Research Article GU J Sci, Part A, 13(2): 710-732 (2026) 10.54287/gujsa.1905999

Gazi University
Journal of Science
PART A: ENGINEERING AND INNOVATION
http://dergipark.org.tr/gujsa

Analysis of Deep Learning-Based Object Detection Across Infrared Bands for UAV Imagery in Military Applications

Yusuf Furkan YUCESOY[1] Cagri SAHIN[1*]

[1] Department of Computer Engineering, Gazi University, Ankara, Türkiye

Keywords

Image Processing
Military Applications
Deep Learning
Infrared Bands

Abstract

In modern military operations, object detection plays an important role in enhancing situational awareness and tactical decision-making. However, the use of object detection in complex military environments faces challenges, including poor visibility and the need for real-time decision-making. Infrared imaging technologies, which operate outside the visible spectrum, are effective for detecting military targets, particularly in low-visibility conditions. Despite the advantages of infrared imaging for object detection, comparison of performance across different infrared spectra in military applications remains limited. This limitation is primarily due to the varying object visibility across different infrared bands and the lack of accessible military datasets, which are crucial for analyzing how these bands influence military object detection and the performance of deep learning models under such conditions. This study provides an analysis of object detection across various infrared bands, including near-infrared, short-wavelength infrared, mid-wavelength infrared, and long-wavelength infrared, using popular deep learning-based object detection models such as YOLOv8, YOLOv9, YOLO Ghost P2, and RT-DETR. The models were trained separately on each infrared band using simulated data and evaluated for their ability to recognize military tanks, as this capability is crucial for unmanned aerial vehicles in modern defense applications. Our findings reveal that the near-infrared band is the most suitable for detecting tanks, and the YOLOv8 model provides the best performance.

Cite

Yucesoy, Y. F., & Sahin, C. (2026). Analysis of Deep Learning-Based Object Detection Across Infrared Bands for UAV Imagery in Military Applications. *GU J Sci, Part A*, *13*(2), 710-732. doi:10.54287/gujsa.1905999

Author ID (ORCID Number)

0009-0000-0204-2560 Yusuf Furkan YUCESOY
0000-0001-5800-6243 Cagri SAHIN

Article Process

Submission Date 09.03.2026
Revision Date 10.04.2026
Accepted Date 05.06.2026
Published Date 30.06.2026

1. INTRODUCTION

In modern military operations, integrating advanced technologies is essential to maintaining a strategic advantage and operational efficiency. Among these technologies, object detection has become a crucial tool, significantly enhancing decision-making, situational awareness, and the effectiveness of tactical operations (Aziz et al., 2020). Object detection systems enable the automatic identification and classification of enemy forces, vehicles, infrastructure, and potential threats in real-time. These systems support a broad range of applications, from the autonomous navigation of unmanned aerial vehicles (UAVs) to surveillance, target acquisition, and security operations.

However, the military's use of object detection encounters several challenges, primarily due to the complex and unpredictable environments in which operations occur. Visibility can be significantly reduced by factors

*Corresponding Author, e-mail: cagrisahin@gazi.edu.tr

This work is licensed under Creative Commons Attribution-ShareAlike 4.0 International.

such as poor lighting, which can hinder the effectiveness of detection systems. Additionally, military environments often contain a variety of objects, making it challenging to distinguish between key targets and irrelevant background noise. In such situations, quick and accurate decision-making is crucial for military success. To address these challenges, infrared (IR) imaging technology, which operates in a different wavelength range than visible light, offers significant opportunities in the field of object detection for military applications. Especially in modern conflict environments, the quick and accurate detection of enemy tanks is critical to the success of military strategies. Therefore, the detection of tanks using unmanned aerial vehicles is considered an essential element of military defense plans (Liu et al., 2022; Soumya et al., 2023).

There are different bands in the infrared spectrum, and these bands can cause various types of objects to appear differently. The detected bands in the infrared spectrum can affect the effectiveness of the detection process. Some objects appear more distinct in the infrared spectrum, while others may be less distinct or even invisible. This can complicate the analysis of infrared images and the detection of objects. Therefore, to use infrared images effectively, the effects of different bands in the infrared spectrum on tank detection need to be analyzed in order to optimize detection capabilities.

Although there are studies investigating the performance of object detection algorithms in infrared images, there remains a lack of comprehensive research comparing the performance across different infrared spectra, particularly in military applications (S. Li et al., 2021; Zhang et al., 2022; Tham et al., 2023). The reasons for this are the limited availability of publicly accessible infrared image datasets across different infrared bands, and infrared datasets used in military applications are not available due to security concerns.

The primary aim of this study is to analyze the performance of various deep learning models for military tank detection, to investigate the impact of variations in the infrared spectrum on detection accuracy, and to evaluate the influence of training duration (epochs) on model performance. To achieve these objectives, we analyze object detection for recognizing military tanks across various infrared spectra, including near-infrared (NIR), short-wavelength infrared (SWIR), mid-wavelength infrared (MWIR), and long-wavelength infrared (LWIR) by using You Only Look Once version 8 (YOLOv8), You Only Look Once version 9 (YOLOv9), You Only Look Once Ghost Phase 2 (YOLO Ghost P2), and Real-Time DEtection TRansformer (RT-DETR) deep learning models on simulated UAV imagery. During the analysis, we tested the models that were trained separately for each infrared band using test data from their own bands. After that, we created a common test dataset by combining the test data from each infrared band and evaluated how all the models performed on this test dataset. We also investigated the impact of the different epochs (100, 200, 300, 400) on model accuracy to determine how training duration affects detection capabilities in each IR spectrum. Note that, in our previous conference work, we only used the YOLOv8 and YOLOv9 models for object detection in different IR spectra, with each trained for just 100 epochs (Yucesoy & Sahin, 2024).

The results of our study demonstrate the following:

- For all performance metrics, which are box loss, generalized intersection over union (GIoU), and mean average precision at IoU thresholds from 0.5 to 0.95 (mAP50-95), the considered deep learning models using NIR exhibit the best performance.
- Among the selected deep learning models, YOLOv8 is the most suitable for recognizing military tanks in the IR spectrum.
- The performance of newer models is not always superior, as demonstrated by YOLOv8's consistent advantage over YOLOv9.
- Increasing training epochs can improve model performance, but it may also introduce variability and declines in specific infrared bands.
- As the wavelength increases in the IR spectrum, detection success decreases.
- The transformer-based RT-DETR model takes around 1.8x longer to train than YOLO-based models for the same number of epochs.

The remainder of this article is organized as follows: The next two sections provide general background information on infrared bands, deep learning models, and related work. The methodology section then explains how the scenario and dataset were generated, including the tools used, as well as details on model training and performance metrics. This is followed by the presentation of results, accompanied by a discussion of the findings. After addressing potential threats to validity, we conclude with our findings and propose directions for future research.

2. BACKGROUND

2.1. Infrared Imaging

One of the key features of infrared technology is that it can operate under conditions that the human eye is unable to detect. With this capability, IR imaging can be used in fields such as medicine, search and rescue, security, industrial inspections, environmental monitoring, building inspections, and agriculture. In the military, IR imaging provides a strategic advantage by detecting hidden and camouflaged objects.

IR images obtained from sensors typically operate in different spectra, which are divided into four main categories: Near Infrared (NIR, 0.75–1.4 μm), Short-Wave Infrared (SWIR, 1.4–3 μm), Mid-Wave Infrared (MWIR, 3–8 μm), and Long-Wave Infrared (LWIR, 8–12 μm) (EMF Corporation, 2024). In the context of IR imaging, “μm” (micrometers) represents a unit of length used to measure wavelengths of electromagnetic radiation, which are used to categorize the infrared spectrum. IR radiation is typically linked to two main sources: reflected radiation, which originates from external light sources, and the emission of thermal energy from objects, which is directly related to their temperature. Overall, the main differences between IR spectra lie in their wavelength ranges and how they interact with materials, which affects their applications and usefulness.

