IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 20 December 2024, accepted 28 February 2025, date of publication 5 March 2025, date of current version 24 March 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3548108

RESEARCH ARTICLE

Near Miss Detection Using Distancing Monitoring and Distance-Based Proximal Indicators

LEK MING LIM[1], LU YANG[2,3], WEN ZHU[1], AHMAD SUFRIL AZLAN MOHAMED[2], AND MAJID KHAN MAJAHAR ALI[1]

[1]School of Mathematical Sciences, Universiti Sains Malaysia (USM), Pulau Pinang, Penang 11800, Malaysia
[2]School of Computer Sciences, Universiti Sains Malaysia (USM), Pulau Pinang, Penang 11800, Malaysia
[3]Chengyi College, Jimei University, Xiamen, Fujian 361000, China

Corresponding author: Majid Khan Majahar Ali (majidkhanmajaharali@usm.my)

This work was supported in part by the School of Mathematical Sciences, Universiti Sains Malaysia; and in part by the Research University Team (RUTeam), Universiti Sains Malaysia, under Grant R502-KR-RUT001-0000000406-K134.

ABSTRACT Despite efforts to improve road safety, accidents persist due to insufficient evidence from manual police reporting, non-optimized detection algorithms, and technical limitations in real-time video processing and modelling. This study focuses on detecting and tracking vehicles within a monitoring system and analyzing near-miss incidents (black spot and unseen area), specifically examining the influence of video quality on detection performance using advanced model detectors (YOLOv4-tiny, YOLOv5, YOLOv7, and YOLOv7+CNeB). The experiment employed methods for vehicle detection through the monitoring system. Near-miss detection was conducted using two approaches: manual observation (Social Distancing Monitoring and Bird's Eye View) and automatic calculation (using DN indicators). Statistical methods, including descriptive statistics, and one-way ANOVA, were applied to compare datasets obtained from these indicators. The study concludes that YOLOv7+CNeB is effective for vehicle detection and near-miss analysis when video quality is considered in system design and implementation. YOLOv7+CNeB significantly reduces the time required to collect evidence from specific roads, provides visual reports, and addresses technical limitations in current algorithms. Future research should explore additional factors contributing to near-miss events, such as road environment, lane changes, and driver behaviours.

INDEX TERMS Near miss events, object detection, machine learning, transportation.

I. INTRODUCTION

Pulau Pinang has introduced several measures to improve traffic safety and encourage sustainability, such as installing closed-circuit television (CCTV) systems on the mainland and island. These CCTV systems are essential for observing traffic flow, detecting incidents, and enforcing traffic regulations, thereby enhancing road safety. The state government emphasizes implementing smart solutions and green technology as part of the Penang 2030 vision for a Green and Smart City. Policies like the Green Economy Policy for Industry (2020), Public Transit Oriented Transportation (2020), Renewable Energy Policy (2030), and Green Public Procurement (2030) reflect this commitment. Technologies, including the Internet of Things (IoT) and data analytics, are employed to optimize traffic management, reduce congestion, and improve road safety, while initiatives designed to reduce carbon emissions and increase energy efficiency promote an eco-friendly and sustainable transport system.

Moreover, Pulau Pinang's transport and traffic safety-related Sustainable Development Goals (SDGs) align with the state's policies to construct comprehensive plans addressing road safety and sustainable city development. This involves upgrading public transit, developing infrastructure, and establishing laws that support environmentally friendly transportation methods like walking and cycling. The state's SMEs also contribute to these initiatives by integrating sustainable practices into their operations, such as reducing waste, introducing energy-efficient technology, and encouraging eco-friendly transportation options for employees. The long-term objective of Pulau Pinang's transportation system is to enhance the quality of life for its citizens by focusing on

The associate editor coordinating the review of this manuscript and approving it for publication was Wenbing Zhao.

safety, sustainability, and innovation, with the state government collaborating with various stakeholders to achieve these goals. This initiative aligns with the Penang 2030 project, promoting environmental sustainability and improving living standards.

The problem statement highlights several key challenges in accident monitoring and prediction in Penang. The accidents reported in Penang are typically documented manually using POL 37 data in police records, which results in incomplete evidence and data, making effective analysis and decision-making difficult. Despite the widespread use of CCTV surveillance, advanced monitoring techniques for real-time observation and forecasting are insufficient.

One significant limitation is the variability in the functionality and quality of CCTV systems used for traffic monitoring. In real-world deployments, many CCTV cameras may not be fully operational due to technical malfunctions, limited maintenance, or inadequate coverage, leading to inconsistencies in data collection. Additionally, the quality of recorded video varies significantly across different locations, influenced by factors such as resolution, frame rate, and environmental conditions. Previous researches do not highlight the impact of video quality on detection accuracy, emphasizing the need for further investigation into how different levels of resolution, lighting conditions, and motion blur affect model performance. Furthermore, real-time application remains a challenge due to the computational demands required for processing high-volume video streams efficiently.

The development of reliable monitoring and prediction systems is hindered by the prevalent use of videos for insurance claims and technological limitations such as RAM shortages and high costs of video storage. The 45-day automatic deletion period of video footage stored by MBPP further restricts access to historical data. The quality of video footage also varies, with low-quality footage introducing noise and distortions that affect the accuracy of predictive models.

The Pulau Pinang state government promoted ''Penang 2030'' to build a smart city to reduce road fatalities. Despite the implementation of many regulations, the incidence of accidents in Pulau Pinang persists. The existing detection models like RCNN, CNN, and YOLO are not optimized for specific roadways in Penang, necessitating thorough investigations to create targeted preventative methods. Current algorithms for object recognition and data collection from CCTV videos are incomplete, requiring powerful processors and lacking the capability to capture data about vehicles on specific roads, which is crucial for traffic management and accident prevention.

Research gaps also emerge in the context of near-miss incidents, which are precursors to accidents. These incidents, influenced by factors like unseen areas, driver behavior, and lighting conditions, are often unreported and unrecorded. According to the safety triangle theory, a high frequency of near misses (300 incidents) typically precedes serious injuries and fatalities. However, traditional accident reporting systems fail to capture this data, limiting the understanding of hidden risks contributing to road congestion and accidents. Additionally, feature selection, parameter optimization, and predictive modeling for identifying black spots, unseen areas, and unsafe lane changes remain underexplored in Penang.

Given these gaps, the research motivation stems from the pressing need to develop a comprehensive system that leverages advanced object detection algorithms and predictive models to improve traffic safety and sustainability. This study addresses these issues through the following contributions:

(1) To determine and identify all possible factors (vehicle size, shape, colour, speed, camera resolution and environment factors) and parameters contributing to road accidents in Penang, including optimization of advance detection models.

(2) To classify and optimize the times by implementing an advanced monitoring approach using an object detection algorithm to enhance the effectiveness of low-and high-quality video footage of CCTV surveillance.

(3) To detect, measure and perform the accuracy of the near misses (unseen area and black spot) between vehicles in CCTV videos to capture near-miss events for better analysis and decision-making in road safety initiatives using Social Distancing Monitoring, Bird's Eye View and Distance-Neighbours (DN).

In the organization of this paper, we introduce the gaps in manual reports, traditional methods to identify accidents, vehicle detection and near-miss incidents in previous research in Section II. Then, the experiment data, the implementation of object detector models and manual observation method and the distance-based proximal indicators near miss detection method are discussed in Section III. In Section IV, the details of identifying possible parameters of vehicle detection and near miss detection and calculation are explained. Lastly, Section V is the concluding remarks that complete the paper.

II. MOTIVATION

This study introduces an advanced object detection framework tailored to the unique traffic patterns and constraints of Penang's roadways, setting a benchmark for context-specific traffic analysis. By integrating state-of-the-art YOLO models, road-specific data, and environmental parameters, the framework transforms low- and high-quality CCTV footage into actionable insights, enabling real-time identification of near-miss incidents and optimizing traffic safety measures. This system integrates Social Distancing Monitoring, Bird's Eye View perspectives, and a distance-based proximal indicator to quantify proximity dynamics, trajectory overlaps, and near misses, offering unprecedented precision in predicting collision likelihoods and recording traffic patterns. Furthermore, our research pioneers a data-driven evaluation of traffic networks, revealing their dual role in mitigating urban traffic congestion and reducing carbon footprints. This integrated approach reduces dependency on high-cost hardware, bridges the gap in underreported traffic risks, and provides a scalable

model for cities aiming to achieve environmental sustainability and enhanced mobility.

III. LITERATURE REVIEW

A. MANUAL REPORTS

The main concern in the transportation issue is regarding the accuracy, recent (real-time based) and reliable data. Currently, there is too much dependence on observing manual data compared to auto-calculating real-time data [1]. For instance, the baseline data, POL 37 which is a police report are the only manual confidential data that will be used by most of the insurance companies to make claims to theirs's clients. However, the data was not used to understand the road condition due to not solely enough information because of the inaccurate and missing data.

Reference [2] uses manual reports to assess the quality of raw data in vessel operations. Despite limitations such as being time-consuming and prone to human error, manual reports provide detailed and accurate context that enhances the overall data quality assessment process.

In the study of [3], the researchers utilize manual reports to supply detailed annotations and contextual data, enhancing the training and accuracy of deep learning models for automatic image caption generation. Although manual reports can be labor-intensive and prone to human error, they significantly improve data quality, leading to better model performance and capturing insights that automated methods alone might miss.

The manual reports typically document basic information such as the number of drivers involved, victims, the time and location of the accident, and the sequence of events leading to the accident. However, important parameters crucial for comprehensive analysis are often missing from these reports, rendering the data incomplete for researchers' needs.

The majority of researchers have substantiated that manual reports are time-consuming and prone to human error. These reports have often failed to convincingly establish the factors contributing to road accidents, often attributing accidents to driver behaviours without sufficient evidence. Therefore, integrating manual reports with visual documentation is necessary to provide robust evidence. This gap underscores the necessity of integrating manual reports with visual documentation, such as CCTV footage, to provide robust, objective evidence that can enhance accident analysis and prevention efforts.

In summary, the shortcomings of existing studies highlight the limitations of manual data collection methods, including inaccuracies, missing information, and susceptibility to human error. While manual reports can provide valuable contextual details, their labor-intensive nature and limited scope make them insufficient for comprehensive and real-time analysis. These limitations necessitate a hybrid approach that integrates manual reports with advanced technologies, such as automated data processing and visual evidence, to overcome these challenges and improve the quality and reliability of transportation data.

B. THE TRADITIONAL METHOD IDENTIFIES ACCIDENTS

Traditionally, parameter scaling during blackspot, unseen area, and other environmental evaluations was based on manually collecting information and basic statistical calculations. These approaches usually involve identifying factors like traffic volume, road structure, visibility conditions, and accident records. The parameters are then scaled using empirical data and local traffic limitations. Despite their methodical approach, traditional approaches frequently face limitations in capturing dynamic environmental elements, demanding the use of modern data analytics and real-time monitoring technology for thorough analysis and predictive modelling.

A black spot is a road location with a high incidence of traffic accidents due to factors like poor visibility, obscured intersections, and inadequate warning signs, necessitating enhanced safety measures and better road design. The study of [5] utilizes black spot analysis to identify high-risk locations on roads. The method involves analyzing historical data on road traffic accidents to pinpoint areas with a concentrated occurrence of incidents. The advantages of this approach include targeted interventions and improvements in road safety measures at identified black spots, potentially reducing accident rates. However, limitations may include reliance on historical data that may not capture current traffic patterns or emerging risks, necessitating regular updates and supplementary real-time monitoring for comprehensive safety improvements.

An unseen area is an area around a moving vehicle that mirrors can't show, posing a significant collision risk, especially during lane changes, and requiring drivers to check this zone manually to prevent accidents and injuries. The study of [6] utilizes a vehicle safety system designed for identifying unseen areas and navigating hilly areas. The systems may utilize sensors, cameras, or radar technologies to monitor unseen areas around the vehicle. The advantages likely include improved driver awareness and enhanced safety through real-time monitoring and alerts. Limitations may involve cost, technological complexity, and potential challenges in adverse weather conditions or rugged terrain affecting sensor performance.

Besides that, several environmental factors such as vehicle speed, steering angle, relative lane position, and camera angle adjustments are critical for accurate lane change detection and affect vehicle collision in dynamic conditions. The study of [7] utilizes driving simulator tests to evaluate lane change identification and camera angle adjustment methods. While simulator tests offer controlled conditions for systematic evaluation, findings should be interpreted cautiously due to potential differences from real-world driving environments and behaviours. The study underscores the importance of both accurate lane change detection and optimal camera positioning in enhancing driver awareness and safety in varying driving conditions.

