

**INSTITUTE OF MANAGEMENT SCIENCES (Pak-Aims)**

**Title of Synopsis:**

Real Time Car Parking Occupancy Prediction System

**Name of Student:**

Hassan Arif

**Registration Number:**

193880

**Title of Degree:**

MPHIL-CS

**Name of Supervisor:**

Dr. Tauqir Ahmad

**SYNOPSIS**

**Title**

Real Time Car Parking Occupancy Prediction System

**Introduction**

In urban areas, detecting parking occupancy is a big concern. Because the number of cars on the road is constantly increasing around the world, this problem is projected to intensify in the future years. Related scientific works reveal that this subject has drew scientific attention for a long time, but no adequate answer has been found. Different ways with hardware sensors on each parking space have been presented, but we believe that only Computer Vision (CV) would address the problem.

CV has gained a lot of traction in recent years, and algorithms have become much more exact. Simultaneously, data models are smaller, allowing detection to occur on Edge AI-capable devices.

The latest Edge AI Camera devices are built to match the performance of pricey server-side components. AI and GPU modules with several trillion operations per second (TOPS) are now included in Edge AI Cameras, as well as support for image sensors with up to 200 Megapixels. We suggest employing these methods to develop a platform that will capture 3D models of the seen location using the latest Edge AI camera devices paired with updated CV algorithm performances.

In Pakistan there are lot issues regarding car parking due to increasing in number of cars. There are no proper parking systems available in Pakistan to manage car parking especially if we talk about street parking, most of the people used to park their cars on a wrong place that creates the problem for other people. In our research proposal, we are presenting the idea to overcome this issue.

**Background**

1. **Parking Occupancy Detection or Prediction**

The parking occupancy detection problem determines whether or not a certain parking space is currently occupied. This can be used in conjunction with both visual and non-visual solutions. Visual solutions make advantage of surveillance camera footage, which is frequent in parking lots. To determine whether a parking place is occupied, non-visual methods use sensors other than cameras, such as magnetic field sensors. Visual sensors are used in conjunction with magnetic sensors in some ways. This is owing to magnetic sensors' increased energy usage when compared to visual sensors. They use sight sensors to turn on the magnetic sensor, which detects whether or not a parking lot is occupied. Another sensor combination strategy is to employ magnetic and distance sensors, which compensates for certain vehicles' modest magnetic footprint. This method has a 99 percent accuracy rate. However, the necessity for a specialized sensor for each parking space is a problem, resulting in higher installation and maintenance costs. Furthermore, these sensors can alter conditions in parking lots such as water and oil.

The use of security cameras already installed in parking spaces is another typical method of detecting parking occupancy. As opposed to non-visual approaches, this eliminates the need for extra sensors. However, using computer vision technologies to generate reliable findings in a variety of weather and lighting circumstances, as well as vehicle types, is difficult. A parked vehicle may also block the view of neighboring parked vehicles and visual objects.

Many present systems treat the position of parking lots in each image as a constant feature. They manually partition the image into parking spaces and classify them as occupied or not. The condition of three adjoining parking places is classified as a unit using an upgraded color vectors feature-based Support Vector Machine (SVM) approach, eliminating the influence of occlusion. A hierarchical Bayesian generator concept is also introduced, which is more robust than earlier SVM-based solutions.

In the PKLot dataset, a deep learning solution based on Convolutional Neural Network (CNN) classifiers was also developed and shown an accuracy of above 90%. Later on, an end-to-end solution that included a camera system and a front-end application was offered.

For the PKLot dataset, this solution had an AUC over 0.99. M-RCNN (Mask Region-based Convolutional Neural Networks) are also used to determine the availability of parking spaces by matching vehicle-like items to designated parking lots. All of these systems, however, relied on manually segmenting the image into parking spots. Furthermore, there is a scarcity of study on the effectiveness of object detection as a method for detecting vacant parking places without the need for prior manual input.

