Journal of Physics: Conference Series

PAPER • **OPEN ACCESS**

Vehicle and Parking Space Detection Based on Improved YOLO Network Model

To cite this article: Xiangwu Ding and Ruidi Yang 2019 *J. Phys.: Conf. Ser.* **1325** 012084

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

This content was downloaded from IP address 184.174.59.144 on 07/11/2019 at 13:08

Vehicle and Parking Space Detection Based on Improved YOLO Network Model

Xiangwu Ding[1], Ruidi Yang[1*]

[1] Donghua University, shanghai, shanghai, 201620, China

[*]ruidyyang@163.com

Abstract. YOLO has a fast detection speed and is suitable for object detection in real-time environment. This paper is based on YOLO v3 network and applied to parking spaces and vehicle detection in parking lots. Based on YOLO v3, this paper adds a residual structure to extract deep vehicle parking space features, and uses four different scale feature maps for object detection, so that deep networks can extract more fine-grained features. Experiment results show that this method can improve the detection accuracy of vehicle and parking space, while reducing the missed detection rate.

1. Introduction

With the increasing number of cars, the number of parking lots is also increasing. In many parking lots, ground sensors are used to determine the state of various spaces. Traditional parking space detection methods are based on ultrasonic [1], geomagnetic [2], infrared ray [3-4]. This requires the installation and maintenance of sensors in each parking space, especially in parking lots with a large amount of available space, which may be expensive. Although this method can obtain higher accuracy, it is expensive. In order to reduce the cost, this paper proposes a object detection model based on YOLO v3, which can improve the accuracy of detection and achieve real-time detection without losing the detection speed.

The paper is organized as follows: Section II introduces the related work in detail. Section III discusses the model method proposed in this paper. Section IV introduces the data sources and the experimental results of this paper. And section V summarizes the research results of this paper and the planning of future research work.

2. Related Work

Object detection is one of the basic tasks in computer vision, which refers to the determination of the existence of specific features in images [5].Teoh and Bräunl[6] use a multi-size symmetric search window to extract each symmetric region, which is then validated by the Adaboost classifier.Teutsch and Kruger [7] use sliding windows to find candidate vehicles from airborne cameras, and then use AdaBoost classifier based on motion features to verify them.

In recent years, target detection methods based on deep learning have gradually gained importance. K.Simonyan and A.Zisserman [8] proposed a very deep CNN convolutional network to construct a VGG network for object classification. RCNN [9] (Region Proposal-Convolutional Neural Network) is a target detection method combining Region Proposal and Convolutional Networks. This is the first time that deep learning has been applied to traditional target detection tasks. S. Ren et al. [10] proposed Faster R-CNN which is more effective than RCNN. Faster R-CNN abandoned the selective search in RCNN

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

and introduced the RPN network, which enabled the region proposal, classification, and regression to share common convolution features, thus further speeding up the detection. However, the Faster RCNN is still divided into two steps: first, determine whether there is a target for the area frame, and then identify the target. YOLO [11] combines target discrimination and target recognition in one, further improving the speed of detection. YOLO v3 [12] uses multi-scale features to detect objects, and achieves 57.9% effect of mAP in 51 ms time on VOC data sets. It can be seen that YOLO v3 can guarantee both accuracy and detection rate in the field of target detection, and achieve better detection results.

In the management of parking lots, besides vehicle detection, it is also a challenging task to identify whether parking spaces are vacant or not. M. Ahrnbom et al. [13] extracts features such as color and gradient size in LUV space, and then trains a SVM-based classifier to classify parking lots into vacant or occupied states. Giuseppe Amato et al. [14] used deep CNN (convolutional neural network) to train detectors to detect parking spaces and their states based on LBP features. Tom Thomas et al. [15] constructed a convolution neural network of binary classifier to judge whether parking lot is occupied or not. Cheng-Fang Peng [16] extracted three new features for each parking space for occupancy state judgment, namely vehicle color characteristics, local gray-scale variation features and corner features. The deep neural network is then trained to determine the occupancy status of each parking space based on the above three characteristics. The system [14] established in periodically captures an image of a portion of the parking lot, and for each parking space, the occupancy status is determined by using the trained CNN. The pictures captured by the camera are filtered through a mask that identifies various parking spaces. But the mask is built manually by people. That is, you need to manually create a mask for each parking space in different parking lots.

