MODIFICATION OF ALEXNET ARCHITECTURE FOR DETECTION OF CAR PARKING AVAILABILITY IN VIDEO CCTV

Evan Tanuwijaya, Chastine Fatichah

Department of Informatics Engineering, Institut Teknologi Sepuluh Nopember Surabaya
Keputih, Sukolilo, Surabaya 60111, East Java, Indonesia

E-mail: evantanuwijaya.18051@mhs.its.ac.id, chastine@if.its.ac.id

Abstract

The difficulty of finding a parking space in public places, especially during peak hours is a problem experienced by drivers. To assist the driver in finding parking space availability, a system is needed to monitor parking availability. One study to detect the availability of parking lots utilizing CCTV. However, research on the availability of parking spaces on CCTV data has several problems, detecting parking slots that are done manually to be inefficient when applied to different parking lots. Also, research to detect the availability of parking lots using the Convolution Neural Network (CNN) method with existing architecture has many parameters. Therefore, this study proposes a system to detect the availability of car parking lots using You Only Look Once (YOLO) V3 for marking the parking space and proposed a new architecture CNN called Lite AlexNet which has few parameters than other methods to speed up the process of detecting parking space availability. The best accuracy of the marking stage using YOLO V3 is 92.31% where the weather was cloudy. For the proposed Lite AlexNet get the best time training average which is 7 second compare to other existing methods and the average accuracy in every condition is 92.33% better than other methods.
Keywords: *Detection of car parking availability, CNN, YOLO V3, Lite AlexNet*

Abstrak

Sulitnya menemukan tempat parkir di tempat-tempat umum, terutama saat jam sibuk adalah masalah yang dialami oleh pengemudi. Untuk membantu pengemudi dalam menemukan ketersediaan tempat parkir, diperlukan sistem untuk memantau ketersediaan parkir. Satu studi untuk mendeteksi ketersediaan tempat parkir memanfaatkan CCTV. Namun, penelitian tentang ketersediaan ruang parkir pada data CCTV memiliki beberapa masalah, mendeteksi slot parkir yang dilakukan secara manual menjadi tidak efisien ketika diterapkan pada tempat parkir yang berbeda. Selain itu, penelitian untuk mendeteksi ketersediaan tempat parkir menggunakan metode Convolution Neural Network (CNN) dengan arsitektur yang ada memiliki banyak parameter. Oleh karena itu, penelitian ini mengusulkan sistem untuk mendeteksi ketersediaan tempat parkir menggunakan You Only Look Once (YOLO) V3 untuk menandai tempat parkir yang ada dan mengusulkan CNN arsitektur baru yang disebut Lite AlexNet yang memiliki beberapa parameter daripada metode lain untuk mempercepat proses mendeteksi ketersediaan ruang parkir. Akurasi terbaik dari tahap penandaan menggunakan YOLO V3 adalah 92,31% di mana cuaca mendung. Untuk Lite AlexNet yang diusulkan dapatkan rata-rata waktu pelatihan terbaik yaitu 7 detik dibandingkan dengan metode lain yang ada dan akurasi rata-rata di setiap kondisi adalah 92,33% lebih baik daripada metode lain.
Kata Kunci: *Deteksi ketersediaan parkir mobil, CNN, YOLO V3, Lite AlexNet*

1. Introduction

Looking for a parking space in a mall or office is difficult especially during busy hours. For mall, the busy hour is at weekends around 6 - 8 pm and companies are on weekdays around 8 am – 5 pm. This problem costs time, money, and fuel as well as pollution and traffic congestion [1]. To solve this problem, research about the detection of the availability of parking space was conducted. When there is a parking space available, drivers will fight over it [2]–[4]. The use of ultrasonic sensors for detecting the availability of parking space has been widely used by both companies and malls. However, the sensor has a weakness where one sensor can only detect one parking space so the number of sensors must be equal to the total number of parking spaces in the parking lot and this will cost a lot. Therefore, Sensors can be replaced by CCTV placed in several parts of the parking lot. The advantage of CCTV is that CCTV can capture several parking spaces while sensors can only capture one parking space. By utilizing CCTV, researchers can apply object recognition using

deep learning to detect cars in several parking spaces captured by CCTV.