2.2. Deep Learning Models

One of the fundamental tasks in computer vision is object detection, which focuses on identifying and locating objects within images (Gomaa et al., 2020). In deep learning, object detection models are designed to both classify and pinpoint the locations of objects by producing class labels and bounding boxes. Deep learning-based object detection involves feature extraction, bounding box prediction, object classification, and non-maximum suppression. Key models include region proposal methods like Faster R-CNN, which offer accuracy but slower performance, single-stage networks like YOLO for faster real-time detection, and transformer-based models like DETR, which replace CNNs with transformers and eliminate post-processing steps like non-maximum suppression (Shah & Tembhurne, 2023)

In this study, we used 4 deep learning models: YOLOv8, YOLOv9, YOLO Ghost P2, and RT-DETR. YOLOv8 delivers strong real-time object detection performance, balancing accuracy and speed (Sohan et al., 2024). YOLOv9 is the newer version of YOLO and enhances feature extraction and small object detection. When quick processing is important, YOLO Ghost P2 can be used, as it relies on innovative layers to boost performance while requiring low computational resources. Finally, RT-DETR uses an optimized transformer-based approach, enhancing detection capabilities and speed in real-time applications (Zhao et al., 2024). It provides superior detection performance in complex scenarios by leveraging attention mechanisms to better understand object relationships and contexts. However, RT-DETR's transformer-based architecture requires longer training and generally achieves lower inference speeds (FPS) compared to YOLO-based models, creating a trade-off between accuracy and real-time performance.

3. RELATED WORK

Object recognition studies in infrared imagery play a critical role in various fields, including security, military applications, healthcare, and space research. Research in the literature indicates that, due to the ability of the infrared spectrum to detect objects that are invisible to the human eye, effective object recognition can be achieved even in low-light conditions or completely dark environments (Bao et al., 2023; Sun et al., 2024). In recent years, machine learning and deep learning algorithms have greatly improved object recognition in infrared images.

Several studies provide a comprehensive overview of research conducted on infrared images using various object recognition methods. For example, Sun et al. (2024) focus on the fusion of visible and infrared images for object detection, showing how infrared thermal imaging addresses challenges in low light, adverse weather, and obstructed environments. Wu et al. (2020) propose a real-time UAV tracking approach in infrared videos, enhancing a fully convolutional classifier with a feature attention module and target search strategy, demonstrating robustness and effectiveness in challenging environments and handling target displacement. Yang et al. (2025) discuss the challenges and future directions of infrared object segmentation, highlighting

issues such as low signal-to-noise ratios and small, blurred objects in infrared images, while also suggesting deep learning-based methods for overcoming these challenges.

To enhance infrared small target detection, X. Wang et al. (2023) propose an attention-based method inspired by the human visual system, improving object detection in complex backgrounds like the sea and sky. Similarly, a study on pedestrian detection in infrared highlights its advantages in dark environments and addresses dataset limitations with a transfer learning approach (Hu et al., 2020). Building on these advancements, Moradi et al. (2020) developed a method for detecting small targets in search and tracking systems, tackling the challenges of high gray-density backgrounds. Furthermore, Jinke (2023) improves infrared object detection by incorporating spatial and temporal information, further enhancing accuracy in challenging conditions. Sadic et al. (2025) provide a comprehensive review of the use of deep learning for object detection in infrared images, highlighting the challenges posed by low contrast and non-uniformity in such images, and suggesting possible directions for future research.

In our experiments, we used the YOLO series models, which stand out for their efficiency in speed and performance, alongside the transformer-based object recognition model known as RT-DETR. Terven et al. (2023) conducted a comprehensive study on the evolution of YOLO architectures in computer vision, highlighting their importance in real-time object detection systems and applications. YOLO-based models enable automatic labeling, with a customized YOLOv8 fine-tuned for better moving object detection and classification (Gomaa, 2024). Zhao et al. (2024) pointed out YOLO's speed and accuracy limitations due to Non-Maximum Suppression (NMS) and introduced RT-DETR, which improves performance via a hybrid encoder and uncertainty-minimal query selection, surpassing earlier YOLO models. Zhou et al. (2025) apply YOLOv8 with super-resolution reconstruction to enhance UAV detection in infrared images, achieving high accuracy and excellent detection performance, showing the potential of YOLO models in real-world applications.

In addition to these studies, several studies have investigated the combination of YOLO algorithms with IR imagery, highlighting its potential to enhance object detection capabilities across various applications (S. Li et al., 2021; L. Li et al., 2022; Shen et al., 2023; Tu et al., 2023; Y. Wang et al., 2024). The potential of integrating UAV and infrared imaging technologies has also been explored for urban road safety (Aibibu et al., 2024). Additionally, these technologies enable rapid and effective search operations during disasters or accidents (Lagman et al., 2022). Some other studies compare the performance of various deep learning models for infrared object detection and evaluate their effectiveness. For example, a comparative study using deep learning-based infrared object detection models like YOLOv3-4GB, YOLOv5s, and YOLOv7 on the FLIR dataset showed that YOLOv7 achieved the best performance (Aboalia et al., 2024). Tanchuk and Kolobrodov (2024) proposed that YOLOv8 and DETR demonstrate better object detection performance than YOLOv5 and Faster R-CNN. Another study proposes the lightweight YOLO-SMUG model, which addresses high computational demands and improves the detection of small objects in UAV-based infrared imagery,

outperforming existing models in both accuracy and efficiency (Luo & Zhu, 2025). Shchetynin (2025) also presents a model based on YOLO5, which uses the Bi-FPN network to enhance object detection accuracy in infrared images obtained by UAVs.

Furthermore, studies in object detection have primarily focused on utilizing specific infrared bands, which have been used individually in various studies across different topics, such as people, animals, and vehicles (Lee et al., 2015; Ghenescu et al., 2018; Price & Price, 2019; Pavlović et al., 2022). A common factor among these studies is the use of self-collected data.

Although these studies demonstrate object detection in infrared imaging using deep learning, there is a lack of comparisons between different infrared bands, particularly concerning military IR imagery.

4. METHODOLOGY

Our methodology involves generating infrared images of military tanks from a custom scenario created using the Stage application, which simulates realistic environments and military operations. The scenario includes various tank models and a UAV for surveillance. Real-time data from the simulation is processed using the Ondulus IR library to generate infrared images across multiple spectral bands: NIR, SWIR, MWIR, and LWIR. These images are then stored in JPEG format and manually annotated for object detection tasks. The dataset is split into training, validation, and test subsets, with models trained using transfer learning on the YOLOv8, YOLOv9, YOLO Ghost P2, and RT-DETR models. The performance evaluation of the models is based on box loss, generalized intersection over union (GIoU) loss, and mean average precision (mAP50-95) metrics across different infrared spectra. The remaining part of this section provides detailed information about scenario creation, IR image generation and annotation, model training, and performance evaluation metrics.

4.1. Scenario Creation

Since our goal is to provide an analysis of object detection across different infrared spectra, especially for recognizing military tanks, which are crucial for strategic advancement, we need IR images at different wavelengths for the same view. However, such datasets are not publicly available that contain tank image scenes with the same viewing angle in different infrared bands. Therefore, we first generated a scenario that consists of a realistic simulation of military tanks and an UAV using the Stage application to create the necessary data. Stage application is a powerful tool designed for creating realistic and complex scenarios commonly used in military simulations. This software allows users to create dynamic and natural environments for air, land, and sea operations (Presagis, 2024a).

Figure 1 shows an aerial view of the scenario we generated. The scenario includes 3 Challenger2, 1 Challenger3, 1 Leopard2, 1 M1A3, 1 Merkava Mark, 3 T-72, 4 T-84 battle tanks, along with 1 MQ-9A Predator-B class UAV. In the scenario, tanks with predefined routes, speeds, and maneuvers perform their movements in the field while the MQ-9A Predator-B UAV, flying in the sky, conducts reconnaissance and

surveillance activities to track the tanks. Additionally, in order to process scenario information in real-time when the scenario is executed, a Distributed Interactive Simulation (DIS) gateway setting has been configured on the Stage software, and the Stage software has been set up to send scenario data in real-time via the specified IP (Open DIS, 2024).

