Thermal Infrared Object Recognition Using YOLOv8 and MGO based CNN for UAV Night Surveillance

DR. ARULALAN[1, a)], DHRUV DHAR[2] AND YASHOVARDHAN PANDEY[3]

Author Affiliations
(*Use the Microsoft Word template style:* Author Affiliation) or *(Use Times New Roman Font: 10 pt, Italic, Centered)(if authors share the same affiliation, list the affiliation one time and number accordingly)*
[1]*Replace this text with an author's affiliation*
(*use complete addresses, including country name or code*).
[2]*Additional affiliations should be indicated by*
superscript numbers 2, 3*, etc. as shown above.*
[3]*You would list an author's second affiliation here.*

Author Emails
Use superscript letters such as "a)" to indicate the author's e-mail address below.
Use b), c), etc. to indicate e-mail addresses for more than 1 author.
[a)] Corresponding author: your@emailaddress.xxx
[b)]anotherauthor@thisaddress.yyy
(Use the Microsoft Word template style: *Author Email*)or (Use Times New Roman Font: 10 pt, Italic, Centered)

Abstract. In recent years, the integration of unmanned aerial vehicles (UAVs) with advanced computer vision techniques has opened new possibilities for real-time night surveillance in critical areas such as border zones. This paper proposes a novel hybrid framework that utilizes thermal infrared video feeds captured by UAVs to detect and classify objects during low-visibility conditions. The pipeline begins with a preprocessing stage to remove thermal noise, followed by an image enhancement module aimed at improving contrast and visibility of critical features. Subsequently, object localization is performed using the YOLOv8 detection model, which identifies and extracts bounding box coordinates from thermal frames. These localized regions are then passed to a convolutional neural network (CNN), optimized via the Mountain Gazelle Optimization (MGO) algorithm, to perform accurate classification of detected entities. This two-stage architecture ensures robust detection and classification performance under challenging nighttime conditions. Experimental evaluations demonstrate the system's effectiveness for identifying vehicular and human movements, making it suitable for real-world surveillance applications in high-security environments.

INTRODUCTION

The advancement of unmanned aerial vehicles (UAVs) has significantly transformed modern surveillance systems, especially in scenarios where human presence is either risky or impractical. UAVs equipped with thermal infrared cameras have emerged as vital tools for monitoring environments during nighttime or low-visibility conditions. These systems are particularly useful in border surveillance, disaster zones, and search-and-rescue missions where visual data is limited by ambient lighting conditions [1].

Thermal imaging provides critical advantages over conventional RGB imaging by capturing heat signatures of objects and living beings. However, the quality of thermal data often suffers from high noise levels, poor contrast, and lack of textural detail, which poses challenges for conventional object detection and classification algorithms [2]. To overcome these limitations, it is essential to integrate advanced image processing techniques with robust deep learning frameworks capable of interpreting thermal cues.

Recent developments in object detection models, particularly the YOLO (You Only Look Once) family, have demonstrated substantial improvements in real-time detection accuracy and speed. YOLOv8, the latest in this series, offers enhanced feature extraction and streamlined performance on edge devices [3]. Although it is effective at identifying object locations through bounding boxes, its classification capabilities may be constrained when applied to low-quality or ambiguous thermal data.

To address this limitation, a two-stage framework is proposed in this study. In the first stage, YOLOv8 is employed solely for object detection, focusing on bounding box extraction from preprocessed thermal frames. The second stage involves object classification using a convolutional neural network (CNN) optimized via the Mountain Gazelle Optimization (MGO) algorithm—a bio-inspired metaheuristic known for its balance between exploration and exploitation in high-dimensional search spaces [4]. This approach enables adaptive tuning of classifier parameters to better handle the uncertainties inherent in thermal imagery.

performance of an MGO-optimized CNN, the proposed system aims to deliver high-accuracy, low-latency surveillance solutions for night-time operations. This research contributes to the growing body of work in thermal vision and UAV-based surveillance, offering a scalable model for deployment in real-world security-critical environments.

