
YOLO-Vacant Parking Spot Detector

Andrew Park

Electrical and Computer Engineering
A17406465

Minsang Kim

Electrical and Computer Engineering
A16636382

Abstract

Not only within UCSD campus, there are numerous issues related to finding parking spots nearby. Therefore, the technology to annotate the available parking spot at a specific parking lot is demanded. To implement the artificial intelligence model, CNN is implemented to annotate the available parking spot. During implementation, we used the dataset for image-based parking space occupancy to collect data for parking structure to train the architecture. We will display the resulting image, the overall accuracy, and the loss of the model. The model can accurately detect the parking spot and determine whether the spot is empty or not.

1 Introduction

Finding a vacant parking space in a given lot is a tedious and costly task: a person must drive through the whole parking lot in order to find a vacant space. This ineffective process will waste gasoline and time while searching for a parking space. Furthermore, this will cause congestion in the parking lot. Therefore, by displaying the occupancy of spaces in a given parking lot will significantly reduce the time cost, traffic congestion, and gasoline consumption. YOLO-Parking will aid this wasteful process by displaying vacant and occupied parking spaces. Although there are existing methods to detect the occupancy of a parking space, such as car detector on every parking space and other object detector architectures that detect the occupancy, most of them are expensive or slow in detecting parking spots [2]. Furthermore, there are several traditional models to detect the instance segmentation of a vacant parking spot, but they are slow in detection for videos.

In this paper, we propose a new model to detect vacant parking spots: You Only Look Once (YOLO). YOLO architecture is one of the fastest architectures that accurately performs object detection [3]. Additionally, several cameras mounted in the parking lot to detect vacancy of a space is relatively cheaper than placing a car detector on every parking space in a parking lot [2].

2 Related Work

R-CNN R-CNN [2] is a widely used object detecting CNN model with relatively high accuracy. The first step is to regional proposal using selective search methods to propose bounding box. Then, for each region proposal, the model warped the proposed regions to use them as inputs for a convolutional neural network. The CNN classifies each input proposed region using SVM specific to a class [1]. However, the training rate is slow due to large number of parameters to learn, and the inference rate is also slow. In one of the previous approaches in detecting parking spots, Martin Marek proposed R-CNN: the model first pool proposed parking spot patches in the original image and then a CNN produces a binary output whether the space in each patch is occupied or not [2]. The authors provided two methods for pooling for the first step in R-CNN: the interpolation of pixels to construct quadrilateral bounding boxes and construction of minimal bounding boxes [2].

Faster R-CNN FPN The input image first undergoes a feature pyramid network (FPN). Then, the authors extract the features of the parking spots and pass them through respective classification head for classification. However, the input image must be resized, thus limiting the output image from

being full resolution [2]. Both R-CNN and R-CNN is still slow in inference for object detection in videos. Faster R-CNN can use one of the two methods of pooling mentioned above [2].

These two models are merely the baseline of parking vacancy object detector using CNN. In this paper, we intend to improve the model in inference time by using YOLO model.

3 Method

3.1 Model

We began the project with instance segmentation method to determine the pixels of the parking spots. However, it was extremely difficult to find the appropriate dataset for instance segmentation of parking lots. Furthermore, it was not necessary to classify each pixel to detect vacant space. Rather, we merely wanted to count the number of vacant spaces available in the given parking structure and relative spot of vacancy in a given parking lot. This is the reason why we chose neural network models that are capable of object detection.

We decided to use the popular model proposed by Joseph et al. [3] You Only Look Once because of its speed and accuracy. Compared to the speed of the YOLO architecture in object detecting, the model itself is surprisingly simple. Given an image, the model first resizes the image to a resolution of 448×448 . The resized image is divided into $S \times S$ grid and is passed through convolutional neural network. The model consists of 24 convolutional layers with two fully connected layers following the convolutional neural network. During the process of convolutional network, the model simultaneously draw bounding boxes and predicts the probability that each grid contains a class. The prediction of the model is an $S \times S \times (B * 5 + C)$ tensor. In our case, we are evaluating whether the space is occupied or not, so the number of classes $C = 2$. The size of the grid for division $S = 7$ and the number of bounding boxes is B . Each prediction for bounding box consists of 5 values: center (x, y) , height, width, the confidence value. Since there are $B = 2$ bounding boxes, the prediction tensor contains $2 * B = 10$ values. The other $C = 2$ values are the conditional probability that the given object is the i -th class: $P(C_i | Object)$. The confidence probability is calculated by the IOU of each bounding box [3].

