ISSN: 0974-5823 Vol. 7 No. 2 February, 2022

International Journal of Mechanical Engineering

Comparison study about Car Detection and Type Recognition Techniques

Asst. Prof. Dr. Ali Abdulazeez Mohammedbaqer Qazzaz[1] and MSc. student Ahmed Younus Abdulkadhim[2]

1,2 Department of Computer Science, Faculty of Education, University of Kufa, Najaf, Iraq

2 University of AL-kafeel, Najaf, Iraq

Abstract— **Vehicle detection and identification technology plays a vital effect in some intelligent systems for transport nowadays, and this field gets great interesting efforts depending on some related factors like the enhancement of important algorithms for processing image, deep learning and recognition of desired Pattern, vehicle detection and vehicle type prediction technology based on learning algorithms get raised increasing concern that play a vital effort in developing these algorithms. This paper introduces some methods for detecting and automatically telling the vehicle types depending on learning algorithms like CNN techniques, versions of YOLO and versions of SSD, which used in detecting desired objects in real-time environment, and have relatively accepted efficiency. Selected algorithms will be analysed in some manner to introduce the main advantages and drawbacks of each method by the calculating of the performance metrics for each one of them to determine the best model for vehicle type recognition.**

Keywords— Convolution-NN, Vehicle detection, classification, Deep learning, Yolo.

I. INTRODUCTION

Vehicle detection is the base for the execution of a huge types of critical functions. Nowadays, vehicle identification, distance measurement, and statistics on traffic characteristics play a critical role in a variety of applications including traffic flow and density, vehicle location and monitoring, and traffic data mining.[l]

Three different types of car detecting algorithms exist:

- Algorithms depend on motion,
- Algorithms depend on features,
- Algorithms depend on convolution networks.

The process of subtracting Frames in optical flow, as well as the process of subtracting backdrop are the foundation of all techniques depending on motion. This approach is characterized by straightforward computation, and it is not recommended in motions which recognized excessively as rapid or excessively slowly and time-consuming.

Feature-based techniques developed by hand include the techniques like (HOG) Oriented Gradients Histogram and (SIFT) scale invariant feature transform. Before the great success of techniques that based on convolution network, approaches based on features such as (DPM) the models based on deformable part. These techniques have a low representation of features.[2]

Object recognition and tracking technology is highly successful and used many important applications such as self-driving automobiles, medical diagnostics, ball tracking in sports, and video surveillance systems, among others. Object detection techniques locate items inside a digital picture or video frame and surround them with a bounding box labelled with the object's class.

Advances in deep learning and neural networks have resulted in the development of cutting-edge models such as R-CNN[3], YOLO[4], and SSD[5], among others.

Convolutional Networks are a subclass of NNs that are often employed for processing image and object identification. As indicated in image, a NN is a network that collect neurons layers (1), it acts similarly, to how the human brain does. Techniques based on Convolution-NN shown a considerable degree of representational power and have produced perfect outcomes. R-Convolution-NN prepare CNNs for detection operations by using proposals object produced by selective search. Fast- R-Convolution-NN improve performance within the R- Convolution-NN models by creating proposals region on the feature map; these approaches need the processor just once. R-CNN trains faster and more accurately when it employs region proposal networks rather than selective search. R-FCNN utilizes position-sensitive score maps to attempt to minimize calculation time. Due to its great efficiency, the one-stage technique has garnered considerably more attention in recent years. YOLO identifies desired object classes and positions directly using a single convolutional network, which is highly fast. SSD retrieves anchors with a variety of aspect ratios and sizes from a collection of feature maps. It is capable of producing competitive detection findings and operating at a faster rate for example, SSD runs at 58 frames per second on an NVIDIA X with 300× 300 inputs, which is nine times faster that Faster R-CNN.

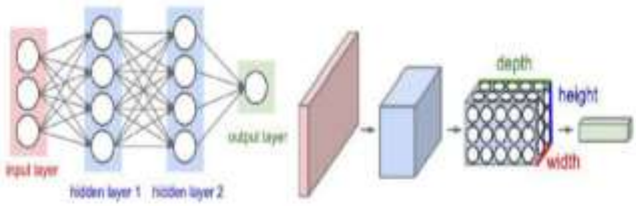


Fig. (l) Neural network compare to Convolution Network Model.

