INTERNATIONAL JOURNAL OF AUTOMOTIVE AND MECHANICAL ENGINEERING (IJAME)
ISSN: 2229-8649 e-ISSN: 2180-1606
VOL. XX, ISSUE XX, XXX – XXX
DOI: https://doi.org/10.15282/ijame.v#i#.###

ORIGINAL ARTICLE

Improving Vehicle Assistance Systems: Evaluation of Augmented Capabilities through Infrared Thermal Camera Integration

M.S.Beg[1], I.M. Yusri[1,2*], and N.H. Badrulhisam[1,2,3], Ibnu Siswanto[4], Gunadi[4]

[1] Faculty of Mechanical & Automotive Engineering Technology Kampus Pekan, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia.
[2] Centre for Automotive Engineering (AEC), Universiti Malaysia Pahang Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia.
[3] Centre for Sustainable Advance Engineering Design, Faculty of Engineering, Built Environment, and Information Technology, SEGi University, 47810 Petaling Jaya, Selangor, Malaysia.
[4] Department of Automotive Engineering Education, Universitas Negeri Yogyakarta.

ABSTRACT – Nighttime driving is difficult owing to low visibility and lights. Nighttime accidents are more dangerous due to reduced obstacle detection, poor vision, and trouble evaluating distances. Knowing the causes and dynamics of nighttime accidents is essential for improving road safety and preventing collisions when natural light is limited. This study proposes using an infrared thermal sensor to assist drivers in mitigating the issue of inadequate light at night, with the ultimate goal of preventing crashes under such circumstances. The investigation compared the infrared thermal camera sensor with the normal camera visual to evaluate how well it worked at night. The testing has been done on the road in Pekan, Pahang. Yolov8 deep learning has been integrated with both cameras to detect items like cars, motorcycles, and traffic lights. The test findings demonstrated how temperature variations can be utilized to precisely detect items on different types of roadways. The study showed that infrared thermal sensors are impressive at detecting traffic lamps, motorcycles, and vehicles. The infrared camera's actual detection on confusion matrices was 0.98 for traffic lamps and 0.87 for motorcycles and vehicles. This shows how well the infrared thermal camera works in dark conditions, with a faster frame rate of 64.94 fps than regular cameras at 55.25 fps. The results of this study demonstrate that using infrared technology can enhance object detection capabilities and, hence, enhance nighttime road safety.

ARTICLE HISTORY
Received: xxxx
Revised: xxxx
Accepted: xxxx

KEYWORDS
Object detection
Normal camera,
Infrared camera
Deep learning
Collision avoidance

1.0 INTRODUCTION

Based on the data in recent years, more than 50 million people have been injured, while the death toll due to road traffic accidents was 1.35 million annually [1], [2]. Injuries sustained in motor vehicle accidents are consistently ranked among the top 10 significant causes of mortality on a worldwide scale, and they now hold the top spot as the leading cause of death [3]. Distracted driving, speeding, drunk driving, reckless driving, fatigue, and drowsiness are some of the factors that lead to driving accidents. Other factors include bad weather, poor vehicle maintenance, inexperienced drivers, running stop signs and red lights, and situation-specific contextual factors [4]. However, one of the significant contributors to the severity of accidents is notably influenced by late-night driving and adverse weather conditions compared to other contributing factors. The primary cause of many accidents, fatalities, and injuries has been driving in conditions of limited visibility and precipitation during the nighttime [5]; apart from that, the likelihood of unfavorable outcomes resulting from collisions was nearly three times higher in low-light conditions as compared to those occurring in daylight. The increased probability of severe injuries in nighttime accidents can be attributed to various factors, including driver fatigue, substance use, and excessive vehicle speed [6].

Diverse strategies have been adopted to augment road safety, including policies, regulations, law enforcement, road infrastructure or system alterations, and other measures with extensive implications that affect traffic circumstances, driving conduct, and general travel. The interventions above are designed to mitigate the hazards linked to vehicular collisions and enhance the safety of individuals utilizing roadways [7]. However, road safety at night needs more attention from the government, which has resulted in significant accidents [8]. Several factors that make the government not interested in developing road safety at night are as follows: There is a perception among governmental entities that there is a lower volume of traffic at night compared to daytime [9], [10]. Consequently, nighttime road safety initiatives may receive a reduced allocation of resources and attention, as they may be perceived as a lower priority. The allocation of adequate resources for nighttime road safety initiatives may pose a challenge, particularly when faced with competing priorities. Implementing road safety measures tailored to nighttime conditions may be perceived as a more costly endeavor by governmental bodies. Potential expenditures include costs associated with upgraded lighting systems, improved signage, and heightened law enforcement presence during evening periods. Governments may allocate resources towards

*CORRESPONDING AUTHOR | I. M. Yusri | ✉ yusri890@yahoo.com

addressing road safety issues deemed more urgent or having a higher frequency of occurrence, such as impaired driving, distracted driving, or exceeding speed limits. Prioritizing certain aspects may lead to unintended consequences, such as a potential decrease in the focus on ensuring road safety during nighttime hours.

To enhance road safety, drivers typically need to install devices to mitigate potential accidents personally. Several drivers utilized dashcam technology to ensure road safety. The presence of a dashcam is of utmost importance due to its potential to provide vital evidence and documentation about road incidents, accidents, and unforeseen events. Utilizing this evidence and documentation can facilitate the identification of culpability, mitigate the risk of fraudulent claims, and improve road safety. In recent years, the Visible Light Communication (VLC) receiver that uses Complementary Metal Oxide Semiconductor (CMOS) camera technology has attracted the attention of the stakeholders due to their excellent dependability and simplicity of integration [11].

Nevertheless, a significant obstacle for VLC systems dependent on camera technology is attaining a substantial data transmission rate [12]. Also, visible light cameras cannot always pick up different types of light, which makes it hard to get good pictures of different parts of a vehicle when there is not much light or when there is a lot of backlighting [13]. Another technology is lidar, a sensing method capable of detecting. Following the mirroring process, the laser bursts are detected by photodetectors known as avalanche photodiodes (APDs) or single-photon avalanche diodes (SPADs). Distance measurements are conducted by comparing the laser reflections to the emitted beams. Because the LiDAR's position and direction (pose) are known, the one-dimensional range readings can be used to find the three-dimensional Cartesian coordinates [14]. In addition, it should be noted that lidar sensors may exhibit vulnerability to unfavorable weather conditions that can impair their accuracy and overall operational efficiency, like heavy rain, snowfall, or fog [15]. Despite these sensors' demonstrated safety and efficacy, a limitation arises in their performance during nighttime conditions. To address these constraints, the present study seeks to investigate the utilization of infrared thermal (IR) cameras in the context of vehicle detection, with a specific emphasis on nighttime driving conditions.

Due to the intricate lighting circumstances, identifying vehicles at night presents a substantial obstacle for vehicle detection systems. Despite the considerable research in this field, a universally applicable or flawless method to tackle this issue has yet to be developed. The nighttime imaging quality could be better, leading to reduced visibility and the inability to discern vehicle outlines and colors. The primary cause of this phenomenon can be attributed to the complex characteristics of nighttime illuminations, including the presence of excessive reflected light and the interference arising from opposing lanes. Donà Riccardo et al. [16], in their work, applied thermal cameras aiming for pedestrian detection by using the thermal camera during low light conditions and detecting the pedestrian through infrared radiation. The authors found that their integration significantly improved the detection of pedestrians, especially in rural areas.

In one of the notable studies by Wang et al. [17] for a vehicle detection application, the author employed millimeter-wave radar infrared cameras, citing their unique adaptability in challenging weather conditions. While the radar system's resolution was not very high, the combination of millimeter-wave radar and infrared imagery made it easier to quickly and accurately find the specific area of interest on the vehicle. In their study, a set of four novel Haar-like templates was created to increase the accuracy of vehicle detection. The templates above were purposefully created to capture the unique characteristics of vehicles in infrared images. This makes them a valuable asset in the overall process of extracting vehicle features. The application of these techniques yielded a significant enhancement of 2.9% in vehicle detection accuracy, as indicated by the obtained results. The experiments yielded a detection speed of 43 frames per second and a notable detection accuracy of 92.4% in various demanding traffic scenarios.

Slimani et al. [18] Identification of vehicles at night is accomplished by analyzing their taillights and headlights. Achieving this involves applying the two-dimensional discrete wavelet transform (2D-DWT) and background subtraction techniques to the input image. The regions of interest are then extracted using the connected component technique. The classification of these regions is subsequently performed through a hybrid approach involving the 2D-DWT and convolutional neural network (CNN) classifiers. Based on the study's findings, the proposed method notably enhanced precision and recall accuracy. Regarding video number 1, there was an improvement in the accuracy of precision and recall, which reached 98.24%. In the second video, there was a notable enhancement in precision and recall accuracy, with improvements of 98.78% and 98.49%, respectively. Furthermore, the night detection technique demonstrated a processing capacity of 23.22 frames per second in the two videos captured at night. This method of CNN is proven to be one of the systematic and proven methods to be implemented for YOLO algorithms and deep learning object detection algorithms [19].

Xiaobiao et al. [20] introduced an approach for estimating distances and detecting pedestrians, both of which play a critical role in ensuring safe driving practices and developing effective autonomous vehicle control strategies. The approach employs a multi-task region-based convolutional neural network that integrates an enhanced network architecture. The study collected data from a near-infrared camera equipped with two near-infrared fill-light devices, which were used to capture authentic nighttime road scenarios. Accurate measurements of pedestrian distances were obtained using LIDAR technology to train the model. The algorithm under consideration, which incorporates a component

for estimating distances, demonstrates a processing speed exceeding seven frames per second. The pedestrian detection system had an average absolute distance estimation error rate of less than 5% and an accuracy rate of about 80%.