There have been many studies conducted related to the detection of the availability of car parking spaces using the image. A research conducted by [5] was proposed to detect the availability of parking space using the modified AlexNet by reducing the convolution layer and named it mini AlexNet that has higher accuracy in detecting parking space availability. However, the research did not explain how to mark the position of parking space. Another research was carried out by [6] who proposed a method for detecting parking space using moving drones. This research utilizes features such as vehicle colors, gradients, and Harris Corners of the captured image to detect the availability of parking spaces and then uses Support Vector Machine (SVM) to do the classification. Other research was conducted by [7], proposing a binary classifier CNN that can determine parking availability. This research conducts pre-processing on binary images and then classified using CNN. Based on research that has been done, many ways can be done to detect the availability of parking spaces. However, research on the availability of parking spaces using CCTV has several problems, that is to detect parking spaces that are done manually is inefficient when applied to different parking lots. Research to detect the availability of parking space using the Convolution Neural Network (CNN) with existing architecture is also inefficient because CNN has many parameters and layers.

Therefore, this research proposes a system to automatically detect the availability of car parking spaces using YOLO V3 [8] and Lite AlexNet which is a modification of AlexNet architecture so it can detect objects quickly and accurately on image or video from CCTV located in parking lots. By utilizing the bounding box from YOLO V3 to detect objects, this system will take the property of the bounding box to detect all available parking locations by taking the bounding box from the cars that have been parked. All bounding box properties obtained from YOLO V3 are then used as markers to detect the availability of cars in the parking lot using the modified AlexNet by reducing the number of parameters and architecture and name it Lite AlexNet. AlexNet is a CNN with an architectural model that has the smallest number of layers and parameters compared to other architectural models such as VGG. AlexNet can classify 1000 classes of objects contained in images such as clothes, rulers, sandals, and others. The detection of the availability of parking space only has two classes, that is, available or occupied. Therefore, researchers use AlexNet architecture and modify its architecture. So that the architecture that was originally able to classify 1000 classes is now a binary classification that can classify only two classes by training the model from scratch [5]. By doing this research, this method is expected to optimize the detection of the availability of parking space quickly so that the driver can find a parking space optimally.

For this research, the researcher uses only one CCTV and focused on building and testing the model for detecting the availability of parking space

This research paper is composed as follows; Section 2 presents the related works of this research. Section 3 discusses data and methods in this research. Section 4 shows the result of the experiment and the analysis of it. The conclusion is made in Section 5.

2. Related works

Research on the detection of the availability of parking spaces has been carried out, ranging from using sensors to cameras. This sub-chapter will explain some of the previous studies regarding the detection of empty parking with image or video sources

In research conducted by [7], proposed a binary classifier for CNN which can determine whether the parking space is empty or not. First the researchers used data from the camera. The data obtained is an RGB image which is then converted to HSV to take the value of V, which is the value
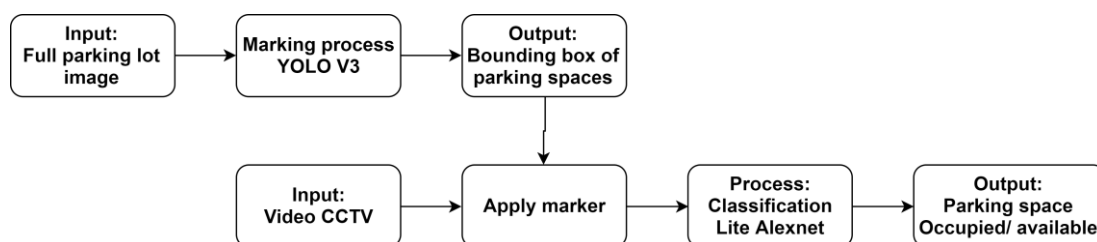

Figure.1. This is the system architecture of the detection the availability of the parking space.

of the color. After that, the logarithm transformation is performed to transform the grey value to be wider and performed the Second derivative algorithm to sharpen the margins in the image. The results of this process are converted into binary images and put into CNN. The results of CNN can classify with 100% accuracy. However, this research didn't explain the CNN architecture and it also didn't explain how the system detects parking space is empty or not.

