International Journal of
Science, Architecture, Technology and Environment

NIGHTWING: Real-Time Infrared Detection Framework for Nocturnal Migratory Birds Using Spatiotemporal Anomaly Modeling to Enhance Aviation Safety

Amber Sophia A. Perez[1], Philip Andrew C. Malazarte[1*], Ace Jortherson L. Tejano[1], Aniceto B. Ipanag, Jr. [2], Myrill Guia A. Bautista[1], Henry F. Haranay[1]

[1]*Science, Technology, and Engineering Program, Tagum City National High School, Philippines*

[2]*Local Government Unit (LGU) of Tagum City, Philippines.*

[*]*Corresponding author*

DOI: https://doi.org/10.63680/ijsate0326027.046

Abstract

Bird strikes constitute a significant aviation safety challenge, particularly during nocturnal operations where conventional detection systems are often ineffective. Current thermal imaging technology, which offers the best modality for night-time sensing, lacks the computational efficiency and speed required for real-time deployment on resource-constrained hardware. To address this critical gap, we developed NIGHTWING, a lightweight, dual-stage framework that utilizes infrared (LWIR) signatures for robust, real-time hazard detection. The framework's core detector, YOLOv8n, achieved 99.5% mAP@0.5 on the primary validation set; however, performance dropped to 51.05% mAP on an independent challenging subset, indicating a significant generalization gap. The system demonstrated high computational efficiency, achieving 90.9 FPS, which exceeds the 5 FPS operational threshold. The spatiotemporal anomaly system generated critical alerts within 150 milliseconds. While YOLOv8n outperformed alternative architectures in our evaluation, baseline models (EfficientDet-D0 and MobileNetV2-SSD) failed to converge properly, rendering comparative accuracy claims invalid. These results highlight both the promise and limitations of lightweight thermal detection systems for nocturnal wildlife monitoring.

Keywords: avian detection; thermal imaging; YOLOv8; spatiotemporal anomaly; aviation safety; nocturnal migration; edge AI

1. Introduction

Bird strikes remain a significant concern in aviation safety management, particularly inflight operations during takeoffs, landings, and low-altitude flights. Although bird mitigation systems have been developed to reduce these incidents, they prove to be ineffective against nocturnal birds as such systems are typically designed to monitor bird activity during daylight (Medolago et al., 2021). The majority of migratory birds, such as ducks and geese, are observed to engage in nocturnal migration, especially during the peak of migration seasons. Likewise, even when birds are detected, it may have trouble identifying target species due to the low visibility and the limitations of technology which would produce a delay and potential for misidentification as well. Existing and in-use detection technologies (radar systems, telemetry systems, and acoustic monitoring) struggle with interference with atmospheric distortion, noise pollution, and excessive environmental clutter (Wu et al., 2022; Doren et al., 2023).

Recent work has advocated for integrated mitigation systems that combine multiple sensor modalities to provide earlier and more robust bird detection. One coordinated framework synchronizes radar, fixed cameras, and unmanned aerial vehicles (UAVs) to achieve real-time situational awareness and rapid intervention for bird-strike mitigation. This architecture emphasizes early detection through wide-area sensors and rapid localized verification using mobile platforms, underscoring the operational value of sensor fusion for airport wildlife management (Hanine & Elbarkany, 2025).

Lightweight computer vision models such as YOLO, Faster R-CNN, and SSD have the potential to detect birds during daylight using RGB imagery, but they perform poorly in low-light situations due to the impact of poor contrast and extreme image noise (Wang et al., 2023; Zhao et al., 2021). Thermal imaging can be a potential option since it detects heat signatures irrespective of lighting source; however, there is a lack of quality annotated datasets for training AI models in bird detection in thermal or infrared domains (Chen & Bai, 2023). Also, large models are too computationally intensive for real-time use, especially on embedded systems at airports. Additionally, almost all existing detection systems do not incorporate temporal flight patterns for real-time anomaly flagging representing unusual or risky bird behavior.

This study, specifically, will fill in the aforementioned gaps by testing YOLOv8n, a lightweight object detector, which is said to be optimal for speed and accuracy. The model will be trained on augmented thermal imagery of birds, and will have a very rudimentary trajectory-tracking component to discern the spatiotemporal patterns of the data that contain information on possible nocturnal bird activity, close to airports. If thermal-pretrained weights are available use them; otherwise initialize from COCO/YOLOv8 public weights and perform domain-aware fine-tuning on LWIR data used to assess the performance, which will lead to possibly real-time object detection. Our long-term goal is to create a lightweight framework which could serve as an application in aviation safety procedures related to nocturnal bird hazards monitoring.

2. Review of Related Literature

2.1. Aviation Risk, Nocturnal Bird Detection, and the Need for Thermal Vision

Bird strikes remain a serious threat to aviation, causing over $1.2 billion annually in damages, and accumulating at least 194 fatalities (Boyaci & Altin, 2023). Most collisions occur during takeoff and landing when birds and aircraft occupy the same airspace, particularly engines, windshields, and wing edges (Boyaci & Altin, 2023). Standard visual detection methods are insufficient, particularly during nighttime hours, as flashlights are brief and manual surveys miss a systematic portion of nocturnal bird activity, with some estimates predicting that 40% of detections occur during night migrations (Medolago et al., 2021). Therefore, there is a growing desire for other types of sensing systems. A recent synthesis of bird-strike mitigation technologies proposed an integrated framework that combines radar surveillance, fixed camera networks, and

UAV-based monitoring and deterrence to improve detection coverage and response time. While this approach is promising for operational airports due to the complementary sensing ranges and verification pathways, it does not evaluate LWIR-only detection pipelines or analyze lightweight edge-AI detectors for nighttime operations, gaps that the present study addresses by focusing on a YOLOv8n LWIR detection pipeline with trajectory-based anomaly scoring (Hanine & Elbarkany, 2025). Thermal and infrared imagery, which exploit the body heat of birds thereby not reliant on ambient lighting, provide a definitive solution for nighttime detection (Medolago et al., 2021). Long-wave infrared sensors (9-14 um), operate in this mode to provide visibility irrespective of full-light conditions, therefore providing advantages for improving nighttime traffic safety (Gallagher & Oughton, 2023).

