International Journal of Advances in Applied Sciences (IJAAS)
Vol. 15, No. 2, June 2026, pp. 451~461
ISSN: 2252-8814, DOI: 10.11591/ijaas.v15.i2.pp451-461

Comparison of thermal and non-thermal images for tomato fruit detection

Sulfayanti Faharuddin Situju[1], Wawan Firgiawan[1,2], Hironori Takimoto[3]
[1]Department of Informatics, Faculty of Engineering, Universitas Sulawesi Barat, Mamuju, Indonesia
[2]Artificial Intelligence Study Center, Universitas Sulawesi Barat, Mamuju, Indonesia
[3]Department of Information and Communication Engineering, Faculty of Computer Science and Systems Engineering, Okayama Prefectural University, Soja, Japan

Article Info

Article history:

Received Dec 27, 2024
Revised Mar 5, 2026
Accepted Apr 22, 2026

Keywords:

Object detection
RGB image
Thermal imaging
Tomato
You only look once version 8

ABSTRACT

Farmers use manual observation to sort, grade, and estimate tomato production results to meet market demands. However, this method requires a lot of energy and time, making it unsuitable for large-scale tomato cultivation utilizing the detection process. This study aims to develop the automatic tomato detection technology in an industrial environment based on a conveyor belt by using thermal or non-thermal imaging and you only look once version 8 (YOLOv8). The dataset consists of 570 images obtained from each thermal and non-thermal camera and has undergone augmentation techniques to enrich the data variety. The model was trained and validated using 640×640-pixel images for 40 epochs. In this paper, we conduct a comparative analysis of the tomato detection result using YOLOv8 on thermal and non-thermal imaging. The results indicate that the model trained with thermal data significantly outperformed the non-thermal model, achieving 99% precision, 98% recall, 98% F1-score, and 99% mean average precision (mAP)50 during validation. The thermal model received a 99% accuracy rate during validation, while the non-thermal model attained 94% accuracy, exhibiting a slightly poorer performance and committing several mistakes in detection. The use of thermal cameras on moving automation systems has demonstrated its capability and effectiveness, making it more optimal for application in the agricultural industry.

Corresponding Author:

Sulfayanti Faharuddin Situju
Department of Informatics, Faculty of Engineering, Universitas Sulawesi Barat
Prof. Dr. Baharuddin Lopa, SH Street, Talumung, 91412 Majene, West Sulawesi, Indonesia
Email: sulfayanti@unsulbar.ac.id

1. INTRODUCTION

Tomato plants, known by their Latin name *Lycopersicon esculentum* Mill, are annual shrubs belonging to the Solanaceae family [1]. Tomatoes can be consumed directly. However, in Indonesia, tomatoes are a food ingredient widely used to make typical Indonesian dishes, such as chili sauce, soup, and other traditional foods. In addition, the content of protein, carbohydrates, fat, minerals, and vitamins that are useful for health also adds to the utility of this plant [2]. Food consumption data shows increased tomato consumption from 2017 to 2021 [3]. This then gave birth to efforts to increase tomato production in order to meet the needs of the community for this commodity.

Tomato cultivation is carried out from pre-harvest to post-harvest. Post-harvest cultivation is carried out to maintain the quality of tomatoes to be marketed because, during the storage period, there can be changes in conditions or even a decrease in quality; for example, tomatoes become overripe or too soft.

The quality of tomatoes is usually sorted immediately after the tomatoes are picked by selecting good quality to be marketed and discarding abnormal tomatoes. The tomatoes to be marketed will be selected again through grading (level of ripeness) based on color and size for specific market purposes or consumer preferences. The process of sorting, grading, and estimating tomato production results by farmers is carried out through manual observation to ensure market needs. This method requires a lot of energy and time, so it is unsuitable for large-scale tomato cultivation [4]. Therefore, technology such as automatic tomato sorting conveyor tools can help farmers sort, grade, and count tomatoes so that each process can be more effective and efficient. Technology can also reduce post-harvest process errors caused by humans [5].

The tomato detection and counting method underpins the creation of an automated tomato sorting conveyor, as object detection is the initial phase in the mechanization of the agricultural sector, aimed at identifying target objects and categorizing them appropriately [4], [6]. These methods have utilized many computer vision and deep learning methods by exploring images to detect and count tomatoes in agricultural environments [4]–[12]. The deep learning method is used to detect objects such as tomatoes because of its ability to handle the complexity of the detected object environment and various lighting conditions, such as a mask-region-based convolutional neural network (Mask-R-CNN) [8], [11] Faster-R-CNN [9] and you only look once (YOLO) [4]–[7], [10], [13]. YOLO's dominance is inseparable from its role, which has been proven to detect objects quickly and efficiently [14]. These studies utilize red, green, blue (RGB) images, such as: reference [4] improved the performance of YOLOv8 in the online automated maturity grading and counting process by adopting the multi-head self-attention (MHSA) attention mechanism in YOLOv8's backbone, [5] developed an automatic tomato detection and classification system that can identify tomato ripeness in real time, [11] research utilized Mask-R-CNN to distinguish tomatoes from stems in the development of a harvesting robot, [6] compared the results of tomato detection and counting using YOLOv4, YOLOv4-tiny, YOLOv3, and YOLOv3-tiny, and [13] also demonstrated tomato classification results using YOLOv8 on tomato image augmentation data. Other research combined RGB images with other image types to improve the performance of tomato detection and counting. In [9], an image-based approach was introduced that utilizes RGB and spectral images to evaluate the number of ripe tomatoes ready for harvest. In [7], combines object detection with multiple objects tracking and specific tracking region counting to enhance tomato cluster counting ability. In [10], carries out a target tracking network that can identify and count tomatoes at different growth stages, which is referred to as YOLO-DeepSORT.

