






Deep Learning Model for Unmanned Aerial Vehicle-based Object Detection on Thermal Images

Moch Arief Soeلمان^{*}, Catur Supriyanto, Purwanto

Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia

Corresponding Author Email: arief2802@dsn.dinus.ac.id

Copyright: ©2023 IETA. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ria.370608>

ABSTRACT

Received: 9 September 2023

Revised: 12 October 2023

Accepted: 21 November 2023

Available online: 27 December 2023

Keywords:

unmanned aerial vehicle, thermal images, deep learning, yolo, HIT-UAV

This research is related to the use of deep learning models on thermal images for object detection using Unmanned Aerial Vehicles (UAVs). Thermal imaging proves to be efficient in environments with minimal light and during nighttime, as it operates based on emitted heat rather than visible light. The ability to detect objects in thermal images enhances surveillance and security measures, particularly in dark conditions. During search and rescue missions, especially in settings with restricted visibility, thermal imaging aids in the identification of individuals or animals by detecting their heat signatures. This facilitates the process of locating them in diverse conditions. Detecting objects at night is challenging due to the lack of illumination. Experiments were conducted using the publicly available HIT-UAV dataset, consisting of 2898 images. This dataset includes several classes such as Person, Car, Bicycle, Other Vehicle, and DontCare. This study proposes the use of both YOLOv5 and YOLOv8 for object detection on this dataset. The YOLOv8 model is the latest model currently available. Both YOLOv5 and YOLOv8 variants were developed by Ultralytics. The experiment used five Yolo models: nano (n), small (s), medium (m), large (l), and extra-large (x). By using YOLOv8m, we achieved a mean Average Precision at IoU threshold 0.5 (mAP@0.5) of 0.855. The performance exceeds that of several previously proposed models, such as YOLOv4-tiny, Faster-RCNN, and YOLOv4, which yielded mAP scores of 0.504, 0.768, and 0.847, respectively.

1. INTRODUCTION

In recent years, the use of unmanned aerial vehicles (UAVs) has gained significant attention in various fields, including object detection. One of the key challenges in object detection is the ability to accurately detect objects in various environmental conditions. In this regard, the utilization of infrared thermal data has emerged as a promising solution. This research paper focuses on the importance of using infrared thermal dataset for UAV-based object detection. The use of infrared thermal data in object detection offers several advantages. Firstly, infrared thermal data provides a unique perspective by capturing the thermal energy emitted by objects, allowing for better differentiation and identification of objects in various lighting and weather conditions [1]. Additionally, it enables the detection of objects that are not visible in traditional visual imagery, such as those hidden behind obstacles or in low-light environments. Moreover, infrared thermal data enhances the accuracy of object detection by minimizing false positives and false negatives, thereby improving the overall performance of object detection systems. Furthermore, the use of UAVs for collecting infrared thermal data presents distinct advantages.

UAVs offer a flexible and cost-effective means of collecting data over large areas, enabling comprehensive coverage and high-resolution data acquisition. The mobility and

maneuverability of UAVs also allow for targeted data collection in specific areas of interest, enhancing the efficiency and effectiveness of object detection tasks. Additionally, UAVs provide a safe and non-invasive method of data collection, eliminating the need for human intervention in potentially hazardous or inaccessible environments. However, the use of infrared thermal data for object detection is not without challenges and limitations. Firstly, the interpretation and analysis of infrared thermal data require specialized knowledge and expertise. The complex nature of thermal signatures and the variations in thermal characteristics of different objects pose challenges in accurately identifying and classifying objects. Moreover, the availability and quality of infrared thermal datasets can be limited, hindering the development and evaluation of object detection algorithms. Additionally, the integration of infrared thermal data with other sensor modalities and data fusion techniques presents technical challenges that need to be addressed for optimal object detection performance.

