YOLO5PKLot: A Parking Lot Detection Network Based on Improved YOLOv5 for Smart Parking Management System

Duy-Linh Nguyen[0000−0001−6184−4133], Xuan-Thuy Vo[0000−0002−7411−0697], Adri Priadana[0000−0002−1553−7631], and Kang-Hyun Jo[0000−0002−4937−7082]

Department of Electrical, Electronic and Computer Engineering, University of Ulsan, Ulsan 44610, South Korea
ndlinh301@mail.ulsan.ac.kr, xthuy@islab.ulsan.ac.kr, priadana@mail.ulsan.ac.kr, acejo@ulsan.ac.kr

Abstract. In recent years, the YOLOv5 network architecture has demonstrated excellence in real-time object detection. For the purpose of applying in the smart parking management system, this paper proposes a network based on the improved YOLOv5, named YOLO5PKLot. This network focus on redesigning the backbone network with a combination of the lightweight Ghost Bottleneck and Spatial Pyramid Pooling architectures. In addition, this work also resizes the anchors and adds a detection head to optimize parking detection. The proposed network is trained and evaluated on the Parking Lot dataset. As a result, YOLO5PKLot achieved 99.6% mAP on the valuation set with only fewer network parameters and computational complexity than others.

Keywords: Convolutional neural network (CNN) · Ghost Bottleneck · Smart parking management system · Parking lot detection · Parking lot dataset · YOLOv5.

1 Introduction

Currently, there are about 1.45 billion cars in the world and it is increasing every year. The report in [18] predicts that in 2023 the number of vehicles sold is about 71 million units. The rapid increase both in the number and type of vehicles has led to the expansion of parking lots in supermarkets, shopping malls, city offices, etc. Automated operations to manage and distribute parking spaces are essential. For a long time ago, researchers and engineers have been designing automated parking management systems. These techniques are mainly based on various types of sensors to determine the status of parking spaces such as ultrasonic [19], infrared [5], geomagnetic [21], and wireless [20]. This type of parking usually requires the installation and maintenance of each sensor per parking space. Therefore, these methods increase the cost quite a lot when deployed in large-scale parking lots. In general, the sensing methods achieve high prediction but have a large cost. From the above analysis along with the development of the computer vision field, this paper proposes a vision-based parking lot detection network. This work improves the famous object detection network YOLOv5 by focusing on redesigning the backbone network and adding a new detection head to increase the object detection ability. This network uses lightweight architectures in Ghost Bottleneck (Ghost) to greatly reduce network parameters and computational complexity, serving real-time applications on low-computing devices. The paper provides several main contributions as follows:

1 - A modified Ghost Bottleneck block is proposed to apply to the backbone of YOLOv5.

2 - Redesigns YOLOv5 backbone network with a combination of lightweight Ghost Bottleneck and Spatial Pyramid Pooling (SPP) architectures.

3 - Adds a detection head with new anchor sets to improve the prediction task.

The remainder of the paper is distributed as follows: Section 2 introduces the techniques related to parking lot detection. Section 3 details the proposed techniques. Section 4 presents and analyzes the experimental results. Section 5 concludes the issue and future development orientation.

2 Related work

2.1 Traditional-based method

These methods are implemented through two main steps, feature extraction, and parking classification. The feature extraction process generates one or more feature vectors using traditional techniques. Specifically, [1, 8] used Local Phase Quantization (LPQ) and Local Binary Patterns (LBP) as feature extractors and classifiers using Support Vector Machine (SVM). Later, the authors developed new methods based on the change of camera and parking areas [3, 2]. [4] applies Quaternionic Local Ranking Binary Pattern (QLRBP) for feature extraction, Support Vector Machine (SVM), and k-nearest neighbors (k-NN) are used for classification. In [10] the LBP and the Histogram of Oriented Gradients (HOG) were used as feature extractors for the SVM classifier. The advantage of the above methods is that it is easy to implement, but the accuracy is not high.

