Improving Wild Pig Detection through Data Augmentation and Thermal Imagery: A Comparative Study of Model Performance

Alhim Vera **Donghoon Kim**
Department of Aerospace Engineering & Engineering Mechanics
University of Cincinnati
Cincinnati, OH 45221
veragoaa@mail.uc.edu, donghoon.kim@uc.edu

Abstract

This study presents a comparative analysis of wild pig detection using a multimodal fusion framework, integrating YOLOv8 with both RGB and thermal imagery. We investigate the impact of data augmentation techniques and GroundDINO-based auto-labeling on detection accuracy, aiming to enhance the model's robustness in diverse environmental conditions. Our results demonstrate significant improvements in precision and recall when incorporating thermal imagery and augmentation, especially in challenging environmental scenarios, such as low light and dense vegetation.

1 Introduction

Wild pigs, also known as feral hogs or wild boars, have emerged as one of the most invasive species globally, causing profound ecological disruption and significant economic damage. Their high reproductive rates, adaptability to diverse environments, and destructive behaviors, such as uprooting vegetation and preying on native species, have led to widespread biodiversity loss and ecological imbalance. Additionally, wild pigs are vectors for diseases that can infect livestock and humans, making their management a critical concern for conservationists and agricultural stakeholders [1, 2, 3].

Conventional wild pig monitoring methods, including manual tracking and camera trapping, are labor-intensive and often fail to scale effectively. The need for real-time, accurate, and scalable monitoring solutions has driven research into machine learning and computer vision techniques. convolutional neural networks, in particular, have been instrumental in advancing automated animal detection systems, enhancing the efficiency of wildlife monitoring and management strategies [4, 5]. The integration of RGB and thermal imagery has shown promise for improving detection rates, especially in low-light or occluded environments where RGB imagery alone may fail [6, 7].

However, current approaches often lack specificity in detecting wild pigs across diverse habitats. Given the variability in their behavior and environmental conditions, there is a growing need for a more tailored solution. This research seeks to bridge this gap by optimizing multimodal computer vision models that fuse RGB and thermal data for more accurate wild pig detection.

In this work, we employ YOLOv8, a cutting-edge object detection model renowned for its speed and accuracy, and evaluate its performance with and without thermal imagery and data augmentation. Data augmentation techniques, combined with advanced auto-labeling tools like Grounding DINO, are leveraged to improve model robustness and accuracy [8, 9]. The central aim is to assess how the integration of thermal imagery and data augmentation can enhance detection precision, recall, and overall performance in real-time wildlife monitoring scenarios.

Latinx in AI @ NeurIPS-24 (LXAI 2024).

The contributions of this paper are threefold: (1) we present a robust multimodal detection framework optimized for wild pig detection; (2) we evaluate the efficacy of data augmentation and auto-labeling in improving detection performance; and (3) we demonstrate the practical applicability of these techniques in real-world settings, contributing to advancements in ecological AI and multimodal fusion systems.

2 Methodology

In this study, we compare the impact of incorporating thermal imagery and data augmentation techniques into the wild pig detection process. The methodology illustrated in Figure 1, begins with data acquisition of RGB and thermal images, followed by comprehensive preprocessing steps, including noise reduction and normalization. Next, we apply data augmentation techniques to further enhance model generalization. Finally, we train two versions of YOLOv8—one using only RGB images and another using both RGB and thermal images and compare their performance using precision, recall, and F1-score metrics.

