

Efficient Residual YOLOv12-L Atrous Squeeze LightGBM Optimized Fuzzy CatBoost Network for Vehicle Identification in Traffic Environments

R Kiranmai

kiranmair352@gmail.com

SRM Institute of Science and Technology Ramapuram

R Deeptha

SRM Institute of Science and Technology Ramapuram

Article

Keywords: Vehicle detection, YOLOv12-L, TDA-LightGBM, CatBoost, Convolutional Block Attention Module, Gradient Optimized Fuzzy Classifier and Atrous Convolution

Posted Date: September 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10342681/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Efficient Residual YOLOv12-L Atrous Squeeze LightGBM Optimized Fuzzy CatBoost Network for Vehicle Identification in Traffic Environments

***¹Kiranmai R**

¹kiranmair352@gmail.com

Department of Computer Science and Engineering

SRM Institute of Science and Technology

Ramapuram, Chennai, Tamil Nadu 600089

²Deeptha R

²deepthar66@outlook.com

Department of Information Technology

SRM Institute of Science and Technology

Ramapuram, Chennai, Tamil Nadu 600089

Abstract: The detection of vehicles, pedestrians, and cyclists in traffic environments plays a key role in traffic surveillance, however, existing methods have limitations in segmenting distant objects and detecting high illumination and fragmented background. To address this issue, a Residual YOLOv12-L Atrous Squeeze LightGBM Optimized Fuzzy CatBoost Network (RYAS-LOFCN) is proposed, which provides accurate segmentation and detection of vehicles. In traditional models, different pixel density levels and different structural clarity levels between its foreground and its background, caused by perspective compression effect, makes it difficult to accurately discriminate the features of the distant objects. To address this issue, a novel Residual Atrous Squeeze Attention Network with Feature Pyramid Network (RASAN-FPN) is introduced, which collectively addresses the perspective compression problem by restructuring how spatial and contextual information is captured and emphasized across scales for the traffic objects. Moreover, existing models are not able to handle mask coalescence, scale imbalance, and inter-class ambiguity in dense traffic scenes, leading to poor classification of visually similar objects, due to anchor assignment mechanism. To address this issue, a novel YOLOv12-L Based TDA-LightGBM Gradient-Optimized Fuzzy CatBoost Network (YTG-FCN) is introduced, which

allows effective multi-scale feature extraction of segmentation mask inputs, simultaneously capturing the structural features of all three types of traffic objects. The result obtained from the simulated results showed that the model has an increased accuracy of 98.63% and precision of 98.56%, which is better than the existing models.

Keywords: *Vehicle detection, YOLOv12-L, TDA-LightGBM, CatBoost, Convolutional Block Attention Module, Gradient Optimized Fuzzy Classifier and Atrous Convolution.*

1. Introduction

Identifying vehicle in traffic environments from input videos is essential for traffic monitoring, accident prevention and other applications as well. Vehicle detection is the process of detecting and locating parts of the road traffic from surveillance camera images, road sensors or unmanned surveillance imaging system. Traffic scenes often present dramatic illumination changes, occlusions, motion blur, scale variation, dense traffic and perspective distortion in real-world traffic conditions, making the detection process quite difficult. In highway or urban traffic, especially, the distant objects seem compressed in space and structurally weak in the images and the closer objects occupy more image space with better feature representation. Such variations have a tremendous impact on the ability of computer vision systems to effectively extract foreground traffic objects from highly complex road backgrounds to enable practical and reliable vehicle identification, which remains a challenging research subfield in intelligent traffic analysis and control [1-4].

In the segmentation process, objects such as vehicles, pedestrians, and cyclists are separated from the surrounding road background at the pixel level to generate structurally meaningful regions. Effective segmentation highly influences the reliability of the subsequent detection and classification stages. Good segmentation allows for correct boundary extraction, consistent shape for objects and minimises unwanted background regions interfering with vehicle identification. In conventional approaches, threshold-based segmentation and background subtraction methods were commonly utilized to distinguish moving vehicles and pedestrians. These methods are efficient and easy to implement in a static environment with traffic, but are not suitable when traffic changes, shadows interfere, or when there is changing light intensity in the background, resulting in incomplete foreground extraction. Similarly, edge detection and contour-based segmentation

approaches were employed to identify object boundaries by extracting structural edge information from traffic scenes. In addition, watershed segmentation and region growing techniques were introduced to improve object separation in dense traffic environments by grouping neighboring pixels into meaningful regions. However, these approaches often suffer from over-segmentation and mask coalescence when adjacent vehicles, pedestrians, and cyclists appear spatially close to each other in congested traffic scenes [5-8].

In the detection stage, segmented traffic objects, including vehicles, pedestrians and cyclists, are recognized, located and labelled via bounding boxes and category labels in the detection phase. Effective detection provides accurate localization and classification of traffic participants for traffic monitoring. Haar Cascade classifiers and template matching methods were widely used in traditional methods to detect vehicles and pedestrians by learning predefined structural appearance patterns from training images. These methods provide faster detection performance in simple traffic environments, but they struggle to handle large appearance variations, pose changes, and partial occlusions in real-world traffic conditions. Similarly, for pedestrian and vehicle detection, Histogram of Oriented Gradients (HOG) is extensively used with Support Vector Machines (SVM) which are classifiers that are extracted based on shape descriptors obtained from gradients. Although HOG-SVM improves object representation and classification accuracy, it performs poorly for small-scale and distant traffic objects due to weak feature visibility and insufficient spatial information [9-11].

In traffic analysis systems several deep learning/ intelligent methods are proposed for enhancing its segmentation and detection performance. To tackle the issue of multi-scale feature extraction and separation of foreground markers in intricate traffic scenarios, CNN-based segmentation techniques like U-Net, Mask R-CNN, and Feature Pyramid Networks (FPN) were used. But they suffer from feature suppression in far-field traffic regions and difficulty preserving precise boundaries for small-scale pedestrians and cyclists. Similarly, object detection frameworks such as YOLO, SSD, Faster R-CNN, and EfficientDet were introduced to achieve real-time object localization and classification; however, these models often face reduced detection sensitivity for small objects and localization degradation in crowded traffic scenes. In addition, ensemble learning and fuzzy logic-based classifiers were integrated with deep learning frameworks to improve classification reliability and manage structural ambiguity in traffic objects. Nevertheless, these hybrid methods increase computational complexity, require extensive parameter optimization and

experience performance degradation during structurally inconsistent caused by dense traffic, illumination variation and overlapping object boundaries [12-15]. These challenges in existing techniques highlight the need for a novel technique with effective segmentation and detection for vehicle identification in traffic environments. This shows that an intelligent deep learning method is needed to effectively identify distant and small objects, while accurately detecting vehicles in high illumination and a fragmented background.

1.1 Contributions of the Proposed Methodology

The contributions of the proposed Residual YOLOv12-L Atrous Squeeze LightGBM Optimized Fuzzy CatBoost Network in effectively classification and detection of vehicle in traffic environments are as follows:

- To address the limitations of capturing distant and small objects in a traffic environment, a novel RASAN-FPN is introduced, which captures the features of distant and small objects, preserves hierarchical features, suppressed background noise and refines feature localization.
- To overcome the limitations in handling mask coalescence, scale imbalance, and inter-class ambiguity in dense traffic, a novel YTG-FCN is introduced, which deals with residual boundary errors, handles ambiguity in similar objects, and counters the light variation-induced mask fragmentation.

The organization of the paper is as follows: Section 2 presents the comprehensive review of the literature survey; Section 3 presents the overall methodology of the proposed work; Section 4 presents the results and discussion part of the proposed method, along with comparative analysis; finally, Section 5 concludes the paper.

2. Literature Survey

Tallat *et al.* [16] presented an improved YOLO-based deep learning model for dynamic traffic flow optimization and real-time vehicle recognition in smart traffic management systems. The model integrated the AI and computer vision techniques to asses the real-time traffic surveillance videos, accurate detect vehicle types, and compute traffic density. The model used YOLOv11 to process video in real-time to identify objects such as obstructions, cars and pedestrians in the traffic. Following that, the prediction was made using the intelligent system's predictive model that was based on AI. According to the results obtained from the model, the computer vision along with the

AI techniques properly enhanced the accuracy rate of detection of the system. The model, however, was built on a limited scope of data and as such has decreased applicability in real-world situations. Wu *et al.* [17] optimized segmentation while maintaining computational efficiency by integrating a multi-stage pipeline encompassing contrast enhancement and region-based clustering. An RNN optimized via a hybrid loss function combining cross-entropy and mean squared error was used for vehicle detection. This approach improved the localization and confidence estimation of the model during vehicle estimation. Then, the model utilized SURF and BRISK descriptors integrated with a Swin Transformer backbone for classification and a multi-object tracking stream employing DeepSORT for trajectory association. Yet, the model was not adaptable to fast dynamic changing environments.

Hanzla *et al.* [18] introduced a pioneering approach for robust vehicle analysis using UAV-captured images, incorporating a novel six-stage framework. In the beginning, the model applied a GAN-based Super-Resolution (ESRGAN) to enhance the quality of the image to address the interference of noise in the image. Then, a U-Net-based architecture was used for the segmentation of the image to ensure precise extraction of the features from complex background. After that, a YOLOv10 model was utilized for image detection and finally, a hybrid of CNN, HOG, and Wavelet Transform was utilized for feature extraction. A particle Swarm Optimization technique was also used for optimizing the extracted features from the process. Still, the model faced significant limitations during detection in dense traffic conditions.

Li *et al.* [19] combined deep learning models with metric learning models and provided a multi-stage traffic flow model. The model utilized language-prompting to automate segmentation by leveraging the Segment Anything Model for vehicle detection. This improved the adaptability of the model to complex and heavy traffic scenarios. The model also employed StrongSORT algorithm for vehicle tracking, which was integrated with mask-based tracking to enhance the recognition of the vehicle in congested conditions. Moreover, the PP-OCR module accurately extracted timestamps to support interval speed measurement across multiple viewpoints. Nevertheless, the model had limited scalability in regions with different vehicle distributions and high traffic conditions.

Medina *et al.* [20] introduced a novel method that integrated YOLO for vehicle detection, Voronoi analysis for spatial segmentation, and DBSCAN for clustering. The model effectively identified and distinguished the intrinsic characteristics through drone-captured video signals. In this, the

challenges such as missed detections and high-speed anomalies caused by off-center bounding boxes were addressed by a custom tracking solution. The introduced model also conducted a vehicle traffic inventory of the interchange. This comprehensive strategy proved helpful both in more accurately estimating traffic and giving insight for urban planners and traffic managers. Furthermore, the model did not take into account pedestrian and cyclist movement and environmental factors.

Li *et al.* [21] presented a surveillance video-assisted federated digital twin (SV-FDT) framework to empower an intelligent transport system with pedestrian and vehicle-in-the-loop. A comprehensive pedestrian-vehicle interaction was built by SVFTD by leveraging multi-source traffic surveillance videos. It had three layers, where the first layer collected traffic surveillance videos from multiple sources. The edge layer provided segmentation-based visual understanding and twin-agent-based interaction modelling. Finally, the cloud layer integrated LDTs across different regions to construct a global DT model in real-time. However, the model had limited reliability under different weather conditions and dense traffic regions.

Yang *et al.* [22] developed a novel approach for identifying vehicle types by leveraging spatial-semantic information from trajectory data. The introduced model initially detected staying points from trajectory data, then a DBSCAN clustering was utilized to adaptively segment original vehicle trajectories into various microtrips in a steady-moving pattern. Finally, the calculated movement and geo-semantic features were utilized for two classification tasks using three typical classification models. The results of the model showed that the introduced framework achieved reliable performance in identifying various vehicle types. Yet, the model did not differentiate vehicles from pedestrians.

