IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, February 2023
Impact Factor: 7.301

Survey of Deep Learning Techniques for Vehicle Detection

Dr. Benila S[1], Karan Kumar R[2], Karthikraja N[3], Kavimukilan M[4]
Assistant Professor, Department of Computer Science & Engineering[1]
UG Scholar, Department of Computer Science & Engineering[2,3,4]
SRM Valliammai Engineering College, Chengalpattu, Tamil Nadu, India

Abstract: *Machine learning techniques have advanced quickly, making it a more crucial tool for object detection. The machine learning-based object detection methods can learn both low-level and high-level picture characteristics, unlike conventional manually built feature-based methods. The machine learning-based image features are more representative than the manually created features. As a result, while the conventional object detection techniques will also be briefly discussed, this review paper concentrates on object recognition algorithms based on machine learning and other deep convolutional neural networks. The following sections of this study are comprised through the review and analysis of machine learning-based object identification algorithms in recent years: traditional object detection architectures, backbone networks, loss functions, and training procedures; difficult issues; datasets; evaluation metrics; applications; and future development. Nowadays, "Unmanned Aerial Vehicles" (UAVs) are utilised for a variety of surveillance purposes. Particularly, due to its potential in applications like traffic control, parking lot management, and simplifying rescue operations in disaster zones and difficult terrains, the detection of on-ground cars from UAV photos has gained substantial attention.*

Keywords: Vehicle, Detection, Machine Learning, Deep Learning, Highway.

I. INTRODUCTION

No Transportation officials are currently searching for more efficient ways to reduce traffic congestion due to the continuous increase in the number of automobiles on the road. There has been interest in Automatic Vehicle Detection (AVD) and traffic information systems for real-time observation and control of traffic in light of this enormously growing amount of traffic. This calls for the collection of exact vehicle data as well as the use of powerful vehicle detection algorithms. For the proper operation of the traffic, constant traffic statistics are required not only in times of crisis (such as when there is traffic congestion or when there is rerouting due to vehicle accidents), but also under regular circumstances.

Locating and identifying things are the two main goals of object detection, and these tasks are accomplished using rectangular bounding boxes. Object categorization, feature extraction, and instance segmentation are related to object detection in various ways. Detecting objects, which includes things like faces, texts, pedestrians, logos, videos, vehicles, and medical images, is a significant topic of computer vision with significant applications in both academic research and real-world industrial production. The creation and use of deep neural networks have been constrained in recent decades due to the limitations of processing power, datasets, and fundamental ideas. Consequently, the conventional object detection techniques were still widely used in the field of computer vision. Traditional object identification techniques include NLPR-HOGLBP, Oxford-MKL, Selective Search, and DPM, among others. The region selector, feature extractor, and classifier make up the core architecture of conventional object identification methods.

Machine learning-based object recognition is becoming more and more popular for applications such as safety in the development of autonomous vehicles and traffic monitoring. The motion of a vehicle is used in conventional machine vision techniques to distinguish it from a stationary backdrop picture. The backdrop details are utilised to create a background model once the video's picture has been repaired. After that, the backdrop model is used to compare each frame image, and the moving item may also be split.

A neural network predicts items in a picture and identifies them using bounding boxes in object detection, a sophisticated type of image classification. Thus, the term" object detection" refers to the identification and location of items in an image

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, February 2023
Impact Factor: 7.301

that fall under one of several established classes. Object detection, also known as object recognition, is a crucial subfield in computer vision because tasks like detection, recognition, and localization have broad applications in real-world contexts. YOLO proposes the usage of an end-to-end neural network that provides predictions of bounding boxes and class probabilities all at once as opposed to the strategy used by object detection algorithms before it, which repurpose classifiers to do detection.

In this article, we give an overview of deep learning methods for detecting vehicles in films, as well as the history and categorization of research projects. It also condenses the main points of various approaches. It talks about the research studies' methods for improving accuracy and considers them in terms of their optimization goals. Additionally, it discusses the methods for minimising computing overhead and concludes by outlining some potential directions for future study.

