International Journal of Computational Science and Engineering Research
ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/
Regular Issue , Vol. 3, Issue. 4 (October – December) , 2026 , Pages: 1 - 12
Received: 29 January 2026 ; Accepted: 22 September 2026 ; Published: 03 October 2026
https://doi.org/10.63328/IJCSER-V3RI4P1
https://www.youtube.com/watch?v=TRSyeJPdxxE
https://archive.org/details/jsp-ijcser-v3ri4p1
Original Paper : Applied Research – Experimental Research Paper

An Improved Yolov3 Framework For Real-Time Object Detection In Autonomous Vehicles

Venkata Krishna Reddy Atchi

School of Informatics, Computing, and Cyber Systems, Northern Arizona University Arizona, United States of America

** Corresponding Author: Venkata Krishna Reddy Atchi ; avkreddy0607@gmail.com*

Abstract: Optical vision is a crucial part of self-driving cars. Vehicle, street, building, pedestrian, and road sign detection that is accurate could help self-driving cars drive as safely as human beings. Object detection, on the other hand, has been a tough task for years due to the effects of lighting, topology, and occlusion on photographs of objects in the real world. Many Convolutional Neural Networks (CNN) based algorithms have improved the detection performance in larger samples in recent years. These approaches have sluggish identification speed, on the other hand, limit their use in real-time scenarios. The Resnet50, revolutionary network architecture with exceptional efficiency acted as the backbone network and the addition of CAR (Concatenation and residual) architecture in the YOLO model is proposed in this research. This architecture uses a topology similar to that of a residual network, but with a few more layers and a few skip connections between layers for which feature reuse is not necessary. This enhanced network is used to detect real-time objects in the You Look Only Once version 3 (YOLO V3) detector. COCO, ImageNet, and PASCAL VOC databases as well as custom datasets are used to train this network for Indian Vehicles. The proposed network outperforms state-of-the-art CNNs, according to the results of the experiments.

Keywords: Deep Learning, CNN , Resnet50, Self-Driving Vehicles, Skip Connections, YOLO V3.

1. Introduction

For the past few decades, automated driving assistance has been a hot topic in computer vision research. The current increase in transportation and traffic has produced a massive logistical difficulty and burden on society. Autonomous vehicles have the potential to provide enormous benefits to the public, including the prevention of traffic accidents, increased comfort, and convenience. Over 1.2 million people are killed and 50 million people have wounded in road traffic accidents annually, according to the World Health Organization [1]. Drivers are responsible for 60% of all traffic accidents and injuries [2], passengers for 20%, and pedestrians for 20%. As a result, the autonomous driver assistance system should adhere to traffic police rules, traffic signals, vehicle indicators, lane markings, vehicle speed, and traffic signs in a consistent manner. Humans are prone to distractions, weariness, and boredom when driving for long periods. As a result, there are many accidents, some of which are fatal. As a result, the focus of current research is on supporting automated vehicles. Even though autonomous driving is becoming increasingly important, the technology is still in its early stages. Our cities and highways are highly volatile ecosystems in which various players interact, such as pedestrians, animals, street furniture, and other cars. As a result, Autonomous Vehicles will require sophisticated perception systems to correctly hold the environment and, be able to interpret what is happening in the environment and act accordingly. Substantial progress has been made in the fields of computer vision, machine learning, and mobile robotics research in recent years. Deep Learning and, in particular, Convolutional Neural Networks (CNNs), have seen significant advancements, with promising results. CNNs are believed to have made a significant advance, outperforming current systems on tasks including object detection and classification, as well as semantic segmentation and interpretation (even surpassing human capabilities [3]). CNN's are attracting a lot of attention these days, whether as a novel technique of extracting robust features to replace hand-crafted ones or as an end-to-end trainable system. This backdrop drives to contribute to the study and development of novel and resilient Convolutional Neural Network-based algorithms that will replace current computer vision approaches in autonomous vehicle perception systems. The goal of this

Jack Sparrow Publishers © 2026, IJCSER, All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 4 , 2026 , https://doi.org/10.63328/IJCSER-V3RI4P1

project is to develop a method for developing an automated driver support system that will be used in automated vehicles that operate in the Indian traffic environment. This study primarily focuses on detecting things on the road, such as vehicles and pedestrians.

The YOLO V3 object detector, which employs densenet50 to identify the items, inspired this research. Even though DenseNet not only solves the vanishing-gradient problem [4], it also improves feature reuse. DenseNet, on the other hand, uses a lot of GPU RAM due to the concatenation operations. We feel that dense connectivity, while more efficient, causes redundancy when early characteristics aren't needed in subsequent levels. This research proposes a novel method for removing inefficient layer connections and replacing them with a more efficient architecture. Furthermore, other benchmarks, including Caltech [8], KITTI [9], ImageNet [10], and MS COCO [11], have made significant contributions to the object detection field.

2. Literature Survey

2.1. Vehicle detection algorithm based on manual feature design

There are many different types of vehicles, and when considering important visual components, textures are significant. This enables the use of algorithms based on pattern recognition to detect vehicles. The researchers used a variety of feature extraction approaches to discover these traits. Common extraction techniques include the Haar feature [12], histogram of oriented gradients feature [13], Haar feature [13], and pattern-focused features like local binary pattern (LBP) and local ternary pattern (LTP) They are not specifically made for vehicle detection, but the study's suggestion that these three qualities apply to the vehicle recognition method based on the picture properties of the vehicle and statistical technique learning. Hu [14] introduced the local binary pattern (LBP) and multi-resolution directional gradient histogram (HOG) features. The HOG features from the training data are recovered and overlaid to extract the LBP features from the training samples. To extract the LBP features from the training data, the HOG features are recovered and stacked over three distinct resolutions of the training samples. A feature vector is created by combining the overlay HOG and LBP features. The detection rate is much higher when compared to a vehicle detection system with just one attribute. The excellent classification speed and effectiveness of the histogram intersection kernel support vector machine are used in this method to train and classify the features. The combination of the Haar and HOG Features in a vehicle detection system was suggested by Wei [15]. The potential area of interest (ROI) was identified using Haar features, and the target cars were found using the inherent benefits of HOG features. Additionally, the cascade structure A trained support vector machine (SVM) was used to retrieve the more precise findings after the AdaBoost approach had selected the HOG properties of the ROI regions. Experimental results demonstrate that the method provides efficient and accurate identification under real-world conditions. Huang [16] presents a method for detecting vehicles and estimating the distance between them. On urban and suburban roads, their work performs wonderfully. In their research, vehicle properties are extracted using HOG features, and vehicle area is identified using an SVM classifier. Surprisingly, the algorithm recognizes cars based on the bottom-left shadows of the vehicles.

