A Brief Review of Convolutional Neural Networks Based Solutions for Smart Parking Systems

Praveen Meduri
Electrical and Electronics Engineering
California State University, Sacramento
Sacramento, USA
praveen.meduri@csus.edu

Andres Estebanez
Department of Computer Engineering
California State University, Sacramento
Sacramento, USA
andresxestebanez@csus.edu

CSCI-ISPC SHORT PAPER

***Abstract*—Parking issues are becoming more significant as the number of drivers in crowded urban centers increases. At the same time, new breakthroughs in the fields of internet-of-things, computer vision and machine learning have opened up new opportunities for maximizing parking efficiency. In this paper, we present a comparative study of the existing approaches that use convolutional neural networks based smart parking solutions. We will explore how well various experiments using convolutional neural networks have dealt with the complex problems surrounding parking space identification, such as occlusions, different lighting conditions, and difference in car orientations.**

Keywords—smart parking, convolutional neural networks (CNN), object recognition, computer vision

I. INTRODUCTION

Our world is becoming increasingly urban with around 54% of world's population currently living in cities. By 2050, this figure is projected to compound up to 66% [1]. To ensure that quality of life is not degraded for its denizens, a future city should be well equipped with numerous smart solutions to the multitude of problems that arise from scaling. The particular problem of parking management is a persistent one. Researchers at UCLA reported that an average of 30% of drivers struggle to find parking during busy times of the day [2].

Convolutional Neural Networks have demonstrated excellent capabilities in solving high-level tasks in Computer Vision applications like object detection, classification, segmentation, etc. In this paper, we will review some of the state-of-the-art convolutional neural network (CNN) based approaches for parking space occupation detection. Our goals in this paper are: (i) to discern which CNN techniques for parking space detection are the most accurate and effective, (ii) to provide a comparative analysis of the existing approaches, (iii) to identify important variables that affect the accuracy of the detection algorithms.

Convolutional Neural Networks have become increasingly common for image recognition problems since AlexNet won the ImageNet image classification competition in 2012 [4]. AlexNet was trained on 1.2 million images to classify into 1,000 categories, and achieved the lowest error rate in the competition with only 15.2%, with the second best being 26.2% [5]. This was no accident, as CNNs possess a variety of characteristics which make them well-suited for image recognition problems.

- They are very robust. CNNs deal with problems such as occlusions, different weather conditions, and operating at different times of day in different levels of sun light very well, with little loss of accuracy [3].
- The classification portion is not computation heavy [4] and can be implemented on a Raspberry Pi. Amato et al. [4] accomplished a classification speed of 15 seconds per iteration using a Raspberry Pi model 2. Another effective way of implementing a CNN-based classifier is the Field-Programmable Gate Array (FPGA), which have been shown to implement CNN-based classifiers very well [6].
- The architecture of CNNs is very conducive to image recognition problems as the receptive fields and spatial subsampling take advantage of the spatial locality of image-based problems [7]. Local correlations in the data, meaning ones that have a close proximity in space, will form features. The CNN takes advantage of this in a way that a fully connected neural network cannot, as the numerous connections are not as efficient for image-based problems.
- Improvements in the arithmetic units of computers have made convolutional neural network training times to become tractable[8]. Specifically, improvements made in GPU technology have brought about an age where the training of extremely large and deep CNNs on extremely high resolution images is now possible.

The rest of the paper is organized as follows. Section 2 will discuss the general structure of a CNN and present a few architectures. Section 3 will discuss the smart parking solutions presented in literature using CNNs. Section 4 will be discussions and section 5 presents conclusions.

II. CONVOLUTION NEURAL NETWORKS (CNNS)

The use of CNNs for parking space detection problem has yet to be explored in-depth. The earliest investigations of the problem were conducted in 2016, by two research papers [4][9]. This is therefore a very young problem. In [3], A Haar-Cascade based approach for Smart Parking System is presented.

Two very popular architectures that were used in the majority of research on the subject are the aforementioned AlexNet, and LeNet, which first developed in 1998 to recognize characters from handwriting [8]. Although other networks have

This work is partly funded through CSUS Research and Creative Activity grant.