2.2 Machine learning-based method

With the remarkable development of object detection networks in the computer vision field, smart parking management systems are also developed based on popular networks. [9] refines the YOLOv3 architecture by adding residual blocks to the original network to improve feature extraction for parking classification. [6] designs a lightweight version of YOLOv3 with MobileNetV2 architecture to improve parking classification. A faster R-CNN two-stage detection network was applied in [15] with different camera angles and parking changes. [17, 16] exploit the power of the Mask R-CNN network to extract individual cars and then classify parking conditions. Generative Adversarial Networks (GANs) are also used to directly detect occupancy and vacancies using drone imagery [14]. The advantage of machine learning methods is high detection and classification accuracy but requires networks to reach a certain depth and complexity to ensure operation in real parking conditions.

3 Methodology

Fig. 1 details the proposed parking lot detection network. This network is refined based on the original YOLOv5 architecture [13] with three main modules: backbone, neck, and head.

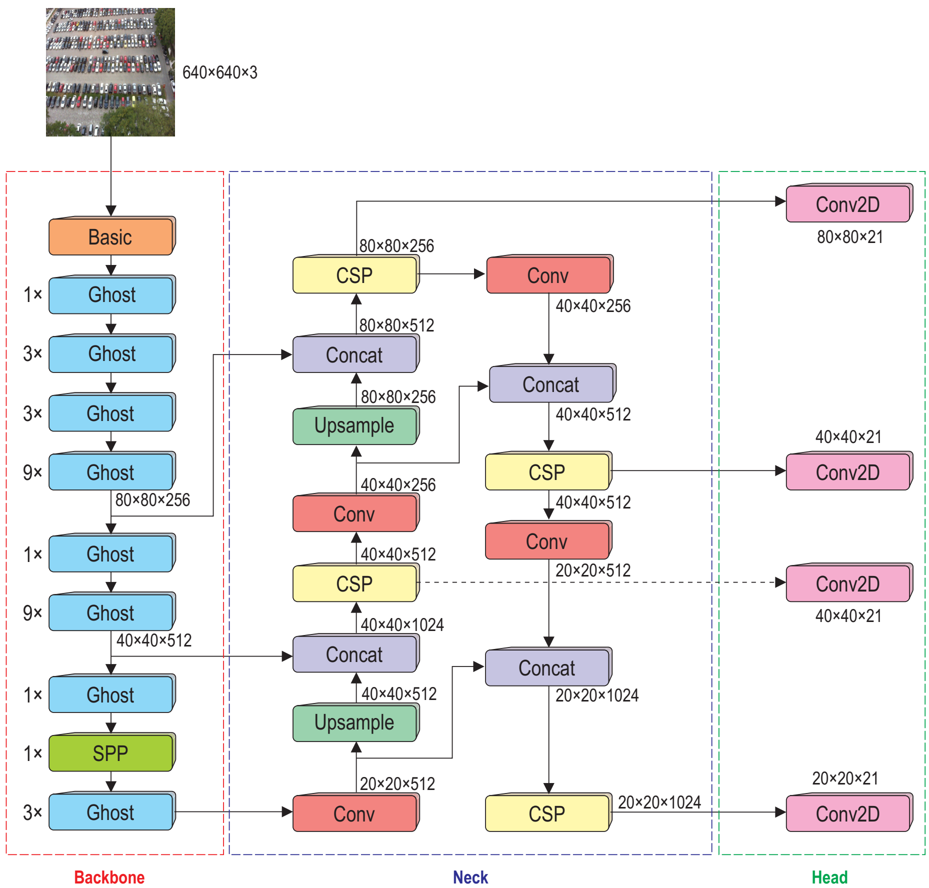

Fig. 1. The proposed parking lot detection network (YOLO5PKLot).