Sukanto *et al.* [23] presented a machine learning-based approach to automatically detect auto-rickshaws in traffic images. The system used YOLOv8 model for real-time model detection to detect the vehicle in the traffic images. A real-time inference pipeline using a trained YOLOv8 model was utilized in this study for the detection of the vehicle. A set of 1,730 annotated images that were captured under various traffic conditions was used in this model for training the framework for effective detection. The result obtained from the model showed that the framework provided reliable detection in real-time detection of the vehicle under traffic scenarios. Still, the model was limited to a single class of vehicle.

Onsu *et al.* [24] leveraged the LLaVA visual grounding multimodal large language model (LLM) for traffic monitoring tasks on the real-time Quanser Interactive Lab simulation platform, covering scenarios like intersections, congestion, and collisions. The multimodal LLM model, LLaVA, an advanced framework integrating visual and textual inputs, was fine-tuned for real-time traffic monitoring system. Then, a real-time segmentation model using YOLOv11 was utilized for segmentation of the image. The segmentation model identified key elements in the input image from a pre-installed camera, such as vehicles and pedestrians. Moreover, the model had limited performance under large-scale traffic scenarios with increases number of vehicles.

Liu *et al.* [25] developed a novel method based on YOLOv8n for efficient traffic recognition and tracking systems. Initially, the model developed a SCC_Detect based on SCConv to reduce the computation of redundant features in attention heads. Then, a dual convolutional kernel was used in the place of convolutional kernel to construct a lightweight deep neural network. At the same time, a loss function based on Focaler-EIoU is introduced to improve the accuracy of the model in detection. The embedded BotSORT tracker achieved better results in accurate and stable recognition of vehicles in traffic scenarios. The introduced model had a high practical value and application prospects in traffic detection and deployment. Nevertheless, the model was not able to perform well in highly dense weather conditions and complex urban scenarios.

From the survey, it is clear that the model in [16] had limited generalizability in real-world scenarios, [17] was not adaptable to dynamic changing environments, and [18] faced challenges during detection in dense traffic conditions. Moreover, [19] had limited scalability in regions with different vehicle distributions, [20] did not cover factors such as pedestrian and cyclist movement, [21] was not reliable under different weather conditions and [22] was not able to differentiate pedestrians from vehicles. Additionally, [23] was limited to a single class of data during detection, [24] had limited performance under large-scale traffic scenarios and [25] had limited performance in dense weather conditions. These challenges in existing techniques necessitate the need for a new intelligent approach with effective classification and detection techniques for vehicle detection in traffic environments. Table 1 shows the summary of the literature survey, demonstrating the significance and limitations of existing models.

Table 1: Novelty Comparison Table

Method	Significance	Limitation
--------	--------------	------------

STMS [16]	Used YOLOv11 to process video in real-time to identify objects such as obstructions, cars and pedestrians in the traffic.	The model had limited generalizability in real-world scenarios.
SASSC-RNN [17]	An RNN optimized via a hybrid loss function combining cross-entropy and mean squared error was used for vehicle detection.	The model was not adaptable to fast dynamic changing environments
ESRGAN-YOLOv10 [18]	Utilized ESRGAN to enhance the quality of image, U-Net for segmentation and YOLOv10 for image detection.	The model was not suitable for dense traffic conditions.
SAM [19]	Utilized language-prompting to automate segmentation by leveraging the Segment Anything Model for vehicle detection.	The model had limited scalability in regions with different vehicle distributions.
YOLO-DBSCAN [20]	Integrated YOLO for vehicle detection, Voronoi analysis for spatial segmentation, and DBSCAN for clustering	The model model did not cover factors such as pedestrian and cyclist movement.
SV-FDT [21]	Leveraged multi-source traffic surveillance videos for pedestrian-vehicle detection.	The model had limited reliability under different weather conditions.
DBSCAN [22]	Utilized DBSCAN to adaptively segment original vehicle trajectories into various microtrips.	The model was not able to differentiate vehicles from pedestrians.
YOLOv8 [23]	Utilized YOLOv8 model for real-time model detection.	The model was limited to a single class of vehicle.
LLaVa-YOLOv11 [24]	YOLOv11 was utilized for segmentation of the image.	The model had limited performance under large-scale traffic scenarios.
YOLOv8-SCConv [25]	Developed a SCC_Detect to reduce the computation of redundant features in attention heads.	The model was not able to perform well in complex urban scenarios.
Proposed Model	Utilized RASAN-FPN for accurately segmenting distant objects and a YOLO based architecture for accurate detection under complex traffic scenarios.	The performance of the model was highly affected by adverse weather conditions.

Table 1 shows the novel comparison between the proposed model and existing models, showing the significance and limitations of the models. From this table, it is clear that the existing models

face a lot of issues in segmentation and detection of vehicle in traffic environments. This shows the need for the development of a new intelligent with effective segmentation and detection techniques.

2.1 Motivation of the Research

The detection of vehicle, pedestrians, and cyclist in heavy traffic scenarios is one of the important things for traffic control in real-time, which helps in traffic surveillance, autonomous driving, highway surveillance, accident prevention and urban traffic management. Several deep learning and machine learning-based techniques are introduced for reliable vehicle detection; however, the models face significant issues in classification and detection, making them unsuitable in real-world scenarios. These limitations of the existing techniques served as a motivation for developing an efficient framework for vehicle detection in traffic environments. The problems and limitations of the existing techniques in segmentation and detection that served as a motivation for developing the proposed model are as follows:

2.1.1. Problems in Segmentation

Real-world traffic scenes exhibit significant variability in the appearance of vehicles, pedestrians, and cyclists in terms of perspective, where distant objects are found to have a very compressed appearance with minimal spatial information, whereas nearby vehicles have more structural information. This causes a *perspective compression effect*, where the feature normalization in the convolutional hierarchy suppresses weak feature activations in distant vehicles due to depth information, causing structural information in these distant objects to be dominated by surrounding contextual information. This produces an extreme unbalanced condition which results in different pixel density levels and different structural clarity levels between its foreground and its background section. In such high-density images, distant objects are found to have spatial heterogeneity, where the boundaries of these distant vehicles, pedestrians, and cyclists tend to fuse with the road due to a lack of scale representation, a phenomenon known as boundary vanishing. This causes coalescence in the mask, where the centroid of the vehicle is unstable due to feature responses from different spatially proximate instances collapsing onto a single feature response manifold. Moreover, due to sub-pixel changes in the boundaries of the vehicle, there is a non-linear shift in the geometric center of the vehicle, even in cases of stable motion. The generated

segmentation masks become incomplete because they exhibit two types of diffusion and one type of spatial misalignment. Thus, due to the *spatial-depth conflict*, the model fails to accurately segment the individual vehicle, pedestrians, and cyclists boundaries in the compressed region of the image, since the discriminative power for vehicle boundaries is found to be degraded in spatially compressed far-field regions.

2.1.2. Problems in the Detection of Vehicles, Pedestrians, and Cyclists

After segmentation, the segmented image becomes the input of the detection stage, the error residual boundaries and mask coalescence that are produced by the segmentation stage directly degrade the accuracy of detection. During dense traffic scenarios, masks of adjacent objects cause the detection model to either not detect an object or falsely enclose multiple objects in a single bounding box proposal. The large-scale difference between object classes in the same scene, with a large vehicle taking up most of the mask space and a distant pedestrian or cyclist taking up just a few pixels, leads to an anchor assignment mechanism that is biased in favor of large-scale objects, depriving the detector of enough gradient updates in small-scale classes. This systematic training imbalance makes a detection bias of pedestrians and cyclists chronically under-detected despite their correctly generated segmentation mask. Moreover, ambiguity in the inter-class boundary between visually similar categories, like pedestrians and cyclists, coupled with their non-rigid and highly variable appearance structures, results in common inter-class confusion in the detection head. Mask relics due to illumination, including fragmented foreground areas as a result of shadow and light contrast in various traffic environments, further lowers the structural consistency of the mask input to the extent that consistent object localization and classification cannot be achieved in all three categories of traffic objects.

3. Proposed Methodology

In order to overcome the challenges associated with segmentation and detection of vehicle pedestrians, and cyclists on traffic environments, a novel “**Residual YOLOv12-L Atrous Squeeze LightGBM Optimized Fuzzy CatBoost Network (RYAS-LOFCN)**” is proposed. In the segmentation stage, the backbone architecture is the “*Residual Atrous Squeeze Attention Network with Feature Pyramid Network (RASAN-FPN)*”, which incorporates Atrous convolution, Squeeze-and-Excitation (SE), and a Convolutional Block Attention Module (CBAM), which collectively

address the perspective compression problem by restructuring how spatial and contextual information is captured and emphasized across scales for the traffic objects such as vehicles, pedestrians, and cyclists. In the beginning, Atrous convolution is incorporated to increase the receptive field, but keeps the spatial resolution intact. This is useful in the sense that the model captures the features of distant and small objects such as far-field pedestrians and cyclists, while still maintaining the resolution required to segment them from the background. Moreover, the feature pyramid network is used to preserve the feature hierarchies across scales, which is useful for simultaneously representing the highly variable scales of vehicles, pedestrians, and cyclists in the same scene, and prevents the dominant features in the images from overwhelming the features of the small objects. Within this hierarchy, SE blocks perform channel-wise recalibration, adaptively amplifying weak but semantically relevant activations corresponding to pedestrians and cyclists while suppressing redundant background responses that contribute to boundary ambiguity. Simultaneously, the CBAM is employed to refine the feature localization by selectively focusing the model on the structurally consistent regions of each object class. As a result, the combined effect of receptive field expansion, channel prioritization, and spatial attention prevents feature collapse in densely populated far-field regions, reduces mask coalescence between adjacent instances, and stabilizes boundary reconstruction across these three object categories, thereby enabling more precise and spatially consistent segmentation masks that serve as trustworthy input for the next detection stage.

To address the difficulties of vehicle, pedestrian, and cyclist detection using segmentation masks in the complex traffic conditions, a new YOLOv12-L Based TDA-LightGBM Gradient-Optimized Fuzzy CatBoost Network (YTG-FCN) is proposed, where TDA-LightGBM, Gradient-Optimized Fuzzy Classifier, and CatBoost are integrated into the hidden layers of YOLOv12-L. This YOLOv12-L adopted a Residual Efficient Layer Aggregation Network and area-based attention mechanism, which allowed effective multi-scale feature extraction of segmentation mask inputs, simultaneously capturing the structural features of all three types of traffic objects. The first hidden layer has the TDA-LightGBM to deal with residual boundary errors and mask propagated coalescence inherited during the segmentation step. The processing property is used to cross two parallel pixel streams and topological feature streams so that the model differentiates properly bounded object regions and coalescent or boundary-eroded mask artifacts that have different topological shape signatures. Subsequently, a Gradient-Optimized Fuzzy Classifier is incorporated

into the second hidden layer to manage systematic assigning bias to large-scale vehicles, and inter-class confusion between visually similar pedestrians and cyclists. The fuzzy membership functions optimized using a Gradient Descent-Optimized Classifier, graded scores of the confidence of the class, which are used to model structural ambiguity as continuous membership scores, and accentuate detection cues of chronically underrepresented small-scale classes. Lastly, the third hidden layer incorporates the CatBoost to counter the light variation-induced mask fragmentation due to shadows and light contrasts. Its ordered boosting structure avoids gradient estimation bias when training on class-imbalanced and structurally inconsistent data, and gives consistent class assignment even when structural integrity is compromised by masking. Therefore, the combination of these three approaches in YOLOv12-L systematically addresses a problem, allowing dependable and precise identification of vehicles, pedestrians, and cyclists in various traffic conditions.