Some research about object detection is also applied to RGB images taken using a regular camera, and images taken via infrared [15]–[18]. In [15], offers a rapid detection and recognition model using a YOLO model with a multi-scale self-attention mechanism, specifically designed for small road targets in infrared detection contexts. In [16], utilizes both the original image and its corresponding encoded version, derived through a textural descriptor method, are utilized as inputs to a deep learning model that incorporates target attention and size attention mechanisms for object detection. According to [17], a network using RetinaNet on visible and thermal infrared (TIR) cameras from unmanned aerial vehicles (UAVs) to detect and classify ground objects. While, [18] conducted an analysis of the current state of research on the fusion of visible and infrared images for object detection. Infrared images in object detection show faster detection capabilities in challenging environments, such as low light intensity or when detection is required for small objects in the image. Object detection is also done by utilizing images obtained from a thermal camera [19]–[22]. In [19], introduces a neural network model designed to detect small and tiny objects using the vehicle detection in aerial imagery (VEDAI) dataset [23], which is the thermal imaging captured by UAVs, utilizing the YOLOv5 architecture enhanced with a transformer encoder at the final stage. In [20], supports automated animal detection and classification in thermal images by introducing a novel algorithm for extracting thermal features. Subsequently, a thermal signature for each identified object is computed using morphological operations. In the study by [21], a Faster-R-CNN is applied to thermal and visual spectrum imagery on 4-wheeler, 2-wheeler, traffic light, and human objects to perform a comparative evaluation, revealing that thermal camera images have outperformed results compared to visible spectrum images. In [24], proposed a tomato detection method in a complex natural environment based on the based on residual and C-SPP enhanced YOLOv4 (RC-YOLOv4), and resulted a 88% accuracy in a natural environment. According to [22], visible and thermal images, specifically car, person, bicycle, and dog as the dataset, and adopts RetinaNet as the baseline model to integrate features from multiple data sources for object detection. A thermal camera is valuable for detecting heat signatures, allowing them to work effectively in low visibility conditions, such as at night, where conventional cameras have difficulty.

The application of thermal imaging technology in agriculture is feasible due to temperature differentials that facilitate object detection, which can subsequently be employed for numerous reasons, including prediction and identification [25]–[31]. These works illustrate the trend of thermal use in edge computing for field robotics [25], [29], lower-cost sensor deployment [26], [31], and climate-resilient agriculture [26]–[30]. Shalash *et al.* [29] indicate that artificial intelligence–thermal imaging integration facilitates the prompt detection of fluctuations due to pests or diseases, enabling timely treatment and mitigation. Meanwhile, Awais *et al.* [26] demonstrate that thermal imaging may discern variations in water status and the crop water stress index (CWSI). On the other hand, [25] indicates that thermal imaging enhances fruit grading processes, ensures food safety, and reduces post-harvest losses. Additionally, thermal imaging is an environmentally sustainable technology that generates no waste or hazardous emissions [31]. These studies illustrate the effectiveness of thermal imaging in distinguishing between stressed and non-stressed plants and in acquiring detailed data on canopy temperature. This information can facilitate the comprehension and interpretation of canopy temperature [32]. This method possesses significant potential for application in tropical agriculture, given that tropical regions experience high temperatures and intense sunlight [33], [34].

Research in [35] integrated thermal and visible light imagery with depth data to create a system for the remote detection of plants afflicted by the tomato powdery mildew fungus. This research utilizes the handcrafted feature method for feature extraction and a support vector machine (SVM) for classification. Deep learning adapted to process thermal imaging provides robust object detection, even in complex environments. Thus, the combination of thermal imaging and deep learning allows for high accuracy in detecting and distinguishing objects based on temperature [19], [20]. In [36], proposed a method for the classification of plant and crop diseases with a thermal camera and a CNN. This study works on 28 plants and 6 crop disease classifications. Previous studies show that RGB images are used for tomato detection, and integrated thermal technology is especially used for tomato disease detection. Meanwhile, this study attempts to develop technology for detecting tomatoes on a moving conveyor machine using the YOLOv8 model. This research involved creating a library of non-thermal and thermal images of tomatoes, implementing YOLOv8 for tomato detection in moving conditions, demonstrating the reliability of tomato detection using a combination of thermal technology and deep learning, and presenting a comparative analysis of the tomato detection outcomes utilizing thermal and non-thermal imaging. Non-thermal images are the images obtained from an RGB camera, while thermal images are the images obtained from a thermal camera. The YOLOv8 model is used to obtain detection accuracy according to industrial conditions. Therefore, this research contributes to demonstrating the comparison of the tomato detection results using YOLOv8 in non-thermal images and thermal camera images. Generally, this study emphasizes the development of advanced technology to improve the accuracy and efficiency of tomato detection in various field conditions, such as fruit transfer machines or belt conveyors.

2. PROPOSED METHOD

This study proposes using the YOLOv8 algorithm for tomato detection, using a thermal and non-thermal images dataset to determine the performance of each technology in detecting moving objects on a belt conveyor machine. The YOLOv8 model was chosen for this task due to its superiority in real-time object detection and its efficiency in image processing [37]. YOLO detector is a widely used single-stage object detection network that balances accuracy and speed well. This study uses the latest version of YOLOv8 (web reference) for tomato detection. Mosaic data augmentation is used in YOLOv8. This data augmentation method applies several augmentation methods to new training images, such as flip, rotation, scaling, and color change. This can improve the model's accuracy during the training process compared to not using it. This augmentation aims to simulate various environmental and lighting conditions during the detection process, as suggested in the study [38].

YOLOv8 comprises three primary components: backbone, neck, and head. The backbone is tasked with extracting various features from the input. It operates by dividing feature maps into two halves. The Neck function integrates the features from the backbone. The Head in YOLOv8 identifies the object by directly predicting its center. Figure 1 illustrates the main architecture of YOLOv8.

3. METHOD

The method in this research involves several main stages: data collection, data preprocessing, including augmentation, image labeling, and split data; generating a model; model validation from the training and testing phase to produce the detection results. These stages apply to both types of images used, namely non-thermal and thermal images. The stages of this study can be seen in Figure 2.

