The Analysis of Deep Learning Based Vehicle Classification, Tracking and Speed Estimation Systems

Authors: Chaitanya Anand, Arkadeepto Majumder

Mentor: Prof. Ardhendu Kundu

Abstract

With the increase in traffic volume due to rapid advancements of technology and increasing urban population, traffic congestion and traffic accidents are becoming social problems. Computer Vision is a useful tool to get information regarding such traffic incidents, however, vehicle detection and tracking based on camera is highly dependent on the environmental factors and technological factors and thus they can be inefficient. Therefore, these varying environmental factors and technological limitations have given way to deep learning based algorithms used for object detections that are employed in vehicle classification and tracking, along with speed detection in order to classify and track vehicles in complex and diverse environments. The rapid increase of computational capability of modern devices have lead to a widespread use of these deep learning techniques such as CNNs, RNNs, Siamese networks, and object detection algorithms like YOLO and Faster R-CNN, etc to enhance traffic safety and efficiency.

Introduction

As the urban population increases exponentially coupled with rapid technological advancments, the number of vehicles has increased significantly and the capacity of exixting transporation systems are stretched thin causing severe traffic congestions in many highly populated cities with inefficient transport systems [1]. This situation is not one that can be solved by simply building more highways and subways, as not only the resources for such infrastructural development are scarce, but at the same time, these steps have been proven to be moot in solving this problem of traffic management [2].

An effective alternative to deal with traffic congestion is a traffic management and monitoring system that is a fully automated system to collect traffic data such as the number of vehicles, type of vehicles and vehicle speed. The collection of these data is important for such traffic management system to perform its functions which is to perform analysis on the traffic to better utilize the roadway systems and facilitate the safety of transportation [3]. The goal of such automated traffic surveillance system is to remove the need for human labour for simple vision tasks that can be performed efficiently by computer systems.

Currently, various data collection and processing technologies in the field of Internet of Things, a lot of traffic monitoring cameras have been placed at congestion prone areas of intersections, highways, etc [4]. However, most of the monitoring systems still work in the tradition way of collecting raw video data and transmitting and storing it. This video data is then analysed manually that presents a significant problem due to the sheer volume of the data that need to be processed and analysed [5]. To tackle this inefficiency, there are various machine learning algorithms that have been developed to extract such data from the video data without manual intervention.

With great progress of in the field of machine learning and artificial intelligence(AI), various methods were proposed to tackle the problem of vehicle tracking and classification. The traditional vehicle detection and classification models were mainly based on the following methods: i) Scale Invariant Feature Transform (SIFT) for feature matching and extraction; ii) Gaussian Mixture model for moving vehicle detection; iii) License Plate extraction method; iv) Histogram of Oriented Gradient(HOG) and Support Vector Machine(SVM) [6]. Modern methods of tracking and classification are mostly based of deep learning techniques and models such as Convolutional Neural Networks (CNNs), Recurrent Convolutional Neural Networks (RCNNs), You Only Look Once (YOLO), etc. This survey aims at analysing the existing methods and models used to achieve the goal of tracking and classification of vehicles.

Survey

Traditional Methods (Computer Vision, GMMs, SVMs)

Tarun Kumar et al. [14] propose an elementary but efficient approach for detection and speed estimation for moving vehicles using a single point CCTV camera using image processing. The proposed approach is implemented by using OpenCV and JavaCV with MySQL being used as database. Although the calibration of camera is strict to suit this model of detection, this flaw can be easily handled through modern tools and models. Regardless, the innovation of this approach lies in the selection of Region of Interest for vehicle detection. The accuracy of detection through this proposed method was found to be 87.7% with the maximum accuracy found to be 98.3%.

Chen et al. focus on effectively capturing a car image from video footage [15]. The authors adopt the background Gaussian Mixture Model (GMM) and the shadow removal algorithm [16] to reduce the negative impacts on vehicle classification caused by shadow, camera vibration, illumination changes, etc. The Kalman filter [17] is used for vehicle tracking and SVM is used to perform vehicle classification. Experiments were performed with real video footage obtained from cameras deployed in Kingston upon Thames, UK. Vehicles were classified into five categories, i.e., motorcycles, cars, vans, buses, and unknown vehicles. The classification accuracy for these vehicle types was 94.6%.

