Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://journal.nsps.org.ng/index.php/jnsps

J. Nig. Soc. Phys. Sci. 6 (2024) 2198

Journal of the Nigerian Society of Physical Sciences

# Data safety prediction using YOLOv7+G3HN for traffic roads

Lek Ming Lim[a], Lu Yang[b], Ahmad Sufril Azlan Mohamed[b], Majid Khan Majahar Ali[a,*]

[a]School of Mathematical Sciences, Universiti Sains Malaysia, 11800 USM Pulau Pinang, Malaysia
[b]School of Computer Sciences, Universiti Sains Malaysia, 11800 USM Pulau Pinang, Malaysia

## Abstract

Pulau Pinang has introduced several measures to enhance traffic safety and promote sustainability, including the installation of CCTV systems and the implementation of smart solutions and green technology as part of the Penang 2030 vision, aligning with the Sustainable Development Goals (SDGs). However, despite these efforts, road accidents persist due to non-optimised detection models, incomplete data from manual reporting, and technological constraints in real-time video analysis and predictive modelling. This study evaluates the effectiveness of the YOLOv7+G3HN framework for vehicle detection and near-miss analysis, with a focus on the influence of video quality on detection performance. The research aims to understand how high- and low-quality video inputs affect the accuracy and computational efficiency of detection algorithms. High-quality videos resulted in significantly faster computation times for vehicle detection than low-quality videos, highlighting the importance of video resolution in optimising detection processes. Despite the robustness of the algorithm, with no errors detected in both video qualities, higher miss detection rates in low-quality videos suggest that lower resolution may compromise detection accuracy and the effectiveness of monitoring systems. Near-miss analysis revealed that high-quality videos had a lower probability of near-miss occurrences than low-quality videos, highlighting the importance of video resolution for detection efficacy. These findings emphasise the critical role of high-resolution video inputs in enhancing detection accuracy and reliability, advocating for their implementation to optimise vehicle detection and improve road safety. Additionally, YOLOv7+G3HN outperforms YOLOv7 in both accuracy and speed. The study concludes that the YOLOv7+G3HN framework is effective for vehicle detection and near-miss analysis, provided that video quality is considered in system design and implementation.

DOI:10.46481/jnsps.2024.2198

*Keywords:* Vehicle detection, Near miss detection, YOLOv7+G3HN, Machine Learning

*Article History :*
Received: 16 June 2024
Received in revised form: 10 July 2024
Accepted for publication: 17 July 2024
Published: 10 August 2024



Communicated by: B. J. Falaye

## 1. Introduction

Pulau Pinang has introduced several measures to enhance traffic safety and promote sustainability, including the installation of CCTV systems to monitor traffic flow and enforce regulations. As part of the Penang 2030 vision, the state is implementing smart solutions and green technology to optimise traffic management, reduce congestion, and improve road safety through IoT and data analytics. The state's initiatives align with the Sustainable Development Goals (SDGs) and aim to create a sustainable transport system by enhancing public transit, infrastructure, and eco-friendly transportation methods. Collaboration with small- and medium-sized enterprises (SMEs) to integrate sustainable practices further supports these efforts, ultimately improving the quality of life for citizens and promoting

*Corresponding author Tel. No.: +60-149-543-405.
*Email address:* majidkhanmajaharali@usm.my (Majid Khan Majahar Ali )

environmental sustainability.

The Pulau Pinang state government's "Penang 2030" initiative aims to reduce road fatalities and create a smart city, but accidents persist despite numerous regulations. Current detection models like RCNN, CNN, and YOLO are not optimised for Penang's unique road conditions, necessitating specific preventive strategies. Manual accident reporting leads to incomplete data, and existing CCTV systems lack advanced real-time monitoring and sufficient historical data due to storage limitations. Additionally, near misses often go unreported, contributing to accidents, and technological constraints hinder the development of effective real-time video analysis and predictive models.

While challenges persist in traditional accident reporting methods and the limitations of existing CCTV systems hinder real-time monitoring, advancements in object detection technology offer promising solutions. The model of YOLO streamlines real-time object detection by analysing entire images at once. By efficiently predicting bounding boxes and class probabilities, YOLO models can potentially aid in detecting and preventing accidents on particular roadways in Pulau Pinang, where traditional detection models have struggled to adapt.

According to the YOLO (You Only Look Once) model, a single neural network architecture is used to analyse the entire input image at once, instead of dividing it into smaller sections or using sliding windows like other object detection algorithms. This unified approach allows for real-time object detection by directly predicting bounding boxes and class probabilities from the full image. The model divides the input image into a grid and then predicts bounding boxes and class probabilities for each grid cell. Each bounding box comes with a confidence score indicating the probability of containing an object, along with class probabilities for the predicted object categories. Finally, the object detection was computed by YOLO models.

In line with the unified approach advocated by the YOLO model, vehicle detection commonly employs image processing techniques to analyse images captured by cameras [1]. This methodological synergy facilitates vehicle monitoring, counting, speed calculation, and classification [2]. Researchers often select image processing methods for vehicle detection, a strategy that parallels the holistic image analysis approach utilised by YOLO models. The vehicle detection system recognises and tracks the vehicle using monitoring videos and displays visualisation reports. In the previous research, the researchers connected the model or algorithm method to object detection. The main reason is that the results will be shown in a monitoring system, also known as a visualised report. Monitoring systems help researchers analyse the data from image or video processing.

Rouf *et al.* [3] implemented a real-time vehicle detection, tracking, and counting system using the YOLOv7 algorithm, showcasing its capability to process video streams efficiently. The advantage of YOLOv7 lies in its ability to perform simultaneous detection, tracking, and counting of vehicles in real-time, which is essential for traffic management and surveillance applications. However, the study highlighted potential limitations such as the need for robustness in handling occlusions and varying lighting conditions, as well as the challenge of accurately distinguishing between different types of vehicles in crowded traffic scenarios.

Zeng *et al.* [4] experimented using an improved YOLOv7 algorithm, named YOLOv7-UAV, to enhance object detection in images captured by unmanned aerial vehicles (UAVs). The advantage of this improved algorithm includes superior accuracy and efficiency in detecting objects from aerial imagery, which is critical for applications such as surveillance and environmental monitoring. However, the study identified limitations such as the increased computational load and the need for high-quality datasets for optimal performance.

The connection between vehicle detection and near misses is crucial for understanding road safety. Vehicle detection technologies, using sensors or computer vision, recognise vehicles near each other. Near misses happen when vehicles almost collide but manage to avoid it, emphasising the need for precise and quick detection to prevent accidents. Studying near misses can alert drivers to dangers in traffic and enable improvements to vehicle detection systems for safer roads. Essentially, effective vehicle detection plays a key role in reducing the probability of near miss events on the road.

Traditionally, researchers get their empirical data from surveys or questionnaires. The empirical data is then analysed using mathematical models such as regressions, binomials, and variances, or software such as Excel and SPSS. Makizako *et al.* [5] proposed a study that employs a survey method to investigate the relationship between recent experience with near-miss traffic accidents among elderly Japanese drivers and analyses the data using the Bonferroni correction method and logistic regression. Terum & Svartdal [6] shows that the study aims to investigate concerns about safe driving behaviour and drivers' experiences with traffic accidents and near accidents. The frequency of accidents and near-accidents were used as variables for the driver safety index in an analysis of variance (ANOVA).

Following enhancements from traditional observation methods for near misses, the development of near miss calculation has been realized. According to the research of Lim *et al.* [7], near misses are detected through a monitoring system using YOLOv3, YOLOv3-tiny, YOLOv4, YOLOv4-tiny, YOLOv5, and Faster RCNN. The monitoring system uses advanced approaches such as YOLOv3, YOLOv3-tiny, YOLOv4, YOLOv4-tiny, YOLOv5, and Faster RCNN to identify near misses. Improvements have been made in the detection of near misses, increasing the length of video analysis from 80 seconds to 3 minutes while considering different video qualities [8].

The research of Wang *et al.* [9] developed YOLOv7 to enhance the architecture and optimization techniques to improve performance in object detection tasks. YOLOv7 has been utilized in a range of industries, such as surveillance, industrial automation, and autonomous vehicles, where accurate and timely object detection is critical [10]. The study of Yang *et al.* [11] combined YOLOv7 and Kalman filter to improve the accuracy and precision in object tracking and perform trajectory prediction.