3.1. Data acquisition

Data acquisition was carried out using thermal and non-thermal cameras. Data were taken while tomatoes ran on a belt conveyor machine, simulating industrial conditions. Data collection was initially conducted by operating the conveyor at a controlled speed to ensure a consistent flow of tomatoes. The RGB and thermal cameras were positioned with a consistent distance and angle for data collection, specifically perpendicular to the conveyor surface for data uniformity. Throughout the acquisition process, tomatoes were randomly positioned on the conveyor, displaying tomato colors, dimensions, and orientations. The resultant data comprised RGB and thermal images of the tomatoes, respectively. The data were subsequently organized systematically to enhance the annotation process and analysis that followed. The data collection and testing scenarios can be seen in Figure 3.

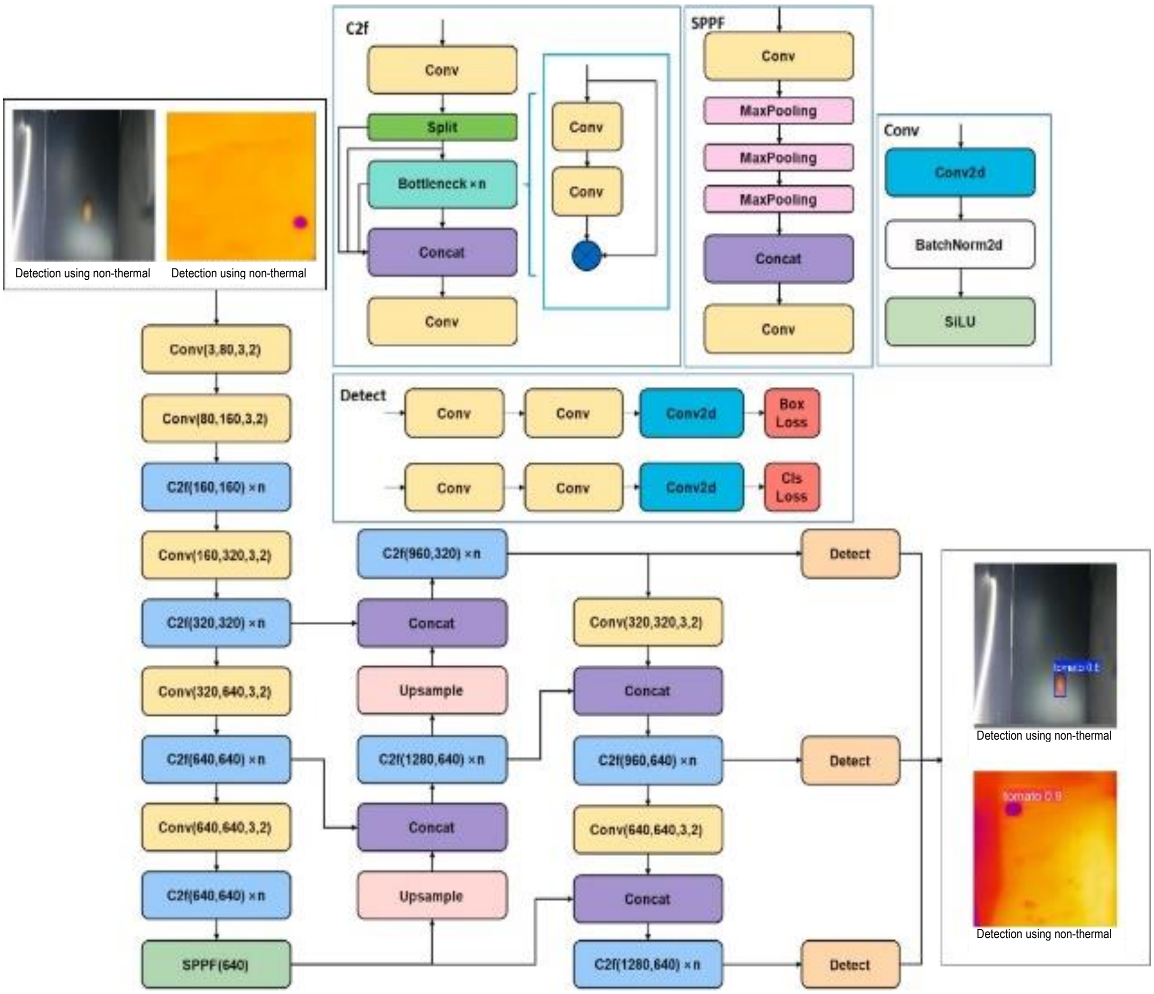


Figure 1. The YOLOv8 architecture for thermal and non-thermal images of tomato detection

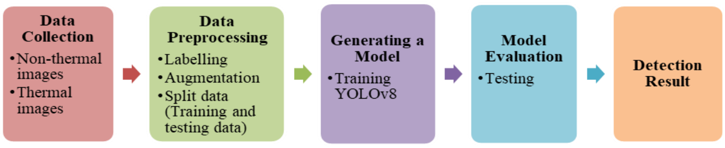


Figure 2. The stages in the research

Figure 3. Data collection process on the conveyor belt machine

3.1.1. Camera specifications

Table 1 shows the differences in specifications between the RGB camera for non-thermal data and the thermal camera for thermal imaging data. The MobIR air thermal camera utilizes a vanadium oxide (VOx) microbolometer sensor in the long-wave infrared (LWIR) spectrum, offering 120×90-pixel IR resolution, a temperature measurement ranges from 20 °C to 120 °C, and a 25 Hz frame rate for precise heat detection. Meanwhile, the Logitech C270 RGB camera employs a CMOS sensor for visible light, delivering higher HD 720p (1280×720-pixel) resolution at up to 30 fps, though it lacks any thermal measurement capabilities.

