IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PERFORMANCE EVOLUTION OF VEHICLE DETECTION AND TRACKING USING YOLO'S

[1]K. Nagi Reddy [2]S Indumathi, [3]Sk. Reshma , [4]R. Eswar, [5]Sk. Sofia, [6]T. Harsha sai

[1,2] Professor,[3,4,5,6] under graduate students

[1]Department of ECE,NBKR Institute of Science & Technology, Vidyanagar, Tirupati (Dist),AndhraPradesh,India

ABSTRACT

Vehicle detection and Tracking are gaining importance in traffic management and transportation. However, due to the various types of vehicles, detection remains a difficulty, which directly impacts the accuracy of vehicle. This project introduces a Multiple Objective Vehicle Detection and Tracking using the YOLO Framework. YOLO (You Only Look Once) is a popular object detection algorithm that has revolutionized the field of computer vision. It is fast and efficient, making it an excellent choice for real-time object detection tasks. Object detection is a computer vision task that uses deep learning techniques to detect objects in images and videos. It is a valuable tool for object detection tasks, YOLO is able to detect multiple objects in a single image, while many other CNN-based algorithms can only detect one object at a time. This makes YOLO ideal for real world applications such as self-driving cars and video surveillance. Hence ,It reduces the false detection rate (i.e. Accuracy) of vehicle targets caused by occlusion, an improved method of vehicle detection in different traffic scenarios based on an improved YOLOv8 network is proposed.

Keywords—Yolov3,Yolov5,Yolov8,Video Mp4.

I. INTRODUCTION

Vehicle object detection is an important branch of computer vision, which is also the foundation of driverless, intelligent transportation, vehicle tracking and other fields. In the current practical application scenarios, the main challenge of vehicle object detection is to explore the relationships of the accuracy, efficiency and actual demand of detection algorithm. How to detect vehicle objects in complex traffic scenes quickly and accurately has been the research fields of computer vision and transportation related interdisciplinary. Early traditional vehicle detection algorithms were mainly based on artificial feature extraction. Van [1]. As an essential part of autonomous driving, environmental perception must have the capability of fast and accurate object detection in real-world condition, in order to ensure safe and correct driving behavior and decision [2], [3]. In other words, an object detector applicable for autonomous driving should satisfy the following two prerequisites. First, highly accurate and robust detection performance is required, With the introduction of deep learning, object detection has experienced remarkable development and progress, and much more promising performance has been further reported by adopting advanced convolution architecture.

Currently, CNN-based detectors can be roughly categorized into two classes: two-stage and one-stage object detector. To be specific, two-stage detector, e.g., R-CNN (Region Convolution Neural Network) model [3]–[5], generally reports accurate detection performance by the guidance of region proposal and bounding box (bbox) refinement. However, slow inference speed hinders its application in real-time system due to heavy computational

cost in region proposal generation. In contrast, one-stage detector, i.e., YOLO model [6]–[8], has extremely fast detection speed in a single inference that formulates object detection as a simple regression problem. To investigate a fast but accurate object detector suitable for autonomous-driving system, this work mainly focus on one-stage method and attempt to explore its potentials for performance improvement. As mentioned above, onestage detector is generally less accurate than a two-stage one, and there are three main reasons as follows.

Firstly, the ubiquitous problem of extreme class imbalance in one-stage detector damages detection accuracy a lot. To solve this problem, RetinaNet [9] with focal loss is proposed to eliminate the effect of foreground-background imbalance problem, but it hardly eradicates its imbalance problem. More importantly, discriminative feature and their relation are difficult to discover and capture in a single inference, and consequently, irrelevant or negative feature information causes inaccurate box regression and mislocalization. This is an essential problem resulting in poor detection performance, and aim to improve one-stage detection performance from this perspective. Inspired by human vision mechanism, a fast and accurate object detector termed as self-attention YOLOv3 (SAYOLOv3) is proposed in this work. Based on efficient YOLOv3 [8] . YOLOv5 is a family of single-stage deep learning-based object detectors that are capable of more than realtime object detection with state-of-the-art accuracy. It is the latest release of the YOLO family and is a group of compound-scaled object detection models trained on the COCO dataset used for object detection[10].Besides, YOLOv5 has been used as a solution for long-distance vehicle detection under night conditions for military operations [11]. . The YOLOv5 is approached to compare with YOLOv8 performance since its architecture is an improved version of the former. The model's performance evaluation is based on a comparative test using vehicle images in different conditions, such as position, lighting, and distance. The lack of an appropriate label dataset has been circumvented by building a dataset of 4,075 labeled images This work is structured in five sections. Section 2 introduces some YOLO concepts and focusing on YOLOv5 and YOLOv8. Furthermore, the training and validation methodology is presented in Section 3, including the accurated dataset. The performance evaluation is presented in Section 4 with a discussion about the finds related to the YOLOv5 and YOLOv8 metrics. Section 6 is results and discussion followed by section 7 concludes the paper