Therefore, the primary focus of this study is to identify the vehicles in the mixed vehicle traffic and evaluate the near-miss events. This evaluation is conducted using a state-of-the-art, real-time object detection algorithm. In this study, a finely tuned YOLOv7 model is purposefully combined with the algorithm that tracks the vehicles and calculates the distance between vehicles to evaluate near miss events over a longer duration of experiment videos. Besides that, the near miss events are calculated automatically through the algorithm.

## 2. Methodology

### *2.1. Flowchart of study*

Figure 1 provides a summary overview of the entire research presented in this study. This project begins with Closed-Circuit Television (CCTV), which is given by Majlis Bandaraya Pulau Pinang (MBPP). The CCTV has two qualities of video: high-quality video and low-quality video.

After obtaining the videos, the videos are used as input to train and label the images; this process is called image labelling. Then, the convolutional neural network in the YOLOv7+G3HN model conducts the feature extraction. Feature extraction transforms raw data into a lower-dimensional form by identifying and retaining the most relevant information, enhancing model performance and efficiency. It reduces noise and simplifies data, making it more suitable for machine learning and pattern recognition tasks.

Next, the extracted important characteristics go into the detection layers, which are the backbone of YOLOv7+G3HN, to identify and classify the objects. The output comes out with multiple bounding boxes. Therefore, post-processing is employed where Non-Max Suppression (NMS) reduces redundancy in object detection by eliminating overlapping bounding boxes, ensuring each object is represented by a single, best-fitting box.

After training the YOLOv7+G3HN, the experiment conducts vehicle detection, which only detects cars and motorcycles in mixed-vehicle traffic. Besides that, near miss detection is also applied in this experiment to monitor and calculate the near-miss between vehicles.

### *2.2. Data description*

The road traffic videos in Penang were provided by MBPP. The videos were recorded in Bayan Lepas, which is an industrial area located in Penang. The high-quality video was captured on a school holiday, while the low-quality video was taken during the Chinese New Year. The quality of a video can be affected by its frame rate. A higher frame rate can increase a video's perceived smoothness and clarity, but a lower frame rate might make it choppy and blurry, affecting perception and processing [12].

The difference between using different lengths of videos in image processing experiments is primarily in the amount of visual information being presented in the outcome, the duration of the task, and the cognitive demands in the experiment. In a prior study by Lim *et al.* [13], video lengths of 20s, 40s, 60s, and 80s were used due to the limitations of computational resources. As the length of the video increases, real-time processing becomes more challenging due to the increased amount of data that needs to be processed, requiring more computational resources.

In this experiment, the 3-minute experiment videos are used to identify cars and near miss events in traffic and develop strategies to address them. The first step is to watch the video and identify any instances where two vehicles almost collide [14]. The next step is to analyse the causes of these events, such as driver error [15], poor road conditions [16], or other factors, and develop strategies such as improving driver training, installing better signage or road markings, and implementing new traffic regulations. Finally, it is important to evaluate the effectiveness of strategies by reviewing video footage and analysing traffic data [17]. By carefully analysing the video footage and developing strategies to prevent near miss events from occurring in the future, it can also help improve road safety and reduce the risk of accidents.

### *2.3. YOLOv7 model*

According to Wang *et al.* [9], the researchers introduced YOLOv7 as the latest version in the well-known series of real-time object detection models called YOLO (You Only Look Once). YOLOv7 follows a one-stage detection approach, performing object localization and classification simultaneously. This method enhances computational efficiency and makes it more suitable for real-time applications [18]. YOLOv7's architecture consists of three main components: the backbone, the neck, and the head. Figure 2 shows the structure of YOLOv7.

In object detection tasks, the accuracy and efficiency of YOLOv7 are significantly enhanced by the Extended Efficient Layer Aggregation Network (E-ELAN). This network plays a crucial role in improving feature fusion and gradient flow, which allows the model to learn more effectively. By addressing issues such as vanishing gradients, E-ELAN ensures that YOLOv7 can maintain strong performance even as the depth of the network increases [11]. According to the research of Zhao *et al.* [19], Hierarchical fusion is improved by Extended Efficient Layer Aggregation Network (ELAN-H). This enables better gradient flow, resulting in more robust training of intricate structures. In this study, the transformer of YOLOv7 (E-ELAN and ELAN-H) is replaced to develop a YOLOv7+G3HN model to improve the accuracy.

### *2.4. YOLOv7+G3HN Model*

The YOLOv7+G3HN model was combined by YOLOv7 with a Hornet transformer [20]. The backbone network of YOLOv7 is replaced by a model named G3HN [21]. G3HN utilises recursive gated convolution (gn Conv) and HorBlock, which provides several features of transformer and CNN, like input matching, long-range and high-level spatial interactions. Still, it can be efficiently realised within a convolutional framework; for instance, one can perform this using attention mechanisms without much extra cost on computation [22].

In the Hornet part, the HorNet Vision Transformer, specifically the high-order spatial attention, takes a more effective strategy for executing spatial interactions that are achieved by a fusion of convolution layers and fully connected layers [23].

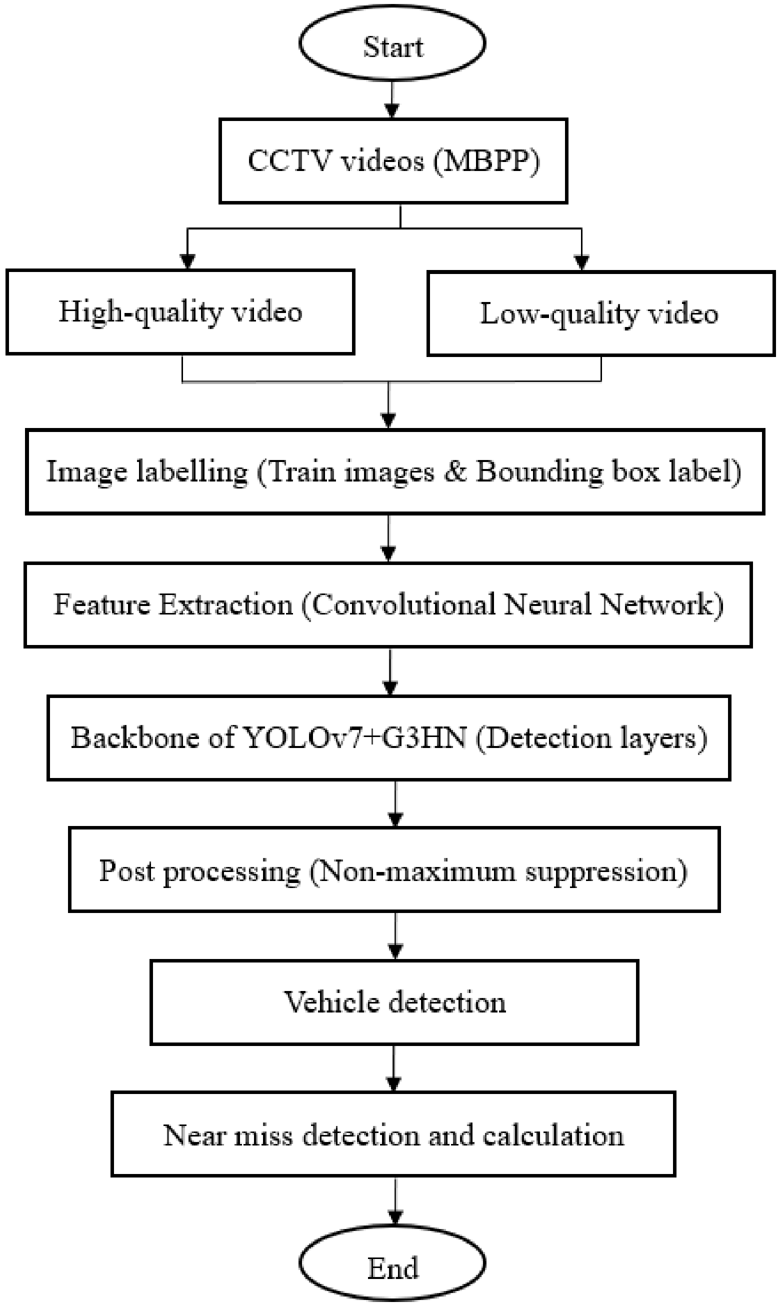


Figure 1. Methodology Flowchart.

