BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

AI-POWERED OBJECT AND THREAT DETECTION APPROACHES IN DRONE SURVEILLANCE SYSTEMS FOR DEFENCE USE

Ware Prajwal Anil

Student
Department of Artificial Intelligence and Data Science Engineering
Ajeenkya DY Patil School of Engineering, Pune, India

Prof. Sandeep Jayswal

Professor
Department of Artificial Intelligence and Data Science Engineering
Ajeenkya DY Patil School of Engineering, Pune, India

Abstract

Nowadays, unmanned aerial vehicles (UAVs) play a vital role in surveillance activities of a country's defense. They allow, at almost no risk for the operators, to monitor the target area for long period of time. Unfortunately manual analysis of UAV footages is not suitable for border or combat operations at all. It is extremely slow, prone to a very high level of operator fatigue and also very inconsistent. Recently, deep learning-based computer vision has brought to the table detection and tracking of various types of threats in real-life drone footage. In this paper, we review the work on AI-based object and threat detection in drone surveillance and we pay particular attention to YOLO-variant and feature-fusion networks. From the ten representative papers on detection methods, we draw a conclusion of their results, highlight the limitations set by drone cameras, such as very small object size, motion blur, very changing altitude, lighting changes, and touch upon threat classification and deployment at the system edge. After that, we describe a prototype of a five-stage pipeline which brings together detection, classification, tracking, alerting and human confirmation. Based on our analysis, a majority of the positive results come from civilian benchmarks like VisDrone, the multi-sensor fusion has hardly any studies to its name, and the testing of robustness and reliability under a battlefield or border setting is still an open problem. Our paper being purely a literary review, makes no claims about the performances of the experiments; it concludes with outlining a plan for conducting the tests in the real implementation scenario.

Keywords: Deep Learning, Drone Surveillance, Edge AI, Object Detection, Threat Classification, Unmanned Aerial Vehicles (UAV), YOLO

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Introduction

Throughout the last ten years unmanned aerial vehicles (UAV/Drones) evolved from merely reconnaissance devices to be the mainstay of a modern defense's surveillance. They ensure continuous extensive area monitoring on the borders, camps, and battlefields without putting the soldiers at a direct risk. At one point, human operators were the major source in providing the surveillance through a great length of drone videos. Such surveillance is not efficient and is a time-consuming, tire-inducing, and inconsistent way of operation.

As the result of introducing deep learning, the landscape has changed. Nowadays drones can discover classify and track the people, vehicles, weapons, or any personnel looking suspicious while the video is being recorded at that moment, thereby transforming surveillance into a kind of intelligent, reactive defense tool, rather than just a passive logging task of recording (Lu et al., 2025; Zhu et al., 2023). The YOLO type of detectors were primarily responsible for the breakthrough, they have been developed further by incorporating feature- pyramids, multi-scale fusion mechanisms, and attention modules so as to cope with detecting targets on aerial images which are very small, moving at speed, and being a part of a clutter background, (Liu et al., 2025; Lu et al., 2025; Tian et al., 2025; Wang & Yu, 2025a; Wei & Wang, 2025):

This paper focuses on investigating that area. It explores how current AI & ML frameworks manage drone imagery object detection, threats identification post-detection, what issues drone-based detection poses, and at which points the existing research is not meeting what a defense-secure system would require (Kaur et al., 2026). One of the main objectives is a separation of those detection methods that are nearing practical use level from those which are still at an experimental stage because most drone detectors have so far been tested only on datasets of civilian origin and their performance under the conditions of a defensive system is not known.

Survey Literature section will show a survey over previous related studies, and it will also highlight missing aspects in this area of study. Then, the paper defines the design of our method and problem description. After that, conceptual model followed with methodology. Design features are demonstrated through the means of a real border-monitoring use case, and also, the authors discuss the major concerns, difficulties which design will be faced with and a feature-wise comparison between existing techniques. Ethical, societal and environment aspect of the study in the section, and finally, the conclusion of the whole work. All of these are mentioned at the end of the paper.

