Turkish Journal of Science & Technology
17(2), 395-403, 2022

Research Paper
https://doi.org/10.55525/tjst.1080861

Optimized YOLOv4 Algorithm for Car Detection in Traffic Flow

Alzubair Alqaraghuli[1], Oğuz Ata[2*]
[1] Information Technology, Computer Engineering, Altinbas University, Istanbul, Turkiye
[2] Information Technology, Computer Engineering, Altinbas University, Istanbul, Turkiye
[1]zubairsk53@gmail.com, [2*] oguzata@gmail.com

(Geliş/Received: 30/05/2022; Kabul/Accepted: 02/08/2022)

Abstract: The vehicle detection accuracy and actual in images and videos appear to be very tough and critical duties in a key technology traffic system. Specifically, under convoluted traffic conditions. As a result, the presented study proposes single-stage deep neural networks YOLOv4-3L, YOLOv4-2L, YOLOv4-GB, and YOLOv3-GB. After optimizing the network structure by adding more layers in the right positions with the right amount of filters, the dataset will be repaired and the noise reduced before being sent to the mentoring. This research will be applied to YOLOv3 and YOLOv4. In this study the OA-Dataset is collect and used, the data set is manually labeled with the care of different weathers and scenarios, as well as for end-to-end training of the network. Around the same time, optimized YOLOv4 and YOLOv3 demonstrate a significant degree of accuracy with 99.68 % and precision of 91 %. The speed and detection accuracy of this algorithm are found to be higher than that of previous algorithms.

Key words: Car Detection, The Traffic Flow, YOLOv4, Deep learning.

Trafik Akışında Araba Algılama için Optimize Edilmiş YOLOv4 Algoritması

Öz: Görüntülerde ve videolarda araç algılama doğruluğu ve gerçekliği, önemli bir teknoloji trafik sisteminde çok zor ve kritik görevler gibi görünmektedir. Özellikle, kıvrımlı trafik koşulları altında. Sonuç olarak, sunulan çalışma tek aşamalı derin sinir ağları YOLOv4-3L, YOLOv4-2L, YOLOv4-GB ve YOLOv3-GB önermektedir. Doğru miktarda filtre ile doğru pozisyonlara daha fazla katman ekleyerek ağ yapısını optimize ettikten sonra, veri seti onarılacak ve mentorluğa gönderilmeden önce gürültü azaltılacaktır. Bu araştırma YOLOv3 ve YOLOv4'e uygulanacaktır. Bu çalışmada OA-Dataset toplanmış ve kullanılmış, veri seti farklı hava koşulları ve senaryolar dikkate alınarak ve ayrıca ağın uçtan uca eğitimi için manuel olarak etiketlenmiştir. Aynı zamanda, optimize edilmiş YOLOv4 ve YOLOv3, %99,68 ve %91 hassasiyetle önemli derecede doğruluk gösterir. Bu algoritmanın hızı ve algılama doğruluğu, önceki algoritmalardan daha yüksek bulunmuştur.

Anahtar kelimeler: Araba algılama, Trafik akışı, YOLOv4 modeli, Derin öğrenme.

1. Introduction

With all of the changes in people's lives, city growth has accelerated, resulting in a large increase in the number of private cars. As a result, traffic congestion has become a major issue that impacts people's lives. The intelligent transportation system (ITS), the traffic with a good accuracy predictions may present basic information to make a decision to the traffic administration. As a result, in smart cities, high-accuracy traffic transmission estimate is seen as a critical component of evolving smart transmission systems.The main task in traffic prediction is the accurate and rapid detection of cars in traffic videos or images, Therefore, it is really substantial to find an algorithm that can produce correct detection and real-time recognition of cars [1]. The presented study aims to sort the conventional detection methods in the following: - Xu et al. found out distinct factors that are able to pull out the features from the zone of interest which are nominated in the pictures then by training a classifier, will apply the detection, Unfortunately, due to the sensor's complexity, these techniques will decrease accuracy[2]. Qiu et al. used visual detection, which is based on inter-frame and visual flow variations. Visual flow has good accuracy, but the problem with visual flow is the slow detection speed, even though the inter-frame difference process is fast but not very accurate[3]. Felzenszwalb et al. suggested a method called “the classification of sliding window”, the first step of this method is sliding the windows to pull out features of the region of interest after that the second step is applying a classifier to get detection on the target by using the support-vector machine (SVM). This method requires a lot of calculation that leads to slow speed in detection [4].while , Kenan Mu et al.

