Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 2, No. 4, February : 2024

ROAD TRAFFIC VEHICLE DETECTION AND TRACKING THROUGH DEEP LEARNING USING BOTH PUBLICLY AVAILABLE AND CUSTOM-COLLECTED DATASETS

Mrs.T.SRUJANA Assistant professor, Department of ECE, Sree Dattha Institute Of Engineeringand Science, Hyderabad

Abstract : Intelligent vehicle detection and counting are becoming increasingly important in the fieldof highway management. However, due to the different sizes of vehicles, their detection remains achallenge that directly affects the accuracy of vehicle counts. To address this issue, this paper proposes a vision-based vehicle detection and counting system. A new high-definition highway vehicle dataset with a total of 57,290 annotated instances in 11,129 images is published in this study. Compared with the existing public datasets, the proposed dataset contains annotated tiny objects in the image, which provides the complete data foundation for vehicle detection based on deep learning. Therefore, this work is focused on implementation of You Only Look Once (YOLO-V4) based DeepSORT model for real time vehicle detection and tracking from video sequences. Deep learning based Simple Real time Tracker (Deep SORT) algorithm is added, which will track actualpresence of vehicles from video frame predicted by YOLO-V4 so the false prediction perform by YOLOV4 can be avoid by using DeepSort algorithm. The video will be converted into multiple frames and give as input to YOLO-V4 for vehicle detection. The detected vehicle frame will be further analysed by DeepSort algorithm to track vehicle and if vehicle tracked then DeepSort will put bounding box across tracked vehicle and increment the tracking count. The proposed model istrained with three different datasets such as COCO, Berkeley and Dash Cam dataset, where Dash Cam dataset is the custom collected dataset.

IndexTerms **–** Vehicle detection, tracking-by-detection, YOLO, DeepSORT, road traffic data

I.INTRODUCTION

Vehicle detection and tracking is a common problem with multiple use cases. Government authorities and private establishment might want to understand the traffic flowing through a place to better develop its infrastructure for the ease and convenience of everyone[1]. A road widening project, timing the traffic signals and construction of parking spaces are a few examples where analysing the traffic is integral [2]. Traditionally, identification and tracking has been carried out manually. A person will stand at a point and note the count of the vehicles and their types. Recently, sensors have been put into use, but they only solve the counting problem. Sensorswill not be able to detect the type of vehicle.

A fundamental source of the economic growth of any nation depends on well-planned and resilient transportation systems based on spatial information. Regardless, most cities around the world are still facing a rampant increase in traffic volume and complications in traffic management, resulting in poor quality of life in modern cities[3]. However, recent advancements in internet bandwidth, artificial intelligence, and sensing technologies have minimized these difficulties by collaboratively bringing forward location intelligence for public safety. Automation in location intelligence in road environments using sensing technologies allow authorities to achieve resiliencein road safety, controlled commutes, and assessments of road conditions[4].

II.RESEARCH METHODOLOGY

Metrics that describe, he quality and key characteristics in numerous object tracking systems must bestudied and compared in accordance to carefully analyse and evaluate their performance. Regrettably, there has yet to be agreement on such a range of generally valid measures. They present two new measures for evaluating MOT systems inthis paper. Multiple objects tracking precision (MOTP) as well as multiple objects trackingaccuracy (MOTA) are

suggested benchmarks that can be used for a variety of monitoring activities & permit for objective contrast of tracking systems' primary features, like accuracy at locating targets, precision at recognizing target configurations, but also way todetect targets on consistent bases. They put the proposed metrics to the test in a series of global evaluation workshops to see how useful and expressive they were. The CLEAR workshops in 2006 and 2007 featured a wide range of monitoring activities whereby a bignumber of models were tested & evaluated. Their studies findings reveal that its suggestedmeasures accurately reflect the numerous methods' qualities and shortcomings in a simpleand direct manner, helps in easy evaluation in performance, thus relevant toward a wide range of circumstances[1].

