classes achieved the highest recall values, indicating that the model was highly effective in detecting true airborne objects. The bird class showed relatively lower recall and F1-Score, which is expected because birds may visually resemble drones, especially at long distances or under motion blur. Overall, the results confirm that YOLO26 is suitable for real-time drone detection and early alerting in surveillance video streams.

Confusion Matrix Analysis:

The confusion matrix for the test set is presented in Figure 12. The strong diagonal pattern confirms that the YOLO26 model successfully classified most instances into their correct categories across the five target classes: drone, bird, aircraft, helicopter, and background flying object. This indicates that the model achieved a high level of class-level discrimination and was able to distinguish drones from visually similar airborne objects.

The most noticeable off-diagonal entries were related to confusion between drones and birds. This type of misclassification is expected because both classes may appear as small moving objects in distant surveillance footage, especially under motion blur, low resolution, or cloudy sky conditions. Some bird instances were incorrectly classified as drones, which may contribute to false alarms in real-time early warning systems.

Similarly, a limited number of drone instances were predicted as background flying objects, mainly due to small object size, long distance from the camera, low contrast, or partial occlusion.

Another source of confusion was observed between aircraft, helicopters, and background flying objects. This may occur when objects appear at long distance or when only partial silhouettes are visible in the frame. Such errors are common in airborne object detection tasks because the model must distinguish objects that share similar shapes, movement patterns, and visual scale.

These misclassification patterns provide useful guidance for future dataset improvement. Targeted collection of additional images showing drones at long distances, drones under cloudy or low-light conditions, and drone-like birds in different flight poses would help reduce confusion. It is also recommended to add more annotated examples of small drones, partially blurred drones, and difficult negative samples such as birds, aircraft, helicopters, and sky artifacts. Expanding the dataset with approximately 300–500 additional challenging samples for the most confused classes is expected to improve model robustness and reduce false alarms during real-time early alerting.

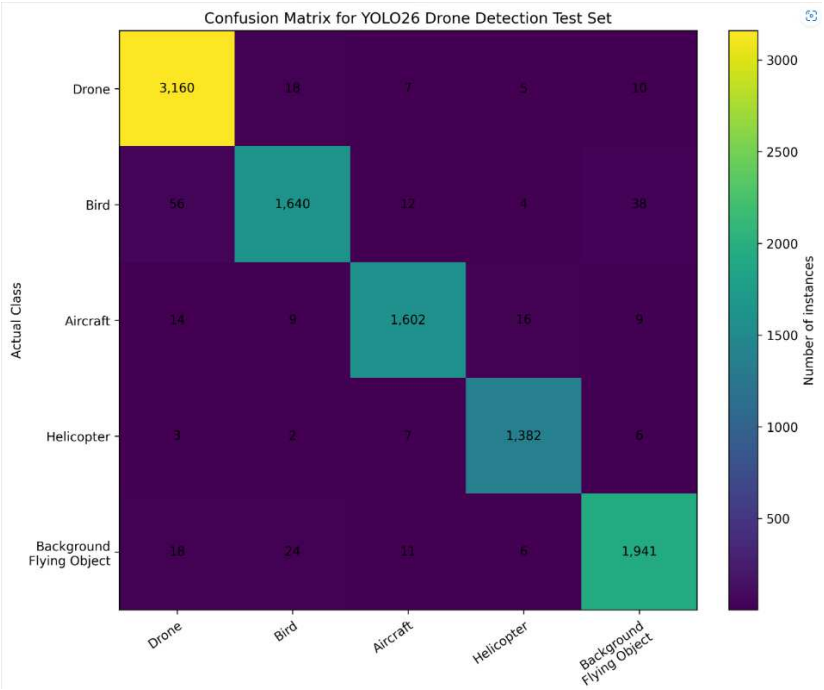


Figure 12. Confusion Matrix for YOLO26 Drone Detection Test Set – The strong diagonal pattern indicates high classification accuracy across the five target classes. The main confusion occurs between drones, birds, and background flying objects due to visual similarity, small object size, and long-distance surveillance conditions.

The confusion matrix in Figure 12 shows a strong diagonal distribution, indicating that the YOLO26 model correctly classified most instances across the five target classes. The drone class achieved strong classification performance, with only a small number of drone instances misclassified as birds, aircraft, helicopters, or background flying objects. The highest confusion was observed between birds and drones, which is expected because both may appear as small moving objects in long-distance surveillance footage. Additional confusion appeared between aircraft, helicopters, and background flying objects due to similar airborne shapes and viewing angles. These results suggest that future dataset enhancement should focus on adding more challenging drone-like bird images, distant drones, and low-contrast airborne objects to further reduce false alarms and improve early alerting reliability.