The previous research primarily focuses on connecting blackspots, unseen areas and environmental factors with accidents but does not extend this correlation to include

near-miss incidents. Next, the data quality and completeness due to inconsistent data collection, underreporting, and lack of detailed accident information. Spatial and temporal resolution issues arise from aggregating data over large areas and long periods, obscuring specific trends and patterns. Manual data collection and analysis processes introduce time-consuming tasks and potential human errors. Additionally, these methods often fail to leverage advanced technologies like GIS, machine learning, and real-time data analytics, which could enhance their accuracy and effectiveness.

To summarize, the limitations of previous studies include a heavy reliance on historical data, manual collection methods, and basic statistical analysis, which often fail to capture dynamic environmental factors or emerging risks. Many studies lack the integration of advanced technologies like GIS, machine learning, and real-time monitoring, limiting their ability to provide comprehensive and predictive insights. Furthermore, there is insufficient exploration of near-miss incidents, inconsistent data collection, and the aggregation of data over large spatial and temporal scales, which obscure specific trends and patterns. These limitations highlight the need for a hybrid approach that combines traditional methods with modern technologies to improve the reliability, scalability, and precision of traffic safety evaluations.

C. VISION TRANSFORMERS

Vision Transformers are a type of deep learning model designed for image processing tasks, inspired by transformer architectures that have been highly successful in natural language processing. The key difference between traditional CNNs and Vision Transformers is that Vision Transformers treat an image as a sequence of patches, similar to how transformers handle text as a sequence of words. This allows Vision Transformers to capture long-range dependencies between pixels across the image, potentially leading to better performance on complex tasks.

The study of [84] focuses on multimodal data fusion for remote sensing image classification, specifically combining hyperspectral images (HSI) and LiDAR data. The proposed method, modality fusion vision transformer, utilizes a stackable modality fusion block, including multimodal cross-attention and spectral self-attention modules, to improve feature fusion and address issues of incomplete category representation and disjointed spectral and spatial information in HSI. The method improves classification accuracy by better utilizing both HSI and LiDAR data but may face challenges in computational complexity and model efficiency due to the integration of multiple attention modules.

The study of [85] focuses on traffic sign recognition in the field of autonomous driving, investigating whether Vision Transformers can outperform traditional convolutional neural networks (CNNs) for this task. The study evaluates the potential of Transformer-based models to achieve state-of-the-art performance in traffic sign recognition, similar to their success in other computer vision tasks. The approach may offer improved accuracy over CNNs but could face challenges in terms of model complexity and computational requirements.

The study of [86] focuses on automated brain tumor classification in the field of medical image analysis, specifically using Vision Transformers to diagnose brain tumors from T1-weighted Magnetic Resonance Imaging (MRI) scans. The method involves using pretrained and fine-tuned Vision Transformers models for a three-class classification task, evaluating their performance on a brain tumor dataset. While the approach leverages the powerful capabilities of transformer models, it may face challenges in terms of computational requirements and the need for large, high-quality datasets for training.

To summarize, the limitations of previous studies include the Vision Transformers have demonstrated significant potential across various domains, including remote sensing, autonomous driving, and medical image analysis. While these models show promise in improving classification accuracy and performance, they also face challenges such as high computational complexity, model efficiency, and the need for large, high-quality datasets. These limitations highlight the importance of addressing the computational demands and scalability of Vision Transformers for broader, practical applications.

D. VEHICLE DETECTION AND FEATURE EXTRACTION

Vehicle detection was a process that applied an image processing method to the vehicles captured in the camera. This procedure was used for vehicle monitoring, counting, calculating the average speed of each vehicle, and determining the category of the vehicle detected. Most of the researchers chose to use image processing to conduct vehicle detection.

The study of [8] employs a filter-based feature selection methodology to improve vehicle/non-vehicle classification. This method enhances computational efficiency and classification accuracy by selecting the most relevant features from the dataset. However, it assumes feature independence and may miss important feature interactions, potentially leading to suboptimal feature sets for complex classification tasks.

The study of [9] employs a method that integrates Feature Refinement with dung beetle optimizer (DBO) to optimize the Random Forest Regression Classifier for autonomous vehicle detection. This approach enhances detection accuracy and efficiency through adaptive and dynamic feature selection, although it may involve high computational complexity and technical implementation challenges. The method's success depends significantly on the quality of the training data used in the optimization process.

Many researchers integrate models or algorithms with software primarily for displaying results in monitoring systems, assisting in data analysis from image or video processing. However, gaps such as imbalanced datasets and underrepresented vehicle types significantly limit algorithm

performance in accurately detecting vehicles. Urban environments present substantial challenges with complex backgrounds, occlusions, and variable lighting conditions, hindering reliable detection methods. Real-time processing, crucial for applications like autonomous driving, faces obstacles due to computational intensity and latency issues. Traditional methods also struggle with robustness to severe occlusions, impacting their effectiveness in challenging scenarios.

In summary, the shortcomings of existing studies include reliance on methods that assume feature independence, which may overlook critical interactions; computational complexity and dependency on high-quality training data; and limitations posed by imbalanced datasets and underrepresented vehicle types. Urban environments introduce challenges like occlusions, variable lighting, and complex backgrounds, which traditional detection methods fail to address comprehensively. Additionally, real-time applications face obstacles due to high computational demands and latency issues. These limitations emphasize the need for robust approaches, such as integrating Convolutional Neural Networks (CNN) and Region-Based Convolutional Neural Networks (RCNN), to overcome these challenges and enhance detection accuracy and efficiency.

E. NEAR MISS

Near misses happen when vehicles almost collide but manage to avoid them, emphasizing the need for precise and quick detection to prevent accidents. By studying near misses, drivers can be alert to the 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.

The study of [15] utilizes a VR-integrated simulation approach to assess bicycle-vehicle conflicts at urban intersections, aiming to identify near misses and blackspots. While offering advantages such as safety in controlled environments and detailed analysis capabilities, the approach faces challenges related to simulation realism, resource intensity, and the need for validation against real-world data and ethical considerations.

The study of [16] investigates dangerous scenarios involving vehicles turning right and pedestrians in left-hand traffic environments. To identify near misses and unseen areas, the researchers likely employed observational methods, possibly using video analysis or simulation techniques to analyse scenarios where collisions were narrowly avoided. The advantages of this approach are capturing real-world interactions and providing insights into critical safety issues at intersections. By the way, the method may be limited by the availability of detailed and comprehensive observational data, as well as challenges in accurately quantifying near misses. This study faces difficulties in replicating complex traffic scenarios and ensuring the generalizability of findings across different urban settings and traffic conditions.

The study of [17] utilizes a video-based tool for evaluating near-miss crashes, focusing on lane change identification. This method offers detailed visual analysis and objective data collection, although it is dependent on recording quality and may have limitations related to the field of view and privacy concerns.

There have been limited studies combining the near-miss events with the YOLO series and Faster RCNN. In addition, no research is being undertaken in Penang. The majority of researchers conduct surveys and questionnaires to make data gathering easier. The lack of accident records and statistics is the basis behind this. As a result, to complete the data, they can only collect people's experiences. Therefore, this study will show reports in visualization reports instead of manual reports.

To summarize, the limitations of previous studies include challenges with simulation realism, resource intensity, and the need for validation against real-world data. Additionally, many studies are limited by the availability of comprehensive and detailed observational data, the difficulty in accurately quantifying near misses, and issues with the generalizability of findings across different traffic conditions. Further limitations include dependency on recording quality and privacy concerns, and the lack of integration between near-miss detection and modern object detection algorithms like YOLO and Faster RCNN. This study aims to address these gaps by providing visual reports instead of relying on manual reports and focusing on the specific context of Penang, where research on near misses is currently lacking.

IV. EXPERIMENTAL SETUP

The simulation was meticulously crafted to thoroughly investigate the performance of various object detector models across a range of video quality conditions. This detailed examination aimed to assess the models' effectiveness in both vehicle detection and the identification of near-miss incidents. By utilizing videos of different resolutions and quality levels, the study sought to understand how these factors impact the accuracy and reliability of the detection algorithms.

A. EXPERIMENT DATASETS

The baseline data will be obtained from POL 37 records. By analysing this baseline data, types of accidents such as near misses, fatalities, crash analyses, and black spots can be identified. Next, Jabatan Kerja Raya (JKR) and MKJR will collect additional evidence to justify and secure funding from the federal government. Following this, discussions will take place in the next Rancangan Malaysia (RMK). This process typically takes about 5 to 10 years to complete or repair the specified roads. Using current technology, such as data visualization, is expected to reduce costs, time, and resources. The database contains POL 37 records and serves as the baseline data. However, POL 37 is a manual report that is often incomplete or has significant missing data.

Therefore, MBPP collaborated with USM to solve traffic issues by using an advanced approach or machine learning to form updated data in visual reports. The road traffic videos in Penang were provided by MBPP. The high-quality video was recorded in Bayan Lepas, which is an industrial area located in Penang while the low-quality video was captured at the junction near Jalan Sungai Dua, USM. Both videos are captured during peak sessions where 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.

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 [1], video lengths of 20s, 40s, 60s, and 80s were used due to 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.

The studies of [19] and [20] have noted that longer videos or higher frame rates can significantly increase computational demands, which can pose a challenge for real-time video analysis. The studies suggest that specialised hardware or optimised algorithms may be required to process longer videos in real-time. The limitations of the hardware resources of the laptop used in the experiment caused the algorithms to become increasingly computationally intensive and slow down over time. In contrast, this current project employs 3-minute videos to demonstrate the advancements in the algorithm's performance.

The experiment video with a 3-minute duration might be used to train object detection models such as the YOLO [21] and RCNN series [22] to recognise and categorise objects in traffic such as vehicles, pedestrians, bicycles, and other things. These models may be used to detect near-miss incidents in traffic by evaluating video data and recognising objects that are in close proximity or likely to be on a collision track [23]. However, it is critical to thoroughly construct and evaluate these algorithms to ensure their accuracy and usefulness in real-world scenarios.

The 3-minute experiment videos can be used to identify 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 [24]. The next step is to analyse the causes of these events, such as driver error [25], poor road conditions [26], 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 [27]. 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. Table 1 shows the details of the experiment data.

TABLE 1. Experiment data.

High-quality video	**Quality of videos**	Low-quality video
18th December 2018 (Tuesday)	**Date**	5th February 2019 (Tuesday)
6:35:00 p.m. – 6:38:00 p.m.	**Time**	6:35:00 p.m. – 6:38:00 p.m.
30 fps	**Frame per second**	25 fps
Persimpagan Jalan Kampung Jawa - Lebuhraya Tun Dr Lim Chong Eu	**Place**	Persimpangan Jalan Sungai Dua dekat USM

B. YOLOV4-TINY

YOLOv4-tiny is a downsized version of YOLOv4. Based on YOLOv4, YOLOv4-tiny is proposed to simplify the network architecture and minimise parameters, making it suitable for implementation on smartphones and Internet of Things (IoT) systems [30]. The study of [31] states that YOLOv4-tiny is more rapid in training and detection. It contains only two YOLO heads, as compared to three YOLO heads in YOLOv4, and it was developed using 29 pre-trained convolutional layers, as compared to 137 in YOLOv4. Figure 1 shows the network structure of YOLOv4-tiny.

C. YOLOV5S

Unrelated to the official YOLO versions, YOLOv5s is an independent implementation of the YOLO algorithm created by another group of developers [4]. It differs from YOLOv3-tiny and YOLOv4-tiny because of its unique architecture and codebase [32].

The modular design of YOLOv5s is one of its outstanding features since it enables users to easily customize and modify the model to their requirements [33]. Due to their widespread use and ease of deployment, pre-trained models are now easily accessible for a range of object identification tasks [34]. Labelled data was supplied during training to improve the algorithm's ability to recognize objects [35]. The initial step was to refine the pre-trained models for the particular dataset [36]. YOLOv5s achieves competitive accuracy and inference speed, making it suitable for object identification experiments and real-time applications [37].

