

Effective Image-Based Parking Occupancy Detection using Masked Region Based Convolutional Neural Network

MSc Research Project
Data Analytics

Ronu Skariah
Student ID: x21159840

School of Computing
National College of Ireland

Supervisor: Noel Cosgrave

National College of Ireland
Project Submission Sheet
School of Computing



Student Name:	Ronu Skariah
Student ID:	x21159840
Programme:	Data Analytics
Year:	2022
Module:	MSc Research Project
Supervisor:	Noel Cosgrave
Submission Due Date:	15/12/2022
Project Title:	Effective Image-Based Parking Occupancy Detection using Masked Region Based Convolutional Neural Network
Word Count:	8487
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	15th December 2022

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	<input type="checkbox"/>
Attach a Moodle submission receipt of the online project submission , to each project (including multiple copies).	<input type="checkbox"/>
You must ensure that you retain a HARD COPY of the project , both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	<input type="checkbox"/>

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Effective Image-Based Parking Occupancy Detection using Masked Region Based Convolutional Neural Network

Ronu Skariah
x21159840

Abstract

As the field of machine learning is shaping the world completely the usage of the techniques in machine learning can be effectively utilized to finding out a solution to the parking space identification problem which is less expensive and easier to implement. This research implements a dynamic, simple, and less expensive algorithm for the identification of parking spaces using machine learning techniques. This algorithm uses a deep learning network for the identification of all the parking spaces in any parking lot and performs the intersection over union technique to identify all the empty and filled space in any parking lot. This dynamic algorithm was able to perform on an average scale over images of different parking lots collected over sunny, rainy, and overcast weather. The object detection deep learning network Mask-RCNN which uses instance segmentation for the identification of vehicles was performed with a map score of 0.901 after performing the hyperparameter tuning-based training over 2,000 images.

1 Introduction

1.1 Background and Motivation

Rapid enhancements in the vehicle intelligence industry have been increasing very drastically towards the direction of artificial intelligence and networking area. Artificial intelligence has taken over the vehicle industry as self-driving vehicles with accurate driving technology have taken over the roads. People around the world always prefer personal transportation for the journey, and as a result count of vehicles on road increases every day. Since fossil fuel is the major renewable energy source most vehicles use, their combustion causes pollution growth. Pollution is caused by traffic congestion which is formed by an increase in vehicle count on roads and the long queue of vehicles formed for looking free space for parking. Major cities around the world face the issue of parking space availability, even in paid parking lot people suffer difficulty in finding a free spot even after they are paying for a space. A long queue of vehicles means a waste of fuel and time. Random studies specify that people in the USA spend an average of 7.8 minutes looking for a parking spot in cities. Properly optimized utilization of parking spots can help in the development of smart cities and hence the concepts of parking guidance and parking spot detection come under intelligent parking system.

Most of the approaches that still exist can be classified as sensor-based approaches and image-based approaches. A Parking Guidance and Information System (PGI) (Chen and Chang; 2011) was the first wireless sensor-based system developed to solve the issue of parking space detection and this technique reduces the time and effort required to find an empty parking space. Another conventional technique was using Histogram Oriented Gradients for the identification of parking spaces, this concept was smart and relatively faster compared to the previous model. The issue with these conventional methods is that they are expensive to set up and maintain. Some Image-based approaches use image processing and deep learning concept to perform the identification of empty and free lot.

This study proposes an effective algorithm for the detection of parking space occupancy in any indoor and outdoor parking that uses machine learning techniques. The initial stage of the algorithm is to identify the total parking spaces in a parking lot automatically. This will be a major task in the algorithm which can be handled by a deep learning network. Some object detection algorithms use a Convolutional Neural Network (Acharya et al.; 2018) that performs object detection by sliding windows that slide over an image to identify features. To detect any motion of an object in a different orientation it needs to scan the same image multiple times. The deep learning network Yolo V3 (Ojha et al.; 2021) also performs object detection using a bounding box technique, but they are not faster. This study proposes the detection of parking space occupancy using a Mask Region-based Convolutional Neural Network which is a computer vision deep learning network that comes under the RCNN family which performs instance segmentation for object detection and produces bounding boxes along with a mask around the object.

Using this proposed deep learning network in the algorithm total parking spaces in a parking lot can be identified faster and accurately. Vacant and occupied parking spaces can be detected accurately by performing the evaluation metrics of intersection over the union. As the demand for intelligent parking systems increases, this study can have a good impact on the transportation sector, and vehicle intelligence industry and benefit all people by reducing time and energy consumption for finding a parking space. Computer vision which is a field in artificial intelligence and object detection is an exciting and easy-to-implement concept with proper data. Difficulties in finding a parking space these days have motivated the concept of using this object detection in computer vision to find a better solution that is less expensive, easy to implement, and precise in all terms.

1.2 Research Question and Objective

The research question that has been outlined in this study is “How deep learning networks and techniques can be used to develop a dynamic algorithm for indoor and outdoor parking space identification problems?”. To solve the research question the specific set of objectives that have been followed includes:

- Identification of state-of-art deep learning network to perform object detection.
- Identification of a dynamic technique for detection of free and occupied parking spaces.
- Evaluation of the state-of-art model in object detection based on map score and evaluation of the algorithm performance in identifying vacant and filled spaces.

1.3 Limitations and Challenges

Some of the challenges faced for this implementation was the limitation in the usage of a live surveillance feed because it requires additional permission. This issue was replaced by using a parking space image dataset (Amato et al.; 2017). Another challenge was finding out the vehicle detection dataset for training the Mask R-CNN model for object detection. This was solved by using an image dataset from the Roboflow public data repository and annotations were done in XML format using an open-source annotation tool. One of the most difficult challenges faced was the evaluation of the model because to evaluate the model, predictions should be compared with ground truth. The current dataset is not a labeled dataset which causes difficulties in the evaluation of the algorithm accuracy. Occlusion in an image which is obstruction of an object by another object may cause failure in the identification of the object by the model. Once the algorithm does not identify an object then it won't identify that particular space as empty or occupied. Panoramic view and night vision images distort image pixels which cause low image detection.

2 Related Work

In this section conventional method that was used in parking space detection along with some promising work in deep learning technique that has been performed in the same area and object detection concepts are discussed.

