Preprocessed Mask RCNN for Parking Space Detection in Smart Parking Systems

Ahmad Afiif Naufal[1]   Chastine Fatichah[1]*   Nanik Suciati[1]

[1]*Department of Informatics Engineering, Institut Teknologi Sepuluh Nopember, Indonesia*
* Corresponding author's Email: chastine@if.its.ac.id

Abstract: This research developed a smart parking system through video data analysis using deep learning techniques that automatically determine the availability of vacant parking spaces. This system has two main stages. The first is the stage of marking the parking position on the image of a parking lot captured by the camera. This research proposes a Preprocessed Region-based Convolutional Neural Network (Mask R-CNN) to mark the parking position on the input image of a full parking lot. The preprocess that combining contrast enhancement using the Exposure Fusion framework, aims to overcome the problem of lighting variations in images taken in an open area. In the second stage, each parking position is examined whether the position is vacant or not using mAlexNet. A series of trials on images with varying light conditions indicate that the Preprocessed Mask R-CNN can improve marking the parking positions with an accuracy of Intersection over Union (IoU) reach 85.80%. The result of marking the parking position is then used in the trial of the availability of parking space on video data using mAlexNet, and achieving an accuracy of 73.73%.

Keywords: Smart parking systems, Car parking spaces detection, Convolutional neural network, Mask R-CNN, Image enhancement.

1. Introduction

Car drivers in big cities have difficulty finding without in because of limited parking space and a large number of cars. The Smart City concept is applied to complement the city and improve society's quality by increasing the application of the latest information technology in the fields of the Internet of Things (IoT), Computer Vision, Machine Learning, 4G Networks, and others. From different applications related to Smart City (e.g., traffic analysis [1], vehicle detection [2], etc.), research related to the development of smart parking systems can optimize the use of parking lots in big cities.

A smart parking system is used widely in the transportation system to analyze and show vacant parking spaces. A smart parking system can easily help the driver and the administrator check and extract information about the parking space. The systems monitor the availability of vacant parking spaces and display them in a strategic location to be seen by customers and parking attendants. Car

detection in the traditional approach uses several sensors, such as inductive loop vehicle detectors, radar, and laser [3]. A parking lot must have a sensor in this approach. Therefore, a large number of sensors are needed to overcome the parking area with a large number of parking spaces for the scheme so that it will cost a lot.

Car detection is a fundamental step of many applications, such as intelligent drivers, intelligent transportation systems, and so on. Existing methods mainly include motion-based methods, shallow learning-based methods, and deep learning-based methods. Among these methods, deep learning can learn data representations more effectively [4]. Deep learning-based ways are better than traditional methods and make extraordinary progress in the development of car detection. Object classification using deep learning-based methods has attracted many researchers because of the good results [5]. One of the ways to obtain data from the parking area is visual monitoring [6]. Visual monitoring used to monitor the situation around the parking area to know

whether it is filled or vacant by using closed-circuit television (CCTV). One camera can monitor some parking spaces therefore, the usage of visual monitoring can be decrease costs compared to a sensor approach. The process of visual monitoring on the smart parking system was already done and got a good result when applied to a parking area that the parking slot is marked manually [7]. However, another issue that occurs on a smart parking system with a visual monitor is when providing annotations for the parking spaces that are still done manually [8]. The object car was not well detected when detecting parking space because every parking area is different. In addition, the open parking area captures different lighting varieties that depend on weather conditions [9]. Sometimes, the car can resemble the background, which is the parking area.

This research develops an automatic initialization strategy to detect parking spaces by detecting an object or the car in the parking area using Preprocessed Mask R-CNN. We used the CCTV, which captures the parking area that is filled with cars to obtain the parking area. The problem might occur if the car is located in an open parking area when the lighting variation due to the constant change in weather. To deal with this problem, we propose the preprocess method that combining contrast enhancement using the Exposure Fusion framework before applying Mask R-CNN. The output of marking the parking position on parking lot image is the region of interest (ROI) of each parking space obtained by determining the class and confidence value. In the next stage, each parking position is examined whether the position is vacant or not using mAlexNet. Data used in this research are images taken by CCTV, which obtained by capturing the situation around the parking area for 30 minutes from sunrise until sunset. We evaluate the results using the IoU values to determine how well the parking space detection results compared to the ground truth. The results of this research are expected to determine parking spaces automatically in different parking area environments so that they can improve the classification of parking availability statuses properly.

This paper is organized as follows, section 2 discusses related works on parking spaces detection in general. Section 3 presents the proposed methodology for the detection of parking spaces in CCTV camera captured images. And section 4 explains the experiments that have been carried out and shows the results, conclusion, and future work from this research presented in section 5.

2. Literature review

Detection of parking spaces availability will provide information related to the number of available parking spaces. A problem arises when the expected outcome from detecting parking space availability is not the same as the real situation. Research using CCTV captured images states that parking location annotations are a significant challenge in analyzing parking lots for the next process. There are some researches about detecting parking spaces availability with visual monitoring.

The method using binary classifiers for Convolutional Neural Network (CNN) used for the smart parking system was already suggested [10]. The data obtained is an RGB image converted to HSV to take the value of V, which is value. The transformation logarithm is applied to the gray value so that it becomes even broader, and Second Derivative performed to sharpen the outlines in the image. This process's results are converted into binary images and entered into the CNN. The research's deficiencies are that the researcher does not explain how the system can detect each empty or filled parking spaces.