Jinlong et al. [21] have discussed the problem of insufficient manual data for nighttime images captured by urban surveillance cameras. This issue arises because most labeled systems predominantly concentrate on scenarios occurring during the daytime. In order to improve vehicle recognition in unlabeled nighttime photos from the target domain, their study employed domain adaptation approaches to determine the most effective strategy to use tagged daytime photographs from the source domain. Furthermore, the researchers developed a method for estimating traffic flow parameters based on traffic flow theory. To assess the efficacy of their proposed vehicle detection methodology, the researchers employed a dataset comprising 2,200 traffic images. One thousand two hundred images were obtained during daylight hours, while the remaining 1,000 were captured at night. The dataset comprised a total of 57,059 vehicles. In the computation analysis during daylight hours, the Faster region-based convolutional neural network (RCNN) model was employed to assess the average mean absolute error (MAE) and average root mean square error (RMSE). The MAE was determined to be 1.87 km/h, while the RMSE was 3.00. Faster R-CNN demonstrated an average mean absolute error (MAE) of 5.07 kilometers per hour and an average root mean square error (RMSE) of 8.77 in nocturnal situations. Nevertheless, the errors decreased upon implementing the suggested approach, resulting in an average root mean square error (RMSE) value of 7.61 and an average mean absolute error (MAE) of 4.22 km/h. The suggested method demonstrated a mean reduction in the percentage of error (PER) of 15.13%.

Wu et al. [22] proposed a blind spot warning system (BSWS) consisting of camera models that undergo algorithms for blind spot detection (BSD) and dynamic calibration specifically designed to operate efficiently in both diurnal and nocturnal conditions. The suggested solution uses the Horizontal Edge and Shadow Composite Region (HESCR) technique to identify the region of interest and determine the exact location of the shadows cast by the cars in question during the day. In detecting vehicles in nighttime road scenes, the system employs a methodology that involves identifying luminous objects and subsequently recognizing the paired headlights of the specific vehicles of interest. This approach serves the purpose of enabling blind spot detection (BSD) functionality. The system under consideration exhibits operational capabilities at frame rates of 20 and 50 frames per second, thereby ensuring its functionality under varying lighting conditions, including both daytime and nighttime scenarios. A series of experimental video sequences with a cumulative duration exceeding 30 minutes were recorded in road environments during both daytime and nighttime conditions to conduct a quantitative evaluation of the vehicle detection performance. The system exhibits a notable degree of precision in identifying vehicles, attaining detection accuracies of 97.22% and 91.11% during daylight and nighttime scenarios, respectively.

Tampuu et al. [23] examine the effectiveness of using pictures as inputs for a self-driving artificial neural network to perform road-following in actual driving situations. Their findings demonstrate that, in the studied scenarios, the performance obtained utilizing these photos as input is, at the very least, comparable to camera-based models. The evaluation's driving data set consists of more than 500 km of difficult country roads. The collection contains information from LiDAR, camera sensors, and centimetre-level GNSS trajectory data. The dataset includes a range of meteorological situations, including wintry snowfall. LiDAR models may rely on the depth and intensity channels while driving at night since they are active sensing systems that are not reliant on outside light sources. The reflectivity of surfaces does not affect the depth channel, offering reliable results even when snow is covered. Even in training settings, it has been shown that depth alone is inadequate for safe end-to-end driving, although depth may still provide trustworthy information to improve overall performance.

Li et al. [24] have proposed the development of a convolutional neural network that incorporates the Light Enhancement Net. The network effectively improves visibility in low-light conditions by applying its robust nonlinear fitting capabilities. Furthermore, it has been observed that implementing the neural network leads to a substantial decrease in the computational time necessary for processing. The data additionally indicates that the network enhances the identification of pedestrians and other vehicles under low-light circumstances, rendering them more apparent.

Xu Liu [25] identified unstructured road surfaces in visually challenging lighting scenarios, including low-light conditions, backlighting, and hazy environments. The proposed method leverages thermal infrared imagery to achieve improved segmentation of unstructured roads using thermal infrared images. The network architecture of URTSegNet draws inspiration from BiSeNet V 2 and is composed of two branches, namely the Detail Branch and the Semantic Branch. To assess the efficacy of URTSegNet, the researchers gathered a dataset consisting of infrared, unstructured road imagery. This dataset encompasses a range of demanding scenarios, including complete darkness during the night, diminished lighting conditions, backlighting, and atmospheric haze. The dataset was utilized to evaluate the model's performance, which yielded an Intersection over Union (IoU) score of 97.5%. Additionally, the model demonstrated an inference speed of 129 frames per second (FPS) when executed on a single NVIDIA GTX 1660S graphics card.

Choi and Kim [26] developed a system under consideration that integrates a thermal infrared camera and a LiDAR sensor to enhance the capabilities of object detection and identification in various lighting conditions, including both daytime and nighttime scenarios. The researchers devised and produced a three-dimensional calibration target to

externally calibrate the two sensors and acquire their external parameters to accomplish this objective. The researchers conducted experiments to verify the performance of the system. The findings indicated that the proposed sensor system and fusion algorithm effectively detected and identified objects, even in environments with limited visibility. The vision camera and LiDAR sensor combo had an average object detection precision of 56.167% throughout the day. Conversely, it was 55.914% with the combination of the LiDAR sensor and thermal infrared camera.

Zou et al. [27] present a novel model that centers on applying thermal infrared analysis in a specialized manner for excavators, while Ikram et al. [28] used a transformer-based multimodal object detection. This article presents a novel algorithm for detection and perception that utilizes multisensory fusion techniques to improve the performance of excavators. Focusing on directly combining information from various sensors, the research focuses on optimizing the data acquired from each sensor. This methodology is of utmost importance when excavators function in environments with low illumination levels, presenting possible hazards to the operator and individuals walking nearby. To address this matter, the scholars leverage the functionalities of three disparate sensors: infrared cameras, RGB cameras, and Light Detection and Ranging (LiDAR) sensors. The authors present a novel approach for detecting obstacles that demonstrate efficacy in seamlessly transitioning between diurnal and nocturnal conditions. The neural network methodology utilized in this study exhibits enhanced performance compared to the existing detection algorithms. Furthermore, the integration of three-dimensional data allows the sensors to capture and transmit the spatial coordinates of different obstacles to the excavator's operator more efficiently.

Much literature has been published on optimizing various sensors and multiple types of cameras, namely infrared cameras, lidars, radar, and millimeter-wave sensors, with integrated algorithms to enhance and detect vehicles and obstacles. The primary aim of this study is to tackle the challenges above by developing a vehicle detection system for at-night vehicles. To the best of the authors' knowledge, there needs to be more existing literature that discusses optimizing vehicle detection analysis performance in nighttime scenarios using an infrared thermal sensor combined with the Yolov8 algorithm, mainly through the analysis of temperature details. The main advantage of applying infrared thermal sensors through temperature distribution analysis is that they can detect and capture thermal signatures emitted by objects, including vehicles, regardless of lighting conditions. This allows for better visibility and detection of vehicles in low-light or completely dark environments where traditional visible-light cameras may struggle. This study aims to investigate the vehicle detection accuracy of infrared thermal sensors through Yolov8 algorithms, mainly through temperature analysis. The study will focus on the roads in Pekan, Pahang. It aims to develop a more reliable and efficient driver assistance system that enhances road safety and reduces collisions in areas without road lamps. In addition to the utilization of infrared camera testing, this study will also incorporate the use of a conventional visual camera to assess and compare the nocturnal performance of the camera with the capabilities of the infrared camera sensor. This analysis will offer a significant understanding of the extent of the visual capacity and further substantiate the necessity of infrared cameras for nighttime applications.

2.0 MATERIAL AND METHODOLOGY

An extensive description of the experimental platform is given in this section., including a detailed discussion of the sensors employed, their respective types, and associated parameters. Furthermore, this paper thoroughly elucidates the relevant ideas and methodologies utilized in the experiment, which can be divided into three primary sections. Normal and infrared camera sensors are evaluated under actual road conditions in Pekan Pahang, encompassing a 20-kilometer distance. Additionally, the research will be undertaken to optimize and enhance the integration of infrared camera sensors with the YOLOv8n algorithm. Finally, the chapter examines practical implementations of combining the infrared camera sensor with the YOLOv8n algorithm in low-light conditions.

2.1 Vehicle and Camera Preparation

The present study employed a vehicle equipped with a 2.0-liter engine capacity. Figure 1 indicates the vehicle setup for the testing. The conventional and infrared cameras were mounted to the roof of the vehicle frame and linked to a monitoring screen. The primary function of the monitoring display is to visually represent the perspective captured by the infrared thermal camera sensor. The infrared camera was also linked to a 5-volt power source to supply the required power for its functioning.

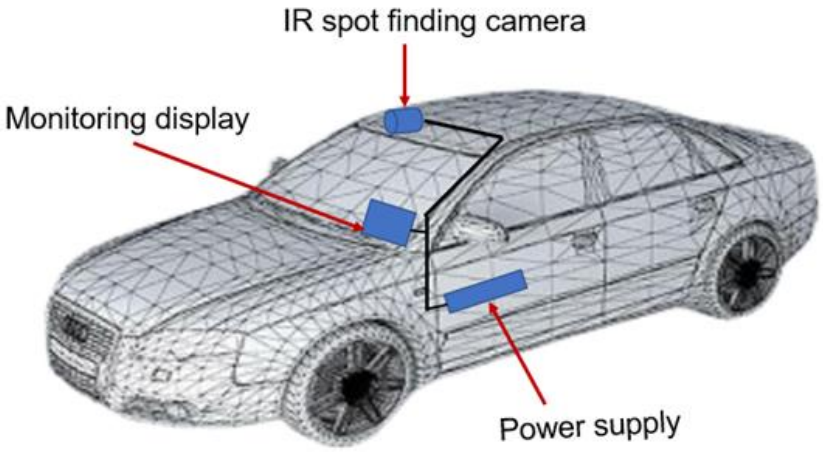


Figure 1. Vehicle and camera setup.

Figure 2 shows the route for the visual collection for both day and night. The drive covers two segments of the route, commencing at McDonald's Pekan and concluding at Terminal Bas Pekan, covering a total distance of 8.2 kilometers. Subsequently, the testing phase continues from Pantai Serandu and extends over a distance of 5.6 kilometers.

