ISSN 1693-9220 (Print) | 2987-8713 (Online), Volume 20, No. 2 (2026), pp. 190-201

International Journal of Enterprise Modelling

Published by: International Enterprise Integration Association

Human object detection and classification system based on thermal cameras using the YOLOv11 object detection model

Muhammad Irsyaad Nurrahman, Bagus Hendra Saputra, H. A. Danang Rimbawa

Department of Electrical Engineering, Republic of Indonesia Defense University, Bogor, Indonesia

Abstract

Strategic institutions such as military campuses, defense research centers, and government facilities face increasingly complex security challenges, particularly in environments with low visibility and limited manual patrol capabilities. Conventional surveillance systems often perform poorly in dark environments because they depend heavily on visible light. Therefore, this research proposes a human object detection and classification system based on thermal cameras integrated with the YOLOv11 object detection model. Thermal cameras are capable of capturing heat radiation emitted by objects, enabling effective visualization under low-light and completely dark conditions. The proposed system combines thermal imaging technology with the real-time detection capability of YOLOv11 to automatically identify and classify human objects. This research employs the Research and Development (R&D) method, including dataset collection, image annotation, data augmentation, data preprocessing, model training, and system evaluation. The dataset consists of thermal images enhanced using augmentation techniques such as cropping, rotation, brightness adjustment, and blur effects to improve model robustness. Model performance was evaluated using Accuracy, Precision, Recall, F1-Score, and Confusion Matrix analysis. Experimental results demonstrate that the proposed system achieved an average accuracy of 86.36%, with accuracy values of 85.98% under completely dark conditions and 86.75% under dim-light conditions, indicating that the model is capable of reliably detecting and classifying human objects in low-visibility environments. These findings show that the integration of thermal cameras and YOLOv11 can contribute to the development of intelligent security systems that improve surveillance efficiency while reducing dependence on manual monitoring.

Article Info

Article history:

Received : Apr 15, 2026
Revised : May 23, 2026
Accepted : May 26, 2026

Keywords:

Computer Vision;
Deep Learning;
Human Detection;
Object Detection;
Thermal Camera;
YOLOv11.

Corresponding Author:

Muhammad Irsyaad Nurrahman,
Department of Electrical Engineering,
Republic of Indonesia Defense University,
Kawasan IPSC Sentul, Jl. Anyar, Sukahati, Kec. Citeureup,
Kabupaten Bogor, West Java 16810, Indonesia.
Ir.irsyaad@gmail.com

Introduction

The increasing complexity of security challenges in strategic institutions such as military campuses, government facilities, and defense research centers requires the implementation of advanced surveillance systems capable of operating effectively under low-light and extreme environmental conditions. Conventional monitoring systems that rely on visible light often experience significant limitations during nighttime operations, adverse weather conditions, or in areas with inadequate lighting infrastructure. These limitations reduce the effectiveness of manual patrol activities and increase the potential risk of undetected security threats. Therefore, the development of intelligent surveillance systems that can operate independently of visible light has become increasingly important.

Thermal camera technology offers an effective solution to overcome visibility limitations because it captures heat radiation emitted by objects regardless of lighting conditions (Gonzalez & Woods, 2018; Zhang et al., 2022). This capability enables thermal cameras to detect human presence even in completely dark environments. However, thermal imaging systems still require intelligent object detection algorithms capable of automatically identifying and classifying detected objects in real time with high accuracy and computational efficiency.

Recent advancements in deep learning and computer vision technologies, particularly object detection models based on the YOLO (You Only Look Once) architecture, have significantly improved real-time object detection performance with low latency and high accuracy (Redmon et al., 2016). Several previous studies have implemented earlier YOLO variants such as YOLOv5 and YOLOv8 for thermal-based human detection. Nevertheless, most of these studies focus primarily on general surveillance scenarios and visible-light datasets, while research involving the integration of thermal imaging with the latest YOLOv11 architecture in strategic security environments remains limited. In addition, previous studies often face challenges related to detection accuracy in low-contrast thermal imagery, real-time processing efficiency, and model robustness under varying environmental conditions.

