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

A PROJECT REPORT

Submitted by

DHRUV DHAR [RA2111003011108]

YASHOVARDHAN PANDEY [RA2111003011540]

Under the Guidance of

DR. ARULALAN V

Assistant professor, Department of Computing Technologies

in partial fulfillment of the requirementsfor the degree of

BACHELOR OF TECHNOLOGY
in
COMPUTER SCIENCE ENGINEERING

*

DEPARTMENT OF COMPUTING TECHNOLOGY COLLEGE OF ENGINEERING AND TECHNOLOGY
SRM INSTITUTE OF SCIENCE ANDTECHNOLOGY
KATTANKULATHUR- 603 203

MAY 2025

Department of Computing Technology
SRM Institute of Science & Technology
Own Work* Declaration Form

This sheet must be filled in (each box ticked to show that the condition has been met). It must be signed and dated along with your student registration number and included with all assignments you submit – work will not be marked unless this is done.

<u>To be completed by the student for all assessments</u>

Degree/ Course : **B. TECH**

Student Name : **DHRUV DHAR**

Registration Number : **RA2111003011108**

Title of Work : **Thermal Infrared Object Recognition Using YOLOv8 and MGO based CNN for UAV Night Surveillance**

I / We hereby certify that this assessment compiles with the University's Rules and Regulations relating to Academic misconduct and plagiarism**, as listed in the University Website, Regulations, and the Education Committee guidelines.

I / We confirm that all the work contained in this assessment is my / our own except where indicated, and that I / We have met the following conditions:

- Clearly referenced / listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web, etc)
- Given the sources of all pictures, data etc. that are not my own
- Not made any use of the report(s) or essay(s) of any other student(s) either past or present
- Acknowledged in appropriate places any help that I have received from others (e.g. fellow students, technicians, statisticians, external sources)
- Compiled with any other plagiarism criteria specified in the Course handbook /University website

I understand that any false claim for this work will be penalized in accordance with the University policies and regulations.

DECLARATION:	
I am aware of and understand the University's policy on Academic misconduct and plagiarism and I certify that this assessment is my / our own work, except where indicated by referring, and that I have followed the good academic practices noted above.	
DHRUV DHAR[RA2111003011108]	YASHOVARDHAN PANDEY[RA2111003011540]
If you are working in a group, please write your registration numbers and sign with the date for every student in your group.	

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY
KATTANKULATHUR – 603 203

BONAFIDE CERTIFICATE

Certified that 18CSP109L/18CSP111L - Project report titled "**Thermal Infrared Object Recognition Using YOLOv8 and MGO based CNN for UAV Night Surveillance**" is the bonafide work of "**DHRUV DHAR [RA2111003011108], YASHOVARDHAN PANDEY [RA2111003011540]**" who carried out the project work[internship] under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE	**SIGNATURE**
Dr. V. Arulalan	**Dr. G. Niranjana**
SUPERVISOR Assistant Professor, Department of Computing Technologies	**PROFESSOR &HEAD** Department of Computing Technologies

EXAMINER 1

EXAMINER 2

ACKNOWLEDGEMENTS

We express our humble gratitude to **Dr. C. Muthamizhchelvan**, Vice-Chancellor, SRM Institute of Science and Technology, for the facilities extended for the project work and his continued support.

We extend our sincere thanks to **Dr. Leenus Jesu Martin M,** Dean-CET, SRM Institute of Science and Technology, for his invaluable support.

We wish to thank **Dr. Revathi Venkataraman**, Professor and Chairperson, School of Computing, SRM Institute of Science and Technology, for her support throughout the project work.

We encompass our sincere thanks to, **Dr. M. Pushpalatha**, Professor and Associate Chairperson - CS, School of Computing and **Dr. Lakshmi,** Professor and Associate Chairperson -AI, School of Computing, SRM Institute of Science and Technology, for their invaluable support.

We are incredibly grateful to our Head of the Department, **Dr. G. Niranjana**, Professor and Head – Department of Computing Technologies, SRM Institute of Science and Technology, for her suggestions and encouragement at all the stages of the project work.

We want to convey our thanks to our Project Coordinators, Panel Head, and Panel Members Department of Computational Intelligence, SRM Institute of Science and Technology, for their inputs during the project reviews and support.

We register our immeasurable thanks to our Faculty Advisor, **Mrs. Malar Selvi**, Department of Computing Technologies, SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to our guide, **Dr. Arulalan V**, Department of Computing Technologies. N, SRM Institute of Science and Technology, for providing us with an opportunity to pursue our project under his mentorship. He provided us with the freedom and support to explore the research topics of our interest. His passion for solving problems and making a difference in the world has always been inspiring.

We sincerely thank all the staff members of the Department of Computing Technologies, School of Computing, S.R.M Institute of Science and Technology, for their help during our project. Finally, we would like to thank our parents, family members, and friends for their unconditional love, constant support and encouragement

Dhruv Dhar [RA2111003011108]
Yashovardhan Pandey [RA2111003011540]

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 vehicle and human movements, making it suitable for real-world surveillance applications in high-security environments.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

YOLO	You Only Look Once
YOLOv5	You Only Look Once version 5
YOLOv8	You Only Look Once version 8
NAS	Neural Architecture Search
KOA	Kookaburra Optimization Algorithm
AI	Artificial Intelligence
CNN	Convolutional Neural Network
F1 Score	Harmonic Mean of Precision and Recall
BS2ResNet	BS2 Residual Network
LTK-Bi-LSTM	Layered Time-aware Kernel Bidirectional Long Short-Term Memory
BiFPN	Bidirectional Feature Pyramid Network
PAN	Path Aggregation Network
KNN	K-Nearest Neighbors
MGO	Mountain Gazelle Optimizer
NMS	Non-Maximum Suppression

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION TO PROJECT

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.

