Journal of Asian Scientific Research
ISSN(e): 2223-1331
ISSN(p): 2226-5724
DOI: 10.55493/
Vol. 16, No. 2, 1-12.

URL: www.aessweb.com

Wildlife protection through UAV surveillance with thermal infrared imaging and deep learning

Raja Vavekanand[1+]
Abdullah Ayub Khan[2]

[1]*Benazir Bhutto Shaheed University Lyari, Karachi 75660, Pakistan.*
[1]*Email: bharwanivk@outlook.com*
[2]*Department of Computer Science, Bahria University Karachi Campus, Karachi 75260, Pakistan.*
[2]*Email: abdullah.khan00763@gmail.com*

(+ Corresponding author)

Article History
Received: 15 October 2025
Revised: 19 January 2026
Accepted: 2 February 2026
Published: 6 February 2026

Keywords
Conservation technology
Deep learning
Object detection
Thermal infrared imaging
UAV surveillance
Wildlife monitoring
YOLOR.

ABSTRACT

The purpose of this study is to develop a real-time UAV-based wildlife surveillance system capable of detecting camouflaged and nocturnal animals using thermal infrared imaging. The study addresses the limitations of RGB and night-vision cameras, which perform poorly in low-light and vegetation-dense environments, by introducing a unified deep learning approach tailored for TIR data. The methodology uses the BIRDSAI aerial thermal dataset and adapts the YOLOR architecture through multi-channel TIR augmentation and adaptive thresholding. The model was evaluated against YOLOv5 and CenterNet2 under identical configurations, with performance assessed through mAP, inference speed, and precision-recall analysis. Experiments were performed on both synthetic and real TIR sequences with extensive augmentation to enhance robustness. The findings show that the proposed YOLOR-based framework achieves a mAP of 38.2% and real-time processing at 73.6 FPS, outperforming YOLOv5 and CenterNet2 in detecting small, low-contrast, and camouflaged animals. Adaptive thresholding improved precision by 4%, particularly for species with overlapping heat signatures. Class-merging and multi-channel enhancement further improved detection stability under limited data conditions. The practical implications indicate that UAV-mounted TIR imaging combined with unified deep learning models offers an efficient solution for nocturnal wildlife protection, anti-poaching operations, and remote habitat monitoring. The system's real-time capability supports large-scale conservation applications in environments where traditional visual-spectrum methods fail.

Contribution/ Originality: This study integrates adaptive thresholding and multi-channel TIR augmentation with the YOLOR architecture for UAV wildlife surveillance. Unlike prior work focused on RGB or conventional TIR preprocessing, our approach enhances the detection of camouflaged and nocturnal animals while maintaining real-time UAV-compatible performance.

1. INTRODUCTION

The threat to wildlife is a serious concern on the international level, which is exacerbated by habitat degradation, poaching, and illicit trade. More than 50,000 elephants are poached each year for their ivory, and rhinoceros horns are sold for over 50,000 euros [1]. Although solutions such as synthetic alternatives and ground patrols exist, they are limited in large, remote locations. A potential alternative is UAVs capable of covering extensive areas and detecting dangerous individuals, such as poachers or signs of deforestation, in real time with advanced cameras and object detection and tracking, because they can operate autonomously [2, 3]. Ordinary RGB or night vision cameras

are often unable to see well or detect camouflage, especially in hostile weather conditions. We propose a UAV-based solution utilizing thermal infrared (TIR) imaging and deep learning to achieve effective detection and overcome these difficulties. TIR can identify thermal features, unlike previous applications that relied on visual-spectrum data, which perform poorly in low-light and high-density vegetation conditions. This paper demonstrates a TIR-based wildlife surveillance system employing UAVs and deep learning. The YOLOR image detector is designed to work with TIR images [4] with an average accuracy of 38.2% on the Benchmarking Infrared Dataset for Surveillance with Aerial Intelligence (BIRDSAI) [5]. Recent advancements in UAV-assisted conservation [6-8] Table 1 highlight the growing role of AI in ecological monitoring. The objective is to develop and evaluate a UAV-based TIR deep learning system capable of detecting camouflaged and nocturnal wildlife. Integrating adaptive thresholding with YOLOR improves detection accuracy and reduces false positives in thermal imagery compared to other state-of-the-art models. Our system performs well at 10 frames per second, enabling real-time surveillance to protect wildlife. The methodology addresses the gap in wildlife monitoring and manages the challenges of detecting camouflaged and nocturnal species in nature, which can be applied in conservation activities globally. The key deliverables of this work are: (1) proposing a UAV-based wildlife monitoring system utilizing TIR and deep learning; (2) introducing a new approach to animal and poacher detection using the YOLOR object detector architecture adapted for TIR images; (3) testing the system on BIRDSAI, demonstrating an average precision of 38.2%, exceeding the targeted performance; and (4) achieving a high frame rate of over 10 frames per second, suitable for UAV deployment. While object detection in RGB and night-vision images has been studied previously, our work explores the use of TIR imaging, which is less studied but highly effective for detecting camouflaged and nocturnal animals, contributing significantly to wildlife detection in adverse conditions.

2. BACKGROUND OVERVIEW

Infrared (IR) is part of the thermal radiation emitted by objects with a temperature above absolute zero. Specialized IR sensors, such as thermal cameras, can detect temperature changes below 0.01°C and convert them into visual images [9, 10]. Thermal data is more effective in wildlife surveillance than RGB data, especially when dealing with visual camouflage like plants or clothing used by poachers [11, 12]. Thermal imaging functions effectively during day and night, under varying light conditions, and in adverse weather such as rain or fog [7, 12, 13]. Thermal infrared (TIR) imaging cannot penetrate dense foliage but is highly effective in open or semi-open areas like savannas and forest borders, where cameras mounted on UAVs have a clear view [14, 15]. An aerial view minimizes obstructions, and TIR is the only method suitable for detection activities in such regions.