II. BACKGROUND AND OVERVIEW

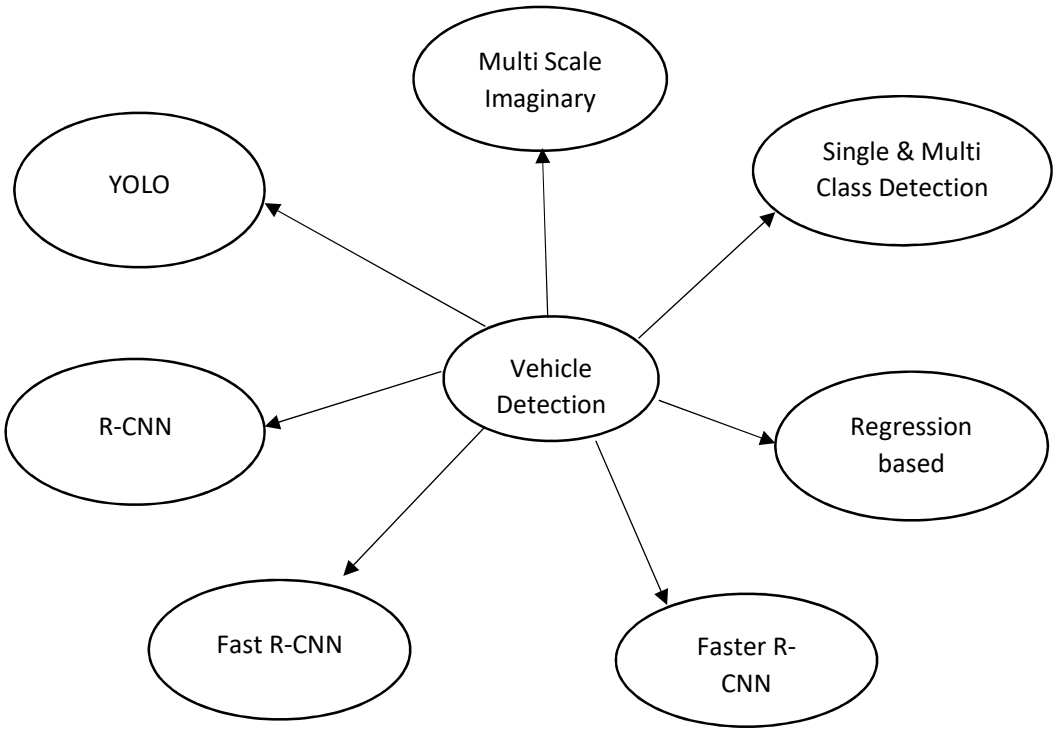

Figure 1: Concept in Vehicle Detection

R-CNN: A bounding box must be created around the target item in order for the object detection technique to succeed. The length of the output layer may vary since the number of objects in various photos may differ. To detect all the target items in an image, a lot of regions are needed because the position of the object and its aspect ratio may vary across photos. The R-CNN algorithm utilises a "selective search" technique to locate area suggestions in order to solve this issue. These RPs are square-wrapped and sent to a CNN so it can create an output feature vector. R-drawback CNN's is that it costs a lot of money to categorise RPs for each image.

Fast R-CNN: Some of the shortcomings of the R-CNN algorithm are addressed with the "Fast R-CNN" algorithm. It provides the input picture to the CNN for creating a CONV fmap rather of feeding it the RPs. This is used to identify RPs and warp them into squares. Once they have been reduced to a predetermined size, they are fed into an FC layer via a RoI pooling layer. The object class is then predicted using a soft max layer. "Fast R-CNN" does not constantly feed RPs to the CNN. Instead, a map is retrieved from each picture after just one CONV operation is carried out on it. Even though "Fast R-CNN" is more effective than RCNN, the employment of RPs still causes a bottleneck.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, February 2023
Impact Factor: 7.301

Faster R-CNN (FRCNN): Both R-CNN and "Fast R-CNN" generate RPs using the "selective search" technique. Instead of using it, the FRCNN technique allows the model to learn the RPs using a different network. The "Fast R-CNN" detector is placed after an RPN in the FRCNN. A fully-CONV network is RPN. The RPN creates a list of possible regions, referred to as region proposals, for a certain input picture. These region ideas are rectangular and come in a variety of sizes and aspect ratios. Additionally, it generates a confidence rating for the existence of the target object in each of these areas. The produced region suggestions and their related box coordinates are ranked using this value.

These area suggestions are improved using the "Fast R-CNN" detector. Using the RoI pooling layer, it projects each area suggestion onto a map with a predetermined size. Every map goes through a series of FC layers before reaching a classification layer and a "box regression layer" for fine-tuning the coordinates of the region recommendations. The backbone network used by the RPN and "Fast R-CNN" detectors is typically VGG16, and both use the same CONV layers. They can be trained together since the layers are shared.

Networks based on regression: Regression-based single-step networks like YOLO and SSD use a single CNN to predict both class probabilities and boxes simultaneously. They use contextual data regarding item classes and appearance to achieve excellent object detection accuracy.