2.2. Detection Algorithm- Dual-Stage Detection

The detection procedure is split into two phases by the two-stage detection approach: developing potential regions and categorizing them related to location. Convolutional neural networks are used by the well-known object identification method R-CNN [117,18,19] to extract vehicle features (CNN). The strategy has a two-stage resolution for detecting the issue. The selective search algorithm creates zones of interest proposition first (ROI). Secondly, identify the region and item type that is the most accurate using the Support vector machine. The ROI pooling layer in Fast RCNN [20] solves the duplication calculation issue that existed in the region convolutional neural network (R-CNN). Additionally, softmax is used instead of the SVM classifier in the Fast R-CNN method to improve the detection results. Both R-CNN and Fast RCNN have the selective search method that can be replaced with RPN networks in Faster RCNN [21], which is an upgraded version of Fast R-CNN. The addition of the Smooth L1 loss function is another invention in the Faster RCNN, which smooths the training process and enhances detection performance. The R-FCN is the integration of Faster RCNN and FCN, the region-based fully convolutional network (R-FCN) [22] seeks to enhance object detection results.

2.3. Detection algorithm -Single-stage detection.

The core concept of single-stage detection is to uniformly do massive sampling in various sizes and aspect ratios of the pictures, and apply CNN to extract features before immediately classifying and doing theregression, all in one step. Because of this, uniform intensive sampling, although having the benefit of being rapid, also has the drawback of being more difficult to train, which leads to marginally worse model detection and classification accuracy. A single shot detector (SSD) creates a consistent, end-to-end regression problem from the inspection work, and the location and classification are acquired concurrently through a single procedure [23]. To find and categorize targets simultaneously, a comparable Prior box based on Anchor in Faster RCNN was presented, as well as a Pyramidal Feature Hierarchy-based detection method that predicts targets on a feature map of different sensory fields was also added. These

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 4 , 2026 , https://doi.org/10.63328/IJCSER-V3RI4P1

suggestions for converting detection into regression were taken from YOLO. With the aid of a CNN network, YOLO [24–26] implements detection; the algorithm is quick and simple; both the training and prediction processes are end-to-end, and It has the advantage of a bigger field of vision during detection and making it harder to misjudge the background because YOLO does the convolutional calculation of the full image. Here the whole convolutional layer is utilized by the attention module. Additionally, YOLO has strong model durability and good generalization power while migrating. Additionally, the performance of YOLO detection is enhanced by the new Multi-scale training, an adjustable anchor box, and the Darknet-19 feature extraction network. RetinaNet [27], a One-Stage algorithm model with an accuracy equivalent to the Two-Stage detection approach, addresses the problem that the loss of simple samples can mask the loss of a considerable number of complicated instances. RetinaNet's inference time is a bit slower than YOLO and SSD's.

3. Object Detection Using YOLO V3

3.1. Bounding Box Prediction

This method predicts bounding, boxes using dimension clusters as anchor boxes, similar to YOLO [5]. To make the bounding box, the network predicts four coordinates for each

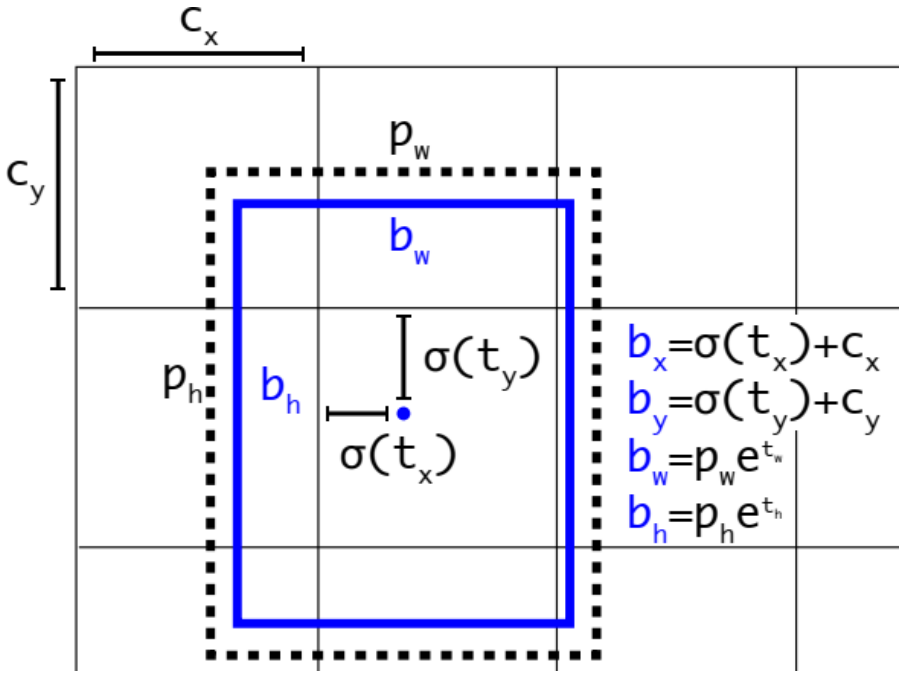


Figure. 1 Bounding boxes with location and dimension priors

3.2. Class Prediction on objects

In each bounding box, multilabel classification is used to predict the corresponding classes of the bounding box could include. Independent logistic classifiers are utilized instead of the softmax layer, which is not essential for maximum performance. Binary cross-entropy loss is used to forecast classes during training. For complex image datasets like the Open Images Dataset, this approach is helpful [6].

There are several overlapping labels in this collection which means women and person. You must presume that each box has exactly one class when using a softmax, even though this isn't always the case. The approach of multiple labels of data modeling is more precise. By utilizing logistic regression, YOLOv3 calculates an abjectness score for each bounding box.

3.3. Predictions Across Scales

Boxes are predicted by YOLOv3 using three distinct scales. This method uses a feature pyramid network-like approach to extract features from those scales [7]. We added several convolutional layers to the fundamental feature extractor. A 3-d tensor is then used to forecast the bounding box, objectness, and class.

3.4. The Architecture OF YOLOv3

The first step of the YOLOv3 algorithm is to divide an image into a grid. The amount of boundary boxes (also known as anchor boxes) that will be placed around objects that perform well in the predetermined categories is predicted for each grid cell. Each bounding box only identifies one item and includes a confidence rating that indicates how accurate the forecast should be. The ground truth boxes' dimensions from the original dataset are grouped to find the most typical shapes and sizes before being used to build the border boxes. R-CNN (Region-based Convolutional Neural Networks), Fast R-CNN (an

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 4 , 2026 , https://doi.org/10.63328/IJCSER-V3RI4P1

R-CNN improvement reported in 2017), and Mask R-CNN are further similar techniques that can accomplish the same objective. Unlike systems like R-CNN and Fast R-CNN, YOLO is trained to do both classification and bounding box regression simultaneously.