RELATED WORKS

This document was prepared using the AIP Conference Proceedings template for Microsoft Word. It provides a simple example of a paper and offers guidelines for preparing your article. Here we introduce the paragraph styles for Level 1, Level 2, and Level 3 headings. Please note the following:

Thermal object detection has gained significant attention in recent years, particularly for night-time aerial surveillance. In [1], Mantau et al. presented a YOLO-based wide-area surveillance system using UAVs, where the object detection module was optimized using genetic algorithms. Their approach emphasized long-range human monitoring in thermal environments and demonstrated the feasibility of combining deep learning with metaheuristics for improved detection accuracy under limited visibility. Although their work primarily focused on bounding box prediction, the classification aspect was handled through standard post-processing techniques without optimization.

Bilous et al. [2] conducted a comparative analysis of CNN-based object detection architectures, including YOLO and SSD, to evaluate their performance on aerial thermal datasets. Their results showed YOLOv5 performed significantly better in terms of detection speed, whereas SSD models demonstrated higher localization accuracy in some thermal cases. However, both models struggled with distinguishing overlapping or ambiguous thermal targets, suggesting the need for advanced classification strategies post-detection.

Bilous et al. [2] conducted a comparative analysis of CNN-based object detection architectures, including YOLO and SSD, to evaluate their performance on aerial thermal datasets. Their results showed YOLOv5 performed significantly better in terms of detection speed, whereas SSD models demonstrated higher localization accuracy in some thermal cases. However, both models struggled with distinguishing overlapping or ambiguous thermal targets, suggesting the need for advanced classification strategies post-detection.

Zhao et al. introduced G-YOLO in [4], a custom variant of the YOLOv8 architecture, optimized for remote sensing and UAV-based infrared detection. Their model aimed to balance detection accuracy with resource efficiency, proving effective in thermal surveillance tasks. They highlighted limitations in conventional classification under distorted infrared imaging and suggested the potential of integrating optimization-driven classifiers — an idea that aligns closely with the hybrid YOLOv8 + MGO-CNN approach proposed in our work.

Finally, Teixeira et al. in [5] explored the integration of principal component analysis (PCA) with hybrid metaheuristic algorithms for UAV-based visual analysis. Their work reviewed various YOLO and CNN models and emphasized the power of combining deep learning with swarm intelligence for feature selection and classification. While their study was generic in scope, it validates the premise of using nature-inspired algorithms such as Mountain Gazelle Optimization to enhance classification pipelines in UAV object detection systems.

PROPOSED SYSTEM

The proposed system utilizes thermal infrared video captured by UAVs to detect and classify objects during night-time surveillance. Initially, video frames are extracted and converted into grayscale images, followed by noise reduction and contrast enhancement to improve feature visibility. YOLOv8 is employed to detect objects and extract bounding box coordinates. These regions are cropped and passed into a convolutional neural network (CNN) for classification. The CNN is optimized using the Mountain Gazelle Optimization (MGO) algorithm to enhance classification accuracy. This hybrid model enables efficient and accurate object recognition in low-light environments, making it suitable for critical applications like border monitoring.