3.1 Proposed network architecture

The backbone module follows the design of the backbone in the YOLOv5 architecture, but this work changes a few essential modules. The techniques are applied to reduce a lot of the network parameters and computational complexity but still ensure good feature extraction. Specifically, the Focus module in YOLOv5 is replaced by the Basic module shown in Fig. 2. This module is designed with two main branches. One branch consists of 3 Conv blocks contiguously, the another branch is a Max pooling layer that is attached after the first Conv block in the first branch. The output feature maps from these two branches are concatenated and followed by another Conv block.

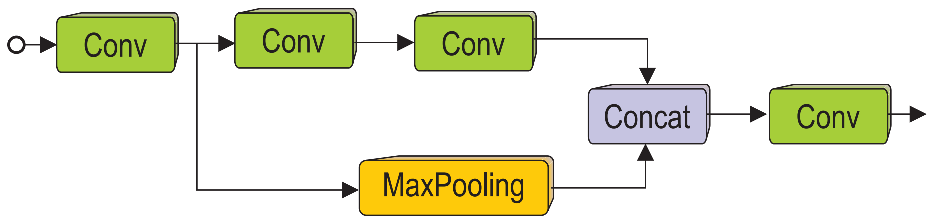

Fig. 2. The Basic module architecture.

The design of the Conv block is described in Fig. 3 with one convolution operation (Conv2D), one batch normalization (BN), and one Sigmoid Linear Unit (SiLU) activation function.

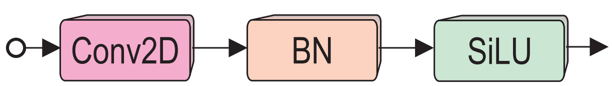

Fig. 3. The Conv block design.

Next, this design replaces all CONV blocks and Cross Stage Partial modules (Bottleneck CSP) in YOLOv5 with an architecture inspired by Ghost Bottleneck [11], named Ghost. This module is built on top of the GhostConv module (Fig. 4(a)) combined with the Squeeze and Excitation (SE) attention module [12] for stride of 1 (Fig. 4(b)) and added with the depthwise separable convolution layer (DWConv) [7] for stride of 2 (Fig. 4(c)).

Finally, this study also changes the Kernel size of the Maxpooling in the Spatial Pyramid Pooling (SPP) module from 5×5, 9×9, and 13×13 to 3×3, 5×5, and 7×7. The changed SPP module is shown as in Fig. 5.

The neck module still reuses YOLOv5's Path Aggregation Network (PAN) architecture to combine the current feature maps with previous feature maps in the first stage. Multi-scale feature maps are generated with enriched information.

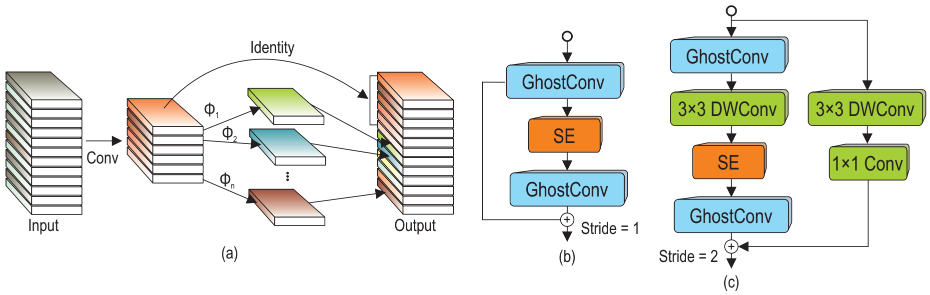

Fig. 4. (a) Ghost convolution, (b) Ghost Bottleneck module with the stride of 1, and (c) Ghost Bottleneck module with the stride of 2.

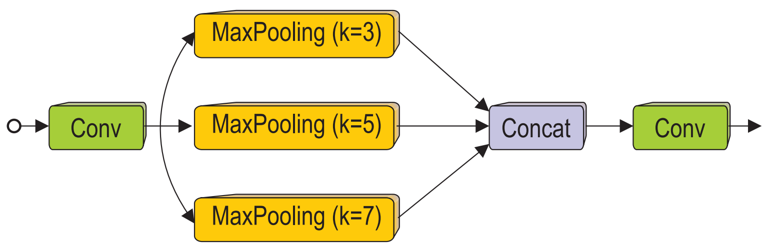

Fig. 5. The Spatial Pyramid Pooling (SPP) module.