The next research about car detection is from [6]. In this study, a method is proposed to detect car parking through cameras on drones. First, the parking lot is detected by giving four points at each parking space. Then the results of the image are processed using lens correction to transform the lines and created a straight line and then the line's coordinates were stored. To classify a car, data from each parking space that is available or not was used as training for the SVM classifier with the Histogram of Oriented Gradients feature with eight orientations, Density of corners in the slot, and the color of the vehicle. To determine whether there is parking, a line model is used where the parked car will form a straight line then if there is an empty parking space, the line will be broken. The broken line will be marked as an empty parking space. The accuracy value obtained from this study was 97.6%. This research has an effective method for drones only so that if it is applied to the camera it will still have deficiencies that cannot form a line to detect the availability of parking space.

Other studies were taken from [5]. This research proposes a CNN method that uses a smaller architecture and parameter than AlexNet called a mini AlexNet to detect empty parking spaces. By using a mini AlexNet and dataset that has been made by the researcher, this research succeeded in detecting parking lots with an accuracy of above 90%. However, this study does not explain how to take the marking of each parking space so there are still some questions whether the mini AlexNet can find the parking spot automatically from the whole picture or whether the researchers conducted a marking process to find out the parking space location.

3. The proposed system for Detection of Car Parking Availability

This section will discuss the system for detecting the availability of parking space. To detect car parking spaces, researchers developed a system. We based this research from [2], [5], [9], and make improvements by detecting car space automatically and the smaller architecture of AlexNet. In this study, we have 2 main stages. The first stage is the marking stage where the system

will be given an image of a parking lot full of parked cars to be marked. The next stage is the classification stage. In this stage, the parking space that had been marked from the marking process will be classified as available or not with input from image or video from CCTV. Therefore, this section will discuss the system that will be created. The whole system is illustrated in Figure 1. As described previously. The system of this research has two main stages namely marking and classification which will be described in the following sub-section.

3.1 Dataset Description

In this study, the first two data needed are the image of a parking lot that has been fully parked by a car for the marking stage and the second is a video recording from CCTV to detect the availability or occupied of parking spaces. We are using images and videos from CCTV located outdoor. This research will use training data obtained from[5], namely CNR Park data. This data set is a collection of images of outdoor parking lots that have been labeled.

Figure.2. These images are an example of the dataset to train the Lite AlexNet model obtained from CNR Park.

Figure.3. These images are an example of a parking lot image to test the proposed system obtained from CNR Park.

In this data, there are various kinds of weather conditions ranging from sunny and rain with a size of 150 × 150 each according to Figure 2. For testing data will be used data from CNR Park which has a 1000 × 750 size image that captures the entire parking lot as in Figure 3.

3.2 Marking stage

The marking stage is a stage to find out all positions of the existing parking spaces so that this stage does not need to be repeated. The process that the researcher build is described in Figure 4. Parking lots that have been filled with parked cars will be photographed and then the photo will be

used as input for the marking stage. Photos that have been entered will be processed by using the pre-trained You Only Look Once (YOLO) V3 [8], [10]–[12] method to detect all the cars in the photo. We choose one image from the CCTV. The image must be a parking lot filled with a fully parked car. Using this image, we're going to detect every car in the image and generate a bounding box of each car. With the assumption that the position of cars is the same with the parking space, we're going to save the property of bounding box which contains x position and y position as well as width and height of the bounding box. The result of the marking process is in Figure 5.
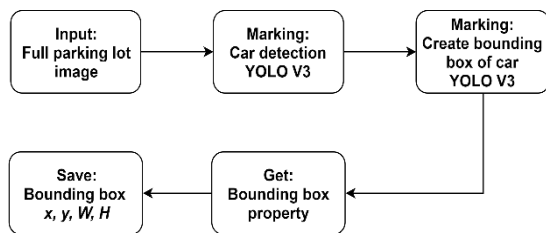

Figure 4. The marking stage process begins with a full parking lot image and using YOLO V3 to detect all the cars in the image and create a bounding box on each car.

Figure 5. Result of the marking stage using YOLO V3 and create a bounding box on each car.