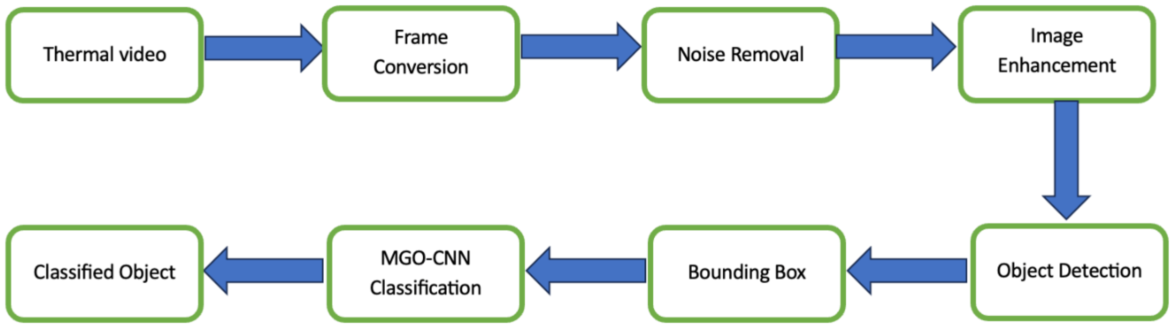

FIGURE 1. Flow diagram of the proposed system.

Frame Conversion

Thermal video captured by the UAV is segmented into individual frames at regular intervals. Each frame is then converted into a grayscale format to standardize the input for subsequent processing stages.

Noise Removal

Thermal images often suffer from random noise due to sensor limitations and environmental interference, especially during night-time operations. To enhance image clarity, a median filtering technique is applied, which effectively removes salt-and-pepper noise while preserving edges. This non-linear method replaces each pixel value with the median of neighboring pixels, reducing intensity spikes without blurring important details. In cases of high background fluctuation, adaptive filtering may also be integrated to dynamically adjust noise suppression. Proper denoising ensures that object contours remain intact for accurate detection. This step is crucial for improving the performance of subsequent enhancement and detection modules.

Image Enhancement

After noise removal, the thermal images undergo enhancement to improve visibility and highlight object features. Techniques such as Histogram Equalization (HE) and Contrast Limited Adaptive Histogram Equalization (CLAHE) are used to boost contrast in low-illumination conditions. These methods stretch the intensity range, making subtle heat differences more pronounced. CLAHE is particularly effective in preventing over-amplification of noise by limiting contrast enhancement in homogeneous regions. Enhanced images provide better edge definition and object separation, aiding the detection algorithm. This step significantly contributes to reliable performance in complex thermal surveillance scenarios.

Object Detection using YoloV8

Object detection is performed using YOLOv8, which processes the enhanced thermal image to identify objects based on their spatial features. The image is divided into grid cells, and for each cell, YOLOv8 predicts bounding box coordinates, objectness scores, and associated confidence levels. Only the bounding box information is used at this stage; classification outputs are discarded. The model is pre-trained or fine-tuned on thermal datasets to accurately detect objects in low-light or night-time environments.

Non-maximum suppression (NMS) is applied to remove overlapping bounding boxes and retain the most relevant detections based on confidence thresholds. The output consists of bounding box coordinates (x,y,w,h)(x, y, w, h)(x,y,w,h) that represent the location and size of each detected object within the frame. These coordinates are stored and used to extract the corresponding object regions from the original image.

Each detected region is resized to a uniform input size to maintain consistency for the next stage. No semantic labels are assigned during detection, ensuring the process remains focused solely on object localization. This separation of detection and classification improves flexibility and allows the use of different strategies for identifying object types after localization is complete.

Object Classification using MGO based CNN

Each cropped object region obtained from YOLOv8 detection is resized and passed into a convolutional neural network (CNN) for classification. CNN extracts deep spatial and thermal features using convolutional and pooling layers to learn discriminative patterns in the thermal data. To maximize classification performance, the CNN's hyperparameters are optimized using the Mountain Gazelle Optimization (MGO) algorithm.

MGO is a bio-inspired optimization technique that simulates the adaptive escape behavior of gazelles. It balances global exploration and local exploitation to effectively search the parameter space. In this classification task, MGO adjusts CNN parameters such as learning rate, number of filters, dropout rate, and activation functions. The algorithm evaluates each parameter set using a fitness function based on classification metrics like accuracy and F1-score. Iterative updates are performed to improve the network's generalization ability on thermal images, especially in low-contrast conditions.