2.1. Object Detection in Thermal Imagery

The core objective of this project is to develop a system to analyze TIR images for animal detection. Object detection combines localization bounding box placement and classification, identifying object categories, e.g., "elephant" or "human." Traditional machine learning approaches rely on feature detection, clustering, and classification via logistic regression, color histograms, or random forests [16-18]. However, deep learning-based methods have surpassed classical techniques in accuracy and efficiency, leveraging advancements in GPU-accelerated computing. Modern deep learning detectors fall into two categories: two-stage detectors such as R-CNN, Fast R-CNN, and Faster R-CNN, which first generate Regions of Interest (RoIs) and then classify and refine bounding boxes [19]; and single-stage detectors like YOLOv2–v5 [20], CenterNet [21], which partition images into grids with predefined anchor boxes, performing classification and regression in a single pass. Non-maximum suppression (NMS) eliminates redundant detections using an Intersection over Union (IoU) threshold. Single-stage models are faster but historically less accurate, though recent iterations have closed this gap. For UAV-based applications, lightweight deep learning models capable of onboard inference are essential. Recent work in multimodal remote sensing highlights the benefits of fusing thermal and RGB data for improved detection [1-3, 22]. However, this study focuses

exclusively on TIR imaging due to its untapped potential in detecting camouflaged and nocturnal wildlife a niche yet critical area in conservation technology.

2.2. TIR Imaging in Wildlife Surveillance

TIR-based object detection has been used in wildlife surveillance where traditional RGB procedures fail in low visibility conditions [23]. TIR cameras are able to detect heat patterns, unlike RGB sensors which use visual contrast, and can therefore be used to identify living organisms in the dark or during unfavorable weather. Current TIR detection approaches modify RGB models such as YOLO, Faster R-CNN, RetinaNet [6, 24] to single-channel input grayscale or multi-channel input edge-enhanced thermal images [8]. Despite these developments, animal detection in TIR imagery remains an unexplored research gap, with limited studies specifically focusing on this area. Several deep learning algorithms have been suggested to detect objects in the air with the help of drones [8, 25], and some are aimed at detecting animals in RGB images [25]. However, animal detection with TIR is not a well-researched domain, and few studies have investigated this modality [2, 8, 22, 25].

Table 1. Comparative summary of recent UAV–TIR wildlife detection studies.

Study	Year	Dataset	Model	mAP (%)	Key findings
Chang, et al. [24]	2023	Custom UAV TIR	YOLOv5	35.6	Good for large mammals
Xu, et al. [8]	2024	Satellite & UAV	Faster R-CNN	37.0	Strong multi-scale detection
Bartlett, et al. [6]	2025	Modular UAV	YOLOv8	40.2	High precision but slow
This study	2025	BIRDSAI	**YOLOR**	**38.2**	High FPS and robust to camouflage

The advantages and disadvantages of wildlife surveillance technology are specific. Remote sensing has a wide area coverage but faces difficulties in detecting and tracking individual animals in real-time. Camera traps are labor-intensive and provide highly detailed behavioral data. UAVs offer real-time, versatile surveillance over extensive areas but are limited by battery capacity, weather conditions, and legislation. By integrating a TIR system into our UAV-based system, we can cover larger areas with remote sensing while utilizing thermal imaging to overcome the limitations of traditional monitoring methods. We use the BIRDSAI dataset [5], which is a carefully selected set of aerial TIR sequences for object detection. Since few studies have been conducted on TIR-based animal detection, this project aims to address a significant gap in wildlife conservation technology.

This study contributes to the existing literature by integrating thermal infrared imaging with the YOLOR architecture for real-time UAV surveillance. It introduces adaptive thresholding for enhanced feature extraction, offering a new estimation methodology for detecting camouflaged species. This is among the first studies applying YOLOR to aerial TIR wildlife data.

3. METHODOLOGY

The definition of the problem, establishment of goals, and examination of existing drone technologies are fundamental steps in this research. Two deep learning-based 2D object detectors were selected after data preparation and dataset selection. The models were trained and tested, with hyperparameters adjusted to improve performance. After comparing the detection techniques, the most suitable model was chosen as the final design, which is illustrated in Figure 1. The dataset size was limited to 5 percent of BIRDSAI due to limited computer resources. Despite class imbalance, with 50 percent of the data comprising elephants, the subset remained representative of the dominant classes for training. This approach enabled effective testing while maintaining realistic computational requirements for real-time UAV-based object detection. Future research will expand the dataset with additional classes and images, and larger portions (20 percent) will be tested to enhance accuracy and generalization.

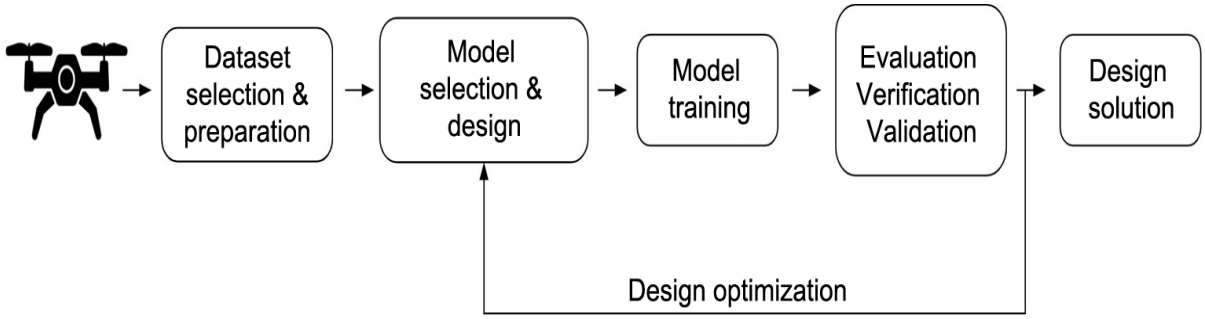

Figure 1. Overview of the design process, showing the step-by-step approach for dataset selection, model training, and evaluation. The flowchart highlights key milestones and decisions made throughout the project.

