Development of Traffic Light and Road Sign Detection and Recognition Using Deep Learning

Towards Safe and Robust Sensor-Perception System of Autonomous Vehicle Development Research

Joseph M. De Guia[1], Madhavi Deveraj[2]
School of Information Technology (SOIT), Mapua University, Manila, Phillippines[1, 2]
Energy Research Institute (ERI@N), Nanyang Technological University, Singapore[1]

***Abstract*—Traffic light and road sign violations significantly contribute to traffic accidents, particularly at intersections in high-density urban areas. To address these challenges, this research focuses on enhancing the accuracy, robustness, and reliability of Autonomous Vehicle (AV) perception systems using advanced deep learning techniques. The novelty of this study lies in the comprehensive development and evaluation of real-time traffic light and road sign detection systems, comparing state-of-the-art models including YOLOv3, YOLOv5, and YOLOv7. The models were rigorously tested in a controlled offline environment using the Nvidia Titan RTX, followed by extensive field testing on an AV test vehicle equipped with sensor suite and Nvidia RTX GPU. The testing was conducted across complex urban driving scenarios at the CETRAN proving test track, JTC Cleantech Park, and NTU Singapore campus. The traffic light detection and recognition (TLR) results demonstrate that YOLOv7 outperforms YOLOv5 and YOLOv3, achieving a mean Average Precision (mAP@0.5) of 93%, even under challenging conditions like poor lighting and occlusions. While the traffic road sign detection (TSD) mAP@0.5 of 96%. This superior performance highlights the potential of YOLOv7 in enhancing AV safety and reliability. The conclusions underscore the effectiveness of YOLOv7 for real-time detection in AV perception systems, offering crucial insights for future research. Potential implications include the development of more robust and accurate AV systems, capable of safely navigating complex urban environments.**

Keywords—Artificial intelligence; autonomous vehicle; traffic light recognition; road sign detection; YOLO; real-time object detection

I. Introduction

Traffic light and road sign violations are a significant concern in Singapore, contributing to the rising number of traffic accidents, particularly at intersections. In 2023, over 31,815 red-light running violations were recorded, leading to numerous collisions and injuries at intersections [1]. Reports from the Singapore Police Force and Channel NewsAsia highlight the growing trend of drivers running red lights, resulting in dangerous collisions and fatalities [2]. The WHO Global Status Report on Road Safety 2023 also emphasizes that traffic signal violations are a major cause of road traffic injuries and fatalities globally [3]. These violations are especially problematic in high-density areas like the Central Business District (CBD) and busy residential and school campus zones, where the volume of both vehicular and pedestrian traffic is high. The failure to obey traffic signals in these areas can lead to hazardous situations, such as collisions with pedestrians at crossings or crashes involving multiple vehicles at junctions.

To address these challenges, recent research in AV development has focused on the integration of advanced perception systems that can accurately detect and classify traffic lights [4] and road signs in real-time [5]. These systems, leveraging deep learning algorithms and high-precision sensors, are designed to operate effectively even under challenging conditions such as poor lighting or heavy rain. By ensuring that AVs can reliably recognize and respond to traffic lights and signs, these technologies hold the potential to significantly reduce traffic violations, prevent accidents, and enhance overall road safety. Refer to Fig. 1 for the demonstration of the detection capability of the proposed TLR and TSD system.

(a) (b)

Fig. 1. Traffic light and road sign detection results. (a) The system accurately identifies a green light (95% confidence) and a red right arrow (94% confidence) at night. (b) The system detects a zebra crossing sign with 85% confidence in daylight. These results demonstrate the system's ability to detect and classify signals and signs across varying lighting conditions, essential for AV safety.

Previous studies have explored various approaches to traffic light and road sign detection, including traditional image processing techniques, machine learning models, and more recently, deep learning methods. While traditional methods offered limited success due to their inability to generalize across different environments and lighting conditions, deep learning has emerged as a powerful tool to improve the accuracy and robustness of real-time object detection and classification [4][5].

Notable advancements include the use of convolutional neural networks (CNNs) in models such as Fast R-CNN [6], SSD [7], and YOLO (You Only Look Once) [8], which have demonstrated superior performance in detecting a wide range of objects in images.

However, despite these advancements, challenges remain in achieving high recognition accuracy and stability under diverse and dynamic environmental conditions. Environmental variability, including changing lighting and weather conditions, as well as occlusions by other objects, can hinder accurate detection [9]. Real-time processing constraints, such as the need for high computational efficiency and low latency, further complicate the deployment of these systems [10]. Additionally, generalizing detection models across diverse geographical regions, handling data annotation and training challenges, and ensuring robustness against adversarial attacks and physical manipulation are ongoing issues [11].

Considering these challenges, this research focuses on advancing real-time detection and recognition of traffic lights and road signs using state-of-the-art deep learning techniques. By exploring the performance of different YOLO variants - YOLOv3 [12], YOLOv5 [13], and YOLOv7 [14], this study aims to improve the recognition accuracy and stability of AV perception systems. The research is particularly focused on high-risk areas such as junctions and pedestrian crossings, where the need for precise and reliable detection is paramount.

The novelty of this work lies in its comprehensive approach to evaluating and improving AV perception systems under real-world conditions. By conducting extensive testing in complex urban environments, this research contributes to filling critical gaps in current AV technologies. The study's objectives include the development and deployment of an optimal model for traffic light detection and recognition (TLR) and traffic road sign detection (TSD), evaluating system performance to enhance accuracy and reliability, and reporting the results of real-world testing. Our threefold contribution to the research:

- Performance Evaluation and Enhancement: The study conducted comprehensive evaluations of the perception system, leading to improvements in detection accuracy, robustness, and reliability, particularly under challenging environmental conditions.
- Comparative Analysis of YOLO Variants: The research compared the performance of multiple YOLO variants - YOLOv3, YOLOv5, YOLOv7 in real-time detection tasks, providing insights into their strengths and weaknesses in the context of AV perception systems.
- Real-World Testing and Validation: The research involved testing, verification, validation, and self-assessment of the deployed model in real-world test environments, including diverse urban scenarios in Singapore.
    - Ultimately, this research aims to contribute to the advancement of AV technologies, ensuring their safe and effective integration into public roadways.
    - The paper is divided into six sections. Section II describes the related works to TLR and TSD recognition research. In Section III we present in details of the TLR and TSD perception development in the AV research platform. Section IV is the Methodology describing our design, development, and evaluation of TLR and TSD systems in the AV research platform sensor and perception systems using custom-trained YOLO variant models. The results and discussion of the evaluation and assessment is presented in Section V. Finally, in conclusion, lesson learnt, and future research direction to further enhance the features and capabilities of the perception system is presented in Section VI.