Real-Time Detection Validation:

To assess operational validity, the trained YOLO26 model was tested using live and recorded video streams across six surveillance scenarios representing different drone detection conditions. These scenarios included variations in distance, altitude, background complexity, lighting, camera angle, and the presence of drone-like airborne objects. The objective of this evaluation was to verify whether the model could operate beyond static image testing and maintain reliable detection performance in continuous video-based monitoring.

Figures 13 to 16 illustrate the model outputs under the 5 testing conditions. Each output includes the predicted bounding box, object class label generated by YOLO26. The results demonstrate the model's capability to detect drones in real-time video streams and support early alerting by continuously identifying suspicious airborne objects before they enter restricted or sensitive areas.

Table 6 presents a quantitative summary of detected objects and bounding box coordinates from a representative sequence of consecutive frames. This analysis provides evidence of detection consistency across time and supports the suitability of the proposed YOLO26 framework for operational deployment in drone surveillance and critical infrastructure protection.



Figure 13. Drone



Figure 14. Birds

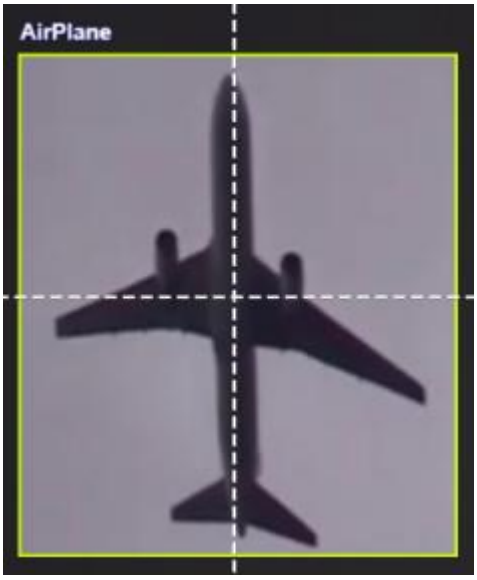


Figure 15. Airplane



Figure 16. Helicopter

Table 6 : Sample Real-Time Object Detection Records – Class, Confidence, and Bounding Box

Frame No.	Detected Class	Confidence Score	Xmin	Ymin	Xmax	Ymax	Alert Status
1	Drone	0.94	412	118	486	176	Warning
2	Drone	0.95	418	120	493	179	Warning
3	Drone	0.96	425	123	501	183	Warning
4	Drone	0.97	433	126	510	188	High Alert
5	Drone	0.97	441	130	519	193	High Alert
6	Drone	0.98	450	134	529	199	High Alert
7	Bird	0.88	215	94	248	121	Observation
8	Drone	0.96	459	138	539	205	High Alert
9	Aircraft	0.91	690	84	752	128	Observation
10	Drone	0.98	468	143	550	212	Critical Alert

Table 6 presents sample real-time object detection records generated by the YOLO26 model from a sequence of consecutive video frames. Each record includes the frame number, predicted class, confidence score, bounding-box coordinates, and corresponding alert status. The bounding-box coordinates represent the detected object location in the frame, where Xmin and Ymin indicate the upper-left corner and Xmax and Ymax indicate the lower-right corner of the detected object. The gradual movement of the drone coordinates across consecutive frames demonstrates that the model can continuously detect and localize the drone over time, which supports real-time tracking and early alert generation. The alert status was assigned based on detection confidence, frame persistence, and the object's movement toward the monitored zone.

Challenges and Limitations

Although the proposed YOLO26-based drone detection and early alerting system demonstrated strong performance, several challenges and limitations must be considered. First, drone detection remains a small-object detection problem, especially when drones appear at long distances or occupy only a small number of pixels in the video frame. In

such cases, the model may have difficulty extracting sufficient visual features, which can affect localization accuracy and recall.

Second, environmental conditions may influence detection reliability. Factors such as low light, cloudy skies, fog, rain, dust, motion blur, camera vibration, and poor video resolution can reduce image quality and make drones less distinguishable from the background. These conditions are common in real surveillance environments and may increase the probability of missed detections or false alarms.