2.1 Conventional Methods for Parking Space Identification

One of the most convenient parking space detection techniques (Suhr and Jung; 2016) is based on the combination of multiple approaches that includes a free space-based and parking slot mapping-based. The open spaces are identified by identifying the proper coordinates of parked cars and pillars, whereas the parking spaces are identified using parallel line pairs. The two concepts for detecting free space by using the ultrasonic sensor-based technique and recognizing all the open areas by making predictions about the placement of all the surrounding obstacles by the use of pillar data. The ultrasonic sensor which is placed at the different sections of vehicles is used to assess the distance of nearby obstacles. This method finds openings by gathering ultrasound sensor information and recognizing rapid distance changes. The secondary methodology is a pillar-based system that detects pillars using AVM images and ultrasound sensors and creates open areas based on their placements. Identifying pillars improves open space recognition and parking slot monitoring. This method achieved a 97.4% recall rate and a 99.2% accuracy rate in underground and indoor conditions. The recommended fusion strategy has a high recall since it recognizes parking slot markings when there are no nearby obstructions and free spaces if the parking lines are of low quality.

Another method (Kergoat et al.; 1993) uses a modified Mean Shift technique which is an image segmentation algorithm that performs accurate segmentation on an image for the identification of parking space problem. This method was evaluated with a proper detection rate of 97% on both outdoor and indoor parking lots. The limitation of the concept was that it was unable to identify non-vehicle items and also it cannot find a vehicle if the color of the vehicle and the ground are similar. This limitation was later overcome using the LBP for extracting the parking texture features. The LBP approach

extracts the texture of the grey scale images around a region immediately and measures it. Using these invariant properties of greyscale this method was able to perform well even with uneven lighting and the effect of shadow. A binary SVM classifier was used by LBP to properly classify an empty and vacant spot. Histogram-based techniques (Choeychuen; 2013) was the better technique so far used to enhance the parking lot mapping accuracy a fine-tuning has been performed on top. All the test results were better when testing was performed for this approach on a simulation and real-time parking lot. In order to perform object detection automated threshold and fine-tuning were used which enhanced the detection accuracy. Using the predicted calculation of the parking lot vehicle motion is analyzed and if a vehicle is estimated to be parked then the system will include that in a spatial-temporal histogram which is used for mapping the parking lot. This approach was not handled well for an outside parking lot and they were a more static approach which is costly.

2.2 Deep Learning Methods for Parking Space Identification

Parking detection classification (Nyambal and Klein; 2017) using a convolutional neural network along with Caffe and Nvidia DiGits was a sprout in deep learning methodology-based solution in the vehicle industry. This system was able to perform robustly on any of the foreign parking lot datasets according to the test results. The final convolutional neural network model for the approach Alexnet with Stochastic Gradient Descent as solver was determined by comparing the test results obtained for models AlexNet and LetNet. The data collection was the most difficult section carried out in this implementation. The whole system is performed on the basis of prediction and classification. The training was properly carried out using the Digit framework and the Caffe model was used for the prediction of vacant spaces in a parking garage. Classification using an Alexnet with Stochastic Gradient Descent as solver had performed best based on this prediction. The prediction was performed based on the fifth frame from the video feed and test results were showing an accuracy of 99% for any dataset.

Using a random forest and an artificial feed-forward network (Martín Nieto et al.; 2019) parking spaces can be identified. For this, the data was collected in real-time using multi-camera feeds, and the results show that the feed-forward network outperforms the random forest model. The NSE, MASE, and MAE value for the random forest is twice that of the feed-forward network. Data pre-processing part consist of resampling at one-minute intervals and an occupancy lookback window was introduced. For every 5-minute period, the prediction intervals were programmed to forecast up to 90 minutes in advance. With the help of hyperparameter adjustment, these models were improved.

A convolution using pixel skipping (Khalfi and Guerroumi; 2021) to cover a wider input is a dilated convolution layer and this dilation layer can increase the accuracy of a CNN model. The CNN model used here is the Alexnet without the final layer. After applying linear rectification (ReLU) max pooling has been applied to all the convolutional layers. Each of the convolutional layers has a dilation parameter from the first layer to the second layer rate of the dilation parameter changes and the first layer filter count is made at 64, the second layer at 128, and the final layer is at 256. The model parameters were set up using a dilated convolution layer. By reducing the number of filters, an architecture was created with fewer training parameters, allowing for the provision of a model that is more optimal than alternative models and along the entire categorization, the model accuracy was 92%

When image processing technique (Holdsworth et al.; 2001) is been compared with a machine learning technique for detecting unoccupied and occupied parking spaces. It was found that the convolutional network achieved 99.9% occupancy and 99.4% vacancy accuracy while the color-based image processing approach has 22.6% occupancy accuracy and 99.1% vacancy accuracy. The convolutional network with seven convolution layers classifies unoccupied and occupied parking spaces more accurately, even with a low-resolution video feed. The CNN model uses 4 layers of 5x5 filters to extract image features and a max pooling layer to minimize image size. The fifth layer is a flattening layer, flattening the preceding four layer’s values into one-dimensional output. The sixth dense layer splits one-dimensional data into 16 units. While the color-based parking detection uses each parking spot’s RGB mode value. The mode value is the image’s most frequent value and classification is based on the mode value. These approaches had a good performance but they are not reliable for any dynamic environments. While (Amato et al.; 2016) uses images of parking spaces that are masked and segmented into different patches and CNN classifies them as unoccupied or populated. The CNN is trained using the Caffe model using gradient descent with momentum. This strategy was finalized after comparing the mAlexnet with three convolutional layers and two fully connected layers against mLetnet with four convolutional layers and one fully connected layer. mAlexnet performed better than mLetnet.

The street parking occupancy detection (Gkolas and Vlahogianni; 2019) was performed based on the concept of the average length of a car is 4.5 meters, and a parking spot is free if the area is 5 meters long. Based on the vehicle speed and a video capture rate of 10 images per second, 1 image for every 2 meters was taken into consideration. To prevent a vehicle from missing a 5-m gap as it passes, 7 meters are needed. For the CNN model after max-pooling layers, a LetNet-5 model with two convolutional layers, one stride, and zero padding was used. This model was developed using gradient descent with momentum and trained on a dataset gathered by using video mounted on a moving car. In order to compare the performance of CNN for parking space recognition and prediction, various SVM models were trained on the same dataset, and with the help of the two convolutional layer networks, 90% accuracy was attained.