Figure 2 describes the configuration of YOLOv5s. Three main components contribute to the core of the YOLOv5s

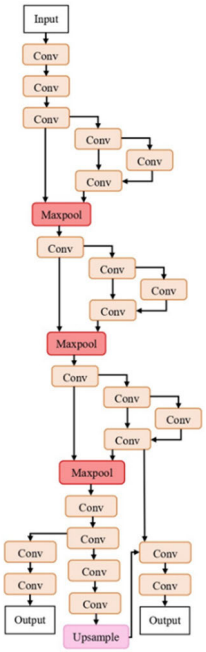


FIGURE 1. Backbone of YOLOv4-tiny.

network model: the backbone, the neck, and the YOLO head. Features from multiple images of different scales are extracted by the backbone network [38]. Features from various scales are combined by the neck, which is the feature pyramid network, and sent to the detection network [39]. Using the features of the images, the YOLO head, also referred to as the detection network, generates the object bounding box and predicts the category of the object within it [40].

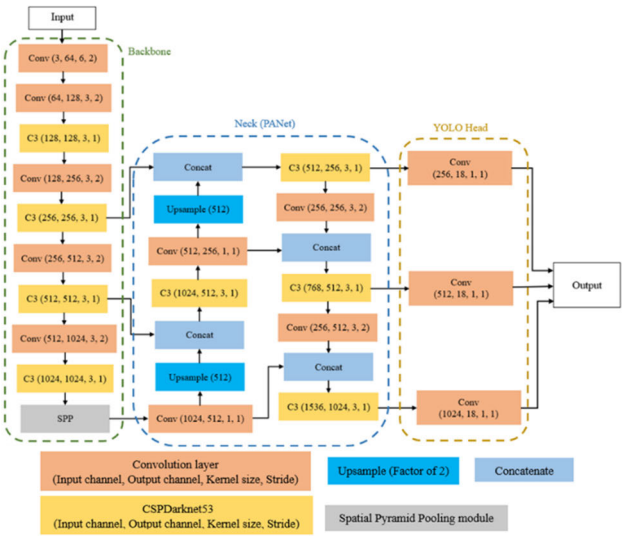


FIGURE 2. YOLOv5S network structure.

A module in YOLOv5s is an individual building block that manages data input. The number of input feature channels is determined by the input channels, which affects the level of data complexity that the module can handle [35]. The quantity of output feature channels, which controls the number and variety of learnt features, is determined by output channels [41]. The kernel size of the convolutional filter determines the dimensions of the convolutional filter, influencing the size of the receptive field and the types of information that may be retrieved [40]. Stride affects the spatial resolution of the output feature map and the model's capacity to detect specific details or large trends by determining the step size of the kernel as it moves over the input [42].

In the YOLOv5s configuration, several key elements support the comprehensive model understanding of images. A basic approach called convolution scans and extracts important visual features, assisting edge identification and pattern recognition [37]. The CSPDarknet53 Module improves data understanding by separating complex elements into manageable parts and repeating them at different rates [36]. The SPP (Spatial Pyramid Pooling) allows for the simultaneous analysis of several scales in an image, capturing both the larger context and specific details [32]. Similar to zooming in on images, upsampling increases the size of the image to reveal finer details [43]. Concatenation combines many data sets, much like putting together a puzzle to gain a comprehensive understanding [4].

D. YOLOV7

In this study, lightweight models such as YOLOv4-tiny and YOLOv5s are applied to perform car detection. Additionally, advanced or complex models, such as YOLOV7 and fine-tuned YOLOv7, are introduced. YOLOV7 and YOLOv7+CNeB are employed for both car detection and near-miss detection in this study.

According to the research of [44], the researchers developed the YOLOv7 as the latest iteration in the famous collection of real-time object detection models called YOLO (You Only Look Once). YOLOv7 tracks the one-stage detection method, showing that it localizes objects and does classification at the time, which promotes computational efficiency and real-time applications better [45]. YOLOv7 contains three parts: backbone, neck and head. Figure 3 shows the structure of YOLOv7.

YOLOv7's backbone network is a modified version of Efficient Convolutional Neural Network (EfficientNet) design which is well suited for quick detection of important features [46]. At different scales, features from the input image are taken out by the backbone which then allows the neck and head to detect objects and classification.

The YOLOv7 network structure consists of three main components: the Backbone, Neck, and Head. The Backbone, responsible for feature extraction, employs multiple E-ELAN and CBS modules, interspersed with MP layers, to enhance spatial information retention and hierarchical feature representation [44]. The Neck, incorporating SPPCSPC and ELAN-H modules, refines the extracted features through

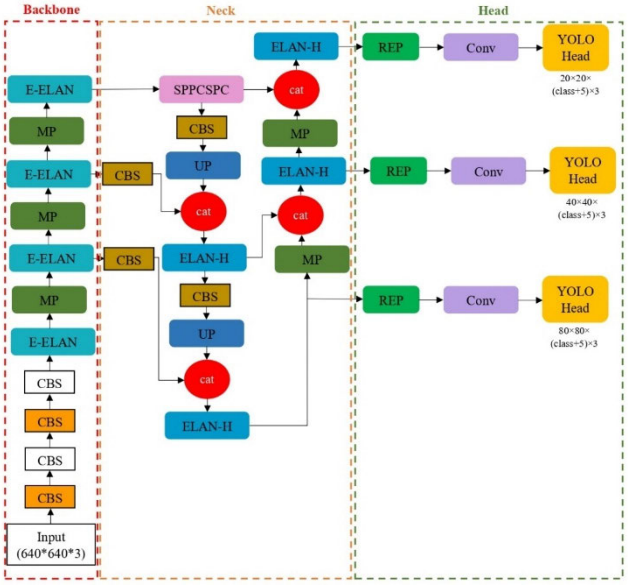


FIGURE 3. YOLOv7 network structure.

up-sampling, concatenation, and CBS operations to facilitate multi-scale feature fusion [49]. Finally, the Head comprises REP and convolutional layers that process the refined features and generate multi-scale predictions at different resolutions (20×20, 40×40, and 80×80), enabling accurate object detection across varying object sizes [48]. This structured approach enhances YOLOv7's detection performance by efficiently leveraging feature hierarchies and multi-scale representations.

The deep convolutional neural network of the YOLOv7 Backbone is designed to effectively extract features from input images and incorporates improvements to enhance its performance [47]. The Multi-scale Processing (MP) model enables the network to deal with objects of different sizes by extracting features from multiple scales, often utilizing methods like Feature Pyramid Networks (FPNs) and Spatial Pyramid Pooling (SPP). In object detection tasks, the accuracy and efficiency of YOLOv7 are improved by the Extended Efficient Layer Aggregation Network (E-ELAN) which is responsible for enhancing feature fusion and gradient flow enabling the model to learn more and addressing problems like vanishing gradients [48].

For better object detection, YOLOv7's neck refines features from the backbone. The Spatial Pyramid Pooling with Contextual Spatial Pyramid Convolution (SPPCSPC) module combines SPP with contextual convolutions enhancing multi-scale feature extraction and space context understanding [67]. Equation (1) shows Spatial Pyramid Pooling (SPP):

$$O_{SPP} = [P_1(X), P_2(X), \ldots, P_n(X)]. \tag{1}$$

where O_{SPP} is the output of SPP and $P_i(X)$ represents pooling operation at scale i. Equation (2) shows Contextual Spatial Pyramid Convolution (CSPC) as follows:

$$O_{CSPC} = \sum_{i=1}^{n} C_i(P_i(X)). \tag{2}$$

where O_{CSPC} is the output of CSPC, C_i represents convolution operation at scale i and $P_i(X)$ pooled feature map at scale i.

In more complex surroundings, the Contextual Attention Transformer (CAT) is perfect for seizing contextual information and enhancing detection accuracy by focusing on significant features [49]. Equation (3) shows the self-attention mechanism formula [50]:

$$Attention(Q, K, V) = softmax\left(QK^T/\sqrt{d_k}\right)V. \tag{3}$$

where Q, K, V are the query, key, and value matrices, respectively. The d_k representing the key vector dimension. The attention weights are computed by the softmax function by scaling the dot product of Q and K^T by $\sqrt{d_k}$ which helps prevent gradients from becoming too small during training.

According to the research [35], Hierarchical fusion is improved by the Extended Efficient Layer Aggregation Network (ELAN-H). This enables better gradient flow, resulting in more robust training of intricate structures. Equation (4) shows the formula of Feature Aggregation in ELAN-H:

$$F_{out} = \sum_{i=1}^{n} W_i \cdot F_i. \tag{4}$$

where F_{out} are the aggregated features, W_i represents the weight for the feature map from layer i and F_i is the feature map from layer i.

The head parts of YOLOv7 enhance the features of the neck into output for bounding boxes, scores of confidences, and class probabilities. For better representation of the features, Residual Encoding Pooling (REP) pools feature preserving important information by encoding residual information while reducing dimensionality [54]. Equation (5) shows Residual Encoding Pooling (REP) as follows:

$$F_{REP} = P(F) + F. \tag{5}$$

where F_{REP} is output from REP, $P(F)$ is the pooled feature map and F is the original feature map (residual connection).

Convolutional layers perform convolutional processing to aggregate the properties, refine them for the final prediction phase and extract these features from the pooled output for further processing [54]. Equation (6) shows convolutional layer output is

$$O = \left[\frac{(I - K + 2P)}{S}\right] + 1. \tag{6}$$

where O is the output dimension, I is the input dimension, K is the kernel size, P is the padding and S is the stride.

After the image is divided into a grid, the final predictions are made by the YOLO head such as bounding box coordinates, confidence scores, and class probabilities among others which help in ensuring both correctness and efficiency in object detection [46]. Therefore, the YOLO head made the prediction result as the dimension with filter. The filter

number is configured directly in the layer of the convolutional network [51]. Equation (7) shows the formula of the filter number is

$$\begin{aligned} Filters &= (classes + w + h + x + y + Confidence\ score) \\ &\quad \times num \\ &= (classes + 1 + 1 + 1 + 1 + 1) \times num \\ &= (classes + 5) \times num. \end{aligned} \tag{7}$$

where 5 represents the conditions (x, y - position of boundary box, w, h - width and height of the image and C - the Confidence score), c is the class probability. Equation (8) shows the YOLO detect 3 boxes per grid cell, so the

$$Filters = (classes + 5) \times 3. \tag{8}$$

E. YOLOV7+CNEB

In the study of [52], the researchers created the ConvNeXt Block (CNeB) in 2021 to enhance visual recognition. In this study, the YOLOv7+CNeB are employed to boost object detection ability. Figure 4 represents the architectural framework of YOLOv7+ CNeB, which embodies the fusion of YOLOv7 with the ConvNeXt Block module.

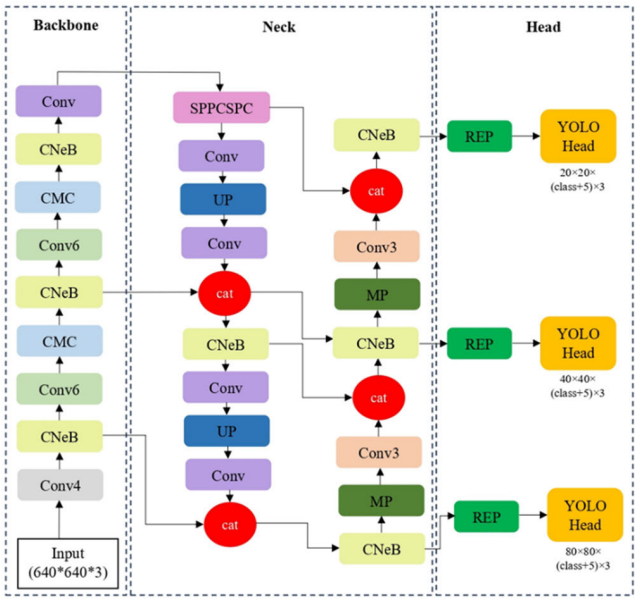


FIGURE 4. YOLOv7+CNeB architecture.

The integration of the ConvNeXt Block (CNeB) into YOLOv7 enhances the feature extraction process by improving spatial and channel-wise feature representation while optimizing computational efficiency. Unlike the original YOLOv7 feature extraction module, which relies on group convolution for channel splitting before separate convolutions, CNeB employs depthwise convolution for each channel, allowing for more efficient feature aggregation without the need for complex normalization or activation functions [53]. Additionally, CNeB incorporates LayerNorm (LN) to stabilize inputs and GELU activation to facilitate learning of complex patterns [54]. This design refines feature hierarchies, enhances the extraction of fine-grained spatial details, and improves the model's ability to capture long-range dependencies. By embedding CNeB into YOLOv7, the model achieves a more robust feature representation, leading to improved detection accuracy, particularly for complex object structures and occluded instances.