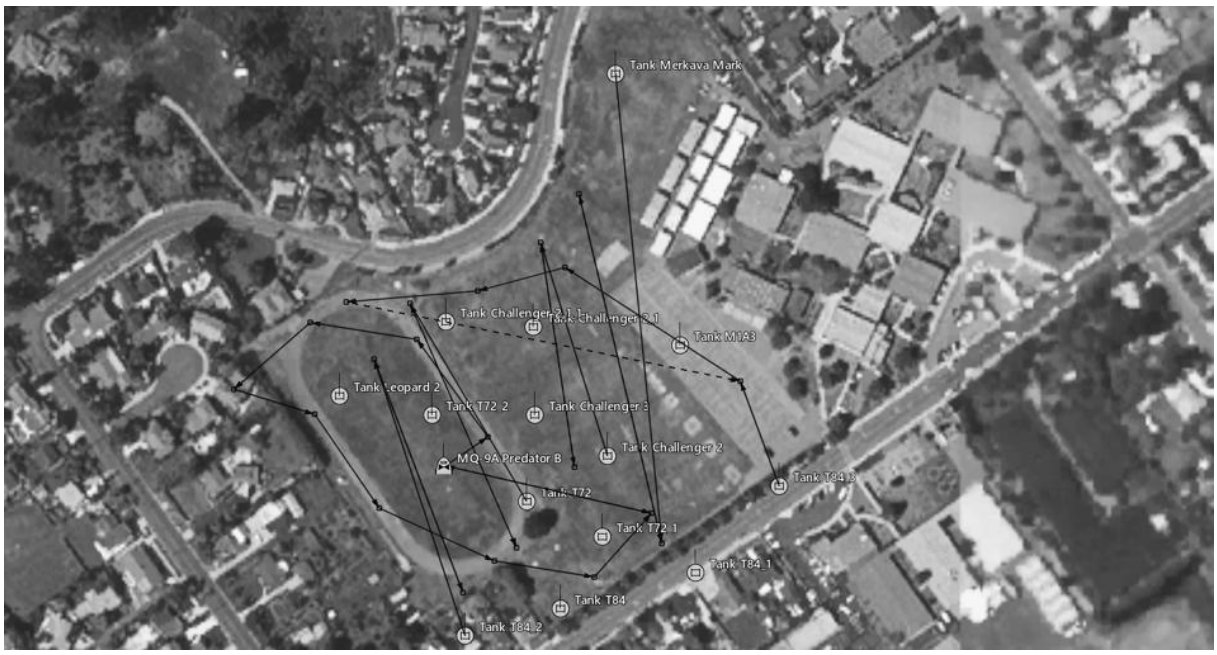


Figure 1. *Scenario creation*

4.2. IR Image Generation

IR image generation involves capturing images using infrared radiation emitted or reflected by objects, which is beyond the detection capability of the human eye. These images represent the infrared radiation emitted by objects. Specialized infrared cameras and sensors detect the emitted heat, and once the radiation is captured, it is converted into electrical signals. These signals are directly proportional to the intensity of the detected infrared radiation. Due to the limited availability of public datasets, particularly for military tank IR images captured by UAVs, we generated synthetic infrared images that closely mirror the real-world scenario of tanks from the custom scenario we created.

To generate infrared images at different wavelengths of the same view, we used three software libraries in the custom desktop application we integrated. Since the DIS software library is used for real-time communication with the Stage application, we rely on the Ondulus IR library to process real-time asset information and generate the infrared image of the simulated sensor based on the scenario's current environmental conditions (Presagis, 2024b). We chose these because they play a critical role in simulations, particularly for military training. They enable the creation of complex mission scenarios and detailed analysis using real-time, high-quality sensor data. Finally, the OpenCV library is used to capture and save the infrared images generated by Ondulus IR in JPEG format (OpenCV, 2024). This allows for the storage of images and facilitates further analysis and integration into different systems.

Figure 2 presents a high-level view of our IR image generation process. From our scenario created by the Stage application, we obtain entity information such as tanks, UAVs, and time via the DIS Gateway to use this information in our custom application. Then, our application generates infrared images based on real-time asset data and the scenario's current conditions using the Ondulus IR library and saves these images in JPEG format, as illustrated in the figure. Overall, we generated 2100 IR images for each spectral range - NIR, SWIR, MWIR, and LWIR - each depicting the same scene across the different spectra, resulting in a total dataset of 8400 images. These images were generated using a shared simulation scenario, in which platform trajectories, target configurations, viewpoints, and environmental conditions were kept identical. Within this scenario, different infrared sensors for each band were simulated independently using the Ondulus IR Sensor Simulator, adjusting only the band-specific sensor and spectral parameters. This approach ensures that all images share the same geometry, viewing conditions, and scene dynamics, differing only in their infrared spectral characteristics, thereby enabling a fair comparison across infrared bands. During the generation of the IR image dataset, no additional augmentation was applied to the infrared images, as the synthetic data generated in our study was already diverse and closely resembled the data captured by real UAVs in operational environments. The generated IR images were then manually annotated to identify and label tank objects within them.

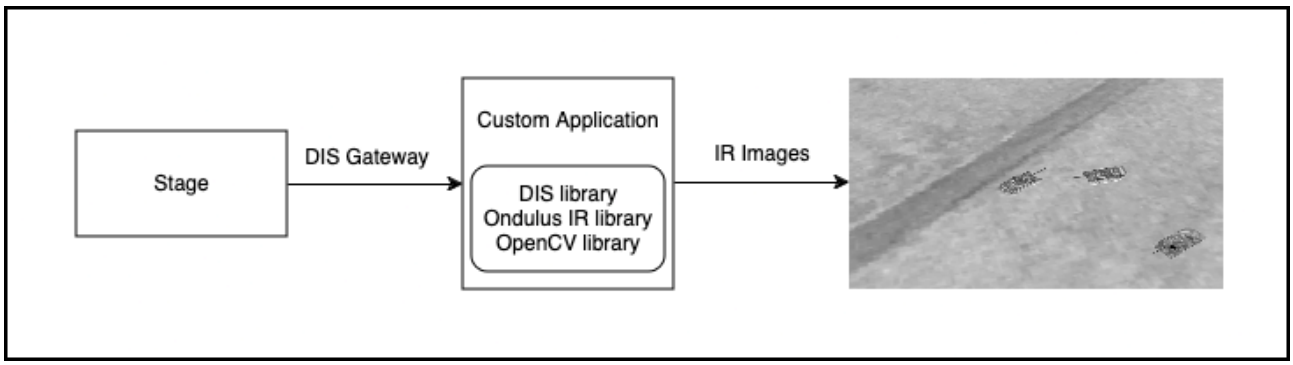


Figure 2. *IR image generation*

4.3. Experimental Design

We divided our dataset of generated infrared images into training, validation, and test subsets. Specifically, 70% of the images were randomly allocated for training, 15% for validation, and 15% for testing. This 70-15-15 split is a commonly adopted practice in deep learning, providing a balanced distribution between model training, evaluation, and testing, while also being well-suited to the size of our dataset. Specifically, for each infrared band, this resulted in 1,500 images for training, 300 images for validation, and 300 images for testing.

To train our models, we configured the computer with the Windows 11 operating system, an AMD Ryzen 5 5500 CPU, and an Nvidia RTX 3060 GPU to efficiently process and analyze the generated IR images. During training, we installed PyTorch 2.2.1 and CUDA 12.1 in virtual Anaconda environments on the Nvidia RTX 3060 GPU to facilitate the training process. This setup ensured that a suitable software environment was prepared for training the models on the GPU.

4.4. Model Training

To adapt the features learned from a visible light dataset to the unique characteristics exhibited by infrared images, a training approach utilizing transfer learning is used during the model training phase. This approach not only shortens training times but also considerably reduces the consumption of system resources during the training process. As a result, pre-trained models on the COCO dataset with visible light data have been selected for fine-tuning, specifically YOLOv8 series models such as YOLOv8m and YOLO Ghost P2, as well as RT-DETR series model rtdetr-l and YOLOv9 series model YOLOv9c.

The reason for selecting these three models from the relevant YOLO series is their comparable number of parameters. YOLOv8m has approximately 25.9 million parameters, while YOLOv9c has about 25.3 million parameters. Similarly, the YOLO Ghost P2 model also has around 25.9 million parameters. This similarity facilitates an equitable comparison among the models. Additionally, the inclusion of the RT-DETR model in the training process is due to the need to evaluate the performance of infrared bands in a transformer-based object detection model. Another factor in selecting these four models is their alignment with the hardware capacity constraints for training.