Method to detect parking spaces availability by using a drone camera has already been proposed [11]. First, the parking area detected by giving four points in each parking area. After that, we need to make corrections on the result images to transform the lines created into a straight line to save its coordinates. To determine the availability of parking spaces, a line model used where the already parked car will form a straight line. If there is a parking space available, then the line will be disconnected, then the disconnected line will be marked as an available parking space. This research has an effective method for drones only, so if this method is applied to a camera, it will still have deficiencies in which it cannot form a line to show parking spaces availability.

The CNN method using smaller size Alexnet (mini Alexnet) to detect parking space availability has already been proposed [6]. This research is done to automatically detect parking spaces using a CCTV camera made by the researcher. Using mini Alexnet and dataset made by the researcher, this research succeeded in detecting parking spaces with accuracy above 90%. Unfortunately, this research method did not explain how to retrieve each parking space's location, so some questions were left. The mini Alexnet can automatically find the parking spaces from the whole image or whether the researcher makes a mark to find out the location of the parking spaces.

A method using You Only Look Once (YOLO) V3 and Lite Alexnet, which has few parameters to speed up the detection of parking spaces availability, has already been proposed [9]. The research conducted automatic parking spaces detection by utilizing the boundary box of YOLO V3. Classification of parking space availability using CNN with Alexnet architecture has modified the number of layers and parameters called Lite Alexnet. Research has a constraint that is needed to help marks in the form of lines on the parking spaces to find the parking spaces from the whole image automatically.

An approach using the Mask R-CNN model for counting vehicles in parking areas [12]. The proposed methodology does not require manually entered information on parking spot locations, allowing for a simple 'plug-and-play' installation. The proposed car detection allows the input image to have various perspectives, illuminations, and occlusions. The decline in performance is slight and responds well to changes in perspective and illumination in this research. However, the number of vehicles in parking locations tend to be underestimated.

The use of the Mask R-CNN method for the detection of parking space has already been proposed [13]. Bounding box from detection is compared with manually annotated bounding box to determine the classification of parking space status. This research scenario is still limited because it is necessary to initialize manual annotations for each parking space. This research can detect parking spots correctly, about 90%. Miss classification occurs because of camera angles. The research scenario has no obstruction in the parking lot so that the research obtains high results.

Another research was conducted by using Mask R-CNN to help solve the multi-angle parking issue [14]. This research proves that the Mask R-CNN model can run well in low angle CCTV areas with the highest accuracy value of 86%. The model has shown to adapt to various parking scenarios and is ideal for real-world use. However, a problem found in this research still has issues related to lighting, and car detection failure occurs if there are other objects in front of it (such as trees or light poles, or even other cars).

3. Materials and method

In this research, several processes are conducted to detect parking space. Those processes are contrast enhancement, automatic marking of coordinate points, and selection of the bounding box of each car as the detected object. The process carried out at the parking space detection stage can be seen in Fig. 1.

3.1 Datasets

Two datasets are needed in this research. First are images of the parking area from the stage of marking the area and a video from CCTV record to detect parking space availability in the parking area. The datasets that be used are images and videos from CCTV located in the outdoor parking area. In this research, data training that we used to detect parking space is CNRPark published in 2016 [6].

This dataset contains images of the outdoor parking area, that already segmented and labeled. In this dataset, there are various weather conditions, from sunny to rainy, each image size 150 ×150 pixels. Data collected from CNRPark and CCTV will be used to detect parking area automatically and detect parking space availability. The size of data testing used to capture all corners of the parking area is 1000 × 750 pixels for CNRPark and 1920 × 1080 pixels for video from CCTV on the parking area at Informatics Department ITS.

CNRPark dataset contains 4.081 images taken from 9 angles of different parking areas in sunny, cloudy, and rainy days in the morning, afternoon, and night. Data for videos are taken from CCTV in the parking area at Informatics Department. It consists of 30 videos with a total duration of 129 minutes taken from 1 angle of the parking area with different light conditions.

3.2 Exposure fusion framework

Exposure Fusion framework is a contrast improvement method by putting an RGB image. The purpose of increasing contrast is to solve bad lighting problems. Contrast increasing method that has been done was successfully solved low contrast problem and light distortion [15]. Increasing exposure can show less exposed areas, but at the same time, areas that is well exposed can become overexposed. Fusion Exposure framework's formula to get an image with all pixels that are well-positioned to increase exposure is defined Eq. (1), and the operator ∘ means element-wise multiplication. Well-exposed pixels are given a large weight, whereas poorly exposed pixels are given a small weight applied to the three color components that are not uniform for all pixels, so normalized weights is defined Eq. (2). We give a list of notations used in this paper in Table 1.

$$R^c = \sum_{i=1}^{N} W_i \circ P_i^c \qquad (1)$$

$$\sum_{i=1}^{N} W_i = 1 \qquad (2)$$

258

The problem faced is that images with other exposure settings are not available for image enhancement problems. That is because the image only captured once, and there are no additional images for different exposures. In contrast to images taken with different exposures, the problem related to image enhancement can be overcome properly so that it is highly correlated between the two images. Eq. (3) states the input image P can be mapped to the set of exposures. Brightness Transform Function (BTF) is a mapping function between two images that only differ in exposure to overcome the problem of different exposure settings [16].