II. Related Works

The integration of sensors and perception algorithms is crucial in AV systems for accurately identifying and categorizing objects like traffic lights, traffic signs, and road markings on public roads. These elements are essential for road safety and traffic management. However, real-time detection of these elements is challenging due to varying environmental conditions, such as changes in lighting, occlusions, image quality issues, motion blur, and glare. These challenges highlight the need for selecting appropriate sensor modalities (e.g., cameras, LiDAR, RADAR) and optimizing perception algorithms to enhance the reliability of AV systems. Several studies have focused on overcoming these difficulties, emphasizing the importance of real-time automation in AVs, particularly at traffic junctions, pedestrian crossings, and roundabouts, where accurate detection and recognition are crucial for preventing collisions and ensuring pedestrian safety.

Early research in TLR relied on rule-based and classical machine learning detectors, which were eventually outperformed by deep learning-based methods. These learning-based detectors significantly improved precision and recall, demonstrating the potential of deep learning for enhancing detection accuracy [15]. The use of stereo cameras has also been proposed to enhance tracking capabilities, particularly for object localization and motion estimation [16]. Addressing data imbalance issues in training datasets is crucial, as seen with the LARA and LISA traffic light databases, where uneven data distribution across traffic light states requires data-centric approaches to improve model performance [17].

Deep learning methods, especially Convolutional Neural Networks (CNNs) and YOLO detectors, have shown promise in TLR. For example, YOLOv3 achieved an AUC of 90.49% and a precision of 50.32% on the LISA dataset [18]. Suggestions for improvement include testing models under nighttime conditions and using ensemble methods like SSD and R-FCN [19]. Advanced methods have been explored, such as using CNNs for traffic light color recognition and integrating Faster R-CNN with k-means clustering, achieving an average precision (AP) of 83%, which increased to 90% for objects larger than 8 pixels [20]. However, challenges remain, including high false positive rates and ensuring consistent classification [21]. The development of large-scale, high-variance datasets like the DriveU Traffic Light (DTLD) dataset, recorded across 11 cities, has been a significant contribution, providing valuable resources for training and evaluating CNN-based models [22].

Traffic road signs provide crucial information that is essential for the decision-making process and safety of autonomous vehicles (AVs). These signs, including speed limits, danger warnings, and directional guidance, are vital for ensuring safe navigation. For instance, when an AV detects a "School ahead" or "Hospital ahead" sign, it can adjust its speed and exercise increased caution. However, detecting traffic signs is challenging due to factors like lighting variations, changes in scale, weather conditions, occlusions, and rotations. Various approaches, including traditional object detectors like Support Vector Machines (SVMs) and pattern matching techniques, have been explored to address these challenges [23]. Despite their application, these methods often struggle with detecting small-scale traffic signs or performing well under difficult conditions [24].

Recent advancements in traffic sign detection have leveraged CNNs and YOLO detectors. While region-based networks and one-stage detectors face limitations in detecting small-scale signs, Region-Proposal Networks (RPNs) have shown superior performance [25]. The integration of Inception V2 for feature extraction has led to competitive results in benchmarks like the German Traffic Sign Detection Benchmark (GTSDB) [26]. A notable method involved generating traffic sign proposals using a color probability model and the Maximally Stable Extremal Region (MSER) detector, followed by an SVM classifier to filter out false positives and categorize signs [27]. Another approach introduced "Capsule Networks," which capture complex spatial relationships, enhancing detection accuracy [28]. Additionally, a model using a single CNN to estimate the location and boundary of traffic signs has improved performance in detecting small and occluded signs [29].

III. Traffic Light and Traffic Road Sign Detection Development for AV Research Platform

The TLR and TSD systems are integral components of the perception system in the AV test vehicle. TLR processes camera or sensor data to detect and interpret traffic light signals, coordinating with the vehicle's path planning and decision-making systems to ensure appropriate actions like stopping or proceeding. TSD captures and processes images to identify and interpret road signs, influencing the vehicle's driving behavior. These systems were tested and validated in the AV research platform's test vehicle at the CETRAN [30] proving test track and along public roads in Cleantech Park and NTU Singapore campus.

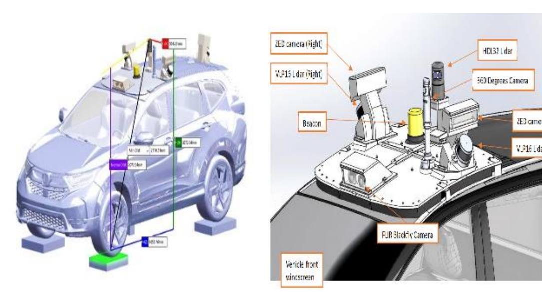
A. *AV Research Platform, Test Vehicle & Test Region*

The AV research platform uses a Honda CR-V Hybrid Electric Vehicle (HEV) as a medium-size SUV test vehicle to develop and test the AV prototype's sensor and perception systems. The platform integrates high-performance, reliable hardware components with reference autonomous driving software (ADS), ensuring compatibility and robustness. Refer to Fig. 2 for the AV prototype research platform test vehicle.

1) Hardware: The selection process prioritizes commercial off-the-shelf (COTS) items from OEM manufacturers, certified for AV development. It emphasizes CPU and GPU capabilities for efficient parallel processing of sensor data and decision-making tasks. The custom-built industrial PC with an Intel Core i9, 64GB DDR4 RAM with NVIDIA GPU RTX 3080 and AGX Orin is recommended for handling deep learning-based perception algorithms and real-time image processing. The GPU's energy-efficient design and small form factor are well-suited for complex algorithms, while its compatibility with ROS ensures seamless integration into the vehicle.

2) Sensors suite: The sensor perception system integrates various key sensors, including LiDAR, cameras, GNSS+RTK, IMU, and ultrasonic sensors, to enable comprehensive environmental awareness. The LiDAR provides 360º 3D images with high accuracy and long-range sensing, while the GNSS system combined with the local RTK network SiReNT, ensures precise positioning. The IMU offers reliable measurements of angle, angular velocity, and acceleration. Visual perception is achieved through a FLIR Blackfly camera mounted on the front view of the vehicle, providing short- and long-range 2D images, and a Mynteye/ZED stereo camera for full-field 3D measurements. Ultrasonic sensors enhance distance detection, contributing to the AV's robust perception capabilities, particularly in environments with complex traffic lights and road signs. Refer to Table I for the Vision Sensors Filed of View (FoV) device measurements.

3) Software stack: The software stack, built on ROS and running on Ubuntu 18.04, includes sensing, perception, planning, and control software packages for ADS, enabling SAE level 3 autonomy. It processes real-time data from front-facing cameras, LiDAR, and GNSS+IMU+RTK for environmental awareness, employing deep learning algorithms like YOLOv3 for object and traffic light signal recognition. Object tracking uses 2D and 3D data fusion from vision and LiDAR detectors to prevent collisions. The perception system predictions guide the ADS in making decisions regarding objects, obstacles, and traffic signals.