During training, we set batch=-1, val=True, lr0=0.01, and lrf=0.01, while all other hyperparameters were kept at their default values. Setting batch=-1 allows the batch size to be automatically determined based on GPU memory, maximizing resource utilization. Enabling val=True ensures that the model is periodically evaluated on the validation dataset, providing performance feedback during training. The initial learning rate, lr0=0.01, controls the speed at which model weights are updated at the start of training, while lrf=0.01 gradually reduces the learning rate to 1% of its initial value by the end of training, facilitating stable convergence. To ensure reproducibility and avoid potential bias introduced through manual hyperparameter tuning, we consistently used the default, out-of-the-box configurations provided by the official YOLO and RT-DETR frameworks, including standard settings for input resolution, learning rate schedules, and optimizers. While these default hyperparameters may not be optimal for our simulated military tank dataset, they enable a fair comparison of the models, since these settings were originally optimized for different datasets and tasks.

To facilitate a comprehensive comparison of the performance of object detection models on infrared bands, the selected models were trained for 100 epochs across all infrared bands using simulated military tank images. The training of the RT-DETR model for 100 epochs took approximately 1.8 times longer than that of the YOLO-based models, which completed the same training in a shorter duration. Additionally, to observe the impact of the number of epochs on performance, the top-performing model is subjected to training sessions with different numbers of epochs.

Figure 3 presents examples of generated infrared images with different spectra, illustrating the diversity within our dataset used for model training and performance evaluation. Each image represents a unique spectral range, enabling a comprehensive evaluation of the models' capabilities across various IR bands. The figure shows

NIR in the upper left, SWIR in the upper right, MWIR in the lower left, and LWIR in the lower right, ordered by wavelength from low to high.

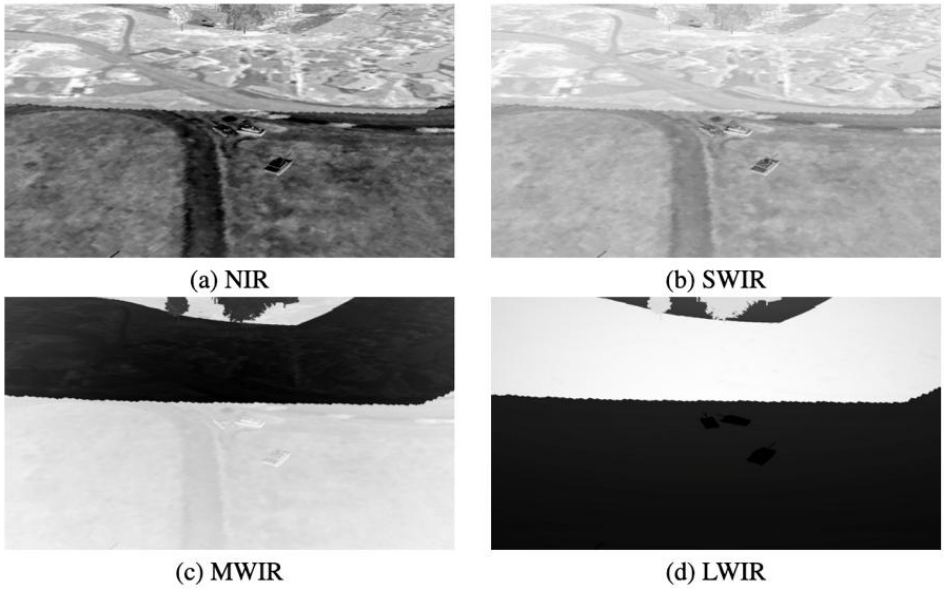


Figure 3. *Examples of generated IR images with different spectra*

A key consideration is the inherent modal divergence between visible-spectrum (RGB) imagery and infrared data, which ideally would be addressed through pre-training on large-scale infrared datasets to reduce the resulting domain gap. However, the lack of publicly available datasets covering all four evaluated infrared bands (NIR, SWIR, MWIR, LWIR) prevents effective domain-specific pre-training. Moreover, training high-capacity models such as YOLOv8 and RT-DETR from scratch on our limited datasets (1,500 images per band) would likely lead to severe overfitting and unstable convergence. Therefore, COCO pre-training was adopted as a practical initialization strategy. Despite the modality gap, it provides useful low-level feature representations (e.g., edges, contours, and textures) that improve convergence stability and facilitate effective fine-tuning on infrared imagery.

4.5. Training and Evaluation Metrics

In this study, we use three metrics to evaluate model accuracy on generated IR images of tanks from the same scene.

Box Loss Value: The loss function is a method that measures the difference between the predictions made by the model and the actual bounding box. Box loss is a loss function used in object detection algorithms to ensure that the predicted bounding boxes are accurately positioned and sized (He et al., 2019). It sums the errors for each example in the training or validation sets. If the model makes a perfect prediction, the loss value is zero; any deviation from perfection increases the loss. A high loss value usually indicates that the model's outputs are incorrect, while a low loss value suggests that the model is making fewer mistakes. During the model's training process, efforts are made to minimize this loss value, thereby making the positions and sizes of the detected objects more accurate.

Generalized Intersection over Union (GIoU) Loss: This loss function is an extension of the classic Intersection over Union (IoU) metric used to improve the performance of object detection models in the field of computer vision. The classic IoU calculates the overlap ratio between two object bounding boxes by providing the ratio of the intersection area to the union area; however, it does not have a meaningful gradient in cases where the boxes do not touch each other. In contrast, GIoU takes into account the smallest enclosing box between the two boxes in addition to the classic IoU, aiming to minimize the difference between the boxes even when the IoU is zero. This enhances the model's learning process and allows for more accurate positioning of the bounding boxes.

Mean Average Precision (mAP50-95): This metric calculates the average precision by varying the IoU threshold values between 50% and 95%. This metric evaluates how well the model performs in scenarios requiring both low and high threshold values. For example, mAP50 represents the average precision calculated at a 50% threshold, while mAP95 is calculated at a 95% threshold. mAP50-95 averages the performance across different thresholds, measuring the overall accuracy and consistency of the model in object detection. This metric is critical for the development and comparison of object detection models, as it reveals how the model performs not just at a specific threshold, but across a broad range (Reis et al., 2024). To calculate the mAP50-95, the precision and recall must be computed at various confidence thresholds. While precision measures how many of the examples predicted as positive by the model are actually positive, recall assesses how many of the true positive examples the model correctly identifies as positive.

5. RESULT AND DISCUSSION

This section presents our findings on the evaluation of the deep learning models trained on different IR bands to assess their performance in detecting tanks in IR imagery for military applications. We discuss the results based on box value loss, GIoU, and mAP at thresholds ranging from 50 to 95. Each of these metrics provides valuable insights into the models' performance. Box value loss helps us understand how accurately the models predict bounding boxes around objects, while GIoU measures the quality of these predictions by taking into account both the area of overlap and the overall size of the bounding boxes. Lastly, mAP at various thresholds allows us to evaluate the models' precision and recall across different levels of confidence, offering a comprehensive view of their effectiveness in detecting objects within infrared images.

5.1. Box Loss Value

We examined the performance of the YOLOv8, YOLO Ghost P2, and YOLOv9 models by analyzing the box loss values across different infrared bands (NIR, SWIR, MWIR, LWIR) over 100 epochs. This evaluation not only highlights the comparative effectiveness of each model but also provides insights into how various IR spectral bands influence object detection performance.

Figure 4 shows the results for YOLOv8, YOLO Ghost P2, and YOLOv9, with each model shown in a separate plot. In each plot, the y-axis represents the box loss value, ranging from 0.2 to 1.7, and the x-axis represents

the number of epochs, ranging from 5 to 100. Based on the results, NIR consistently shows the highest performance across all models, making it the most effective choice for our applications. It is followed by SWIR, then MWIR, and finally LWIR. This trend indicates that as we move to longer infrared wavelengths, the effectiveness of detection decreases, highlighting the challenges associated with using the longer IR spectrum for our study. For the MWIR in the YOLO Ghost P2 and YOLOv9 models, the box loss values start higher than those for LWIR, indicating that initially, MWIR may face more challenges in accurately detecting military tanks compared to LWIR. However, as training progresses, the performance of MWIR may improve, leading to a decrease in box loss values and better performance compared to LWIR. In contrast, this trend is not observed in YOLOv8, which highlights the unique performance characteristics of YOLOv8 in handling various spectral bands, particularly MWIR. In addition, the performance of NIR and SWIR is more closely aligned compared to the other spectral bands.