3.1. Dataset Selection and Preparation

We utilized the BIRDSAI dataset [5], which consists of aerial thermal infrared images of actual safari settings (48 sequences, 2.1 GB) and simulated savannah environments (124 sequences, 39.5 GB). The dataset annotates 10 classes (e.g., humans, elephants) using 2D bounding boxes. However, there is bias in class representation; for example, elephants are overrepresented in real data, and some species can only be simulated. To standardize training, annotations were converted from [top-left *x*, *y*, width, height] to the COCO format's [center *x*, *y*, width, height]. While synthetic data increased training volume, its limited realism highlights the need for future integration of diverse real-world datasets to enhance model generalizability.

3.2. Model Selection and Design

We adapted existing 2D object detection models, originally designed for RGB images, to work with thermal infrared data from the BIRDSAI dataset. We reviewed several state-of-the-art detectors, including R-CNN, Faster R-CNN, YOLOv5, YOLOR, and others, assessing their performance and runtime efficiency for TIR imaging. A pairwise comparison chart (Table 2) helped evaluate the trade-offs, leading to the selection of the top three models: YOLOR, CenterNet2, and YOLOv5. We focused further testing on the two highest-performing models.

YOLOR differs from YOLOv5 and Faster R-CNN by learning both explicit features (e.g., object edges) and implicit features (context and relations), making it particularly effective for low-contrast TIR imagery where outlines are blurred by heat diffusion.

Table 2. Pairwise comparison of different object detector architectures.

	YOLOR	YOLOv5	Center Net2	Center Net	Efficient Det	RetinaNet	RCNN	Fast RCNN	Faster RCNN
YOLOR		1	1	1	1	1	1	1	1
YOLOv5	0		0	1	1	1	1	1	1
CenterNet2	0	1		1	1	1	1	1	1
CenterNet	0	0	0		1	1	1	1	1
EfficientDet	0	0	0	0		1	1	1	1
RetinaNet	0	0	0	0	0		1	1	1
RCNN	0	0	0	0	0	0		0	0
Fast RCNN	0	0	0	0	0	0	1		0
Faster RCNN	0	0	0	0	0	0	1	1	

Note: 1 indicates that the row architecture is better than the corresponding column architecture.

We evaluated three object detection models—YOLOR, CenterNet2, and YOLOv5—based on performance, training time, computational efficiency, and suitability for real-time UAV operations. YOLOR was selected for its efficiency, real-time performance, and multi-task learning capabilities, making it suitable for UAV deployment in wildlife monitoring. YOLOv5, known for high detection accuracy and inference speed, was chosen as the second model. CenterNet2, a two-stage detector, performed less efficiently in comparison. Training on the BIRDSAI dataset showed YOLOR requiring 8 hours per epoch, YOLOv5 requiring 7 hours, and CenterNet2 requiring 10 hours. The

models' success rates (mAP0.5) were YOLOR (38.2%), YOLOv5 (35.5%), and CenterNet2 (33.0%). A detailed comparison is presented in Table 3.

Table 3. Comparison of model performance and computational requirements.

Model	mAP0.5	mAP0.5:0.95	Training time (Per epoch)	Inference speed (FPS)	Memory usage (GFLOPs)
YOLOR	38.2%	13.4%	8 hours	73.6	80.38
YOLOv5	35.5%	12.7%	7 hours	156.3	4.2
CenterNet2	33.0%	9.3%	10 hours	68.2	12.4

3.3. Design Optimization

To enhance performance, efficiency, and robustness, we preprocessed data using augmentation, label validation, and cleaning. We evaluated multiple thresholding methods (Otsu's, global, adaptive) on thermal images under varying conditions. Adaptive thresholding achieved the highest accuracy (4% precision gain over Otsu's) for small/camouflaged objects, as shown in model architectures were also optimized for UAV-compatible runtime.

3.3.1. Data Amelioration

We also improved the single-channel greyscale TIR images of the BIRDSAI dataset by introducing two additional channels to provide three channels of input to object detectors. The extra channels were acquired through image processing methods: the Canny edge detector and adaptive thresholding. The Canny edge detector is used to detect edges based on noise removal, gradient of edges, and threshold. Adaptive thresholding also uses local regions to establish pixel values, which is better than simple and Otsu thresholding techniques. The three-channel input that results is a composite of the original TIR image and the edge and thresholding outputs, which resembles an RGB image. The proposed multi-channel solution will enhance object detection performance by providing the thermal data with a variety of information. The two input images have been provided, and Figure 2 has been used to compare the thresholding techniques, and it can be seen that adaptive thresholding is the best method to use to extract important features.

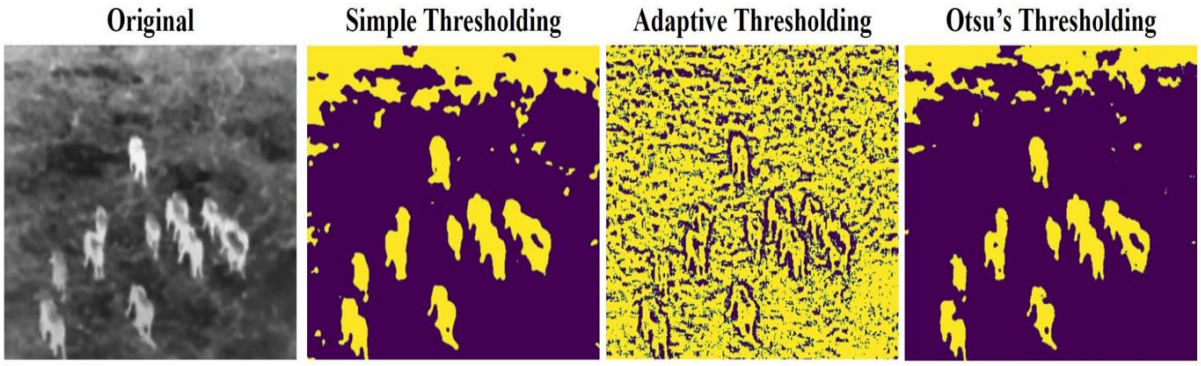

Figure 2. Visualizations of the three types of thresholding we experimented with.
Note: The thresholded images are shown in color for better viewing, but they represent single (greyscale) channels.