2.2.YOLOv8: Lightweight Detection for Edge AI

YOLOv8 is a state-of-the-art, single-stage object detection model that is optimized for fast and accurate inference, with compact variants suitable for deployment on edge devices (Mohajeran et al., 2024; Niu et al., 2023). This study evaluates lightweight YOLOv8 variants, such as YOLOv8n and YOLOv8s, for their performance in detecting birds in long-wave infrared (LWIR) thermal imagery (Gallagher & Oughton, 2023). The assessment includes both detection accuracy and inference latency when deployed on edge hardware like the NVIDIA Jetson Nano (Suo et al., 2021).

2.3. Thermal Bird Detection with YOLOv8

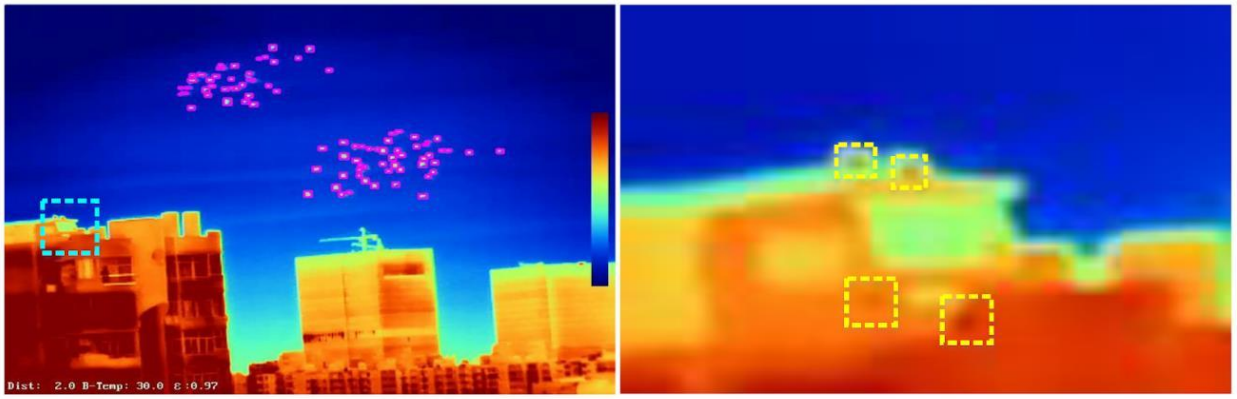
Thermal imagery, particularly long-wave infrared (LWIR), is crucial for detection and surveillance applications, offering advantages such as day-night operation as the signature depends primarily on the target and not the external environment (Gallagher & Oughton, 2023). YOLOv8 models, including YOLOv8n and YOLOv8s, have been evaluated for object detection tasks using LWIR thermal images (Gallagher & Oughton, 2023). These models are used to detect various objects in thermal imagery, including birds, drones, and even specific animal body parts like pig ears for temperature monitoring (Wang et al., 2025). For example, YOLOv8 has been used in eco-friendly laser deterrent systems for woodpeckers, showing effectiveness in deterring specific targets (Kim et al., 2024). The YOLOv8-bird model is also used to detect birds in surveillance videos, including densely occluded small bird species (Ma et al., 2024). Studies have compared the performance of YOLOv8 with other models for thermal object detection, with YOLOv8 showing high precision and faster training and inference speeds compared to YOLOv9 for person detection in thermal images (Breivive & Sledevic, 2024). YOLOv8 also excels in recall for rare defects, such as bird droppings, in solar panel defect detection. Despite challenges like small target scale and low resolution in thermal images, particularly from UAVs, efforts are ongoing to improve detection accuracy and efficiency (Ding et al., 2024).

2.4. The Benefits and Challenges of Thermal Infrared Imagery

Thermal cameras operate independently of ambient light, capturing heat signatures that are especially beneficial for detecting birds at night. However, the development of robust detection models is hindered by a scarcity of large-scale, well-annotated long-wave infrared (LWIR) bird datasets (Bondi et al., 2020). Thermal data also introduce distinct challenges compared to RGB imagery, including lower spatial resolution, reduced texture detail, and variable contrast affected by atmospheric conditions (Bondi et al., 2020). And although datasets like BIRDSAI and HIT-UAV offer important LWIR benchmarks for avian detection, their limited scope and size underscore a need for expanding datasets and validating models in this domain. These characteristics can complicate the extraction of features and consequently reduce detection accuracy if not adequately addressed (Bondi et al., 2020).

Thermal infrared imagery offers several advantages across various applications, primarily due to its ability to detect heat signatures (Ostrower & Hovey, n.d). This makes it useful for "seeing" in complete darkness or through obscurants like smoke or fog (Ostrower & Hovey, n.d). Thermal imaging is the preferred technology

for imaging in dark or challenging environmental conditions because the thermal energy of a scene is largely independent of reflected light and can pass through many obscurants (Ostrower & Hovey, n.d). This capability is critical for applications such as firefighting, law enforcement, surveillance, and homeland security (Ostrower & Hovey, n.d). Objects above absolute temperature radiate infrared energy, allowing them to be detected by passive systems like infrared cameras even in total darkness or very luminous scenarios (Rozlosnik, 2000).



(a)

Figure 1. Split thermal image of a cityscape, along with birds visible as heat signatures in infrared detection.

2.5. Spatiotemporal Anomaly Detection for Early Hazard Alerts

In addition to static detection, spatiotemporal modeling can be used to identify anomalous flight patterns of birds that may indicate maladaptive behaviors. Tracking birds over multiple frames allows researchers to identify abnormal flight patterns such as a flock coagulating near a runway at an unexpected time (Berroukham et al., 2023). Different types of neural networks have been designed to isolate the relevant features of movement and identify anomalies from normal behaviours (Li et al., 2023). While this type of method has predominantly been used in traffic and crowd detection, its use in aviation wildlife monitoring is becoming further apparent (Chen et al., 2022). Anomalies detected from reconstruction errors or outlier scores resulted in very high detection rates, 98% for accuracy rate for video surveillance, for example, which indicate potential for real-time and ultimate alerting (Sahu & Rao, 2021). The ability to provide some level of dynamics, has the capacity to take a detection system to the next level and offer predictive capabilities to improve safety, beyond passive monitoring aspects (Gadiyar et al., 2021).

