VPDS: An AI-based automated vehicle occupancy and violation detection system

Rishabh Singh*, Aishwarya Gupta*, Abhinav Kumar*[1], Mayank Gupta*, Lalitha KS*, Bishal Santra*[2], Manasa Kolla*

Computer Vision and Media Analytics Group, Conduent Labs *

Abstract

The High Occupancy Vehicle/High Occupancy Tolling (HOV/HOT) lanes are operated based on voluntary HOV declarations by drivers. A majority of these declarations are wrong to leverage faster HOV lane speeds illegally. It is a herculean task to manually regulate HOV lanes and identify these violators. Therefore, an automated way of counting the number of people in a car is expedient for fair tolling and for violator detection.

In this paper, we propose a Vehicle Passenger Detection System (VPDS) which works by capturing images through Near Infrared (NIR) cameras on the toll lanes and processing them using deep Convolutional Neural Networks (CNN) models. Our system has been deployed in 3 cities over a span of two years and has served roughly 30 million vehicles with an accuracy of 97% which is a remarkable improvement over manual review which is 37% accurate. Our system can generate an accurate report of HOV lane usage which helps policy makers pave the way towards de-congestion.

Introduction

Intelligent Transportation Systems (ITS) improve safety and mobility through the integration of sensing, computational power and advanced communications into the transportation infrastructure (Xu et al. 2014). Such systems enable efficient management of lanes by incorporating various aspects like carpooling, tolling, traffic management and transit in a multi-purpose roadway. This creates novel avenues for agencies in terms of congestion pricing in order to generate revenue and manage demand dynamically. Population growth has resulted in heavy congestion on the 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 USA (sta 2018).

*Equal Contribution by all authors

[1]School of Computing, University of Utah

[2]Indian Institute of Technology Kharagpur

*Conduent Labs, manasa.manasa@conduent.com

HOV lanes are standard car-pool lanes where a minimum of two (HOV2+) or three (HOV3+) vehicle occupants are required to legally use the lane, making them lesser congested and enabling efficient rapid transit (Daley et al. 2011). In order to avoid congestion, encourage car-pooling and to address the concerns related to air pollution, agencies have come up with a system which also allows the car with only one occupant to use the carpool lanes. These lanes are known as High Occupancy Toll (HOT) lanes and have a variable toll. Generally, agencies try to maintain a minimum speed of 45 miles per hour (mph) on such lanes, so they need to monitor the number of single occupant vehicles allowed to enter HOV lanes. This is done typically by charging a toll adjusted to the dynamic congestion in the lane. The toll price is increased if the average traffic speed in the HOT lane decreases below the accepted minimum.

However, to gain the benefits offered by HOV/HOT lanes, the entry rules (number of occupants in the vehicle) need to be be enforced vigilantly. The declaration of the occupancy status of a vehicle is a voluntarily compliance by the driver. But in most cases, this declaration is falsified in order to avoid the toll. The current practice is to rely on visual inspection by road-side officers to enforce these rules, but the process is found to be inefficient, costly and potentially dangerous (Artan et al. 2016). Typical violation rates can exceed 50-80%, while manual enforcement rates are typically less than 10% (Schijns and Mathews 2005). While tagging genuine passengers as violators causes discomfort among the consumer base, allowing too many violators in an HOV lane nullifies its purpose. In either case, the loss is incurred by the transportation agency providing the service. Therefore, an automated way of counting the number of occupants of a car is extremely necessary for fair tolling and violator detection.

In this paper, we address the above challenge by proposing a Vehicle Passenger Detection System (VPDS) - a deep neural network based framework for counting the number of passengers inside a vehicle by processing the images of its front and rear view. Our contribution to the problem is two-fold. We first apply a state-of-the-art object detection technique, YOLOv3, for the

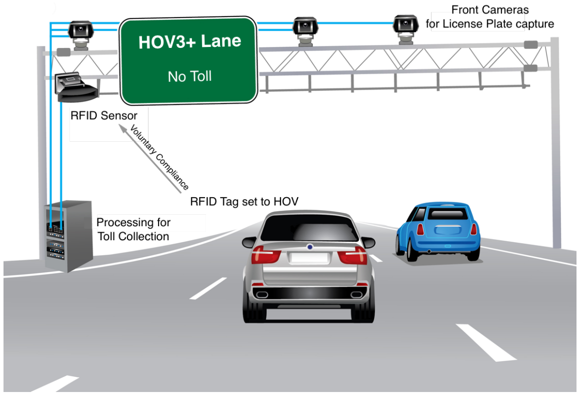


Figure 1: Present: Counting done with an RFID pass.

front and the rear window detection. This eliminates the need for doing any manual parameter tuning which is often required for Deformable Parts based Models (DPM). Next, we perform classification (to count number of people) separately for the front and rear regions of interest, i.e. windows using GoogleNet-based models. Further, we combine the outputs to arrive at the decision of HOV/not-HOV for a vehicle. We showcase the effectiveness of the proposed solution in terms of superior performance to existing approaches within the constraints of inferior image quality and significant external factors like illumination, occlusion and traffic congestion.