3.3 Classification stage

The classification stage is a stage of determining the parking space that has been marked using markers from the marking stage. The classification will determine whether the parking is available or not. The process of the classification stage is illustrated in Figure 6. The video from the CCTV camera will be used as input. The input video will be processed by marking the parking location based on the bounding box obtained at the marking stage. After the bounding box is applied to CCTV video, each bounding box must detect whether there is a car in the parking space by processing each bounding box with Lite AlexNet. The use of Lite AlexNet in the classification stages is to process video quickly and accurately.

The architecture of the Lite AlexNet was built from the AlexNet model in Table 1 to Lite AlexNet in Table 2. The researcher used AlexNet because architecture AlexNet was able to make a good prediction surpass the state-of-the-art and AlexNet architecture has a small layer and parameter compared to the state-of-the-art now. Lite AlexNet will detect the car in each bounding box and determine whether the parking lot is available or not. The architecture of Lite AlexNet is based on the AlexNet architecture itself with the number of layers and parameters reduced. This reduction aims to speed up the computing time of the classification process. Another reason we proposed Lite AlexNet is the number of classes that want to predict is only 2 classes while AlexNet can predict 1000 classes which make AlexNet inefficient. The number of layers in AlexNet is 8 layers consisting of 5 convolution layers and 3 fully connected layers while in Lite AlexNet, the number of layers becomes 4, which is 3 convolution layers and 1 layer fully connected. On Lite AlexNet, the size of the kernel in the convolution layer used the same kernel size on AlexNet, which is $11 \times 11, 5 \times 5$, and 3×3. On the fully connected layer, Lite AlexNet uses only one layer and for the output layer it will produce one value that represents two classes that is available or occupied. The researcher only modified the architecture and train it from scratch without using a pre-trained model.
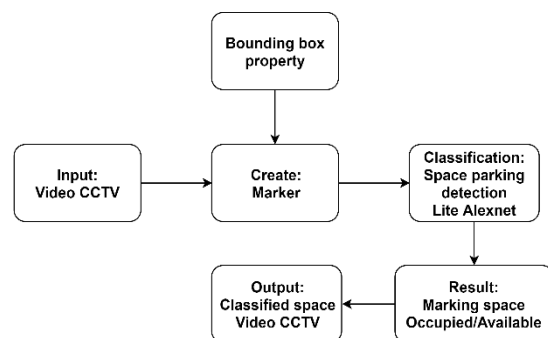

Figure 6. The classification stage process begins with a video from CCTV and the create marker using bounding box location and use Lite AlexNet to classify each bounding box.

The architecture of Lite AlexNet is the result of architectural modification and the number of parameters reduced from AlexNet [4], [5], [13]–[20]. Lite AlexNet accepts input images with a size of 151×151 and has an RGB color so before entering the image in a bounding box, the image will be resized to adjust the size of the input layer. Each convolution layer has a different filter size and has its stride. The activation function used is the Rectified Linear Unit (ReLU). ReLU replaces the negative value to 0 with equation (1).

$$(1)$$

$$f(k) = \max(0, k)$$

Each convolution layer ends with a maxpooling layer according to AlexNet where maxpooling will draw the highest value from the specified filter size. Each convolution layer is going to visualize object features in the image such as the angle, edge, and texture of the target [21]. Modifications made will reduce the number of parameters [21] to make it more efficient to detect the availability of parking spaces. The reduction in the number of parameters is intended so that the classification process can be done quickly with an accuracy value that is not much different from AlexNet.

TABEL 1
ARCHITECTURE ALEXNET

Name	Filter size	Stride	Output size
Input			$227 \times 227 \times 3$
Convolution 1	11×11	4	$55 \times 55 \times 96$
Maxpool 1	3×3	2	$27 \times 27 \times 96$
Convolution 2	5×5	1	$27 \times 27 \times 256$
Maxpool 2	3×3	2	$13 \times 13 \times 256$
Convolution 3	3×3	1	$13 \times 13 \times 384$
Convolution 4	3×3	1	$13 \times 13 \times 384$
Convolution 5	3×3	1	$13 \times 13 \times 256$
Maxpool 3	3×3	2	$6 \times 6 \times 256$
FC 1			2048
FC 2			1024
Output			1000