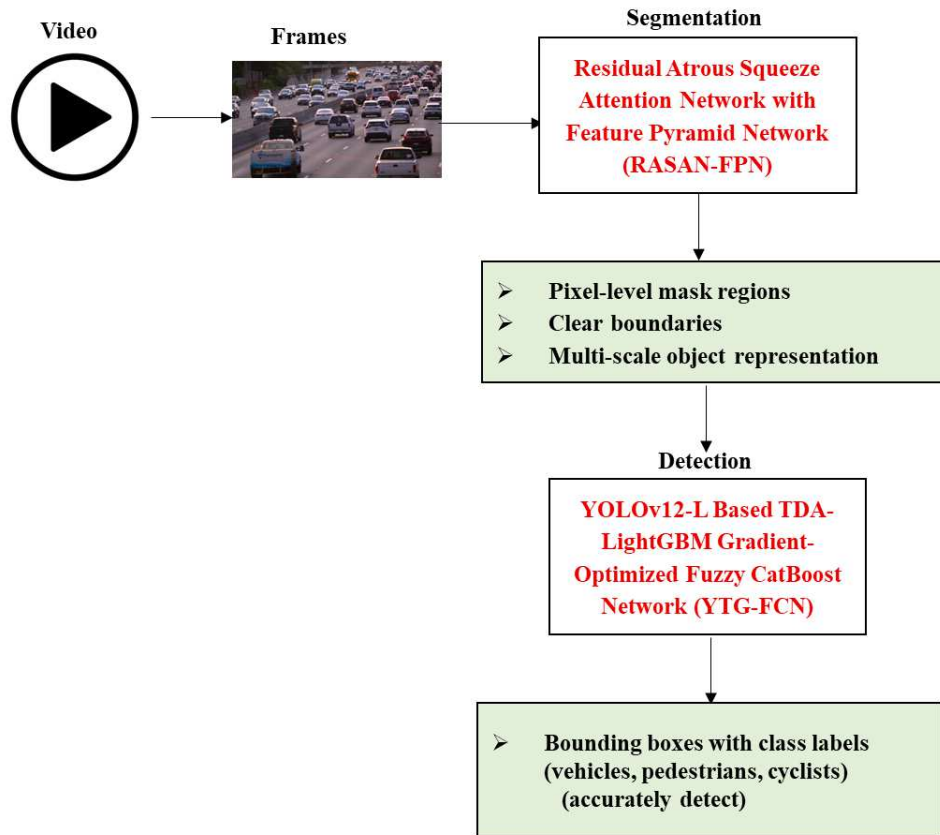


Figure 1: Block Diagram of the RYAS-LOFCN

Figure 1 shows the block diagram of the RYAS-LOFCN for vehicle, pedestrian, and cyclist detection in traffic environments. Initially, the video of the traffic scenario is divided into separate frames and are provided as input to the segmentation model. In the segmentation model, the input frames are segmented into pixels and masks are provided in the image for multi-scale feature representation while capturing distant objects. Then, the segmented masked frames are forwarded to the YOLO-based detection layer, which precisely identifies different objects such as cyclist, pedestrian and vehicle in the traffic environments.

3.1. Residual Atrous Squeeze Attention Network with Feature Pyramid Network (RASAN-FPN) for segmentation

A traffic scenario contains a lot of objects such as vehicles, pedestrians and cyclists, in which the existing methods struggle to preserve distant objects that appear small in the frame, as the repeated pooling operation reduces spatial resolution and suppresses weak object features. Besides, traditional approaches have limitations in terms of providing information for both small-scale and surrounding objects, which ultimately results in inaccuracies in the generation of masks in the segmented image. To overcome the limitations of the existing techniques, a novel Residual Atrous Squeeze Attention Network with Feature Pyramid Network (RASAN-FPN) is introduced. The model integrated Atrous Convolution, FPN, SE block and CBAN into a residual segmentation framework for effective segmentation, which is clearly explained in the below sub-sections step-by-step.

3.1.1. Atrous Convolution Module

Large-scale contextual informations of the image are expanded by the Atrous convolution module without reducing the actual spatial resolution. This module preserves fine-grained structure of the image by capturing small-scale distant features, thereby addresses the limitations of existing models in the disappearance of distant pedestrian and cyclist. The initial raw video input is divided into number of frames and each frame is passed into the Atrous module as input. The input image from traffic video is given in equation (1) as follows,

$$I \in \mathbb{R}^{H \times W \times C} \quad (1)$$

where, H represents the height of the input image, W represents the width of the input image and C represents the number of input channel. The, the Atrous convolution expands the receptive field

of the image by performing convolution without reducing the resolution of the image, which is represented in equation (2) [26].

$$F_a(p) = \sum_k I(p + r \cdot k)w(k) \quad (2)$$

In which, $F_a(p)$ is the output feature at pixel p , I is the input image, r is the dilation rate, k is the convolution kernel index and $w(k)$ is the convolution kernel weight. This increases the receptive using the dilation rate without enlarging the parameter count or reducing the resolution. The higher value of dilation rate allows the module to capture wider range of contextual information of the distant objects in the frame. These kernel weights of the respective object or feature determine its contribution to the segmentation process. This operation captures long-range contextual information while preserving distant objects. Generated Atrous feature maps are fed into the FPN network for maintaining multi-scale hierarchical representation.

3.1.2. Feature Pyramid Network

The FPN receives the Atrous feature maps generated from the Atrous convolution to produce multi-scale hierarchical representation of features. The FPN simultaneously handles large-scale vehicle and small-scale pedestrian by combining feature maps from different scales. The single-scale traditional feature selection models always misrepresent the small-scale features of the objects. However, by taking hierarchies of features, the FPN retains both small-scale features and retains large-scale semantic features. The pyramid fusion operation of FPN in multis-scale feature representation is shown in the equation (3) [27].

$$P_l = Conv(F_l) + Up(P_{l+1}) \quad (3)$$

Here, P_l is the pyramid feature map at level l , F_l is the feature map from layer l , $Up(\cdot)$ is the upsampling operation and $Conv(\cdot)$ is the convolution operation. In this, through the upsampling and convolution operation, the model combines both large-scale and small-scale features. The upsampling process rescues fine details of small object features, removing visual distortions, and convolves the semantic information of strong features at the same scales, and hence represents both strong features and weak features. These hierarchical features produced from FPN are forwarded to the SE block for channel-wise attention and recalibration.

3.1.3. Squeeze and Excitation Block

The SE block receives the hierarchical feature maps from FPN and computes channel-wise importance weights. The SE block identifies the most important features related to the detection of vehicles, pedestrians and cyclists by adaptively recalibrating feature importance to respective features, unlike traditional CNN models that treat all channels equally. The global pooling operation is computed as per in equation (4) [28].

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W P_c(i, j) \quad (4)$$

where, z_c is the channel descriptor, H and W are height and weight of the channels, $P_c(i, j)$ is the feature value at position i, j . Instead of keeping all spatial pixel values, the average global pooling computes a single representative value for the entire channel by averaging all the channel activation features across the height and width of the image. Thus, compresses the spatial equation into channel descriptors, in which the channel attention weights are computed as in equation (5).

$$s = \sigma(W_2 \delta(W_1 z)) \quad (5)$$

where s denotes the channel attention weight, W_1, W_2 are learnable weight matrices, δ is the ReLU activation function and σ is the sigmoid activation function. The channel descriptor vector is passed through two learnable transformers to generate attention scores for each feature channel, as given in equation (6) for performing feature recalibration.

$$\hat{P}_c = s_c \cdot P_c \quad (6)$$

where, $\hat{P}_c$ is the recalibrated feature map. The feature map is recalibrated by multiplying each feature with their respective channel weights, which amplifies the important channels needed for detection and suppresses irrelevant channels. The recalibrated feature maps are passed to the CBAM module for refining both channel and spatial attention features.

3.1.4. Convolutional Block Attention Module

The CBAM models refine both channel and spatial attention features from the recalibrated feature maps received through the SE block by selectively focusing on structurally important object regions. The segmentation mask is refined in this module by emphasizing relevant spatial features and suppressing background noises, which is computed using equation (7) [29].

$$M_s = \sigma(f^{7 \times 7}([AvgPool(\hat{P})MaxPool(\hat{P})])) \quad (7)$$

In which, M_s is the spatial attention map, $f^{7 \times 7}$ is the convolution operation, $AvgPool$ is the average pooling, $MaxPool$ is the maximum pooling and $\hat{P}$ is the recalibrated feature map. The spatial regions with important features for detection are identified by the attention maps and averaging pooling captures the presence of a global object, while the maximum pooling captures the strong local activation features. Then, the refined segmentation features are computed using the equation (8).

$$F_s = M_s \otimes \hat{P} \quad (8)$$

Here, F_s is the refined segmentation feature and $\otimes$ is the element-wise multiplication. The object boundaries are enhanced and irrelevant background regions are suppressed by the element-wise multiplication operation. Finally, segmented mask image is generated using equation (9).

$$S_m = Softmax(F_s) \quad (9)$$

where, S_m is the final segmented mask image. The generated mask image is passed to the detection for stage for vehicle detection in traffic environments. Figure 2 shows the overall process of RASAN-FPN in generating segmented mask images from the raw input image.

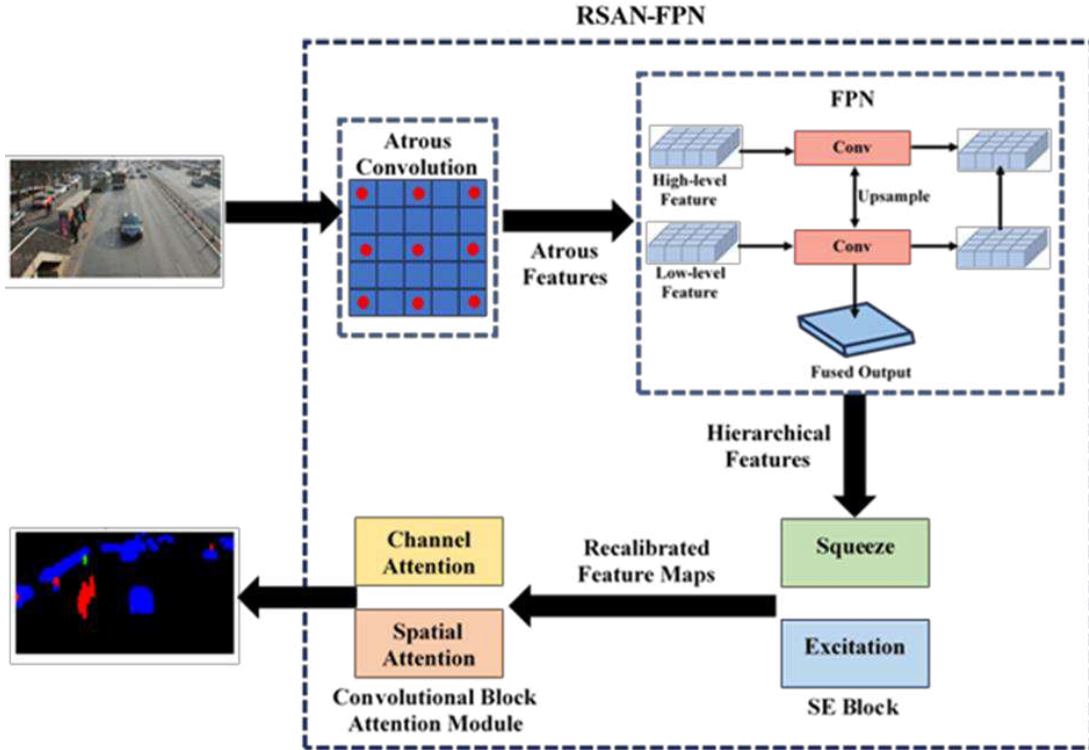


Figure 2: Architecture Diagram of RASAN-FPN

Figure 2 shows the overall process of the RASAN-FPN based Segmentation. Firstly, the Atrous convolution module is used to generate Atrous feature maps from the raw image, and FPN is used for the hierarchical extraction of features. The SE block then recalibrates the features and lastly, the CBAM refines the features and yields the final segmented mask image.