Infrared thermal data plays a crucial role in object detection, particularly in scenarios where visual light cameras may fall short. One of the key advantages of using infrared thermal images is that they capture information outside the spectrum of the human eye, providing data that is not visible to us [1]. While thermal infrared (TIR) images may not contain the detailed information present in visual RGB images, they still

hold value for object detection purposes. TIR allows object detection to overcome challenges posed by changes in light intensity that can affect color perception by the human eye. Additionally, TIR is sensitive to temperature changes, making it useful for detecting thermal variations in objects. In fact, studies have shown that infrared thermal cameras are better at identifying car and bicycle objects compared to visual light cameras, particularly in nighttime operations [2]. This highlights the importance of using infrared thermal data for object detection, as it provides superior performance compared to visual light cameras in identifying objects at night. Furthermore, the use of infrared thermal data contributes to the development of drone-based object detection tasks and has been shown to enhance the performance of detection models with limited image data [3]. The availability of datasets such as the High-altitude Infrared Thermal Unmanned Aerial Vehicle (HIT-UAV), which comprises high-altitude UAV-based infrared thermal images, further emphasizes the significance of using this type of data for object detection and offers valuable resources for training and testing object detection algorithms. In summary, incorporating infrared thermal data into object detection processes allows for improved detection capabilities, especially in low-light or challenging environmental conditions, and opens up new possibilities for UAV-based applications such as search and rescue missions at night [4].

In recent years, the YOLO (You Only Look Once) method has become popular in object detection using UAVs or drones [5]. YOLO is a real-time object detection algorithm that enables UAVs to efficiently detect objects in images or videos directly. YOLO is designed to provide real-time object detection results, which are crucial in UAV applications such as monitoring and surveillance. With its ability to process images quickly, UAVs can detect and respond to objects rapidly.

Certainly, the YOLO algorithm has outperformed alternative object detection algorithms like the Region-Based Convolutional Neural Network (R-CNN) and the Single-Shot Multi-box Detector (SSD) due to its exceptional real-time detection accuracy [6]. YOLO excels in both precision and speed [7]. YOLO has significant development leading up to YOLOv8. Each new version of the YOLO model brings improvements in accuracy, speed, and object detection capabilities. YOLOv8 is one of the latest iterations of this model, which is likely to continue enhancing its features and performance.

Some studies such as Jia et al. [8] presents a forest fire detection strategy using pre-processed datasets and UAV-captured images, with a focus on YOLOv8 technology. YOLOv8 offers the best balance between accuracy and speed. The proposed model, based on YOLOv8, accurately identifies fires and aids in mitigating forest resource damage. Also in Serrano and Bandala [9] which employs deep learning algorithms within the YOLO architecture, including YOLOv5, YOLOv6, YOLOv7, and YOLOv8, to classify terrain types based on aerial images. In simulations, YOLOv8 achieved the highest mean average precision (mAP@0.5:0.95) of 89.1 and an F1 score of 90.8, outperforming YOLOv5, YOLOv6, and YOLOv7. This demonstrates that YOLOv8 is superior in terrain classification based on mAP and F1 scores.

In the previous study [10], several models such as YOLOv4, Faster-RCNN, and SSD-512 were employed on the HIT-UAV dataset. In that research, SSD-512 appeared to be more accurate than other models. YOLOv4 exhibited limitations in

accurately identifying small objects [11]. In terms of speed, YOLO outperformed SSD-512. For real-time applications, both accuracy and speed are crucial. This study proposes the use of the latest version of YOLO, YOLOv8, for the HIT-UAV dataset.

Several parameters, such as epochs, batch size, and image size, are defined to test the model. Performance comparison is conducted with previous models, namely SSD-512, Faster-RCNN, YOLOv4, and YOLOv4 tiny.

2. RELATED WORKS

Suo et al. [10] presents the HIT-UAV dataset, which is a high-altitude infrared thermal dataset designed for object detection on Unmanned Aerial Vehicles (UAVs). The dataset contains 2,898 infrared thermal images extracted from videos captured by UAVs in various scenarios. Each image is annotated with object instances using bounding boxes of two types to handle the challenge of object overlap in aerial images. The dataset also includes flight data for each image, such as altitude and camera perspective. The authors trained and evaluated well-established object detection algorithms on the dataset and found that the algorithms performed exceptionally well compared to visual light datasets. They believe that the HIT-UAV dataset will contribute to various UAV-based applications and research.

