Real-time image-based parking occupancy detection using deep learning

Debaditya Acharya
acharyad@student.unimelb.edu.au

Weilin Yan
jayan@student.unimelb.edu.au

Kourosh Khoshelham
k.khoshelham@unimelb.edu.au

Infrastructure Engineering, The University of Melbourne

Abstract

Parking Guidance and Information (PGI) systems have a potential to reduce the congestion in crowded areas by providing real-time indications of occupancy of parking spaces. To date, such systems are mostly implemented for indoor environments using costly sensor-based techniques. Consequently, with the increasing demand for PGI systems in outdoor environments, inexpensive image-based detection methods have become a focus of research and development recently. Motivated by the remarkable performance of Convolutional Neural Networks (CNNs) in various image category recognition tasks, this study presents a robust parking occupancy detection framework by using a deep CNN and a binary Support Vector Machine (SVM) classifier to detect the occupancy of outdoor parking spaces from images. The classifier was trained and tested by the features learned by the deep CNN from public datasets (PKLot) having different illuminance and weather conditions. Subsequently, we evaluate the transfer learning performance (the ability to generalise results to a new dataset) of the developed method on a parking dataset created for this research. We report detection accuracies of 99.7% and 96.7% for the public dataset and our dataset respectively, which indicates the great potential of this method to provide a low-cost and reliable solution to the PGI systems in outdoor environments.

1 Introduction

People spend on average 7.8 minutes in cruising for a parking spot. This accounts for about 30% of the traffic flows in cities (Arnott and Inci, 2006), and contributes to traffic congestion during the peak hours. To alleviate this issue and save time and effort in finding a vacant parking space, PGI systems (Chen and Chang, 2011) have been developed. PGI systems require accurate and up-to-date information on the occupancy of parking spaces to be able to provide the users with reliable guidance to vacant spots.

The advantages of using camera-based PGI systems as compared to the other existing systems are threefold (Ichihashi et al., 2009; Bong et al., 2008; True, 2007). Firstly, there is no requirement for additional infrastructure, provided that the facility is equipped with CCTV surveillance cameras covering the parking spaces. Secondly, camera-based systems provide the exact location of the vacant parking spaces which is a requirement for navigation of the vehicles to the vacant parking spaces. Thirdly, camera-based methods are highly applicable to on-street and residential parking spaces.

Image-based parking occupancy detection essentially involves the detection of vehicle objects in parking spaces. In the literature, object detection has been mostly performed by extracting hand-crafted visual features, such as Scale-Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF) and Histogram of Oriented

Gradients (HOG) from the images (Girshick et al., 2014) and their subsequent classification. The drawback of using the hand-crafted features is the limited ability of such features to adapt to variations of the object appearance that is highly non-linear, time-varying and complex (Yilmaz et al., 2006; Chen et al., 2016). Deep CNNs overcome this limitation by learning features that optimally describe the image content. It has been shown that CNNs pre-trained by large image datasets yield a remarkable performance in a variety of image recognition and object detection tasks (Acharya et al., 2017; Donahue et al., 2014; Hong et al., 2015; Wang et al., 2015).

The hypothesis of this research is that features extracted by a pre-trained CNN can be used directly to train an SVM classifier for the detection of parking occupancy in a CCTV image sequence. This is usually referred to as transfer learning, which is an active area of research in machine learning. To test this hypothesis, we use a pre-trained CNN to extract features and train an SVM classifier from a publicly available dataset of parking images. The trained SVM classifier is subsequently used to classify the occupancy of a dataset created for the purpose of this research, which includes a sequence of images captured by a camera overlooking a street with marked parking bays. The results are compared to the state-of-the-art methods that fine-tune a pre-trained CNN for the classification task. The main contributions of the present work are the following:

- A transfer learning approach to parking occupancy detection is proposed and its performance is evaluated by using visual features extracted by a deep CNN directly
- A detailed accuracy analysis is performed to identify the parameters that affect the accuracy of the framework

We report results that indicate the potential of the method in terms of accurate transfer learning and robustness. The developed framework is suitable for real-time applications with a simple desktop computer and can operate out-of-the-box. Thus, this method has the potential to provide a reliable solution to the PGI systems for outdoor and on-street parking occupancy determination at no additional cost.

