Vehicle Classification and Counting From Recorded Video using using YOLOv8 and DeepSORT

Md. Oleraj Hossin[1], Ishmoth Ura Nuri[2], Joy Adhikary[3], Bushrat Jahan Nisha[4], Nazrul Islam[5]
[1,6]Department of Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka, Bangladesh
*Corresponding Author: 18192103072@cse.bubt.edu.bd

***Abstract*—In everyday we use different types of transportation medium. The Collections of those transportation medium data of traffic and parking are essential for transportation and parking agencies to make informed decisions regarding planning, supervising, security and operation of a safe and efficient transportation system. In every country there are many type of transportation medium such as bus, truck, private car, cycle, bike and etc. So some time, we need to take decision-making step for some purpose. But the collection of transportation data is crucial to data-driven decision-making. So using Artificial Intelligence for such data collection efforts could save substantial time and cost and help to improve transportation safety and efficiency. Collecting, processing, and interpreting data often require significant manual efforts. Furthermore, the collection of real-time information about vehicles, traffic and parking conditions also be a crucial part of this project. Artificial Intelligence/machine learning (AI/ML) and computer vision-based data collection could provide a safer, faster, and more cost-effective alternative. The availability of many open-source object detection algorithms like YOLO, DeepSORT, in recent years has made it possible to detect and classify real time object detection.**

***Index Terms*—Vehicle Classification, Vehicle counting, Deep Learning, YOLOv8, DeepSORT**

I. Introduction

Video surveillance equipment has become an important element in applications for traffic control and parking management due to the rapid improvements in multimedia capture technology. In order to facilitate effective traffic management, researchers have developed a wide range of vehicle detection and tracking algorithms based on video surveillance footage.

Many traditional approaches for object detection have been explored, S. P. Dhole et al. in [1] is one such widely used technique in object detection aimed at detecting regions of images with consistent properties. The requirement for a distinct background and foreground separation as well as lighting circumstances limit this type of detection. Blob extraction for object detection has several disadvantages, one of which is the unpredictability of detections because shadows and neighboring objects frequently tend to be associated with the item.

For many years, object tracking has also been a topic of attention. Among the conventional methods of object tracking found in the literature, optical flow-based tracking techniques, as proposed by L. Kurnianggoro et al. in [2], have been frequently used to estimate the instantaneous velocity of pixel motion utilizing spatial and temporal brightness fluctuations. The assumptions of brightness constancy and spatial coherence limit this approach to object tracking. According to S. M. K. C. S. B. Egodawela et al., the primary limitation of optical flow is that it can only identify two orthogonal directions of motion.

Researchers are now paying more attention to deep learning-based frameworks because, despite being data-hungry, they show promise in modeling complex nonlinearities, which are prevalent in most real-world situations. Convolutional neural networks (CNNs), which are constructed using multiple building blocks such as convolution layers, pooling layers, and fully connected layers. In contrast, as mentioned earlier, traditional methods of object detection and tracking utilize traditional video processing techniques which perform poorly under ambient lighting changes, occlusions etc.

In 2001, the first real-time detection of human faces was developed (Viola and Jones 2001), which was much faster and more accurate than earlier algorithms (Zou et al. 2019) [3]. However, object detection started to evolve at an unprecedented speed when Regions with Convolutional Neural Network (R-CNN) was proposed in 2014 (Girshick et al 2014) [4]. Nowadays, it can be done in milliseconds using a single neural network trained to detect multiple objects.

A cutting-edge object recognition system called YOLO, introduced by J. Redmon et al. in [4], is based on a single CNN and is capable of predicting multiple bounding boxes. Given sufficient training data, coordinates and class probabilities can be calculated with an acceptable degree of precision. YOLO views the detection of objects as a regression issue. It divides the input image into grid cells and forecasts bounding boxes, confidence intervals for those bounding boxes, and conditional class probabilities for each grid cell. The class-specific confidence scores for each box are calculated using the confidence and the conditional class probabilities. According to experimental findings, YOLO is more than twice as accurate as earlier real-time detection studies. The previous YOLO frameworks served as the foundation for YOLOv2, YOLOv3, YOLOv4, YOLOv5 and YOLOv7, each of which improved upon them in terms of speed and accuracy. Current version of YOLO is YOLOv8 which is more accurate and more faster than all YOLO version.