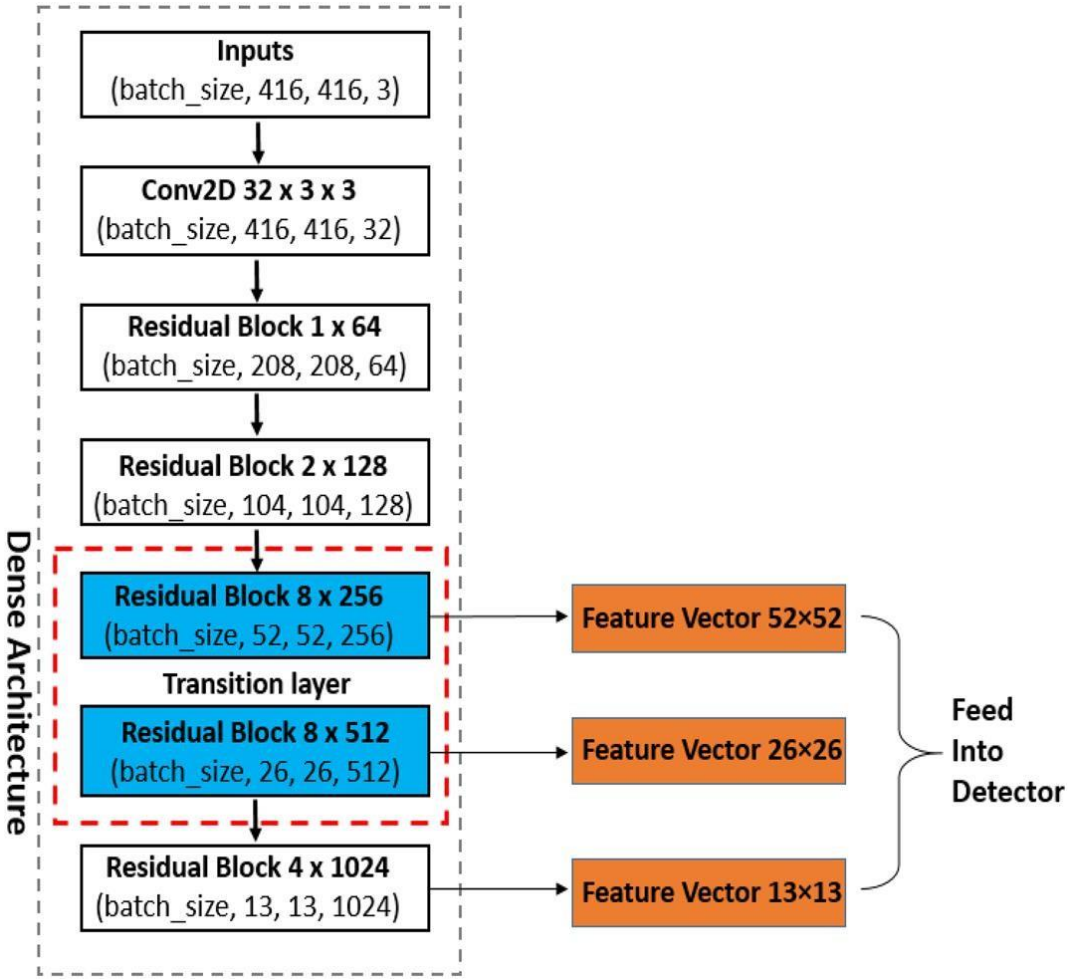


Figure. 2 Yolo-V3 model

In terms of mean average precision (mAP) and intersection over union (IOU) values, YOLOv3 is also quick and precise. Compared to other comparable detection techniques, it is substantially quicker (thus the moniker - You Only Look Once). Figure 2 displays the YOLOv3 darknet model. Furthermore, without the requirement for retraining, By altering the model's size, one can easily strike a balance between accuracy and speed.

3.5. Disadvantages of YOLOv3 vs. Other Algorithms

since the RetinaNet training period is longer than that of yolov3, The yolov3 app does suggest a trade- off when utilizing a Yolo between speed and precision. however, with a bigger dataset, yolov3 can match the retinanet's accuracy in object recognition, Consequently, it is a strong choice for models that benefit from large-scale data training. This is demonstrated by typical detection models like traffic detection, which can be trained using a large amount of data because of the abundance of photos of different types of vehicles. on the other hand, when large datasets are hard to come by, yolov3 might not be the greatest option for specialized models.

4. Architecture of Proposed Model – Car Yolo Model

The proposed model architecture is shown in the following figure 3 with the number of convolutional networks connected with one or more skip connections along with two extra layers of skip convolutional networks. The CAR model is the connection which is the combination of concatenation and residual connections applied in all three feature resolutions from the backbone network. A densely linked architecture proposed by Huang et al. [28] was implemented based on the YOLOv3 design depicted in Figure 3 for increased feature reuse and representation. This design allows for more compact andprecise detection models.

Figure 4 depicts an overview of the updated Vehicle detection model for two classes. The YOLO-Vehicle model the dense architecture was designed to replace the residual blocks 8×256 and 8×512. This is to improve the

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 4 , 2026 , https://doi.org/10.63328/IJCSER-V3RI4P1

detecting scale outlet's deeper network. For each thick layer, a 1×1 bottleneck layer 23 and a 3×3 convolutional layer were put together. To streamline the model, a transition layer was added between the two thick layers. The major goal of the improvements was to make it possible to recognize several feature maps from different network layer sources. This will make it possible to recognize smaller cars with accuracy in a range of situations. The concatenated features of 26 26 768 and 13x13x384 in the feature pyramid network (FPN) of the YOLO-Vehicle model grow to 26x26x28x16 and 13xx13x1408 characteristics in the FPN of YOLO-Vehicle. The YOLO-improved Vehicle's characteristics help to maintain the LWYS technique's ability to detect smaller vehicles with greater precision. To create a more precise and quicker YOLO-Vehicle real-time detection model, The effects of various front detection layer (FDL) reductions and activation functions must be considered. Then the addition of concatenation and residual connection (CAR) in the three feature map layers additionally added the features. The concatenation combining the feature and the residual is removing null features so the addition of features and deletion of null feature maps improves the detection.