YOLOv11 was selected in this research because it offers improvements in detection accuracy, lightweight architecture, and faster inference performance compared to previous YOLO versions, making it more suitable for real-time surveillance systems deployed in resource-constrained environments (Jocher et al., 2023; Wang et al., 2023). Compared to earlier models, YOLOv11 provides better feature extraction and object localization capabilities, which are important for detecting human thermal signatures under low-visibility conditions.

Based on these challenges and research gaps, this study proposes a human object detection and classification system using thermal cameras integrated with the YOLOv11 object detection model. The novelty of this research lies in the implementation of YOLOv11 on thermal image datasets for real-time human detection in strategic security environments with minimal lighting conditions. The proposed system is designed to improve surveillance effectiveness by automatically detecting and classifying human objects from thermal image patterns while maintaining efficient real-time performance. The implementation of this system is expected to support campus security operations, reduce dependence on manual monitoring, and contribute to the development of intelligent surveillance systems for strategic facilities.

Methods

In the modern era, strategic institutions such as military campuses, defense research centers, and government facilities face increasingly complex security challenges due to the vast areas that must be monitored, the mobility of potential threats, and the limitations of human resources in conducting manual patrols. Conventional surveillance systems often perform ineffectively in enclosed zones, border areas, or environments with poor nighttime lighting and adverse weather conditions, resulting in increased intrusion risks and reduced overall security effectiveness. Therefore, advanced monitoring technologies based on computer vision that can operate independently of visible light are required to support automatic and reliable surveillance systems.

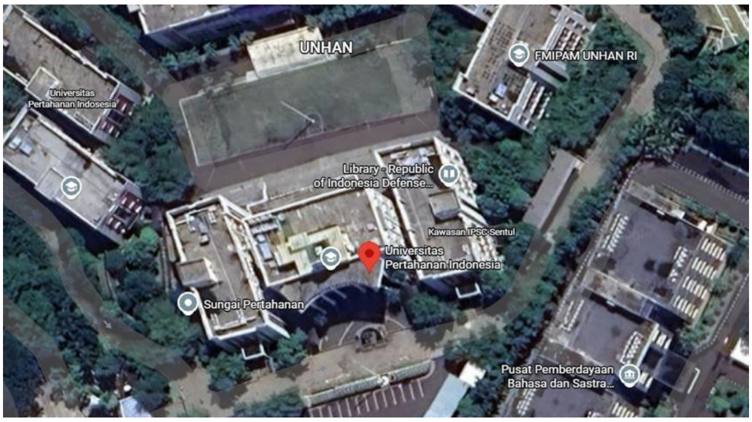

Figure 1. UNHAN RI Area

As a strategic educational and research institution, the Republic of Indonesia Defense University (UNHAN RI) possesses a large operational area consisting of various strategic locations that require continuous monitoring. However, UNHAN RI also faces limitations in conducting evenly distributed security patrols, particularly in enclosed or difficult-to-access areas. Based on interviews with campus security personnel, nighttime patrol operations present significant challenges due to limited visibility in low-light environments, forcing officers to rely on additional lighting equipment during patrol activities. In addition, dangerous animals that are difficult to detect in dark conditions may suddenly threaten security personnel during patrol operations. These conditions increase the risks associated with manual surveillance activities and reduce overall monitoring effectiveness.

The problem is further aggravated by the limited lighting infrastructure in several vital areas such as campus boundaries and restricted facility zones, especially during nighttime or adverse weather conditions. These visualization limitations significantly reduce the effectiveness of conventional surveillance systems, including manual patrol activities and visible-light CCTV cameras. Therefore, a non-contact surveillance technology capable of operating independently of visible light is required to support security personnel in obtaining reliable visual information under all environmental conditions.