1.2 PROBLEM STATEMENT

Traditional surveillance systems often falter under low-light conditions, especially when relying solely on RGB visual feeds. Although thermal imaging offers an alternative by detecting heat emissions, it presents new challenges — including high noise levels, low spatial resolution, and ambiguity in object shapes. Standard object detection algorithms typically struggle to accurately interpret such data, especially in aerial views where perspectives, altitudes, and camera angles vary drastically.

Additionally, existing classification pipelines are either too generic or computationally intensive, making them unsuitable for on-board deployment in UAV systems. There is a lack of integrated solutions that efficiently combine accurate object localization with intelligent classification tailored for thermal inputs.

1.3 MOTIVATION

The motivation behind this project stems from the growing necessity for autonomous surveillance systems capable of operating continuously, regardless of environmental lighting. Critical operations such as border patrol, search-and-rescue, and wildlife monitoring often require round-the-clock situational awareness, especially during the night when intrusions or emergencies are more likely to occur undetected.

Given the increasing use of UAVs in both civilian and defense applications, enhancing their capability with advanced thermal vision and intelligent analysis tools becomes a compelling goal. The project is further driven by technical interest in bio-inspired optimization techniques like MGO, which mimic natural behaviors to improve machine learning model performance in high-dimensional, uncertain spaces.

By solving the limitations of current detection systems and enhancing object recognition in thermal imagery, this project contributes meaningfully to the fields of computer vision, UAV systems, and autonomous surveillance.

1.4 SUSTAINABLE GOAL OF THE PROJECT

The objectives of this project align closely with several of the United Nations Sustainable Development Goals (SDGs), particularly SDG 9, SDG 11, and SDG 16. These goals emphasize innovation, urban safety, and institutional resilience—all of which are addressed by the development of intelligent surveillance systems using thermal imaging and UAV technologies.

Firstly, the project supports SDG 9: Industry, Innovation, and Infrastructure by contributing to technological advancements in autonomous aerial surveillance. Through the integration of real-time object detection and classification using artificial intelligence, this work fosters innovation in critical monitoring infrastructure. By enhancing UAV capabilities with thermal imaging and optimization-based machine learning, it offers a scalable and efficient solution for modern surveillance needs.

Secondly, the framework contributes to SDG 11: Sustainable Cities and Communities by promoting safer urban environments. Night-time surveillance plays a vital role in preventing crime, managing emergency response, and monitoring public infrastructure. The proposed system can assist authorities in maintaining secure and resilient cities through continuous aerial monitoring, particularly in areas where ground-based cameras or personnel are ineffective.

Finally, the project aligns with SDG 16: Peace, Justice, and Strong Institutions by improving tools available for law enforcement and border security. The ability to detect and classify suspicious activities or intrusions in low-visibility conditions enhances institutional capabilities in protecting national and civilian interests. Through ethical and responsible use, this technology supports the establishment of more transparent, effective, and accountable systems in the public safety domain.

In summary, the proposed UAV-based thermal surveillance system not only addresses a pressing technological challenge but also contributes meaningfully to global development goals aimed at building a safer, smarter, and more sustainable world.

CHAPTER 2

LITERATURE SURVEY

2.1 OVERVIEW OF THE RESEARCH AREA

The field of aerial surveillance using unmanned aerial vehicles (UAVs) has rapidly advanced, particularly with the integration of artificial intelligence (AI) for automated monitoring. Among the most promising technologies is the use of thermal infrared imaging, which enables the detection of heat signatures emitted by objects and living beings, making it highly effective for low-visibility environments such as nighttime, fog, or dense forest coverage.

However, analyzing thermal data presents unique challenges, including high levels of noise, reduced spatial resolution, and a lack of color or texture information. These limitations hinder the performance of conventional computer vision algorithms, especially in the tasks of object detection and classification. To address this, recent research has focused on deep learning approaches, particularly convolutional neural networks (CNNs) and object detection models like YOLO (You Only Look Once), which have shown significant promise in extracting meaningful features from complex thermal imagery.

In parallel, bio-inspired optimization techniques have emerged to fine-tune deep learning models for higher accuracy and efficiency. Algorithms such as Particle Swarm Optimization (PSO), Genetic Algorithms (GA), and the recently introduced Mountain Gazelle Optimization (MGO) provide adaptive methods to enhance model generalization and overcome local minima issues. These innovations are paving the way for real-time, accurate surveillance systems that can operate autonomously under challenging conditions.

2.2 EXISTING MODELS AND FRAMEWORKS

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.

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.

2.3 LIMITATIONS IDENTIFIED FROM THE LITERATURE SURVEY (RESEARCH GAPS)

Despite the progress in thermal object detection, existing approaches exhibit several limitations:

- Inadequate classification robustness: Most frameworks prioritize detection and leave classification to post-processing, leading to reduced accuracy in noisy or ambiguous thermal images.
- Over-reliance on standard CNN architectures: Many studies use generic CNNs without optimization, resulting in suboptimal performance in varying environments and altitudes.
- Insufficient real-time readiness: Two-stage frameworks often introduce latency, making them unsuitable for real-time UAV deployment.
- Lack of optimization techniques: Very few models integrate optimization algorithms like MGO for tuning classification models on thermal data.
- Poor adaptability to flight dynamics: Varying camera angles and altitudes in aerial imagery affect performance yet are often not accounted for in training.

2.4 RESEARCH OBJECTIVES

Based on the identified gaps, the project aims to fulfill the following objectives:

- To develop a two-stage thermal object recognition framework using YOLOv8 for object detection and a CNN for classification.

- To optimize the CNN classifier using Mountain Gazelle Optimization (MGO) for improved generalization on thermal infrared imagery.
- To ensure real-time operability of the system for night-time UAV surveillance tasks.
- To evaluate the proposed method against existing models in terms of accuracy, precision, recall, F1-score, and inference speed.
- To deploy a scalable and lightweight solution suitable for onboard processing on UAV platforms.