1. **Detection of Object**

Detecting occurrences of semantic items in a given image is defined as object detection. Classic object detectors, two-stage detectors, and single-stage detectors are the three primary types of solutions. The sliding window method is used in traditional object detectors, where a classifier is applied to a region specified by a sliding window. A CNN or classic machine learning methods such as SVM could be used as this classifier.

For two-stage object detection, several approaches have been utilized. The first phase produces a small number of candidate proposals that should include all of the items. The second stage divides the proposals into two categories: foreground and background. R-CNN is a method that uses CNN for both stages, followed by a number of enhancements. The most recent technology is faster R-CNN. In standardized object detection tests like Common Objects in Context (COCO), state-of-the-art two-stage approaches surpass all other methods. The Pascal VOC data format, on the other hand, provides a standardized picture dataset for object detection, with a distinct structure for the boundary box. When compared to single-stage detectors, the fundamental disadvantage of two-stage detectors is speed. In terms of inference and training, they are slower than single-stage detectors. This time is usually spent on the region proposal stage, which is CPU intensive in many circumstances. R-CNN is a faster version of R-CNN. The introduction of the Region Proposal Network addressed this issue (RPN). Quicker RCNN may now be performed entirely on a GPU or a deep neural network accelerator, making it faster and more accurate than the rest of the R-CNN family.

In summary, the two-state technique filters the backdrop to forecast object locations before classifying the objects. As a result, R-CNN supports a high level of accuracy. However, real-time object detection is computationally expensive. Single-stage algorithms, on the other hand, which are the most recent state-of-the-art object detection techniques based on deep learning, allow for quick processing. As a result, it is appropriate for real-time object detection. However, when compared to two-stage approaches, the accuracy is low. Thus, by rescaling the loss function to enhance accuracy, the single state-based RetinaNet may be utilized to repair the focus loss in object detection.

In several existing research, the object detection element is located in the image's foreground. The unoccupied parking spaces instances, on the other hand, are part of the backdrop in the parking occupancy detection scenario. For the inference service, a single-stage detector, RetinaNet, and a two-stage detector, Faster R-CNN, were chosen. In pictures, these detectors have been utilized for similar tasks such as detecting road markings and detecting road damage. They were the state-of-the-art single and two-stage detectors with the greatest average precision scores in the COCO dataset, which is commonly considered as the benchmark for object detection, at the time of this work.

**Objectives of the Study**

* Car detection using 3D bounding boxes.
* Total availability of parking lots in left and right lane.
* Indication of exact available parking lot.
* Improvement in the accuracy of detection.

**Research Questions**

* How much accuracy needed for exact object detection?
* How many cameras could be used for detection of car?
* How cameras could detect the parking lot if the car is parked on wrong lot?
* Are the devices enough capable to detect the car and parking lot in lower illumination?
* Are the devices enough capable to detect exact available parking lot?

**Research Strategy**

In our research, our major goal is to detect exact parking lot to provide easiness for the driver. In research strategy, we will divide this research in many phases. In the first phase dataset will be collected to train the model for car and parking lot detection. In the second phase model will be selected either deep learning or may be computer vision to achieve better accuracy. In the third phase collected dataset will be applied on selected model and train the model. Data analysis will be performed to calculate the accuracy achieved and compare the accuracy from previously work done.

1. **Review of literature**

To classify the parking spot, the suggested solution in Paper [1] uses Convolutional Neural Networks (CNNs) onboard a Raspberry Pi camera. Each camera may keep track of up to 50 parking places. For parking space detection, previously generated filters were utilized to extract parking spaces, and a neural network was employed to classify them as free or occupied. AlexNet (mAlexNet) and LeNet-5 were utilized as convolutional neural networks, both of which were modified versions of existing neural networks (mLeNet).