DeepSORT [5] One of today's most well-liked cutting-edge object tracking frameworks, an upgraded version of SORT. In order to provide feature vectors for use as a deep association measure, DeepSORT has integrated a pre-trained neural network. While DeepSORT was created with the Motion Analysis and Re-identification Set (MARS) dataset in mind, a sizable dataset for video-based human reidentification, it makes use of a feature extractor that was built for humans but does not work well for automobiles.

The key contributions of our paper are:

- Firstly, Collect survilance video footage and extracted the data related to transporataion medium.
- Secondly, Determine the types of vehicle and and find pre-trained dataset to train YOLO Model.
- Then, predicted the vehicles from the video that collected from servilance camera Using YOLO and DeepSORT.
- Process a pre-train model and run it with YOLO and DeepSORT algorithm to train and classify the vehicles and count the specific types of vehicles.
- Finally, Got the output that identify and keep record of vehicles.

The remainder of the research work is arranged as follows: Section II depicts the related work of our study and III provide framework of our study and full methods for the proposed YOLO and DeepSORT architecture used to Classify vehicle. Section IV contains the empirical results and analysis, and Section V brings the work to be concluded.

II. Related Work

Cheng-Jian Lin [6] study presents a real-time traffic monitoring system based on a virtual detection zone, Gaussian mixture model (GMM), and YOLO to increase the vehicle counting and classification efficiency. GMM and a virtual detection zone are used for vehicle counting, and YOLO is used to classify vehicles.In this study, the Montevideo Audio and Video Dataset (MAVD), the GARM Road-Traffic Monitoring data set (GRAM-RTM), and our collection data sets are used to verify the proposed method. YOLOv4 achieved the highest classification accuracy of 98.91% and 99.5% in MAVS and GRAM-RTM data sets, respectively. Moreover, the proposed method with YOLOv4 also achieves the highest classification accuracy of 99.1%, 98.6%, and 98% in daytime, night time, and rainy day, respectively. But the average absolute percentage error of vehicle speed estimation with the proposed method is about 7.6% .and vehicles appearing in the video are assumed to be inside the virtual detection zone; thus, the width of the virtual detection zone should be sufficiently large for counting the vehicles. In the future work, we will focus on algorithm acceleration and model simplification.
"An Improved YOLO v2 for Vehecle Detection" [7] In this paper, by improving YOLOv2, a model called YOLOv2-Vehicle was proposed for vehicle detection. To obtain better anchor boxes, the vehicle bounding boxes on the training data set were clustered with k-means clustering, and six anchor boxes with different sizes were selected. Next, the loss function was improved with normalization to decrease the influence of the different scales of the vehicles. Based on the experimental results, them YOLOv2-Vehicle could reach 94.78%. In future work, we will collect more actual vehicle data to further study how to improve the accuracy and speed of vehicle detection.
Zhang et al. [8] research study used YOLOv4 object detection and DeepSORT tracking models to count and classify vehicles into 13 FHWA vehicle classes from video footages from existing roadside cameras. YOLOv4 was also used to detect the presence of rumble strips from roadway images and used to create a rumble strip inventory map. The trained vehicle counting/classification model achieves a counting accuracy of 97% at daytime and 91% at nighttime. The rumble strip detection model has 95% accuracy.
In Tanvir Ahmad. [9] paper, a modified YOLOv1 based neural network is proposed for object detection. 'e new neural network model has been improved in the following ways. Firstly, modification is made to the loss function of the YOLOv1 network.Secondly, a spatial pyramid pooling layer is added; thirdly, an inception model with a convolution kernel of 1 1 is added, which reduced the number of weight parameters . results are discussed and the network performance is checked using t-SNE visualisation tool, showing the extent to which the new network is able to extract rich features from images.Our modified network average detection rate is 65.6% and 58.7% on the Pascal VOC 2007 and 2012 dataset.YOLOv1 neural network based object detection by modifying loss function and adding spatial pyramid pooling layer and inception module with convolution kernels of 1 1. 'e new network is trained on an end-to-end method, and the extensive experiment on a challenging Pascal VOC dataset, 2007/2012, shows the effectiveness of the improved new network, with the detection results being 65.6% and 58.7%, respectively. 'e results of the proposed network have been compared with those of R-CNN and YOLOv1, from which the effectiveness of the proposed method is demonstrated. The difficulties of this paper is here small object can't be detected.
Mohammed Thakir Mahmood et al. Vehicle counting is an important process in the estimation of road traffic density to evaluate the traffic conditions in intelligent transportation systems. This form of information collection in intelligent systems is faced with low detection accuracy, inaccuracy in vehicle type detection, slow processing speeds. The YOLO mechanism can apply different machine or deep learning algorithms for accurate vehicle type detection. In this study researcher propose an infrared based technique to combine with YOLO for vehicle detection in traffic. This method will be compared with a machine learning technique of K-means++ clustering algorithm, a deep learning mechanism of multitarget detection and infrared imagery using convolutional neural network. [10] The technique was evaluated in classical methods. The other techniques have been used to detect vehicle in normal images. The findings of this survey show that the is potential in the future to develop various techniques based on YOLO to detect vehicle on infrared images.

