ATMOS (Autonomous Traffic Monitoring and Observing System): A Deep-learning Based Real-Time Traffic Monitoring System using CCTV footage

Myles Antonette Solatorio, and Pauline Joy Acebedo

***Abstract*—The paper proposes a vehicle detection, classification and counting system that is capable of running in real-time. By using this computer vision-based system, complex traffic scenes can be analyzed and traffic interactions can be quantified. The system uses a deep-learning based detection algorithm called You Only Look Once (YOLO). Specifically, the project will use the YOLOv3, which has shown promising results in similar studies especially in complex environments. The algorithm is also much faster than other popular algorithms which makes it ideal for real-time deployment. The system will use footage collected from CCTV cameras installed along roads to monitor traffic. The vehicles will be counted according to their class. The data will then be shown real-time in the video-feed and in a dashboard.**

***Index Terms*—deep learning, traffic monitoring, CCTV, convolutional neural network, traffic congestion, traffic flow.**

I. Introduction

IN the last decade or so, authorities in the Philippines have increasingly depended om CCTV cameras to monitor real-time traffic especially in metropolitan areas. However, the utilization of these systems leaves so much to be desired. These camera systems are manually monitored by operators and usually only used for monitoring accidents and congestion. Quantifying the information provided by these systems would be helpful in optimizing traffic mobility and improving traffic flow. The apparent lack of comprehensive traffic flow and network analysis in the Philippines, which is crucial for urban planning and road infrastructure management is rather concerning. The most recent report of Numbeo published last January 2021 shows that the country scored 192.88 for the traffic index and 243.20 for the inefficiency index. The use of computer vision in traffic monitoring has been steadily gaining traction over the years. These advancements have shown remarkable results and achievements. Recent researches have shown that CNN-based detection and classification networks such as YOLO perform well even in complex urban settings e.g. low lighting conditions, bad weather. Thus, in an effort to improve our understanding of traffic interactions in Philippine roads, the researchers have proposed a vehicle detection and counting system. The proposed project uses a deep-learning based system to detect, classify and count vehicles in real-time. The project aims to provide a systematic solution to counting vehicles and monitoring congestion using computer vision-based techniques.

mds
August 26, 2015

II. Significance/Rationale

A. Economy

Road congestion costs our economy a significant portion of the gross domestic product. High levels of traffic congestion and density cause huge delays in transport of goods and delivery of services, and increased fuel wastage.

B. Environment

The long time spent by vehicles on the road also contributes to the increase in emission of pollutants. Thus, minimizing the time spent by vehicles on the road can help lessen their carbon footprint.

C. Efficiency

Plenty of Filipinos spend a few hours each day stuck in traffic. Philippines is always among the top countries with longest traffic wait times. Commuters lose valuable time that can instead be spent on productive activities. Traffic costs people income opportunities. Filipinos can do away with delayed transactions and the mental and physical fatigue caused by traffic.

D. Safety

Efficient traffic networks also translate to safer commute and roads. With better road networks, we can significantly reduce the risk of accidents and hazards in roads.

E. Society

If the project is deployed successfully, especially in highly urbanized areas, it can help local governments to employ efficient routes, implement better policies, and plan effective road infrastructure projects. This could also provide road users with vital information that would help them understand traffic flow and interactions better, and consequently, navigate the roads better and more safely.

III. General Objective

The main objective of this project is to create a deep-learning based monitoring system that will analyze CCTV footage of traffic flow to detect real-time congestion based on the weighted count of vehicles in roads.

A. Specific Objectives

- Deploy the system in roads with varying congestion levels.
- Test the system in several types of environments, lighting conditions, and weather conditions.
- Publish a dashboard/web app that will show the real time data (timelines, etc.)