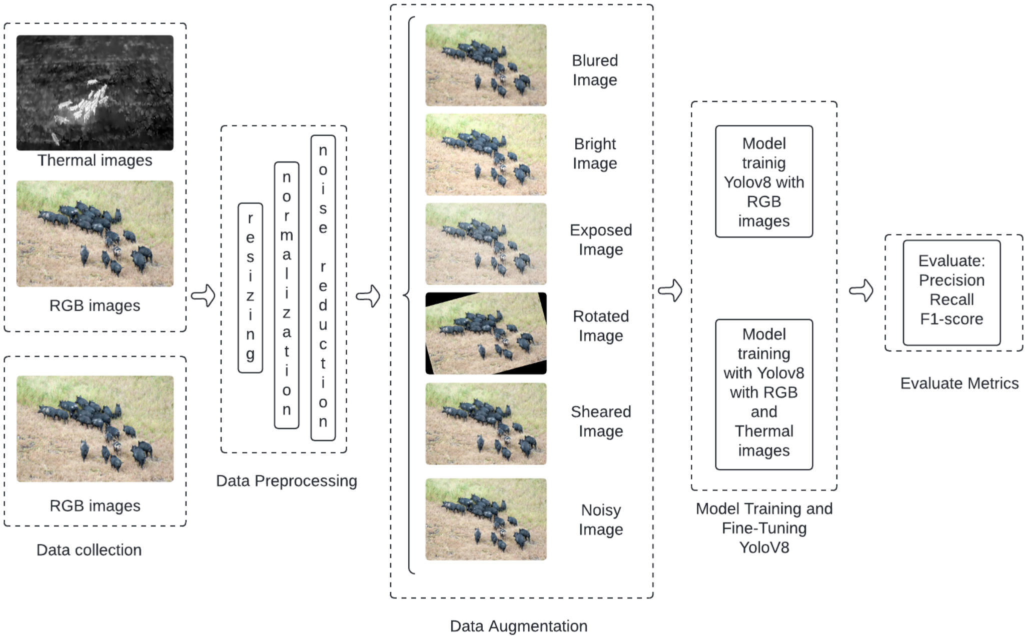

Figure 1: Overview of the full research methodology for multimodal wild pig detection, from data acquisition to model training and analysis.

2.1 Data Collection

The dataset is composed of 6,580 RGB images and an additional 2,819 thermal images, totaling 9,399 multimodal images. The RGB images were sourced from multiple wildlife datasets, including the 'Labeled Wildlife Dataset' [10, 11], and others, providing a diverse range of wild pig scenarios. For the thermal imagery, we manually conducted web scraping across various online wildlife monitoring sources. The thermal images were extracted from publicly available videos and processed frame-by-frame to obtain the necessary thermal data for analysis. This approach focused on environments with low visibility and challenging conditions. This expanded dataset provided critical insights, as the combination of RGB and thermal modalities captures a broader range of environmental variations, particularly during nocturnal or low-visibility monitoring.

To efficiently label this additional dataset, we used the Grounding DINO model [12]. Grounding DINO operates by linking natural language descriptions with image regions, allowing for precise auto-labeling of the objects within the thermal and RGB data. The architecture employs a transformer module for context analysis and a grounding module that aligns textual descriptions with image features.

During the labeling process, we tested multiple textual prompts, including “wild-pig,” “pigs,” “feral hogs,” and “boars,” to determine which description provided the most reliable results. After thorough evaluation, we found that the prompt “wild-pig” consistently yielded the best detection accuracy and alignment with our dataset. This strategic prompt selection, combined with Grounding DINO’s robust multi-modal capabilities, significantly reduced manual annotation efforts by automatically labeling 2,819 thermal images with high precision and consistent quality.

Following the automated labeling, we conducted a manual review process where human annotators verified the labeled images to ensure correctness and consistency. This combination of automated labeling using Grounding DINO and subsequent human verification ensured the dataset was both comprehensive and accurate, enhancing its reliability for model training.

2.2 Data Processing and Data Augmentation

To prepare the dataset for model training, images were resized to 640×640 pixels for consistency, normalized to a 0–1 range for standardized inputs, and noise was reduced using a Gaussian blur filter to enhance feature preservation, particularly in thermal images, improving model accuracy.

To enhance model generalization, various data augmentation techniques were applied, including horizontal flips, random rotations ($\pm 15^\circ$), and random cropping (0–20%) to introduce orientation and scale variability. Brightness, hue, and saturation were adjusted within $\pm 15\%$ and $\pm 25\%$ ranges to simulate different lighting conditions, while Gaussian noise was added to mimic environmental noise. These augmentations created a more diverse and robust dataset, increasing its size to 44,236 images and significantly improving the model’s adaptability to real-world scenarios. Figure 2 illustrates these transformations on a sample image, showcasing the achieved visual diversity.

