

Vision-Based Adaptive Traffic Signal Control Using Deep Learning for Real-Time Traffic State Estimation

Swapna Raghunath, Department of Electronics and Communications Engineering, G.Narayanamma Institute of Technology and Science, Hyderabad, India

swapna.raghunath@gnits.ac.in

Yerrarapu Sravani Devi, Assistant Professor, Computer Science and Engineering, GITAM School of Computer Science and Systems Engineering, GITAM Deemed to be University, Hyderabad, India.

syerrara@gitam.edu

Meeniga Vijaya Lakshmi, Department of Electronics and Communications Engineering, G.Narayanamma Institute of Technology and Science, Hyderabad, India

m.vijayalakshmi@gnits.ac.in

Bestha Venkateshulu, Department of Electronics and Communications Engineering, G.Narayanamma Institute of Technology and Science, Hyderabad, India

b.venkateshulu@gnits.ac.in

Abstract

Urban traffic congestion at signalized intersections is intensified by highly variable and heterogeneous traffic demand, making fixed-time signal control inefficient under changing conditions. This study proposes a vision-based adaptive traffic signal control framework that integrates deep-learning-based vehicle detection, multi-object tracking, real-time traffic-state estimation, and adaptive green-time allocation. Video data collected from three signalized intersections in Hyderabad were processed using YOLOv8 to detect and classify five vehicle categories: cars, two-wheelers, auto-rickshaws, buses, and heavy vehicles. Vehicle detections were subsequently used to estimate traffic count, occupancy, queue length, and waiting time, which were integrated into a Traffic Demand Index (TDI) for adaptive signal control. The proposed approach was evaluated against fixed-time and conventional density-based control strategies. The reported detection results achieved 94.1% precision, 92.7% recall, 93.4% F1-score, and 95.2% mAP@50. The overall mean absolute percentage error (MAPE) was 4.5%. Compared with fixed-time control, the proposed method reduced average waiting time from 96.4 to 62.7 s and average queue length from 71.5 to 45.2 m, while increasing throughput from 2486 to 2832 vehicles/h. It also achieved lower waiting time and queue length than density-based control. These results indicate that integrating deep-learning-based visual perception with an interpretable traffic-demand-driven controller can provide responsive intersection-level traffic management under heterogeneous traffic conditions.

1. INTRODUCTION

Urban traffic congestion has become a constant issue in the rapidly developing cities under the impact of rapid urbanization, rising vehicle ownership and the growing demand for urban transportation. This is important especially at signalized intersections, where traffic flows can significantly differ from approach to approach and from day to day. In Hyderabad this is an example challenge. As per the 2025 TomTom Traffic Index, Hyderabad had an average level of congestion at 55.5 per cent with traffic users spending around 123 hours during rush hour. The time spent travelling for a 10 km journey averaged 32 min 37 s and congestion during the evening rush time was 85.7% [1]. The following conditions suggest that upgrading road

infrastructure alone may not be enough and that there is a need to upgrade the management of existing signalized intersections.

Signalized intersections are important nodes in an urban road network because traffic volumes on the different approaches are seldom balanced and/or constant. Traditional fixed-time control signals provide fixed green, yellow and red intervals without consideration of the actual traffic demand at the instant. Therefore, the approach with less traffic will be receiving green time while the approach with more traffic will continue to have increasing queue length and delay. This is an inefficient way to allocate the road space and can cause more delays, multiple stops and starts, higher fuel usage and lower intersection flow. This restriction is greater in mixed traffic where vehicles such as cars, two-wheeled vehicles, auto-rickshaws, buses and heavy vehicles share limited road space and have significantly different sizes, manoeuvrability and driving habits. Traffic volume and queue formation at signalized intersections with mixed traffic and non-lane-based traffic has also been reported as being significantly different in field investigations conducted in Hyderabad [2], [3]. Figure 1 depicts a typical traffic congestion scenario and uneven traffic distribution at an urban signalized intersection.



Fig. 1. Typical Traffic congestion and uneven traffic demand at an urban signalized intersection

Predicted traffic conditions have therefore become an important alternative to the predetermined signal timing technique, and adaptive traffic signal control (ATSC) has developed as an important alternative strategy [4]. In recent years, research on machine learning and Deep Reinforcement Learning (DRL) has increased in the area of dynamically identifying signal-control policies [5]. For instance, Haddad et al. (2021) created a cooperative deep reinforcement-learning system for several intersections and demonstrated that the agents at the intersections can be coordinated to enhance traffic-performance metrics [6]. Long et al. (2022) explored the DRL-based transit signal priority, which showed the feasibility of learning-based control for traffic conditions that change dynamically [7]. Later, Yazdani et al. (2023) introduced Intelligent Vehicle Pedestrian Light framework, which considers the interaction between vehicles and pedestrians in signal control decision making [8]. Liu et al. (2023) investigated a teacher-student reinforcement-learning framework, which simplifies the complexity of training traffic signal-control policies [9]. Recent studies further developed adaptive control to multi-objective and network-level optimization. The study by Zhang et al. (2024) delved into the multi-objective DRL approach, which incorporated elements of traffic

efficiency, safety, and decarbonization, and Cai and Wei (2024) presented an improved DRL-based signal controller that featured advancements in training and robustness [10] [11].

Other research works have focused on co-evolutionary and multi-agent reinforcement-learning solutions for time-varying traffic demand and multi-intersection coordination [12] [13] [14]. The developments reflect the capabilities of learning-based control, but also increased complexity of state representations, training environments, optimization procedures and computational requirements. When deploying these approaches in the field, the complexity of these approaches may become a practical consideration, especially where the individual intersections where they are deployed lack sensing infrastructure.

Another crucial requirement for adaptive signal control is the ability to well estimate the dominant traffic state. The camera-based traffic monitoring approach is a non-intrusive way to gain traffic information without extensive pavement-based sensors [15]. The traditional image-processing methods used to obtain vehicle counts, such as grayscale conversion, thresholding, background subtraction, and foreground segmentation, can be used to get the basic vehicle number, but their accuracy can be reduced by the influence of changes in illumination, shadows, vehicle appearances, occlusion, and the differences in traffic flow [16]. This is especially pertinent for urban traffic in India, where image regions are often partially or completely covered by two wheelers and auto rickshaws.

A better alternative is to implement deep-learning-based object detection, which can recognize individual vehicles and their classes directly from traffic images. The YOLO family of detectors has therefore been the subject of much research for intelligent transportation systems, where the goal is to simultaneously locate and classify objects, along with relatively efficient inference [17] [18]. The detection of YOLO-based method can get more information than just foreground segmentation, such as vehicle class, position, and spatial distribution. Together with temporal tracking, these detections can also help estimate vehicle count, vehicle occupancy, vehicle queue length and waiting time. Existing vision-based research, though, tends to focus mainly on vehicle detection or traffic monitoring, and the application of adaptive signal-control systems are often conducted in simulated environment or in highly complex reinforcement-learning frameworks. The relationship between real world camera-based traffic perception and an interpretable adaptive signal-control mechanism is less explored.