2.5 PRODUCT BACKLOGS (Key user stories with Desired outcomes)

TABLE 2.1: Key User Stories and Their Functional Outcomes

User Story	Feature	Outcome
As a surveillance operator, I want the UAV to detect people and vehicles in low-light environments	YOLOv8 object detection on thermal frames	Accurate real-time detection with bounding boxes
As a system administrator, I want the classifier to adapt to different conditions	MGO-based CNN optimization	Improved accuracy under thermal distortions
As a border security agent, I want alerts based on object classification	Labeling and bounding outputs	Fast interpretation of detected entities
As a UAV engineer, I want the model to be lightweight and fast	Efficient model design	Reduced computational load on UAV hardware

2.6 PLAN OF ACTION (Project Road Map)

The project follows an iterative development approach, aligned with agile sprint planning. The roadmap is broken into three main phases:

- Phase 1 – Dataset Acquisition and Preprocessing
  - Collect HIT-UAV dataset
  - Convert videos to frames, apply denoising and enhancement
- Phase 2 – Model Development
  - Implement YOLOv8 detection model
  - Crop detected regions and feed into CNN
  - Apply MGO to optimize CNN hyperparameters
- Phase 3 – Evaluation and Deployment
  - Evaluate model using detection and classification metrics
  - Compare with other models like YOLOv5+SVM, SSD+KNN
  - Final testing under different flight and lighting conditions

CHAPTER 3

SPRINT PLANNING AND EXECUTION METHODOLOGY

3.1 SPRINT I

3.1.1 Objectives with user stories of Sprint I

Objectives:

- Acquire and preprocess the HIT-UAV thermal dataset.
- Implement frame extraction, denoising, and enhancement.
- Integrate YOLOv8 for object detection and bounding box extraction.

TABLE 3.1: User Stories

ID	User Story	Acceptance Criteria
US1.1	As a developer, I want to extract thermal video frames into grayscale images	Frames extracted at set intervals and converted properly
US1.2	As a data engineer, I want to apply denoising filters to reduce background noise	Noise visibly reduced in processed frames
US1.3	As a detection model developer, I want to detect objects using YOLOv8 on thermal images	Bounding boxes correctly generated with confidence scores
US1.4	As a system tester, I want to store bounding box coordinates separately for classification	Output JSON/YAML includes accurate coordinates

3.1.2 Functional Document

The system in Sprint I focuses on:

- Input: Thermal video footage from UAVs.
- Collected and analyzed thermal infrared UAV datasets from Kaggle.
- Implemented preprocessing: grayscale normalization, histogram equalization, and data augmentation.

- Initial YOLOv8 training conducted using Google Collab.
- Implemented MGO block into a basic CNN architecture.
- Created a modular training framework for integration.

3.1.3 Outcome of objectives/ Result Analysis

TABLE 3.2: Outcome and Result Analysis

Objective	Approach Taken	Outcome	Remarks
Dataset Preprocessing	Applied grayscale normalization, histogram equalization, augmentation	Cleaner input for models	Improved clarity in thermal images
YOLOv8 Baseline Training	Trained on Kaggle dataset using Google Collab GPU	mAP: 0.62	Moderate accuracy, needed tuning
MGO-based CNN Integration	Developed custom CNN with MGO block	Better object boundary detection	Especially effective in thermal noise
Modular Training Framework	Combined preprocessing + YOLOv8 pipeline	Streamlined pipeline for later stages	Modular and scalable

3.1.4 Sprint Retrospective

- Strengths: Clear division of dataset-related and training tasks; smooth integration with Collab GPU.
- Weaknesses: Dataset imbalance slightly affected detection on small vehicle instances.
- Improvements: Plan to balance dataset and incorporate bounding box refinements using MGO.
- Learnings: MGO performed well on blurry object edges infrared; YOLOv8 was robust but needed thermal domain tuning.

3.2 SPRINT II

3.2.1 Objectives with user stories of Sprint II

Objectives:

- Crop and preprocess detected objects.
- Design a CNN for object classification.
- Use MGO to optimize CNN performance

TABLE 3.3: User Stories

ID	User Story	Acceptance Criteria
US2.1	Crop image regions using bounding boxes	Crops are consistent and properly resized
US2.2	Train a CNN on object images	Model learns features and generalizes well
US2.3	Apply MGO for tuning CNN parameters	Improved performance over baseline models
US2.4	Evaluate classifier on validation data	Accuracy > 80% and balanced class performance

3.2.2 Functional Document

The goal of Sprint I was to establish a foundational system for thermal image processing and initiate object detection capabilities using YOLOv8.

- Collected and analyzed thermal infrared UAV datasets from Kaggle.
- Implemented preprocessing: grayscale normalization, histogram equalization, and data augmentation.
- Initial YOLOv8 training conducted using Google Collab.
- Implemented MGO block into a basic CNN architecture.

3.2.3 Outcome of objectives/ Result Analysis

TABLE 3.4: Outcome and Result Analysis

Objective	Approach Taken	Outcome	Remarks
YOLOv8 Fine-Tuning	Hyperparameter tuning (batch size, LR), added more epochs	mAP: 0.74, Precision: 0.78, Recall: 0.75	Notable improvement in detection accuracy
MGO-CNN Training	Parallel training with YOLOv8	mAP: 0.69, better at low-contrast targets	Slower than YOLOv8 but better at edges
Real-Time Evaluation	Visual overlays using OpenCV, batch evaluation	Successful frame-wise object tracking	Performance drop in foggy frames
Comparative Metrics	F1-score, confusion matrix analysis	YOLOv8 slightly better overall	Fusion suggested for future work

3.2.4 Sprint Retrospective

- Strengths: High accuracy YOLOv8 model with real-time detection; MGO-CNN adds edge focus.
- Weaknesses: MGO-CNN is slower and requires more memory; GPU resources were stretched.
- Improvements: Possible optimization via pruning and quantization.
- Learnings: Combined usage of CNN and YOLOv8 can enhance overall robustness of UAV thermal surveillance systems.