$$P_i = g(P, k_i) \qquad (3)$$

In the phase of increasing contrast, the input image is combined with the image itself with other exposures to reduce complexity. Eq. (4) explains the formula for image merging. The problem of image enhancement can be divided into three parts, namely determining the values of W, g and k.

$$R^c = W \circ P^c + (1 - W) \circ g(P^c, k) \qquad (4)$$

The key to a successful algorithm for increasing an area that is not exposed and maintaining an area that has been well-exposed is the design for calculating W weights. Under-exposed pixels are assigned a small weight value, and well-exposed pixels get a significant weight value. The lighting of the scenery and the results of the weight matrix will be positively correlated. A high weight value is given to well-exposed pixels to maintain contrast (very bright areas). Eq. (5) expresses the weight calculation carried out.

$$W = T^\mu \qquad (5)$$

Optimization is needed to get an estimation illumination scene map T. Optimization is done by refining T through the optimization equation in the formula Eq. (6), where $|*|_2$ and $|*|_1$ are the l_2 and l_1. $\nabla_h T$ (horizontal) and $\nabla_v T$ (vertical) are the results of the first order derivative filter ∇. Eq. (6) aims to minimize the difference between the initial map L and the refined map T, while at the same time maintaining the smoothness of T. For initial estimation of illumination for each x, we using the lightness component as defined in the Eq. (7). Another vital thing for improving the illumination map is the design determination of M. We design the weight matrix as defined in the Eq. (8), where $|*|$ is the absolute value operator.

$$T^o = \min_T \|T - L\|_2^2 + \lambda \|M \circ \nabla T\|_1 \qquad (6)$$

$$L(x) = \max_{c \in \{R,G,B\}} P_c(x) \qquad (7)$$

$$M_d(x) = \frac{1}{|\sum_{y \in w(x)} \nabla_d L(y)| + \epsilon}, d \in \{h, v\} \qquad (8)$$

To reduce the complexity of the proposed closed-form solution, we approximate Eq. (6) as in [17]. Eq. (9) expresses the results of our approximate, but the problem occurs because many quadratic operations are involved. Let m_d, l, t and $\nabla_d l$ denotes vector versions of each of M_d, L, T and $\nabla_d L$. Eq. (10) expresses a linear function to obtain the proposed solution, where $\oslash$ is the division of elements wisely, $Diag(v)$ is an operation to construct a diagonal matrix using v, and D_d. The M design is the main difference between the illumination map estimation method used in our research and previous research [24]. The model we use is more straightforward and can produce the same results. The W can generate using other illumination decomposition techniques in the Retinex-based method.

$$T^o = \min_T \sum_x \left(\left(T(x) - L(x)\right)^2 + \right.$$
$$\left. \lambda \sum_{d \in \{h,v\}} \frac{M_d(x)(\nabla_d T(x))^2}{|\nabla_d L(x)| + \epsilon} \right) \qquad (9)$$

$$I + \lambda \sum_{d \in \{h,v\}} \left(D_d^T Diag(v) D_d\right) t = 1,$$
$$where \; v = m_d \oslash (|\nabla_d l| + \epsilon) \qquad (10)$$

3.3 Mask R-CNN

Mask R-CNN is a method expanded from Faster R-CNN by adding branches to predict mask objects parallel with the existing branch to detect square limit [18]. The main point is to separate the classification task and pixel-level mask prediction.

Based on the Faster R-CNN framework [19], a third branch was added to predict object masks in parallel with existing branches for classification and localization. Branch mask is a fully-connected small network that is applied to each RoI, predicting pixel-to-pixel mask segmentation. The additional mask makes it possible to choose pixels wisely on each object segment and extract each object separately without any background (which is not possible with semantic segmentation)

Mask R-CNN prediction results on the Common Objects in Context (COCO) dataset can identify object boundaries at the pixel level in detail. The

Mask R-CNN method is a simple method to be done and only adds a small overhead to get it to run on five frames per second. This method successfully outperforms all single model entries in every task given. This method also the winner of the COCO 2016 competition. The result given by the Mask R-CNN method is a square limit and segmentation mask in every object that face the instance segmentation.

Because pixel-level segmentation requires more fine-grained alignment than boundary squares, the Mask R-CNN enhances the RoI pooling layer (RoIAlign) so that it can be better and more precisely mapped to the area of the original image. The R-CNN model processes the input image provided to the network by carrying out a selective search. Then, it uses the result area of the selective search for feature extraction and classification using CNN that has already trained.

The Fast R-CNN model still uses a selective search algorithm to get the proposal region [20], but the Region of Interest (ROI) Pooling module is added to this model. The Fast R-CNN model can extract fixed-size windows from feature maps and use these features to get the final class label and boundary box. The advantage of the Fast R-CNN model is that the network can now be trained end to end. In the Faster model, R-CNN introduces the Regional Proposal Network (RPN), which makes regional proposals directly into the architecture to increase the need for selective search algorithms.