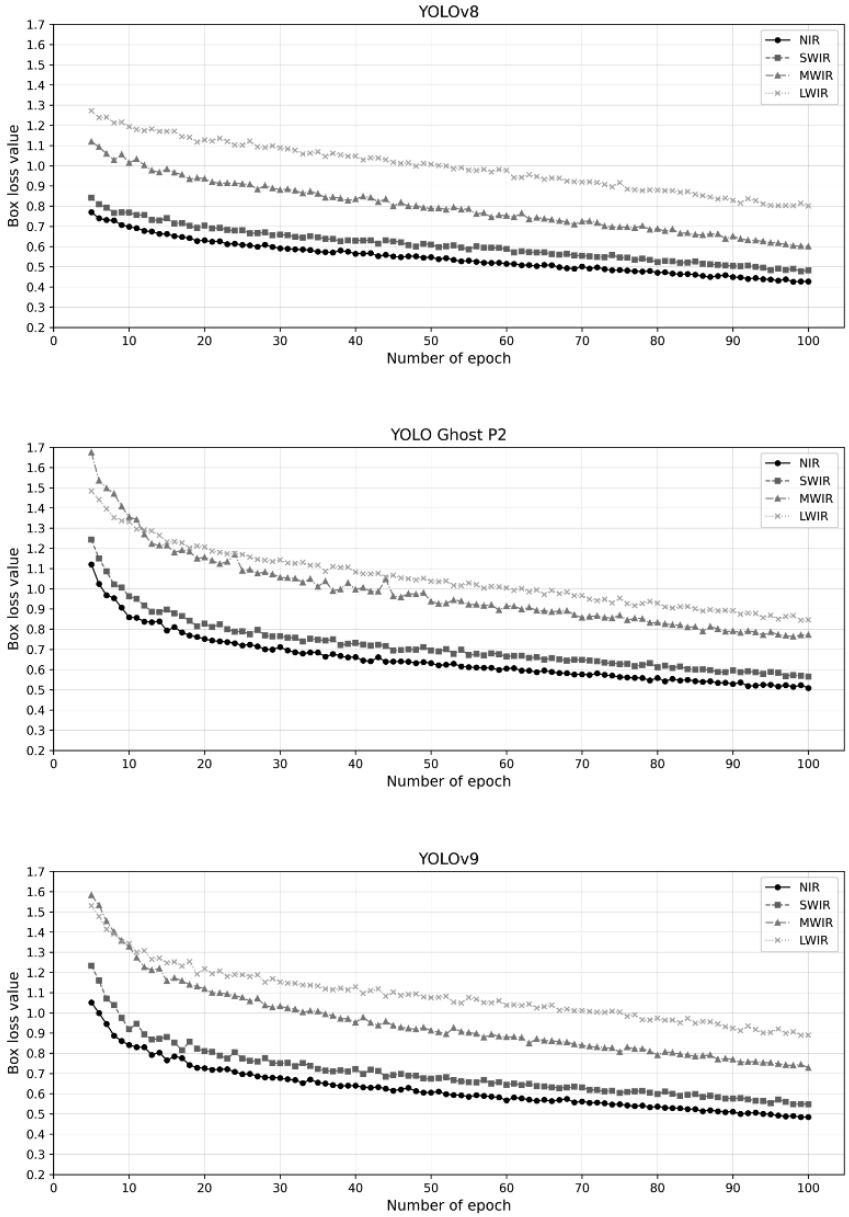


Figure 4. *Comparison of the box loss values over 100 epochs*

To compare model performance in the IR bands, we calculated the average values from the last 10 epochs, specifically from epochs 90 to 100, as all models demonstrated similar trends across the different IR bands, allowing for an effective comparison. Table 1 shows the average box loss values for each model across the various IR spectra, providing a clear overview of their performance in military object detection tasks. Based on the data, the YOLOv8 model shows better performance in all infrared bands (NIR, SWIR, MWIR, LWIR). It has the lowest average box loss values, which means it is more effective at detecting tanks compared to the other models. This highlights that newer versions of YOLO are not always guaranteed to have better performance.

Table 1. *Comparison of YOLO models across spectral bands (Average box loss value)*

Model	NIR	SWIR	MWIR	LWIR
YOLOv8	0.437	0.493	0.621	0.813
YOLO Ghost P2	0.522	0.581	0.778	0.865
YOLOv9	0.496	0.562	0.750	0.909

5.2. GIoU Loss

As RT-DETR leverages a transformer-based architecture, it utilizes the GIoU loss metric to analyze performance across IR bands, unlike YOLO-based models. YOLO-based models use the box loss value because their structure predicts bounding box offsets relative to predefined positions, while RT-DETR uses GIoU loss, as it treats detection as a set prediction problem in which each query is matched to a ground truth box. GIoU is used specifically during training rather than for evaluation, as it provides meaningful gradients that help the model improve bounding box predictions, even when the predicted boxes do not overlap with the ground truth. Consequently, RT-DETR's transformer-based architecture generally requires a longer training time. Figure 5 illustrates the GIoU loss values for the RT-DETR model over 100 epochs across each spectral band.

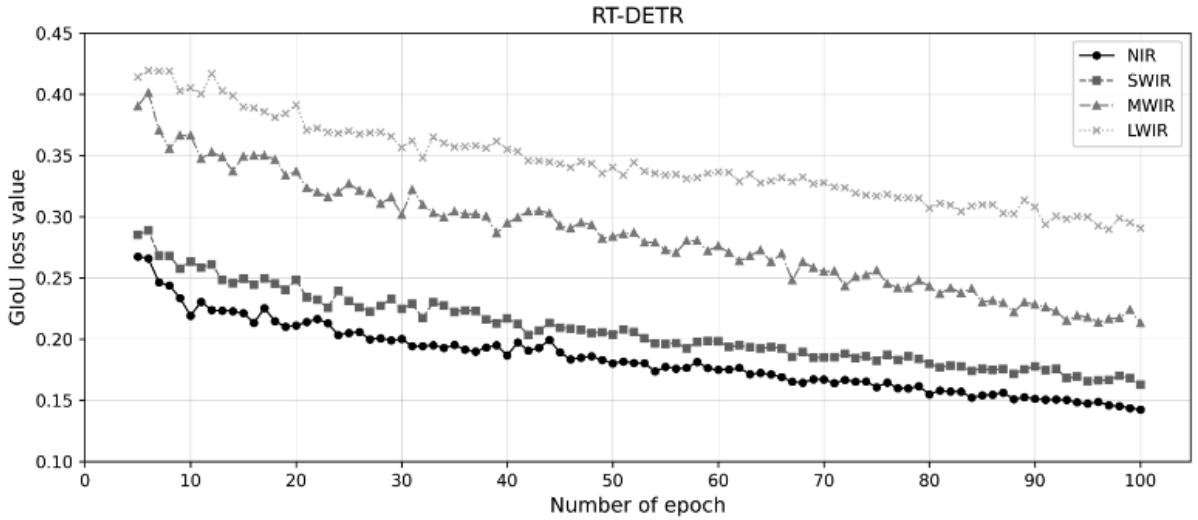


Figure 5. *GIoU loss values over 100 epochs*

As shown in the figure, the loss curves across the IR bands exhibit a similar pattern when compared to the YOLO-based models, following the sequence of NIR, SWIR, MWIR, and LWIR. While the absolute loss values of the YOLO-based models and RT-DETR are not directly comparable due to their differing definitions, their relative convergence behaviors provide meaningful insights. RT-DETR exhibits a smoother decay and reaches a steady state more rapidly than the YOLO-based models, suggesting a more stable optimization landscape across the IR bands. This comes at the cost of higher computational demand, with approximately 1.8× greater latency.

Comparing models optimized with different loss functions (e.g., Box Loss in YOLO variants versus GIoU in RT-DETR) introduces a limitation in achieving strict algorithmic parity during convergence analysis. While standardizing the optimization objective across all models could provide a more uniform baseline, we intentionally evaluate each architecture using its natively optimized loss configuration. The loss formulations in modern YOLO architectures are highly tuned (e.g., CIoU and Distribution Focal Loss), and replacing them with GIoU could disrupt their intended optimization dynamics and lead to non-representative performance. Accordingly, our convergence analysis focuses on relative optimization stability and decay trends rather than absolute loss values. For fair and quantitative comparison of detection performance, we consistently rely on standardized mAP metrics across all models.