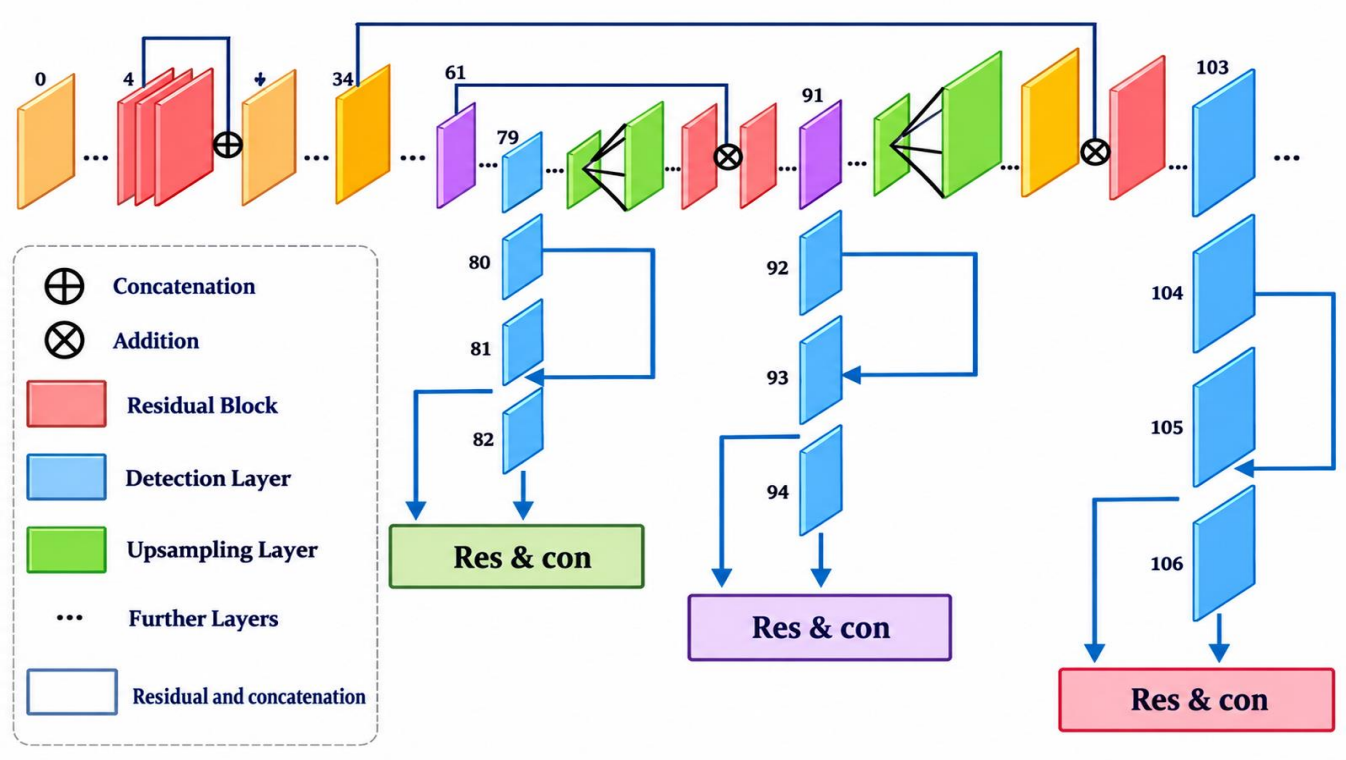


Figure. 3 Proposed model (CAR connection)

4.1. Addition of CAR connection in YOLO-v3

The CAR connection is the combination of concatenation and residual connection which is applied to all the three feature resolutions.

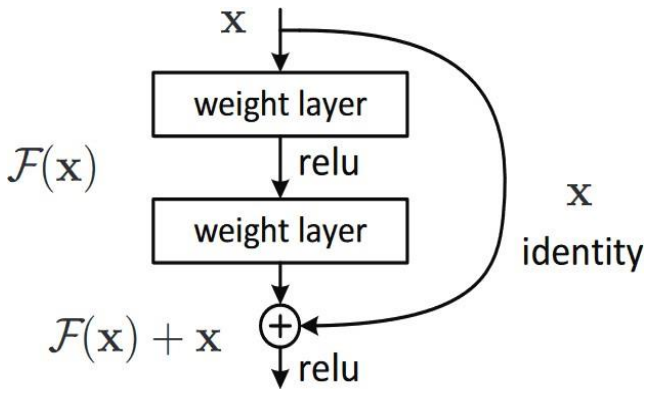


Figure. 4 Skip connections

The residual connection is the previous layer output feature maps values if added with the subsequent layer output, so if the present layer fails to learn, the last outcome layer will come in the place of the current layer The identity connection or skip connection is shown in figure 4.

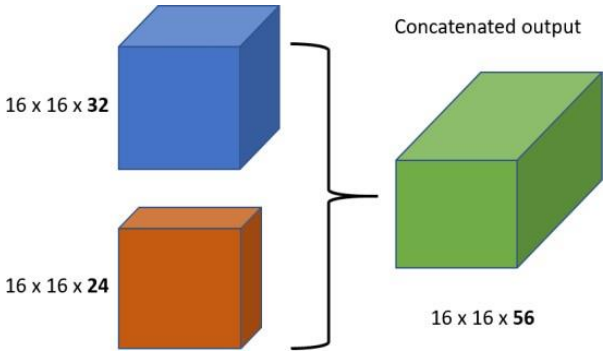


Figure. 5 Concatenation connection

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 4 , 2026 , https://doi.org/10.63328/IJCSER-V3RI4P1

The previous layer is connected to the net layer output so that at least the previous features can be used in the next layer if the current layer fails to learn. This skip or residual connection improves the feature learning process by the network. The residual network 50 is acting as the skip connection network here.

4.1.1. Concatenation Connection

The feature concatenation method yielded dense multiscale information from input images for vehicle classification. is displayed in figure 5. The sequence layers, which consisted of 3 feature layers from the backbone network, finally concatenate the output of different resolutions and the residual connection output.

4.2. Feature Extractor-Resnet 50 for YOLO-V3

The YOLO detector uses Resnet-50 for feature extraction, and two intermediate and final layer features are given to the detector to detect the vehicles. Shortcut connections in ResNet 50 used to skip two layers, but now they skip three levels, and 1 * 1 convolution layers have been added.

The following components are part of the Resnet-50 architecture, as illustrated in table 1: Convolution yields a single layer with a stride of size 2, a kernel size of 7 * 7, and 64 different kernels. Max pooling with a stride size of two comes next. The following convolution starts with a 1 * 1,64 kernel, then moves on to a 3 * 3,64 kernel, and eventually a 1 * 1,256 kernel. We have nine levels in this phase thanks to the repetition of these three layers a total of three times. A kernel of 1 * 1,128, a kernel of 3 * 3,128, and a kernel of 1 * 1,512 are the next three kernels. We went through this process four times for a total of 12 layers. The next kernels are 1 * 1,256, 3 * 3,256, and 1 * 1,1024; they are repeated six times for a total of 18 layers. There were then two more 3 * 3,512 and one additional 1 * 1,2048 kernels added, for a total of nine layers.

A total of 50 layers are created after performing an average pool, a fully connected layer with 1000 nodes, a SoftMax function, and an average pool. Convolutional deep network. Figures 4 and 5 illustrate the skipped connections.

