Deep Learning-based Detection System for Heavy-Construction Vehicles and Urban Traffic Monitoring

Sreelatha R[1]

Department of Information Science
BMS College of Engineering
Bangalore, Karnataka, India

Dr. Roopa Lakshmi R[2]

Department of Computer Science
Manipal Institute of Technology
Manipal, Karnataka, India

Abstract—In this intelligent transportation systems era, traffic congestion analysis in terms of vehicle detection followed by tracking their speed is gaining tremendous attention due to its complicated intrinsic ingredients. Specifically, in the existing literature, vehicle detection on highway roads are studied extensively while, to the best of our knowledge the identification and tracking of heavy-construction vehicles such as rollers are not yet fully explored. More specifically, heavy- construction vehicles such as road rollers, trenchers and bulldozers significantly aggravate the congestion in urban roads during peak hours because of their deadly slow movement rates accompanied by their occupation of majority of road portions. Due to these reasons, promising frameworks are very much important, which can identify the heavy-construction vehicles moving in urban traffic-prone roads so that appropriate congestion evaluation strategies can be adopted to monitor traffic situations. To solve these issues, this article proposes a new deep-learning based detection framework, which employs Single Shot Detector (SSD)-based object detection system consisting of CNNs. The experimental evaluations extensively carried out on three different datasets including the benchmark ones MIO-TCD localization dataset, clearly demonstrate the enhanced performance of the proposed detection framework in terms of confidence scores and time efficiency when compared to the existing techniques.

Keywords—Intelligent transportation systems; heavy-construction vehicles detection; traffic monitoring and SSD-based CNN component; deep learning

I. INTRODUCTION

Now-a-days, Machine-learning based transportation systems are gaining more popularity due to the incorporation of deep-learning technology for various traffic domains such as traffic monitoring, speed measurement, density estimation and so on [1]. Further, due to the presence of upgraded visual surveillance systems along with GPS, makes it possible to generate enormous volume of traffic information, which can be used for future processing. As a result, traffic monitoring by means of analyzing traffic congestion using movement of vehicles is acquiring enormous attention in the recent years due to its complicated intrinsic factors. Specifically, the state-of-the art highly sophisticated traffic surveillance cameras are capturing traffic flow along with rich-set of traffic parameters, which in turn can be effectively employed for detection followed by tracking of targeted vehicles in the given traffic environment.

Generally speaking, the performance of any traffic monitoring system is primarily dependent on two critical factors -speed of moving vehicle and density of road-traffic, which may vary from minimum to a greater extent. Specifically, in the present literature, urban traffic monitoring is implemented by means of analyzing traffic congestion of roads based on vehicle speed and road category aspects [1].It is possible to detect the slow moving vehicles in urban traffic using Robust Visual Features. Authors have proposed a new visual feature to detect the vehicles suing SURF based Features [2]. However, in the real-world road scenarios, the speed of the given vehicle mainly depends upon the type of vehicle such as slow-moving Excavator vs fast moving sports car. Due to these reasons, the classification followed by the exact identification of target moving vehicles in given traffic scenarios is very much important, so that the reliability of given traffic monitoring system can be guaranteed to a greater extent [3]. In addition to that accurate identification of target moving vehicles is very much essential in order to reduce the false alarms in target vehicle classification systems.

From another perspective, in the existing road traffic situations, one of the critical factors for heavy-traffic congestion is the presence of construction vehicles such as Front-end loaders and Rollers, which generally move at dead-slow speed. Precisely, the construction vehicles while moving dominate a huge portion of road due to their bigger size when compared with other vehicles such as cars. Further, the construction vehicles generally move at a low-paced manner and fail to follow with the average speed of other moving vehicles, due to which they affect the flow of moving traffic to a greater extent. In other words, on road construction vehicles are becoming a main hindrance in the present urban traffic situations, due to their severe impact in terms of reducing the average traffic flow on the specified urban roads. In this way, the existence of heavy construction vehicles on the urban traffic scenes significantly affects the traffic flow characteristics to a larger extent. In order to solve these issues, the accurate detection followed by tracking of heavy-construction vehicles in urban traffic roads are very much compulsory so that the traffic in congested urban roads can be controlled as well as monitored up to certain extent on-road traffic situations. The organization of the article is described as follows. Section I gives the introduction of the topic. Section II explains the existing literature survey. Section III gives the methodology frame work of the proposed Single Shot Detector (SSD)-based object detection framework, which can detect the

construction vehicles in traffic. Section IV shows the results and discussion. Section V concludes the paper with future work.