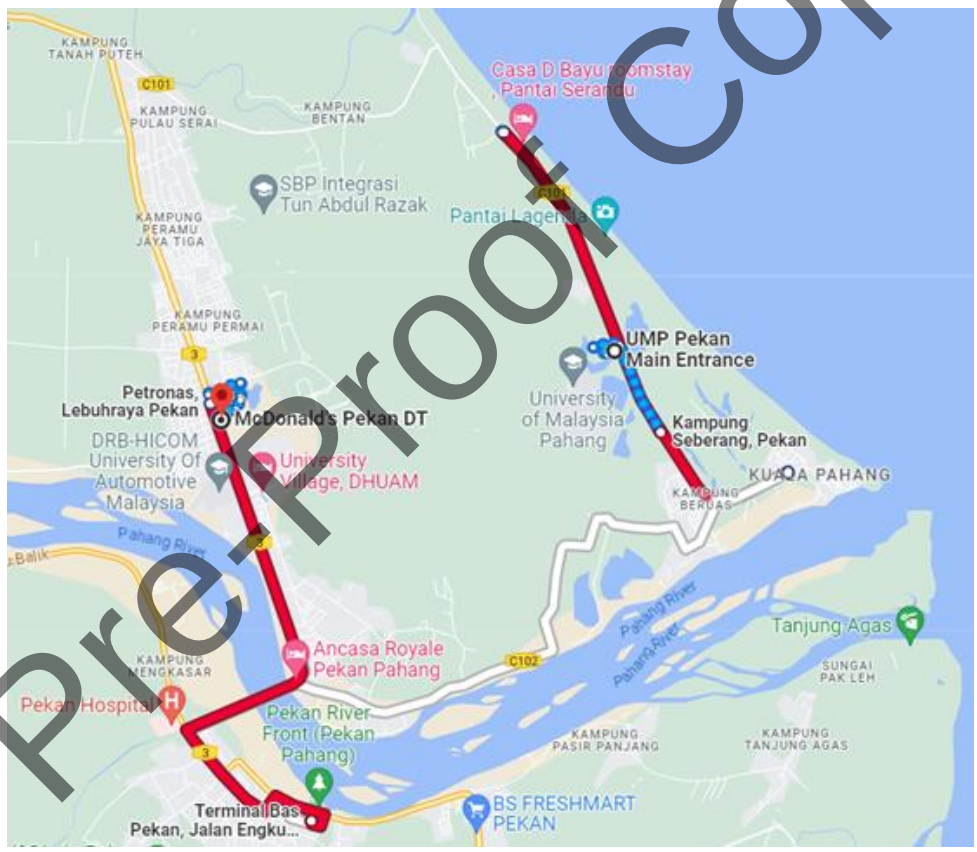


Figure 2. Vehicle tested route for the image collection.

Table 1. General specifications of a normal camera.

Specification	Description
Image Sensor	12-megapixel image sensor
Video Recording	HD at up to 60fps
Field of View (FOV)	170-degree wide-angle lens
Display	2-inch LCD screen
Connectivity	Wi-Fi for wireless file transfer and remote control
Battery Life	Removable 1050mAh rechargeable battery, approximately 90 minutes of continuous recording time
Additional Features	Time-lapse recording, burst photo mode, adjustable exposure, loop recording, built-in adjustable filters
Storage	Supports MicroSD cards up to 64GB

Table 2. General specifications of the infrared camera sensor.

Specification	Description	
Signal output object temperature	0 - 5 V	
Lower limit signal output	0 V	
Upper limit signal output	5 V	
Temperature unit	℃	
Thermocouple	20 Ω	
Digital interfaces	USB, RS232, RS485, Profibous DP, Modbus RTU, Ethernet	
	Sensing head	Electronic box
Environmental rating	IP65 (NEMA-4)	-
Material	Stainless steel	Die Casting
Dimensions	100 mm x 50 mm, M48x 1.5	89 mm x 70 mm x 30 mm
Ambient temperature cable	Max. 105 °C [High-temperature cable (optional): 180 °C]	-
System accuracy	± 1 °C or ± 1%	± 1,5 °C or ± 1,5%

2.2 System workflow

Figure 3 provides a comprehensive depiction of the proposed system. The custom dataset comprises various entities, including vehicles, pedestrians, and traffic lights, recorded during the Pekan Pahang's daytime and nighttime periods on the road. In the initial stage, the system extracts consecutive frames from the input video, segregating them based on whether they belong to the day or night. This enables independent processing to be conducted on each frame. The determination of the sampling rate, which establishes the quantity of frames per second (fps), is based on specific requirements. Each frame contains items identified and categorized using a convolutional neural network (CNN) model. The YOLOv8n model is the object detection model for system implementation and development. The YOLOv8n model has been trained to accurately classify the abovementioned objects, with the vehicle perspective serving as the basis for identification. Consequently, the resulting image encompasses bounding boxes that indicate the detected objects. Ultimately, evaluating the system's performance hinges on its efficacy in detecting objects under varying lighting conditions, encompassing both daytime and nighttime scenarios.

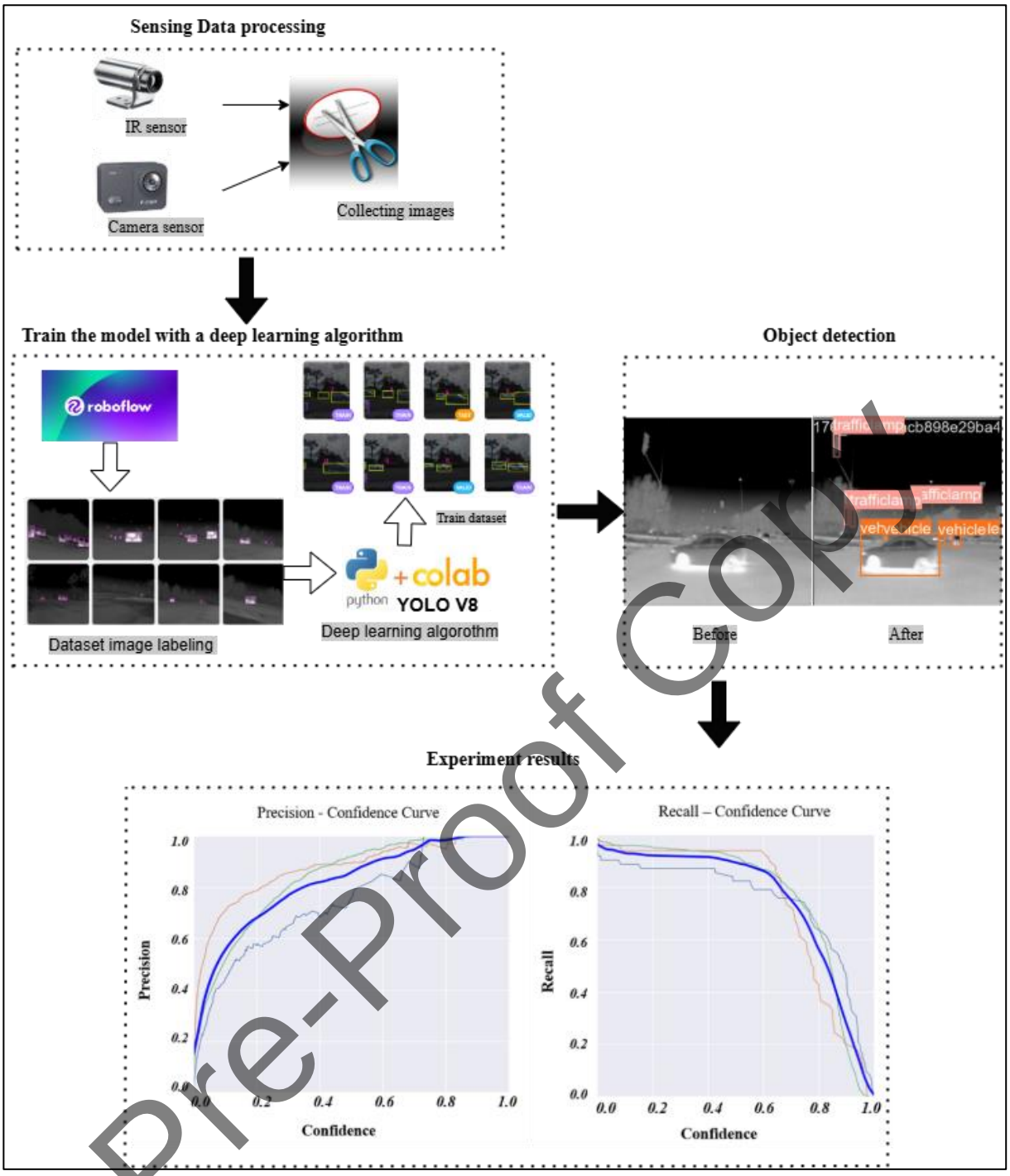


Figure 3. System workflow.

In 2015, University of Washington academics Joseph Redmon and Ali Farhadi created the sophisticated object recognition system YOLO (You Only Look Once) [15]. According to Zhao et al. [9], Typically, object detection models consist of three phases: object classification, feature extraction, and ROI selection. The initial version of YOLO exhibited rapid object recognition capabilities yet encountered challenges in accurately localizing smaller objects. Nevertheless, the utilization of YOLO in our study is manageable due to the absence of tiny images in the dataset employed for our specific problem. The YOLO algorithm functions by partitioning input images into a grid of dimensions M × M. The detection of an object is attributed to a specific grid cell if the object's center is situated within said cell. Both the bounding box coordinates and the associated confidence scores are included in the predictions produced by YOLO. The confidence score in the YOLO algorithm is determined by multiplying the probability of an object being present (Pr(Object)) with the Intersection over Union (IOU) metric. Pr(Object) represents the likelihood of an object being detected, while IOU quantifies the degree of overlap between the predicted bounding box as indicated in Equation (1). The grid cell provides conditional class probabilities as Pr(Class|Object).

$$Class_i \times IOU_{pred}^{truth} = Class_i|Object \times Object \times IOU_{pred}^{truth} \quad (1)$$

Normalization is applied to ensure that the bounding box coordinates fall within the range of 0 to 1. To incorporate non-linearity, the subsequent layers employ the leaky rectified linear activation function, as specified, except the final layer, which utilizes a linear activation function. Equation (2) shows the methodology employed to derive class-specific confidence scores for individual bounding boxes in the testing phase.

$$f(x) = \begin{cases} x, & if x > 0 \\ 0.1x & otherwise \end{cases} \quad (2)$$

2.3 Data analysis procedure

The methodology begins with obtaining a suitable dataset for training and testing, such as the Roboflow dataset manager, which should contain annotated images of labeled vehicles, motorcycles and traffic lights. Integrating camera and infrared (IR) sensors involves creating a connection between these sensors and the vehicle's system, ensuring a smooth integration of hardware and software components. The camera sensor captures real-time images or videos by collecting infrared (IR) sensor data. The camera and infrared sensor data that have been obtained are subsequently analyzed to detect and classify potential objects such as vehicles, motorcycles and traffic lights. Next, the dataset is preprocessed by resizing images (640x640), converting annotations to YOLO format if necessary, and dividing it into training and testing sets. A Google Colab platform is utilized to conduct deep learning experiments to establish the necessary environment. Ultralytics YOLOv8.0.111; Python-3.10.11 torch-2.0.1+cu118 CUDA:0 (Tesla T4, 15102MiB). Use of a Persistence-M GPU, two CPUs, 12.7 GB of RAM, and a 23.3/78.2 GB HDD. The network architecture and optimizer settings are initialized accordingly. Subsequently, the YOLOv8n model is trained on the annotated images or videos from the training dataset. The model's performance is evaluated during training using mean average precision (mAP) or mean intersection over union (mIOU). If necessary, the model is fine-tuned to enhance its performance further. During the iterative improvements stage, the YOLOv8-n model's results are contrasted with the ground truth information obtained from the camera and infrared (IR) sensors. Parameters such as precision, accuracy, recall, and F1 score are computed to assess the integrated system's efficacy. The analysis identifies potential improvement areas by examining discrepancies between the model's predictions and the sensor data.