These are the input of the detection heads. The CONV block in this module is replaced by the new Conv described above.

The detection head module utilizes three heads from the YOLOv5 architecture with feature maps from PAN neck including $80 \times 80 \times 256$, $40 \times 40 \times 512$, and $20 \times 20 \times 1024$. To increase detection ability, this work adds a detection head at a feature map of size $40 \times 40 \times 1024$ in the early stage of the PAN module. The study also resizes all anchor sizes to be suitable for the size of the objects in the PKLot dataset. The details of the detection heads and the anchor's designs are shown in Table 1.

Table 1. Heads and anchors design.

Head	Input	Anchors	Ouput	Object
1	$80 \times 80 \times 1024$	(4, 5), (8, 10), (13, 16)	$80 \times 80 \times 21$	Small
2 (Added)	$40 \times 40 \times 1024$	(10, 13), (16, 30), (33, 23)	$40 \times 40 \times 21$	Medium
3	$40 \times 40 \times 512$	(30, 61), (62, 45), (59, 119)	$40 \times 40 \times 21$	Medium
4	$40 \times 40 \times 1024$	(116, 90), (156, 198), (373, 326)	$20 \times 20 \times 21$	Large

3.2 Loss function

The loss function used in this paper is defined as follows:

$$Loss = \lambda_{box} L_{box} + \lambda_{obj} L_{obj} + \lambda_{cls} L_{cls} \tag{1}$$

Where L_{box} is the bounding box regression loss using CIoU loss, L_{obj} is the object confidence score loss using Binary Cross Entropy loss, and L_{cls} is the classes loss also using Binary Cross Entropy loss to calculate. λ_{box}, λ_{obj}, and λ_{cls} denote balancing parameters.

4 Experiments

4.1 Dataset

This experiment uses the Parking Lot Dataset [8] to train and evaluate the performance of the proposed network. This dataset is proposed by authors from the Federal University of Parana. The PKLot dataset contains 12,416 high-resolution images (1280×720 px) extracted from cameras of three different parking lots. The images were taken in sunny, cloudy, and rainy day conditions. Parking spaces are labeled as occupied and empty classes. To perform the experiment, this dataset was split into three subsets: training (8,691 images), evaluation (2,483 images), and testing (1,242 images). To be adaptive to the training process, this work reduces the image size to 640×640 px and converts the standard PKLot dataset format to YOLOv5 format.

4.2 Experimental setup

This study uses the original code of YOLOv5 [13] to generate modifications based on the Python programming language and the Pytorch framework. The proposed network is trained, evaluated, and tested on a GeForce GTX 1080Ti 11GB GPU. The Adam optimization is used. The learning rate is initially set to 10^2 and the final by 10^5. The momentum start at 0.8 and then increased to 0.937. The training process goes through 300 epochs with a batch size of 32. The balancing parameters λ_{cls}=0.5, λ_{box}=0.05, and λ_{obj}=1, respectively. Several data augmentation methods are applied such as flip up-down, flip left-right, mixup, and mosaic. The speed testing process conducts on the PKLot test set with the image size of 640×640, a batch size of 32, a confidence threshold of 0.5, and an IoU threshold of 0.5.

4.3 Experimental result

To evaluate the performance, this experiment performs training and evaluation from scratch YOLOv5 (n, s, m, l, x) versions and the proposed network. Besides, this work also compares with other previous networks that have been conducted on the PKLot dataset. As a result, the proposed network achieves

99.6% mean Average Precision (mAP). The results shown in Table 2 demonstrate that the network outperforms previous networks with 4.9% mAP when compared to the best competitor (GAN in [14]). When compared with tiny versions of YOLOv5, the proposed network achieves comparable performance to YOLOv5s and YOLOv5n while the network parameter (4,155,700 parameters) is only half that of YOLOv5s and more than two times that of YOLOv5n. In terms of computational complexity, the YOLO5PKLot network is only 2.8 GFLOPs, the smallest of all the comparison networks. The proposed network also achieves the best inference speed of 2.9 ms on the PKLot test set. The qualitative results of the proposed network on the PKLot dataset are shown in Fig. 6.