2. RELATED WORK

2.1 **Vehicle Detection**

Vehicle detection and statistics in highway monitoring video sequences are critical to intelligent traffic management and highway control. Data is collected for massive database of traffic video footage for analysis. In general, a more distant road surface can be evaluated at a high viewing angle. At this viewing angle, the object size of the car varies dramatically, and the detection accuracy of a small object far away from the road is low. In the face of complicated video sequences, it is critical to properly address and implement the difficulties. This paper focus on the aforementioned concerns to provide a viable solution. The discovery delicacy of a small object far down from the road is low. In the face of complicated video tape sequences, it's critical to duly address and apply the forenamed challenges. This paper concentrate on the forenamed challenges to feasible results, the vehicle discovery results to multi-target discovery and shadowing vehicle counting. Advanced CNN has achieved good results in object discovery. Still, CNN is sensitive to gauge changes in object discovery [11, 12]. The one-stage system is one-stage to prognosticate objects, and the grid's spatial constraints make it insolvable to have advanced perfection with the two-stage approach, especially for small objects. The two-stage system uses the region of interest pooling to member seeker regions into blocks according to given parameters, and if the seeker region is lower than the size of the given parameters, the seeker region is padded to the size of the given parameters. In this way, the characteristic structure of a small object is destroyed and its discovery delicacy is low. The same system is used to deal with the same type of object, which will also lead to inaccurate discoveries.The use of image conglomerations or multi-scale input images can break the below problems, although the computation conditions are large.

2.2 **Vehicle Tracking**

Advanced vehicle object discovery operations, similar to multi-object shadowing, are also critical ITS tasks [13] Utmost multi-object shadowing styles use discovery-grounded styles. Detection-Based Tracking (DBT) and Discovery-Free Tracking (DFT) for object initialization. The DBT system uses background modeling to descry moving objects in videotape frames before tracking them. The DFT system needs to initialize the object of the shadowing but cannot handle the addition of new objects and the departure of old objects. The Multiple Object Tracking algorithm needs to consider the similarity of intra-frame objects and the associated problem of inter-frame objects. Normalized cross-correlation (NCC) can be used to measure the similarity of intra-frame objects.

Presently, discovery- position rejection or lineposition rejection can break this problem. To break the problems caused by scale changes and illumination changes of moving objects,[14] used SIFT points for object shadowing. The Sphere point discovery algorithm [15] is proposed for use in this work. The sphere can gain better birth point points at a significantly advanced speed than SIFT. The perceptivity of Convolutional Neural Networks to gauge changes makes small object discovery inaccurate. A large-scale high-definition dataset of traffic vehicles is established that can provide many different vehicle objects fully annotated in various scenes captured by traffic areas. The dataset can be used to evaluate the performance of many vehicle detection algorithms when dealing with vehicle scale [16]. A method for detecting small objects in traffic scenes is used to improve vehicle detection accuracy. The traffic road surface area is extracted and divided into the remote area and the proximal area is placed into the convolution network for vehicle detection[14]. A multi-object tracking and analysis method for traffic scenes is proposed. The detection object feature points are extracted and matched by the ORB algorithm and the road discovery line is determined to count the vehicle movement direction and the business inflow[17].

3.METHODS

3.1 YOLOV3 Model

YOLO (You Only Look Once) is a popular and general one stage object detection algorithm, which has been developed to the third generation and is called YOLOv3 [18]. The algorithm structure of YOLOv3 is shown in Figure 1. The main Objective of this YOLOv3 for recognition, classification and positioning are transformed into regression problems, which is the core idea of the algorithm. Only one convolutional network is used to predict the classes and location of the object. Figure 2 shows the architecture of the dense road detection system based on YOLOv3, which includes three parts: the YOLOv3 backbone network, a transfer training unit, and optimization of network parameters.