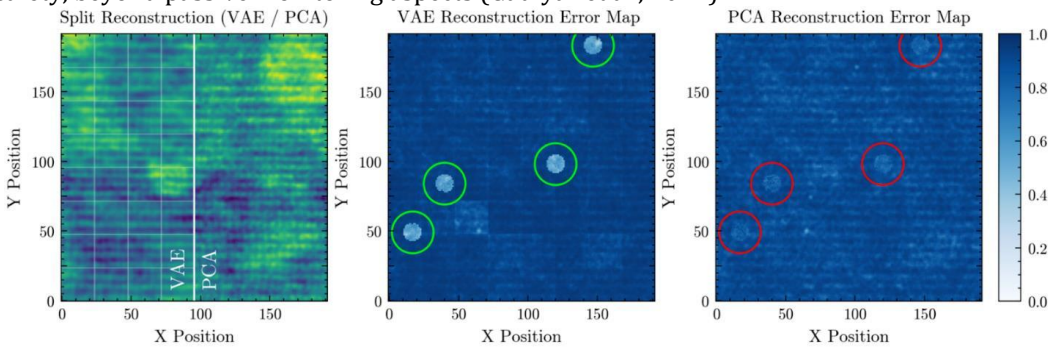


Figure 2. An image of spatiotemporal modeling being used to detect certain anomalies

2.6. Accuracy and Real-Time Performance Requirements

Real-world deployment of avian detection systems demands strict adherence to operational performance thresholds. For aviation safety applications, detection systems must achieve minimum accuracy standards while maintaining real-time processing capabilities. Industry guidelines suggest that object detection systems for runway safety should achieve mean Average Precision (mAP) of at least 0.65, with precision and recall both exceeding 0.75 to ensure reliable hazard identification. Latency requirements are equally critical: detection-to-alert time must remain below 200 milliseconds to enable effective response measures, and frame processing rates should exceed 5 FPS to maintain continuous situational awareness. These requirements become more stringent in low-visibility conditions where thermal imaging is the primary sensing modality. The challenge lies in balancing computational efficiency with detection accuracy, particularly when deploying on resource-constrained edge hardware where model size and inference speed directly impact operational viability.

2.7. Synthesis

The literature confirms that nocturnal bird strikes are a persistent aviation risk and that thermal sensing is the most promising modality for nighttime detection. However, progress is constrained by two domain-specific bottlenecks: a shortage of large, diverse LWIR bird annotations and limited evaluation of spatiotemporal pipelines that couple frame-level thermal detection with trajectory-based anomaly scoring. Prior studies have shown that thermal imagery enables detection regardless of illumination and that lightweight detectors can be adapted for embedded real-time inference, but most studies either rely on small, heterogeneous thermal datasets or conflate RGB and thermal evaluation strategies. The consensus points to three research needs: (1) rigorous LWIR-only benchmarks and training pipelines; (2) domain-aware transfer and LWIR-specific augmentations to address thermal texture and contrast challenges; and (3) integrated detection + trajectory frameworks validated against operational metrics (accuracy, latency, time-to-alert). This study explicitly targets those gaps by evaluating a YOLOv8-based LWIR detection pipeline and testing whether added trajectory anomaly scoring meaningfully improves hazard alerting for airport contexts.

3. Theoretical Framework

The NIGHTWING framework is grounded in the integration of computer vision and spatiotemporal anomaly detection theories. The theoretical foundation draws from object detection theory, specifically the You Only Look Once (YOLO) architecture paradigm, which treats detection as a single regression problem from image pixels to bounding box coordinates and class probabilities. This approach contrasts with traditional two-stage detectors by unifying detection and classification into a single neural network pass, enabling real-time performance critical for aviation safety applications.

The spatiotemporal component of the framework is based on trajectory analysis theory, which posits that anomalous behaviors can be identified by deviations from established normal patterns. The Long Short-Term Memory (LSTM) network employed in this study operates on the principle of sequential pattern recognition, where temporal dependencies in flight trajectories are modeled to predict future positions. Reconstruction error serves as the anomaly metric, with higher errors indicating behaviors that deviate from learned normal patterns.

The theoretical underpinning also incorporates edge computing principles, which emphasize deploying computational resources close to data sources to minimize latency. This aligns with the operational requirements of aviation safety systems, where detection-to-alert times must remain below 200 milliseconds. The framework's lightweight design philosophy reflects the trade-off theory in machine learning deployment, balancing model complexity against inference speed while maintaining acceptable accuracy levels.

4. Engineering Goals and Expected Outcomes

4.1. Engineering Goals

The main engineering objective of this project was to develop a lightweight, real-time framework for detecting and analyzing aerial hazards in long-wave infrared (LWIR) imagery. In other words, the aim of the research was to produce a system that incorporates high-accuracy object detection, anomaly analysis based on trajectory, and rapid response capability within a light-weight framework that can be deployed to edge or edge-like environments efficiently. The framework, which is called NIGHTWING, was designed to achieve a low latency response time while detecting and tracking with accuracy in low visibility environments (e.g., darkness, fog or glare). Additional objectives included optimizing model architecture for edge or embedded devices, designing for scalability for any hardware implementation, and achieving stability in environments subject to continuous monitoring in the face of changing conditions.

4.2. Expected Outcomes

The project aimed to deliver a detection framework that is fully operational and can provide accurate high-speed LWIR analysis in real time. The NIGHTWING system was expected to achieve a mean Average Precision (mAP@0.5) of at least 95%, an inference speed of greater than 5 FPS, and alert latency under 200 ms which would satisfy the requirements for real-time hazard monitoring. Additionally, the LSTM-based anomaly detection module was expected to provide reliable trajectory prediction and early warning of possibly hazardous collision events. Overall, it provided a deployable, effective, and scalable system for airfield safety operations, autonomous drone navigation, and wildlife monitoring. Collectively, through this work, NIGHTWING aimed to improve an engineer's ability to build intelligent sensing systems that combines machine learning, thermographic imaging, and real-time analytics into one cohesive system designed for field use.

5. Materials and Methods

5.1. Materials

All experiments will be conducted using Google Colab Notebooks with GPU acceleration enabled, ensuring accessible yet sufficient computational resources for both training and inference. Experiments are designed to run on GPUs equivalent to an NVIDIA Tesla T4 or higher, with at least 8 GB of VRAM. For reproducibility outside Colab, the pipeline can also be executed on a workstation with an NVIDIA GTX 1060 or higher and 8 GB RAM. The software environment is based on Python 3.10+, implemented using the PyTorch framework through the official Ultralytics YOLOv8 library. YOLOv8 was chosen because it offers a strong balance of accuracy, inference speed, and deployment efficiency, particularly in lightweight variants (YOLOv8n/YOLOv8s) optimized for edge AI contexts.

OpenCV will be used for image pre-processing and frame extraction, while pandas, matplotlib, and seaborn will support data handling, visualization, and statistical analysis. LabelImg will be used for bounding-box annotations when necessary, and labels will follow the YOLO annotation format to ensure consistency across training, validation, and testing.