[*] Corresponding author: oguzata@gmail.com ORCID Number of authors: [1] 0000-0002-6117-8051, [2] 0000-0003-4511-7694

used a method called edge detection, but the process of edge detection can be affected by noise and background intervention which means inaccurate detection [5]. The above-mentioned target detection approaches necessitate a lot of calculations, which results in slow detection speeds, as well as a lack of generality in the field of feature extraction.

In the last few years, with the fast developments in artificial intelligence technologies and computer vision, new methods of object detection algorithms have been investigated, these methods are based on deep learning, the convolutional neural networks (CNN) is considered to be very convenient and the feature extraction of the images has strong generalization [6]. These days, there are two major methods of object detection in deep learning: the first is the algorithm that combines candidate region suggestions and convolutional neural networks, exemplified by spatial pyramid pooling (SPP)-net [7]. and Region-Based Convolutional Neural Networks (R-CNN) [8]. the second one is You Only Look Once (YOLO)[9-12] [10][11][12]. and Single Shot MultiBox Detector chain (SSD) [13]. These algorithms convert detection issues into case of regression by employing deep learning. The R-CNN algorithm uses feature extractor to select region suggestion boxes, which improves object detection accuracy. However, due to the scaled candidate box and the large number of calculations, the process takes a long time, resulting in the loss of image feature information. In the R-CNN, there is a problem with the size of the fixed input layer, the SPP-net algorithm solved this problem by using a pyramid pooling layer. But this will lead to cumbersome training steps and every step generate a ratio of errors due to the SVM classification and the convolutional neural network should be trained separately. As a result of that, it will take a long time for training and a large space from the hard disk due to the large number of feature files that are saved after the training. The rapid R-CNN algorithm which is merging the specifications of SPP-net into the R-CNN has solved some problems like test time, long training and large space occupation, etc. But it still depends on the selective search method to find the extraction of the particular box, so that means the problem of time-consuming still exists [14]. By replacing the selective search (SS) with the RPN(region proposal networks), the Faster R-CNN back end and the region frame extraction candidate are integrated with convolutional neural network model, as a result, the candidate region extraction time will be short in the Faster R-CNN [15]. The first truly end-to-end object detection algorithm is Faster R-CNN. However, it is real-time object detection speed is far from the requirement.YOLO chain is a network algorithm based on regression which uses a full map for training and it returns both object category and object frame at diverse positions. To train the network, the algorithm of YOLO employs a method rely on the frame area algorithm of the candidate. For the training, YOLO employs the full image then returns both object category and object frame at various positions, using the steps that the researcher mentioned makes it a lot simpler to fastly differentiate the background area from objects however it is susceptible to errors of position to enhance the structure of the YOLO network model YOLOv2 employ methods, and due to these methods the speed of detection improved. The network of YOLOv2 is simple, even with it enhancing the speed but it does not make the detection more accurate. YOLOv3 employs the idea of Feature Pyramid Networks(FPN) to achieve multi-scale predictions [16] and for extracting the image features, YOLOv3 uses the ideas of the deep residual network(ResNet) to achieve equilibrium between the speed of detection and the detection accuracy [17]. YOLOv4 Is much better than the last versions of YOLO in speed and performance. YOLOv4 uses a bag of freebies and it's a method that is used to enhance the training without any cost from the hardware, also to improve the accuracy of detection, YOLOv4 use collection of modules. These modules excess the inference cost by little amount but it has a good effect on the detection accuracy. The presented study suggests some changes to the structure of the YOLOv4 to optimize it, in this study, the researcher investigate optimized YOLOv4-2L, YOLOv4-3L, YOLOv4-GB, and YOLOv3-GB for tracking and detecting vehicles in traffics. Optimized modules of YOLO is based on the YOLOv4 algorithm, the researcher adds more layers and fixes the data set befor training the model to enhance the YOLOv4 detection system performance on traffics.

2. Principle and Composition of Traffic Detectoin System

Many researchers in field indicates that the characterization of traffic state studies can be gathered into some categories like determination of average speed, determination of traffic flow parameters, and detection of the congestion location [18]. The detection system of traffics consists of an image preprocessing module, the module of video image acquisition, the module of vehicle flows statistics, and the module of vehicle detection and identification. All the detection system modules are displayed in figure 1.