2 Background and related work

The existing PGI systems are classified into four categories (Ichihashi et al., 2009; Bong et al., 2008), based on the detection methods: 1) counter-based systems, 2) wired sensor-based system, 3) wireless magnetic sensor-based and 4) image or camera-based systems. Counter-based systems rely on sensors at the entrance and exit point of the parking lots. Counter-based systems can only provide information on the total number of vacant spaces rather than guiding the drivers to the exact location of the parking spaces, and such systems cannot be applied to on-street parking bays and residential parking spaces. Wired sensor-based and wireless magnetic sensor-based systems rely on ultrasonic, infrared light or wireless magnetic-based sensors installed on each parking space (Ichihashi et al., 2009). Both systems have been applied in practical commercial use especially in indoor environments like mega shopping malls. However, such methods require the installation of costly sensors ($\approx$ \$40, True (2007)) in addition to processing units and transceivers for wireless technologies (Bong et al., 2008). Sensor-based systems enjoy a high degree of reliability, but their high installation and maintenance cost limits their use for wide applications. Compared to the sensor-based systems, camera-based technologies are relatively cost efficient because both functions of general surveillance and parking lot occupancy detection can be performed simultaneously (Ichihashi et al., 2009).

In the literature, different approaches to parking occupancy detection have been proposed. Funck et al. (2004) use an algorithm to compare the reference image and input datasets to calculate the vehicle to parking space pixel area using principal component analysis. Tsai et al. (2007) train a Bayesian classifier to verify the detections of vehicles using corners, edges, and wavelet features. True (2007) adopts a combination of vehicle feature point detection and colour histogram classification. The Car-Park Occupancy Information System (COINS) (Bong et al., 2008) integrates advanced image processing techniques including seeding, boundary search, object detection and edge detection together for reliable parking occupancy detection. ParkLotD (Ichihashi et al., 2009) uses edge features for the detection of parking occupancy. Huang et al. (2013) use a Bayesian framework based on a 3D model of the parking spaces for the detection of occupancy that can operate day and night. Jermsurawong et al. (2014) use customised neural networks that are trained to determine parking occupancy based on extracted visual features from the parking spaces. del Postigo et al. (2015) detects the occupancy by combining background subtraction using a mixture of Gaussian to detect and track vehicles and for creating a transience map to detect the parking and leaving of vehicles. de Almeida et al. (2015) train SVM classifiers on multiple textural features and improve the performance of detection using ensembles of SVMs. Similar to COINS, Masmoudi et al. (2016) carry out trajectory analysis using real-time videos and temporal differencing in images to identify whether the parking space is occupied or vacant. The methods mentioned above are based on hand-crafted features (such as edges, colour, texture) and background subtraction, which makes these methods susceptible to the different weather conditions and illumination variation.

The CNNs (Lecun et al., 1998) are a machine learning algorithm that uses the local spatial information in an image and learns a hierarchy of increasingly complex features, thus automating the process of feature construction. Recently, CNN-based frameworks have achieved state-of-the-art accuracies in image classification and object detection (Krizhevsky et al., 2012). Valipour et al. (2016) demonstrate the practicality of a deep CNN (VGGNet-f) in the application of parking space vacancy identification. The network was fine-tuned to yield a binary classifier with overall accuracy better than 99%. They evaluate the transfer learning ability of the trained classifier on another dataset and reported an accuracy of approximately 95%. Amato et al. (2016) develop a decentralised solution for visual parking space occupancy detection using a deep CNN and smart cameras. The authors train and fine-tune a miniature version of AlexNet (Krizhevsky et al., 2012), mAlexNet for binary classification and report an accuracy of 90.7% for transfer learning process. Similar work has been performed by Amato et al. (2017), where the authors extends the CNRPark dataset (Amato et al., 2016) and compare the results of mAlexNet with AlexNet. The results indicate the achievable accuracy for transfer learning for AlexNet and mAlexNet are in the range of 90.52 - 95.60% and 82.88 - 95.28% respectively. Xiang et al. (2017) use a Haar-AdaBoosting cascade classifier to detect the vehicles in gas stations and validate the true positives with a deep CNN and report an accuracy of greater than 95%.

In summary, there is clear evidence in the literature that feature learning by deep CNNs outperform the conventional methods using hand-crafted features for the detection of parking occupancy in terms of accuracy, robustness and transfer learning. However, all the CNN-based systems mentioned above fine-tune the existing pre-trained networks, which is an additional training step requiring additional effort. In this work, we propose a transfer learning approach to parking space occupancy detection based on a pre-trained CNN without fine tuning. We train a binary SVM classifier using the features extracted by the pre-trained model and evaluate its performance in determining parking space occupancy.