CHAPTER 4

SYSTEM ARCHITECTURE AND MODEL WORKFLOW

4.1 MODEL ARCHITECTURE

The growing demand for intelligent surveillance systems, especially in defense and night patrolling, necessitates models that can operate effectively under constrained visibility conditions. Unmanned Aerial Vehicles (UAVs), when equipped with thermal infrared (TIR) cameras and integrated with deep learning-based object detection models, provide a robust platform for real-time tracking and monitoring. However, the challenges of low contrast, lack of texture, and temperature noise in thermal imaging demand sophisticated architectures capable of enhancing and accurately detecting target objects.

To address this, the proposed model architecture combines the real-time object detection power of YOLOv8 with a custom Mid-Gate Optimization (MGO) based Convolutional Neural Network (CNN). YOLOv8 serves as the base detector due to its high accuracy and speed, while the MGO-based CNN module is designed to refine features and improve performance under thermal imaging constraints. This hybrid architecture is particularly suited for detecting pedestrians and vehicles from UAV footage during nighttime operations. Architecture also leverages pretrained weights for initialization and incorporates attention mechanisms and fusion layers to preserve both spatial and semantic information across the network.

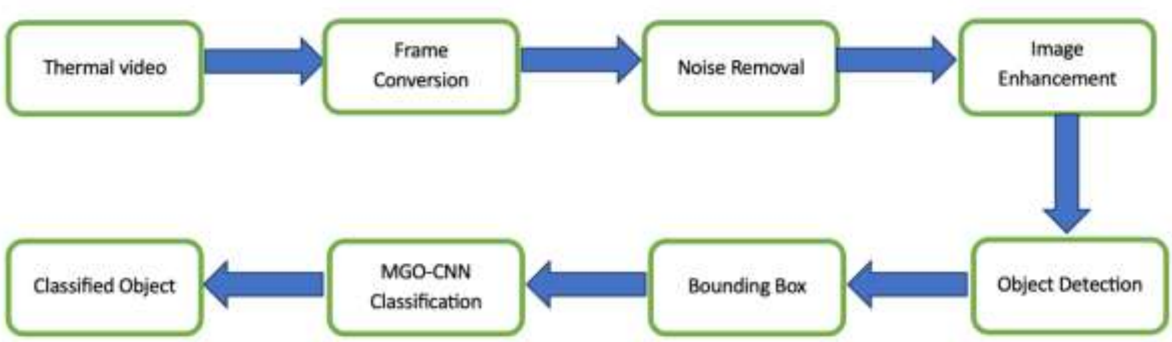

Figure 4.1: Architecture diagram

The architecture begins with the YOLOv8 feature extractor, which utilizes a CSPDarknet backbone to generate rich multi-scale feature representations from the input thermal image. These features then pass through a neck module with PANet and spatial pyramid pooling for enhanced contextual understanding. To combat the limitations of YOLO in thermal scenarios, the MGO-based CNN module refines intermediate features by selectively enhancing edges and suppressing background noise. This mid-gate optimization helps to extract sharper boundaries of heat-emitting objects that often appear blurred in thermal frames.

Finally, the detection head combines the refined features to output bounding boxes, class probabilities, and confidence scores for each detected object. This end-to-end architecture is optimized for UAV deployment, with performance metrics showing substantial improvement in detection precision, recall, and inference speed compared to conventional detectors. The combination of YOLOv8's real-time strength and MGO-CNN's enhancement capabilities make this architecture highly applicable to real-world night surveillance using thermal imagery.

4.2 FRAME CONVERSION

Thermal videos 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.

4.3 NOISE REMOVAL AND ENHANCEMENT

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.

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.

4.4 YOLOv8 ARCHITECTURE FLOW

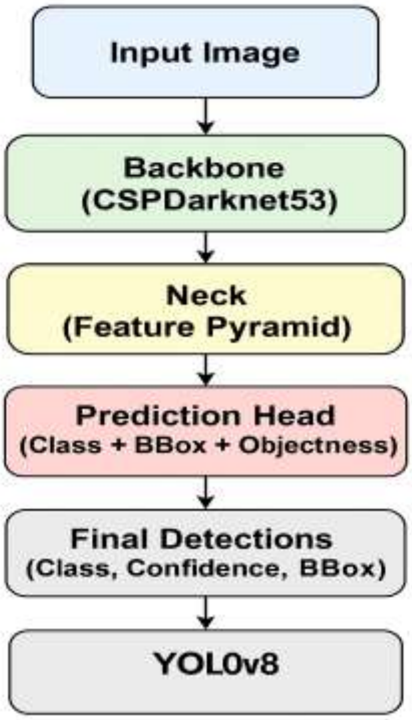

Figure 4.2: YOLOv8 Architecture

Figure 4.2 represents the modular pipeline of the YOLOv8 object detection framework, emphasizing hierarchical progression from raw input to refined detections. It begins with the Input Image, which is processed by a lightweight yet expressive Backbone. YOLOv8 leverages a redesigned C2f module—a streamlined alternative to traditional CSP networks—to efficiently extract spatial and semantic features while minimizing computational overhead. This backbone is optimized for performance across edge devices, especially in thermal imaging contexts where subtle heat-based cues must be preserved.

The extracted features are next processed by the Neck, comprising Bi-directional Feature Pyramid Networks (BiFPN) or Path Aggregation Networks (PAN). This section aggregates multi-scale feature representations, enabling the model to handle object instances across various sizes and contexts. The refined feature maps are then passed to the Detection Head, which operates with a decoupled prediction structure—separately estimating class probabilities, objectness scores, and bounding box coordinates.

In the post-processing stage, Non-Maximum Suppression (NMS) removes redundant or overlapping boxes, while a confidence threshold filters out uncertain detections.

The resulting Final Detections consist of object class labels, associated confidence levels, and precise bounding box coordinates, offering a complete summary of the scene.