Single-class and multi-class detection: A single-class detection approach finds only one class of items, such as cars. For instance, it may identify both automobiles and buses as the vehicle even if it could detect both. A multi-class detection approach identifies items that belong to numerous classes and labels them as such.

Multi-scale imagery: It alludes to a picture collection that consists of photographs at various sizes. In the context of this essay, it alludes to pictures shot at various heights.

YOLO: The YOLO algorithm divides the image into N grids, each of which has an equal-sized SxS area. These N grids are each in charge of finding and locating the thing they contain. Accordingly, these grids forecast the item label, the likelihood that the object will be present in the cell, and B bounding box coordinates relative to their cell coordinates.

This method significantly reduces computation since cells from the picture are used for both detection and recognition, but it generates a lot of duplicate predictions because many cells may predict the same item with different bounding box predictions. Non-Maximal Suppression is used by YOLO to address this problem.

YOLO suppresses all bounding boxes with lower probability scores in non-maximal suppression. To do this, YOLO looks at the likelihood scores connected to each choice and selects the largest one. The bounding boxes with the biggest Intersection over Union with the current high probability bounding box are then suppressed. Up until the final bounding boxes are achieved, this phase is repeated.

III. KEY IDEAS

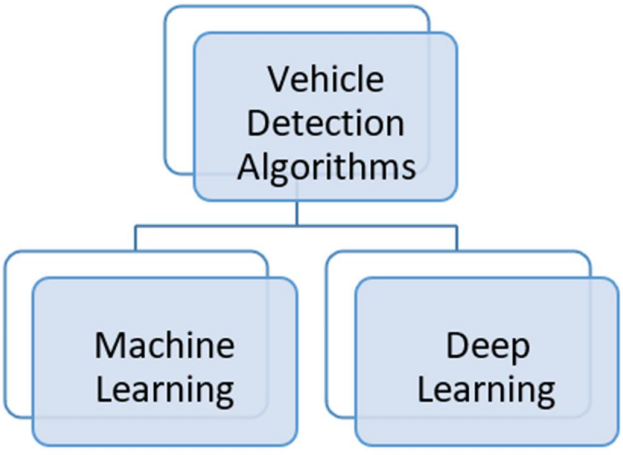

Figure 2: Key Ideas

3.1 Machine Learning

D. Chattopadhyay el al. (2022) [1] proposed a technique based on machine learning for real time vehicle detection. With the creation of bounding boxes around each vehicle, our data can now be utilised to train YOLOv3 and TinyYOLOv3,

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, February 2023
Impact Factor: 7.301

which need high-quality training data. The data was divided into training, validation, and test sets, and four separate versions of the YOLOv3 and TinyYOLOv3 models were trained. Before training, pre-trained weights are used to initialise all model versions. According to the initial YOLOv3 technical report, the pre-trained weights were trained on the Microsoft COCO dataset and stored on the YOLO website. As we just want to identify autos, we modify our model architecture to predict just 1 class as the original YOLOv3 and TinyYOLOv3 models were trained on the COCO dataset, which includes 80 classes.

In order to train the model, several sets of hyper-parameters were used. We train our models using a unique set of hyper-parameters for comparison in addition to the original YOLOv3 and TinyYOLOv3 hyper-parameter configurations. With a score of 55.56%, YOLOv3 Version 2 has the greatest mean average precision, but at the expense of achieving a high detection speed. Tiny YOLOv3, which is YOLOv3's smaller sibling, offers a 3x quicker detection rate.

Y. Ding et al. (2022) [2] says that Military actions have made long-distance object detection technologies a key research issue. It is quite difficult to identify objects at night, especially when there is poor lighting. Traditional sensors can't find things, are impacted by ambient light, and have a limited functioning range. However, the advantages of excellent interference rejection and quick reaction are seen in lidar sensors. Convolutional Neural Networks (CNN) model, one of the most widely used models, has produced positive results in object detection in recent years. One-stage algorithms and two-stage algorithms are two categories into which object identification techniques based on CNN can be subdivided. While preserving high detection precision, the one-stage method performs better in real-time than the two-stage technique.