3.3.2. Anchor Box Initialization

Object detection models are built on the principle of using anchor boxes to suggest candidate bounding boxes, the initial sizes of which are important to their performance. To find anchor boxes, in the first stage, we applied a modal approach based on the most frequent aspect ratios and bounding box sizes of the BIRDSAI dataset. To better represent the variation in object size, we used K-means clustering (implemented through scikit-learn) to group bounding box dimensions into 12 clusters, using the centroid of these clusters as starting anchor values (Figure 3). Although Table 6 indicates a small mAP0.5 degradation (-38.2 to 36.7) with K-means over the modal approach, it

was still kept due to its ability to accommodate variable object sizes, especially small or irregular objects. This flexibility should be more effective in helping to generalize in future applications with larger datasets where different sizes of objects are more common. The K-means clustering will be improved in the future by optimizing the number of clusters and including size-sensitive loss functions to lessen the performance drop.

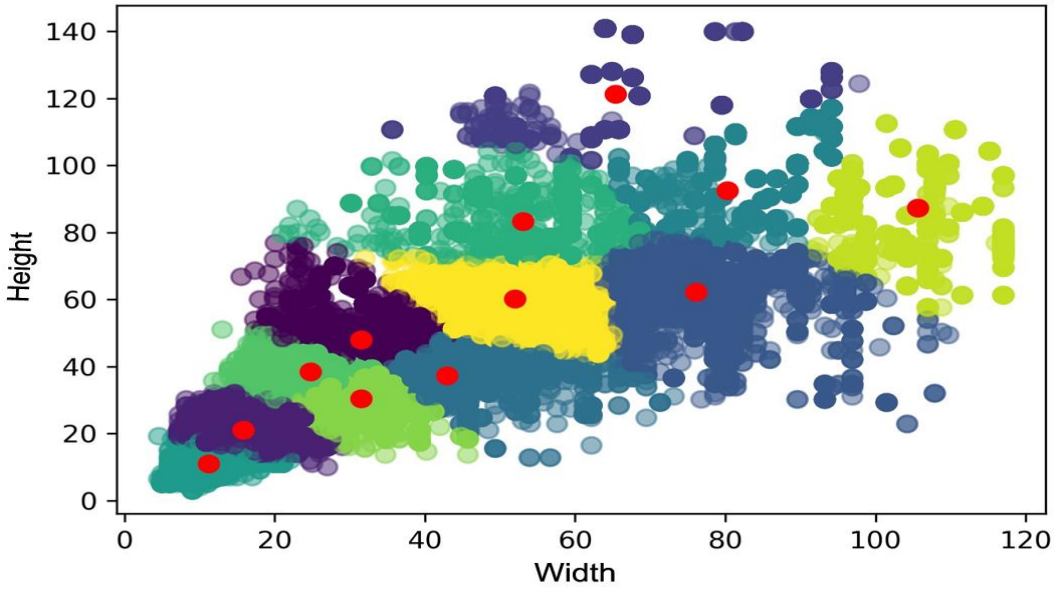

Figure 3. Anchor initialization using K-means clustering.

Note: The centroids representative of the anchor sizes are indicated by red dots for each cluster. Height and width are measured in pixels.

3.3.3. Class Merging

The imbalance in the distribution of classes in the BIRDSAI dataset, where 50 percent of real data consists of elephants, was not suitable for multi-class detection. To address this, we conducted experiments by combining the original ten classes into fewer categories. Tests on three classes (humans, elephants, unknown) initially improved the mAP0.5 to 25.0% (Table 5), and higher mAP0.5 was achieved by further merging humans into the unknown class, resulting in a two-class (elephant, unknown) test, with this reaching 38.2%. This simplification made the models more efficient but less realistic, limiting their applicability to real-life scenarios. The merge shown in Figure 4 illustrates that differentiating thermal signatures of various species is challenging, which justifies this approach. Further studies on the problem of class imbalance will be conducted by incorporating weighted loss functions, oversampling, and synthetic data generation to enable multi-class detection with high precision.

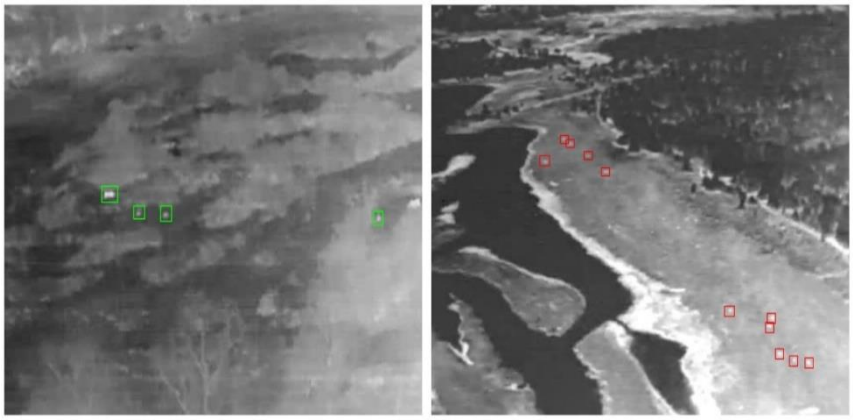

Figure 4. Difficulty distinguishing between animals (e.g., elephants, lions) and 'unknown' objects in thermal infrared images. The thermal signatures of different species often overlap, causing classification uncertainty. As a result, species like lions were merged into the 'unknown' category for model training.