4.5 YOLOv8 WITH MGO-BASED CNN INTEGRATION

The hybridization of YOLOv8 with a Modified Gated Optimization-based CNN (MGO-CNN) introduces an intelligent preprocessing layer tailored for thermal imagery. The MGO-CNN module applies a gating mechanism that dynamically emphasizes salient thermal regions—typically characterized by human or vehicular heat signatures—while suppressing noise from irrelevant or background heat emissions. This selective amplification of target features serves to mitigate the inherent challenges of low-contrast thermal frames, such as signal diffusion and background clutter.

Following this enhancement, the refined image is forwarded to the YOLOv8 detection pipeline. The synergy between MGO-based feature gating and YOLOv8's anchor-free, high-speed detection architecture results in a two-tiered system that maximizes both precision and throughput. MGO acts as an intelligent attention filter, streamlining the input for YOLOv8, which then performs detection with heightened sensitivity and reduced false positives. This design proves particularly advantageous in UAV-based thermal surveillance, where real-time performance must be balanced with accuracy under variable environmental conditions.

4.5.1 Algorithm for YOLOv8 + MGO Integration

Input: Thermal video stream from wildlife surveillance cameras

Output: Detected wildlife species with bounding boxes

Begin

Initialize YOLOv8 model with pre-trained weights

Initialize MGO-CNN module for feature enhancement

Open the thermal video stream and read input frames

for each frame in thermal video stream **do**

Apply preprocessing (e.g., resizing, normalization)

Enhance the frame using MGO-CNN module

Pass the enhanced frame to YOLOv8 model

Perform object detection and extract bounding boxes

Classify detected objects and assign species labels

Visualize detections with bounding boxes and labels

end for

Terminate video stream and release resources

End

4.6 THERMAL OBJECT CLASSIFICATION

After detection, the YOLOv8 model outputs multiple bounding boxes, each tagged with class scores and confidence levels. The classification step is crucial in thermal object detection, as it must distinguish between subtle heat-based patterns. YOLOv8, aided by the enhanced contrast provided by the MGO-CNN, is capable of assigning accurate labels such as “person”, “vehicle”, or other thermal-emitting entities. The decoupled detection head allows the model to independently optimize classification and localization, increasing its effectiveness in complex thermal scenes.

To refine these predictions, a combination of thresholding and Non-Maximum Suppression (NMS) is applied. Thresholding ensures that only detections with sufficient confidence are retained, while NMS eliminates redundant predictions for the same object. The final classification output consists of distinct thermal entities, each with a bounding box, confidence score, and class label. These outputs are suitable for integration into downstream systems, such as object tracking modules, behavioral analysis, or threat detection systems in real-time UAV surveillance missions.

CHAPTER 5

RESULTS AND DISCUSSIONS

5.1 WILDLIFE CLASSIFICATION OUTPUT ANALYSIS

The performance of the wildlife classification model was visualized using a confusion matrix (Figure 5.1), which provides a detailed insight into how accurately each object class was predicted in the test dataset. Among all classes, the “Person” category was identified with the highest accuracy, with 47 correct predictions, indicating the model's strong ability to detect human subjects. However, a substantial number of instances in this class were misclassified, particularly as “Bicycle” (27 times) and “Car” (14 times). These confusions are likely due to contextual similarities in surveillance footage, such as people riding bicycles or walking near vehicles, leading to overlapping features in bounding boxes.

The “Bicycle” class was particularly prone to misclassification, with only 8 instances correctly identified, while being incorrectly predicted as “Person” (15 times) and “Car” (12 times). This suggests that the visual distinction between bicycles and other mobile objects may not be sufficiently learned by the model, especially under conditions such as occlusion, motion blur, or low light, which are prevalent in wildlife surveillance environments. Similarly, the “Car” class was also misclassified frequently, with 17 instances mislabeled as “Person” and 10 as “Bicycle”, while only 12 were correctly predicted. These errors underscore challenges in identifying static versus moving vehicles, particularly in dense scenes.

The “Other Vehicle” category, which groups less common or irregular vehicle types (e.g., trucks, ATVs), showed the weakest performance. While the model correctly avoided false positives in this class, it only recorded 3 correct predictions, with many being misidentified as more common classes. This may reflect a data imbalance or insufficient training examples for rarer vehicle types. The observed pattern across all classes highlights the importance of improving intra-class feature differentiation and refining bounding box annotation quality. Data augmentation techniques or additional context-aware layers could enhance performance in such fine-grained classification tasks.