In this study, the aforementioned lack is addressed by proposing an integrated vision-based adaptive traffic signal control framework tailored for the heterogeneous urban traffic. Three signalized intersections in Hyderabad (Suchitra Junction, Crown Cafe Junction and KBR Park Junction) were chosen as experimental intersections to represent mixed traffic, unequal approach loading and traffic demand. These have also been analyzed in earlier field studies in Hyderabad under mixed traffic and non-lane-based conditions, creating a relevant real time environment for studying traffic at intersections. This study employs direct observation video footage of traffic captured at the selected intersections from fixed-position cameras that are equipped with RGB sensors. The camera view is set up so that it will show the oncoming traffic, the stop line, and the line of cars. The three intersections are analysed during approximately 3 hours of traffic video for each of the 7 days of observation, representing the morning peak, afternoon/off-peak and evening peak times. Based on these recordings, about 24,000 frames are manually annotated for use in model development and assessment.

The proposed framework employs YOLOv8 for vehicle detection and classification into five categories: car, two-wheeler, auto-rickshaw, bus, and truck/heavy vehicle [19]. The detected

vehicles are subsequently tracked across consecutive frames to obtain temporal information required for traffic-state estimation. Vehicle count, occupancy, queue length, and waiting time are estimated for each traffic approach and integrated into a Traffic Demand Index (TDI). Unlike a controller based only on vehicle count, the TDI represents multiple characteristics of traffic demand and congestion. The resulting traffic state is used by the adaptive controller to dynamically allocate green time while maintaining predefined minimum and maximum green-time constraints and preventing excessive waiting on individual approaches.

The effectiveness of the proposed framework is evaluated against two reference strategies: conventional fixed-time signal control and density-based adaptive control [20] [21]. Evaluation is performed using both perception-level and traffic-management measures, including precision, recall, F1-score, mAP, vehicle-counting error, queue length, waiting time, delay, throughput, vehicles served per cycle, and signal utilization. The study therefore establishes a complete perception-to-control framework in which real-world traffic video is transformed into vehicle-level information, vehicle-level information into traffic-state indicators, and traffic-state indicators into adaptive signal-control decisions.

The principal contribution of this work is not simply the replacement of conventional image processing with deep-learning-based vehicle detection. Rather, it is the integration of deep-learning-based visual perception, temporal vehicle tracking, multi-parameter traffic-state estimation, and interpretable adaptive signal control within a single closed-loop framework. By using real traffic observations from heterogeneous Hyderabad intersections, the study aims to provide a practical and scalable approach to intersection-level traffic management while avoiding the complexity of fully learning-based signal-control policies.

2. METHODOLOGY

The proposed methodology establishes a vision-based closed-loop framework for adaptive traffic signal control in which real-time traffic video is converted into vehicle-level information, traffic-state parameters, and subsequently adaptive signal-control decisions. The methodology integrates traffic-video acquisition, dataset preparation, annotation, video preprocessing, deep-learning-based vehicle detection and classification, multi-object tracking, traffic-state estimation, Traffic Demand Index (TDI) computation, adaptive green-time allocation, and comparative performance evaluation. The methodology is specifically tailored to the heterogeneous nature of urban traffic, which is characterised by varying traffic demand, queue configuration and waiting situations between approaches and over time. The proposed framework simultaneously uses multiple parameters associated with traffic states to calculate the signal priority of each approach, unlike traditional image processing methods, which rely mainly on the number of vehicles. The overall processing process is shown in Fig. 2.

2.1 Overall System Architecture

The proposed system is designed to be a continuous system of perception to control. Video images of traffic captured by stationary video cameras at selected signalized intersections are initially analyzed to identify and classify individual cars. Temporal information of movement and stopping behaviour of the detected vehicles is then obtained by tracking them across consecutive frames.

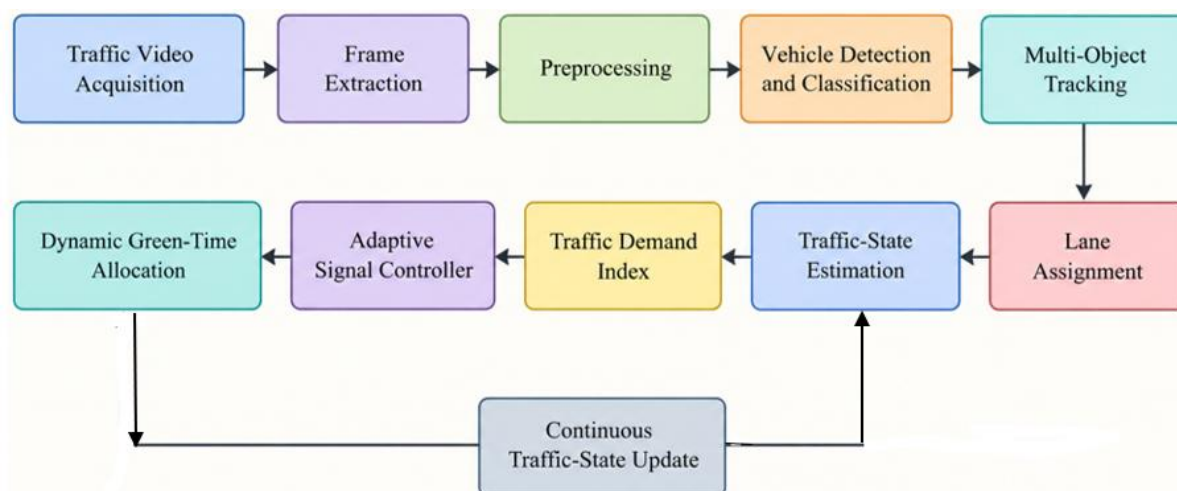


Fig. 2. Process flow of vision-based closed-loop framework for adaptive traffic signal control

The tracked vehicle information is used to estimate the state of the traffic for each approach by the number of vehicles, traffic density, lane occupancy, queue length and waiting time. These parameters are then normalised and combined into a Traffic Demand Index (TDI) that expresses the instant traffic demand for each approach [22]. The adaptive controller uses the relative TDI values to allocate green time whilst adhering to a minimum and maximum green-time constraint and avoiding unnecessary waiting on individual approaches. The entire processing procedure is depicted as in Figure 2. Its closed-loop architecture enables the controller to continually adjust the signal decision as other traffic observations are inputted. Therefore, the signal timing is not based on a predetermined fixed-time schedule, but rather depends on the current traffic conditions.