978-1-7281-1360-9/18/$31.00 ©2018 IEEE
DOI 10.1109/CSCI46756.2018.00092

been developed, often-times implementing more convolutional layers, these two remain popular and laid the groundwork for CNN architectures that would come later [7].

LeNet was one of the earliest of the popular CNNs; to be implemented in 1998. LeNet, or more specifically LeNet-5, has an architecture of 7 total layers, not counting the input layers. They are organized as follows: there is first a convolution layer, a pooling layer, a second convolution layer, a second pooling layer, a third convolution layer, a fully connected layer, and an output layer. These convolutional and down-sampling layers are followed by a fully-connected layer, and then a final output layer. The activation function for LeNet is the sigmoid, or 1/(1+e^-x) [8]. Features of LeNet, such as (i) alternating convolutional layers with filters to discern certain features from the data, and (ii) down-sampling layers which reduce the size of the matrix in which the data is stored while keeping relevant information, are common in later CNNs such as AlexNet. More recent CNN architectures like VGGNet and GoogleNet inherit many aspects of these older architectures, while innovating in some aspects [7].

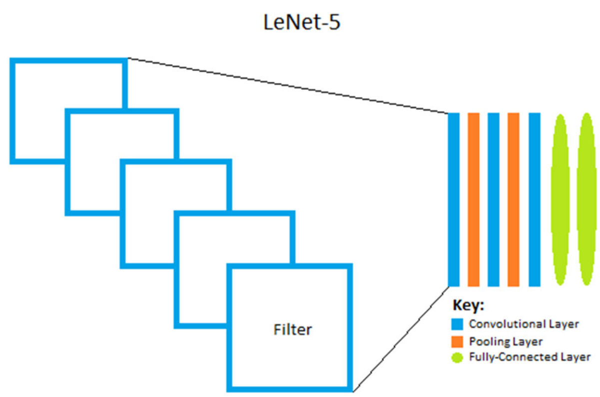

Figure. 1: The architecture of LeNet-5 [8].

AlexNet is another very popular CNN architecture that is used in smart parking systems. The most salient difference between AlexNet and LeNet is that AlexNet is significantly deeper than LeNet, owing to the increased computational power available almost 15 years later. AlexNet contains 5 convolutional layers followed by down-sampling layers in the form of max-pooling. There are then 3 fully-connected layers.

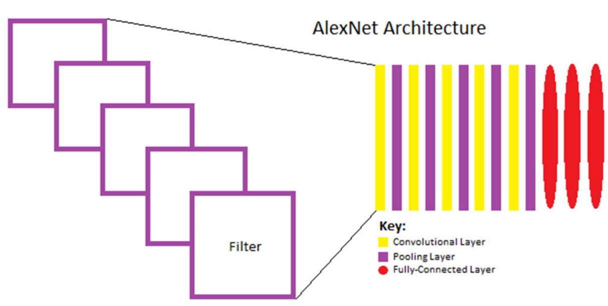

Figure. 2: The architecture of AlexNet [5].

Another important attribute of AlexNet is its use of the ReLU activation function, which significantly lowers the computation needed for gradient descent relative to other activation functions such as tanh, and therefore training time in general [5].

III. Existing CNN Based Solutions to Smart Parking

One of the earliest attempts at using CNNs for smart parking systems was by Amato et al. 2016. They used two CNN architectures dubbed mAlexNet and mLeNet, where the “m” stands for mini. This is an apt title, as they both are diminished in size. This is done in order to improve computation time on the embedded device on which they hope to deploy the classifier, and reportedly did not diminish accuracy. Amato et al. [4] here show that edge computing for smart-parking is feasible with a CNN-based classifier. mAlexNet has three convolutional layers each followed by max-pooling instead of five, with one fully connected layer featuring soft-max, which normalizes the outputs to a probabilistic value of one or less, where all outputs sum to one. mLeNet features two convolutional layers, followed by max pooling, and two fully connected layers with soft-max. ReLU was the selected activation function in both CNNs[4].