Application Description

A roadside observer is often posted near HOV/HOT lanes to manually enforce the rules, but the accuracy and reliability of such a method is abysmally low (Artan et al. 2016). The rate of passage of vehicles through an HOV lane can be as high as 1 vehicle per second, which makes for too many cars for a human observers to process. Therefore, manual enforcement of carpooling lanes is difficult on urban highway systems.

This is specially relevant because of the voluntary compliance system in place in most car pooling lanes. As shown in Figure 1, even HOT lanes operate based on an RFID tag and self compliance by the driver. The statistics reveal that 80% of the times the driver gives a wrong declaration to avoid toll (Schijns and Mathews 2005). Therefore, automated methods to identify and fine the violators have to be developed which will result in better compliance in the usage of car pooling lanes and hence, lesser congestion.

Computer vision with Artificial Intelligence (AI) is the most effective way of developing an automated vehicle occupancy counting system. Figure 2 shows the entire pipeline for identifying HOV violators. At first, the front and the rear seat[1] images are captured by two cameras, one of which is aimed at the oncoming traffic,

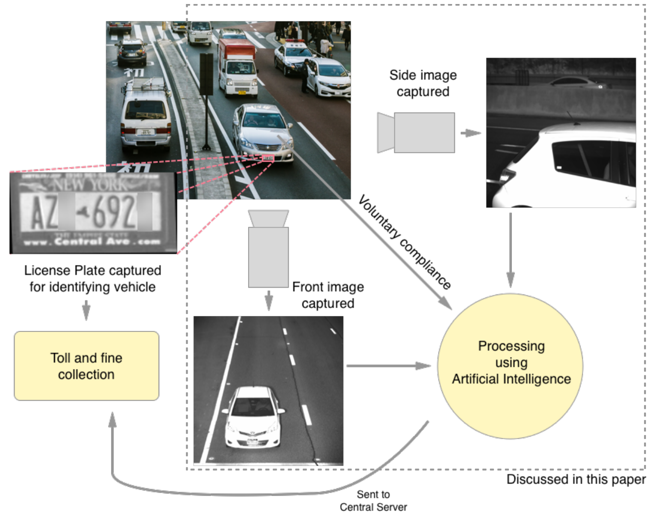


Figure 2: Proposed vehicle occupancy count processing pipeline. This involves capturing and localizing the front and the rear images by two distinct cameras and processing them using AI.

while the other is set perpendicular to it. These two images are then processed using AI based approaches. In case, the system finds a mismatch between the number of passengers in the vehicle declared during voluntary compliance and the number detected by the system, the driver is fined with the help of the license plate information of the vehicle. License plate recognition is a separate problem which has been tackled in (Bulan et al. 2017). For this paper, the problem we address is to count the number of people seated in a vehicle and classify the vehicle as an HOV3+ violator or a non-violator.

Background

The images captured by the hardware contains information from the entire scene. The front and side views of a vehicle form a limited region of the image only. The rest of the image is redundant background noise such as the road, street lights and other vehicles present in the scene. This superfluous information can hinder the performance of classification. As such, we need to crop the relevant portion of the images which is helpful towards the task of counting passengers. Thus, the first step in the framework is extraction of region of interest (ROI), precisely, the windows from both the images.

Previous works (Xu et al. 2014) in occupancy detection use DPM (Felzenszwalb et al. 2010) for ROI extraction, but there are sizable limitations to it. DPM relies heavily on image preparation, cropping and in general the saliency of the vehicle in the pre-processed image. Therefore, in this paper, we propose ROI extraction using YOLOv3, (Redmon and Farhadi 2018) which does not have such heavy dependencies on image pre-processing.

The next step in the pipeline is to count the number of people in vehicle using these ROIs. We extensively

[1] The terms "rear" (rear seats) and "side" have been used interchangeably for the side view of the vehicle.

