DEEP CNN FRAMEWORK FOR OBJECT DETECTION AND CLASSIFICATION SYSTEM FROM REALTIME VIDEOS

Janakiraman[1], Kothibalija Sowmya[2], B Sandeep[2], Vennapusa Vishnuvardhan Reddy[2]

[2]UG Scholar, [1,2]Department of Computer Science and Engineering

[1,2]Kommuri Pratap Reddy Institute of Technology, Ghatkesar, Hyderabad, Telangana.

ABSTRACT

Accurately counting and classifying vehicles in real-time has become essential for effective traffic management, surveillance, and transportation systems. This capability is crucial for optimizing road infrastructure, enhancing safety measures, and making informed decisions for traffic planning. With increasing traffic congestion and road safety concerns, there is a growing demand for robust, automated vehicle counting and classification systems. Traditionally, these tasks relied on manual sensor deployment or fixed cameras at specific locations. However, these methods struggle with complex traffic scenarios, real-time processing, and varying environmental conditions, as well as occlusions and diverse vehicle types. Recent advancements in deep learning have revolutionized object detection, making real-time vehicle counting and classification feasible. One notable model is the YOLO (You Only Look Once) algorithm based on the Darknet framework. Utilizing the power of this model, a real-time vehicle counting and classification system has been developed with the OpenCV library. This system employs a pretrained YOLO model to detect the number of vehicles in a given video and classify each vehicle type. By doing so, it eliminates the need for extensive human intervention and ensures automated and accurate vehicle counting in real-time. Moreover, the system effectively handles varying traffic conditions and different vehicle types, enhancing its accuracy and reliability. The proposed system offers numerous benefits. It provides valuable data for traffic analysis, enabling better traffic management strategies and improved infrastructure planning. With this system, authorities can efficiently address traffic congestion, implement targeted safety measures, and optimize traffic flow. Furthermore, the integration of the YOLO algorithm within the Darknet framework opens new possibilities for real-time traffic management. By leveraging deep learning, this system offers a reliable and efficient solution to the challenges posed by modern traffic scenarios, contributing to safer and more organized road networks for everyone.

Keywords: YOLO, Object Detection, Road traffic, vehicle counting, Transfer Learning, Darknet framework, leveraging deep learning

1. INTRODUCTION

Real-time vehicle counting and classification is a cutting-edge technology and system designed for monitoring and analyzing traffic on roads, highways, and other transportation networks. This technology leverages various sensors, cameras, and computer vision algorithms to capture, analyze, and categorize vehicles as they pass through designated monitoring points. The primary goal of real-time vehicle counting and classification is to provide critical data and insights to transportation authorities, urban planners, and traffic management systems. Here is an overview of this technology in a detailed paragraph Real-time vehicle counting and classification systems utilize a network of strategically positioned cameras and sensors equipped with advanced image processing and machine learning capabilities. These devices continuously capture video or images of the traffic flow in real-time, allowing for the automatic detection and tracking of vehicles. Through computer vision algorithms, the system can identify and classify vehicles based on their type, such as cars, trucks, buses, motorcycles, and more. It can also estimate vehicle speed, direction, and lane occupancy. The applications of real-time vehicle counting and classification are diverse and impactful. Firstly, it provides transportation authorities with critical traffic data, including traffic volume, congestion patterns, and vehicle composition. This data is invaluable for optimizing traffic flow, making informed decisions about road maintenance and expansion, and enhancing road safety measures. Furthermore, these systems are essential for toll collection, parking management, and traffic law enforcement. By automatically categorizing vehicles, authorities can apply appropriate toll rates, manage parking spaces efficiently, and enforce rules and regulations effectively. Additionally, real-time vehicle counting and classification play a crucial role in intelligent transportation systems (ITS). ITS integrates data from these systems to facilitate dynamic traffic management, improve traffic signal timing, and enhance the overall transportation experience for commuters. In smart cities, this technology is an integral part of efforts to reduce traffic congestion, minimize environmental impact, and promote sustainable urban development. So, real-time vehicle counting and classification is a sophisticated technology that enhances traffic management, infrastructure planning, and road safety. It provides real-time insights into traffic conditions, allowing authorities to make data-driven decisions and optimize transportation networks for the benefit of both commuters and the environment. As urbanization continues to grow, the adoption of these systems becomes increasingly important for creating efficient and sustainable transportation systems in modern cities.