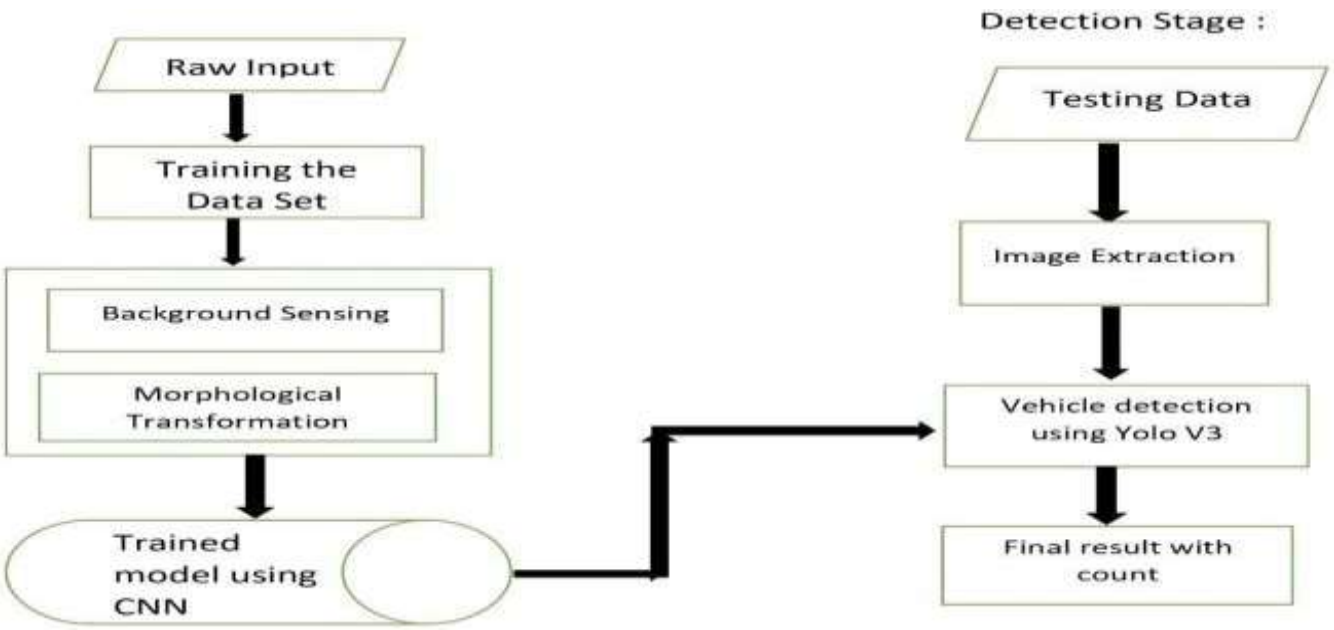


Fig.1. Structure of YOLOv3

VOC 2012, and COCO datasets are selected for YOLOv3 network pre-training; then, the images containing vehicles are extracted from the VOC 2007 dataset and re-labeled to form the special vehicle dataset, and the transfer training-based YOLOv3 model is transferred and trained in this dataset; finally, the test dense road pictures or videos are input into the proposed model, and the output is obtained by feature extraction, multi-scale detection, and Non-Maximum Suppression (NMS) processing.