The process first creates a three-channel 64 64 III picture from the 64 64 single channel intensity image I acquired by GmAPD lidar before training the YOLOv5s network. The YOLOv5s network is then upgraded to the LIDARYOLO network, and the model is trained, all in accordance with the picture data characteristics. Finally, the 64 64 III pictures are converted into the 512 512 BC IIR image by using closest neighbour interpolation, merging distance channel data, and improving image contrast. Our enhanced LIDAR-YOLO network is retrained using this picture to enable the identification and localisation of long-distance dynamic vehicles. The technique may be used to identify other things and enable real-time object identification and location in addition to using it to detect vehicles in military operations. LIDARNet was developed to retain detection precision and boost detection speed in order to meet the goal of real-time detection, and the network's basic structure was further refined. The findings show that the detection precision P is 93.40 and the detection frames per second FPS is 64.77 respectively. The test results revealed that P increased by 0.56% and R increased by 0.13%, indicating an improvement in the network's detection ability for multi-scale object characteristics. Lidar pictures produce better results for vehicle detection than the initial network. Comparing our network against the most recent state-of-the-art detection algorithms reveals how fast and effective it is.

X. Zhang et al. (2022) [3] discusses daytime traffic monitoring, including vehicle classification, detection, and tracking. However, due to the low brightness at night, it is more difficult to detect and track moving vehicles using visible light cameras. At night, typical vehicle attributes like colour, shape, or the licence plate may not be seen. Although numerous image sensors, including infrared and LiDAR cameras, have demonstrated promising performance in a number of applications for monitoring night time traffic, their extensive costs and sparse deployments have hindered their widespread use. The goal of this study is to design an autonomous vehicle recognition and tracking system that uses solely visual cameras because visual camera infrastructures are so common. the location where the specific vehicle light feature is found and tracked. The only noticeable and trustworthy features for vehicle location at night are the headlights and taillights. It is nevertheless crucial to use these observed salient vehicle light attributes to infer vehicle positions at night.

To record traffic dynamics, this system uses two stationary cameras that are positioned at front and rear viewpoints. The front and rear cameras are positioned so that they can primarily capture the vehicle's headlights and taillights. The main goal of this system is to use the stationary rear camera to identify and track various vehicle contours. The inputs to this system are RGB video sequences taken at night from the front and rear views, and the outputs are polygonal rear-view vehicle contour trajectories. Two iPhone 8 devices with 1920 1088 resolutions and 30 frames per second are used to film the front and rear views of the night time traffic surveillance movies. We instal this front-view camera on the side of a motorway so that it may record the majority of the vehicle headlights without occlusion because installing a "front view" camera on the highway would be difficult due to access restrictions. This multi-camera network's configuration is

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, February 2023
Impact Factor: 7.301

displayed. Prior to data acquisition, the cameras are synchronised. We record movies in both sparse and congested traffic environments to show the suggested detection and tracking system's sturdiness. Additionally, two lighting situations, such as low and dark, are individually offered within each night time traffic scenario by modifying the iPhone lens exposure time. These four traffic video sets, each of which include synchronised front and back camera streams, thereby represent our night time datasets.

Four exemplary frames are included in the first two rows, which are arranged sequentially. The first two rows also display the original input frames and their accompanying identified ground facts. The sixth row displays the outcomes of this suggested multi-camera system with high FP no system with all VLI, VLGR, and VCR (no noise). The nearly similar visualisations are not displayed because the VLI only makes up about 1% AMOT poly of the system performance. This multi-camera system with all VLI, VLGR, and VCR achieves higher performance in both vehicle contour identification and tracking, as can be seen.

Y. Fang et al (2019) [4] proposed a vision-based methods utilising an on board camera mounted on the host vehicle have been thoroughly researched for the identification and tracking of nearby cars in the development of ADAS and autonomous driving systems. But in addition to more general challenges like shifting lighting, shadows, background clutter, and a wide range of object sizes and shapes, tracking objects in busy traffic environments is challenging due to frequently occurring changes in appearance brought on by partial observations or shifting target viewpoints. Part-based approaches have received a lot of attention as an alternate strategy since they were first developed. The particle filter's sampling efficiency can be substantially increased while the movement coherence of the various sections can be enforced. Parts with higher confidence will offer more helpful information for monitoring in accordance with their look and deformation cost. A monocular camera was mounted on the roof of the host car. It took pictures of the reference car as it was being driven normally around the host. GPS was used to record the paths of both cars, with the reference vehicle's ground-truth bounding boxes being automatically approximated at timed image frames. Bounding boxes are manually labelled on the sequences of various other vehicles in order to ensure diversity of the data set A part model is made up of a geometric model Mg and an appearance model Ma. Since learning part models is not the main focus of this research, we directly build Ma using the results of and build Mg using the methods. Ma is constructed specifically from the vehicle models for the five most common views, each of which has N = 8 pieces. Structured roads frequently have predictable patterns of traffic. In order to anticipate a vehicle's status more accurately and effectively, prior information of the distributions of vehicle locations and speeds can be very beneficial. Vehicle trajectories on the same road as the test data but on separate days are collected using the system. As explained below, these trajectories are applied to training road models. We can get the centre position distributions of a vehicle from five typical viewpoints by first consistently tessellating an orthogonal road space and then projecting the grid cells onto the picture plane.