2.2 Traffic Data Collection and Video Frame Preprocessing

The experimental study is performed on actual traffic videos captured at three intersections of Hyderabad city, India, namely Suchitra Junction, Crown Cafe Junction and KBR Park Junction. These intersections represent heterogeneous urban traffic conditions with different levels of traffic demand and unequal loading across approaches. The selected locations are also reported in earlier field investigations of mixed and non-lane-based traffic conditions in Hyderabad, providing a relevant field setting for evaluating traffic-state estimation under heterogeneous traffic.

The intersections are selected based on clear visibility of approaching traffic, visibility of the stop line and queue region, stable camera position, adequate field of view covering the major traffic approaches, presence of heterogeneous traffic and occurrence of cars, two-wheelers, auto-rickshaws, buses, and heavy vehicles. The observations are conducted during three representative periods: Morning peak: approximately 08:00–09:00 h; Afternoon/off-peak: approximately 13:00–14:00 h; and Evening peak: approximately 17:00–18:00 h. Each observation session is approximately 60 min in duration. Data are collected over seven observation days for each intersection, resulting in approximately 63 hours of raw traffic video across the three study locations and observation periods. The use of multiple intersections, days, and time periods provides variation in traffic volume, vehicle composition, queue formation, and illumination conditions.

The primary dataset consists of directly captured RGB traffic video obtained using fixed-position cameras overlooking the selected intersections. The cameras are positioned to capture

the approaching traffic, stop line, and queue region within the field of view. The videos were recorded at approximately 1920×1080 pixels and 25–30 frames/s to provide sufficient spatial and temporal information for vehicle detection and tracking. Raw video recordings are kept for further analysis. In order to prevent the excessive redundancy between successive frames, representative frames are extracted from the continuous video with a certain sampling interval. The recordings yielded over 28,000 candidate frames when sampled at approximately 8 s intervals.

After quality screening and exclusion of unsuitable and highly repetitive frames, approximately 24,000 frames remained for manual annotation. The vehicle detection and classification algorithms are based on the YOLOv8 deep-learning detection system. The whole processing chain comprises image processing, detection, tracking, traffic state estimation, computation of TDI and real-time traffic-state estimation. MATLAB software was used to implement the algorithm. The final data set consists of ~24,000 annotated traffic frames for the three intersections, seven observation days and three observation periods. Selected frames represent low, medium and high traffic volume situations. Five categories of vehicles are taken into account: Car, Two-wheeler, Auto-rickshaw, Bus and Truck/heavy vehicle.

Variations in vehicle scale, vehicle partial occlusion, traffic density, illumination, shadows, road position and vehicle orientation are present in the selected frames. This variability is essential to enable the detector to learn traffic characteristics and not just a traffic pattern that is specific to a particular scene. The dataset consists of the recorded videos of the intersections.

The annotated 24,000 frames are split into 16,800 training frames (70%), 3,600 validation frames (15%), and 3,600 testing frames (15%). Instead of the adjacent frames being split randomly, it is done at the level of the recording session. The frames that come from the same video sequence will not be assigned to the training and testing sets at the same time. This avoids time leakage and a more realistic evaluation of model generalisation to unseen traffic observations.

The manually annotated frames are annotated by marking the bounding box on the visible vehicles of the 5 predefined classes. Each of the annotations includes a Vehicle Class, Bounding box coordinates, and a Frame Id. Each of the vehicles is correlated to each traffic approach or lane for subsequent traffic-state estimation, if required. Bounding-boxes have been drawn tightly to the visible size of each vehicle. If enough visual information is available to identify a partially occluded vehicle, the vehicle will be annotated as such. Images with significant obstruction, high blur, damaged areas or lack of information are omitted. After creating the data set, it is checked for wrong labels, wrong annotations, repeat frames and class imbalance before training the model. The preprocessing module transforms the obtained traffic frames into a standard format which can be used for deep learning-based vehicle detection system. This is in contrast to traditional methods that transform the images taken from the traffic into grayscale images for thresholding and background subtraction, which help identify the vehicle, but their colour, texture, shape and spatial information are not used.

The first step in the process of frame extraction from the video is to filter out those severely affected by motion blur, occlusion, excessive overexposure, underexposure, and the like that can interfere with vehicle identification. The images selected are resized to fit the input dimensions of the selected YOLOv8 detector and still preserve the aspect ratio of the images. Geometric distortion is minimized by the use of padding where needed. For each traffic approach a region of interest (ROI) is defined. The ROI covers the roadway region relevant to

traffic movement and excludes irrelevant regions such as buildings, sky, advertisements, and surrounding structures. ROI-based processing reduces unnecessary computational operations and limits false detections from non-road objects. Limited brightness and contrast adjustment may be applied where moderate illumination variations are present. Data augmentation is applied only to the training set and include scaling, cropping, brightness variation, contrast variation, and moderate geometric perturbation. No augmentation is applied to validation or testing images.

2.3 Deep-Learning-Based Vehicle Detection and Classification

A YOLO-based object-detection architecture is employed to detect and classify vehicles in each traffic frame. YOLOv8 is selected as the primary detection model because it provides a suitable balance between detection accuracy and computational efficiency for object-detection applications. Recent work has also demonstrated the applicability of YOLO-based models to traffic monitoring and intelligent transportation systems.

For an input frame I_t , the detector produces a set of vehicle detections represented by equation (1).

$$\mathcal{D}_t = \{(x_i, y_i, w_i, h_i, c_i, p_i)\}_{i=1}^{M_t} \quad (1)$$

where $\mathcal{D}_t$ represents the complete set of detections obtained from frame t ; M_t denotes the total number of detected vehicles in that frame; x_i and y_i denote the centre coordinates of the bounding box of vehicle i ; w_i and h_i represent the width and height of the bounding box, respectively; c_i denotes the predicted vehicle class; and p_i represents the confidence score associated with the detection. A confidence threshold is applied to eliminate low-confidence detections. Non-maximum suppression is subsequently used to remove duplicate bounding boxes corresponding to the same vehicle. The detector is trained using the annotated training dataset and optimized using the validation dataset. Its final performance is evaluated only on the independent testing dataset.

Each detected vehicle is assigned to one of the five predefined vehicle categories: car, two-wheeler, auto-rickshaw, bus and truck/heavy vehicle. Vehicle classification is important because different vehicle categories occupy different amounts of roadway space and contribute differently to traffic demand. For example, a bus or truck generally occupies substantially more roadway space than a two-wheeler. Therefore, vehicle class information is retained for subsequent traffic-state estimation rather than using only a total vehicle count. The classification performance is assessed using class-wise precision, recall and F1-score, together with overall mean average precision.

2.4 Multi-Object Vehicle Tracking and Lane Assignment

Frame-level vehicle detection alone is insufficient for estimating vehicle trajectories, waiting time and queue development. Therefore, multi-object tracking is incorporated into the proposed framework. Each tracked vehicle is assigned a unique identification number that remains associated with the vehicle while it remains within the observable region. The tracking stage provides the temporal information required to determine vehicle entry time, exit time, movement trajectory and stopping duration. It also prevents the same vehicle from being counted repeatedly across consecutive frames.