Furthermore, a residual connection through concatenation is also present in CNeB enhancing gradient flow and preserving features. On the whole, G3HN mixes complexity and efficiency on the one hand, while on the other hand, CNeB offers better provision of features at a probable increased computational expense [55]. Figure 5 shows the structure of CNeB which is implemented in YOLOv7.

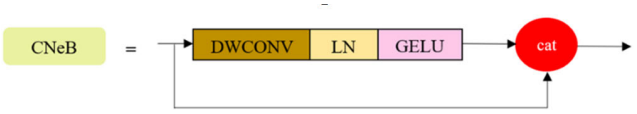


FIGURE 5. Structure of CNeB.

The CNeB module in YOLOv7 is designed to improve feature extraction and processing by utilizing depth-wise convolution (DWCONV) for the efficient processing of channels [56]. Equation (9) displays the depth-wise convolution and applies a separate convolutional filter to each input channel.

$$Output_{i,j,k} = \sum_{m,n} Input_{i+m,j+n,k} \cdot Filter_{m,n,k}. \tag{9}$$

where $Output_{i,j,k}$ is the output feature map at the position (i, j) for channel k, $Input_{i+m,j+n,k}$ is the input feature map at the position $(i + m, j + n)$ for channel k and $Filter_{m,n,k}$ is the depth-wise filter applied to channel k.

Layer Normalization (LN) for stabilizing and accelerating training, and the Gaussian Error Linear Unit (GELU) nonlinear activation to enhance learning of complex patterns [57]. Equation (10) and (11) shows that GELU applies a smooth non-linear activation function.

$$\Phi(x) = 0.5 * \left(1 + erf\left(x/\sqrt{2}\right)\right), \tag{10}$$

$$GELU(x) = x \cdot \Phi(x). \tag{11}$$

where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution and erf is the error function. Residual connection introduced by concatenation (CAT) operation preserves crucial features leading to enhanced gradient [52]. Equation (12) displays the concatenation operation that combines the output of GELU activation with the original input.

$$CAT(X, Y) = [X; Y]. \tag{12}$$

where X is the original input to the CNeB module, Y is the output from the GELU activation and $[X; Y]$ denotes the concatenation of X and Y along the feature dimension.

Algorithm 1 Pseudo code of the Social Distancing Monitoring

FOR each frame in input frame DO
  Detect only vehicle in frame and get centroid
  FOR each pair of vehicles DO
    Calculate distance between 2 vehicles using horizontal and vertical unit length
    IF distance less than 50 cm **THEN**
      Draw bounding red boxes around vehicles
    ELSE IF distance between 50 cm AND 180 cm **THEN**
      Draw bounding yellow boxes around vehicles
    ELSE IF distance more than 180 cm **THEN**
Draw bounding green boxes around vehicles
    END IF
  END FOR
END FOR
END

Algorithm 2 Pseudo code of the Bird's Eye View

FOR each frame in the input frame DO
  Get Region of Interest (ROI)
  Get Bird's Eye View of ROI with centroid of object
  Calculate the horizontal and vertical unit length from points marked in first frame
  Detect only vehicle in frame and get centroid
  Project detected points in Bird's Eye View
  Calculate distance between 2 vehicles using horizontal and vertical unit length
  FOR each pair of vehicles DO
    IF distance less than 50 cm **THEN**
      Draw red dot in Bird's Eye View
    ELSE IF distance between 50 cm AND 180 cm **THEN**
      Draw yellow dot in Bird Eye View
    ELSE IF distance more than 180 cm **THEN**
      Draw green dot in Bird Eye View
    END IF
  END FOR
END FOR
END

F. MANUAL OBSERVATION METHOD

The manual observation method which uses Social Distancing Monitoring is an approach that employs CCTV footage to observe and calculate the distance between vehicles [58]. Algorithm 1 depicts the flowchart for Social Distancing Monitoring that recognises each vehicle in videos and shows the distance between vehicles through different coloured bounding boxes.

Bird's Eye View is a method for detecting vehicles and converting them into points. It shows the distance between the points in the specifically drawn area [58]. The various coloured points show the distance between vehicles. Algorithm 2 illustrates the flowchart of the Bird's Eye View on vehicle detection.

G. DISTANCE-BASED PROXIMAL INDICATORS

In object tracking, 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 (13).

$$\theta = arctan\sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}. \tag{13}$$

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 θ. The appropriate values for α and R in this study can efficiently determine the front vehicle's position based on the displacement direction [59].

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 [60]. 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 [61]. Equation (14) displays the calculation of the velocity, V involves dividing the distance by the time interval between the two frames.

$$V = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \times W(u)\left[(x, y) \times \frac{fps}{5} \times 3.6 \times 3\right]. \tag{14}$$

Through this study, 1 second can collect 30 frames in high-quality video and 25 frames in low-quality video. Equation (15) 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.

$$w_{ij} = \left(v_{xi} - v_{xj}\right) / \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}. \tag{15}$$

Equation (16) until (18) display X_j calculates the difference in length both horizontally and vertically between the focused vehicles, c_0 and the vehicle, c_j surrounding it [62].

$$S_x^{0j} = \left|x_j - x_0\right| - 0.5 * \left(L_j + L_0\right) \tag{16}$$

$$S_y^{0j} = \left|y_j - y_0\right| - 0.5 * \left(W_j + W_0\right), \tag{17}$$

$$X_j = \left[\frac{S_x^{0j}}{0.5} * \left(\left|v_{xj}\right| + \left|v_{x0}\right|\right), S_y^{0j}/0.5 * \left(\left|v_{yj}\right| + \left|v_{j0}\right|\right)\right]^T. \tag{18}$$

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.

Algorithm 3 Pseudo code of the Distance Neighbors

Initialization
Load each frame in input frame
Select the bounding box with the target
Select the pixels inside the box as the potential target
For each pixel on the inside do
Function IPM
Get central point for each target
For each target
If the target is moving
Calculate the distance among targets
If less than 50 cm
  Set Near miss
Else if more than or equal to 50 cm
  Set Safe
End if
End if
End for

Equation (19) shows the speed difference between two vehicles in real-life situations contributes to the distance between them, denoted as DN [63]. A more accurate evaluation of vehicle distance is created by this method of evaluation.

$$DN = f\left(X_j\right) / Sigmoid\left(w_{ij}\right). \tag{19}$$

Algorithm 3 shows the distance-based proximal indicators applied in vehicle detection and tracking.

V. RESULTS AND DISCUSSION

The primary objective of this experiment is to analyze the performance of identifying potential factors and parameters in vehicle detection. This includes evaluating manual observation methods for near-miss detection and employing distance-based proximal indicators with the proposed object detection model. Subsequently, the study aims to explore the one-way ANOVA analyses to assess their statistical significance and relationships.

A. POSSIBLE FACTORS AND PARAMETERS

In the context of vehicle size and shape, the analysis showed that cars were consistently detected with higher accuracy compared to smaller cars that were farther from the CCTV cameras [64]. The size and shape of the cars appear to play a crucial role in the detection process, as cars provide more distinctive features that make them easier to detect accurately [12]. As shown in Figure 6, the cars that appear near CCTV cameras are bigger and easier to detect by YOLOv4-tiny or vice versa.

Vehicles with bright and distinct colors, such as red and yellow, showed higher detection rates in comparison to vehicles with dull or neutral colors. The color and texture of vehicles have a significant impact on their visibility in traffic surveillance footage. Algorithms could be fine-tuned to prioritize certain colors and textures for improved detection accuracy [11]. Figure 5 illustrates that dull color cars are easy to miss detected by applying YOLOv4-tiny compared to bright color cars.

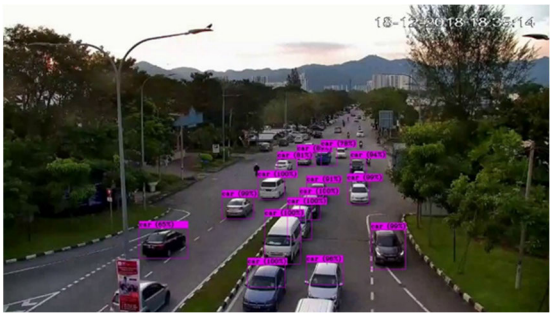


FIGURE 6. Size and shape of cars in car detection.

The detection accuracy varied depending on the direction and speed of vehicle movement. Vehicles moving away from the camera lens had a higher rate of missed detections compared to those moving towards or across the camera's field of view [78]. The speed and direction of vehicle movement present challenges for detection algorithms, particularly in high-speed scenarios or when vehicles are moving away from the camera [20].

In the context of environmental factors, vehicle detection accuracy was observed to increase significantly during clear weather conditions with good visibility and minimal precipitation, demonstrating the highest detection rates [65]. The impact of weather conditions on vehicle detection is evident from the results, as adverse weather can obscure vehicle features and affect the accuracy of detection algorithms [18]. Moreover, the presence of sunny days hindered the algorithms' ability to detect vehicles accurately. Lighting conditions significantly influence the success of vehicle detection, with algorithms performing well in daytime environments [10]. Additionally, road conditions can contribute to detection accuracy, as they remain consistent on well-maintained and smooth road surfaces [66]. However, irregular road conditions may pose challenges, requiring adjustments to the algorithms to handle such variations effectively [13]. Figure 7 displays that both videos were taken from a place in good weather, visible lighting conditions and a smooth road surface environment.

Video quality and camera resolution play a crucial role in detection performance, as higher-quality videos and higher-resolution cameras typically provide clearer, more detailed frames that enable detection models to more accurately identify objects, movements, and patterns [13]. High-resolution videos (e.g., 1080 × 720) with minimal noise and good lighting allow models to discern fine details such as edges, textures, and smaller objects, improving accuracy in tasks like vehicle detection, pedestrian tracking, and traffic sign recognition. For instance, higher-resolution cameras tend to yield higher precision in detecting smaller cars that are farther from the camera, enhancing the model's ability

High-quality video

Low-quality video

FIGURE 7. The environment of the experimental video.

to identify distant objects. In contrast, lower-quality videos, characterized by lower resolution (e.g., 640 × 480) or poor lighting, blur key features and reduce scene clarity, making it harder for models to distinguish between objects or identify specific details. This degradation in video quality results in an increase in false positives, missed detections, or incorrect classifications, ultimately affecting the overall performance of detection systems [10]. Thus, camera resolution and video quality directly impact the detection accuracy, highlighting the need for high-quality video input for optimal performance in real-world applications.

In terms of the camera's angle of view, the detection accuracy varied depending on the viewing angle [20]. Frontal or near-frontal views of vehicles exhibited the highest accuracy. The angle of view has a significant impact on vehicle detection performance, with frontal views providing the most straightforward detection scenarios [14]. Figure 6 shows the both resolutions and angle view of the CCTV camera [67].

B. COMPARISON MODELS RESULT

To provide a thorough evaluation of the different object detection models used in this study, several key metrics were analyzed. These metrics included mean Average Precision (mAP) at 0.5 and 0.95 Intersection over Union (IoU) thresholds, which assess the models' accuracy in detecting and localizing objects. Additionally, the training speeds (measured in hours per training session) and model sizes (in megabytes) were compared to determine the efficiency and practicality of each model. Table 2 compares how well seven

TABLE 2. Comparison of models.

Model	mAP@0.5	mAP@0.95	Times train/h	Model sizes/MB
YOLOv4-tiny	72.8	41.2	15.2	41.7
YOLOv5s	83.2	53.3	14.9	**34.1**
YOLOv7	96.4	**64.4**	**8.3**	74.8
YOLOv7+CNeB	**98.9**	62.9	13.2	66.6

models perform to their mAPs at 0.5 and 0.95, IoU thresholds and the training speeds and model sizes.

The mAP@0.5 shows the accuracy of the model in detecting objects with at least 50% overlap with ground truth bounding boxes [68]. The YOLOv7+CNeB model achieves the highest score of 98.9%, indicating excellent object detection accuracy. The mAP@0.95 measures accuracy at a much stricter 95% overlap threshold [69], and again, YOLOv7 tops with a score of 64.4%, though scores are generally lower than at 0.5 IoU.

The times train represents the training speed in terms of iterations or steps per hour, while the model size gives the file size or memory footprint of the trained model in megabytes [62]. Models like YOLOv7 and YOLOv7+CNeB achieve high accuracy but have larger model sizes, whereas smaller models like YOLOv4-tiny are faster to train but less accurate [70].