Shaniya et al. [12] focuses on using drones and a combination of RGB and thermal infrared (TIR) images for detecting small objects, particularly humans, from aerial perspectives. The researchers train the YOLOv4 object detection model on both RGB and TIR datasets captured by drones. They demonstrate that YOLOv4 accurately detects humans in both types of images. The study highlights the potential of this technology for improving surveillance and search-and-rescue missions. The YOLOv4 model is enhanced with additional layers and methods, achieving faster and more accurate detection compared to other models. The research utilizes the VisDrone 2021 RGBT dataset, showcasing YOLOv4's successful performance in overcoming real-world challenges. Overall, this study contributes to saving lives and enhancing security measures through the advancement of drone-based object detection.

Perdana et al. [13] used special cameras called thermal cameras on drones to find and rescue people during disasters. The researchers use a computer program called deep learning Convolutional Neural Network (CNN) to help the cameras detect people accurately, even in difficult situations. By changing the structure of the deep learning network, they improve accuracy without needing a lot of computer power. They train the model using various datasets and annotate the data using an image annotation tool. The study shows that their approach can locate victims more effectively and save lives during disasters. The research was supported by the Ministry of Research, Technology, and Higher Education in Indonesia, and references various sources for learning about thermal camera detection and deep learning techniques.

Mantau et al. [14] demonstrates the utilization of the advanced object-detection technique known as YOLOv5. This method is applied to a dataset comprising visual images captured from a UAV (RGB imagery) combined with TIR for the purpose of poacher detection. The research employs seven distinct training approaches involving both RGB and thermal infrared data to identify the most effective model, which will

subsequently be deployed on the Jetson Nano module. The experimental outcomes reveal that a novel model, employing transfer learning with a pre-trained model from the MS COCO dataset, enhances the ability of YOLOv5 to detect humans and objects within the RGBT image dataset.

Wang et al. [15] focuses on improving UAV object detection, where small objects and resource constraints pose challenges. The proposed UAV-YOLOv8 model optimizes YOLOv8 for aerial photography scenarios. They introduce Wise-IoU v3 for better localization, employs the BiFormer attention mechanism, and designs the FFNB feature processing module to enhance feature integration. The model achieves superior detection performance, with 7.7% higher accuracy compared to the baseline model and outperforming other mainstream models. While it excels in small object detection, future research will address further improvements for feature-less objects like bicycles. Wu et al. [16] propose a robust and real-time tracking algorithm for infrared drones, incorporating a feature attention module and an expansion strategy for searching the target. The algorithm is designed to track drones in real-time, addressing the challenges of real-infrared scenes with high efficiency. The proposed algorithm is based on the Anti-UAV infrared dataset, which has been used to analyze the performance of thermal and infrared (TIR) tracking comparisons. The paper aims to track the drone in a video, considering the drone's size and texture, as well as the presence of buildings, trees, and false similar targets. The proposed algorithm is expected to be more efficient than existing object tracking algorithms, making it a valuable tool for detecting and tracking UAVs in various applications.

A deep learning-based approach is proposed by Ding et al. [17] for detecting and tracking small targets in infrared images. The authors enhance the network architecture of the Single Shot MultiBox Detector (SSD) specifically for infrared small target detection, introducing a modified version called Single Shot MultiBox Detector for Small Target (SSD-ST). By eliminating low-resolution layers and enhancing high-resolution layers, the performance of the network is improved. To further refine the detection results, the researchers introduce an Adaptive Pipeline Filter (APF) that utilizes temporal correlation and motion information. This filter effectively corrects the detection outcomes. The proposed method is evaluated using a dataset comprising 16,177 infrared images and 30 trajectories. The results demonstrate a recall rate exceeding 90% and a precision exceeding 95%. These findings indicate that the proposed method outperforms traditional approaches in complex scenes, successfully accomplishing the task of detecting and tracking infrared small targets. The utilization of infrared imaging has gained significant attention due to its affordability, resistance to interference, and capability to operate in various weather conditions. Nonetheless, the detection and tracking of small UAV targets in infrared images pose considerable challenges.

The researchers also address the difficulties associated with atmospheric cloud radiation and imaging noise, which lead to a relatively low signal-to-noise ratio (SNR).