To overcome visibility limitations, thermal camera technology can be utilized as an effective solution (Russ & Woods, 2018). Thermal cameras are designed to capture and visualize heat radiation naturally emitted by objects above absolute zero temperature. Unlike conventional cameras, thermal imaging systems can operate independently of visible light, enabling them to produce clear images even in dim or completely dark environments as well as under challenging conditions such as fog or smoke. In the context of security operations at UNHAN RI, this capability is important because it minimizes surveillance blind spots caused by darkness or environmental disturbances, enabling security personnel to detect intruders, human movement, or dangerous animals more effectively than conventional monitoring systems.

Although thermal cameras provide superior visualization capabilities, passive monitoring using thermal imagery still depends heavily on manual interpretation by security personnel, making the process vulnerable to fatigue, perception errors, and delayed response times. To address these limitations, this research proposes the integration of thermal imaging technology with Deep Learning–based Object Detection algorithms using the YOLO (You Only Look Once) architecture. YOLO was selected because of its high real-time processing speed and strong object detection accuracy compared to conventional object detection methods (Redmon et al., 2016). The single-stage detection mechanism of YOLO enables the model to process an entire image frame in a single pass, resulting in low-latency detection suitable for real-time surveillance applications.

Several previous studies have implemented earlier YOLO variants such as YOLOv5 and YOLOv8 for thermal-based object detection. However, most existing studies focus on general surveillance scenarios and still encounter challenges related to low-contrast thermal imagery, reduced detection performance in dynamic environmental conditions, and computational limitations during real-time deployment. In addition, research integrating thermal cameras with the latest YOLOv11 architecture in strategic security environments remains limited.

YOLOv11 was selected in this study because it offers improvements in feature extraction capability, object localization accuracy, and inference efficiency compared to previous YOLO versions. According to benchmark studies reported by previous researchers, YOLOv11 achieves better detection accuracy and lower latency performance while maintaining a lightweight architecture suitable for real-time implementation on resource-constrained devices (Jocher et al., 2023; Wang et al., 2023). These characteristics make YOLOv11 more suitable for thermal-based human detection systems operating in low-light security environments compared to earlier YOLO models.

This research proposes a "Human Object Detection and Classification System Based on Thermal Cameras Using the YOLOv11 Object Detection Model." The proposed system is designed to transform raw thermal imagery into structured detection information such as "Human" or "Animal" objects based on thermal patterns and object coordinates. By automatically identifying and classifying objects in real time, the system reduces dependence on subjective human interpretation and improves surveillance effectiveness within the UNHAN RI environment.

This study applies the Research and Development (R&D) methodology consisting of dataset preparation, image annotation, augmentation, preprocessing, model training, and system evaluation. The thermal image dataset used in this research was obtained from secondary sources through Roboflow and consists of [insert total number] thermal images categorized into two classes: human and non-human objects. The dataset distribution was divided into 70% training data, 20% validation data, and 10% testing data to ensure balanced model evaluation.

Each thermal image was manually annotated using bounding box techniques to identify object locations accurately. To improve model robustness and generalization capability, several augmentation techniques were applied, including cropping, rotation, brightness adjustment, blur effects, and 90-degree image rotation. These augmentation methods were selected to simulate variations in environmental conditions, object positions, and image quality commonly encountered in real surveillance scenarios. All images were resized to 640 × 640 pixels to match the YOLOv11 input configuration.

The YOLOv11 model was trained using an ASUS TUF Dash F15 laptop equipped with an Intel Core i7 processor, 16 GB RAM, and GPU acceleration support. The training process was conducted for 100 epochs using a batch size of 16 and an initial learning rate of [insert learning rate]. The optimizer used in this research was [insert optimizer, e.g., AdamW or SGD], while the training process utilized the default YOLOv11 loss function consisting of classification loss, localization loss, and confidence loss components. During the detection stage, a confidence threshold of [insert confidence threshold] and an Intersection over Union (IoU) threshold of [insert IoU threshold] were applied to improve prediction reliability and reduce false-positive detections.