3 Methodology

The research focuses on determining the occupancy of parking spaces from the images obtained by surveillance cameras considering the cost-efficient characteristics of camera-based systems. The present framework adopts ImageNet-VGG-f model (Chatfield et al., 2014), which is a pre-trained deep CNN trained on the ImageNet dataset (Deng et al., 2009). The architecture of the pre-trained deep CNN consists of 5 convolutional layers having 11x11, 5x5, 3x3, 3x3 and 3x3 image kernels respectively, that stride over the whole image, pixel by pixel (except the first layer where the stride is 4 pixels) to generate 3D volumes of feature maps. The width of the first convolution layer is 64, and 256 for the rest of the layers. A max-pooling layer follows the first, second and last convolution layer. The last convolution layer is followed by three fully connected layers having 4096, 4096 and 1000 neurons respectively and the final output consists a layer of a soft-max classifier. The architecture of the network is very similar to that shown in Figure 1. Figure 2 shows the simplified layout of the framework which consists of training an SVM classifier and evaluation of the classifications results.

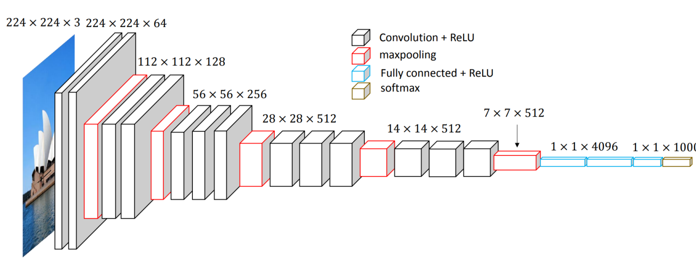

Figure 1: The architecture of a VGGNet CNN (after Wang et al., 2017)

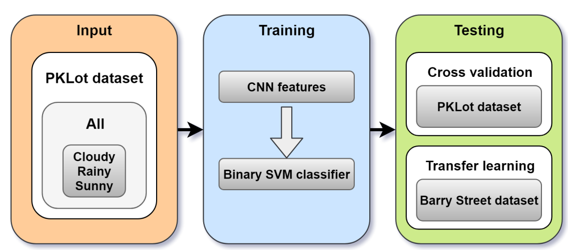

Figure 2: A simplified layout of the framework

Support Vector Machines (Cortes and Vapnik, 1995) are a machine learning technique, which transforms a non-linear separable problem into a linearly separable problem by projecting data into the feature space and then finding the optimal separate hyperplane. The separating hyperplane is a global optimum solution, and hence, the generalising ability of the SVM classifier is higher as opposed to the fully connected (FC) layers in the CNN. The FC layers can yield a local-minima during the training by back-propagation algorithm. A CNN-SVM system compensates the limits of the CNN and the SVM classifiers by incorporating the merits of both the classifier and have demonstrated best classification results for pedestrian detection (Szarvas et al., 2005) and recognizing handwritten digits (Niu and Suen, 2012). Inspired by the results of CNN-SVM systems, we use the features from a CNN and perform classification using a linear SVM classifier.

3.1 Experimental design

The experimental framework consists of two main stages: 1) training a binary SVM classifier using the features extracted by the CNN from the PKLot dataset 2) evaluation of the classification accuracy by cross validation on the PKLot dataset and the transfer learning ability on the Barry Street dataset. Donahue et al. (2014) state that the activations of the neurons in the late layers of a deep CNN serve as robust features for a variety of object recognition tasks. Hence, the features of each image are extracted from the 21^{st} layer of the CNN, that is the last layer before the classification, which consists of a vector containing 1000 elements. Consequently, the extracted features from images of the PKLot datasets were used to train and test four binary SVM classifiers using the ground truth labels v.i.z. 1) cloudy weather 2) rainy weather 3) sunny weather 4) whole dataset (0.67 million images) containing images of cloudy, rainy and sunny weather together. Subsequently, the accuracy assessment of the trained classifiers was performed by 5-fold cross-validation, to eliminate any biasing from the datasets. To evaluate the transfer learning performance of the method, the classifier that was trained using the whole PKLot dataset, was tested on segmented images (Figure 4) of Barry street dataset, which was created for the purpose of this research.

