Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/14/20230772

Analysis of reversing image recognition problems based on lightweight model

Dewei Dai [1, †], and Jingyi Yang [2, 3, †]

[1]Gezhi High School Shanghai (Fengxian Campus), Shanghai, Province 200000, China
[2]Jinan Foreign Language School, Jinan, Province 250000, China

[3]2021004160@poers.edu.pl
[†]These authors contributed equally.

Abstract. In recent years, people's living quality has improved, so the price of cars has fallen, and the number of cars in the world has been on the rise. In cities where land is expensive, there are fewer places to park, fewer parking Spaces, and parking itself is a difficult technology to learn. Therefore, the number of accidents caused by parking increases year by year. It is urgent to solve the safety problem of parking. Although some auxiliary astern tools have emerged as The Times require, these tools still have their shortcomings. For example, the commonly used astern radar and astern image cannot see objects behind, and astern image is a wide-angle lens that makes it difficult for the driver to judge the distance between the car and obstacles. In this environment, the current situation needs to be improved. In order to improve the accuracy of obstacle judgment, this paper chooses semantic segmentation as the function. This paper chooses lightweight model to speed up the process of deliver information, such as ShuffleNetV2.

Keywords: deep learning, assisted parking, lightweight model.

1. Introduction

Technology is developing at a rapid pace today, and people are no longer content to operate on bulky, immobile devices. However, because mobile phones don't have enough performance, if you use cloud computing, you can make mistakes in different regions. In order to make work and learning more productive, people began to create ways to make the transfer and transmission of models more rapid, convenient and efficient.

For example, as people's living standard gets better, the number of cars shows a spurt growth, parking has become a difficult problem for daily travel. Under the development trend of artificial intelligence and automobile integration, artificial intelligence technology based on neural network, deep learning and other algorithms has become a hot and difficult topic in the research field of intelligent driving industry [1]. As cars move towards autonomous driving, pedestrian and obstacle recognition is becoming more and more important. Now the function of identifying obstacles is still a little slow, the prompt given is late, will cause some accidents caused by the failure to brake in time when reversing. This paper study the mobile end model to assist parking in order to avoid as much as possible the traffic accidents caused by the slow function of reversing to recognize obstacles, and reduce unnecessary casualties.

Our research topic is to improve the mobile end deep learning model for identifying obstacles to accelerate the speed of identifying obstacles and achieve the goal of reducing the occurrence of reversing

DOI: 10.54254/2755-2721/14/20230772

accidents. Deep learning is a gradually developed method for artificial intelligence image classification. Deep learning methods have great advantages over traditional machine learning methods in image classification. Its core advantage lies in the ability to automatically learn complex and abstract deep features from shallow features, thereby making classification and decision-making more efficient and accurate.

The requirements of obstacle information extraction industry are essentially the ground object classification of each pixel of the image. In depth learning, it belongs to the pixel-based image classification method, also known as image semantic segmentation or image labelling, which is to divide the image into several semantic classification identification areas based on semantic units. Image semantic segmentation is the synthesis of target classification recognition and segmentation. Around the theme of image semantic segmentation, predecessors have done a lot of research. LeCun et al. established the basic model idea of convolution neural network (CNN) in 1990 and proposed CNN's classical architecture lenet5 in 1998. In 2012, Alex et al. proposed a AlexNet model based on CNN structure, classified massive natural images and achieved breakthrough results. Since then,CNN has become one of the most widely used in-depth learning models in the field of graphic classification. In 2015, JLong et al. proposed the full convolution neural network (FCN). T can not only recognize the category of pixel, but also restore the position of pixel in the original graph, and truly realize the pixel classification of image. Image classification based on FCN is an end-to-end architecture, which enables users to input a single image of any size to be classified, and finally outputs a classification map of the same size with accurate target boundary and correct annotation, that is, the combination of positioning accuracy and recognition accuracy. U-Net model is an improved model in FCN, which combines the characteristics of deconvolution network and jump network and has been applied in the field of image classification in recent years [2].

Introduction is the first part of this paper, case description is the second part, analysis on the problems is the third part , suggestion is the fourth part, the fifth part is the conclusion, the sixth part is the reference.