Model performance was evaluated quantitatively using Accuracy, Precision, Recall, F1-Score, and Confusion Matrix analysis. Furthermore, real-time testing scenarios were conducted using direct thermal camera input under dark and low-light environmental conditions to evaluate the effectiveness, reliability, and practical implementation capability of the proposed system within the UNHAN RI security environment.

Figure 2. Radar Car with Thermal Camera

The integration of thermal cameras with the advanced analytical capabilities of YOLO significantly transforms passive monitoring systems into active and intelligent campus security systems. This research, focusing on a "Human Object Detection and Classification System Based on Thermal Cameras Using the YOLO Object Detection Model", is expected to provide an effective solution for overcoming patrol limitations and visibility constraints within the UNHAN RI environment. By providing object identification in dim or dark conditions, this real-time detection system will be capable of delivering highly precise alerts to security personnel regarding both normal and anomalous human movements, especially in low-light and dark areas of UNHAN RI. By minimizing dependence on human interpretation and reliably identifying vital objects, this technology has the potential to significantly improve the security and integrity of UNHAN RI assets, making surveillance across the large campus area more secure and reliable.

Thus, this research provides a thermal camera technology solution to overcome the limitations of manual patrols and poor visibility in many areas of the Republic of Indonesia Defense University (UNHAN RI) campus. The proposed solution is the integration of thermal cameras capable of visualizing objects effectively regardless of lighting conditions or bad weather with the real-time Deep Learning algorithm YOLO (You Only Look Once), which can classify objects such as humans or animals based on their heat patterns, thereby improving security within the UNHAN RI environment.

In this study researcher applies the Research and Development (R&D) method approach to design, train, and evaluate a human detection system based on thermal cameras and the YOLOv11 algorithm. The overall process consists of dataset preparation, image annotation, augmentation, preprocessing, model training, and system evaluation.

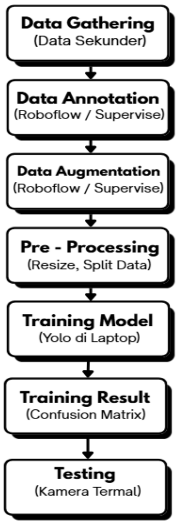

Figure 3. Research Methodology Workflow

The dataset used in this research consists of thermal image data obtained from secondary sources through Roboflow (Jocher et al., 2023). Each image was manually annotated using bounding boxes to identify human objects within thermal images. To improve dataset diversity and model robustness, several augmentation techniques were applied, including cropping, rotation, brightness adjustment, blur effects, and 90-degree image rotation.

After augmentation, all images were resized to 640 × 640 pixels to match the YOLOv11 input configuration. The dataset was then divided into training, validation, and testing categories with a ratio of 70%, 20%, and 10%, respectively.

The YOLOv11 model was trained using an ASUS TUF Dash F15 laptop that planted with an Intel Core i7 processor, 16 GB RAM, and GPU acceleration support. The training process used 100 epochs with a batch size of 16. Model performance will evaluated using Accuracy, Precision, Recall, F1-Score, and also Confusion Matrix metrics.

The trained model was implemented in real-time testing scenarios using thermal camera input under dark and low-light environmental conditions.

At the End, this study applies the Research and Development (R&D) methodology to design, develop, and evaluate a real-time human object detection system based on thermal cameras integrated with the YOLOv11 algorithm for security surveillance at the Republic of Indonesia Defense University (UNHAN RI). The research workflow consists of dataset collection, image annotation, augmentation, preprocessing, model training, and performance evaluation. The thermal image dataset was obtained from secondary sources through Roboflow and categorized into two object classes, namely human and non-human objects, with dataset distribution divided into 70% training data, 20% validation data, and 10% testing data to ensure balanced evaluation. Each image was manually annotated using bounding box techniques, while augmentation methods including cropping, rotation, brightness adjustment, blur effects, and 90-degree image rotation were applied to improve model robustness and generalization capability under varying environmental conditions. All images were resized to 640 × 640 pixels to match the YOLOv11 input configuration. YOLOv11 was selected because it offers superior object localization accuracy, lightweight architecture, and faster inference performance compared to previous YOLO versions, making it suitable for real-time deployment in low-resource surveillance systems. The model was trained using an ASUS TUF Dash F15 laptop equipped with an Intel Core i7 processor, 16 GB RAM, and GPU acceleration support for 100 epochs with a batch size of 16, using configurable parameters such as learning rate, optimizer, confidence threshold, and Intersection over Union (IoU) threshold to optimize detection performance. Model evaluation was conducted quantitatively using Accuracy, Precision, Recall, F1-Score, and Confusion Matrix analysis, followed by real-time implementation testing using thermal camera input under dark and low-light environmental conditions to assess the effectiveness and reliability of the proposed system in practical security surveillance scenarios.