5.3. mAP50-95

To evaluate the models, we used two different testing approaches. Initially, the performance of each model on the images from its specific IR band was measured, allowing us to understand how effective each model was in its specialized band. After this initial evaluation, a common test dataset was created from the test data of each infrared band. This newly formed test set was constructed by combining data from different bands, and the performance of all models was examined on this dataset. This way, the models were evaluated not only on the data from their specific bands but also on the overall performance in IR images. Evaluations using the general test set showed how effective the models were across a wide variety of IR images. This allowed us to see not only how well the trained models performed in their specific bands but also in general IR image processing tasks.

Table 2 presents the mAP50-95 values for the models trained for 100 epochs and tested across their respective bands. The YOLOv8 model outperforms the others in all bands, achieving the highest mAP50-95 scores of 91.2 in NIR, 90.0 in SWIR, 87.5 in MWIR, and 79.6 in LWIR. This indicates that it is particularly effective for tasks requiring accurate detection across these spectral bands. However, despite its lower GIoU loss, RT-DETR exhibits slightly lower mAP50-95 values across all spectral bands compared to the YOLO models. This suggests that while RT-DETR excels in bounding box accuracy, the YOLO models may still offer better overall detection performance in terms of mAP metrics. Interestingly, YOLO Ghost P2 is known for being lightweight and built on earlier versions, while also offering competitive performance compared to YOLOv9. Furthermore, all models demonstrate a decline in performance from the NIR to LWIR bands. This trend highlights the

challenges associated with detecting military tanks at longer infrared wavelengths. The decrease in mAP scores, especially pronounced in LWIR, indicates a potential area for further optimization and improvement.

Table 2. *mAP50-95 values for respective test dataset*

Model	NIR	SWIR	MWIR	LWIR
YOLOv8	91.2	90.0	87.5	79.6
YOLO Ghost P2	90.8	88.5	85.1	79.1
YOLOv9	90.0	88.0	86.4	78.0
RT-DETR	88.4	87.1	80.0	73

In a similar manner, Table 3 presents the mAP50-95 values obtained from a common test dataset encompassing the entire IR spectrum by testing the YOLOv8, YOLOv9, YOLO Ghost P2, and RT-DETR models, all trained for 100 epochs. The columns (NIR, SWIR, MWIR, LWIR) represent the training band of each model, while all evaluations are conducted on the same combined test dataset. As expected, model performance decreases when comparing the tests across their respective bands and the common dataset. However, it is clear that RT-DETR performs better on the common dataset, demonstrating higher mAP values compared to YOLOv9 and YOLO Ghost P2. Additionally, YOLOv8 does not have the best performance for LWIR, and this is the only case where it falls short compared to other models.

Table 3. *mAP50-95 values for the common test dataset*

Model	NIR	SWIR	MWIR	LWIR
YOLOv8	81.7	76.8	76.5	32.8
YOLO Ghost P2	76.8	67.9	68.3	41.0
YOLOv9	77.7	69.2	66.4	37.8
RT-DETR	78.4	72.5	70.7	46.7

5.4. Epoch Impact

To investigate the impact of epochs on the accuracy of detecting military tanks in IR images, we analyzed the mAP50-95 values for YOLOv8, which demonstrated the best performance among the models considered. We determined the number of epochs by examining the box loss values for YOLOv8, which had almost stopped changing at 400 epochs, indicating that the model had largely converged. Accordingly, we conducted training with 100, 200, 300, and 400 epochs to study how training duration influences detection accuracy across different IR bands. It should be noted that this analysis was not intended as hyperparameter tuning using the validation set, but rather to observe training behavior and convergence characteristics.

Table 4 and Table 5 show the mAP50-95 values across different epochs for the respective test dataset and the common test dataset of the YOLOv8 model for different IR spectra, respectively. Table 4 demonstrates that increasing the number of training epochs generally leads to improved mAP values across all infrared bands, highlighting the effectiveness of training duration in enhancing model performance. In contrast, Table 5 reveals

that while the NIR and SWIR bands show improvement with increased epochs, the MWIR band experiences variability, and the LWIR band suffers a notable decline in performance. This indicates that while additional training may benefit some bands, it could lead to overfitting or performance issues in others, particularly in the LWIR band for the common test dataset. The observed performance degradation and rapid overfitting in the LWIR spectrum can be attributed to both physical and technical factors. Physically, unlike shorter wavelengths (NIR/SWIR), which capture reflected radiation and preserve geometric texture, LWIR (8–12 μm) is dominated by self-generated radiation from objects, resulting in images with low spatial frequency and limited structural detail. Technically, this shallow feature space allows deep learning models to quickly learn the available representations during early epochs. However, with prolonged training, the network exhausts generalizable features and begins to memorize dataset-specific noise and background artifacts, leading to a noticeable decline in mAP on the test set.

Table 4. *mAP50-95 values across different epochs for the respective test dataset of YOLOv8*

Epoch	NIR	SWIR	MWIR	LWIR
100	91.2	90.0	87.5	79.6
200	91.4	90.4	88.1	83.1
300	91.8	90.6	88.2	83.7
400	92.6	91.3	90.1	84.1

Table 5. *mAP50-95 values across different epochs for the common test dataset of YOLOv8*

Epoch	NIR	SWIR	MWIR	LWIR
100	81.7	76.8	76.5	32.8
200	81.7	77.2	75.2	31.4
300	81.9	76.5	76.4	30.7
400	83.7	81.5	79.9	22.9

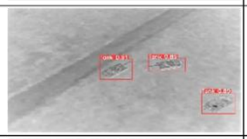
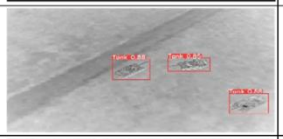
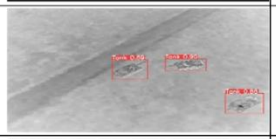
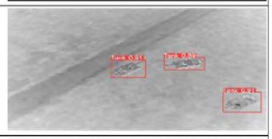
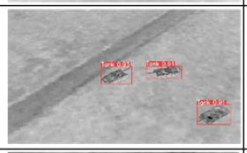
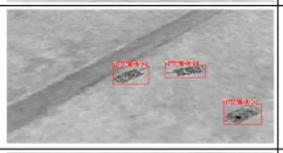
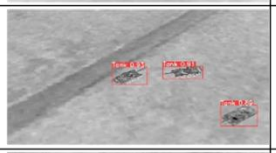
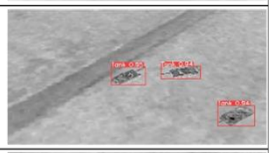
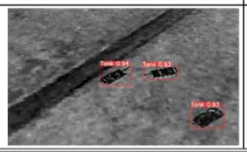
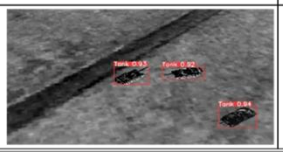
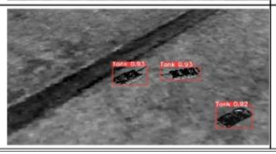
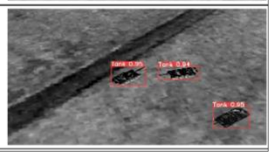
5.5. Qualitative Error Analysis

To gain a deeper understanding of the behavioral differences between the evaluated architectures, we conducted a qualitative analysis of the detection results. Table 6 presents a visual matrix comparing the performance of YOLOv8, YOLOv9, YOLO Ghost P2, and RT-DETR across the LWIR, MWIR, SWIR, and NIR spectral bands. This visual assessment clarifies how each model processes the unique characteristics of infrared imagery and adapts to the progressive loss of spatial information.

As illustrated in Table 6, environmental context and target signatures degrade significantly as the wavelength increases. In the NIR and SWIR bands, background terrain and road structures are clearly visible, providing rich spatial features. In these conditions, CNN-based architectures—particularly YOLOv8—demonstrate

excellent bounding box precision and high confidence scores, leveraging their localized feature extraction capabilities. However, in the LWIR band, background texture is largely lost, and the visual information is reduced to thermal emissions (i.e., hot spots against a dark background).