2. Case description

Parking is inherently a relatively difficult skill, because people will need a series of difficult techniques such as reversing in the process of parking. The reason why reversing is more difficult is because the car is opaque, so people sitting in the car will block their sight because of the doors, the A and B pillars of the front windshield, and other structures of the car itself, resulting in blind spots in the field of vision. For example, when reversing, there will be a blind spot of 2-3 square meters of vision caused by the rear windshield.

In recent years, due to the improvement of people's quality of life and the decline of car prices, the global car ownership has been steadily increasing. However, the infrastructure of some cities, such as parking lots and road width, has not increased with the increase in car ownership, resulting in serious traffic congestion and insufficient parking spaces in many cities. As a result, the probability of minor accidents such as paint wear and vehicle collisions increase [3], [4].

Therefore, in order to help the driver in the car obtain more information about the parking space and reduce the interference of blind spots, many auxiliary reversing technologies have been invented, among which parking sensors and vehicle backup camera are commonly used. However, the parking sensors cannot identify the type of the objects behind; the vehicle backup camera with a wide-angle lens is difficult to judge the distance from obstacles, so that the driver cannot fully obtain the environmental conditions around the car body, so safety accidents are very prone to occur. Therefore, how to improve the existing assisted reversing technology has become a problem to be solved.

The type of obstacle could be determined through semantic segmentation. Then, the type of obstacle inferred by the program will be displayed on the vehicle screen to help driver in the stage of parking. In general, semantic segmentation uses a classification at the pixel level to classify all pixels belonging to the same class into one class and mark them with one color, so as to realize the recognition of image content [5].

DOI: 10.54254/2755-2721/14/20230772

In order to reduce the delay, protect the driver's privacy, and reduce the burden of cloud storage, lightweight models which is specially designed for mobile terminals are used. These models, such as MobileNetV3, ShuffleNet, and GhostNet, has relatively small parameters and low hardware requirements. It is quantitatively compared which of the three models is more suitable for assisted driving through two aspects of accuracy and time complexity.

However, MobileNetV3 and ShuffleNetV2 are designed for image classification tasks, so in order to apply these two models to semantic segmentation, it is necessary to add a network (such as FCN network) that can produce a marked imagine instead of a probability. At the same time, the design of some original models also needs to be modified. Ensure program accuracy.

3. Analysis on the problem

3.1. Semantic segmentation task based on MobileNetV3 and Fully Convolutional Networks(FCN)

3.1.1. Introduction of FCN. Jonathan Long proposed the use of Fully Convolutional Networks, which is also called FCN, in [6] to implement supervised training semantic segmentation for the first time. FCN can classify each pixel from abstract features, making image classification extend from the image level to the pixel level. The Figure 1 is architecture of FCN-8S.

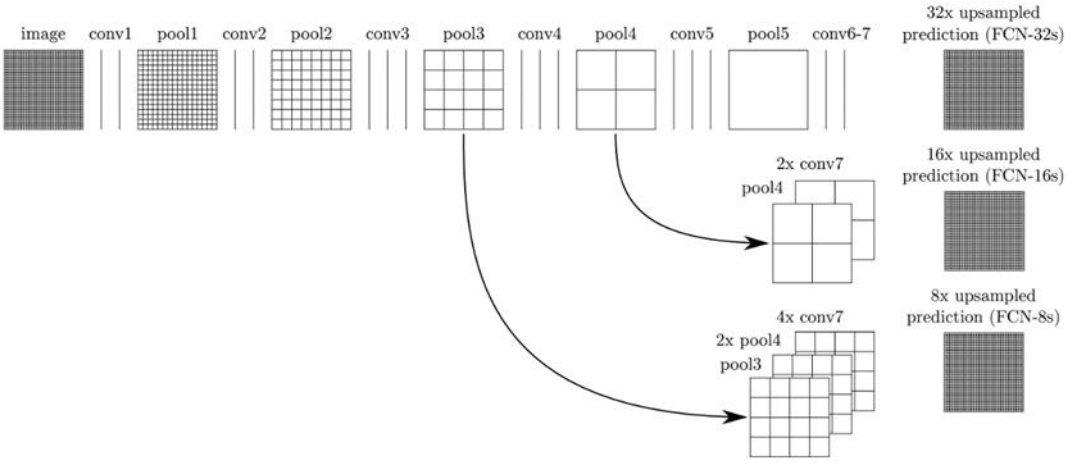

Figure 1. FCN network architecture.