The regular neural network consists of

- input layer,
- hidden layer and
- output layer in one dimension,

while Convolution Network contains neurons in 3- dimensions. As illustrated in figure(1). [7]

II. **OBJECT RECOGNITION BASEDON ON DEEP LEARNING**

The ultimate objective of machine learning and computer vision is to develop technique for recognizing objects can detect the existence of an item in a picture. Previously conducted object recognition research has included the selection of object characteristics on the basis of human knowledge the detection of items depending on their characteristics, as depicted in figure (2).[8]

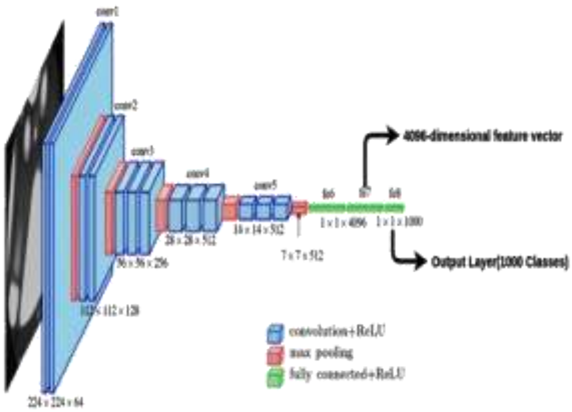
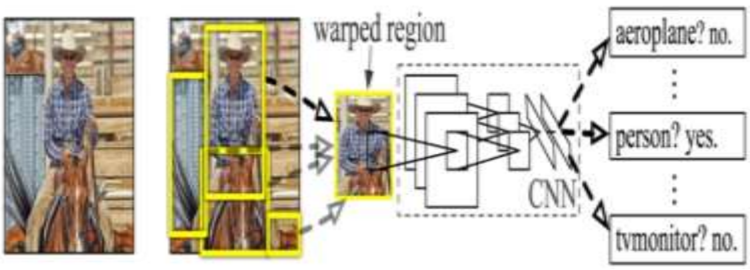


Fig. (2) Convolutional Neural Network (CNN)

Several studies on object identification based on deep learning have been undertaken, and large number of models with accepted performance have proposed. Multiple object recognition techniques have been created. Additionally, research is being conducted to build improved systems capable of identifying things using just a portion of them. Numerous deep learning-based object identification systems are available, including R-C-NN, Fast R-C-NN, Faster RCNN, YOLO models, and SSD versions. [9] As illustrated in Figure (3), two steps make up the R-CNN detector: Regional categorization and proposal. To begin, the detector uses Selective Search to locate 2,000 boxes that correspond to the area of the target item. Then, all Bounding Boxes are classified using CNN. As the cost of calculation grows, the processing speed slows.



a.input image   b-extract region   c-CNN features   d. classify

Fig. (3) R-CNN

Faster RCNN compensates for processing speed by performing after passing through CNN, detect objects once in the output feature map. As seen in Figure (4), a faster RCNN overcomes the bottleneck created by the selective search method by using the Region Proposal Network. Faster R Convolution Network is more than hundred times faster than R-Convolution Network when processing speed is compared. The downside of the R-CNN series' object detection technology is its poor processing speed, which makes it incompatible with real-time applications. [l0]

All prior object identification techniques relied on areas to determine the item's position within the image. The network does not consider the entire image, rather than that, some areas are contain the object. (YOLO) You Only Look Once is a radically different approach for object recognition than the region-based algorithms previously mentioned. YOLO predicts output boxes and their class probabilities using a single NN. YOLO models divide a desired image into grids of size (S x S) grids and then generates a number of bounding boxes (BBs) inside each grid, for each BB, the network outputs the probability of the class and the value of offset. The BBs with probability larger than or equal to a predetermined threshold are isolated and used to determine the wanted object in the input image. YOLO is orders of magnitude quicker than conventional object detection algorithms (45 frames per second). The YOLO algorithm's weakness is that it has difficulty recognizing tiny objects inside a picture, as demonstrated in figure (5). [4]

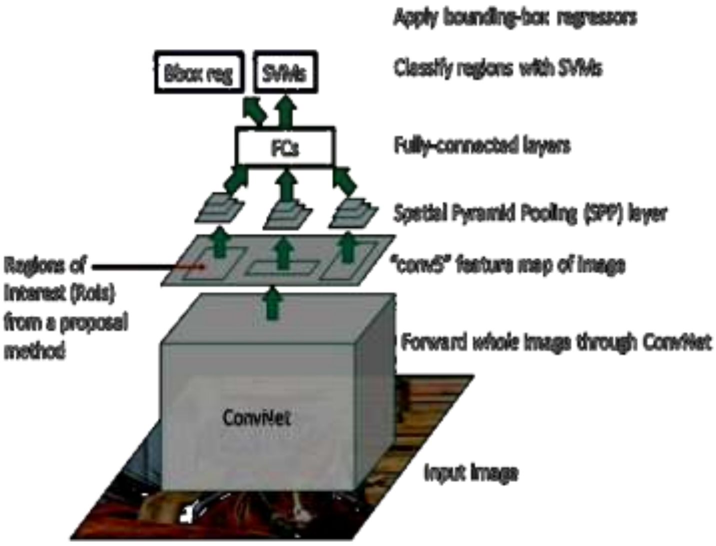


Fig. (4) Faster R-CNN

Unlike the other region proposal algorithms, such as Faster R-CNN, which require two distinct processes to identify items, one to estimate region proposals and another to detect objects within each region proposal, SSD detects a large number of objects in an image in a single shot, as illustrated in figure (6). It creates feature maps through the use of the VGG-16 model architecture. A (3× 3) convolutional filter is used to compute the bounding boxes and class scores of objects using these feature maps. SSD makes use of precisely scaled feature maps to recognize both large and small things in a photograph. Low-resolution feature maps may suffice for larger things.