For instance, YOLOv4-tiny improved with better accuracy and robustness, especially in detecting smaller objects, but comes with increased computational demands. YOLOv5, while not officially part of the YOLO series, is highly optimized for performance, offering a balance between speed and accuracy but may have some limitations in handling very high-resolution images. YOLOv7 and YOLOv7+CneB enhances performance by improving accuracy and speed, particularly in detecting small objects, and incorporates advanced techniques like Transformer-based attention mechanisms, but it requires higher computational resources.

The YOLOv4-tiny, YOLOv5s, YOLOv7, and YOLOv7+CNeB models were selected for this investigation based on the following factors: First of all, these models are adaptable enough to be used in a variety of computer environments for performance evaluation, ranging from lightweight to advanced complexity [71]. Second, a variety of application situations, including real-time processing and high-precision operations, might benefit from these models' adaptability [72]. Thirdly, these models address the specific demands of real-time monitoring systems by investigating the adaptability of models with different levels of complexity [73].

The comparison of the YOLO models, detailing their specific advantages and limitations in the context of the study, is presented in Table 2, with additional information provided in the Appendix. Table 2 summarizes the Performance Average Precision (AP) metrics for YOLOv1 through YOLOv7, highlighting how advancements in architecture and optimization techniques have progressively improved detection

TABLE 3. Near miss manual observation method result.

	Type of models	YOLOv4-tiny	YOLOv5s
High-quality video	Number of frames, A	3810	3734
	Percentage, $A \div 5400 \times 100\%$	**70.56 %**	**69.15 %**
Low-quality video	Number of frames, B	2253	2123
	Percentage, $B \div 5400 \times 100\%$	**50.07 %**	**47.18 %**

accuracy and efficiency. The table and appendix offer a comprehensive overview, allowing for clear comparisons of each model's strengths.

YOLOv7 is chosen over newer versions like YOLOv8, YOLOv9, and YOLOv10 due to its optimized balance between backbone design, network architecture, and accuracy, making it particularly suitable for applications like vehicle detection and near-miss detection in traffic scenarios. The backbone of YOLOv7 is lightweight and efficient, enabling real-time processing with high precision, while its network focuses on practical implementation without overly complex enhancements that may increase computational demand [87]. In contrast, YOLOv8, YOLOv9, and YOLOv10, while offering improved accuracy through advanced features like deeper backbones and additional layers, often come with increased computational complexity, higher memory usage, and potential overfitting risks on smaller datasets [88]. These limitations make them less practical for real-world traffic systems where hardware constraints and real-time performance are critical.

C. NEAR MISS DETECTION

In this study, experiments were conducted using YOLOv4-tiny and YOLOv5s models for near-miss detection through manual observation methods. In contrast, the YOLOv7 and YOLOv7+CNeB models were employed for the automatic calculation of near-miss incidents. The purpose of this dual approach was to compare the effectiveness and accuracy of manual versus automated detection techniques. The findings from these near-miss detection experiments are detailed in the subsequent sections, providing a comprehensive analysis of each model's performance.

D. MANUAL OBSERVATION METHOD

YOLOv4-tiny and YOLOv5s utilize CCTV to detect near-miss events and produce videos showcasing these occurrences in the monitoring system without explicit near-miss calculations, categorizing this approach as manual observation for near-miss detection. Table 3 displays the near-miss detection results on the high-quality video using YOLOv4-tiny with the same duration.

The researchers concentrated on near misses since there are insufficient records regarding near-miss events and the incidents that happened in real life [74]. Therefore, the videos were taken from the black spot location to identify the near-miss events using the approach of Bird's Eye View and Social Distancing Monitoring [75].

The near-miss detection in the videos is achieved by employing YOLOv4-tiny and YOLOv5s. Notably, YOLOv4-tiny exhibits the highest percentage of near misses compared to YOLOv5s. Conversely, YOLOv5s demonstrates a higher percentage of near misses than YOLOv4-tiny in both high- and low-quality videos.

Figures 8 show high and low-quality video results by applying Social Distancing Monitoring and a Bird's Eye View.

Social Distancing Monitoring is used to monitor the distance between vehicles and predict the chance of accidents occurring in the video [76]. Bird's Eye View is applied to show the distances between points (vehicles) in a particular frame box and estimate Near Miss occur. The two methods above are implemented in this study to show how Near Miss happen and to estimate the distance for Near Miss between vehicles in video and real life. The Social Distancing Monitoring and Bird's Eye View are conducted by using YOLOv3 to detect the vehicles [77].

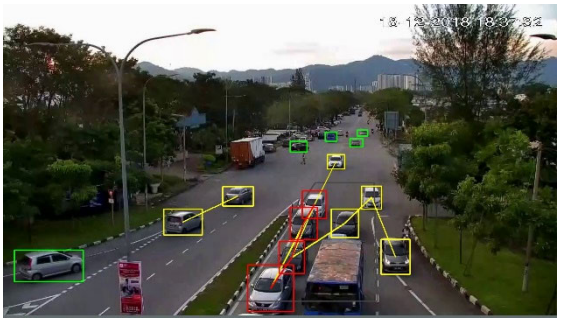
Overall, Following the comparison of accuracy and speed among the three models for object detection of manual observation, the best model, YOLOv4-tiny, was selected due to its high speed and accuracy in near-miss detection. Followed by the YOLOV5 which is ranked the second choice for near-miss detection

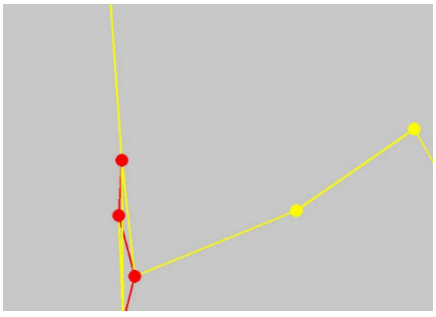
E. DISTANCE-BASED PROXIMAL INDICATORS

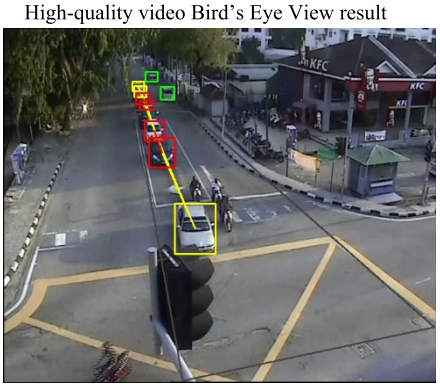
Social Distancing Monitoring and Bird's Eye View techniques represent manual observation-based methods for near-miss detection, relying on lightweight YOLO models such as YOLOv4-tiny and YOLOv5s. These models estimate potential collision risks based on predefined distance thresholds; however, their detection process requires manual observation through a monitoring system. While YOLOv4-tiny and YOLOv5s offer efficiency and low computational requirements, they are inherently limited in capturing dynamic interactions in real-world traffic conditions, making them less suitable for fully automated near-miss detection. Now, the experiment introduces advanced near-miss detection techniques using automatic calculation methods by employing enhanced models or fine-tuned YOLOv7 models.

This study explores the investigation of the near-miss incident using DN indicators. However, all the data analysis relies on the data collected during object detection. After that data has been acquired, object tracking as well as locating the target centre of gravity needs to be performed [78]. Object tracking will lead to determining a similar vehicle in various frames to find out the speed of that vehicle [63].

The more advanced YOLOv7 and YOLOv7+CNeB models utilize DN proximal indicators to systematically track and detect vehicles in real-life scenarios. This method goes beyond static distance measurements by incorporating motion dynamics and spatial-temporal relationships,


High-quality video Social Distancing Monitoring result


High-quality video Bird's Eye View result


Low-quality video Social Distancing Monitoring result

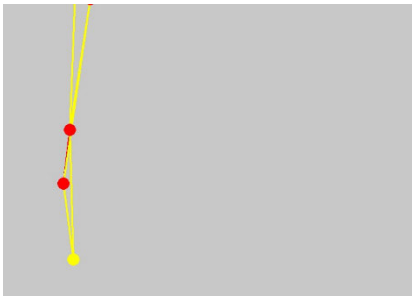

Low-quality video Bird's Eye View result

FIGURE 8. Social distancing monitoring and bird's eye view.

enabling a more accurate and automated near-miss detection system.

The effectiveness of the DN indicator algorithm in detecting near-miss events was evaluated based on its ability to track vehicle movements, label trajectories, and measure inter-vehicle distances. The algorithm continuously monitored vehicle motion and computed the DN value at each time step. A near-miss event was detected when the DN value fell below the predefined threshold of 50 cm, indicating a high probability of collision risk. The probability of near-miss incidents is raised by the lower DN values, which indicate greater proximity to risks [60]. Besides that, 50 cm is the threshold value for DN. This process effectively accounted for variations in vehicle speed and direction, demonstrating the robustness of the DN algorithm in identifying critical near-miss scenarios in dynamic traffic environments. Table 4 displays the near-miss events between cars in different quality videos using YOLOv7 and YOLOv7+CNeB.

TABLE 4. Distance-based proximal indicators near miss detection results.

Type of model	Quality of videos	Results
YOLOv7	High-quality video	**24.95%**
	Low-quality video	**8.18%**
YOLOv7+CNeB	High-quality video	**21.74%**
	Low-quality video	**2.64%**

For high-quality videos, YOLOv7 (24.95%) shows a higher probability of near misses compared to YOLOv7+CNeB (21.74%). Similarly, in low-quality videos, YOLOv7 (8.18%) also exhibits a higher probability of near misses compared to YOLOv7+CNeB (2.64%).

The proportion of high-quality video among YOLOv7 and YOLOv7+CNeB is nearly identical. This consistency is attributed to the ability of high-quality video to facilitate easier detection of the necessary parameters by the models. Conversely, the results for low-quality videos among YOLOv7 and YOLOv7+CNeB differ significantly. This variance arises from the MBPP's objective to minimize costs by employing advanced models to identify near misses in low-quality video rather than enhancing the video quality.

The proposed near-miss detection framework leverages YOLOv7+CNeB module to enhance detection accuracy and robustness in real-life traffic scenarios. YOLOv7+CNeB improves upon traditional object detection by incorporating contextual spatial relationships between detected objects, allowing for more precise tracking of vehicle interactions. This approach enables the identification of near-miss events by analyzing dynamic vehicle movements and proximity patterns, making it well-suited for complex urban environments where rapid and accurate detection is critical.

This study evaluates the applicability of the proposed near-miss detection system in urban environments with varying video quality and complex traffic conditions. By analyzing different surveillance footage, the ability of models to detect near-miss incidents under diverse conditions, such as low resolution is assessed. The findings highlight the model's potential to assist traffic management authorities in predicting vehicle movement trends and implementing proactive safety measures. Ultimately, the proposed approach enhances urban traffic safety by identifying high-risk areas and enabling timely interventions to prevent accidents.

This study evaluates near-miss detection using YOLO models, showing that YOLOv4-tiny identifies the highest

percentage of near misses, while YOLOv5s performs better in both
high- and low-quality videos. Additionally, YOLOv7 outperforms YOLOv7+CNeB in detecting near misses, with significant differences in low-quality footage due to cost-saving strategies by MBPP. Compared to prior studies, Study [62] emphasizes distance-based metrics for near-miss detection, aligning with this distance-based proximal indicators.

F. ONE-WAY ANOVA RESULTS

YOLO is employed to detect and classify near-miss incidents by analyzing video footage from CCTV cameras. While YOLO effectively provides object detection outputs, statistical validation is required to determine the significance of key influencing factors, such as the distance between vehicles and vehicle speed, on near-miss occurrences. To achieve this, ANOVA is applied to evaluate whether the variations in detected near-miss incidents are statistically significant or merely due to random chance. By integrating YOLO's detection capabilities with ANOVA's statistical analysis, this study ensures a more comprehensive assessment of the factors contributing to near-miss incidents. Therefore, establishing the relationship between YOLO and ANOVA is crucial for enhancing the reliability and accuracy of the analysis.

An increased risk of driving is related to low DN where 50 cm is the threshold value for DN [79]. The threshold value is fixed to calculate the number of near-miss events from the results of the labelled analysis of the dataset. To calculate the statistics, one-way ANOVA is applied to compare the values between the datasets.

ANOVA is a technique that can be used to compare the means of two or more samples (using the F distribution). The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of two or more independent (unrelated) groups [80]. One-way ANOVA analysis tests the equality of means and variances of the independent variable with dependent variables that include more than two values [81]. This study will use the following hypothesis.