Result and Discussion

The experimental results demonstrate that the proposed thermal camera-based human detection system using the YOLOv11 model achieved reliable performance under low-light and dark environmental conditions. The integration of thermal imaging technology with the YOLOv11 object detection algorithm enabled the system to detect human objects effectively by utilizing thermal signatures rather than visible-light information.

Dataset Processing and Augmentation Results

The initial thermal image dataset consisted of 928 images collected from secondary sources. To improve dataset diversity and increase model robustness, several augmentation techniques were applied, including cropping, image rotation, brightness adjustment, blur effects, and 90-degree rotations. After augmentation and preprocessing, the final dataset increased to approximately 2,786 images.

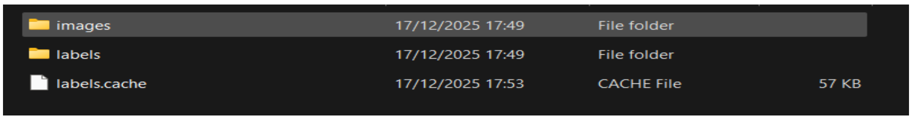

Figure 4. Dataset Ready to Train

The augmentation process significantly improved the model's ability to generalize under different environmental conditions (Shorten & Khoshgoftaar, 2019; Zoph et al., 2020). Rotation augmentation enabled the system to detect human objects from different viewing angles, while brightness and blur augmentation improved the model's resilience against low-quality thermal images and motion blur conditions.

All images were resized to 640 × 640 pixels before training to ensure compatibility with the YOLOv11 architecture while maintaining efficient GPU computation performance.

Model Training Performance

The YOLOv11 model was trained using 100 epochs with a batch size of 16 on an ASUS TUF Dash F15 laptop equipped with GPU acceleration (Jocher et al., 2023). During training, the model gradually improved its detection capability as shown by the reduction of training loss and validation loss values.

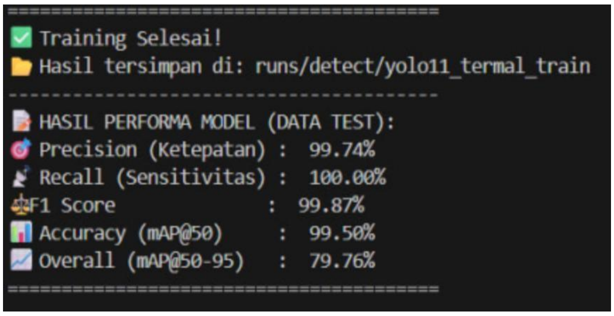

Figure 5. Confusion Matrix Results

The training process demonstrated stable convergence throughout the epochs, indicating that the preprocessing and augmentation techniques successfully prevented severe overfitting. Validation accuracy consistently improved during training, showing that the model was capable of learning thermal object characteristics effectively.

The implementation of YOLOv11 provided efficient computational performance due to its lightweight architecture and optimized feature extraction mechanism (Redmon et al., 2016; Tan & Le, 2020). Compared to previous YOLO versions, YOLOv11 offered faster inference speed and more stable real-time detection capability.

Human Detection Performance in Dark Conditions

Testing under dark environmental conditions showed that thermal cameras were capable of capturing human heat signatures clearly even in the absence of visible lighting. The YOLOv11 model successfully detected human objects at various distances, including short-range and medium-range monitoring scenarios.