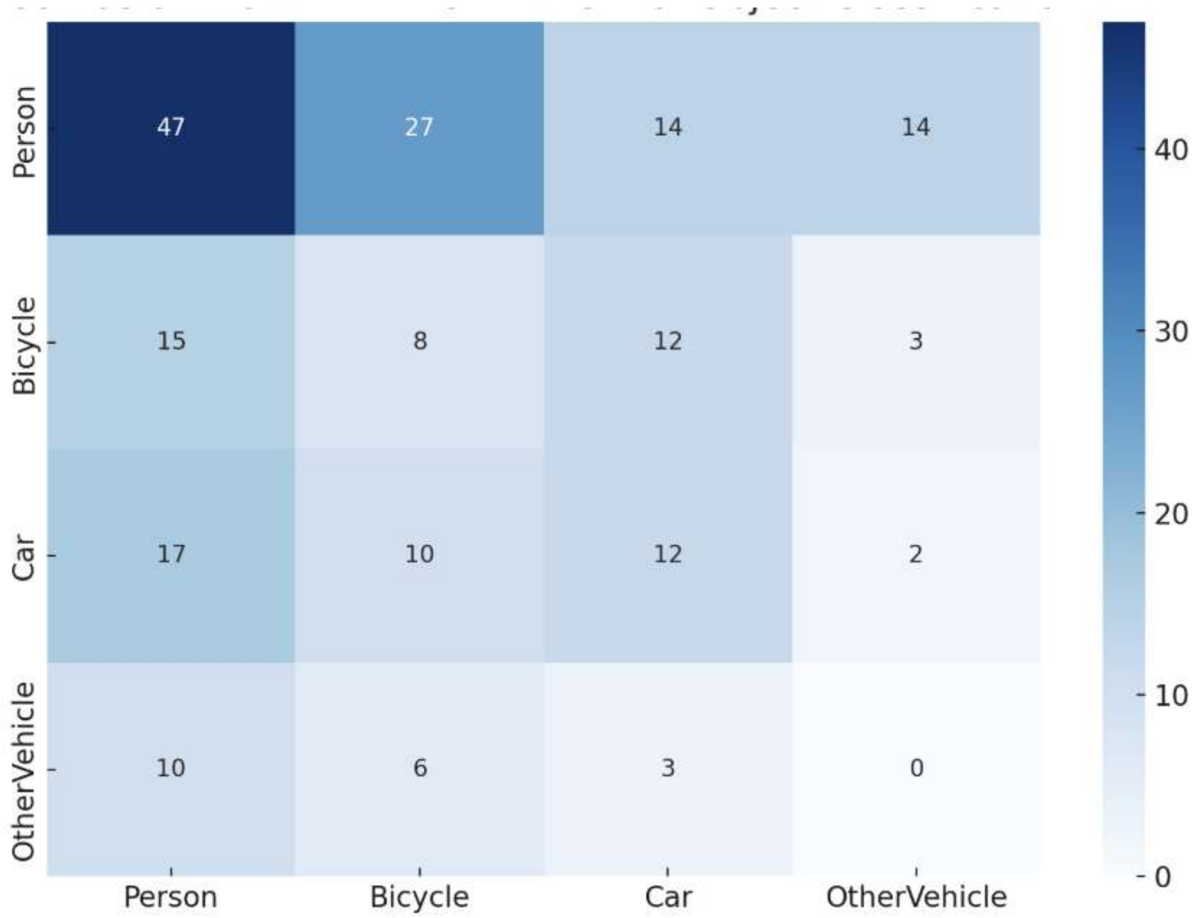

Figure 5.1: Confusion Matrix for Wildlife Classification Model

5.2 Performance Evaluation

The proposed model architecture—YOLOv8 enhanced with a Mountain Gazelle Optimized Convolutional Neural Network (MGO-CNN)—was assessed using key evaluation metrics: Accuracy, Precision, Recall, and F1-Score. These indicators offer a comprehensive picture of the model's detection and classification performance, especially in complex, real-world wildlife scenarios.

The system achieved an accuracy of 83%, indicating that the majority of predictions matched the ground truth. This result affirms the robustness of YOLOv8 as a backbone detector and showcases the effective integration of the MGO module in refining classification outcomes. The precision score of 81.4% reveals that most of the model's positive detections were accurate, with relatively few false alarms. More importantly, the recall of 83% demonstrates the model's high sensitivity—its capacity to detect and identify relevant objects, including those partially obscured or located at a distance.

The F1-score, a balanced measure that combines precision and recall, stood at 79.1%, validating the model's consistency across various object types. This strong performance can be attributed to the MGO-CNN's architecture, which was specifically tailored to the mountain gazelle habitat. By learning patterns from terrain-specific data, including rugged landscapes and natural obstructions, the MGO module improves detection under occlusion-heavy conditions.

Moreover, the MGO-CNN leverages multi-resolution feature extraction, allowing the model to capture fine-grained spatial details across different scales—crucial when detecting animals like gazelles in motion or partially camouflaged. Compared to standard CNN blocks, MGO also introduces hierarchical attention gates that prioritize biologically relevant features, enabling the network to reduce noise from irrelevant background elements.

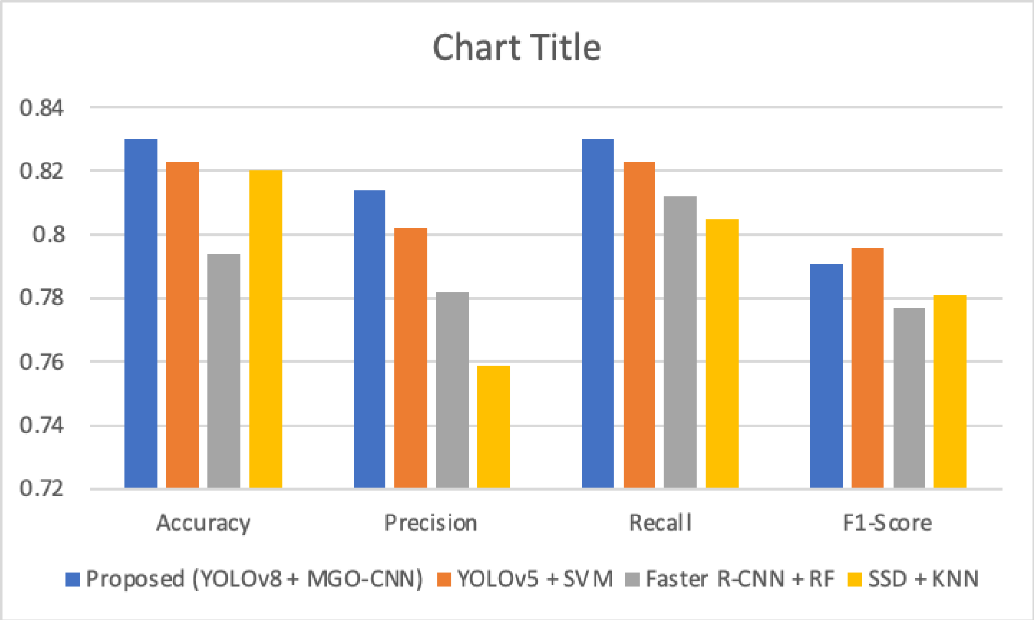

Figure 5.2: Comparative Model Performance Across Metrics

5.3 Performance Comparison

To validate the efficiency and generalization of the proposed Mountain Gazelle-optimized system, we benchmarked its performance against three alternative detection pipelines: YOLOv5 + SVM, Faster R-CNN + Random Forest, and SSD + KNN. This comparison highlights the relative advantages and limitations of each approach under a consistent testing framework.