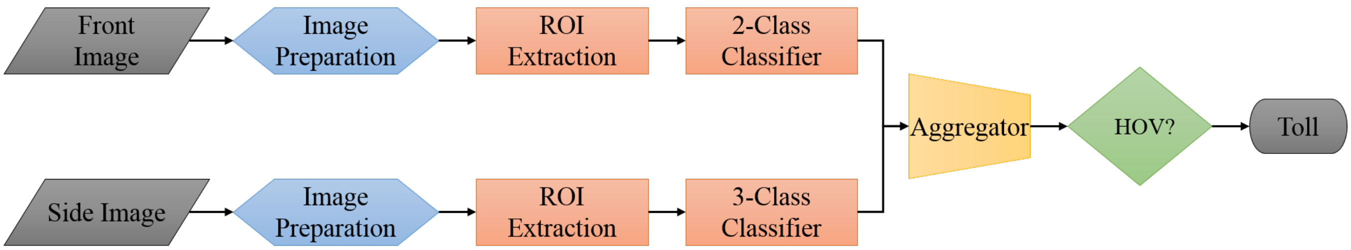


Figure 3: Flowchart of passenger counting and HOV violator detection using Computer Vision techniques. ROI extraction from front and rear images is followed by individual classifiers, results of which are aggregated for the final decision.

surveyed approaches proposed in the literature for addressing the problem of occupancy detection. These methods can be broadly classified into three main categories viz, detection, feature and density based methods as explained below.

Detection-based methods Face detection has been extensively explored in this domain for counting people by detecting passenger faces using a pixel threshold (Wang, Xu, and Paul 2015). Skin detection (Hao, Chen, and Li 2006) has also been explored in which front image of the vehicle is processed using a color skin model to coarsely detect the facial region. The number of occupants can also be estimated by detecting the empty seats in a vehicle (Fan et al. 2013). However, passengers in a vehicle do exhibit arbitrary poses. Especially in side view images, the visibility of faces is a prominent issue given frequent occlusion. In such cases, detection based methods do not yield convincing results.

Feature-based methods These methods try to synthesize features which can capture the difference between a passenger and his/her surroundings. Some of the popular features are Histogram of Oriented Gradients (HOG) (Dalal and Triggs 2005), Bag of Words, Fisher Vectors (Perronnin and Dance 2007) or a combination of these (Skaff et al. 2014). Some approaches have explored distance-based metrics between descriptors in order to discriminate between images having only the driver or both the driver and the passenger in the front image of a vehicle (Xu, Paul, and Perronnin 2017). Fisher Vectors with classifiers are proven to have superior performance to DPM based models in terms of accuracy and capturing bad illumination, pose-variability and occlusion (Artan and Paul 2013). These features, however, fail to capture the variability in low resolution/lesser informative images, which is even more prominent in a real setting.

Density-based methods These methods aim to estimate the count of people in an image by learning to create a density map of the input image (Sindagi and Patel 2018). This entire spectrum of approaches fail to estimate people count in low density scenarios and given the limit on the number of people seated in a vehicle, these methods fail in our problem.

Part of the framework included in this paper were earlier reported in (Xu et al. 2014; Artan et al. 2016; Wshah et al. 2016). This paper claims sufficient novelty and improvements over these previous works. (Xu et al. 2014) uses DPM for localization and Support Vector Machines (SVM) for classification while (Artan et al. 2016) uses DPM for windshields extraction from images and feature-specific models for classification. (Wshah et al. 2016) proposes a solution for detecting whether there is a passenger present in a vehicle or not, which is essentially a binary classification and thus can only be used for figuring HOV2+ violations. Our work not only improves detection, but also counts the exact number of people in a vehicle by incorporating multi-class classification in the rear images. This makes our solution suitable even for detecting HOV3+ violations which is a lot more generic and modular than their solution.

Usage of AI Technology

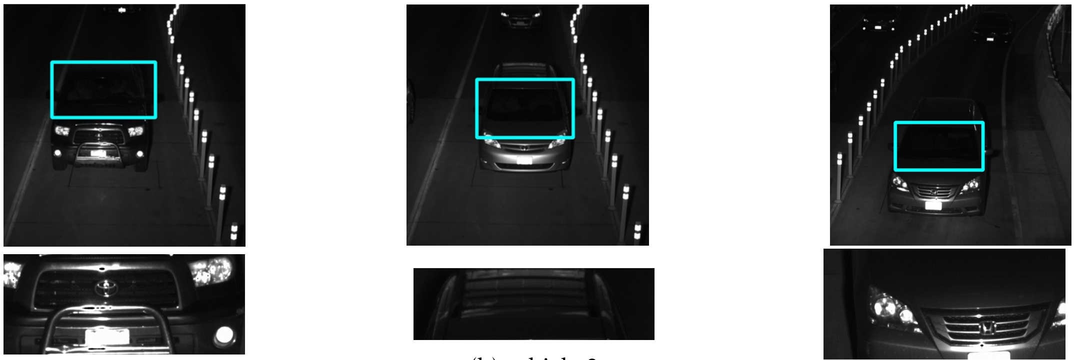
Figure 3 describes the vehicle occupancy counting system formulated as a computer vision problem.