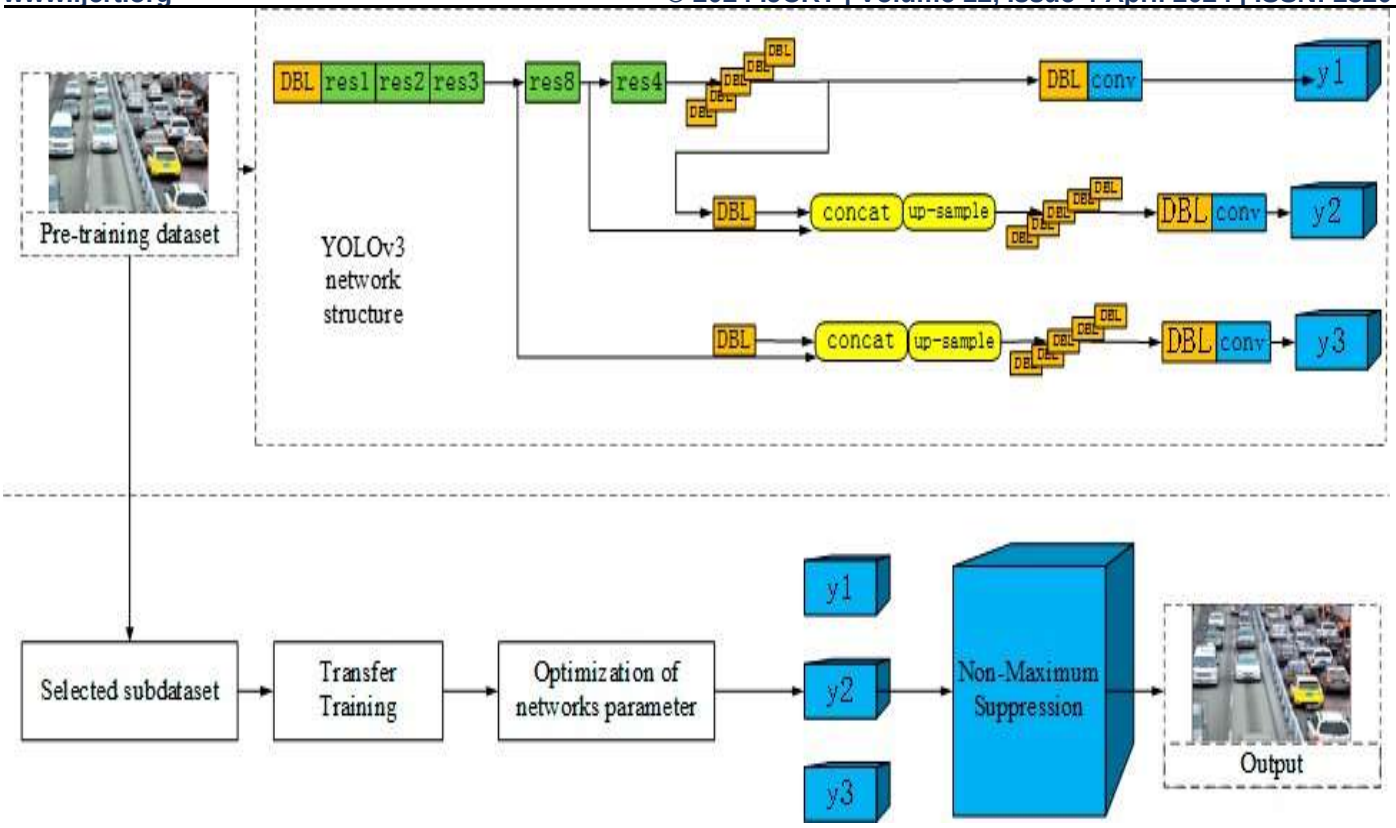


Fig2.Architcture of Detection model

Network Compared with the YOLOv2 network, the backbone portion of the YOLOv3 network has evolved from Darknet-19 to Darknet-53, consequently expanding the number of network layers [19] and adding the cross-layer sum operation in the residual network; YOLOv3 network has fifty-three convolutional layers (ResNet, Residual Network). Darknet-53 is an entirely convolutional network comprised of 3 × 3 and 1 × 1 convolutional layers, including 23 residual modules and layers of detection channels that are completely interconnected. As depicted in Fig 2, the convolutional layers are interconnected by quick link [20]. This SC(Shortcut Connections) structure can greatly enhance the computation performance of the network, enabling the network to obtain faster detection speed in a limited number of network layers. In the detection architecture, YOLOv3 separates three channels for feature detection into distinct grid sizes. These channels include feature maps with grid sizes of 52 × 52, 26 × 26, and 13 × 13, which correspond to the detection of large-scale (y1), medium-scale (y2), and small-scale y3) picture features, respectively. Thereby, The YOLOv3 can provide a higher detection accuracy with fewer network parameters and fewer superfluous network layers, enabling it to improve both the detection speed and the detection accuracy. By comparison, the conventional R-CNN relies on deepening the network structure to enhance the recognition rate.

3.2 YOLOv5 Model

The primary structure of the vehicle discovery and counting system is described in this section. The videotape data from the business scene is the first input. The face area of the road is also removed and resolved. Eventually, to complete multi-object shadowing and gather vehicle business information, Sphere point birth is done on the linked vehicle box. As in Fig. 3, the road face segmentation system is employed to prize the high- way road area. Grounded on the position of the camera, the road is divided into two sections a remote region and a near area. This algorithm can ameliorate the small object discovery effect and break the problem of delicate object discovery caused by a sharp change in object scale.

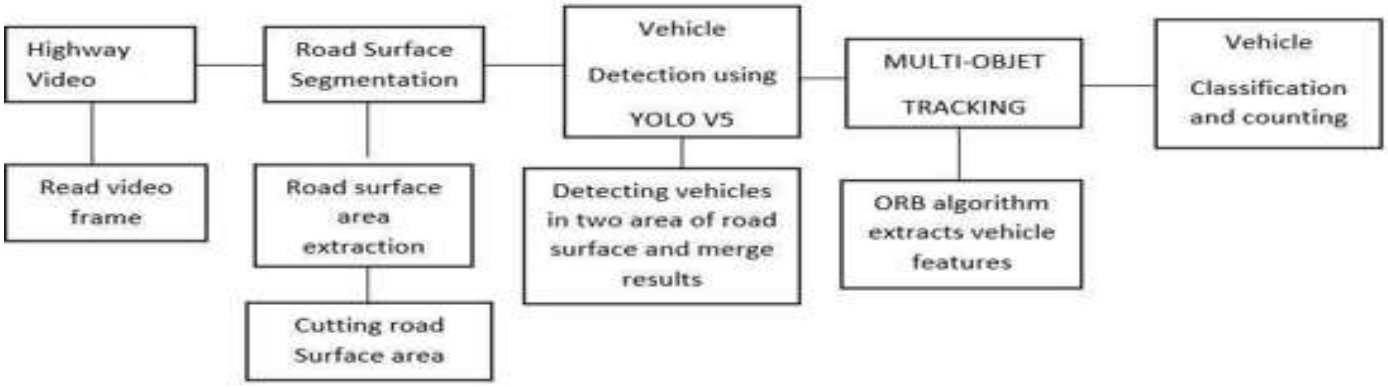


Fig.3.Overall flow of Method

The sphere system is also used to track several objects. To negotiate a correlation between the same item and multiple videotape frames, the Sphere algorithm excerpts and matches the characteristics of the detected box. The object tracking line is created, the vehicle driving direction is established, and business data similar to the number of buses in each order is gathered. From a trace surveillance videotape viewpoint, this system enhances object recognition delicacy and creates a discovery shadowing and business information collection strategy that covers the whole field of view[21].

3.3 YOLOv8 Model

YOLOv8 is the latest version of the object detection model architecture shown in Fig.4, succeeding YOLOv5. YOLOv8 introduces improvements in the form of a new neural network architecture [22]. Two neural networks are implemented, namely the Feature Pyramid Network (FPN) and the Path AggregationNetwork (PAN), along with a new labeling tool that simplifies the annotation process.

This labeling tool contains several useful features, such as automaticlabeling, shortcut labeling, and customizable hotkeys. The combination of these features makes it easier to annotate images for training the model.FPN works by gradually reducing the spatial resolution of the input image while increasing thenumber of feature channels. This results in the creation of a feature map that is capable of detecting objects at different scales and resolutions.Consequently, the network can capture features more effectively at various scales and resolutions, which is crucial for accurately detecting objects of differentsizes and shapes [23].