2.4 Confusion Matrix

The evaluation of a classification model or classifier's performance on known test data is commonly conducted using a widely utilized table known as a confusion matrix. While the confusion matrix is generally easy to understand, its terminology can be confusing.

Table 3 depicts a confusion matrix about a binary classification problem, exhibiting the correctly predicted choices (cells along the diagonal) and the incorrectly classified modal choices. The confusion matrix offers three essential metrics for assessing the predictive performance of the algorithm: The total prediction error is determined by summing the absolute differences between the predicted and observed volumes for each mode and then dividing this sum by the total trip volume. The fraction of observed mode selections that are incorrectly predicted is known as the classification error. The correctly predicted rate represents the percentage of observations correctly identified and positioned in the diagonal cells. The indicators, as mentioned earlier, are essential for assessing the algorithm's predictive ability. The following is the nomenclature used in the confusion matrix:

Table 3. Crucial terms used in the confusion matrix.

		Predicted	
Actual		NO	YES
	NO	TN	FP
	YES	FN	TP

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (3)$$

$$True\ positive\ rate = \frac{TP}{TP + FN}$$

3.0 RESULTS AND DISCUSSION

This paper uses two tools to show accurate results in a clear and meaningful way: the temperature distribution analysis and the confusion matrix. It also included an analysis of visual images using grey-level analysis to demonstrate the quality of the infrared visual condition. The confusion matrix compares the expected and actual choices and shows how prediction errors are distributed among the observed categories. The research analysis was shown in three different ways; (a) The hardware analysis uses an infrared thermal sensor to see temperature and grey levels. (b) They test and train to predict accuracy for motorcycles, traffic lamps, and vehicles using regular cameras and thermal infrared visualization. (c) analyzing the ability to make predictions based on training and testing data.

3.1 Grey-Scale Level Analysis of the Infrared Sensor

Based on the grey-level infrared image, the average, middle value, spread, and shape are used to understand the thermal patterns. The results have helpful information about the image's strength and spread. The grey-level infrared image had an average value of 98.89 and a middle value of 115.0. The median is higher than the mean, meaning the image has lower grey-level values. This suggests that there are colder regions in Figure 4. The skewness value is close to zero at 0.015, which means the distribution is almost symmetrical with a slight tendency towards colder regions. This information shows that the scene has different temperatures, with slightly colder areas in Figure 4.

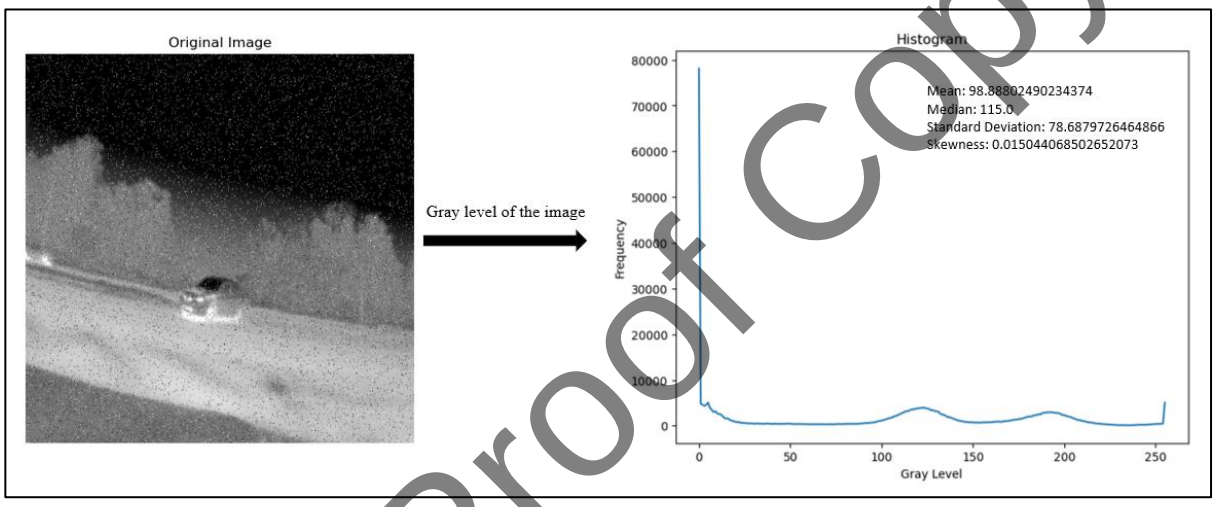


Figure 4. Grey-scale Level Analysis

3.2 Analysis of Object Temperature through Infrared Visualization

This section presents the analysis findings of the temperature differences of detected objects in various road conditions, including T-junction, junction, and straight road scenarios. Significant variations were observed among different object classes. As shown in Figure 5(b), the temperatures of human subjects, motorcycles, and vehicles were measured at 26.5 °C, 28 °C, and 25.8 °C at the T-junction. The temperature at a random cursor point in this area was recorded at 25.1 °C. Similar analyses were performed for the junction and straight road conditions. As shown in Figure 6(b), at the junction, the recorded temperatures for vehicles and traffic lamps were 28.5 °C and 27.5 °C, respectively, while the random cursor point temperature was 23 °C. On the straight road, as in Figure 7(b), the motorcycle and vehicle temperatures were measured at 30.6 °C and 28.6 °C, respectively. The results demonstrate that object temperature can serve as a distinguishing factor between different objects and help identify them accurately. The temperature profiles in Figure 6(d) and Figure 7(d) illustrate the temperature changes over time. This continuous process reveals the highest and lowest temperature points. In the T-junction scenario in Figure 5, temperatures ranged from 20 °C to 32 °C, while at the junction in Figure 6, the range was 23.3 °C to 28.3 °C. For the straight road condition in Figure 7, temperatures varied between 26 °C and 31.5 °C. Notably, the temperature profiles are influenced by the specific road conditions, leading to varying temperature ranges. The temperature-time diagram graphically presents the selected objects' temperature trends, providing further insights into their behavior. Additionally, the histogram depicts the percentage temperature differences for each road condition. At the T-junction, the temperature variations were approximately 11.4%; at the junction, they were 8%; and on the straight road, they reached 12.1%.

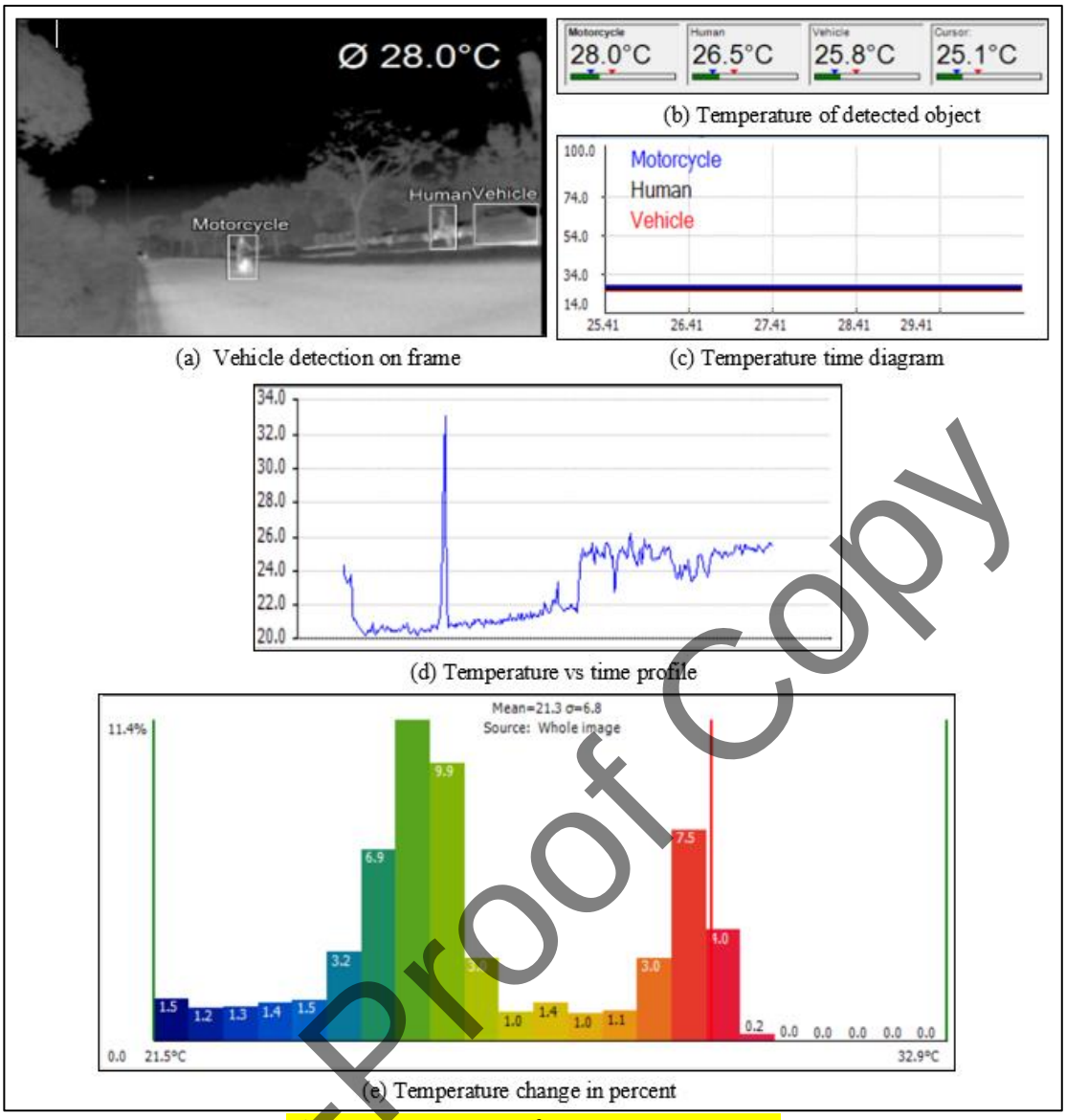


Figure 5. IR Sensor Performance at T- junction.