I. Hypothesis 1: DN indicator and models with various quality video

To determine the effect of models on the DN indicators, it was chosen three different models with different quality videos [82]. According to the hypothesis 1, H_0 = There is no significant difference in DN indicator and models with various quality videos and H_1 = There is a significant difference in DN indicators and models with various quality videos. Table 5 shows the ANOVA of the DN indicator among the three models with various quality videos.

For YOLOv7+CNeB with low-quality video, the comparison to the DN indicator of YOLOv7 with high-quality video shows p-values of 0.024, which are less than $\alpha = 0.05$ [83]. Therefore, the null hypothesis is rejected, indicating a statistically significant difference in performance.

TABLE 5. ANOVA of DN indicator.

Compared with YOLOv7 HQV	Significant difference
YOLOv7 LQV	0.143
YOLOv7+CNeB HQV	0.231
YOLOv7+CNeB LQV	0.024

The p-value of 0.024 indicates a statistically significant difference in the DN indicator between YOLOv7+CNeB using low-quality video and YOLOv7 using high-quality video. The significant result suggests that the YOLOv7+CNeB model, even when operating on low-quality video, performs differently from the YOLOv7 model with high-quality video. This could be due to the advanced techniques or enhancements in the YOLOv7+CNeB model, which are robust enough to extract meaningful information from lower-quality inputs, leading to a different DN indicator compared to the baseline YOLOv7 model with high-quality video.

In contrast, comparisons involving YOLOv7 with low-quality video, YOLOv7+CNeB with high-quality video to the DN indicator of YOLOv7 with high-quality video present p-values of 0.143 and 0.231, respectively, which are greater than $\alpha = 0.05$. Therefore, the null hypothesis is accepted for these comparisons, indicating no statistically significant difference in performance. This implies that, while these models perform adequately, they do not exhibit a significant improvement over YOLOv7 with high-quality video, underscoring the need to select the most optimal configurations based on specific use-case requirements [83].

The one-way ANOVA analysis was conducted to assess the statistical significance of performance differences among detection models when applied to high- and low-quality videos. The results revealed a significant variance in detection accuracy, indicating that certain models performed better in specific video quality conditions. Post-hoc analysis further identified which models exhibited robustness in high-quality videos and which were more effective in low-resolution scenarios. These findings influenced the selection of detection models by ensuring that the most suitable algorithm was applied based on video quality, thereby optimizing detection accuracy and reliability across varying visual conditions.

VI. CONCLUSION

In this paper, we propose four models (YOLOv4-tiny, YOLOv5s, YOLOv7 and YOLOv7+CNeB) that were applied to conduct vehicle detection on videos of varying quality (low and high). The experiment also addressed error detection and missed detection during the vehicle detection process. The comparison of computational time, error detection, and missed detection revealed that YOLOv7+CNeB outperformed the other models. Following that YOLOv7, YOLOv5s and YOLOv4-tiny. Since the YOLOv7, and YOLOv7+CNeB are more advanced YOLO than YOLOv4-tiny and YOLOv5s, the YOLOv7 series show higher accuracy and precision than others.

Furthermore, near-miss detection was performed using two approaches: manual observation with YOLOv4-tiny, and YOLOv5s, and automatic calculation using DN indicators with YOLOv7, and YOLOv7+CNeB. Lastly, statistical analyses, including descriptive statistics and one-way ANOVA, were applied to compare the datasets obtained from the automatically calculated indicators.

VII. FUTURE RESEARCH

Future enhancements to the DN indicator model will focus on integrating environmental factors such as weather conditions and road surface quality to improve near-miss prediction accuracy. While the current study was conducted under sunny weather and flat road conditions, future research should explore performance under more challenging scenarios, including rain, fog, sloped roads, and complex junctions. The adverse weather can affect vehicle dynamics and visibility, while variations in road surface quality impact stability and braking efficiency, both of which influence near-miss events. By incorporating these factors, the DN model can evolve into a more robust and context-aware detection system, ensuring its effectiveness across diverse real-world traffic conditions.

By exploring the latest YOLO series for comparison with YOLOv7 could also enhance the accuracy and precision of the system, as these models feature advanced training algorithms. Besides that, the future research should explore the integration of ensemble learning by combining YOLOv7+CNeB with other detection models or using the latest version of YOLO to achieve a balance between computational efficiency and detection robustness. Ensemble approaches can leverage the strengths of multiple models, enhancing feature extraction capabilities and improving detection accuracy, particularly in complex or occluded scenarios. The system can possibly adapt to different environmental conditions while minimizing performance trade-offs by fusing models with varying architectures and optimization strategies.

Additionally, the integration of near-miss event identification and calculation into the monitoring system would provide actual values for comparison with algorithm-generated indicators, enabling detailed error analysis and predictive modeling. However, implementing such systems in real-world settings presents significant challenges, including high hardware requirements for real-time processing, computational costs, and the need for compatibility with existing traffic management systems. To satisfy the needs of real-world applications, thorough testing and refinement are also required to ensure accuracy across a variety of environmental conditions.

Moreover, future research could also explore advanced architectures and applications in diverse fields which drawing inspiration from existing studies to address the recent development of multi-modal learning to enhance to road safety. The previous research [89] presented the integrating multi-modal data, such as weather, market trends, and consumer behaviour which improved accuracy and efficiency in electricity price forecasting. In the study of [90] leveraged hyper-relational interactions with map elements, dynamics, and sensor data in Intelligent Connected Vehicles (ICVs) enhances trajectory prediction for safer autonomous driving.

TABLE 6. Evolution of YOLOV1 to YOLOV7.

YOLO	Backbone	Network	Performance Average Precision (AP)
YOLOv1 [22]	Darknet24	Used a simple structure with 24 convolutional layers and 2 fully connected layers to produce a fixed-size feature map.	Achieved an AP of 63.4% on the PASCAL VOC2007 dataset.
YOLOv2 [93]	Darknet24	Improved detection by adding Batch Normalization, anchor boxes, a high-resolution classifier, and multi-scale input during training	Achieved an average AP of 78.6% on the PASCAL VOC2007 dataset.
YOLOv3 [94]	Darknet53	The model utilizes Feature Pyramid Network (FPN) for multi-scale object detection	Achieved an AP of 36.2% on the COCO dataset.
YOLOv4 [95]	CSPDarknet-53	The model employs Mish activation and DropBlock regularization for improved performance.	Achieved an AP of 43.5% on the COCO dataset.
YOLOV5 [96]	yolov5CSPDarknet	The model utilizes the C3 module for residual learning and SPPF structure for feature integration and parameter aggregation.	Achieved an AP of 55.8% on the COCO dataset.
YOLOv6 [97]	EfficientRep	The model features a feature pyramid in the neck for fusion and a decoupled head for classification and regression.	Achieved an AP of 52.5% on the COCO dataset.
YOLOv7 [45]	yolov7Backbone	The model introduces the E-ELAN with SPPCSPC for efficient feature extraction, reduced computation, and faster processing.	Achieved an AP of 56.8% on the COCO dataset.

Similarly, [91] showed hybrid CNN-GNN frameworks in mobile traffic prediction have shown promise, and incorporating real-time user behaviour and network conditions could further refine forecasts. These directions highlight the potential of multi-modal learning to address complex challenges by effectively integrating diverse data modalities.

In future work, integrating the proposed method into the latest YOLO version, such as YOLO11, could further enhance detection speed and accuracy, leveraging its state-of-the-art performance in real-time object detection. Additionally, domain-specific large models (VSLM) have demonstrated significant success, particularly in scenarios with limited labeled data, offering a promising direction for improving model generalization in specialized tasks. Furthermore, addressing potential human labeling errors through approaches like meta-learning-guided label noise correction could enhance the reliability of training data. A thorough review of these advancements in the literature would provide deeper insights into their applicability, and future research should explore their integration to further optimize detection performance and robustness.

Near-miss detection using machine learning depends on efficient data processing, transmission, and model adaptation, closely aligning with advancements in wireless communication and federated learning. Key developments in non-orthogonal multiple access (NOMA) schemes, low-resolution analog-to-digital converters (ADCs), and IoT applications optimize signal processing and transmission, addressing challenges in dynamic environments. Similarly, advanced quantization techniques and bandwidth-efficient federated learning (FL) enhance data efficiency and accuracy under resource constraints. Secure FL, adaptive MEC resource allocation, and optimization techniques improve real-time decision-making and secure data transmission, while channel state information (CSI) feedback compression and adaptive neural networks enable better model adaptation and information recovery. These advancements collectively enhance the reliability and efficiency of near-miss detection in real-world applications.

APPENDIX

See Table 6.

ACKNOWLEDGMENT

The authors would like to express their gratitude to Springtech Ventures Sdn Bhd for their valuable support and collaboration throughout this research. The resources, technical guidance, and practical insights provided by Ts. Lee Jian Aun were instrumental in the successful completion of this work.

REFERENCES

[1] L. L. Ming, M. K. M. Ali, M. T. Ismail, and A. S. A. Mohamed, "Data safety prediction using bird's eye view and social distancing monitoring for Penang roads," *Pertanika J. Sci. Technol.*, vol. 30, no. 4, pp. 2563–2587, Sep. 2022, doi: 10.47836/pjst.30.4.15.

[2] G. Chen, J. Cai, N. G. M. Rytter, and M. Lützen, "A practical data quality assessment method for raw data in vessel operations," *J. Mar. Sci. Appl.*, vol. 22, no. 2, pp. 370–380, Jun. 2023, doi: 10.1007/s11804-023-00326-w.

[3] X. Yin, X. Wei, and Q. Tang, "Automatic generation of pantograph image caption based on deep learning," in *Proc. Int. Conf. Elect. Inf. Technol. Rail Transp.* Singapore: Springer, 2023, pp. 163–172, doi: 10.1007/978-981-99-9315-4_18.

[4] Z. S. Chen and J. S. L. Lam, "Life cycle assessment of diesel and hydrogen power systems in tugboats," *Transp. Res. D, Transp. Environ.*, vol. 103, Feb. 2022, Art. no. 103192, doi: 10.1016/j.trd.2022.103192.

[5] I. Karamanlis, A. Nikiforiadis, G. Botzoris, A. Kokkalis, and S. Basbas, "Towards sustainable transportation: The role of black spot analysis in improving road safety," *Sustainability*, vol. 15, no. 19, p. 14478, Oct. 2023.

[6] B. Hajare, S. Bhute, D. Sakle, K. Ruhatiya, D. Sable, and K. Kahar, "Vehicle safety system for blind spot and hilly areas," *SSGM J. Sci. Eng.*, vol. 1, no. 1, pp. 40–43, 2023. [Online]. Available: https://ssgmjournal.in/index.php/ssgm/article/view/62

[7] C. Bernhard, R. Reinhard, M. Kleer, and H. Hecht, "A case for raising the camera: A driving simulator test of camera-monitor systems," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 65, no. 2, pp. 321–336, Mar. 2023, doi: 10.1177/00187208211010941.

[8] A. Islam, S. Mallik, A. Roy, M. Agrebi, and P. K. Singh, "A filter-based feature selection methodology for vehicle/non-vehicle classification," in *Measurements and Instrumentation for Machine Vision*. Boca Raton, FL, USA: CRC Press, 2024, pp. 137–156.

[9] R. Kannamma, M. M. Y. Devi, S. Madhusudhanan, and R. Sethuraman, "Feature refinement with DBO: Optimizing RFRC method for autonomous vehicle detection," *Intell. Service Robot.*, vol. 17, no. 3, pp. 489–503, May 2024, doi: 10.1007/s11370-024-00520-x.

[10] J. Fernández, J. M. Cañas, V. Fernández, and S. Paniego, "Robust real-time traffic surveillance with deep learning," *Comput. Intell. Neurosci.*, vol. 2021, no. 1, Jan. 2021, Art. no. 4632353, doi: 10.1155/2021/4632353.

[11] P. Jagannathan, S. Rajkumar, J. Frnda, P. B. Divakarachari, and P. Subramani, "Moving vehicle detection and classification using Gaussian mixture model and ensemble deep learning technique," *Wireless Commun. Mobile Comput.*, vol. 2021, no. 1, Jan. 2021, Art. no. 5590894, doi: 10.1155/2021/5590894.

[12] A. Taheri Tajar, A. Ramazani, and M. Mansoorizadeh, "A lightweight tiny-YOLOv3 vehicle detection approach," *J. Real-Time Image Process.*, vol. 18, no. 6, pp. 2389–2401, Dec. 2021, doi: 10.1007/s11554-021-01131-w.