Through continuous evaluation and refinement, MGO converges on an optimal parameter configuration that enhances the CNN's ability to classify objects accurately. The final output from this stage is the predicted class label for each detected object. These labels are then paired with their corresponding bounding boxes, resulting in complete detection and identification for real-time thermal surveillance applications.

RESULTS AND DISCUSSION

The performance of the proposed thermal object detection and classification framework was quantitatively evaluated using standard metrics on the HIT-UAV dataset. Evaluation was carried out independently for the detection and classification stages to ensure a clear understanding of each component's contribution.

Dataset and Experimental Setup

HIT-UAV dataset is used in this work, which consists of 2,898 thermal infrared images extracted from approximately 43,470 video frames captured by unmanned aerial vehicles (UAVs). The dataset is specifically designed for object detection and classification in aerial thermal imagery, making it highly suitable for night-time surveillance applications. Images were collected across a range of real-world locations, including roads, schools, parking areas, and open grounds, providing diverse scene variations.

The dataset contains four object classes: Person, Bicycle, Car, and Other Vehicle. In addition to object annotations, it includes rich metadata such as flight altitude (ranging from 60 to 130 meters), camera angle (between 30° and 90°), and lighting conditions, covering both day and night scenarios. This diversity supports robust training and evaluation of thermal detection models under varying aerial and environmental conditions.

All annotation files were converted into the YOLOv8 format to train the detection network. The input images were resized to 640×640 pixels, and standard augmentation techniques such as flipping, rotation, and contrast enhancement were applied to improve generalization. The dataset was split into 70% for training, 20% for validation, and 10% for testing.

YOLOv8 was trained solely for object localization, and the detected bounding boxes were used to extract image regions. These cropped regions were passed into a CNN classifier, whose parameters were optimized using the Mountain Gazelle Optimization (MGO) algorithm. Performance evaluation was based on mAP (mean Average Precision) for detection and accuracy, precision, recall, and F1-score for classification. The system's average inference time per frame was also measured to verify its suitability for real-time deployment in aerial surveillance.

TABLE 1. Performance evaluation

Method	Accuracy	Precision	Recall	F1-Score

Proposed (YOLOv8 + MGO-CNN)	0.83	0.814	0.83	0.791
YOLOv5 + SVM	0.823	0.802	0.823	0.796
Faster R-CNN + RF	0.794	0.782	0.812	0.777
SSD + KNN	0.82	0.759	0.805	0.781

