Challenges in Infrared Small-Target Detection: A Benchmark of YOLO Models on UAV and Bird Infrared Imagery

Swapnil Shahapurkar[1], Pratiksha Angadi[1], Chandni Kumari[1], Prashant Uppar[1], Prema T. Akkasaligar[1]
[1] *Department of Computer Science and Engineering, KLE Tech, University's Dr. MSSCET, Belagavi, India*
premaakkasaligar.mss@kletech.ac.in

2025 10th International Conference on Signal Processing and Communication (ICSC) | 979-8-3315-1189-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICSC64553.2025.10968859

***Abstract*—Detecting small objects in infrared (IR) images, such as birds and Unmanned Aerial Vehicles (UAVs), presents significant challenges due to their reduced size, limited pixel representation, and lack of distinct features. These challenges are further complicated in real-world environments where complex backgrounds, such as skies, clouds, and vegetation, combined with low signal-to-noise ratio (SNR) and low signal-to-clutter ratio (SCR), obscure the targets. Existing detection models often struggle to accurately detect small targets due to low contrast and indistinct edges. This highlights the need for detection techniques tailored to small object detection in infrared imagery. In this study, we evaluate the performance of various You Only Look Once (YOLO) models, focusing on their applicability and limitations for detecting small infrared objects. The novelty of this work lies in systematically analyzing the impact of preprocessing techniques and fine-tuning on detection performance. Fine-tuned YOLO models are evaluated for small target detection, with their strengths and weaknesses identified under diverse conditions. Insights are provided for adapting these models to enhance real-time surveillance systems in critical applications.**

Key Words: Infrared imagery, small target detection, real time processing, UAV, YOLO Model

I. Introduction

Infrared dim small target detection is vital for applications like UAV and bird surveillance, missile guidance, and maritime rescue. Infrared imagery is effective in low visibility conditions, but poses challenges due to dim targets with minimal pixel representation and complex backgrounds. This paper focuses on improving YOLO based models for improved detection performance in these scenarios. Accurate detection systems are critical for airspace monitoring and small object tracking in rescue and defense. However, current methods often fail to meet mean average precision requirements, especially for low contrast and small targets, emphasizing the need for innovative solutions. The modern YOLO models such as YOLOv3, YOLOv5, and YOLOv10 offer real time efficiency. Their application to infrared dim small target detection is limited by challenges like low signal-to-noise ratio (SNR), cluttered backgrounds, and dataset scarcity. This study focuses on analyzing how different object detection algorithms, particularly YOLO based models, perform under various data preprocessing techniques and evaluates their limitations in infrared small object detection. By fine tuning pre trained YOLO models with strategies like data augmentation, transfer learning, and hyperparameter tuning, the study examines the impact of preprocessing methods on detection performance. The novelty lies in identifying algorithmic strengths and weaknesses to provide insights into adapting models for challenging conditions, advancing real time surveillance systems and other critical applications.

II. Related Works

Advancements in object detection using thermal and infrared imaging have enabled applications such as traffic monitoring and UAV surveillance. YOLO frameworks and deep learning models are widely used for detecting objects like pedestrians, vehicles, UAVs, and animals. This study evaluates YOLO models for detecting small infrared targets such as UAVs and birds, addressing a key research gap in complex environments, especially at night and in privacy-sensitive contexts.