[13] X. Zhang, B. Story, and D. Rajan, "Night time vehicle detection and tracking by fusing vehicle parts from multiple cameras," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 8136–8156, Jul. 2022, doi: 10.1109/TITS.2021.3076406.

[14] M. Humayun, F. Ashfaq, N. Z. Jhanjhi, and M. K. Alsadun, "Traffic management: Multi-scale vehicle detection in varying weather conditions using YOLOv4 and spatial pyramid pooling network," *Electronics*, vol. 11, no. 17, p. 2748, Sep. 2022, doi: 10.3390/electronics11172748.

[15] Z. Xu, N. Zheng, D. B. Logan, and H. L. Vu, "Assessing bicycle-vehicle conflicts at urban intersections utilizing a VR integrated simulation approach," *Accident Anal. Prevention*, vol. 191, Oct. 2023, Art. no. 107194, doi: 10.1016/j.aap.2023.107194.

[16] Y. Matsui and S. Oikawa, "Characteristics of dangerous scenarios between vehicles turning right and pedestrians under left-hand traffic," *Appl. Sci.*, vol. 13, no. 7, p. 4189, Mar. 2023, doi: 10.3390/app13074189.

[17] A. Kondyli, S. D. Schrock, and F. Tousif, "Evaluation of near-miss crashes using a video-based tool," Kansas Dept. Transp., Bur. Res., Topeka, KS, USA, Tech. Rep. K-TRAN:KU-21-4, 2023.

[18] F. Wang, Y. Li, Y. Wei, and H. Dong, "Improved faster RCNN for traffic sign detection," in *Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2020, pp. 1–6.

[19] Z. Wu, M. Bai, J. Zheng, Y. Zhan, X. H. Xia, J. Lei, and Z. Chen, "An applied research of improving Kalman filter in dual mode navigation," in *Proc. 29th Chin. Control Decis. Conf. (CCDC)*, May 2017, pp. 3977–3982, doi: 10.1109/CCDC.2017.7979196.

[20] N. Abid, T. Ouni, and M. Abid, "Vehicle detection for intelligent traffic surveillance system," in *Proc. 5th Int. Conf. Adv. Technol. Signal Image Process. (ATSIP)*, Sep. 2020, pp. 1–5, doi: 10.1109/ATSIP49331.2020.9231936.

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[22] Y. Ren, C. Zhu, and S. Xiao, "Object detection based on fast/faster RCNN employing fully convolutional architectures," *Math. Problems Eng.*, vol. 2018, pp. 1–7, Jan. 2018, doi: 10.1155/2018/3598316.

[23] H. Makizako, H. Shimada, R. Hotta, T. Doi, K. Tsutsumimoto, S. Nakakubo, and K. Makino, "Associations of near-miss traffic incidents with attention and executive function among older Japanese drivers," *Gerontology*, vol. 64, no. 5, pp. 495–502, 2018, doi: 10.1159/000486547.

[24] H. Kataoka, T. Suzuki, S. Oikawa, Y. Matsui, and Y. Satoh, "Drive video analysis for the detection of traffic near-miss incidents," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 3421–3428, doi: 10.1109/ICRA.2018.8460812.

[25] N. N. R. N. Mahdi, N. Mohamed, and M. N. Shafei, "Risk factors for near miss incident among long distance bus drivers in Malaysia," *Iranian J. Public Health*, vol. 43, no. 3, pp. 117–124, 2014.

[26] A.-S. Karakaya, J. Hasenburg, and D. Bermbach, "SimRa: Using crowdsourcing to identify near miss hotspots in bicycle traffic," *Pervas. Mobile Comput.*, vol. 67, Sep. 2020, Art. no. 101197, doi: 10.1016/j.pmcj.2020.101197.

[27] M. Leni Siregar, H. R. Agah, and F. Hidayatullah, "Near-miss accident analysis for traffic safety improvement at a 'channelized' junction with U-turn," *Int. J. Saf. Secur. Eng.*, vol. 8, no. 1, pp. 31–38, Jan. 2018, doi: 10.2495/safe-v8-n1-31-38.

[28] G. Yasmine, G. Maha, and M. Hicham, "Overview of single-stage object detection models: From YOLOv1 to YOLOv7," in *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, Jun. 2023, pp. 1579–1584, doi: 10.1109/iwcmc58020.2023.10182423.

[29] J. Terven and D. Cordova-Esparza, "A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS," 2023, *arXiv:2304.00501*.

[30] Z. Jiang, L. Zhao, S. Li, and Y. Jia, "Real-time object detection method based on improved YOLOv4-tiny," *J. Netw. Intell.*, vol. 7, no. 1, pp. 1–11, Jan. 2020.

[31] K. Ntzelepi, M. E. Filippakis, M. E. Poulou, and A. Angelakis, "Performance evaluation of YOLOv4 and YOLOv4-tiny for real-time face-mask detection on mobile devices," *Int. J. Artif. Intell. Appl.*, vol. 13, no. 3, pp. 31–47, May 2022, doi: 10.5121/ijaia.2022.13303.

[32] X. Dong, S. Yan, and C. Duan, "A lightweight vehicles detection network model based on YOLOv5," *Eng. Appl. Artif. Intell.*, vol. 113, Aug. 2022, Art. no. 104914, doi: 10.1016/j.engappai.2022.104914.

[33] M. Kasper-Eulaers, N. Hahn, S. Berger, T. Sebulonsen, Ø. Myrland, and P. E. Kummervold, "Short communication: Detecting heavy goods vehicles in rest areas in winter conditions using YOLOv5," *Algorithms*, vol. 14, no. 4, p. 114, Mar. 2021, doi: 10.3390/a14040114.

[34] Y. Wang, H. Ma, and L. Li, "Road traffic vehicle detection method using lightweight YOLOv5 and attention mechanism," in *Proc. 7th Int. Conf. Image, Vis. Comput. (ICIVC)*, Jul. 2022, pp. 201–207, doi: 10.1109/ICIVC55077.2022.9886525.

[35] K. Zhang, C. Wang, X. Yu, A. Zheng, M. Gao, Z. Pan, G. Chen, and Z. Shen, "Research on mine vehicle tracking and detection technology based on YOLOv5," *Syst. Sci. Control Eng.*, vol. 10, no. 1, pp. 347–366, Apr. 2022, doi: 10.1080/21642583.2022.2057370.

[36] P. Zhu, B. Chen, B. Liu, Z. Qi, S. Wang, and L. Wang, "Object detection for hazardous material vehicles based on improved YOLOv5 algorithm," *Electronics*, vol. 12, no. 5, p. 1257, Mar. 2023, doi: 10.3390/electronics12051257.

[37] Q. Zhao, W. Ma, C. Zheng, and L. Li, "Exploration of vehicle target detection method based on lightweight YOLOv5 fusion background modeling," *Appl. Sci.*, vol. 13, no. 7, p. 4088, Mar. 2023, doi: 10.3390/app13074088.

[38] J. Wang, Y. Dong, S. Zhao, and Z. Zhang, "A high-precision vehicle detection and tracking method based on the attention mechanism," *Sensors*, vol. 23, no. 2, p. 724, Jan. 2023, doi: 10.3390/s23020724.

[39] Z. Wu, H. Xie, X. Hu, J. He, and G. Wang, "Lightweight vehicle detection and recognition method based on improved YOLOv5 in SAR images," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput. (ICSPCC)*, Oct. 2022, pp. 1–6, doi: 10.1109/ICSPCC55723.2022.9984375.

[40] Y. Wang, G. Yang, and J. Guo, "Vehicle detection in surveillance videos based on YOLOv5 lightweight network," *Bull. Polish Acad. Sci. Tech. Sci.*, vol. 70, no. 6, Nov. 2022, Art. no. e143644, doi: 10.24425/bpasts.2022.143644.

[41] Y. Fan, Q. Qiu, S. Hou, Y. Li, J. Xie, M. Qin, and F. Chu, "Application of improved YOLOv5 in aerial photographing infrared vehicle detection," *Electronics*, vol. 11, no. 15, p. 2344, Jul. 2022, doi: 10.3390/electronics11152344.

[42] M. H. Hamzenejadi and H. Mohseni, "Fine-tuned YOLOv5 for real-time vehicle detection in UAV imagery: Architectural improvements and performance boost," *Expert Syst. Appl.*, vol. 231, Nov. 2023, Art. no. 120845, doi: 10.1016/j.eswa.2023.120845.

[43] L. Shao, H. Wu, C. Li, and J. Li, "A vehicle recognition model based on improved YOLOv5," *Electronics*, vol. 12, no. 6, p. 1323, Mar. 2023, doi: 10.3390/electronics12061323.

[44] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.

[45] D. Hu, S. Li, and M. Wang, "Object detection in hospital facilities: A comprehensive dataset and performance evaluation," *Eng. Appl. Artif. Intell.*, vol. 123, Aug. 2023, Art. no. 106223, doi: 10.1016/j.engappai.2023.106223.

[46] O. E. Olorunshola, M. E. Irhebhude, and A. E. Evwiekpaefe, "A comparative study of YOLOv5 and YOLOv7 object detection algorithms," *J. Comput. Social Informat.*, vol. 2, no. 1, pp. 1–12, Feb. 2023, doi: 10.33736/jcsi.5070.2023.

[47] N. D. T. Yung, W. K. Wong, F. H. Juwono, and Z. A. Sim, "Safety helmet detection using deep learning: Implementation and comparative study using YOLOv5, YOLOv6, and YOLOv7," in *Proc. Int. Conf. Green Energy, Comput. Sustain. Technol. (GECOST)*, Oct. 2022, pp. 164–170, doi: 10.1109/GECOST55694.2022.10010490.

[48] L.-D. Quach, K. N. Quoc, A. N. Quynh, and H. T. Ngoc, "Evaluating the effectiveness of YOLO models in different sized object detection and feature-based classification of small objects," *J. Adv. Inf. Technol.*, vol. 14, no. 5, pp. 907–917, 2023, doi: 10.12720/jait.14.5.907-917.

[49] K. Jiang, T. Xie, R. Yan, X. Wen, D. Li, H. Jiang, N. Jiang, L. Feng, X. Duan, and J. Wang, "An attention mechanism-improved YOLOv7 object detection algorithm for hemp duck count estimation," *Agriculture*, vol. 12, no. 10, p. 1659, Oct. 2022, doi: 10.3390/agriculture12101659.

[50] A. Vaswani, P. Ramachandran, A. Srinivas, N. Parmar, B. Hechtman, and J. Shlens, "Scaling local self-attention for parameter efficient visual backbones," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 12889–12899.

[51] L. A. Silva, H. S. S. Blas, D. P. García, A. S. Mendes, and G. V. González, "An architectural multi-agent system for a pavement monitoring system with pothole recognition in UAV images," *Sensors*, vol. 20, no. 21, p. 6205, Oct. 2020, doi: 10.3390/s20216205.

[52] Z. Chen, X. Chen, and K. Ren, "An improved network for pedestrian-vehicle detection based on YOLOv7," *Proc. Int. Conf. Artif. Life Robot.*, vol. 28, pp. 803–807, Feb. 2023, doi: 10.5954/ICAROB.2023.OS31-3. [Online]. Available: https://alife-robotics.co.jp/members2023/icarob/data/html/data/OS/OS31/OS31-3.pdf

[53] G. Yan, J. Guo, D. Zhu, S. Zhang, R. Xing, Z. Xiao, and Q. Wang, "A flame detection algorithm based on improved YOLOv7," *Appl. Sci.*, vol. 13, no. 16, p. 9236, Aug. 2023, doi: 10.3390/app13169236.

[54] C. Li, J. Lin, Z. Li, C. Mai, R. Jiang, and J. Li, "An efficient detection method for litchi fruits in a natural environment based on improved YOLOv7-litchi," *Comput. Electron. Agricult.*, vol. 217, Feb. 2024, Art. no. 108605, doi: 10.1016/j.compag.2023.108605.

[55] R. Chang, S. Zhou, Y. Zhang, N. Zhang, C. Zhou, and M. Li, "Research on insulator defect detection based on improved YOLOv7 and multi-UAV cooperative system," *Coatings*, vol. 13, no. 5, p. 880, May 2023, doi: 10.3390/coatings13050880.

[56] Y. Hui, J. Wang, and B. Li, "STF-YOLO: A small target detection algorithm for UAV remote sensing images based on improved SwinTransformer and class weighted classification decoupling head," *Measurement*, vol. 224, Jan. 2024, Art. no. 113936, doi: 10.1016/j.measurement.2023.113936.