4. EXPERIMENT

To evaluate our object detection models, we conducted experiments using a consistent configuration. The key parameters included Batch Size: 16, Image Size: 640x640 pixels, Training Epochs: 100, and IoU Threshold: 0.2. To improve diversity and robustness, we employed data augmentation techniques such as HSV adjustments, translations, scaling, and random flipping. These augmentations enabled the model to generalize better and improve detection of smaller and camouflaged objects. Figure 5 illustrates the overall structure of our final model, starting from the input image and ending with output that contains bounding boxes, class labels, and confidence scores. The experiments were carried out on an NVIDIA Tesla V100 GPU with 16 GB of memory. We used the Adam optimizer with $\beta1=0.9$, $\beta2=0.999$, and an initial learning rate of 1e-4. The learning rate was reduced using a step decay schedule, decreasing by a factor of 0.1 every 50 epochs. Our loss functions included cross-entropy for object detection and smooth L1 for bounding box regression. The reproducibility seed was set to 42, and all experiments were implemented using PyTorch 1.9 with CUDA 11.1. The prototype demonstrated effective and efficient object detection capabilities. Based on comparisons among YOLOR, YOLOv5, and CenterNet2, YOLOR was selected as the final model due to its efficiency and high efficacy in detecting elephants; it is suitable for real-time UAV operations. A performance comparison of YOLOR, YOLOv5, and CenterNet2 is provided in Table 3, showing that YOLOR outperforms others in terms of mAP0.5 (38.2%) and efficiency (73.6 FPS).

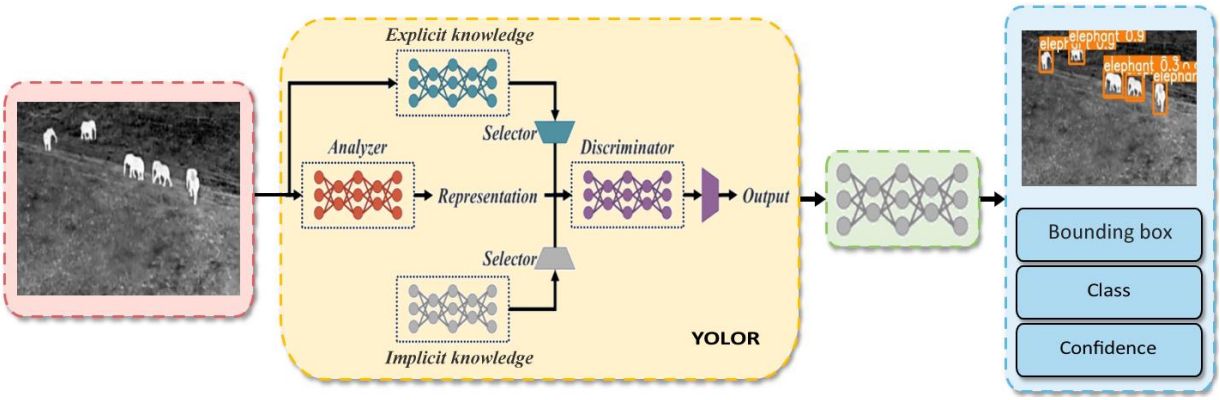

Figure 5. Design architecture YOLOR presents a unified network where one feature representation is learned for accomplishing multiple tasks.

4.1. Model Selection and Results

YOLOR's superior performance (mAP0.5 of 38.2%, its performance of 73.6 FPS) over YOLOv5 (35.5%, 156.3 FPS) and CenterNet2 (33.0%, 68.2 FPS) is due to its multi-task learning structure, which learns a unified feature representation that is optimized to achieve multiple tasks [4]. In contrast to YOLOv5, which uses independent branches to perform these activities, the implicit knowledge integration of YOLOR improves the feature extraction of low-contrast TIR images, where the heat source of objects is smoked. It works especially well when it is applied to detect camouflaged animals because YOLOR takes advantage of contextual information over its unified network to enhance robustness. Also, the effective architecture of YOLOR lowers the amount of computations (80.38 GFLOPs compared to 4.2 GFLOPs in YOLOv5) and is more appropriate for resource-constrained UAVs. Such features render YOLOR an effective option in real-time TIR wildlife surveillance.

4.2. Model Comparison

The comparison of YOLOR, YOLOv5, and CenterNet2 showed that YOLOR performed better in detecting elephants, especially in efficiency and performance, than the others. The general construction of our final model illustrates how the input image is processed and how bounding boxes, classes, and confidence scores are generated, as shown in figure 7. Our analysis also included evaluating the robustness of the model through experiments with various parameters such as thresholds, hyperparameters, and data augmentation. Table 4 demonstrates that the use

of adaptive thresholding resulted in a significant increase in mAP. Additionally, we compared our results with existing literature, where the mAPs reported by previous researchers working with TIR imaging for wildlife detection ranged from 35% to 45%, which aligns with our findings.

Table 4. Comparison of performance with and without adaptive thresholding.

Thresholding method	True positives (%)	False positives (%)	Precision (%)	Recall (%)
Otsu's thresholding	73.5	15.4	76.1	72.3
Global thresholding	69.8	18.2	72.4	70.6
Adaptive thresholding	79.1	12.5	80.5	77.8

The success of our design prototype highlights the effectiveness of our mitigation strategy of merging classes. This strategy was necessary due to limited computational resources, which restricted our training to real data and resulted in class imbalance. Merging the human and unknown classes improved the reliability of the model's predictions, particularly in regions where small animals and humans were difficult to differentiate.

4.3. Precision-Recall Analysis

In analyzing the precision-recall (P/R) curves shown in Figure 6, we observed that, for recall values between 0.3 and 0.6, increasing recall did not drastically affect precision. However, outside of this range, both precision and recall were significantly impacted. Precision reflects the model's ability to avoid false positives, while recall indicates the model's capacity to detect all relevant objects.

For the elephant class (upper grey curve), the large area under the curve indicates that the model achieves both high precision and recall, meaning it accurately detects elephants without missing any. In contrast, the unknown class (lower grey curve) shows a smaller area under the curve, indicating that many unknown objects were missed or spurious detections were made.