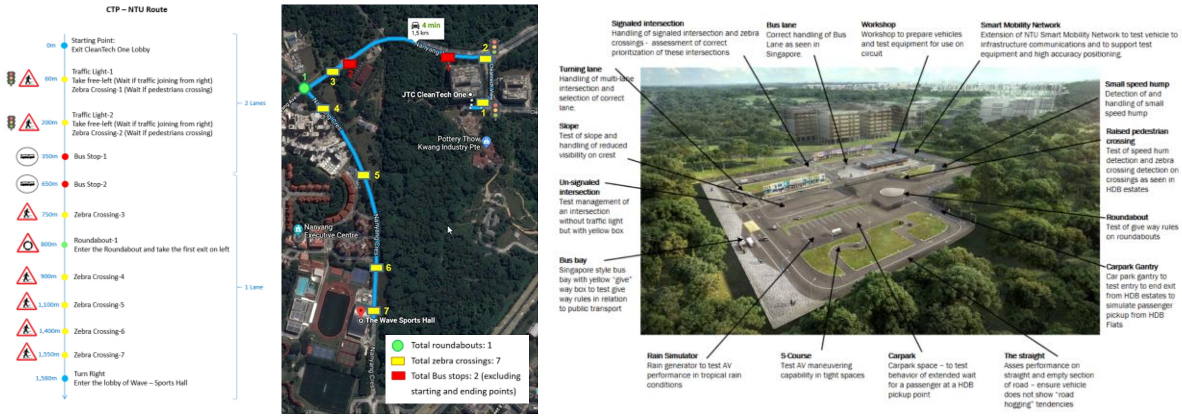


(a) (b)

Fig. 2. (a) The AV research test vehicle equipped with (b) roof-mounted sensor suite for detecting obstacles, pedestrians, traffic lights, road signs, and vulnerable road users. This sensor data supports decision-making, navigation, and control in complex environments.

TABLE I. FIELD OF VIEW (FOV) OF VISION SENSORS

Sensor	FoV Vertical	FoV Horizontal
LIDAR - HDL 32	+10 to -30 Degrees	360 Degrees
LIDAR - VLP16	+15 to -15 Degrees	360 Degrees
CAMERA - FLIR Blackfly	45 Degrees	60 Degrees
CAMERA - ZED 2	120 Degrees	120 Degrees



(a) (b)

Fig. 3. The image shows designated AV test regions at NTU's campus and CETRAN proving ground. (a) The map highlights key testing locations such as zebra crossings, bus stops, and intersections along the NTU route. (b) The CETRAN facility showcases urban driving scenarios like S-curves, carpark gantries, and smart mobility networks, critical for evaluating AV performance.

4) Testing region. The CETRAN test track facility, managed by NTU at the NTU Smart Campus Cleantech Park in Jurong Innovation District replicates urban road conditions in Singapore. This facility includes the traffic lights and road signs, bus stops, pedestrian crossings, and tropical weather scenarios like heavy rain. This facility allows for controlled, realistic testing of AVs, providing the flexibility to experiment with various AV features without on-road traffic risks. It has a proving ground for assessing AV performance and safety, validating designs for transport, and guiding AV development and certification. The site supports the NTU AV project team's progress toward Level 3 autonomy, with additional trials conducted on selected NTU campus roads approved by the transportation regulator. These trials are part of a milestone testing regime required before public road trials, focusing on mixed bi-directional traffic routes at NTU Clean Tech Park (CTP) and NTU The Wave Sports Centre. The trials routes shown in Fig. 3 covered areas (a) NTU - CTP - NTU The Wave Sports Centre, (b) CETRAN.

B. *Traffic Lights Detection and Recognition*

The TLR system is vital for the safe navigation of AVs especially in urban environments. It primarily uses camera detection, with sensor data fusion to enhance reliability and reduce false positives. The system applies pre-processing steps like color segmentation and edge detection to images before isolating regions of interest (ROI) to focus on likely by LiDAR and radar for obstacle traffic light locations, reducing computational load inputs, supplemented computational load inputs, supplemented. Initially, YOLOv3 was used for real-time detection and classification of traffic lights, balancing speed, and accuracy. YOLOv3 predicts bounding boxes and classifies traffic light states with confidence scores to minimize false positives, with temporal smoothing algorithms ensuring consistent recognition across frames. The detection process integrates with the path planning module to send a stop or go commands based on traffic light status.

To further enhance detection accuracy and performance, newer versions of YOLOv5 and YOLOv7 were introduced. YOLOv5 improved feature extraction and accuracy for small objects, while YOLOv7 offered enhanced detection capabilities with advanced backbone architectures and layer aggregation, improving performance in complex environments. Both versions provided improved computational efficiency, crucial for real-time operations, and better generalization across diverse conditions, ensuring high reliability. Their modular design also allows for future enhancements, making the system adaptable to evolving AV technology needs.

C. *Traffic Road Sign Detection*

The TSD system is vital for autonomous driving, enabling vehicles to accurately interpret and respond to road signs, ensuring safe and compliant navigation. The process starts with integrating camera sensors that capture real-time video feeds for detecting road signs. These images undergo pre-processing, such as text and color normalization and edge enhancement, to emphasize features relevant to road signs. The system isolates regions of interest (ROI) to optimize the detection process by focusing computational resources on areas likely to contain road signs.

We run an experiment for YOLOv5 and was chosen for its efficiency in real time object detection, offering a good balance between speed and accuracy. YOLOv5 predicts bounding boxes and classifies road signs into categories like stop signs and speed limits. The recognition module interprets these signs and passes the information to the decision-making module, which adjusts the vehicle's actions accordingly, such as modifying speed based on detected speed limit signs. To further improve detection accuracy and reliability, YOLOv7 was considered, offering enhanced performance and better generalization across different environmental conditions, making the detection system more robust in diverse scenarios. The detection and decision-making process involves sensor integration, pre-processing, detection architecture (starting with YOLOv5 and progressing to YOLOv7), and recognition, with each stage contributing to the system's overall efficiency and performance in AV) applications.

IV. METHODOLOGY

This section outlines the methodology employed for the development and evaluation of TLR and TSD and recognition systems in the AV research platform sensor and perception systems using custom-trained YOLO variant models. The methodology is divided into five key components: Dataset Selection, Preparation, and Preprocessing; Model Implementation; Training and Optimization; Performance Evaluation; and Comparative Analysis; and Experiment.

A. Dataset Selection, Preparation, and Preprocessing

The dataset selection, preparation, and preprocessing involved utilizing the nuances images with 2D annotations [32] and the Singapore Traffic Road Sign Dataset [33]. These datasets provide a comprehensive set of images capturing various Singapore public roads, featuring traffic lights and road signs under different environmental conditions, including daytime, night time, and light rain [34]. This diversity in the dataset is crucial for ensuring the model can generalize across various scenarios, enhancing the robustness of the AV perception system.