The layer design is in line with previous transforms, where it replaced the self-attention sub-layer with Conv. This gn Conv consists of standard convolution; elemental product and linear projection enabling it to easily replace spatial mixture layers in the Vision Transformer [24].

$$\left[p_0^{HW*C}, q_0^{HW*C}\right] = \emptyset_{in} \in \mathbb{R}^{HW*C}. \quad (1)$$

$$p_1 = f(q_0) \odot p_0 \in \mathbb{R}^{HW*C}, \; y = \emptyset_{out}(p_1) \in \mathbb{R}^{HW*C}. \quad (2)$$

$$p_{k+1} = f_k(q_k) \odot g_k(p_k). \quad (3)$$

The basic operation of HorNet is $g^n$ Conv, where x $\in \mathbb{R}^{HW*C}$ is the input feature and $y = gConv(x)$ is the output. This is done through Equation (1) and (2) such that the linear projection layers corresponding to $\emptyset_{in}$ and $\emptyset_{out}$ perform channel mixing and $f$ is the depth convolution [25]. Afterwards, gating convolution is performed by the recursion equation (3) which then sends the final recursive step back to the projection layer to get $g^n$ Conv. From the recursive Equation (3), it can be seen that each step causes $p_k$to increase by 1. This shows that the $g^n$ Conv has nth-order spatial interactions [26]. The final step usually includes using f for deep convolution of modelling the feature concatenation for easier implementation and improved efficiency on the

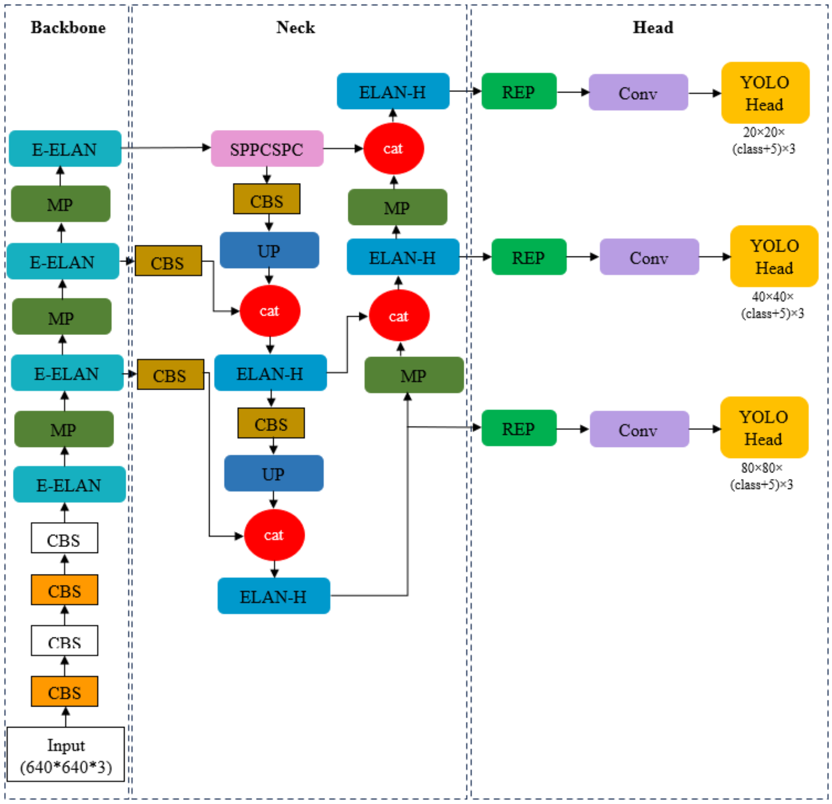


Figure 2. YOLOv7 structure.

GPU [27]. Figure 3 shows the YOLOv7+G3HN architecture which is defined by the incorporation of several elements designed to enhance the accuracy of object detection.

All these elements complement each other to enable the accurate detection of objects in complex scenes by the model. Using different parts and techniques makes this model better suited for the detection of objects than other versions of YOLOv7. Figure 4 shows the combination of each component in these elements.

These components include the idea behind the Contextual Attention Transformer (CAT) module, which is to focus on the attention of different parts of the scene [28]. This will make sure that the algorithm can highlight significant characteristics when identifying subjects. The Multi-scale Processing (MP) model is responsible for dealing with characteristics that occur in various sizes [29]. On the other hand, it is the responsibility of the Upsampling (UP) process to increase feature map dimensions so that small objects can be effectively recognised [30].

This G3 Convolution (g3Conv) convolutional layer incorporates the G3HN mechanism into the feature extraction process, improving model representation learning abilities [31]. Gated Grouping Hierarchical Network (G3HN) presents hierarchical characteristic aggregation plus gating group methods that ease contextual understanding loss while object linking in meanings [32]. In complex traffic scenarios, the Hornet system enhances its capacity for object recognition by focusing its efforts on critical areas such as occlusions and clutter.

The backbone of any object detection model is Convolution (Conv), which is typically a conventional convolutional layer that takes input pictures and extracts characteristics from them [33]. The convolutional layers of the VGG network design included Conv 3, 4, and 6, each with its capacity to hold features at different stages. For example, Conv3 holds three layers of convolution layers. The Convolutional-Maxpool-Convolutional (CMC) is a module in the architecture that consists of a sequence of a convolutional layer, followed by a max-pooling layer, and then another convolutional layer. It is used to extract features while reducing the spatial dimensions of the input feature maps [34].

The Spatial Pyramid Pooling with Contextual Spatial Pyramid Convolution (SPPCSPC) module was designed to effectively capture multi-scale features [35]. The module of Residual Encoding Pooling (REP) encodes the residual features and aggregates them spatially. Thus, the model improves its ability to handle feature representations.

The overall effectiveness of YOLOv7 + G3HN is contributed to by each of these components in detecting objects in traffic scenes with occlusions and varying scales. Nevertheless, even after these improvements, this model may have some drawbacks regarding the complexity of calculations as well as the variety of teaching samples that might have an impact on its capability to apply certain rules learned from one dataset onto a different dataset.

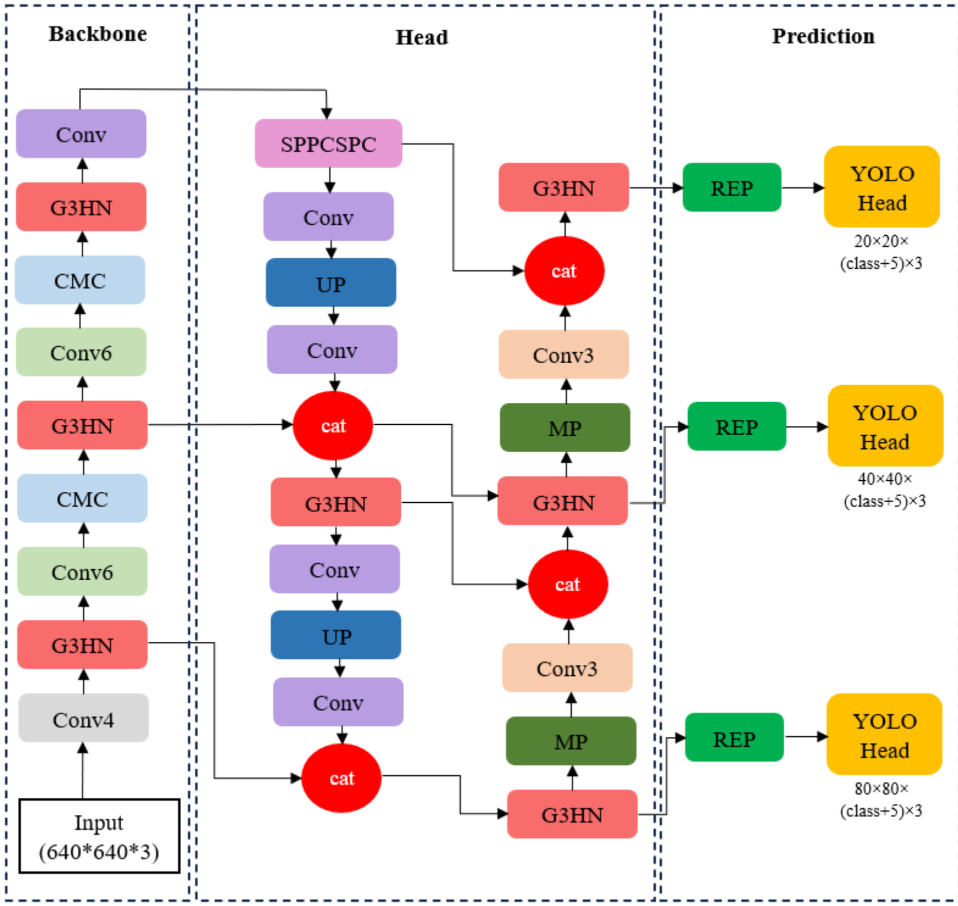


Figure 3. YOLOv7+G3HN structure.