For each tracked vehicle k , its trajectory is represented by the sequence as shown in equation (2).

$$\mathcal{T}_k = \{(x_k^t, y_k^t)\}_{t=t_s}^{t_e} \quad (2)$$

where $\mathcal{T}_k$ denotes the trajectory of vehicle k ; x_k^t and y_k^t represent the vehicle position at time t ; t_s denotes the time at which the vehicle first enters the observation region; and t_e denotes the

time at which it leaves the observation region. The trajectory information is subsequently used for lane assignment, queue estimation and waiting-time calculation.

Following detection and tracking, each vehicle is assigned to the corresponding traffic approach or lane. The traffic state of every approach is then estimated using multiple complementary parameters rather than relying solely on vehicle count. The number of vehicles present in approach i at time t is represented as given by equation (3).

$$N_i(t) = \sum_{k=1}^{M_t} \mathbf{1}(k \in i) \quad (3)$$

where $N_i(t)$ denotes the number of vehicles detected in approach i at time t ; M_t represents the total number of tracked vehicles observed at time t ; and $\mathbf{1}(k \in i)$ is an indicator function that takes the value 1 when vehicle k belongs to approach i and 0 otherwise. Equation (3) therefore provides the instantaneous vehicle count for each traffic approach. The traffic density of each approach is estimated from the number of vehicles occupying the corresponding roadway region shown in equation (4)

$$D_i(t) = \frac{N_i(t)}{A_i} \quad (4)$$

where $D_i(t)$ denotes the estimated traffic density of approach i ; $N_i(t)$ represents the number of vehicles observed in that approach; and A_i denotes the effective roadway area included within the predefined region of interest for approach i . The density measure allows traffic conditions to be compared across approaches with different observable roadway areas.

Lane occupancy represents the proportion of the roadway region occupied by vehicles. It is calculated from the spatial extent of the detected vehicle bounding boxes within the predefined roadway region. The normalized occupancy of approach i is represented as equation (5).

$$O_i(t) = \frac{A_{V,i}(t)}{A_i} \quad (5)$$

where $O_i(t)$ denotes the occupancy of approach i ; $A_{V,i}(t)$ represents the combined image-plane area occupied by detected vehicles in that approach at time t ; and A_i represents the effective image-plane area of the roadway region associated with that approach. A higher value of $O_i(t)$ indicates that a greater proportion of the available roadway region is occupied by vehicles.

Queue length is estimated from the spatial positions of vehicles waiting upstream of the stop line. If $y_{SL,i}$ represents the image-coordinate position of the stop line for approach i , and $y_{Q,i}(t)$ represents the position of the rearmost vehicle in the queue, the estimated queue length in image coordinates is expressed as equation (6).

$$Q_i(t) = |y_{SL,i} - y_{Q,i}(t)| \quad (6)$$

where $Q_i(t)$ represents the estimated queue length for approach i ; $y_{SL,i}$ denotes the position of the stop line; and $y_{Q,i}(t)$ denotes the position of the rearmost queued vehicle. Where camera calibration and roadway geometry are available, the image-coordinate distance can be transformed into an approximate physical distance in metres.

The waiting time of a tracked vehicle is determined from its trajectory. A vehicle is considered to be waiting when its displacement or estimated speed remains below a predefined threshold. For vehicle k , the waiting duration is calculated as given by equation (7).

$$W_k = t_{r,k} - t_{s,k} \quad (7)$$

where W_k represents the waiting time of vehicle k ; $t_{s,k}$ denotes the beginning of the waiting interval; and $t_{r,k}$ represents the time at which the vehicle resumes movement above the

predefined movement threshold. The average waiting time for approach i is subsequently calculated from the waiting times of all vehicles associated with that approach. The traffic-state variables have different units and numerical ranges. Vehicle count, queue length, waiting time and occupancy therefore cannot be directly combined without normalization. For each traffic variable X , min-max normalization is applied as shown in equation (8).

$$\hat{X}_i(t) = \frac{X_i(t) - X_{\min}}{X_{\max} - X_{\min}} \quad (8)$$

where $\hat{X}_i(t)$ represents the normalized value of variable X for approach i ; $X_i(t)$ denotes the observed value at time t ; $X_{\min}$ denotes the minimum value of the corresponding variable in the calibration dataset; and $X_{\max}$ denotes its maximum value in the calibration dataset. The normalization process transforms the traffic-state variables to a common scale between 0 and 1, allowing them to be combined in the subsequent traffic-demand calculation. The normalized traffic-state variables are integrated to obtain a Traffic Demand Index (TDI) for each approach. The TDI represents the instantaneous traffic demand and serves as the principal input to the adaptive signal controller. The TDI for approach i is defined as equation (9).

$$TDI_i(t) = \alpha \hat{N}_i(t) + \beta \hat{Q}_i(t) + \gamma \hat{W}_i(t) + \delta \hat{O}_i(t) \quad (9)$$

where $TDI_i(t)$ denotes the traffic-demand index of approach i at time t ; $\hat{N}_i(t)$ denotes the normalized vehicle count; $\hat{Q}_i(t)$ denotes the normalized queue length; $\hat{W}_i(t)$ denotes the normalized waiting time; and $\hat{O}_i(t)$ denotes the normalized lane occupancy. The coefficients α , β , γ and δ represent the relative contribution of vehicle count, queue length, waiting time and occupancy, respectively. The coefficients satisfy equation (10).

$$\alpha + \beta + \gamma + \delta = 1 \quad (10)$$

where each coefficient is non-negative and represents the relative importance assigned to the corresponding traffic-state parameter. The weighting coefficients are determined through calibration using the training/development data rather than being arbitrarily assigned. This formulation allows the proposed controller to consider both the number of vehicles and the severity of congestion represented by queue formation and waiting time.

2.5 Adaptive Traffic Signal-Control

The calculated traffic-demand index is supplied to the adaptive signal controller. The controller allocates green time according to the relative traffic demand of the approaches while maintaining minimum and maximum permissible green durations. The initial demand-proportional green-time allocation for approach i is expressed as equation (11).

$$G_i(t) = G_{\min} + (G_{\max} - G_{\min}) \frac{TDI_i(t)}{\sum_{j=1}^n TDI_j(t)} \quad (11)$$

where $G_i(t)$ denotes the green duration allocated to approach i ; $G_{\min}$ denotes the minimum permissible green duration; $G_{\max}$ denotes the maximum permissible green duration; $TDI_i(t)$ denotes the traffic-demand index of approach i ; $TDI_j(t)$ represents the traffic-demand index of the j -th approach; and n denotes the total number of traffic approaches.