3. METHODS

3.1 Dataset

The HIT-UAV is a public dataset, which gathered from numerous videos recorded in public spaces like schools,

parking lots, streets, and play areas. The dataset includes 2,898 infrared thermal images captured by unmanned aerial vehicles (UAVs) across different settings, such as schools, parking lots, roads, playgrounds, and more. The HIT-UAV dataset offers dual types of labeled bounding boxes for every object depicted in the images: oriented and standard. The oriented bounding box addresses the challenge of substantial overlap among object instances in aerial images, while the standard bounding box facilitates efficient utilization of the dataset. The dataset includes five object types: People, Cars, Bicycles, Other Vehicles, and DontCare, including a total of 24,899 annotated objects. The category labeled DontCare is reserved for objects that could not be precisely classified by the annotators. Annotation files in XML and JSON formats are provided, aligning with the VOC and MS COCO dataset formats, respectively. Figure 1 shows the sample of HIT-UAV thermal images. The dataset consists of 2,029 images for training, 290 images for validation, and 579 images for testing.



Figure 1. Samples of HIT-UAV Thermal Images

3.2 Experimental design

For object detection, the YOLOv8 model is employed in multiple versions, including YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x. YOLOv8n, the nano version of Yolo, could be a smaller and lighter version of Yolo, suitable for resource-constrained environments such as edge or IoT devices. YOLOv8s is usually refers to a smaller version of the Yolo model, designed for faster processing on less powerful hardware. YOLOv8s may sacrifice some accuracy for speed. YOLOv8m is likely a mid-sized configuration that aims to balance between accuracy and speed. It might be suitable for applications where real-time processing is necessary, but with a bit more emphasis on accuracy compared to smaller versions. YOLOv8l is a larger and more complex version of Yolo, likely designed to achieve higher accuracy at the cost of increased computational resources. It might be suitable for applications where accuracy is critical and computational power is less of a constraint. The last version is YOLOv8x, which refers to an even larger or more feature-rich version of Yolo, potentially with additional capabilities or improvements over the large version. These five Yolo versions are used to determine which model is suitable for the dataset.

Figure 2 depicts the YOLOv8 network structure, which comprises mostly of a backbone, neck, and head. The experiment was conducted utilizing the NVIDIA DGX A100 Server equipped with a GPU capacity of 40 gigabytes (GB). NVIDIA DGX A100 is designed specifically to provide high computational performance, especially in deep learning workloads. By default, the dataset follows the COCO format, where the information for each object is stored within a JSON file. To convert this format into YOLO format, the annotations are saved in a TXT file, with each image in the dataset having a corresponding single text file.

The YOLO model was trained using an image size of 512x512 with 150 epochs and 16 batch size. Table 1 shows the parameter value of YOLO model. This research employs a larger number of epochs compared to previous study [10] because involving more epochs provides the model with more opportunities to adapt to patterns and features present in the training data. This can enhance the model's ability to comprehend more abstract and complex representations. In some cases, involving more epochs can aid in addressing overfitting issues. By extending the training process, the

model has more chances to adapt to the training data without generating poor generalization on the test data.