Table 1. Camera specifications

Technical parameters	MobIR air (thermal camera)	Logitech C270 (RGB camera)
Sensor types and spectrum	VO × microbolometer (long range infrared/LWIR)	CMOS (visible light/RGB) sensor
Resolution	120×90-pixel (IR resolution)	HD 720p (1280×720-pixel)
Temperature measurement range	20 °C to 120 °C	Not applicable
Frame rate	25 Hz	Up to 30 fps

3.1.2. Imaging conditions

The research was performed in a tropical setting characterized by daytime temperatures of 31 °C to 34 °C and relative humidity levels of 70% to 90%. Image data from thermal was acquired outdoors under a direct sunlight environment on sunny, rainless days to reduce interference from infrared radiation caused by water vapor. The acquisition process occurred from 9:30 to 11:00 AM to mitigate the impact of intense sun radiation, which could distort tomato temperature measurements. The camera was positioned above the conveyor, maintaining a fixed distance from the object to guarantee data consistency. The gathered data comprised video footage of tomatoes positioned on the moving conveyor.

The dataset consists of two types of images, namely, images from thermal and regular (non-thermal). Thermal images are used to overcome the challenges of poor lighting and ensure consistent detection in various environmental conditions, as studies have shown that the use of thermal is very effective in detecting objects in low or complex lighting situations [39]. Using thermal imaging also helps distinguish objects with similar colors or textures from the background, thereby improving the quality of the data obtained [35]. The results indicated that the thermal imaging could effectively display tomato objects, as they discernible than other objects in the background. This condition can support the development of a plant sorting system using an outdoor conveyor, as in [40].

3.2. Preprocessing

Data preprocessing is essential to prepare the acquired data for utilization in the model training process. Preprocessing is implemented on both thermal and non-thermal datasets. The acquired video data was extracted into image frames. Subsequently, image data obtained from videos recorded with thermal and non-thermal cameras is labeled or annotated. Moreover, augmentation techniques are employed to enhance the variability of the dataset. Each camera type has 570 images that have undergone an augmentation process to enhance data variability and improve model training quality. The augmentation methods encompass rotation, flipping, saturation, exposure, and blurring. Augmentation enhances the model's diversity and improves its generalization to novel data. The final stage involves partitioning each dataset into training and validation. Table 2 presents the visual representations derived from the labelling and augmentation methods.

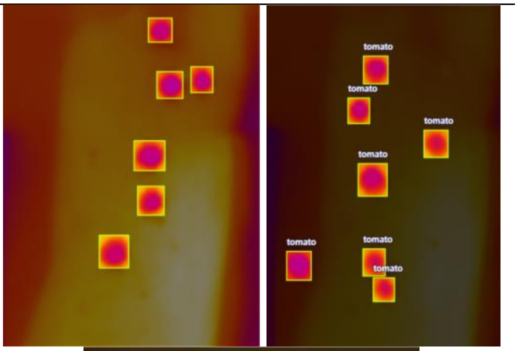
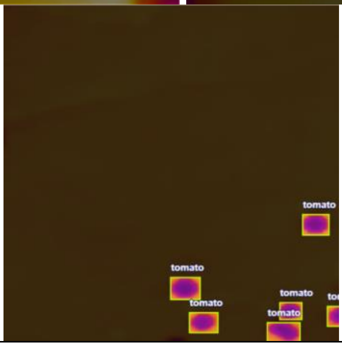
3.3. Detection algorithm

The YOLOv8 model from Ultralytics was trained using a dataset that has 640×640-pixel images from non-thermal and thermal images dataset respectively. The training process involves fine-tuning the parameters using the Adam optimization algorithm, effectively accelerating convergence and reducing loss during training. Fine-tuning aims to optimize the model performance to achieve the best results in detecting tomatoes on conveyor belt machines. The training phase in this study utilizes the Tesla T4 graphics processing unit (GPU). It is excellent for deep learning tasks such as image processing with the YOLOv8 model and object detection. With 15 GB of memory and CUDA version 12.2, the Tesla T4 is excellent for processing large datasets in industrial environments that require real-time processing. These specifications ensure fast and accurate model training, making it applicable in industrial situations such as automatic tomato sorting. Following 40 epochs of training, a validation process was conducted to assess the efficacy and performance of the trained model [41], [42].

3.4. Evaluation metrics

Model testing is performed after training to assess the model's performance in detecting tomatoes. Evaluation is performed using a separate test dataset, which the model never sees during training, to ensure that the model can generalize well to new data. Test data in the form of a video of tomatoes on a moving conveyor. The metrics used to measure model performance include precision, recall, F1-score, mean average precision (mAP), and intersection over the union (IoU). Precision and recall measure the model's accuracy in detecting relevant objects, while the precision and recall of the F1-score balance [43]. Specifically, precision is a metric to measure how accurate the model's predictions are by comparing the correct predictions (true positive (TP)) to the total positive predictions (TP + false positive (FP)). Recall is a metric to measure how many objects the model successfully detects out of all the objects present (detected and missed). F1-score is the harmonic mean of precision and recall, balancing accuracy and completeness. In addition, mAP provides a comprehensive picture of the model's detection performance under various object overlap conditions. mAP is the average of the average precision (AP) values for all classes and the IoU threshold, while AP is calculated as the area under the precision-recall curve for each class. IoU is a metric that quantifies the overlapping area between a predicted bounding box and a ground truth bounding box. On a conveyor, part of the tomato's surface is covered, and the tomato can change position due to movements such as rotation. This condition requires very precise bounding box detection, so the model must do more than just detect the tomato. Therefore, mAP is necessary because it serves a crucial function in assessing the precision of object localization. This metric has also been widely used to determine the results of object detection in the agricultural sector [7], [9], [10], [14].

Table 2. Represent images from preprocessing

	Non-thermal	Thermal
Image labelling		
Image augmentation		

Additionally, inference time measures a model's time to process a single image or batch of images. This time is usually measured in milliseconds (ms) or frames per second (FPS). There is no fixed formula for inference, but this metric indicates the model's ability to detect tomatoes effectively within a system.