By ensuring the approach with the higher traffic demand is allocated a larger portion of green time, this condition in equation (11) will ensure that high traffic demand approaches receive more green time. Meanwhile, the minimum green-time requirement means that an approach cannot be barred altogether. The controller also adds additional "maximum waiting time" and "no starvation" constraints. Thus, when an approach has been in the red phase for a long enough period of time, it can gain priority in turn-by-turn traffic, in a graduated manner, while maintaining a low instantaneous traffic demand. So, the ultimate signal decision is based on the current traffic demand at the time of the signal change as well as the waiting conditions of in the vehicles.

The adaptive controller is constrained with a set of predetermined signal-safety constraints to ensure its practical operation. These are minimum green time, maximum green time, yellow interval and clearance interval. During every signal cycle the controller makes sure that the green time granted is within the limits as shown in equation (12).

$$G_{\min} \leq G_i(t) \leq G_{\max} \quad (12)$$

where $G_i(t)$ denotes the green duration assigned to approach i , $G_{\min}$ represents the minimum permissible green duration and $G_{\max}$ represents the maximum permissible green duration. The yellow and clearance intervals are maintained independently of the adaptive green-time calculation to ensure safe transition between successive signal phases.

In the fixed-time approach, each traffic approach is assigned a predetermined green duration independent of the current traffic demand. The same signal cycle is maintained under different traffic conditions. The conventional density-based approach determines signal timing primarily from the number of vehicles detected in each approach. This represents the fundamental principle of the earlier image-processing-based system, in which vehicle count was obtained through image processing and the green duration was increased when the estimated traffic density was higher.

The proposed method differs from the conventional density-based approach by incorporating deep-learning-based vehicle detection and tracking and by using multiple traffic-state parameters, namely vehicle count, queue length, waiting time and lane occupancy. These parameters are combined into the Traffic Demand Index and used for dynamic green-time allocation. The three approaches are evaluated under comparable traffic conditions to determine whether the proposed multi-parameter adaptive strategy can reduce traffic delay and queue formation while increasing intersection throughput.

2.6 Performance Evaluation

The proposed framework is evaluated at both the vehicle-detection level and the traffic-control level. The deep-learning detector is evaluated using precision, recall, F1-score, mAP@50 and mAP@50–95. These metrics measure the ability of the model to correctly identify vehicles, classify them into the appropriate categories and localize their bounding boxes. Automated vehicle counts are compared with manually verified counts. The Mean Absolute Error (MAE) is calculated as equation (13).

$$MAE = \frac{1}{N} \sum_{r=1}^N |C_r - \hat{C}_r| \quad (13)$$

where MAE denotes the mean absolute counting error; N represents the number of evaluated observations; C_r denotes the manually verified vehicle count for observation r ; and $\hat{C}_r$ denotes the vehicle count produced by the proposed system for the same observation.

A lower MAE indicates greater accuracy in vehicle counting.

The effectiveness of the signal-control system is evaluated using the following indicators: average vehicle waiting time, maximum vehicle waiting time, average queue length, maximum queue length, average vehicle delay, intersection throughput, number of vehicles served per signal cycle, signal utilization and vehicle-counting error. For each indicator, the proposed adaptive controller is compared with the fixed-time and conventional density-based approaches under equivalent traffic conditions. Lower waiting time, lower queue length and lower delay indicate improved traffic-control performance, whereas higher throughput and a greater number of vehicles served per cycle indicate more efficient use of the intersection.

The experimental observations are categorized according to traffic intensity, including low-, medium- and high-density conditions. The performance of the three signal-control approaches is subsequently compared under these different traffic states. Repeated observations are made at the intersections selected, on observation days and throughout the observation time periods for each of the performance indicators. For each control strategy, mean values are reported for the performance indicators.

The testing procedure is aimed at testing the central question of the study: whether a better estimation of the traffic-state via deep-learning based vehicle detection and tracking can lead to a more responsive signal-time allocation as compared to a fixed-time and conventional density-based signal-time allocation, which can reduce vehicle waiting time and queue length and improve intersection throughput. The proposed approach introduces a framework for the perception-to-control paradigm for adaptive traffic signal control. A large, diverse database of 24,000 frames of manually annotated videos from several signalized intersections is created from real traffic videos. The dataset is split into training, validation and testing sets, and the recordings are separated from each other to avoid temporal data leakage when training, testing and validation a vehicle-detection model with YOLO. Vehicles detected are classified and tracked from one frame to the next, allowing for estimation of the number of vehicles, vehicle density, lane occupancy, queue length and queue waiting time at each vehicle approach. The three signal control strategies were evaluated offline using the traffic states derived from the recorded video data. No live modification of the signal controllers at the study intersections was performed.

The estimated traffic state variables are normalized and then aggregated into a Traffic Demand Index, which represents a single unified traffic demand estimate from the real time traffic state variables. This index is used by the adaptive controller to dynamically distribute the green time without starving the approaches with less signal traffic volume and dynamically allocate green time while satisfying minimum and maximum signal timing constraints and preventing low-volume approaches from being starved. The obtained system is compared to fixed-time and traditional density-based signal-control strategies for the vehicle-detection performance assessment as well as for traffic-management performance evaluation based on waiting time, queue length, delay and intersection throughput. The methodology can thus be said to provide a full closed loop process in which traffic video is converted to vehicle-level information, vehicle level information is converted to traffic-state information, and traffic-state information is converted to adaptive signal-control decisions.

3. RESULTS AND DISCUSSION

The proposed vision-based adaptive traffic signal-control framework was tested with the traffic video recorded on three traffic signalized intersections namely Suchitra Junction, Crown Cafe Junction and KBR Park Junction in Hyderabad. The data collection took place during 7 days of consecutive running (from 8th February 2026 to 14th February 2026) for both working days and weekends. The seven days were chosen to allow to observe the transportation demand of a day-to-day nature, instead of just a single day.

Three one-hour observation sessions were considered on each day: 08:00–09:00 h, 13:00–14:00 h, and 17:00–18:00 h. The morning session is the first part of the morning commuter peak while the evening session is the evening return trip and mixed urban traffic demand. This afternoon session was added as an off-peak "reference" session as a way to test the adaptive controller's performance under relatively lower and more variable demand. Consistency is also

achieved when observation windows are kept the same over the seven-day period that the traffic is being observed, providing consistency when comparing traffic between days and intersections.

Therefore, each intersection supplied ~21 hours of traffic video (7 days x 3 observation periods per day). The entire data set thus amounted to ~ 63 hours of raw video. Sampling interval was defined in the methodology and representative frames were extracted, and the extracted frames were subjected to quality screening and annotation. Final set of data consisted of about 24,000 manually marked images which included cars, two-wheelers, auto rickshaws, buses and heavy vehicles. The selected observation schedule provides variation in traffic intensity while maintaining a consistent measurement protocol. The inclusion of Sunday, 8 February, through Saturday, 14 February, also enables the dataset to capture differences between weekend and working-day traffic patterns. This is particularly useful for evaluating an adaptive controller because its decision mechanism should respond to the observed traffic state rather than to a fixed assumption about the time of day.