TABLE 2
ARCHITECTURE LITE ALEXNET

Name	Filter size	Stride	Output size
Input			$151 \times 151 \times 3$
Convolution 1	11×11	2	$71 \times 71 \times 20$
Maxpool 1	3×3	2	$35 \times 35 \times 20$
Convolution 2	5×5	3	$11 \times 11 \times 25$
Maxpool 2	3×3	1	$9 \times 9 \times 25$
Convolution 3	3×3	2	$4 \times 4 \times 30$
Maxpool 3	3×3	1	$2 \times 2 \times 30$
FC 1			30
Output			1

On the first layer, the convolution layer has a filter size of 11×11 with 20 total filters and with 2 strides. After that, we use the maxpooling layer with filter sizes 3×3 and stride 2. On the second layer will receive the value from the maxpooling layer on the first layer and then process with the layer convolution with size 5×5 with the number of kernels 25 and stride 3 and then maxpooling with filter size 3×3 with stride 1. On the third

layer, it has a convolution layer with size 3×3 with the number of filters 30 and strides 2 followed by maxpooling with size 3×3 and stride 1. After convolution is done and gets features from the input image, the next layer is a fully connected layer where the convolution layer with $2 \times 2 \times 30$ size features will be flattened into a 1D vector. The size of the fully connected layer is 30 which will then be processed by the fully connected layer with the RELU activation function and dropout of 0.4. Lite AlexNet's output layer is generating 1 value using the Sigmoid activation function in equation (2)

$$Sigmoid(x) = \frac{1}{(1 + e^{-x})} \quad (2)$$

where x is the result of the multi-layer perceptron and the Binary Cross-Entropy loss function which produces a value of 0 or 1. If there are cars in the bounding box, then in the next process, the bounding box will turn red. Conversely, if there are no cars in the bounding box, the bounding box will be green like in Figure. 7. The result of this process is the CCTV video that has a bounding box that can determine whether the spaces are empty or not.

4. Experiment and analysis

In this section, we will discuss the experiment result and analysis of this proposed system. This system was build using Python programming language and we use Keras to implement the CNN. We are using Google Colaboratory to write our Python and using GPU that was provided.

4.1 Result of the marking stage

The marking stage uses the YOLO V3 method to detect the position of the car in the parking lot. YOLO V3 method can classify 1000 classes of objects in an image. Therefore, the results of the YOLO V3 method is processed to get car objects only and then get bounding boxes from each car position. The data used to carry out this marking process is a parking lot image that has been parked fully by a car with a size of 1000×750 pixels with an RGB color of 6 images that have different weather and time conditions. Detection results using YOLO V3 can be seen in Table 3.

TABLE 3
RESULT OF YOLO V3 DETECTION MARKING PARKING SPACE

Image	Detected	Not Detected	Accuracy	Condition
1	36	3	92.31%	Cloudy
2	29	1	96.67%	Cloudy
3	24	18	57.14%	Rain
4	14	17	45.16%	Night
5	47	7	87.04%	Sunny
6	45	8	84.91%	Sunny

For the detection process using YOLO V3 obtained good results when the weather is cloudy. YOLO V3 can detect objects with a high degree of accuracy in cloudy conditions because the brightness is not bright and not dark, so the car is detected properly. when it is sunny, there is a decrease in inaccuracy. This is because when the weather is sunny, white cars become blurry due to high lighting levels. Then the researchers tried to use the image of the parking lot at night. At night, the accuracy of the YOLO V3 becomes 45.16% due to low lighting levels and only cars that have bright colors like red or white can be detected. Researchers tried to use the parking lot when it was raining, and the accuracy value was 57.14%. This accuracy value is higher compared to the state of the night because there is still a moderate level of lighting and there is a splash of rainwater to create interference in the form of salt and pepper in the image.

4.2 Result of classification stage

The classification stage uses Lite AlexNet to classify parking spaces that have been marked as available or not. This stage will produce value 1 for occupied and value 0 for available. The Lite AlexNet will be tested for under two climates in Indonesia, namely sunny and rainy, then the researcher divides the time by in the morning, afternoon and evening respectively to detect what kind of condition this model can classify cars accurately when we place it in a real parking lots in mall or office that have a different level of brightness. Each situation will use 10 frames from video CCTV. Besides, Lite AlexNet results will be compared with other architectures namely AlexNet, Mini AlexNet, and VGG16. For measuring the success of this method, we're using the False Positive Rate, False Negative Rate, and Accuracy of each method[20].