III. Methodology

In this section, stages of the proposed system of vehicle classify and tracking algorithm are discussed. Video data were collected from a surveillance camera and training and validation pretrained data sets were collected officail website. YOLOv4 is used with COCO Dataset, an open-source neural network framework written in C and CUDA, for vehicle localization and identification. Now Discuss about proper methodology in this context.

A. Outline of the Architecture

In our study, we collect survilances camera footage and collected the pre-trained dataset. The we applied YOLO and DeepSORT combination alogorithm with pre-trained weight value of coco dataset. Then accounding to pre-trained model we recived the output with the better accurecy level. The overall architecture of our study is represented in Fig. 4.

B. Dataset Analysis

The COCO (Common Objects in Context) or Darknet or AlexNet dataset is a widely used large-scale image recognition dataset for object detection, segmentation, and captioning tasks. It contains over 330,000 images, each annotated with bounding boxes around the objects in the image, as well as segmentation masks for some objects and captions describing the scene. Pre-trained models trained on the dataset are often used as a starting point for many computer vision tasks. There are several popular deep learning frameworks, such as Yolo, TensorFlow, PyTorch, and MXNet, that provide pretrained models trained on the COCO dataset. These models can be downloaded and used to perform object detection, segmentation, or captioning on new images or videos.

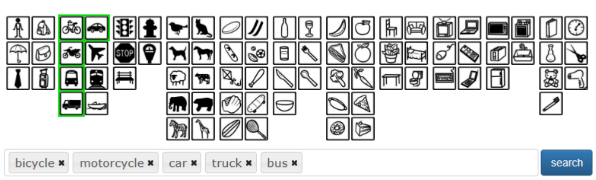

Fig. 1: Coco Pre Trained Dataset Object List

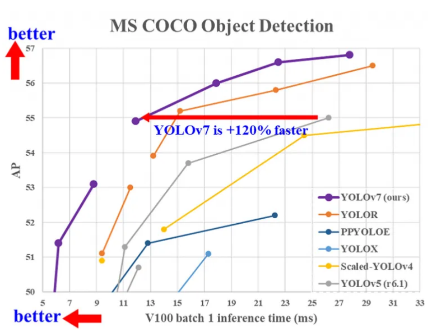

Fig. 2: Object Detection Benchmark on YOLOv7

C. FWHA Vehicles Class

The Federal Highway Administration (FHWA) classifies vehicles into 13 classes, as shown in Figure 2 (FHWA 2016). Therefore, to classify vehicles, the YOLOv4 model needs to be trained with vehicle image data annotated as one of the 13 FHWA classes.