The front and rear image capture are two separate processing streams whose results are combined to obtain the HOV decision. Both these streams have ROI extraction and classification (human counting) modules in common. Each of these modules are explained in the following subsections.

ROI Detection

As previously mentioned, ROI extraction is a necessary step before classification to remove redundant background information.

We evaluate DPM and YOLOv3 on images of a particular day without any pre-processing (crop, rotation or scaling) of the data. We found that DPM was able to find the correct ROIs in 52% of the cases, while YOLOv3 was correct in more than 96% cases. The correctness of ROIs was determined based on an intersection over union (IOU) threshold with the ground truth labels. Further, forward pass on YOLOv3 is 10 times faster than on DPM. Figure 4 clearly depicts that YOLOv3 detects the windshields accurately despite the presence of other similar entities like sunroofs, which in the case of DPM



(a) vehicle 1 (b) vehicle 2 (c) vehicle 3

Figure 4: Comparison of ROI detection by YOLOv3 (top) and DPM (bottom). Clearly, the DPM based method fails.

have to be removed manually by setting crop parameters for each lane. This was one of the main drawbacks of VPDS methods like (Wshah et al. 2016) which use DPM for ROI extraction. Even after manual parameter setting, the detection accuracy of DPM based models is lesser than YOLOv3. Thus, YOLOv3 is the clear victor for real-time applications like transportation in terms of both time and accuracy.

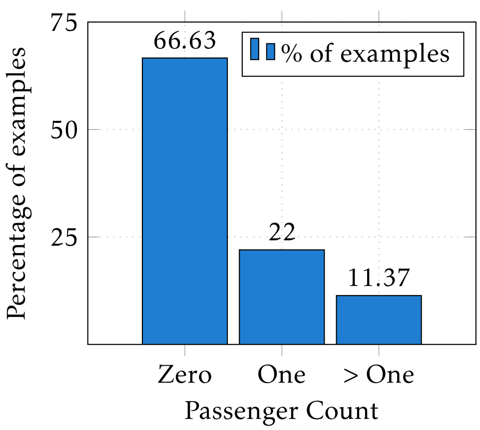


Figure 5: Class distribution of number of passengers seated in the rear images of the vehicle. Note that the dataset has most of the images in which the rear seat is empty while a very few number of images have more than one passenger seated in the rear of the vehicle.

Person Counting

The number of passengers in a vehicle can be divided into front and rear occupants. The ROIs extracted for each view are processed by two separate CNNs. The front row can contain only one person apart from the driver, thus the problem turns into a binary classification of presence/absence of a passenger. We use GoogleNet (Szegedy et al. 2015) in conjunction with YOLOv3 as ROI detector for this task which gives an accuracy of more than 97%. Our method is a significant improvement over the DPM-enabled classifier which has an accuracy of 94.6% when used along with lane specific pre-processing of images according to manually preset parameters.

The rear classification is a more challenging task because of several reasons. Firstly, it is a four-class classification problem since the number of passengers can range from zero to three. However, for HOV3+ predictions, three classes (zero, one and more than one classes) are needed. Secondly, a heavily skewed dataset results in a bias in predictions. The skewness of the rear classes is shown in Figure 5. The rear seat is empty in majority of the images while a fewer number of them have more than one passenger seated. Lastly, since the images are from the side, the passengers are quite susceptible to occlusion by one another. These challenges, which could result in huge errors in identification of a person by existing methods, demand an AI-engine robust to such externalities. Thus, we investigated three popular CNN architectures viz. GoogleNet (Szegedy et al. 2015), ResNet (He et al. 2016) and VGGNet (Simonyan and Zisserman 2014) after oversampling the examples of class 1 (1 passenger) and class 2 (2 or more passengers) to match the number of class 0 (no passenger) examples. The front and rear count of passengers, as obtained from their respective deep CNNs, are eventually accumulated to get an estimate of the count of occupants in a vehicle. This count can be compared to a rule, say 3 people, to arrive at the decision of HOV3 violation or not.