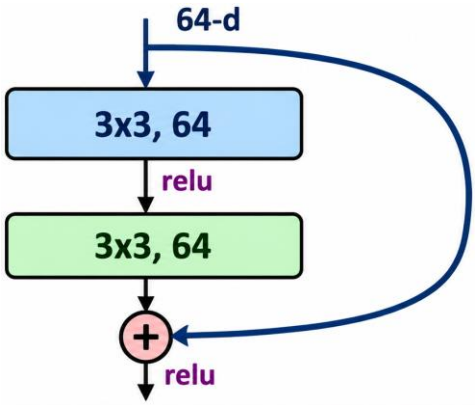


Figure. 4 2-layer Skip connection

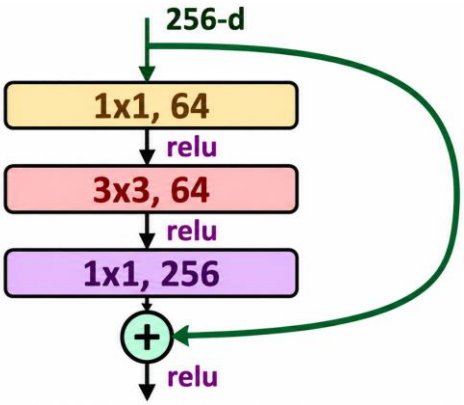


Figure. 5 3-layer Skip connection

YOLO is the name of a Convolutional Neural Network for Object Detection (CNN). Systems that use classifiers are CNNs. that can spot patterns in incoming pictures and translate them into well-organized data arrays. The benefit of YOLO is that it is faster than other networks without compromising accuracy. It enables the model to evaluate the entire image at test time, enabling it to generate predictions based on the image's complete context. The resemblance of areas to predetermined categories is taken into account when scoring regions using YOLO and other convolutional neural network techniques. High-scoring regions emphasize positive detection of the class that most closely resembles them.

Depending on whether portions of the video do better for particular vehicle classes, YOLO might be used, for instance, to identify various automobile kinds in a live traffic stream.

4.3. Training Details

4.3.1. Loss function

The reduced function shows the system's ability to accurately recognize class labels. The loss function is cross-entropy. The loss function minimization equation is 3.

$$L(w) = \sum N\ 4c{=}1{-}yic\ log fc(xi){+}\in \|w\|2 \quad (3)$$

fc(xi) − predicted probability of class c for image x

4.3.2. Gradient Decent optimizer

Slope drop or gradient descent is generally utilized as an advancement calculation that iteratively refreshes the organization's learnable boundaries, i.e., portions and loads, to limit the misfortune. The angle of the misfortune work gives us the course wherein the capacity has the steepest pace of increment, and each

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 4 , 2026 , https://doi.org/10.63328/IJCSER-V3RI4P1

learnable boundary is refreshed the negative way of the slope with a discretionary still up in the air dependent on a hyperparameter called the learning rate.

4.3.3. ***Learning rate (η)***

Exponential decay, constant, and factored are the three techniques used to lower the learning rate during training. In the first step, a constant T is regarded as a reliable parameter that can be used with a step function to manually slow down learning.

The following equation may then be used to modify the learning rate during training:

$$t\eta t = \eta 0 \beta \epsilon \quad (4)$$

$$\eta 0 = 0.0001$$

Where t denotes the learning rate, $\eta 0$ denotes the starting rate, and β denotes the decay factor, with a range of (0,1).

Table. 1 Residual Network Architectures

layer name	34-layer	50-layer	101-layer
conv1	$7 \times 7,64$, stride 2		
	3×3 max pool, stride 2		
conv2_x	$\begin{bmatrix} 3 \times 3,64 \\ 3 \times 3,64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1,64 \\ 3 \times 3,64 \\ 1 \times 1,256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1,64 \\ 3 \times 3,64 \\ 1 \times 1,256 \end{bmatrix} \times 3$
conv3_x	$\begin{bmatrix} 3 \times 3,128 \\ 3 \times 3,128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1,128 \\ 3 \times 3,128 \\ 1 \times 1,512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1,128 \\ 3 \times 3,128 \\ 1 \times 1,512 \end{bmatrix} \times 4$
conv4_x	$\begin{bmatrix} 3 \times 3,256 \\ 3 \times 3,256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1,256 \\ 3 \times 3,256 \\ 1 \times 1,1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1,256 \\ 3 \times 3,256 \\ 1 \times 1,1024 \end{bmatrix} \times 23$
conv5_x	$\begin{bmatrix} 3 \times 3,512 \\ 3 \times 3,512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1,512 \\ 3 \times 3,512 \\ 1 \times 1,2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1,512 \\ 3 \times 3,512 \\ 1 \times 1,2048 \end{bmatrix} \times 3$
	average pool,2048-d fc		

5. Results And Discussion

5.1. Data Augmentation

The Custom dataset was used to train and test the models. we have taken a total of 10 object classes. The training and testing images are separated by an 80-20 ratio In figure 6, the experimental outcomes are displayed. It displays how training and testing data are set up. A system's functionality is calculated by inspecting the classification labels. The specific parameters are mathematically expressed as follows.

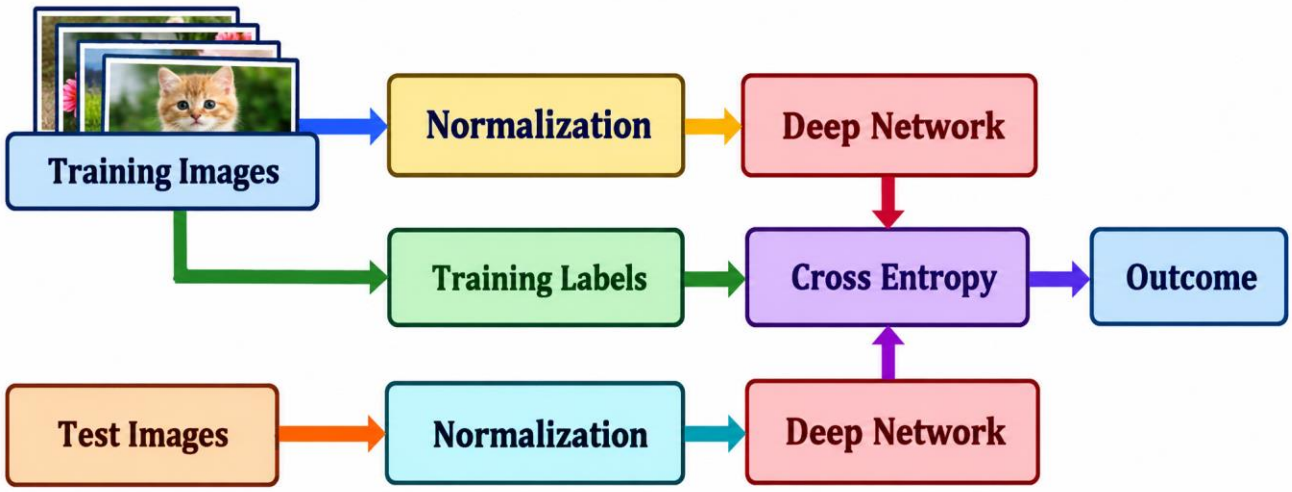


Figure. 6 Vehicle detection experimental setup

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 4 , 2026 , https://doi.org/10.63328/IJCSER-V3RI4P1