3. Model Method

3.1. Yolo v3 network structure

YOLO V3 uses the network structure of Darknet53 and contains 53 convolution layers. Darknet53 consists of five residual blocks, drawing on the idea of the Resnet neural network [12]. Each residual block consists of multiple residual units, and a residual unit is constructed by inputting residual operations with two DBL units. Among them, the DBL unit contains convolution, batch normalization and leaky relu activation functions. By introducing a residual unit, the depth of the network can be deeper and avoid vanishing gradient. The structure diagram is shown in Figure 1.

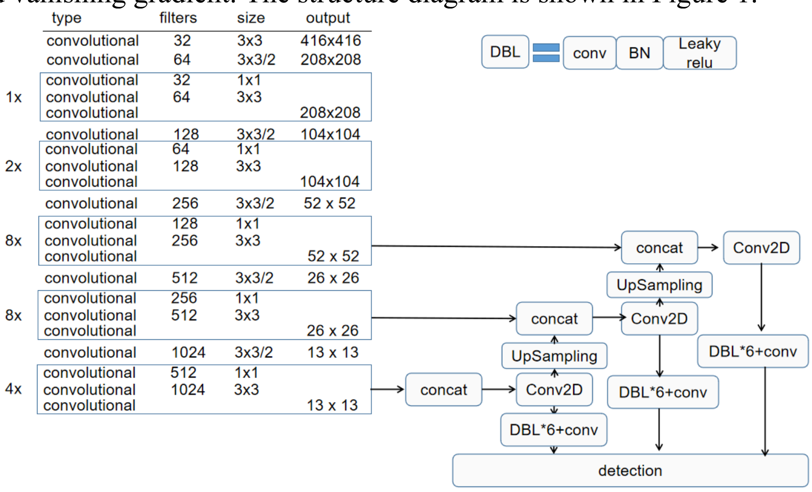


Figure 1. YOLO v3 network structure.

3.2. Improved network structure

Based on YOLO v3, this paper adds a residual blocks to extract deep vehicle parking space features and uses four different scale feature maps for object detection. This allows deeper networks to extract more granular features. The YOLO v3 network uses 6 DBL units and a 1x1 convolution in the target detection output layer. In order to avoid the gradient disappearing from the reuse of enhanced features, this paper turns 6 DBL units into 2 DBL units and 2 resnet units. The improved network structure is shown in Figure 2.

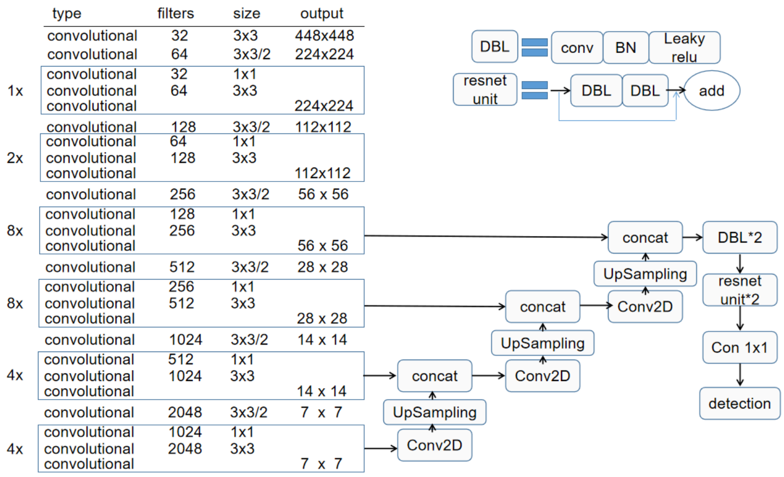


Figure 2. The improved network structure.