The custom and curated datasets used include:

- Traffic Light Dataset: Consists of approximately 700,000 image frames annotated with 17 object labels, including traffic lights with signal color status.
- Traffic Road Sign Dataset: Contains 100,000 image frames with 2,549 labels in 1,778 images across 22 class labels.

1) Data preparation is essential for ensuring effective learning and reliable real-world performance. Annotation Format was converted to YOLO-compatible formats, including class, and bounding box coordinates. Data Splitting involved dividing the datasets into training (75%), validation (15%), and testing (10%) sets, which allowed for the assessment of the model's generalization capabilities on unseen data. Class Balance was analyzed, and techniques like oversampling and class weighting were employed to address imbalances. Strategies included combining similar road sign classes and augmenting underrepresented nighttime images through targeted data augmentation and synthetic data generation. This approach aimed to balance the dataset and enhance the model's generalization ability, resulting in 1,321 training images, 279 validation images, and 178 testing images, which helped reduce the risk of overfitting.

2) Preprocessing steps involved normalization, resizing, and augmentation to further improve the model's generalization capabilities. Normalization scaled pixel values to a common range [0, 1] by dividing them by 255, ensuring consistent input feature scaling, which is crucial for the convergence of gradient-based optimization algorithms. Image Augmentation included transformations such as rotation, scaling, flipping, and color adjustments (brightness, contrast, hue, and saturation) to increase dataset diversity and simulate real-world conditions, including varying lighting and weather scenarios, to improve performance under challenging conditions like low light. A significant challenge during augmentation was the introduction of unrealistic distortions, leading to overfitting, which was mitigated by carefully tuning augmentation parameters and conducting a pilot run to evaluate the effects of each technique on model performance.

B. Model Implementation

The model implementation phase involved transforming the theoretical framework into a functional system by selecting appropriate model architectures, configuring them to meet task-specific requirements, and setting up the training environment.

1) Model architecture customization: The YOLO models—YOLOv3, YOLOv5, and YOLOv7—were selected for their optimal balance between speed and accuracy, making them well-suited for real-time applications. To enhance detection accuracy, predefined anchor boxes were recalculated using k-means clustering to better match the aspect ratios of objects like traffic lights and road signs, resolving issues with mismatched default settings. Adjustments to the number of convolutional layers and filters were made to balance accuracy and inference speed, with overfitting being mitigated by introducing dropout layers and L2 regularization, particularly in YOLOv7. Additionally, modifications to the head and neck structure, including the feature pyramid network (FPN) layers, improved detection performance for objects of varying sizes, such as distant traffic lights. Hyperparameters such as learning rate, batch size, and momentum were fine-tuned, with initial training instability due to a high learning rate being resolved through the use of a cosine annealing learning rate schedule.

2) Pre-training, environmental setup, and model initialization: The models were initialized with pre-trained weights from YOLO models trained on the COCO dataset, which were then fine-tuned on the specific traffic light and road sign dataset. The training environment was set up with GPUs, such as the Nvidia Titan RTX, and included essential software dependencies like PyTorch and CUDA to ensure efficient training. Challenges with GPU memory limitations during large batch size training were addressed using mixed-precision training, which reduced memory usage while maintaining computational efficiency. Proper initialization and checkpointing were also crucial for ensuring stable training throughout the process.

C. Training and Optimization

The goal of the training and optimization process was to maximize model performance, ensuring high accuracy and robustness in detecting and classifying traffic lights and road signs under various conditions. The process involved several key steps:

1) Training process: The training process began with the initialization of pre-trained weights, such as yolov5l.pt, which provided a foundation for fine-tuning the model according to the specific dataset requirements. Data augmentation techniques, including random cropping, scaling, and flipping, were employed to improve the model's generalization across various scenarios. Hyperparameters, such as learning rate, momentum, batch size, and the number of epochs, were meticulously tuned to strike a balance between training speed and model accuracy. The Stochastic Gradient Descent (SGD) optimizer with momentum was used to update model weights, aiding in faster convergence. The loss function was carefully designed with components for object classification, bounding box regression, and objectness score, with balanced weighting to ensure the model-maintained focus on both localization and classification accuracy. A key challenge in this process was preventing the model from disproportionately prioritizing one aspect over the other, which was effectively managed by adjusting the loss function weights through experimentation.

2) Optimization techniques: The optimization techniques employed focused on enhancing convergence speed, stability, and generalization. A cosine annealing scheduler was used for dynamic learning rate adjustments, incorporating a warm-up phase at the start to stabilize early training and prevent gradient explosions. Regularization methods, such as weight decay and dropout, were implemented to mitigate overfitting, especially during later epochs on smaller datasets. Early stopping and the use of strong dropout layers were particularly effective in preserving model performance. Data augmentation through Mosaic was crucial in increasing dataset diversity, addressing the challenge of limited data for rare road signs. Early stopping and checkpointing strategies helped avoid overfitting and preserved optimal model weights when further training risked degrading validation performance. Mixed precision training on Nvidia GPUs efficiently managed memory resources, reduced training time, and maintained high model performance. Additionally, a grid-based hyperparameter search combined with cross-validation was used to identify the best combination of learning rates, batch sizes, and momentum values, resulting in improved model convergence.

D. Performance Evaluation

The performance evaluation phase focused on assessing the effectiveness and reliability of the trained models using a comprehensive set of evaluation metrics and scenario-based testing.

1) Evaluation metrics such as mAP, IoU, and F1 score were used to evaluate the models. mAP provided an overall measure of detection performance across different object classes by averaging the precision-recall curve. IoU quantified the overlap between predicted bounding boxes and ground truth, giving insight into localization accuracy. The F1 score, which balances precision and recall, was critical for assessing the model’s ability to accurately classify and detect objects without missing or falsely detecting them.

2) Scenario-based testing trained models were deployed in various real-world conditions to evaluate their robustness and generalization capabilities. Testing was conducted in diverse environments, including urban settings with varying lighting, weather conditions, and traffic dynamics. The models were tested on the CETRAN proving test track and in multiple regions of the NTU campus, each presenting challenges like mixed bi-directional traffic, complex road layouts, and variable lighting conditions. This approach ensured the models performed reliably across a range of scenarios, reflecting the diverse conditions an AV might encounter in real-world operations.