2. LITERATURE SURVEY

Alpatov et al. considered road situation analysis tasks for traffic control and ensuring safety. The following image processing algorithms are proposed: vehicle detection and counting algorithm, road marking detection algorithm. The algorithms are designed to process images obtained from a stationary camera. The developed vehicle detection and counting algorithm was implemented and tested also on an embedded platform of smart cameras. Song et al. proposed 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 provided the complete data foundation for vehicle detection based on deep learning. Neupane et al. created a training dataset of nearly 30,000 samples from existing cameras with seven classes of vehicles. To tackle P2, this trained and applied transfer learning-based fine-tuning on several state-of-the-art YOLO (You Only Look Once) networks. For P3, this work proposed a multi-vehicle tracking algorithm that obtains the per-lane count, classification, and speed of vehicles in real time.

Lin et al. presented 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. Moreover, the distance and time traveled by a vehicle are used to estimate the speed of the vehicle. 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. Chauhan et al. used the state-of-the-art Convolutional Neural Network (CNN) based object detection models and train them for multiple vehicle classes using data from Delhi roads. This work gets upto 75% MAP on an 80-20 train-test split using 5562 video frames from four different locations. As robust network connectivity is scarce in developing regions for continuous video transmissions from the road to cloud servers, this work also evaluated the latency, energy and hardware cost of embedded implementations of our CNN model-based inferences. Arinaldi et al. presented a traffic video analysis system based on computer vision techniques. The system is designed to automatically gather important statistics for policy makers and regulators in an automated fashion. These statistics include vehicle counting, vehicle type classification, estimation of vehicle speed from video and lane usage monitoring. The core of such system is the detection and classification of vehicles in traffic videos. This work implemented two models for this purpose, first is a MoG + SVM system and the second is based on Faster RCNN, a recently popular deep learning architecture for detection of objects in images.

Gomaa et al. presented an efficient real-time approach for the detection and counting of moving vehicles based on YOLOv2 and features point motion analysis. The work is based on synchronous vehicle features detection and tracking to achieve accurate counting results. The proposed strategy works in two phases; the first one is vehicle detection and the second is the counting of moving vehicles. For initial object detection, this work has utilized state-of-the-art faster deep learning object detection algorithm YOLOv2 before refining them using K-means clustering and KLT tracker. Then an efficient approach is introduced using temporal information of the detection and tracking feature points between the framesets to assign each vehicle label with their corresponding trajectories and truly counted it. Oltean et al. proposed an approach for real time vehicle counting by using Tiny YOLO for detection and fast motion estimation for tracking. This application is running in Ubuntu with GPU processing, and the next step is to test it on low-budget devices, as Jetson Nano. Experimental results showed that this approach achieved high accuracy at real time speed (33.5 FPS) on real traffic videos. Pico et al. proposed the implementation of a low-cost system to identify and classify vehicles using an Embedded ARM based platform (ODROID XU-4) with Ubuntu operating system. The algorithms used are based on the Open-source library (Intel OpenCV) and implemented in Python programming language. The experimentation carried out proved that the efficiency of the algorithm implemented was 95.35%, but it can be improved by increasing the training sample.

Tituana et al. reviewed different previous works developed in this area and identifies the technological methods and tools used in those works; in addition, this work also presented the trends in this area. The most relevant articles were reviewed, and the results were summarized in tables and figures. Trends in the used methods are discussed in each section of the present work. Khan et al. aimed of this work is that a cost-effective vision-based vehicle counting and classification system that is mainly implemented in OpenCV utilising Python programming and some methods of image processing. Balid et al. reported on the development and implementation of a novel smart wireless sensor for traffic monitoring. Computationally efficient and reliable algorithms for vehicle detection, speed and length estimation, classification, and time-synchronization were fully developed, integrated, and evaluated. Comprehensive system evaluation and extensive data analysis were performed to tune and validate the system for a reliable and robust operation.