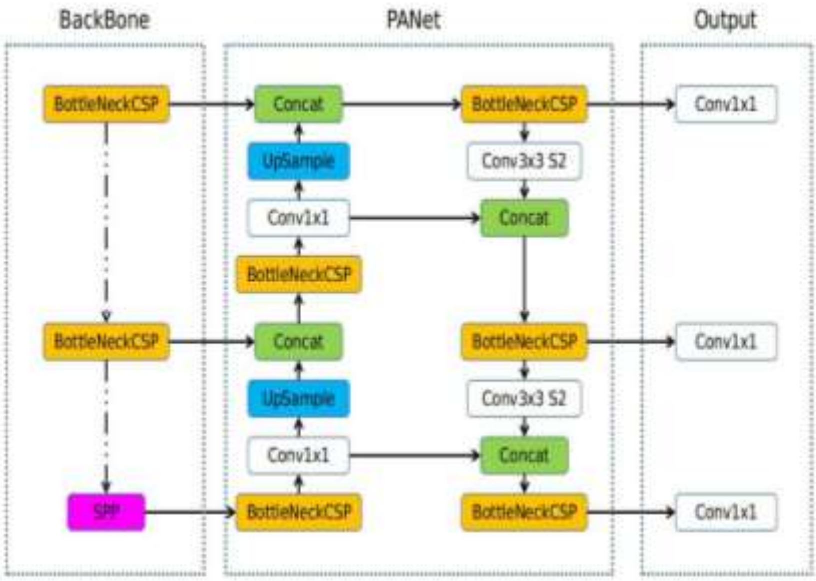


Fig. (5) Yolo v5

However, high-resolution feature maps would be necessary for distinguishing small things. The default bounding boxes have been carefully chosen in advance to encompass a diverse range of real-time objects. The various resolution filters make use of rectangular boxes with differing aspect ratios (1, 2, 3, 1/2, and 1/3). To establish an object's ideal bounding box, a technique called matching strategy is used, this approach stipulates that a default box with IOU larger than a certain threshold (0.5) in respect to the ground truth is a positive match. [11]

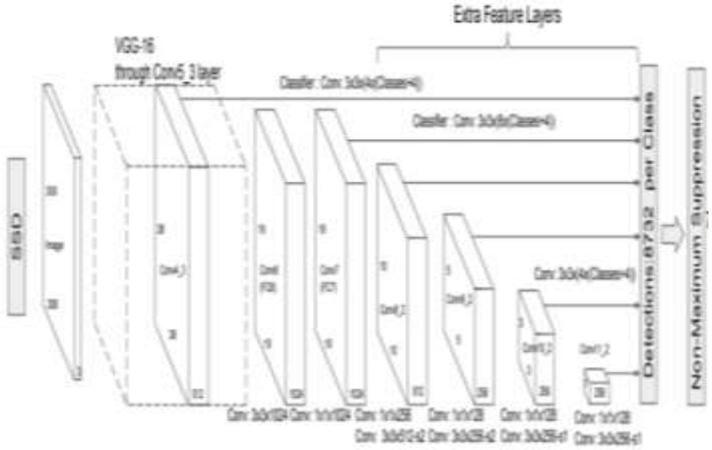


Fig. (6) SSD model

III. **METHODS**

Xinchen Wang, et. al. In [21] introduce a system of vehicle type prediction by using Faster R-CNN as illustrated in figure (7) and table (1). The experimental findings demonstrate that the approach is not only faster, but also more resilient and accurate. Aimed against automobiles and trucks, it achieved an accuracy of 90.65 and 90.51 percent, respectively. Finally, they said that they tested the system on an NVidia Jetson TKl board equipped with (192 CUDA) cores, which is simulated the computing brain of the future in the computer vision techniques, and automatic-driving automobiles. The results indicate that it takes around (0.354) seconds for recognizing image and maintains good rate of accuracy while using installed network on the NVDIA Jetson TKl.[12]

On the basis of the dataset setup, a categorization system for vehicle types was examined. The experiment was conducted using a 64-bit Ubuntu 14.04 system powered by an CPU E5-2630 v3, (64) GB RAM and NVIDIA GTX (1080)GPU.