E. Comparative Analysis

1) Quantitative analysis: The quantitative analysis compared the performance of YOLOv3, YOLOv5, and YOLOv7 using metrics such as mAP, IoU, F1 score, and inference speed (FPS). YOLOv5 and YOLOv7 demonstrated superior mAP scores compared to YOLOv3, particularly excelling in detecting smaller objects and distinguishing between different traffic signal colors and road signs, which are crucial for AV perception systems. YOLOv7 outperformed both YOLOv3 and YOLOv5 in terms of inference speed, making it ideal for real-time applications that require quick decision-making. Additionally, YOLOv5 offered a good balance between speed and accuracy, making it suitable for scenarios where both factors are important. YOLOv7 also exhibited a more optimized trade-off between accuracy and model size, enabling it to run efficiently on AV hardware platforms like the Nvidia Titan RTX and Nuvo-6108GC.

2) Qualitative analysis: The qualitative analysis focused on the models' real-world performance and their ability to generalize to diverse and challenging environments. YOLOv5 and YOLOv7 exhibited better robustness across various environmental conditions, including different lighting and weather scenarios, compared to YOLOv3. YOLOv7 produced fewer false positives and negatives, particularly in cluttered scenes with multiple objects, while YOLOv5 performed well but showed slightly more false negatives in low-light conditions. In scenario-specific performance, YOLOv7 was preferred for scenarios requiring high accuracy, such as detecting smaller, less visible road signs, whereas YOLOv5 provided balanced performance across different test regions like the NTU campus and CETRAN proving test track.

F. Experiment

1) Environmental setup: The environment setup involved both hardware and software configurations to optimize the training and deployment of the models. The models were trained offline using an Nvidia Titan RTX GPU, selected for its ability to manage large datasets and complex models like YOLOv3, YOLOv5, and YOLOv7. For real-time inference and performance evaluation during field testing, the models were deployed on the AV test vehicle's perception system, which is equipped with an Nvidia RTX GPU within the Nuvo-6108GC. The training was conducted using the PyTorch framework, supported by key libraries such as OpenCV, TensorBoard, and YOLO-specific tools for data augmentation and anchor generation. A Linux-based OS, optimized for CUDA operations, was used for both offline and onboard systems. Data management involved storing datasets on high-speed SSDs to reduce loading times during training, and for field testing, model weights and necessary datasets were preloaded onto the AV's onboard system.

2) Model implementation and training: The model implementation and training process began with the utilization of pre-trained YOLOv3, YOLOv5, and YOLOv7 models, originally trained on the COCO dataset. These models served as a foundation and were fine-tuned for specific tasks related to TLR and TSD. Custom training involved loading datasets consisting of 700,000 image frames for traffic lights and 100,000 for road signs into the training environment. The fine-tuning process focused on optimizing hyperparameters and applying data augmentation techniques like mosaic and mixup to enhance model robustness. Continuous validation monitoring was crucial in detecting overfitting and adjusting training strategies as needed. Optimization techniques included the use of a cosine annealing scheduler for dynamic learning rate adjustments, and regularization techniques like weight decay and dropout to prevent overfitting. Early stopping was employed to avoid unnecessary computation, and mixed precision training was utilized to efficiently manage memory resources, allowing for larger batch sizes without compromising training speed or accuracy.

3) Deployment and field testing: The deployment and field-testing phase involved transferring the optimized model weights to the AV perception system, enabling the AV test vehicle to process and respond to visual inputs in real-time during field tests. Field testing was conducted in key regions within NTU's campus, including the CETRAN proving test track and NTU-Clean Tech Park, where the environments presented complex traffic scenarios to evaluate the models' traffic light and road sign recognition capabilities. During these tests, real-time logging and analysis of model predictions were carried out, focusing on critical metrics such as detection accuracy, false positive/negative rates, and inference speed, to assess the performance of the deployed models.

4) Analysis and iteration: The analysis and iteration phase involved post-field-testing analysis, where performance data from each test was reviewed and verified to identify weaknesses and areas for improvement, especially in cases where the model struggled to correctly identify traffic lights or road signs under varying environmental conditions. Based on these analyses, the models were refined further through fine-tuning, retraining with augmented datasets, or architectural adjustments to better align with AV system requirements. Once optimized and validated, the models were deployed for long-term testing and evaluation in the AV test vehicle, with continuous monitoring and updates as new data and scenarios emerged.

V. Results and Discussion

A. Detection Model for Traffic Lights

The training process for the TLR model demonstrated a progressive improvement across multiple metrics as training epochs progressed. The loss curves, including box loss, object loss, and classification loss, reveal the training dynamics and model convergence behavior. The box loss started at a relatively, higher value and gradually decreased over the training epochs, indicating that the model was improving in accurately predicting the bounding boxes for detected objects. The object loss also followed a declining trend, suggesting that the model was becoming better at identifying whether an object exists in each bounding box. Similarly, the classification loss reduced as training continued, which shows that the model's ability to correctly classify the detected objects improved over time.

1) Epochs of significant improvements at *epoch 30-50*: Notable improvements were observed in this range, where the precision and recall metrics showed significant jumps. The loss curves also depicted steeper declines during these epochs, indicating faster convergence. *Epoch 90-110:* Another period of significant improvement occurred, particularly in the mAP metrics. The model's ability to detect objects with higher accuracy at varying IoU thresholds improved markedly.

2) Challenges encountered toward the later stages of training, particularly after epoch 120, signs of overfitting began to emerge. This was indicated by a plateau in validation metrics such as mAP@0.5 and mAP@0.5:0.95, while the training metrics continued to improve. Overfitting was also suggested by the increase in validation losses despite the continual decrease in training losses. In the early epochs (0-20), the model showed some signs of underfitting, where the precision and recall metrics were relatively low, and the loss values were high. This phase was marked by a slower reduction in losses and a gradual improvement in detection metrics.

The initial training results for the TLR model reveal a detailed analysis of the model's performance in recognizing various traffic light states, including "RedLeft," "Red," "GreenLeft," and "Yellow." The model showed moderate accuracy, with a precision of approximately 86% and a recall of about 80%. The mAP at an IoU threshold of 0.5 (mAP@0.5) reached around 86%, while the more stringent mAP@0.5:0.95 was approximately 65%. These results indicate that while the model performs adequately in many cases, there is room for improvement, particularly in distinguishing between similar traffic light states, which is crucial for making safe and accurate driving decisions in real-world AV applications. Refer to Table II for the summary of the training model results. The confusion matrix shows that the model was relatively accurate in distinguishing between different traffic light states, but there were instances of misclassification, particularly between visually similar states. For example, the model sometimes confused "RedLeft" with "Red" or "GreenRight" with "GreenLeft." The F1-Confidence curve further illustrates the model's precision and recall balance across different confidence thresholds, showing an optimal F1 score of 0.83 a confidence threshold of 0.65, which suggests the model can be further optimized to improve overall performance. Another observation in the initial model tested offline using recorded image bag files, where several issues were identified. This likely due to instance distribution imbalance. Specifically, there was an imbalance in the number of samples for objects and traffic lights, as well as between daytime and night time samples. Although the model was able to detect some of the traffic light status, it failed to identify the green right signal due to its proximity to another green traffic light signal. Another observation on the system fails to identify a traffic light status that is present in the scene. This can be problematic because it means that the detector missed an object it was supposed to detect. This can add to the false negative detection. Refer to Fig. 4 and Fig. 5 for the observation.