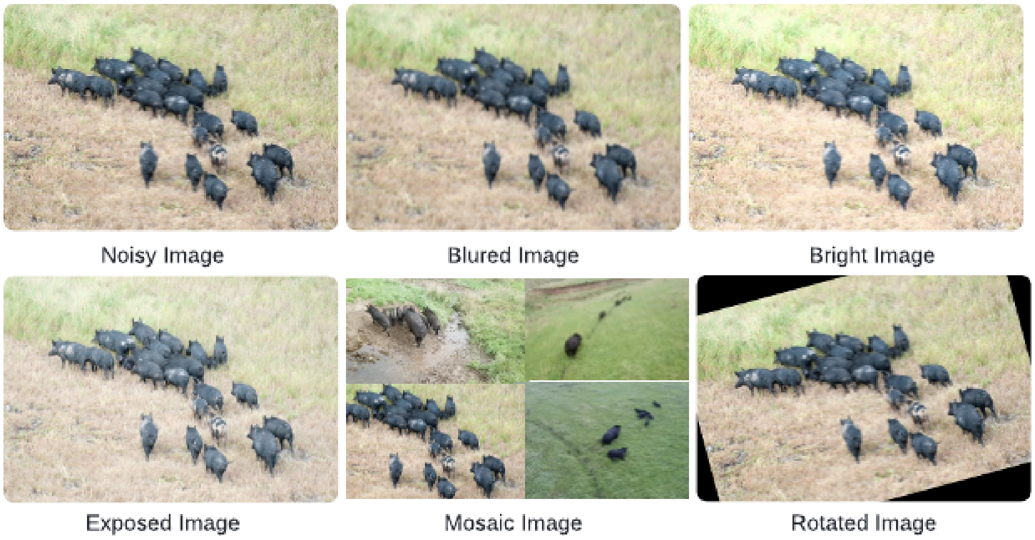

Figure 2: Examples of data augmentations applied to wild pig images: noisy image, blurred image, bright image, exposed image, mosaic image, and rotated image. These augmentations enhance model robustness by simulating various real-world conditions..

2.3 Model Training and Fine-Tuning

YOLOv8 was chosen for its balance between speed and accuracy, making it well-suited for both RGB-only and multimodal RGB-Thermal object detection tasks. To achieve optimal model performance, we meticulously adjusted the training hyperparameters. Below is an in-depth explanation of these hyperparameters, along with their corresponding formulas and justifications.

2.3.1 Optimizer and Learning Rate

The Adam optimizer was employed for its adaptive learning rate properties, making it suitable for complex models like YOLOv8. The update rule for Adam is defined as 1:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t, \tag{1}$$

where θ_t is the model parameter at time step t, η is the learning rate, $\hat{m}_t$ is the bias-corrected first moment estimate, $\hat{v}_t$ is the bias-corrected second-moment estimate, and ϵ is a small constant for numerical stability.

The initial learning rate (LR_{init}) was set to 0.01 based on recommendations for training stability and convergence. The following cosine decay schedule is used to decrease the learning rate over time:

$$LR(t) = \eta_{\text{min}} + \frac{1}{2}(LR_{\text{init}} - \eta_{\text{min}}) \left[1 + \cos\left(\frac{\pi t}{T}\right)\right], \tag{2}$$

where $\eta_{\text{min}} = 0.0001$ is chosen to ensure gradual convergence at later training stages. The parameter T represents the total number of training steps or epochs over which the learning rate schedule is applied, controlling the rate at which the learning rate decays from the initial value LR_{init} to η_{min}.

2.3.2 Momentum

The momentum parameter in Adam, represented by β_1, was set to 0.937. It is a hyperparameter that controls the decay rate of the exponentially weighted moving average of past gradients, balancing recent and historical gradient information to smooth updates and accelerate convergence. Specifically, the first-moment estimate m_t is computed as:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \tag{3}$$

where g_t is the gradient at time step t. For stability, the initial momentum was set to 0.8, transitioning smoothly to 0.937 after the warm-up phase. Notably, Adam includes a bias correction step to produce $\hat{m}_t$, which adjusts m_t by compensating for the initial bias from exponential averaging. This corrected value, $\hat{m}_t$, is applied in the update rule, ensuring unbiased parameter updates during early training stages, thereby enhancing model performance and stability.

2.3.3 Weight Decay

To prevent overfitting and promote model generalization, a weight decay hyperparameter with $\lambda = 0.0005$ was applied via an L2 regularization term added to the model's loss function. This regularization term penalizes large weight values, encouraging the model to favor simpler, more generalizable solutions. The total loss function incorporating weight decay is expressed as:

$$L_{\text{reg}} = \lambda \sum_{i=1}^{N} \theta_i^2. \tag{4}$$

2.3.4 Loss Functions and Gains

To optimize the model's learning across key aspects of object detection, several gain values were applied to specific loss components within the YOLOv8 configuration, including bounding box accuracy, classification balance, and keypoint precision.