II. RELATED WORK

From the past few decades, a huge number of attempts is made in the literature towards identification and tracking of moving vehicles in urban traffic scenarios. For instance, Wang et al. [1] presented spatio-temporal features-based system for vehicle detection followed by the type classification of vehicles. In this approach, first moving objects are detected using spatio-temporal features-based algorithm, which then classifies type of detected objects by utilizing features fusion methods. Though this method performs better in general scenes; yet, it fails for identifying complex vehicle models as well as poorly illuminated traffic scenes. Further, Song et al. [3] introduced a visual feature- based vehicle detection and counting system, which can be employed in highway traffic scenarios. Precisely, in this framework, the defining of the road surface area consists of a remote area and also a proximal area. In each frame, the two road areas are sequentially detected in order to get reasonable detection results in the monitoring field. The authors employed ORB feature extraction algorithm, to predict the position of the object in the image, which is further analysed, so that the vehicle trajectory of different objects can be calculated.

Iwasaki et al. [4] proposed a robust vehicle detection framework, which uses IR thermal camera based thermal images for detecting vehicle positions followed by their respective movements. Praveen et al. [5], introduced a Gaussian mixture models-based approach for vehicle tracking followed by the speed estimation of the moving vehicles in various traffic surveillance kind of applications. However, this approach fails to attempt towards the detection of construction vehicles in traffic-prone urban roads. Further, very recently in 2020, Afrin and Yodo [6] presented a summative survey of various road traffic congestion measures that primarily contribute towards resilient and also sustainable transportation systems. According to their study, it is suggested that the management of traffic in the work zone areas are very much important in order to control congestion especially at peak hours of movements. In this aspect, they suggest that, the work zone should be planned cautiously in terms of ramp meters, computerized lane usage systems, coordinated traffic control plans and controlling traffic signals, which in turn could be useful to reduce the traffic congestions during peak hours.

Recently, Ankit Gupta et al. [3] in 2019, presented a detailed study, which indicated the impact of slowly moving vehicles on the capacity of crowed urban roads. Specifically, in their study, the authors utilized various road links from the urban arterial network of Varanasi, which though having widened road lanes, yet poses congestion challenges regularly due to the poor traffic management aspects in the Varanasi city. Further, they considered passenger car units as the basic unit of measuring highway capacity in terms of experimental and also direct empirical approaches and thereby performed an analysis on the impact of dynamic behavior of passenger car units. However, this study mainly concentrates on different modes of corrosion-induced failures in the reinforced concrete structures and their impacts on service life of the simple rectangular beams. Furthermore, Ji et al. [7] presented a video-based construction vehicles detection framework, which can detect hydraulic excavators and dump trucks on state-owned land areas. Precisely, the authors introduce detection techniques using ROI of inverse valley features of mechanical arm as well as spatial-temporal reasoning for identifying hydraulic excavators. However, their system employs videos captured from stationary cameras and focuses mainly on traffic scenes on state-owned land areas.

To summarize, the existing state-of-the art techniques are focusing primarily towards the detection and tracking the speed of moving vehicles in urban traffic roads [8], [9], yet not much efforts are done for the detection of heavy construction vehicles in traffic-prone roads. From another perspective, construction vehicles such as road rollers, trenchers and bulldozers significantly aggravate the congestion in urban roads during peak hours because of their deadly slow movement rates accompanied by their occupation of majority of road portions. Due to these reasons, promising frameworks are very much important, which can identify the heavy-construction vehicles moving in urban traffic-prone roads so that appropriate congestion evaluation strategies can be adopted to monitor traffic situations. Further, though vehicle detection on highway lanes are studied extensively on the literature, yet to the best of our knowledge, detection frameworks for identifying heavy-construction vehicles on urban traffic scenes are not fully explored.

A. Motivation and Contributions

This article presents a deep-learning based detection framework, which identifies the heavy construction vehicles from the urban traffic scenes by making use of Single Shot Detector[10](SSD)-based object detection using convolutional neural networks. Specifically, the proposed construction vehicle detection framework is named as, "Deep-Learning Based Detection Framework", abbreviated as DLDF, employs SSD technique in order to accurately identify the construction vehicles present in the moderate to heavily congested traffic environments. More specifically, the main contributions of the proposed DLDF are given by,

- A brand-new SSD-based deep learning network with feature extraction and detection modules is created by the combination of SSD and convolutional layers. Precisely, a new SSD network of size 122 layers starting from input layer to till softmax layers is generated in order to detect heavy construction vehicles from traffic scenes.

