American Journal of Neural Networks and Applications
2022; 8(1): 1-5
http://www.sciencepublishinggroup.com/j/ajnna
doi: 10.11648/j.ajnna.20220801.11
ISSN: 2469-7400 (Print); ISSN: 2469-7419 (Online)

Automated Passenger Detection and Toll Processing System Using Convolution Neural Network

Rishabh Dara[1], Alex Sumarsono[2, *]

[1]Department of Industrial Engineering, California State University East Bay, Hayward, USA
[2]Department of Computer Engineering, California State University East Bay, Hayward, USA

Email address:
rdara@horizon.csueastbay.edu (R. Dara), alex.sumarsono@csueastbay.edu (A. Sumarsono)
[*]Corresponding author

To cite this article:
Rishabh Dara, Alex Sumarsono. Automated Passenger Detection and Toll Processing System Using Convolution Neural Network. *American Journal of Neural Networks and Applications.* Vol. 8, No. 1, 2022, pp. 1-5. doi: 10.11648/j.ajnna.20220801.11

Received: March 5, 2022; **Accepted**: March 28, 2022; **Published**: April 9, 2022

Abstract: As a way to incentivize more people to drive multi-passenger vehicles, policies for high-occupancy vehicle (HOV) lanes and congestion toll discount are put in place at various locations. Being able to identify the correct number of people in a vehicle using the lanes is paramount in deciding the toll for that vehicle. In general, these lanes are operated based on voluntary declarations by the drivers, which makes it prone to abuse where vehicles with fewer occupants than required illegally use the HOV lanes. Hence, the capability to detect violators in real time is very critical. However, in many of the cases, vehicle occupancy detection relies on a labor-intensive manual method. This is quite unreliable and costly in terms of significant loss of revenue. This study proposes to remedy this problem by applying an object detection algorithm based on a deep convolutional neural network, known as the YOLO algorithm. This algorithm can automatically detect the number of occupants in a vehicle with very high degree of accuracy. Images are captured through Near Infrared (NIR) cameras on the HOV lanes. With proper fusion, clear signatures or silhouettes of the front passengers' faces are distinguishable from other inanimate objects in the vehicle. Using YOLOv3 the accuracy reaches 96%. This information is then used to charge the express lane toll. It is estimated that up to 95% of potential loss of revenue could be avoided. It is, therefore, a viable and attractive solution.

Keywords: Computer Vision, High Occupancy Vehicle, Machine Learning, Object Detection

1. Introduction

Population growth has resulted in heavy congestion on highway lanes, leading to both economic and environmental concerns. Recent statistics reveal a monotonic increase in the number of vehicles on highways from 193 million in 1990 to approximately 268.8 million vehicles registered in 2016 in the US [1, 2]. High Occupancy Vehicle (HOV) lanes are standard car-pool lanes where vehicles using those lanes legally must be occupied by a minimum of two (HOV2+) or three (HOV3+) people. In order to avoid congestion, yet still encourage car-pooling, transportation agencies allow cars with one occupant to use carpool lanes by paying tolls. They monitor the number of single occupancy and double occupancy vehicles entering HOV lanes and use the information to charge the drivers tolls. Tolls are not incurred on HOV3+ vehicles. However, to gain the benefits offered by HOV lanes, the number of occupants per vehicle rules need to be strictly enforced. The declaration of the occupancy status of a vehicle is a voluntary compliance by the driver [3]. Unfortunately. in many cases, this declaration is falsified in order to avoid the toll. The current practice relies on visual inspection by road-side officers to enforce these rules. This process is found to be inefficient and costly. Typical violation rates range between 50% and 80%, while manual enforcement rates are typically less than 10% [4]. Therefore, an automated way of counting the number of occupants in a vehicle is extremely necessary for fair tolling [5]. In this paper, the above challenge is addressed by proposing a deep neural network based solution for counting the number of passengers inside a vehicle using an object detection algorithm, known as You Only Look Once (YOLO).

The current setup widely used in transportation for HOV lane enforcement is shown in Figure 1 [6]. Users of HOV lanes activate an RFID tag while entering the tolling booth.