3.2. YOLOv12-L Based TDA-LightGBM Gradient-Optimized Fuzzy CatBoost Network for detection

Mask relics due to illumination and fragmented foreground areas produced as a result of shadow and light contrast in various dense traffic environments, reduces the detection accuracy of the model. To address these issues of existing techniques, a novel YOLOv12-L Based TDA-LightGBM Gradient-Optimized Fuzzy CatBoost Network (YTG-FCN) is proposed for effective detection of vehicle in dense traffic environments. The model integrated TDA-LightGBM, Gradient-Optimized Fuzzy Classifier and CatBoost into the hidden layer of the YOLOv12-L architecture for effective detection of objects in traffic environments across all three classes. The proposed detection process is explained clearly in the below sub-sections step-by-step. The architecture diagram of YTG-FCN in detecting vehicles is represented in Figure 3.

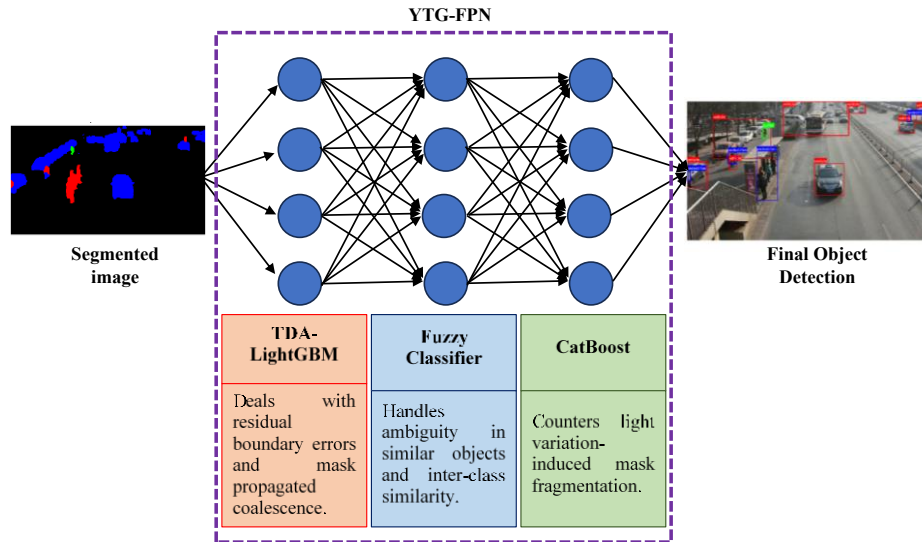


Figure 3: Architecture of YTG-FCN

Figure 3 demonstrates the architecture of YTG-FCAN, which integrated TDA-LightGBM, Gradient Optimized Fuzzy Classifier and CatBoost classifier in the hidden layer of YOLOv12-L architecture for vehicle detection in traffic environments. In the architecture of YTG-FPN, TDA-

LightGBM deals with residual boundary errors and mask propagated coalescence, Fuzzy Classifier handles ambiguity in similar objects and inter-class similarity and CatBoost counters light variation-induced mask fragmentation. By integrating these three modules in the YOLOv12-L architecture, YTG-FPN provides accurate detection of vehicles, pedestrians and cyclists.

3.2.1. YOLOv12-L architecture

YOLOv12-L extracts hierarchical traffic object features from segmentation masks produced by the RASAN-FPN module. Unlike conventional approaches in object detection that struggle to detect large vehicle and small pedestrian at the same, the YOLOv12-L architecture improves multi-scale feature representation through residual feature aggregation. Equation (10) shows the extraction of feature from segmented mask image and the residual aggregation is given in equation (11) [30].

$$F_y = \text{Conv}(S_m, W) \quad (10)$$

$$R_o = F_y + F_{skip} \quad (11)$$

Here, F_y is the extracted feature map, S_m is the segmented input mask, W is the convolution weights, R_o is the residual output and F_{skip} is the skip connection feature map. The object features from the segmented mask images are extracted using the convolution operation and the spatial details are preserved by the residual connection by combining deep semantic features, thereby improving the stability for the detection of distant traffic objects. The extracted features from the segmented mask are forwarded to the TDA-LightGBM module to deal with residual boundary errors.

3.2.2. TDA-LightGBM

The TDA-LightGBM module corrects the segmentation errors by analyzing topological object structures from the extracted multi-scale features. Conventional techniques have limited reliability in detection stage, as they are always affected by the fragmented masks and merged objects. However, the TDA-LightGBM module effectively discriminates continuous object feature structures from the distorted regions in segmentation. Topological representation of features is represented in equation (12) [31].

$$T_f = \phi(R_o) \quad (12)$$

where, T_f is the topological feature representation and $\phi()$ is the topological descriptor extraction function. This topological feature representation extracts information regarding connectivity and shape of the detected image. Then, the boosted prediction is represented in the equation (13).

$$Y_t = \sum_{m=1}^M \gamma_m h_m(T_f) \quad (13)$$

In which, Y_t denotes the boosted output, h_m is the weak learner, γ_m is the weight of the learner and M is the number of boosting stages. The boosted prediction module combines the multiple weak learner to improve the reliability of the prediction, where higher learner weights improve the accuracy in discrimination between valid and invalid masks. These refined features are passed to the Gradient Optimized Fuzzy Classifier.

3.2.3. Gradient Optimized Fuzzy Classifier

The fuzzy classifier handles ambiguity in similar objects and inter-class similarity from the refined features of TDA-LightGBM. Objects like pedestrian and cyclist exhibits similar structural patterns to each other in dense traffic, leading to ambiguity in identification. So, the fuzzy classifier is incorporated in this model, which provides gradual confidence estimation for each feature instead of rigid classification., thereby mitigating the ambiguity associated with object identification in dense traffic environments. The fuzzy membership classifier computes class membership classifier as per in equation (14).

$$\mu_i(x) = \frac{1}{1 + \left| \frac{x - c_i}{\sigma_i} \right|} \quad (14)$$

Here, $\mu_i(x)$ is the fuzzy membership score, c_i is the class center and σ_i is the spread parameter. Objects like pedestrian and cyclist exhibits similar structural patterns to each other in dense traffic, leading to ambiguity in identification. Then, the gradient optimization is applied as per in equation (15) [32].

$$\theta_{t+1} = \theta_t - \eta \nabla L(\theta_t) \quad (15)$$

where, θ_t is the parameter set, η is the learning rate and L is the loss function. The gradient optimization function updates the fuzzy parameters to reduce the errors during classification of the objects. These optimised fuzzy outputs are given to the module CatBoost for final object classification.

3.2.4. CatBoost Module

The CatBoost module performs ordered boosting for robust object classification on optimized fuzzy outputs from Gradient Optimized Fuzzy Classifier. In the absence of structurally consistent training data such as variation in light and class imbalance, standard boosting methods suffer from instability in prediction. However, the ordered boosting of the CatBoost module improves the reliability and consistency of the classification. The ordered prediction function is given in equation (16) [33].

$$F(x) = \sum_{k=1}^K \alpha_k T_k(x) \quad (16)$$

where, $F(x)$ is the final predicted score, $T_k(x)$ is the output of the decision tree, α_k is the weight of the tree and K is the number of trees. For reliable detection of the object, the boosting equation combines the value of multiple weighted decision trees, where the contribution of reliable classifiers is increased by the larger weighted trees. The final object class detection is given in equation (17).

$$O_d = \arg \max (F(x)) \quad (17)$$

Here, O_d is the final detected object class. The class with the most confidence score is selected as the final predicted class of the object. This reliable detection of object stabilizes object classification under inconsistent traffic conditions.

Overall, the integration of these three approaches in YOLOv12-L systematically addresses a problem, allowing dependable and precise identification of vehicles, pedestrians, and cyclists in various traffic conditions. Algorithm 2 shows the overall process of YTG-FCN in detection of vehicles from the segmented mask images.

Algorithm 2: YTG-FCN for detection**Input:** Segmented masked image**Output:** Final detected class image

Step 1: YOLOv12-L architecture

1. Receive the segmented mask image as input.
2. Extract the feature maps from segmented image using equation (10).
3. Perform residual aggregation on the extracted feature using (11).

Step 2: TDA-LightGBM

1. Receive the extracted features from the segmented mask as input.
 2. Extract topological feature representation using equation (12).
-

-
3. Use multiple weighted learners for boosted prediction.
 4. Use weighted boosting operation for refined feature extraction as in equation (13).

Step 3: Gradient Optimized Fuzzy Classifier

1. Get the refined features as input.
2. Compute fuzzy membership score for each class of features using equation (14).
3. Estimate gradual confidence score for each object class.
4. Update the fuzzy classifier parameters.
5. Optimize the fuzzy outputs using gradient optimization as per equation (15)

Step 4: CatBoost

1. Use the optimized fuzzy output as input.
2. Apply ordered boosting using multiple weighted decision trees using equation (16).
3. Select the object class with the maximum score as the output class.
4. Produce the image with output class as output.

The Algorithm 2 shows the overall process of the YTG-FCN in object detection. First, the YOLOv12-L architecture performs residual aggregation on the feature received from the segmented image. Then, TDA-LightGBM refines the feature using gradient boosting and the Gradient Optimized Fuzzy Classifier optimizes the confidence score for the object class. Finally, the CatBoost module applies ordered boosting using multiple weighted decision trees and produces the class with the maximum confidence score as output.

4. Result and Discussion

This section discusses the results of the proposed model in effective segmentation and detection of vehicles in traffic environments. This section contains the performance and simulation results of the proposed model, along with the comparative analysis with existing techniques to justify the better performance of the model.

4.1. System Configuration

The proposed method is simulated in Python for vehicle detection in traffic environments,

Software	: Python
OS	: Windows 11 Pro (64-bit)
Processor	: Intel i5

RAM : 16 GB RAM

The experimental setup is designed to ensure that an accurate performance evaluation is conducted for vehicle identification in traffic environments.

4.2. Dataset Description

The proposed model used a large highway traffic sequence dataset, UA-DETRAC used for training and testing. Original continuous traffic data includes vehicles, pedestrians and cyclists seen through roadside surveillance cameras. Traffic videos are reconstructed from these frame sequences to retain the temporal traffic information by combining consecutive frames. The total of prepared datasets is 100 traffic videos with 70 videos to train the proposed model, and 30 videos for testing the proposed model. During training, each video is further divided into a sequence of frames to train frame-wise feature extraction and object detection, which is a means for effective learning of spatial and temporal traffic features. A variety of traffic scenarios, including dense traffic, distant objects, occlusion, illumination variation, and scale variation, are included in the dataset. Differences in traffic make it more robust and generalized or generic, the proposed framework. The data preparation process is to ensure reliable training and evaluation for real-time, traffic object detection and classification.

UA-DETRAC dataset link: <https://www.kaggle.com/datasets/bratjay/ua-detrac-orig/data>

Vehicle detection and segmentation dataset link:

<https://www.kaggle.com/datasets/ajithdari/vehicle-detection-and-segmentation>

4.3. Simulation Results

This section shows the results of the proposed model obtained through simulation in the segmentation and detection of vehicles in traffic environments.