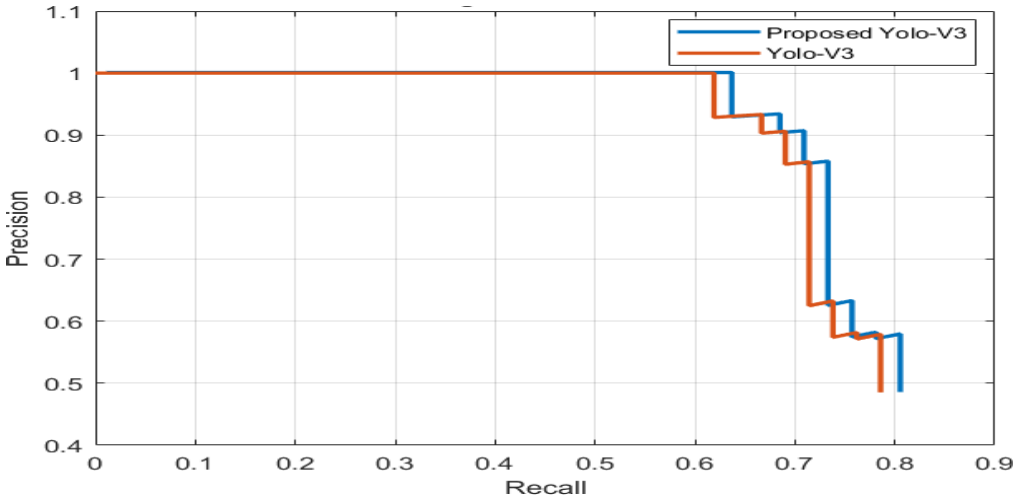


Figure. 7 Map curve for the system



Figure. 8 Sample outcome

The comparison of the Mean Average Precision (MAP) for both the existing YOLO V3 model and the proposed YOLO V3 model is shown in figure 7. It can be seen that the proposed model admits higher MAP values than the classic YOLO V3 model. It is due to the addition of the front-end detection layer. It eventually helps in the appropriate detection of almost all the objects in the image as seen in figure 8.

$$\textbf{Precision} = (Tp)/(Tp + Fp)$$
$$\textbf{Recall} = Tp/(Tp + Fn) \quad (7)$$
$$\textbf{F1 -score} = 2 * (\text{Precision} * \text{Recall})/(\text{Precision} + \text{Recall})$$

After normalizing the input, the performance metrics of all the existing models and the proposed method is observed and tabulated in table 2. The sensitivity of detection is improved which is the most important metric.

Table. 2 With the normalization of inputs

Method	Accuracy	Sensitivity	Specificity
RCNN	91	93	93
Yolov2	93	94	92
Yolov3	93.5	94.3	91
Proposed	94.9	96	95.9

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 4 , 2026 , https://doi.org/10.63328/IJCSER-V3RI4P1

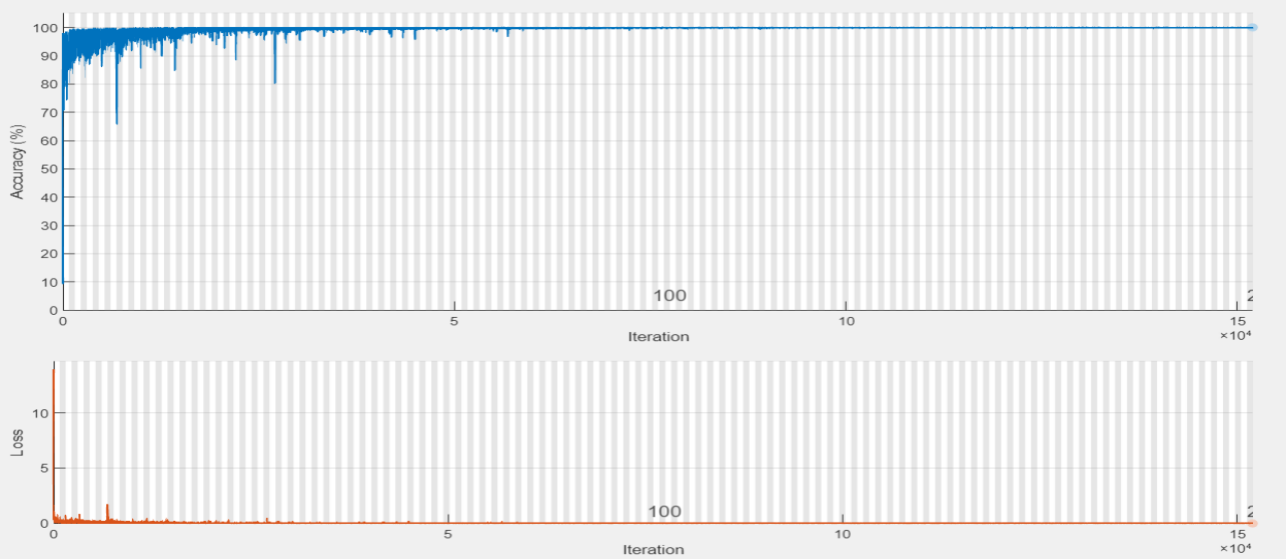


Figure. 9 Training progress plot

The higher accuracy and specificity of the proposed method also make the proposed model highly reliable and robust. Specifically, in comparison to YOLO V3, this proposed customized model with the front-end detection layer always shows better performance. The plot of the metrics such as Acc, Prec, and Sens are found to be improved in comparison to other approaches as shown in figure 10 and figure 11. The training plot displayed in figure 9 shows the accuracy rapidly increasing and touching the full potential in fewer iterations. Also, the loss is suddenly decreasing, which shows the network can extract the features well and learn the model quickly.

Table. 3 Performance Comparison

Input Images	YOLO V3				Proposed Model			
	Accuracy	Sensitivity	Specificity	F1- Score	Accuracy	Sensitivity	Specificity	F1- Score
Test 1	0.86	0.88	0.68	0.90	0.90	0.93	0.83	0.94
Test 2	0.89	0.95	0.80	0.93	0.92	0.96	0.88	0.97
Test 3	0.90	0.89	0.78	0.88	0.91	0.92	0.85	0.93
Test 4	0.75	0.82	0.66	0.85	0.87	0.91	0.78	0.92