The system solely relies on self-compliance of the driver. Statistics reveal that 80% of the times the driver gives a wrong declaration to avoid a toll [7]. Artificial Intelligence based computer vision is the most effective way of deploying an automated vehicle occupancy counting system [8]. By using vehicle license plate information, the driver of the vehicle, which the system finds a mismatch between the number of passengers declared in voluntary compliance and the number detected, can be fined. Solutions to license plate recognition have been proposed [9]. In this paper, the problem addressed is to count the number of people seated in a vehicle and to charge the toll based on the occupancy.

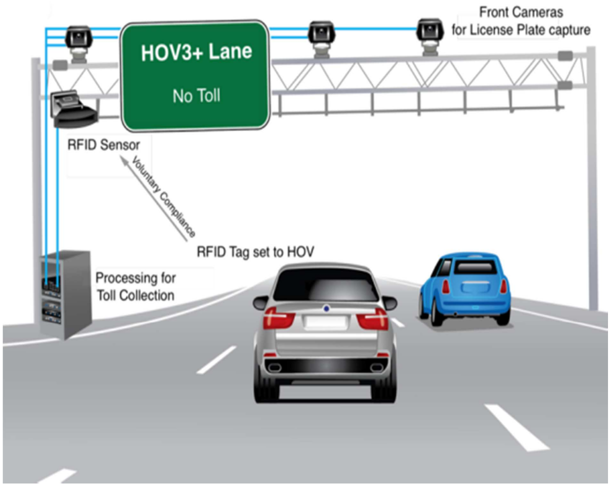

Figure 1. *Existing occupancy counting with RFID.*

The rest of this paper is organized as follows: section 2 describes the methodology of the technique used in this paper, section 3 presents the experiments and the results, and section 4 discusses the monetary implications. The paper ends with some concluding remarks in section 5.

2. Methodology

Features from the entire image are used to find an object in an image [10]. It is accomplished by predicting each bounding box and the associated classes. YOLO first divides the image into an *S*x*S* grid as shown in Figure 2 [11, 12].

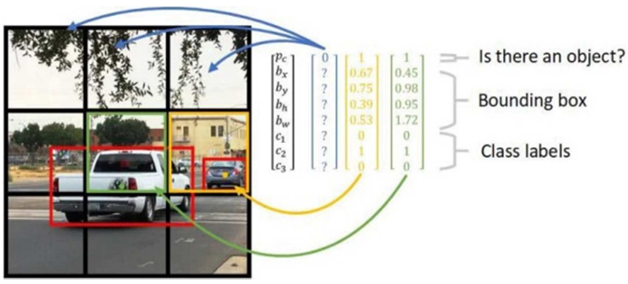

Figure 2. *SxS grid division of image run on YOLO.*

Within each grid cell, YOLO predicts the location, sizes, and confidence scores of the predetermined number of bounding boxes. In other words, it predicts the potential location of an object and its likely class label. If a grid cell contains the center of the object, then the location is correctly predicted by the bounding boxes of that grid cell. Each bounding box contains five predictors: x coordinate, y coordinate, width, height, and confidence score.

The conditional class probabilities and the individual box confidence predictions determine the class-specific confidence scores for each bounding box.

$$\Pr(Class_i|Object) * \Pr(Object) * IOU_{pred}^{truth} = \Pr(Class_i) * IOU_{pred}^{truth} \quad (1)$$

YOLO is run on images captured by the HOV lane cameras. Once a clear silhouette of a passenger's face is obtained from the cameras, predictions are generated in an $S \times S \times (B * 5 + C)$ tensor where each grid cell predicts the location and confidence scores of B number of bounding boxes across C number of classes.

The detection network architecture is shown in Figure 3. It consists of 24 convolutional layers followed by two fully connected layers combined with alternating 1x1 convolutional layers used for feature space reduction.

The class probabilities and bounding box coordinates are predicted by the final layer. The bounding box width and height is normalized with respect to the image width and height, respectively.