IV. Related Literature

The interest towards using computer vision techniques for intelligent traffic systems has been growing. Thus, several vehicle detection/counting techniques have been tested by researchers in similar studies. Some of the popular algorithms used include Gaussian Mixture Model (GMM), Region Convolutional Network (R-CNN), another variation of R-CNN called Faster R-CNN, and You Only Look Once (YOLO). Some studies also used a combination of CNN-based techniques with Support Vector Machine (SVM). Classical approaches such as the background subtraction-based Gaussian Mixture Model are widely used in these types of traffic monitoring systems. The advantages of this algorithm include fast and accurate detection. Background subtraction-based detection methods extract objects by evaluating the difference between the presumed static background and the images to be processed. In a paper by Maqbool et al. (2018), a GMM was used to compute the variance, covariance and mean of every pixel in a frame. As a new frame arrives, these parameters are calculated again. The foreground is determined when the difference between the values for the two frames is larger than the product of actual value and standard deviation. In this method, the model is created and updated based on a singular modal distribution, such as singular Gaussian distribution. While background subtraction is efficient for vehicle detection, Xia et al. (2016) discussed that the performance of the model can suffer in complex environment. The variation in lighting, presence of leaves or wind, and even a slight movement of the camera can affect the effectiveness. A combination of shallow learning and deep learning techniques can also be used. A system developed by Adu-Gyamfi et al. (2017) separated object recognition to two tasks: localization and classification. Selective Search is used to localize region proposals. A Deep Convolutional Neural Network (DCNN) is then used to warp the regions to fixed square sizes and extracting the unique feature descriptors. Then, the classification is executed using a linear SVM scoring system. Another popular model used is the R-CNN. Over the years, several variations and enhancements have been made to the algorithm. This is due to the heavy computing power needed by the model and its relatively slow performance. These caveats are improved by variations of the model called Fast and Faster R-CNN. Unlike regular R-CNNs, these models can achieve real-time detection. A study conducted by Zhang et al (2019), employed the Faster R-CNN algorithm. However, the frame rate achieved was only 12 fps and is therefore not appropriate for real-time deployment. Several techniques can be done to improve the speed of the model. However, the detection accuracy suffered at the expense of these attempts. The study also tested the performance of the Single Shot MultiBox Detector (SSD) algorithm. However, the model neglects the relationship between different layers of the feature pyramid. This leads to the model having a relatively poor performance in the monitoring of small vehicles and is therefore inappropriate for a setting like the Philippines. In another study (Al-Ariny, 2020), the algorithm used was an enhanced variation of the Faster R-CNN called the Mask R-CNN. This is a framework that was first discussed in 2017. The network adds a parallel channel to give a segmentation mask for each detected object in addition to the bounding-box and the class label. Deep-learning based detection algorithms often require greater computing costs compared to classical methods. Thus, they are often inappropriate for real-time deployment. One approach that shows great performance in processing images in real-time is the YOLO. In a project by Oltean et al. (2019), a YOLOv3-tiny technique was proposed to achieve real-time requirements. However, a regular YOLOv3 is much more accurate in complex setting. The YOLO approach can process images in real-time at a frame rate of up to 45 fps. Another paper (Kadim et al, 2020) showed that the YOLO model consistently achieved the highest accuracy and highest detection rate in all videos as compared to the other algorithm used which is SSD. In the YOLO algorithm, a single convolutional network simultaneously predicts multiple bounding boxes and class probabilities for those boxes (Redmon et al, 2016). YOLO trains on full images and directly optimizes detection performance.

V. Methodology

A. Image Acquisition and Pre-processing

The training and testing images will be collected from CCTV cameras detected for traffic surveillance. The footage will be cropped out to include only the segment or side of the road with frontal-view of the vehicles and to eliminate the irrelevant parts of the images.

B. Vehicle Localization and Detection

The model that will be used for the proposed method is the You Only Look Once version 3 or YOLOv3. YOLOv3 unifies the separate components of object detection (classification and localization) into a single neural network. Past works have shown that this particular model performs faster that R-CNNs and works greatly in detecting objects even in low lighting environments. The network consists of 24 convolution neural networks (CNN) layers with 2 fully connected (FC) layers. The weights of the first 20 CNN layer are obtained from darknet53 convolutional network. The network was trained with the Imagenet classification dataset. After the aforementioned layers, 4 CNN layers and 2 FC layers which have been train using the COCO dataset are further added. A region of interest will be defined to avoid the multiple detection and counting of a single vehicle. Then, the each of the vehicles detected within the region will be associated with their respective individual trackers. This is to ensure that vehicles are consistently labeled within the tracking period. Single Kalman filter models (constant velocity) will be used for the tracking. This model was created based on the assumption that most vehicles travel at a constant velocity within the short region of a road. Vehicles detected outside the ROI will be excluded from the tracking. A tracker will be given a constructed gating region to figure the potentially associated vehicles. The tracking association is resolved using the global nearest neighbor, which is based on centroid distance and histogram-based appearance similarity. Tracker trajectories for each vehicle will be decided using the lowest centroid and similarity distances. The tracker will be terminated at the end to avoid unnecessary consumption of processing power and time, and improve the system performance. A tracker that misses a vehicle for a consecutive number of frames will be terminated.

C. Vehicle Classification

There is a tendency for vehicles to be classified to different classes in different frames, especially between similar vehicle classes. Thus, for each vehicle, the history of predicted classes within a time period (before it reaches the LOI) will be kept. The frequency of the detected classes will be recorded and updated in each frame. This will be done using the cumulative confidence value of the class starting from the moment the tracker is initialized. Then, the class with the highest frequency will be decided as the final vehicle class.

D. Vehicle Counting

The counting will be executed using a user defined region-of-interest virtual ROI and virtual line-of-interest (LOI). The ROI will be used to filter out vehicles that are outside the desired area of tracking from being detected. Once a vehicle crosses the LOI, the counter will be updated and incremented. To make sure that each vehicle is only counted once, counted vehicles will be indicated as “already counted”. The estimated overall vehicle volume will be quantified according to the weighted vehicle class. Bigger vehicle classes will be assigned bigger weights.

E. Real-time Data Dashboard

Aside from showing the classifications in the video feed, the real-time count and data will be reflected and published real-time in a web application/dashboard. Several visualizations such as bar (reflecting the vehicle count) and time-series (reflecting the vehicle count trend over time) charts will be published in the dashboard.

VI. Conclusion

The conclusion goes here.

Appendix A

Appendix one text goes here.

Appendix B

Appendix two text goes here.

Acknowledgment

The authors would like to thank...

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.