$$FPR = \frac{vacant\ spaces\ detected\ as\ occupied}{total\ vacant\ spaces} \quad (3)$$

Equation (3) is a False Positive Rate formula that compares the predicted number of available parking spaces detected as occupied compared to the number of available parking spaces. If the FPR value is getting closer to 0 then the model predicts correctly on an empty parking lot.

$$FNR = \frac{occupied\ spaces\ detected\ as\ vacant}{total\ occupied\ spaces} \quad (4)$$

Equation (4) is a False Negative Rate formula that compares the predicted number of occupied parking spaces detected as available compared to the number of occupied parking spaces. If the FNR value is getting closer to 0 then the model predicts correctly on the filled parking lot.

$$ACC = \frac{number\ of\ correct\ detection}{total\ number\ of\ tested\ spaces} \quad (5)$$

Equation (5) is calculating the accuracy of the classification method where the number of correct parking predictions is compared with all marked parking lots. For the training phase of the method, researchers used CNR park data with a total of 12

TABLE 4
RESULT OF TRAINING MODEL FOR CLASSIFICATION

Method	Accuracy (%)	FPR	FNR	Time (Second)	Number parameter
Lite AlexNet	98.00	0.98	0.97	**7**	**30,246**
Mini AlexNet	98.00	**0.99**	0.98	9	32,475
AlexNet	98.00	0.98	0.97	15	28,080,753
VGG16	**99.00**	**0.99**	**0.99**	98	134,264,641

thousand with a comparison of available parking image data of 6 thousand and parking image data that occupied as many as 6 thousand images. For each method trained, we train with epoch 50 with a validation split of 0.1. The optimizer method used is Stochastic Gradient Descent (SGD) which has a learning rate of 0.01. The loss function used is Binary Cross-Entropy. The results of the training can be seen in Table 4. It can be seen from the training phase that the proposed architecture, namely Lite AlexNet has an accuracy, FPR, and FNR value that is not far from other methods. But in terms of time for data training, Lite AlexNet is faster at 7 seconds for each epoch. This proposed model can extract important features with high accuracy that is 0.98 with the smallest train time that is 7 seconds with 30,246 parameters. This proves that to detect the availability of parking

space accurately, the CNN architecture used does not need to use architecture and parameter in large numbers to make more efficient the training time of a model.

The results of the testing phase found that the proposed method, which is Lite AlexNet, has a high degree of accuracy in some cases, namely in the morning, sunny, and morning rain. This is because during these weather conditions the lighting level in the parking lot is not too high so it can be classified. When the weather is dark, the proposed method of Lite AlexNet is inferior to AlexNet and VGG 16 which has a higher number of layers and parameters so that it can draw more features from the marked parking location. Based on the results of this trial, it is found that the Lite AlexNet proposed method can detect the position of the parking lot accurately which can be seen

TABLE 5
RESULT OF THE TESTING MODEL USING SUNNY CLIMATE VIDEO CCTV

| | Sunny | | | | | | | | |
| | Morning | | | Afternoon | | | Night | | |
Method	FPR	FNR	Accuracy (%)	FPR	FNR	Accuracy (%)	FPR	FNR	Accuracy (%)
Lite AlexNet	0.250	0.030	**96.40**	**0.175**	**0.027**	**92.34**	0.134	0.269	91.89
AlexNet	**0.022**	0.071	95.07	0.298	0.033	91.89	**0.054**	**0.027**	**94.14**
Mini AlexNet	0.289	**0.000**	95.95	0.803	0.033	86.94	0.172	0.203	86.49
VGG16	0.250	0.082	95.02	0.319	0.105	87.84	0.081	0.269	93.24