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

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

$$F1 - Score = 2 \times \frac{Precision \text{ x } Recall}{Precision + Recall} \quad (3)$$

$$mAP = \frac{1}{n}\sum_{i=1}^{N} AP_i \quad (4)$$

4. RESULTS AND DISCUSSION

Training and validation were conducted to compare the performance of the YOLOv8 model in detecting tomatoes using thermal data and non-thermal data. Training was conducted for 40 epochs. Tables 3 and 4 present the normalized results of the confusion matrix for tomato detection utilizing YOLOv8 on thermal and non-thermal datasets, respectively. YOLOv8 effectively identifies tomato objects, particularly in thermal datasets, with an error rate of up to 1%. Both confusion matrices reveal that certain backgrounds are misidentified as tomatoes, indicating that YOLOv8 has heightened sensitivity in detecting tomato objects across both dataset types. Thus, YOLOv8 demonstrates sufficient reliability for application in tomato object detection within industrial settings that employ conveyors. The primary evaluation measures, such as precision, recall, and F1-score for each model, come from validation data. The results of the mAP50 metric were measured to assess the model's accuracy in detecting objects. Inferences were obtained from testing using videos from each thermal and non-thermal camera. Table 5 compiles the metric evaluation results, while Figure 4 shows the results of the comparison of thermal (Figure 4(a)) and non-thermal (Figure 4(b)) testing on the testing dataset.

Table 3. Normalized confusion matrix of tomato detection using YOLOv8 with thermal imaging

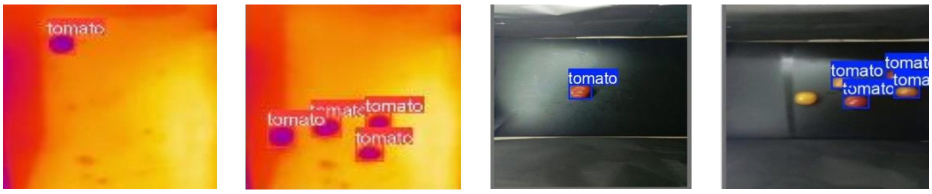
Actual	Prediction	
	Tomato	Background
Tomato	0.99	0.01
Background	1.00	0

Table 4. Normalized confusion matrix of tomato detection using YOLOv8 with non-thermal imaging

Method	Actual	Prediction	
		Tomato	Background
YOLOv8	Tomato	0.94	0.06
	Background	1.00	0

Table 5. Comparison of thermal and non-thermal testing results

Metrics	Thermal	Non-thermal	Combination (thermal and non-thermal)
Precision	0.99	0.84	0.90
Recall	0.98	0.88	0.98
F1-score	0.98	0.86	0.94
Metrics mAP50	0.99	0.77	0.98
Inference	22.0 fps	24.2 fps	-



(a) (b)

Figure 4. Comparison of material conditions for (a) thermal with less and normal amount and (b) non-thermal with less and normal amount

The model trained with thermal images routinely outperforms the model employing solely non-thermal images. The thermal model is more precise and has stronger detection capabilities, particularly when it comes to identifying every tomato object in the image, as evidenced by its higher precision, recall, and F1-score. The thermal model's F1-score is 14.6% higher than the non-thermal model. This suggests that the thermal model is substantially more successful in detection overall, not just marginally better. Utilizing the FP rate ≈1−precision approach, the ratio of non-thermal to thermal FP rates is 16:1. This indicates that the thermal detector commits 16 times fewer errors in identifying nonexistent objects compared to the non-thermal detector. Simultaneously, utilizing the false negative (FN) rate ≈1−recall methodology, the FN rate ratio between non-thermal and thermal is 6. This shows that the Thermal model overlooks six times fewer things than the non-thermal model, rendering it more sensitive and nearly capable of detecting all objects. The comprehensive comparison of the results indicates that thermal is significantly superior to non-thermal. The relative mAP50 difference of thermal to non-thermal is approximately 28%, resulting in a much better overall precision-recall curve in the thermal model compared to the non-thermal model. Figure 5 displays the comparison of mAP graphs during the training process, showing the outcomes of the model training procedure. The training results on the thermal dataset show that the model consistently achieves a high mAP50 across all training epochs compared to the non-thermal dataset. In the first epoch, the model obtained a mAP50 of 0.94712, but then increased in the next epoch. The model continued to show satisfying performance throughout the training process, with mAP50 ranging from 0.94712 to 0.9944. Even after several epochs, there is no discernible decline in mAP50 values' performance during the training process. Thermal not only excels at a singular point but also demonstrates superior consistency and proximity to the ideal across all epochs. This indicates that the model has effectively learned to generalize visual data pertaining to the object's temperature. In contrast, training on non-thermal data shows more significant performance fluctuations than thermal data. In the first epoch, the mAP50 value is much lower than the thermal data.

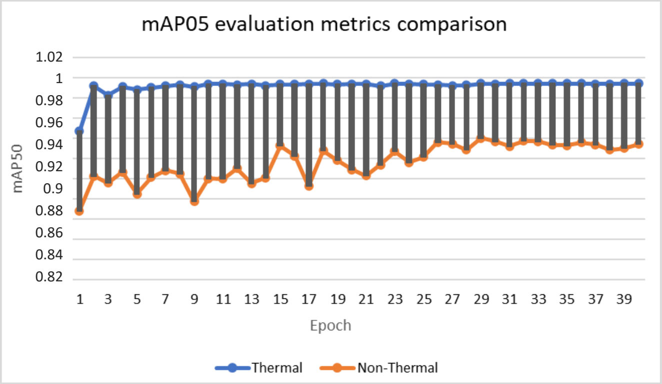


Figure 5. Comparison of mAP graphs in the testing process

Compared to the thermal dataset, the combination dataset exhibits a significantly higher FP rate in its model, approximately ten times greater. The rate indicates that the combination exhibits a higher incidence of false detection, yet maintains a high recall. Each combination and thermal model provide equivalent recall values, indicating that both effectively identify nearly all objects. The combination did not yield a significant result in sensitivity when compared to the use of thermal alone. This also occurs in mAP50. In comparison to using only non-thermal data, the combination of thermal and non-thermal data yielded greater precision, recall, F1-score, and mAP50 values. This demonstrates how tomato detection performance can be enhanced by using thermal images. Simultaneously, the inference values for both models, thermal and non-thermal, produce commendable results, notably ≥20 fps, rendering them appropriate for real-time applications.