Table 1. Parameter value of YOLOv8

Parameter	Value
Image size	512
Epoch	150
Batch	16

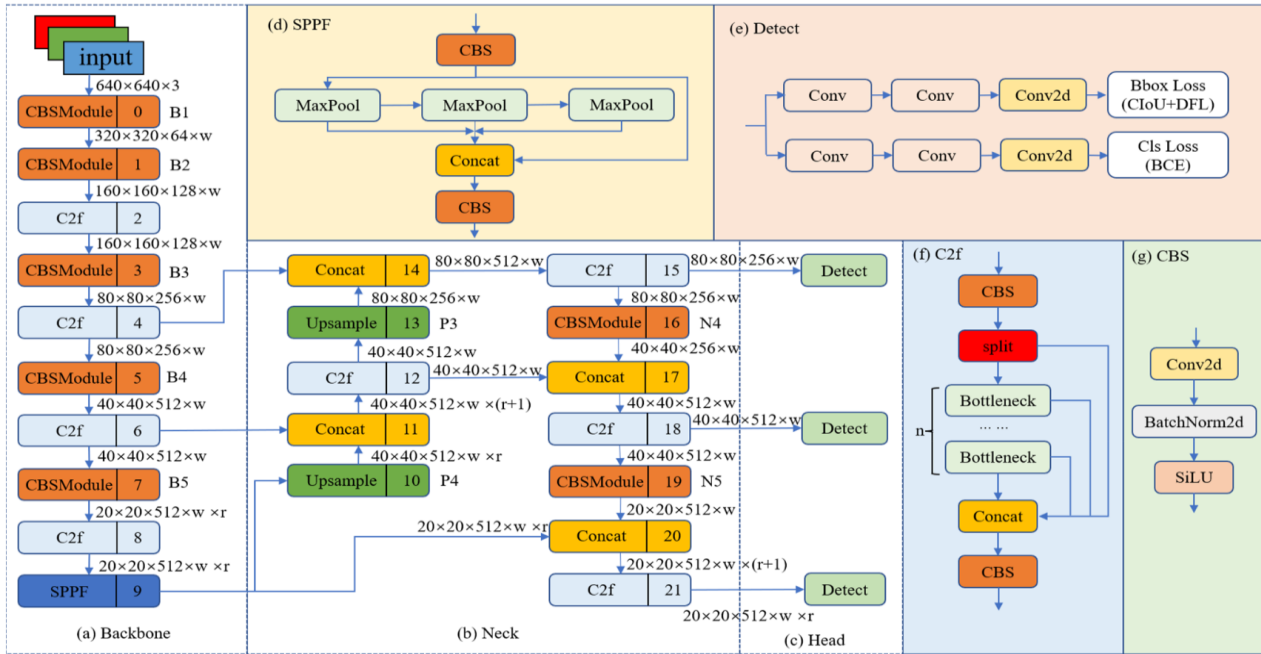


Figure 2. The architecture of YOLOv8 [15]

3.3 Performance metrics

This study evaluates the proposed model using the well-known metrics, such as precision (P), recall (R), and mean average precision (mAP). Recall measures the model's ability to detect all true instances of objects in an image. It is calculated as the number of true detections divided by the total number of actual objects. High recall indicates that the model is less likely to miss actual objects but may produce many false positives.

Precision measures how accurate the model's detections are. It is calculated as the number of true detections divided by the total number of positive detections (both true and false). High precision indicates that the model is less likely to provide many false positives but may miss some actual objects. mAP is a more comprehensive metric used to evaluate the performance of object detection by considering precision at various levels of recall. It measures how accurately the model detects objects at different recall levels. mAP is calculated by computing the area under the Precision-Recall curve and then taking the average of these areas for all object classes. mAP is a useful metric since it provides a better understanding of the model's performance across different object classes and difficulty levels in the object detection task. Using these three metrics give more complete picture of how well the model performs in object detection.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

$$mAP = \frac{1}{N} \sum_{c=1}^N \frac{TP_c}{TP_c + FP_c} \quad (3)$$

where, TP denotes the quantity of accurately identified positive samples as positive, while FP indicates the quantity of incorrectly identified negative samples as positive. FN , on the other hand, signifies the count of erroneously predicted negative samples as negative when they are actually positive.

4. RESULTS

In this section, YOLOv5 and YOLOv8 are compared to previous researchers [1] who used YOLOv4-tiny, FasterRCNN, SSD-512 and YOLOv4 with the same dataset. Table 2 and Table 3 show the performance results of YOLOv5 and YOLOv8 on the testing dataset. In both of these models, x model produced a low mAP@0.5 score 0.818 and 0.812, while m model was able to achieve a high mAP@0.5 score 0.852 and 0.855, respectively on YOLOv5 and YOLOv8.

In YOLOv8, if there is no improvement in performance within 50 epochs, the iteration will stop. In general, the computation time during the training process of YOLOv8 is competitive with YOLOv5. With the same number of epochs, which is 150, YOLOv8n has a training time of 24 minutes,

which is faster than YOLOv5n, which has a training time of 25.62 minutes. Conversely, in Yolo v5l, the training time is faster at 45.66 minutes, compared to YOLOv8l, which has a training time of 46.08 minutes. The faster training process is beneficial during the development of detection models. In some cases, rapid training can lead to the quicker deployment of models into production or practical use. This can reduce the time between model development and implementation.