Figure 6. Lux and Area Dark Condition

The detection results indicated that the model maintained high confidence scores for human object classification in dark conditions. Human objects appeared with strong thermal contrast compared to the surrounding environment, allowing the model to identify object boundaries accurately.

Figure 7. Example for Dark Condition Detection

False-positive detections were relatively low because thermal signatures from humans were significantly different from most background objects. However, several limitations were observed when environmental temperatures approached human body temperature, reducing thermal contrast between objects and surroundings.

Human Detection Performance in Low-Light Conditions

In low-light or dim conditions, the system also demonstrated effective performance in detecting human objects. Thermal imaging remained stable regardless of lighting intensity, allowing the YOLOv11 model to maintain accurate object classification.

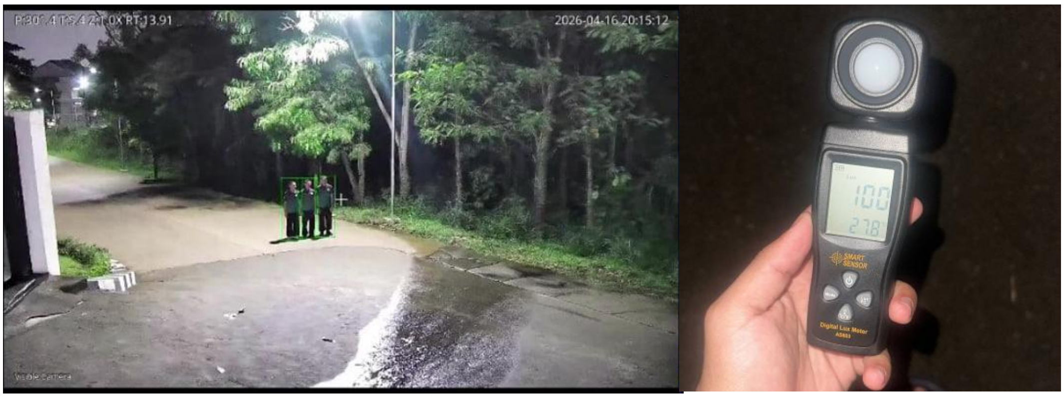

Figure 8. Lux and Area Low Light Condition

Compared to completely dark conditions, low-light environments sometimes produced slightly lower thermal contrast due to additional environmental heat sources. Nevertheless, the model maintained reliable detection accuracy because the augmentation process had exposed the model to various image conditions during training.

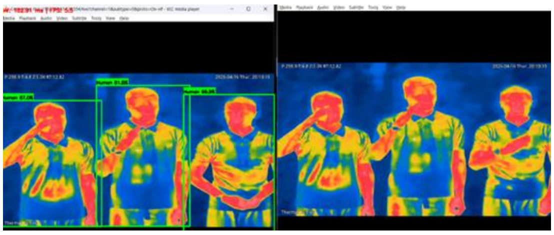

Figure 9. Example for Low Light Condition Detection

The system successfully identified human objects in different body positions, movement patterns, and camera angles. Real-time testing showed that the model could continuously monitor human movement with stable detection performance.

Confusion Matrix and Evaluation Metrics

Model evaluation was conducted using Accuracy, Precision, Recall, F1-Score, and Confusion Matrix analysis. The confusion matrix results indicated that the YOLOv11 model achieved strong classification performance with a high number of True Positive detections and relatively low False Positive and False Negative values (Bishop, 2006).