3.1 Vehicle Detection and Classification Performance

The YOLO-based detector was evaluated using the held-out test set. Precision, recall, F1-score, mAP@50 and mAP@50–95 were used to assess detection performance. The class-wise results are presented in Table 1.

Table 1. Vehicle Detection and Classification Performance

Vehicle class	Precision (%)	Recall (%)	F1-score (%)	AP@50 (%)
Car	96.1	94.8	95.4	97.0
Two-wheeler	93.0	91.2	92.1	94.0
Auto-rickshaw	92.4	90.1	91.2	93.1
Bus	96.5	95.2	95.8	97.4
Truck/heavy vehicle	94.0	92.0	93.0	95.1
Overall	94.1	92.7	93.4	95.2

The detector achieved an overall precision of 94.1%, recall of 92.7% and F1-score of 93.4%, with an mAP@50 of 95.2%. The mAP@50–95 was 84.6%. The relatively high performance across the five vehicle classes indicates that the detector was able to distinguish the principal vehicle categories present in heterogeneous urban traffic. The highest AP@50 was obtained for buses, followed by cars. These vehicles generally occupy larger image regions, making their visual features easier to distinguish. The comparatively lower recall for two-wheelers and auto-rickshaws is associated with their smaller image size, frequent clustering and partial occlusion by neighbouring vehicles. Such conditions are particularly relevant to mixed Indian traffic, where two-wheelers frequently move between larger vehicles and may not maintain clearly separated lanes.

Detection accuracy was further examined at the vehicle-counting level because accurate counting is essential for the subsequent calculation of traffic density, occupancy, queue length and the Traffic Demand Index as shown in Figure 3. The overall mean absolute error was 3.6 vehicles, corresponding to a mean absolute percentage error of approximately 4.5%. The error remained relatively stable across the three traffic conditions. Although the absolute error increased under high-density conditions, the corresponding percentage error remained below 5%. This indicates that the combination of deep-learning-based detection and multi-object tracking provided sufficiently stable vehicle counts for subsequent traffic-state estimation. The higher absolute error under high-density conditions can be attributed primarily to vehicle

overlap and partial occlusion. The tracking stage reduces the possibility of counting the same vehicle repeatedly across successive frames and is therefore important for obtaining temporally consistent vehicle counts.

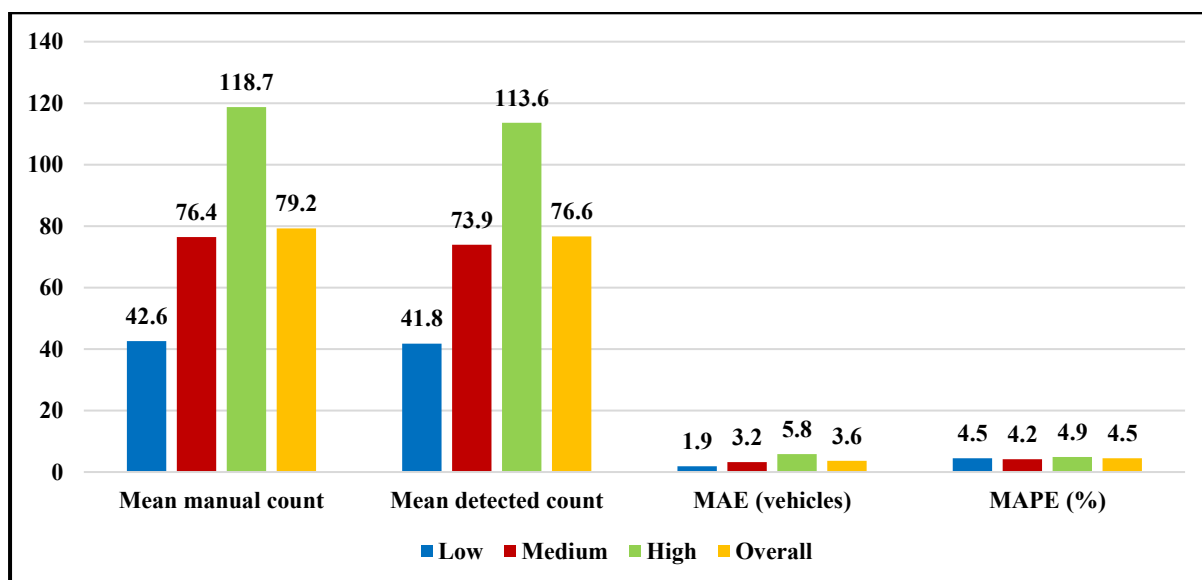


Fig. 3. Vehicle Counting Performance under different Traffic Conditions

3.2 Traffic-State Estimation

The detected and tracked vehicles were used to estimate the principal traffic-state variables for each approach. These included the number of vehicles, lane occupancy, queue length and waiting time. The normalized parameters were subsequently combined to obtain the Traffic Demand Index (TDI) as indicated in Table 2.

Table 2. Traffic-state characteristics under different traffic conditions

Traffic condition	Average vehicles/approach	Occupancy (%)	Queue length (m)	Waiting time (s)	TDI
Low	39	22	24	31	0.29
Medium	72	43	49	67	0.52
High	113	68	78	108	0.78

A progressive increase is observed in all traffic-state indicators as traffic intensity changes from low to high. The average number of vehicles increases from 39 to 113 vehicles per approach, while occupancy rises from 22% to 68%. Similarly, queue length increases from 24 m to 78 m, and average waiting time increases from 31 s to 108 s. The corresponding TDI increases from 0.29 under low traffic conditions to 0.78 under high traffic conditions. This indicates that the TDI provides a more comprehensive representation of traffic demand than vehicle count alone. For example, two approaches may contain a similar number of vehicles but differ considerably in queue length or waiting time. Incorporating multiple traffic-state variables allows the controller to distinguish such conditions and allocate green time according to the overall severity of demand. The three traffic states were subsequently used to evaluate whether the adaptive controller could maintain its performance under different operating conditions.

3.3 Comparative Signal-Control Performance

The proposed adaptive controller was compared with two baseline strategies: fixed-time control and the conventional density-based controller. The same traffic observations were considered for the three control strategies so that differences in performance could be attributed to the signal-control mechanism rather than differences in traffic input. Figures 4 and 5 show the comparative performance of the signal control strategies. The proposed adaptive controller produced the lowest average waiting time, average queue length and average delay among the three strategies. Average waiting time decreased from 96.4 s under fixed-time control to 62.7 s with the proposed controller. This corresponds to a reduction of approximately 35% relative to fixed-time control. The average queue length similarly decreased from 71.5 m to 45.2 m, representing a reduction of approximately 37%.