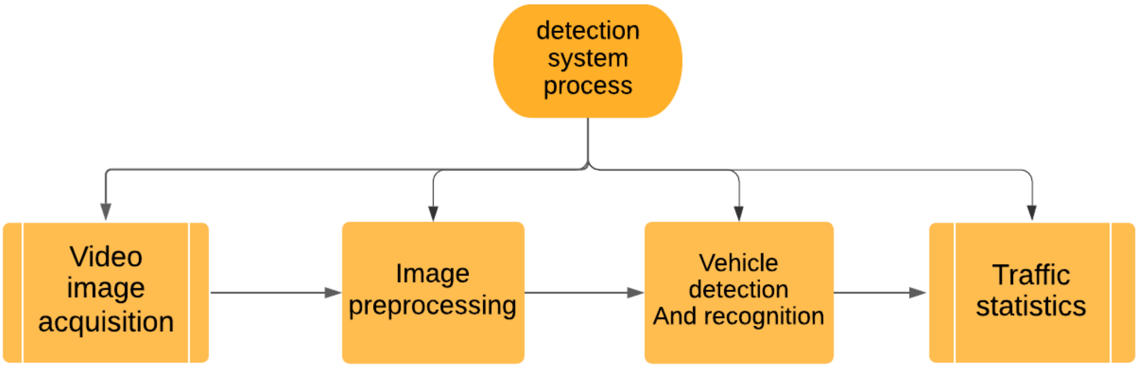

Figure 1. The process of Detection System

The core of the system is the module of vehicle detection and recognition because it recognizes and locates the vehicle in video images. To combine recognition and object position into one, which needs to take the requirements of recognition accuracy and speed detection into consideration, the researcher uses YOLOv4 for vehicles detection and recognization.YOLOv4 is the updated YOLO, so it has advantages in detection accuracy, fast speed, and accurate positioning. On previous versions of YOLO, YOLOv4 uses CSPDarknet53 as Backbone and the CSPDraknet53 is an updated version of Darknet-53. It is based on concepts from CSPNet, they use this kind of the darknet to increase gradient path, reduce memory traffic and balance computation of each layer. YOLOv4 uses Spatial Pyramid Pooling (SPP) as the Neck. And the Head of YOLOv4 is YOLOv3. The present research hotspot is YOLOv4, in figure 2 the YOLOv4 network structure is shown.

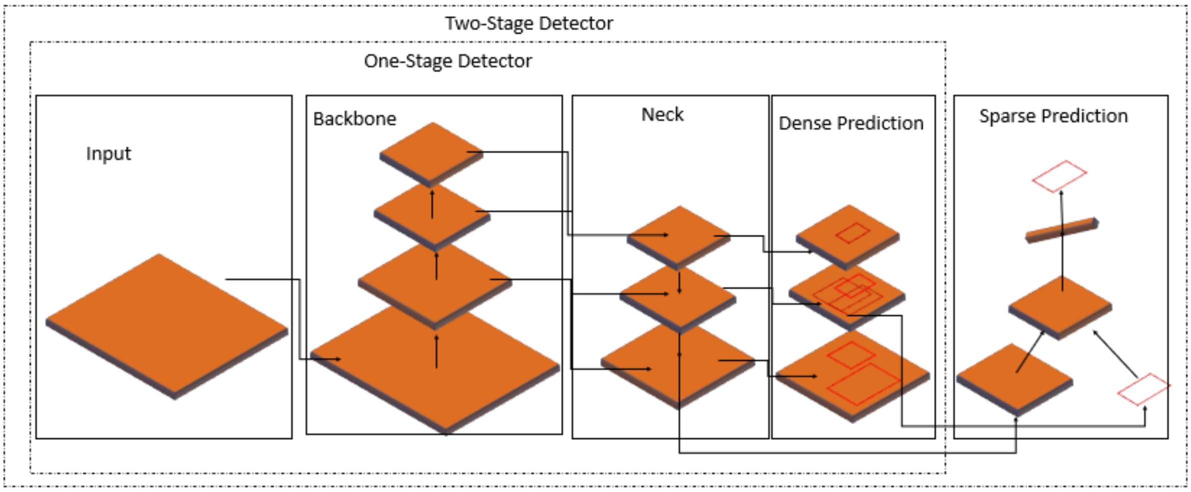

Figure 2. YOLOv4 Network Sturcture

2.1 YOLOv4 Structure Optimization

YOLOv4 uses CSPdarknet-53 and CSPdarknet-53 consists of convolutional layers. So as what has been done in YOLO 9000, the study finds out that increasing the number of layers helps to enhance the accuracy of the detection, the presented study suggests doing the same with YOLOv4. In the structure of YOLO. There are three layers responsible for the detection, so in the first optimization , the researcher adds one layer before the first detection layer and one more layer before the third detection layer. These layers contain a specified amount of filters, and this study revealed that when the model is trained, the detection accuracy improved. In the second optimization, the researcher added three layers, one before the first YOLO layer, which is responsible for detection, and two layers before the third detection layer, and the detection rate increased.