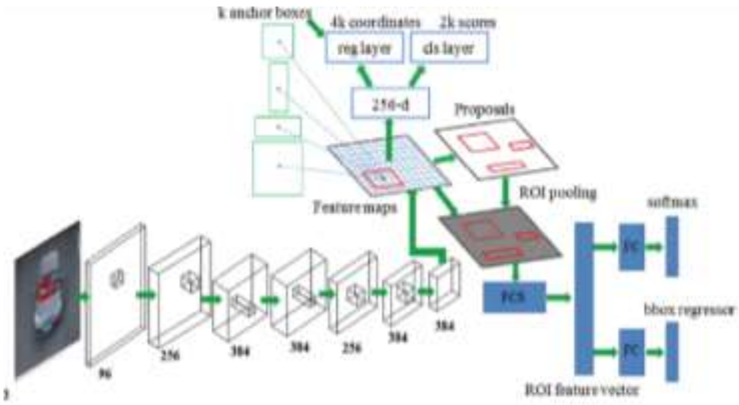


Fig. (7) proposed system for method (l)

The training set contains 37,578 randomly selected sample photos, (l5000) images of autos, (13698) images of trucks, (4805) images of minivans, and (4075) images of buses.

TABLE (1)

PARAMETERS OF CONVOLUTIONAL LAYERS

Names of layers	Input image	Layer1 Conv1	Layer2 Conv2	Layer3 Conv3	Layer4 Conv4	Layer5 Conv5	Layer6 Conv6	Layer7 Conv5
The improved	ZF net							
Mask size	-	(5by5)	(5by5)	(3by3)	(3by3)	(3by3)	(3by3)	(3by3)
strides	-	2	2	1	1	1	1	1
channels	3	96	256	384	384	256	256	384

The results of the proposed system presented in Table (2).

TABLE (2)

RESULTS OF IDENTIFICATION FOR THE TEST SET

Algorithms	types	(AP) (%)	MAP (%)
RPN + ZF shared	Car, Bus, Minivan, Truck	90.6560, 66.3634, 76.6880, 90.5138	81.0553
RPN + ZF unshared	Car, Bus, Minivan, truck	90.6096, 60.3027, 74.0719, 90.3138	78.8245

There are a total of 42,578 images in the experiment that will be trained and assessed. There are 37,578 images used for training and 5000 images utilized for testing. The total number of iterations is 100000. The network training was completed in ten hours due to the GPU acceleration technology. The findings indicate that it performs When convolutional layers are used, has a higher average detection accuracy for automobiles and trucks, but a lower average detection precision for vehicles such as minivans and buses. The outcome might be the consequence of a small training fleet of minivans and buses. [l3] The accuracy of Table illustrates the detecting process (2).

Linkai Chen, et. al. in [22] offer a vehicle's framework recognition and categorization based on video monitoring from traffic cameras that is both efficient and successful. They classify vehicles according to their size and the vehicle's aspect ratio databases. Then, use a convolutional NN to identify the presence of a vehicle. The system employs fusion methods for features to join high and low levels information and detects cars of varying sizes based on their respective attributes. To increase performance, they use a fully convolutional architecture rather than fully connected (FC) layers. Additionally, various approaches have been implemented, including batch-norm, mining of hard examples, and inception Extensive studies on the Jiangsu Highway Dataset (JSHD) confirm our method's competitiveness. The proposed technique outperforms the Faster R-Convolution-NN by (6.5) percent mean average accuracy (MAP).With (1.5) GB of GPU RAM in the phase of testing, this network runs at rate of (15) frames per second, Three times as quick as the R-CNN Faster. It conducts tests on a vehicle dataset to determine the efficacy of the proposed network (JSHD).The trials are conducted on Ubuntu (l6.04) using a GPU (NVIDIA TITAN XP) and an i7 7700 CPU (3). [2]

TABLE (3)

MODULE ABLATION ANALYSIS

	Faster R-CNN				Ours
K-means		√	√	√	√
Feature concatenate?			√	√	√
Detecting on different feature?				√	√
Batch normalization?					√
JSHD (mAP)	69.2	70.8	72.8	73.8	75.7

Izzah Amani Tarmizi, et. al. in [23] develop an area identification method for automatically recognizing vehicles on the road using a Convolutional-NN for autonomous autos [l4] by identifying regions is crucial for the autonomous car to move safely and without colliding with other cars. Through the use of a CNN technique, they develop a system for enhancing vehicle recognition in low-light and poor weather circumstances. They analyzed the (KITTI) Vision, (iROADS) Adverse Driving Scenarios and Matlab datasets. During training phase, the Convolution-NN trained on (177) images and then evaluated on l24 images. The results of simulation explained that the accuracy related to the variety of weather situations. The detection accuracy is (94.3%) in bright weather, (61.4%) in dark weather, (73.4%) and (98.7, respectively, in snowy conditions and they employed a straightforward process, as seen in figure (8). [l4]