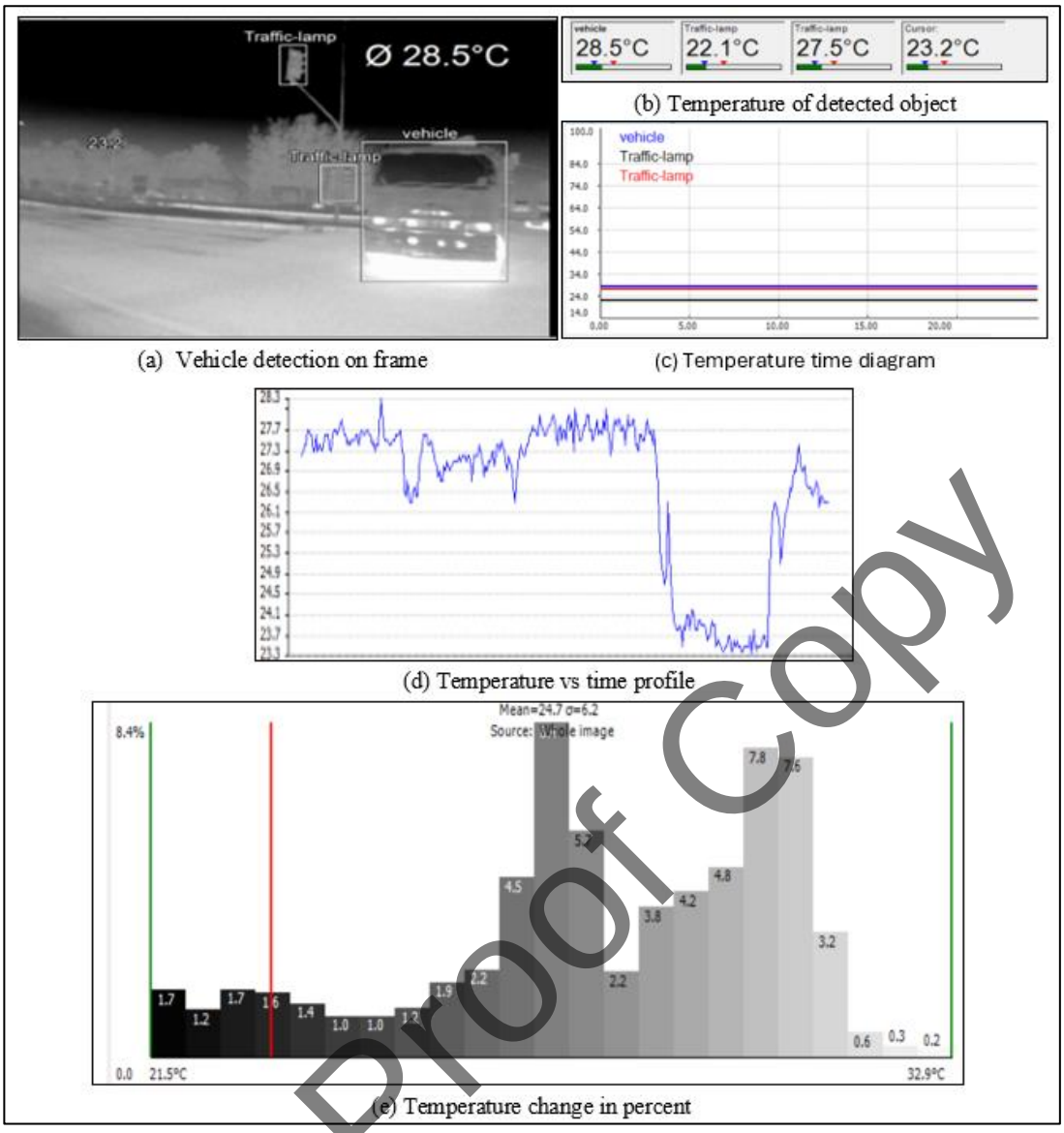


Figure 6. IR Sensor Performance at Junction

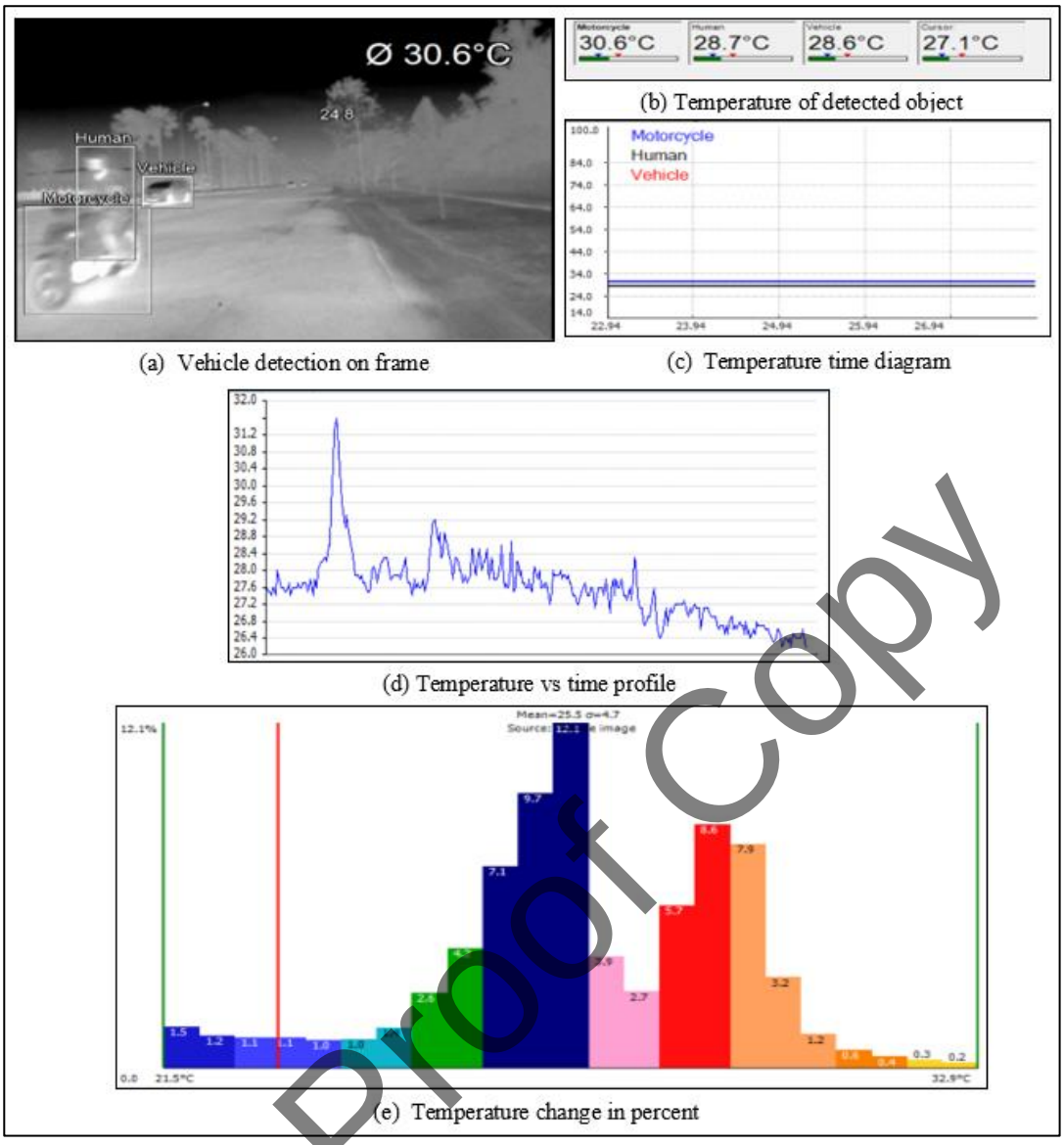


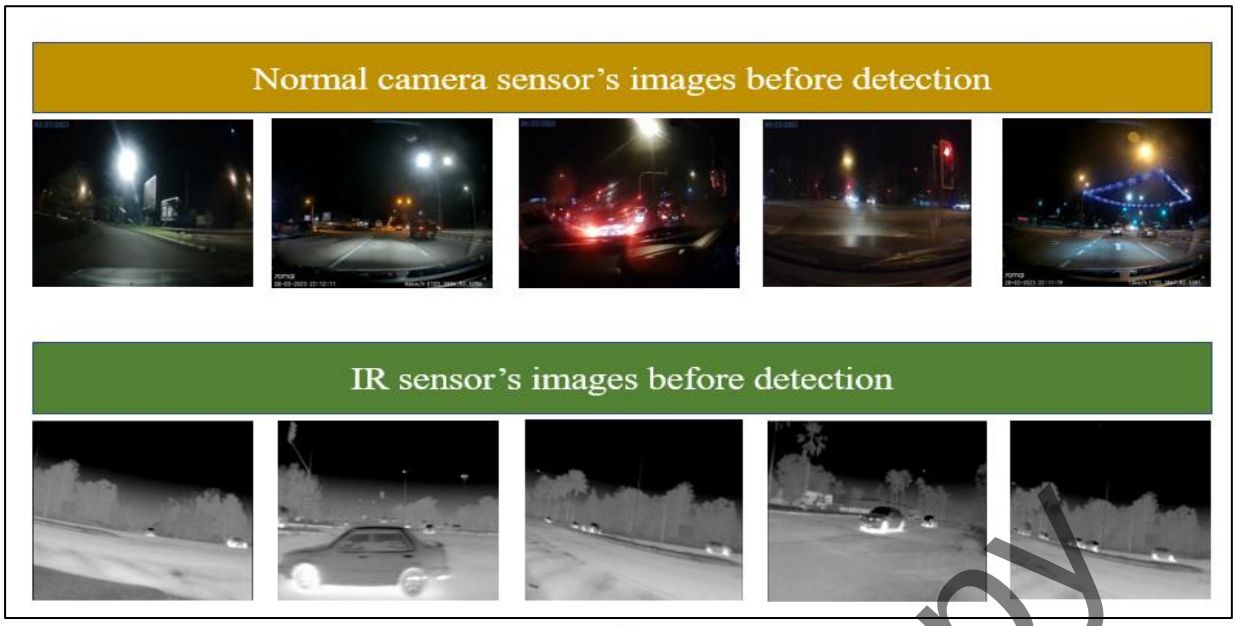
Figure 7. IR Sensor Performance at Straight Road