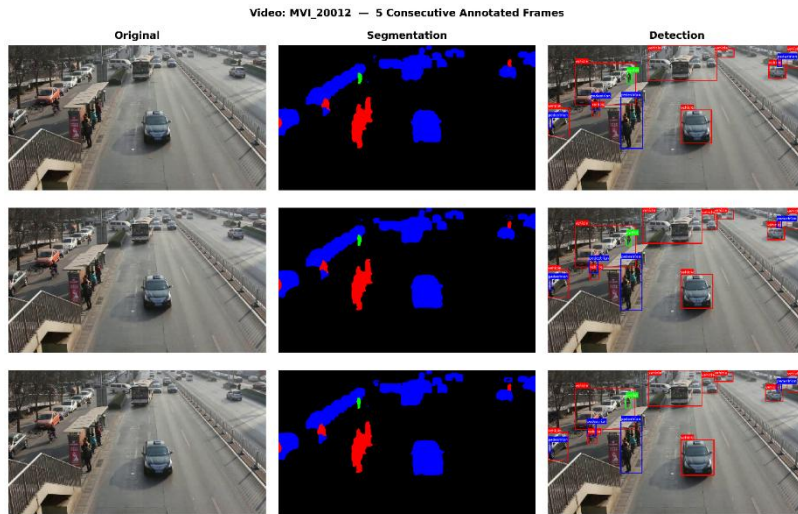


Figure 4: Simulated Result of the First Video

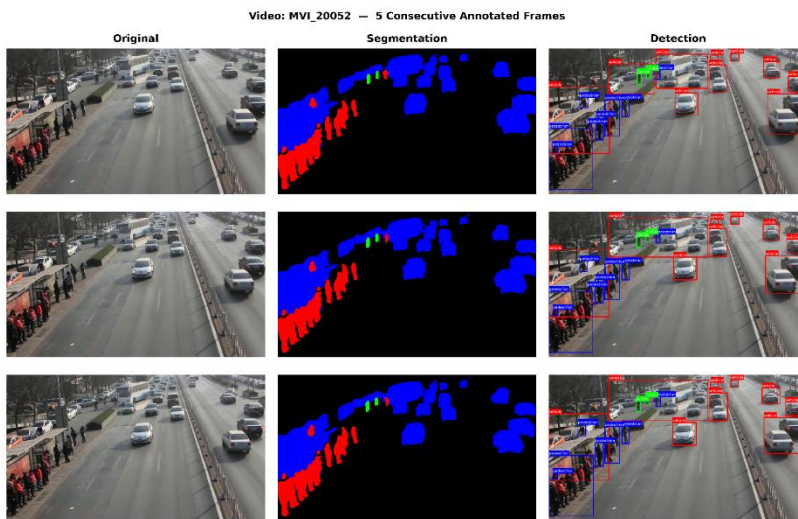


Figure 5: Simulated Result of the Second Video

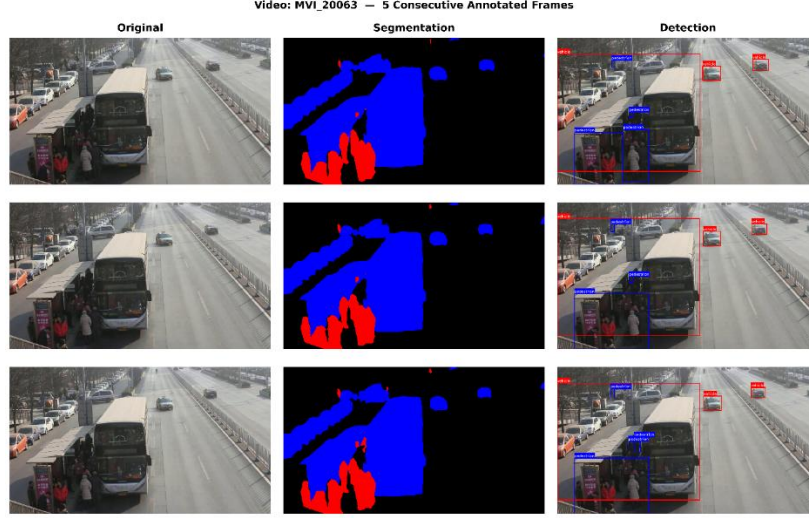


Figure 6: Simulated Result of the Third Video

Figure 4, 5 and 6 shows the simulated results of the proposed model for 3 different videos with 3 continuous frames each. From these, figures it is clear that the proposed model effectively segments and identifies vehicle, pedestrian and cyclist from the given image. In these images, the model was effectively able to capture overlapping object, differentiate pedestrian from cycle and accurately capture distant objects. The fact that the model was able to correctly detect the vehicles in three different videos in 3 continuous frames shows the increased generalizability of the model with low error. In this model, RASAN-FPN provides reliable and effective segmentation of the input image, mitigating the problems involved with distant object and high-speed object detection, while the YTG-FPN model provides reliable detection through its YOLO architecture, mitigating the problems associated with fragmented images and illuminated areas. Together with these, the proposed model provides effective and accurate detection of vehicles in traffic environments with reliability.

4.4. Performance Analysis

Effectiveness and reliability of the proposed model in accurately detecting vehicle, pedestrian and cyclist is discussed in this section, though a performance analysis of the model using different metrics such as, accuracy, precision, recall, F1-Score, $\text{map}@0.5$, $\text{map}@0.95$, GFLOP, dice score, mIoU and inference speed.

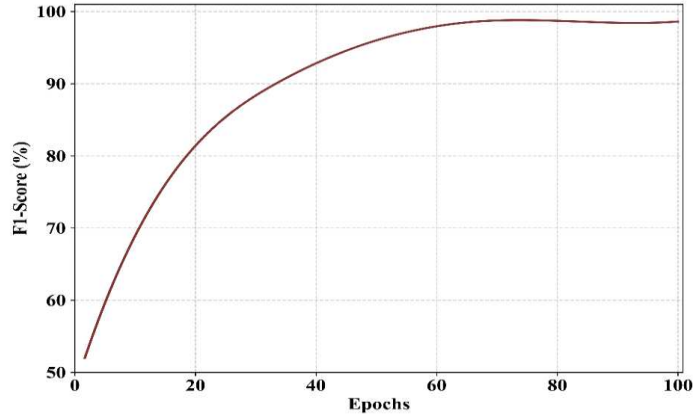


Figure 7: F1-Score of the Proposed Model

F1-Score of the introduced model over 100 epochs is displayed in the Figure 7. The F1-Score of the model starts with a value of slightly above 50% in the initial epoch and increases gradually as the number of epochs increases steeply and reaches a value of around 97% in the 60th epoch. After that, there is a small and steady increase in the value of F1-Score around 98% constantly and reaches a final value of 98.61% at the final epoch. The model has better balance between accurately detecting true classes and reduced detection of false classes, which helps the model in achieving a higher value of F1-Score.

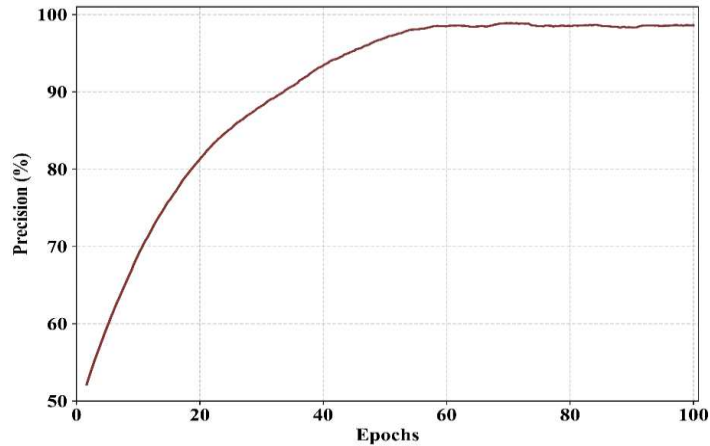


Figure 8: Precision of the Proposed Model

Performance of the proposed model for precision over 100 epochs is displayed in the Figure 8. The model starts at a value of around 50% in the first epoch and as the number of epochs increases, the model learns effectively and quickly, reaching a value of over 97% in the 60th epoch. After that, there is a small fluctuation in the learning rate, still the model has managed to finish at a better

precision of 98.56% at the final epoch. Background noises and false boundary detection are suppressed by SE blocks and CBAM, producing more accurate object prediction, resulting in increased precision of the model.

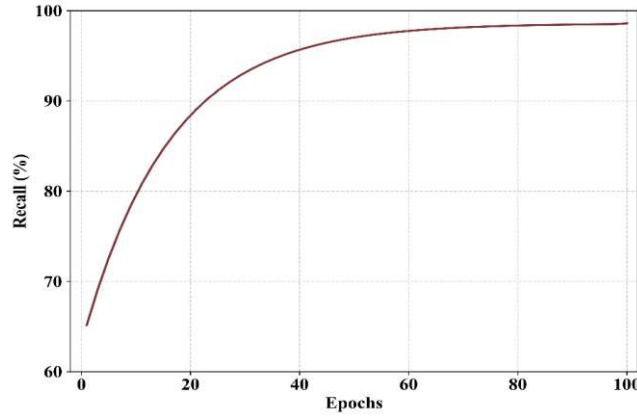


Figure 9: Recall of the Proposed Model

Figure 9 shows the performance of the proposed model based on recall over 100 epochs. The recall of the process also reached a better value of 98.6% at the 100th epoch, even it starts from a value of around 65% in the first epoch. This higher increase in the value of recall percentage over epochs indicate the better learning rate of the model. The RASAN-FPN model for segmentation in the proposed model, effectively captures small and distant objects in traffic environment, resulting in better recall rate of the model.

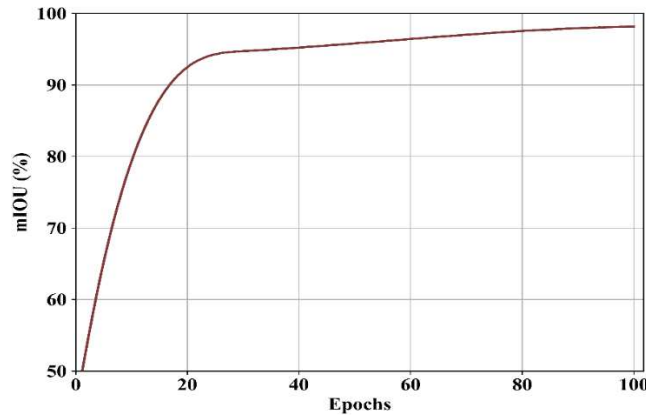


Figure 10: mIoU of the Proposed Model

mIoU performance of the proposed model over 100 epochs is illustrated in the Figure 10. The proposed model starts with a mIoU value of around 50%, which then increases rapidly and reaches

over 90% within the 20th epoch, indicating the better learning rate of the model. After the 20th epoch, the mIoU value of the proposed model increases slowly and steadily, reaching a value of 98/14% at the 100th epoch. The model is able to achieve a higher value of mIoU, due to the incorporation of Atrous convolution and FPN, which produces stable segmentation mask image with clear boundaries.

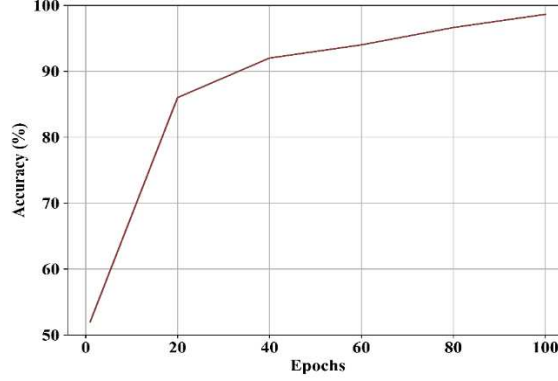


Figure 11: Accuracy of the Proposed Model

Accuracy performance of the proposed model over 100 epochs is demonstrated in Figure 11. The accuracy of the model starts with a value just over 50% at the first epoch, which then increases and reaches a final value of 98.63%. Even though, there are some differences in the learning rate of the model between the epochs, the model is able to achieve a higher value at the end of the training. By integrating RASAN-FPN for reliable segmentation and YTG-FPN for effective detection of vehicles, the model is able to achieve a better accuracy value of 98.63%.

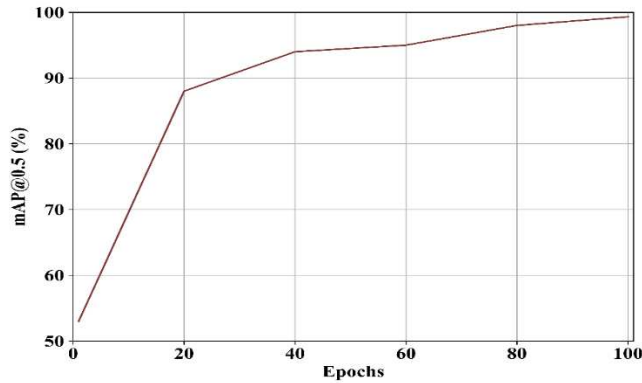


Figure 12: mAP@0.5 of the Proposed Model

mAP@0.5 of the proposed model over 100 epochs is demonstrated in the Figure 12. The model starts with a value of just over 50% at the initial stage, which then tends to increase rapidly and reaches just below 90% at the 20th epoch. After that there is a steady increase in mAP with periodical transitions and the model has managed to attain a value of 99.32% at the 100th epoch. Integration of YTG-FPN for vehicle detection, improves localization and prediction of objects across multiple scales, achieving better accuracy with less false detection, resulting in the higher value of mAP.