Due to occlusion, wide-angle images are much more challenging to detect using any deep-learning algorithm. It has been resolved using an improved Mask R-CNN model (Wu et al.; 2022) that integrates a backbone topology and Selective Kernel Networks. Moreover, it creates a route with clear downward connections from the lowest level to the higher level by substituting Soft-NMS for the Non-Maximum Suppression at the back of Feature Pyramid Networks (FPN). For low-angle camera perspectives, this revised architecture was able to outperform the original Mask R-CNN. In any given feature map, two convolutions with kernel sizes of 3x5 and combining depth-wise convolutions together constitute the three phases of the kernel attentiveness technique. Higher-level feature maps provide greater semantic information for larger resolution, but inadequate spatial location information for Resnet50, which has been substituted by Resnet101. The top-down sampling approach is used to achieve higher-resolution features. With this technique, the NMS is replaced by the Soft-NMS, enabling the detection score for nearby objects to decline rather than be set to 0. The primary dataset has indeed been divided into train, valid, and test based on the 6:2:2 ratio. The learning rate for the model has been set at 0.001, and the thresholds for computing the AP have been set at 0.5 and 0.75.

2.3 Deep Learning Methods for Object Detection

A practical application of Mask R-CNN model on object detection and KLT tracker (Al-Ariny et al.; 2020) for counting the number of automobiles in videos. The segmentation model Mask R-CNN instance locates cars every N frames and the findings demonstrate that the suggested strategy performs well and accurately. Automobile detection will be performed by the Mask R-CNN using an ROI mask and tracking will be performed using a KLT tracker which enhances the counting accuracy. To decrease time complexity, the detecting step is performed over every N frame. Corner points are properly identified and monitored by performing differences in coordinates for each corner point from one frame to the next frame. Once the corner points of detected automobile segments are reviewed, if there are new points that automobile will be spotted for the first time, and it will be labeled. If a recognized segment comprises both old tracked and new detectable corner points, the vehicle object was previously identified in a previous frame if any corner points inside the segment are old tracked points. This model accuracy was evaluated based on precision at 93.3% and 96.43% on two different datasets.

Vehicle identification using an instance segmentation technique (Ojha et al.; 2021) to generate bounding boxes and object masks uses transfer learning to achieve map and mar of 90.27% and 92.38% based on the evaluation. Instance segmentation serves the purpose of handling object detection from images with partial occlusion. The model provides instance segmentation by identifying every instance of an object of interest within an image at the pixel level. Convolution feature maps are generated using ResNet-101 which is a deep convolution neural network and are then sent over to RPN to generate anchor boxes for predicting vehicle-containing locations. These proposals are then given to two concurrent structures, one of which uses fully connected layers to classify the region as either a vehicle or a background and the feature pyramid network is used for determining the coordinates of the bounding box. ROI pool operation is replaced by the ROI Align layer to avoid loss. The pre-trained weight was used to learn the model for transfer learning. The weight is adjusted after each epoch using a learning rate of 0.001, weight decay of 0.0001, and momentum of 0.9.

Even though Mask R-CNN performs best, the Network degradation issue of Mask R-CNN (Chen et al.; 2021) has been overcome using a residual network by dropping the direct fitting function which reduces higher processing complexity and extra parameters. The Residual network used is a 101 layer which has been created by stacking up residual networks of 52 and 152 layers. Mask R-CNN uses the residual network and feature pyramid for feature extraction. In order to avoid the pooling quantization loss the ROI pooling has been replaced by ROI alignment. Mask R-CNN target detection allows for the real-time recognition and segmentation of target objects even in the absence of background or partial occlusion. The recall rate based on the evaluation was 20% and the AP value for object detection was 98%. Night-time vehicle detection utilizing a YoloV3 with MSR image enhancement (Benjdira et al.; 2019) method This suggested method achieved an average accuracy of 93.66% and 6.14% which is higher than Faster R-CNN. It also provides an image improvement approach comprising the MSR algorithm, which is the conversion of RGB color space to YUV color space, and the minimum detectable inverse as the incident image factor. The image was used to train the YOLOV3 model, which is a Darknet-53 with 53 convolutional layers, some 3x3 and 1x1 filters, and some residual connections. The bottleneck layers provide the same function as the pooling levels in a normal CNN. To enable the recognition of multi-scale objects, the last three feature maps

of varying sizes have been integrated. This method outperformed the Faster R-CNN and SSD, with AP and FPS at 93.66% and 30.03 fps, respectively. Compared to the Faster R-CNN and SSD, this recommended method generated 6.14 and 3.21 percent more AP and 20.26 and 4.09 percent greater FPS.

The comparison of the performance of the Haar Cascade classifier and YoloV3 (Manase et al.; 2020) for parking space detection from an image using image processing, was performed based on ten completely distinct situations. YOLOv3 has achieved the best accuracy of 96.88% with a 90% probability while the Haar Cascade Classifier has an accuracy of 63.34 percent. The data has been gathered from a stationary camera that has been placed at a height of 2.5 meters in one of the roadside parking lots. Conversion of video to the image frame, extraction of RGB components from these image frames, and conversion of the images frame to grayscale has been carried forward as a part of image pre-processing. Object detection is conducted using two models. The haar cascade classifier, which consists of the Haar-Like Feature, Integral image, Adaboost (Adaptive Boosting), and Cascade Classifier processes, was used for object recognition in order to identify the car in each frame. YoloV3 was used to recognize the car from each image frame in order to evaluate the performance of this approach. The comparison test shows that the YoloV3 model was able to perform detection faster compared to the haar classifier.

The concept of identification of pedestrian crosswalks (Malbog; 2019) and segment instances was performed using a Mask R-CNN model. The model achieved an object identification accuracy of 97% and generated a mask of multiple colors for different crosswalks within a single image. Crosswalk images were collected and used to carry out this implementation. These images were annotated and split into train and test with an 8:2 ratio. Mask R-CNN with Resnet-101-based backbone was trained at a learning rate of 0.01 for up to 30 epochs. RPN network in Mask R-CNN applies a basic binary classifier to images, after which the system calculates scores for each item in the picture and places those objects in one of many anchors (boxes). An RPN generates targeting by first generating a grid of anchors covering the entire image at different scales, and then computing the Intersection-over-Union of those anchors with the ground truth object. The classifier initiates plans to provide a probability for each class and regression bounds. When a low coincidence filter is applied, detection will result in non-maximal suppression. During the detection step, information from the previous layer is used to produce refined bounding boxes and class IDs. It is mostly used for creating segmentation masks. This model has an issue due to the occlusion in the images which was causing the model detection accuracy.