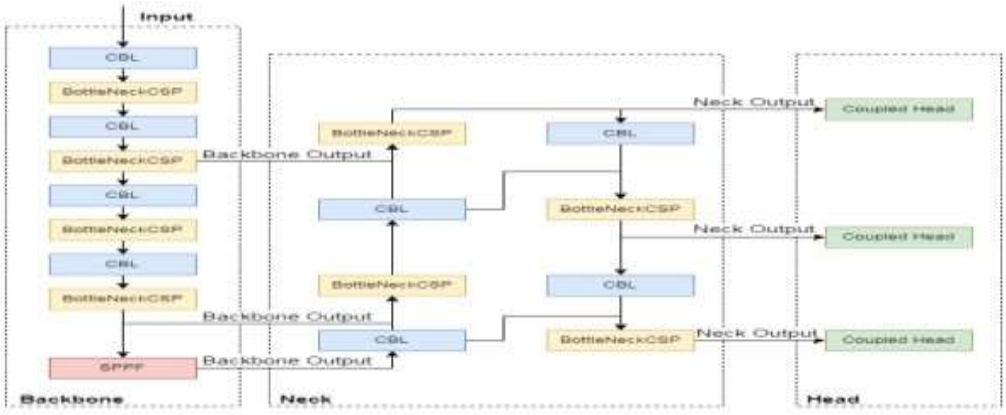


Figure 1. The structure of YOLOv5 [24]

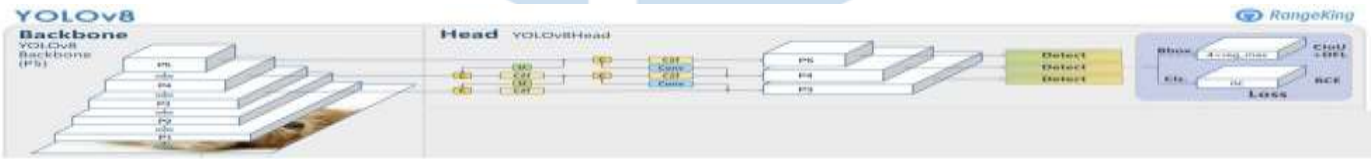


Fig4.Architecture of YOLOv8

4.Experiments

4.1 Dataset

From the perspective of image accession, business image datasets can be divided into three orders images captured by cameras, images captured by surveillance cameras, and images captured by non-surveillance cameras [25]. The KITTI standard dataset [26] contains images from both trace scenes and general road scenes for use in independent driving and working problems similar to 3D object discovery and shadowing. The Tsinghua Tencent business sign dataset [27] has images from cameras covering different lighting and rainfall conditions, but no vehicles are marked. This dataset includes vehicle orders, including the vehicle make, model, and time of manufacture. However, the dataset comprises a large number of images. The 28300 prints show the vehicle's top speed, number of doors, number of seats, relegation, and machine type. The 150,200 prints depict the vehicle's overall look. One illustration is the BIT- Vehicle Dataset[30], which comprises 10,000 prints. This dataset categorizes vehicles into six orders SUV, hydrofoil, minivan, truck, machine, and micro-bus. Still, the firing angle is positive, and the vehicle objects to generalize for CNN training. This section introduces the vehicle dataset from the perspective of the trace surveillance videotape.

As shown in Fig.5, the images have an RGB format and a 1920 x 1080 resolution. It is noted that annotated the lower objects in the proximal road area, and the dataset therefore contains vehicles. An annotated case near the camera has further features, and a case far from the camera has smaller features. Annotated cases of different sizes are salutary to the enhancement of the discovery delicacy of small vehicle objects[24].

Fig.5.Vehicle labeling category of dataset

4.2 Network training and vehicle detection using YOLOv3

YOLO (You Only Look Once) is a popular and general one stage object detection algorithm, which has been developed to the third generation and is called YOLOv3 [38]. The algorithm structure of YOLOv3 is shown in Figure 1.

First a standardized image is used as input to the algorithm. Next the image is divided into S×S grids. Then use these grids to generate class probability map, bounding boxes and confidence score. Finally, the object candidate box with confidence and location is actually output on the image. Object recognition, classification and positioning are transformed into regression problems, which is the core idea of the algorithm. Only one convolutional network is used to predict the classes and location of the object, so as to achieve rapid object detection Shown in fig 6.

In the YOLOv3 algorithm [38], a group of anchor boxes with fixed size are introduced for prediction based on the idea of Faster R-CNN algorithm [39]. There are k initial anchor boxes are obtained by dimension clustering of the height and width of the manually labeled anchor boxes in the dataset through k-means algorithm. For different datasets, YOLOv3 sets 9 initial anchor boxes to obtain 9 clustering results, which are: (10×13); (16×30); (33×23); (30×61); (62×45); (59×119); (116×90); (156×198); (373×326). However, the correlation between general dataset classes and vehicle dataset classes is low. The general dataset is not suitable for real-time vehicle detection. The k-means algorithm randomly selects k initial clustering centre from the sample set, so that causes a low clustering accuracy. From the research, the improved k-means++ algorithm is selected [9]. The threshold ε is set to cluster the anchor box, and the next initial cluster centre is more likely to be selected from a relatively far point. Intersection Over Union (IOU) is the ratio between the prediction result of the anchor box and the intersection area and union area of the ground truth, which has nothing to do with the size of the anchor box. This paper uses IOU as the object cluster analysis of measurement. YOLOv3 uses logistic regression to predict the probability of objects contained in the anchor box.