3.4 mAlexNet

The mAlexNet is a development of Alexnet which used to do binary classification. Alexnet is a famous deep convolutional neural network and used as a reference in many research [21]. Originally, Alexnet was trained on one million image datasets to recognize 1000 different classes. In research related to parking space availability, we only need to differentiate two classes. Smaller deep convolutional neural network architecture needs to be used to make the detection system work well. Simplify network was done because originally, Alexnet was designed to more complex introduction tasks on binary classification problems.

The mAlexNet is a deep convolutional neural network architecture inspired by AlexNet. In AlexNet, three convolutional layers and two fully-connected layers are used, including the output layer. First and second convolutional layers (conv1-2) followed by max pooling, local response normalization (LRN), and rectification linear (ReLU). The third convolutional layer (conv3) doesn't use local response normalization (LRN). The total

amount of conv1-2 filter and amount of neurons in the fully-connected layer (fc4) drastically reduced so that it would be equal to the problem dimension and obtained architecture with about 1340 1 parameters than AlexNet. In fc4 and fc5 (the output layer), no dropout regularization was used.

4. Results and analysis

Wherever Times is specified, Times Roman of Times New Roman may be used. If neither is available on your word processor, please use the font closest in appearance to Times. Avoid using bit-mapped fonts if possible. True-Type 1 fonts are preferred.

4.1 Parking space detection

The marking phase proposed is to determine all parking spaces to be analyzed subsequently. An image parking area that has been filled by the car is used as input for the marking stage. The image will be given a preprocessing step, before the detection of all vehicles using the Mask R-CNN method. An input image of a parking area assumes that the captured parking area has been filled by the car so that it has the same position as the available parking space.

Matterport is a media technology company that provides source code for the Mask R-CNN method written in Python. The Mask R-CNN provided in Matterport has been pre-trained so that researchers can use the results of the pre-trained to be used in this research using the Python programming language with the same architecture to detect cars. The proposed system introduces the preprocessing stage in the Mask R-CNN. The preprocessing step was added to overcome lighting problems in the detection of vehicles using the Mask R-CNN. Fig. 2 is the architecture of the proposed Preprocessed Mask R-CNN. The parking space status detection is shown in Fig. 3.

Parking space detection proposed is applied to the 270 chosen dataset images. The selection was made by taking an image of a parking area filled with cars and meeting all criteria. Obtained object detection results consisting of several classes in each image. The car object class is selected and used for further analysis. Each image consists of more than one car object (only the parking area filled with cars). Car object detection results can have obstacles in different environments. In high lighting conditions, there will be an inaccurate detection of car objects due to the reflection of sunlight. ROI's Ground Truth (GT) data and the number of cars are made from the pre-selected dataset images. The obtained area became the focus of the detection of car objects in
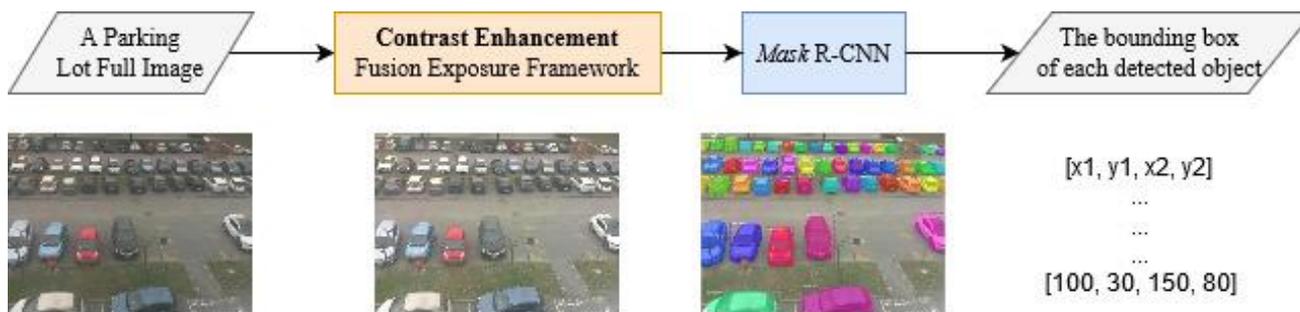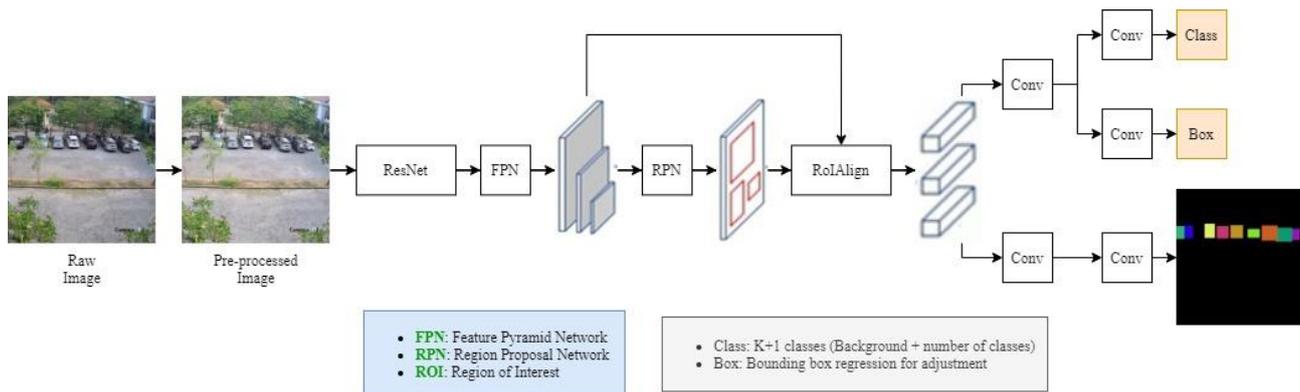
Figure. 1 Parking space detection