Application Use and Payoff

According to the global traffic survey (inr 2017), drivers in New York spent 91 peak hours stuck in traffic. This traffic congestion will cost an average $100 billion over the next five years in the form of wasted time and gas. HOV lanes minimize the delay caused due to traffic congestion by promoting car pooling and hence reduce the number of cars on the highway. These lanes reduce the average travel time by 70% for vehicles moving on HOV

Model	Precision	Recall	Accuracy	Group Accuracy
Deep Classification Models				
ResNet	83.27	77.27	92.57	84.73
VGG-16	88.58	**85.29**	**95.01**	88.32
GoogleNet	**89.01**	83.66	94.82	**88.68**
Person Detection Models				
YOLOv3	87.20	77.93	93.50	86.20

Table 1: Performance comparison (in percentage) of different counting models - HOV3+ and group . (front + rear)

lanes. Additionally, for vehicles on the usual lanes the delay reduces by 50%. Further, the cars which have less number of people can still take HOV lanes during congestion by paying a toll. In fact, 33% of vehicles choose to use HOV lanes. The automated vehicle occupancy counting system as described in this paper proves to be extremely effective for non-intervening functioning for huge volumes of traffic flowing across the highways throughout the day all round the year with negligible maintenance.

Table 1 shows the results of using different CNNs on the person counting task. In addition to the standard precision, recall and accuracy, we also report the *group accuracy*. Group accuracy is calculated when individually both the front and rear passengers are counted correctly by the nets rather than the total count of HOV3+ violators or non-violators.

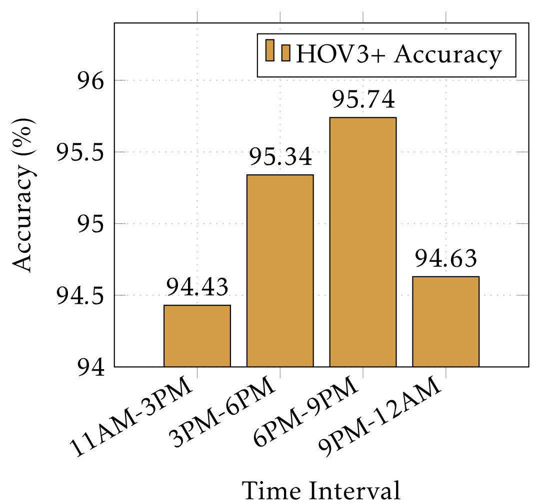


Figure 6: HOV3+ classification accuracy during various intervals of a particular day in October, 2017. HOV3+ classification using the proposed solution VPDS is in the range 94-96%.

Figure 6 depicts the performance of the deployed system during peak hours of congestion. As evident from the bar graph, the accuracy of classification stays fairly constant in the range 94-96%. There is a relative drop in performance during noon and early-night hours. The former can be attributed to significant glare present due to the sun which hinders the visibility of

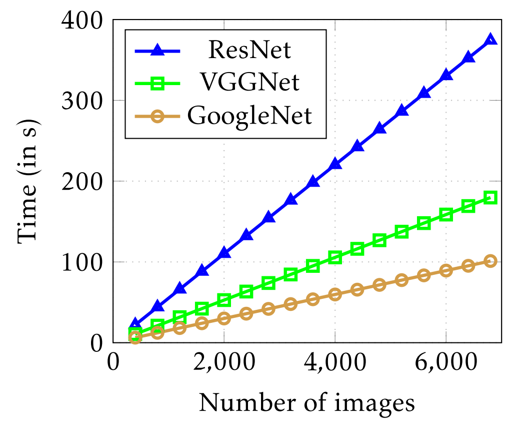


Figure 7: Comparison of latency in decision for different deep CNNs.

human faces, especially in the front window. The latter variation can be safely ignored as the results are within margin of error as the traffic volume falls considerably post 9 PM.

Figure 7 depicts the time taken by the rear AI-engine for processing images in real-time. Our proposed system based on GoogleNet takes approximately half the time taken by VGGNet. This in conjunction with the high accuracy achieved (Table 1) demonstrate the effectiveness of the overall system. The numerical figures corroborate our claim of fast processing without compromising on accuracy of decision. Thus the proposed system is efficient even for high traffic flow.