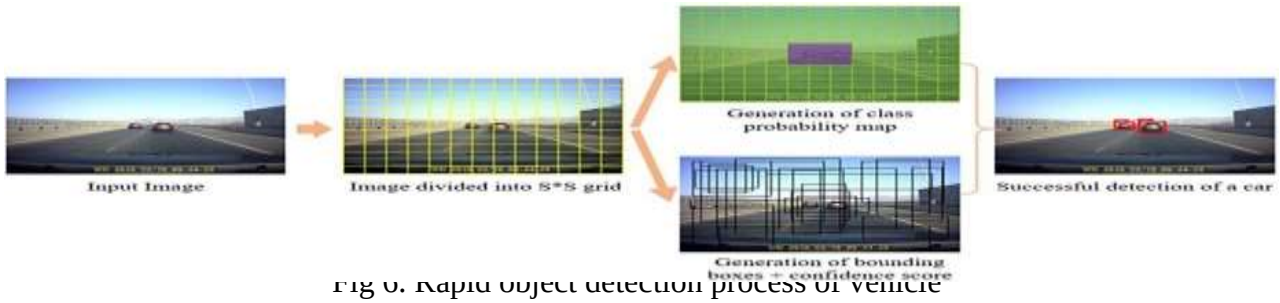


Fig 6. Rapid object detection process of vehicle

4.3 Network training and vehicle detection using YOLOv5

The YOLOv5 network has the advantage of being important faster than other networks of the one-stage family. Also, it achieved similar results to the state-of-the-art and still maintained delicacy, and its prognostications

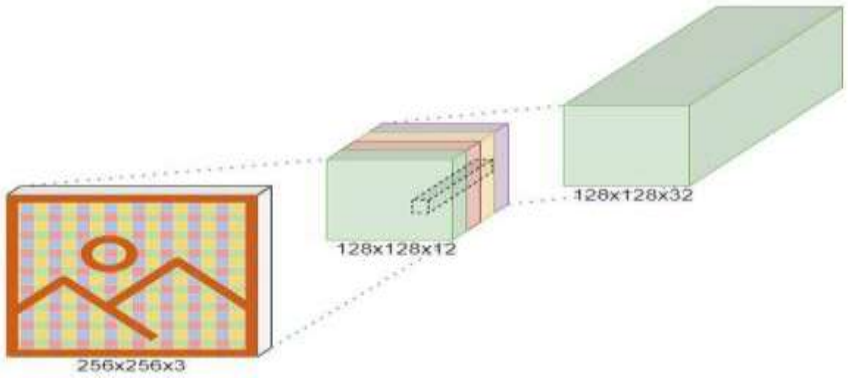


Fig 7 .Standard focus layer of YOLO-v5 model

depended on the global environment of the input image. Accordingly, the proposed model is grounded on the Yolo armature as a birth. The armature of the YOLOv5 network contains numerous layers that connect. The operations performed in each step can summarize the YOLO-v5 network into three different sections. The first section, the backbone (i.e.called csp darknet), is composed, in the case of Yolo-v5, of the most common operations in CNNs (e.g. complications, trains, maximum-pooling) and a simple forwarding medium that's constructed to prize multiple features for the coming section. The concept of the backbone is common content in multiple deep literacy networks of object discovery and is used as a simple base network.

In addition, network copes with the problems of repeated grade information in large-scale chines. It also integrates the grade changes into the point map, yielding a significant reduction in the trainable parameters and floating-point operations per second, which increases the conclusion speed and delicacy. In general, most of the detectors failed to detect objects properly. Thus, the YOLO-v5 model to be more efficient with tiny objects (i.e., vehicles).In the last released version of YOLO is adapted , the Focus layer is one of the important modules of Yolo-v5.The Focus layer first copies the input image size (e.g.,3x256x256) into four copies. The four copies were then sliced into four slices by sampling with a step size of 2 (i.e., 3 ×128 × 128). The four slices are then concatenated in-depth with an output of 12 × 128 × 128, and then passed to the next convolutional layer with 32 kernel filters to generate an output of 3 2 × 1 28 × 1 28, and the result is fed into the next convolutional layer through batch normalization and RELU as an activation function. The focus sub-caste, illustrated in Fig.7 separates the image and transforms spatial information into no qualitative features to mound the RGB information.

4.4 Network training and Vehicle detection using YOLOv8

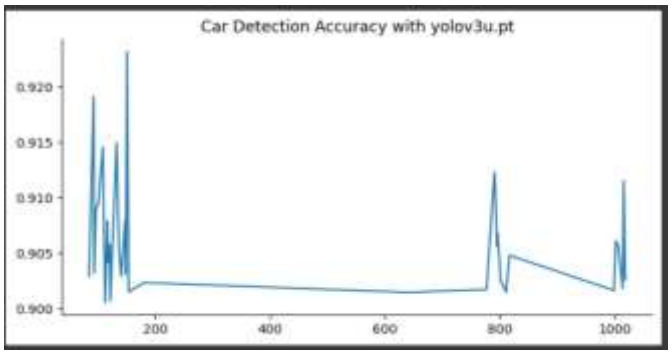
The YOLOv8 model was trained with data belonging to intersection images. Thus, the ability of the model to recognize features such as the dimensions, colors, textures and shapes of vehicles is provided by training. Training and validation processes play an important role in measuring the performance of the model. In addition, these processes are quite challenging in terms of computer hardware. Therefore, the study was carried out on Google Colab, which provides free NVIDIA Tesla T4 GPU support. Various applications and libraries need to be installed to work with the YOLOv8 algorithm.