A box loss gain of 7.5 was applied to improve bounding box accuracy, emphasizing spatial precision in object localization, which is crucial in complex detection scenarios. To balance the impact of classification with the spatial components, the classification loss gain was set to 0.5, ensuring strong classification performance without detracting from localization accuracy. For further refinement of bounding box predictions, a distribution focal loss gain of 1.5 was introduced, enhancing robustness by enabling the model to handle challenging examples and maintain confidence across diverse object scales and appearances. Lastly, a higher pose loss gain of 12.0 was incorporated to prioritize precise keypoint detection, particularly beneficial for RGB-Thermal data fusion tasks, where accurate localization of keypoints is essential for multimodal data applications.

2.3.5 Validation Settings

The dataset was split into training, validation, and test sets with a 70/15/15 distribution to ensure comprehensive model evaluation and hyperparameter tuning.

All hyperparameter adjustments were made based on recommended best practices for improving YOLO model training [13]. This included selecting pre-trained weights from the COCO dataset [14] to leverage transfer learning for better initial performance.

3 Results

This section presents the comparative performance of the two models: one trained using RGB images only and the other incorporating both RGB and thermal images. The evaluation metrics used include precision, recall, F1-score, confusion matrices, and various performance curves that highlight the effectiveness of each approach.

3.1 Precision-Recall Curve

The Precision-Recall (PR) curve, depicted in Figure 3, demonstrates the balance between precision and recall for both models. The model trained with RGB-Thermal data consistently outperforms the RGB-only model, achieving a precision of 0.882 compared to 0.797 for the RGB-only model. This improvement is particularly notable in low-contrast environments, such as nighttime or dense foliage, where thermal imagery provides clearer contrast between wild pigs and their surroundings.

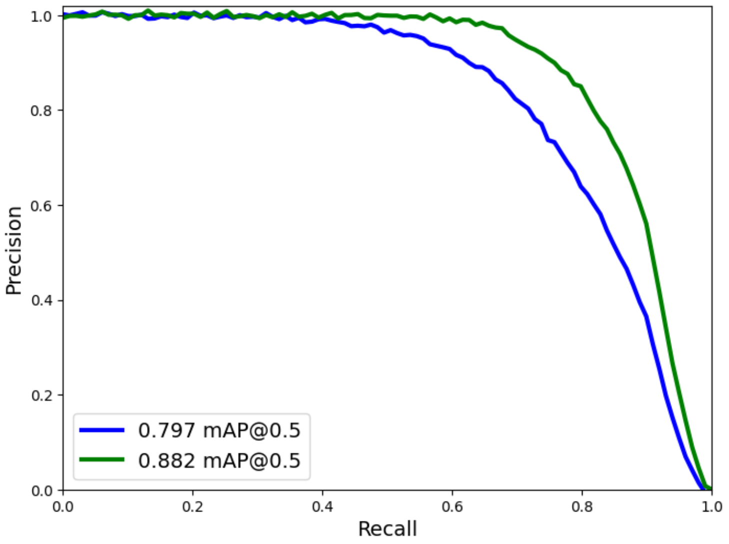

Figure 3: PR curves for RGB-Thermal model (green) and RGB-only model (blue).

The thermal data enables the model to detect objects that are difficult to distinguish based solely on RGB features, such as pigs in shadowed areas or against backgrounds of similar color. The PR curve shows that the multimodal model maintains a higher precision across various recall levels, indicating a reduced false positive rate while preserving detection sensitivity. This leads to more reliable wild pig detection, especially in challenging conditions that often confound RGB-based models. By achieving higher precision without sacrificing recall, the RGB-Thermal model delivers more consistent and reliable results, which is crucial for tasks like wildlife monitoring, where accuracy and robustness are paramount

3.2 F1-Score Curve

Figure 4 illustrates the F1-score curves for both the RGB-only and RGB-Thermal models, providing a single metric that balances precision and recall. The F1-score is crucial in scenarios where both false positives (incorrectly detecting a wild pig) and false negatives (missing a wild pig detection) need to be minimized.