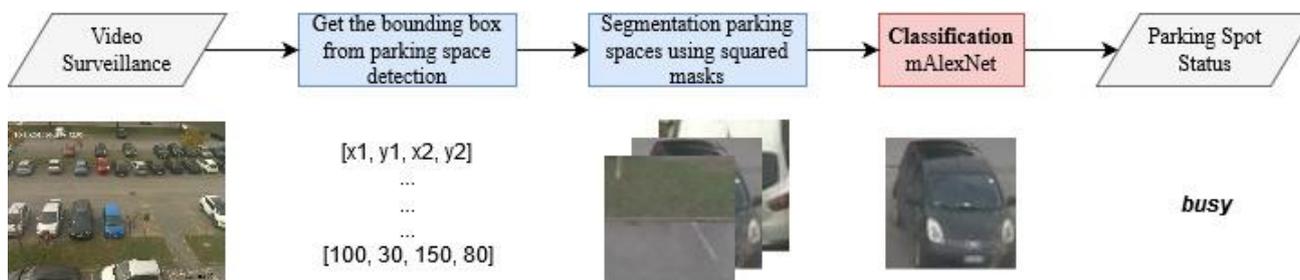Figure. 2 Preprocessed Mask R-CNN architecture

Figure. 3 Parking space status detection

(a)   (b)

Figure. 4 Example of image data: (a) input image and (b) results of detection of car objects as parking spaces

each image.

The result of parking space detection is in the form of a boundary box for each object in the camera. Fig. 4 is an example of the results of the detection of parking spaces for each object in the input image. The results of the parking space detection analysis with the highest IoU value will be references for the next analysis in each of the test scenarios conducted. The bounding box used to cut the image caught by CCTV to determine the status of parking space availability.

After performing a parking space detection analysis, the bounding box of the image with the best results will produce segmented images (patches). The imagery is used to determine the parking space and is classified using a model that has been made for the detection of parking space status.

4.2 Parking space status detection

After obtaining the results of a bounding box for each parking space at the parking detection space, a cut and a status selection made using a classification tool. The detection phase of a parking space is a collection of segmented images of a parking space. These patches are obtained from the detection of a parking space that has been detecting car objects in the parking lot. Fig. 5 is an example of the results of patches and their classes. The total number of labels obtained in the manual annotation results is 144,965, while the automatic parking space detection results
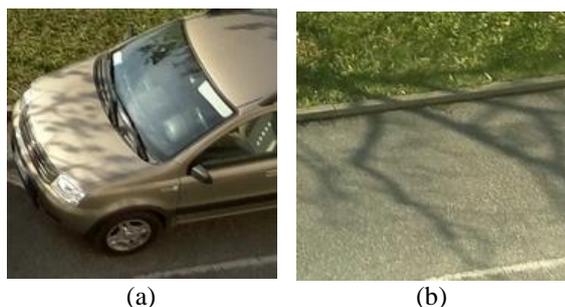
Figure. 5 Patches class label: (a) busy and (b) free

are 191,920.

The classification stage is the stage of determining the parking space that has located from the marking stage. The classification determines whether or not the parking space is available. Fig. 3 explained the classification stage carried out for the status of the parking space. To classify parking space availability at this stage, some CNN architectures that are used to detect parking space availability are Alexnet, and mAlexNet. This CNN will classify two classes, which are occupied (busy) or empty (free). At the training stage of the model, each model is trained with 18 epochs using Gradient Descent (GD), with a learning rate of 0.01 divided in half every six epochs for a batch size of 64, momentum 0.9 and decay weight 0.0005. The data used to conduct model training has a total of 94,493, with a comparison of busy and free classes 50:50. During the training process, the data was separated for use as testing with a split validation level of 0.1.

This research uses CCTV data from a parking lot provided by CNR Park. Then the marking was obtained from the detection of the place, and then the image of the marked area will be processed by the model and make the class busy or free.