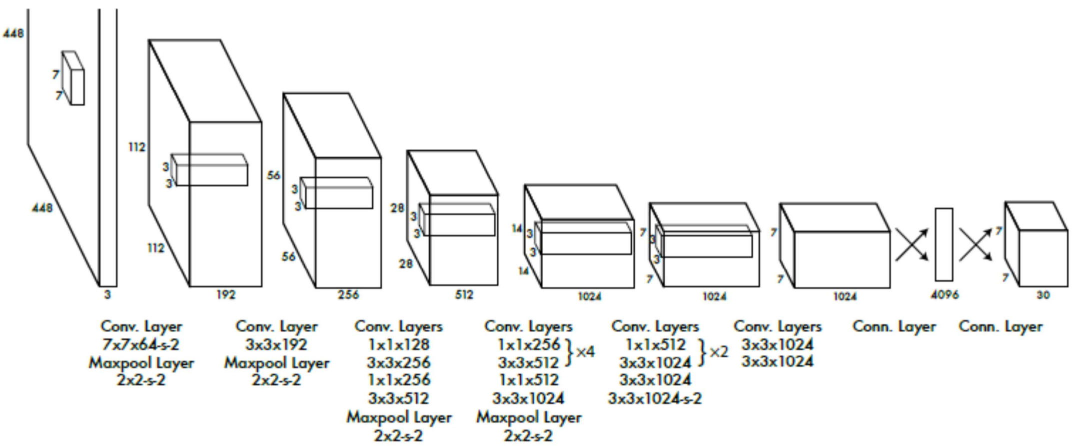

Figure 3. *Bounding boxes with dimension priors and location prediction.*

To minimize the likelihood of running into vanishing gradient problems, a leaky rectified linear activation function is used for all the layers.

$$\phi(x) = x \text{ if } x > 0,\ 0.1 \text{ otherwise} \tag{2}$$

Sum of squared error optimization is used in the output of the model. To avoid model instability, two hyperparameters λ_{coord} and λ_{noobj} are added, where they are set to 5 and 0.5, respectively.

For YOLOv3 the bounding boxes are predicted at three different scales as shown in Figure 4.

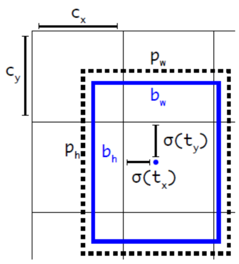

Figure 4. *Bounding boxes with dimension priors and location prediction.*

The overall predictions of the network are given by

$$b_x = \sigma(t_x) + c_x \tag{3}$$

$$b_y = \sigma(t_y) + c_y \tag{4}$$

$$b_w = p_w e^{t_w} \tag{5}$$

$$b_h = p_h e^{t_h} \tag{6}$$

Where t_x, t_y, t_w and t_h are the four coordinates of each bounding box, (c_x, x_y) is the offset from the top left corner and p_w, p_h are the width and height of the bounding box, respectively.

During training binary cross-entropy loss is used for class predictions. A 3D tensor of *N x N x [3*(4+1+80)]* yields the prediction for 4 bounding boxes offsets, 1 object prediction and 80 class predictions.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Figure 5. *Darknet-53.*

YOLOv3 is implemented using 53 convolutional layers known as Darknet-53. It is shown in Figure 5.

Intersection over Union shown in Figure 6 is used as a metric to calculate the confidence score. It indicates the confidence of the model accuracy about a class label within the bounding box and the class membership with respect to the box.

Figure 6. *Confidence score through Intersection of Union.*

A lot of bounding boxes may be created, most of which are irrelevant. To preserve the correct ones, only bounding boxes whose predicted class meets a certain confidence score are preserved. This step allows relevant objects, such as cars and people, to be isolated within the image, as depicted in Figure 7.

Figure 7. *Object detection using YOLO.*

3. Experimental Results

Unlike other CNN-based object detection algorithms, YOLO utilizes a single neural network in the entire detection pipeline. One pass prediction on bounding boxed and class probabilities are obtained directly from the whole image. Therefore, the detection performance can be optimized end-to-end. All these characteristics make this algorithm quite suitable for quick object detection. Within a class of YOLO algorithms, YOLOv3 was chosen for this study. It is an improvement of the initial YOLO algorithm with respect to bounding box prediction, class prediction, and feature extractor.

3.1. Occupancy Detection Capability

The raw data for the experiments was obtained by downloading images of high occupancy lanes with flowing traffic in the San Francisco area [13]. The images on the HOV lanes were captured with Near Infrared (NIR) cameras, which are not affected by tinted windshields. The NIR cameras are modified specifically to work at highway speeds. Images with different vehicle occupancies were selected to test the model. Figures 8, 9, 10 and 11 are examples of the various detection results. Only YOLOv3 is able to successfully detect the occupants inside the vehicle.