- The proposed SSD-based DLDF is trained and tested by employing a rich set of databases (app. 137743 images+) of moving vehicles, which consists of traffic scenes collected from three different datasets including benchmark MIO-TCD localization dataset [11], VISAL Dataset [12] and so on.

- The performance evaluations of the proposed DLDF with state-of-the art technique to demonstrate the better predictions along with time efficiency comparisons.

III. METHODOLOGY OF PROPOSED FRAMEWORK

Fig. 1 shows the block diagram of proposed Single Shot Detector (SSD)-based object detection framework, which can detect the construction vehicles in traffic scenes by making use of deep learning based automatically learned image features. Specifically, the proposed construction vehicle detection framework is named as, "Deep-Learning Based Detection Framework", abbreviated as DLDF, employs SSD technique in order to accurately identify the construction vehicles present in the heavily congested traffic environments. More specifically, the proposed DLD framework consists of two stages namely, training as well as testing stages as shown in Fig. 1.
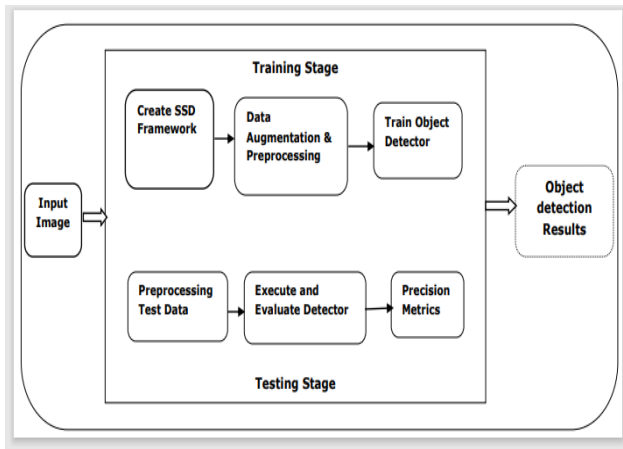

Fig. 1. Block Diagram of Proposed DLDF Framework.

Initially the input image is fed in to the training stage of the framework, which starts with creation of SSD framework module. Precisely, the SSD object detection network is created which consists of two sub-networks namely, a feature extraction network and a detection network. More precisely, the feature extraction network is generated by employing a pretrained CNN such as Mobile Net whereas detection sub-network is developed by composing SSD-specific layers and few convolutional layers. The SSD layers are used in order to specify the several significant input parameters to the proposed SSD network including input size, number of classes, size of training images and so on. After the SSD network is created, data augmentation and preprocessing is carried out in order to enhance the network accuracy by randomly transforming the original training data. Precisely, the transformations such as random flipping, random scaling and jittering image color are carried out during data augmentation and preprocessing module, in order to increase the variety of training samples. After the preprocessing module is completed, the SSD detector is trained as per the training parameters such as max epochs and initial learning rate as specified during the network creation stage.

In the testing stage, initially the preprocessing transformations are applied followed by the execution of detector on the test images. Then the resultant detection results are evaluated by means of precision and recall metrics. Then the object detection results are indicated in the form of outputs containing the bounding boxes, scores, and the labels for vehicles detected in the image.

IV. EXPERIMENTAL SETUP AND DATABASE CREATION

The performance of the proposed DLD framework is evaluated on three different datasets as given by.

- The 2017 MIO-TCD localization dataset [11].

- VISAL Dataset [12].

- Web-source Traffic videos, which are illustrated as follows.

1) 2017 MIO-TCD localization dataset [11]:" Miovision traffic camera dataset" (MIO-TCD) is one of the bench-mark dataset widely used in traffic analysis incorporating motor vehicles. It includes 11 traffic object classes such as buses, trucks, pickup trucks, work vans and pedestrians. It contains 7,86,702 annotated images captured at different timings of the day by hundreds of traffic surveillance cameras that are deployed in Canada and the United States. The 2017 MIO-TCD localization dataset contains 137,743 high resolution images each consisting of one or more foreground objects among the predefined object classes. Specifically, the 110,000 training images and 27,743 testing images from this dataset are utilized in order to evaluate the performance of the proposed framework.