The variation in system level accuracy and yield in terms of the system confidence threshold can be observed in Figure 8. The system level confidence is formulated using the individual GoogleNet classification confidences for front and rear classifiers. This trend facilitates the use of decision threshold as a design parameter to obtain a better violator detection rate in terms of accuracy. This also caters for the transportation agency to have a control over allowed false positive rate (non-violators detected as violators) of the system. This increasing trend of accuracy with respect to the confidence threshold also guarantees the correctness of the classifier confidence. On the other hand the yield curve acts as another system performance measure showcasing the fraction of vehicles classified for

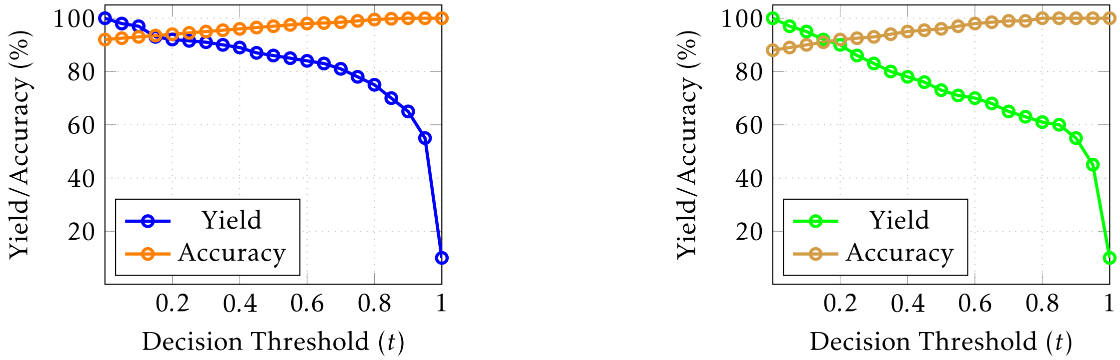


Figure 8: Variation of accuracy and yield of VPDS with threshold on Confidence Score for (a) HOV3+ and (b) Non-HOV3+ predictions. The final confidence score is obtained by multiplying the softmax output of front and the rear classifier

varying confidence thresholds.

We now showcase the results of rear-counting since it is a challenging problem. Table 2 shows the precision, recall and accuracy achieved by three classification CNNs and YOLOv3 as a person detection system. The overall accuracy of rear classification is highest for GoogleNet. The counting of rear images is less accurate since the CNNs get confused between the images with one versus more than one person sitting in the back due to occlusion. This is confirmed by the visualization of the features learned by GoogleNet for the rear classification task which is shown in Figure 9. The blue cluster corresponding to empty seats is well-separated from the highly overlapping red and green clusters. This depicts the high accuracy in detecting the presence/absence of a person by the AI model. However, it gets slightly confused in demarcation between one or more passengers owing to significant occlusion when seen from side.

Model	**Precision**	**Recall**	**Accuracy**
Deep Classification Models			
ResNet	72.21	71.46	86.87
VGG-16	81.22	**79.35**	91.04
GoogleNet	**81.6**	79.32	**91.08**
Person Detection Models			
YOLOv3	79.92	72.77	88.4

Table 2: Performance comparison of the models for the rear passenger count (in percentage)

In addition, we also plot the yield versus accuracy curve for the rear classification models in Figure 10. It is consistent with the table 2 that GoogleNet is more accurate in counting people than VGGNet and Resnet for the same yield.

The proposed system is very robust to certain factors that we showcase. The rear seat looks empty in Figure 11(a), but histogram equalization shows a baby on a safety seat in Figure 11(b). The baby passenger, although not visible to human eyes, is correctly identified as one person by the system.

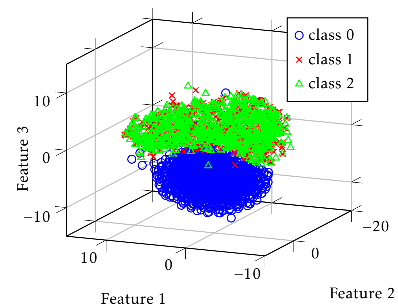


Figure 9: A 3-D t-SNE plot of GoogleNet features for the rear classification task. The plot shows the effect of occlusion on the decision made by GoogleNet.

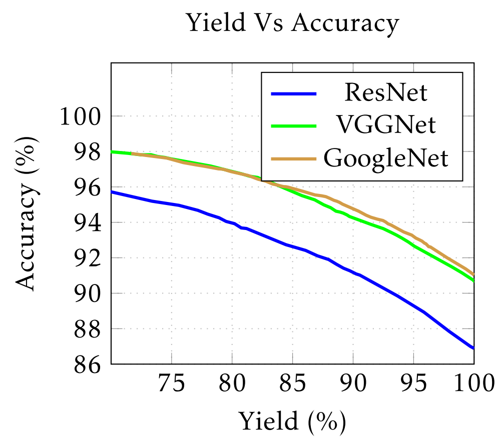


Figure 10: Yield vs Accuracy curve for the rear passenger counting using three popular CNN architectures.