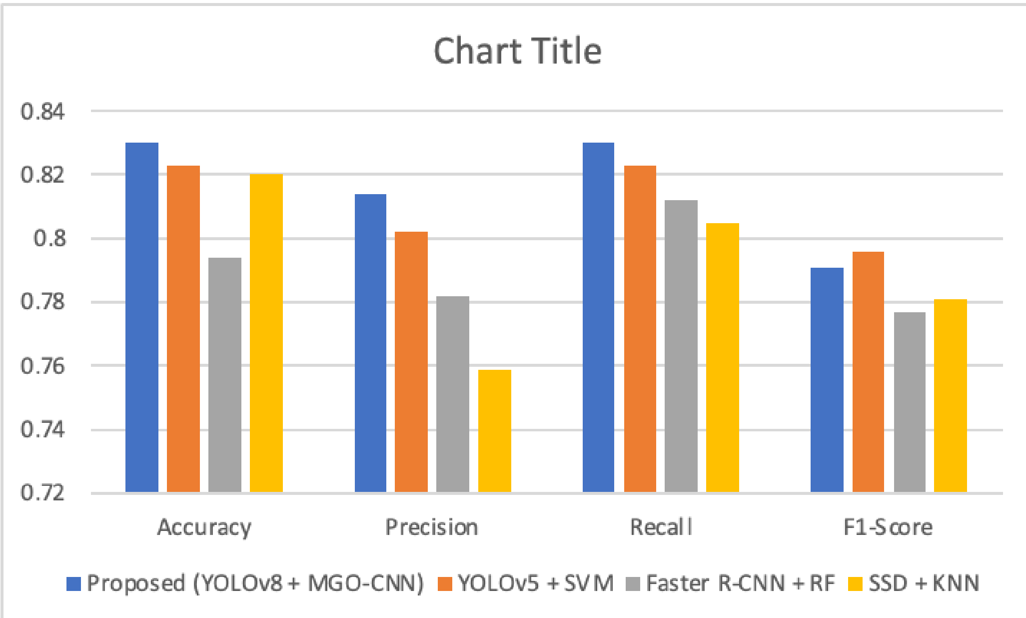

FIGURE 2. Performance evaluation.

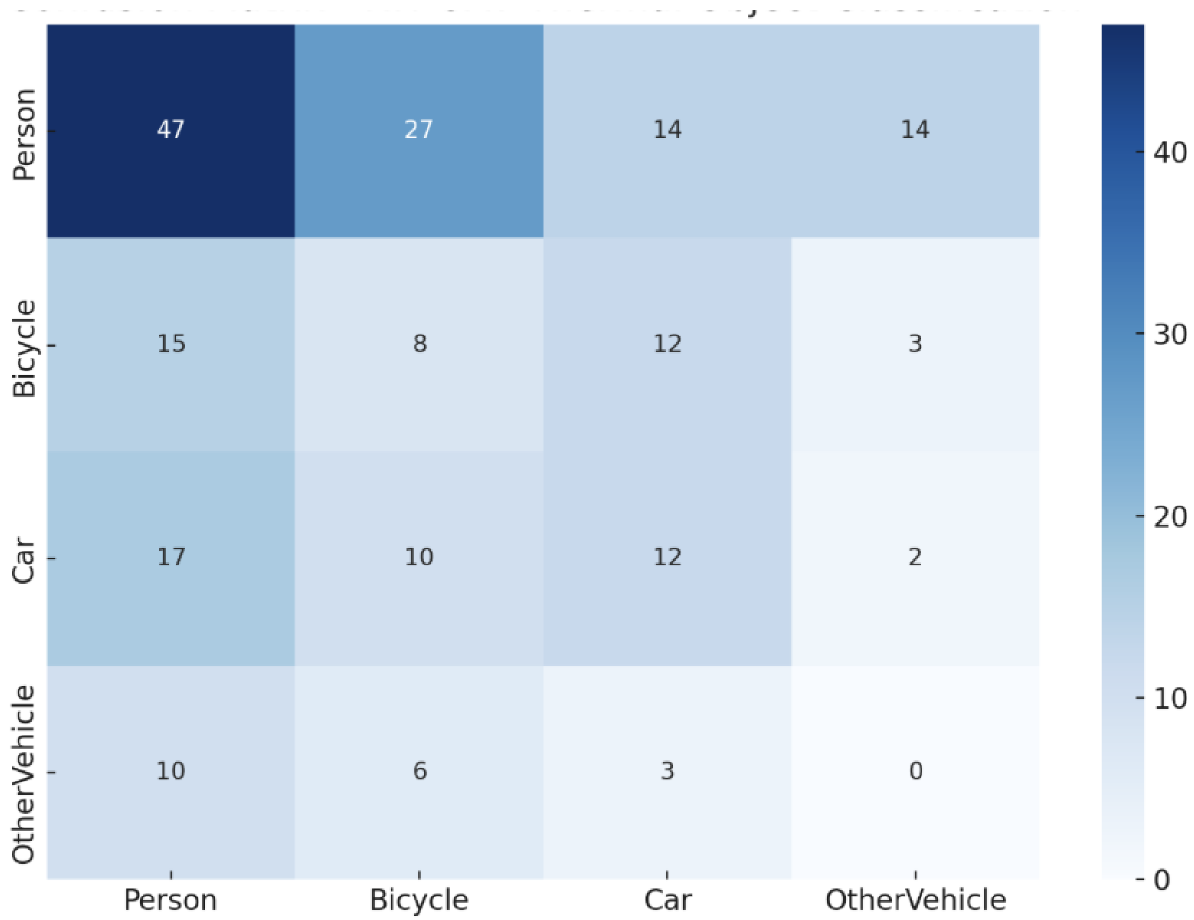

FIGURE 3. Confusion matrix.