3 Methodology

The research focuses on object localization using deep learning techniques. Object localization is performed based on instance segmentation which is a technique that identifies object instances and their boundaries. Instance segmentation will be utilized here to provide the pixel coordinate of the car using semantic segmentation and object detection. Using these pixel coordinates of the car the occupancy of the car inside a parking lot can be identified by calculating the intersection over the union. This makes the methodology dynamic, as the major concept will be the detection of the car in the region of interest. This research follows the KDD methodology.

3.1 Data Description and Acquisition

This research uses two sets of data for the implementation, which includes an image dataset used for training the object detection algorithm. The dataset for object detection (MXK; 2022) will be gathered from the roboflow repository which is a publicly available dataset repository, the dataset has more than 6,000 high-resolution images of a car which are slightly annotated based on the model and brand. These images are obtained from the roboflow server. Since the usage of surveillance cameras is limited, the parking lot images are gathered from the CNR Park-it (Amato et al.; 2017) dataset which has around 1,50,000 images of the parking lot and the images are annotated based on the occupancy of the car in each parking spaces. These images are collected by using more than 5 Raspberry Pi cameras over a particular period of time. This dataset is publicly available to use and is obtained from the CNR Park-it webserver.

3.2 Data Preparation

The object detection was performed using a Masked Region-Based Convolutional Neural network that requires the image annotation (Malbog; 2019) in XML formats for training. The training dataset is annotated using a rectangular bounding box selector in the annotation tools to generate the coordinates of the car in each image and these annotations are stored in XML format. The model learning will be performed based on the bounding box coordinate which is generated. Since the images captured from a camera will be in RGB (Red, Blue, Green) format, conversion of the color channel will not be required and the dimensions of the images are also kept constant. Finally, all these images are stored based on serial numbering format and stored in a directory, Table 1 shows the partition of the object detection dataset.

Table 1: Data split ratio

	Total Size	Ratio
Train	2002	0.8
Validation	220	0.1
Test	226	0.1

3.3 Modelling

Object detection which is part of computer vision will be a major area of this research. Deep learning algorithms like YOLO and R-CNN are the two-object recognition deep learning algorithm families. Mask R-CNN which is an evolution among the R-CNN networks that uses a Region Proposal Network and instance segmentation property for object detection will be used. Based on the comparison of previous work (Benjdira et al.; 2019),(Ojha et al.; 2021),(Nyambal and Klein; 2017) the Mask RCNN can perform object detection faster compared to the Yolo, Faster RCNN, and CNN network, but there will be slight variation in accuracy and hence it has been considered.

3.3.1 Mask R-CNN

Mask R-CNN is a state-of-art deep learning algorithm that performs instance segmentation and uses the basic Faster R-CNN framework. The instance segmentation algorithm

performs both segmentation and object detection together. Basically, an instance segmentation algorithm will first perform object detection which will identify an object and performs segmentation to identify the pixel coordinates of that object. As shown in Figure 1 Mask R-CNN uses two phases the first stage scans the images and identifies the candidate object. The second stage will predict the object class, fine-tunes the bounding box, and creates a mask at the pixel level for the item.

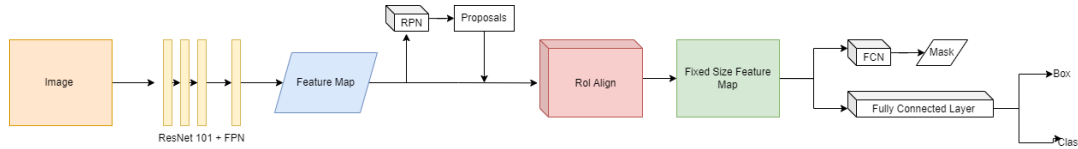


Figure 1: The architecture diagram of Mask R-CNN.

The working of Mask R-CNN (Ojha et al.; 2021) is similar to Faster R-CNN as the region proposal creation stage is the same in both architectures, but the second stage generates a binary mask for each RoI in addition to predicting class and generating bounding boxes in parallel. The backbone is a deep neural network RestNet101 + FPN which is a top-bottom with a lateral connection. It extracts the feature map from the images and there are different layers of RoI generation where this multi-layer feature pyramid network produces ROI at several scales that enhance the performance of the ResNet architecture. Based on the feature pyramid network the size of the feature map will be half or even doubled. The top-bottom approach is a technique used to create resulting feature maps.

It begins at the top feature map and moves down to larger by conducting upscale processes. The 1x1 convolution is used to restrict the number of streams to 256 before upsampling. The outcome of the preceding upsampling cycle is then added gradually to the above. All outputs are subjected to a 3 X 3 convolution layer in order to create the finalized 4 extracted features. The sixth feature map is produced by a max pooling operation from the fifth feature map. A 3x3 convolution layer applies to the entire convolution feature map created by the former layer in the RPN network. It then transfers the output of all these branches, which compute the object score and regress the bounding box coordinates.

3.3.2 ROI Align

As RoI pooling performs quantization in the first and second stages, this results in significant degradation of input if it is fed to the final layer, RoI align has replaced RoI pooling and RoI Align will minimize the data redundancy (Ojha et al.; 2021) since as it does not conduct quantization in any phase. Beginning with the feature map of dimension HxW, the ROI Align divides the feature map of the previous Convolution layer into MxN grids of the same size. With the addition of the segmented branch to the architecture, the Mask R-CNN interpretation performance was increased to double frame per second.

3.3.3 The Mask Representation

Here the Mask R-CNN will use a fully connected network to predict the mask. The reason is, to increase the performance both classification and regression layers outputs

should be collapsed and it requires the pixel correlation with these above layers. The fully connected convolution network produces the output mask of shape $M \times M$ based on the feature fed from the RoI to the network. Among the fully connected layer, the mask enhancement is done by a 1×1 layer of convolution and the channel size will be reduced to 256 after that.

3.4 Alternate Methodology Considered

Mask R-CNN and Yolo are the two real-time object detection deep learning algorithms. Along these algorithms, Yolo uses a single Convolutional Neural Network to perform object detection and it works based on bounding box regression, residual blocks, and Intersection over the union. Mask R-CNN on the other hand uses the concept of image segmentation. Image segmentation is the division of a digital image into different segments. Mask R-CNN uses two types of segmentation that includes semantic segmentation which will divide a digital image into a fixed category and instance segmentation which will split an image into different categories of segments. Based on Benjdira et al. (2019) Yolo has better accuracy compared to Mask R-CNN, but since Mask R-CNN uses (Malbog; 2019) image segmentation it will be faster than Yolo and it also provides the object location and outline in an image. This will be helpful in implementing a dynamic algorithm that uses intersection over the union. The alternate approach of training a model from scratch (Nyambal and Klein; 2017) using the empty and filled parking space category will be a more straightforward technique and the complete algorithm will not be dynamic as the model has to be trained multiple times based on the environment change. Since training a model always required a lump-sum amount of processing power, time, and data this methodology was rejected.