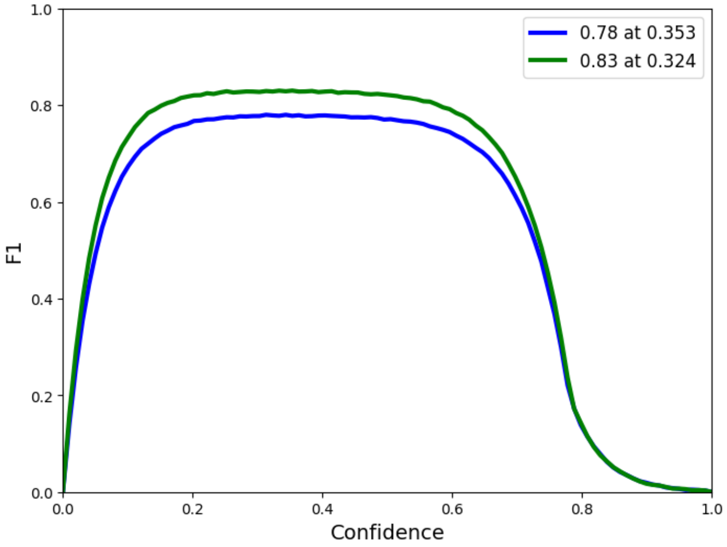

Figure 4: F1-score curves for RGB-Thermal model (green) and RGB-only model (blue).

As observed from the curve, the RGB-Thermal model consistently outperforms the RGB-only model across a wide range of confidence thresholds. Specifically, the RGB-Thermal model achieves a peak F1-score of **0.83** at a confidence threshold of **0.324**, whereas the RGB-only model attains a lower peak F1-score of **0.78** at a confidence threshold of **0.353**.

This comparison highlights several key points:

- **Higher F1-Score**: The RGB-Thermal model's higher F1-score indicates a more effective balance between precision and recall, showcasing its ability to detect wild pigs more accurately and reliably. This reflects the advantages of incorporating thermal data, which enhances performance, particularly in poor visibility environments (e.g., low light or dense vegetation).
- **Lower Confidence Threshold**: The RGB-Thermal model achieves its best performance at a lower confidence threshold (0.324 compared to 0.353 for the RGB-only model). This suggests that the addition of thermal data allows the model to make more confident predictions at lower thresholds, increasing overall sensitivity while maintaining precision.
- **Improved Performance**: The RGB-Thermal model benefits from the fusion of RGB and thermal data, resulting in better detection capabilities in challenging scenarios. Thermal imagery complements RGB data by improving object differentiation, especially when visual cues alone may not be suffice.

Overall, the addition of thermal data has a clear positive impact, reflected in both the higher peak F1-score and the model's ability to operate effectively at a lower confidence threshold. This underscores the value of multimodal data fusion in improving the accuracy and robustness of object detection models, particularly in wild pig detection.

3.3 Confusion Matrix

The confusion matrices for both models Figures 5 further illustrate the improvement in detection accuracy between the RGB-only and the RGB-Thermal models. These matrices were normalized to provide a clearer comparison between classes: wild pig and background.

For the RGB-only model, the wild pig detection accuracy is 81%, with 19% of wild pigs misclassified as background. Background detection, however, is perfect, achieving 100% accuracy. This