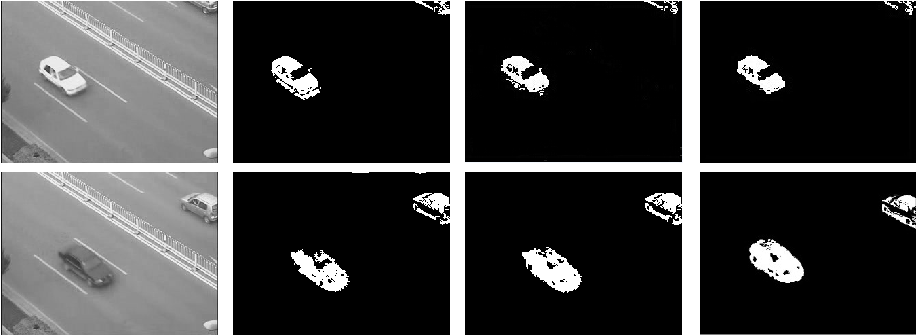


Figure : Moving Object Detection using Gaussian Mixture Algorithm Model

Deep Learning Algorithms (CNNs, RCNNs)

Minglan Sheng et al. [7] suggests one approach is to use convolutional neural networks (CNNs) for feature extraction and classification. CNNs have been shown to be effective in identifying vehicle types from different angles and scenes. However, there are still challenges in accurately identifying vehicles in real traffic situations, such as low image definition, deformation due to large aspect ratios, and imbalanced training datasets. To address these challenges, researchers have explored different techniques, including the use of depth learning for vehicle positioning and classification. Other methods have focused on specific features, such as the outline, windshield, rear-view mirror, and license plate, to classify vehicle types. These traditional methods, however, have limitations in terms of recognition accuracy. Therefore, there is a need for further research to improve the accuracy and robustness of vehicle detection and classification methods.

However, region based CNNs were computationally expensive as originally developed, their cost has been drastically reduced with the introduction of new methods such as Fast R-CNN [11]. Shaoqing Ren et al. [12] proposes a method to unify the fast R-CNNs and RCNs into a fully–convolutional network (FCN) through alternative training scheme between fine-tuning for the region proposal task and then fine-tuning for the object detection, while keeping the proposals fixed. This scheme is used to build the unified convolutional network which was tested against the PASCAL VOC detection benchmarks [13] and it resulted in improvement in region proposal quality and thus object detection accuracy.

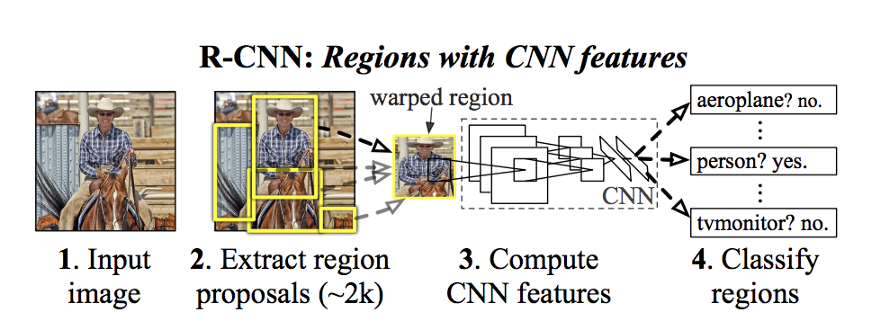


Figure : Basic Steps involved in Object Detection using RCNN Model

A major challenge for applying machine learning algorithms for automated background filter in image processing and feature extraction is that different parts of the vehivle are treated without distinction, which degrades their performance. Zhao et al. focus on this problem that potentially misses the key part of a car image [18]. Their work is motivated by the human vision system that distinguishes the key parts of an image from the background, which is called the multiglimpse and visual attention mechanism. The key idea of their work is thus to exploit the visual attention mechanism to generate a focused image first and provide the image as input to CNN for more accurate vehicle classification. They performed experiments to classify a vehicle into five types, sedans, vans, trucks, SUVs, and coaches, and achieved the classification accuracy of 97.9%.