3.3 Accuracy Analysis

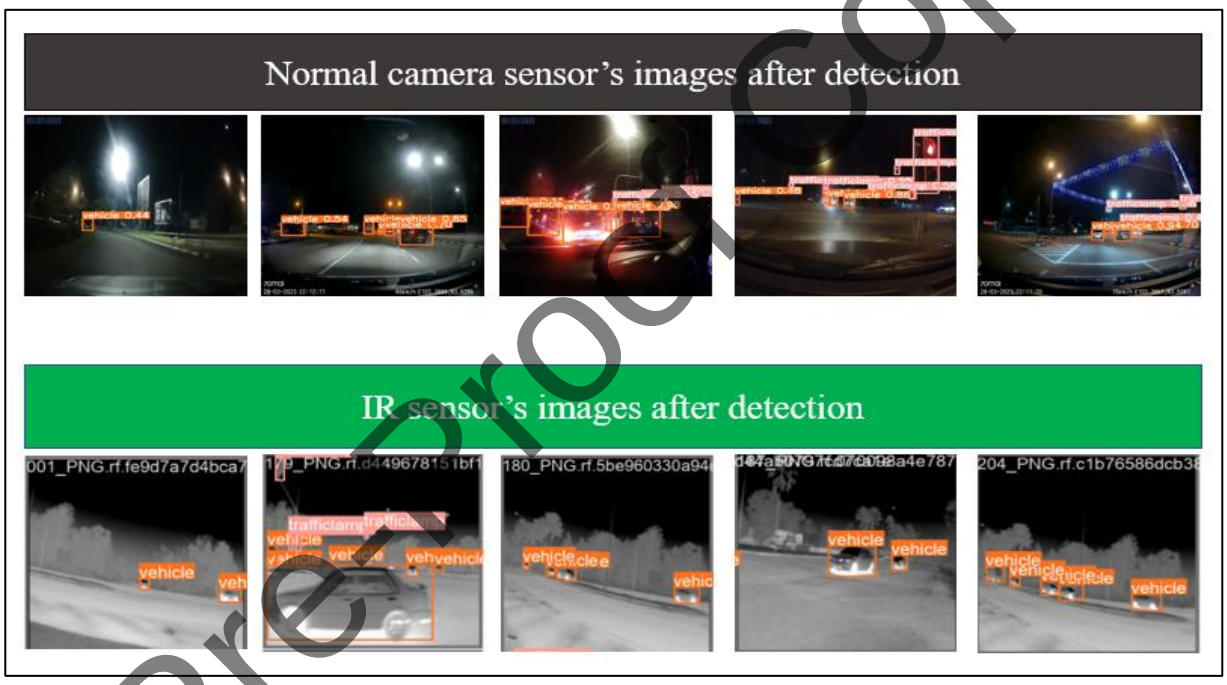
This study compares regular visual and infrared (IR) spot-finding camera sensors in nighttime conditions. Detection performance evaluation is conducted on straight highways, T-junctions, and junctions based on YOLOv8 results. Before that, the experimental setup is complete. The measurement of frames per second (FPS) involves analyzing the time it takes for a model to process inputs and perform post-processing on a single GPU.

The visual camera sensor's preprocessing speed was 3.6 ms, its inference speed was 8.5 ms, and post-processing per image was 6 ms. These measurements indicate an approximate frame rate of 55.25 FPS. The data suggests that performance in nighttime conditions is relatively efficient. The IR spot-finding camera sensor demonstrated a preprocessing speed of 6.4 ms, a postprocessing speed of 5.1 ms, and an inference speed of 3.9 ms, resulting in an estimated frame rate of 64.94 FPS. The mentioned advancement shows a significant enhancement in terms of both speed and efficiency when compared to the standard visual camera sensor. The performance of the IR sensor remained consistent across all scenarios. The preprocessing time slightly increased, while the inference time demonstrated a nearly 50% reduction. That indicates that the sensor can easily analyze infrared data. The performance of the IR sensor remained consistent across all scenarios. The preprocessing time slightly increased while the inference time significantly reduced, indicating the sensor's efficient analysis of infrared data in Figure 8(a).

The successful detection of objects did not significantly impact the speed or frame rate. The system's ability to maintain performance after detection suggests robustness in various nighttime driving scenarios. The performance of the visual camera and IR sensor remained stable after detection. The increased frame rate means a higher responsiveness to environmental changes, particularly in complex junctions in Figure 8(b).



(a)



(b)

Figure 8. Vehicle Detection Images before and after applying the algorithm for Using Normal Camera and Infrared Camera

The IR spot-finding camera sensor is more suited for real-time applications due to its greater frame rate and quicker inference time, particularly in difficult dark situations. Both sensors performed well, but the IR sensor's higher frame rate may provide advantages in detecting fast-moving objects. The study carefully examined important performance metrics, such as box_loss, class_loss, and obj_loss, to compare the IR spot-finding sensor to the regular visual camera sensor. The metrics plots are utilized for training and validation outcomes in the Google colaboratory simulation results.
In object detection, cls_loss sorts objects into groups, box_loss ensures the groups' boundaries correctly include the objects, and dlf_loss (usually a distance-based loss) ensures that features inside an object are correctly located. Together, these three losses make for a complete and accurate detection system. The lower cls_loss and ojb_loss indicate higher accuracy of evaluation metrics.

In Figure 9, the performance metrics of a visual camera sensor are evaluated under nighttime conditions, presenting both training and validation results. After completing 30 epochs, the training values for box_loss, cls_loss, and dfl_loss (objectiveness) have stabilized at 1.6933, 1.1616, and 1.1634, respectively. Simultaneously, there has been a rise in both precision and recall values, reaching 0.596 and 0.53472, respectively. Table 5 discusses the mean average precision, specifically mAP50 and mAP50-95.

Two previous studies illustrate object detection using average visual cameras. Girshick et al. (2014) introduced a method called R-CNN, employing convolutional neural networks (CNN) for both object detection and semantic segmentation, employing the PASCAL VOC 2010 dataset and attaining a mean average precision (mAP) of approximately 58%. Their work laid the foundation for feature extraction, utilizing pre-trained CNNs to scan regions of interest within images. Two years later, Liu et al. (2016) built upon this foundation, proposing the Single Shot MultiBox Detector (SSD), which simultaneously predicts multiple bounding boxes and class scores for those boxes in a single pass. The SSD demonstrated a remarkable precision of 76.8% on the VOC 2007 test. The notable increase in precision illustrates how the field has rapidly advanced, optimizing both speed and accuracy for real-time object detection applications.

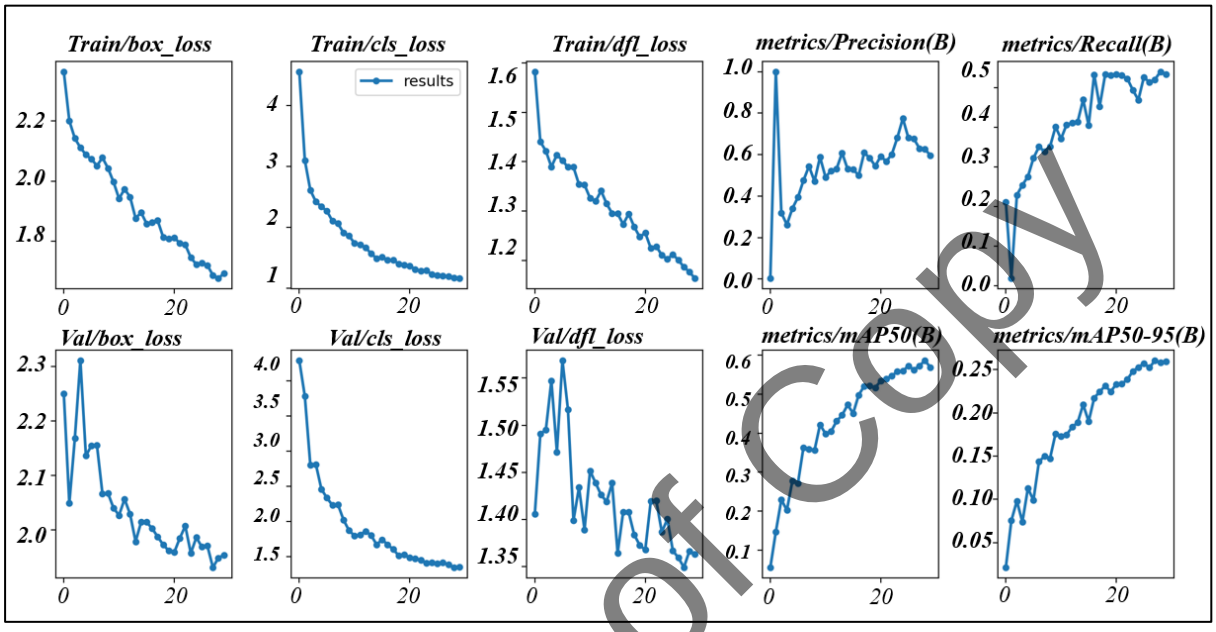


Figure 9. Performance Evaluation Metrics of Visual Camera Senor at Night

Based on an analysis of visual sensor performance evaluation metrics and a review of previous studies, it has been determined that night-time situations present challenges in object detection, especially in low-light conditions. Apart from that, the performance of the IR spot-finding camera is significant. Figure 11 displays all the performance metric plots. After an impressive performance after training for 30 epochs, the values for box_loss, cls_loss, and dfl_loss have become 1.003, 0.47605, and 0.99079, respectively. The precision and recall have also increased to 0.92119 and 0.90534, respectively. The validation results, including the mean average precision values mAP50 and mAP50-95, are discussed in Table 5.