Jahan et al. presented convolutional neural network for classifying four types of common vehicle in our country. Vehicle classification plays a vital role of various application such as surveillance security system, traffic control system. This work addressed these issues and fixed an aim to find a solution to reduce road accident due to traffic related cases. To classify the vehicle, this work used two methods feature extraction and classification. These two methods can straight forwardly be performed by convolutional neural network. Butt et al. proposed a convolutional neural network-based vehicle classification system to improve robustness of vehicle classification in real-time applications. This work presented a vehicle dataset comprising of 10,000 images categorized into six-common vehicle classes considering adverse illuminous conditions to achieve robustness in real-time vehicle classification systems. Initially, pretrained AlexNet, GoogleNet, Inception-v3, VGG, and ResNet are fine-tuned on self-constructed vehicle dataset to evaluate their performance in terms of accuracy and convergence. Based on better performance, ResNet architecture is further improved by adding a new classification block in the network. Gonzalez et al. showed a vision-based system to detect, track, count and classify moving vehicles, on any kind of road. The data acquisition system consists of a HD-RGB camera placed on the road, while the information processing is performed by clustering and classification algorithms.

The system obtained an efficiency score over the 95 percent in test cases, as well, the correct classification of 85 percent of the test objects.

3. PROPOSED SYSTEM

The research work begins with the acquisition of real-time video streams as the primary input source, typically sourced from surveillance cameras or traffic monitoring systems. The first module, Background Subtraction and ROI, plays a crucial role in isolating moving objects, i.e., vehicles, from the stationary background.

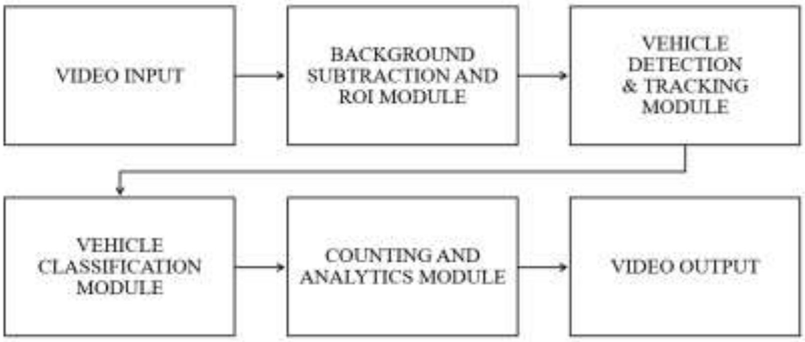

VIDEO INPUT
BACKGROUND SUBTRACTION AND ROI MODULE
VEHICLE DETECTION & TRACKING MODULE
VEHICLE CLASSIFICATION MODULE
COUNTING AND ANALYTICS MODULE
VIDEO OUTPUT

Fig. 1: Block diagram of proposed system.

This is achieved using advanced algorithms to create a Region of Interest (ROI) that narrows down the area for subsequent analysis, reducing computational load and minimizing false positives. The heart of the system lies in the Vehicle Detection and Tracking module, which employs the YOLO (You Only Look Once) deep learning model. YOLO excels at real-time object detection and tracking, enabling the system to identify vehicles within the defined ROI and track their movements across frames, allowing for continuous monitoring. The next step involves the Vehicle Classification module, powered by the Darknet framework, which classifies the detected vehicles into specific categories, such as cars, trucks, or motorcycles. Following this, the Counting and Analytics module quantifies vehicle movements, including counting, speed measurement, and other relevant analytics, providing valuable data for traffic management and research purposes. Finally, the system generates a video output that overlays processed information onto the original video feed, offering a user-friendly visual representation of the vehicle counting and classification results, making it a versatile tool for various applications in traffic monitoring, urban planning, and transportation research.

The detailed operation illustrated as follows:

Step 1: This is the starting point of system. Acquire real-time video streams as input data for the system. These video streams could come from surveillance cameras, traffic cameras, or any source capturing vehicle movements.

Step 2: Background subtraction is a crucial step for isolating moving objects (vehicles) from the stationary background. In this module, use algorithms and techniques to detect the background and create a Region of Interest (ROI) where vehicle detection will occur. This step helps reduce noise and focus on the relevant area.

Step 3: Use YOLO (You Only Look Once), a deep learning-based object detection model, for vehicle detection. YOLO can efficiently detect and locate objects in real-time video frames. It identifies the vehicles within the defined ROI and can track their movements across frames, allowing us to follow vehicles as they move through the video.