3.2 Datasets

3.2.1 PKLot

The PKLot dataset (de Almeida et al., 2015) contains 12,417 images of 3 parking sites (Figure 3), from which 695,899 segmented parking spaces (Figure 3) were generated and labelled in the data package. The image acquisition was made by a 5-minute time-lapse interval over 30 days during the daytime on three weather conditions namely rainy, sunny and cloudy days. The images are captured from various locations and orientations covering vehicles in different angles and sizes. The number of occupied and empty parking spaces account for approximately equal percentages of the whole PKLot dataset, with 48.54% and 51.46% respectively.

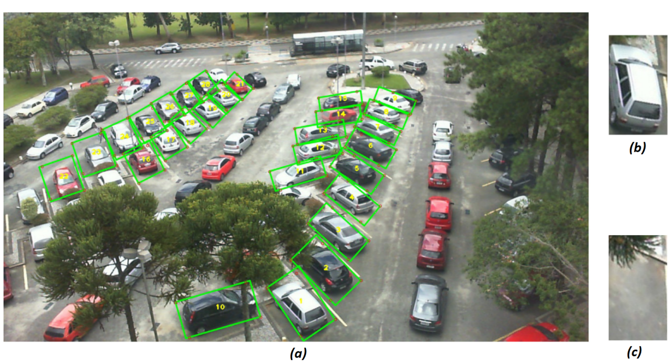

Figure 3: PKLot dataset: (a) 28 delimited spaces, (b) occupied sub-image, and (c) empty sub-image.

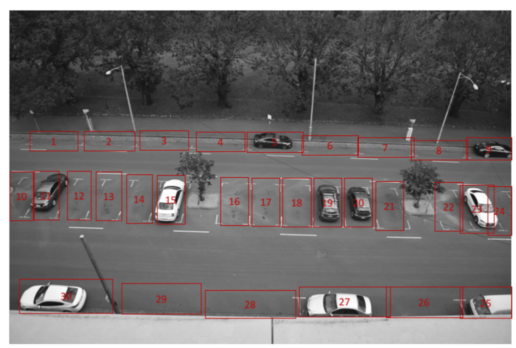

Figure 4: Barry Street dataset: segmentation of the individual 30 parking spaces

3.2.2 Barry street

This dataset was created by the authors by capturing a sequence of images from the rooftop of Faculty of Business and Economics Building, the University of Melbourne overlooking to the 30 on-street parking spaces along the Barry Street, Melbourne, VIC, Australia. The images were captured by a DSLR camera at a fixed angle from 10.18 AM to 18.15 PM with 30-second intervals on a sunny to cloudy day resulting in a total of 810 images. A total number of 24300 segmented parking space images were generated by defining the coverage of each parking space (Figure 4). For the evaluation, a ground truth label set was generated by manually labelling each image segment as either occupied or vacant.

4 Results

4.1 Evaluation criteria

For the evaluation we use three measures: overall accuracy, sensitivity, and specificity, as defined in Equations 1, 2, and 3 respectively. In the equations, TP (True Positive) is the number of occupied sub-images classified as occupied, TN (True Negative) is the number of unoccupied sub-images classified as unoccupied, FP (False Positive) is the number of unoccupied sub-images classified as occupied, and FN (False Negative) is the number of occupied sub-image classified as unoccupied.

$$Overall\ accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1) \qquad Sensitivity = \frac{TP}{TP+FN} \quad (2) \qquad Specificity = \frac{TN}{TN+FP} \quad (3)$$

4.2 Evaluation results

Figure 5 shows the overall performance of the four classifiers trained by the features extracted by the pre-trained CNN on the PKLot dataset (Cloudy, Sunny, Rainy and All) and the performance of the classifier trained by the whole PKLot dataset tested on the Barry street dataset. Note that the number of observations is on a logarithmic scale, which enables the proper visualisation of the results. The three performance measures, accuracy, sensitivity, and specificity, of the classifiers for the different weather conditions are very similar and are in the range of 99.74 - 99.93%, 99.73 - 99.91% and 99.51 - 99.94% respectively. It is worth noting that the classifier trained on the cloudy day dataset achieves the highest accuracy of 99.93%. The processing time for each image segment of the parking spaces is 0.067 seconds on a simple desktop computer (2.5GHz i5 processor). It takes approximately 2 seconds to process all the parking spaces in an image and hence is the framework is suitable for real-time applications without any dedicated hardware.