Table 2. The performances of various YOLOv5 models on testing dataset

Model	P	R	mAP@0.5	mAP	Training Time (min)	Epoch
YOLOv5n	0.895	0.756	0.833	0.555	25.62	150
YOLOv5s	0.862	0.801	0.852	0.569	27.12	150
YOLOv5m	0.877	0.814	0.852	0.565	34.38	150
YOLOv5l	0.878	0.798	0.847	0.572	45.66	150
YOLOv5x	0.881	0.78	0.818	0.543	66.66	150

Table 3. The performances of various YOLOv8 models on testing dataset

Model	P	R	mAP@0.5	mAP	Training Time (min)	Epoch
YOLOv8n	0.885	0.721	0.81	0.528	24	150
YOLOv8s	0.875	0.819	0.853	0.559	23.64	142
YOLOv8m	0.868	0.828	0.855	0.568	34.8	150
YOLOv8l	0.872	0.79	0.836	0.562	46.08	150
YOLOv8x	0.851	0.775	0.812	0.544	50.94	123

	Person	Car	Bicycle	OtherVehicle	DontCare	Background
Person	2330	0	12	0	2	455
Car	2	1274	0	11	0	155
Bicycle	1	9	707	0	2	193
OtherVehicle	0	5	0	21	1	4
DontCare	0	0	0	0	19	13
Background	278	51	77	2	7	0

Figure 3. Confusion Matrix on YOLOv8n

	Person	Car	Bicycle	OtherVehicle	DontCare	Background
Person	2358	0	10	0	2	295
Car	2	1275	2	7	0	80
Bicycle	1	0	727	0	0	113
OtherVehicle	0	3	2	24	0	3
DontCare	1	0	0	0	23	9
Background	249	61	55	3	6	0

Figure 4. Confusion Matrix on YOLOv8m

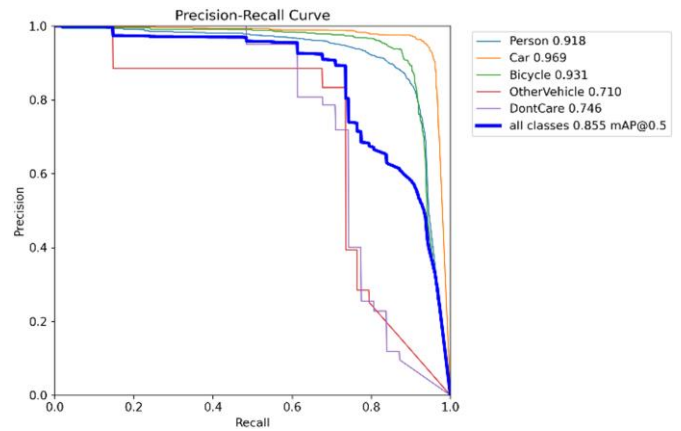


Figure 5. Precision Recall Curve on YOLOv8m

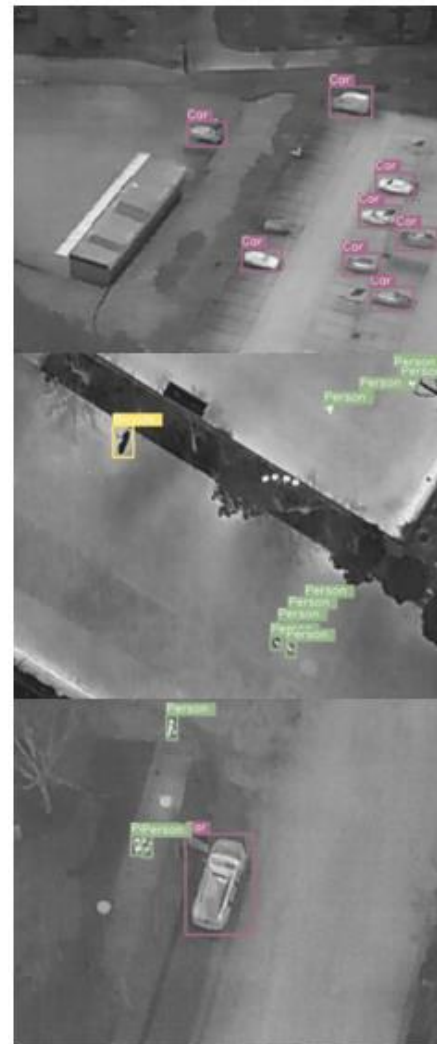


Figure 6. HIT-UAV thermal images with object detection for Car (Purple), Bicycle (Yellow), and Person (Green)