In FCN, compared with CNN, the global average pooling layer and fully connected layer are removed. In addition, transposed convolution and 1*1 convolution are added at the end of the structure. Transposed convolution is used to achieve pixel prediction, while 1*1 convolution is used to reduce the amount of calculation. However, because the network structure with so many parameters is too complex and need a long time to produce a marked graph, FCN is not suitable for direct application to the scene of reversing assistance. Therefore, inserting FCN into other lightweight models is a solution.

3.1.2. Introduction of MobileNetV3. In 2019, MobileNetV3 [7] which is designed for applying CNN model on mobile terminal was proposed. This model has two models of different sizes, MobileNetV3-Large and MobileNetV3-Small. The difference is that MobileNetV3-Large has more layers. The model used in this article is MobileNetV3-Large.

3.1.3. Experimental method. In [8], the author proposed a way to combine FCN with MobileNetV3. The whole model has two parts. The first part is encoder part and the second part is decoder part. The encoder part mainly uses the first 16 layers in MobileNetV3 to extract features and expand the number of channels. Decoder part is deconvolution upsampling structure. Figure 2 provides the structure of the

DOI: 10.54254/2755-2721/14/20230772

network.

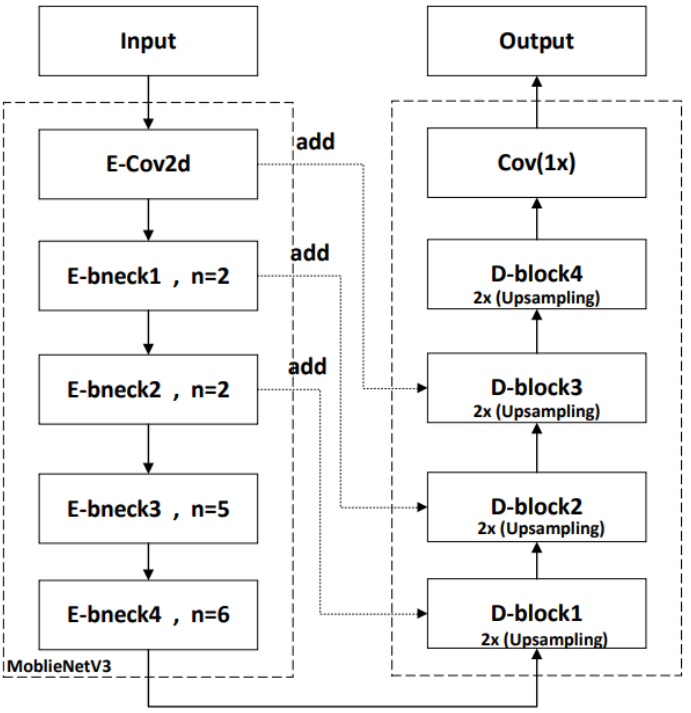

Figure 2. Network architecture of the improved Mobile v3 [8].

3.1.4. Data in experiment and analysis. From the data shown in table 1, it is clear that the time taken of MobileNet+FCN is much smaller than that of FCN-8S, but the segmentation precision has a greater decline compared with the original model.

Table 1. Experiment data for MobileNet+FCN architecture compared to other approaches.

Network	Parameters(Mb)	Running time (ms)	MIoU(%)
FCN-8S	20	478.4	0.3731
MobileNet+FCN	2.13	100.2	0.2557

3.2. Semantic segmentation task based on ShuffleNet and SkipNet

3.2.1. Introduction of ShuffleNet. ShuffleNet[9] is mainly used for image classification. The characteristic of this model is to use group convolution and propose channel shuffle. Group convolution can reduce the amount of computation and thus reduce resource usage and running speed. channel shuffle can be a special fusion of various groups to improve accuracy.