4 Design Specification

There are several concepts that may be used for the detection of parking spaces. In the beginning, parking spots will be manually identified by locating each parking space's coordinates. However, this approach will not be suitable for a dynamic environment. The second concept is to identify parking spaces based on parking meters, this requires an assumption that there will always be a parking lot in front of every parking meter, it cannot be guaranteed that this approach will be effective. There won't be parking meters in all parking lots and locating the parking space can be challenging because it may be either in front or even relocate. An even more appropriate concept will be finding all cars which are not moving and from this, it can be identified as a parking lot. This concept will not be the best one, but this can overcome the limitation of the other two concepts. The algorithm followed for the implementation is shown in the Figure 2.

Detection of the car is performed based on object detection where the class will be the car. Using the deep learning algorithm Mask R-CNN detection of the car from a parking lot image will be performed. For identification of the car, the model will be first trained on an object detection dataset based on a constant learning rate. Then the trained model weight will be used for identifying the car in the parking lot image. Since the Mask R-CNN algorithm performs instance segmentation which is the combination of object detection and semantic segmentation, it will produce a mask around the object and also generate the object coordinate for the detected car class. Using these object

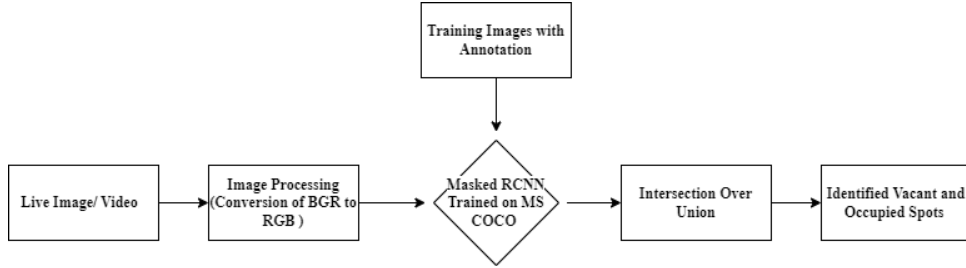


Figure 2: The Algorithm Design flowchart.

coordinate evaluation metrics of Intersection Over Union will be performed to identify the occupancy of the car in a parking space.

5 Implementation

The overall process flow that has been followed in this research for carrying out the complete implementation is shown in Figure 3. A details description is discussed below.

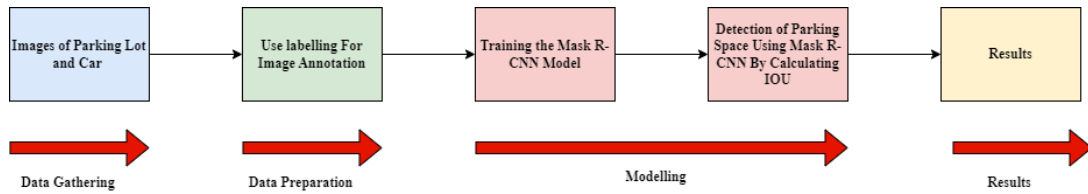


Figure 3: The implementation process flow.

The final work will be carried out based on the two major concepts which include object detection using instance segmentation and intersection over union which will be used for the identification of empty and occupied parking lots based on the concept of bounding box overlap. Normally parking spaces are identified by the detection of a car that does not have any motion in a video and since the video has been replaced by an image here, the above concept can be replaced by the concept of reference imaging where the Mask R-CNN uses an image of a completed filled parking lot to identify all the parking spaces. This process will be performed using the instance segmentation property of the model which has been discussed in the above. Each space will be identified by detection of the vehicle i.e., if the model identifies a vehicle in the parking lot image, then it will be a parking space and these coordinates will be used to perform identification of free or empty lot. Once all the vehicles in the images are detected properly then that will provide the total parking space or total vehicles that can be parked in that parking space. When a new image frame is supplied again the same object detection will be performed to detect vehicles in the parking lot and these coordinates will be compared with the reference image coordinates that have been already identified by the Mask R-CNN to perform the intersection over union computation. This calculation will provide the probability of space in the parking lot that is occupied or empty. If the IOU value obtained based on the calculation is greater or less than 0.5 it will be an empty spot since no vehicle bounding box overlapping is found, it is shown in Figure 4.

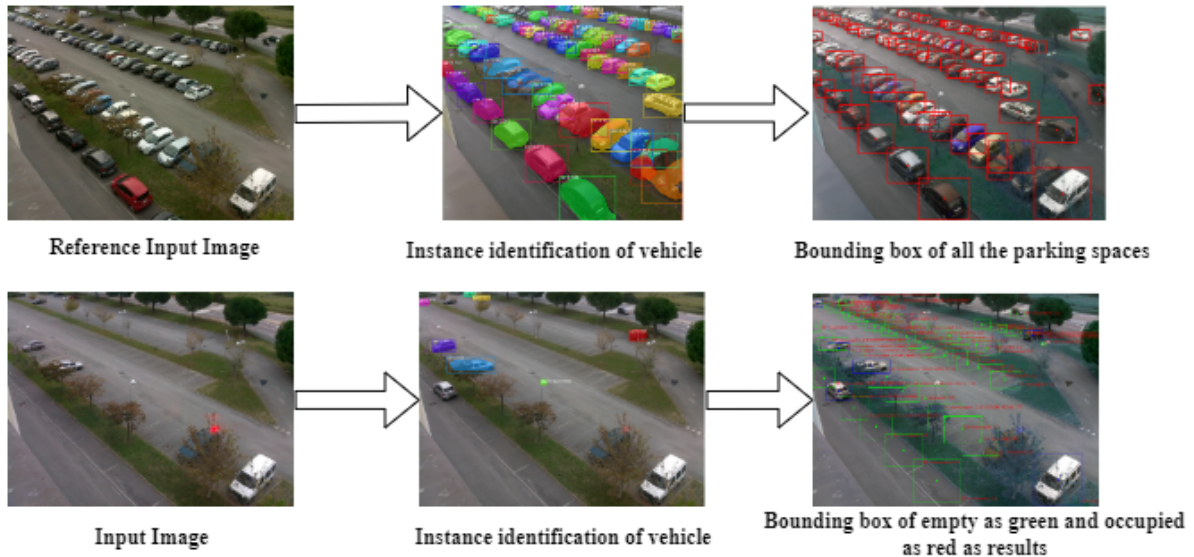


Figure 4: The Working concept of the algorithm.