Table 4. Comparison with the SOTA models on HIT-UAV dataset

Ref	Model	mAP@0.5
Suo et al. [10]	YOLOv4-tiny	0.504
Suo et al. [10]	Faster-RCNN	0.768
Suo et al. [10]	YOLOv4	0.847
Suo et al. [10]	SSD-512	0.856
Ours	YOLOv5m	0.852
Ours	YOLOv8m	0.855

Figure 3 and Figure 4 show the confusion matrices for YOLOv8n and YOLOv8m. In YOLOv8n, there are still error detections, specifically, the Car object being detected as Bicycle. Cars and bicycles can have similar shapes and sizes, especially when viewed from a distance or at certain angles. YOLO relies on the features it has learned during training, and if the differences between these classes are subtle, it may lead to confusion. This issue does not occur in YOLOv8m. Based on the confusion matrix, YOLO is capable of effectively distinguishing between the objects Person, Car, and Bicycle. However, there are still many error detections where these three objects are mistakenly classified as the Background. Thermal images capture heat radiation rather than visible light, and this can introduce complexities that differ from traditional color images. Objects like cars, persons, and bicycles might have limited variability in their thermal signatures, making it difficult for YOLO to distinguish them clearly, especially if the temperature differences between objects and the background are subtle. Figure 5 shows the Precision-Recall Curve for YOLOv8m. The results of object detection using YOLOv8 can be seen in Figure 6.

Table 4 presents a comparison with previous publications on the same dataset. The mAP@0.5 value for Yolo v8 is better compared to YOLOv5 and YOLOv4, although the difference is slight. The difference in mAP@0.5 values for YOLOv8 is quite significant when compared to YOLOv4-tiny and Faster-RCNN, which only yield mAP@0.5 values of 0.504 and 0.768, respectively. However, when compared to SSD-512, the performance of YOLOv8 is still slightly inferior, with a difference of only 0.001.

While the difference is very slight, in some cases, small variations in the performance of detection models can impact the overall accuracy of the application. For instance, in applications that require highly accurate object detection, even slight differences can be critical. In the context of security or surveillance, minor differences in detection capabilities can have serious implications. Errors in detecting specific objects may affect the effectiveness of the security system. Small differences may be more relevant in situations where resource savings (such as computational power or memory) are crucial. Models with nearly equivalent performance but greater efficiency may be a preferable choice.

5. CONCLUSION AND FUTURE WORK

The object detection application through UAV at night not only requires a fast model but also demands high accuracy. YOLO is a model suitable for its real-time processing capabilities; however, the accuracy shown in previous research is still low. In this study, the use of the latest version of YOLO is proposed. In this study, YOLOv5 and YOLOv8 models are proposed for use with a public dataset called HIT-UAV. This dataset consists of a collection of thermal images with various objects within them, such as Cars, Humans (Persons), Bicycles, and Other Vehicles. The experimental results show that YOLOv5 and YOLOv8 achieved mAP@0.5 scores of 0.852 and 0.855, respectively. From the experimental results, the performance of YOLOv8 and SSD is slightly different. YOLOv8 can be a choice when applied to real-time thermal detection applications. Accurate models can enhance the UAV's ability to detect objects or obstacles at night, enabling safer navigation and more effective obstacle avoidance. Accurate models allow UAVs to monitor the

environment at night, including monitoring forest fires, changes in surface temperatures, or detecting air pollution in low-light conditions. For future work, it is essential to provide an explainable AI (XAI) such as SHAP or LIME to explain the results. The use of XAI helps to enhance users' trust in the outputs of intelligent system applications. Thus, AI models, known for being black box systems, can be avoided.

ACKNOWLEDGMENT

Thanks to DRTPM RISTEKDIKTI for funding this research in 2023.