In [1], Leihao Zhao integrated multi-rotor UAVs with YOLOv5 for infrared-based object detection in search and rescue operations, achieving 99% mAP. However, real-world complexity and high resource demands remain challenges. In [2], Rahman *et. al.*, reviewed ML and DL methods for UAV detection, achieving over 99% accuracy, but noted the reliance on large labeled datasets and the need for solutions addressing dataset scarcity. In [3], Seidaliyeva *et. al.*, explored UAV detection using radar, RF, and vision systems, emphasizing sensor fusion for improved accuracy, while recognizing weather variability as a major limitation. In [4], Antonenko *et. al.*, proposed an intelligent UAV-based detection algorithm with high precision, yet highlighted the significant resources required, advocating for more deployable solutions. In [5], Yafoz *et. al.*, discussed UAV applications in disaster management and military operations, noting their high-resolution imagery capabilities, but acknowledged regulatory and environmental deployment challenges. In [6], Brassai *et. al.*, combined thermal and RGB imagery to improve detection accuracy in low light, but faced challenges with dataset preparation and performance in noisy environments. In [7], Zhang *et al.*, introduced a YOLOv5-based one-stage detector with segmentation for limited data challenges in remote sensing imagery, but overfitting risks from transfer learning remain a concern. In [8], Bing Feng *et. al.*, developed a YOLOv5 model for classifying substation equipment in

979-8-3315-1189-0/25/$31.00 ©2025 IEEE

infrared images, achieving high recognition rates despite challenges in low contrast and complex backgrounds. In [9], Yusuf Furkan Yücesoy *et. al*., evaluated YOLOv8 and YOLOv9 for detecting tanks in infrared across different spectra, noting that YOLOv8 outperformed YOLOv9 but faced challenges with low-resolution data. In [10], Maham Misbah *et. al*., enhanced YOLOv5 with MSF-GhostNet for detecting small infrared objects like drones, improving precision but struggling with complex backgrounds and low SNR.

The literature survey highlights advances in object detection, particularly in thermal and infrared imaging, while underscoring the significant challenge of detecting dim small targets in complex backgrounds. YOLO based algorithms excel in general object detection but remain underexplored for small infrared targets such as UAVs and birds. This study systematically evaluates YOLO models on key metrics for detecting these targets, providing information to guide practical implementations and addressing a critical gap in this specialized domain.

III. Proposed Methodology

Detecting small, dim objects such as UAVs and birds in infrared imagery is a critical challenge due to the complex characteristics of these targets and their environments. This project utilizes the SIRST-UAVB dataset, which consists of 640×512 mid wave infrared images captured in the 3–5 μm wavelength range at distances between 100 and 800 meters. With 94.3% of the data comprising small targets, including UAVs and birds, the dataset presents these objects against diverse and cluttered backgrounds such as skies, clouds, buildings, and vegetation. Low signal-to-noise ratio (SNR), low signal-to-clutter ratio (SCR), and significant background noise further complicate detection, as these features obscure the targets and reduce visibility. This study focuses on training YOLO models, including YOLOv3, YOLOv5, and YOLOv10, on the annotated samples within the dataset. The objective is to assess their ability to detect small, streamlined targets under challenging infrared conditions. Initial experiments revealed limited improvements across YOLO versions, highlighting the inherent difficulty of detecting small objects in IR imagery and emphasizing the need for advanced detection techniques tailored to this domain. The main objectives of this work are

- **Benchmark YOLO Algorithms on Infrared Imagery:** To conduct a comparative analysis of YOLOv3, YOLOv5, and YOLOv10 (Nano and Small) to evaluate their performance in detecting UAVs and birds in UAV borne infrared imagery, focusing on metrics such as mean average precision, precision, recall, and computational efficiency.
- **Analyze Detection Challenges in Small Object Domains:** To investigate the limitations and challenges of YOLO based algorithms in detecting dim and small infrared targets amidst complex backgrounds, identifying key factors that impact detection performance.
- **Identify YOLO Models Strengths and Weaknesses:**To examine the strengths and weaknesses of older and newer YOLO models in detecting UAVs and birds, particularly in challenging scenarios such as dim lighting, small target size, and cluttered backgrounds, to inform future research directions.

YOLO Architecture: The You Only Look Once (YOLO) architecture is a popular object detection framework known for its real time speed and high mean average precision. Unlike traditional object detection methods that use region based networks or sliding windows, YOLO treats object detection as a single regression problem, predicting bounding boxes and class probabilities directly from an input image in a single forward pass through the network.