The performance of the system is slightly affected in the presence of infants in the rear seat, as they are often occluded by the safety carrier. Given the ambiguity in the presence of a child, as shown in Figure 12(a), we labelled such data containing only a safety carrier as "No occupant". However, whenever even the slightest features of a child is visible, the system correctly recognizes him/her as a person 12(b). Various other scenarios shown in Figure 13 like a cat present at the back (0 person), occlusion of face, only hair of a passenger being visible or pose variation which lead to failure of manual counting process, are efficiently handled by VPDS with high accuracy.

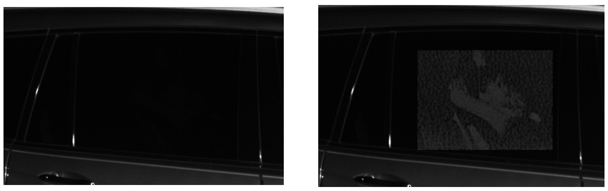


Figure 11: Effect of Tint and Illumination: (a) Tinted window (b) Histogram equalized

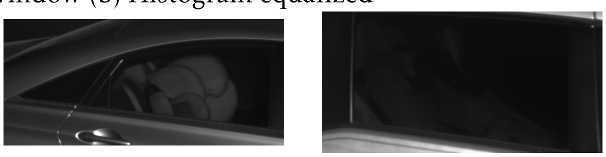
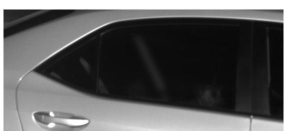


Figure 12: Child-safety seats: (a) Occluded (b) Non-occluded



(a) No Person

(b) Two Persons

(c) Two Persons

(d) Two Persons

Figure 13: Corner cases correctly predicted by VPDS. Faces have been redacted to preserve privacy.

Application Development, Deployment and Maintenance

We use a database of around 35000 raw images of front and rear views of vehicles. This dataset is then split randomly into 3 sets - train, validation and test which contain roughly 24000, 4000 and 7000 images respectively. The train set images are shown to the network while the validation set is used to do the model selection. We evaluate the performance of the model on the test set. We trained our models on an NVIDIA Tesla K80 GPU. We used DarkNet (Redmon 2013 2016) and Caffe (Jia et al. 2014) for training the YOLOv3 and the classifier models respectively.

VPDS has been deployed at more than three sites and functions efficiently at a very high accuracy. The system is tunable based on the site requirements for more sensitive violator detection or less sensitive non-violator ticketing. The system has been successful in ticketing the 80% of violators arising from the voluntary compliance system and proved very beneficial for the transportation law enforcement agencies. VPDS requires a one time installation of the imaging equipment at the site. The images from the site are collected and a site specific model is trained to achieve the best possible accuracy. Once the training is done there is minimal or no requirement of any intervention at the site except in the rare event of any hardware failure. The local AI processing station is also configured one time and does not require monitoring except during a system upgrade. Also, the images collected during the process can be backed up elsewhere to avoid flooding the processing server.

Conclusion and Future Work

The reliability of voluntary compliance is questionable as studies have shown that 80% of the vehicles in an unmonitored HOV lane are in violation of the law. With the ever spreading urban sprawl and an overwhelming dependency of US cities on automobiles, decongestion is one of the highest priorities. We have developed VPDS, an AI based vehicle passenger detection system to effectively enforce HOV/HOT lane movement. VPDS automates and improves identification of HOV violators and assigning fines and tolls to HOV lane users. Moreover, it is extremely fast and takes less than 2s for classifying a vehicle as a violator or a non-violator with 96% accuracy without thwarting the normal traffic flow and using minimal hardware. Over a period of 2 years during which VPDS was deployed at three different sites in US, it has served approximately 30 million passengers. Serving roughly 1800 vehicles in morning and 2300 vehicles in the evening rush hours at one particular site, VPDS achieved an accuracy between 94-96% irrespective of the traffic flow or time of the day.This exemplifies an AI-based system which is highly accurate, consistent, fast, responsive in real-time, robust to externalities and requires little maintenance. In the future, we aim to further improve the efficiency of VPDS by exploring methods such as RFCN (Dai et al. 2016) for better ROI detection and models such as NASNet (Zoph and Le 2016) or Fisher Vectors with neural networks (Perronnin and Larlus 2015) for better classification. We also envision to make a holistic system with vehicle type identification as a sub-module working in conjugation with VPDS to automatically generate toll/fine so that congestion and violation management happen in a seamless manner.