Figure 8. *Trial run using YOLOv5n – unable to detect the driver.*

Figure 9. *Trial run using YOLOv5l – unable to detect the driver.*

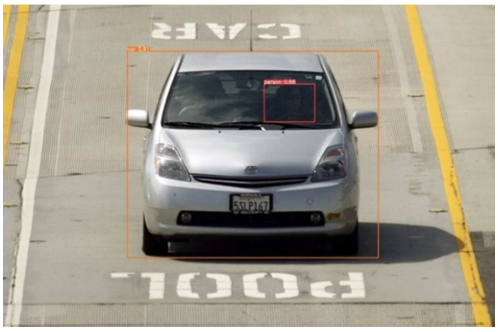

Figure 10. *Single occupancy detection using YOLOv3.*

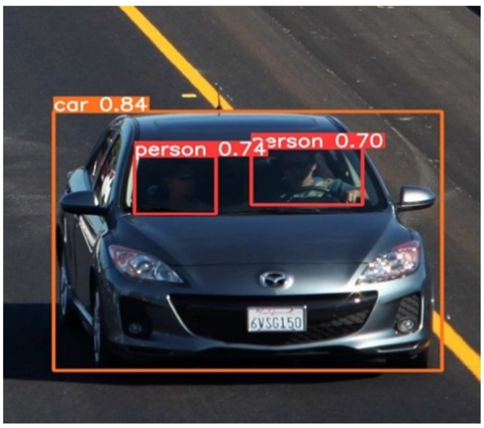

Figure 11. *Double occupancy detection using YOLOv3.*

There is little correlation between the capability to recognize and localize a vehicle and that to detect the occupants inside the vehicle. The scenario in Figure 9 shows a car is found in the image and localized in a bounding box with 90% confidence score. Nevertheless, the driver cannot be detected. On the other hand, in Figure 11 two people inside a car were detected although the detected vehicle had a much lower confidence score. This is due to the independent treatment of each object in YOLOv3, hence, the algorithm is able to decouple an object inside another object without affecting the convergence criteria.

3.2. Classification Performance

In order to find the best version used for this application, the same set of image data shown in Table 1 was used to find the accuracies of different YOLO versions.

Table 1. *Image Dataset.*

Image Name	Size	Content
25.jpg	544x640	1 person, 3 cars, 2 trucks
3.jpg	640x640	2 persons, 1 car
4.jpg	640x576	2 persons, 1 car
5.jpg	480x640	2 persons, 1 car
6.jpg	384x640	1 person, 2 cars, 1 stop sign
7.jpg	384x640	1 person, 2 cars
8.jpg	640x608	1 person, 2 cars
hybrids.jpg	416x640	1 person, 1 car

The performance of different YOLO versions is measured in terms of classification accuracy and classification time. Table 2 below shows the average classification accuracy results.

Table 2. *Average Accuracies for Different YOLO versions.*

Model	Accuracy
YOLOv3	96%
YOLOv5n	76%
YOLOv5x	60%
YOLOv5l	52%

As seen from the table, the results from YOLOv5n, YOLOv5x, YOLOv5l were significantly lower than those from YOLOv3.

Figure 12 shows the plot of the time required to classify an object as a function of the total number of pixels in an image. The range of the classification time is between 1.5 and 2.5 seconds. These numbers fall well within the feasible range of required classification time of moving vehicles, which means the algorithm can be used in real time.

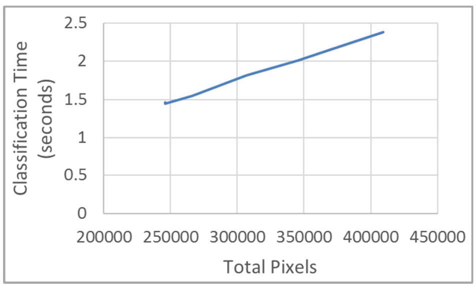

Figure 12. *Classification Time vs Total Number of Pixels.*

Based on the results above, YOLOv3 should be chosen for this application.

4. Monetary Implications

Once the number of people is detected, the toll for using the HOV lane is charged based on the occupancy. The rates for this study are $2.50 for single occupancy and $1.50 for double occupancy. No toll for HOV 3+.