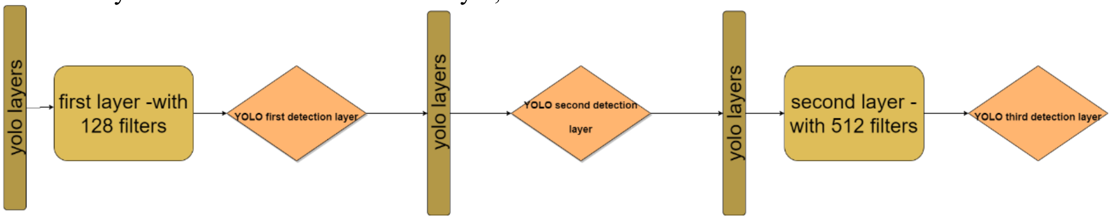
yolo layers
first layer -with 128 filters
YOLO first detection layer
yolo layers
YOLO second detection layer
yolo layers
second layer - with 512 filters
YOLO third detection layer

Figure 3. YOLOv4_2L

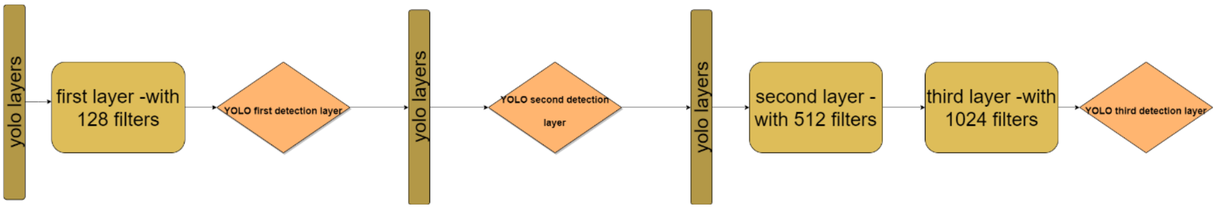
yolo layers
first layer -with 128 filters
YOLO first detection layer
yolo layers
YOLO second detection layer
yolo layers
second layer - with 512 filters
third layer -with 1024 filters
YOLO third detection layer

Figure 4. YOLOv4_3L

figures 3 – 4 shows that the optimization of the structure are YOLO first detection layer, YOLO second detection layer, and YOLO third detection layer refers to the layers that are responsible for detection, the square shape refers to the new layers that the researcher adds, in addition to the original layers that represented by the long shape.

2.2 Dataset Optimization

E.Torpov et al. referred that the low rate of frames makes the detection of vehicle tracking not feasible, and image compression and low image quality can add more noise to the images[19]. The dataset is a really important part that plays an essential role in the YOLO. The presented study decides to clear the images from the noise before labeling and importing them. The procedure is to apply the Gaussian filter on the images using the open-cv library in python, the gaussian filter will remove the noise from the images[20].

2.3 Making the Dataset

If the study aims to examine real-world traffic, it should use images from real-world traffic with data set called (DETRAC). Furthermore, it is a large-scale detection and tracking data set that is used to detect and track vehicles. This data set is chosen since it is derived from real footage of traffic on highways and bridges in locations such as Tianjin and Beijing. A total of 6203 images from the data set are manually tagged. This data set contains a variety of meteorological conditions. Images of the sun, rain, clouds, and night, as well as their height and angle, are all different.

1) The collecting images is done with different weather situations, dusk images, rainy images, and daytime images, the images are taken from the dataset of DETRAC, the number of images is 6203 images;
2) To make training set the researcher randomly extracts 80% of the dataset.
3) To make testing set the researcher randomly extracts 20% of the dataset.

4) 6203 images can be called as the OA. Dataset in the current study.

To label each vehicle in the training set , testing set images and use the labelling tool and when label the vehicle on the images an XML file will be created. The XML file saves the information of labeling for the next training. Five values will be saved in the XML file. The first value is an individual number that refers to the type of the class. So if there are two classes, for example, gender classes, (0) will refer to male and 1 will refer to female, but in our case, there is only one class which is vehicle so the individual number will be 0, the rest values will be decimal numbers. These numbers will refer to the location of the vehicle, the first one is for X center coordinate, the second decimal number refer to Y center coordinate, the third decimal number refer to width, and the final decimal number refers to the height.

3. Analysis and Results of Experiment

3.1. The Paltform

In the presented study, it used Windows 10 system, and PyCharm environment, also python 3.7. is used under the Darknet framework, the YOLOv4 algorithm is applied. NVIDIA RTX 3070 graphic card is used to accelerate training and the processor is Ryzen 7 5800H.