This [9] paper proposes a framework for predicting driver behavior in dilemma zones at signalized intersections using machine learning techniques. The study collected vehicle attribute data at the onset of the yellow indication and developed prediction models using Support Vector Machine (SVM) and Artificial Neural Network (ANN). The models achieved high prediction accuracies and were evaluated based on various performance measures. The study found that the start and end points of the dilemma zone were influenced by approaching speeds and varied by time of day. The proposed framework shows potential for improving intersection signal operations and reducing red-light violations. The findings of this study contribute to the existing research on driver behavior in the dilemma zone and highlight the need for further research to enhance traffic safety and signal timing strategies.

Implementation

The project was implemented using a Python script that utilizes the OpenCV library and the YOLO (You Only Look Once) object detection framework to analyze a video stream, identify vehicles, and calculate their speeds based on their movement between two predefined lines in the video frame. The YOLO model is loaded with pre-trained weights and configuration files, and the COCO dataset is used for class labels. The video stream is obtained from a file ('veh2.avi'), and the output is saved to an MP4 file ('result.mp4').

The script processes each frame of the video, detects vehicles using YOLO, and draws bounding boxes around them. Two horizontal lines are drawn on the video frame to represent positions where vehicle speeds will be calculated. The script tracks the time when each vehicle crosses the first line, and when it crosses the second line, it calculates the vehicle's speed based on the known real-life distance between the two lines. The calculated speed is then displayed on the video frame along with the bounding box and label for each detected vehicle.

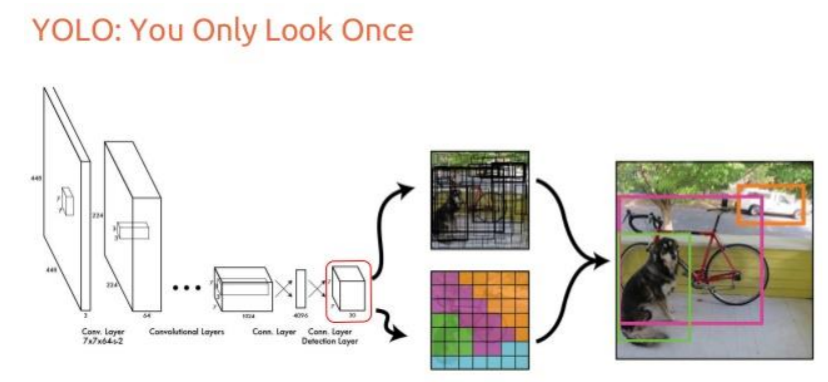


Figure : Basic Overview of Object Detection Process using YOLO Algorithm

The script employs multi-object detection, non-maximum suppression to filter out overlapping bounding boxes, and frame-by-frame processing. The calculated speeds and visual annotations are included in the output video. The processing continues until the user presses 'q' to exit the video playback. This script is useful for traffic monitoring applications, providing a practical example of computer vision and object detection in a real-world context.

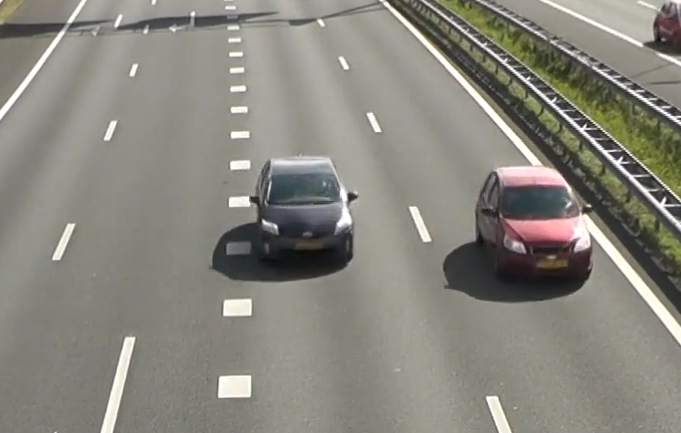
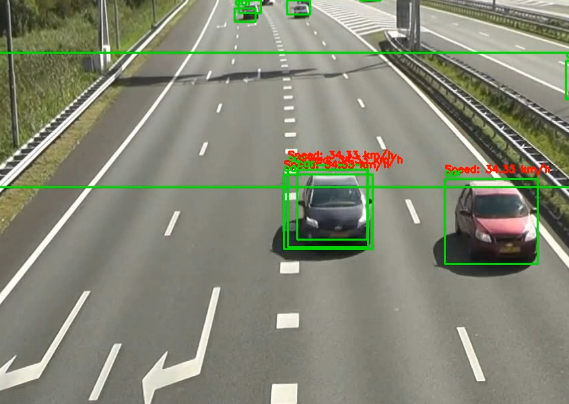