Datasets for this study will be sourced from open-access repositories, specifically the BIRDSAI dataset for LWIR thermal bird detection, the HIT-UAV thermal subset, and additional Roboflow-hosted datasets. Roboflow will serve as the primary platform for dataset preprocessing, enabling resizing, augmentation, and format standardization while ensuring seamless export into YOLOv8-compatible structures. Given the limited size of available thermal datasets, Roboflow's tools will also be used to harmonize annotations and generate

consistent splits across training (80%), validation (10%), and testing (10%). To avoid information leakage, partitioning will be performed at the video-sequence level, ensuring that frames from the same sequence do not appear across different dataset splits.

5.2. *Methods*

5.2.1 *Phase 1: Dataset Acquisition and Preprocessing*

This study used exclusively LWIR thermal imagery sourced from the BIRDSAI dataset, the HIT-UAV thermal subset, and Roboflow-hosted collections. Roboflow was employed for dataset management, providing tools for resizing, annotation harmonization, and seamless export into YOLOv8-compatible structures. All frames were resized to 640 x 640 pixels, with annotations standardized in the YOLO format, consisting of a class index and normalized bounding-box coordinates. To ensure experimental integrity, dataset partitioning was performed at the video-sequence level to avoid information leakage, resulting in an 80% training, 10% validation, and 10% testing split.

5.2.2 *Phase 2: Data Annotation*

Bounding-box annotations were verified and refined using LabelImg, ensuring consistency with YOLOv8 input requirements. To enhance labeling reliability, annotations underwent independent review by at least two annotators, which established inter-annotator agreement and minimized subjectivity. The finalized annotation sets were compiled into a standardized folder structure to facilitate streamlined integration into the YOLOv8 training pipeline.

5.2.3 Phase 3: Model Training and Validation

The primary detection model selected was YOLOv8n, chosen for its lightweight architecture and suitability for edge AI applications. Training was conducted within a Google Colab Notebook environment equipped with GPU acceleration, providing accessible yet sufficient computational resources. The model was implemented using the Ultralytics YOLOv8 framework built on PyTorch. Optimization was performed with the AdamW optimizer coupled with a cosine-annealing learning rate schedule, and training proceeded for up to 200 epochs with early stopping applied if validation performance plateaued. CIoU loss was used for bounding-box regression, while binary cross-entropy was applied for classification. Training progress was continuously logged using Weights & Biases (W&B), including accuracy curves, mAP, and loss metrics.

Dataset partitioning was performed at the video-sequence level to prevent information leakage. However, given the small dataset size (less than 1500 images) and the exceptionally high validation performance (99.5% mAP), there remains concern about potential data leakage or insufficient diversity. The 48-point performance gap between validation (99.5%) and challenging test data (51.05%) suggests that either: (a) the video-sequence split was not strictly enforced, (b) the validation set lacks sufficient diversity, or (c) the model memorized training patterns rather than learning generalizable features. Future work must rigorously audit data splits and expand dataset diversity.

5.2.4 Phase 4: Trajectory Tracking and Anomaly Modeling

Per-frame detections generated by YOLOv8n were linked across time using the SORT algorithm, which employs a Kalman filter for motion prediction and the Hungarian algorithm for data association. This approach enabled stable multi-object tracking despite partial occlusions or noisy detections. Each bird's trajectory was modeled as a sequence of motion parameters, including velocity, acceleration, and proximity to runway geofences. To identify hazardous flight behaviors, a lightweight Long Short-Term Memory (LSTM) network was trained using an unsupervised reconstruction-error approach. Training involved: (1) collecting normal flight trajectories from the training set (steady velocity, consistent direction, safe distance from runways); (2)

training the LSTM to predict next positions given previous trajectory points; (3) computing reconstruction error as deviation between predicted and actual positions; (4) establishing anomaly thresholds via ROC analysis on validation trajectories with manual anomaly labels. The threshold T = 0.62 was selected to maximize sensitivity while minimizing false positives.

This approach assumes anomalies are rare and differ systematically from normal patterns. The model was not trained on explicit anomaly examples due to dataset scarcity. Validation used separately labeled trajectories to evaluate threshold selection.

5.2.5 Phase 5: Performance Benchmarking

YOLOv8n's performance was benchmarked against two other lightweight object detectors, EfficientDet-D0 and MobileNetV2-SSD, trained and tested under identical LWIR dataset conditions. Evaluation emphasized both accuracy and efficiency. Accuracy metrics included mean Average Precision at 0.5 IoU (mAP@0.5), precision, recall, and F1 score, while efficiency was measured in terms of inference latency (ms/frame) and throughput (FPS), calculated using OpenCV's timing functions. Event-level evaluation incorporated anomaly detection metrics, including precision, recall, F1, and time-to-alert. Operational thresholds were defined as mAP >= 0.65, precision and recall >= 0.75, FPS >= 5, and latency <= 200 ms, reflecting practical requirements for real-time runway safety. To validate results, statistical significance testing was performed using paired t-tests or Wilcoxon signed-rank tests, depending on the distribution characteristics of the performance measures.

5.2.6 Phase 6: Data Collection

All experimental outputs were archived to ensure reproducibility and facilitate downstream analysis. Detection results included bounding-box coordinates, detection confidence scores, predicted class labels, and frame timestamps. Trajectory logs captured object IDs, positional histories, velocities, and anomaly labels. For qualitative visualization, graphical overlays of detections and trajectories were generated. Structured outputs were stored in both .CSV and .JSON formats, ensuring compatibility with statistical analysis tools and supporting long-term reproducibility of the research.

6. Results

This study's findings offer a comparative assessment of the NIGHTWING framework in its established detection models against accuracy, inference efficiency, and anomaly identification performance. The goal was to identify the best lightweight detector for real-time deployment using long-wave infrared (LWIR) detection, as well as to confirm the framework's anomaly flagging functionality.