Step 4: After detecting and tracking vehicles, further analyze and classify them using the Darknet framework. Darknet is a neural network framework well-suited for classification tasks. It can classify vehicles into different categories such as cars, trucks, motorcycles, or any other relevant classes.

Step 5: In this step, count and analyze the detected and classified vehicles. We can track the number of vehicles passing through specific points or regions of interest, calculate vehicle speed, and gather other relevant analytics data. This information can be useful for traffic management, surveillance, or research purposes.

Step 6: Finally, the system provides video output with the processed information overlaid on the original video feed. This output can include counted vehicles, their classifications, and any other relevant data. It allows users to visualize and interpret the results of the vehicle counting and classification system.

YOLO-V3 Model

Object detection is a phenomenon in computer vision that involves the detection of various objects in digital images or videos. Some of the objects detected include people, cars, chairs, stones, buildings, and animals. This phenomenon seeks to answer two basic questions:

What is the object? This question seeks to identify the object in a specific image.

Where is it? This question seeks to establish the exact location of the object within the image.

Object detection consists of various approaches such as fast R-CNN, Retina-Net, and Single-Shot MultiBox Detector (SSD). Although these approaches have solved the challenges of data limitation and modeling in object detection, they are not able to detect objects in a single algorithm run. YOLO algorithm has gained popularity because of its superior performance over the aforementioned object detection techniques.

YOLO Definition: YOLO is an abbreviation for the term ‘You Only Look Once’. This is an algorithm that detects and recognizes various objects in a picture (in real-time). Object detection in YOLO is done as a regression problem and provides the class probabilities of the detected images.

YOLO algorithm employs convolutional neural networks (CNN) to detect objects in real-time. As the name suggests, the algorithm requires only a single forward propagation through a neural network to detect objects. This means that prediction in the entire image is done in a single algorithm run. CNN is used to predict various class probabilities and bounding boxes simultaneously. The YOLO algorithm consists of various variants. Some of the common ones include tiny YOLO and YOLOv3.

Importance of YOLO: YOLO algorithm is important because of the following reasons:

- Speed: This algorithm improves the speed of detection because it can predict objects in real-time.
- High accuracy: YOLO is a predictive technique that provides accurate results with minimal background errors.
- Learning capabilities: The algorithm has excellent learning capabilities that enable it to learn the representations of objects and apply them in object detection.

YOLO algorithm working: YOLO algorithm works using the following three techniques:

- Residual blocks
- Bounding box regression
- Intersection Over Union (IOU)

Residual blocks: First, the image is divided into various grids. Each grid has a dimension of S x S. The following Figure 2 shows how an input image is divided into grids. In the Figure 2, there are many grid cells of equal dimension. Every grid cell will detect objects that appear within them. For example, if an object center appears within a certain grid cell, then this cell will be responsible for detecting it.

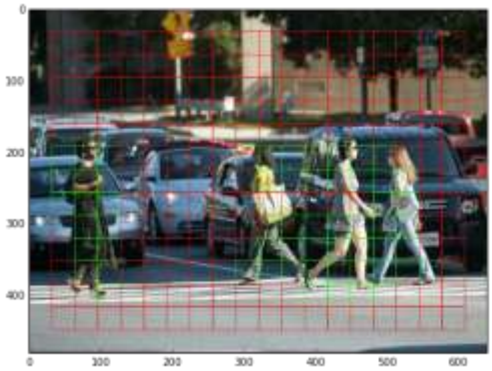


Figure 2. Example of residual blocks.

Bounding box regression: A bounding box is an outline that highlights an object in an image. Every bounding box in the image consists of the following attributes:

- Width (bw)
- Height (bh)
- Class (for example, person, car, traffic light, etc.)- This is represented by the letter c.
- Bounding box center (bx,by)

The following Figure 3 shows an example of a bounding box. The bounding box has been represented by a yellow outline. YOLO uses a single bounding box regression to predict the height, width, center, and class of objects. In the image above, represents the probability of an object appearing in the bounding box.