TABLE II. SUMMARY OF THE INITITAL TRAINING RESULTS TRAFFIC LIGHT DETECTION MODEL – YOLOV3

Metric	Results
Dataset	TLR Dataset
Precision	86%
Recall	80%
F1 score	0.83 at 0.65
mAP @0.5	86%
mAP @0.95	65%
Inference Speed (ms/image)	12 ms/frame
Model Size (Parameters)	44M

(a) (b)

Fig. 4. The image compares traffic light detection with ground truth validation. (a) The original scene without detected objects. (b) The system successfully detects and labels two red traffic lights with confidence scores of 0.63 and 0.53. Ground truth annotations are used to validate the system's accuracy in real-world conditions.

(a) (b)

Fig. 5. The image demonstrates the system's detection and recognition capabilities under challenging conditions. (a) Left: the system detects multiple red traffic lights and a 'mandatory left' sign during heavy rain. (b) Right: in low visibility, the system accurately identifies a construction site and split signs but misses a 'give way' sign and a red traffic light (false negatives highlighted in red).

The model's inference speed is an essential factor, particularly for real-time applications like autonomous driving. The model demonstrated an inference speed of approximately 12 milliseconds per image, which is fast enough for real-time detection tasks. Additionally, the model parameter size is around 44 million (44M), striking a balance between complexity and computational efficiency.

Evaluating the performance of a model for deployment in an AV it is important to consider the trade-offs between accuracy, inference speed, and model size. In this case, the model's inference speed of 12 milliseconds per image is sufficient for real-time applications, but the trade-off is a relatively moderate accuracy, particularly under more stringent evaluation conditions (mAP@0.5:0.95). For deployment in an AV test vehicle, a balance must be struck between accuracy and speed. If higher accuracy is required, more complex models or ensemble methods could be used, though this would likely reduce inference speed. Conversely, if speed is prioritized, a more streamlined model might be necessary, though this could compromise accuracy.

The TLR model exhibits moderate accuracy and a fast inference speed, making it suitable for real-time applications in AV. However, the performance of the model could be improved by addressing class imbalances, enhancing data augmentation techniques to better handle variations in lighting and angles, and fine-tuning the model's hyperparameters. Additionally, considering more advanced architectures or employing ensemble learning methods could help improve accuracy and robustness, particularly in distinguishing between visually similar traffic light states. For deployment in an AV test vehicle, this model could serve as a baseline, but further refinements may be necessary depending on the specific requirements for accuracy and speed in the target application. Balancing the trade-offs between accuracy and inference speed will be critical to ensuring that the model meets the operational demands of real-time autonomous driving. Refer to Table III for the summary of the TLR performance results.

B. Detection Model for Traffic Road Signs

The initial training results for the three YOLOv5 models—YOLOv5s, YOLOv5m, and YOLOv5l—demonstrate different strengths in detecting and recognizing traffic signs, varying in accuracy, speed, and computational demands. As training progresses, there is a consistent decrease in losses related to bounding box regression, object classification, and class prediction, indicating that the models are effectively learning. The reduction in metrics like train/box loss, which starts at 0.11157 and decreases to around 0.0235, suggests improved accuracy in predicting object locations and class labels.

Performance metrics such as precision, recall, and mean Average Precision (mAP) also improve throughout training. Precision stabilizes and exceeds 0.93, while recall shows a steady increase, indicating the model's ability to capture more relevant objects. The mAP metrics, including mAP_0.5, reach approximately 0.96 by the end of training, reflecting the model's effectiveness across various IoU thresholds. Refer to the summary Table IV training and Table V performance.

TABLE III. TRAFFIC LIGHT DETECTION PERFORMANCE RESULTS

Model	Precision (%)	Recall (%)	F1 Score (%)	mAP@0.5	mAP@0.95	Inference Speed (fps)	Model Size
YOLOv3	86	80	83	0.86	0.65	12	240
YOLOv5	90	87	88.5	0.90	0.68	28	140
YOLOv7	93	90	91	0.93	0.70	22	120

TABLE IV. SUMMARY OF TRAINING RESULTS TRAFFIC ROAD SIGN – YOLOV5

Model	Batch Size	Precision	Recall	mAP @0.5	mAP @0.95	Inference Speed (ms/image)	Model Size (Parameters)
YOLOv5s	16	90%	85%	90%	58%	7 ms	7M
YOLOv5m	32	92%	88%	93%	61%	10 ms	21M
YOLOv5l	64	93%	87%	94%	60%	14 ms	47M

TABLE V. TRAFFIC ROAD SIGN DETECTION PERFORMANCE RESULTS

Model	Precision (%)	Recall (%)	F1 Score (%)	mAP@0.5	mAP@0.95	Inference Speed (fps)	Model Size
YOLOv3	92	90	91	0.92	0.70	25	240
YOLOv5	94	92	93	0.94	0.72	28	140
YOLOv7	96	94	95	0.96	0.74	22	120

Validation losses, which decrease consistently alongside training losses, indicate strong generalization to unseen data and a low risk of overfitting. The learning rates, initially set high and gradually reduced, facilitate faster convergence early on and fine-tuning later in the training process.

Despite the strong performance, further improvements could be achieved through advanced augmentation techniques, fine-tuning hyperparameters, or applying model pruning and quantization to reduce model size and inference time. These adjustments would enhance the model's suitability for deployment in resource-constrained environments like an AV test vehicle, where the trade-off between accuracy and inference speed is crucial.

C. Field Testing in Real-World Scenarios

Field testing of TLR and road TSD models was conducted on the AV test vehicle using the Nuvo-6108GC PC environment equipped with an NVIDIA RTX GPU. The goal was to evaluate the model's performance in real-world scenarios reflective of Singapore's urban conditions. Testing occurred across test regions: the CETRAN proving test track, JTC Cleantech Park, and NTU’s campus, each presenting distinct challenges. The CETRAN test track, designed for AV testing, replicates various Singaporean urban road elements, offering a controlled environment for rigorous testing without the risks of live traffic. This site allowed for comprehensive assessment of the AV’s manoeuvrability through city driving scenarios [31].