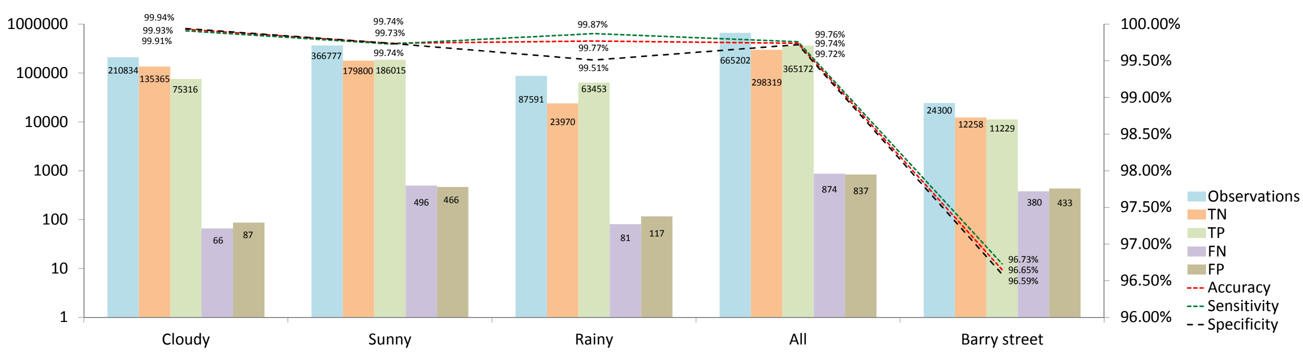

Figure 5: The classification results of the trained classifiers on different weather conditions and Barry St. dataset.

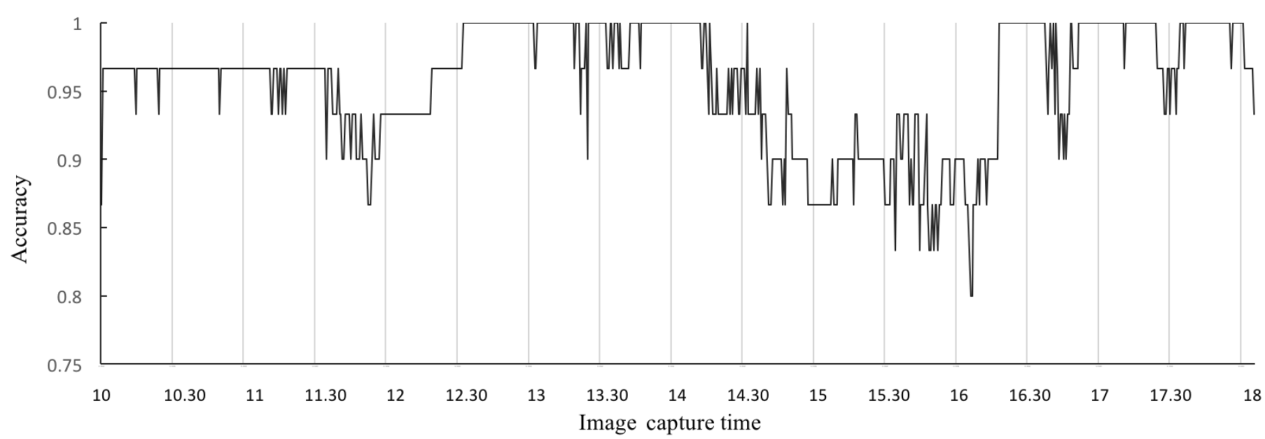

Figure 6: The classification accuracy of Barry street images by the time of the day.

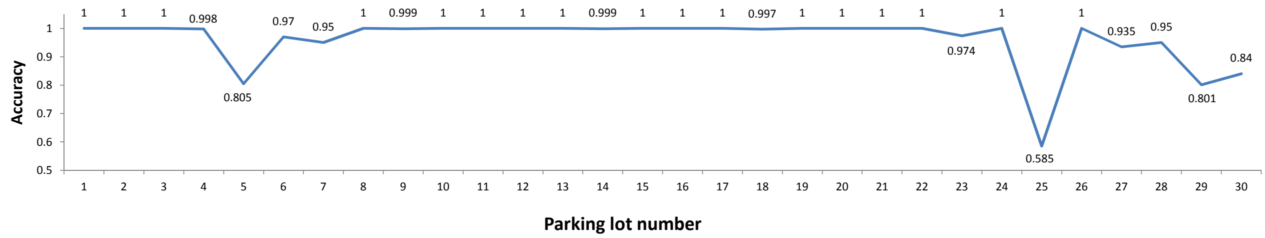

Figure 7: The variation of the accuracy over the whole period by the parking space number.