Table 1. Comparison of high-and low-quality video in vehicle detection.

| Types of videos | High-quality video | Low-quality video |
|---|---|---|
| Computational time (s) | 126 | 139 |
| Error detection | 0 % | 0 % |
| Miss detection | 1.59 % | 9.38 % |

## 3. Vehicle detection

Vehicle detection, a crucial component of transportation and surveillance systems, relies on sophisticated image processing techniques to accurately identify vehicles within images or videos [36]. These techniques utilise machine learning models trained to recognise various types of vehicles. Through extensive labelling and training, these models learn to classify vehicles into distinct categories, including cars, motorcycles, bicycles, and more [37]. Once trained, the models can effectively analyse input images or videos and highlight the detected vehicles by drawing bounding boxes around them, facilitating their identification and tracking [38].

In the realm of vehicle detection, the YOLOv7+G3HN framework emerges as a prominent solution due to its advanced capabilities. YOLOv7, short for "You Only Look Once version 7," represents the latest iteration of the YOLO family of object detection models [39]. Known for its real-time processing speed and accuracy, YOLOv7 employs a unified approach that allows it to analyse the entire input image at once rather than dividing it into smaller sections [40]. Paired with G3HN, an enhancement module designed to improve the detection accuracy of small objects like vehicles, the YOLOv7+G3HN framework offers enhanced performance in vehicle recognition and detection tasks [21].

In the present study, the YOLOv7+G3HN framework serves as the cornerstone for vehicle recognition and detection. By leveraging the robust capabilities of YOLOv7 and the specialised enhancements provided by G3HN, researchers aim to achieve precise and efficient vehicle detection in various scenarios. This approach not only enables the accurate identification of vehicles but also facilitates their tracking and monitoring, thereby contributing to the development of effective transportation management and surveillance systems.

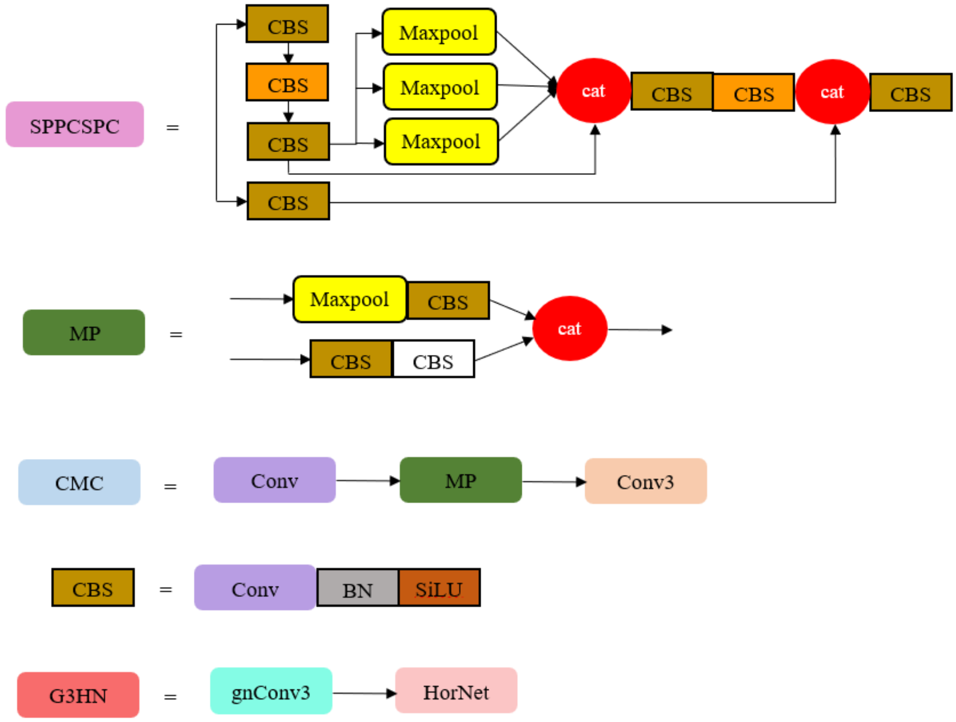


Figure 4. Combination of each component in the elements.

Table 2. Analysis of near miss events by YOLOv7+G3HN.

| Detection type of video | Near miss detection | Error detection | Miss detection |
|---|---|---|---|
| High-quality video | $\frac{2670}{15723} \times 100\% = 16.98\%$ | $\frac{0}{5400} \times 100\% = 0\%$ | $\frac{86}{5400} \times 100\% = 1.59\%$ |
| Low-quality video | $\frac{976}{4996} \times 100\% = 19.54\%$ | $\frac{0}{4500} \times 100\% = 0\%$ | $\frac{422}{4500} \times 100\% = 9.38\%$ |

### 3.1. Near miss

In previous research, near miss events are only detected through a monitoring system but they cannot be calculated automatically. Therefore, this experiment introduces YOLOv7+G3HN which is combined with Distance-Neighbours (DN) to identify near miss events, track vehicles and calculate near miss events through the algorithm. Figure 5 shows the flowchart of near miss calculation.

DN calculates the nearest vehicle in front of the target vehicle can be recognised as the one that is within a 2α degree sector range and at a distance R in front of the aimed vehicle in Equation (4).

$$\theta = arctan\sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}, \tag{4}$$

where the displacement coordinates of the aimed vehicle in two frames are represented by $x_i$ and $x_j$, and the target's displacement direction is determined by $\theta$. The appropriate values for $\alpha$ and $R$ in this study can efficiently determine the front vehicle's position based on the displacement direction [41].

The following formula and process explain the calculation of distance. Assume each vehicle as $R = \{R_k\}_{k=1}^{k}$, $k$ is the number of vehicles, the centre of the vehicle be $c_k = (x_k, y_k)$ and the speed of the vehicle as $\left(v_{xk}, v_{yk}\right)$. The pre-determined actual distance of the vehicle and the detected pixel length of the vehicle are used to calculate the relationship between the real distance and the pixel length [42]. The motion between the two frames is then calculated as the pixel distance using the centre coordinates of the front and back frames of each vehicle. The actual distance travelled by the vehicle between the two frames can be determined using this ratio and the pixel width [43]. Equation (5) shows the calculation of the velocity, $V$ involves dividing the distance by the time interval between the two frames.

$$V = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \times W(u) \,/\, \left[(x, y) \times \frac{fps}{5} \times 3.6 \times 3\right]. \tag{5}$$

Through this study, 1 second can collect 30 frames in high-quality video and 25 frames in low-quality video. Equation (6) illustrates a line connects the object detection results and a weight $\left(w_{ij}\right)$ is assigned to represent the intensity of the connection between the two vehicles. This weight is determined by the speed difference between the two vehicles within a given distance unit.

$$w_{ij} = \frac{v_{xi} - v_{xj}}{\sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}}. \tag{6}$$