B. Xu et al (2019) [5] states that although the COCO data set's item fills more pixels than the automobiles in the aerial picture data set and its background is straightforward and simple to detect, the YOLOv3 framework has produced positive results in this data set. In order to recognise vehicles in aerial images, YOLOv3 cannot be applied simply. The following changes are implemented to make the yolo network more appropriate for aerial vehicles: The YOLOv3 structure has become deeper. By conducting training studies, increasing depth is advantageous to fitting aerial vehicles. Increasing network depth indiscriminately can lengthen computation times yet have unclear effects. Finally, there are 75 convolutional layer networks. Because aerial vehicle pixels are so small and feature information is shrinking during convolution, under-level feature maps have fewer details about the vehicles, so we aim to call high-level information to improve the effect. In comparison to other algorithms, our approach's precision and recall are greater by 2.1% and 0.3%, respectively, than the best results, and it is also higher by 1.34% than the most advanced vehicle recognition algorithm. Aerial vehicles can be found with the detector.

P. Kumar et al. (2021) [6] states a theory which gave a total clarification of utilizing PC vision procedures in vehicle identification and counting framework utilizing sensor security. To get the reason, first and foremost, they utilized the information on deduction procedure to discover frontal area objects for each situation in the video. Another, to finish vehicle disclosure and then some, examination assembled a top-quality vehicle object dataset from the possibility of observing cameras and moved an objective disclosure and following investigation for street checking video scenes. A more proficient ROI the region was gotten by the extraction of the street surface space of the thruway utilizing senor security crossing points, locus hither are deuce badly known strategies, specifically foundation deduction and optical

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, February 2023
Impact Factor: 7.301

stream, and not to be utilized for the identification of a vehicle to stop at a red sign. All things being equal, the three similitude estimations of the space of the finder can be utilized to recognize the presence of coaches in the two aforementioned edges. For a right examination was performed to decide the quantity of units in a sensor field.

An advanced camera has been introduced on the scaffold, and to gather in a data set, by the best go away to illustrations. Seven of the information recordings, of the expressway, around three to seven minutes long, were recorded throughout the span of 10 short period of time, if 2 hours beyond the day. The spatial goal of the recorded recordings is 1280 x 720 pixels goal at 30 casings per-second the extemporization has been executed in C++ regardless of the OpenCV playground. A test was directed to gauge the precision of vehicle identification. If a sign is gotten on the proposed technique, the quantity of vehicles distinguished by the proposed approach are shown and contrasted and the outcomes got by manual counting. This recommends that the eye-opening strategy can misfunction admirably in each of the tried videos. An aggregate of 646 man with 667 at every one of the seven of the mistreatment recordings have been found, bringing about an exactness of any of 96.85%.

Object detection algorithms in Machine learning:

Detector	Number of n layers	FLOPS	FPS	Map Value	Set of Data Used
SSD300	-	-	46	74.3	COCO-data
Fast R-CNN	4	-	0.5	68.4	COCO-data
Faster R-CNN	4	-	7	73.2	COCO-data

Table 1: Machine Learning Techniques

3.2 Deep Learning

Song et al. (2022) [7] sates the methods for object detection employed in this work are described by this theory. The YOLOv3 network was utilised in the highway vehicle detection framework implementation. The fundamental concept of the prior two generations of YOLO algorithms is carried over into the YOLOv3 algorithm. To extract the features, a convolutional neural network is used. Traditional machine vision techniques and sophisticated deep learning techniques are the two main categories of vision-based vehicle object recognition. The motion of a vehicle is used in conventional machine vision techniques to distinguish it from a stationary backdrop image. Two or three consecutive video frames' worth of pixel values are used to determine the variance.

The vehicle object dataset created in the "Vehicle dataset" section was used to test the performance of the methods described in the "Methods" section. The gradient fell reasonably with this strategy, and the loss value decreased. We modified the default anchor box using the k-means++ approach to make it better suited for the dataset annotation box that would be annotated. The highway's road surface area was extracted to produce a more productive ROI area. Based on the annotated highway vehicle object dataset, the YOLOv3 object identification algorithm developed the end-to-end highway vehicle detection model.