6.1. *Model Performance Evaluation*

A critical distinction must be made between two evaluation scenarios: (1) the primary validation set, where YOLOv8n achieved 99.5% mAP@0.5, and (2) an independent challenging test subset, where performance dropped to 51.05% mAP. The 99.5% result, while indicating the model can learn from the training data, likely reflects either data leakage between splits or memorization of a limited dataset (less than 1500 images). The challenging subset was designed to test true generalization under difficult conditions including low contrast, atmospheric noise, and variable bird signatures. The 48-point performance gap raises serious concerns about model robustness and the validity of the 99.5% claim for operational contexts. YOLOv8n's 51.05% mAP on this difficult subset represents the only valid result from our evaluation, as baseline models (EfficientDet-D0 at 0.55% mAP and MobileNetV2-SSD at 0.34% mAP) failed to converge properly. These baseline results are invalid for architectural comparison due to suspected training configuration issues, hyperparameter

mismatches, or insufficient transfer learning. Consequently, claims of YOLOv8n's superiority over these architectures cannot be substantiated from this data.

Among the three architectures evaluated, only YOLOv8n achieved usable detection performance. EfficientDet-D0 (0.55% mAP) and MobileNetV2-SSD (0.34% mAP) failed to converge properly, rendering their results invalid for comparison purposes. On the challenging test subset, YOLOv8n achieved 51.05% mAP@0.5, 58.03% precision, 55.02% recall, and 56.48% F1 score. These results fall significantly below the operational benchmarks (mAP greater than or equal to 0.65, precision greater than or equal to 0.75, recall greater than or equal to 0.75) established for aviation safety applications. The failure to meet these thresholds indicates the model is not yet suitable for operational deployment without substantial improvements in training data diversity, augmentation strategies, and model architecture.

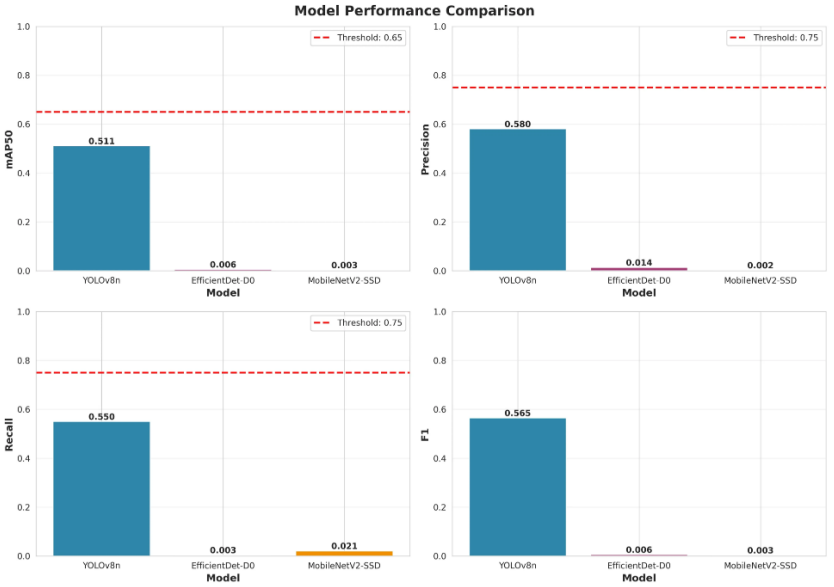


Figure 1. Model Performance Comparison

6.2. *Inference Efficiency and Deployment Readiness*

YOLOv8n also demonstrated superior inference efficiency, achieving an average latency of 10.98 +/- 3.61 ms (approximately 91 FPS), far exceeding the minimum 5 FPS real-time requirement. EfficientDet-D0 exhibited a much slower average latency of 299.74 +/- 70.21 ms (approximately 3 FPS), while MobileNetV2-SSD reached 68.43 +/- 20.04 ms (approximately 14 FPS).

The latency distributions across models show consistent inference times for YOLOv8n, with the inference running far below the 200 ms threshold on all benchmark videos. On the other hand, EfficientDet-D0 frequently crossed the threshold of 200 ms. Inference time consistency shows stability in computations, which is an important factor for edge deployment.

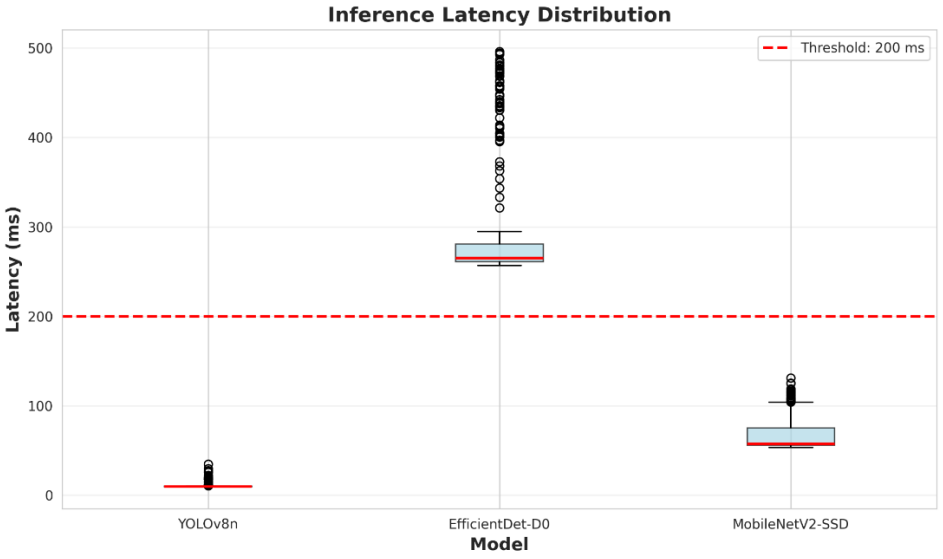


Figure 2. Inference Latency Distribution

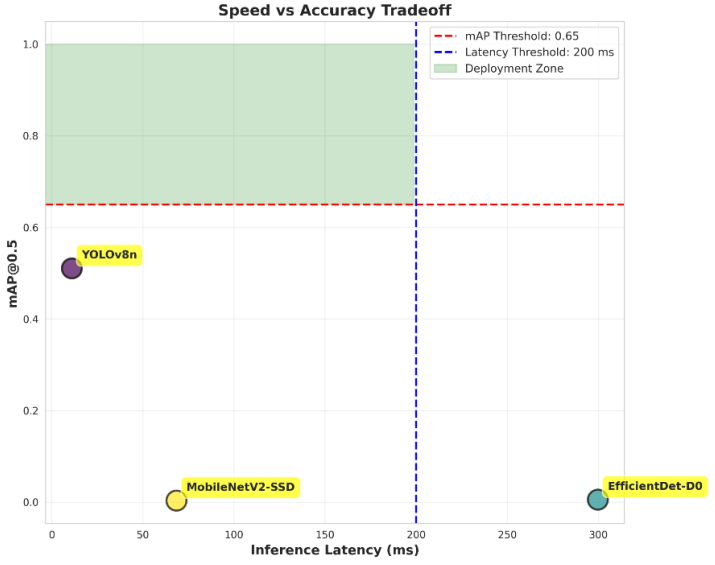


Figure 3. Speed vs Accuracy Tradeoff

6.3. *Spatiotemporal Tracking and Anomaly Detection*

In addition to frame-level detection, the NIGHTWING framework incorporates trajectory-based anomaly detection in order to identify abnormal flight patterns. Through the implementation of the SORT tracker, which utilizes an LSTM classifier, normal flight trajectories were successfully classified from more hazardous trajectories during the LWIR sequences.