Literature Survey

YOLO Extensions for Small-Object Detection

Many recent works of drone-based detection have been focused on adapting YOLO family to target small, dense and low-contrast images. AMFE-YOLO (Wang & Yu, 2025a) develops a novel adaptive multi-path feature extraction (AMFE) module to focus on occluded targets based on YOLOv8s, and also it is a reconfiguration that reduces false positives on small-objects detection head. The authors claim a 5.5% average precision increase and reduction parameter count by about 0.7 M. On the other hand MASF-YOLO (Lu et al., 2025) uses a multi-scale feature aggregation module (MFAM), an improved efficient multi-scale attention module (IEMA), and a dimension-aware selective integration module (DASI) to upgrade YOLOv11. The results show that mAP@0.5 improvement for VisDrone2019

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

validation set amounts to 4.6% and mAP@0.5:0.95 by a factor of 3.5%, compared to YOLOv11-s. Also, the paper mentions the fact that the small-scale variant would use 60% and 65% of the parameters/computation of YOLOv11-m respectively. Liu et al. (2025) introduce a YOLO-based detector that incorporates three innovations: detail-semantic adaptive fusion (DSAF), multi-scale residual spatial attention (MSRSA), and feature pyramid reuse and fusion network. The network adds a unique 160×160 detection head specifically for the tiniest targets. A YOLOv8-based model, MFDA-YOLO (Tian et al., 2025) presents with the dynamic alignment detection head (DADH) which is meant to handle the weakness of the localization classification coordination. The model is said to outperform YOLOv11 and YOLOv13 in VisDrone2019, HIT-UAV, and NWPU VHR-10 datasets by the authors. Another paper, Lai et al. (2025), put forward SRD-YOLOv5, a highly efficient YOLOv5n with scale sequence feature fusion module, a multi-scale feature extraction module, a super tiny target detection layer, a decoupled head, etc., and they claimed that it was the top contender on drone remote-sensing images. Wang et al. (2025b) come up with the idea of a feature fusion framework that uses a combination of downsampling and upsampling to resolve the imbalance problem with the two modules on small targets of UAV images. RFAG-YOLO (Wei & Wang, 2025) is based on a modification of YOLOv8 and it adds before detection head a receptive field network (RFN) block and a scale-aware feature amalgamation The works which are summarized here deal with the evaluation of either civilian or remote-sensing drones. It can be observed that their summaries are not indicating that the models tested have been tested under the defense-specific scenarios.

Real-Time and Lightweight Designs

Deploying on drones is dependent on computation that would fit inside battery-powered hardware. The study of YOLO-Drone published by Zhu et al. (2023) is built upon a Darknet59 backbone, it is a multi-scale spatial pyramid and feature pyramid module (MSPP-FPN), and GIoU loss. As a result, the researchers state that the proposed detector yields mAP improvements of as high as 10.13% on UAVDT and as high as 8.59% on VisDrone over other top methods. They also note that it has a real-time rate of 53 FPS in their test setup and an mAP of 87.71% when they use a special LED lighting for the night operation. Another paper, written by Zhang (2023), presents Drone-YOLO, a model in which the usual downsampling convolutions are replaced with RepVGG modules within a YOLOv8 network. It is the tiny version with 5.25 million parameters which reportedly is competitive to or even better than the baseline of 9.66 millions parameters on VisDrone2019. These speed figures from the authors of the respective models are relative to their own test conditions; and the results presented in summary do not prove the model can run effectively on a battery-powered drone.