Figure .2: After Detection, Classification and Estimation of speed

Figure 4.1: The raw footage fed to Model

Modern Tools and Models (Tensorflow, DeepSORT, YOLO)

Moving away from variations of the convolutional neural networks (CNNs), Muhammad Azhad bin Zuraimi and Fadhlan Hafizhelmi Kamaru Zaman [19] propose a vehicle detection and tracking system using Tensorflow library with DeepSORT algorithm based on YOLOv4 model to perform Real time traffic surveillance with accuracy of 82.08% with the mean average precision of 0.5 with performance speed as 40 fps using GTX 1660ti. With modern CUDA enabled graphical processors or GPUs available through cloud platforms such as Google Colab, this metric could be further improved upon using the proposed methods.

Cheng-Jian Lin, Shiou-Yun Jeng, and Hong-Wei Lioa [10] have presented a real-time traffic monitoring system that addresses the challenges of accurate vehicle detection and classification in traffic flows, particularly dealing with total occlusions. They employ convolutional neural networks, specifically utilizing the You Only Look Once (YOLO) algorithm. Their system, which integrates a virtual detection zone, Gaussian mixture model (GMM), and YOLO, enhances vehicle counting and classification efficiency. Additionally, the method estimates vehicle speed based on distance and time traveled. Experimental results using the MAVD and GRAM-RTM datasets demonstrate the effectiveness of the proposed approach. The highest classification accuracy achieved with YOLOv4 was 98.91% in MAVD and 99.5% in GRAM-RTM datasets. The method also showed strong performance in different environmental conditions (daytime, night time, and rainy day). Moreover, the average absolute percentage error for vehicle speed estimation was approximately 7.6%.

Future Prospects and Challenges

We have witnessed significant development of vehicle classification systems in the past decade. Thanks to recent advances in sensing, machine learning, and wireless communication technologies, the classification accuracy has improved greatly at a significantly reduced cost. However, these emerging vehicle classification systems have left a number of open questions. In this section, we discuss these challenges and several future research directions.

More and more vehicle classification systems depend on machine learning techniques. To achieve high classification accuracy, however, a huge amount of data should be collected to train and create an effective classification model. Especially, the manual labelling process for training the classification model requires a significant amount of time and efforts. It also requires extra efforts for obtaining the ground truth data.

Although we have seen that many classification systems achieve very high classification accuracy, achieving near 100% classification accuracy especially for a large number of vehicle types is still a very challenging task. One possible reason for the difficulty lies in the fact that most solutions rely on a single type of sensor for vehicle classification. If different type of sensors could be used together to collect this data, the task would become much easier because collecting video data is dependent upon physical and environmental factors and thus could be complemented through the use of devices such as infrared sensors [17] which could be used for better detection during night or dark periods of the day.