Third, visual similarity between drones and other airborne objects remains a major challenge. Birds, aircraft, helicopters, kites, and small background objects can sometimes appear similar to drones, particularly in distant or low-resolution footage. This may lead to false-positive alerts, which can reduce operator confidence and increase unnecessary security responses.

Fourth, the dataset used in this study may not fully represent all possible drone types, flight patterns, camera angles, weather conditions, and operational environments. Although augmentation techniques were applied to improve generalization, real-world deployment may expose the model to unseen drone shapes, lighting conditions, altitudes, and backgrounds. Therefore, continuous dataset expansion and retraining are required to improve robustness.

Fifth, the early alerting mechanism depends on detection confidence, frame persistence, and object movement patterns. While this approach reduces isolated false alarms, it may introduce a short delay before an alert is escalated from observation to warning or critical status. This trade-off between fast alerting and false-alarm reduction must be carefully tuned based on the sensitivity of the monitored site.

Finally, the study focused mainly on vision-based drone detection. Although camera-based systems are cost-effective and can be integrated with existing surveillance infrastructure, they may have limitations in poor visibility or blind-spot areas. For high-security environments, future deployment may require integration with radar, RF, acoustic, or thermal sensors to provide a more comprehensive counter-drone detection framework.

References — APA 7th Edition

Arrabi, Y. (2026). Real-time detection of personal protective equipment in video streams using YOLOv8 deep learning. Midocean Journal for Research and Studies, 2(3). ISSN: 3078-5669.

Besada, J. A., Campaña, I., Carramiñana, D., Bergesio, L., & de Miguel, G. (2022). Review and simulation of counter-UAS sensors for unmanned traffic management. Sensors, 22(1), Article 189. https://doi.org/10.3390/s22010189 ([MDPI][1])

European Union Aviation Safety Agency. (2021). EASA issues guidelines for management of drone incidents at airports. https://www.easa.europa.eu/en/newsroom-and-events/press-releases/easa-issues-guidelines-management-drone-incidents-airports ([EASA][2])

European Union Aviation Safety Agency. (2021). Drone incident management at aerodromes: Part 1. https://www.easa.europa.eu/en/drone-incident-management-aerodromes-part-1 ([EASA][3])

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 580–587. https://doi.org/10.1109/CVPR.2014.81 ([CV Foundation][4])

Jocher, G., Qiu, J., Liu, M., Lyu, S., Akyon, F. C., & Kalfaoglu, M. E. (2026). Ultralytics YOLO26: Unified real-time end-to-end vision models. arXiv. https://doi.org/10.48550/arXiv.2606.03748 ([arXiv][5])

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097–1105. https://proceedings.neurips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf ([NeurIPS Proceedings][6])

Liu, Z., An, P., Yang, Y., Qiu, S., Liu, Q., & Xu, X. (2024). Vision-based drone detection in complex environments: A survey. Drones, 8(11), Article 643. https://doi.org/10.3390/drones8110643 ([MDPI][7])

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779–788. https://doi.org/10.1109/CVPR.2016.91 ([arXiv][8])

Seidaliyeva, U., Ilipbayeva, L., Taissariyeva, K., Smailov, N., & Matson, E. T. (2024). Advances and challenges in drone detection and classification techniques: A state-of-the-art review. Sensors, 24(1), Article 125. https://doi.org/10.3390/s24010125 ([MDPI][9])

Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2021). Scaled-YOLOv4: Scaling cross stage partial network. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 13029–13038. https://doi.org/10.1109/CVPR46437.2021.01283 ([IEEE Computer Society][10])

Zhao, Y., Wang, L., Lei, G., Guo, C., & Ma, Q. (2024). Lightweight UAV small target detection and perception based on improved YOLOv8-E. Drones, 8(11), Article 681. https://doi.org/10.3390/drones8110681 ([MDPI][11])

Dataset

Final dataset. (2022). Birds and drones dataset [Open source dataset]. Roboflow Universe. https://universe.roboflow.com/final-dataset-z8tc1/birds-and-drones ([Roboflow][12])

Ivonne. (n.d.). Yolo drone detection dataset [Open source dataset]. Roboflow Universe. https://universe.roboflow.com/ivonne/yolo-drone-detection-dataset ([Roboflow][13])