TABLE 6
RESULT OF THE TESTING MODEL USING RAINY CLIMATE VIDEO CCTV

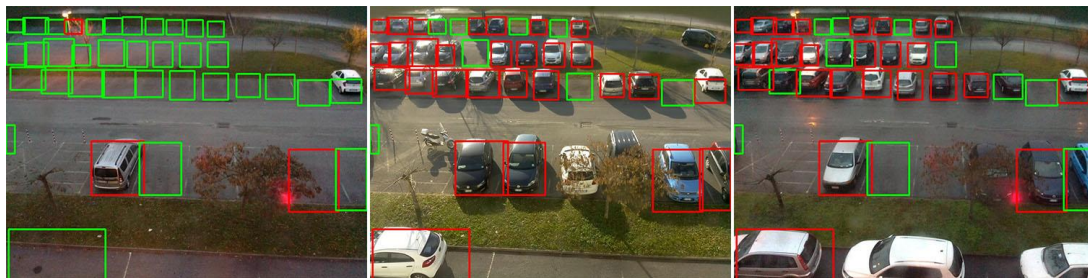| | Rainy | | | | | | | | |
| | Morning | | | Afternoon | | | Night | | |
Method	FPR	FNR	Accuracy (%)	FPR	FNR	Accuracy (%)	FPR	FNR	Accuracy (%)
Lite AlexNet	**0.000**	**0.000**	100.00	**0.000**	0.139	90.53	**0.099**	0.293	85.14
AlexNet	0.021	0.182	96.85	0.146	0.176	86.49	0.093	0.283	86.04
Mini AlexNet	0.135	0.013	90.54	0.021	**0.096**	**92.34**	0.044	0.666	83.78
VGG16	0.015	0.084	94.59	0.111	0.175	86.94	0.121	**0.157**	**87.39**

Figure 7. Result of the classification process that classifies each bounding box as available (green) or occupied (red).

from the values of accuracy, FPR, and FNR, respectively 100%, 0, and 0. When it is raining, the AlexNet Lite model can detect empty spaces well so that it gets a very low FPR value. For weather conditions during the afternoon rain, the method used for the test has decreased accuracy because the input from CCTV experiences blur due to the rain in the image causing salt & pepper noise. The result of the tested method can be seen in Table 5 and Table 6.

5. Conclusion

In this research, we proposed a system to detect the availability of parking space using YOLO V3 for marking the parking space and proposed a new architecture called Lite AlexNet to classify the parking space is available or occupied. From the result of the marking stage, we can detect the position of car space using the position of the car by entering an image from the parking lot that has parked full by car. The best result we get is 96.67% from the image with the condition of cloudy which has intermediate lighting level. For the classification process, the proposed Lite AlexNet can classify the marked parking space. The best result from this proposed method is from condition sunny morning, sunny afternoon, and rainy morning. This proposed method has successfully classified with the intermediate level of brightness using smaller architecture and parameters of CNN. In future work, an improvement of detection at night and when it heavily rains condition is needed so it can be applied to many kinds of the parking lot.

References

[1]   M. Fraifer and M. Fernstrom, "Smart car parking system prototype utilizing CCTV nodes: A proof of concept prototype of a novel approach towards IoT-concept based smart parking," in *2016 IEEE 3rd World Forum on Internet of Things*, 2016, pp. 649–654, doi: 10.1109/WF-IoT.2016. 7845458.

[2]   A. Geitgey, "Snagging Parking Spaces with Mask R-CNN and Python," *medium.com*,2019.https://medium.com/@ ageitgey/snagging-parking-spaces-with-mask-r-cnn-and-python-955f2231c400 (accessed Apr. 22, 2019).

[3]   G. Amato, F. Carrara, F. Falchi, C. Gennaro, and C. Vairo, "Car parking occupancy detection using smart camera networks and Deep Learning," in *IEEE Symposium on Computers and Communications*, 2016, pp. 1212–1217,

doi: 10.1109/ISCC.2016.7543901.

[4]   S. Valipour, M. Siam, E. Stroulia, and M. Jagersand, "Parking-stall vacancy indicator system, based on deep convolutional neural networks," in *2016 IEEE 3rd World Forum on Internet of Things,* 2016, pp. 655–660, doi: 10.1109/ WF-IoT.2016.7845408.

[5]   G. Amato, F. Carrara, F. Falchi, C. Gennaro, C. Meghini, and C. Vairo, "Deep learning for decentralized parking lot occupancy detection," *Expert Systems with Applications*, vol. 72, pp. 327–334, 2017, doi: 10.1016/j.eswa.2016.10.055.

[6]   C. F. Peng, J. W. Hsieh, S. W. Leu, and C. H. Chuang, "Drone-based vacant parking space detection," in *32nd IEEE International Conference on Advanced Information Networking and Applications Workshops,* 2018, pp. 618–622, doi: 10.1109/WAINA.2018.00155.