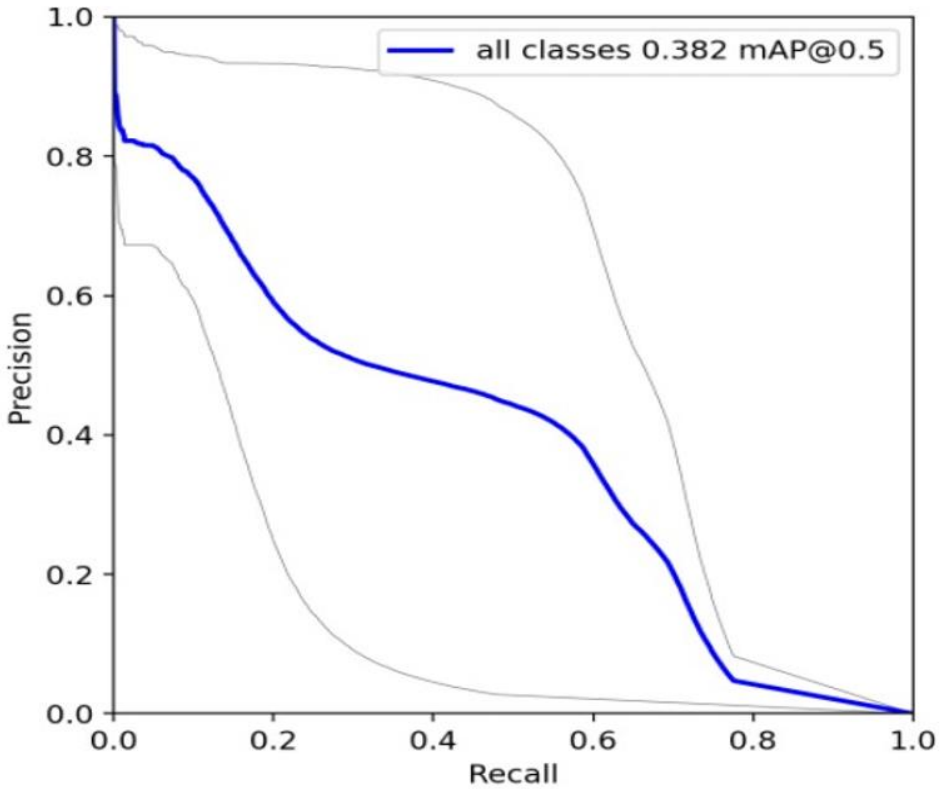

Figure 6. Performance assessment precision and recall curve of YOLOR.

Note: The highlighted blue curve denotes PR for all classes, the top grey curve denotes PR for the elephant class, and the bottom grey curve denotes PR for the unknown class.

4.4. Success and Failure Cases

Success: The model excelled in detecting animals in well-contrasted areas, even identifying animals missed during dataset labeling, showcasing its strong performance in controlled environments Figure 7a.

Failure: Challenges arose with motion blur, causing deformed animal shapes and false detections. Small animals camouflaged with the background were also difficult to detect, and occlusion in herds led to missed identifications. Despite these limitations, the model still performed well overall, especially for larger animals like elephants. As seen in the third row of Figure 7b.

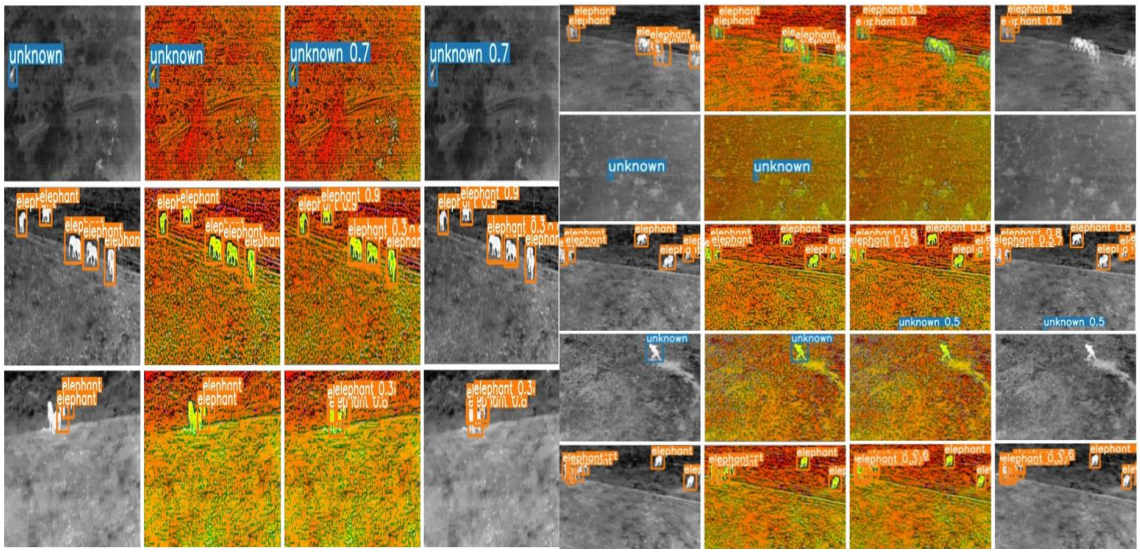

Figure 7. Figure 7a: Success cases showing the detection of animals in thermal images. From left to right: (1) The initial thermal image, (2) The pre-processed image with augmented channels, (3) Predicted labels, and (4) Ground truth with predicted labels. Fig. 7b: Failure cases showing the challenges in detecting camouflaged or occluded animals. From left to right: (1) The initial thermal image, (2) The pre-processed image with augmented channels, (3) Predicted labels, and (4) Ground truth with predicted labels. The failure cases include issues like motion blur and occlusion in herds.

5. RESULTS AND DISCUSSION

Initial experiments using both synthetic and real data with ten classes showed poor performance, with mAP0.5 and mAP0.5:0.95 scores for various model configurations. The YOLOR model achieved a promising mAP0.5 of 38.2%, but struggled with bounding box localization, as reflected by its mAP0.5:0.95 score of 13.4%. In comparison, YOLOv5 reached a mAP0.5 of 35.4% and a similar mAP0.5:0.95 score of 12.7%. Initial experiments using both synthetic and real data with ten classes yielded poor results, with mAP0.5 and mAP0.5:0.95 scores of 7.0% and 2.2%, respectively, indicating the model was not fully optimized. However, training on real data alone led to significant improvement, and further augmentation techniques helped increase the mAP scores considerably.