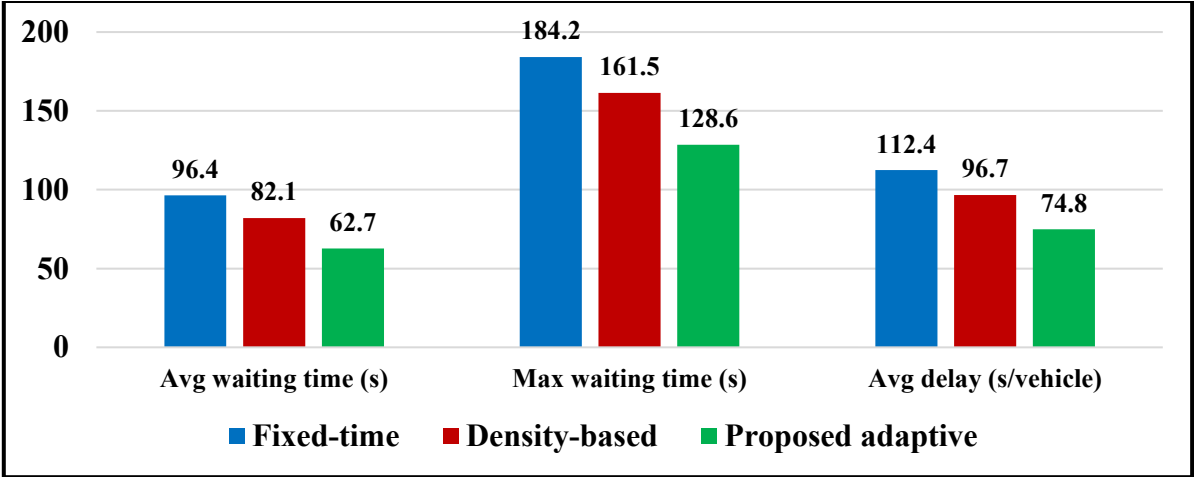


Fig. 4. Comparative performance of Waiting times and Average delay

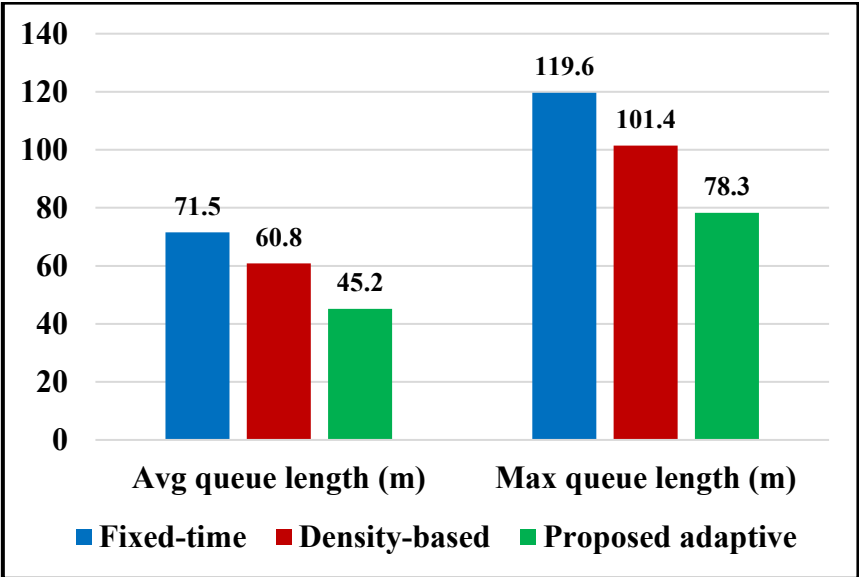


Fig. 5. Comparative performance of Average length and Maximum Queue Lengths

Table. 3. Comparative performance of Throughput, Vehicles served/cycle, Signal utilization and Counting error %

Performance Indicator	Fixed-Time	Density-based	Proposed Adaptive
Throughput (vehicles/hour)	2,486	2,648	2,832
Vehicles served/cycle	78.4	83.7	89.6
Signal utilization (%)	72.1	76.4	82.8
Counting error (%)	11.8	8.1	4.5

This improvement is a result of multi-state traffic control decision based on multiple traffic-state parameters. The fixed-time strategy is not sensitive to fluctuations in the traffic demand at any one time, and the density-based strategy is not sensitive to velocity changes but is primarily sensitive to the number of vehicles. The proposed controller, on the other hand, uses vehicle count, occupancy, queue length and waiting time with the help of the TDI. As a result, an approach with a high instantaneous vehicle count may be given extra green time even if the instantaneous vehicle count is not a complete measure of congestion. Table 3 shows the comparative performance of Throughput, Vehicles served/cycle, Signal utilization and Counting error %. There was a slight rise in the average number of vehicles on each cycle from 78.4 to 89.6, and in the percentage of time that the signal is used from 72.1% to 82.8%. The results show that the adaptive controller could utilize the green time more efficiently and achieve the required signal timing constraints. The throughput also improved from 2,486 vehicles/h at the fixed-time control to 2,832 vehicles/h at the proposed adaptive control. This is a rise of about 13.9%. The proposed adaptive control shows a significant reduction in the counting error percentage from 11.8% for fixed-time technique to 8.1% for the density-based technique to 4.5%.

One of the key characteristics of the proposed system is the continuous exchange of information between the traffic-state estimation and signal control. If the TDI for one approach rises with respect to other approaches that compete for it, the controller increases the amount of green time that it grants the approach while enforcing the minimum and maximum green-time limits. If demand drops, the green allocation can be reduced and allocated to a more demanding approach, on the other hand. This behaviour is important because traffic demand does not remain constant throughout the day. The morning, afternoon and evening observation periods contain substantially different traffic states, while additional variation occurs between the seven observation days. A fixed-time controller cannot respond to these short-term changes because its phase durations remain predetermined. The proposed closed-loop framework instead updates the control decision using the most recently estimated traffic state.

The starvation-prevention constraint also ensures that an approach with temporarily low demand continues to receive service. Thus, the adaptive mechanism does not simply maximize green time for the most congested approach; it balances demand responsiveness with minimum service requirements for competing approaches. The concept of the proposed framework is a non-intrusive way to get information about traffic state with video cameras. Since the detection, tracking and traffic-state estimation is based on the video information, this method can be combined with existing camera systems with minimal installation of new road sensors.

However, it was conducted at three intersections in Hyderabad and for a 7-day span. Detection and tracking performance can be affected by the camera viewpoint, illumination, occlusion by vehicles, and the geometric and traffic composition of intersections. The data set is also limited in time, and does not include all environmental and seasonal conditions. Thus, additional

testing over extended durations and more intersections would need to be done before being deployed on a wider scale.

4. CONCLUSION AND FUTURE SCOPE

The proposed vision-based adaptive traffic signal-control framework combined the vehicle detection, multi-object tracking, traffic-state estimation, and green-time allocation based on Traffic Demand Index (TDI) with the YOLO algorithm. The detector was evaluated on the traffic data collected from three intersections in Hyderabad over 7 days, and its performance has been reported as 94.1% Precision, 92.7% Recall, 93.4% F1-score and 95.2% mAP@50 with an overall MAPE of around 4.5%. The proposed method compared to fixed-time control resulted in a reduction of the average waiting time from 96.4 s to 62.7 s, a reduction in the queue length from 71.5 m to 45.2 m and an increase of the throughput from 2,486 vehicles/h to 2,832 vehicles/h.