The workflow for proposed method is shown in Fig.1 The workflow startes with collecting and preparing infrared images to detect UAVs and birds in complex backgrounds.The techniques like data augmentation and resizing to make the dataset more diverse and compatible with YOLO models are used. Each model is trained to identify and classify UAVs and birds. After training, the models using metrics like losses, precision, recall, and mean average precision (mAP) are compared. The research shows how these models can be used in important areas like improving security in military applications and monitoring wildlife. This workflow helped to build a reliable detection system and highlighted its importance in real world applications.

TABLE I: YOLO Models: Architecture Details

YOLO Model	Input Size	FLOPs	Parameters	Layers
YOLOv3	416x416	65 G	62737160	106
YOLOv5	640x640	24 G	9122996	262
YOLOv10 Nano	640x640	8.4 G	2707820	385
YOLOv10 Small	640x640	24.8 G	8067900	402

The YOLO models architecture details are shown in TABLE I that describes, input size represents the image resolution used in each model processes, influencing its ability to detect small targets like UAVs and birds. Higher resolutions retain more detail but demand greater computational resources. FLOPs measures the computational complexity, reflecting the operations required per inference. Lower FLOPs indicate faster processing, essential for real time applications. Parameters denote the total number of trainable weights in each model. More parameters enable complex feature learning but may lead to higher computational requirements. Layers indicates the depth of the network, affecting its capacity to extract features at various levels of abstraction, critical for detecting small infrared targets. From table II it is observed that different versions of YOLO are tested and adjusted settings like the number of training rounds (epochs) and batch size to improve their performance.

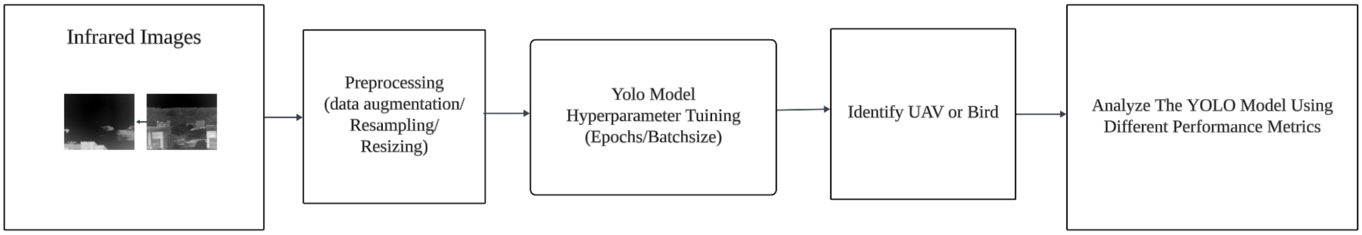

Fig. 1: Proposed Methodology of Small Object Detection in Infrared Images

IV. RESULTS AND DISCUSSION

The implementation is carried out using Google Colab with a Tesla T4 GPU for model training. The dataset, consisting of infrared images annotated with UAV and bird labels, is prepared using Roboflow for tasks like resizing and export in YOLO format. It is then imported into Colab from Google Drive for seamless integration. For training, YOLOv3 is directly implemented in Google Colab, while higher versions like YOLOv5, YOLOv10 Nano, and YOLOv10 Small utilized the structured dataset created with Roboflow. Pre trained models for these versions are downloaded from YOLO's official ultralytics repository and used for transfer learning to speed up the training process. The workflow followed for all experiments is consistent: data preparation with loading a pre trained model, and training on the custom dataset with defined hyperparameters. The results, including model weights and performance metrics, are saved for further analysis. This approach ensured consistency and reproducibility across all models.