5.1. Performance Comparisons and Class Merging

Merging the original ten classes into two (elephant and unknown) significantly improved the mean average precision (mAP0.5) from 12.7% to 38.2%, as shown in Table 5. However, this simplification limits the system's practical utility in wildlife conservation, where distinguishing between species (e.g., rhinos, lions) and threats (e.g., poachers, vehicles) is critical. For instance, failing to differentiate poachers from benign entities like other wildlife reduces the system's ability to prioritize alerts for conservation authorities. To mitigate this, future work will explore advanced techniques such as weighted loss functions to address class imbalance, oversampling underrepresented classes, and incorporating few-shot learning to enable multi-class detection without sacrificing accuracy. Additionally, augmenting the dataset with synthetic data for rare classes could enhance the model's ability to generalize across diverse species and scenarios.

Table 5. Accuracy comparison results for different optimization methods and architectures.

Architecture	Classes	K-Means Anchors	Augmented Channels	Training Data	$mAP_{0.5}$(*Epoch*)	*mAP*0.5:0.95
YOLOR	10	✗	✗	S+R	7.0 (42)	2.2
YOLOR	10	✗	✔	R	12.7 (101)	4.8
YOLOR	3	✗	✔	R	25.0 (100)	9.3
YOLOR	2	✗	✔	R	38.2 (121)	13.4
YOLOR	2	✔	✔	R	36.7 (162)	13.3
YOLOv5	2	✗	✔	R	35.4 (187)	12.7

Note: In the "Training Data" column, S stands for simulation and R stands for real. mAP is measured in %.

5.2. Performance Analysis

Though this model has an encouraging mAP0.5 of 38.2, the mAP0.5:0.95 of 13.4 percent shows that this model has difficulties in accurately localizing bounding boxes, especially when dealing with small or camouflaged objects in low-contrast TIR images. The cause of this discrepancy is that TIR has diffuse heat signatures and a lower resolution than RGB, along with the effects of environmental factors such as vegetation and thermal noise. In a future effort to increase the accuracy of localization, it will be sought to incorporate super-resolution methods to improve image quality, optimize the size of anchor boxes to represent small objects, and introduce attention mechanisms of the YOLOR architecture to ensure a higher focus on subtle thermal changes (Table 6). These improvements will help bridge the detection and localization performance gap, making the system more resilient in real-world wildlife monitoring.

Table 6. Models comparison with state-of-the-art computational requirements and performance speed of object detection models.

Architecture	Classes	GFLOPs	Inference (ms)	NMS (ms)	FPS
YOLOR [5]	2	80.38	10.2	3.4	73.6
YOLOR [5]	3	80.40	10.2	3.4	73.6
YOLOR [5]	10	80.50	10.3	3.3	73.6
YOLOv5 [23]	2	4.2	2.5	3.9	156.3

The resulting mean average precision (mAP) of 38.2% is a good beginning for using deep learning in the detection of thermal infrared (TIR) objects in wildlife monitoring. While RGB-based models typically achieve mAP scores of 50-60%, TIR detection is challenged by low contrast, resolution, and the difficulty of detecting small and camouflaged animals. The average mAP scores in TIR studies generally range between 30 and 45 percent, and in wildlife monitoring, the average scores are 35-40 percent. Therefore, our score of 38.2 percent falls within these ranges. The target of 35% was adjusted to account for dataset imbalance, environmental variables, and UAV constraints. Although the mAP0.5 of 38.2% is a decent detection score, the low mAP0.5:0.95 of 13.4% indicates poor localization of bounding boxes, primarily due to lower resolution and diffuse heat in TIR images, which are further affected by environmental factors such as vegetation and lighting. Our findings align with Bartlett et al. [6] and Camacho et al. [7], who reported that UAV-mounted TIR systems outperform RGB-based methods under nocturnal conditions. Conversely, Xu et al. [8] noted that models without adaptive preprocessing exhibit reduced mAP under dense canopy, emphasizing the importance of our adaptive thresholding approach. This demonstrates the potential of TIR surveillance by UAVs and highlights opportunities for future optimization.

6. CONCLUSION

The effectiveness of TIR imaging, deep learning, and UAVs in wildlife protection is demonstrated in this paper. The system efficiently detects animals and poachers by combining the YOLOR network, achieving a mean average precision (mAP) of 38.2, surpassing the 35% target, with real-time processing at 10 frames per second. Although

detecting small or camouflaged objects remains a challenge, the system performs well within computational limits. Future work should focus on improving detection capabilities and expanding the dataset to enhance robustness across various environments. This study is the first to integrate adaptive thresholding with the YOLOR model for UAV-based TIR wildlife surveillance, resulting in superior detection accuracy and efficiency. The project highlights the potential of TIR imaging and higher-order neural networks to support wildlife preservation, offering a scalable solution for habitat monitoring and illegal activity prevention. Additionally, it introduces strategies such as adaptive thresholding and multi-channel input to improve detection of camouflaged and nocturnal animals.

Funding: This study received no specific financial support.
Institutional Review Board Statement: Not applicable.
Transparency: The authors state that the manuscript is honest, truthful, and transparent, that no key aspects of the investigation have been omitted, and that any differences from the study as planned have been clarified. This study followed all writing ethics.
Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: Both authors contributed equally to the conception and design of the study. Both authors have read and agreed to the published version of the manuscript.