4.3 Performance metrics evaluation

In this research, we measure the performance of detection results quantitatively. For the first performance metric, we use the IoU value to calculate parking space detection accuracy. Thus, the higher the IoU value, the more accurate the detection results. Table 2 is the average accuracy of the IoU of the experiments conducted without preprocessing and the three preprocessing methods according to the scenario. The table states that the best preprocessing for outdoor parking detection is ying. Fig. 6 is an example of the results of increased parking space detection from experiments conducted. From these results, without the method of preprocessing car objects challenging to detect is shown by the yellow circle in Fig. 6 (a). While with some commonly used preprocessing methods, such as histogram
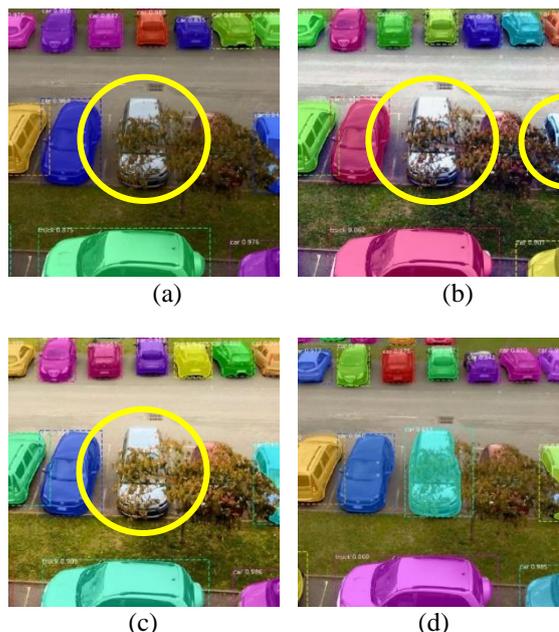

Figure. 6 Example results from increased parking space detection: (a) without preprocessing, (b) he, (c) dhe, and (d) ying
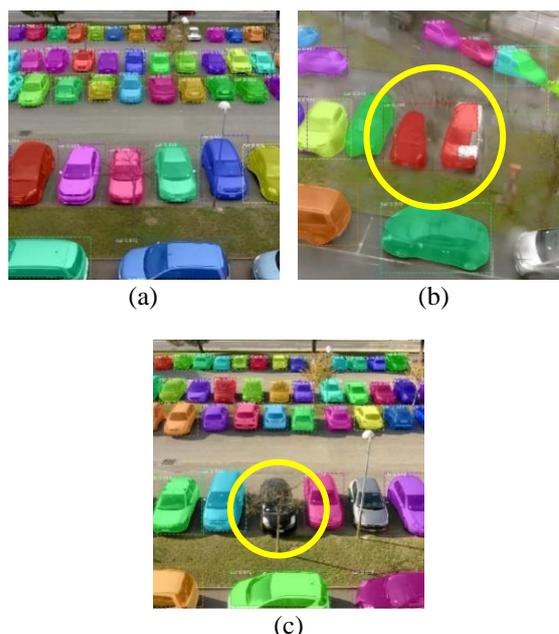

Figure. 7 Example results from increased parking space detection: (a) overcast, (b) rainy, and (c) sunny
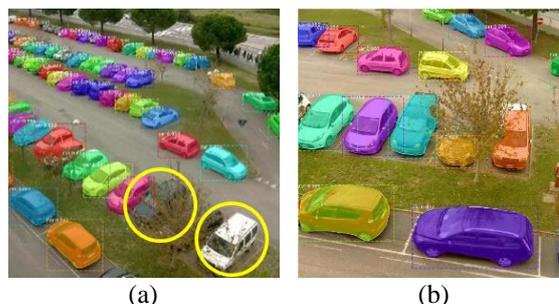

Figure. 8 Example of parking lot detection results on (a) camera1 and (b) camera2

equalization (he) and dynamic histogram equalization (dhe), there is a decrease in the detection performance of car objects, shown by the yellow circle in Fig. 6 (b).

Table 3 is the average accuracy of the IoU of the experiments conducted from five lambdas according to the scenario. The use of lambda value is to determine the closed-form solution in the calculation of the Beta-Gamma Correction model. Detection results can be accurate by choosing the right lambda parameter values. From the table, we can get that the best lambda for outdoor parking space detection is 2.

Table 4 is the average accuracy of the IoU of the experiments conducted from three weather conditions according to the scenario. From this table, we can get that the best weather conditions for outdoor parking detection are overcast. Fig. 7 shows an example of the proposed automatic parking space detection based on this scenario. From these results, it is known that in rainy weather conditions, there is overlapping detection because the camera captures the dew. So the results of poor car object detection performance are shown by the yellow circle in Fig. 7 (b). While in sunny weather conditions, car objects can resemble a background color, so it is difficult to detect, indicated by the yellow circle in Fig. 7 (c).

Table 5 shows the average accuracy of IoU from experiments conducted from nine cameras according to the scenario. From the table, we can get that the best camera for outdoor parking space detection is camera2. Fig. 8 shows an example of the proposed automatic parking space detection based on these scenarios. The point of view and distance of the camera can determine the performance of the detection results. Scenario camera2 got the highest accuracy, and it is because the object of the car object captured at close range. In the camera1 test scenario, many car objects fail to detect because of some cases, and such as car objects that resemble background colors, a two-car can be a one-car object, the yellow circle in Fig. 8 (a) indicates this.