3.2. Network Training

The model multiple times is used to conduct the results, first, the trained model for 30,000 iterations but the results are not good , it found out that the more iteration steps are not a good choice to improve the detection accuracy, so it is used for only 4000 iterations. In the first 1000 iterations, the average loss was so high then it starts to get down slowly and it reaches 2.0 at 1100-1500, at 1500-2500 the average lose was 1.0 then it goes to 0.9 at 3000-3500 and at 3500-4000 it reaches 0.8-0.7, and that was the best average loss in the study, it gives us a really good accuracy. the tests with YOLOv4, YOLOv4-2L, and YOLOv4-3L, with YOLOv4-GB, YOLOv3, and YOLOv3-GB. The researcher uses part of our data set and it contains 200 images, these images contain vehicles in different positions, the training is done with 2000 iterations, it is also found out that if the training increases, it will not make a big difference the results are almost the same.

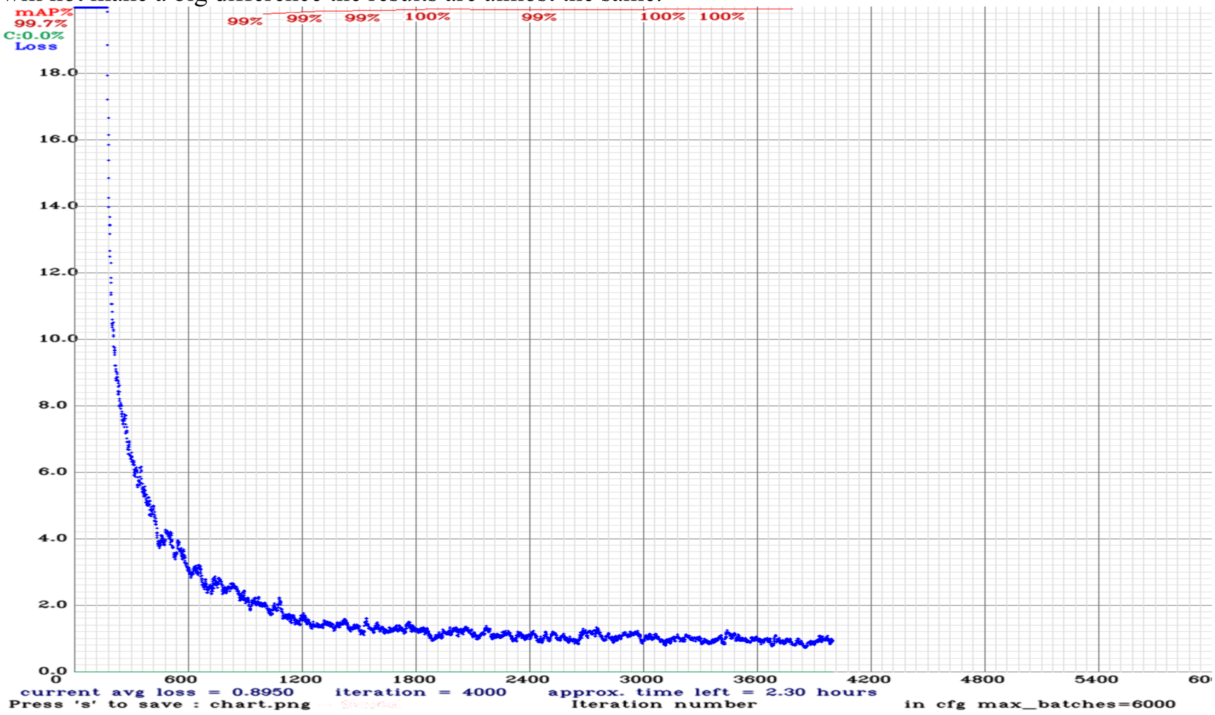

Figure 5. YOLOv4-2L Training Chart

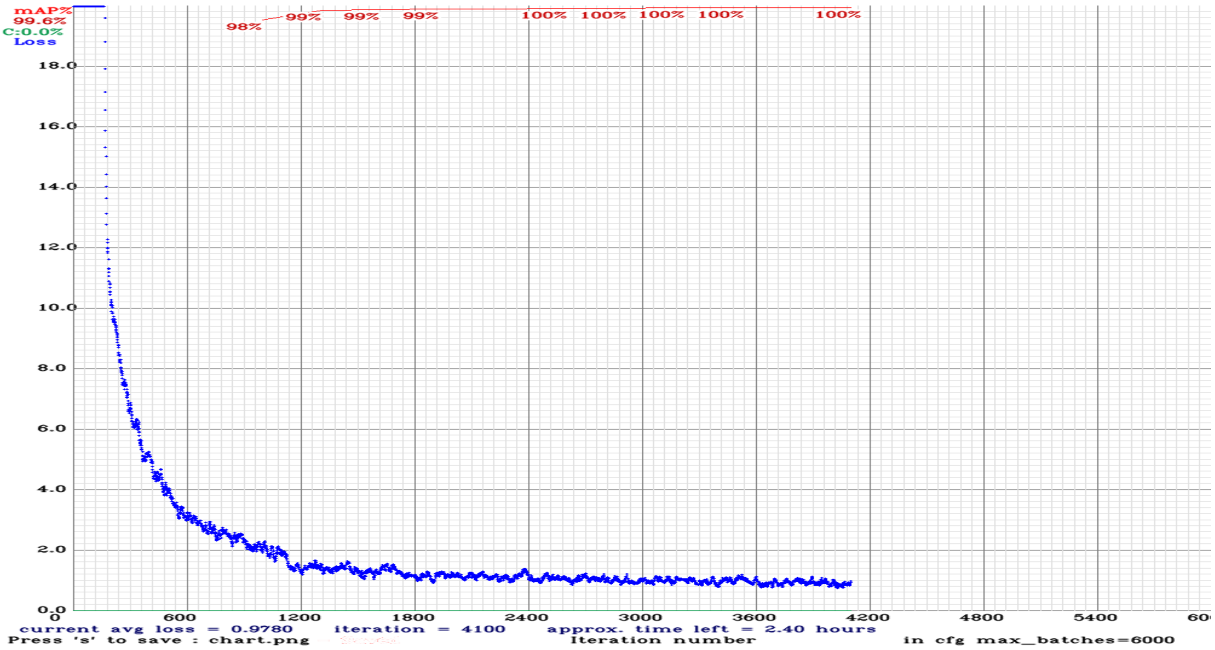

Figure 5. YOLOv4-3L Training Chart

3.3. Evaluation Parameters

To make sure of the efficiency of optimized YOLOv4, the particular tests are used on dataset, analyzing the experimental data, and compression of the experimental results. In the monitoring processes, some problems may occur like false detection and missed detection. As evaluation parameters, Precision, Recall, and mAP(mean average precision) are also used to fulfill the required results. The Precision mean accuracy which refers to the proportion of the accurately detected vehicles to all number of detected vehicles. And the Recall implies the review rate which means the proportion of the number of detected vehicles to the aggregate number of vehicles in the dataset. Equations are displayed below :