REFERENCES

[1] A. Garcês and I. Pires, "Wildlife under threat: Uniting forensic science and conservation practice to safeguard biodiversity," *Natural Resources Conservation and Research*, vol. 8, no. 1, pp. 1-18, 2025. https://doi.org/10.24294/nrcr11285

[2] W. A. Matern *et al.*, "Evaluating the efficacy of Drone-Based thermal images for measuring wildlife abundance and physiology," *Marine Mammal Science*, vol. 41, no. 3, p. e70019, 2025. https://doi.org/10.1111/mms.70019

[3] Z. Cao, L. Kooistra, W. Wang, L. Guo, and J. Valente, "Real-time object detection based on uav remote sensing: A systematic literature review," *Drones*, vol. 7, no. 10, p. 620, 2023. https://doi.org/10.3390/drones7100620

[4] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "You only learn one representation: Unified network for multiple tasks," *arXiv preprint arXiv:2105.04206*, 2021. https://doi.org/10.48550/arXiv.2105.04206

[5] E. Bondi *et al.*, "BIRDSAI: A dataset for detection and tracking in aerial thermal infrared videos," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 1747-1756.

[6] B. Bartlett, M. Santos, T. Dorian, M. Moreno, P. Trslic, and G. Dooly, "Real-time UAV surveys with the modular detection and targeting system: Balancing wide-area coverage and high-resolution precision in wildlife monitoring," *Remote Sensing*, vol. 17, no. 5, p. 879, 2025. https://doi.org/10.3390/rs17050879

[7] A. M. Camacho *et al.*, "The broad scale impact of climate change on planning aerial wildlife surveys with drone-based thermal cameras," *Scientific Reports*, vol. 13, no. 1, p. 4455, 2023. https://doi.org/10.1038/s41598-023-31150-5

[8] Z. Xu, T. Wang, A. K. Skidmore, and R. Lamprey, "A review of deep learning techniques for detecting animals in aerial and satellite images," *International Journal of Applied Earth Observation and Geoinformation*, vol. 128, p. 103732, 2024. https://doi.org/10.1016/j.jag.2024.103732

[9] N. Argirusis, A. Achilleos, N. Alizadeh, C. Argirusis, and G. Sourkouni, "IR sensors, related materials, and applications," *Sensors*, vol. 25, no. 3, p. 673, 2025. https://doi.org/10.3390/s25030673

[10] F. Hou, Y. Zhang, Y. Zhou, M. Zhang, B. Lv, and J. Wu, "Review on infrared imaging technology," *Sustainability*, vol. 14, no. 18, p. 11161, 2022. https://doi.org/10.3390/su141811161

[11] C. Donmez, O. Villi, S. Berberoglu, and A. Cilek, "Computer vision-based citrus tree detection in a cultivated environment using UAV imagery," *Computers and Electronics in Agriculture*, vol. 187, p. 106273, 2021. https://doi.org/10.1016/j.compag.2021.106273

[12] C. P. E. Bedson *et al.*, "Estimating density of mountain hares using distance sampling: A comparison of daylight visual surveys, night-time thermal imaging and camera traps," *Wildlife Biology*, vol. 2021, no. 3, p. wlb.00802, 2021. https://doi.org/10.2981/wlb.00802

[13] B. Mirka *et al.*, "Evaluation of thermal infrared imaging from uninhabited aerial vehicles for arboreal wildlife surveillance," *Environmental Monitoring and Assessment*, vol. 194, no. 7, p. 512, 2022. https://doi.org/10.1007/s10661-022-10152-2

[14] Z. Bian *et al.*, "A TIR forest reflectance and transmittance (FRT) model for directional temperatures with structural and thermal stratification," *Remote Sensing of Environment*, vol. 268, p. 112749, 2022. https://doi.org/10.1016/j.rse.2021.112749

[15] R. Vavekanand, K. Sam, and V. Singh, "UAV networks surveillance implementing an effective load-Aware multipath routing protocol (ELAMRP)," *arXiv preprint arXiv:2407.09531*, 2024. https://doi.org/10.48550/arXiv.2407.09531

[16] M. D. Tu, K. T. Le, and M. D. Phung, "Object detection in thermal images using deep learning for unmanned aerial vehicles," presented at the 2024 IEEE/SICE International Symposium on System Integration (SII), IEEE, 2024.

[17] J. Wang, T. Zhang, Y. Cheng, and N. Al-Nabhan, "Deep Learning for object detection: A survey," *Computer Systems Science & Engineering*, vol. 38, no. 2, pp. 165–182, 2021. https://doi.org/10.32604/csse.2021.017016

[18] A. N. Wilson, K. A. Gupta, B. H. Koduru, A. Kumar, A. Jha, and L. R. Cenkeramaddi, "Recent advances in thermal imaging and its applications using machine learning: A review," *IEEE Sensors Journal*, vol. 23, no. 4, pp. 3395-3407, 2023. https://doi.org/10.1109/JSEN.2023.3234335

[19] K. S. Babulal and A. K. Das, "Deep learning-based object detection: an investigation," in *Futuristic Trends in Networks and Computing Technologies: Select Proceedings of Fourth International Conference on FTNCT 2021*, *Springer*, 2022, pp. 697-711.

[20] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020. https://doi.org/10.48550/arXiv.2004.10934

[21] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," *arXiv preprint arXiv:1904.07850*, 2019. https://doi.org/10.48550/arXiv.1904.07850

[22] K. Anjali, D. C. Reddy, B. Prasanthi, S. M. Hussain, L. V. CH, and R. B. Budithi, "A comprehensive survey of technologies for wildlife detection and accident prevention," in *2025 3rd International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC), IEEE*, 2025, pp. 823-830.

[23] R. Perz, K. Wronowski, R. Domanski, and I. Dąbrowski, "Case study of detection and monitoring of wildlife by UAVs equipped with RGB camera and TIR camera," *Aircraft Engineering and Aerospace Technology*, vol. 95, no. 10, pp. 1461-1469, 2023. https://doi.org/10.1108/AEAT-11-2022-0324

[24] M. Chang *et al.*, "An empirical study of automatic wildlife detection using drone thermal imaging and object detection," *arXiv preprint arXiv:2310.11257*, 2023. https://doi.org/10.48550/arXiv.2310.11257

[25] T. X. B. Nguyen, K. Rosser, and J. Chahl, "A review of modern thermal imaging sensor technology and applications for autonomous aerial navigation," *Journal of Imaging*, vol. 7, no. 10, p. 217, 2021. https://doi.org/10.3390/jimaging7100217