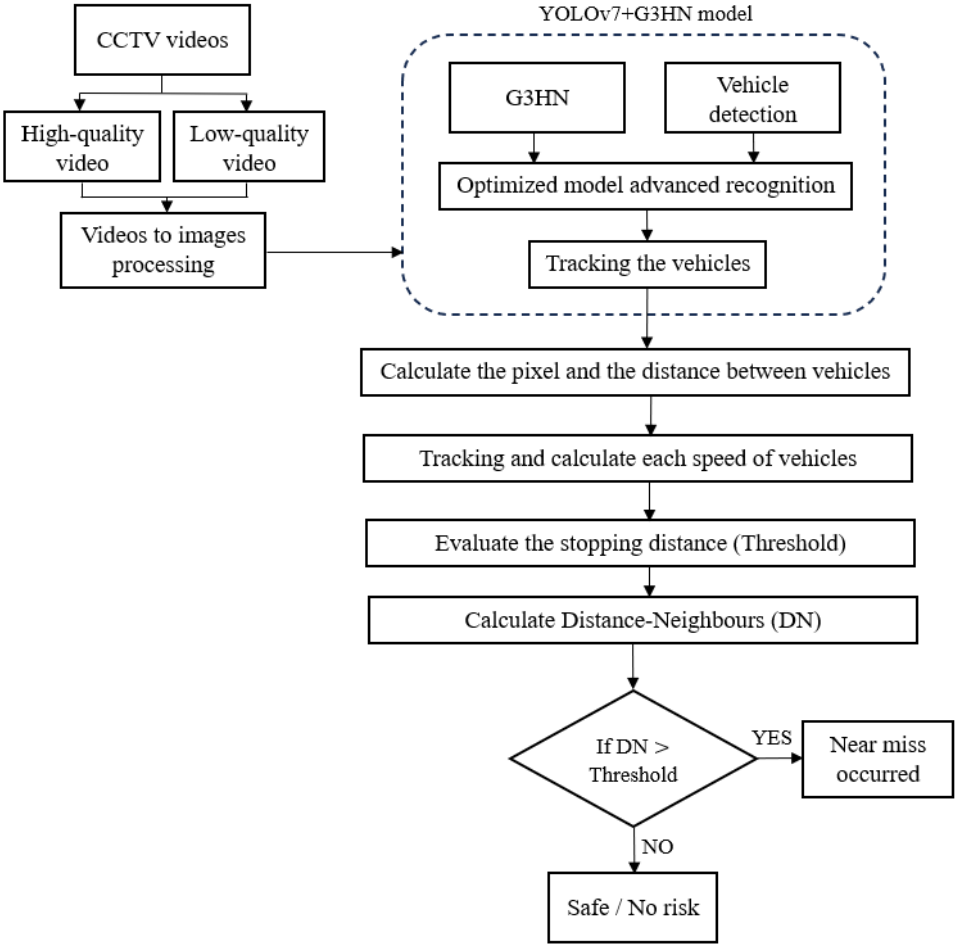


Figure 5. Flowchart of near miss detection and calculation.

Table 3. Comparison of the model result.

| Model | mAP@0.5 | mAP@0.95 | Times train/h | Model size/MB |
|---|---|---|---|---|
| YOLOv7 | 90.8 | 51.6 | 12.83 | 74.8 |
| YOLOv7+G3HN | 92.1 | 59.2 | 10.7 | 104.9 |

Equations (7), (8) and (9) displays $X_j$ calculates the difference in length both horizontally and vertically between the focused vehicles, $c_0$ and the vehicle, $c_j$ surrounding it [21].

$$X_j = \left( \frac{S_x^{0j}}{\frac{|v_{xj}|+|v_{x0}|}{2}}, \frac{S_y^{0j}}{\frac{|v_{yj}|+|v_{j0}|}{2}} \right)^T, \tag{7}$$

$$S_x^{0j} = |x_j - x_0| - \frac{(L_j + L_0)}{2}, \tag{8}$$

$$S_y^{0j} = |y_j - y_0| - \frac{(W_j + W_0)}{2}, \tag{9}$$

where $S_x^{0j}$ and $S_y^{0j}$ represents the longitudinal and the transverse distance between the two vehicles. $L_j$ and $L_0$ are the length of the aimed vehicle while $W_j$ and $W_0$ are the width of the focused vehicle.

Equation (10) shows the speed difference between two vehicles in real-life situations contributes to the distance between them, Distance Neighbour denoted as DN [44]. A more accurate evaluation of vehicle distance is created by this method of evaluation.

$$DN = \frac{f(X_j)}{Sigmoid\,(w_{ij})}. \tag{10}$$

Figure 6 shows the stopping distance is the combination of thinking distance and braking distance. Thinking distance is based on the driver's reaction time of 1.5 seconds assuming the driver's behaviour is normal in this study [45]. Braking distance is the distance from the driver pressing the brake until the vehicle stops [46]. The stopping distance is also used as a threshold value to determine the occurrence of near misses.

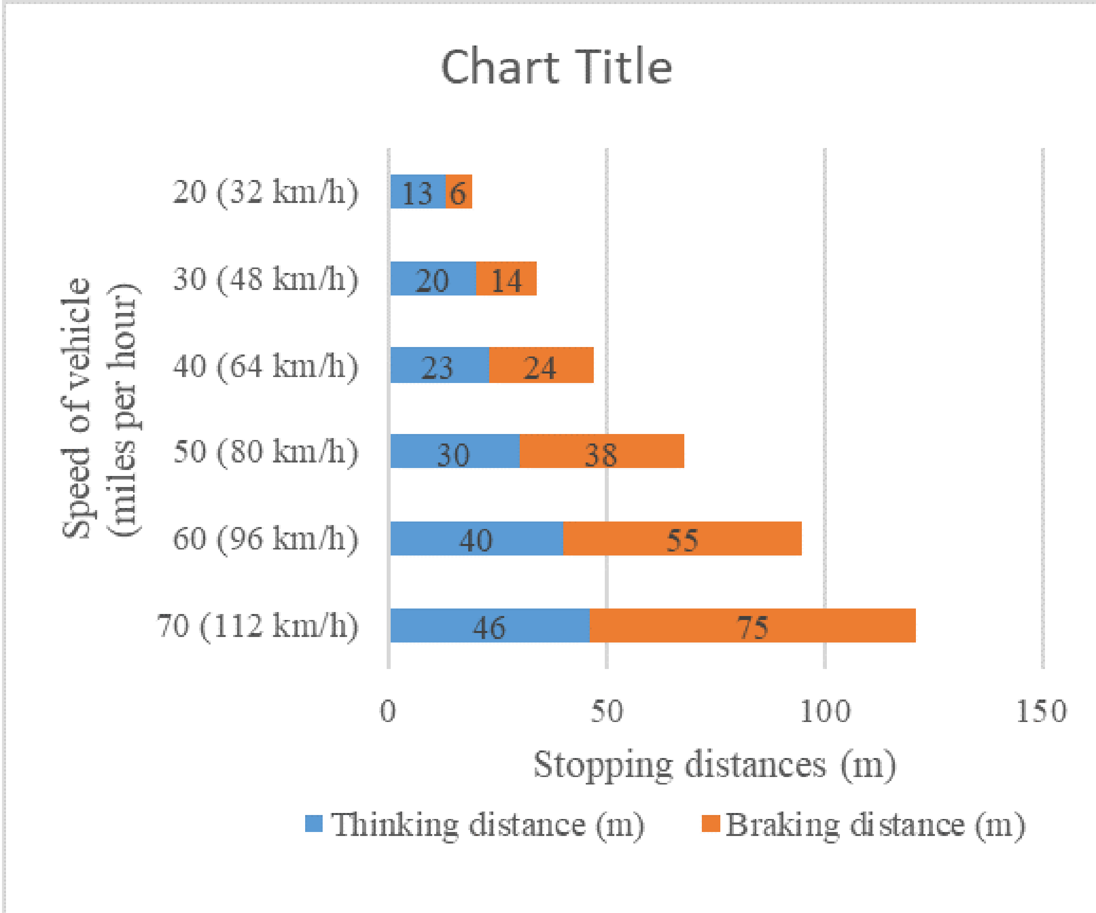


Figure 6. Threshold of stopping distances.

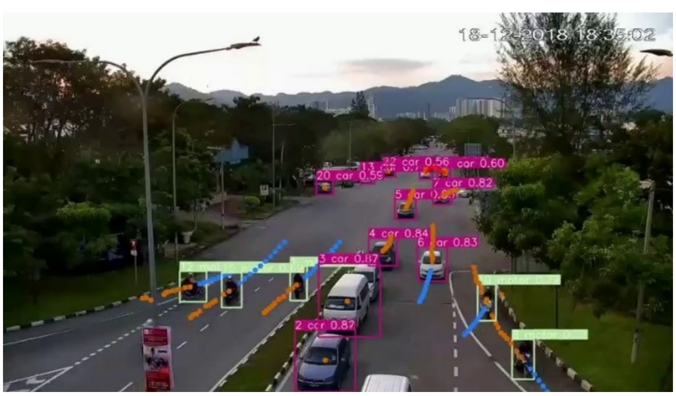

Figure 7. High-quality video.


Figure 8. Low-quality video.

## 4. Results and discussion

### 4.1. Vehicle detection

Vehicle detection was performed on the YOLOv7+G3HN in high-quality video and low-quality video correspondingly, the accuracy and speed of the model are evaluated.