2) VISAL dataset [12]: This Video, Image, and Sound Analysis Lab (VISAL) dataset is created for highway traffic video classification purpose, which consists of set of highway traffic videos ranging from low, medium, or high traffic scenes.

3) Openly available traffic videos: Different images of construction vehicles are extracted from web-based traffic videos are considered for the experimentation purpose in the proposed framework. Specifically, 10+ categories of construction vehicles dumping truck, bulldozers, Excavators, grader and front-end loader images from low-density as well as high-density traffic roads are considered for evaluation purposes.

Fig. 2 shows the snapshot of sample database images that are considered for training and testing stages of the proposed DLD framework as given below. The proposed DLD framework is evaluated in HP -Pro laptop of Intel Core i5-2.71 GHz processor, 8 GB RAM, 64-bit OS with MATLAB environment. Initially, the performance of proposed SSD framework for detecting construction vehicles from traffic scenes is evaluating by means of creation of SSD-based deep learning network with feature extraction and detection modules. Specifically, in the proposed DLD framework, SSD network of size 122 layers starting with input to focalloss and softmax layers is generated in order to detect construction vehicles from traffic scenes. More specifically, Fig. 3(a) shows the snapshot of Layer-Graph (Lgraph) generated by the proposed DLD framework, which clearly indicates the first few layer's pictorial version, name of each layer along with type, activation function, weight and bias values. Similarly, Fig. 3(b) illustrates the snapshot of last few layers along with all other specifications of each layer.

Fig. 2.   Snapshot of Sample Database Images.

V.   RESULT AND DISCUSSION

For evaluation purpose, the proposed deep-learning based Detection framework DLDF is compared with state-of-the art approach introduced by Ji et al. [4] by considering similar kind of experimentation scenarios. Specifically, the efficiency and effectiveness of the proposed DLDF framework is compared with that of the reference method, which are mentioned in subsequent discussions as follows:

1) Method 1: proposed DLDF method -indicated as DLDF Method.

2) Method 2: ROI-based method by Ji et al. [4] - indicated as Ref.Method illustrated as follows.

Fig. 3 and 4 shows the snapshot of the last few layers used in Lgraph with specifications for the analysis used for training. Fig. 5 shows the detection results of proposed DLDF method by means of bounding box and label. Specifically, the heavy construction vehicle named 'Excavator' moving in a field-side road is exactly detected by proposed DLDF technique in terms of bounding box with suitable dimensions as well as confidence level scores. More specifically, the proposed DLDF framework accurately identifies the Excavator vehicle by means of bounding box of dimensions (101, 53,147,114) and confidence level score of value 75.56%. Further the total time taken by the DLDF framework for detection of this vehicle is 7.872 seconds including feature extraction and mapping stages.

Fig. 6 presents the detection results of proposed DLDF method in terms of yellow-colour two bounding boxes and labels. Specifically, the heavy construction vehicle named 'Roller' moving in a field-side road is exactly detected by proposed DLDF technique in terms of two bounding boxes of dimensions (91,28,141,139) and (23,26,177,150) respectively

whereas confidence level scores of corresponding boxes are 83.05% and 67.01% respectively. However, the size of Roller vehicle seen in the input image is more than the threshold boundary of detection boxes; therefore, the complete vehicle is mapped into two bounding boxes, with the primary box indicating higher confidence results when compared with that of secondary bounding box.
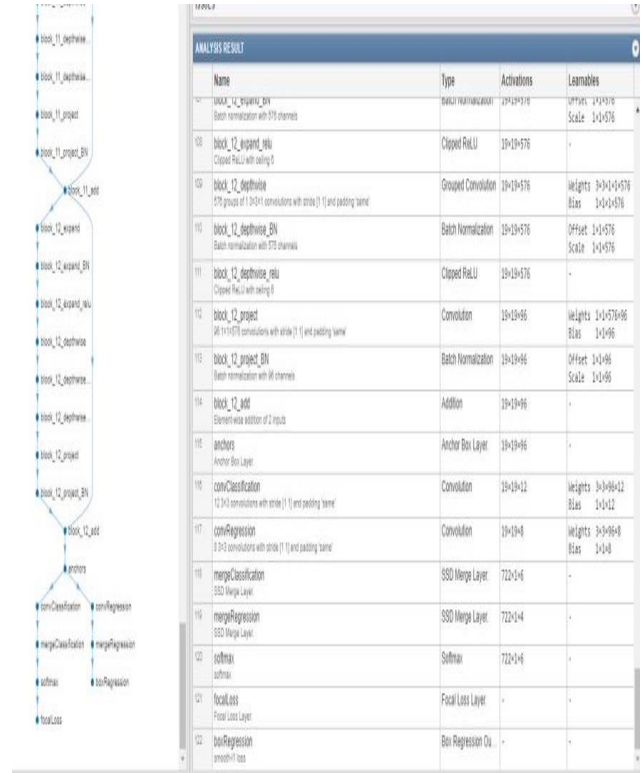