The dataset they used was a combination of two different datasets. One is PKLot [10], which contains over 12,000 distinct images of parked cars from different locations, with differing sun-light and different orientations. PKLot has three sets of images from two different parking lots. Each set comes from one stationary camera in each parking lot, with one of the parking lots having two such cameras. The second dataset is one they created themselves called CNRPark, which consists of over 6,000 images of cars, but with fewer differences in lighting and more occlusions. Their method consisted of training their CNNs on CNRPark datasets and validating them on the PKLot set. They then tested them on different subsets of the CNRPark dataset [4].

The results for the experiment were favorable for the CNN. mAlexNet performed “significantly better” than mLenet. On intra-dataset experiments, accuracy in the high 90% region was achieved, and on testing on different parking lots even mid-80% accuracy was achieved. The results of this test are shown in table-1. This demonstrated the effectiveness of using the CNN for the smart-parking problem.

In the work of Cazamias et al. [9], a similar approach is taken as in Amato et al. [4], where the model is trained on the PKLot [10] data. However, a slightly different CNN architecture was used with 3 convolutional layers followed by max-pooling, ReLU activation, and three fully connected layers with softmax to compute the result. Their results were excellent when training and testing on all parking lots in PKLot, with an accuracy of 99.97%. However, when they attempted to generalize the results from training on one of the parking lots in PKLot and tested on the other, this accuracy fell to about 80%. This would suggest that the angle of cars has a significant effect on the effectiveness of the classifier [9].

Nyambal et al. [11], used AlexNet and LeNet. The AlexNet is implemented with Stochastic Gradient Descent, which dramatically lowers training time by recalculating the gradient with respect to only one training instance rather than every training instance, and the LeNet uses Nesterov's Gradient Descent. They are both then trained on a dataset of their own, consisting 782 images, with the locations of parking spaces once again inputted manually. AlexNet once again achieved greater accuracy with a .9549 accuracy to LeNet's .9364 [11]. However, their experiment was conducted in situations of low occlusion and minimal variations in car orientations.

Xiang et al. [7] took a slightly different approach, and used a Haar-Cascade to identify regions with a car, and then used a CNN to scan over them in order to suppress the false-positive rate. The Haar-Cascade with Adaboost was first proposed by Viola and Jones in their very important work [15]. Other papers have explored using a Haar-Cascade with Adaboost with significant success in [3] and [16].This method was chosen due to the speed with which this could be achieved, as their problem required a fast response-time. The CNN would only be employed on those areas where the Haar-Cascade had identified a car. If the CNN and Haar-Cascade both identified a car in this area, it would be considered a positive; otherwise it would be considered negative. They used VGGNet CNN architecture, which uses 13-convolutional layers and 3 fully connected layers. However, despite the increased depth, the number of weights is lower and therefore it does not have more connections than many shallower nets. Other than this, the CNN also uses ReLU for the activation function and max-pooling, as with all the other networks used. Using their two-part classifier, they achieved an accuracy of 95% with high precision and recall values of 99.4% and 96.1% respectively. Their experiment also produced these results at a speed necessary to implement the system in real-time [7].

Valipour et al. 2016 [12] implemented a CNN using VGGNet-9, which is one of the smallest VGGNet's. It consists of 5 convolutional layers each followed by 3 fully connected layers. Using the PKLot [10] they achieved very high 99.97% accuracy on that dataset. When they used this model on a different parking lot—one on which it was not trained—that accuracy dropped to 95%[12]. Figure-6 shows an ROC curve for their results on the different parking lots.

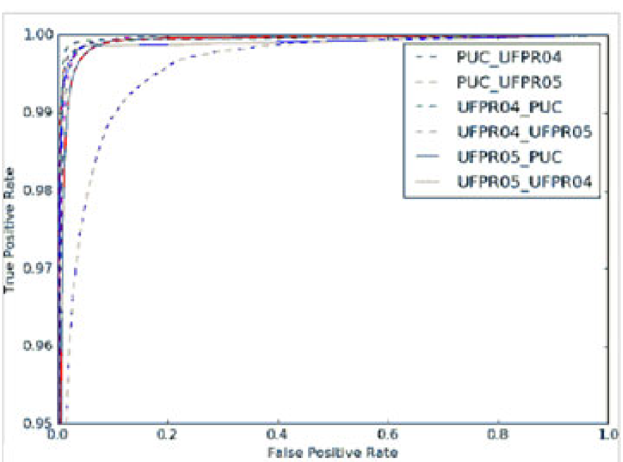

Figure 6 ROC curve for the PKLot set (figure from [12]).