As the performance in object detectors increases, the foundation for a tracker becomes significantly more trustworthy. The problems for a successful tracker havechanged as a result of this, as well as the increased use of higher frame rates. As a result ofthis shift, considerably simpler tracking algorithms may now compete with more complexsystems for a fraction of the processing cost[3]. This paper outlines and illustrates such a method by conducting extensive tests with a range of object detectors. The proposed technique can easily operate at 100K fps on the DETRAC vehicle tracking dataset, beatingthe state-of-the art. The notion of a passive detection filter is used to analyse a very simple tracking technique in this research. Due to its modest computing footprint, the suggested approach can serve as a basic predictive model for other trackers and provide an appraisal of the necessity of additional efforts in the tracking algorithm. It also permits reviewing tracking benchmarks to evaluate if the specific concerns they indicate (for instance, missed detections, frame rate, etc.) are within the capabilities of existing algorithms [4].

A target's visibility plan was learnt and used to infer its spatial attention plan. Characteristics are then weighted using the spatial attention map[5]. Furthermore, the occlusion state can be evaluated using the visibility mapping, thatregulates a continuous updating mechanism using weighted loss upon training samples having varying occlusion states over multiple frames [6]. It's possible to think of it as a temporal attention process. On the rigorous MOT15 and MOT16 benchmark datasets, the suggested approach obtains 34.3 percent and 46.0 percent in MOTA, respectively[7].

According to paper, hierarchical neural networks are difficult to be trained. Theyoffer a residual learning strategy for significantly deeper training models than earlier used models. Researchers specifically reformulate the levels as training residual methods with relation to the level’s inputs, rather than learning unreferenced functions[8]. Researchers presented significant experimental proof suggesting that residual networks are simpler to implement and that increasing complexity can boost performance[9]. Authors used theImageNet data to evaluate residual nets having put to 152 levels of complexity, that are 8 layers deeper to VGG networks whereas have less intricacy. The aggregation of such residue nets scores 3.57 % error just on ImageNet test set. The work won 1st spot with in ILSVRC 2015 classification problem[10]. CIFAR-10 study using 100 - 1000 levels is shown. The complexity in depictions is crucial for many image identification tasks. Only because of their exceptionally deep depiction can they accomplish an 28 % notable improvement on COCO object recognition samples. Their contributions towards ILSVRC & COCO 2015contests used deep residual networks as the basis, and thus won 1st spot in the ImageNet identification,ImageNet localization, COCO recognition, as well as in COCO segmentation tasks[11].

A fundamental source of the economic growth of any nation depends on well- planned and resilient transportation systems based on spatial information. Regardless, mostcities around the world are still facing a rampant increase in traffic volume and complications in traffic management, resulting in poor quality of life in modern cities [12]. However, recent advancements in internet bandwidth, artificial intelligence, and sensing technologies have minimized these difficulties by collaboratively bringing forward location intelligence for public safety. Automation in location intelligence in road environments using sensing technologies allows authorities to achieve resilience in

road safety, controlled commutes, and assessments of road conditions[13]. Vehicle detection and classification using deep learning (DL) and multi-object tracking (MOT) on video streams obtained .The first generation, YOLOv1, unified the methods of feature extraction, object localization, and classification to form a single-stage architecture.This network was SOTA in terms of mean average precision (map) with a fast detection speed. YOLO was a continuous series of convolutional layers with occasional max pool layers at its time. Among other changes, YOLOv2 eliminated the fully-connected layer at the end of YOLOv1, allowing the network to perform independently of image resolution[14].

III PROPOSED METHOD

The use of deep convolutional networks (CNNs) has achieved amazing success in the field of vehicle object detection. CNNs havea strong ability to learn image features and can perform multiple related tasks, such as classification and bounding box regression. The detection method can be generally divided into two categories. The two-stage method generates a candidate box of the object via various algorithms and then classifies the object by a convolutional neural network. The one-stage method does not generate acandidate box but directly converts the positioning problem of the object bounding box into a regression problem for processing.