3.3. Data Set

The current research dataset comes from three datasets, the PASCAL VOC dataset, the COCO dataset, and the PKLot dataset. The PASCAL VOC and COCO data sets are common data sets for object detection, with 20 categories and 80 categories respectively. This experiment only uses the three data sets car, bus, and truck. The PKLot dataset [17] provides an image dataset for parking space classification. The data used in this paper is a picture of the parking lot of the Pontifical Catholic University of Parana (PUCPR) in Brazil. The PKLot dataset PUCPR is included in different days. Images in different lighting conditions, including sunny, rainy, and cloudy conditions. It provides the original image and the corresponding annotation file. The three weathers totaled 4,473 pictures. Summary of the PUCPR subsets characteristics as shown in Table 1.

Table 1. Summary of the PUCPR subsets characteristics.

Parking lot	Weather Condition	#Of days	#Of images	# Of parking spaces		
				Occupied	Empty	Total
PUCPR (100 parking spaces)	Sunny	24	2315	96,762 (46.42%)	111,672 (53.58%)	208,433
	Overcast	11	1328	42,363 (31.90%)	90,417 (68.10%)	132,780
	Rainy	8	830	55037 (45.77%)	27917 (51.46%)	82,954
	Subtotal		4473	194,162	230,006	424,168

3.4. Data Set Processing

The PKLot dataset is not the format of the PASCAL VOC dataset. The processing first extracts the coordinate information of the xml file and converts the information into the format of the VOC dataset.

Although YOLOv3 has achieved good detection results in the daily detection data set in the target detection field, it needs some improvement for the parking lot data set to adapt to the parking lot detection task. The improved YOLO v3 algorithm model is based on the feature pyramid network structure, and the feature maps of different levels are merged and connected to obtain four sets of predicted feature maps, and the position and class predictions are performed on the four sets of predicted feature maps.

3.5. Clustering Analysis of anchor boxes

This paper mainly studies the problem of parking space detection. The anchor box defined by the original network is not suitable for the object studied in this paper. Therefore, the K-means clustering algorithm is used to cluster and classify car parking spaces in the dataset.

YOLO v3 introduces the idea of anchor boxes used in Faster RCNN. anchor boxes are a set of initial candidate boxes with fixed width and height. The selection of initial anchor boxes will directly affect the detection accuracy and speed of the network. YOLO v3 algorithm gets the priori box dimension which is trained on COCO data set, and divides its parameters into three different scales. For the specific task of parking detection, it is necessary to cluster the specific data set and get the corresponding clustering center. The improved algorithm also uses K-means clustering algorithm to cluster the width of target box in PKLot dataset. When the number of clusters is 4, the width and height of the corresponding clustering centers in the PKLot dataset are (25,30), (30,48), (42,62), (58,74).

3.6. Network Training

We use the automotive data on PASCAL VOC and COCO datasets to pre-train the network. As the network trained in the previous one is a universal physical detector, we use the PKLot data set to fine-tune the network to detect vehicles and parking spaces in parking scenes.

The network parameters are fine-tuned by using the pictures of the training set in the PKLot dataset, so that the detection effect of the whole network is optimized. Some experimental parameters are set as shown in Table 2.

Table 2. Description of network parameters.

Parameter Name	Parameter Value
Learning Rate	0.005
Epoch	50
Batch size	32
Momentum	0.9
Weight Decay	0.0005

4. Experiment Results

In this paper, the corresponding experiments are carried out under the system of Ubuntu18.04, CUDA10.0, CUDNN7.5.1, NVIDIA GTX 2080Ti. The test was conducted using the PKLot dataset test set, using the precision and recall as indicators. FP (False Positive) indicates the number of negative samples that are incorrectly marked as positive samples, and FN (False Negative) indicates the number of positive samples that are incorrectly marked as negative samples, calculated according to formula (1) and formula (2). Recall and precision.

$$\text{Recall} = \frac{TP}{\text{TP} + \text{FN}} \tag{1}$$

$$\text{Presicion} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

This paper compares YOLO v3 with the improved network model in the test set of PKLot dataset, and tests the accuracy and recall rate of the model. The experimental results are shown in Table 3.

Figure 3. Shows some experimental results.

Table 3. Comparison of the results of the two algorithms on the test set.