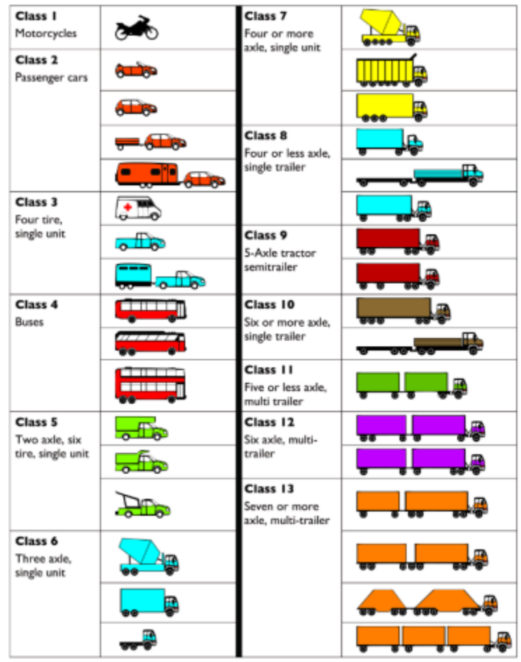

Fig. 3: FWHA 13 Types of Vehicles

D. Model Development

A YOLOv4 detector was trained for vehicle localization and classification. To achieve the best training, this study focused on optimizing the training using the YOLOv8 training loss (YOLO loss). The size of the training set was improved by data augmentation techniques to make the system robust to ambient lighting changes, camera noise , and camera shake. Input resolution of images determines the accuracy as well as training and inference times. Larger pixel resolution may improve the accuracy while increasing the training and inference times. The resolution of the training images was set to fixed pixels to obtain reasonable accuracy. The training was initiated using pre-trained weights (yolov8/chk7) to reduce the time taken for training. This study consists of 6 classes of vehicles and the detector was trained to classify and counting the algorithm. To avoid overfitting, the best weight file was selected from overall best weight file.

A feature vector is a vector that contains information of an object's prominent features i.e., colour, shape, scale, etc. To extract features in detections every YOLOv8 detection is sent through a feature extractor. Here, as an improvement to the existing DeepSORT architecture, COCO is used as the feature extractor replacing the existing network trained on the MARS dataset.