Figure 6 shows the classification accuracy of Barry street images by the time of the day, where the overall accuracy achieved is 96.65%. This visualisation enables us to analyse the variation of the accuracy with factors such as lighting condition, shadows, weather and traffic. Figure 7 shows the variation of the accuracy across different parking spaces which allows us to identify parking spaces that are classified less accurately.

The binary classification using the deep features achieved consistently reliable results with an average accuracy of 99.7% across different weather conditions for the PKLot dataset. This overall accuracy outperforms the other non-image based methods as mentioned in Section 2, and is competitive with the methods that fine-tune the pre-trained CNNs (Valipour et al., 2016; Amato et al., 2017, 2016; Xiang et al., 2017). Transfer learning is a more challenging task because the classifier is now required to recognise unfamiliar images, which eliminates the contingency that occurs regarding the feature classes, image capture perspective or angles. It is noted that there is a performance drop for transfer learning, where the average accuracy is 96.65%. However, the accuracies reported here indicate that our method outperforms the methods that fine-tune the pre-trained CNNs.

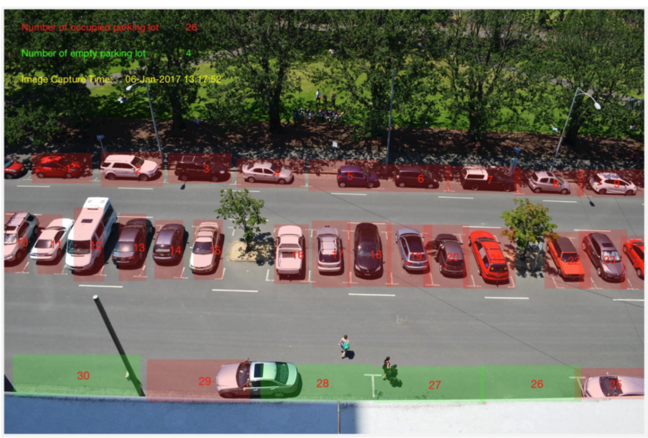

Figure 8: The classification results for a frame of Barry street dataset. The parking spaces classified as occupied and unoccupied are marked in red and green respectively. The building wall (bottom of the frame) occludes the visibility of the parking spaces 25 - 30. There is an ambiguity of parking occupancy detection for the parking spaces 28 - 29.

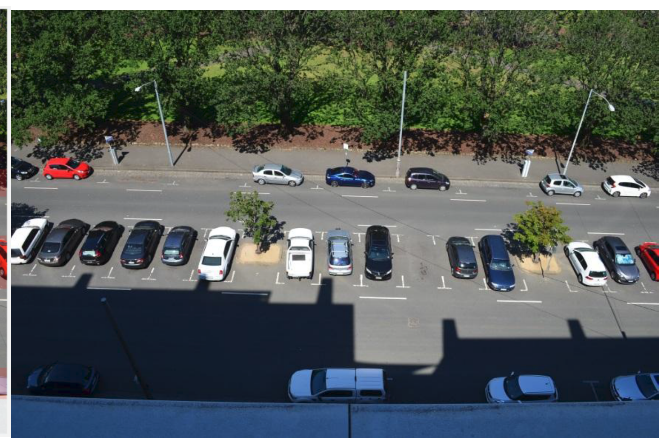

Figure 9: A frame of Barry street dataset showing the effects of building shadows on the parking spaces. Bottom left of the image shows a parked vehicle which is camouflaged by the building's shadow and the color of the vehicle.

It can be seen in Figure 6 that the classification accuracy drops in the time interval 14:30 hrs to 16:15 hrs for the Barry street dataset. After examining the images captured within this time interval two factors were identified as reasons for the lower accuracy of the classifier. Firstly, frequent changes in the occupancy status during this time span (due to office hours) creates an ambiguity in partially occupied parking spaces for the classifier but also during the creation of ground truth, which also accounts for the poor overall accuracy, as shown in Figure 8. Secondly, shadow of the building cast on the parking spaces (Figure 9) reduces the visibility and contrast of the image segments of the parking spaces. Figure 9 is an example image taken within this time interval showing the low visibility and contrast in the lower segments due to shadow.