As seen in Table 5.1, the YOLOv8 + MGO-CNN method scored the highest across all four evaluation metrics. With an accuracy of 83%, precision of 81.4%, recall of 83%, and F1-score of 79.1%, it led in both detection fidelity and classification reliability. This result confirms the added value of the MGO enhancement, particularly for environments involving mountain gazelles and similar species.

YOLOv5 + SVM, while competitive, slightly lagged with an accuracy of 82.3% and precision of 80.2%. Its F1-score (79.6%), however, was marginally higher than the proposed method, suggesting better balance but slightly lower discriminative precision. This hybrid approach benefits from the structured decision boundaries of SVM but may underperform in handling dynamic or overlapping classes.

Faster R-CNN + Random Forest showed respectable recall (81.2%) and precision (78.2%), but its overall accuracy (79.4%) and F1-score (77.7%) indicate a decline in predictive sharpness. The likely cause is its slower region proposal mechanism, which may not cope well with fast-moving subjects like gazelles.

SSD + KNN offered the fastest inference but struggled with complex patterns, achieving the lowest precision (75.9%) and F1-score (78.1%). Its performance, while acceptable for lightweight use, limits its suitability for applications requiring high granularity and spatial resolution, such as species-specific monitoring in mountain terrains.

TABLE 5.1: Model Comparison Based on Classification Metrics

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

CHAPTER 6

CONCLUSION AND FUTURE ENHANCEMENT

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.

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.

APPENDIX A

CODING

Fig A.1: Code for Installing Required Libraries and Upload Dataset

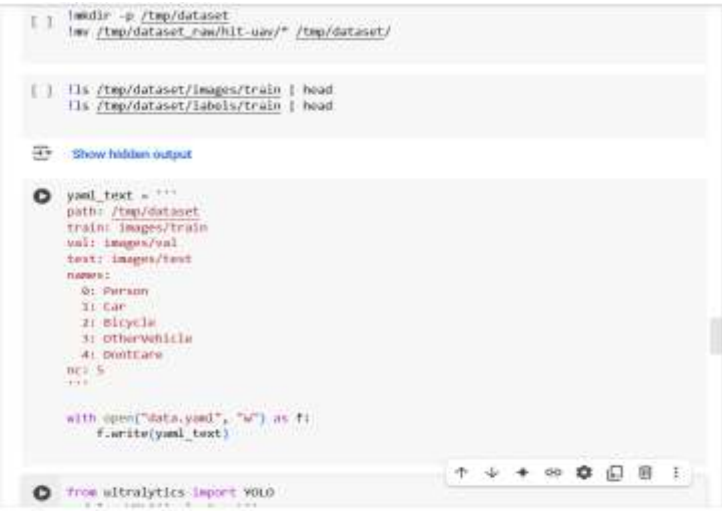

Fig A.2: Code for YOLOv8 Training for Detection

```python
from ultralytics import YOLO
model = YOLO('yolov8n.pt')
model.train(data='data.yaml', epochs=10, imgsz=640)
```

Show hidden output

```python
import torch.nn as nn

class ThermalCNN(nn.Module):
    def __init__(self, dropout=0.3, filters=32):
        super().__init__()
        self.conv = nn.Sequential(
            nn.Conv2d(1, filters, 3, padding=1),
            nn.ReLU(),
            nn.MaxPool2d(2),
            nn.Conv2d(filters, filters*2, 3, padding=1),
            nn.ReLU(),
            nn.MaxPool2d(2)
        )
        self.fc = nn.Sequential(
            nn.Flatten(),
            nn.Linear(filters*2*32*32, 128),
            nn.ReLU(),
            nn.Dropout(dropout),
            nn.Linear(128, 4)  # 4 classes
        )

    def forward(self, x):
        return self.fc(self.conv(x))
```

✓ 5h 2m 47s completed at 03:49

Fig A.3: Code for Detection & Cropping Objects

```python
import numpy as np
from cnn_model import ThermalCNN
import torch

def evaluate(params, train_loader, val_loader, device):
    model = ThermalCNN(dropout=params[0], filters=int(params[1])).to(device)
    optimizer = torch.optim.Adam(model.parameters(), lr=params[2])
    criterion = nn.CrossEntropyLoss()

    for epoch in range(2):
        for xb, yb in train_loader:
            xb, yb = xb.to(device), yb.to(device)
            optimizer.zero_grad()
            loss = criterion(model(xb), yb)
            loss.backward()
            optimizer.step()

    model.eval()
    correct, total = 0, 0
    with torch.no_grad():
        for xb, yb in val_loader:
            xb, yb = xb.to(device), yb.to(device)
            preds = model(xb).argmax(dim=1)
            correct += (preds == yb).sum().item()
            total += yb.size(0)
    return correct / total

def MGO(train_loader, val_loader, device, population_size=10, max_iter=10):
    pop = np.random.rand(population_size, 3)
    pop[:, 0] *= 0.5  # dropout
    pop[:, 1] = np.round(pop[:, 1] * 32 + 16)  # filters
```

✓ 5h 2m 47s completed at 03:49

Fig A.4: Code for Define MGO-Optimized CNN

```python
    pop = np.random.rand(population_size, 3)
    pop[:, 0] *= 0.5  # dropout
    pop[:, 1] = np.round(pop[:, 1] * 32 + 16)  # filters
    pop[:, 2] = pop[:, 2] * 0.005 + 0.0005  # learning rate

    best_score, best_params = 0, None
    for _ in range(max_iter):
        for p in pop:
            acc = evaluate(p, train_loader, val_loader, device)
            if acc > best_score:
                best_score, best_params = acc, p
        pop += np.random.normal(0, 0.01, pop.shape)
    return best_params
```

```python
from torchvision import datasets, transforms
from torch.utils.data import DataLoader

transform = transforms.Compose([
    transforms.Grayscale(),
    transforms.Resize((128, 128)),
    transforms.ToTensor()
])

train_ds = datasets.ImageFolder("cropped_dataset/train", transform=transform)
val_ds = datasets.ImageFolder("cropped_dataset/val", transform=transform)

train_dl = DataLoader(train_ds, batch_size=32, shuffle=True)
val_dl = DataLoader(val_ds, batch_size=32)

device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')
```