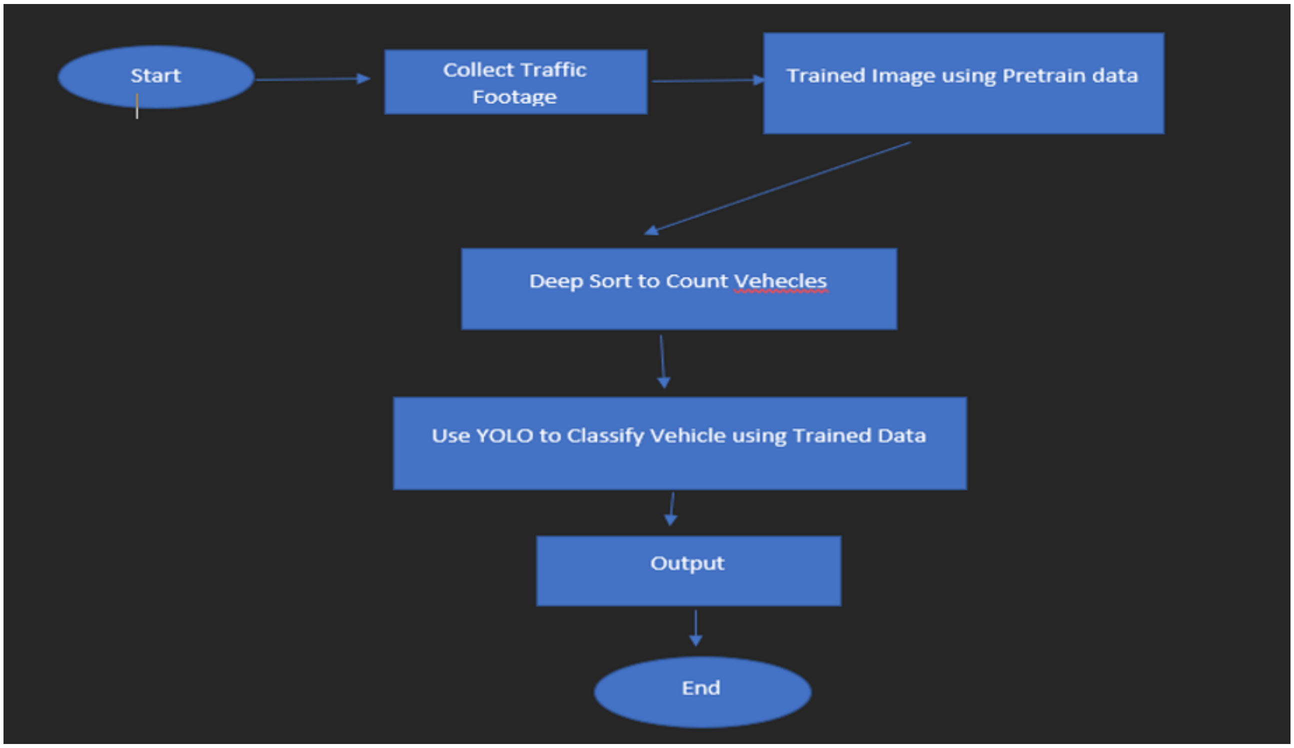

Fig. 4: Workflow of Vehicle classification form pre-trained dataset.

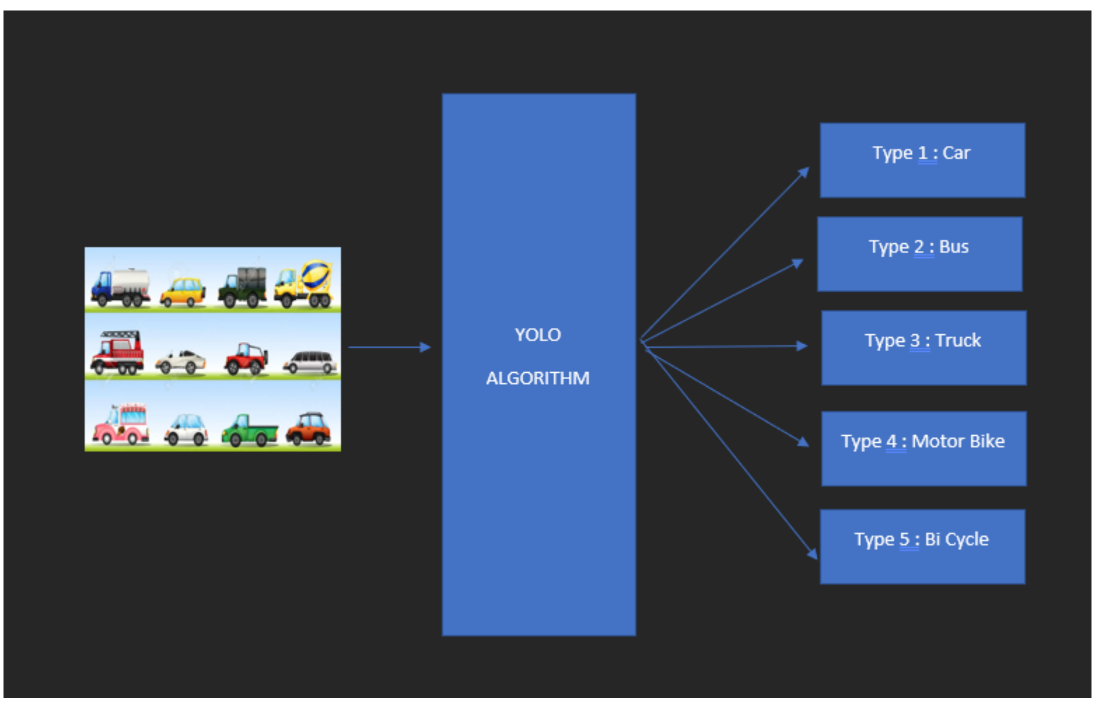

Fig. 5: Simple YOLO model to classify the Vehicles

IV. Result and Evaluation

This part contains information on our study, including the experimental strategy and assessment of the outcom.

A. Performance Evaluation

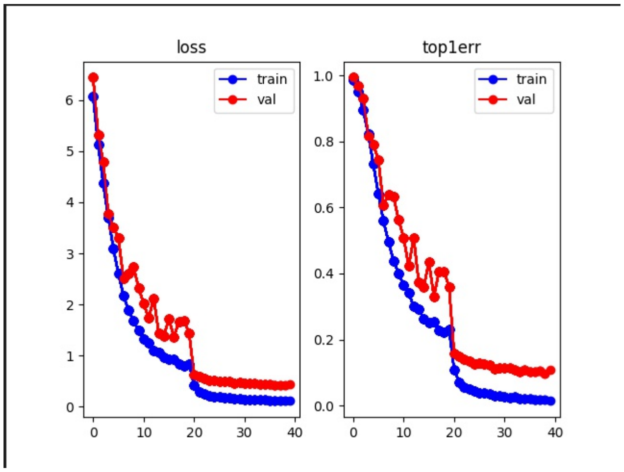

Fig. 6: Accurecy of Pre-Traind YOLO Model

B. Results

The output of YOLOv8 and DeepSORT that can classify and count the vehicles of traffic or parking are. That can take record as like given figure.