Table 2. Comparison result of proposed detection network with retrained YOLOv5 and other networks on PKLot dataset.

Model	**Parameter**	**Weight**	**GFLOPs**	**mAP**	**Inference time**
YOLOv5x	86,224,543	169.3 MB	204.2	99.7	20 ms
YOLOv5l	46,636,735	91 MB	114.3	99.7	16 ms
YOLOv5m	21,060,447	42.5 MB	50.4	99.7	13 ms
YOLOv5s	7,050,367	14.4 MB	15.3	99.6	11 ms
YOLOv5n	1,766,623	3.8 MB	4.2	99.6	3.1 ms
YOLOv3 [9]	N/A	N/A	N/A	93.3	N/A
Faster R-CNN [15]	N/A	N/A	N/A	91.9	N/A
Mask R-CNN [16]	N/A	N/A	N/A	92.0	N/A
GAN [14]	N/A	N/A	N/A	94.7	N/A
YOLO5PKLot	4,155,700	8.6	2.8	99.6	2.9 ms

With the outstanding ability in calculation and inference time as mentioned above, YOLO5PKLot can be applied in parking lot management systems with available low computing devices such as CPU and edge devices. However, during testing, the YOLO5PKLot also revealed several weaknesses that made the accuracy decrease when detecting objects in bad weather conditions such as the parking lot being obscured, the parking lot is far away from the camera, and different camera angles. Several detection mistakes are shown in Fig. 7.

4.4 Ablation study

Ablation study 1 conducted training and evaluation of proposed networks with backbone using the Basic and Ghost modules combined with an SPP module as standard. Then replace the Basic module with the Focus module, replace the SPP with the Spatial Pyramid Pooling - Fast (SPPF) module, and completely remove the SPP architectures for comparisons. The results in Table 3 show that when replacing the Basic module with the original Focus module, it increases the computational complexity to 9.2 GFLOPs while the network parameter reduces a few and just still maintains the mAP at 99.6 %. When replacing an SPP