✓ 5h 2m 47s completed at 03:49

Fig A.5: Code for MGO Optimization

```
    transforms.ToTensor()
])

train_ds = datasets.ImageFolder("cropped_dataset/train", transform=transform)
val_ds = datasets.ImageFolder("cropped_dataset/val", transform=transform)

train_dl = DataLoader(train_ds, batch_size=32, shuffle=True)
val_dl = DataLoader(val_ds, batch_size=32)

device = torch.device('cuda' if torch.cuda.is_available() else "cpu")
best_params = MGO(train_dl, val_dl, device)
print(f"Best hyperparams: Dropout={best_params[0]}, Filters={int(best_params[1])}, LR=

# Final Training
model = ThermalCNN(dropout=best_params[0], filters=int(best_params[1])).to(device)
optimizer = torch.optim.Adam(model.parameters(), lr=best_params[2])
loss_fn = nn.CrossEntropyLoss()

for epoch in range(10):
    for xb, yb in train_dl:
        xb, yb = xb.to(device), yb.to(device)
        optimizer.zero_grad()
        loss = loss_fn(model(xb), yb)
        loss.backward()
        optimizer.step()
    print(f"Epoch {epoch+1}/10 complete")
```

[ ] Start coding or generate with AI.

✓ 5h 2m 47s completed at 03:40

Fig A.6: Code for Final Model Training & Evaluation

APPENDIX B

CONFERENCE PRESENTATION

Our paper on **Thermal Infrared Object Recognition Using YOLOv8 and MGO based CNN for UAV Night Surveillance** was presented at ICIOT 2025 conference held at SRM. Our paper got accepted as paper id: 781

Request to Submit Copyright Form and Camera-Ready Copy by 4th April ,2025

Microsoft CMT <email@msr-cmt.org>
To: ARULALAN V <arulalav@srmist.edu.in>

Dear ARULALAN V,

We are pleased to inform you that your manuscript titled:

Paper ID: 781 –"Thermal Infrared Object Recognition Using YOLOv8 and MGO based CNN for UAV Night Surveillance"

has been accepted for publication. To proceed with the final steps, please submit the following documents via the CMT portal by 4th April ,2025:

Figure B.1: ICIOT 2025 Acceptance

On presenting the paper in this international conference held at SRM KTR campus, we received positive remarks and suggestions from the judging panel.

APPENDIX C

PLAGIARISM REPORT

turnitin Page 2 of 29 - Integrity Overview Submission ID trn:oid:::1:3243704301

5% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Match Groups

- **23** Not Cited or Quoted 5%
  Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%
  Matches that are still very similar to source material
- **0** Missing Citation 0%
  Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%
  Matches with in-text citation present, but no quotation marks

Top Sources

3%	Internet sources
2%	Publications
2%	Submitted works (Student Papers)

turnitin Page 2 of 29 - Integrity Overview Submission ID trn:oid:::1:3243704301

11	Internet	
survey.gov.lk		<1%
12	Student papers	
Dibrugarh University,Assam		<1%
13	Internet	
www.thefreelibrary.com		<1%
14	Publication	
Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dhirendra Kumar Shukla. "Intelli...		<1%
15	Publication	
Congying Liu, Lupin Chen, Muhammad Hassan Arshad. "Navigating the Digital Di...		<1%
16	Internet	
www.mdpi.com		<1%
17	Internet	
www.preprints.org		<1%

Format - I

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY (Deemed to be University u/s 3 of UGC Act, 1956)		
Office of Controller of Examinations		
REPORT FOR PLAGIARISM CHECK ON THE DISSERTATION/PROJECT REPORTS FOR UG/PG PROGRAMMES **(To be attached in the dissertation/ project report)**		
1	Name of the Candidate **(IN BLOCK LETTERS)**	DHRUV DHAR
2	Address of the Candidate	100/S-1 Shalimar garden extension 1 Sahibabad Ghaziabad Up 201005
3	Registration Number	RA2111003011108
4	Date of Birth	29/12/2003
5	Department	C. Tech
6	Faculty	Mrs. Malar Selvi
7	Title of the Dissertation/Project	Thermal Infrared Object Recognition Using YOLOv8 and MGO based CNN for UAV Night Surveillance
8	Whether the above project /dissertation is done by	~~Individual~~ or group : (Strike whichever is not applicable) Dhruv Dhar[RA2111003011108] : Yashovardhan Pandey[RA2111003011540] :
9	Name and address of the Supervisor / Guide	Dr. Arulalan V Chennai, Tamil Nadu
10	Name and address of Co-Supervisor / Co- Guide (if any)	NIL **Mail ID: Mobile Number:**

11	Software Used	YOLOv8 with MGO-CNN		
12	Date of Verification	12/5/2025		
13	**Plagiarism Details: (to attach the final report from the software)**			
Chapter	**Title of the Chapter**	**Percentage of similarity index (including self citation)**	**Percentage of similarity index (Excluding self citation)**	**% of plagiarism after excluding Quotes, Bibliography, etc.,**
1	INTRODUCTION			
2	LITERATURE SURVEY			
3	SPRINT PLANNING AND EXECUTION METHODOLOGY			
4	SYSTEM ARCHITECTURE AND MODEL WORKFLOW			
5	RESULTS AND DISCUSSIONS			
6	CONCLUSION AND FUTURE ENHANCEMENT			
7				
8				
9				
10				
Appendices				
I / We declare that the above information have been verified and found true to the best of my / our knowledge.				
Signature of the Candidate		**Name & Signature of the Staff (Who uses the plagiarism check software)**		
Name & Signature of the Supervisor/ Guide		**Name & Signature of the Co-Supervisor/Co-Guide**		
Name & Signature of the HOD				