Yan Zhou et al. (2022) [8] proposed a model to improve the tracking algorithm's capacity for feature extraction, a deeper network structure is used. To replace the previously introduced candidate area generating approach, split the grid candidate area using the YOLOv2 detection algorithm. This technique can enhance the quality of candidate region generation and enhance the precision of the tracking algorithm, much like the detection algorithm. The YOLOv2-tracker tracking algorithm, which forms the foundation of the first and second points of the aforementioned improvements, is based on the YOLOv2 detection algorithm. The coincidence rate is used to gauge the tracking method's accuracy. The spatial robustness evaluation approach is utilised since there is only a slight change in the geometry of the vehicle during the tracking procedure. When the target car in some movies departs the video screen, the data set gathered for this research to test the tracking algorithm's performance shows that the video is still playing. At this point, the tracking algorithm with the best performance should not mark the "target" in the ensuing video images. Therefore, the method of evaluating the accuracy of the coincidence rate must be modified appropriately in order to evaluate the accuracy of the algorithm. Specifically, when the target leaves the video screen, if the algorithm does not predict the "target," it means that the current frame is successfully tracked; otherwise, it is not. In the test, the algorithm successfully "followed" the target object in a frame of images, meaning it does not When the algorithm "follows" the target, the method will fail because the "blackening" of the current frame necessitates the algorithm to forecast the position of the target vehicle in the current

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, February 2023
Impact Factor: 7.301

frame. if it is determined that the algorithm "follows," or does not predict, the position of the target vehicle in a certain frame. Use the previously saved "blackening" processing mark map to swap out the current frame's "blackening" map. The YOLOv2tracker tracking algorithm flow is explained in detail, single and dual network frameworks are suggested, and the effectiveness of the two networks is tested.

S. Zhou et al. (2022) [9] uses traditional vehicle detection algorithms and deep learning-based vehicle detection algorithms are the two main categories into which vehicle detection systems may be separated. The KITTI dataset is used to evaluate the performance of the YOLOv5-GE model with those of five other target detection networks that are more widely used, including the Faster RCNN, SSD, YOLOv3, YOLOv4, and YOLOv5. Among these, SSD, YOLOv3, YOLOv4, and YOLOv5 are one-stage detection algorithms, whereas Faster RCNN is a two-stage detection technique.

The goal of this ablation comparison experiment was to demonstrate the effectiveness of each improvement module's optimization. According to the data in the table, adding the GAM attention mechanism increases average accuracy by 1.9% but decreases speed by 1ms. The loss function is changed to EIoU, and the inference time is 2ms faster than YOLOv5 despite the accuracy not improving significantly. The average accuracy increases by 1.4% when these two improvements are combined into the model, while the inference time stays the same. Experimental evidence shows that the YOLOv5-GE algorithm can currently recognise vehicles accurately in difficult situations and has the greatest overall performance of all compared models.

To fulfil the requirement for vehicle recognition accuracy and real-time performance in complicated settings, the YOLOv5-GE model outperforms the YOLOv5 model in recognising both tiny and dense targets while guaranteeing the same detection speed. The model may be modified to detect additional kinds of targets and has high generalisation capabilities. The model will be further optimised in the following phase of the research to improve detection performance under the relevant complicated and changing environmental weather conditions while maintaining speed accuracy.

Yin G, et al, (2022) [10] says that foreign academics have proposed the ITS (Intelligent Traffic System), which uses it to actualize the close cooperation of people, cars, and roads, in order to improve transportation efficiency, relieve current traffic congestion, and decrease traffic accidents. All of this is predicated on the ability to efficiently detect and precisely monitor moving vehicles on highways in real time, which is also the central idea and essential technology of the intelligent high speed transportation system. The accuracy of video detection in a perfect environment has increased along with image processing technology. Traditional video-based vehicle detection algorithms, however, have poor detection accuracy and sluggish response times in complicated surroundings, making it challenging to meet actual needs. People have discovered that convolution neural networks are ideal for target recognition thanks to deep learning research. Several traditional object detection algorithm models are used on various datasets to evaluate the model's validity. The testing findings demonstrate the model's effectiveness, but there is a problem with missing detection for some small-scale targets and targets with backdrop clutters.

Computer vision tasks include image categorization tasks, and there are frequently instances where there are more images in certain categories and fewer in others. For instance, during training, the ratio of positive and negative anchors in the target detection method employing the anchor mechanism may be drastically different in the detection job. In essence, it can be categorised as a category imbalance problem; in addition to the category imbalance problem, the classification of picture anchor pixels also faces a straightforward sample overpowering problem. The effective gradient could be diluted if there are too many simple samples. These two issues frequently coexist, and if they are not effectively resolved, the model's performance will suffer. The approach is to focus on the challenging samples during the model training process, which increases the model's capacity for recognition and raises its performance in terms of detection. By including a difficult sample mining strategy in the training phase, perform target detection on challenging samples (which have no targets) in the initial training process, and collect all detected rectangular frames, then use the samples the model incorrectly classified to expand the difficult sample set, and finally add the difficult sample set to the training set to retrain the new model.