[57] C. Lang, X. Yu, and X. Rong, "LSDNet: A lightweight ship detection network with improved YOLOv7," *J. Real-Time Image Process.*, vol. 21, no. 2, pp. 1–14, Apr. 2024, doi: 10.1007/s11554-024-01441-9.

[58] L. M. Lim, S. Sathasivam, M. T. Ismail, A. S. A. Mohamed, O. J. Ibidoja, and M. K. M. Ali, "Comparison using intelligent systems for data prediction and near miss detection techniques," *Pertanika J. Sci. Technol.*, vol. 32, no. 1, pp. 365–394, Dec. 2023, doi: 10.47836/pjst.32.1.20.

[59] M. Ljung, Y.-H. Huang, N. Aberg, and E. Johansson, "Close calls on the road? A study of drivers' near-misses," in *Proc. 3rd Int. Conf. Traffic Transp. Psychol.*, 2004, pp. 1–14. [Online]. Available: http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-

[60] R. Suwanda, Z. Syahputra, and E. M. Zamzami, "Analysis of Euclidean distance and Manhattan distance in the K-means algorithm for variations number of centroid K," *J. Phys., Conf. Ser.*, vol. 1566, no. 1, Jun. 2020, Art. no. 012058. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1566/1/012058/meta

[61] C. Ricotta and S. Pavoine, "A new parametric measure of functional dissimilarity: Bridging the gap between the bray-curtis dissimilarity and the Euclidean distance," *Ecol. Model.*, vol. 466, Apr. 2022, Art. no. 109880, doi: 10.1016/j.ecolmodel.2022.109880.

[62] L. Yang, A. S. A. Mohamed, and M. K. M. Ali, "Traffic conflicts analysis in Penang based on improved object detection with transformer model," *IEEE Access*, vol. 11, pp. 84061–84073, 2023, doi: 10.1109/ACCESS.2023.3299316.

[63] S. M. S. Mahmud, L. Ferreira, M. S. Hoque, and A. Tavassoli, "Application of proximal surrogate indicators for safety evaluation: A review of recent developments and research needs," *IATSS Res.*, vol. 41, no. 4, pp. 153–163, Dec. 2017, doi: 10.1016/j.iatssr.2017.02.001.

[64] M. A. Bin Zuraimi and F. H. Kamaru Zaman, "Vehicle detection and tracking using YOLO and DeepSORT," in *Proc. IEEE 11th IEEE Symp. Comput. Appl. Ind. Electron. (ISCAIE)*, Apr. 2021, pp. 23–29, doi: 10.1109/ISCAIE51753.2021.9431784.

[65] M. Hassaballah, M. A. Kenk, K. Muhammad, and S. Minaee, "Vehicle detection and tracking in adverse weather using a deep learning framework," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4230–4242, Jul. 2021, doi: 10.1109/TITS.2020.3014013.

[66] 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. C, Emerg. Technol.*, vol. 124, Mar. 2021, Art. no. 102946, doi: 10.1016/j.trc.2020.102946.

[67] M. M. Rahman, M. T. Ismail, N. Awang, and M. K. M. Ali, "A new parametric function-based dynamic lane changing trajectory planning and simulation model," *Pertanika J. Sci. Technol.*, vol. 29, no. 1, pp. 8–12, 2021, doi: 10.47836/pjst.29.1.12.

[68] Y. Xu, L. Zhu, Y. Yang, and F. Wu, "Training robust object detectors from noisy category labels and imprecise bounding boxes," *IEEE Trans. Image Process.*, vol. 30, pp. 5782–5792, 2021, doi: 10.1109/TIP.2021.3085208.

[69] K. Zhao, Y. Liu, S. Hao, S. Lu, H. Liu, and L. Zhou, "Bounding boxes are all we need: Street view image classification via context encoding of detected buildings," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5602817, doi: 10.1109/TGRS.2021.3064316.

[70] H. Luo, L. Ji, M. Zhong, Y. Chen, W. Lei, N. Duan, and T. Li, "CLIP4Clip: An empirical study of CLIP for end to end video clip retrieval," 2021, *arXiv:2104.08860*.

[71] P. Wang, H. Liu, X. Zhou, Z. Xue, L. Ni, Q. Han, and J. Li, "Multidimensional evaluation methods for deep learning models in target detection for SAR images," *Remote Sens.*, vol. 16, no. 6, p. 1097, Mar. 2024, doi: 10.3390/rs16061097.

[72] G. S. Patel, A. A. Desai, Y. Y. Kamble, G. V. Pujari, P. A. Chougule, and V. A. Jujare, "Identification and separation of medicine through robot using YOLO and CNN algorithms for healthcare," in *Proc. Int. Conf. Artif. Intell. Innov. Healthcare Industries (ICAIIHI)*, Dec. 2023, pp. 1–5, doi: 10.1109/icaiihi57871.2023.10489407.

[73] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," *Multimedia Tools Appl.*, vol. 82, no. 6, pp. 9243–9275, Mar. 2023, doi: 10.1007/s11042-022-13644-y.

[74] Y. Matsui, M. Hitosugi, T. Doi, S. Oikawa, K. Takahashi, and K. Ando, "Features of pedestrian behavior in car-to-pedestrian contact situations in near-miss incidents in Japan," *Traffic Injury Prevention*, vol. 14, no. 1, pp. S58–S63, Jan. 2013.

[75] V. Vinitha and V. Velantina, "Social distancing detection system with artificial intelligence using computer vision and deep learning," *Int. Res. J. Eng. Technol.*, vol. 7, no. 8, pp. 4049–4053, 2020. [Online]. Available: https://www.irjet.net/archives/V7/i8/IRJET-V7I8698.pdf

[76] Y. Ong, "Near miss vehicle collisions estimation using YOLO," M.S. thesis, Dept. School Comput. Sci., Universiti Sains Malaysia, Pulau Pinang, Malaysia, 2020.

[77] Q. Chen, H. Huang, Y. Li, J. Lee, K. Long, R. Gu, and X. Zhai, "Modeling accident risks in different lane-changing behavioral patterns," *Anal. Methods Accident Res.*, vol. 30, Jun. 2021, Art. no. 100159, doi: 10.1016/j.amar.2021.100159.

[78] L. Yang, A. S. A. B. Mohamed, and M. K. B. M. Ali, "An improved object detection and trajectory prediction method for traffic conflicts analysis," *Promet-Traffic Transp.*, vol. 35, no. 4, pp. 462–484, Aug. 2023, doi: 10.7307/ptt.v35i4.173.

[79] A. R. Bhele and D. S. D. Ghodmare, "Non-motorized transportation analysis of traffic density, pollution using regression anova analysis," *Int. J. Sci. Res. Sci. Technol.*, vol. 2021, pp. 534–542, Aug. 2021, doi: 10.32628/ijsrst218474.

[80] H. Ulvi, M. Uysal, M. Öktem, and H. Onder, "The comparative analysis of urban transport in Ankara by gender and age groups," *Megaron*, vol. 14, no. 4, pp. 1–11, 2019, doi: 10.14744/MEGARON.2019.24572.

[81] V. Masarotto, V. M. Panaretos, and Y. Zemel, "Transportation-based functional ANOVA and PCA for covariance operators," *Electron. J. Statist.*, vol. 18, no. 1, pp. 1887–1916, Jan. 2024, doi: 10.1214/24-ejs2240.

[82] Z. Qiu and H. Cao, "Commuter exposure to particulate matter in urban public transportation of Xi'an, China," *J. Environ. Health Sci. Eng.*, vol. 18, no. 2, pp. 451–462, Dec. 2020, doi: 10.1007/s40201-020-00473-0.

[83] L. M. Lim, L. Yang, A. S. A. Mohamed, and M. K. M. Ali, "Data safety prediction using YOLOv7+G3HN for traffic roads," *J. Nigerian Soc. Phys. Sci.*, vol. 6, no. 3, p. 2198, Aug. 2024, doi: 10.46481/jnsps.2024.2198.

[84] B. Yang, X. Wang, Y. Xing, C. Cheng, W. Jiang, and Q. Feng, "Modality fusion vision transformer for hyperspectral and LiDAR data collaborative classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 17052–17065, 2024, doi: 10.1109/jstars.2024.3415729.

[85] Y. Zheng and W. Jiang, "Evaluation of vision transformers for traffic sign classification," *Wireless Commun. Mobile Comput.*, vol. 2022, pp. 1–14, Jun. 2022, doi: 10.1155/2022/3041117.

[86] S. Tummala, S. Kadry, S. A. C. Bukhari, and H. T. Rauf, "Classification of brain tumor from magnetic resonance imaging using vision transformers ensembling," *Current Oncol.*, vol. 29, no. 10, pp. 7498–7511, Oct. 2022, doi: 10.3390/curroncol29100590.

[87] R. Mahadshetti, J. Kim, and T.-W. Um, "Sign-YOLO: Traffic sign detection using attention-based YOLOv7," *IEEE Access*, vol. 12, pp. 132689–132700, 2024, doi: 10.1109/ACCESS.2024.3417023.

[88] M. Gluč ina, N. Žigulic, D. Frank, I. Lorencin, and D. Matika, "Comparative analysis of YOLOv9 and YOLOv10 algorithms for urban safety improvement," in *Proc. IEEE 22nd Jubilee Int. Symp. Intell. Syst. Informat. (SISY)*, Sep. 2024, pp. 185–190, doi: 10.1109/sisy62279.2024.10737612.

[89] M. He, W. Jiang, and W. Gu, "TriChronoNet: Advancing electricity price prediction with multi-module fusion," *Appl. Energy*, vol. 371, Oct. 2024, Art. no. 123626, doi: 10.1016/j.apenergy.2024.123626.

[90] Y. Lu, W. Wang, R. Bai, S. Zhou, L. Garg, A. K. Bashir, W. Jiang, and X. Hu, "Hyper-relational interaction modeling in multi-modal trajectory prediction for intelligent connected vehicles in smart cites," *Inf. Fusion*, vol. 114, Feb. 2025, Art. no. 102682, doi: 10.1016/j.inffus.2024.102682.

[91] W. Jiang, Y. Zhang, H. Han, Z. Huang, Q. Li, and J. Mu, "Mobile traffic prediction in consumer applications: A multimodal deep learning approach," *IEEE Trans. Consum. Electron.*, vol. 70, no. 1, pp. 3425–3435, Feb. 2024, doi: 10.1109/TCE.2024.3361037.

[92] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525, doi: 10.1109/CVPR.2017.690.

[93] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[94] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[95] Ultralytics. (2020). *YOLOv5*. [Online]. Available: https://github.com/ultralytics/yolov5

[96] Meituan, Inc. (2022). *YOLOv6: A Single-stage Object Detection Framework for Industrial Applications*. GitHub repository. [Online]. Available: https://github.com/meituan/YOLOv6

LEK MING LIM received the B.S. degree in modeling mathematics and the M.S. degree in operation research from Universiti Sains Malaysia, Malaysia, where she is currently pursuing the Ph.D. degree with the School of Mathematical Sciences. Her research interests include statistical and mathematical modeling in detection image processing, machine learning, and near-miss detection.

LU YANG received the B.S. degree in electronic and information engineering from Jimei University, China, and the M.S. degree in computer science from Assumption University, Thailand. She is currently pursuing the Ph.D. degree with the School of Computer Sciences, Universiti Sains Malaysia, Malaysia. Her research interests include vision and image processing, machine learning, and data mining.

WEN ZHU received the bachelor's degree from Taiyuan Normal University, China. She is currently pursuing the master's degree with the School of Mathematical Sciences, Universiti Sains Malaysia. Her research interests include data analysis, deep learning, and object detection.

AHMAD SUFRIL AZLAN MOHAMED received the BIT degree (Hons.) from Multimedia University, Malaysia, the M.Sc. degree from The University of Manchester, U.K., and the Ph.D. degree from the University of Salford, U.K. He is currently with the School of Computer Sciences, Universiti Sains Malaysia, Malaysia. His research interests include image processing, video tracking, facial recognition, and medical imaging.

MAJID KHAN MAJAHAR ALI received the B.Sc. and Ph.D. degrees in mathematics with economics from Universiti Malaysia Sabah. In 2017, he joined the School of Mathematical Sciences, Universiti Sains Malaysia (USM), Malaysia. His research interests include big data, machine learning, metaheuristic, and optimization.