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

The (TP) referes to True Positive which is the correct vehicles detection number, the (TN) mean True Negative which refers to the correct backgrounds detection number, the (FP) referes to False Negative which indicates detection missed number.

3.4 Setup of YOLO Architecture

Here the names of the modified YOLO will be exblined in deteails.

- YOLOv4-2L : this name refers to the optimized YOLOv4 with extra two layers in the structure of YOLOv4, these layers contain filters, 128 filters in the first layer and 512 filters in the second.
- YOLOv4-3L : this name refers to the second optimized YOLOv4 with extra three layers in the structure of YOLOv4, these layers also contain filters, exactly 128, 512 and 1024 filters to the first,second and third layer successively.
- YOLOv4-GB and YOLOv3-GB : these names refers to YOLO with fixed data set.

4. The Analysis of Comparison Results of Different Algorithms

The Precision, Recall, and mAP (Mean Average Precision) of the presented study are compared with other models. results are displayed in Table 1.

Table 1. Comparative Between Presented Study and Other Studies

	The Precision	The Recall	Accuracy
YOLOv4 (Original)[12]	90%	100%	99.67%
YOLOv4-2L(Presented Study)	89%	100%	99.68%
YOLOv4-3L(Presented Study)	91%	100%	99.67%
YOLOv4-GB(Presented Study)	97%	100%	99.98%
YOLOv3(Original)[11]	97%	100%	99.97%
YOLOv3-GB(Presented Study)	97%	100%	99.97%
YOLOv3-DL(Previous Study)[1]	96%	98%	98.83%

From table 1 it appare that the accuracy has increased in comparison with other studies that are indicated that the current study can detect more vehicles in the image and difficult positions like vehicles which are far away from the detection camera.