After an extensive training phase encompassing 400 epochs, various algorithms were evaluated using the FLIR dataset. The results reveal significant insights into the performance of different models. Specifically, the mean average precision at 50% overlap (mAP50%) was recorded as follows: Faster R-CNN demonstrated a mAP50% of 79.3%, YOLOv5m reached 81.9%, YOLO-ACN achieved 82.3%, and YOLO-CIR topped the models with a mAP50% of 84.9%. Their findings highlight that the YOLO-CIR model demonstrates enhanced performance. However, in this particular experiment, the YOLO V8 model performed much better, achieving a remarkable 94.7% mAP50%, as detailed in Table 5.

Following the analysis of performance metrics for the typical visual camera sensor and the IR spot-finding camera sensor, significant differences are apparent in their effectiveness under nighttime conditions. The comparison distinctly emphasizes the enhanced detection abilities of the IR spot-finding sensor, especially in challenging nighttime and low-light situations.

A set of custom images was selected in the dataset to validate the experimental model performance as a part of those 169 images for the visual camera sensor in Table 4 and 171 images for the IR spot-finding camera sensor in Table 5. Normal visual camera sensors achieved a 58.1% mean average precision at mAP50(B). An enhancement of the YOLOv3 algorithm by Y. Wang et al. (2019) to create an end-to-end ship target detection system The study provided detailed results for the mean average precision with a 50% overlap (mAP50) on different versions of YOLOv3. Specifically, the mAP50 for YOLOv3-608 was recorded at 73.7%, while for YOLOv3-416, it was 71.2. Furthermore, the study also reported a mAP50 of 75.1 for FPN and FRCN.

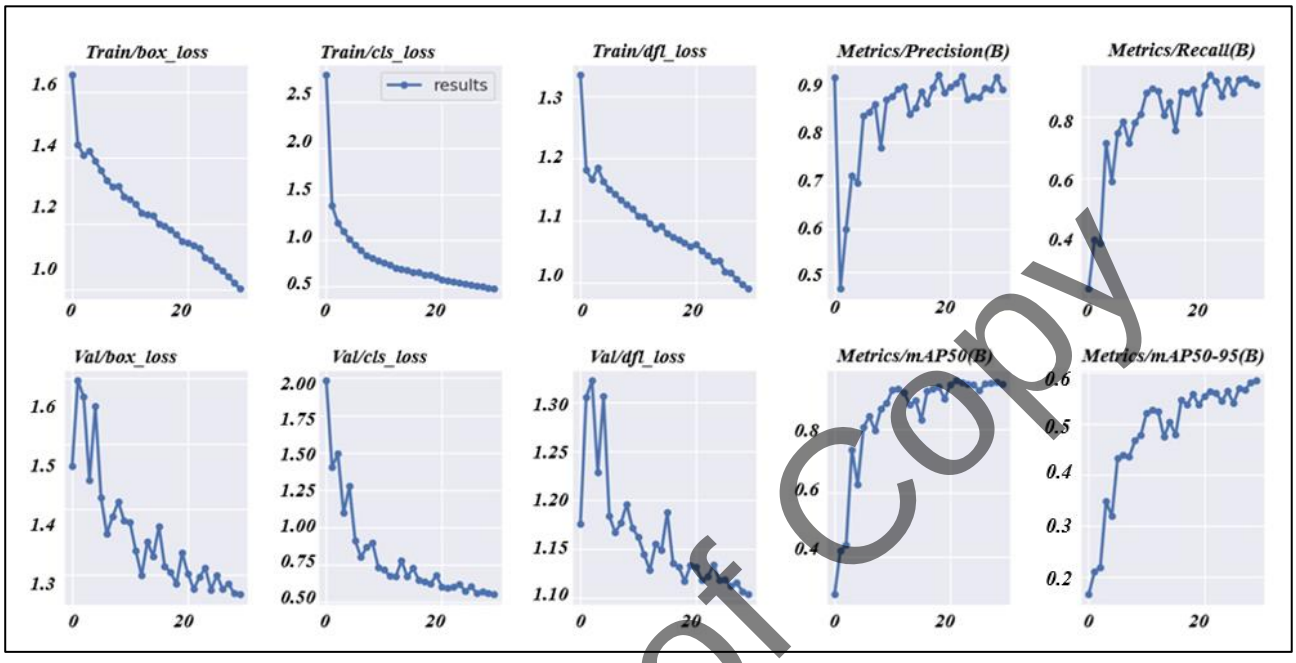


Figure 10. Performance Evaluation Metrics of Infrared Sensor at Night

Table 4. Visual camera sensor performance at night

Class	Images	Instances	Box (P)	R	mAP50(B)	mAP50-95(B)
All	169	637	0.692	0.511	0.581	0.262
Motorcycle	169	54	0.504	0.259	0.315	0.116
Traffic-lamp	169	184	0.827	0.668	0.723	0.303
Vehicle	169	399	0.744	0.605	0.704	0.365

On the other hand, the IR spot-finding sensor achieved 94.7 % mAP50 in Table 5. In all metrics evaluated, the IR sensor performed better than the standard visual camera sensor and other previous studies.

Table 5. IR sensor performance

Class	Images	Instances	Box (P)	R	mAP50(B)	mAP50-95(B)
All	171	662	0.919	0.928	0.947	0.567
Motorcycle	171	62	0.852	0.871	0.901	0.554
Traffic-lamp	171	49	0.969	0.98	0.974	0.497
Vehicle	171	551	0.936	0.935	0.966	0.65

All table results have optimized that the IR sensor significantly performed better in the YOLOv8 simulation test, especially in low-light night conditions. Visual sensor result analysis shows how difficult it is to detect objects at night.

3.4 Confusion Matrix

Object detection performance between IR spot-finding sensors and standard camera sensors during nighttime has been evaluated through a comprehensive comparative analysis of confusion matrices in Figure 11. The study examined true

detection accuracies for three distinct classes: motorcycles, traffic lamps, and vehicles. False detections brought forth further distinctions.

The IR sensor exhibited superior performance, with accuracies nearing perfection for traffic lamps (0.98) and substantial correctness for motorcycles and vehicles (0.89 and 0.96, respectively). That indicates the IR sensor's adeptness at capturing and correctly identifying objects under the constrained luminosity of nighttime. The IR sensor's false detection rates were generally low, except for vehicles as background (0.83). The absence of errors in classifying motorcycles with other classes is particularly striking, showcasing the sensor's precision.

Conversely, the normal camera sensor's performance was markedly lower, with the most significant disparity observed in the motorcycle class (0.31). The reduced accuracy signals the limitations of traditional visual sensors in conditions lacking sufficient illumination. The normal camera sensor exhibited more widespread false detection rates, with pronounced errors such as a 0.63 rate for vehicles being mistaken for the background. These findings indicate a heightened susceptibility to misclassifications, particularly under challenging visual conditions.

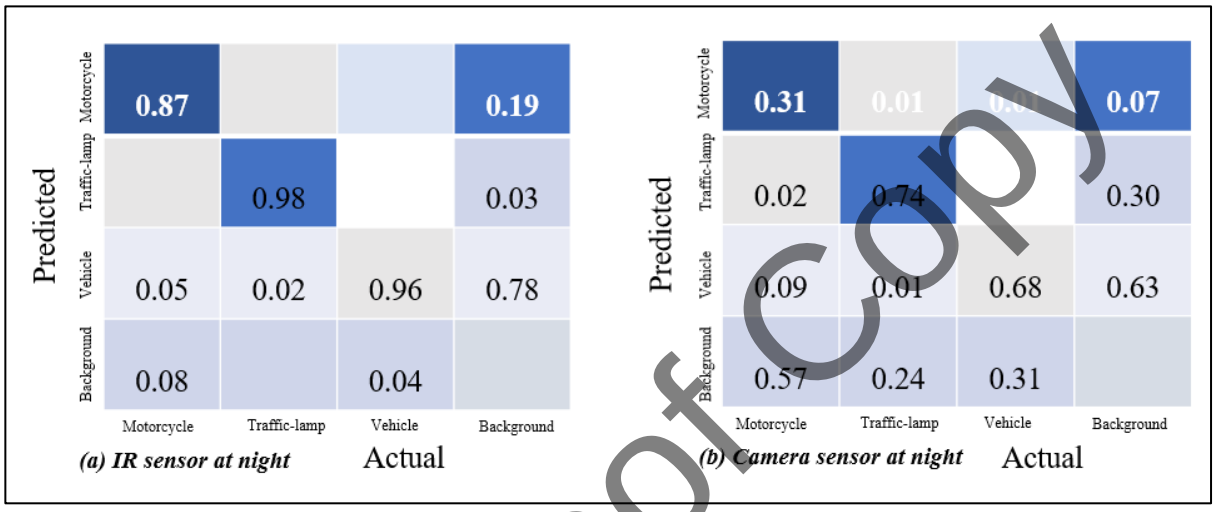


Figure 11. Confusion Matrix of IR Sensor's and Normal Camera Censor's Detection at Night

The IR spot-finding sensor is the optimal choice for nighttime object detection. Its enhanced accuracy and reduced false detection rates make it a more suitable and reliable option, demonstrating the value of leveraging infrared technology for object recognition tasks in nighttime conditions.

4.0 CONCLUSION

Detecting vehicles at night is always hard for vehicle systems, affecting driving safety. This study highlights the role of infrared thermal sensors at night to assist the driver in a collision avoidance system. Therefore, the application of Yolov8 embedded into infrared thermal sensors has been emphasized in this study. In conclusion, these studies show that:

The analysis examined how temperatures varied between objects in different road situations: T-junction, junction, and straight road. Different types of objects showed noticeable differences. At the T-junction, people, motorcycles, and cars showed 26.5°C, 28°C, and 25.8°C, respectively. In the junction situation, the vehicle and traffic lamp temperatures were 28.5°C and 27.5°C, respectively, while the random cursor point was at 23°C. On the straight road, motorcycles and vehicles had temperatures of 30.6°C and 28.6°C, respectively. This study shows that differences in the temperature of objects can help with accurate identification. The temperature graphs in Figures 9 and 10 show how temperatures change over time. They show the highest and lowest temperatures, respectively. Temperature ranges change because of road conditions. The graph and chart display how the temperature of an object changes over time and the percentage differences for different road conditions: T-junction (11.4%), junction (8%), and straight road (12.1%).

The infrared image with different shades of grey was analyzed. The distribution's average, median, spread, and shape were calculated to understand the thermal patterns. The results show information about how strong and spread out the image is. The image's median is higher than the average, which means there are lower grey-level values and colder regions in Figure 7. The skewness value is 0.015, which is close to zero. This means the distribution is almost symmetrical but slightly leans towards colder regions. This data shows different temperatures in the scene, with colder areas being more common.