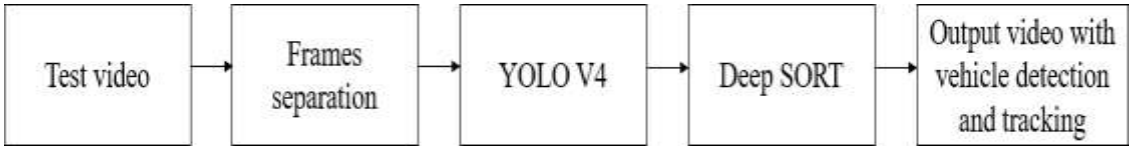

Fig 1: Proposed system block diagram

Therefore, this work is focused on implementation of You Only Look Once (YOLO-V4) based Deep SORT model for real time vehicle detection and tracking from video sequences. Deep learning based Simple Real time Tracker (Deep SORT) algorithm is added, which will track actual presence of vehicles from video frame predicted by YOLO-V4 so the false prediction perform by YOLOV4 can be avoid by using DeepSort algorithm. The video will be converted into multiple frames and give as input to YOLO-V4 for vehicle detection. The detected vehicle frame will be further analyzed by DeepSort algorithm to track vehicle and if vehicle tracked then DeepSort will put bounding box across tracked vehicle and increment the tracking count.The proposed model is trained with three different data such as COCO, Berkeley and Dash Cam dataset. Here, DashCam dataset is the custom collected dataset.

Dataset

COCO dataset: The dataset contains car images with one or more damaged parts. The image/ folder has all 80 images in the dataset. There are three more folders train/,val/ and test/ for training, validation and testing purposes respectively.

Berkeley dataset: Berkeley Deep Drive(link is external) (BDD) and Nexar announced the release of 36,000 high frame-rate videos of driving, in addition to 5,000 pixel-level semantics-segmentedlabeled images, and invited public and private institution researchers to join the effort to develop accurate automotive perception and motion prediction models.

Preprocessing and frame separation

Digital image processing is the use of computer algorithms to perform image processing on digital images. As a subfield of digital signal processing, digital image processing has many advantages over analogue image processing. It allows a much wider range of algorithms to be applied to the input data — the aim of digital image processing is to improve the image data (features) by suppressing unwanted distortions and/or enhancement of some important image features so that our AI-Computer Vision models can benefit from this improved data to work on. To train a network and make predictions on new data, our images must match the input size of the network. If we need to adjust the size of images to match the network, then we can rescale or crop

data to the required size.

YOLO V4

The test video from the datasets is taken to apply the algorithms and splitted into different frames. The spitted frames are applied to YOLOV4 to detect the vehicles and to track the vehicles, frames are applied to DeepSort. Finally, the output video is generated with vehicle detection and tracking.

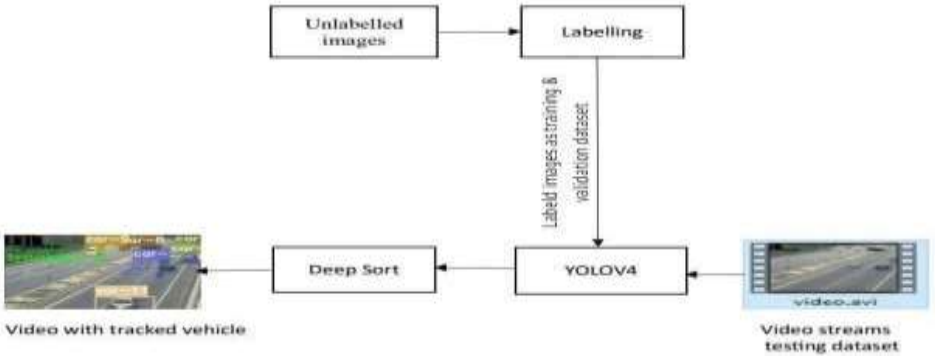

Fig 2: Methodology of proposed system

Tracking vehicles is another aspect of research in transportation. DeepSORT is a recent tracking algorithm, extending SORT (Simple Online and Real-Time) tracking algorithm. The original algorithm was developed considering MOT task. With the main goal of supporting online and real-time applications. This means that the tracker associates detected objects from previous and current frames only.

YOLO is a Convolutional Neural Network object detection system, that handles object detection as one regression problem, from image pixels to bounding boxes with their class probabilities. Its performance is much better than other traditional methods of object detection, since it trains directly on full images. YOLO is formed of 27 CNN layers, with 24 convolutional layers.