Two data sets were employed in the experiment: PKLot [2] and CNRPark, the latter of which was created by the authors of this paper. CNRPark is smaller than PKLot, but it features more difficult examples, such as parking places blocked by trees and lampposts. However, because mAlexNet outperformed other state-of-the-art models based on non-deep machine learning in every situation, it was put to the test against them. In this scenario, mAlexNet outperforms the competition, especially when testing on the CNRPark data set. This demonstrates that CNNs are noise-resistant, as they maintain excellent accuracy even in images with varying light conditions, shadows, and partial occlusions. The data and trained models, as well as the code to replicate the entire experiment, are all freely available.

The authors of paper [3] used the PKLot and CNRPark data sets to test a novel form of convolutional neural network. CarNet is a Dilated Convolutional Neural Network that was put to the test. With CNNs, dilation is a novel approach that entails creating gaps in the convolution matrix. As a result, dilation rate k = 1 denotes a conventional convolution, but k = 2 denotes a convolutional matrix with a one-pixel gap, k = 3 denotes a two-pixel gap, and so on. Figure 1 depicts the dilation rate and convolutional matrix. Dilated convolution was found to be a means to enhance the global view of the network exponentially while keeping the parameters linearly increasing.

Three convolutional layers with max-pooling layers are followed by three fully connected layers in CarNet. The matrix size for each of the convolutional layers is 11 x 11 and the dilation rate is two. There are 96, 192, and 384 filters, respectively. The first and second completely connected layers each have 4096 units; however the output layer has only two units. Because the problem at hand is a binary classification problem, this is the case. To avoid over fitting, dropout regularizations are applied in the fully connected layers.

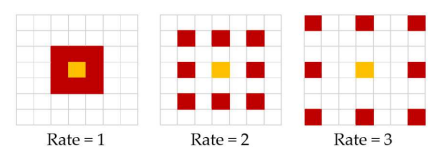


Figure 1. Examples of dilated convolution for different rates

Additional studies were conducted using CarNet but without dilation to demonstrate the influence of dilation. CarNet with dilatation demonstrated higher accuracy during training and validation in all scenarios. The authors used dilated convolution to train the neural network how to learn huge characteristics while disregarding small ones.

Furthermore, different CarNet versions were put to the test. The number of convolutional layers was adjusted, which caused a variation in complexity. The effectiveness of architectures with two to four convolutional layers was investigated, and the architecture with three layers was found to be the most effective, indicating that fewer layers are better for this task.

mAlexNet [1] was used to compare testing on the PKLot and CNRPark data sets. mAlexNet performs poorly while training on PKLot and testing on CNRPark, and vice versa, while CarNet is more trustworthy. When it comes to training and testing on the same data set, mAlexNet and CarNet produce mixed results, outperforming each other in different circumstances with mean accuracy of 96.74 percent and 97.04 percent, respectively.

The paper [4] presents a multi-camera system based on simultaneous processing of each camera and combining the findings. The end result is a multi-camera spot matrix that depicts the parking lot's occupancy. Each camera provides frames to the system's vehicle detector, which employs object detection to create a bounding box for each vehicle identified. Two algorithms were employed in the evaluation of the suggested system. The main technique was the Faster Region-Based Convolutional Neural Network (R-CNN), and the second was the Deformable Parts Model (DPM). To produce bounding boxes for automatic spot mapping, homographic transformation and perspective correction must be performed once they have been obtained. The occupancy spot matrices for each camera are the result of the spot mapping. In the information fusion process, all of these matrices are integrated to produce a parking place occupation matrix. Because there were no parking lot data sets available, the authors created their own, which is now available for research.

Both detectors were evaluated in scenarios with and without a mask in terms of detection outcomes. The results of the masked detector are superior to those of the detector without a mask because it removes potential detections in sections of the parking lot that are not being watched at the time. When utilizing the area under the curve metric to compare the findings to ground realities, the results demonstrate that perspective adjustment is required and results in a significant improvement.