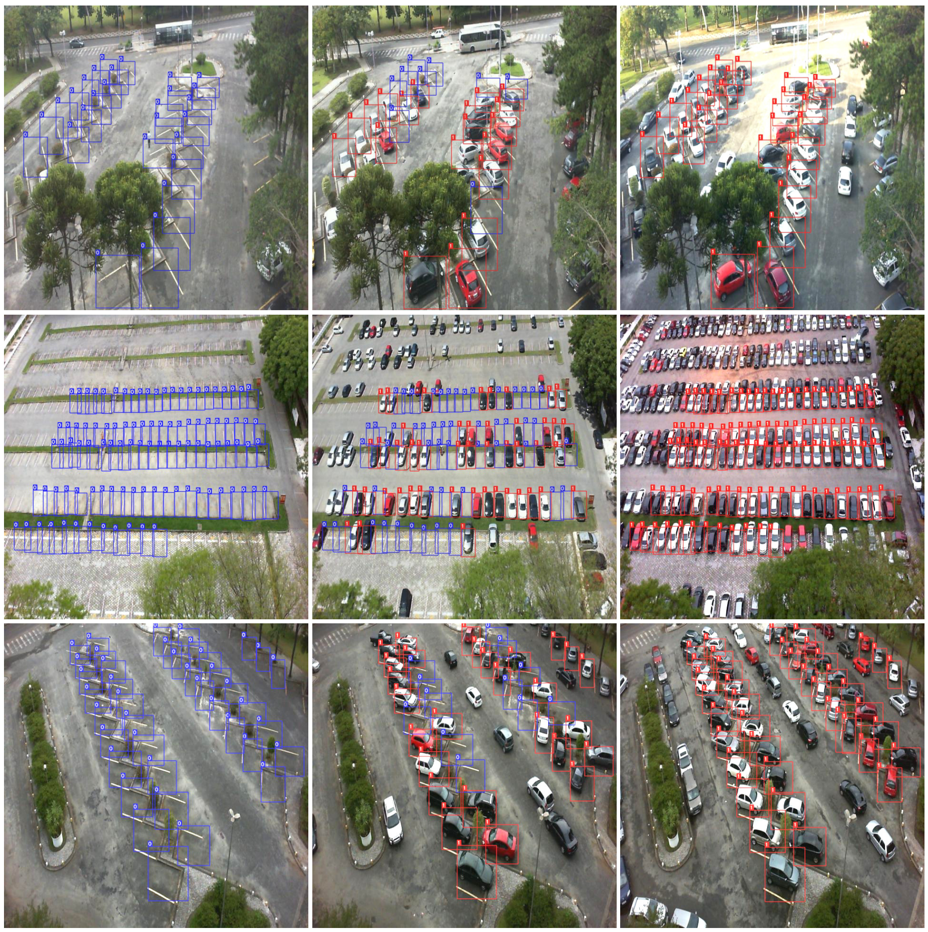

Fig. 6. The qualitative results of the proposed network on the test set of PKLot dataset with IoU threshold = 0.5. The numbers denote the classes: 0 is space-empty, and 1 is space-occupied.