Threat-Specific Classification

No aerial detection necessarily means a drone-related incident. Detection of birds and drones as well as birds vs. drone classification is the main concern of Kaur et al. (2026) work, which resulted in YOLOBirDrone, a system built around YOLO for identifying and classifying birds and drones, together with a dataset called BirDrone. The project stems from recognizing small birds as hard as drones, which could lead to false alarms for a drone detection system. But, the approach is rather limited, as it focuses primarily on birds-drones identification, without covering humans, vehicles, weapons and other anomalies that are part of surveillance objectives.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Benchmark Datasets

The VisDrone dataset and challenge (Du et al., 2019) have been credited with creating the reference point or benchmark in object detection from drones and to date remains VisDrone as the standard against which the majority of later UAV detectors are compared. Such standardization has made it possible to assess fairly the merits but at the same time the VisDrone dataset has been focused mainly on civilian scenes like, say, pedestrians and traffic rather than scenarios of fighting, defending the border or security threat, so their outcome may not really reflect the capability in defending a place.

Summary of Research Gaps

Table 1 shows the methods, obtained results and the major gaps left by the studies reviewed. Several gaps can be identified in the research: For instance, almost all the models have been trained and tested using civilian-related datasets, with no defense-standard benchmark adopted. Then, the use of multi sensor fusion incorporating thermal, audio, or multispectral inputs are mostly unexplored. Finally, robustness against occlusion, motion blur, and variable lighting is still the open issue, which implies that the gap between very good results on the benchmark and the requirements for reliability in the field has not yet been resolved.

Table 1: Summary of Literature Survey and Research Gaps

Ref.	Method	Reported Result	Residual Gap
Wang & Yu, 2025a	AMFE-YOLO: YOLOv8s with the addition of the AMFE Module, along with the modification to the head	+5.5% AP compared to YOLOv8s; -0.7 million parameters	Only civilian datasets; focused on detecting occluded objects
Lu et al., 2025	MASF-YOLO: YOLOv11 + MFAM, IEMA, DASI	+4.6% mAP@0.5 over YOLOv11-s	Based on VisDrone data only; no images of battlefield situations
Liu et al., 2025	Liu et al.: DSAF, MSRSA, 160×160 head	Improved tiny-object detection	UAV benchmarking; no combat scene data
Tian et al., 2025	MFDA-YOLO: YOLOv8 + DADH	Beats YOLOv11 and YOLOv13 on three UAV sets	Civilian and remote-sensing data
Kaur et al., 2026	YOLO BIR DRONE: Bird vs. Drone	Improved bird–drone discrimination	Scope: limited; other threat categories excluded;
Wei & Wang, 2025	RFAG-YOLO: YOLOv8 + RFN block, SAF	Beats YOLOv7/v8/v10/v11 on VisDrone	No standard defense dataset
Zhu et al., 2023	Zhu et al.: Darknet59, MSPP-FPN, GIoU	+10.13% UAVDT, +8.59% VisDrone mAP; 53 FPS	Speed based on the authors' configuration; no deployment test on embedded systems

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Ref.	Method	Reported Result	Residual Gap
Zhang, 2023	Drone-YOLO: YOLOv8 + RepVGG downsampling	The tiny architecture variant (5.25 million parameters) can rival a 9.66 million parameter architecture in performance.	Civilian VisDrone only
Du et al., 2019	VisDrone dataset and challenge	Standard UAV benchmark	Civilian scenes; threat categories omitted
Lai et al., 2025	SRD-YOLOv5: YOLOv5n + SSFF, MSFE, ESTDL	Outperforms other detection methods for aerial images;	Benchmarks based on remote sensing but not battlefield data;
Wang et al., 2025b	Joint feature upsampling with downsampling fusion	More accurate determination of location and identification of small objects	Three residual gaps are recurring over the papers.

First of all, it is true for the majority of models that they have not only been trained and tested on the civilians datasets but also have not used a defense-grade benchmark standardized at the industry level. The second and third points are related to multi-sensor fusion and robustness in adverse situations of occlusion, motion blur, or low lighting. In both cases, these aspects are hardly researched. In other words, it is possible to say that despite the top rankings of the benchmarks, the level of reliability at the defense grade hasn't yet been reached which points to the fact that the gap hasn't been closed.