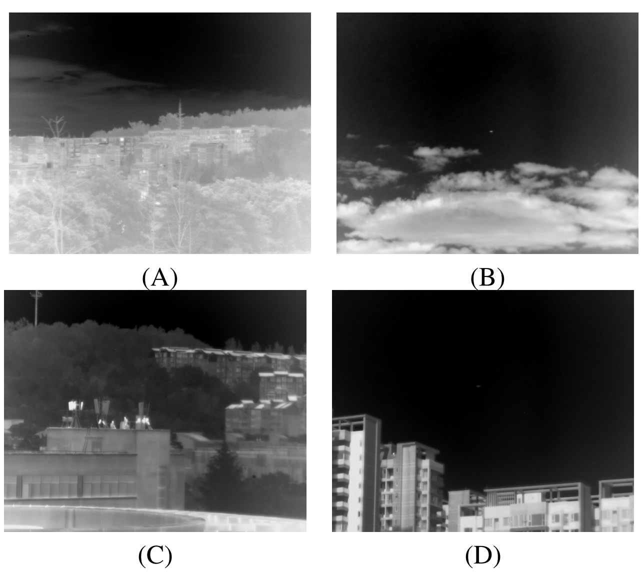

Fig. 2: Sample infrared images

The model is evaluated using following hyperparameters: Epochs determines how many times the dataset is passed through the model during training. Longer training captures fine patterns, while shorter runs provide quick performance insights. Batch size is number of images processed before updating model weights. Smaller batch sizes suit limited GPU memory, while larger sizes stabilize training and reduce gradient noise. Learning rate controls the step size for weight adjustments during training. Default YOLO schedules are used, balancing fast convergence with avoidance of overshooting optimal values. Image size is resized inputs ensure small targets remain detectable while maintaining data detail for effective learning. The dataset contains infrared images of UAVs and birds captured in various challenging scenarios. It has more samples of UAVs compared to birds, reflecting a class imbalance that can influence model training and evaluation. The sample images are shown in Fig.2. The input image for testing is shown in Fig.2(D). The provided image is an infrared (IR) scene featuring urban infrastructure with a background of the sky. The primary objective of using this image is to train machine learning models for small target detection, specifically focusing on identifying UAVs and birds in IR imagery. The image showcases typical challenges in IR based target detection, including low contrast, potential noise, and the presence of cluttered backgrounds. The UAV is visible as a small, bright target against the darker sky, making it an ideal sample for testing and improving the mean average precision of models like YOLO for detecting streamlined and small objects in real world scenarios.

The YOLOv3 model is applied to the original dataset to evaluate its performance on detecting UAVs and birds.The Fig.3 shows UAV and birds detection performed by YOLO v3 model. The model achieved an mean average precision of 75.59% for UAV detection and 38.59% for bird detection.The YOLOv3 performed reasonably well in detecting UAVs, it struggled with bird detection due to the smaller number of bird samples and the less defined features of these targets.

The YOLOv10 nano and small models are trained on the same dataset to analyze their performance in comparison to YOLOv3. The Fig.(4 and 5) represents object detection by YOLO v10 nano and small.YOLOv10 nano is trained for 100 epochs and achieved 72.9% mean average precision for UAV detection and 43.5% for bird detection. YOLOv10 Small, trained for 50 epochs, achieved 71.4% for UAVs and 35.4% for birds. These results showed that YOLOv10 models have similar results as compared to YOLOv3 for UAV detection and also encountered similar challenges with bird detection. From Fig.(4 and 5), it is observed that bird detection is challenging due to dataset imbalance and the inherent complexity of bird features.

The dataset is resampled to balance the UAV and bird samples, and both YOLOv10 Nano and YOLOv5 models are trained. Despite resampling, the mean average precision remained unchanged. YOLOv10 Nano achieved 71.6% for UAVs and 36.9% for birds, while YOLOv5 reached 73.6% for UAVs and 37.6% for birds. This suggests that resampling had no effect on the results and highlights the difficulty YOLO algorithms face when detecting small, less defined targets like birds in challenging infrared conditions, as seen in Figures (6 and 7).