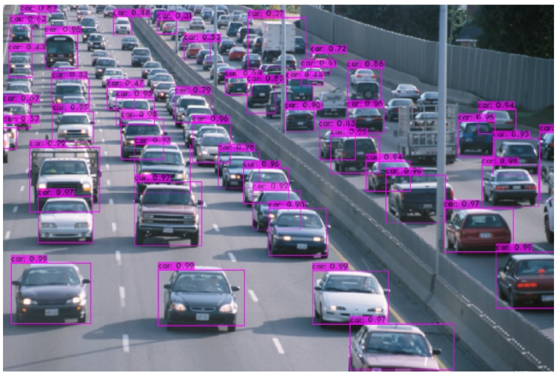

Figure 7. YOLOv4

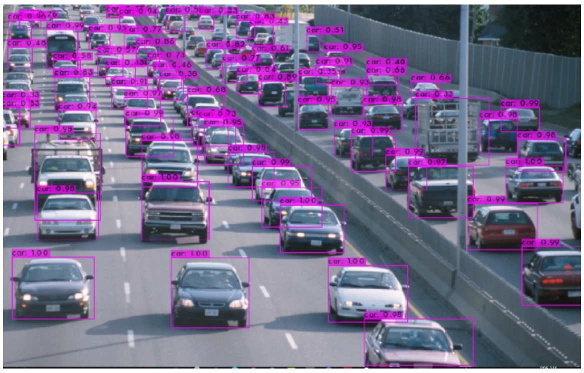

Figure 8. YOLOv4-2L

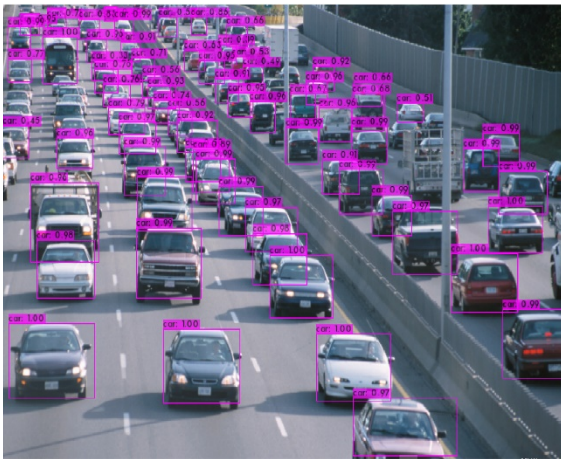

Figure 9. YOLOv4-3L

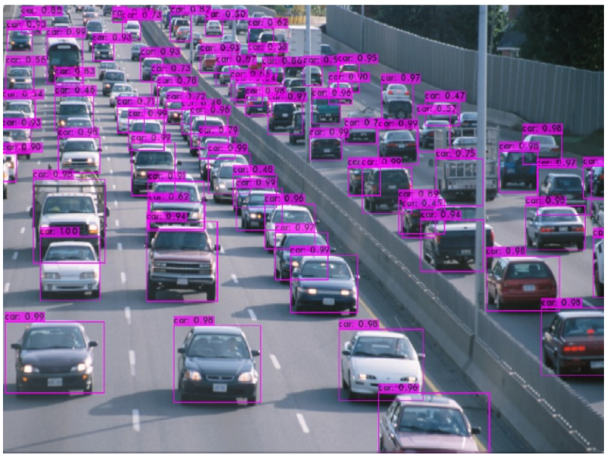

Figure 10. YOLOv4-GB

The figures(7-8-9-10) above shows the results of YOLO object detection method and the optimized YOLOv4 with indicating that the presented study shows massive accuracy compared to the original study.

4.1 Video Analysis and Comparison

To experience the detection accuracy of the presented study in a video stream, the study is applied on a vehicle driving video and the length of the video is 30s while, the test results are shown as a follow in Table2. The presented studies's accuracy rate is higher than other studies.

Table 2. Video Comparison

	AVG-FPS
YOLOv4 (Original)[12]	62.0
YOLOv4-2L (Presented Study)	62.5
YOLOv4-3L (Presented Study)	60
YOLOv4-GB (Presented Study)	62.2
YOLOv3 (Original)[11]	63.3
YOLOv3-GB (Presented Study)	63.1

The test results shows that it took a different amount of time to count the vehicles per frame depending on the algorithm used within study, furthermore, the presented study is nearly similar to the real traffic. The current study is compared to other studie's results like YOLOv3 and YOLOv4, that are indicated the video monitoring accuracy rate of the traffic flow has increased.

5. Conclusion

Due to the limitation in YOLOv4, the current study finds out that the detecting far-away object is hard with noticing that YOLOv4 may miss the detection if there are more vehicles or when the vehicles are different in size, and that will affect the accuracy rate of the traffic flow prediction and statistics information. By using the presented study, the high accuracy rate of traffic flow statistics could be produced, and the results are conducted by using the presented study is ultimately based to the real number of vehicles in images. The results show that the YOLOv4 can be improved in the accuracy rate of traffic monitoring and in real-time.

References

[1] Y. Q. Huang, J. C. Zheng, S. D. Sun, C. F. Yang, and J. Liu, "Optimized YOLOv3 algorithm and its application in traffic flow detections," *Appl. Sci.*, vol. 10, no. 9, May 2020, doi: 10.3390/app10093079.