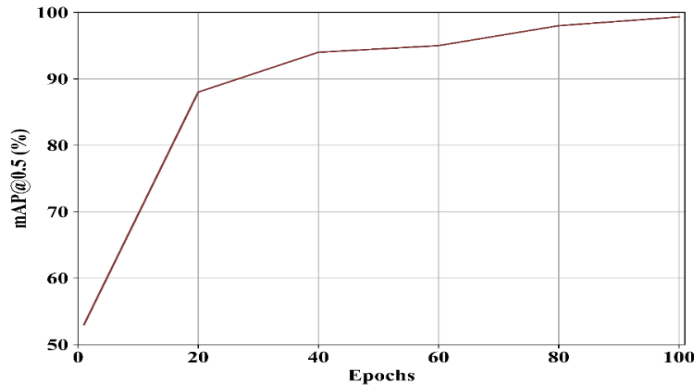


Figure 13: mAP@0.5 -0.95 of the Proposed Model

mAP@0.5-0.95 of the proposed model over 100 epochs is demonstrated in the Figure 13. The model starts with a mAP value of just over 50% at the first epoch, which then increases rapidly and reached a value of just near 90% at the 20th epoch. After that, the model learns at a smooth and reaches a value of 92.46% at the 100th epoch. Integration of YTG-FPN for vehicle detection, improves localization and prediction of objects across multiple scales, achieving better accuracy with less false detection, resulting in the higher value of mAP.

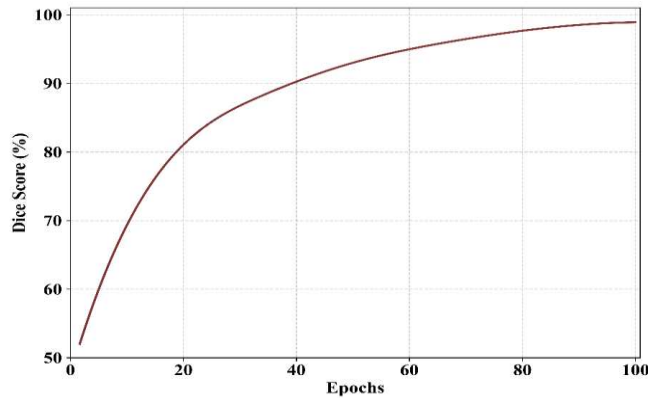


Figure 14: Dice score of the Proposed Model

Figure 14 shows the dice score of the proposed model over 100 epochs. The model starts at the first epoch with a normal dice score value just over 50%, which then tends to increase with the number of epochs due to the better learning rate of the model and reaches a value near 90% just at the 20th epoch. Then, the learns smoothly with periodic changes in learning rate and the value of dice score also increases, reaching a higher value of 98.91% at the 100th epoch. The RASAN-FPN helps the proposed model in generating precise and stable segmentation mask, resulting in the higher dice score of the model.

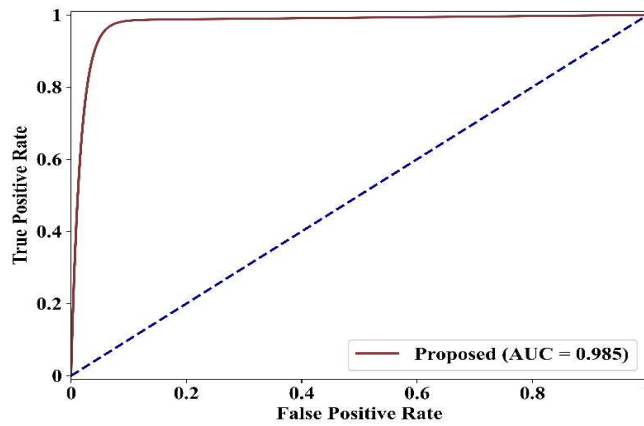


Figure 15: AUC of the Proposed Model

Figure 15 shows the AUC of the proposed model with true positive rate in the y-axis and false positive rate in the x-axis. The model attains a better AUC value of 0.985, indicating the better performance of the model in accurately identifying vehicle in traffic environments. Sharp increase in true positive rate shows the ability of the correctly identify vehicles in complex environments and low false positive rate indicate the lower error of the model. The model achieves better discrimination capability due to the improved separation between vehicles and background by the Gradient-Optimized Fuzzy Classifier and CatBoost, helping the model in accurate detection, increasing the AUC value.

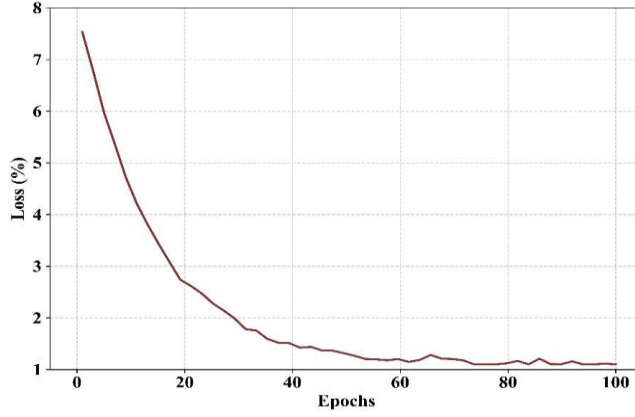


Figure 16: Loss of the Proposed Model

Loss of the proposed model over 100 epochs is shown in Figure 16. In this, the model starts with a loss value of over 7% at the first epoch, which then continues to decrease sharply and reaches a value of below 2% around 30th epoch, demonstrating better learning rate. After that, the loss value decreases smoothly and reaches a value of 1.1% at the final 100th epoch. The model has improved feature learning due to the incorporation of Atrous convolution, attention mechanisms, TDA-LightGBM, and CatBoost, resulting in the lower loss value of the model.

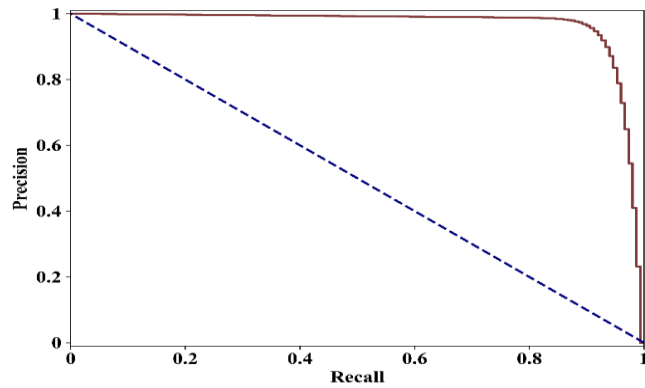


Figure 17: Precision-Recall of the Proposed Model

Precision-Recall curve of the proposed model with precision in y-axis and recall in x-axis is displayed in Figure 17. The proposed model has a higher precision-recall value of 0.984, indicating the better detection capability of the model at different confidence thresholds. The higher value demonstrates that model accurately detects vehicle, pedestrian and cyclist with few false detections, through attention mechanisms and optimized classifiers.

4.5. Comparative Analysis

Performance of the proposed model is compared with the performance of the existing techniques ESRGAN-YOLOv10 [18], STMS [16], SASSC-RNN [17], SSD [19], YOLOv8 [19], SAM [19], YOLOv7-tiny [25], YOLOv8n [25] and YOLOv8-SCConv [25] in vehicle detection based on metrics such as accuracy, precision, recall, F1-Score, map@0.5, map@0.95, weight and GLOP is discussed in this section. This analysis shows that the performance of the proposed model is better than the existing models in vehicle detection in traffic environments.

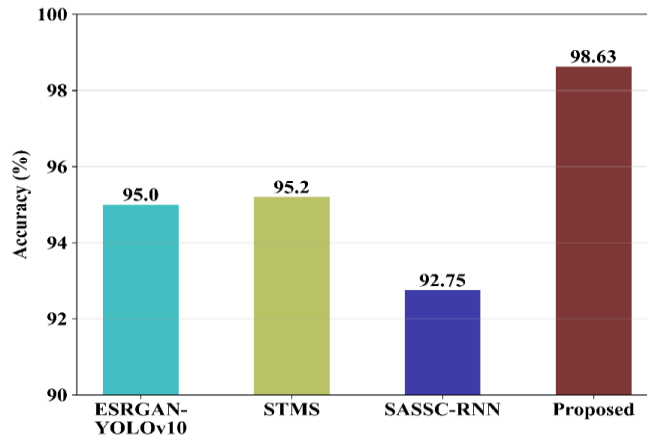


Figure 18: Accuracy comparison of the Proposed Model

Figure 18 shows the comparative analysis of the proposed model with existing techniques such as, ESRGAN-YOLOv10, STMS and SASSC-RNN based on accuracy. In this, existing methods such as, ESRGAN-YOLOv10 has achieved accuracy of 95%, STMS has achieved accuracy of 95.2% and SASSC-RNN has achieved accuracy of 92.75%. In contrast, the proposed model has achieved a higher accuracy of 98.63%, representing its better performance than the traditional models. This higher accuracy of the model is due to the integration of RASAN-FPN, which accurately captures distant and small objects, and YTG-FPN, which mitigates the problems associated with fragmented images and variations in illumination.

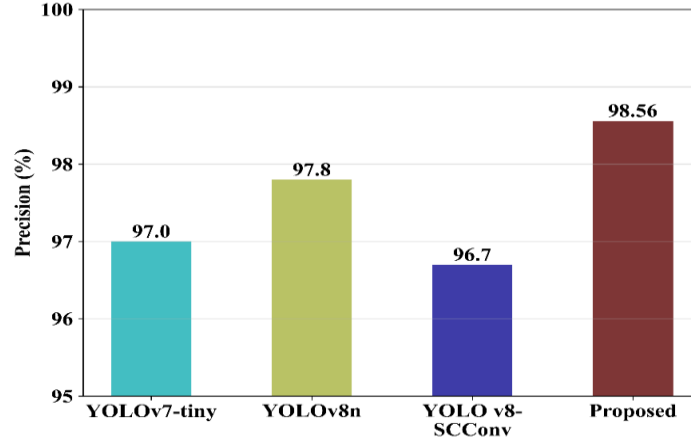


Figure 19: Precision Comparison of the Proposed Model

Comparison of the proposed model with existing techniques based on precision is shown in the Figure 19. In this, existing methods such as YOLOv7-tiny, YOLOv8n and YOLOv8-SCConv has achieved precision of 97% and 96.7%, while YOLOv8n has achieved better precision of 97.8%, comparatively, the proposed model is able to achieve a better precision of 98.56%. Background noises and false boundary detection are suppressed by SE blocks and SBAM, producing more accurate object predicting, resulting in increased precision of the model than the existing techniques.

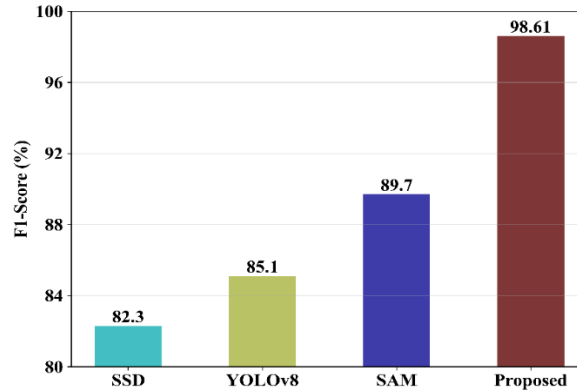


Figure 20: F1-Score Comparison of the Proposed Model

Figure 20 shows the comparison of the proposed model based on F1-Score with existing models such as, SSD, YOLOv8 and SAM. In this, proposed model has achieved a higher F1-Score value of 98.61% due to the better balance between accurate detection reduces false classification. While,

the compared models have shown relatively less values of 82.3% for SSD, 85.1% for YOLOv8 and 89.7% for SAM, indicating the superiority of the proposed over the compared techniques.