Problem Statement and Design Objectives

Drone detection is still, in large part, relying on manpower to analyze the videos. The automated detection technologies have, so far, been created and tested using civilian datasets. There is, therefore, a demand for a machine-learning system that can simultaneously detect and classify the threats and then provide the result within the energy and time constraints of drone hardware and maintain the authority of a human operator over the decision process. Below is the list of goals that we followed through this research:

(1) To study currently available AI and ML architectures for object detection in aerial/drone imagery, focusing primarily on YOLO variants and feature-fusion networks. (2) To evaluate methods for threat classification of humans, vehicles, weapons and anomalous behavioral acts. (3) To assess problems inherent to detection with drones such as small object size, motion blur, variable altitude and light conditions. (4) To see how possible a solution in real-time and low-latency is defense-wise. (5) Lastly, to pinpoint the areas of the investigation which, upon further exploration, may give rise to innovative and valuable research.

Proposed Conceptual Architecture

A five-stage collaborative system constitutes the framework. Aerial data acquisition, a stage, records videos and drone's thermal images. Other sources can provide images or data if the drone lacks them. Frame pre-processing and detection at the onboard side normalize each video frame, improve low-light areas, and deploy a YOLO detector for locating objects with bounding boxes.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

The stages are arranged as a linear sequence of pipeline operations; detection, classification, tracking, and raising alerts are carried out by each new frame while the results are relayed to the operator of the control room. The detector could be embedded on the aircraft or placed close to the aircraft hence, the main limitations are latency and onboard compute power. The main sequence of the pipeline can be expressed as Detect, Classify, Track, and Verify. This conceptual design serves for analysis without describing a system that has been realized or assessed.

Methodology

Detection and Classification

The image is first resized and normalized before the low-light enhancement is applied, depending on the level of illumination. A YOLO variant-based detector generates bounding box candidates accompanied by confidence scores. Each of these detections is then classified under one of five labels: people, vehicles, weapons, birds, or 'unknown'. We decided on including a bird label due to the fact that birds-drones confusion is a well-documented source of false alarms (Kaur et al., 2026), whereas we included the unknown label to refrain from being forced into threat classification of low-confidence items.

Tracking and Motion Analysis

By matching detections in successive frames we form tracks and study the motion of each track. Tracking is an aid to two major requirements. In the first place it retains the identity of an object through time so as detection of the same person over and again will not be seen as separate events, and, secondly, it furnishes a source of movement characteristics such as direction and speed that help to estimate threat level ranking.

Alert Generation

Ranking is just a part of decision-making, not a tool to substitute for the operator's discretion, only after a sufficient number of errors have been found in real-world testing, one can set the calibration parameters.

Procedure

Algorithm 1 is a description of the steps in the system. From the input, which is a live video feed from a drone, it gives the output as a list of threat alerts and corresponding video frames that can be reviewed by a human expert.

Algorithm 1: Threat Detection Pipeline

Input: UAV video stream F = {f1, f2, ..., fT}

Final Output: threat alerts A supported by a sequence of frames

1: Frame f_t gets acquired from a drone camera.

2: Frame processing includes resizing, normalization, low-light enhance

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

3: A detection using YOLO-variant detector -> resulting in boxes and confidences

4: Assign each detection to a class:

men / women | cars / motorcycles | guns / swords | ducks / hawks | other

5: Match detections over frames -> tracks; Motion analyzed

6: If a track looks like a threat, then an alert is produced

with its relevant frames

7: Operator sees alerts and decides on the confirmation

Conceptually: A = f(D, C, T, V)

human | vehicle | weapon | bird | unknown

Figure 1 presents the same procedure as a flowchart.