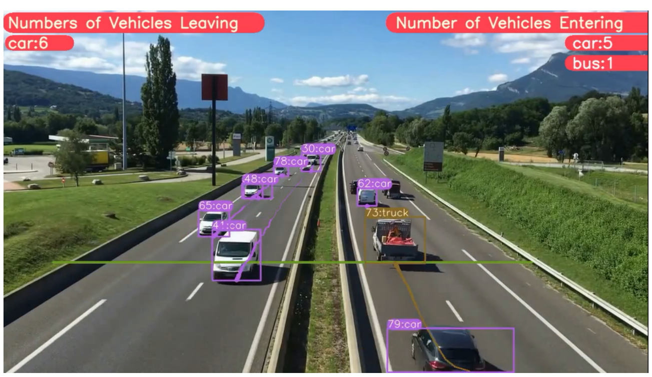

Fig. 7: Accurecy of Pre-Traind YOLO Model

V. Conclusion

Vehicle classification is an important task that has many applications in transportation, public safety, and other fields. In recent years, there have been significant advances in the use of deep learning algorithms, such as YOLO and Deep SORT, for vehicle classification. An open-source object detection algorithm, namely YOLOv4, is used to successfully detect, count, and classify vehicles into different classes. Image annotations were performed using CVAT, and Deep SORT was used for vehicle tracking. In our proposed system we could detect and classify the different types of vehicles. Together, YOLO and Deep SORT can be used to accurately detect and classify vehicles in real-time. However, there are also many challenges associated with vehicle classification, such as variability in appearance, occlusion, and lighting conditions.

Overall, vehicle classification using YOLO and Deep SORT is a promising area of research that has important applications in a variety of fields. Further research and development in this area could lead to significant improvements in transportation safety and efficiency.

References

[1] D. Raut R. Chandwadkar S. P. Dhole, V. R. Gadewar. ”object detection using blob extraction”. *ISSN:2349-5162, Vol.1*, 6, November-2014.

[2] K.H. Jo L. Kurnianggoro, A. Shahbaz. ”dense optical flow in stabilized scenes for moving object detection from a moving camera[c]”. *IEEE*, pages 704–708, 2016.

[3] Shi Z. Guo Y. Zou, Z. and J. Ye. ”object detection in 20 years: A survey”. *IEEE*, 2019.

[4] R. Girshick J. Redmon, S. Divvala and A. Farhadi. ”you only look once: Unified, real-time object detection”. *IEEE*, page 779–788, 2016.

[5] D. Paulus N. Wojke, A. Bewley. ”simple online and realtime tracking with a deep association metric”. *IEEE*, pages 3645–3649, 2017.

[6] Hong-Wei Lioa Cheng-Jian Lin, Shiou-Yun Jeng. ”a real-time vehicle counting, speed estimation, and classification system based on virtual detection zone and yolo”. *Hindawi*, 2021.

[7] Pei Guo Haibo Hu Hong Xiang Qian Zhang Bin Cai Jun Sang, Zhongyuan Wu. ”an improved yolov2 for vehicle detection”. *MDPI*, 4 December 2018.

[8] Medha Pujari Eddie Y. Chou Rabin Subedi, Pratik Shrestha. ”vehicle classification, rumble strips detection, and mapping using artificial intelligence”. *ASCE*, August 2022.

[9] Muhammad Yahya Belal Ahmad Tanvir Ahmad, Yinglong Ma. ”object detection through modified yolo neural network”. *Hindawi*, 06 Jun 2020.

[10] Mohammed Rashid Ahmed Ahmed Mohammed Thakir Mahmood, Saadaldeen Rashid Ahmed. ”detection of vehicle with infrared images in road traffic using yolo computational mechanism”. *Hindawi*, 2020.