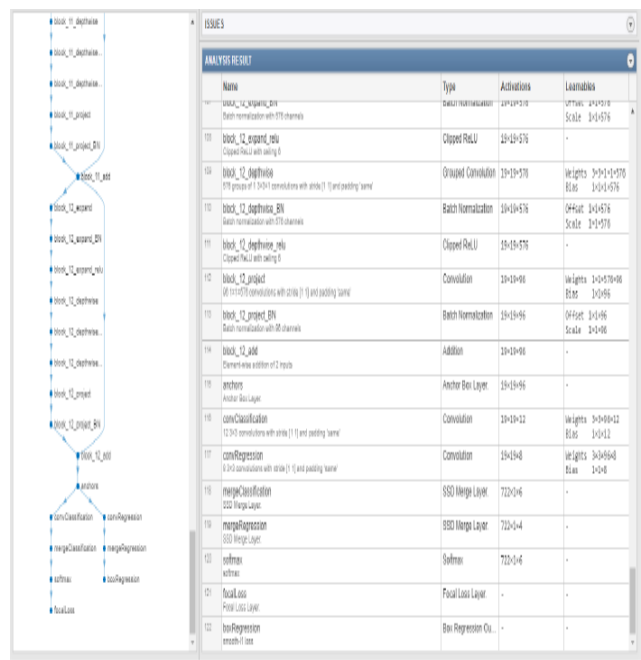Fig. 3.   Snapshot of Layer-Graph (Lgraph).

Fig. 4.   Snapshot of Last Few Layers of Lgraph with Specification.

Fig. 5.   Detection Results for Excavator Vehicles.

Fig. 7 indicates the construction vehicle detection results of proposed DLDF framework in terms of 'Vehicle' label followed by a bounding box. Precisely, the heavy construction vehicle named UNAC 'Trencher' moving on road-side is exactly detected by proposed DLDF technique in terms of bounding box with suitable dimensions as well as confidence level scores. More precisely, the proposed DLDF framework accurately identifies the Trencher vehicle by means of Bounding box of dimensions (32, 43,156,128) and confidence score of value 64.02%. There is a slight decrease in confidence scores are due to the presence of external objectives within scope of observation and also the movement of vehicle at considerably farther distance.

Fig. 6.   Detection Results for Roller Vehicles.

Fig. 7.   Detection Results of UNAC' Trencher 'Vehicles.

Fig. 8.   (a) Ground Truth Traffic Scene (b) Detection Results of Ref.Method (c) Detection Results of DLDF Method.

Fig. 8 illustrates detection performance of the proposed Deep-learning based detection framework (DLDF) by comparing with the corresponding results of ROI-based method reference method [4]. Specifically, Fig. 8(a) shows the ground truth traffic scene taken from the experimental dataset, which depicts slightly busy traffic road with various kinds of moving vehicles including cars and a construction truck vehicle. Fig. 8(b) indicates the detection results of reference method in terms of White-colored regions on the resultant image. Fig. 8(c) shows the detection results of the proposed DLDF method, in which the construction vehicle is exactly detected, even though it is moving at a quite reasonably at a farther distance. In this way, the better detection results of proposed DLDF method can be clearly observed when compared with the reference method by means of labeling and bounding boxes.