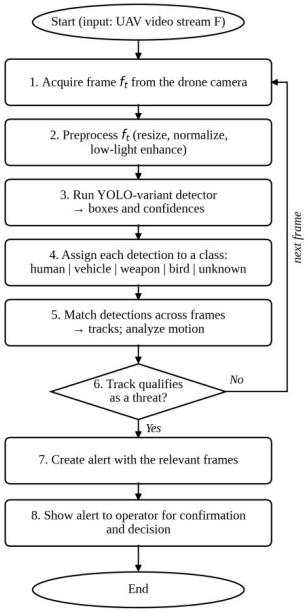


Figure 1: Flowchart of the Threat Detection Pipeline (Algorithm 1)

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

This equation: A = f(D, C, T, V) describes the phases of the entire system in simple words: the set of alerts A is made up from detections D (i.e., the objects found), their categorization C (i.e., what type of object they are), the set of tracks T maintained from frame to frame over time, and, finally, human verification step V. The function is descriptive rather than something that has been trained and validated with actual experiments.

Representative Border-Surveillance Scenario

To get the feel for how the system works, take an airborne unit surveillance example. Imagine the airborne unit surveillance of one side of border at night by varying height drone. Initially, an enhancement operation increases the quality of a weak area next to a fence. A detected entity will be indicated by the system, and the object will be characterized as a human by the classifier module. The next few frames will see the object's identity still being maintained by the tracker, even though the image is blurry due to winds, and the movement study suggests that the object is coming closer to the wire fence. Some bird-related detections will show up if we consider another scene with a flock flying near by but the birds label in this case will prevent them coming up as threats. A single warning with a few best frames is sent out by the system and the staff then look over them.

This example aims at demonstrating how the system will work and not reporting any result. It reveals the nature of a collaborative pipeline and the position the human operator plays in such a system. Real-world reliability of such a behaviour, obviously, is conditional upon the factors affecting performance that are explained in the next sections.

Expected Challenges and Discussion

Various studies show that each technique individually addresses certain aspects of the problem. Multi-scale and attention modules lead to a better discovery of small objects (Lai et al., 2025; Liu et al., 2025; Lu et al., 2025; Tian et al., 2025; Wang & Yu, 2025a; Wang et al., 2025b; Wei & Wang, 2025). On the other hand, lightweight networks show great speeds in drone benchmark tests (Lai et al., 2025; Zhang, 2023; Zhu et al., 2023). Specialized bird classifiers greatly reduce the number of bird-related false alarms (Kaur et al., 2026). Nevertheless, all of these studies have been done in the context of the civilian datasets. Getting a gain on the VisDrone dataset is not the same as achieving results on battlefield or border territory, since performance under the latter conditions is quite different from that on the former dataset. In addition, the studies of the performance of VisDrone, mentioned above, have focused on civilian rather than defense conditions. Therefore, the research literature indicates that each of the stages considered is feasible but the paper does not support the reliability of the whole chain.

Three types of difficulties arise due to these constraints. Very small objects at a long range correspond to very few pixels. Hence the reliability of detecting them goes down. Blurred moving objects, darkness, rain/wind can lead to deterioration of both detection and tracking. In addition, neural networks that have been trained on civilian scenes may make incorrect decisions in the areas of combat and border scenes which may be quite different from civilian ones in both the type of objects and the kind of background (e.g. snow/rock/dry terrain). Every of the three problems outlined is addressed by the respective works but none of them has been successfully handled in a defense setup.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

Comparison with Existing Approaches

Table 2 shows how our proposed analysis relates with four different kinds found in the literature survey on a comparative basis and highlights their objectives, pros, and cons.