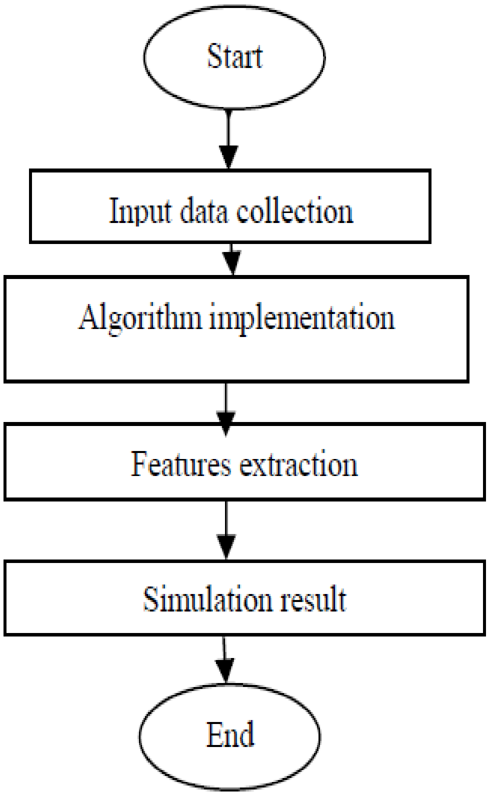


Fig. (8) search (3) methodology

The outcome is shown by the value in the enclosing box area. As seen in figure (9) Positive detection is defined as having a detection accuracy more than 0.5, whilst negative detection having less than 0.5.

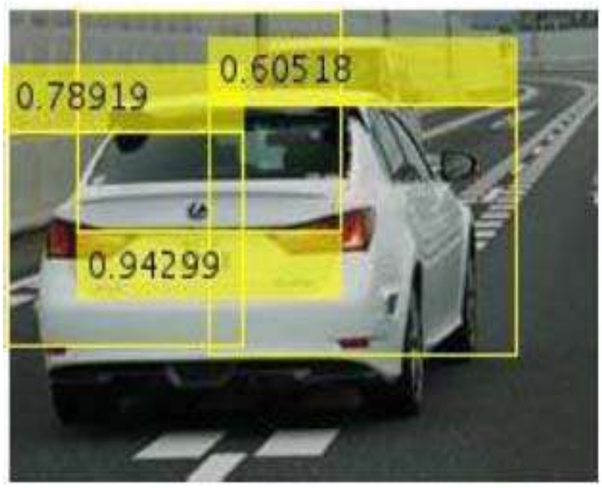


Fig.(9) Result in a sunny weather

As can be seen, CNN discovered three distinct locations from various vantage points. The first region outputs a value of (60.5 %) detection accuracy, whereas the second region outputs a value of (94.3 %) detection accuracy, (78.2 %) is recorded in the third area. Each ROI exhibits positive detection, indicating that the CNN implementation is capable of providing high-accuracy detection.

Aleksa Ćorović et.al. in [24] training the network to recognize five object types(truck, vehicle, pedestrian, lights, and traffic signs) and showing the approach's effectiveness under a variety conditions of driving. The software solution is written in the C programming language and depend on the Darknet-NN design. The method utilized to solve the problem is shown in Figure (10). The video feed is taken via the car's forward-facing camera. As the YOLO network is entered, the camera frames are scaled and sent.[l5]

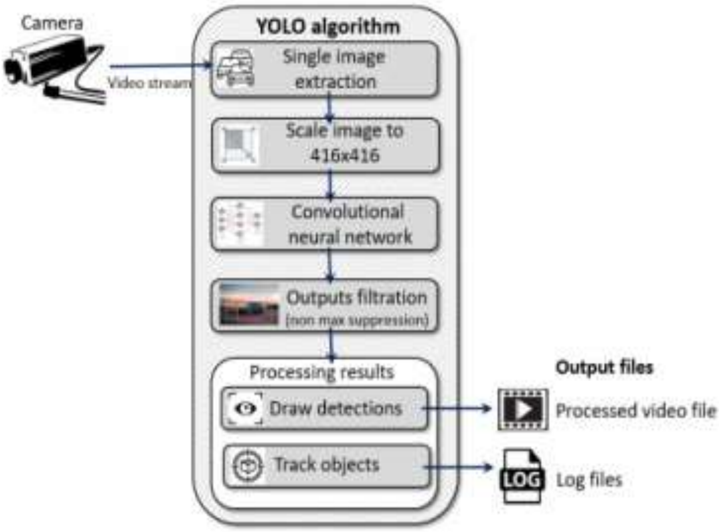


Fig. (10) overview of Yolo algorithm

The YOLO3 network was evaluated by using a processor of Nvidia GTX (1060) GPU, the Berkley Deep Drive dataset was utilized for training. It trains by (70000) images and (30000) validation pictures. A small dataset of (300) photos depicting diverse traffic circumstances from the Novi Sad city, Serbia was created and categorized manually for purposes of testing and validation. Along with it, dashboard camera footage was captured in a variety of weather and lighting settings as illustrated in figure (11). The NN trained for (120) epochs which took two weeks.