Future research efforts include testing over extended time periods and more intersections under different geometric, traffic and weather conditions. The identification of partially covered two-wheelers and auto-rickshaws, integration of pedestrian and priority vehicle information, and coordinated control of multiple intersections will also be discussed. The computational latency, communication needs, and scalability of real-time deployment can also be explored, as well as compared to more sophisticated traffic-control strategies.

REFERENCES

- [1] "<https://www.tomtom.com/>," TomTom International BV, 2025. [Online]. Available: <https://www.tomtom.com/traffic-index/city/hyderabad/>. [Accessed 20 May 2026].
- [2] Jithender Jathoth, Arpan Mehar, "Estimation of Queue Length at Signalized Intersection Under Non-Lane Based Traffic Conditions," *Periodica Polytechnica Transportation Engineering*, vol. 51, no. 1, pp. 31-39, 2023.
- [3] Mohammed Ibrahim, Karanam Madhu Kumar, Srinivasa Kumar R., "Saturation Flow Concept at Signalized Intersection for Non Lane Based Traffic," *International Journal of Innovative Research in Science Engineering and Technology*, vol. 6, no. 1, pp. 801-808, 2017.
- [4] Filia Sologia, Yuli Adam Prasetyo, "Adaptive traffic control systems in intelligent transportation systems: a systematic literature review," *International Journal of Advanced Technology and Engineering Exploration*, vol. 13, no. 140, pp. 210-235, 2026.
- [5] Marwa Elharoun, Sherif M. El-Badawy, Elsayed Abd-Elazem Shwaly, Usama Elrawy Shahdah, "Adaptive traffic signal control using deep reinforcement learning: A multi-objective approach for single and multi-intersection scenarios," *International Association of Traffic and Safety Sciences Research*, vol. 49, no. 4, pp. 481-492, 2025.
- [6] Tarek Amine Haddad, Djalal Hedjazi, Sofiane Aouag , "An IoT-Based Adaptive Traffic Light Control Algorithm for Isolated Intersection," in *International Conference on Computing Systems and Applications*, Algiers, Algeria, 2021.
- [7] Meng Long, Xiexin Zou, Yue Zhou, Edward Chung, "Deep reinforcement learning for transit signal priority in a connected environment," *Transportation Research Part C: Emerging Technologies*, vol. 142, no. 103814, 2022.

- [8] Mobin Yazdani, Majid Sarvi, Saeed Asadi Bagloee, Neema Nassir, Jeff Price, Hossein Parineh, "Intelligent vehicle pedestrian light (IVPL): A deep reinforcement learning approach for traffic signal control," *Transportation Research Part C: Emerging Technologies*, vol. 149, no. 103991, 2023.
- [9] Junxiu Liu, Sheng Qin, Min Su, Yuling Luo, Shunsheng Zhang, Yanhu Wang, Su Yang,, "Traffic signal control using reinforcement learning based on the teacher-student framework," *Expert Systems with Applications*, vol. 228, no. 120458, 2023.
- [10] Gongquan Zhang, Fangrong Chang, Jieling Jin, Fan Yang, Helai Huang, "Multi-objective deep reinforcement learning approach for adaptive traffic signal control system with concurrent optimization of safety, efficiency, and decarbonization at intersections," *Accident Analysis & Prevention*, vol. 199, no. 107451, 2024.
- [11] Changjian Cai, Min Wei , "Adaptive urban traffic signal control based on enhanced deep reinforcement learning," *Scientific Reports*, vol. 14, no. 14116, pp. 1-11, 2024.
- [12] Wubing Chen , Shangdong Yang , Wenbin Li , Yujing Hu, Xiao Liu , Yang Gao, "Learning Multi-Intersection Traffic Signal Control via Coevolutionary Multi-Agent Reinforcement Learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 11, pp. 15947-15963, 2024.
- [13] Kai Hu, Mingyang Li, Zhiqiang Song, Keer Xu, Qingfeng Xia, Ning Sun, Peng Zhou, Min Xia, "A review of research on reinforcement learning algorithms for multi-agents," *Neurocomputing*, vol. 599, no. 128068, 2024.
- [14] Karima Benhamza, Hamid Seridi, Meriem Agguini, Amel Bentagine , "A multi-agent reinforcement learning based approach for intelligent traffic signal control," *Evolving Systems*, vol. 15, no. 1, p. 2383–2397, 2024.
- [15] Sudipta Das, Sujata Chakravarty, Satyabrata Dash, Rabindranath Nanda, "Enhancing Traffic Monitoring Systems through Image Analysis and Rule-Based Machine Learning," *Procedia Computer Science*, vol. 258, pp. 2119-2129, 2025.
- [16] Kusworo Adi, A P Widodo, Catur Edi Widodo, A Pamungkas, "Automatic vehicle counting using background subtraction method on gray scale images and morphology operation," *Journal of Physics Conference Series* , vol. 1025, no. 1, pp. 1-6, 2018.
- [17] Ays,e Aybilge Murat, Mustafa Servet Kiran, "A comprehensive review on YOLO versions for object detection," *Engineering Science and Technology, an International Journal*, vol. 70, no. 102161, pp. 1-19, 2025.
- [18] Xuefang Lv, Tianyu Chen, Caixia Song, Chaoran Yang, Tao Ping, "Application of YOLO Algorithm for Intelligent Transportation Systems: A Survey and New Perspectives," *International Journal of Distributed Sensor Networks*, vol. 2025, no. 2859040, pp. 1-33, 2025.
- [19] M. Bakirci, "Utilizing YOLOv8 for enhanced traffic monitoring in intelligent transportation systems (ITS) applications," *Digital Signal Processing*, vol. 152, no. 104594, 2024.
- [20] Dang Viet Anh Nguyen , Carlos Lima Azevedo, Tomer Toledo, Filipe Rodrigues, "A framework for benchmarking traffic signal control robustness under incidents: comparative study of reinforcement learning-based methods," *European Transport Research Review* , vol. 85, pp. 1-24, 2026.

- [21] S. Omefe, "Assessment of the effectiveness of adaptive traffic signal control systems in reducing pedestrian-vehicle conflicts at high-risk crossings," *GSC Advanced Research and Reviews* , vol. 24, no. 1, pp. 140-153, 2025.
- [22] Nikhil Nigam, Dharendra Pratap Singh, Jayatrilok Choudhary, Surendra Solanki, "An Efficient Model for Real-Time Traffic Density Analysis and Management Using Visual Graph Networks," *IEEE Access*, vol. 13, pp. 140413-140439, 2025.