Table 2: Comparison of Approach Categories

Approach	Main Purpose	Advantages	Limitations
Accuracy-oriented YOLO extensions (Lu et al., 2025; Wang & Yu, 2025a; Wei & Wang, 2025)	Detect small targets	Outstanding improvements on UAV benchmarks	Civilian data only
Feature alignment and fusion (Liu et al., 2025; Tian et al., 2025; Wang et al., 2025b)	A much better ability of detecting the very small objects.	Enhanced performance with various UAV datasets	No defense-specific evaluation
Real-time and lightweight designs (Lai et al., 2025; Zhang, 2023; Zhu et al., 2023)	A solution with a good balance of low hardware requirements and high speed for UAV platforms	High frame rate reported; small model (fewer parameters)	This is only an approximation as the actual speed varies with conditions.
Specialized classifiers (Kaur et al., 2026)	Reduce bird false alarms	This article also gives an overview of another major source of false-positive detections that could be solved.	Narrow; no defense classes
Benchmark datasets (Du et al., 2019)	Standardized evaluation	Enables fair comparison	Only civilian scenes were included with no defense classes
Proposed review	Link detection, classification, edge deployment, and gaps	Identifies defense-relevant gaps	Literature-based; no experiments

Ethical, Societal and Sustainability Considerations

When using surveillance videos that include people not involved in the surveillance activity, privacy concerns need to be addressed. The personal data of these people should only be stored and available under very tight controls. Mistaking someone is just as serious: mislabelling an innocent civilian as a potential threat could have serious consequences so the error rate should be understood well enough before using the system. An imbalanced training dataset might be a reason for bias where environments, clothing styles, or populations are under-represented. People who operate the machines and organizations who install them have a share in the responsibility. ewline In addition to this, another possible misuse is to use surveillance methods solely for defense and security without taking measures for other purposes which is not in keeping with human rights nor with the law in which case it will be necessary to define policies to prevent misuse.

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

AI-enhanced drone security has an enormous potential as it may reduce danger exposure while maintaining a high degree of safety. However, such wide scale use of drone technology could lead to a serious violation of privacy and individual liberty, and relying purely on automated systems could result in a reduction of people's ability to make decisions, and therefore there is an increasing need for clear guidelines on how surveillance videos should be handled so that the public can remain on board with the changes.

Besides economic and technical aspects, sustainability was highlighted as an important factor in drone development that was addressed in the work. Training of deep learning models takes a lot of time and electricity, which is particularly a concern for small-scale drones running on batteries. Smaller and more compact models with onboard computation would significantly cut down power usage and at the same time eliminate sending massive videos. Since there are several ways drones and people can collaborate as well as threats and situations change, continuous upgrades as well as maintenance of these systems must therefore be planned for. It is aligned with the objectives of the SDG 9, Industry, Innovation and Infrastructure, as the proposal is based on the use of AI to drive development in drone technology, and with SDG 16, Peace, Justice and Strong Institutions, as long as the application of the system remains to be open, accountable, and the rights and freedoms of human beings are guaranteed.

Evaluation Plan and Future Work

We recommend that in a real world scenario the detection precision for small objects, the frame rate of edge hardware, and the false-alarm rate specially for birds should be investigated. Performance test scenarios involving occlusion, lowlight, and motion blur should also be considered. The above are the research directions recommended but it must be made clear that the present paper does not make a claim for performance. Different models must be tested on the same protocol so that they can be reliably compared.

From a practical viewpoint, the review points out several issues that still need attention. These include, for one thing, the need for the provision of comprehensive defense-oriented datasets with defense-relevant classes, situations, and environments. In the second place, one may combine visual light with thermal and audio as well as multispectral for detecting targets in all weather conditions. In the third place, it is the capability of a system to resist being tricked by occlusion, motion blur, low light, or cluttered backgrounds that needs more investigation. Fourth, we must train models to work on Jetson-class hardware in near-real time. For the fifth point, threat detection is not restricted to bird-drone differentiation but should also encompass identification of people, vehicles, weapons, and other threats. Last point, it is vital to have defined the requirements for criteria, human-verification processes, and governance of AI-assisted surveillance before implementing these technologies.