Table 1. Comparison of preprocessing methods for parking space detection

Preprocessing	Processing Time (s)	Number of Car Accuracy (%)	IoU Accuracy (%)
without	4.88	91.78	84.36
he	4.86	91.50	84.38
dhe	4.91	93.80	84.38
ying	**4.82**	**92.30**	**85.80**

Table 2. Comparison of lambda parameters (ying) for parking space detection

Lambda	Processing Time (s)	Number of Car Accuracy (%)	IoU Accuracy (%)
1	4.78	93.94	85.74
2	**4.82**	**92.30**	**85.80**
3	4.67	92.85	85.49
4	4.67	93.45	84.90
5	4.67	92.27	84.54

Table 3. Comparison of weathers for parking space detection

Weather	Processing Time (s)	Number of Car Accuracy (%)	IoU Accuracy (%)
OVERCAST	**4.82**	**92.30**	**85.80**
RAINY	4.68	87.98	80.63
SUNNY	4.80	88.97	81.87

Table 4. Comparison of cameras for parking space detection

Camera	Processing Time (s)	Number of Car Accuracy (%)	IoU Accuracy (%)
camera1	8.94	60.83	69.58
camera2	**4.82**	**92.30**	**85.80**
camera3	4.78	89.72	76.16
camera4	4.78	95.77	75.80
camera5	4.94	94.88	77.11
camera6	4.94	84.97	73.01
camera7	5.07	95.84	77.66
camera8	5.05	91.56	82.27
camera9	4.87	90.82	83.88

Table 5. Comparison of classification models for parking space status detection

Model	Proposed Patches Result		
	Accuracy (%)	Precision (%)	Recall (%)
AlexNet	67.01	60.57	97.50
mAlexNet	**73.73**	**66.69**	**94.81**

For the second metric performance, we use a confusion matrix to get information about the classifier for the detection of parking space status. Information about the classifier can be obtained from the confusion matrix, such as accuracy, precision, and recall. Accuracy is a measure of experimental results that calculates the truth ratio of all data. Precision shows the relevance of the data to the predicted value.

Table 6. Results of the proposed system on video data

Video Filename	Accuracy (%)	Precision (%)	Recall (%)
data10-1.mp4	86.34	86.34	100.00
data10-2.mp4	83.52	83.52	100.00
data10-3.mp4	72.28	72.28	100.00
data10-4.mp4	76.20	76.20	100.00
data10-5.mp4	84.26	84.26	100.00
data10-6.mp4	93.44	93.44	100.00
data10-7.mp4	98.44	98.44	100.00
data10-8.mp4	87.95	87.95	100.00
data2-2.mp4	72.99	72.99	100.00
data2-3.mp4	80.42	80.42	100.00
data2-4.mp4	70.72	75.17	92.27
data23-1.mp4	72.79	72.79	100.00
data23-2.mp4	75.00	75.00	100.00
data23-3.mp4	71.75	71.75	100.00
data23-4.mp4	80.95	80.95	100.00
data23-5.mp4	87.50	87.50	100.00
data23-6.mp4	90.00	90.00	100.00
data23-7.mp4	93.75	93.75	100.00
data24-1.mp4	93.75	93.75	100.00
edit_25.mp4	92.12	92.12	100.00
Average	**83.21**	**83.43**	**99.61**

Recall, or sensitivity is the sum of the proportions of data whose class classification result prediction is true of all data, which classed as true to ground truth.

In this test, we use the pre-trained BVLC Caffe model to evaluate the performance of parking space status detection. In Table 6, we can see the value of accuracy, precision, and recall to detect the status of each patch image from parking space detection results. Classification of each patch resulting from parking space detection has been done well, namely achieving an accuracy value of 67.01% for the AlexeNet model and 73.73% for the mAlexNet model.

In Table 7, we can see accuracy, precision, and recall value to detect the status of each parking space that has been obtained from the parking space detection results for each video data. Furthermore, testing was conducted on video data received from CCTV installed in the ITS Information Department parking lot, achieved the results of parking space status detection with an average value of accuracy, precision and recall of 83.21%, 83.43% and 99.61%.

4.4 Approach evaluation

The proposed system aims to automatically detect existing parking spaces in parking lots, where most

Table 7. Comparison with existing approaches

Method	Need a Marker	Using Predefined Space	Support Automated System
A drone-based [11]	Yes	No	No
Classify with predefined space [6], [9], [10], [13]	No	Yes	No
Proposed method	No	No	Yes

researches still require manual segmentation of parking spaces on video data. Another contribution is the use of the Preprocessed Mask R-CNN to improve the accuracy and efficiency of the results. From the system evaluation results, our approach can find previously invisible parking spaces without a particular configuration on the parking lot. The system can detect filled parking spaces and calculate the number of empty parking spaces by subtracting the number of filled parking spaces from the total number of parking spaces.

A comparison of the related approaches for parking occupancy detection, it is shown in Table 8. An automated system of the proposed method can replace the manual use of sensor signals. As a result, our solutions can be applied directly to new parking lots without the costly manual configuration.

5. Conclusion

This research developed an automatically detecting the availability of vacant parking spaces through video data analysis in the outdoor parking area. The object detection technique detects parking spaces through the proposed approach. This system has two main stages, the first stage is marking the parking position on the image of a parking lot. We propose a Preprocessed Mask R-CNN to mark the parking position on the input image of a full parking lot. The preprocess that combining contrast enhancement using the Exposure Fusion framework, aims to overcome the problem of lighting variations in images taken in an open area. The next step is marking, which used the Mask R-CNN method. A marking label on the car object was used for further analysis of the parking area. The second stage is feature extraction based on the region of interest (ROI) of each marking on the parking area for classification of the ROI whether vacant or not using the mAlexNet model.

Experiments for parking space detection are evaluated using the accuracy of IoU. The best preprocessing method from the experimental results is Ying's Exposure Fusion with a lambda of 2 that can detect well and able to achieve an IoU accuracy of 85.80%. In the classification method, the best classification result of the system of parking space detection status obtained was when using the mAlexNet model from the proposed parking space detection method, which is 73.73%.