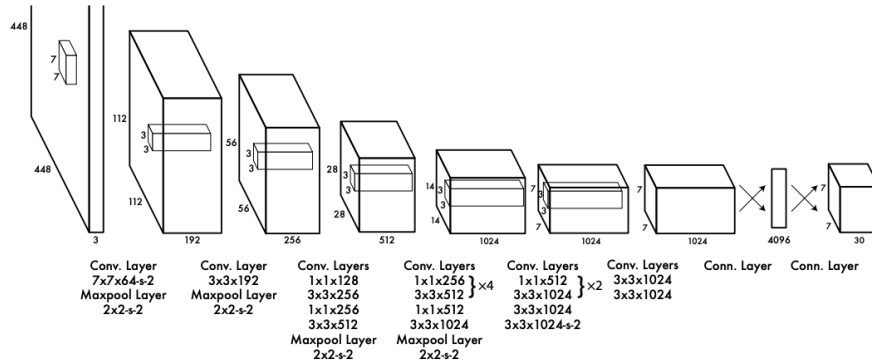


Figure 1: YOLO Architecture

To reduce the training time, we collected pretrained parameter in the backbone of the YOLO architecture. This is a unconventional architecture for vacant parking spot detection. This architecture is faster than other conventional model, such as R-CNN and Faster R-CNN [3]. Therefore, this model can be used in video cameras mounted in the parking lot to detect parking spaces in live footages.

- training algorithm, testing algorithm.
- what is the new proposed techniques compared to previous work, and the reason and strength of choosing the method.

4 Experiments

4.1 Dataset

For the dataset, Parking Space Image Dataset is used. The dataset consists of parking lot images and

4.2 Results

The following figures are the results after training the model with a learning rate of 0.001 and with Adam optimizer with 5 epoch:

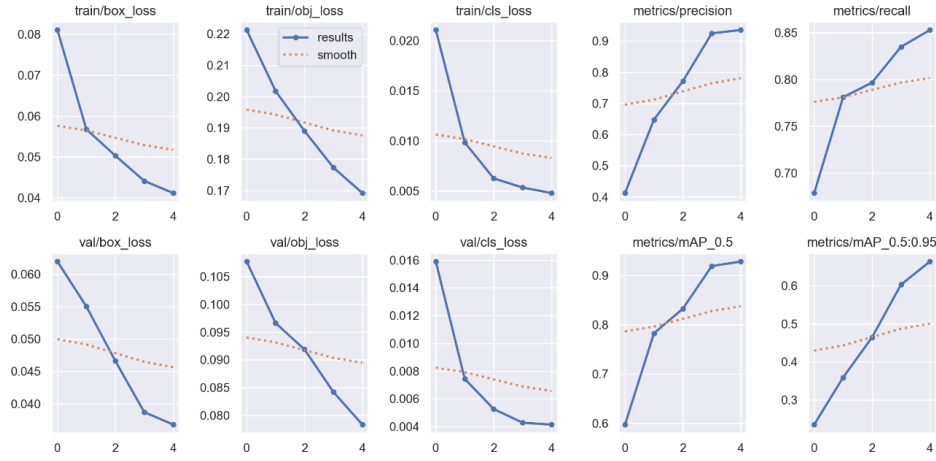


Figure 2: Loss Function Over Epochs

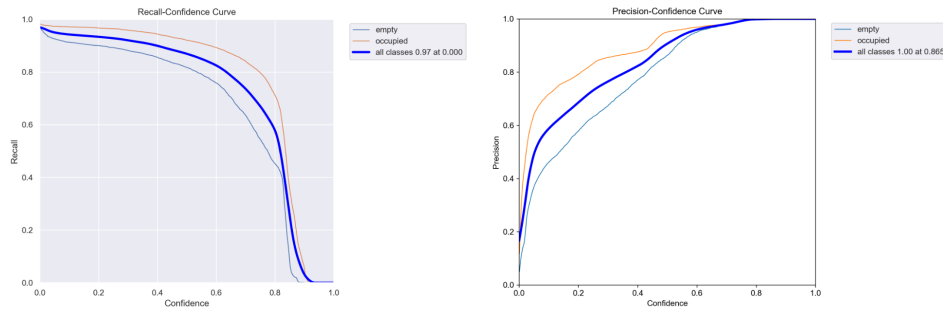


Figure 3: Recall-Confidence Graph (left) and Confidence-Precision Graph (right)

The inference image is shown below using the same model.



Figure 4: Inference Image

References

- [1] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation, 2014.
- [2] Martin Marek. Image-based parking space occupancy classification: Dataset and baseline. *CoRR*, abs/2107.12207, 2021.
- [3] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection, 2016.