The advent of self-driving cars [20] have led to opening of a whole new avenue with regards to the management of traffic systems. Autonomous Traffic Management (ATM) [21] is one of the most vigorously pursued areas of research in the ITS now a days which promise to reduce traffic problems through a well-connected and coordinated infrastructure. For this, AVs and their connected environment need to manage through dynamically changing traffic situations intelligently, and road conditions [22]

Conclusion

We presented a review of the modern traffic surveillance systems, focusing on the key functionality of vehicle detection, tracking and speed calculation systems. By categorizing the vehicle classification systems into 3 categories based on the tools used by authors in implementing the system, mainly the traditional methods (SVMs, GMMs, SIFT, Computer Vision), the Deep Learning Techniques (CNNs, RCNNs, RPMs, ANNs) and the modern methods that provide an inbuilt set of tools to aid in classification and detection such as TensoFlow, Colab, YOLO, etc. We also discussed some of the limitations of these proposed solutions as well as some future research directions.

References

1. M. Won, T. Park and S. H. Son, "Toward mitigating phantom jam using vehicle-to-vehicle communication", IEEE Trans. Intell. Transp. Syst., vol. 18, pp. 1313-1324, May 2017.

2. Muhammad Tahir Masood PhD, P. E. "Transportation problems in developing countries Pakistan: a case-in-point." International Journal of Business and management 6.11 (2011): 256.

3. Ahmad Arinaldi, Jaka Arya Pradana, Arlan Arventa Gurusinga, “Detection and classification of vehicles for traffic video analytics”, Procedia Computer Science, Volume 144, 2018, Pages 259-268, ISSN 1877-0509

4. Lu, H., Sun, Z., Qu, W.: Big data and its applications in urban intelligent transportation system. J. Transp. Syst. Eng. Inf. Technol. 15(5), 45–52 (2015)

5. Ravish, Roopa, and Shanta Ranga Swamy. "Intelligent traffic management: A review of challenges, solutions, and future perspectives." Transport and Telecommunication Journal 22.2 (2021): 163-182.

6. Sheng, M., Liu, C., Zhang, Q., Lou, L., & Zheng, Y. (2018). Vehicle Detection and Classification Using Convolutional Neural Networks. 2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS). doi:10.1109/ddcls.2018.8516099

7. M. Sheng et al. “Vehicle Detection and Classification Using Convolutional Neural Networks”. In: (2018). doi: 10.1109/ddcls.2018.8516099.

9. M. Rahman, M.-W. Kang, and P. Biswas. “Predicting time-varying, speed-varying dilemma zones using machine learning and continuous vehicle tracking”. In: Transportation Research Part C: Emerging Technologies (2021). doi: 10.1016/j.trc.2021.

10. Teen-Hang Meen et al. “A Real-Time Vehicle Counting, Speed Estimation, and Classification System Based on Virtual Detection Zone and YOLO”. In: Mathematical Problems in Engineering 2021 (2021), p. 1577614. issn: 1024-123X. doi: 10.1155/2021/1577614.

11. R. Girshick. Fast R-CNN. arXiv:1504.08083, 2015

12. Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems 28 (2015).

13. M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results, 2007

14. Kumar, T., & Kushwaha, D. S. (2016). An Efficient Approach for Detection and Speed Estimation of Moving Vehicles. Procedia Computer Science, 89, 726–731. doi:10.1016/j.procs.2016.06.045

15. Z. Chen, T. Ellis and S. A. Velastin, "Vehicle detection tracking and classification in urban traffic", Proc. 15th Int. IEEE Conf. Intell. Transp. Syst., pp. 951-956, Sep. 2012.

16. Z. Chen, N. Pears, M. Freeman and J. Austin, "Background subtraction in video using recursive mixture models spatio-temporal filtering and shadow removal", Proc. Int. Symp. Vis. Comput., pp. 1141-1150, 2009.

17. M. Won, "Intelligent Traffic Monitoring Systems for Vehicle Classification: A Survey," in IEEE Access, vol. 8, pp. 73340-73358, 2020, doi: 10.1109/ACCESS.2020.2987634.

18. D. Zhao, Y. Chen and L. Lv, "Deep reinforcement learning with visual attention for vehicle classification", IEEE Trans. Cognit. Develop. Syst., vol. 9, no. 4, pp. 356-367, Dec. 2017.