From Figure 7, it is evident that the classification accuracy for parking spaces 5, 25, and 27 - 30 is poor, as compared to the other spaces, especially parking space 25 with an overall accuracy of only 58%. A few factors were identified accounting for the lower accuracy of the classifier. Firstly, the segmentation of the parking spaces is not clear for the parking spaces 25 - 30. Hence, vehicles hence were not parked consistently inside the whole segmentation box but across two slots instead, as shown in Figure 8. Secondly, the visibility of the vehicles in the parking spaces 25, 27 - 30 is partial due to the occlusion of the parking spaces by a building wall (Figure 8). Thirdly, the type and shape of the vehicles that were parked in spaces 25, 27 - 30 are differed from those seen in the PKLot dataset and this biases the classifier to wrongly classify the vehicles of different appearances. Fourthly, the coverage of the parking space 25 in the camera view is partial and less than 50% (Figure 8). Lastly, on a closer look at the classification results of occupancy in the parking space 5, it is observed that the accuracy drop can be attributed to strong solar reflections from the vehicle parked in that space. It is also observed that the accuracy of the parking space is high during cloudy weathers, where there are no reflections.

5 Potential for commercialisation

The beauty of the transfer learning is that, a framework like this can be implemented in any on-street and residential parking space without any training and can start working right from the minute of the installation. The achievable accuracy suggests the great potential of this framework for commercial use. However, for a practical PGI system, several aspects of the proposed framework can be improved. Firstly, the model was not trained or tested in low-light conditions such as night time, which may limit its accountability and make it less persuasive for future commercial use. Secondly, in practice, it should be able to detect the pre-defined areas of the parking spaces automatically rather than manually identifying the boundaries. The parking spaces can be easily be detected by integrating a framework that can detect the parking spaces automatically. Thirdly, the framework should be tested on images from real-time surveillance to examine the applicability of live camera feed for the framework. Fourthly, while training the classifier, images of vehicle types of diverse geographical regions should be used to remove any bias created due to repetitive vehicle types of a specific geographical region. Fifthly, the ambiguity caused by partial occupancy of the parking spaces can be improved by a dynamic segmentation method. Sixthly, the effect of shadow and strong solar reflection on the classification results can be reduced by radiometric pre-processing of individual image patches before extracting the features using the CNN. Lastly, the framework can be accelerated to achieve real-time performance with a low-end cheap Graphics Processing Unit (GPU) for an increased number of parking spaces.

6 Conclusion

An image-based framework is developed in this paper for identifying parking space occupancy in outdoor environments using features extracted by a pre-trained deep CNN and their subsequent classification by an SVM classifier. The framework achieved a high accuracy of 99.7% on the training dataset, and a transfer learning accuracy of 96.6% on an independent test dataset, which indicates its suitability for mass applications in all weather conditions. The framework can potentially provide a cheap and reliable solution to the PGI systems in outdoor environments. However, there are a few challenges limiting the performance in transfer learning including the shadows of the buildings on the parking spaces, strong solar reflection from the vehicles, vehicles parked outside or in between the designated bays by the drivers and the bias of the training data used. The performance evaluation of the framework for parking occupancy detection in the night time remains a topic of future research.

Acknowledgements

This research was supported by a Research Engagement Grant from the Melbourne School of Engineering and the Melbourne Research Scholarship.

References

Acharya, D., Khoshelham, K., and Winter, S. (2017). Real-time detection and tracking of pedestrians in cctv images using a deep convolutional neural network. In *Proc. of the 4th Annual Conference of Research@Locate*, volume 1913, pages 31–36.

Amato, G., Carrara, F., Falchi, F., Gennaro, C., Meghini, C., and Vairo, C. (2017). Deep learning for decentralized parking lot occupancy detection. *Expert Systems with Applications*, 72(Supplement C):327 – 334.

Amato, G., Carrara, F., Falchi, F., Gennaro, C., and Vairo, C. (2016). Car parking occupancy detection using smart camera networks and deep learning. In *2016 IEEE Symposium on Computers and Communication (ISCC)*, pages 1212–1217.

Arnott, R. and Inci, E. (2006). An integrated model of downtown parking and traffic congestion. *Journal of Urban Economics*, 60(3):418 – 442.

Bong, D., Ting, K., and Lai, K. (2008). Integrated approach in the design of car park occupancy information system (coins). *IAENG International Journal of Computer Science*, 35(1):7–14.