Further, the efficiency of the proposed DLDF framework is evaluated by considering the total time consumption of all activities followed by the subsequent comparisons with the reference method, which is detailed Table I. Table I results are indicated as follows, the time taken by the proposed DLDF method for feature extraction is 2.480 seconds followed by total prediction time is 4.875 seconds whereas the time taken by Ref. method for feature extraction is 2.466 seconds followed by total detection time is 4.553 seconds respectively. Though, the time taken by the proposed DLDF method is slightly high when compared with that of Ref. method, yet it shows considerably better performance in terms of detection results as shown in Fig. 8(c). Furthermore, the confidence level of proposed framework scores 50.93% for the given traffic scene, even capturing at a faraway distance, while the ref. method achieves only 35.60% for the same scene. In this way, it is observed from the detection results that the proposed framework performs reasonably better compared to the reference technique in terms of good confidence scores.

TABLE I. TIME EFFICIENCY COMPARISON RESULTS (IN SECONDS)

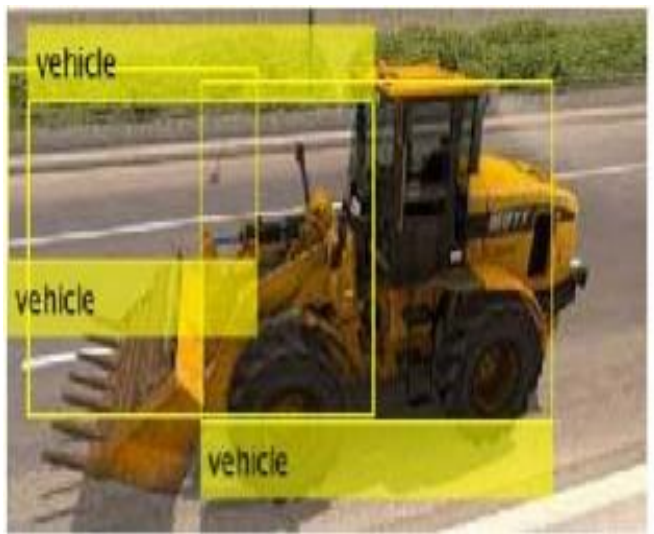| Task	Proposed DLDF	Ref. Method
Features extraction	2.480	4.875
Total Time	2.466	4.553

Fig. 9. Detection Results of Front-End Loader Vehicle.

Fig. 9 shows the vehicle detection results of proposed DLDF framework in terms of 3 'Vehicle' labels followed by a respective bounding box. Specifically, the heavy construction vehicle named 'Front End Loader' moving on road-side is exactly detected by proposed DLDF technique in terms of 3 bounding boxes with suitable dimensions and confidence level scores as shown in Table II. More specifically, it is observed in Table II results that the bounding box with dimensions [11,29,167,99] identifies larger portion of the vehicle, which results in confidence score of 66.98%, whereas remaining two boxes cover slightly lesser portions of vehicle, hence result in lesser confidence scores respectively. Since in the middle bounding box, the vehicle coverage is more due to which the detection confidence scores are also increased as shown in Table II.

Fig. 10 indicates the vehicle detection results of proposed DLDF framework by means of a label followed by the respective bounding box. Precisely, clustered scene with two different construction vehicles named - 'Bulldozer' and 'TLB', which are moving on a hilly-road side is considered for evaluation purpose. Though the test image includes two different construction vehicles, yet the proposed DLDF method is able to identify the vehicles at a reasonably good confidence level of 63.97%. However, it can be observed that, since the input image is slightly cluttered in terms of including body parts of two different Vehicles, which may complicate the detection process. Due to these reasons, the proposed DLDF detects it as single vehicle, since it combines the front portion of one vehicle with side portion of another type vehicle.

Fig. 11 shows the vehicle detection results of proposed DLDF framework for the test image, in which blurred version of vehicle can be observed. Precisely, it can be noticed that, in the test image, the image quality is very low followed by a lot

of occurrence of overlapping on the construction vehicle, due to which the proposed DLDF fails to detect the construction vehicle present in the image. If noise and overlapping are eliminated from the test image, then the proposed DLDF can be employed to detect the presence of vehicle.

TABLE II. DETECTION RESULTS

Bounding boxes- dimensions	Confidence scores (in %)
[1 19 121 61]	54.40
[11 29 167 99]	66.98
[94 23 171 107 99]	54.55

Fig. 10. Detection Results of Bulldozer, TLB Vehicles.

Fig. 11. Detection Results of Blurred Version of Vehicles.