19. Bin Zuraimi, M. A., & Kamaru Zaman, F. H. (2021). Vehicle Detection and Tracking using YOLO and DeepSORT. 2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE). doi:10.1109/iscaie51753.2021 .9431784

20. A. Mushtaq, I. U. Haq, M. U. Imtiaz, A. Khan and O. Shafiq, "Traffic Flow Management of Autonomous Vehicles Using Deep Reinforcement Learning and Smart Rerouting," in IEEE Access, vol. 9, pp. 51005-51019, 2021, doi: 10.1109/ACCESS.2021.3063463.

21. S. El Hamdani and N. Benamar, "Autonomous traffic management: Open issues and new directions", Proc. Int. Conf. Sel. Topics Mobile Wireless Netw. (MoWNeT), pp. 1-5, Jun. 2018.

22. M. V. Rajasekhar and A. K. Jaswal, "Autonomous vehicles: The future of automobiles", Proc. IEEE Int. Transp. Electrific. Conf. (ITEC), pp. 1-6, Aug. 2015.

Appendix

import cv2

import numpy as np

import time

net = cv2.dnn.readNet('yolov3.weights', 'yolov3.cfg')

with open('coco.names', 'r') as f:

    classes = f.read().strip().split('\n')

cap = cv2.VideoCapture('veh2.avi')

fps = cap.get(cv2.CAP\_PROP\_FPS)

# Defining video writer for MP4 format

fourcc = cv2.VideoWriter\_fourcc(\*'mp4v')  # Codec for MP4

output = cv2.VideoWriter('result.mp4', fourcc, fps, (int(cap.get(3)), int(cap.get(4))))

line1\_y = 100

line2\_y = 300

line\_distance = 10

start\_time = None

while True:

    ret, frame = cap.read()

    if not ret:

        break

    blob = cv2.dnn.blobFromImage(frame, 0.00392, (416, 416), (0, 0, 0), True, crop=False)

    net.setInput(blob)

    outs = net.forward(net.getUnconnectedOutLayersNames())

    class\_ids = []

    confidences = []

    boxes = []

    for out in outs:

        for detection in out:

            scores = detection[5:]

            class\_id = np.argmax(scores)

            confidence = scores[class\_id]

            if confidence > 0.5:

                center\_x = int(detection[0] \* frame.shape[1])

                center\_y = int(detection[1] \* frame.shape[0])

                w = int(detection[2] \* frame.shape[1])

                h = int(detection[3] \* frame.shape[0])

                x = int(center\_x - w / 2)

                y = int(center\_y - h / 2)

                class\_ids.append(class\_id)

                confidences.append(float(confidence))

                boxes.append([x, y, w, h])

    indices = cv2.dnn.NMSBoxes(boxes, confidences, 0.5, 0.4)

    cv2.line(frame, (0, line1\_y), (frame.shape[1], line1\_y), (0, 255, 0), 2)

    cv2.line(frame, (0, line2\_y), (frame.shape[1], line2\_y), (0, 255, 0), 2)

    for i, box in enumerate(boxes):

        x, y, w, h = box

        vehicle\_midpoint = (x + w // 2, y + h // 2)

        if y < line1\_y and vehicle\_midpoint[1] > line1\_y:

            start\_time = time.time()

        if y < line2\_y and vehicle\_midpoint[1] > line2\_y and start\_time is not None:

            crossing\_time = time.time()

            vehicle\_speed = (line\_distance / (crossing\_time - start\_time))\*3.6

            cv2.putText(frame, f'Speed: {vehicle\_speed:.2f} km/h', (x, y - 10), cv2.FONT\_HERSHEY\_SIMPLEX, 0.5, (0, 0, 255), 2)

        cv2.rectangle(frame, (x, y), (x + w, y + h), (0, 255, 0), 2)

        label = str(classes[class\_ids[i]])

        cv2.putText(frame, label, (x, y - 5), cv2.FONT\_HERSHEY\_SIMPLEX, 0.5, (0, 255, 0), 2)

    cv2.imshow('Result', frame)

    output.write(frame)

    if cv2.waitKey(1) & 0xFF == ord('q'):

        break

cap.release()

output.release()

cv2.destroyAllWindows()