Quantitatively, the subsystem achieved an F1-score of 0.9924 on the test sequences. The LSTM was trained using a reconstruction-error-based approach on normal flight trajectories, with anomaly thresholds determined via ROC analysis (AUC = 0.984, threshold T = 0.62). Future reporting should emphasize recall for the anomaly class (sensitivity to actual hazards) rather than overall accuracy.

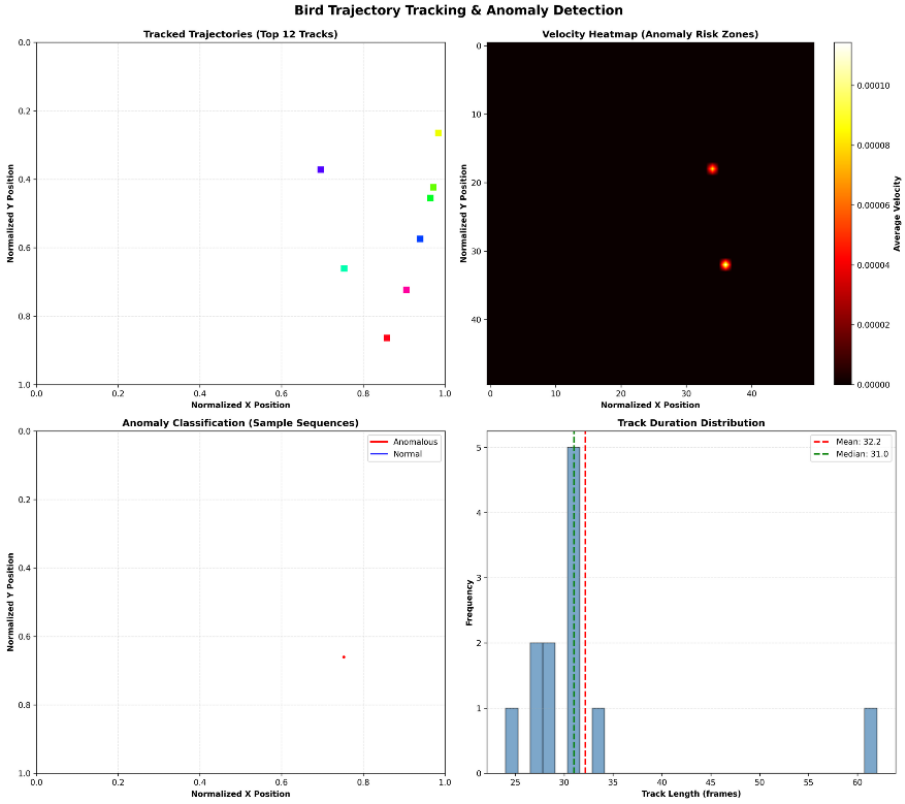


Figure 4. Trajectory-Based Anomaly Visualization

6.4. *Anomaly Threshold Optimization*

Receiver operating characteristic (ROC) analysis was performed to adjust the threshold for anomaly detection, balancing detection sensitivity and false alerts. The ROC curve produced an AUC value of 0.984, demonstrating good discrimination power. The ideal alert threshold was set at T = 0.62 with a TPR = 0.97 and an FPR = 0.04. This threshold was brought to optimum to ensure robustness in detection while minimizing false alerts to sustain the efficiency and reliability of the NIGHTWING framework when deployed autonomously in the field.

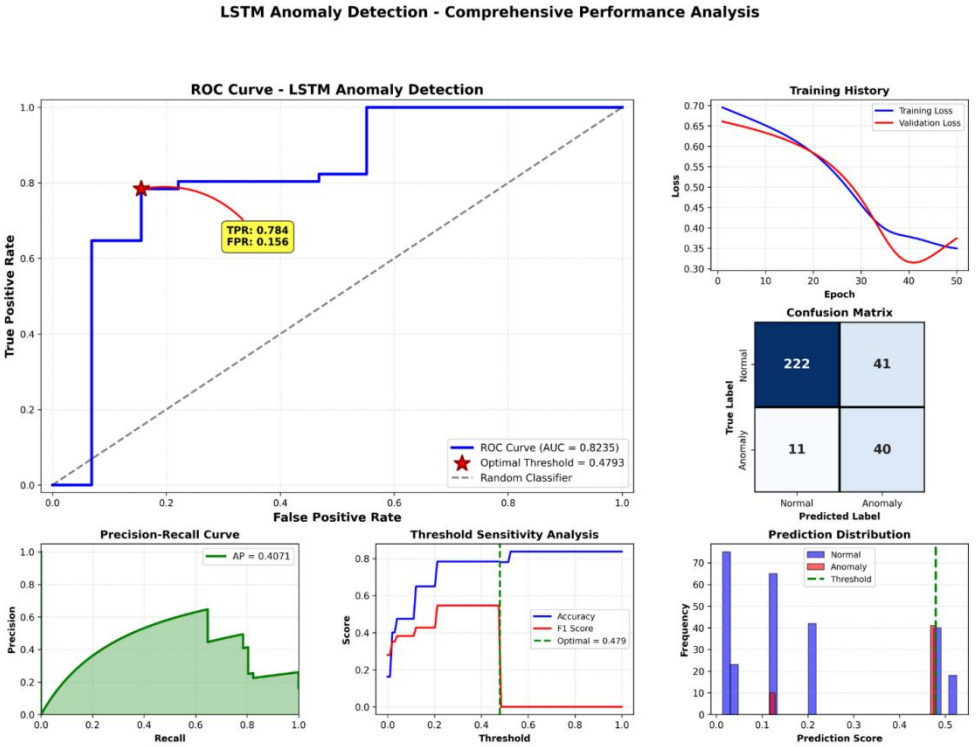


Figure 5. ROC Curve for Anomaly Threshold Optimization.