VI. CONCLUSION AND FUTURE WORK

In this article, a new deep-learning based detection framework, which employs Single Shot Detector (SSD)-based object detection system consisting of CNNs is proposed for detecting heavy-construction vehicles on urban traffic scenes. The experimental results carried out on three different datasets including the benchmark ones, clearly demonstrate the enhanced performance of the proposed detection framework in terms of confidence scores and time efficiency when compared to the existing techniques. In future, the proposed framework can be successfully employed on intelligent transportation systems for monitoring congested conditions of urban traffic situations.

REFERENCES

[1] Yu Wang, X. Ban, H. Wang, DI Wu, Hao Wang, Shouqing Yang, Sinuo Liu and Jinhui Lai, Detection and Classification of Moving Vehicle from Video Using Multiple Spatio-Temporal Features, in IEEE Access, special section on recent Advances in Video Coding and Security, Vol.7, Pages 80287-80299, https://doi.org/10.1109/ACCESS.2019.2923199 2019.

[2] Sreelatha R and Roopa Lakshmi R, A Novel Framework for Detecting Slow Moving Vehicles in Urban Traffic Using Robust Visual Features, R, Sreelatha and R, Roopa Lakshmi, A Novel Framework for Detecting Slow Moving Vehicles in Urban Traffic Using Robust Visual Features, in proc. ICSMDI 2021- available at SSRN digital library: http://dx.doi.org/10.2139/ssrn.3854305.,pages 1-10, 2021.

[3] H. Song, H. Liang, H. Li, Z. Dai and X. Yun, "Vision-based vehicle detection and counting system using deep learning in highway scenes", in European Transport Research Review Journal, Vol. 11:51, Pages 1-16, https://doi.org/10.1186/s12544-019-0390-4, 2019.

[4] Y. Iwasaki, M. Misumi and T. Nakamiya, Robust Vehicle Detection under Various Environments to Realize Road Traffic Flow Surveillance Using an Infrared Thermal Camera, The Scientific World Journal, Hindawi Publishing Corporation, Vol.1, Pages 1-12, https://doi.org/10.1155/2015/947272 ,2015.

[5] Praveen M Dhulavvagol, Abhilash Desai and Renuka Ganiger, Vehicle Tracking and Speed Estimation of Moving Vehicles for Traffic Surveillance Applications, in proc. of ICCTCEEC-2017, Pages73-377, https://doi.org/10.1109/CTCEEC.2017.8455043 2017.

[6] Tanzina Afrin and Nita Yodo," A Survey of Road Traffic Congestion Measures towards a Sustainable and Resilient Transportation System", in MDPI journal of sustainability, Vol 12, Pages 1- 23, https://doi.org/10.3390/su12114660, 2020.

[7] Wenyang Ji, Lingjun Tang, Dedi Li, Wenming Yang, Qingmin Liao," Video-based construction vehicles detection and its application in intelligent monitoring system", in proc. of Elsevier CAAI Transactions on Intelligence Technology, vol.1, pages 162-172, https://doi.org/10.1016/j.trit.2016.09.001, 2016.

[8] Kahlil Muchtar, Nasaruddin, Afdhal and IndraNugraha, "Attention-based Approach for Efficient Moving Vehicle Classification", in proc. of Elsevier Procedia computer science, ICCSCI-2019,Pages683-691, doi /10.1016/j.procs.2019.08.217, 2019.

[9] Huansheng Song, Haoxiang Liang, Huaiyu Li, Zhe Dai and Xu Yun, Vision-based vehicle detection and counting system using deep learning in highway scenes, in the journal of European Transport Research Review, 2019.

[10] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng Yang Fu, and Alexander C. Berg." SSD: Single shot multibox detector." In 14th European Conference on Computer Vision, ECCV 2016. Springer Verlag, 2016.

[11] Zhiming Luo, Frdric Branchaud-Charron, Carl Lemaire, Janusz Konrad, Shaozi Li, Member, Akshaya Mishra, Andrew Achkar, Justin Eichel, and Pierre-Marc Jodoin, MIO-TCD: A New Benchmark Dataset for Vehicle Classification and Localization, in IEEE Trans. on Image Processing, Vol. 27, No.10,pages5129-5141, doi.org/10.1109/TIP.2018.2848705 2018.

[12] Antoni B. Chan and Nuno Vasconcelos, Probabilistic Kernels for the Classification of Auto-regressive Visual Processes, In IEEE Conference on Computer Vision and Pattern Recognition(CVPR), San Diego, Jun 2005