The authors of [5] proposed a technique for detecting parking occupancy using a real-time video feed. Camera nodes, IoT devices, cellular data transmission modules, and a centralized server make up the system. The authors were the first to use the MIOvision Traffic Camera Dataset (MIO-TCD) [6] for training purposes. MobileNet Single-Shot Detection (SSD) and Background-Based Detection were employed as vehicle detection methods.

Combining these two algorithms yielded greater results, with precision reaching 99 percent at night. The overall detection accuracy of the suggested method was 95.6 percent. This includes indoor and outdoor environments, as well as wet, sunny, foggy, noon, and midnight scenarios. This study shown that it is possible to achieve good results by specializing pre-trained models on night photos of cars. It also shown that using Edge AI, a reliable parking surveillance system can be developed.

Using Haar-AdaBoosting and convolutional neural networks, Xiang et al. [7] developed a unique technique for real-time parking occupancy detection. The Haar-AdaBoosting cascade classifier extracts potential vehicle regions from the frame, which are then sent to CNN, which filters out non-vehicle regions and suppresses false positives. Gentle AdaBoost is utilized in this paper because it has a lower computational complexity than other AdaBoost algorithms.

A camera is mounted on the gas station to monitor the fuelling of parking spaces in the experiment. The proposed system's accuracy rate is greater than 94 percent. The findings of the experiment reveal that if the illumination around parking spaces is stable at night, the detection accuracy is not affected.

Detecting some particular types of vehicles, such as tankers and trailers, is one of the areas where the model fails. This can be addressed by increasing the amount of photos of specific categories used in training. Similarly, if a parking space is occupied illegally for less than 30 seconds, the system will not detect it.

The approach developed by Stojanovi et al. [8] revolutionizes the understanding of on-street parking occupancy and the prediction of available parking spaces. The moment has come to build a platform that will leverage Location Awareness based on 3D representations of parking surroundings, they used Edge AI capable camera devices and Computer Vision algorithms. The proposed approach makes use of computational photography, which takes advantage of the camera's capacity to adapt to changing ambient conditions while still capturing excellent photos.

1. **Research Methodology**

In this research, we are going to work on Car Detection and Parking Occupancy Prediction. For that purpose Computer Vision might be used to achieve more accuracy. Deep learning approaches may also be a better approach for the object detection in real time. We split the work in many phases. In the first phase object will be detected either it is car or other type of vehicle. In our research, we are restricting the cars only in parking lot. Object detection might be performed using high accuracy cameras. To detect the car’s dimension, we may generate 3D bounding boxes to calculate the dimension of the cars more accurately.

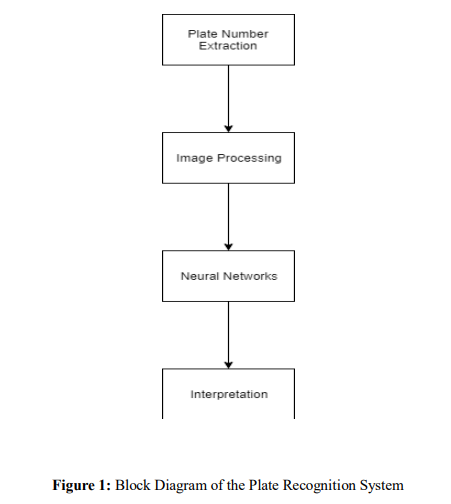
After detection of object, it will allow the object which is car in our case to enter in the parking area. In the second phase, it will specify the driver about the total available parking space in left and right lots. For example, 50 out of 100 spaces available in left lane and 60 out of 100 spaces available in right lane.

In the third phase, which is our main research area, it will specify the driver about exact available parking lots in parking area.