TABLE(4)

Training YoloV3 Network

Epoch	Precision	Recall	F1 score	mAP value	Average IOU
40	0.37	0.35	0.36	18.98	24.19
47	0.39	0.37	0.38	21.98	26.12
56	0.37	0.39	0.38	23.49	25.44
75	0.40	0.48	0.44	30.98	28.12
90	0.58	0.53	0.56	44.06	44.06
109	0.60	0.54	0.57	44.53	43.65
120	0.63	0.55	0.59	46.60	45.98

Using previously disclosed technology, YOLO network analysed a selected stream of video from the camera with a resolution of frame (l920 × l080) with rate of (23) frames / second. [l6]

Fig. (11). Examples of car detection in different weather conditions

Muhammad Azhad et. al. in [25], used the methodology illustrated in figure (12) that describes the architecture of the system for detecting vehicles in order to control traffic. Around (1500) photos, each representing a distinct form of vehicle, including automobiles, motorbikes, buses, and trucks. The whole training set consists of (7319) photos that were retrieved from Google. Additionally, (750) photos are included in the validation datasets for each class's kinds. After creating the weight file, it may be examined to determine the model's quality.

mAP (mean Average Precision) is an assessment measure for object identification that combines recall and precision, as seen in table (5). To review categorization and localization, mAP (mean Average Precision) should be measured was important. [17]

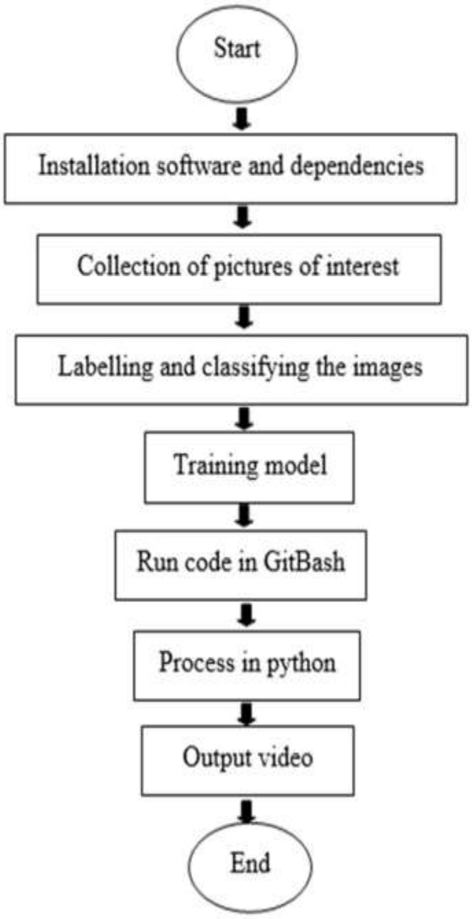


Figure (l2) proposed flowchart

The suggested technique for vehicle recognition and tracking makes use of the Tensor-Flow framework and the Deep-SORT method based on theYolo-v4 model [l7]

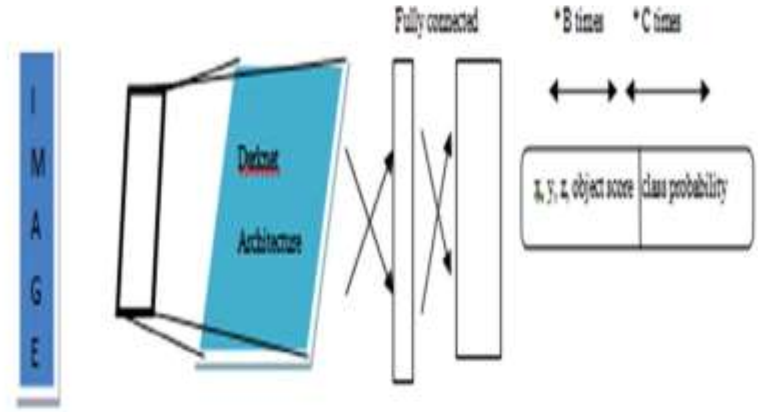
TABLE (5)

mAP Performance

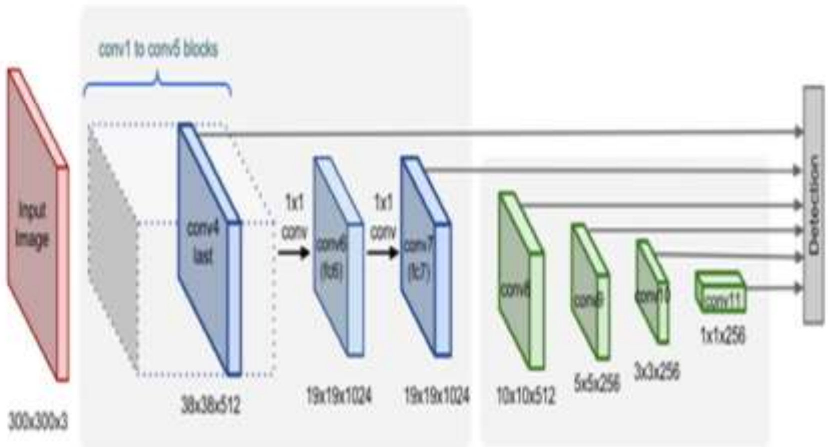
model	mAP @0.5	Classes	Validati0n dataset
YOLOv4(custom)	82.08	4	Custom dataset
YOLOv4-tiny(custom)	76.14	4	Custom dataset
YOLOv3(custom)	80.32	4	Custom dataset
YOLOv3-tiny(custom)	66.03	4	Custom dataset
YOLOv4-(alexAB)	67.9	80	Coc0 validation dataset 2017
YOLOv4-tiny(alexAB)	40.2	80	Coco validation dataset 2017
YOLOv3(pjreddie)	57.9	80	Coco test dataset 2017
YOLOv3-tiny(pjreddie)	33.1	80	Coco test dataset 2017