Fig. 3: Detection performed by YOLO v3 on original dataset

Fig. 4: Detection performed by YOLO v10 nano on original dataset

Fig. 5: Detection performed by YOLO v10 small on original dataset

Fig. 6: Detection performed by YOLO v10 nano on resampled dataset

Fig. 7: Detection performed by YOLO v5 on resampled dataset

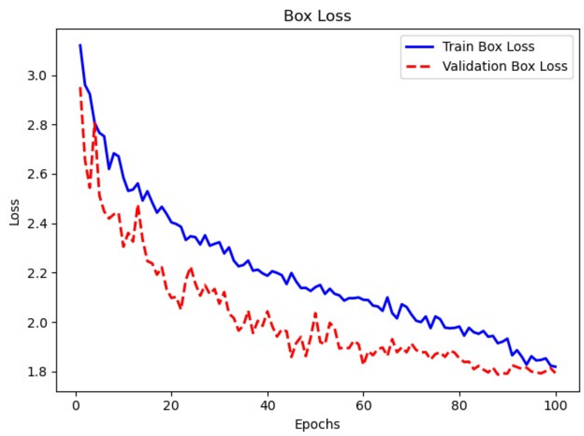

Fig. 8: Training metrics showing box loss for YOLO v10 nano on original dataset

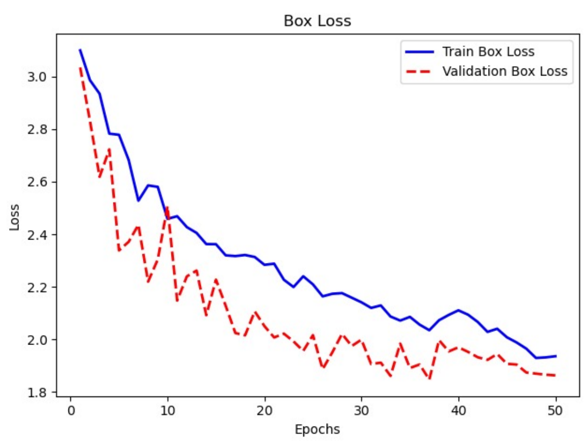

Fig. 9: Training metrics showing box loss for YOLO v10 small on original dataset

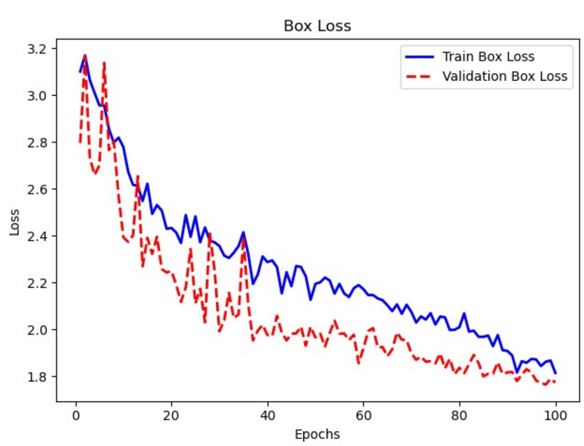

Fig. 10: Training metrics showing box loss for YOLOv10 nano on resampled dataset

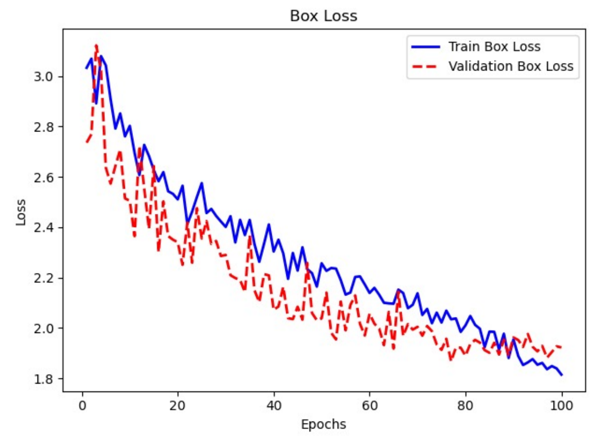

Fig. 11: Training metrics showing box loss for YOLO v5 on resampled dataset

The graphs in Figures(8,9,10 and 11), shows the box loss for the YOLO models trained in this study on both original and resampled dataset. Box loss quantifies the model's mean average precision in predicting bounding box coordinates for detected objects. From Figures(6,8,10 and 12), it is observed that a decreasing trend in both training and validation box loss across epochs reflects the model's improved learning. The slight gap between training and validation curves highlights generalization capability. This analysis helps evaluate the models' performance in detecting small, dim infrared targets like UAVs and birds.