	YOLO	ours
Precision	91.6%	93.3%
Recall	87.2%	90.9%

Both models were pre-trained on the automotive dataset and then finetuned 50 epoch on the PKLot dataset. The model val_loss values are compared as shown in Figures 2 and 3 (Figure 3. is YOLOv3 and Figure 4. is the improved model):

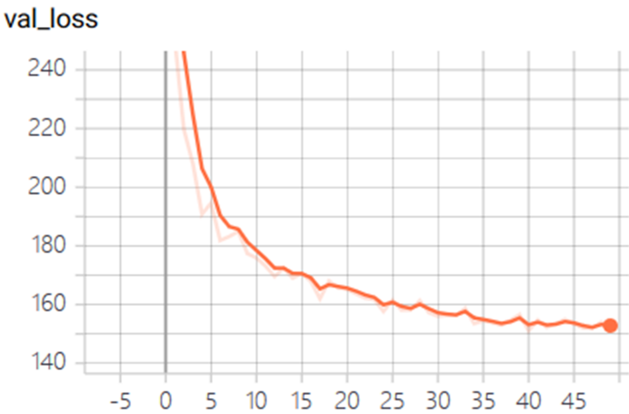


Figure 4. The val_loss of the YOLOv3 model on the PKLot test data set.

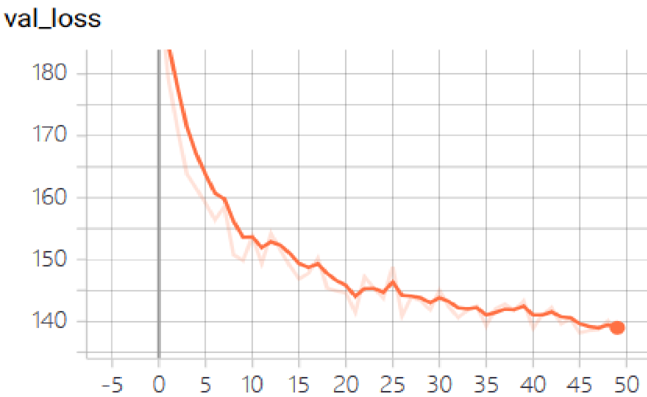
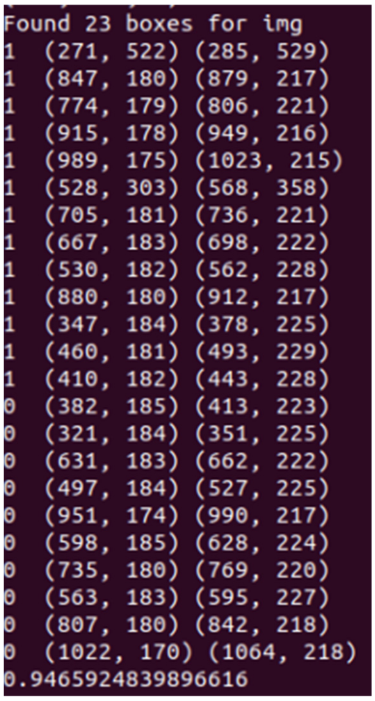


Figure 5. The val_loss of the ours model on the PKLot test data set.



```
Found 23 boxes for img
1  (271, 522) (285, 529)
1  (847, 180) (879, 217)
1  (774, 179) (806, 221)
1  (915, 178) (949, 216)
1  (989, 175) (1023, 215)
1  (528, 303) (568, 358)
1  (705, 181) (736, 221)
1  (667, 183) (698, 222)
1  (530, 182) (562, 228)
1  (880, 180) (912, 217)
1  (347, 184) (378, 225)
1  (460, 181) (493, 229)
1  (410, 182) (443, 228)
0  (382, 185) (413, 223)
0  (321, 184) (351, 225)
0  (631, 183) (662, 222)
0  (497, 184) (527, 225)
0  (951, 174) (990, 217)
0  (598, 185) (628, 224)
0  (735, 180) (769, 220)
0  (563, 183) (595, 227)
0  (807, 180) (842, 218)
0  (1022, 170) (1064, 218)
0.9465924839896616
```

Figure 6. The detection time.

The detection time used for the improved model is shown in Figure 5. In this figure, 23 parking spaces and vehicles were detected. The detection time used was 0.947 seconds, and the average detection time for a target is 43 milliseconds.