1) Test regions and specific challenges. CETRAN facility provided a geofenced, controlled environment showcasing urban scenarios like S-curves and carpark gantries, crucial for evaluating AV performance. Traffic lights and road signs were readily available for evaluation without traffic interference. NTU-Clean Tech Park (CTP) - The Wave Sports and Recreation Centre featured mixed bi-directional traffic, creating a complex environment to test the model’s ability to accurately detect and classify traffic lights and road signs amidst dynamic vehicle movements and varying traffic density. The Nuvo-6108GC, with its NVIDIA RTX GPU, supported the real-time processing needs of advanced deep learning models, achieving an average inference speed of 12ms per frame, suitable for urban traffic scenarios. The model demonstrated robust detection, maintaining high precision and recall metrics across various conditions, including challenging lighting and complex traffic scenarios.

2) Challenges observed during field testing. The model performed well overall, but low-light scenarios occasionally reduced detection accuracy for certain road signs. The model navigated intersections and recognized pedestrian crossings effectively, though closely spaced traffic lights sometimes caused minor detection delays. The model maintained detection capabilities in light rain, but heavy rain introduced reflections that occasionally confused the perception system.

D. Lessons Learned

The performance evaluation of the model development and deployment in the AV research platform focused on enhancing safety and robustness, particularly in image recognition, motion speed profiles, and obstacle detection at critical areas like junctions and pedestrian crossings. Several test scenarios, including encounters with pedestrians and other vehicles, provided insights that guided improvements in design and safety approaches.

1) Model development and training revealed that the choice of model architecture significantly impacted the balance between detection accuracy and inference speed. YOLOv7 excelled in complex urban environments but required higher computational resources. Data augmentation techniques like mosaic and mixup were essential in improving the model's robustness across diverse conditions, though issues with data imbalance, particularly in night time scenarios, highlighted the need for a balanced training dataset.

2) Deployment and real-time inference showed that fine-tuning was necessary to optimize real-time performance. While YOLOv7 offered the highest accuracy, its computational demands resulted in slightly slower inference times, particularly in high-traffic environments where quick decision-making was crucial. This trade-off emphasized the need for further optimization to handle scenarios with high visual complexity.

3) Field testing observations across the CETRAN proving test track, NTU-Clean Tech Park, and NTU campus highlighted YOLOv7's superior detection accuracy, especially in identifying over 95% of traffic lights and road signs. However, the model's inference speed sometimes lagged in complex traffic scenarios, making YOLOv5 a more balanced choice in environments requiring rapid processing. False positives and negatives were noted, particularly in varying lighting conditions, underscoring the need for improved robustness against environmental noise and challenging conditions like heavy rain and glare. Our team also tested to other testing site to verify the performance of the object detection using other use case in the golf range [35] and integrated wildlife recreation area [36] how the model performs in a different field testing and observation.

4) Post-field-testing analysis involved a detailed review of model failures, particularly under low-light and adverse weather conditions, which led to higher false-negative rates. The analysis stressed the importance of enhancing data augmentation and possibly integrating additional sensors like LiDAR. While YOLOv7 achieved impressive metrics under ideal conditions (precision of 95%, recall of 93%, F1 score of 94%, and mAP@50 of 0.95), these dropped under challenging conditions, indicating a need for further refinement. Inference speed was 22 fps for YOLOv7, compared to 28 fps for YOLOv5, highlighting the trade-offs between speed and accuracy.

VI. Conclusion

This research made significant advancements in AV perception systems by evaluating performance, comparing YOLO variants, conducting real-world testing, and developing a comprehensive testing framework. YOLOv7 emerged as the best-performing model, achieving a mAP@0.5 of 93% for Traffic Light Recognition (TLR) and 96% for Traffic Sign Detection (TSD), even in challenging environments such as low-light and occlusion scenarios. Its superior precision and recall, with F1 scores of 91% for TLR and 95% for TSD, demonstrated its suitability for real-time AV applications. YOLOv5, while slightly less accurate, provided a strong balance between speed and accuracy, making it adaptable for various conditions.

Real-world testing in urban environments, including the CETRAN proving test track and NTU campus, validated YOLOv7's reliable performance and its readiness for deployment in AV systems. A significant contribution of the study was the development of a scenario-based testing framework that included continuous performance monitoring and model refinement. This framework helped identify and correct model weaknesses, enhancing performance and robustness.

The research offers critical insights for AV developers and researchers, particularly on balancing speed, accuracy, and robustness in real-world applications. Future work could focus on exploring hybrid models that integrate YOLO with Transformer-based architectures, testing in diverse environments such as rural areas and highways, incorporating multi-modal data (LiDAR, RADAR), and exploring adversarial robustness testing.

Acknowledgment

This research acknowledges the AV research team of Energy Research Institute (ERI@N) Nanyang Technological University Singapore.

References

[1] Singapore Police Force, "Annual Road Traffic Situation 2021," [Online].Available:https://www.police.gov.sg/media/D4435F72157942 D3B323EE4A507D4CFB.ashx. [Accessed: 11-Aug-2024].

[2] D. Lee, "The Big Read: Amid rising number of traffic accidents, a road culture that's proving hard to change," Channel NewsAsia, 23-Aug-2023.[Online].Available:https://www.channelnewsasia.com/singapore/big-read-rising-traffic-accidents-road-culture-4328841.[Accessed:11-Aug-2024].

[3] World Health Organization, "Global status report on road safety 2023," 2023.[Online].Available:https://www.who.int/publications/i/item/97892 40045747. [Accessed: 11-Aug-2024].

[4] M. B. Jensen, M. P. Philipsen, A. Møgelmose, T. B. Moeslund and M. M. Trivedi, "Vision for Looking at Traffic Lights: Issues, Survey, and Perspectives," in IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 7, pp. 1800-1815, July 2016, doi: 10.1109/TITS.2015.2509509.

[5] R. Zhang, K. Zheng, K., P. Shi, Y. Mei, H. Li, T. Qiu, “Traffic Sign Detection Based on the Improved YOLOv5,” Appl. Sci. 2023, 13, 9748. https://doi.org/10.3390/app13179748.

[6] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2015, pp. 1440-1448, doi: 10.1109/ICCV.2015.169.

[7] W. Liu, et al. , “SSD: Single Shot MultiBox Detector” In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science(), vol 9905. Springer, Cham. https://doi.org/10.1007/978-3-319-46448-0_2.

[8] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91

[9] E. Marti, M. A. de Miguel, F. Garcia and J. Perez, "A Review of Sensor Technologies for Perception in Automated Driving," in IEEE Intelligent Transportation Systems Magazine, vol. 11, no. 4, pp. 94-108, winter 2019, doi: 10.1109/MITS.2019.2907630

[10] H. Zhao et al., "Driving Scenario Perception-Aware Computing System Design in Autonomous Vehicles," 2020 IEEE 38th International Conference on Computer Design (ICCD), Hartford, CT, USA, 2020, pp. 88-95, doi: 10.1109/ICCD50377.2020.00031.