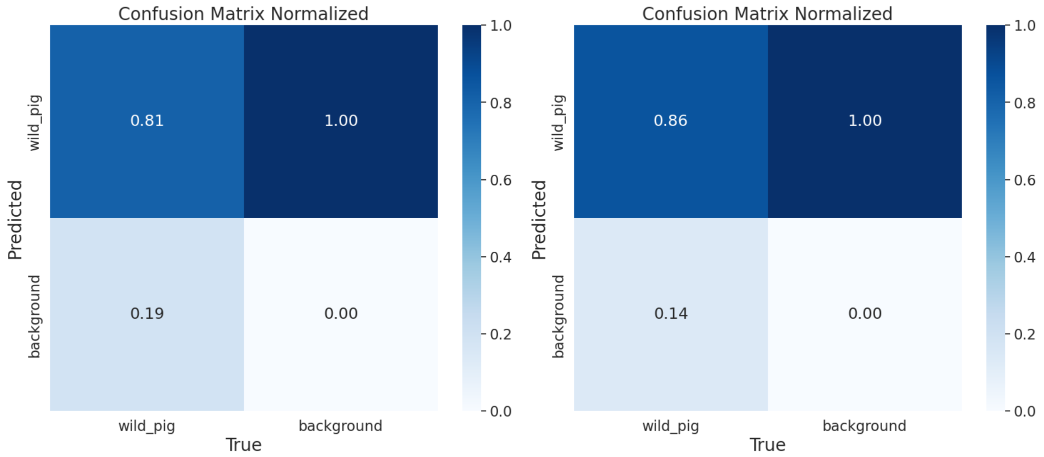

Figure 5: Confusion matrices for RGB-only model (left) and RGB-Thermal model (right).

implies that the model effectively distinguishes between the background and wild pigs in simpler environmental conditions but struggles to detect wild pigs in more complex scenarios.

In contrast, the RGB-Thermal model demonstrates improved performance. Wild pig detection accuracy increases to 86%, reducing the misclassified wild pigs to 14%. Similar to the RGB-only model, background detection remains flawless at 100% accuracy. This indicates that thermal imagery enhances the model's performance, especially in challenging environments like low light or dense vegetation, where heat signatures help distinguish wild pigs from the background.

The improvement in wild pig detection accuracy from the RGB-only to the RGB-Thermal model highlights the significance of incorporating thermal data to reduce false negatives, ensuring a more robust wildlife monitoring system.

3.4 Summary of Results

The inclusion of thermal imagery, along with data augmentation techniques, has proven to be a pivotal factor in enhancing the model's performance, particularly under challenging environmental conditions such as low visibility and dense vegetation. The RGB-Thermal model shows consistent improvements over the RGB-only model across multiple performance metrics.

As shown in Table 1, the precision, recall, and F1-score for the RGB-Thermal model demonstrate significant improvements. The precision increased from 0.82 to 0.88, indicating fewer false positives, while recall improved from 0.90 to 0.96, showcasing the model's enhanced ability to detect wild pigs accurately. This improvement is also reflected in the F1-score, with the RGB-Thermal model achieving 0.83 compared to 0.78 for the RGB-only model.

Table 1: Comparison of Performance Metrics Between RGB-only and RGB-Thermal Models

Metrics	RGB-only Model	RGB-Thermal Model
Precision	0.82	0.88
Recall	0.90	0.96
F1-Score	0.78	0.83

Moreover, the final dataset, after augmentation and the inclusion of thermal imagery, comprised 44,236 images. This larger and more diverse dataset provided the model with a robust foundation for detecting wild pigs across a variety of environmental conditions, ensuring improved generalization to real-world scenarios.