The thermal detection accuracy demonstrated superior results compared to non-thermal camera data. The thermal camera captures heat and generates color waves that correspond to the object's temperature, allowing Tomato Edge to effectively detect and precisely determine the shape of the tomato, differentiating it from surrounding objects as previously mentioned in [44]. The utilization of a non-thermal for tomato detection yielded inadequate results due to the occasional blurriness of moving tomatoes on the conveyor. The image's blurriness limits the ability to identify the edges of the tomato, resulting in an inaccurate representation of its shape. This reduced the model's ability to accurately identify tomatoes. Consequently, our study demonstrated that thermal cameras possess benefits in identifying moving objects relative to non-thermal cameras. However, some additional considerations must be addressed: the study found that the influence of light circumstances does not a significantly differentiate the detection outcomes of thermal and non-thermal data; the roundness feature of the tomato object is the primary attribute that distinguishes it from the background, this perhaps leading to the misidentification of other round objects as tomatoes; a thermal camera can identify objects by their heat when there is a notable temperature disparity between the object and its surroundings, while RGB cameras depend on ambient light so performance drops sharply in dusk, night, or shaded areas.

The analysis of the obtained findings indicates the need to integrate the thermal camera with sensors, cameras, or other devices as a hybrid detection method so that the thermal camera can work at room temperature. Nonetheless, this study has demonstrated that the thermal camera is effective in aiding automatic detection systems within industrial environments.

5. CONCLUSION

This study has presented the comparison of thermal and non-thermal using YOLOv8-based for object detection. The system greatly enhances tomato detection performance. The analysis results show that the model trained using thermal data has very good consistency during the training process, without experiencing a significant decrease in performance. This is achieved by applying a dataset that has undergone augmentation techniques and is divided for training, validation, and testing purposes. This model outperformed the one trained on non-thermal camera data in terms of precision, recall, F1-score, and mAP50 values. On the other hand, the non-thermal model was unstable during the first few epochs and had higher performance fluctuations and mAP50 findings. Based on the confusion matrix of the validation data, the thermal model achieved 99% detection precision. In contrast, the non-thermal model showed higher errors, with some background cases being wrongly detected, such as tomatoes, and vice versa. As a result, thermal cameras have been shown to be more effective at supporting the tomato detection system. YOLOv8 has also demonstrated proficiency in detecting tomato items with an inference rate exceeding 20 fps across both dataset types, indicating its capability to specifically identify features of moving objects. When it comes to industrial automation activities that take place after harvest, such as the use of conveyor belts, this thermal technology can increase precision and efficiency. Future work is expected to focus on testing in more complex environments, including system integration for real-time tomato sorting machines. It might be valuable to conduct experiments that account for potential non-target objects or negative objects (such as leaves or other fruits or vegetables) on the conveyor belt. It is desirable to evaluate the robustness of the method in cases where objects other than tomatoes appear by mistake. In addition, inference speed optimization will be a top priority so that the system can run more efficiently and effectively in real-time applications. In the future, the insights gained from the proposed method could contribute to the development of a robust tomato detection system for automatic monitoring of tomato fields by robots, capable of tracking ripeness and fruit count under varying conditions.

FUNDING INFORMATION

This research was funded by the 2024 research grants from the Directorate of Research, Technology, and Community Service.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Sulfayanti Faharuddin Situju	✓			✓	✓	✓	✓		✓	✓		✓	✓	✓
Wawan Firgiawan	✓	✓	✓			✓	✓	✓	✓	✓	✓		✓	✓
Hironori Takimoto				✓	✓					✓				

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [SFS], upon reasonable request.

REFERENCES

[1] T. Zeng, S. Li, Q. Song, F. Zhong, and X. Wei, “Lightweight tomato real-time detection method based on improved YOLO and mobile deployment,” *Computers and Electronics in Agriculture*, vol. 205, 2023, doi: 10.1016/j.compag.2023.107625.

[2] I. Rafani and F. B. M. Dabukke, “The performance and development of Indonesian tomatoes,” ap.fftc.org.tw. 2024. Accessed: Oct. 15, 2024. [Online]. Available: https://ap.fftc.org.tw/article/3523.

[3] G. M. Yolanda, D. H. Darwanto, and M. K. Ardhi, “Consumers’ attitude and preference toward fresh tomatoes in Special Region of Yogyakarta, Indonesia,” *Agraris*, vol. 8, no. 2, pp. 123–138, Oct. 2022, doi: 10.18196/agraris.v8i2.12150.

[4] P. Li, J. Zheng, P. Li, H. Long, M. Li, and L. Gao, “Tomato maturity detection and counting model based on MHSA-YOLOv8,” *Sensors*, vol. 23, no. 15, 2023, doi: 10.3390/s23156701.

[5] S. Ruparelia, M. Jethva, and R. Gajjar, “Real-time tomato detection, classification, and counting system using deep learning and embedded systems,” in *Proceedings of the International e-Conference on Intelligent Systems and Signal Processing*, 2022, pp. 511–522, doi: 10.1007/978-981-16-2123-9_39.

[6] V. Priyanka, P. Sandhya, and V. Vani, “Deep learning based object detection and counting in tomato field,” in *AIP Conference Proceedings*, 2024, vol. 2802, no. 1, doi: 10.1063/5.0182358.