REFERENCES

- [1] Teju, V., Bhavana, D. (2020). An efficient object detection using OFSA for thermal imaging. *The International Journal of Electrical Engineering & Education*, 60(1S): 1957-1978. <https://doi.org/10.1177/0020720920944434>
- [2] Cheng, P., Xiong, Z., Bao, Y., Zhuang, P., Zhang, Y., Blasch, E., Chen, G. (2023). A deep learning-enhanced multi-modal sensing platform for robust human object detection and tracking in challenging environments. *Electronics*, 12(16): 3423. <https://doi.org/10.3390/electronics12163423>
- [3] Rietz, J., van Beeck Calkoen, S.T., Ferry, N., Schlüter, J., Wehner, H., Schindlitz, K.H., Lackner, T., Hoermann, C.V., Conraths, F.J., Müller, J., Heurich, M. (2023). Drone-based thermal imaging in the detection of wildlife carcasses and disease management. *Transboundary and Emerging Diseases*, 2023: 5517000. <https://doi.org/10.1155/2023/5517000>
- [4] Banuls, A., Mandow, A., Vázquez-Martín, R., Morales, J., García-Cerezo, A. (2020). Object detection from thermal infrared and visible light cameras in search and rescue scenes. In *2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR)*, Abu Dhabi, United Arab Emirates, pp. 380-386. <https://doi.org/10.1109/SSRR50563.2020.9292593>
- [5] Baidya, R., Jeong, H. (2022). YOLOv5 with convMixer prediction heads for precise object detection in drone imagery. *Sensors*, 22(21): 8424. <https://doi.org/10.3390/s22218424>
- [6] Shi, Q., Li, J. (2020). Objects detection of UAV for anti-UAV based on YOLOv4. In *2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT)*, Weihai, China, pp. 1048-1052. <https://doi.org/10.1109/ICCASIT50869.2020.9368788>
- [7] Redmon, J., Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv Preprint arXiv: 1804.02767*. <https://doi.org/10.48550/arXiv.1804.02767>
- [8] Jia, X., Wang, Y., Chen, T. (2023). Forest fire detection and recognition using YOLOv8 algorithms from UAVs images. In *2023 IEEE 5th International Conference on Power, Intelligent Computing and Systems (ICPICS)*, Shenyang, China, pp. 646-651. <https://doi.org/10.1109/ICPICS58376.2023.10235675>
- [9] Serrano, K.K.D., Bandala, A.A. (2023). YOLO-based terrain classification for uav safe landing zone detection. In *2023 IEEE Region 10 Symposium (TENSYP)*, Canberra, Australia, pp. 1-5.

- <https://doi.org/10.1109/TENSYMP55890.2023.10223656>
- [10] Suo, J., Wang, T., Zhang, X., Chen, H., Zhou, W., Shi, W. (2023). HIT-UAV: A high-altitude infrared thermal dataset for unmanned aerial vehicle-based object detection. *Scientific Data*, 10(1): 227. <https://doi.org/10.1038/s41597-023-02066-6>
- [11] Liu, H., Fan, K., Ouyang, Q., Li, N. (2021). Real-time small drones detection based on pruned yolov4. *Sensors*, 21(10): 3374. <https://doi.org/10.3390/s21103374>
- [12] Shaniya, P., Jati, G., Alhamidi, M.R., Caesarendra, W., Jatmiko, W. (2021). Yolov4 rgbt human detection on unmanned aerial vehicle perspective. In 2021 6th International Workshop on Big Data and Information Security (IWBIS), Depok, Indonesia, pp. 41-46. <https://doi.org/10.1109/IWBIS53353.2021.9631856>
- [13] Perdana, M.I., Risnumawan, A., Sulistijono, I.A. (2020). Automatic aerial victim detection on low-cost thermal camera using convolutional neural network. In 2020 International Symposium on Community-Centric Systems (CcS), Tokyo, Japan, pp. 1-5. <https://doi.org/10.1109/CcS49175.2020.9231433>
- [14] Mantau, A.J., Widayat, I.W., Leu, J.S., Köppen, M. (2022). A human-detection method based on YOLOv5 and transfer learning using thermal image data from UAV perspective for surveillance system. *Drones*, 6(10): 290. <https://doi.org/10.3390/drones6100290>
- [15] Wang, G., Chen, Y., An, P., Hong, H., Hu, J., Huang, T. (2023). UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios. *Sensors*, 23(16): 7190. <https://doi.org/10.3390/s23167190>
- [16] Wu, H., Li, W., Li, W., Liu, G. (2020). A real-time robust approach for tracking uavs in infrared videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, pp. 1032-1033. <https://doi.org/10.1109/CVPRW50498.2020.00524>
- [17] Ding, L., Xu, X., Cao, Y., Zhai, G., Yang, F., Qian, L. (2021). Detection and tracking of infrared small target by jointly using SSD and pipeline filter. *Digital Signal Processing*, 110: 102949. <https://doi.org/10.1016/j.dsp.2020.102949>