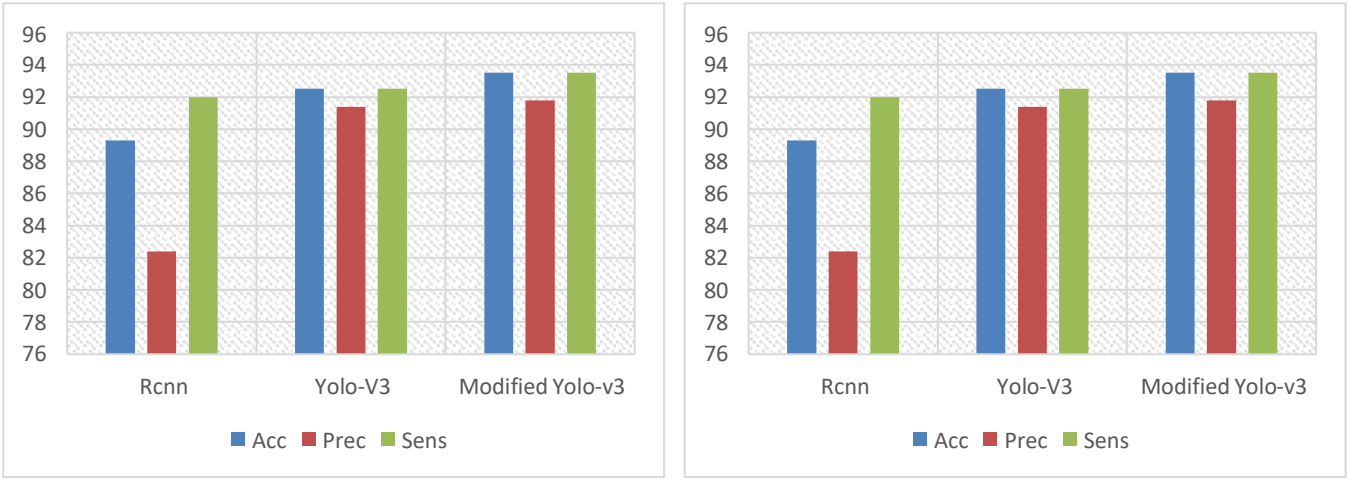


Figure. 10 Detection With Normalisation

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 4 , 2026 , https://doi.org/10.63328/IJCSER-V3RI4P1

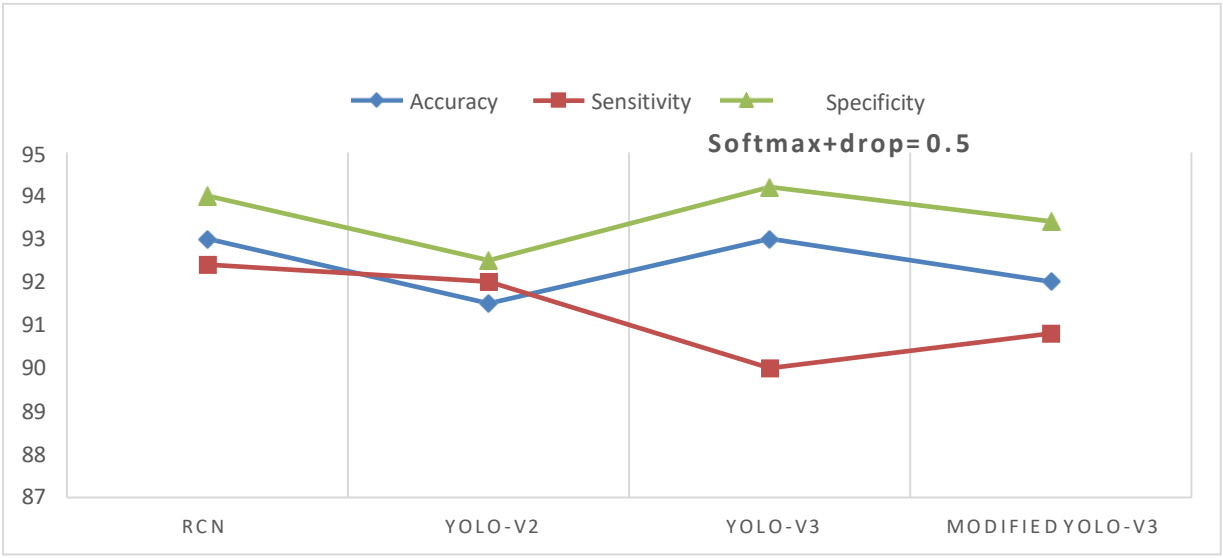


Figure. 12 In case of Dropout 0.5

Figure 12 shows the performance regarding the dropout layer. The significance of the drop layer is examined, for dropout 0,5 performance improved compared drop out 0.7

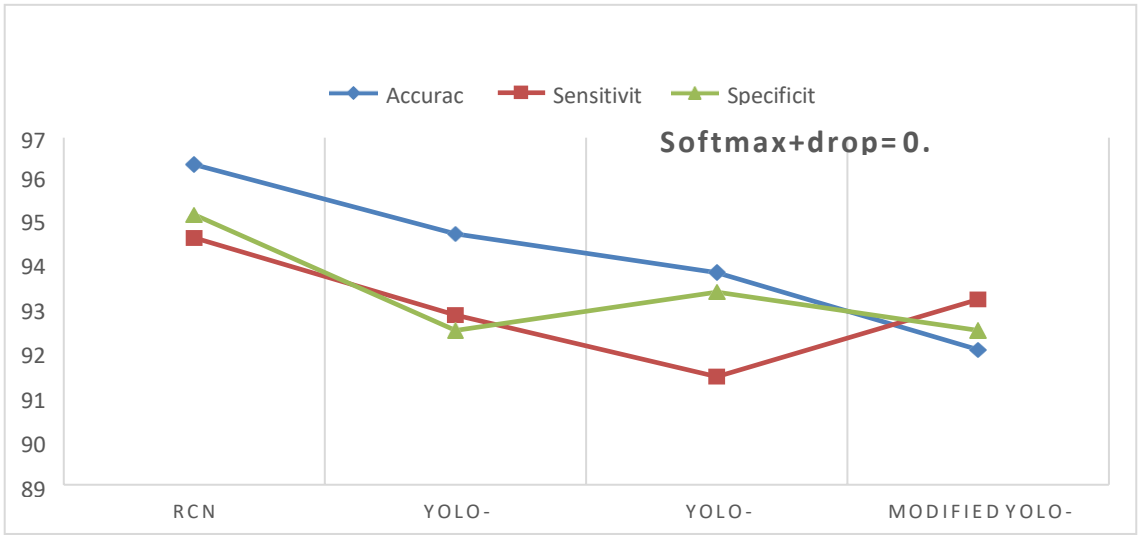


Figure. 13 In case Drop out 0.7

Figure 13 shows the performance regarding the dropout layer with 0.7. The significant drop in the performance compared drop out 0.5

Table 4 Performance comparison of the proposed system with the different backbone network

Different Backbone	YOLO V3				Modified YOLO V3			
	Accuracy	Sensitivity	Specificity	F1-Score	Accuracy	Sensitivity	Specificity	F1-Score
Darknet-53	0.86	0.85	0.72	0.87	0.91	0.93	0.83	0.90
Resnet 18	0.90	0.88	0.81	0.89	0.92	0.96	0.88	0.92
Squuze net	0.91	0.89	0.84	0.87	0.92	0.92	0.85	0.91
Resnet 50	0.92	0.92	0.87	0.83	0.935	0.92	0.90	0.92

Table 5 displays the comparison results of different existing machine-learning models. The proposed YOLOv3 model performed well compared to all other accuracies, sensitivity, and specificity methods. The accuracy got a 1 % improvement, and the specificity got an edge over the base Yolov-3 model.