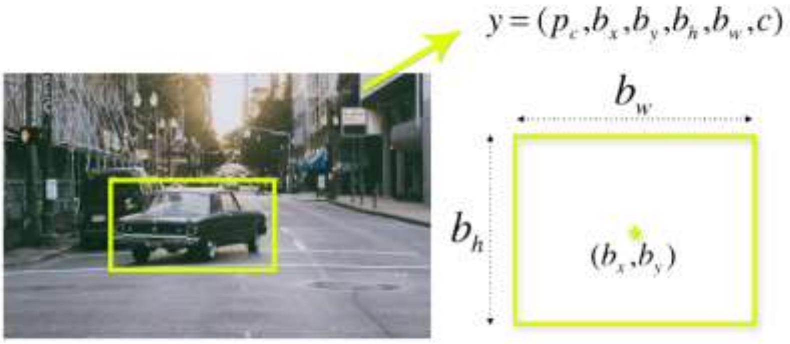


Figure 3. Bounding box regression

Intersection over union (IOU): Intersection over union (IOU) is a phenomenon in object detection that describes how boxes overlap. YOLO uses IOU to provide an output box that surrounds the objects perfectly. Each grid cell is responsible for predicting the bounding boxes and their confidence scores. The IOU is equal to 1 if the predicted bounding box is the same as the real box. This mechanism eliminates bounding boxes that are not equal to the real box.

Combination of the three techniques: The following image shows how the three techniques are applied to produce the final detection results.

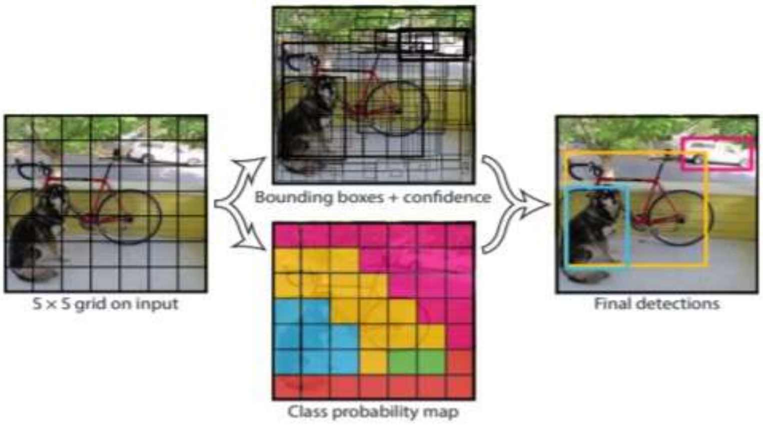
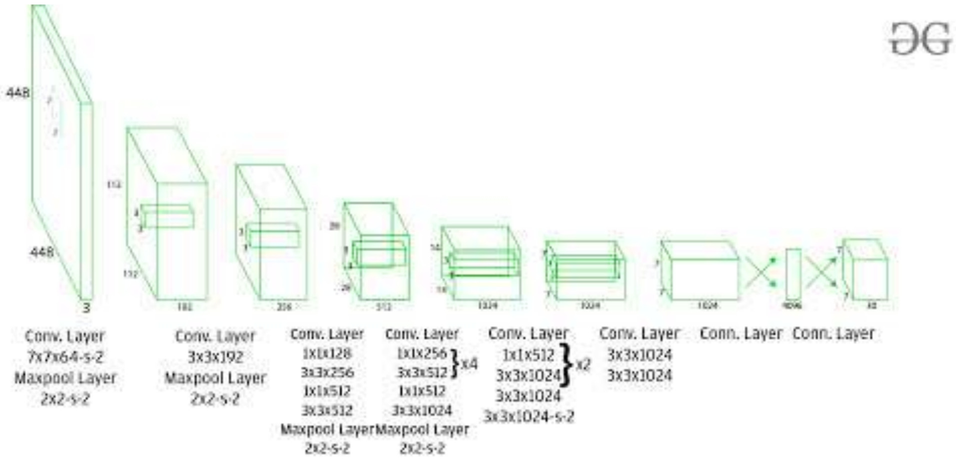


Figure 4. Combination of three modules.

First, the image is divided into grid cells. Each grid cell forecasts B bounding boxes and provides their confidence scores. The cells predict the class probabilities to establish the class of each object. For example, we can notice at least three classes of objects: a car, a dog, and a bicycle. All the predictions are made simultaneously using a single convolutional neural network. Intersection over union ensures that the predicted bounding boxes are equal to the real boxes of the objects. This phenomenon eliminates unnecessary bounding boxes that do not meet the characteristics of the objects (like height and width). The final detection will consist of unique bounding boxes that fit the objects perfectly. For example, the car is surrounded by the pink bounding box while the bicycle is surrounded by the yellow bounding box. The dog has been highlighted using the blue bounding box.