It has been shown that utilizing Yolov-v4 as well as yolo-v4-tiny is more efficiency and quicker than the old method. It may be used in real-time cameras for surveillance on highways or in video recording to determine the number of cars passing by from the start of the recording to the end. According to the results of this search the use the YOLOv4 model instead of YOLO-v3 model (when the problem requires the accuracy and speed with perfect states) is correct decision, where the YOLO-v4 model can achieve at (82.08%) mAP is (0.5), a (2 %) increase over the previous model in mAP is (0.5). [l7]

Pooja Gupta et. al. in [26], As illustrated in figure (13) they investigated two algorithms: (YOLO-v3) and (SSD). Two operations carried out, the first one for item recognition and the another one for object counting using a video dataset. The effectiveness of these two systems is assessed in terms of counting and detection, and a comparison is offered. The findings indicate that SSD achieves a greater accuracy, recall, and F1 value than YOLO-v3. [l8]



a.YOLO3 system



b. Singular Shot Detection System

Fig. (l3) proposed system for search (6)

For Python Language, (l5.6) in HD WLED touch screen (l366 × 768), 10th Generation Intel Core (i7-lo65G7 l.3) GHz up to (3.9) GHz (8) GB DDR SDRAM (2666) MHz, (5l2) GB SSD, They employed system on two datasets: a picture dataset and a video dataset. They utilize image data sets for item identification and video data sets for object counting. Images from the (SGSITS) college celebration and another events for family. When the YOLOv3 and SSD findings are compared, the SSD results demonstrate more accuracy than for the YOLOv3. as illustrated in table (6).[l8]

TABLE (6)

performance metrics

Data	proposed date sets			training dataset in SSD			tasting dataset in SSD		
	Precision	recall	F1	Precision	recall	F1	Precision	recall	F1
Image1	0.90	0.86	0.89	0.98	0.93	0.96	0.96	0.82	0.92
Image2	0.87	0.82	0.85	0.99	0.91	0.95	0.96	0.86	0.91
Image3	0.88	0.83	0.86	0.99	0.88	0.92	0.95	0.88	0.89
Video1	0.89	0.81	0.85	0.98	0.89	0.93	0.95	0.84	0.93
Video2	0.91	0.85	0.88	0.98	0.84	0.94	0.96	0.85	0.92
Video3	0.87	0.81	0.84	0.99	0.87	0.95	0.95	0.83	0.91

Jeong-ah Kim and co-workers in [27], proposed A technique for recognizing vehicle kinds was developed in this work using learning model. Faster-R-Convolution-NN, SSD and YOLO algorithms with a good accuracy that can be used in conditions of real time With an identification rate of (93%), the Yolov4 model outperforms earlier approaches. Classification System and Data Base for Vehicle Types The categorization of vehicles used for type of vehicle identification was given by the Korea Expressway Corporation. Light vehicle (marked as small), vehicle carrying (9) passengers (called automobile), vehicle carrying (9-25) passengers (called minivan), and vehicle carrying (25) passengers or more (called bigvan). Freight cars are categorized into two types: (dubbed micro trucks) and (dubbed trucks) [19]. To assess the performance of several deep learning-based models for vehicle type identification, the performance of each approach was examined as depicted in table (7). Following model training, we selected the model with good processing time and perfect accuracy. The test included (4500) photos that were not included in the set of training [20].

TABLE (7)

Performance of YOLO, Faster R-CNN and SSD Models

Label	Perf. of YOLO-v4			Perf. of Faster R-CNN			Perf. of SSD		
	Pr.	TP	FP	Pr.	TP	FP	Pr.	TP	FP
Car	98.1	273	25	98.1	263	17	92.7	257	34
Mini-Van	94.9	52	5	87.2	52	41	84.3	29	1
Big-Van	100.	8	0	100.	8	1	87.5	7	0
Mini-Truck	99.0	162	4	99.7	164	16	97.9	151	0
Truck	98.5	27	5	100.	27	2	94.9	21	5
Compact	98.6	36	1	80.3	25	10	85.8	32	15

IV. CONCLUSIONS

Some of the researches select system to recognize small number of classes like four and this state help in improving system accuracy, as well as the selecting images with one or few objects effect in simplification calculation and reducing detecting time as a result. Each mentioned strong trait for any system must be explained in detail, the object detection projects used in various applications a comparison of the two existing models for object detection is two difficult task because of many logic reasons like the difference in exploit data sets in each study, The (YOLO-v3) and (SSD) model has been analyzed with different photo datasets for detecting objects as well as videos for many purposes like object detection, counting and interpreting. The results have been carried out on different datasets and the performance metrics will be calculated to improve the accuracy of the proposed algorithms. Several object detection methods for the purpose of recognizing car model was introduced and analyzed in order to compare performance metrics for all by calculating calculation speed and accuracy, it can be seen clearly that the proposed model, YOLO-v4, has the best features while the SSD network is fast.