5.1 Data Preprocessing and Preparation

Normally a Mask R-CNN can take input images of any standard size across the RGB channel for feature extraction, since all the images used for training, and testing the model and the images from the CNR Parkit dataset are in RGB format the requirement for preprocessing is been ignored. Only image annotation is been done for training the model on the training dataset as a part of data preparation. The dataset obtained from roboflow for performing detection of the car using Mask R-CNN has annotation which was done based on the car brand and model. Proper annotations are required where the annotations should be in Pascal VOC XML format. To perform these an image annotation tool has been used. All the images are renamed as a sequential whole number properly and annotated using a bounding box. The reason for using a bounding box is that the research concept mainly focuses on object detection of a single class and hence bounding boxes will be an easier concept on top of polygon selectors as the learning will be easy for the model. The generated annotations are stored in an XML format for each image. The whole annotated dataset for training the Mask R-CNN model is divided as train, test, and valid based on the ratio as stated in Table 1.

5.2 Object Detection Based on Model Training

The first part of the implementation phase is training the Mask R-CNN model to detect car class. As Mask R-CNN performs instance segmentation for identifying the objects in an image Ojha et al. (2021) with a greater accuracy using the CNN and RPN network and uses an ROI align to avoid feature loss to produce the pixel coordinates of the object along with an outline mask. To perform the object detection phase the model needs to be trained properly, for training the annotated dataset with proper training, testing, and validation have been used. The critical parameter which is the number of classes will be set to two since we are performing training based on a single object and the background will be an object hence car will be the object here and that makes the total class two. The mode parameter is set to training to indicate that training will be performed on the

model. When the model is loaded for training, there are more input layers than when it is loaded for inference alone. The additional layers include the input images and bounding boxes. The training will be performed for two batches for five epochs with pre-trained weights that have been obtained by performing model training on the COCO dataset over different classes and this has provided a better accuracy for the training. The final model weight which has the best loss function was saved using Keras save weight functionality to be used for the final phase of the implementation. The final trained weight was also evaluated against the test data to obtain the map score for object detection.

5.3 Parking Space Detection

The final stage of the research is the phase where parking space identification is performed based on the concept of bounding box overlap. This is where the object detection using the trained Mask R-CNN is used, a fully occupied parking lot image from a t second will be used as an implication image for identifying all the parking spots which can be occupied at a time. Using the Mask R-CNN model object detection is performed and the output results that are obtained from the model provide the measure of the degree of confidence for the object detection. It is the detection rate for a car, where a car detected in a parking lot means a parking space is occupied. The X/Y pixel coordinates of the bounding box for the object and a mask that is made of bits shows which of the pixels in the bounding box belong to the object and which are not. This output is utilized as an implication for identifying empty spaces.

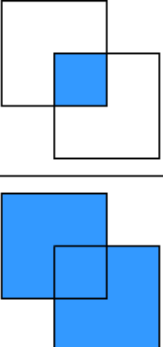
$$\text{IOU} = \frac{\text{Area of Intersection}}{\text{Area of Union}}$$


Figure 5: The formulae for IoU calculation.

When an image from t+1 second has been provided, then the same model performs object detection on the image to identify all the occupied space which is the detection of the car. To separate the spaces which are empty the concept of degree of overlap was performed. The degree of overlap between two objects is calculated using intersection over union (IoU) shown in Figure 5, and the pixel coordinates that are obtained by performing object detection on the implication are utilized as the foundation for assessing if that is an empty space. The `mrcnn.utils.compute.overlaps()` method in Opencv may be used to determine the IoU value. The IoU score is calculated based on newly identified pixel coordinates for the bounding box of the car detected by MaskR-CNN of the newly supplied image against the pixel coordinates for the bounding box of the car detected from the implication image. If IoU is closer to 1, the degree of overlap between the two boundaries exists, and when IoU is closer to 0, there is almost no overlap between the

two borders. The degree of overlap between the vehicle border and the parking space boundary is determined by the system model using the IoU. If the IoU is low, the parking spot indicates a large amount of available space which means there are empty spaces.

6 Evaluation

The complete implementation was done based on the design that the parking lot will be identified by the detection of non-moving cars. This concept requires object detection to identify all vehicles. Using the image of a fully occupied parking lot object detection will be carried out on the image in order to identify all the parking spaces in the parking lot. The Figure 6 shows an overview of the object detection performed for the identification of total parking spaces.



Figure 6: The overview of parking spot detection.

6.1 Case 1: Object Detection using Mask RCNN

For performing object detection the Mask RCNN model was trained on the vehicle image dataset using proper annotation. The training was performed with a learning rate of 0.001 for two batches, among the two batches the first batch was performed for 5 epochs with 2002 train images and 220 validation set images. Table 2 shows the map score obtained for the batches. The mean accuracy precision (map) score used here is the average of the area under the precision curve (AP). AP is calculated by the concept of IOU (intersection over union) shown in Figure 5 it compares the overlapping of the predicted bounding box of a class car with the actual bounding box of the same class. If the overlapping is above a threshold value which is set as 0.5, then the prediction is True positive, which means the prediction is a car. When overlapping is below the threshold it is a false positive prediction which shows that is not a car. Figure 7 shows the concept of overlapping of the bounding box which is used for the calculation of map score. The first image shows an IoU score of 0.781 where the predicted bounding box (green) overlaps almost the same as the actual bounding box (red). While the second image the predicted bounding box (green) has overlap compared to the actual bounding box (red) and hence the IoU score is almost 0.422. Once all the AP values have calculated the mean of the AP will be computed to produce the mean accuracy precision.



Figure 7: The overlapping of bounding box.

Table 2: The Mean Accuracy Precision of the model.

Case	Image Count	Epoch	Map Score
Train	2002	5	0.905
Test	395	5	0.919

6.1.1 Hyper Parameter Optimization

As a part of Hyperparameter optimization, the backbone CNN models are ResNet101 and ResNet50, the ResNet101 was used even though it has a slower training time as the accuracy in the feature extraction is higher compared to ResNet50. In order to reduce the false positive results from the prediction the detection confidence interval was maintained at 0.6 as this will be a min score that shows that model will use predictions which are having an accurate score greater than 0.7. The Train ROIs Per Image which is the number of ROI for the Regional Proposal Network will be another hyperparameter that is set to 512 as the number of classes to be identified is limited to 1. Keeping the value of the Max GT instance to 256 has reduced the false positive prediction and also reduced the total training time of the model. After performing the hyperparameter there was a slight change in the mean average of the IoU values that shows the precision of the model Table 2 show the map value obtained. The best training weight was selected after optimization of the model for the detection of the vehicle from the parking lot images.