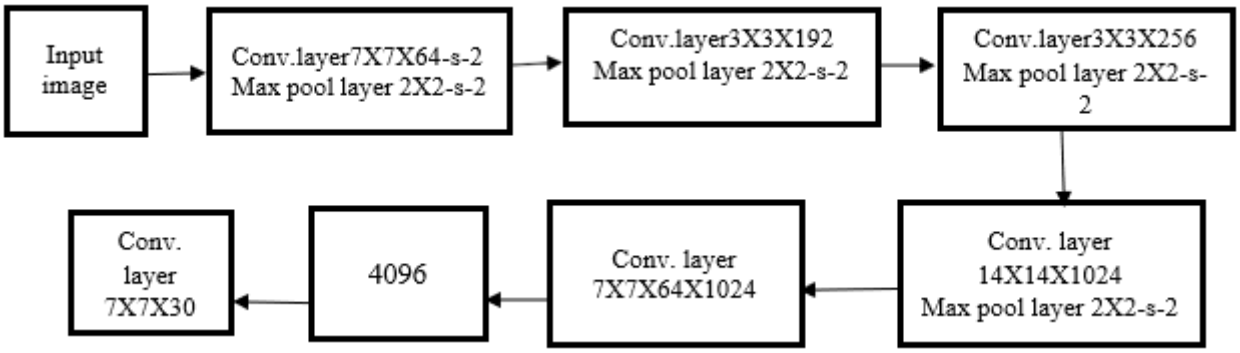

Fig.3 Structure of YOLOV4 algorithm

DeepSORT

DeepSORT algorithm is deployed for the purpose of tracking. The enhanced version of the algorithm where the association metric is substituted by an informed metric integrating motion and appearance information using Convolutional Neural Network. The algorithm takes the detection outputs from the previous stage and run tracking for each detected object. In tracking by detection scheme, the accuracy of tracking is based on the quality of detection results. The Kalman filter an important role in deep sort. It identifies noise in detecting and uses previous states to predict the closed frame surrounding the object best suited. Each time it detects an object it creates a track containing all the necessary information of that object it also tracks and deletes track with detection time exceeds a given threshold due to objects are out of frame. In addition to eliminate duplicates they set a minimum threshold value for detection in the first frame the next problem lies in association between new objects and new predictions from the Kalman filter.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 2, No. 4, February : 2024

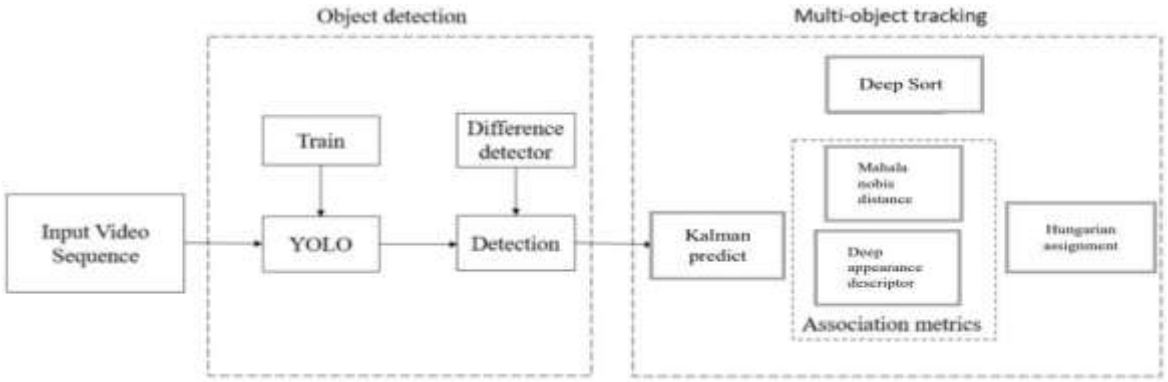

Fig. 4 DeepSort algorithm for multi-object Tracking.

DeepSORT which is an improved version of SORT is one of the most popular state-of-the-art objects tracking frameworks today. DeepSORT has integrated a pre-trained neural network to generate feature vectors to be used as a deep association metric. Since DeepSORT was developed focusing on the Motion Analysis and Re-identification Set (MARS) dataset, which is a large-scale video-based human reidentification dataset, it uses a feature extractor trained on humans which does not perform well on vehicles. Several state-of-the-art object detection and tracking algorithms including SORT and DeepSORT were deployed detect and track different classes of vehicles in their region of interest and it has been stated that the trackers did not perform ideally at predictingvehicle trajectories which resulted in ID switches during occlusions. A vehicle tracking fuses the prior information of the Kalman filter to solve the problem of vehicle tracking under occlusion. But it has been stated that the proposed method does not perform well if the target is lost for a longer period.