[7]   T. Thomas and T. Bhatt, "Smart Car Parking System Using Convolutional Neural Network," in *Proceedings of the International Conference on Inventive Research in Computing Applications,* 2018, pp. 172–174, doi:10.1109/ICIRCA. 2018.8597227.

[8]   J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," 2018, [Online]. Available: https://pjreddie.com/ media/files/papers/YOLOv3.pdf.

[9]   S. Nurullayev and S. W. Lee, "Generalized parking occupancy analysis based on dilated convolutional neural network," *Sensors*, vol. 19(2), pp. 1-25, 2019, doi: 10.3390/s19020277.

[10]   W. Lan, J. Dang, Y. Wang, and S. Wang, "Pedestrian detection based on yolo network model," in *Proceedings of 2018 IEEE International Conference on Mechatronics and Automation,* 2018, pp. 1547–1551, doi: 10.1109/ICMA.2018. 8484698.

[11]   R. Zhang, Y. Yang, W. Wang, L. Zeng, J. Chen, and S. McGrath, "An algorithm for obstacle detection based on YOLO and light filed camera," in *Proceedings of the International Conference on Sensing Technology, ICST*, Jan. 2019, vol. 2018-December, pp. 223–226, doi: 10.1109/ ICSensT.2018.8603600.

[12]   A. Ammar, A. Koubaa, M. Ahmed, and A. Saad, "Aerial Images Processing for Car Detection using Convolutional Neural Networks: Comparison between Faster R-CNN and YoloV3," 2019, [Online]. Available: https://arxiv.org/pdf/1910.

07234v1.pdf.

[13] D. Tran, H. Manh Do, W. Sheng, H. Bai, and G. Chowdhary, "Real-time detection of distracted driving based on deep learning," *IET Intelligent Transport Systems*, vol. 12(10), pp. 1210–1219, 2018, doi: 10.1049/iet-its.2018.5172.

[14] Y. Zhang, K. Shang, J. Wang, N. Li, and M. M. Y. Zhang, "Patch strategy for deep face recognition," *IET Image Processing*, vol. 12, no. 5, pp. 819–825, 2018, doi: 10.1049/iet-ipr.2017.1085.

[15] A. K. Rangarajan, R. Purushothaman, and A. Ramesh, "Tomato crop disease classification using pre-trained deep learning algorithm," *Procedia Computer Science*, vol. 133, pp. 1040–1047, 2018, doi: 10.1016/j.procs.2018.07.070.

[16] A. A. Almisreb, N. Jamil, and N. M. Din, "Utilizing AlexNet Deep Transfer Learning for Ear Recognition," in *4th International Conference on Information Retrieval and Knowledge Management: Diving into Data Sciences,* 2018, pp. 8–12, doi:10.1109/INFRKM.2018.8464769.

[17] A. Şeker, "Evaluation of Fabric Defect Detection Based on Transfer Learning with Pre-trained AlexNet," in *2018 International Conference on Artificial Intelligence and Data Processing,* 2019, doi: 10.1109/IDAP.2018. 8620888.

[18] S. Liawatimena *et al.*, "A Fish Classification on Images using Transfer Learning and Matlab," in *1st Indonesian Association for Pattern Recognition International Conference,* 2018, pp. 108–112,doi:10.1109/INAPR.2018.8627007.

[19] S. Zhou, C. Chen, G. Han, and X. Hou, "Deep convolutional neural network with dilated convolution using small size dataset," in *Chinese Control Conference, CCC*, 2019, pp. 8568–8572, doi: 10.23919/ ChiCC.2019.8865226.

[20] H. T. Vu and C. C. Huang, "Parking Space Status Inference upon a Deep CNN and Multi-task Contrastive Network with Spatial Transform," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29(4), pp. 1194–1208, 2018, doi: 10.1109/TCSVT.2018.2826053.

[21] R. Wang, J. Xu, and T. X. Han, "Object instance detection with pruned Alexnet and extended training data," *Signal Processing: Image Communication*, vol. 70, pp. 145–156, 2019, doi: 10.1016/ j.image.2018.09.013.