Based on the comparison of how well regular and infrared cameras detect objects at night, The camera sensor can process images in 3.6 milliseconds and make decisions in 8.5 milliseconds. This means it can capture about 55.25 frames

per second. On the other hand, the IR sensor showed better efficiency. It had preprocessing and inference speeds of 6.4 ms and 3.9 ms, which resulted in an estimated frame rate of 64.94 FPS. The IR sensor works quickly in real-time, especially at night.

In comparison to the normal visual camera sensor and the IR spot-finding camera's performance metrics at night, the IR spot-finding camera's impressive performance after evaluating the plot metrics, especially precision, recall, and mean average precision, shows that this camera is much better for nighttime operations than the normal visual camera sensor. The IR spot-finding camera demonstrated lower loss values and achieved higher precision and recall rates, signaling its ability to detect and classify objects more efficiently and accurately. This decision stems from its demonstrated capability to achieve 94.7% mean average precision compared to the 58.1% achieved by the normal visual camera sensor.

Using the confusion matrix, the IR sensor had a high accuracy of 0.98 for traffic lamps and a reasonable accuracy of 0.89 and 0.96 for motorcycles and vehicles. This shows that the sensor works well in low-light conditions. The IR sensor mostly had low false detections, except when vehicles were in the background (0.83). The IR sensor was precise in classifying motorcycles. On the other hand, the regular camera sensor did not perform as well, especially for motorcycles (0.31), showing that it has some drawbacks in dark environments.

The infrared camera sensor is the best and most effective choice when detecting objects at night. The improved accuracy and fewer false detections make it a better and more reliable choice for these situations. This improvement shows how important it is to use infrared technology to recognize objects in dark places. The IR sensor can improve safety and efficiency when driving at night. It is more accurate and reduces false alarms compared to visual sensors. This result shows that using infrared-based solutions is practical and effective for dealing with the challenges of detecting objects in low-light conditions.

5.0 ACKNOWLEDGEMENT

The authors would like to thank Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) for providing financial support under Internal Research Grant Scheme No. RDU230379 and laboratory facilities.

6.0 REFERENCES

[1] A. Mesic *et al.*, "Generating consensus on road safety issues and priorities in Ghana: A modified Delphi approach," *Injury*, vol. 54, no. 9, p. 110765, 2023, doi: https://doi.org/10.1016/j.injury.2023.04.052.

[2] S. Tinnes, Q. Noir, G. Delubac, J. Matias, R. Dona, and B. Ciuffo, "Automatic emergency braking: how can affordable thermal camera improve reliability and extend use cases to nighttime conditions," in *Infrared Technology and Applications L*, G. F. Fulop, M. H. MacDougal, D. Z. Ting, and M. Kimata, Eds., SPIE, 2024, p. 130460Y. doi: 10.1117/12.3013738.

[3] S. Chen, M. Kuhn, K. Prettner, and D. E. Bloom, "The global macroeconomic burden of road injuries: estimates and projections for 166 countries," *Lancet Planet Health*, vol. 3, no. 9, pp. e390–e398, Sep. 2019, doi: 10.1016/S2542-5196(19)30170-6.

[4] J. J. Rolison, S. Regev, S. Moutari, and A. Feeney, "What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records," *Accid Anal Prev*, vol. 115, pp. 11–24, 2018, doi: https://doi.org/10.1016/j.aap.2018.02.025.

[5] H. Li *et al.*, "The Effect of Rainfall and Illumination on Automotive Sensors Detection Performance," *Sustainability*, vol. 15, no. 9, 2023, doi: 10.3390/su15097260.

[6] J. Yang, S. Liu, H. Su, and Y. Tian, "Driving assistance system based on data fusion of multisource sensors for autonomous unmanned ground vehicles," *Computer Networks*, vol. 192, p. 108053, 2021, doi: https://doi.org/10.1016/j.comnet.2021.108053.

[7] Y. Xu *et al.*, "Hybrid modulation scheme for Visible Light Communication using CMOS camera," *Opt Commun*, vol. 440, pp. 89–94, 2019, doi: https://doi.org/10.1016/j.optcom.2019.01.047.

[8] I. Abdulrashid, R. Zanjirani Farahani, S. Mammadov, and M. Khalafalla, "Transport behavior and government interventions in pandemics: A hybrid explainable machine learning for road safety," *Transp Res E Logist Transp Rev*, vol. 193, p. 103841, 2025, doi: https://doi.org/10.1016/j.tre.2024.103841.

[9] Z. Eusofe and H. Evdorides, "Assessment of road safety management at institutional level in Malaysia: A case study," *IATSS Research*, vol. 41, no. 4, pp. 172–181, 2017, doi: https://doi.org/10.1016/j.iatssr.2017.03.002.

[10] J. Park, B. K. Thota, and K. Somashekar, "Sensor-Fused Nighttime System for Enhanced Pedestrian Detection in ADAS and Autonomous Vehicles," *Sensors*, vol. 24, no. 14, 2024, doi: 10.3390/s24144755.

[11] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object Detection With Deep Learning: A Review," *IEEE Trans Neural Netw Learn Syst*, vol. 30, no. 11, pp. 3212–3232, 2019, doi: 10.1109/TNNLS.2018.2876865.

[12] Z. Liu, Y. He, C. Wang, and R. Song, "Analysis of the Influence of Foggy Weather Environment on the Detection Effect of Machine Vision Obstacles," *Sensors*, vol. 20, no. 2, 2020, doi: 10.3390/s20020349.

[13] N. Yaghoobi Ershadi and J. M. Menéndez, "Vehicle Tracking and Counting System in Dusty Weather with Vibrating Camera Conditions," *J Sens*, vol. 2017, no. 1, p. 3812301, 2017, doi: https://doi.org/10.1155/2017/3812301.

[14] H. Li, D. J. Graham, H. Ding, and G. Ren, "Comparison of empirical Bayes and propensity score methods for road safety evaluation: A simulation study," *Accid Anal Prev*, vol. 129, pp. 148–155, 2019, doi: https://doi.org/10.1016/j.aap.2019.05.015.

[15] M. Kutila, P. Pyykönen, H. Holzhüter, M. Colomb, and P. Duthon, "Automotive LiDAR performance verification in fog and rain," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018, pp. 1695–1701. doi: 10.1109/ITSC.2018.8569624.

[16] R. Donà *et al.*, "Thermal Cameras and Their Safety Implications for Pedestrian Protection: A Mixed Empirical and Simulation-Based Characterization," *Transp Res Rec*, vol. 0, no. 0, p. 03611981241278346, doi: 10.1177/03611981241278346.

[17] Z. Wang, J. Zhan, Y. Li, Z. Zhong, and Z. Cao, "A new scheme of vehicle detection for severe weather based on multi-sensor fusion," *Measurement*, vol. 191, p. 110737, 2022, doi: https://doi.org/10.1016/j.measurement.2022.110737.

[18] I. Slimani, A. Zaarane, and I. Atouf, "Traffic Monitoring System for Vehicle Detection in Day and Night Conditions," *Transport and Telecommunication*, vol. 24, no. 3, pp. 256–265, Jun. 2023, doi: 10.2478/ttj-2023-0020.

[19] P. Tsirtsakis, G. Zacharis, G. S. Maraslidis, and G. F. Fragulis, "Deep learning for object recognition: A comprehensive review of models and algorithms," *International Journal of Cognitive Computing in Engineering*, vol. 6, pp. 298–312, 2025, doi: https://doi.org/10.1016/j.ijcce.2025.01.004.

[20] X. Dai *et al.*, "Multi-task faster R-CNN for nighttime pedestrian detection and distance estimation," *Infrared Phys Technol*, vol. 115, p. 103694, 2021, doi: https://doi.org/10.1016/j.infrared.2021.103694.

[21] J. Li, Z. Xu, L. Fu, X. Zhou, and H. Yu, "Domain adaptation from daytime to nighttime: A situation-sensitive vehicle detection and traffic flow parameter estimation framework," *Transp Res Part C Emerg Technol*, vol. 124, p. 102946, 2021, doi: https://doi.org/10.1016/j.trc.2020.102946.

[22] B.-F. Wu, H.-Y. Huang, C.-J. Chen, Y.-H. Chen, C.-W. Chang, and Y.-L. Chen, "A vision-based blind spot warning system for daytime and nighttime driver assistance," *Computers & Electrical Engineering*, vol. 39, no. 3, pp. 846–862, 2013, doi: https://doi.org/10.1016/j.compeleceng.2013.03.020.

[23] A. Tampuu, R. Aidla, J. A. van Gent, and T. Matiisen, "LiDAR-as-Camera for End-to-End Driving," *Sensors*, vol. 23, no. 5, 2023, doi: 10.3390/s23052845.

[24] G. Li, Y. Yang, X. Qu, D. Cao, and K. Li, "A deep learning based image enhancement approach for autonomous driving at night," *Knowl Based Syst*, vol. 213, p. 106617, 2021, doi: https://doi.org/10.1016/j.knosys.2020.106617.

[25] X. Liu, J. Wang, and J. Li, "URTSegNet: A real-time segmentation network of unstructured road at night based on thermal infrared images for autonomous robot system," *Control Eng Pract*, vol. 137, p. 105560, 2023, doi: https://doi.org/10.1016/j.conengprac.2023.105560.

[26] J. D. Choi and M. Y. Kim, "A sensor fusion system with thermal infrared camera and LiDAR for autonomous vehicles and deep learning based object detection," *ICT Express*, vol. 9, no. 2, pp. 222–227, 2023, doi: https://doi.org/10.1016/j.icte.2021.12.016.

[27] M. Zou, J. Yu, Y. Lv, B. Lu, W. Chi, and L. Sun, "A Novel Day-to-Night Obstacle Detection Method for Excavators Based on Image Enhancement and Multisensor Fusion," *IEEE Sens J*, vol. 23, no. 10, pp. 10825–10835, 2023, doi: 10.1109/JSEN.2023.3254588.

[28] S. Ikram, I. S. Bajwa, A. Ikram, M. Abdullah-Al-Wadud, and H. Pk, "A Transformer-Based Multimodal Object Detection System for Real-World Applications," *IEEE Access*, vol. 13, pp. 29162–29176, 2025, doi: 10.1109/ACCESS.2025.3539569.

.