Fig8. Examples of predictions made by the trained mode

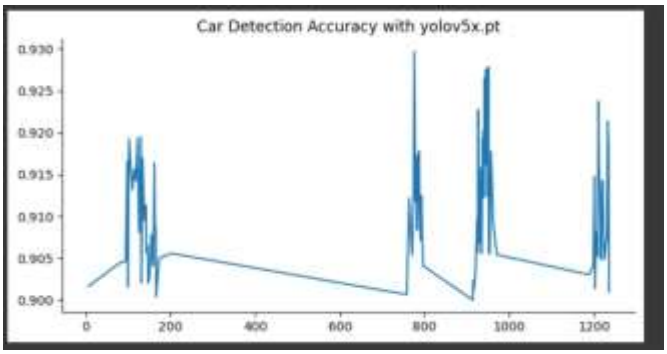
Some of the results of the predictions made by the trained model are shown in Figure 8. Even though no extra improvement was made on the data with an image processing method, the model's performance under different light conditions is quite good.

5.0 Results and Discussion

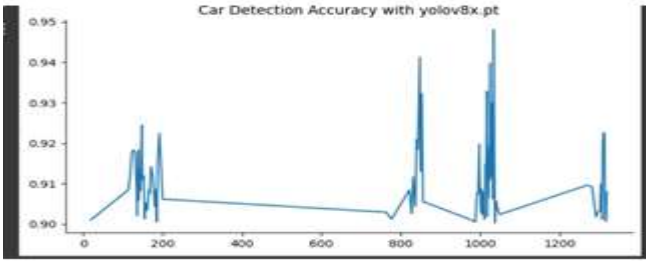
Various vehicles like two wheelers, four-wheelers , heavy vehicles and other vehicles are tested and obtained the results . The Fig. 9 a, b ,and c indicates the accuracies obtained for various YOLO's for same vehicle for reference, similarly, accuracies of the remaining vehicles like bus and truck and tabulated in the table 1 ,2 and 3 respectively. .



a



b



c

Fig.9 a) Detection of accuracy of car for YOLOV3. b) Detection of accuracy of car for YOLOV5. c) Detection of accuracy of car for YOLOV8

5.1.Evaluation of YOLOv3

Table 1:Vehicle accuracies of YOLOv3

S.no	Name of the vehicle	Accuracy(%)
1.	Car	92
2.	Bus	91
3.	Truck	87

It is inferred from the Table1 that Using YOLOv3 Network different types of vehicles (i.e. car, bus and Truck) are to be detected with the accuracy of 92% ,91% and 87%. Respectively.

5. 2 Evaluation of YOLOv5

From the Table2, it is observed that using YOLOv5 Network different types of vehicles(i.e. car, bus and Truck)are to be detected with the accuracy of 93%,92% and 90%.YOLOv5 gives better accuracy when compared with the version of YOLOv3

Table 2:Vehicle accuracies of YOLOv5

S.No	Name of the vehicle	Accuracy(%)
1.	Car	93
2.	Bus	92
3.	Truck	90

5.3 Evaluation of YOLOv8

From the Table3, it is obtained that using YOLOv8 Algorithm different types of vehicles (i.e. car, bus and Truck) are to be detected with the accuracy of 95%,93% and 92%.YOLOv8 gives Best accuracy when compared with the version of YOLOv3,YOLOv5

Table3:Vehicle accuracies of YOLOv5

S.No	Name of the vehicle	Accuracy(%)
1.	Car	95
2.	Bus	93
3.	Truck	92

Table 4. Accuracy comparison of three versions of YOLO's

S.No	Name of vehicle	Accuracy for different YOLO versions		
		v3	v5	v8
1.	Car	0.92	0.93	0.95
2.	Bus	0.87	0.92	0.93
3.	Truck	0.85	0.89	0.94

From the table 4 ,Among all the three versions YOLOv8 gives high accuracy while compared with the other two versions(i.e. YoLOv3,YOLOv5).By using this model only few vehicles are tested and detected and results are obtained**.**

In case of YOLOv3 the vehicle is detected at 800s with 91% accuracy. In the case of YOLOv5 the vehicle is detected at 800s with 93% accuracy. As a result the YOLOv5 is better than the YOLOv3 version. Similarly YOLOv8 is also detected at 800s with 95% accuracy. Hence YOLOv8 version gives more accuracy compare with the YOLOv3,YOLOv5**.**

6.0 CONCLUSION

A novel YOLOv8 algorithm based on the YOLOv3,YOLOV5 algorithm is proposed in this paper in order to improve the real-time accuracy of large-scale vehicles detections. Experimental results prove that the improved YOLOv8 model has a accuracy of 95% and a speed of 59FPS, which are better than the traditional general object detection YOLOv3,YOLOv5 algorithm. Therefore, the YOLOv8 algorithm has better performance and popularization prospect in vehicle object detection