IV RESULTS AND DISCUSSION

To implement this project, we have designed following modules

1) Generate & Load YOLOv4-DeepSort Model: using this module we will generate and load YOLOV4-DeepSort model.
2) Upload Video & Detect Car & Truck: using this module we will upload test video and then apply YOLOV4 to detect vehicle and this detected vehicle frame will be further analyse by DeepSort to track real vehicles.

Fig 5. Detection and tracking of cars and trucks using YOLOV4 with DeepSORT algorithms.

Fig 6. Tracking of vehicles using Frame Per Second and application till the end of the video.

IV. CONCLUSION

In conclusion, the vehicle detection and tracking method presented uses TensorFlow library with DeepSORT algorithm based on YOLOv4 model. It can be proven that using YOLOv4 and YOLOv4-tiny is acceptable and faster than previous one. It can be use in Realtime surveillance camera in the highway or recording video to evaluate the number of vehicles pass by according to what time it started recorded to last recorded. This data then can be used for traffic management by implementing answer if the place proven a lot of congestion or not. It is the best to use YOLOv4model than previous model YOLOv3 if the system wants the highest accuracy with acceptable speed. If the system wants the best accuracy with the highest speed as possible because limitation in hardware or to process it in real-time, it is recommended to use YOLOV4-tiny model which it can achieve higher accuracy. This system can be improved to be more adaptable for vehicle detection if using several suggestion ideas. A vehicle tracking algorithm based on the framework suggested in DeepSORT which is capable of tracking the nonlinear motion of vehicles with a highlevel of accuracy. The proposed algorithm utilizes YOLOv4 with Darknet, an open-source neural network framework, for vehicle localization and identification. The number of detection errors wasminimized by optimizing the training of the detector through hyperparameter optimization and dataaugmentation.

REFERENCES

[1] Bernardin K, Stiefelhagen R (2008) Evaluating multiple objects tracking performance: the CLEAR MOT metrics. EURASIP Journal on Image and Video Processing, 2008, 1–10
[2] Bewley A, Ge Z, Ott L, Ramos F, Upcroft B (2016) Simple online and realtime tracking. In 2016 IEEE International Conference on Image Processing (ICIP) (pp. 3464– 3468). IEEE
[3] Bochinski E, Eiselein V, Sikora T (2017) High-speed tracking-by-detection without using image information. In 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 1–6). IEEE.
[4] Chen Y, Wang J, Chen X, Sangaiah AK, Yang K, Cao Z (2019) Image super-resolution algorithm based on dual-channel convolutional neural networks. Appl Sci 9(11):2316
[5] Chen Y, Wang J, Xia R, Zhang Q, Cao Z, Yang K (2019) The visual object tracking algorithm research based on adaptive combination kernel. J Ambient Intell Humanized Comput 10(12):4855–4867
[6] Chu Q, Ouyang W, Li H, Wang X, Liu B, Yu N (2017) Online multi-object tracking using CNN-based single object tracker with spatial-temporal attention mechanism. In Proceedings of the IEEE International Conference on Computer Vision (pp. 4836–4845)
[7] Ciaparrone G, Sánchez FL, Tabik S, Troiano L, Tagliaferri R, Herrera F (2020) Deep

learning in video multi-object tracking: A survey. Neurocomputing 381:61–88
[8] Fan, D. P., Wang, W., Cheng, M. M., & Shen, J. (2019). Shifting more attention to video salient object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 8554–8564).
[9] He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770–778).
[10] R. WEIL, J. WOOTTON AND A. GARCIA-ORTIZ Advanced Development Center, Systems & Electronics Inc. 201 Evans Lane, Saint Louis, MO 63121-1126, U.S.A.
[11] L. Bhaskar, A. Sahai, D. Sinha, G. Varshney, and T. Jain, "Intelligent traffic light controller using inductive loops for vehicle detection," Proc. 2015 1st Int. Conf. Next Gener. Comput. Technol. NGCT 2015, no. September, pp. 518–522, 2016.
[12] M. Aqib, R. Mehmood, A. Alzahrani, and I. Katib, "InMemory Deep Learning Computations on GPUs for Prediction of Road Traffic Incidents Using Big Data Fusion," 2020, pp. 79–114.
[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.
[14] C. Wang, A. Musaev, P. Sheinidashtegol, and T. Atkison, "Towards Detection of Abnormal Vehicle Behavior Using Traffic Cameras," in Big Data -- BigData 2019, 2019, pp. 125–136.