The YOLO algorithm takes an image as input and then uses a simple deep convolutional neural network to detect objects in the image. The architecture of the CNN model that forms the backbone of YOLO is shown below.



YOLO Layers.

The first 20 convolution layers of the model are pre-trained using ImageNet by plugging in a temporary average pooling and fully connected layer. Then, this pre-trained model is converted to perform detection since previous research showcased that adding convolution and connected layers to a pre-trained network improves performance. YOLO's final fully connected layer predicts both class probabilities and bounding box coordinates.

YOLO divides an input image into an S × S grid. If the center of an object falls into a grid cell, that grid cell is responsible for detecting that object. Each grid cell predicts B bounding boxes and confidence scores for those boxes. These confidence scores reflect how confident the model is that the box contains an object and how accurate it thinks the predicted box is.

YOLO predicts multiple bounding boxes per grid cell. At training time, we only want one bounding box predictor to be responsible for each object. YOLO assigns one predictor to be “responsible” for predicting an object based on which prediction has the highest current IOU with the ground truth. This leads to specialization between the bounding box predictors. Each predictor gets better at forecasting certain sizes, aspect ratios, or classes of objects, improving the overall recall score.

One key technique used in the YOLO models is non-maximum suppression (NMS**)**. NMS is a post-processing step that is used to improve the accuracy and efficiency of object detection. In object detection, it is common for multiple bounding boxes to be generated for a single object in an image. These bounding boxes may overlap or be located at different positions, but they all represent the same object. NMS is used to identify and remove redundant or incorrect bounding boxes and to output a single bounding box for each object in the image.

4. RESULTS AND DISCUSSION

Figure 5 represents a single frame from a video. In this frame, there are two vehicles detected: a bus and a car. The object detection model has identified a car in the frame, and it has assigned a high confidence score to this classification, indicating a 97% accuracy in identifying it as a car. Similarly, the model has also identified a bus in the frame, but with a slightly lower confidence score, indicating a 61.74% accuracy in identifying it as a bus. Figure 6 represents another frame from the same video. In this frame, there are two objects detected: a person and a motorbike. The object detection model has identified a person in the frame, and it has assigned a high confidence score to this classification, indicating a 94% accuracy in identifying it as a person. Similarly, the model has also identified a motorbike in the frame, but with a somewhat lower confidence score, indicating a 64% accuracy in identifying it as a motorbike. Figure 7 represents yet another frame from the same video. In this frame, there are multiple objects detected, including persons, cars, and buses. The object detection model has identified persons in the frame with a high confidence score, indicating a 94% accuracy in identifying them as persons. The model has also detected cars in the frame, and it is highly confident in this classification, indicating a 96% accuracy in identifying them as cars. Furthermore, the model has detected buses in the frame with an extremely high confidence score, indicating a 99% accuracy in identifying them as buses.

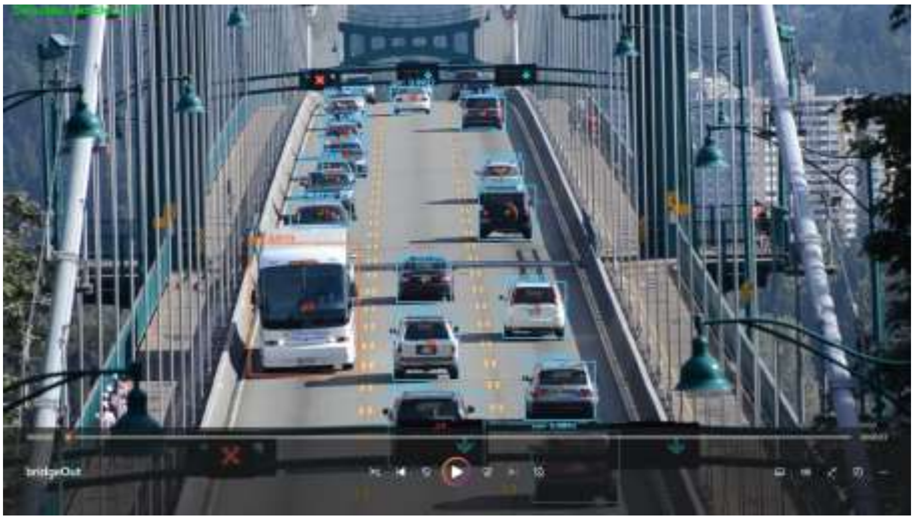


Figure 5: Frame with Bus and car vehicle classification.