REFERENCE

1. Van Pham H, Lee B-R. Front-view car detection and counting with occlusion in dense traffic flow. Int J Control Autom Syst. 2015 Oct;13(5):1150–60.
2. B. Wu, A. Wan, F. Iandola, P. H. Jin, and K. Keutzer, "SqueezeDet: Unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jul. 2017, pp. 129–137
3. X. Dai, "HybridNet: A fast vehicle detection system for autonomous driving," Signal Process., Image Commun., vol. 70, pp. 79–88, Feb. 2019
4. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2014, pp. 580–587.
5. R. Girshick, "Fast R-CNN," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 1440–1448.
6. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards realtime object detection with region proposal networks," in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2015, pp. 91–99.
7. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 779–788.
8. J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 7263–7271
9. J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, arXiv:1804.02767. [Online]. Available: http://arxiv.org/abs/1804.02767
10. T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 2980–2988
11. https://learnopencv.com/custom-object-detection-training-using-yolov5.
12. Y. Ding, Y. Qu, D. Du, Y. Jiang, H. Zhang, B. Song, X. Zhou, and J. Sun, "Long-distance vehicle dynamic detection and positioning based on gm-apd lidar and lidar-yolo," IEEE Sensors Journal, vol. 22, no. 17, pp. 17 113–17 125, 2022.
13 A. A. Nielsen, "The regularized iteratively reweighted mad method for change detection in multi- and hyperspectral data," IEEE Transactions on Image processing, vol. 16, no. 2, pp. 463–478, 2007. [12] D. Rosenbaum, F. Kurz, U. Thomas, S. Suri, and P. Reinartz, "Towards automatic near real-time traffic monitoring with an airborne wide angle camera system," European Transport Research Review, vol. 1, no. 1, pp. 11–21, 2009
14 J. Canny, "A computational approach to edge detection," IEEE Transactions on pattern analysis and machine intelligence, no. 6, pp. 679–698, 1986.
15. P. Negri, X. Clady, S. M. Hanif, and L. Prevost, "A cascade of boosted generative and discriminative classifiers for vehicle detection," EURASIP Journal on Advances in Signal Processing, vol. 2008, pp. 1–12, 2008.
16. Q. Fan, L. Brown, and J. Smith, "A closer look at faster r-cnn for vehicle detection," in 2016 IEEE intelligent vehicles symposium (IV). IEEE, 2016, pp. 124–129
17. H. Asaidi, A. Aarab, and M. Bellouki, "Shadow elimination and vehicles classification approaches in traffic video surveillance context," Journal of Visual Languages & Computing, vol. 25, no. 4, pp. 333–345, 2014. [17] Q.-L. Li and J.-F. He, "Vehicles detection based on three-frame-difference method and cross-entropy threshold method," Computer Engineering, vol. 37, no. 4, pp. 172–174, 2011.
18. Redmon J, Farhadi A. YOLOv3: An Incremental Improvement. arXiv:180402767 [cs] [Internet]. 2018 Apr 8 [cited 2020 Mar 11];
19. Huang, K.Y.; Chang, W.L. A neural network method for prediction of 2006 World Cup Football Game. In Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, 18–23 July 2010; pp. 1–8
20.]. Oyedotun, O.K.; El Rahman Shabayek, A.; Aouada, D.; Ottersten, B. Training very deep networks via residual learning with stochastic input shortcut connections. In Proceedings of the International Conference on Neural Information Processing, Guangzhou, China, 14–18

November 2017; Springer: Cham, Switzerland, 2017; pp. 23–33. 21. Zhu, B.; Huang, M.F.; Tan, D.K. Pedestrian Detection Method Based on Neural Network and Data Fusion. Automot. Eng. 2020, 42, 37–44

21. Z. Zhu, D. Liang, S. Zhang, X. Huang, B. Li, and S. Hu, "Traffic-sign detection and classification in the wild," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2110–2118

22 J. Terven and D. Cordova-Esparza, "A Comprehensive Review of YOLO: From YOLOv1 and Beyond," ACM Comput Surv, Apr. 2023, [Online].Available: http://arxiv.org/abs/2304.00501

23. H. Liang, J. Chen, W. Xie, X. Yu, and W. Wu, "Defect detection of injection-molded parts based on improved- YOLOv5," in Journal of Physics: Conference Series, Institute of Physics, 2022. doi: 10.1088/1742-6596/2390/1/012049

24. H. Liang, J. Chen, W. Xie, X. Yu, and W. Wu, "Defect detection of injection-molded parts based on improved- YOLOv5," in Journal of Physics: Conference Series, Institute of Physics, 2022. doi: 10.1088/1742-6596/2390/1/012049

25. W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu, and T.-K. Kim, "Multiple object tracking: A literature review," Artificial Intelligence, vol. 293, p. 103448, 2021

26. J. Xing, H. Ai, and S. Lao, "Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses," in 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009, pp. 1200–1207

27. H. Zhou, Y. Yuan, and C. Shi, "Object tracking using sift features and mean shift," Computer vision and image understanding, vol. 113, no. 3, pp. 345–352, 2009

28.] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in 2011 International conference on computer vision. Ieee, 2011, pp. 2564–2571

29. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Trans Pattern Anal Mach Intell. 2017 Jun 1;39(6):1137–1149.