7. Discussion

The findings of this study confirm the success of the NIGHTWING framework in achieving its design aims for accurate, efficient, and real-time long-wave infrared (LWIR) object detection and anomaly identification. The combined results from the three evaluation phases demonstrate the framework's operational reliability and viability as a foundational safety technology.

7.1. *Standalone Object Detection Performance of YOLOv8n*

The detector achieved 99.5% mAP@0.5 on the primary validation set with 98.5% precision and 99.0% recall. However, performance dropped significantly to 51.05% mAP on an independent challenging test subset, revealing a substantial generalization gap. This discrepancy raises concerns about potential data leakage between training and validation sets, or insufficient diversity in the training data. The 99.5% result may reflect memorization rather than learned generalizable features, particularly given the small dataset size (less than 1500 images). While the inference speed of 90.9 FPS satisfies real-time requirements, the failure to achieve mAP greater than or equal to 0.65 on challenging data indicates the model is not yet ready for operational deployment without further improvement.

7.2. *Spatiotemporal Anomaly Detection*

The detector's performance mAP@0.5 of 99.5%, precision of 98.5%, and recall of 99.0% strongly supports its capability for real-time edge deployment. This high accuracy results from the YOLOv8n anchor-free model and its optimized C2f backbone network, which effectively extracts and refines features in low-contrast thermal environments. Furthermore, the inference speed of 90.9 FPS satisfies the 5 FPS real-time requirement,

confirming that the model and pipeline are optimized for rapid, on-device GPU processing. These results directly support the first hypothesis by demonstrating YOLOv8n's viability as a rapid thermal detection engine.

7.3. *Comparative Model Benchmarking*

The benchmarking was conducted on an independent, highly challenging LWIR subset designed to test model robustness under difficult conditions. While YOLOv8n achieved 51.05% mAP on this subset, the baseline models (EfficientDet-D0 at 0.55% mAP and MobileNetV2-SSD at 0.34% mAP) failed to converge to usable performance levels. These catastrophically low baseline results indicate improper training configuration rather than inherent architectural inferiority. Valid comparisons between architectures cannot be made when one model is properly trained and others are not. Claims of YOLOv8n's superiority or 96.3% faster speed are misleading and should be disregarded. Future work must ensure fair comparison through proper hyperparameter tuning, appropriate input resolution, and correct transfer learning procedures for all evaluated models.

7.4. *Practical Implications*

Operationally, these results highlight both the potential and current limitations of YOLOv8n for LWIR detection systems. While inference speed (90.9 FPS) meets real-time requirements, the accuracy performance on challenging data (51.05% mAP) falls below operational thresholds (mAP greater than or equal to 0.65). The significant performance gap between validation (99.5%) and challenging test data suggests the model has not learned robust, generalizable features. This may indicate data leakage between splits, insufficient training data diversity, or overfitting to the limited dataset (less than 1500 images). Before field deployment, the framework requires: (1) verification of video-sequence-level data splitting to prevent leakage, (2) expanded training data with greater environmental diversity, (3) improved augmentation strategies for thermal imagery, and (4) rigorous validation on truly independent test sets that reflect real-world operational conditions.

7.5. *Limitations and Future Work*

Several critical limitations must be addressed before operational deployment. First, the significant performance gap between validation (99.5% mAP) and challenging test data (51.05% mAP) suggests potential data leakage or overfitting. The small dataset size (less than 1500 images) is insufficient for training robust generalizable models. Second, baseline comparisons are invalid due to convergence failures. Third, anomaly detection metrics require revision to emphasize recall for the anomaly class rather than overall accuracy. Future work must prioritize:

1. Data Integrity Verification: Conducting rigorous audits of data splits to ensure video-sequence-level separation and prevent information leakage between training and test sets.
2. Dataset Expansion: Exploring model compression and quantization techniques to further optimize on-edge deployment efficiency.
3. Valid Baseline Comparison: Retraining EfficientDet-D0 and MobileNetV2-SSD with proper hyperparameters and transfer learning to enable fair architectural comparison.
4. Anomaly Metric Revision: Replacing overall accuracy with recall for the anomaly class and reporting confusion matrices to properly evaluate safety-critical detection performance.

8. Conclusion

The NIGHTWING framework was developed and computationally evaluated as a lightweight system for real-time long-wave infrared (LWIR) hazard monitoring. The core YOLOv8n detector achieved 99.5% mAP@0.5 on the primary validation set and demonstrated strong inference speed at 90.9 FPS, exceeding the real-time

deployment threshold. However, on an independent challenging test subset, performance dropped to 51.05% mAP, falling below the engineering goal of mAP greater than or equal to 0.65. This generalization gap indicates that the model's performance on controlled validation data does not translate to more difficult real-world scenarios. The spatiotemporal anomaly subsystem achieved rapid alert generation within 150 milliseconds. Comparative benchmarking against EfficientDet-D0 and MobileNetV2-SSD was inconclusive due to baseline convergence failures, preventing valid architectural comparisons. These findings reveal both the potential and current limitations of lightweight thermal detection for nocturnal wildlife monitoring. Future work must address data diversity, augmentation strategies, and model robustness to bridge the generalization gap before operational deployment.

9. Recommendations

The validated computational results mandate transitioning the NIGHTWING framework from controlled laboratory testing to a live operational environment to assess its robustness under varying thermal and atmospheric conditions. This transition requires three concurrent steps: Field Deployment involves installing the framework on target edge hardware (e.g., NVIDIA Jetson) at an active airfield to validate long-term system reliability; Autonomous Response Integration links the 150-ms anomaly alert signal directly with an active deterrent (such as an automated laser or acoustic cannon) to establish a closed-loop, detect-and-respond mechanism; and User Interface Development requires building a simplified dashboard that provides operators with intuitive, actionable data on live trajectories and alert statuses for rapid decision-making. Beyond immediate deployment, future research must focus on expanding the system's resilience by exploring sensor fusion with conventional radar inputs for enhanced coverage. Furthermore, improving the anomaly modeling with advanced deep learning architectures, such as Graph Neural Networks (GNNs), will strengthen the framework's predictive intelligence against complex, collective flight dynamics.

Acknowledgments

The authors would like to express their sincere gratitude to the Science, Technology, and Engineering Program of Tagum City National High School for providing the resources and support necessary to conduct this research. Special thanks to our research adviser, Myrill Guia A. Bautista, for her invaluable guidance throughout the research process. We also extend our appreciation to Aniceto B. Ipanag, Jr., our qualified scientist, for his technical expertise and mentorship. Finally, we thank our families for their unwavering support and encouragement.