Vu et al. 2017 sought to improve upon already existing techniques and create a system that was immune to differences in car sizes and vehicle displacement within the parking spaces. This was done by incorporating a Spatial Transform Network (STN) in order to account for differences in sizes and displacement. The STN itself is a module which can be put into a CNN which will warp the input images using certain spatial transforms [14]. This should help the classifier to identify different car orientations as well as sizes. They also grouped parking spaces into sets of three and had their classifier identify all three units together. The classifier is thus divided in stages with an STN, the CNN which takes the three spaces as input, and a final logistic regression layer. The CNN architecture consists of an input layer, 3 convolution layers with ReLU, max-pooling, 2 convolution layers with ReLU, max-pooling, 2 convolution layers with ReLU, max-pooling, 2 fully-connected layers with ReLU, and a final fully-connected layer with a sigmoid activation function and 8 outputs. The 8 outputs represent the 8 states that the three parking spaces can be in. Sigmoid is used to output a probabilistic value to each of the 8 states. The final logistic regression layer then assesses the occupancy of the middle space, which is the most difficult to identify. They then assessed their accuracy using combinations of only a basic CNN implementation, the three-parking spot-input CNN, and the STN. All methods achieved over 96% accuracy, but the proposed method which brought all of the elements achieved 99.23% [13]. Below is a receiver operating characteristic (ROC) curve for their data in figure – 7.

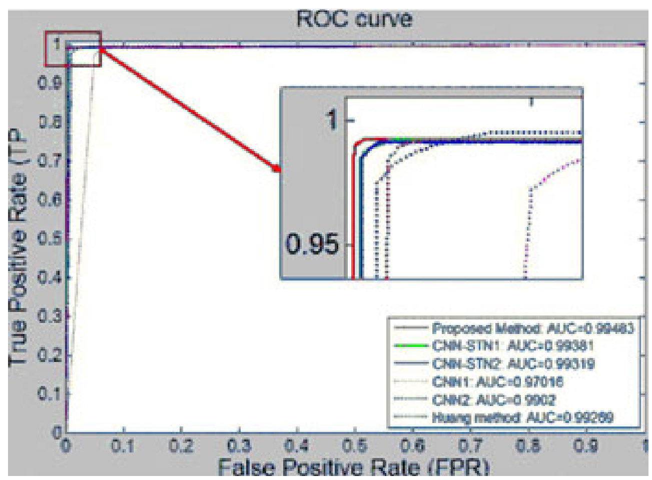

Figure - 7 ROC curve for the different CNN-based implementations (figure from [13]).

This experiment is of particular interest because it was performed on a data-set that incorporated high degree of variability in terms of noise, occlusions, car sizes, and orientations and still achieved very high accuracy.

IV. Discussion

It is clear that although many experiments have been carried out with great success under certain controlled circumstances, the generalizability of these smart-parking solutions is still questionable. The ability of these classifiers to perform well hinges on the similarity of the conditions between the tested parking lots and those upon which the classifiers were trained.

At best, Valipour et al. 2018's classifier still could not identify 5% of the cars when applied to different parking lots. Other experiments suffered even greater performance degradation.

Table. 1: Comparison of CNN-based smart parking results

Experiment:	Best Accuracy:	
	Intra-dataset	Inter-dataset
Cazamias et al. 2016	99.97%	80.70%
Valipour et al. 2018	99.97%	95.00%
Vu et al. 2017	99.23%	N/A
Xiang et al. 2017	95.00%	N/A
Nyambal et al. 2017	95.49%	N/A
Amato et al. 2016	99.60%	90.70%

The differing conditions of the different parking lots are a large obstacle towards creating a general smart-parking system Variables such as noise in the form of weather conditions, lighting, occlusions, different car sizes, orientation variations, and displacement within parking spaces are all significant factors.