6.2 Case 2: Identification of Empty and Occupied Parking Space

The parking space detection accuracy was evaluated by performing a comparison between the total number of parking space that is predicted as occupied and empty in an image by the algorithm against the total number of parking space which is identified as occupied and empty in the same image based on a human observer. For performing this comparison random samples of images were gathered from the dataset and these images are labeled properly which shows the total occupied space and empty space. The occupancy detection performed over random images sample over sunny, rainy, and overcast climates is shown

in the Table 3.

Table 3: Actual vs Prediction results

Sample	Actual Occupied spots	Actual Empty spots	Predicted Occupied spots	Predicted Empty spots
Sunny				
Image 1	10	97	3	39
Image 2	50	57	19	23
Image 3	22	85	6	36
Rainy				
Image 1	17	94	13	58
Image 2	31	80	21	50
Image 3	8	103	6	65
Overcast				
Image 1	10	82	5	49
Image 2	11	81	10	44
Image 3	31	61	23	54

From the comparison of the predicted result with actual results, it was evident that the model is performing on an average scale even after hyperparameter optimization. But comparing it with the prediction scenario over a sunny climate it is seen that the model performance is below average and it must be because of the occlusion of an image that might have caused the dropping in the prediction results. Another problem with the performance being an average for the overall algorithm is that the Mask RCNN was showcasing a good accuracy score as the training dataset was smaller, but the detection accuracy got reduced when it was performing on a real-time feed due to image occlusion and lighting issues. This cause the downgrade of the overall performance of the algorithm. The overall accuracy was performing well over the three scenarios.

6.3 Discussion

This research paper is well-designed for the identification of occupied and free parking spaces in a parking lot. There is recent research that has been done in the field using different deep learning algorithms like YoloV3, Faster RCNN, and Mask RCNN. The reason for considering the Mask RCNN model over all the other well-maintained models was because of its object detection capability using instance segmentation which will provide the outline and pixel coordinates of the object it identifies. The computation time taken by the Mask RCNN model compared to another model was also lower. Based on the evaluation results the model was showing a better performance in object detection as it identified the class car properly. Mask RCNN is better for object detection compared to the Faster RCNN model, the Mask RCNN produces three outputs while the Faster RCNN produces two outputs. The third output is the mask that outlines the entire object to produce the pixel coordinates. While Yolo V3 (Benjdira et al.; 2019) is better than any Faster RCNN model based on its accuracy, the processing time will be the factor that gives an upper hand to Mask RCNN over it. The train and test map score calculated for the model showcase that the model has an accuracy of 91% when it was trained on the dataset containing more than 2000 images. The model was able to identify a good number of cars from a panoramic image even with occlusion. If training can be better with more data, multiple classes, and enhanced Mask RCNN (Wu et al.; 2022) then the model can achieve better accuracy by using selective network and backbone topology.

Based on this object detection accuracy the algorithm was able to perform the identification of total parking space within a parking lot from the image frame of the parking

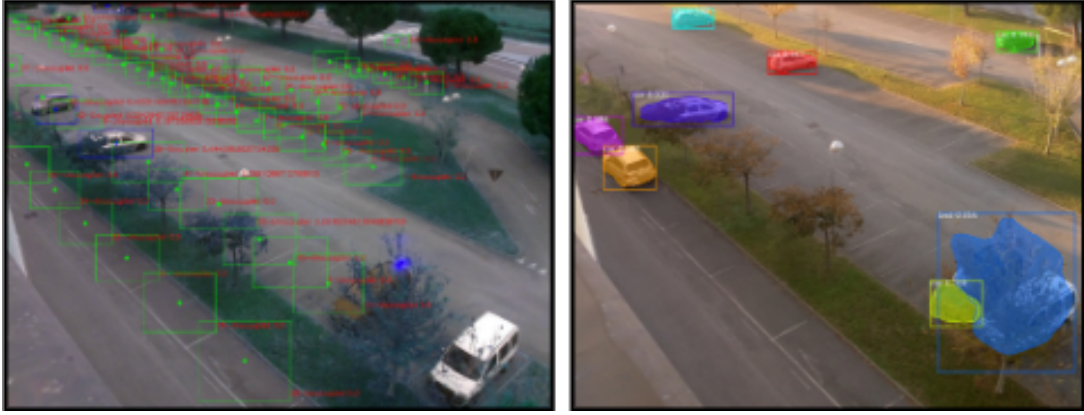


Figure 8: The final results.

lot. Vacant spaces are identified by comparing the bounding box of a current frame with the previous frame and the end results are shown in Table 3. Even though the evaluation test was done based on the comparison between the total empty and occupied spaces identified by the algorithm against a human observer. These results show that the algorithm can perform well in identifying vacant and occupied parking spots. The whole concept really depends on the object detection accuracy of the Mask RCNN model. The algorithm was able to perform well on an average scale for the detection of an empty and occupied parking spot. This research shows a novel approach when compared to the (Gkolias and Vlahogianni; 2019) by replacing a CNN that uses a classification approach with an object detection approach using instance segmentation makes this research more dynamic. As this research approach does not require separate training of the model based on any parking lot images and this makes the research to be implemented in any environment. The key observation that has been found while carrying out the experiments are :

- A single fed image which is having a panoramic view is fed to the model, due to occlusion caused by overlapping of another object like a tree or any parking sign in front of the car will not make the detection of car, as the Mask R-CNN model was only trained to identify the object car. This causes the loss of a parking space count.
- Image with low lighting is another drawback in the entire algorithm as the Mask R-CNN model could not showcase a good detection accuracy on low-lit images and it causes a reduced performance of the algorithm to identify a parking space.
- Major drawback identified with the algorithm was that the images used in the dataset have a parking lot that is close to the road which has the motion of the car. The model just identifies the cars which were on the road for a particular time frame, and it performs the analysis for the vacant and occupied spot for that identified car as well. This causes the algorithm to showcase a vacant spot in the next frame shown in Figure 8.
- The empty and occupancy detection varies based on different climate changes due to exposure to sunlight. The detection is more accurate in rainy climates compared

to the other two as the light exposure is on average which will not make too much distortion in the images.