[11] M. J. Shafiee, A. Jeddi, A. Nazemi, P. Fieguth and A. Wong, "Deep Neural Network Perception Models and Robust Autonomous Driving Systems: Practical Solutions for Mitigation and Improvement," in IEEE Signal Processing Magazine, vol. 38, no. 1, pp. 22-30, Jan. 2021, doi: 10.1109/MSP.2020.2982820.

[12] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018. [Online]. Available: https://arxiv.org/abs/1804.02767

[13] G. Jocher, J. Nishimura, and L. Tao, "YOLOv5," Ultralytics, 2020. [Online]. Available: https://docs.ultralytics.com/yolov5/. [Accessed: Jan. 20, 2024

[14] 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," 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 2023, pp. 7464-7475, doi: 10.1109/CVPR52729.2023.00721.

[15] J. G. Wang and L. B. Zhou, “Traffic Light Recognition with High Dynamic Range Imaging and Deep Learning,” IEEE Trans. Intell. Transp. Syst., vol. 20, no. 4, pp. 1341–1352, Apr. 2019. DOI: 10.1109/TITS.2018.2859226.

[16] M. Bach, D. Stumper, and K. Dietmayer, “Deep Convolutional Traffic Light Recognition for Automated Driving,” in IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 2018, vol. 2018-November, pp. 851–858. DOI: 10.1109/ITSC.2018.8569907.

[17] Z. Shi, Z. Zou, and C. Zhang, “Real-time traffic light detection with adaptive background suppression filter,” IEEE Trans. Intell. Transp. Syst., vol. 17, no. 3, pp. 690–700, Mar. 2016. DOI: 10.1109/TITS.2015.2480700.

[18] J. Levinson, J. Askeland, J. Dolson, and S. Thrun, “Traffic light mapping, localization, and state detection for autonomous vehicles,” in Proceedings - IEEE International Conference on Robotics and Automation, 2011, pp. 5784–5791. DOI: 10.1109/ICRA.2011.5980138.

[19] M. Guo, X. Zhang, D. Li, T. Zhang, and L. V. An, “Traffic light detection and recognition for autonomous vehicles,” The Journal of China Universities of Posts and Telecommunications, vol. 22, no. 1, pp. 50–56, Feb. 2015. DOI: 10.1016/s1005-8885(15)60624-0.

[20] S. Saini, S. Nikhil, K. R. Konda, H. S. Bharadwaj and N. Ganeshan, "An efficient vision-based traffic light detection and state recognition for autonomous vehicles," 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 2017, pp. 606-611. DOI: 10.1109/IVS.2017.7995785.

[21] Y. Lu, J. Lu, S.-H. Zhang, and P. M. Hall, “Traffic signal detection and classification in street views using an attention model,” Computational Visual Media, vol. 4, no. 3, pp. 253–266, Aug. 2018. DOI: 10.1007/s41095-018-0116-x.

[22] M. P. Philipsen, M. B. Jensen, A. Møgelmose, T. B. Moeslund and M. M. Trivedi, "Traffic Light Detection: A Learning Algorithm and Evaluations on Challenging Dataset," 2015 IEEE 18th International Conference on Intelligent Transportation Systems, Gran Canaria, Spain, 2015, pp. 2341-2345. DOI: 10.1109/ITSC.2015.378.

[23] A. Fregin, J. Muller, U. Krebel and K. Dietmayer, "The DriveU Traffic Light Dataset: Introduction and Comparison with Existing Datasets," 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD, Australia, 2018, pp. 3376-3383. DOI: 10.1109/ICRA.2018.8460737.

[24] M. Weber, P. Wolf, and J. M. Zollner, “DeepTLR: A single deep convolutional network for detection and classification of traffic lights,” in IEEE Intelligent Vehicles Symposium, Proceedings, 2016, vol. 2016-August, pp. 342–348. DOI: 10.1109/IVS.2016.7535428.

[25] Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu, “Traffic-Sign Detection and Classification in the Wild,” in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, vol. 2016-December, pp. 2110–2118. DOI: 10.1109/CVPR.2016.231.

[26] I. M. Creusen, R. G. J. Wijnhoven, E. Herbschleb, and P. H. N. De With, “Color exploitation in HOG-based traffic sign detection,” in Proceedings - International Conference on Image Processing, ICIP, 2010, pp. 2669–2672. DOI: 10.1109/ICIP.2010.5653893.

[27] G. Wang, G. Ren, Z. Wu, Y. Zhao, and L. Jiang, “A robust, coarse-to-fine traffic sign detection method,” in Proceedings of the International Joint Conference on Neural Networks, 2013. DOI: 10.1109/IJCNN.2013.6706844.

[28] H. Luo, Y. Yang, B. Tong, F. Wu, and B. Fan, “Traffic Sign Recognition Using a Multi-Task Convolutional Neural Network,” IEEE Trans. Intell. Transp. Syst., vol. 19, no. 4, pp. 1100–1111, Apr. 2018. DOI: 10.1109/TITS.2017.2784095.

[29] H. H. Aghdam, E. J. Heravi, and D. Puig, “A practical approach for detection and classification of traffic signs using Convolutional Neural Networks,” Robotics and Autonomous Systems, vol. 84, pp. 97–112, Oct. 2016. DOI: 10.1016/j.robot.2016.07.003.

[30] CETRAN, "Centre of Excellence for Testing & Research of Autonomous Vehicles," CETRAN. [Online]. Available: https://cetran.sg/. [Accessed: Jan. 20, 2024].

[31] SAE International, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," SAE Standard J3016_202104, Apr. 2021. DOI: 10.4271/J3016_202104.

[32] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. Ionasescu, A. Y. Zhou, J. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuScenes: A multimodal dataset for autonomous driving," in Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 11618-11628. DOI: 10.1109/CVPR42600.2020.01163.

[33] E. Yan, "Classification and Detection of Singapore Road Traffic Signs,"GitHubRepository,https://github.com/eugeneyan84/Classification-and-Detection-of-Singapore-Road-Traffic-Signs. [Accessed: Jan. 20, 2024].

[34] Y. Yang, H. Luo, H. Xu, and F. Wu, "Towards Real-Time Traffic Sign Detecton and Classification," in IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 7, pp. 2022-2031, July 2016.

[35] J. De Guia et al., "Autonomous Golf Ball Picker: The First in Singapore Golf Environment," 2023 IEEE 12th Global Conference on Consumer Electronics (GCCE), Nara, Japan, 2023, pp. 900-901, doi: 10.1109/GCCE59613.2023.10315528.

[36] J. De Guia et al., "Advancing Safety and Robustness: Perception-Planning System of an Autonomous Vehicle Last Mile Delivery," 2024 IEEE Conference on Artificial Intelligence (CAI), Singapore, Singapore, 2024, pp. 113-118, doi: 10.1109/CAI59869.2024.00260.