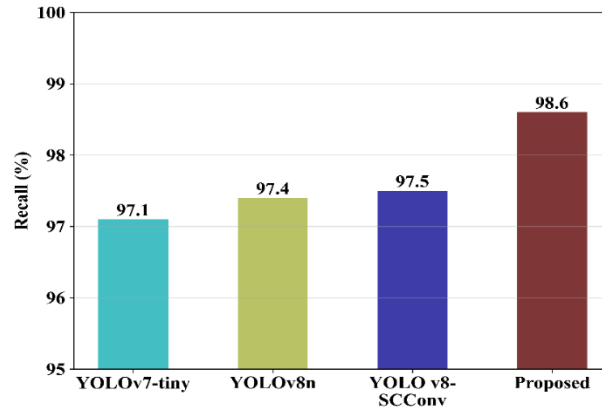


Figure 21: Recall Comparison of the Proposed Model

Comparison of the proposed model with existing models such as YOLOv7-tiny, YOLOv8n and YOLOv8-SCConv based on recall is shown in the Figure 21. From the graph it is clear that YOLOv7-tiny has achieved of 97.1%, YOLOv8n has achieved 97.4% and YOLOv8-SCConv has achieved 97.5% of recall values. Comparatively, the proposed model has achieved a better recall value of 98.6%, which is achieved through the incorporation of RASAN-FPN model for segmentation, which effectively captures small and distant objects in traffic environment, resulting in better recall rate of the model.

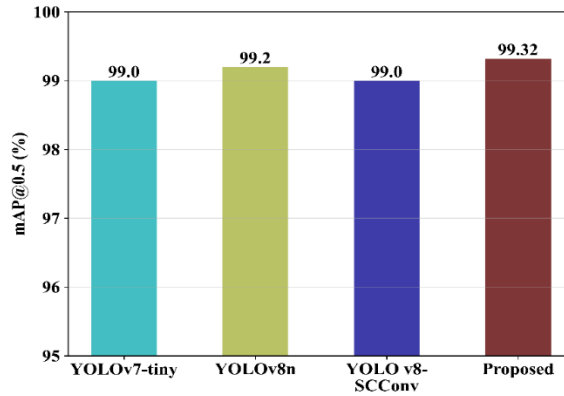


Figure 22: Comparative Analysis of the Proposed Model on mAP@0.5

Figure 22 shows the comparative analysis of the proposed model with existing models based on mAP with at least 50% overlap. In this, YOLOv7-tiny has achieved 99%, YOLOv8 has achieved 99.2% and YOLOv8-SCConv has achieved 99% of mAP scores. Compared to the existing models,

the proposed model has achieved a slightly higher mAP value of 99.3% with at least 50% overlap value. Integration of YTG-FPN for vehicle detection, improves localization and prediction of objects across multiple scales, achieving better accuracy with less false detection, resulting in the higher value of mAP for the proposed model.

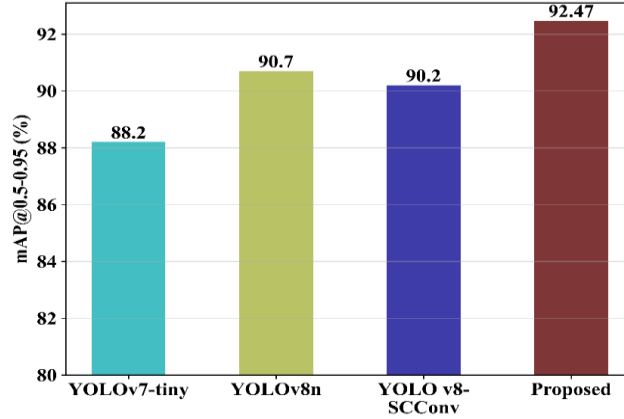


Figure 23: Comparative Analysis of the Proposed Model on mAP@0.5-0.95

Figure 23 shows the comparative analysis of the proposed model with existing models based on mAP with 50%- 95% overlap. In this, YOLOv7-tiny has achieved 88.2%, YOLOv8 has achieved 90.7% and YOLOv8-SCConv has achieved 90.2% of mAP scores. Compared to the existing models, the proposed model has achieved a slightly higher mAP value of 92.47% with 50%-95% overlap. Integration of YTG-FPN for vehicle detection, improves localization and prediction of objects across multiple scales, achieving better accuracy with less false detection, resulting in the higher value of mAP.

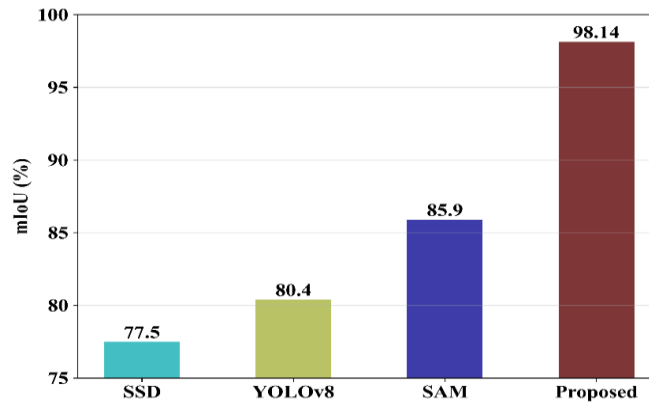


Figure 24: mIoU Comparison of the Proposed Model

Comparison of the proposed model with existing methods such as SSD, YOLOv8 and SAM based on mIoU is shown in the Figure 24. In this, existing methods such as, SSD, YOLOv8 and SAM have managed to attain a mIoU value of 77.5%, 80.4% and 85.9%. In contrast with them, the proposed model has achieved a better mIoU value of 98.14%, indicating its better segmentation capability than the compared models. The model is able to achieve a higher value of mIoU, due to the incorporation of Atrous convolution and FPN, which produces stable segmentation mask image with clear boundaries.

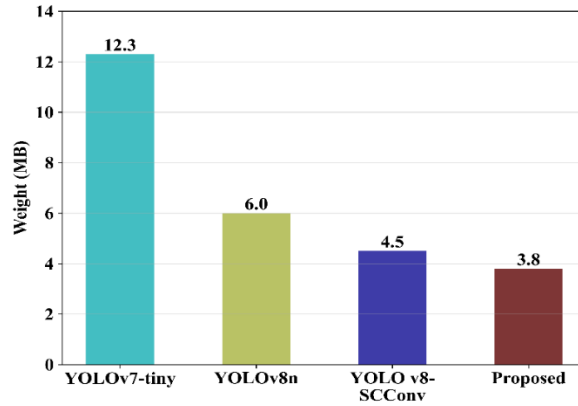


Figure 25: Comparative Analysis of the Proposed Model on weight

Figure 25 shows the comparative analysis of the proposed model with existing methods such as, YOLOv7-tiny, YOLOv8 and YOLOv8-SCConv based on weight of the trained model. Existing methods such as, YOLOv7-tiny, YOLOv8 and YOLOv8-SCConv have weight size of 12.3MB, 6.0MB and 4.5MB, while the proposed model has a lower weight size of 3.8MB for the trained model. This lower weight of the proposed model indicates adaptability of the model in light-weight and low-end devices. Effective feature aggregation and optimized learning mechanism of the proposed model helps the model in reducing the redundant feature representation, resulting in a lower weight size.

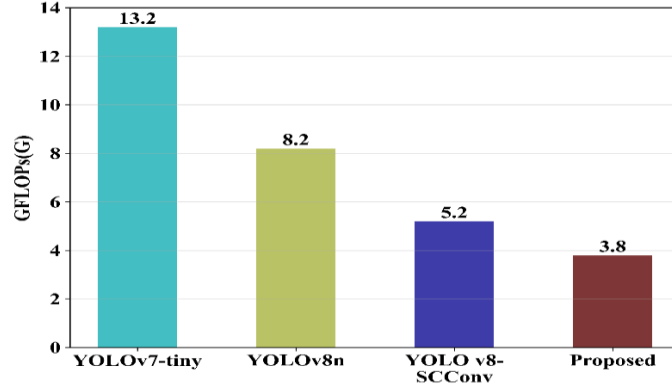


Figure 26: Comparative Analysis of the Proposed Model on GFLOPS

Figure 26 shows the comparative analysis of the proposed model with existing techniques such as, YOLOv7-tiny, YOLOv8n and YOLOv8-SCConv based on GFLOPS. In this YOLOv7-tiny has GFLOPS of 13.2G, YOLOv8n has GLOPS of 5.2G and YOLOv8-SCConv has GFLOPS of 5.2G. Compared to them, the proposed model has a low GFLOPS value of 3.8G, which shows that the proposed model requires fewer floating point operations than the compared model, thereby improving the computational complexity of the model. The model has reduced computational complexity due to efficient feature extraction, attention-based learning and optimized hidden-layer processing.

4.6. Ablation Study

Ablation study of the proposed model is shown in this section to demonstrate the importance of each module in the full process of vehicle detection in traffic environments.

Table 2: Ablation Study

Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Without Atrous Convolution	96.82	96.10	96.18	96.14
Without FPN	97.05	96.42	96.50	96.46
Without SE Block	97.28	96.70	96.82	96.76
Without CBAM Attention	97.51	96.95	97.02	96.98
Without YOLOv12-L	97.74	97.18	97.26	97.22

Without TDA-LightGBM	98.02	97.56	97.64	97.60
Without Gradient Optimized Fuzzy System	98.24	97.88	97.96	97.92
Without CatBoost	98.41	98.16	98.22	98.19
Full Proposed Framework	98.63	98.56	98.60	98.61

Table 2 demonstrates the ablation study of the proposed model, indicating the importance of each model. From the ablation study, it is understandable that even if one of the methods is removed, there is a significant reduction in the values of accuracy, precision, F1-Score and recall. In this, removal of modules results in accuracy ranging from 96.82%-98.41%, precision ranging from 96.1%-98.16%, recall ranging from 96.18%-98.22% and F1-Score ranging from 96.14%-98.19%. However, the whole model with the integration of all the model has achieved a higher accuracy of 98.63%, precision of 98.56%, recall of 98.60% and F1-Score of 98.61%. This shows that, each module plays an important role in the overall process of vehicle detection in traffic environments and even the reduction of one module results in a significant loss of performance.

4.7. Discussion

Performance analysis of the proposed model show that the model is better than existing techniques in vehicle detection in traffic environments. The proposed model has achieved better accuracy of 98.63%, precision of 98.56%, recall of 98.6%, F1-Score of 98.61%, dice score 98.91%, mAP@0.5 of 99.32%, mAP@0.5-0.95 of 92.57%, mIoU of 98.14%, PR of 0.984 and AUC of 0.985, with a low loss of 1.1%. In the proposed model, Atrous convolution captured the features of distant and small objects, FPN preserved hierarchical feature scales, SE block suppressed redundant background noise and CBAM refined feature localization. Together with these, the proposed model provided effective segmentation, while accurately detecting distant and small objects. Then, TDA-LightGBM deals with residual boundary errors and mask propagated coalescence, Gradient Optimized Fuzzy Classifier handles ambiguity in similar objects and inter-class similarity and CatBoost counters the light variation-induced mask fragmentation. Together with these, the proposed model provided effective detection of vehicle under complex traffic scenarios and

fragmented images. Comparative analysis of the proposed model with existing techniques is shown in Table 3.