[2] Y. Xu, G. Yu, Y. Wang, X. Wu, and Y. Ma, "A hybrid vehicle detection method based on viola-jones and HOG + SVM from UAV images," *Sensors (Switzerland)*, vol. 16, no. 8, 2016, doi: 10.3390/s16081325.

[3] Q. J. Qiu, L. Yong, and D. W. Cai, "Vehicle detection based on LBP features of the Haar-like Characteristics," *Proc. World Congr. Intell. Control Autom.*, vol. 2015-March, no. March, pp. 1050–1055, 2015, doi: 10.1109/WCICA.2014.7052862.

[4] P. F. Felzenszwalb, R. B. Girshick, D. Mcallester, and D. Ramanan, "Object Detection With Partbase," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, 2010.

[5] K. Mu, F. Hui, X. Zhao, and C. Prehofer, "Multiscale edge fusion for vehicle detection based on difference of Gaussian," *Optik (Stuttg).*, vol. 127, no. 11, pp. 4794–4798, 2016, doi: 10.1016/j.ijleo.2016.01.017.

[6] K. S. Choi, J. S. Shin, J. J. Lee, Y. S. Kim, S. B. Kim, and C. W. Kim, "In vitro trans-differentiation of rat mesenchymal cells into insulin-producing cells by rat pancreatic extract," *Biochem. Biophys. Res. Commun.*, vol. 330, no. 4, pp. 1299–1305, 2005, doi: 10.1016/j.bbrc.2005.03.111.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, 2015, doi: 10.1109/TPAMI.2015.2389824.

[8] R. Girshick, J. Donahue, T. Darrell, J. Malik, U. C. Berkeley, and J. Malik, "1043.0690," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, p. 5000, 2014, doi: 10.1109/CVPR.2014.81.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection." [Online]. Available: https://goo.gl/bEs6Cj.

[10] J. Redmon and A. Farhadi, "YOLO9000: Better, Faster, Stronger." [Online]. Available: http://pjreddie.com/yolo9000/.

[11] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement." [Online]. Available: https://pjreddie.com/yolo/.

[12] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," 2020, [Online]. Available: http://arxiv.org/abs/2004.10934.

[13] V. Thakar, H. Saini, W. Ahmed, M. M. Soltani, A. Aly, and J. Y. Yu, "Efficient Single-Shot Multibox Detector for Construction Site Monitoring," *2018 IEEE Int. Smart Cities Conf. ISC2 2018*, no. 1, p. 77, 2019, doi: 10.1109/ISC2.2018.8656929.

[14] J. Liu, Y. Huang, J. Peng, J. Yao, and L. Wang, "Fast Object Detection at Constrained Energy," *IEEE Trans. Emerg. Top. Comput.*, vol. 6, no. 3, pp. 409–416, 2018, doi: 10.1109/TETC.2016.2577538.

[15] S. Ren, K. He, and R. Girshick, "Faster R-CNN : Towards Real-Time Object Detection with Region Proposal Networks," pp. 1–9.

[16] F. B. Tesema, J. Lin, J. Ou, H. Wu, and W. Zhu, "Feature Fusing of Feature Pyramid Network for Multi-Scale Pedestrian Detection," *2018 15th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process. ICCWAMTIP 2018*, no. 1, pp. 10–13, 2019, doi: 10.1109/ICCWAMTIP.2018.8632614.

[17] V. Sangeetha and K. J. R. Prasad, "Syntheses of novel derivatives of 2-acetylfuro[2,3-a]carbazoles, benzo[1,2-b]-1,4-thiazepino[2,3-a]carbazoles and 1-acetyloxycarbazole-2- carbaldehydes," *Indian J. Chem. - Sect. B Org. Med. Chem.*, vol. 45, no. 8, pp. 1951–1954, 2006, doi: 10.1002/chin.200650130.

[18] Y. Liu, "Big Data Technology and Its Analysis of Application in Urban Intelligent Transportation System," *Proc. - 3rd Int. Conf. Intell. Transp. Big Data Smart City, ICITBS 2018*, vol. 2018-Janua, pp. 17–19, 2018, doi: 10.1109/ICITBS.2018.00012.

[19] E. Toropov, L. Gui, S. Zhang, S. Kottur, and J. M. F. Moura, "TRAFFIC FLOW FROM A LOW FRAME RATE CITY CAMERA Electrical and Computer Engineering Pittsburgh , PA , USA Instituto Superior Técnico Instituto de Sistemas e Robótica Lisbon , Portugal," *Int. Conf. Image Process.*, pp. 3802–3806, 2015.

[20] "Gaussian filtering • Significant values," pp. 18–32, 2010.