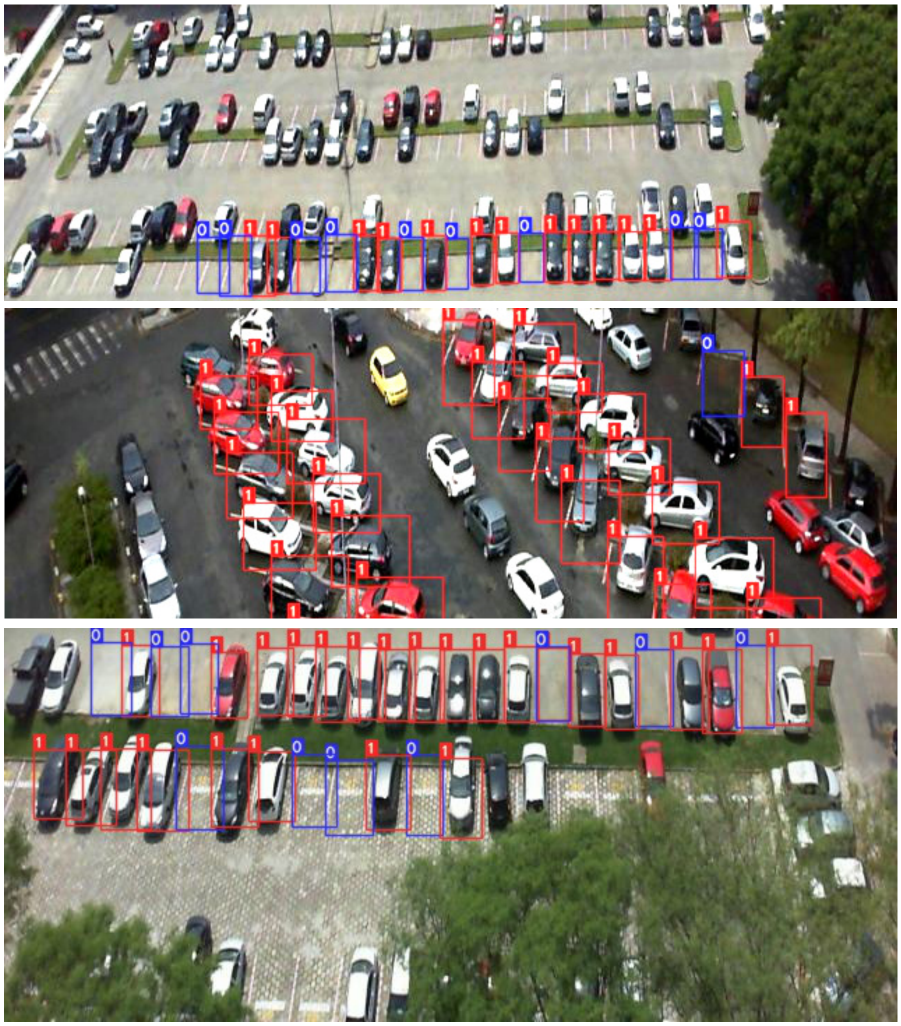

Fig. 7. Several detection mistakes in parking lot detection. (a) The parking lots are at a far distance, (b) The parking lots are obscured from each other, and (c) The parking lots are obscured by other objects (trees).

module with an SPPF module, the efficiency is similar at 99.6% mAP with the same parameters. When using a Basic module and all Ghost modules in the backbone, the network performance decreases by 0.2% mAP.

Table 3. Ablation studies with different backbone designs on the PKLot dataset.

Module	**Proposed network**			
Basic	✓		✓	✓
Focus		✓		
Ghost	✓	✓	✓	✓
SPPF			✓	
SPP	✓	✓		
Parameter	4,155,700	4,150,964	4,155,700	3,635,615
Weight (MB)	8.6	8.7	8.6	7.6
GFLOPs	2.8	9.2	2.8	2.8
mAP	99.6	99.6	99.6	99.4

In another ablation study, this work compared the performance of three and four detection heads. From the results in Table 4, it can be seen that adding a detection head increases the detection ability by 0.1% mAP. The network parameter increased slightly and the computational complexity remained the same at 2.8 GFLOPs.

Table 4. Ablation studies with different head numbers on the PKLot dataset.

Head	**Parameter**	**Weight (MB)**	**GFLOPs**	**mAP**
3	4,150,303	7.6	2.8	99.5
4	4,155,700	8.6	2.8	99.6

5 Conclusion

This paper presents a method to improve the YOLOv5 architecture for parking lot detection in smart parking management systems. The proposed network consists of three main parts: backbone, neck, and head modules. The backbone is redesigned using lightweight architectures to reduce network parameters and computational complexity. The neck is optimized with the addition of activation functions behind convolution operation and BN. The head module added a new detection head and resized the anchors to increase the detection performance of the network. In the future, this network will be further developed with attention modules to address the network's weaknesses when detecting far-distant parking spaces and adapting to different camera angles.

Acknowledgement

This results was supported by "vanishing Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(MOE)(2021RIS-003)

References

1. Almeida, P., Oliveira, L.S., Silva, E., Britto, A., Koerich, A.: Parking space detection using textural descriptors. In: 2013 IEEE International Conference on Systems, Man, and Cybernetics. pp. 3603–3608 (2013). https://doi.org/10.1109/SMC.2013.614
2. Lisboa de Almeida, P.R., Oliveira, L.S., Souza Britto, A.d., Paul Barddal, J.: Naïve approaches to deal with concept drifts. In: 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). pp. 1052–1059 (2020). https://doi.org/10.1109/SMC42975.2020.9283360
3. Almeida, P.R., Oliveira, L.S., Britto, A.S., Sabourin, R.: Adapting dynamic classifier selection for concept drift. Expert Systems with Applications **104**, 67–85 (2018). https://doi.org/https://doi.org/10.1016/j.eswa.2018.03.021, https://www.sciencedirect.com/science/article/pii/S0957417418301611
4. Antoni Suwignyo, M., Setyawan, I., Wirawan Yohanes, B.: Parking space detection using quaternionic local ranking binary pattern. In: 2018 International Seminar on Application for Technology of Information and Communication. pp. 351–355 (2018). https://doi.org/10.1109/ISEMANTIC.2018.8549756
5. Chen, H.C., Huang, C.J., Lu, K.H.: 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). pp. 297–298 (2017). https://doi.org/10.1109/ICCE-China.2017.7991113
6. Chen, W., Sheu, Peng, Wu, L., Tseng: Video-based parking occupancy detection for smart control system. Applied Sciences **10**, 1079 (02 2020). https://doi.org/10.3390/app10031079
7. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2017)
8. de Almeida, P.R., Oliveira, L.S., Britto, A.S., Silva, E.J., Koerich, A.L.: Pklot – a robust dataset for parking lot classification. Expert Systems with Applications **42**(11), 4937–4949 (2015). https://doi.org/https://doi.org/10.1016/j.eswa.2015.02.009, https://www.sciencedirect.com/science/article/pii/S0957417415001086
9. Ding, X., Yang, R.: Vehicle and parking space detection based on improved yolo network model. Journal of Physics: Conference Series **1325**, 012084 (10 2019). https://doi.org/10.1088/1742-6596/1325/1/012084
10. Dizon, C.C., Magpayo, L.C., Uy, A.C., Tiglao, N.M.C.: Development of an open-space visual smart parking system. In: 2017 International Conference on Advanced Computing and Applications (ACOMP). pp. 77–82 (2017). https://doi.org/10.1109/ACOMP.2017.29
11. Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., Xu, C.: Ghostnet: More features from cheap operations. CoRR **abs/1911.11907** (2019), http://arxiv.org/abs/1911.11907

12. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. CoRR **abs/1709.01507** (2017), http://arxiv.org/abs/1709.01507
13. Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., NanoCode012, Kwon, Y., Michael, K., TaoXie, Fang, J., imyhxy, Lorna, Yifu), Wong, C., V, A., Montes, D., Wang, Z., Fati, C., Nadar, J., Laughing, UnglvKitDe, Sonck, V., tkianai, yxNONG, Skalski, P., Hogan, A., Nair, D., Strobel, M., Jain, M.: ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation (Nov 2022). https://doi.org/10.5281/zenodo.7347926, https://doi.org/10.5281/zenodo.7347926
14. Li, X., Chuah, M.C., Bhattacharya, S.: Uav assisted smart parking solution. In: 2017 International Conference on Unmanned Aircraft Systems (ICUAS). pp. 1006–1013 (2017). https://doi.org/10.1109/ICUAS.2017.7991353
15. Martín Nieto, R., García-Martín, , Hauptmann, A.G., Martínez, J.M.: Automatic vacant parking places management system using multicamera vehicle detection. IEEE Transactions on Intelligent Transportation Systems **20**(3), 1069–1080 (2019). https://doi.org/10.1109/TITS.2018.2838128
16. Mettupally, S.N.R., Menon, V.: A smart eco-system for parking detection using deep learning and big data analytics. In: 2019 SoutheastCon. pp. 1–4 (2019). https://doi.org/10.1109/SoutheastCon42311.2019.9020502
17. Sairam, B., Agrawal, A., Krishna, G., Sahu, S.P.: Automated vehicle parking slot detection system using deep learning. In: 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC). pp. 750–755 (2020). https://doi.org/10.1109/ICCMC48092.2020.ICCMC-000140
18. Scotiabank: Number of cars sold worldwide from 2010 to 2022, with a 2023 forecast (in million units). https://www.statista.com/statistics/200002/international-car-sales-since-1990/, note = Accessed: Jan. 01, 2023. [Online]. Available: https://www.statista.com/statistics/200002/international-car-sales-since-1990/
19. Shao, Y., Chen, P., Tongtong, C.: A grid projection method based on ultrasonic sensor for parking space detection. pp. 3378–3381 (07 2018). https://doi.org/10.1109/IGARSS.2018.8519022
20. Yuan, C., Qian, L.: Design of intelligent parking lot system based on wireless network. In: 2017 29th Chinese Control And Decision Conference (CCDC). pp. 3596–3601 (2017). https://doi.org/10.1109/CCDC.2017.7979129
21. Zhou, F., Li, Q.: 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. pp. 268–271 (2014). https://doi.org/10.1109/DCABES.2014.58