3.2.2. Experiment. In [10], the author proposes a model with two parts based on ShuffleNet and SkipNet. Encoder part mainly uses ShuffleNet to reduce the number of channels and prevent overfitting. Decoder mainly uses upsampling structure in SkipNet. Figure 3. shows the algorithm structure based on the improved ShuffleNet network.

DOI: 10.54254/2755-2721/14/20230772

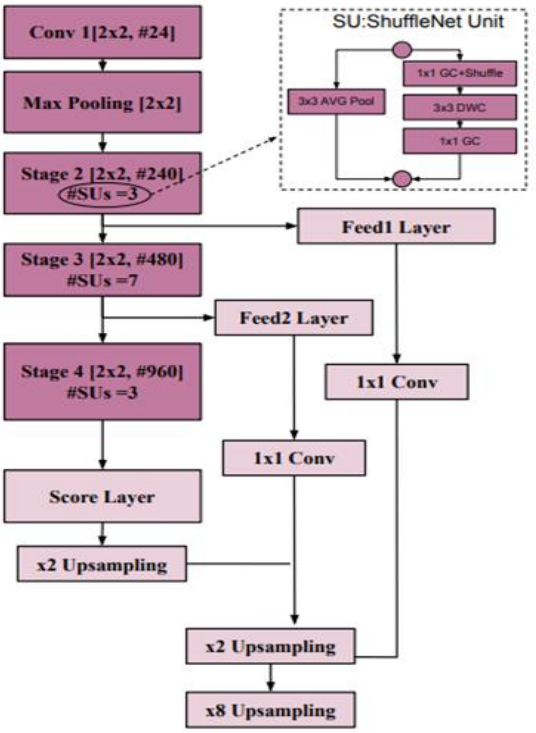

Figure 3. Improved network structure for ShuffleNet [10].

3.2.3. Experimental data and analysis. From table 2, we can see that ShuffleSeg which is the network based on still has high accuracy when the number of parameters is reduced a lot.

Table 2. Experiment data for ShuffleSeg architecture compared to other approaches.

model	GFLOPs	Class IoU	Class iIoU	Category IoU	Category iIoU
SegNet	286.03	56.1	34.2	79.8	66.4
ENet	3.83	58.3	24.4	80.4	64.0
ShuffleNet and SegNet	2.03	58.3	32.4	80.2	62.2

3.3. Semantic segmentation task based on GhostNet and U-Net

3.3.1. Introduction of GhostNet. The goal of GhostNet[11] is to generate as many feature maps as possible with fewer parameters, so as to discover the information required by the features. The author proposes the Ghost module, which generates multiple ghost feature maps based on the feature maps generated by convolution through linear operations with few parameters.

3.3.2. Experiment. In [12], the whole improved model structure is an asymmetric encoder-decoder architecture. The encoder part is mainly composed of the improved Ghost-Net, and a U-Net is used for feature expanding. The main shortcut is removed in the improved Ghost-Net, but skip connection is added in the proposed bottleneck. As for decoder part, Double Conv layer was proposed with two convolution layers, BN and LeackyReLU.

DOI: 10.54254/2755-2721/14/20230772

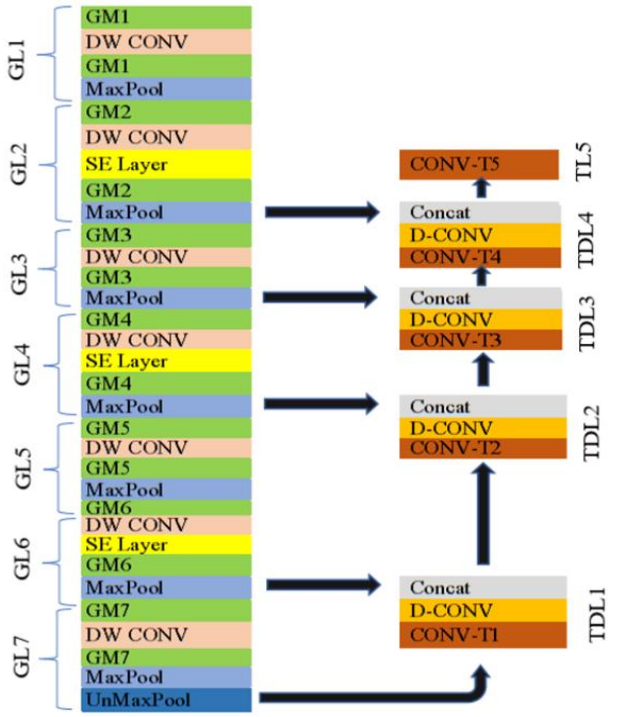

Figure 4. Network structure for the combination of GhostNet and U-Net [12].