REFERENCES

[1] L. LIU, Z., ET AL. SOFT PROTOTYPING CAMERA DESIGNS FOR CAR DETECTION based on a convolutional neural network. in Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019.

[2] Chen, L., et al., An algorithm for highway vehicle detection based on convolutional neural network. Eurasip Journal on Image and Video Processing, 2018. 2018(1): p. 1-7.

[3] Girshick, R., et al. Rich feature hierarchies for accurate object detection and semantic segmentation. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

[4] Redmon, J., et al. You only look once: Unified, real-time object detection. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[5] Liu, W., et al. Ssd: Single shot multibox detector. in European conference on computer vision. 2016. Springer.

[6] Traore, B.B., B. Kamsu-Foguem, and F. Tangara, Deep convolution neural network for image recognition. Ecological Informatics, 2018. 48: p. 257-268.

[7] Gupta, A., et al., Non-Gaussian information from weak lensing data via deep learning. Physical Review D, 20l8. 97(10): p. 103515.

[8] Tarr, M.J. and H.H. Bülthoff, Image-based object recognition in man, monkey and machine. Cognition, l998. 67(1-2): p. 1-20.

[9] Wang, N., Y. Wang, and M.J. Er, Review on deep learning techniques for marine object recognition: Architectures and algorithms. Control Engineering Practice, 2022. 118: p. 104458.

[10] Sun, P., et al. Sparse r-cnn: End-to-end object detection with learnable proposals. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[11] Erhan, D., et al. Scalable object detection using deep neural networks. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.

[12] Hassan, S.U., et al. Detection and Classification of Vehicles based on Deep Semantic Features of Faster R-CNN. in 2019 International Conference on Innovative Computing (ICIC). 2019. IEEE.

[13] Cui, C., et al. Vehicle re-identification by fusing multiple deep neural networks. in 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA). 2017. IEEE.

[14] Tarmizi, I.A. and A. Abd Aziz. Vehicle Detection Using Convolutional Neural Network for Autonomous Vehicles. in 2018 International Conference on Intelligent and Advanced System (ICIAS). 2018. IEEE.

[15] Kadam, S., et al. Freight Analysis Using YOLOv2. in Proceedings of International Conference on Communication and Information Processing (ICCIP). 2019.

[16] Ghosh, S. and D.L. Reilly. Credit card fraud detection with a neural-network. in System Sciences,1994. Proceedings of the Twenty-Seventh Hawaii International Conference on. 1994. IEEE.

[17] Zuraimi, M.A.B. and F.H.K. Zaman. Vehicle Detection and Tracking using YOLO and DeepSORT. in 2021 IEEE llth IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE). 2021. IEEE.

[18] Gupta, P., V. Sharma, and S. Varma, People detection and counting using YOLOv3 and SSD models. Materials Today: Proceedings, 2021.

[19] James, N., et al., Automated Checkout for Stores: A Computer Vision Approach. REVISTA GEINTEC-GESTAO INOVACAO E TECNOLOGIAS, 2021. 11(3): p. 1830-1841.

[20] Kim, J.-a., J.-Y. Sung, and S.-h. Park. Comparison of Faster-RCNN, YOLO, and SSD for real-time vehicle type recognition. in 2o2o IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia). 2020. IEEE.

[21] Wang, X., et al. (2019). "Real-time vehicle type classification with deep convolutional neural networks." Journal of Real-Time Image Processing 16(1).

[22] Chen, L., et al. (2018). "An algorithm for highway vehicle detection based on convolutional neural network." Eurasip Journal on Image and Video Processing 2018(1).

[23] Tarmizi, I. A. and A. Abd Aziz (2018). Vehicle Detection Using Convolutional Neural Network for Autonomous Vehicles. 2018 International Conference on Intelligent and Advanced System (ICIAS), IEEE.

[24] Ćorović, A., et al. (2018). The real-time detection of traffic participants using YOLO algorithm. 2018 26th Telecommunications Forum (TELFOR), IEEE.

[25] Zuraimi, M. A. B. and F. H. K. Zaman (2021). Vehicle Detection and Tracking using YOLO and DeepSORT. 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), IEEE.

[26] Gupta, P., et al. (2021). "People detection and counting using YOLOv3 and SSD models." Materials Today: Proceedings.

[27] Kim, J.-a., et al. (2020). Comparison of Faster-RCNN, YOLO, and SSD for real-time vehicle type recognition. 2020 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia), IEEE.