Journal of Physics: Conference Series **1325** (2019) 012084 doi:10.1088/1742-6596/1325/1/012084

5. Conclusion and Future Work

This paper proposes a vehicle and parking space detection method based on the improved YOLO v3 algorithm. The improved algorithm is applied to vehicle and parking space detection and it has achieved good results. Based on YOLO v3, this paper adds a Residual structure to extract deep vehicle parking space features and uses four different scale feature maps for object detection. This allows deeper networks to extract more granular features. Experiments show that the algorithm improves the accuracy of vehicle and parking space detection in the parking lot. However, under the influence of factors such as illumination and weather, it has certain influence on the detection effect of the algorithm, and it is necessary to further improve the algorithm. With the rise of edge computing, the network structure needs to be reduced. Applying the algorithm to a camera with a Raspberry Pi which is also the key problem to be solved by the author in the following research.

References

[1] Shao, Y., Chen, P. and Cao, T. (2018) A Grid Projection Method Based on Ultrasonic Sensor for Parking Space Detection. In: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. Valencia. pp. 3378-3381.
[2] Zhou, F. and Li, Q. (2014) Parking Guidance System Based on ZigBee and Geomagnetic Sensor Technology. In: 2014 13th International Symposium on Distributed Computing and Applications to Business, Engineering and Science. Xian Ning. pp. 268-271.
[3] Chen, H., Huang, C. and Lu, K. (2017) Design of a non-processor OBU device for parking system based on infrared communication. In: 2017 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-TW). Taipei. pp. 297-298.
[4] Yuan, C. and Qian, L. (2017) Design of intelligent parking lot system based on wireless network. In: 2017 29th Chinese Control And Decision Conference (CCDC). Chongqing. pp. 3596-3601.
[5] Dhawad, S.R., Itkarkar, R.R. (2016) Car Detecting Method using high Resolution images. International Journal on Recent and Innovation Trends in Computing and Communication, 4:197-203.
[6] Teoh, S., Bräunl, T. (2012) Symmetry-based monocular vehicle detection system, Machine Vision and Applications, 23: 831-842.
[7] Teutsch, M., Krüger, W. (2015) Robust and fast detection of moving vehicles in aerial videos using sliding windows. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston. pp. 26-34.
[8] Simonyan, K., Zisserman, A. (2014) Very deep convolutional networks for large-scale image recognition. https://arxiv.org/abs/1409.1556.
[9] R. Girshick, J. Donahue, T. Darrell, J. Malik. (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: Computer Vision and Pattern Recognition (CVPR). Columbus. pp. 580-587.
[10] Ren S., He, K., Girshick, R. and Sun, J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39: 1137-1149.
[11] Redmon, J., Divvala, S., Girshick, R. and A. Farhadi. (2016) You Only Look Once: Unified, Real-Time Object Detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas. pp. 779-788.
[12] Redmon, J., Farhadi, A. (2018) YOLO V3: An Incremental Improvement. IEEE Conference on Computer Vision and Pattern Recognition, 2018:89-95.
[13] Ahrnbom, M., Åström, K. and Nilsson, M. (2016) Fast Classification of Empty and Occupied Parking Spaces Using Integral Channel Features. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Las Vegas. pp. 1609-1615.
[14] Amato, G., Carrara, F., Falchi, F., Gennaro, C. and Vairo, C. (2016) Car parking occupancy detection using smart camera networks and Deep Learning. In: 2016 IEEE Symposium on Computers and Communication (ISCC). Messina. pp. 1212-1217.

[15] Thomas, T. and Bhatt, T. (2018) Smart Car Parking System Using Convolutional Neural Network. In: 2018 International Conference on Inventive Research in Computing Applications (ICIRCA). Coimbatore. pp. 172-174.
[16] Peng, C., Hsieh, J., Leu, S. and Chuang, C. (2018) Drone-Based Vacant Parking Space Detection. In: 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA). Krakow. pp. 618-622.
[17] de Almeida, P. R., Oliveira, L. S., Britto, A.S., Silva, E.J., Koerich, A. L. (2015) PKLot - a robust dataset for parking lot classification. Expert Systems with Applications, 42: 4937-4949.