[7] J. Rong, H. Zhou, F. Zhang, T. Yuan, and P. Wang, “Tomato cluster detection and counting using improved YOLOv5 based on RGB-D fusion,” *Computers and Electronics in Agriculture*, vol. 207, 2023, doi: 10.1016/j.compag.2023.107741.

[8] M. Afonso *et al.*, “Tomato fruit detection and counting in greenhouses using deep learning,” *Frontiers in Plant Science*, vol. 11, pp. 1–12, 2020, doi: 10.3389/fpls.2020.571299.

[9] I. T. Chen and H. Y. Lin, “Detection, counting and maturity assessment of cherry tomatoes using multi-spectral images and machine learning techniques,” in *VISIGRAPP 2020 - Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2020, vol. 5, pp. 759–766, doi: 10.5220/0008874907590766.

[10] Y. Ge *et al.*, “Tracking and counting of tomato at different growth period using an improving YOLO-deepsort network for inspection robot,” *Machines*, vol. 10, no. 6, 2022, doi: 10.3390/machines10060489.

[11] S. Haggag, M. Veres, C. Tarry, and M. Moussa, “Object detection in tomato greenhouses: a study on model generalization,” *Agriculture*, vol. 14, no. 2, 2024, doi: 10.3390/agriculture14020173.

[12] M. Yaqoob, S. Sharma, and P. Aggarwal, “Imaging techniques in agro-industry and their applications, a review,” *Journal of Food Measurement and Characterization*, vol. 15, no. 3, pp. 2329–2343, 2021, doi: 10.1007/s11694-021-00809-w.

[13] W. Firgiawan, Sulfayanti, Nurhalisa, N. Inayah, N. Iriansyah, and Mahmuddin, “Improved Tomato classification in industrial environment using YOLOv8 and data augmentation techniques,” in *Proceedings of 2024 IEEE International Conference on Internet of Things and Intelligence Systems, IoTaIS 2024*, 2024, pp. 231–236, doi: 10.1109/IoTaIS64014.2024.10799424.

[14] R. Sapkota, D. Ahmed, and M. Karkee, “Comparing YOLOv8 and mask R-CNN for instance segmentation in complex orchard environments,” *Artificial Intelligence in Agriculture*, vol. 13, pp. 84–99, 2024, doi: 10.1016/j.aiia.2024.07.001.

[15] P. Liu, Y. Zhang, G. Guo, and J. Ding, “Enhanced detection and recognition of road objects in infrared imaging using multi-scale self-attention,” *Sensors*, vol. 24, no. 16, pp. 1–22, 2024, doi: 10.3390/s24165404.

[16] M. Parhizkar, G. Karamali, and B. A. Ravan, “Object tracking in infrared images using a deep learning model and a target-attention mechanism,” *Complex and Intelligent Systems*, vol. 9, no. 2, pp. 1495–1506, 2023, doi: 10.1007/s40747-022-00872-w.

[17] F. Wu, G. Zhou, J. He, H. Li, Y. Liu, and G. Yang, “Efficient object detection and classification of ground objects from thermal infrared remote sensing image based on deep learning,” in *Pattern Recognition and Computer Vision (PRCV 2021)*, 2021, pp. 165–175, doi: 10.1007/978-3-030-88013-2_14.

[18] Y. Sun, Y. Meng, Q. Wang, M. Tang, T. Shen, and Q. Wang, “Visible and infrared image fusion for object detection: a survey,” in *Proceedings of International Conference on Image, Vision and Intelligent Systems 2023 (ICIVIS 2023)*, 2024, pp. 236–248, doi: 10.1007/978-981-97-0855-0_24.

[19] M. D. Tu, K. T. Le, and M. D. Phung, “Object detection in thermal images using deep learning for unmanned aerial vehicles,” in *2024 IEEE/SICE International Symposium on System Integration, SII 2024*, 2024, pp. 687–692, doi: 10.1109/SII58957.2024.10417611.

[20] P. Christiansen, K. A. Steen, R. N. Jørgensen, and H. Karstoft, “Automated detection and recognition of wildlife using thermal cameras,” *Sensors*, vol. 14, no. 8, pp. 13778–13793, 2014, doi: 10.3390/s140813778.

[21] U. Mittal, S. Srivastava, and P. Chawla, “Object detection and classification from thermal images using region based convolutional neural network,” *Journal of Computer Science*, vol. 15, no. 7, pp. 961–971, 2019, doi: 10.3844/jcssp.2019.961.971.

[22] H. Zhou, M. Sun, X. Ren, and X. Wang, “Visible-thermal image object detection via the combination of illumination conditions and temperature information,” *Remote Sensing*, vol. 13, no. 18, 2021, doi: 10.3390/rs13183656.

[23] S. Razakarivony and F. Jurie, “Vehicle detection in aerial imagery: a small target detection benchmark,” *Journal of Visual Communication and Image Representation*, vol. 34, pp. 187–203, 2016, doi: 10.1016/j.jvcir.2015.11.002.

[24] Y. Wang, Q. Rong, and C. Hu, “Ripe tomato detection algorithm based on improved YOLOv9,” *Plants*, vol. 13, no. 22, 2024, doi: 10.3390/plants13223253.

[25] R. S. Kale and S. Shitole, “Non-destructive fruit quality assessment : a review on emerging trends in thermal imaging technology,” *Journal of Computational Analysis and Applications*, vol. 33, no. 8. pp. 326–342, 2024.

[26] M. Awais *et al.*, “UAV-based remote sensing in plant stress imagine using high-resolution thermal sensor for digital agriculture practices: a meta-review,” *International Journal of Environmental Science and Technology*, vol. 20, no. 1, pp. 1135–1152, Jan. 2023, doi: 10.1007/s13762-021-03801-5.

[27] H. S. Ndlovu, J. Odindi, M. Sibanda, and O. Mutanga, “A systematic review on the application of UAV-based thermal remote sensing for assessing and monitoring crop water status in crop farming systems,” *International Journal of Remote Sensing*, vol. 45, no. 15, pp. 4923–4960, Aug. 2024, doi: 10.1080/01431161.2024.2368933.