Table 6. *Tank Detection Examples on the Same Scene*

	YOLOv8	YOLOv9	YOLO Ghost P2	RT-DETR
LWIR				
MWIR				
SWIR				
NIR				

While all models successfully detect targets in this sample, their behavioral differences become more apparent under these challenging conditions. YOLO-based models, relying heavily on localized spatial gradients, exhibit slight variations in bounding box tightness and confidence as local textures diminish in the LWIR spectrum. Conversely, the Transformer-based RT-DETR utilizes global attention mechanisms to process the entire image context. This architectural advantage allows RT-DETR to maintain robust detection performance and stable confidence levels even when local geometric details are absent, although it can sometimes produce slightly less tightly fitted bounding boxes compared to YOLOv8. These qualitative observations are consistent with our quantitative findings, highlighting the fundamental trade-off between the localized precision of YOLO and the global contextual modeling capabilities of Transformer-based architectures.

6. THREATS TO VALIDITY

One of the most significant threats to validity is the lack of existing datasets. While we created our own dataset, limited access to tools such as the STAGE application and Ondulus IR restricts our ability to generate a wider range of scenarios. Additionally, inconsistencies in the scenarios and environmental conditions simulated within the STAGE environment may introduce variability, despite efforts to construct realistic settings. Certain environmental factors may still have been overlooked, which could impact the generalizability of the results.

Although synthetic infrared images generated from simulated scenarios enable controlled experimentation, this approach may not fully represent real-world IR imaging conditions. Synthetic data inherently introduces a domain gap due to simplified physical modeling and the absence of real sensor noise, atmospheric variability,

and environmental complexity. As a result, models trained on such data may not generalize effectively to real operational environments, and the limited diversity of synthetic scenes may bias the learning process toward unrealistic patterns.

Furthermore, the STAGE simulation environment, while enabling structured IR data generation across spectral bands, is constrained by simplified thermal and physical modeling assumptions. It does not explicitly incorporate environmental noise sources such as sensor-level noise, atmospheric scattering, weather effects (e.g., fog, rain, humidity), or diurnal thermal cycles, including sunrise and sunset crossover effects. These omissions may lead to a relatively sterile dataset that does not fully capture real-world infrared imaging complexity. Consequently, a domain gap between synthetic and operational data may exist, potentially limiting the generalization performance of both YOLO-based and RT-DETR models in real deployment scenarios.

Another potential limitation is the exclusive focus on tank objects during evaluation, which may reduce the applicability of the results to other types of military targets. Additionally, all images are generated from a single simulation scenario, which may lead to temporally adjacent frames appearing across the training, validation, and test splits. To preserve alignment of corresponding scenes across different infrared bands, a fixed random seed is used, ensuring a fair and consistent comparison of model performance despite this limitation.

7. CONCLUSION

Object detection is crucial in modern military operations, enhancing decision-making and situational awareness for tasks like surveillance and autonomous navigation. Challenges arise in complex environments with poor visibility and distinguishing targets from background noise. Infrared imaging helps, particularly in detecting enemy tanks using UAVs, a key aspect of modern defense. However, research comparing performance across different infrared spectra in military applications is limited due to the scarcity of public datasets and security concerns around military infrared data.

In this study, we thoroughly examined how well different deep learning models can detect military tanks using various types of IR images, including YOLOv8, YOLOv9, YOLO Ghost P2, and RT-DETR. We not only assessed the performance of these models but also identified which IR band is most suitable for military tank recognition. By training and testing the models on different bands, we aimed to determine optimal detection conditions and gain insights for military applications. Combining data from all bands gave us a comprehensive view of model performance. We also explored how training durations (100, 200, 300, and 400 epochs) impacted model accuracy, revealing how training time affects tank detection.

Our analysis revealed several key insights regarding the performance of deep learning models for military tank detection in IR images. Firstly, across all performance metrics -- box loss value, GIoU, and mAP50-95 -- the models trained on NIR data demonstrated the best results. Among the models evaluated, YOLOv8 proved to

be the most effective for military tank detection in the IR spectrum, also showing that newer models do not always guarantee better performance, as evidenced by YOLOv8 consistently outperforming YOLOv9. We also noticed that additional training may not always yield further benefits and can lead to variability in certain infrared bands, particularly in the LWIR band, when evaluated on a common test dataset. Additionally, we found that as the wavelength increases in the infrared spectrum, detection effectiveness decreases. Finally, the transformer-based RT-DETR model needed about 1.8x more training time than the YOLO-based models, even when both were trained for 100 epochs.

These findings provide actionable insights for military applications. Specifically, NIR sensors are recommended for UAV-based tank detection due to their superior performance. When selecting sensors, trade-offs such as cost, availability, and operational conditions across different IR bands need to be considered. Moreover, YOLOv8 offers a strong balance of accuracy and computational efficiency, making it suitable for real-time UAV operations.

As such, we believe that our study is important in the field of military applications, particularly due to the lack of comprehensive datasets. By providing a detailed analysis of different deep learning models across various IR bands, we contribute valuable insights that can help improve detection capabilities, particularly for military tanks. Our findings can help direct future research, leading to better military surveillance and reconnaissance efforts.

Future research may explore how seasonal changes like snow and leaf cover affect IR images and detection performance. A qualitative analysis of challenging samples can highlight the approach's strengths and limitations. Additionally, integrating data from multiple sensors (e.g., visible spectrum, radar, LIDAR) could improve detection accuracy by leveraging complementary information.

AUTHOR CONTRIBUTIONS

Y.F.Y. and C.S. designed the study and methodology. Y.F.Y. collected the data and performed the analysis. C.S. supervised the research process. Both authors contributed to the overall structure of the manuscript, made revisions, and finalized it for publication.

AI Exclusion: AI tools are strictly prohibited from being listed as authors.

Legal Consent: All human authors confirm they have reviewed and legally accepted the final version under the **CC BY-SA 4.0** framework.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Aboalia, H., Hussein, S., & Mahmoud, A. (2024). *Infrared multi-object detection using deep learning*. In: Proceedings of the 14th International Conference on Electrical Engineering (ICEENG) (21-23 May 2024, pp. 175-177), Cairo, Egypt. https://doi.org/10.1109/ICEENG58856.2024.10566390

Aibibu, T., Lan, J., Zeng, Y., Lu, W., & Gu, N. (2024). An efficient rep-style Gaussian–Wasserstein network: Improved UAV infrared small object detection for urban road surveillance and safety. *Remote Sensing*, *16*(1), 25. https://doi.org/10.3390/rs16010025

Aziz, L., Haji Salam, Md. S. B., Sheikh, U. U., & Ayub, S. (2020). Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review. *IEEE Access*, *8*, 170461-170495. https://doi.org/10.1109/ACCESS.2020.3021508

Bao, C., Cao, J., Hao, Q., Cheng, Y., Ning, Y., & Zhao, T. (2023). Dual-YOLO architecture from infrared and visible images for object detection. *Sensors*, *23*(6), 2934. https://doi.org/10.3390/s23062934

EMF Corporation. (2024). *Optical material for infrared optics*. URL

Ghenescu, V., Barnoviciu, E., Carata, S.-V., Ghenescu, M., Mihaescu, R., & Chindea, M. (2018). *Object recognition on long range thermal image using state of the art DNN*. In: Proceedings of the Conference Grid, Cloud & High Performance Computing in Science (ROLCG) (17-19 October 2018, pp. 1-4), Cluj-Napoca, Romania. https://doi.org/10.1109/ROLCG.2018.8572026

Gomaa, A. (2024). *Advanced domain adaptation technique for object detection leveraging semi-automated dataset construction and enhanced YOLOv8*. In: Proceedings of the 6th Novel Intelligent and Leading Emerging Sciences Conference (NILES) (19-21 October 2024, pp. 211-214), Giza, Egypt. https://doi.org/10.1109/NILES63360.2024.10753164

Gomaa, A., Abdelwahab, M. M., & Abo-Zahhad, M. (2020). Efficient vehicle detection and tracking strategy in aerial videos by employing morphological operations and feature points motion analysis. *Multimedia Tools and Applications*, *79*(35-36), 26023-26043. https://doi.org/10.1007/s11042-020-09242-5

He, Y., Zhu, C., Wang, J., Savvides, M., & Zhang, X. (2019). *Bounding box regression with uncertainty for accurate object detection*. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (15-20 June 2019, pp. 2883-2892), Long Beach, California, USA. https://doi.org/10.1109/CVPR.2019.00300

Hu, J., Zhao, Y., & Zhang, X. (2020). *Application of transfer learning in infrared pedestrian detection*. In: Proceedings of the IEEE 5th International Conference on Image, Vision and Computing (ICIVC) (10-12 July 2020, pp. 1-4), Beijing, China. https://doi.org/10.1109/ICIVC50857.2020.9177438

Jinke, S. (2023). *Enhancing object detection in infrared videos through temporal and spatial information* MSc Thesis, KTH, School of Electrical Engineering and Computer Science.