M. Umair Arif et al. , "A Comprehensive Review of Vehicle Detection Techniques under Varying Moving Cast Shadow Conditions Using Computer Vision and Deep Learning" (2022) [9]

M. Umair et al, (2022) [11] states a system where computer-assisted study of vehicular traffic was performed by researchers using ITS assistance for automated monitoring and control. The studies are often carried out using video feeds from security cameras that have been placed in various traffic zones. These cameras' data allow for easier scene

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, February 2023
Impact Factor: 7.301

comprehension in terms of things like vehicle detection, classification, and re-identification. Numerous practical uses exist for this, including automatic number plate recognition, queue estimation, speed detection, and the detection of various anomalies, such as traffic bottlenecks and accidents. It is necessary for vehicles to be successfully detected in each frame in order to obtain correct information about them. Using a continual stream of photos, some traditional techniques include background subtraction or moving object detection. Self-shadows and thrown shadows are the two different kinds of shadows. Cast shadows fall on the ground or any other surrounding object, while self-shadows fall on the object itself when it blocks light from a light source. Self-shadows and cast shadows are distinguished in Figure 3 by the use of red and blue borders, respectively. Because both forms of shadows are similar in nature, it is frequently challenging for algorithms to tell them apart.

There are two further categories for statistical approaches: parametric and non-parametric. The three different sorts of parameters, namely spatial, spectral, contextual, and temporal, may all be used in parametric techniques. The spatial parameter specifies whether a single pixel value will be used to create the feature vector or whether a region or frame of many pixels will be utilised to collect higher order statistical data, such as marginal in terms of means and central moments of rows, etc. Frequency data is discussed in the spectral parameters. For problems involving object detection and image classification, classical methods and machine learning techniques have demonstrated distinct performances. Traditional deep learning architectures (such Decision Trees, SVM, K-means, and Naive Bayes, among others) as well as cutting-edge deep learning architectures (like ResNet101, VGG16, and R-CNN). Supervised learning and unsupervised learning are the two main methods used in machine learning. They differ significantly in terms of labelled data. Unsupervised learning and supervised learning are two categories for learning-based approaches.

Vehicle classes have been predicted using the YOLOv5 model both with and without shadow removal. The evaluation of the outcomes is provided below. The majority of the vehicles' classification accuracy has increased using the front view dataset (see Figure 7 a), but the accuracy of the automobile in the front has remained constant at 82%. However, after the shadow was removed, the rickshaw, which was originally classified as a truck, was wrongly identified as a car. The discussion of prior contributions in this area allows for a comparative assessment of moving cast shadow detection techniques. To provide a more comprehensive overview of the work done in this field, a total of 70 articles over the last three decades that contain results of urban traffic scenes have been shortlisted. Existing approaches for detecting moving cast shadows are grouped. Benchmark datasets utilised for the specific conditions relevant to traffic analysis are supplied, along with a definition of cast shadow characteristics. In contrast to the pre-trained models, the YOLO model can be investigated to generate more accurate results. The research project can potentially be expanded by utilising even more complicated datasets with significant weaker shadows.

K. Guo et al, (2021) [12] states that there are more automobiles on the road, they're moving faster, the effectiveness of maintaining the roads is declining, and there is more traffic. to gradually address the aforementioned issues and lessen the harm done to society. Researchers are working to develop a method for real-time vehicle recognition and management since the technology for artificially intelligent vehicles that aid with driving offers a new approach. Intelligent transportation systems were made possible by the enormous advances in deep learning, image processing, and other study areas in machine vision. The intelligent transportation system is frequently utilised in autonomous vehicles, intelligent transportation, and other areas. Sensor production is expensive and cannot be done on a large scale. From the perspective of computer vision, the employment of algorithms for processing video data is acceptable as long as they are effective at increasing detection efficiency and accuracy. such that it can be applied widely. With the help of its numerous feature maps, SSD is able to accurately anticipate objects of various resolutions and perform multi-scale object identification, but it is still unable to recognise small objects with adequate accuracy.