Chatfield, K., Simonyan, K., Vedaldi, A., and Zisserman, A. (2014). Return of the devil in the details: Delving deep into convolutional nets. *Computing Research Repository*, abs/1405.3531.

Chen, M. and Chang, T. (2011). A parking guidance and information system based on wireless sensor network. In *2011 IEEE International Conference on Information and Automation*, pages 601–605.

Chen, Y., Yang, X., Zhong, B., Pan, S., Chen, D., and Zhang, H. (2016). Cnntracker: Online discriminative object tracking via deep convolutional neural network. *Applied Soft Computing*, 38:1088 – 1098.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.

de Almeida, P. R., Oliveira, L. S., Britto, A. S., Silva, E. J., and Koerich, A. L. (2015). Pklot a robust dataset for parking lot classification. *Expert Systems with Applications*, 42(11):4937 – 4949.

del Postigo, C. G., Torres, J., and Menndez, J. M. (2015). Vacant parking area estimation through background subtraction and transience map analysis. *IET Intelligent Transport Systems*, 9:835–841(6).

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255.

Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., and Darrell, T. (2014). Decaf: A deep convolutional activation feature for generic visual recognition. In *International Conference on Machine Learning*, pages 647–655.

Funck, S., Mohler, N., and Oertel, W. (2004). Determining car-park occupancy from single images. In *IEEE Intelligent Vehicles Symposium, 2004*, pages 325–328.

Girshick, R., Donahue, J., Darrell, T., and Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587.

Hong, S., You, T., Kwak, S., and Han, B. (2015). Online tracking by learning discriminative saliency map with convolutional neural network. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pages 597–606.

Huang, C. C., Tai, Y. S., and Wang, S. J. (2013). Vacant parking space detection based on plane-based bayesian hierarchical framework. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(9):1598–1610.

Ichihashi, H., Notsu, A., Honda, K., Katada, T., and Fujiyoshi, M. (2009). Vacant parking space detector for outdoor parking lot by using surveillance camera and fcm classifier. In *2009 IEEE International Conference on Fuzzy Systems*, pages 127–134.

Jermsurawong, J., Ahsan, U., Haidar, A., Dong, H., and Mavridis, N. (2014). One-day long statistical analysis of parking demand by using single-camera vacancy detection. *Journal of Transportation Systems Engineering and Information Technology*, 14(2):33 – 44.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 1097–1105.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

Masmoudi, I., Wali, A., Jamoussi, A., and Alimi, M. A. (2016). Trajectory analysis for parking lot vacancy detection system. *IET Intelligent Transport Systems*, 10(7):461–468.

Niu, X.-X. and Suen, C. Y. (2012). A novel hybrid cnnsvm classifier for recognizing handwritten digits. *Pattern Recognition*, 45(4):1318 – 1325.

Szarvas, M., Yoshizawa, A., Yamamoto, M., and Ogata, J. (2005). Pedestrian detection with convolutional neural networks. In *IEEE Proceedings. Intelligent Vehicles Symposium, 2005.*, pages 224–229.

True, N. (2007). Vacant parking space detection in static images. *University of California, San Diego*, 17.

Tsai, L. W., Hsieh, J. W., and Fan, K. C. (2007). Vehicle detection using normalized color and edge map. *IEEE Transactions on Image Processing*, 16(3):850–864.

Valipour, S., Siam, M., Stroulia, E., and Jagersand, M. (2016). Parking-stall vacancy indicator system, based on deep convolutional neural networks. In *2016 IEEE 3rd World Forum on Internet of Things (WF-IoT)*, pages 655–660.

Wang, N., Li, S., Gupta, A., and Yeung, D. (2015). Transferring rich feature hierarchies for robust visual tracking. *Computing Research Repository*, abs/1501.04587.

Wang, S., Wang, Y., Tang, J., Shu, K., Ranganath, S., and Liu, H. (2017). What your images reveal: Exploiting visual contents for point-of-interest recommendation. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, pages 391–400, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Xiang, X., Lv, N., Zhai, M., and Saddik, A. E. (2017). Real-time parking occupancy detection for gas stations based on haar-adaboosting and cnn. *IEEE Sensors Journal*, 17(19):6360–6367.

Yilmaz, A., Javed, O., and Shah, M. (2006). Object tracking: A survey. *ACM Computing Surveys*, 38(4):1–45.