Another approach has been used to change the weight of the Mask RCNN model. Instead of using the trained weight, a pre-trained weight was used which was the weight obtained from training the Mask R-CNN model on the COCO dataset. This weight provides multi-class detection of the object and hence they showed good accuracy in the prediction of the car using these prediction results the algorithm was able to identify the empty and filled parking spaces. The detection results were almost similar to the previous approach even though the weight used had higher training accuracy. This was caused as the pre-trained weight can be used to detect multiple classes the occlusion in the image causing the identification of the wrong class based on overlapping caused by occlusion in the image. The algorithm has shown a good performance for the given set of images for the identification of empty and vacant parking spaces. The Mask RCNN model showed a map score of 0.905 for the training dataset and a map score of 0.919 for the test dataset.

7 Conclusion and Future Work

To conclude the research which is using deep learning networks and techniques to develop a dynamic algorithm for indoor and outdoor parking space identification problems was able to implement successfully using a Mask RCNN algorithm which uses instance segmentation for the detection of the car in parking lot. Using the technique of intersection over union the algorithm was able to detect empty and occupied parking spaces. The object detection Mask RCNN was able to perform object detection with an accuracy of 0.91% after performing hyperparameter optimization on a car detection image dataset that consists of 2000 images. Using the concept of bounding box overlap the algorithm was able to use the IoU calculation based on the detected car by the Mask RCNN model to identify the occupied and empty parking space. The algorithm prediction was compared for the parking lot images which are labeled occupied and empty to showcase that the algorithm was able to perform well based on the detection count of empty and occupied parking spaces. There are some limitations in the concept that can be caused by the detection mismatch of results that includes occlusion in the image that blocks the visibility of an object from the model or overlapping of an object over an object also causing missing out of the object. Even after hyperparameter optimization, the algorithm was able to perform to an average only it shows that further explorations and changes are required for this implementation to fully incorporate into real-world scenarios but the algorithm was dynamic and can be implemented over any environment.

This work can be further developed by using a more advanced object detection algorithm like Yolo V7 which is the latest object detection algorithm that can detect objects more accurately and faster. By using an image enhancement framework to overcome the image quality issue which causes mislabelling of the detected object by the model. Instead of using an image, processing a live video feed can help to constantly monitor any parking space to identify the probability of that space being empty or occupied.

References

- Acharya, D., Yan, W. and Khoshelham, K. (2018). Real-time image-based parking occupancy detection using deep learning., *Research@ Locate* **4**: 33–40.
- Al-Ariny, Z., Abdelwahab, M. A., Fakhry, M. and Hasaneen, E.-S. (2020). An efficient vehicle counting method using mask r-cnn, *2020 International Conference on Innovative Trends in Communication and Computer Engineering (ITCE)*, pp. 232–237.
- Amato, G., Carrara, F., Falchi, F., Gennaro, C., Meghini, C. and Vairo, C. (2017). Deep learning for decentralized parking lot occupancy detection, *Expert Systems with Applications* **72**: 327–334.
- Amato, G., Carrara, F., Falchi, F., Gennaro, C. and Vairo, C. (2016). Car parking occupancy detection using smart camera networks and deep learning, *2016 IEEE Symposium on Computers and Communication (ISCC)*, pp. 1212–1217.
- Benjdira, B., Khursheed, T., Koubaa, A., Ammar, A. and Ouni, K. (2019). Car detection using unmanned aerial vehicles: Comparison between faster r-cnn and yolov3, *2019 1st International Conference on Unmanned Vehicle Systems-Oman (UVS)*, pp. 1–6.
- Chen, M., Bai, F. and Gerile, Z. (2021). Special object detection based on mask rcnn, *2021 17th International Conference on Computational Intelligence and Security (CIS)*, pp. 128–132.
- Chen, M. and Chang, T. (2011). A parking guidance and information system based on wireless sensor network, pp. 601–605.
- Choeychuen, K. (2013). Automatic parking lot mapping for available parking space detection, *2013 5th International Conference on Knowledge and Smart Technology (KST)*, pp. 117–121.
- Gkolias, K. and Vlahogianni, E. I. (2019). Convolutional neural networks for on-street parking space detection in urban networks, *IEEE Transactions on Intelligent Transportation Systems* **20**(12): 4318–4327.
- Holdsworth, K., Taylor, D. and Pullman, R. (2001). On combined equalization and decoding of multilevel coded modulation, *IEEE Transactions on Communications* **49**(6): 943–947.
- Kergoat, R., Labrune, M., Miltat, J., Valet, T. and Jacquet, J. (1993). Initial rotational susceptibility of exchange coupled permalloy/copper/cobalt multilayers, *1993 IEEE International Magnetics Conference (INTERMAG)*, pp. AC–AC.
- Khalfi, A. and Guerroumi, M. (2021). *Roadside Parking Spaces Image Classification Using Deep Learning*, pp. 323–333.
- Malbog, M. A. (2019). Mask r-cnn for pedestrian crosswalk detection and instance segmentation, *2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS)*, pp. 1–5.

- Manase, D. K., Zainuddin, Z., Syarif, S. and Jaya, A. K. (2020). Car detection in road-side parking for smart parking system based on image processing, *2020 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, pp. 194–198.
- Martín Nieto, R., García-Martín, , Hauptmann, A. G. and Martínez, J. M. (2019). Automatic vacant parking places management system using multicamera vehicle detection, *IEEE Transactions on Intelligent Transportation Systems* **20**(3): 1069–1080.
- MXK (2022). Car model detection dataset, <https://universe.roboflow.com/mxk/car-model-detection>. visited on 2022-12-14.
URL: <https://universe.roboflow.com/mxk/car-model-detection>
- Nyambal, J. and Klein, R. (2017). Automated parking space detection using convolutional neural networks, *2017 Pattern Recognition Association of South Africa and Robotics and Mechatronics (PRASA-RobMech)*, pp. 1–6.
- Ojha, A., Sahu, S. P. and Dewangan, D. K. (2021). Vehicle detection through instance segmentation using mask r-cnn for intelligent vehicle system, *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 954–959.
- Suhr, J. K. and Jung, H. G. (2016). Automatic parking space detection and tracking for underground and indoor environments, *IEEE Transactions on Industrial Electronics* **63**(9): 5687–5698.
- Wu, Y., Sun, Y., Jia, Y. and Liao, F. (2022). Parking-lot vehicles detection from a low-angle camera perspective based on improved mask r-cnn, *2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML)*, pp. 571–575.