Lagman, J. K. D., Evangelista, A. B., & Paglinawan, C. C. (2022). *Unmanned aerial vehicle with human detection and people counter using YOLO v5 and thermal camera for search operations*. In: Proceedings of the IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS) (25 June 2022, pp. 113-118), Shah Alam, Malaysia. https://doi.org/10.1109/I2CACIS54679.2022.9815490

Lee, Y.-S., Chan, Y.-M., Fu, L.-C., & Hsiao, P.-Y. (2015). Near-infrared-based nighttime pedestrian detection using grouped part models. *IEEE Transactions on Intelligent Transportation Systems*, *16*(4), 1929-1940. https://doi.org/10.1109/TITS.2014.2385707

Li, L., Jiang, L., Zhang, J., Wang, S., & Chen, F. (2022). A complete YOLO-based ship detection method for thermal infrared remote sensing images under complex backgrounds. *Remote Sensing*, *14*(7), 1534. https://doi.org/10.3390/rs14071534

Li, S., Li, Y., Li, Y., Li, M., & Xu, X. (2021). YOLO-FIRI: Improved YOLOv5 for infrared image object detection. *IEEE Access*, *9*, 141861-141875. https://doi.org/10.1109/ACCESS.2021.3120870

Liu, H., Yu, Y., Liu, S., & Wang, W. (2022). A military object detection model of UAV reconnaissance image and feature visualization. *Applied Sciences*, *12*(23), 12236. https://doi.org/10.3390/app122312236

Luo, X., & Zhu, X. (2025). YOLO-SMUG: An Efficient and Lightweight Infrared Object Detection Model for Unmanned Aerial Vehicles. *Drones*, *9*(4), 245. https://doi.org/10.3390/drones9040245

Moradi, S., Moallem, P., & Sabahi, M. F. (2020). Fast and robust small infrared target detection using absolute directional mean difference algorithm. *Signal Processing*, *177*, 107727. https://doi.org/10.1016/j.sigpro.2020.107727

Open DIS. (2024). *Distributed interactive simulation*. URL

OpenCV. (2024). *OpenCV*. URL

Pavlović, M. S., Milanović, P. D., Stanković, M. S., Perić, D. B., Popadić, I. V., & Perić, M. V. (2022). Deep learning based SWIR object detection in long-range surveillance systems: An automated cross-spectral approach. *Sensors*, *22*(7), 2562. https://doi.org/10.3390/s22072562

Presagis. (2024a). *Ondulus IR*. URL

Presagis. (2024b). *Stage*. URL

Price, S. R., & Price, S. R. (2019). *iECO learned matched filters for automatic target recognition in synthetic midwave infrared imagery*. In R. I. Hammoud, & T. L. Overman (Eds.), Proceedings of the Automatic Target Recognition XXIX, (15-18 April 2019, Vol. 10988, p. 1098816), Baltimore, Maryland, United States. https://doi.org/10.1117/12.2518741

Reis, D., Kupec, J., Hong, J., & Daoudi, A. (2024). Real-time flying object detection with YOLOv8. https://doi.org/10.48550/arXiv.2305.09972

Sadic, N. M., Shalaby, W. A., El-Dolil, S., Abd El-Samie, F. E., Dessouky, M. I., & Elkaffas, S. M. (2025). Utilization of infrared images for object detection: a survey. *Journal of Optics*. https://doi.org/10.1007/s12596-024-01868-0

Shah, S., & Tembhurne, J. (2023). Object detection using convolutional neural networks and transformer-based models: A review. *Journal of Electrical Systems and Information Technology*, *10*(1), 54. https://doi.org/10.1186/s43067-023-00123-z

Shchetynin, E. Yu. (2025). Model of object detection in infrared spectrum images obtained by unmanned aerial vehicles. *Journal of Applied Informatics*, *20*(1), 16-27. https://doi.org/10.37791/2687-0649-2025-20-1-16-27

Shen, L., Lang, B., & Song, Z. (2023). Infrared object detection method based on DBD-YOLOv8. *IEEE Access*, *11*, 145853-145868. https://doi.org/10.1109/ACCESS.2023.3345889

Sohan, M., Sai Ram, T., & Rami Reddy, Ch. V. (2024). *A Review on YOLOv8 and Its Advancements*. In: I. J. Jacob, S. Piramuthu, & P. Falkowski-Gilski (Eds.), Proceedings of the Data Intelligence and Cognitive Informatics (ICDICI 2023) (27-28 June 2023, pp. 529-545), Tirunelveli, India. Springer. https://doi.org/10.1007/978-981-99-7962-2_39

Soumya, D. S., Aishwarya, Ch., & Vasavi, S. (2023). *FPGA-based military vehicles detection and classification from drone videos using YOLOv5*. In: J. R. Szymanski, C. K. Chanda, P. K. Mondal, & K. A. Khan (Eds.), Proceedings of the 4th International Conference on Energy Systems, Drives and Automations (ESDA 2021) (30-31 December 2021, pp. 265-276), Kolkata, West Bengal, India. https://doi.org/10.1007/978-981-99-3691-5_22

Sun, Y., Meng, Y., Wang, Q., Tang, M., Shen, T., & Wang, Q. (2024). *Visible and infrared image fusion for object detection: A survey*. In: P. You, S. Liu, & J. Wang (Eds.), Proceedings of the International Conference on Image, Vision and Intelligent Systems 2023 (ICIVIS 2023) (16-18 August 2023, pp. 236-248), Baoding, China. https://doi.org/10.1007/978-981-97-0855-0_24

Tanchuk, V., & Kolobrodov, V. (2024). Analysis of existing object detection models for UAV detection on thermal imagery. *Bulletin of Kyiv Polytechnic Institute. Series Instrument Making*, *68*(2), 5-10. https://doi.org/10.20535/1970.68(2).2024.318080

Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023). A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction, 5*(4), 1680-1716. https://doi.org/10.3390/make5040083

Tham, M.-L., Cham, T. H., & Kwan, B.-H. (2023). *Detecting non-injured and injured humans in thermal imaging using YOLOv8*. In: Proceedings of the International Conference on Computer and Applications (ICCA) (28-30 November 2023, pp. 1–5), Cairo, Egypt. https://doi.org/10.1109/ICCA59364.2023.10401590

Tu, X., Yuan, Z., Liu, B., Liu, J., Hu, Y., Hua, H., & Wei, L. (2023). An improved YOLOv5 for object detection in visible and thermal infrared images based on contrastive learning. *Frontiers in Physics*, *11*. https://doi.org/10.3389/fphy.2023.1193245

Wang, X., Lu, R., Bi, H., & Li, Y. (2023). An infrared small target detection method based on attention mechanism. *Sensors*, *23*(20), 8608. https://doi.org/10.3390/s23208608

Wang, Y., Piao, Y., Wang, H., Zhang, H., & Li, B. (2024). An improved forest smoke detection model based on YOLOv8. *Forests*, *15*(3), 409. https://doi.org/10.3390/f15030409

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 (CVPRW) (14-19 June 2020, pp. 4448–4455), Seattle, WA, USA. https://doi.org/10.1109/CVPRW50498.2020.00524

Yang, Z., Yu, H., Zhang, J., Tang, Q., & Mian, A. (2025). Deep learning based infrared small object segmentation: Challenges and future directions. *Information Fusion*, *118*, 103007. https://doi.org/10.1016/j.inffus.2025.103007

Yucesoy, Y.F., & Sahin, C. (2024). Object detection in infrared images with different spectra. In: Proceedings of the International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) (23-25 May 2024, pp. 1–6), Istanbul, Turkiye. https://doi.org/10.1109/HORA61326.2024.10550753

Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., & Chen, J. (2024). *DETRs beat YOLOs on real-time object detection*. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (16-22 June 2024, pp. 16965–16974), Seattle, WA, USA. https://doi.org/10.1109/CVPR52733.2024.01605

Zhou, G., Liu, X., & Bi, H. (2025). Recognition of UAVs in Infrared Images Based on YOLOv8. *IEEE Access*, *13*, 1534-1545. https://doi.org/10.1109/access.2024.3500583