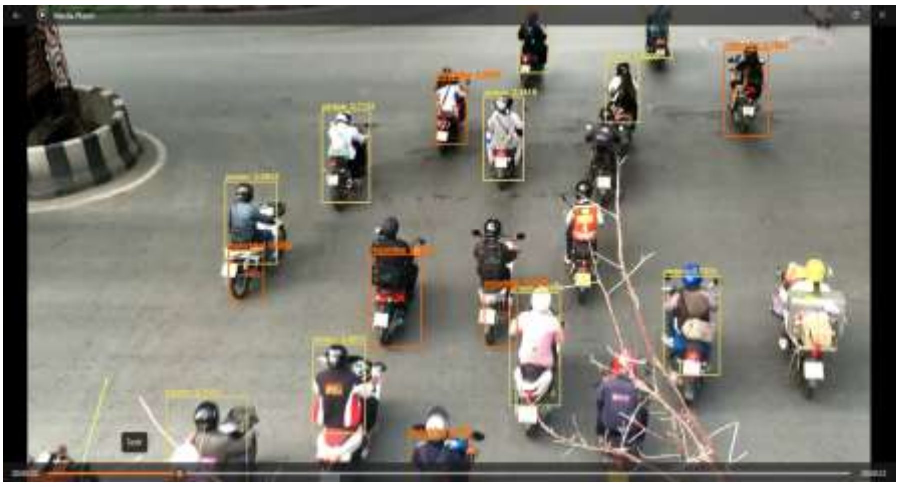


Figure 6: video frame with classification of both person and motor bike.

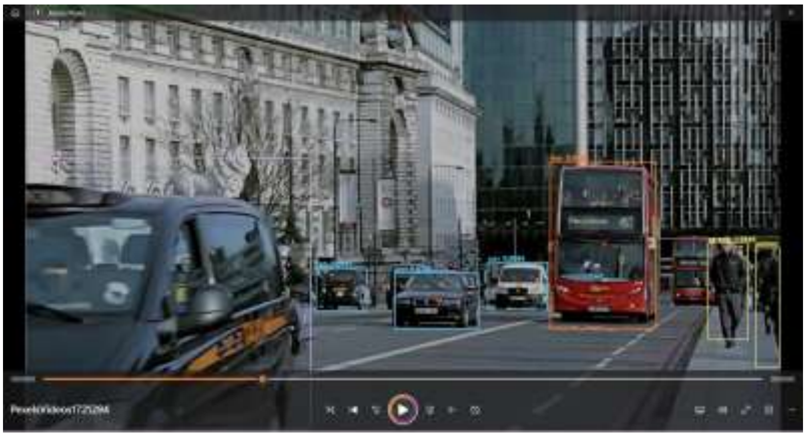


Figure 7: video frame with classification of persons, cars & buses with an accuracy of 94% ,96% & 99%

Figure 8: Shows the input video given to the system.

Figure 9: Shows the result given in the output screen and objects are classifier

5. CONCLUSION

The proposed solution is implemented on python, using the OpenCV bindings. The traffic camera footages from variety of sources are in implementation. A simple interface is developed for the user to select the region of interest to be analyzed and then image processing techniques are applied to calculate vehicle count and classified the vehicles using machine learning algorithms. From experiments it is apparent that CC method outperforms than BoF and SVM method in all results and gives more close classification results to the ground truth values.

REFERENCES

[1] Alpatov, Boris & Babayan, Pavel & Ershov, Maksim. (2018). Vehicle detection and counting system for real-time traffic surveillance. 1-4. 10.1109/MECO.2018.8406017.

[2] Song, H., Liang, H., Li, H. et al. Vision-based vehicle detection and counting system using deep learning in highway scenes. Eur. Transp. Res. Rev. 11, 51 (2019). https://doi.org/10.1186/s12544-019-0390-4.