Conclusion

We conducted a study reviewing techniques related to AI-based object and threat detection in drone surveillance, as part of the defense applications. YOLO-variant and feature-fusion models allow the detection of small airborne objects at very high speed, and the existence of edge hardware makes it possible to process onboard. However, most publications in the field base themselves primarily on civilian datasets, only a very few are discussing multisensor fusion, and in general, they do not give adequate information about robustness to common adverse factors like occlusion, low light, and motion blur. Our paper also suggests a five-step process pipeline linking the stages of detection, recognition,

BODHIVRUKSHA JOURNAL OF DIVERSE DISCIPLINE
E-ISSN : 3139-1486
An Open Access, Peer-Reviewed, Multidisciplinary Scholarly Journal
Bi-monthly | Multilingual (English, Hindi, Marathi) | Academic Research
@ Volume 2, Issue 5 (September-October 2026) (Special Issue 1) ISSN: 3139-1486 (Online)

tracking, warning, and human confirmation plus the proposal for a research plan that can be used once these stages have been developed. Since the study is based on available literature, one should not expect much beyond what the papers under discussion say. Defense datasets, multi-sensor approaches, and evaluations in realistic operating conditions are required to ensure practical reliability of such systems.

Acknowledgment

Author would like to convey sincere thanks to Prof. Sandeep Jayswal for providing guidance and sharing thoughts during the seminar. Also, Author is grateful to the Technical Seminar Coordinator, faculty members, Head of the Department, and Principal of Department of Artificial Intelligence and Data Science Engineering, Ajeenkya DY Patil School of Engineering, Pune, for the help and support from their side.

References

1. Du, D., Zhu, P., Wen, L., Bian, X., Ling, H., Hu, Q., & Zhu, R. (2019). VisDrone-DET2019: The vision meets drone object detection in image challenge results. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (pp. 213–226). https://doi.org/10.1109/ICCVW.2019.00030
2. Kaur, D., Battish, N., Bhavsar, A., & Poddar, S. (2026). YOLOBirDrone: Dataset for bird vs drone detection and classification and a YOLO based enhanced learning architecture. arXiv. https://doi.org/10.48550/arXiv.2601.08319
3. Lai, D., Kang, K., Xu, K., Ma, X., Zhang, Y., Huang, F., & Chen, J. (2025). Enhancing UAV object detection with an efficient multi-scale feature fusion framework. PLOS ONE, 20(10), Article e0332408. https://doi.org/10.1371/journal.pone.0332408
4. Liu, C., Wang, Y., Cao, Q., Zhang, C., & Cheng, A. (2025). Rethinking adaptive contextual information and multi-scale feature fusion for small-object detection in UAV imagery. Sensors, 25(23), 7312. https://doi.org/10.3390/s25237312
5. Lu, L., He, D., Liu, C., & Deng, Z. (2025). MASF-YOLO: An improved YOLOv11 network for small object detection on drone view. In Proceedings of the 2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC). https://doi.org/10.1109/SMC58881.2025.11342487
6. Tian, D., Wang, X., Liu, D., & Hao, Y. (2025). MFDA-YOLO: A multiscale feature fusion and dynamic alignment network for UAV small objects detection. PLOS ONE. https://doi.org/10.1371/journal.pone.0337810
7. Wang, Q., & Yu, C. (2025a). AMFE-YOLO: A small object detection model for drone images. IET Image Processing, 19(1), e70110. https://doi.org/10.1049/ipr2.70110
8. Wang, X., Peng, Y., & Shen, C. (2025b). Efficient feature fusion for UAV object detection. arXiv. https://doi.org/10.48550/arXiv.2501.17983
9. Wei, C., & Wang, W. (2025). RFAG-YOLO: A receptive field attention-guided YOLO network for small-object detection in UAV images. Sensors, 25(7), 2193. https://doi.org/10.3390/s25072193
10. Zhang, Z. (2023). Drone-YOLO: An efficient neural network method for target detection in drone images. Drones, 7(8), Article 526. https://doi.org/10.3390/drones7080526
11. Zhu, L., Xiong, J., Xiong, F., Hu, H., & Jiang, Z. (2023). YOLO-Drone: Airborne real-time detection of dense small objects from high-altitude perspective. arXiv. https://doi.org/10.48550/arXiv.2304.06925