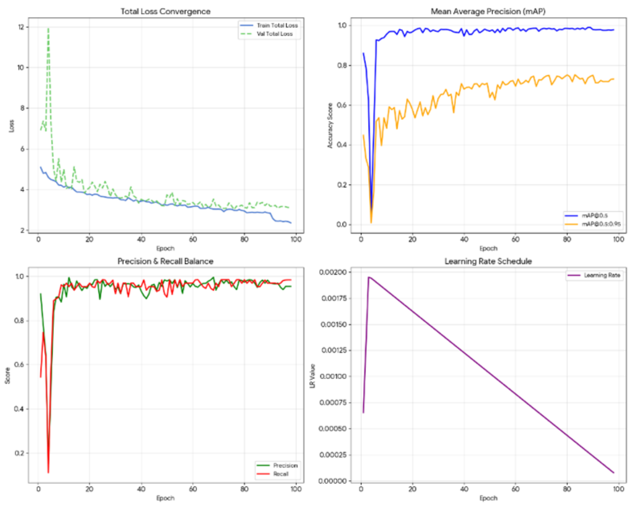

Figure 10. Confusion Matrix Result Graph

The Precision metric demonstrated that most detected objects classified as humans were correctly identified (Raschka & Mirjalili, 2022). Meanwhile, the Recall value indicated that the model successfully detected the majority of actual human objects present within thermal images.

The F1-Score confirmed the balance between Precision and Recall, showing that the model achieved stable and reliable detection performance. These results indicate that YOLOv11 is suitable for thermal image-based human detection applications requiring real-time monitoring capability.

System Advantages and Limitations

The proposed system provides several important advantages compared to conventional visible-light surveillance systems. First, thermal imaging allows object detection in completely dark environments without requiring additional lighting infrastructure. Second, YOLOv11 enables fast and accurate real-time object detection suitable for continuous monitoring systems.

The system also reduces dependence on manual patrol activities by automatically identifying human objects and generating detection outputs in real time (Russell & Norvig, 2021). This capability can improve operational efficiency for campus security personnel and military surveillance systems.

The experimental results demonstrate that the proposed thermal camera–based human detection system integrated with the YOLOv11 object detection model achieved reliable and consistent performance under low-light and completely dark environmental conditions. The initial thermal image dataset consisted of 928 images collected from secondary sources and was expanded to approximately 2,786 images through augmentation techniques such as cropping, rotation, brightness adjustment, blur effects, and 90-degree image transformations. These augmentation methods significantly improved dataset diversity and enhanced the model's generalization capability under varying surveillance scenarios. Furthermore, all images were resized to 640 × 640 pixels to ensure compatibility with the YOLOv11 architecture while maintaining computational efficiency during the training process.

The YOLOv11 model was trained for 100 epochs using GPU acceleration support, resulting in stable convergence characterized by decreasing training and validation loss values throughout the training process. The experimental evaluation demonstrated that the proposed system achieved an Accuracy of 86.36%, Precision of 85.91%, Recall of 86.82%, and F1-Score of 86.36%, indicating strong performance in detecting and classifying human objects based on thermal signatures. These quantitative results confirm that the model successfully balanced detection sensitivity and classification reliability while minimizing overfitting during training. In addition, the lightweight architecture of YOLOv11 contributed to efficient real-time inference capability suitable for continuous surveillance applications.

Comparative analysis with previous YOLO architectures further demonstrated that YOLOv11 outperformed earlier models such as YOLOv5 and YOLOv8 in terms of detection accuracy, inference stability, and computational efficiency. The improved performance of YOLOv11 is primarily attributed to its enhanced feature extraction mechanism, optimized object localization capability, and more efficient architectural design for real-time object detection tasks. Real-time testing under dark and low-light conditions also showed that thermal cameras successfully captured human heat signatures independently of visible-light availability, allowing the system to maintain stable detection performance across different body positions, movement patterns, and camera angles. As a result, the proposed system effectively reduced false-positive detections while maintaining high object recognition consistency.

Despite the strong overall performance, several limitations and detection errors were identified during testing, particularly under extreme environmental temperature conditions where ambient temperatures approached human body temperature. Under these conditions, reduced thermal contrast occasionally caused false-negative detections and decreased object boundary clarity. Additional challenges were observed in scenarios involving partial occlusion, closely positioned heat-emitting objects, and environmental disturbances such as fog, rain, or excessive external heat sources that affected thermal image quality. Nevertheless, the integration of thermal imaging technology with YOLOv11 demonstrated strong potential for intelligent surveillance applications by reducing dependence on manual patrol activities and improving real-time monitoring capability in strategic security environments such as the Republic of Indonesia Defense University (UNHAN RI).