Therefore, the Preprocessed Mask R-CNN as the proposed method provides an accurate automatic parking space detection technique and can reduce unnecessary maintenance installation costs. Our solution does not require additional information for manual segmentation of video data like other studies, making it more flexible. Thus the proposed approach can be used further for better, without a detection of parking space status.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

As the first author, Ahmad Afiif Naufal contributed to the formation of the paper, including the formulation of methods, the implementation of methods, and the conduct of experiments. Chastine Fatichah supervised in the formulation of the proposed method and article preparation. Nanik Suciati supervised in article preparation and assured the process while the research is conducted.

Acknowledgments

References

[1] J. F. Raymond, "Traffic analysis: Protocols, attacks, design issues, and open problems", *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Vol. 2009, pp. 10–29, 2001.

[2] S. Battiato, G. M. Farinella, A. Furnari, G. Puglisi, A. Snijders, and J. Spiekstra, "An integrated system for vehicle tracking and classification", *Expert Syst. Appl.*, Vol. 42, No. 21, pp. 7263–7275, 2015.

[3] T. Celik and H. Kusetogullari, "Solar-powered automated road surveillance system for speed violation detection," *IEEE Trans. Ind. Electron.*, Vol. 57, No. 9, pp. 3216–3227, 2010.

[4] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning", *Nature*, Vol. 521, No. 7553, pp. 436–444, 2015.

[5] H. T. Vu and C. C. Huang, "Parking Space Status Inference Upon a Deep CNN and Multi-Task Contrastive Network with Spatial Transform", *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 29, No. 4, pp. 1194–1208, 2019.

[6] G. Amato, F. Carrara, F. Falchi, C. Gennaro, C. Meghini, and C. Vairo, "Deep learning for decentralized parking lot occupancy detection", *Expert Syst. Appl.*, Vol. 72, pp. 327–334, 2017.

[7] G. Amato, F. Carrara, F. Falchi, C. Gennaro, and C. Vairo, "Car parking occupancy detection using smart camera networks and Deep Learning", In: *Proc. of- IEEE Symp. Comput. Commun.*, Vol. 2016-Augus, No. Dl, pp. 1212–1217, 2016.

[8] P. Almeida and L. S. Oliveira, "PKLot - A Robust Dataset for Parking Lot Classi cation The PKLot Dataset", pp. 1–6.

[9] E. Tanuwijaya and C. Fatichah, "Modification of Alexnet Architecture For Detection of Car Parking Availability In Video CCTV", *J. Ilmu Komput. dan Inf.*, Vol. 13, No. 2, pp. 47–55, 2020.

[10] T. Thomas and T. Bhatt, "Smart Car Parking System Using Convolutional Neural Network", In: *Proc. of Int. Conf. Inven. Res. Comput. Appl. ICIRCA 2018*, No. Icirca, pp. 172–174, 2018.

[11] C. F. Peng, J. W. Hsieh, S. W. Leu, and C. H. Chuang, "Drone-based vacant parking space detection", In: *Proc. of - 32nd IEEE Int. Conf. Adv. Inf. Netw. Appl. Work. WAINA 2018*, vol. 2018-Janua, pp. 618–622, 2018.

[12] L. Ciampi, G. Amato, F. Falchi, C. Gennaro, and F. Rabitti, "Counting Vehicles with Cameras," *CEUR Workshop* In: *Proc. of.*, Vol. 2161, 2018.

[13] J. Ahmad, Z. Lewis, P. Duraisamy, and T. McDonald, "Parking Lot Monitoring using MRCNN", In: *Proc. of 2019 10th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2019*, pp. 10–13, 2019.

[14] T. Agrawal and S. Urolagin, "Multi-Angle Parking Detection System using Mask R-CNN", *ACM Int. Conf. Proc. Ser.*, pp. 76–80, 2020.

[15] Z. Ying, G. Li, Y. Ren, R. Wang, and W. Wang, "A new image contrast enhancement algorithm using exposure fusion framework", *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Vol. 10425 LNCS, pp. 36–46, 2017.

[16] Z. Ying, G. Li, Y. Ren, R. Wang, and W. Wang, "A New Low-Light Image Enhancement Algorithm Using Camera Response Model,"

Proc. - 2017 IEEE Int. Conf. Comput. Vis. Work. ICCVW 2017, Vol. 2018-Janua, No. 2014, pp. 3015–3022, 2017.

[17] X. Guo, "LIME: A method for low-light image enhancement", *MM 2016 -* In: *Proc. of 2016 ACM Multimed. Conf.*, pp. 87–91, 2016.

[18] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN", In: *Proc. of IEEE Int. Conf. Comput. Vis.*, Vol. 2017-Octob, pp. 2980–2988, 2017.

[19] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN : Towards Real-Time Object Detection with Region Proposal Networks", pp. 1–14.

[20] R. Girshick, "Fast R-CNN", in *Proc. of the IEEE International Conf. on Computer Vision*, 2015, Vol. 2015 Inter, pp. 1440–1448.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks", *Commun. ACM*, Vol. 60, No. 6, pp. 84–90, 2017.