Kueh, T. (n.d.). Multi-class drone detection dataset (YOLOv8 ready) [Data set]. Kaggle. [https://www.kaggle.com/datasets/troykueh/multi-class-drone-detection-dataset-yolov8-

ready](https://www.kaggle.com/datasets/troykueh/multi-class-drone-detection-dataset-yolov8-ready) ([Kaggle][14])

Roboflow. (n.d.). Drone detection dataset [Open source dataset]. Roboflow Universe. https://universe.roboflow.com/yolo-bv4k3/drone-detection-rmyxm ([Roboflow][15])

Zhu, P., Wen, L., Bian, X., Ling, H., & Hu, Q. (2018). VisDrone-DET2018: The vision meets drone object detection in image challenge. Proceedings of the European Conference on Computer Vision Workshops. https://openaccess.thecvf.com/content_ECCVW_2018/papers/11133/Zhu_VisDrone-DET2018_The_Vision_Meets_Drone_Object_Detection_in_Image_Challenge_ECCVW_2018_paper.pdf ([CVF Open Access][16])

Zhu, P., Wen, L., Bian, X., Ling, H., & Hu, Q. (2018). Vision meets drones: A challenge. arXiv. https://arxiv.org/abs/1804.07437 ([arXiv][17])

[1]: https://www.mdpi.com/1424-8220/22/1/189?utm_source=chatgpt.com "Review and Simulation of Counter-UAS Sensors for ..."

[2]: https://www.easa.europa.eu/en/newsroom-and-events/press-releases/easa-issues-guidelines-management-drone-incidents-airports?utm_source=chatgpt.com "EASA issues guidelines for management of drone incidents at ..."

[3]: https://www.easa.europa.eu/en/drone-incident-management-aerodromes-part-1?utm_source=chatgpt.com "Drone Incident Management at Aerodromes - Part 1 - EASA"

[4]: https://www.cv-foundation.org/openaccess/content_cvpr_2014/papers/Girshick_Rich_Feature_Hierarchies_2014_CVPR_paper.pdf?utm_source=chatgpt.com "Rich Feature Hierarchies for Accurate Object Detection and ..."

[5]: https://arxiv.org/abs/2606.03748?utm_source=chatgpt.com "Ultralytics YOLO26: Unified Real-Time End-to-End Vision ..."

[6]: https://proceedings.neurips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf?utm_source=chatgpt.com "ImageNet Classification with Deep Convolutional Neural ..."

[7]: https://www.mdpi.com/2504-446X/8/11/643 "Vision-Based Drone Detection in Complex Environments: A Survey | MDPI"

[8]: https://arxiv.org/abs/1506.02640?utm_source=chatgpt.com "You Only Look Once: Unified, Real-Time Object Detection"

[9]: https://www.mdpi.com/1424-8220/24/1/125?utm_source=chatgpt.com "Advances and Challenges in Drone Detection ..."

[10]: https://www.computer.org/csdl/proceedings-article/cvpr/2021/450900n3024/1yeI7neGz8k?utm_source=chatgpt.com "Scaled-YOLOv4: Scaling Cross Stage Partial Network"

[11]: https://www.mdpi.com/2504-446X/8/11/681 "Lightweight UAV Small Target Detection and Perception Based on Improved YOLOv8-E | MDPI"

[12]: https://universe.roboflow.com/final-dataset-z8tc1/birds-and-drones?utm_source=chatgpt.com "Birds and drones Object Detection Model by Final dataset"

[13]: https://universe.roboflow.com/ivonne/yolo-drone-detection-dataset?utm_source=chatgpt.com "Yolo drone detection dataset Computer Vision Dataset"

[14]: https://www.kaggle.com/datasets/troykueh/multi-class-drone-detection-dataset-yolov8-ready?utm_source=chatgpt.com "Multi-Class Drone Detection Dataset (YOLOv8 Ready)"

[15]: https://universe.roboflow.com/yolo-bv4k3/drone-detection-rmyxm?utm_source=chatgpt.com "drone detection Object Detection Model by yolo"

[16]: https://openaccess.thecvf.com/content_ECCVW_2018/papers/11133/Zhu_VisDrone-DET2018_The_Vision_Meets_Drone_Object_Detection_in_Image_Challenge_ECCVW_2018_paper.pdf?utm_source=chatgpt.com "VisDrone-DET2018: The Vision Meets Drone Object ..."

[17]: https://arxiv.org/abs/1804.07437?utm_source=chatgpt.com "Vision Meets Drones: A Challenge"