TABLE II: Comparison of YOLO models for UAV and Bird detection on normal dataset.

Model Name	mAP50 UAV	mAP50 Birds	Precision	Recall
YOLOv3	75.59	38.59	72	70
YOLOv10 Nano	72.9	43.5	74.2	51.8
YOLOv10 Small	71.4	35.4	74.9	45.8

TABLE III: Comparison of YOLO Models for UAV and Bird Detection on Resampled Dataset.

Model Name	mAP50 UAV	mAP50 Birds	Precision	Recall
YOLOv10 Nano	71.6	36.9	69.2	47.2
YOLOv5	73.6	37.6	71.8	46.3

The TABLE (II and III) shows, performance parameters used to evaluate the model. The mAP@50 (Mean Average Precision at IoU 50) measures overall detection mean average precision at a fixed IoU threshold of 50%, balancing precision and recall. It evaluates the model's ability to detect UAVs and birds accurately in infrared imagery. Precision reflects the proportion of objects that are correctly detected and classified as UAVs or birds among all detected objects, ensuring minimal false positives, which are objects incorrectly identified as UAVs or birds in cluttered infrared backgrounds. Recall indicates the proportion of actual UAVs and birds present in the images that the model successfully detects, ensuring it captures as many relevant targets as possible in complex scenarios. From Table (II and III), it is observed that the challenge is observed across all YOLO versions tested, including YOLOv3, YOLOv5, and YOLOv10. The challenge is observed across all YOLO versions tested, including YOLOv3, YOLOv5, and YOLOv10. Despite extensive tuning and training, no significant improvements in mean average precision are achieved, highlighting the difficulty of detecting small targets in infrared imagery. The results suggest that YOLO models, while effective in certain scenarios, face difficulties when dealing with small, low contrast objects against complex backgrounds. The experiments revealed that YOLO models generally excel at detecting larger, well defined objects, such as UAVs, but struggle with small, streamlined objects like birds.

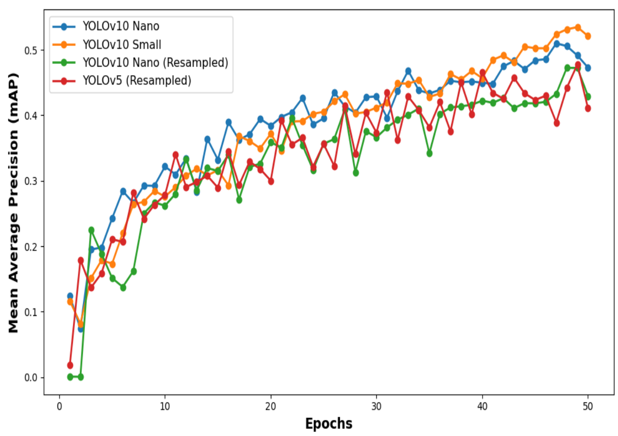

Fig. 12: mAP50 vs epochs

The graph in Fig 12, demonstrates that all the YOLO models, both older and newer versions, exhibit a gradual and slow growth in mean average precision over epochs. Despite the differences in their architectures, all the models perform similarly on dataset, showing almost identical trends in mAP@50. This suggests that the performance improvements are steady but not drastic across different YOLO versions.