The issue of real-time vehicle object detection was investigated. The driving recorder's real-shot video was used as the data set for the detection algorithm in order to imitate the genuine traffic scenario in the city. Through data augmentation methods, the variety of data was enriched. The standard and fixed learning rate was replaced with the learning rate adaptive adjustment technique, which was then used during the model training phase to create a vehicle multi-object detector. When compared to the conventional detection technique, the upgraded SSD detection algorithm performed better. The testing findings demonstrated that the enhanced SSD detector has a detection time of 55.6 MS, which is faster than other conventional detection techniques at the same resolution. The issue of real-time vehicle object detection was investigated. The detection result is good, especially for multi-object, faraway tiny object, and overlapping objects. This

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, February 2023
Impact Factor: 7.301

overcomes the drawbacks of classic detection techniques, such as false detection and missing detection, and satisfies the real-time detection requirements of practical application.

Detector	Number of n-layers	FLOPS	FPS	Map Value	Set of Data Used
YOLOv2	32	62.95	40	48.20	COCO-data
YOLOv3	106	140.70	20	57.80	COCO-data
YOLOv5s	-	17.1	113	36.9	COCO-data

Table 2: Deep Learning Techniques

IV. FUTURE WORK

In upcoming work, a YOLO (You Only Look Once) model will be developed to detect the vehicles present on the highway and classify them as car, bike, truck, or bus based on the video. The distance will be calculated to know how far the vehicle is from the user. The distance will be calculated to know how far the vehicle is from the user. Then the lane of that vehicle will be tracked to know that whether the user car and the vehicle are travelling on the same lane.
Other benefits of using YOLOv5 are as follows:

- It is about 88% smaller than YOLOv4 (27 MB vs 244 MB)
- It is about 180% faster than YOLOv4 (140 FPS vs 50 FPS)
- It is roughly as accurate as YOLOv4 on the same task (0.895 mAP vs 0.892 mAP)

V. CONCLUSION

A computer that does vehicle detection identifies whether a vehicle is present in each picture or video and pinpoints its location. There are still numerous issues with vehicle detection that need to be resolved in the real operation, such as occlusion, illumination, and object form changes. This study analyses the accuracy and real-time performance of various algorithms for vehicle recognition from the perspectives of deep learning and vision.
The tiny object, occlusion, and real-time capabilities of the existing vehicle identification system need to be enhanced despite its relative maturity. Research in several domains is still focused on how to leverage the current vehicle detection technologies to identify small item vehicles. Future automobiles will become more intelligent and practical as the foundational sciences of artificial intelligence and visual computing theory advance, along with sensor technology advancements, cost-performance improvements, and vision-based vehicle identification and tracking technologies

REFERENCES

[1]. D. Chattopadhyay, S. Rasheed, L. Yan, A. A. Lopez, J. Farmer and D. E. Brown, "Machine Learning for Real-Time Vehicle Detection in All-Electronic Tolling System" (2022).
[2]. Y. Ding et al., "Long-Distance Vehicle Dynamic Detection and Positioning Based on Gm-APD Lidar and LIDAR-YOLO" (2022).
[3]. X. Zhang, B. Story and D. Rajan, "Night Time Vehicle Detection and Tracking by Fusing Vehicle Parts From Multiple Cameras" (2022).
[4]. Y. Fang, C. Wang, W. Yao, X. Zhao, H. Zhao and H. Zha, "On-Road Vehicle Tracking Using Part-Based Particle Filter" (2019).
[5]. B. Xu, B. Wang and Y. Gu, "Vehicle Detection in Aerial Images Using Modified YOLO," (2019).
[6]. P. Kumar and S. Sharma, "A Computer Vision Based on Vehicle Detection and Counting System Using Sensor Security," (2021).
[7]. Song, H., Liang, H., Li, H. et al. "Vision-based vehicle detection and counting system using deep learning in highway scenes" (2022).
[8]. Yan Zhou, Jun Zhou and Fangli Liao "Research on Vehicle Tracking Algorithm Based on Deep Learning" (2022).
[9]. S. Zhou, Y. Zhao and D. Guo, "YOLOv5-GE Vehicle Detection Algorithm Integrating Global Attention Mechanism" (2022).
[10]. Yin, G., Yu, M., Wang, M. et al., "Research on highway vehicle detection based on faster R-CNN and domain

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, February 2023
Impact Factor: 7.301

adaptation" (2022).

[11]. M. Umair Arif, M. U. Farooq, R. H. Raza, Z. U. A. Lodhi and M. A. R. Hashmi, , "A Comprehensive Review of Vehicle Detection Techniques under Varying Moving Cast Shadow Conditions Using Computer Vision and Deep Learning" (2022).

[12]. K. Guo, X. Li, M. Zhang, Q. Bao and M. Yang, "Real-Time Vehicle Object Detection Method Based on Multi-Scale Feature Fusion" (2021).