3.3.3. Experimental data and analysis. From table 3, it shows that the mIoU of Ghost-UNet are still high when the amount of model parameters is greatly reduced.

Table 3. Experiment data for semantic segmentation using Ghost-UNet and Seg-Net.

Model	params	mIoU(%)
Seg-Net	29.5M	57
Ghost-UNet	5.8M	74

4. Suggestion

To sum up, Ghost-UNet performs better than the other two models in the field of semantic segmentation, because it has mIoU(mean Intersection over Union)closer to 1.

mIoU refers to the average intersection and union ratio, which is an important indicator for evaluating the performance of semantic segmentation. Its formula is:

$$mIoU = \frac{1}{k}\sum_{i=1}^{k}\frac{P \cap G}{P \cap G}$$

If mIoU is closer to 1, it means that the predicted value is closer to the real value, and the accuracy of semantic segmentation is higher. Therefore, Ghost-UNet with relatively small parameters has higher accuracy on semantic segmentation and is more suitable for assisted parking technology.

5. Conclusion

This paper proposes an improvement scheme for the traffic accidents caused by the blind field of vision of the driver when reversing in parking and the respective defects of the two kinds of auxiliary parking commonly used in the market. The rear object cannot be seen by the astern radar and the astern image

DOI: 10.54254/2755-2721/14/20230772

cannot make the driver know the distance from the rear object. First of all, the most important thing for auxiliary parking is to prompt you to keep up with the astern speed of the car, which requires that the response of the model must be fast, so we choose lightweight models MobileNetV3 and ShuffleNetV2. Secondly, these two models need to analyze the image if they are to be used, so we use semantic segmentation. Of course, there are drawbacks to this approach. Therefore, we finally found that the semantic segmentation accuracy of Ghostle-unet is higher. In the future, we hope that the subject of assisted parking will be studied by more people. The safety of reversing can be improved to minimize the casualties caused by reversing.

References

[1] Information on https://new.qq.com/rain/a/20221212A01HW700
[2] Information on https://www.doc88.com/p-99639059124626.html
[3] B. Wang, C. Shao, J. Li, D. Zhao, and M. Meng, ‘‘Investigating theinteraction between the parking choice and holiday travel behavior,’’ Adv.Mech. Eng., vol. 7, no. 6, pp. 1–11, Jun. 2015.
[4] M. Roca-Riu, E. Fernández, and M. Estrada, ‘‘Parking slot assignment for urban distribution: Models and formulations,’’ Omega, vol. 57,pp. 157–175, Dec. 2015.
[5] Cakir, S., Gauß, M., Häppeler, K., Ounajjar, Y., Heinle, F., & Marchthaler, R. (2022). Semantic Segmentation for Autonomous Driving.
[6] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).
[7] Howard, Andrew, et al. "Searching for mobilenetv3." Proceedings of the IEEE/CVF international conference on computer vision. 2019.
[8] Zhang, Y., & Chen, X. (2020, December). Lightweight semantic segmentation algorithm based on MobileNetV3 network. In 2020 International conference on intelligent computing, automation and systems (ICICAS) (pp. 429-433). IEEE.
[9] Zhang, X., Zhou, X., Lin, M., & Sun, J. (2018). Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 6848-6856).
[10] Gamal, M., Siam, M., & Abdel-Razek, M. (2018). Shuffleseg: Real-time semantic segmentation network. arXiv preprint arXiv:1803.03816.
[11] Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., & Xu, C. (2020). Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 1580-1589).
[12] Kazerouni, I. A., Dooly, G., & Toal, D. (2021). Ghost-UNet: An asymmetric encoder-decoder architecture for semantic segmentation from scratch. IEEE Access, 9, 97457-97465.