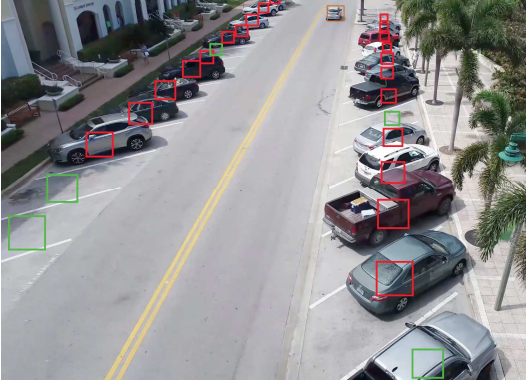
**Detection and Prediction Using Computer Vision**

To log plates at the parking area's entry, an existing image recognition system was employed. The system extracts the license plates of any vehicle entering the parking lot using photos captured by a camera at the entrance. The system's block diagram was shown in Figure 1.

The plate number is first extracted by snapping a photo with a camera. After that, the image is digitally pre-processed to remove skewness and segment characters from each image unit. The plate is encoded and interpreted after being sent into an Artificial Neural Network (ANN) for training, validation, and testing. Following that, the image recognition system sends the license plate information to a central computer, which assigns the vehicle a parking slot. The above image recognition system is used by a camera installed at the dock of each parking spot to verify that the car parked is the vehicle allotted to the spot. If the verification is successful, the central computer system is alerted, and the vehicle's parking time is logged. This might be used to implement variable pricing systems based on space consumption. If the verification fails, the system sounds a short-term alarm, and if the car does not leave the parking area in a short amount of time, the license plate is tagged in the central system. This could be used to inflict managerial punishments. The time it takes a car to find a parking spot under the system will be measured and compared to the time it takes a vehicle to find a parking spot without the system to see how effective the parking management system is compared to traditional techniques.



If we compare 2D and 3D approaches. 2D approach produce rectangle around a specific vehicle as their output. But if we talk about 3D approach, it creates 3D bounding boxes around observed cars. Because solutions that rely mainly on 2D images for recognition do not employ knowledge about 3D space, they have a difficult time comprehending and extracting information about the seen area.

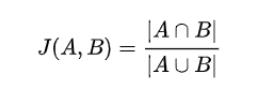
Parking slots detected by 2D rectangle model 3D model draws a bounding box around the detected vehicle

We may use data from the GPS to understand 3D objects and locations, such as the camera's exact location (latitude and longitude) and height. The lens characteristics of the camera will have to be taken into consideration during camera calibration and perspective correction. After a camera has been calibrated, the ratio of pixels and observed objects can be used for additional calculations. Compass data might be used to determine the camera's direction, while gyroscope and accelerometer data might be used to determine the camera's angle (tilt). It will also be able to better recognize the sort of detected vehicle, allowing it to more accurately calculate its size and position in 3D space.

**Detection and Prediction Using Deep Learning**

Mask R-CNN is the best model for building a multi-angle car parking detection. A deep neural network called the Mask R-CNN creates a segmented instance of an object.

The Mask R-CNN model accomplishes this by combining two major CNNs: the Faster R-CNN and the FCN (Fully Convolutional Network). The Faster R-CNN is in charge of object detection, which entails identifying objects of various kinds. Similarly, the FCN does instance segmentation, which is the process of assigning a class/object name to each pixel in an image. As a result, the Mask R-CNN combines these two algorithms to produce a segmented instance of an item by generating a classification of the identified regions of interest, as well as a bounding box around the detected object and the likelihood of its prediction. The model then builds a mask with float values for all selected locations of interest. The final objective is to determine if a parking space is occupied or vacant, given that the cars have been detected and the positions of the parking slots have been determined. We employ the measure known as IoU, which stands for Intersection over Union, to help the model execute this task. The IoU, commonly known as the Jaccard Index, is calculated by dividing the number of pixels covered by both items by the number of pixels where two objects overlap. In our situation, the bounding box of an unoccupied parking space may be partially occupied by a car from a nearby parking space due to the camera's angle; thus, the IoU measure prevents the wrong labeling of such a parking spot. It accomplishes so by calculating the typical overlap area between the bounding boxes of a parking place and the bounding box of a car and normalizing it with the total number of pixels filled by both bounding boxes to produce a value between 0 and 1, which ranges from 0 to 1.