According to the research by Lim *et al.* [8], it was previously concluded that the location with the highest probability of accidents is Lebuhraya Tun Dr. Lim Chong Eu. As a result, a high-quality video from that location was selected as one of the experimental videos. The high-quality video was recorded from 18:35:00 to 18:38:00 on 18th December 2018 at Lebuhraya Lim Chong Eu, and it captured various vehicles including cars, vans, motorcycles, and buses. As part of the experiment, a low-quality video situated at the junction near UMS was selected. Unlike the high-quality video, the low-quality video was recorded from a different day and location, spanning from 18:35:00 to 18:38:00.

Since the study can only detect cars, vans are considered as cars. Motorcycles and buses are not included in the car category. Figures 7 and 8 show the high-quality video and low-quality video which applied YOLOv7+G3HN for vehicle detection.

Table 1 presents a comparison of the time taken, error detection and miss detection for car detection by using YOLOv7+G3HN during the three minutes of experimental videos. Process time is a vital consideration when assessing the computational performance of different object detection models, especially in situations where prompt analysis is necessary or in real-time applications. High-quality video uses less time to compute vehicle detection than low-quality video.

Error detection is the percentage of chance the detection algorithm incorrectly classifies an object as a vehicle when it is not [47]. When comparing the error detection rates, the models of YOLOv7+G3HN show no errors in detecting cars in both quality videos. This shows the accuracy of YOLOV7+G3HN is very high and is not affected by the quality of the videos.

Additionally, miss detection refers to the percentage of instances the detection algorithm fails to recognise actual cars that are present on the scene [48]. The percentage of false negatives has an immediate impact on how accurate vehicle-detecting algorithms are overall. An increased rate of achievement in precisely detecting cars is shown by a decreased false negative rate [49]. The low-quality video displays a blur or unseen objects when the vehicle is far apart from the CCTV camera. This quality of videos could decrease the accuracy of the model detection. Therefore, the miss detection in low-quality video is higher than in high-quality video.

Aqqa *et al.* [50] emphasized the significance of video quality as a crucial factor, yet it often tends to be overlooked. In their research, they conducted tests using Faster RCNN, SSD, YOLO, and RetinaNet for object detection under various levels of video compression. The aim was to investigate the quality distortion caused by compression artefacts during video capture. After conducting a comparison, it is evident that the high-quality video is more suitable for car detection when compared to the low-quality video. The high-quality video demonstrates fewer error detections during the detection process [13].

### *4.2. Near miss detection*

In the previous studies regarding near miss events, the researcher used a questionnaire or survey form method to collect the near miss data because they lacked visual evidence [51]. Therefore, the researchers only can obtain the near miss record from people's experiences. After that, in the research of Lim *et al.* [7], the researcher used the algorithm to detect the vehicles and calculate the number of cases of near miss events from the monitoring system but there is a limitation where the calculation of near miss events is difficult.

Now, the experiment introduces advanced near-miss detection techniques using automatic calculation methods by employing enhanced models or fine-tuned YOLOv7 models. Table 2 shows the analysis of near miss risks and events in different types of videos using various indicators in YOLOv7+G3HN.

The near-miss detection is calculated automatically through the system, which is combined with the YOLOv7+G3HN models. However, machine learning has limitations; it can learn and detect particular objects but cannot identify errors detected and missed detection objects in the system. Therefore, the error detection and missed detection of near misses are calculated manually through the results of videos. The vehicle detection and near-miss detection results are combined into an outcome video, which is then split into frames [7, 8]. Finally, the number of error detections and missed detections is calculated manually.

In the analysis of error detection, an intriguing pattern emerged whereby both high- and low-quality videos exhibited no errors detected by the detection algorithm. This finding suggests a high level of reliability and accuracy in error detection across varying video qualities [52]. However, a notable contrast arose in miss detection, with the low-quality video registering a substantially higher percentage (9.38%) compared to the high-quality counterpart (1.59%). This disparity underscores the impact of video quality on the efficacy of miss-detection algorithms, highlighting the need for further investigation into factors influencing detection performance [53]. Additionally, since both vehicle detection and near miss detection are integrated into the same algorithm, similar trends were observed in vehicle detection, further emphasizing the significance of video quality in detection outcomes.

Upon closer examination of near miss detection outcomes, a compelling trend emerged wherein high-quality videos exhibited a lower probability (16.98%) compared to their low-quality counterparts (19.54%). This discrepancy suggests a nuanced relationship between video quality and the effectiveness of near miss detection algorithms. Notably, the absence of error-detected objects in the near-miss detection analysis underscores the pivotal role of missing detection in shaping detection outcomes [54]. These findings underscore the importance of considering video quality as a critical determinant in the development and evaluation of detection algorithms, with implications for enhancing detection accuracy and reliability in real-world scenarios [48].

From the analysis of near miss events on both-quality videos, this experiment can identify the blind spot's location through a monitoring system which contributes to near miss incidents and prevents this location become the blackspot location. Therefore, the YOLOv7+G3HN models play an important role in detecting and tracking cars and calculating the distance between the vehicles using DN indicators.

### 4.3. Comparison of models

To show that YOLOv7+G3HN has better accuracy than YOLOv7, the comparison models are conducted. Table 3 displays the accuracy, speed and size of the YOLOv7 and YOLOv7+G3HN.

This mAP@0.5 metric measures the model's accuracy in detecting objects with a threshold Intersection over Union (IoU) of 0.5. It reflects how well the model identifies and localizes objects. The higher the mAP@0.5 indicates better performance in object detection. YOLOv7+G3HN shows an improvement (94.1) compared to YOLOv7 (90.8), suggesting enhanced detection capabilities [55].

This mAP@0.95 metric measures the accuracy of the model with a more stringent IoU threshold of 0.95, requiring even more precise localization. The higher the mAP@0.95 values indicate superior precision in object detection. YOLOv7+G3HN achieves a better score (59.2) compared to YOLOv7 (51.6), demonstrating its improved precision and localization abilities [56].

This train time shows the time taken for the model to be trained. The lower values indicate longer training times per epoch or iteration. YOLOv7+G3HN has a lower value (10.7 hours) compared to YOLOv7 (12.83 hours), suggesting it takes longer to train. This could be due to the added complexity and enhancements in the YOLOv7+G3HN model that require more computational effort during training [57].

The model size represents the storage size of the trained model in megabytes. YOLOv7+G3HN has a larger size (104.9 MB) compared to YOLOv7 (74.8 MB). The increase in size is likely due to additional parameters and layers introduced in YOLOv7+G3HN to improve performance, which requires more storage [58].

The improvements in YOLOv7+G3HN result in superior detection accuracy, as indicated by higher mAP@0.5 scores, mAP@0.95 scores and shorter training times. However, these enhancements come with the drawbacks of increased model size. These trade-offs are generally acceptable in situations where high detection accuracy is crucial and sufficient computational resources are available to handle the additional complexity of the model. In such scenarios, the benefits of improved accuracy outweigh the costs of additional storage requirements.

## 5. Conclusion

This study aims to investigate the valuable insights garnered regarding the relationship between video quality and detection outcomes in the assessment of vehicle detection and near-miss analysis utilizing the YOLOv7+G3HN framework. The implementation of this framework facilitated a comprehensive examination of both the efficiency and accuracy of vehicle detection across diverse video qualities. Notably, findings revealed that high-quality videos demonstrated expedited computation times for vehicle detection relative to their low-quality counterparts, highlighting the pivotal role of video resolution in influencing the speed and efficacy of detection algorithms within monitoring systems.

Furthermore, the examination of error and miss detection mechanisms provided additional elucidation on the influence of video quality on detection precision. The results show no instances of error detection were identified across both high- and low-quality videos, indicating the robustness of the detection algorithm across varying video qualities. However, discernible disparities were observed in miss detection, with low-quality videos manifesting a heightened frequency of miss detections in comparison to high-quality counterparts. This observation suggests that diminished video quality may impede the accuracy of detection processes, potentially compromising the effectiveness of monitoring systems in the identification and tracking of vehicles.

Moreover, an analysis of near-miss events yielded compelling insights into the impact of video quality on detection outcomes. High-quality videos exhibited a diminished probability of near-miss occurrences when contrasted with low-quality counterparts, suggesting a plausible correlation between video quality and the incidence of near-miss events. The integration of error and miss detection within the YOLOv7+G3HN framework underscores the interrelated nature of these detection processes, underscoring the imperative for comprehensive evaluation and refinement of monitoring systems to ensure precise and reliable detection outcomes across varying video qualities. Overall, these findings underscore the pivotal role of YOLOv7+G3HN in enhancing detection accuracy and reliability, while emphasizing the necessity of factoring video quality as a critical determinant in monitoring system design and implementation.