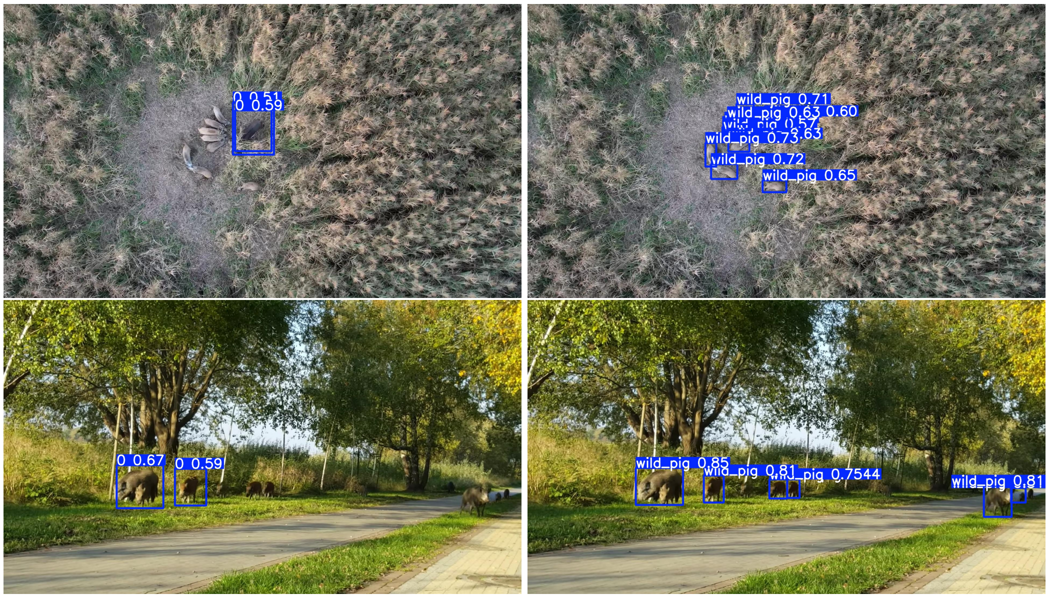

Figure 6: Comparison of detection results for RGB-only (left column) and RGB-Thermal models (right column) across different scenarios. The top row illustrates detection performance in challenging perspective angles, while the bottom row shows detection in more conventional settings.

3.5 Visual Results: RGB vs. RGB-Thermal Detection

The visual comparison between the RGB-only and the RGB-Thermal models highlights the improvement in detection accuracy across diverse environmental conditions. As demonstrated in Figures 6, the RGB-Thermal model significantly enhances the visibility of wild pigs, particularly in scenarios where lighting is poor or the contrast between the pigs and the background is minimal.

In these images, the RGB-only model struggles to differentiate wild pigs from their surroundings, particularly in shaded areas or at a distance, while the RGB-Thermal model detects the pigs with higher confidence and provides clearer boundaries between the animals and their environment. This is especially apparent in dense foliage and low-contrast conditions where thermal imagery helps reveal heat signatures that the RGB sensor aloe might miss. The comparison highlights the robustness of multimodal approaches, improving precision and recall across various environmental challenges and supporting the practical application of these models in real-time wildlife monitoring, especially in scenarios where traditional RGB detection methods falter due to environmental limitations.

4 Conclusion

This research highlights the significant benefits of integrating RGB and thermal imagery for wild pig detection using the YOLOv8 architecture. The multimodal fusion approach consistently outperformed the RGB-only model, particularly in low-light and challenging environments, as evidenced by improvements in precision, recall, and F1-score. By leveraging data augmentation and auto-labeling techniques, the final model demonstrated robust detection capabilities across diverse scenarios, making it a valuable tool for wildlife monitoring. Future work will focus on optimizing this system for deployment on edge devices, ensuring real-time detection in field applications, and exploring its potential for monitor ing other species in complex environments.

References

[1] J. L. Corn and M. J. Yabsley. *Diseases and Parasites That Impact Wild Pigs and Species They Contact*, pages 83–126. CRC Press, 2020.

[2] S. S. Ditchkoff, J. C. Beasley, and J. J. Mayer. *The Future of Wild Pigs in North America*, pages 465–469. CRC Press, 2020.

[3] J. J. Mayer and I. L. Brisbin. *Wild Pig Reproductive Biology*, pages 51–69. CRC Press, 2020.

[4] Y. Zhao and J. Linhoss. Practices and applications of convolutional neural network-based computer vision systems in animal farming: A review. *Sensors*, 21(4):1492–1505, 2021.

[5] John Linhoss and Yang Zhao. Enhanced camera-based individual pig detection and tracking for smart farming using cnns and machine vision. *Applied Sciences*, 13(12):6997–7012, 2023.

[6] A. Brown, B. Williams, and J. Smith. Integrating rgb and thermal imagery for improved wildlife detection. *Journal of Wildlife Management*, 85(3):456–465, 2021.

[7] B. Williams, J. Smith, and H. Zhang. Challenges and advancements in wildlife detection using multi-modal imagery. *Ecological Informatics*, 62:101287, 2022.

[8] T. Wu and Y. Dong. Yolo-se: Improved yolov8 for remote sensing object detection and recognition. *Applied Sciences*, 13(24):12977, 2023.

[9] Ghazala Ahmar and Shariq Siddiqui. Enhanced thermal-rgb fusion for robust object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 123–132, 2023.

[10] Images.cv. Wild Boar Image Classification Dataset, 2024. Accessed: 2024-09-20.

[11] myyyyw. Ntlnp: Night-time light dataset for navigation and perception, 2024. Accessed: 2024-09-20.

[12] Shilong Liu et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In *ECCV 2024*, 2024. `https://arxiv.org/abs/2303.05499`.

[13] Glenn Jocher, Jing Qiu, and Ayush Chaurasia. Ultralytics YOLO, January 2023.

[14] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. *arXiv preprint arXiv:1405.0312*, 2014.