  Low-angle parking frame with all parking spaces occupied Top angle parking frame with IoU<=0.2

**Dataset Samples**

We collected 'Cars Overhead with Context' (COWC) data set and Microsoft's COCO - Common Objects in Context dataset for the second category for training the multi-angle car parking detection model. This dataset was appropriate for training the low angle Mask R-CNN model.



Sample image from COCO dataset



Sample image from COWC dataset

For testing the model, in order to simulate real-life parking situations, we used the PKLot dataset, which contains more than half a million images captured with three different camera views over two parking lots

****

Sample image from PKLot dataset

**Tools and Technologies**

We may use Python and its packages of Keras and TensorFlow for the implementation of the deep learning or Computer Vision model. During all experiments, we may use GeForce RTX 2070 with 16GB RAM or more as our graphical accelerated processing (GPU), for the acceleration of the convolution operations and reduce the training time of deep learning models.

1. **Bibliography**
2. Amato, G., Carrara, F., Falchi, F., Gennaro, C., & Vairo, C. (2016, June). Car parking occupancy detection using smart camera networks and deep learning. In 2016 IEEE Symposium on Computers and Communication (ISCC) (pp. 1212-1217). IEEE..
3. De Almeida, P. R., Oliveira, L. S., Britto Jr, A. S., Silva Jr, E. J., & Koerich, A. L. (2015). PKLot–A robust dataset for parking lot classification. Expert Systems with Applications, 42(11), 4937-4949.
4. Nurullayev, S., & Lee, S. W. (2019). Generalized parking occupancy analysis based on dilated convolutional neural network. Sensors, 19(2), 277.
5. Nieto, R. M., García-Martín, Á., Hauptmann, A. G., & Martínez, J. M. (2018). Automatic vacant parking places management system using multicamera vehicle detection. IEEE Transactions on Intelligent Transportation Systems, 20(3), 1069-1080.
6. Ke, R., Zhuang, Y., Pu, Z., & Wang, Y. (2020). A smart, efficient, and reliable parking surveillance system with edge artificial intelligence on IoT devices. IEEE Transactions on Intelligent Transportation Systems.
7. Luo, Z. et al. (2018). MIO-TCD: A new benchmark dataset for vehicle classification and localization. IEEE Transactions on Image Processing, 27(10), 5129-5141.
8. Xiang, X., Lv, N., Zhai, M., & El Saddik, A. (2017). Real-time parking occupancy detection for gas stations based on Haar-AdaBoosting and CNN. IEEE Sensors Journal, 17(19), 6360-6367.
9. Stojanović, Nikola, Vladan Damjanović, and Srđan Vukmirović. "Parking Occupancy Prediction using Computer Vision with Location Awareness." In 2021 20th International Symposium INFOTEH-JAHORINA (INFOTEH), pp. 1-5. IEEE, 2021.
10. Padmasiri, Heshan, Ranika Madurawe, Chamath Abeysinghe, and Dulani Meedeniya. "Automated Vehicle Parking Occupancy Detection in Real-Time." In 2020 Moratuwa Engineering Research Conference (MERCon), pp. 1-6. IEEE, 2020.
11. Africa, Aaron Don M., Antonio Miguel Sarmiento Alejo, Grant Lewis Milan Bulaong, Samantha Maxine Ronquillo Santos, and Jerrick Spencer Kehyeng Uy. "Computer vision on a parking management and vehicle inventory system." International Journal of Emerging Trends in Engineering Research 8, no. 2 (2020): 323.
12. Agrawal, Tushar, and Siddhaling Urolagin. "Multi-angle parking detection system using mask r-cnn." In Proceedings of the 2020 2nd International Conference on Big Data Engineering and Technology, pp. 76-80. 2020.