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 4 , 2026 , https://doi.org/10.63328/IJCSER-V3RI4P1

Table. 5 Performance comparison of existing systems

Method	Accuracy	Sensitivity	Specificity	mAP
NN	0.917777	0.882614	0.928877	0.72
SVM	0.921114	0.84568	0.928599	0.75
RCNN	0.91689	-	-	0.78
Fast-Rcnn	0.92721	0.96134	0.987033	0.82
Yolo-v2	0.93352	-	-	0.88
Yolov-3	0.93124	-	-	0.91
Proposed Yolov3	0.9412	0.97166	0.9896	0.93

6. Conclusion and Future Work

This study supported the use of YOLO-Vehicle models for vehicle identification based on a modified YOLOv3 model. Small Vehicle datasets obtained from complex environmental conditions are used, along with the label, YOLOv3's dense architecture allows for the reuse of features for well-generalized Reduce missed detections and inaccurate results using vehicle detection, Mish activation, and spatial pyramid pooling(SPP). The addition of the CAR architecture model gives more robustness to the system for vehicle detection. These methods are employed to give the detector human intelligence. The experimental results demonstrate that the proposed methodologies outperform current cutting-edge techniques, particularly in terms of average precision (AP). The performance of YOLO-Vehicle models increases the detection accuracy in terms of average precision (AP). Overall, the YOLO-Vehicle models show enhanced real-time vehicle detection and generalization, making them appropriate for autonomous cars.

References

[1]. Tabish Hussain, L Shu, T Sosorburan, A Seid Adji, Ali Hassan Khan, and AF Raja, "Road traffic accidents: an observational and analytical study exploring the hidden truths in pakistan and south east asian countries," Healthline, vol. 2, no. 1, pp. 52–7, 2011.

[2]. Margie Peden et al., "World report on road traffic injury prevention," 2004.

[3]. K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," arXiv preprint arXiv:1502.01852, 2015.

[4]. G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In CVPR, 2017.

[5]. Krasin, T. Duerig, N. Alldrin, V. Ferrari, S. Abu-El-Haija, A. Kuznetsova, H. Rom, J. Uijlings, S. Popov, A. Veit, S. Belongie, V. Gomes, A. Gupta, C. Sun, G. Chechik, D. Cai, Z. Feng, D. Narayanan,and K. Murphy. Openimages: A public dataset for large-scale multi-label and multi-class image classification. Dataset : rom https://github.com/openimages, 2017.

[6]. T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2117–2125, 2017.

[7]. P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, pp. 743–761, April2012.

[8]. A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3354–3361, June

[9]. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," International Journal of Computer Vision, vol. 115, pp. 211–252, Dec 2015.

[10]. M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," International Journal of Computer Vision, vol. 88, pp. 303– 338, Jun 2010.

[11]. Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]. Computer vision and pattern recognition, 2005: 886-893.

[12]. Mita T, Kaneko T, Hori O. Joint Haar-like features for face detection[C]. International conference on computer vision, 2005: 1619-1626.

[13]. HU Qing-xin, JIAO Wei, GU Ai-hua. Vehicle detection method based on multi-features fusion and intersection kernel SVM [J]. Journal of hefei university of technology (natural science), 2016.

[14]. Wei Y, Tian Q, Guo J. Multi-vehicle detection algorithm through combining Harr and HOG features [J]. Mathematics and Computers in Simulation, 2018: 130-145.

[15]. Huang D, Chen C, Chen T. Vehicle detection and inter-vehicle distance estimation using single-lens video camera on urban/suburb roads [J]. Journal of

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com

ISSN: 3107 - 8605 (Online) , http://www.ijcser.com/ , Vol. 3, Issue 4 , 2026 , https://doi.org/10.63328/IJCSER-V3RI4P1

Visual Communication and Image Representation, 2017: 50-259.

[16]. N.Mo and L.Yan, "Improved Faster RCNN Based on FeatureAmplification and Oversampling Data Augmentation for OrientedVehicle Detection in Aerial Images", Remote Sensing, 2020.

[17]. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards RealTime Object Detection with Region Proposal Networks", 2015, arXiv:1506.01497v3.

[18]. Jia Y, Shelhamer E, Donahue J. Caffe: Convolutional Architecture for Fast Feature Embedding[C].acm multimedia, 2014: 675-678.

[19]. Girshick R. Fast R-CNN[C]. International conference on computer vision, 2015: 1440-1448.

[20]. Ren S, He K, Girshick R. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[21]. Dai J, Li Y, He K. R-FCN: object detection via region-based fully convolutional networks[C]. Neural information processing systems, 2016: 379-387

[22]. Liu W, Anguelov D, Erhan D. SSD: Single Shot MultiBox Detector[C]. european conference on computer vision, 2016: 21-37.

[23]. Redmon J, Divvala S K, Girshick R. You Only Look Once: Unified, Real-Time Object Detection[C]. Computer vision and pattern recognition, 2016: 779-788.

[24]. Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]. Computer vision and pattern recognition, 2017: 6517-6525.

[25]. Redmon J, Farhadi A. YOLOv3: An Incremental Improvement [J]. arXiv: Computer Vision and Pattern Recognition, 2018.

[26]. Lin T, Goyal P, Girshick R. Focal Loss for Dense Object Detection[C]. International conference on computer vision, 2017: 2999-3007.

How To Cite:

Venkata Krishna Reddy Atchi ," An Improved Yolov3 Framework For Real-Time Object Detection In Autonomous Vehicles ", International Journal of Computational Science and Engineering Research , Volume 3, issue 4 , pp 1 - 12, October 2026.
CrossRef DOI: https://doi.org/10.63328/ijcser-v3ri4p1

Declaration

Conflicts of Interest: The authors declare no conflict of interest.

Author Contribution: All authors wrote the main manuscript text and also consent to the submission.

Ethical approval: Not applicable.

Consent to Participate: All authors consent to participate.

Funding: Not applicable, and No funding was received

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Personal Statement: We declare with our best of knowledge that this research work is purely Original Work and No third party material used in this article drafting. If any such kind material found in further online publication, we are responsible only for any judicial and copyright issues.

Generative AI Statement: The authors declare that generative AI was not used in the preparation or creation of this manuscript. The alternative text (alt text) accompanying the figures in this article was generated by Frontiers with the assistance of artificial intelligence. Reasonable efforts have been made to ensure the accuracy and appropriateness of the generated alt text, including review by the authors wherever possible. If you identify any inaccuracies or issues, please contact the editorial team.

Publisher's Note: The statements, opinions, and claims presented in this article are exclusively those of the authors and do not necessarily reflect the views of their affiliated institutions, the publisher, editors, or reviewers. Any products discussed or evaluated, as well as any claims made by their manufacturers, are not warranted, approved, or endorsed by the publisher.

Acknowledgements

We thank everyone who inspired our work

Jack Sparrow Publishers © 2026, IJCSER , All Rights Reserved
www.jacksparrowpublishers.com