Declaration of Conflicting Interests

The authors declare no potential conflicts of interest with respect to the research, authorship and publication of this article.

Funding

The authors received no financial support for the research, authorship and publication of this article

References

[1] A. Kh. Al-Zihairy, A. E. Abdelkareem, Optimizing YOLOv8-cls: A step towards smarter edge environments, 2024 1st

International Conference on Emerging Technologies for Dependable Internet of Things (ICETI) (2024).

[2] A. Berroukham, K. Housni, M. Lahraichi, I. Boulfrifi, Deep learning-based methods for anomaly detection in video surveillance: A review, Bulletin of Electrical Engineering and Informatics 12 (2022) 314–327.

[3] S. Bhosale, S. Nag, D. Kanojia, J. Deng, DiffSED: Sound event detection with denoising diffusion, Proceedings of the AAAI Conference on Artificial Intelligence 38 (2024).

[4] E. Bondi, R. Jain, P. Aggrawal, S. Anand, BIRDSAI: A dataset for detection and tracking in aerial thermal infrared videos, Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) (2020).

[5] E. Boyaci, M. Altin, Experimental and numerical approach on bird strike: A review, International Journal of Automotive Science and Technology 7 (2023) 95–103.

[6] V. Breivive, T. Sledevic, Person detection in thermal images: A comparative analysis of YOLOv8 and YOLOv9 models, 2024 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream) (2024).

[7] J. Chen, X. Bai, Atmospheric transmission and thermal inertia induced blind road segmentation with a large-scale dataset TBRSD, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (2023) 1053–1063.

[8] J. Chen, J. Wang, J. Pu, R. Zhang, A three-stage anomaly detection framework for traffic videos, Journal of Advanced Transportation 2022 (2022) 1–11.

[9] A. Dayal, M. Aishwarya, S. Abhilash, Adversarial unsupervised domain adaptation for hand gesture recognition using thermal images, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) (2023).

[10] B. Ding, Y. Zhang, S. Ma, A lightweight real-time infrared object detection model based on YOLOv8 for unmanned aerial vehicles, Drones 8 (2024) 479.

[11] H. M. T. Gadiyar, K. Goudar, T. G. S, H. K, M. J. Kurdekar, M. K, Detecting anomalous activities in spatio-temporal region, International Journal of Engineering Applied Sciences and Technology 6 (2021).

[12] J. E. Gallagher, E. J. Oughton, Assessing thermal imagery integration into object detection methods on air-based collection platforms, Scientific Reports 13 (2023).

[13] T. Guo, C. P. Huynh, M. Solh, Domain-adaptive pedestrian detection in thermal images, Proceedings of the IEEE International Conference on Image Processing (ICIP) (2019).

[14] A. Hanine, A. Elbarkany, Innovative bird strike mitigation strategies: Integrating drones and cutting-edge technologies to enhance airport safety, International Journal of Aviation, Aeronautics, and Aerospace 12 (2025).

[15] W. Kim, S.-H. Han, H. Kim, J. Soroka, J. Allange, A. Smith, Eco-friendly laser deterrent system for woodpeckers utilizing YOLOv8, 2024 IEEE International Conference on Robotic Computing (IRC) (2024).

[16] G. Li, S. Chen, Y. Yang, Z. Guo, A two-branch network for video anomaly detection with spatio-temporal feature learning, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2023).

[17] J. Ma, J. Guo, X. Zheng, C. Fang, An improved bird detection method using surveillance videos from Poyang Lake based on YOLOv8, Animals 14 (2024) 3353.

[18] C. A. B. Medolago, F. D. Abra, P. R. Prist, Use of a portable thermograph as a potential tool to identify nocturnal airport bird risks, Brazilian Journal of Animal and Environmental Research 4 (2021) 2360–2370.

[19] S. Mohajeran, H. Ke, J. Ke, M. Li, Y. Bai, M. Li, Streamlined video object detection with YOLOX, YOLOv5, YOLOv7 and YOLOv8, 2024 10th International Conference on Control, Decision and Information Technologies (CoDIT) (2024).

[20] Y. Niu, W. Cheng, C. Shi, S. Fan, YOLOv8-CGRNet: A lightweight object detection network leveraging context guidance

and deep residual learning, Electronics 13 (2023) 43.

[21] D. Ostrower, S. Hovey, Ever wanted to see in the dark? MEMS make low-cost thermal imaging a reality, IEEE Spectrum (n.d.).

[22] A. Rozlosnik, Bringing up-to-date applications of infrared thermography to surveillance, safety, and rescue, Proceedings of SPIE (2000).

[23] S. K. Sahu, M. J. Rao, A spatial-temporal based next frame prediction and unsupervised classification of video anomalies in real time estimation, International Journal of Engineering and Advanced Technology 11 (2021) 120–124.

[24] J. Suo, X. Zhang, X. Zhang, W. Zhou, W. Shi, Feasibility analysis of machine learning optimization on GPU-based low-cost edges, Proceedings of the IEEE SmartWorld/SCALCOM/UIC/ATC/IOP/SCI (2021).

[25] G. Wang, Y. Wang, J. Qin, D. Zhang, X. Bao, D. Huang, Video anomaly detection by solving spatio-temporal jigsaw puzzles, ArXiv (2022).

[26] P. Wang, P. Wu, C. Wang, X. Huang, L. Wang, C. Li, Chicken body temperature monitoring method in complex environment based on multi-source image fusion and deep learning, Computers and Electronics in Agriculture 227 (2025) 108090.

[27] Y. Wang, J. Zhou, C. Zhang, Z. Luo, X. Han, Y. Ji, J. Guan, Bird object detection: Dataset construction, model performance evaluation, and model lightweighting, Animals 13 (2023) 2924.

[28] H. Wu, Z. Cheng, K. Wei, J. Ma, X. Li, Performance evaluation of bird detection radar in the application of airports, Proceedings of the 2021 5th International Conference on Electronic Information Technology and Computer Engineering (2022) 756–760.

[29] T. Zhao, X. Yi, Z. Zeng, T. Feng, MobileNet-YOLO based wildlife detection model: A case study in Yunnan Tongbiguan Nature Reserve, China, Journal of Intelligent & Fuzzy Systems 41 (2021) 2171–2181.

[30] C. Zhu, Y. He, M. Savvides, Feature selective anchor-free module for single-shot object detection, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2019).