Another issue was also observed with regard to the generalizability of smart-parking solutions. We have noted that none of the surveyed research papers had schemes to detect the positions of the parking spaces themselves; all of them had the location of these spaces hard-coded into the algorithm. In order to deploy a smart-parking system on a large scale, this presents a significant obstacle, as identifying the topologies of these parking lots and entering their coordinates is a laborious and time-consuming task.

V. Conclusion

CNNs are very effective for image recognition problems and can achieve high accuracy in many of the experiments reviewed. We presented a comparative study of the existing approaches that use convolutional neural networks based smart parking solutions. Main variables that adversely affect the performance of each of the approaches are identified. Issues in the generalizability of the smart-parking solutions are discussed.

References

[1] "World Urbanization Prospects", United Nations Organization, 2014. [Online]. Available: http://www.un.org/en/development/desa/news/population/world-urbanization-prospects-2014.html. [Accessed 09 30 2018].

[2] D. C. Shoup, "*The high cost of free parking*", Chicago: Planners Press, American Planning Association, 2005.

[3] *P. Meduri and E. Telles, "A Haar-Cascade classifier based Smart Parking System", 2018 International Conference IP, Computer Vision, and Pattern Recognitio, pp. 66-70.*

[4] *Giuseppe Amato, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, and Claudio Vairo, "Car parking occupancy detection using smart camera networks and Deep Learning," 2016 IEEE Symposium on Computers and Communication (ISCC), June 2016.*

[5] *A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks", Advances in neural information processing systems, pp. 1097-1105, 2012.*

[6] *K. Abdelouahab, M. Pelcat, F. Berry, and J. Sérot, "Accelerating CNN inference on FPGAs: A Survey," HAL Id: hal-01695375. [Online].*

Available: https://hal.archives-ouvertes.fr/hal-01695375/file/hal-accelerating-cnn.pdf. [Accessed: 30-Sep-2018].

[7] *X. Xiang, N. Lv, M. Zhai, and A. El Saddik, "Real-Time Parking Occupancy Detection for Gas Stations Based on Haar-AdaBoosting and CNN," IEEE Sensor Journal, vol. 17, no. 19, pp. 6360–6367, Oct. 2017.*

[8] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient Based Learning Applied to Document Recognition," http://yann.lecun.com/. [Online]. [Accessed: 01-Oct-2018].

http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf.

[9] J. Cazamias and M. Marek, "Parking Space Classification and Convolutional Neural Networks," stanford.edu. [Online]. Available: http://cs231n.stanford.edu/reports/2016/pdfs/280_Report.pdf.

[10] Almeida, P., Oliveira, L. S., Silva Jr, E., Britto Jr, A., Koerich, A., "PKLot – A robust dataset for parking lot classification", Expert Systems with Applications, 42(11):4937-4949, 2015.

[11] J. Nyambal and R. Klein, "Automated parking space detection using convolutional neural networks ," 2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech) , Nov. 2017.

[12] S. Valipour, M. Siam, E. Stroulia, and M. Jagersand, "Parking Stall Vacancy Indicator System Based on Deep Convolutional Neural Networks," arxiv.org. [Online].

Available: https://arxiv.org/pdf/1606.09367.pdf. [Accessed: 01-Oct-2018].

[13] H. T. Vu and C.-C. Huang, "A multi-task convolutional neural network with spatial transform for parking space detection," 2017 IEEE International Conference on Image Processing (ICIP) .

[14] K. Danilyuk, "ConvNets Series. Spatial Transformer Networks – Towards Data Science," Towards Data Science, 19-Sep-2017. [Online].

[15] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 2001, pp. I-511-I-518 vol.1.

[16] R. Fusek, K. Mozdˇreˇn, M. Surkala, and E. Sojka, "AdaBoost for Parking Lot Occupation Detection" https://pdfs.semanticscholar.org/2963/3712a36c3efc77ce3a9844a2e9a029daf310.pdf