Violations of occupancy declaration has real monetary implications. The estimated loss of revenue is directly proportional to vehicle occupancy, which can be determined either by manual observation or by YOLOv3-based detection algorithm. As previously mentioned, manual observation is able to achieve 10% policy reinforcement. On the other hand, YOLOv3 can detect violations with 96% accuracy yielding much better results.

Assuming $2.50 toll charge for single occupancy and an average of 1,190,136 vehicles passing through toll payment points during the month of February 2022 [14], the estimated loss of revenue is shown in Table 3 where the violation rates are between 50% to 80% [4].

Table 3. *Estimated Monthly Loss of Revenue.*

Violation Rates	Estimated Loss of Revenue	
	Manual Reinforcement	YOLOv3-based Reinforcement
50%	$1,338,903	$59,507
60%	$1,606,684	$71,408
70%	$1,874,464	$83,310
80%	$2,142,245	$95,211

It is clear that employing a machine learning based detection method can lower the loss of revenue by more than 20 times. Hence, from a purely monetary perspective, this should be a strong motivation to deploy such a method.

5. Conclusion

Voluntary declarations by drivers are often abused to leverage the use of HOV lanes illegally. The current vehicle occupancy detection is inefficient and costly. To solve this problem, we applied a powerful yet readily available object detection algorithm, YOLOv3. The experiments show that this algorithm is the most suitable model as it produces the most accurate results. The information can be used to determine the toll amount. From this research we show that a machine-learning based vehicle occupancy counting system is an effective solution that can improve the efficiency and accuracy for determining the number of occupants in a vehicle that uses an HOV lane, therefore, avoiding up to 95% of the potential loss of revenue.

Acknowledgements

This work was supported in part by Department of Computer Engineering at California State University East Bay.

References

[1] Number of cars in the US, (2019). [Online]. https://www.statista.com/statistics/183505/number-of-vehicles-in-the-united-states-since-1990/

[2] Department of Motor Vehicle, (2016). [Online]. https://www.dmv.ca.gov/portal/

[3] Pavlidis, L., Symosek, P., Morellas, V., Fritz, B., Papanikolopoulos, R. & Sfarzo, R. (1999). Automatic passenger counting in the HOV lane. University of Minnesota, USA.

[4] Xu, B., P. Paul, P., Artan, Y. & Perronnin F. (2014). A machine learning approach to vehicle occupancy detection. Proceedings of the 17th International IEEE Conference on Intelligent Transportation Systems, China. https://doi.org/10.1109/ITSC.2014.6957856

[5] Lee, J., Byun, J., Lim J., & Lee, J. (2020). A framework for detecting vehicle occupancy based on the occupant labeling method. Journal of Advanced Transportation. https://doi.org/10.1155/2020/8870211

[6] Schijns S. & Mathews, P. (2005). A breakthrough in automated vehicle occupancy monitoring systems for hov/hot facilities. 12th HOV Systems Conference.

[7] Kumar, A., Gupta A., Santra, B., Kolla, M., Gupta, M. & Singh, R. (2019). VPDS: An AI-based automated vehicle occupancy and violation detection system. Proceedings of the AAAI Conference on Innovative Applications of Artificial Intelligence, USA. https://doi.org/10.1609/aaai.v33i01.33019498

[8] Bulan, O., Kozitsky, V., Ramesh P. & Shreve, M. (2017). Segmentation and annotation free license plate recognition with deep localization and failure identification. IEEE Transactions on Intelligent Transportation Systems. https://doi.org/10.1109/TITS.2016.2639020

[9] Jang, J. (2021). High-occupancy vehicle lane enforcement system. The Open Transportation Journal. http://dx.doi.org/10.2174/1874447802115010194

[10] Szegedy, C., Liu, W., Jia Y. & Sermanet, P. (2015). Going deeper with convolutions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, USA. https://doi.org/10.1109/CVPR.2015.7298594

[11] Redmon J. & Farhadi, A. (2017). YOLO9000: better, faster, stronger. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, USA. https://doi.org/10.1109/CVPR.2017.690

[12] Redmon, J. & Farhadi, A. (2015). YOLOv3: An incremental improvement. University of Washington, USA.

[13] Caltrans. (2021). High-occupancy vehicle systems. [Online]. https://dot.ca.gov/programs/traffic-operations/hov/

[14] The Tool Roads of Orange County (2022). [Online]. https://thetollroads.com/about/investor/transactions