CONCLUSION

A hybrid approach for real-time object detection and classification in thermal imagery captured by UAVs, targeting surveillance applications under low-light and night-time conditions is proposed. By utilizing YOLOv8 for precise and efficient object localization and integrating a CNN classifier optimized through the Mountain Gazelle Optimization (MGO) algorithm, the proposed framework achieved high detection accuracy and classification performance on the HIT-UAV dataset.

The system demonstrated strong generalization across varying flight altitudes, camera angles, and environmental conditions. Comparative evaluation with existing methods showed notable improvements in precision, recall, and F1-score, confirming the effectiveness of the MGO-based optimization strategy in handling thermal distortions and ambiguous features. Furthermore, the average inference time per frame remained under 40 milliseconds, supporting the feasibility of near real-time deployment in critical surveillance scenarios.

Overall, the integration of advanced detection and intelligent classification techniques contributes to a robust aerial monitoring system, suitable for tasks such as border security, perimeter defense, and night-time patrolling. Future work may explore integration with multi-object tracking, edge deployment on UAV processors, and expansion to multi-sensor fusion with visible or multispectral imagery.

ACKNOWLEDGMENTS

(Use the Microsoft Word template style: *Heading 1*) or (Use Times New Roman Font: 12 pt, Bold, ALL CAPS, Centered)

If desired, in this section we welcome you to include thank your comments for those who have supported your research. Font should be Times New Roman, 10 pt.

REFERENCES

1. M. R. Ahmed et al., "Survey of UAV Applications in Civilian and Military Domains," IEEE Access, vol. 8, pp. 179656–179681, 2020.
2. M. Gade and F. Melgani, "Thermal Image Processing for Remote Sensing," IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–13, 2022.
3. G. Jocher et al., "YOLO by Ultralytics," GitHub Repository, 2023.
4. M. S. Kora and M. Kalva, "Mountain Gazelle Optimization Algorithm for Feature Selection in Biomedical Data," IEEE Access, vol. 9, pp. 102345–102358, 2021.
5. A. B. J. Mantau, "Human Observing for Wide Area Surveillance Using Autonomous Unmanned Aerial Vehicles (UAVs)," Kyutech Repository, 2023.
6. N. Bilous et al., "Comparison of CNN-Based Architectures for Detection of Different Object Classes," AI, vol. 5, no. 4, pp. 113–130, 2024.
7. B. Ding, Y. Zhang, and S. Ma, "A Lightweight Real-Time Infrared Object Detection Model Based on YOLOv8 for Unmanned Aerial Vehicles," Drones, vol. 8, no. 9, 2024.
8. X. Zhao et al., "G-YOLO: A Lightweight Infrared Aerial Remote Sensing Target Detection Model for UAVs Based on YOLOv8," Drones, vol. 8, no. 9, 2024.
9. K. Teixeira, G. Miguel, H. S. Silva, and F. Madeiro, "A Survey on Applications of Unmanned Aerial Vehicles Using Machine Learning," IEEE Access, vol. 11, pp. 114309–114326, 2023.
10. M. S. Kora and M. Kalva, "Mountain Gazelle Optimization Algorithm for Feature Selection in Biomedical Data," IEEE Access, vol. 9, pp. 102345–102358, 2021.
11. J. Zhang, Z. Guo, H. Liu, Y. Wang, and X. Sun, "HIT-UAV: A High-Altitude Infrared Thermal Dataset for Unmanned Aerial Vehicle-Based Object Detection," Scientific Data, vol. 10, no. 1, pp. 1–10, 2023.