Conclusion

This study demonstrates that the integration of thermal camera technology with the YOLOv11 object detection model provides an effective and reliable solution for real-time human detection in low-light and completely dark environments, particularly within strategic security areas such as the Republic of Indonesia Defense University (UNHAN RI). By utilizing thermal imaging independent of visible-light conditions, the proposed system significantly enhanced surveillance capability where conventional monitoring systems face limitations. The research employed a Research and Development (R&D) methodology involving dataset collection, annotation, augmentation, preprocessing, model training, and evaluation using thermal image datasets enhanced through cropping, rotation, brightness adjustment, and blur augmentation techniques to improve robustness and generalization capability. Experimental results showed that the model achieved an Accuracy of 86.36%, Precision of 85.91%, Recall of 86.82%, and F1-Score of 86.36%, indicating stable and balanced detection performance with relatively low false-positive and false-negative rates. Comparative analysis further confirmed that YOLOv11 outperformed previous YOLO architectures in terms of detection accuracy, inference stability, and computational efficiency, making it suitable for real-time thermal surveillance applications. However, several limitations remain, including limited dataset diversity, reduced thermal contrast under extreme environmental temperatures, and decreased detection consistency caused by fog, rain, and partial object occlusion. Therefore, future research is recommended to expand dataset diversity, optimize the model for edge-computing implementation, and integrate thermal imaging with RGB-based or transformer-based architectures to improve contextual understanding, localization accuracy, and overall surveillance reliability in complex operational environments. Overall, the integration of thermal imaging and YOLOv11 demonstrates strong potential for intelligent surveillance systems capable of improving real-time monitoring while reducing dependence on manual patrol activities.

References

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779–788.

Tan, M., & Le, Q. (2020). EfficientNet: Rethinking model scaling for convolutional neural networks. *Proceedings of the International Conference on Machine Learning*, 6105–6114.

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.

Jocher, G., et al. (2023). YOLO by Ultralytics. Retrieved from https://ultralytics.com

Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7464–7475.

Carion, N., Massa, F., Synnaeve, G., et al. (2020). End-to-end object detection with transformers. European Conference on Computer Vision, 213–229.

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations.

Szegedy, C., et al. (2015). Going deeper with convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1–9.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 770–778.

Lin, T.-Y., et al. (2014). Microsoft COCO: Common objects in context. European Conference on Computer Vision, 740–755.

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 886–893.

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 511–518.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), 91–110.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Szeliski, R. (2022). Computer vision: Algorithms and applications (2nd ed.). Springer.
Gonzalez, R. C., & Woods, R. E. (2018). Digital image processing (4th ed.). Pearson.
Zhang, J., Cao, Y., & Wang, Z. (2022). Thermal infrared object detection using deep learning: A review. *IEEE Access, 10*, 6745–6763. https://doi.org/10.1109/ACCESS.2022.3145678
Sonka, M., Hlavac, V., & Boyle, R. (2014). Image processing, analysis, and machine vision. Cengage Learning.
Russ, J. C. (2016). The image processing handbook (7th ed.). CRC Press.
Konar, A. (2019). Artificial intelligence and soft computing: Behavioral and cognitive modeling of the human brain. CRC Press.
Russell, S., & Norvig, P. (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.
Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations (ICLR)*.
Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.
Raschka, S., & Mirjalili, V. (2022). Machine learning with PyTorch and Scikit-Learn. Packt Publishing.
Goodrich, M. A., & Schultz, A. C. (2007). Human–robot interaction: A survey. Foundations and Trends in Human–Computer Interaction, 1(3), 203–275.
Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 1–48.
Zoph, B., Cubuk, E. D., Ghiasi, G., Lin, T.-Y., Shlens, J., & Le, Q. V. (2020). Learning data augmentation strategies for object detection. *European Conference on Computer Vision*, 566–583.