[28] R. A. P. Hernanda, H. Lee, J. I. Cho, G. Kim, B. K. Cho, and M. S. Kim, “Current trends in the use of thermal imagery in assessing plant stresses: a review,” *Computers and Electronics in Agriculture*, vol. 224, Sep. 2024, doi: 10.1016/j.compag.2024.109227.

[29] O. Shalash *et al.*, “Fusion of robotics, AI, and thermal imaging technologies for intelligent precision agriculture systems,” *Sensors*, vol. 25, no. 22, Nov. 2025, doi: 10.3390/s25226844.

[30] Z. Zhou, Y. Majeed, G. D. Naranjo, and E. M. T. Gambacorta, “Assessment for crop water stress with infrared thermal imagery in precision agriculture: a review and future prospects for deep learning applications,” *Computers and Electronics in Agriculture*, vol. 182, Mar. 2021, doi: 10.1016/j.compag.2021.106019.

[31] P. Pathmanaban, B. K. Gnanavel, S. S. Anandan, and S. Sathiyamurthy, “Advancing post-harvest fruit handling through AI-based thermal imaging: applications, challenges, and future trends,” *Discover Food*, vol. 3, no. 1, 2023, doi: 10.1007/s44187-023-00068-2.

[32] I. Ismail, D. Budiman, and Z. R. Aidha, “The smart agriculture based on reconstructed thermal image,” *Journal of Information Technology and Computer Engineering*, vol. 6, no. 01, pp. 8–13, Mar. 2022, doi: 10.25077/jitce.6.01.8-13.2022.

[33] M. F. Aragão *et al.*, “Evaluation of crop water status of melon plants in tropical semi-arid climate using thermal imaging,” *Revista Brasileira de Engenharia Agricola e Ambiental*, vol. 27, no. 6, pp. 447–456, Jun. 2023, doi: 10.1590/1807-1929/AGRIAMBI.V27N6P447-456.

[34] S. Elsayed, M. Elhoweity, H. H. Ibrahim, Y. H. Dewir, H. M. Migdadi, and U. Schmidhalter, “Thermal imaging and passive reflectance sensing to estimate the water status and grain yield of wheat under different irrigation regimes,” *Agricultural Water Management*, vol. 189, pp. 98–110, Jul. 2017, doi: 10.1016/j.agwat.2017.05.001.
[35] S. E. A. Raza, G. Prince, J. P. Clarkson, and N. M. Rajpoot, “Automatic detection of diseased tomato plants using thermal and stereo visible light images,” *PLoS ONE*, vol. 10, no. 4, pp. 1–20, 2015, doi: 10.1371/journal.pone.0123262.
[36] G. Batchuluun, S. H. Nam, and K. R. Park, “Deep learning-based plant classification and crop disease classification by thermal camera,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, pp. 10474–10486, 2022, doi: 10.1016/j.jksuci.2022.11.003.
[37] C. Tan, J. Sun, A. H. Paterson, H. Song, and C. Li, “Three-view cotton flower counting through multi-object tracking and RGB-D imagery,” *Biosystems Engineering*, vol. 246, pp. 233–247, 2024, doi: 10.1016/j.biosystemseng.2024.08.010.
[38] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.
[39] R. Ippalapally, S. H. Mudumba, M. Adkay, and N. H. R. Vardhan, “Object detection using thermal imaging,” in *2020 IEEE 17th India Council International Conference, INDICON 2020*, 2020, pp. 1–6, doi: 10.1109/INDICON49873.2020.9342179.
[40] P. V. Balabanov, A. G. Divin, A. S. Egorov, and V. A. Yudaev, “Mechatronic system for fruit and vegetables sorting,” in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 734, no. 1, doi: 10.1088/1757-899X/734/1/012128.
[41] Ultralytics, “Explore ultralytics YOLOv8,” docs.ultralytics.com. 2024. Accessed: Aug. 14, 2025. [Online]. Available: https://docs.ultralytics.com/models/yolov8/
[42] R. Varghese and M. Sambath, “YOLOv8: a novel object detection algorithm with enhanced performance and robustness,” in *2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems, ADICS 2024*, 2024, pp. 1–6, doi: 10.1109/ADICS58448.2024.10533619.
[43] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Information Processing and Management*, vol. 45, no. 4, pp. 427–437, 2009, doi: 10.1016/j.ipm.2009.03.002.
[44] J. E. Gallagher and E. J. Oughton, “Assessing thermal imagery integration into object detection methods on air-based collection platforms,” *Scientific Reports*, vol. 13, no. 1, May 2023, doi: 10.1038/s41598-023-34791-8.

BIOGRAPHIES OF AUTHORS

Sulfayanti Faharuddin Situju is a lecturer in the Department of Informatics, Faculty of Engineering, Universitas Sulawesi Barat since 2020. She graduated with a doctoral degree in 2020 from Okayama Prefectural University and a master's degree in 2016 from Hasanuddin University. Her research interests are deep learning and computer vision application on natural environment. She can be contacted at email: sulfayanti@unsulbar.ac.id.

Wawan Firgiawan is a lecturer from Universitas Sulawesi Barat in the Department of Informatics, Faculty of Engineering. He finished his master's degree in 2023 from the Department of Informatics, Hasanuddin University, and his bachelor's degree in 2021 from the Department of Informatics. Universitas Sulawesi Barat. His research is in artificial intelligence. He can be contacted at email: wawanfirgiawan@unsulbar.ac.id.

Hironori Takimoto is a professor at the Department of Information and Communication Engineering in Okayama Prefectural University. He received the doctoral degree in 2007 from the Graduate School of Engineering, The University of Tokushima. He is an expert in deep learning, image processing, and perception information processing. He can be contacted at email: takimoto@c.oka-pu.ac.jp.