V. Conclusion

The paper presents primary challenges addressed in detecting UAVs and birds in infrared imagery captured by UAV borne systems. The task is difficult due to the small size, irregular shapes, and streamlined nature of the targets, which can easily blend with complex backgrounds.Then experimented with several versions of the YOLO object detection framework, including YOLOv3, YOLOv5, YOLOv10 nano, and YOLOv10 small. The models are trained on a dataset consisting of infrared images with annotated UAV and bird targets. Multiple approaches are explored, including resampling the dataset to improve bird detection performance. Experimenting with multiple YOLO versions and implementing resampling techniques, the results indicate that both older and newer YOLO models performed similarly on the dataset. Even after resampling is used to address the low number of bird samples, no significant improvement in detection mean average precision is observed. The YOLO models consistently demonstrated strong performance in detecting UAVs, but their performance remained suboptimal for birds, especially due to their streamlined nature, small size, and tendency to blend with the background. The state-of-the-art (SOTA) object detection models are optimized for general detection tasks The focused approach on detecting UAVs and birds in infrared imagery highlighted the limitations of YOLO models for this specific application. The experiments confirmed that the proposed YOLO models are effective for detecting UAVs, they struggle to detect birds due to the challenges posed by their small size and streamlined shape, which makes them harder to distinguish from the environment. Future work will focus on addressing the challenges posed by small, streamlined objects in complex infrared environments through further advancements in model development and training techniques.

References

[1] L. Zhao, "Uav infrared image human and vehicle object detection based on yolov5," *International Journal of Computer Science and Information Technology*, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:271388408

[2] M. H. Rahman, M. A. S. Sejan, M. A. Aziz, R. Tabassum, J.-I. Baik, and H.-K. Song, "A comprehensive survey of unmanned aerial vehicles detection and classification using machine learning approach: Challenges, solutions, and future directions," *Remote Sensing*, vol. 16, no. 5, 2024. [Online]. Available: https://www.mdpi.com/2072-4292/16/5/879

[3] U. Seidaliyeva, L. Ilipbayeva, K. Taissariyeva, N. Smailov, and E. T. Matson, "Advances and challenges in drone detection and classification techniques: A state-of-the-art review," *Sensors*, vol. 24, no. 1, 2024. [Online]. Available: https://www.mdpi.com/1424-8220/24/1/125

[4] A. Antonenenko, D. Skudnyi, V. Haleta, Y. Mishkur, V. Huminiuk, A. Burachynskyi, S. Vostrikov, A. Balvak, D. Korotin, A. Tverdokhlib, O. Gryhorovych, S. Popereshnyak, O. Tkachenko, O. Golubenko, and D. Tuzhilin, "Object recognition systems using intelligent technologies in uav," in *Prospective Global Scientific Trends 2024*. European Science, 2024, pp. 50–82.

[5] A. Yafoz, "Drones in action: A comprehensive analysis of drone-based monitoring technologies," *Data and Metadata*, vol. 3, p. .364, Sep. 2024. [Online]. Available: https://dm.ageditor.ar/index.php/dm/article/view/364

[6] S. T. Brassai, R. B. Ambarus, A. Hammas, and A. Németh, "Experiments on detecting and monitoring objects based on thermal imaging," in *2024 25th International Carpathian Control Conference (ICCC)*, 2024, pp. 1–6.

[7] J. Zhang, Z. Hong, X. Chen, and Y. Li, "Few-shot object detection for remote sensing imagery using segmentation assistance and triplet head," *Remote Sensing*, vol. 16, no. 19, 2024. [Online]. Available: https://www.mdpi.com/2072-4292/16/19/3630

[8] B. Feng, Y. Jin, Z. Yin, Y. Liu, X. Wang, and Y. Zhao, "Infrared image recognition and classification of typical electrical equipment in substation based on yolov5," in *2023 2nd Asian Conference on Frontiers of Power and Energy (ACFPE)*, 2023, pp. 140–144.

[9] Y. F. Yucesoy and C. Sahin, "Object detection in infrared images with different spectra," in *2024 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2024, pp. 1–6.

[10] M. Misbah, M. U. Khan, Z. Kaleem, A. Muqaibel, M. Z. Alam, R. Liu, and C. Yuen, "Msf-ghostnet: Computationally-efficient yolo for detecting drones in low-light conditions," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, pp. 1–12, 2024.