In future research, YOLOv7 will be employed for vehicle detection and near-miss detection. Additionally, trajectory prediction will be incorporated to identify blind spots and reduce the likelihood of black spot occurrences. This approach aims to enhance overall traffic safety by leveraging advanced object detection and predictive analytics to proactively address potential hazards.

## Acknowledgement

The authors would like to thank the GFUSM, Universiti Sains Malaysia for their continuous support in this research.

## References

[1] A. Ajmal & I. M. Hussain, *Vehicle detection using morphological image processing technique*, proceedings of the 2010 International Conference on Multimedia Computing and Information Technology (MCIT), 2010 pp. 65–68. https://doi.org/10.1109/MCIT.2010.5444851.

[2] R. A. Hadi, G. Sulong & L. E. George, "Vehicle detection and tracking techniques: a concise review", Signal & Image Processing **5** (2014) 1. https://doi.org/10.48550/arXiv.1410.5894.

[3] M. A. Rouf, Q. Wu, X. Yu, Y. Iwahori, H. Wu & A.Wang, "Real-time vehicle detection, tracking and counting system based on YOLOv7", Embedded Selforganising Systems **10** (2023) 4. https://doi.org/10.14464/ess.v10i7.598.

[4] Y. Zeng, T. Zhang, W. He & Z. Zhang, "Yolov7-uav: an unmanned aerial vehicle image object detection algorithm based on improved yolov7", Electronics **12** (2023) 3141. https://doi.org/10.3390/electronics12143141.

[5] H. Makizako, H. Shimada, R. Hotta, T. Doi, K. Tsutsumimoto, S. Nakakubo & K. Makino, "Associations of near-miss traffic incidents with attention and executive function among older Japanese driver", Gerontology **64** (2018) 495. https://doi.org/10.1159/000486547.

[6] J. A. Terum & F. Svartdal, "Lessons learned from accident and near-accident experiences in traffic", Safety Science **120** (2019) 672. https://doi.org/10.1016/j.ssci.2019.07.040.

[7] L. M. Lim, S. Sathasivam, M. Ismail, A. Mohamed, A. Sufril, O. J. Ibidoja & M. K. M. Ali, "Comparison using intelligent systems for data prediction and near miss detection techniques", Pertanika Journal of Science & Technology **32** (2023). https://doi.org/10.47836/pjst.32.1.20.

[8] L. M. Lim, M. K. M. Ali, M. T. Ismail & A. S. A. Mohamed, "Data safety prediction using bird's eye view and social distancing monitoring for penang roads", Pertanika Journal of Science & Technology **30** (2022) 2563. http://dx.doi.org/10.47836/pjst.30.4.15.

[9] C. Y. Wang, A. Bochkovskiy & H. Y. M. Liao, *YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors*, In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 7464–7475. https://doi.org/10.48550/arXiv.2207.02696.

[10] G. Yasmine, G. Maha & M. Hicham, "Overview of single-stage object detection models: from Yolov1 to Yolov7", In *2023 International Wireless Communications and Mobile Computing (IWCMC)*, IEEE, 2023, pp. 1579–1584. https://doi.org/10.1109/IWCMC58020.2023.10182423.

[11] L. Yang, A. S. A. Mohamed & M. K. M. Ali, "An improved object detection and trajectory prediction method for traffic conflicts analysis", Promet – Traffic & Transportation **35** (2023) 462. https://doi.org/10.7307/ptt.v35i4.173.

[12] L. M. Wilcox, R. S. Allison, J. Helliker, B. Dunk & R. C. Anthony, "Evidence that viewers prefer higher frame-rate film", ACM Transactions on Applied Perception (TAP) **12** (2015) 1. https://doi.org/10.1145/2810039.

[13] L. M. Lim, A. F. M. Sadullah, M. Ismail, N. Awang & M. K. M. Ali, "Penang data safety prediction using database development and alternative road identification using Dijkstra approach", In AIP Conference Proceedings **2465** (2022) 040007. https://doi.org/10.1063/5.0078252.

[14] H. Kataoka, T. Suzuki, S. Oikawa, Y. Matsui & Y. Satoh, *Drive video analysis for the detection of traffic near-miss incidents*, 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018, pp. 3421–3428. https://doi.org/10.1109/ICRA.2018.8460812.

[15] P. M. Salmon, M. G. Lenné, N. A. Stanton, D. P. Jenkins & G. H. Walker, "Managing error on the open road: The contribution of human error models and methods", Safety science **48** (2010) 1225. https://doi.org/10.1016/j.ssci.2010.04.004.

[16] A. S. Karakaya, J. Hasenburg & D. Bermbach, "SimRa: using crowdsourcing to identify near miss hotspots in bicycle traffic", Pervasive and Mobile Computing **67** (2020) 1. https://doi.org/10.1016/j.pmcj.2020.101197.

[17] M. L. Siregar, H. R. Agah & F. Hidayatullah, "Near-miss accident analysis for traffic safety improvement at a 'channelized' junction with U-turn", International Journal of Safety and Security Engineering **8** (2018) 31. https://doi.org/10.2495/SAFE-V8-N1-31-38.

[18] D. Hu, S. Li & M. Wang, "Object detection in hospital facilities: a comprehensive dataset and performance evaluation", Engineering Applications of Artificial Intelligence **123** (2023) 106223. https://doi.org/10.1016/j.engappai.2023.106223.

[19] L. D. Quach, K. N. Quoc, A. N. Quynh & H. T. Ngoc, "Evaluating the effectiveness of YOLO models in different sized object detection and feature-based classification of small objects", Journal of Advances in Information Technology **14** (2023) 907. https://doi.org/10.12720/jait.14.5.907-917.

[20] K. Zhao, L. Zhao, Y. Zhao & H. Deng, "Study on lightweight model of maize seedling object detection based on YOLOv7", Applied Sciences **13** (2023) 7731. https://doi.org/10.3390/app13137731.

[21] Y. Rao, W. Zhao, Y. Tang, J. Zhou, S. N. Lim & J. Lu, "Hornet: Efficient high-order spatial interactions with recursive gated convolutions", Advances in Neural Information Processing Systems **35** (2022) 10353. http://arxiv.org/abs/2207.14284.

[22] Y. Fang, W. Wang, B. Xie, Q. Sun, L. Wu, X. Wang, T. Huang, X. Wang & Y. Cao, *Eva: Exploring the limits of masked visual representation learning at scale*, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 19358–19369. https://doi.org/10.48550/arXiv.2211.07636.

[23] Z. Chen, Y. Duan, W. Wang, J. He, T. Lu, J. Dai & Y. Qiao, *Vision transformer adapter for dense predictions*, In The Eleventh International Conference on Learning Representations, 2022. https://doi.org/10.48550/arXiv.2205.08534.

[24] J. Zheng, H. Wu, H. Zhang, Z. Wang & W. Xu, "Insulator-defect detection algorithm based on improved YOLOv7", Sensors **22** (2022) 8801. https://doi.org/10.3390/s22228801.

[25] W. Wang, H. Bao, L. Dong, J. Bjorck, Z. Peng, Q. Liu, K. Aggarwal, O. K. Mohammed, S. Singhal, S. Som & F. Wei, *Image as a foreign language: BEIT pretraining for vision and vision-language tasks*, In IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 19175-19186. https://doi.ieeecomputersociety.org/10.1109/CVPR52729.2023.01838.

[26] W. Yu, C. Si, P. Zhou, M. Luo, Y. Zhou, J. Feng, S. Yan & X. Wang, "Metaformer baselines for vision", IEEE Transactions on Pattern Analysis and Machine Intelligence **1** (2023) 1. https://doi.org/10.1109/TPAMI.2023.3329173.

[27] Q. Hou, C. Z. Lu, M. M. Cheng & J. Feng, "Conv2former: a simple transformer-style convnet for visual recognition", IEEE Transactions on Pattern Analysis & Machine Intelligence **01** (2022) 1. https://doi.org/10.48550/arXiv.2211.11943.

[28] J. Chen, S. Bai, G. Wan, & Y. Li, "Research on YOLOv7-based defect detection method for automotive running lights", Systems Science & Control Engineering **11** (2023) 2185916. https://doi.org/10.1080/21642583.2023.2185916.