References

Artan, Y., and Paul, P. 2013. Occupancy detection in vehicles using fisher vector image representation. *arXiv preprint arXiv:1312.6024*.

Artan, Y.; Bulan, O.; Loce, R. P.; and Paul, P. 2016. Passenger compartment violation detection in hov/hot lanes. *IEEE Trans. Intelligent Transportation Systems* 17(2):395–405.

Bulan, O.; Kozitsky, V.; Ramesh, P.; and Shreve, M. 2017. Segmentation-and annotation-free license plate recognition with deep localization and failure identification. *IEEE Transactions on Intelligent Transportation Systems* 18(9):2351–2363.

Dai, J.; Li, Y.; He, K.; and Sun, J. 2016. R-fcn: Object detection via region-based fully convolutional networks. In *Advances in neural information processing systems*, 379–387.

Dalal, N., and Triggs, B. 2005. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, 886–893. IEEE.

Daley, W.; Arif, O.; Stewart, J.; Wood, J.; Usher, C.; Hanson, E.; Turgeson, J.; Britton, D.; et al. 2011. Sensing system development for hov/hot (high occupancy vehicle) lane monitoring.

Fan, Z.; Islam, A. S.; Paul, P.; Xu, B.; and Mestha, L. K. 2013. Front seat vehicle occupancy detection via seat pattern recognition. US Patent 8,611,608.

Felzenszwalb, P. F.; Girshick, R. B.; McAllester, D.; and Ramanan, D. 2010. Object detection with discriminatively trained part-based models. *IEEE transactions on pattern analysis and machine intelligence* 32(9):1627–1645.

Hao, X.; Chen, H.; and Li, J. 2006. An automatic vehicle occupant counting algorithm based on face detection. In *Signal Processing, 2006 8th International Conference on*, volume 3. IEEE.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

2017. INRIX Global Traffic Scorecard. `http://inrix.com/scorecard/`. [Online; accessed 4-Sep-2018].

Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; and Darrell, T. 2014. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*.

Perronnin, F., and Dance, C. 2007. Fisher kernels on visual vocabularies for image categorization. In *2007 IEEE conference on computer vision and pattern recognition*, 1–8. IEEE.

Perronnin, F., and Larlus, D. 2015. Fisher vectors meet neural networks: A hybrid classification architecture. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3743–3752.

Redmon, J., and Farhadi, A. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Redmon, J. 2013–2016. Darknet: Open source neural networks in c. `http://pjreddie.com/darknet/`.

Schijns, S., and Mathews, P. 2005. A breakthrough in automated vehicle occupancy monitoring systems for hov/hot facilities. In *12th HOV Systems Conference*, volume 1.

Simonyan, K., and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Sindagi, V. A., and Patel, V. M. 2018. A survey of recent advances in cnn-based single image crowd counting and density estimation. *Pattern Recognition Letters* 107:3–16.

Skaff, S.; Xu, B.; Paul, P.; Saunders, C.; and Perronnin, F. 2014. Occupancy detection for managed lane enforcement based on localization and classification of windshield images. US Patent 8,824,742.

2018. Number of cars in the US. `https://www.statista.com/statistics/183505/number-of-vehicles-in-the-united-states-since-1990/`. [Online; accessed 4-Sep-2018].

Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; and Rabinovich, A. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1–9.

Wang, Y. R.; Xu, B.; and Paul, P. 2015. Determining a pixel classification threshold for vehicle occupancy detection. US Patent 9,202,118.

Wshah, S.; Xu, B.; Bulan, O.; Kumar, J.; and Paul, P. 2016. Deep learning architectures for domain adaptation in hov/hot lane enforcement. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, 1–7. IEEE.

Xu, B.; Paul, P.; Artan, Y.; and Perronnin, F. 2014. A machine learning approach to vehicle occupancy detection. In *Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on*, 1232–1237. IEEE.

Xu, B.; Paul, P.; and Perronnin, F. 2017. Vehicle occupancy detection using passenger to driver feature distance. US Patent 9,760,783.

Zoph, B., and Le, Q. V. 2016. Neural architecture search with reinforcement learning. *arXiv preprint arXiv:1611.01578*.