[3] Neupane, Bipul et al. "Real-Time Vehicle Classification and Tracking Using a Transfer Learning-Improved Deep Learning Network." Sensors (Basel, Switzerland) vol. 22,10 3813. 18 May. 2022, doi:10.3390/s22103813.

[4] C. J Lin, Shiou-Yun Jeng, Hong-Wei Lioa, "A Real-Time Vehicle Counting, Speed Estimation, and Classification System Based on Virtual Detection Zone and YOLO", Mathematical Problems in Engineering, vol. 2021, Article ID 1577614, 10 pages, 2021. https://doi.org/10.1155/2021/1577614.

[5] M. S. Chauhan, A. Singh, M. Khemka, A. Prateek, and Rijurekha Sen. 2019. Embedded CNN based vehicle classification and counting in non-laned road traffic. In Proceedings of the Tenth International Conference on Information and Communication Technologies and Development (ICTD '19). Association for Computing Machinery, New York, NY, USA, Article 5, 1–11. https://doi.org/10.1145/3287098.3287118.

[6] A. Arinaldi, J. A. Pradana, A. A. Gurusinga, "Detection and classification of vehicles for traffic video analytics", Procedia Computer Science, Volume 144, 2018, Pages 259-268, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2018.10.527.

[7] Gomaa, A., Minematsu, T., Abdelwahab, M.M. et al. Faster CNN-based vehicle detection and counting strategy for fixed camera scenes. Multimed Tools Appl 81, 25443–25471 (2022). https://doi.org/10.1007/s11042-022-12370-9.

[8] G. Oltean, C. Florea, R. Orghidan and V. Oltean, "Towards Real Time Vehicle Counting using YOLO-Tiny and Fast Motion Estimation," 2019 IEEE 25th International Symposium for Design and Technology in Electronic Packaging (SIITME), 2019, pp. 240-243, doi: 10.1109/SIITME47687.2019.8990708.

[9] L. C. Pico and D. S. Benítez, "A Low-Cost Real-Time Embedded Vehicle Counting and Classification System for Traffic Management Applications," 2018 IEEE Colombian Conference on Communications and Computing (COLCOM), 2018, pp. 1-6, doi: 10.1109/ColComCon.2018.8466734.

[10] D. E. V. Tituana, S. G. Yoo and R. O. Andrade, "Vehicle Counting using Computer Vision: A Survey," 2022 IEEE 7th International conference for Convergence in Technology (I2CT), 2022, pp. 1-8, doi: 10.1109/I2CT54291.2022.9824432.

[11] A. Khan, A., Sabeenian, R.S., Janani, A.S., Akash, P. (2022). Vehicle Classification and Counting from Surveillance Camera Using Computer Vision. In: Suma, V., Baig, Z., K. Shanmugam, S., Lorenz, P. (eds) Inventive Systems and Control. Lecture Notes in Networks and Systems, vol 436. Springer, Singapore. https://doi.org/10.1007/978-981-19-1012-8_31.

[12] W. Balid, H. Tafish and H. H. Refai, "Intelligent Vehicle Counting and Classification Sensor for Real-Time Traffic Surveillance," in IEEE Transactions on Intelligent Transportation Systems, vol. 19, no. 6, pp. 1784-1794, June 2018, doi: 10.1109/TITS.2017.2741507.

[13] N. Jahan, S. Islam and M. F. A. Foysal, "Real-Time Vehicle Classification Using CNN," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020, pp. 1-6, doi: 10.1109/ICCCNT49239.2020.9225623.

[14] M. A. Butt, A. M. Khattak, S. Shafique, B. Hayat, S. Abid, Ki-Il Kim, M. W. Ayub, A. Sajid, A. Adnan, "Convolutional Neural Network Based Vehicle Classification in Adverse Illuminous Conditions for Intelligent Transportation Systems", Complexity, vol. 2021, Article ID 6644861, 11 pages, 2021. https://doi.org/10.1155/2021/6644861.

[15] P. Gonzalez, Raul & Nuño-Maganda, Marco Aurelio. (2014). Computer vision based real-time vehicle tracking and classification system. Midwest Symposium on Circuits and Systems. 679-682. 10.1109/MWSCAS.2014.6908506.