[29] B. Tu, Q. Ren, Q. Li & W. He, "Hyperspectral image classification using a superpixel-pixel-subpixel multilevel network", IEEE Transactions on Instrumentation and Measurement **72** (2023) 3271713. https://doi.org/10.1109/TIM.2023.3271713.

[30] X. Gong, Q. Liu, R. Fu, R. Zhou, T. Xie & X. Chen, "GAM-YOLOv7-tiny and Soft-NMS-AlexNet: improved lightweight sheep body object detection and pose estimation network", Research Square (2024). https://doi.org/10.21203/rs.3.rs-4008441/v1.

[31] Q. Zhu, K. Ma, Z. Wang & P. Shi, "YOLOv7-CSAW for maritime target detection", Frontiers in neurorobotics 17 (2023) 1210470. https://doi.org/10.3389%2Ffnbot.2023.1210470.

[32] K. Liu, Q. Sun, D. Sun, L. Peng, M. Yang & N. Wang, "Underwater target detection based on improved YOLOv7", Journal of Marine Science and Engineering **11** (2023) 677. https://doi.org/10.3390/jmse11030677.

[33] L. Yang, A. S. A. Mohamed & M. K. M. Ali, "Traffic conflicts analysis in penang based on improved object detection with transformer model", IEEE Access **11** (2023) 84061. https://doi.org/10.1109/ACCESS.2023.3299316.

[34] R. Chang, S. Zhou, Y. Zhang, N. Zhang, C. Zhou & M. Li, "Research on insulator defect detection based on improved YOLOv7 and multi-UAV cooperative system", Coatings **13** (2023) 880. https://doi.org/10.3390/coatings13050880.

[35] Z. Pan, J. Cai & B. Zhuang, "Fast vision transformers with hilo attention", Advances in Neural Information Processing Systems **35** (2022) 14541. https://dl.acm.org/doi/abs/10.5555/3600270.3601327.

[36] Z. Yang, & L. S. Pun-Cheng, "Vehicle detection in intelligent transportation systems and its applications under varying environments: a review", Image and Vision Computing **69** (2018) 143. https://doi.org/10.1016/j.imavis.2017.09.008.

[37] Z. Luo, F. Branchaud-Charron, C. Lemaire, J. Konrad, S. Li, A. Mishra, A. Achkar, J. Eichel & P. M. Jodoin, "MIO-TCD: A new benchmark dataset for vehicle classification and localization", IEEE Transactions on Image Processing **27** (2018) 5129. https://doi.org/10.1109/TIP.2018.2848705.

[38] A. Jazayeri, H. Cai, J. Y. Zheng & M. Tuceryan, "Vehicle detection and tracking in car video based on motion model", IEEE transactions on intelligent transportation systems **12** (2011) 583. https://doi.org/10.1109/TITS.2011.2113340.

[39] M. Ahmed, N. El-Sheimy, H. Leung, & A. Moussa, "Enhancing object detection in remote sensing: a hybrid yolov7 and transformer approach with automatic model selection", Remote Sensing **16** (2023) 51. https://doi.org/10.3390/rs16010051.

[40] N. F. Alhussainan, B. Ben Youssef & M. M. Ben Ismail, "A deep learning approach for brain tumor firmness detection based on five different YOLO Versions: YOLOv3–YOLOv7", Computation **12** (2024) 44. https://doi.

org/10.3390/computation12030044.

[41] M. Ljung, Y. H. Huang, N. Åberg & E. Johansson, *Close calls on the road: a study of drivers' near-misses*. Proceedings of the 3rd International Conference on Traffic and Transportation Psychology, University of Nottingham, UK, 2004. https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-14258.

[42] R. Suwanda, Z. Syahputra & E. M. Zamzami, "Analysis of euclidean distance and manhattan distance in the K-means algorithm for variations number of centroid K", Journal of Physics: Conference Series IOP Publishing **1566** (2020) 012058. https://iopscience.iop.org/article/10.1088/1742-6596/1566/1/012058/meta.

[43] C. Ricotta & S. Pavoine, "A new parametric measure of functional dissimilarity: bridging the gap between the Bray-Curtis dissimilarity and the Euclidean distance", Ecological Modelling **466** (2022) 109880. https://doi.org/10.1016/j.ecolmodel.2022.109880.

[44] S. S. Mahmud, L. Ferreira, M. S. Hoque & A. Tavassoli, "Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs", IATSS research **41** (2017) 153. https://doi.org/10.1016/j.iatssr.2017.02.001.

[45] Q. Chen, H. Huang, Y. Li, J. Lee, K. Long, R. Gu & X. Zhai, "Modeling accident risks in different lane-changing behavioral patterns", Analytic methods in accident research **30** (2021) 100159. https://doi.org/10.1016/j.amar.2021.100159.

[46] T. Tang, K. Anupam, C. Kasbergenn & A. Scarpas, "Study of influence of operating parameters on braking distance", Transportation research record **2641** (2017) 139. https://doi.org/10.3141/2641-16.

[47] M. A. Bin Zuraimi & F. H. Kamaru Zaman, *Vehicle detection and tracking using YOLO and DeepSORT*, 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), 2021, pp. 23–29. https://doi.org/10.1109/iscaie51753.2021.9431784.

[48] P. Jagannathan, S. Rajkumar, J. Frnda, B. Divakarachari & P. Subramani, "Moving vehicle detection and classification using gaussian mixture model and ensemble deep learning technique", Wireless Communications and Mobile Computing **2021** (2021) 5590894. https://doi.org/10.1155/2021/5590894.

[49] A. Taheri Tajar, A. Ramazani & M. Mansoorizadeh, "A lightweight Tiny-YOLOv3 vehicle detection approach", Journal of Real-Time Image Processing **18** (2021) 2389. https://doi.org/10.1007/s11554-021-01131-w.

[50] M. Aqqa, P. Mantini & S. K. Shah, "Understanding how video quality affects object detection algorithms", Science and Technology Publications **5** (2019) 96. https://doi.org/10.5220/0007401600960104.

[51] K. D. Johnson, S. R. Patel, D. M. Baur, E. Edens, P. Sherry, A. Malhotra & S. N. Kales, "Association of sleep habits with accidents and Near Misses in United States Transportation Operators", Journal of Occupational & Environmental Medicine **56** (2014) 510. https://doi.org/10.1097/JOM.0000000000000132.

[52] S. Chikkerur, V. Sundaram, M. Reisslein, & L. J. Karam, "Objective video quality assessment methods: a classification, review, and performance comparison", IEEE transactions on broadcasting **57** (2011) 165. https://doi.org/10.1109/TBC.2011.2104671.

[53] V. A. Adewopo & N. Elsayed, "Smart city transportation: Deep learning ensemble approach for traffic accident detection", IEEE Access **12** (2024) 59134. https://doi.org/10.1109/ACCESS.2024.3387972.

[54] A. Rothblum, D. Wheal, S. Withington, S. A. Shappell, D. A. Wiegmann, W. Boehm & M. Chaderjian, *Improving incident investigation through inclusion of human factors*, Proceedings of the $2^{nd}$ International Workshop On Human Factors In Offshore Operations '02, Huston, Texas, 2002. https://digitalcommons.unl.edu/usdot/32/.

[55] Y. Xu, L. Zhu, Y. Yang & F. Wu, "Training robust object detectors from noisy category labels and imprecise bounding boxes", IEEE Transactions on Image Processing **30** (2021) 5782. https://doi.org/10.1109/TIP.2021.3085208.

[56] P. Wang, H. Liu, X. Zhou, Z. Xue, L. Ni, Q. Han & J. Li, "Multidimensional evaluation methods for deep learning models in target detection for sar images", Remote Sens. **16** (2024) 1097. https://doi.org/10.3390/rs16061097.

[57] T. Diwan, G. Anirudh & J. V. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications", Multimedia Tools and Applications **82** (2023) 9243. https://doi.org/10.1007/s11042-022-13644-y.

[58] G. S. Patel, A. A. Desai, Y. Y. Kamble, G. V. Pujari, P. A. Chougule & V. A. Jujare, *Identification and separation of medicine through Robot using YOLO and CNN algorithms for healthcare*, In International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), IEEE, 2023, pp. 1–5. https://doi.org/10.1109/ICAIIHI57871.2023.10489407.