Table 3: Comparative Analysis of the Proposed Model

Metric	Precision (%)	Recall (%)	mAP @0.5 (%)	mAP @0.95 (%)	Weight (MB)	GFLOPS (G)
YOLOv7-tiny	97	97.1	99.0	88.2	12.3	13.2
YOLOv8n	97.8	97.4	99.2	90.7	6.0	8.2
YOLOv8-SCConv	96.7	97.5	99.0	90.2	4.5	5.2
Proposed Model	98.56	98.6	99.32	92.47	3.8	3.8
	F1-Score (%)			mIoU (%)		
SSD	82.3			77.5		
YOLOv8	85.1			80.4		
SAM	89.7			85.9		
Proposed Model	98.61			98.14		
	Accuracy (%)					
ESRGAN-YOLOv10	95%					
STMS	95.2%					
SASSC-RNN	92.75%					
Proposed Model	98.63					

Table 3 shows the comparative analysis of the proposed model with existing techniques such on metrics such as, accuracy, precision, recall, F1-Score, mAP@0.5, mAP@0.5-0.95, mIoU, weight

and GFLOPS. Comparative analysis revealed that in each of the evaluated metrics the proposed model performs better than current models. This shows that the proposed model provides effective segmentation and detection of vehicle in traffic environments than the traditional models.

5. Conclusion

The proposed RYAS-LOFCN provides effective segmentation and detection of vehicle in traffic environments, while mitigating the issues of traditional models in distant and small object detection, and fragmented foreground areas. In the proposed model, Atrous convolution captured the features of distant and small objects, FPN preserved hierarchical feature scales, SE block suppressed redundant background noise and CBAM refined feature localization. Together with these, the proposed model provided effective segmentation, while accurately detecting distant and small objects. Then, TDA-LightGBM deals with residual boundary errors and mask propagated coalescence, Gradient Optimized Fuzzy Classifier handles ambiguity in similar objects and inter-class similarity and CatBoost counters the light variation-induced mask fragmentation. Together with these, the proposed model provided effective detection of vehicle under complex traffic scenarios and fragmented images. Overall, the proposed model provides a more reliable and accurate detection of vehicle in traffic environments under complex conditions. The results obtained from the proposed model showed that, the model has a higher accuracy of 98.63%, precision of 98.56%, recall of 98.6% and F1-Score of 98.61% with a low loss of 1.1%. Furthermore, the results obtained from comparative study of the proposed model against the previous models verified that the proposed model is a better model for accurate detection of vehicles in traffic environments.

5.1. Future Scope

Even though the proposed model provides accurate detection of vehicle in traffic environments, still the model can be improved further in terms of accuracy and detection, which are as follows:

- Transformer-based temporal attention framework can be integrated in the framework to extend the model to long-term tracking of vehicles and estimation of speed under dense traffic conditions.
- The model will be deployed on edge-enables intelligent transport system, to achieve real-time traffic monitoring with reduced computational cost and low latency.

- Multi-sensor data such as, LiDAR, radar and thermal imaging can be integrated in the model to improve the performance of the model in vehicle detection under adverse weather conditions.

ACKNOWLEDGEMENT

Funding

No Funding

Conflict of Interest

Not Applicable

Competing Interest

no Competing Interests

Ethics, Consent to Participate, and Consent to Publish declarations

Not Applicable

Data Availability

The datasets generated and/or analysed during the current study are available in the [kaggledataset] repository, [<https://www.kaggle.com/datasets/bratjay/ua-detrac-orig/data>, <https://www.kaggle.com/datasets/ajithdari/vehicle-detection-and-segmentation>]

Ethics declaration

Not Applicable

Author Contribution

Kiranmai R- Writing main draft, conceptualization, investigation

Deeptha R- Editing, visualization, supervision, Review

Reference

- [1] Wei, Z., Zou, Y., Xu, H. and Wang, S., 2025. Small Object Detection in Traffic Scenes for Mobile Robots: Challenges, Strategies, and Future Directions. *Electronics*, 14(13), p.2614.
- [2] Kovačovič, P., Pirník, R., Tichý, T., Kafková, J., Gašpar, G. and Kuchár, P., 2026. A Comprehensive Analysis of Incident and Object Detection in Traffic Environments. *Smart Cities*, 9(3), p.41.
- [3] Yang, M., Ewe, L.S., Yew, W.K., Deng, S. and Tiong, S.K., 2025. A Survey of Data Augmentation Techniques for Traffic Visual Elements. *Sensors*, 25(21), p.6672.
- [4] Zhou, W., Yang, L., Zhao, L., Zhang, R., Cui, Y., Huang, H., Qie, K. and Wang, C., 2025. Vision technologies with applications in traffic surveillance systems: A holistic survey. *ACM Computing Surveys*, 58(3), pp.1-47.
- [5] Feilong, Q., Khan, N.A., Jhanjhi, N.Z., Ashfaq, F. and Hendrawati, T.D., 2025. Improved YOLOv5 Lane Line Real Time Segmentation System Integrating Seg Head Network. *Engineering Proceedings*, 107(1), p.49.
- [6] Adhikari, R. and Chetri, A., A REVIEW OF IMAGE SEGMENTATION TECHNIQUES AND APPLICATIONS.
- [7] Matias, J., Pinto, F.C. and Couto, P., 2025. Computer Vision Methods for Vehicle Detection and Tracking: A Systematic Review and Meta-Analysis. *Applied Sciences*, 15(22), p.12288.
- [8] Pandey, A., Pandey, S. and Sanghi, A., 2025, July. Integrating Emotion Recognition, Vehicle Detection & Predictive Safety: A Comprehensive Survey for Intelligent Transportation Systems. In *2025 International Conference on Innovations in Intelligent Systems: Advancements in Computing, Communication, and Cybersecurity (ISAC3)* (pp. 1-6). IEEE.
- [9] Wang, J. and Su, J., 2025. A Review of Object Detection Techniques in IoT-Based Intelligent Transportation Systems. *Computers, Materials & Continua*, 84(1).
- [10] Telikani, A., Sarkar, A., Du, B., Santoso, F., Shen, J., Yan, J., Yong, J. and Yap, E., 2025. Autonomous Aerial Vehicle-Aided Intelligent Transportation Systems: Vision, Challenges, and Opportunities.
- [11] El Asri, S., Zebbara, K., Aftatah, M., Azaz, A., Ait Lhoussaine, A., Lahcen, K.A.S., Baarar, M. and Boubrine, O., 2025. Computer Vision for Vehicle Detection: A Comprehensive Review. *Data and Metadata*, 4, pp.873-873.

- [12] Marcelleno, D.J. and Putra, M.P.K., 2025. Performance Evaluation of Yolov8 in Real-Time Vehicle Detection in Various Environmental Conditions. *Jurnal Teknik Informatika (Jutif)*, 6(1), pp.269-279.
- [13] Dolatyabi, P., Regan, J. and Khodayar, M., 2025. Deep learning for traffic scene understanding: A review. *IEEE Access*, 13, pp.13187-13237.
- [14] Ayachi, R., Said, Y., Afif, M., Alshammari, A., Hleili, M. and Abdelali, A.B., 2025. Assessing YOLO models for real-time object detection in urban environments for advanced driver-assistance systems (ADAS). *Alexandria Engineering Journal*, 123, pp.530-549.
- [15] Wei, J., As'arry, A., Rezali, K.A.M., Yusoff, M.Z.M., Ma, H. and Zhang, K., 2025. A review of yolo algorithm and its applications in autonomous driving object detection. *IEEE Access*.
- [16] Talaat, F.M., El-Balka, R.M., Sweidan, S., Gamel, S.A. and Al-Zoghby, A.M., 2025. Smart traffic management system using YOLOv11 for real-time vehicle detection and dynamic flow optimization in smart cities. *Neural Computing and Applications*, 37(24), pp.19957-19974.
- [17] Alshehri, M., Wu, T., Almujaally, N.A., AlQahtani, Y., Hanzla, M., Jalal, A. and Liu, H., 2025. UAV-based intelligent traffic surveillance using recurrent neural networks and Swin transformer for dynamic environments. *Frontiers in Neurorobotics*, 19, p.1681341.
- [18] Mujtaba, G., Hanzla, M. and Jalal, A., 2025, February. Drone surveillance for intelligent multi-vehicles monitoring and classification. In *2025 6th International Conference on Advancements in Computational Sciences (ICACS)* (pp. 1-6). IEEE.
- [19] Li, R., 2025. A multi stage deep learning approach for real-time vehicle detection, tracking, and speed measurement in intelligent transportation systems. *Scientific reports*, 15(1), p.22531.
- [20] Medina, A., Garon, A., Díaz-Bedoya, D. and González-Rodríguez, M., 2025. Aerial Vehicle Detection and DBSCAN-Based Interchange Region Identification for Urban Traffic Analysis. *IEEE Access*.
- [21] Li, X., Wei, J., Wang, H., Dong, L., Chen, R., Yi, C., Cai, J., Niyato, D. and Shen, X., 2025. Towards intelligent transportation with pedestrians and vehicles in-the-loop: a surveillance video-assisted federated digital twin framework. *IEEE Network*.
- [22] Zhang, Y., Xie, Y., Shi, C., Li, Q., Yang, B. and Hao, W., 2025. Identifying vehicle types from trajectory data based on spatial-semantic information. *Geo-Spatial Information Science*, 28(4), pp.1757-1773.

- [23] Sukanto, S.D., Roy, D., Shakil, F., Singha, N., Asik, A., Joarder, A., Fuad, M.M.N. and Ibrahim, M., 2025. Detecting unauthorized vehicles using deep learning for smart cities: A case study on bangladesh. *arXiv preprint arXiv:2510.26154*.
- [24] Onsu, M.A., Lohan, P., Kantarci, B., Syed, A., Andrews, M. and Kennedy, S., 2025, July. Leveraging multimodal-LLMs assisted by instance segmentation for intelligent traffic monitoring. In *2025 IEEE Symposium on Computers and Communications (ISCC)* (pp. 1-6). IEEE.
- [25] Liu, Y. and Shen, S., 2025. Vehicle detection and tracking based on improved YOLOv8. *IEEE Access*.
- [26] Ankareddy, R. and Delhibabu, R., 2025. Dense segmentation techniques using deep learning for urban scene parsing: A review. *IEEE Access*.
- [27] Zheng, Z., Zhao, J., Fan, J., Bai, R., Zhao, J. and Liu, J., 2025. A complex roadside object detection model based on multi-scale feature pyramid network. *Scientific Reports*, *15*(1), p.15992.
- [28] PANATAGAMA, R.M., 2026. Sistem Deteksi Lajur Jalan pada Malam Hari Menggunakan Deeplabv3+ pada Sistem Autonomous Car= Road Lane Detection System at Night Using Deeplabv3+ in an Autonomous Car System.
- [29] Araújo, V.M., Rili, I., Gisiger, T., Gambs, S., Vasseur, E., Cellier, M. and Diallo, A.B., 2025. AI-powered cow detection in complex farm environments. *Smart Agricultural Technology*, *10*, p.100770.
- [30] Zhang, Y., Wang, W., Ye, M., Yan, J. and Yang, R., 2025. LGA-YOLO for vehicle detection in remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *18*, pp.5317-5330.
- [31] Debbarma, S., Ratankumar, W. and Biswas, S., 2025. Lateral shifting behavior of vehicles at horizontal curves and its influencing factors: application of LightGBM and SHAP. *Traffic Safety Research*, *8*, pp.e000122-e000122.
- [32] Suhermi, N., Aisy, R.R., Rohmah, A.A., Nurhayati, A.A., Pramesty, A.N., Ardanika, A.L. and Isnaini, F.N., 2026. LEVERAGING XGBOOST, LIGHTGBM, AND CATBOOST FOR ENHANCED CUSTOMER SEGMENTATION IN THE AUTOMOTIVE INDUSTRY. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, *20*(3), pp.2281-2298.
- [33] Alotaibi, Y., Nagappan, K., Thanarajan, T. and Rajendran, S., 2025. Optimal deep learning based vehicle detection and classification using chaotic equilibrium optimization algorithm in remote sensing imagery. *Scientific Reports*, *15*(1), p.17921.