International Core Journal of Engineering Volume 7 Issue 5, 2021
ISSN: 2414-1895 DOI: 10.6919/ICJE.202105_7(5).0068

Recognition of Floating Objects on Water Surface

--Based on Improved Residual Network

Jiasong Bao[1,a], Xixia Huang[2,b] and Ke Ji[1]

[1]Shanghai Maritime University, Shanghai 201306, China;

[2]Key Laboratory of Marine Technology and Control Engineering Ministry of Communications, Shanghai 201306, China.

[a]baojiasong96@163.com, [b]961370568@qq.com

Abstract

In order to recognize the recovery of floating objects on water surface by unmanned sweeper, a recogition model is proposed based on an improved residual neural network. The model improves the network performance by adjusting the structure of the origional residual neural network model, increasing the convolution output appropriately, and reducing residual unit, so as to improve the ability of the residual neural network to extract objection features. Experimental results show that the proposed model has a better recognition performance than the exist residual network both in the self-made MyFloats dataset and CIFAR-10 dataset.

Keywords

Residual Neural Network; Floating Objects on Water Surface; Objection Recognition; Myfloats Dataset.

1. Introduction

China has always been attaching importance to the water environment management due to the environment caused by the floating objects on water suface such as water hyacinth, plastic products and metal products[1]. With the proposal of the national garbage classification policy, it is particularly important to accurately identify floating objects on the water surface and classify and recycle them.

At present, the methods used to detect and identify floating objects on water surface mainly include artificial identification, remote sensing detection[2] and machine vision identification. Deng Lei et al. [3] used BP neural network to build a classifier which takes the edge, gray level and texture of the input image as features, designing a discrimination model for floating plastic bottles and branches. Li Ning et al. [4] used the AlexNet network model to realize the identification of plastic bags and plastic bottles after fine-tuning of small samples. Lei Liyi et al. [5] proposed a small floating object dataset containing aquatic plants and fallen leaves, and made fine-tuning on the pre-trained Faster RCNN and SSD models and achieved certain recognition effect. In 2015, Microsoft research put forward the networks with residuals[6], which improved the identification accuracy to 95.06% in ImageNet Large Scale Visual Recognition Challenge(ILSVRC) and has potential to recognize the floating objects on water suface.

On the basis of existing residual neural network models, combined with the idea that wide neural network can also have good performance proposed by Zagoruyko S[7] et al. This paper propose an improved residual neural network of 14 levels containing 6 residual blocks and have the advantage of being wide and relatively shallow. Through the experiment, the improved model has excellent network peformance, and stronger feature extraction ability for objects in complex environment.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202105_7(5).0068

2. Basic theory

2.1 Residual network

The residual neural network was proposed by KaiMing He[8] et al. in the 2015 ImageNet Large Scale Visual Recognition Challenge. With the increasing number of convolution layers in the neural network, when updating parameters with back propagation, simple convolution stacking is easy to cause problems such as gradient disappearance, saturation or degradation of network performance. The introduction of residual learning solves the problem of network performance saturation or degradation caused by the deepening of network layers, and also improves the accuracy of test dateset. The structure of residual network is shown in the figure 1.

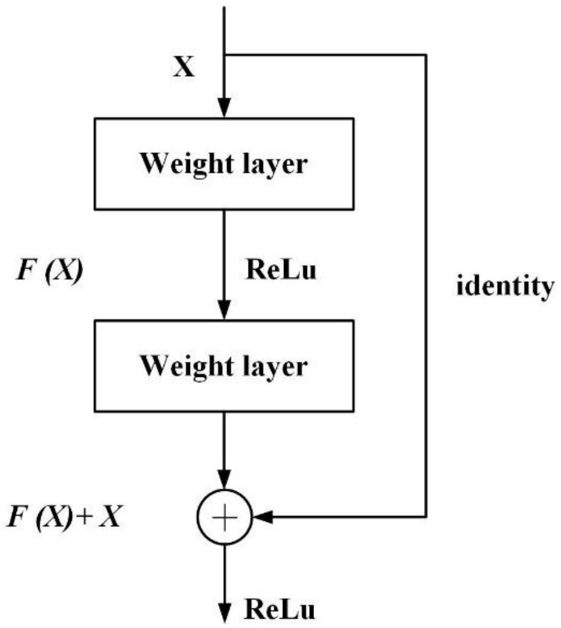

Figure 1. The structure of residual network

The residual model is expressed as follows:

$$H(x) = F(x) + x \tag{1}$$

Among them, x represents the output of the previous layer, $F(x)$ represents the output after residual learning, $H(x)$ represents the overall output of the residual model, weight layer represents the convolutional operation in Convolutional neural network(CNN).

2.2 Improved residual network

Inspired by the wide neural network that proposed by Zagoruyko S[7] et al., it is obvious that there is no guarantee that the gradient can flow through the weight of each residual element when the gradient is in back propagation with the increasing of the depth of the model, leading to a poor learning effect. During the whole training process, only a few residual units can be used to learn useful feature expression. As a result, in this paper, a shallower but wider model is adopted, which makes the residual elements play more roles to improve the model performance more effectively. This structure is realized mainly by reducing the number of residual units and increasing the number of feature graphs of the convolutional layer in the residual units. It has been proved by experiments that increasing the width of the model can improve the performance of the model, but it cannot be completely considered that the width is better than the depth. Therefore, the network model based on ResNet-32 and the idea of wide residual neural network are adopted to improve the 14-layer wide residual neural network model. The improved model is referred to as Wide-ResNet-14, the network structure is shown in figure 2.

ISSN: 2414-1895
DOI: 10.6919/ICJE.202105_7(5).0068

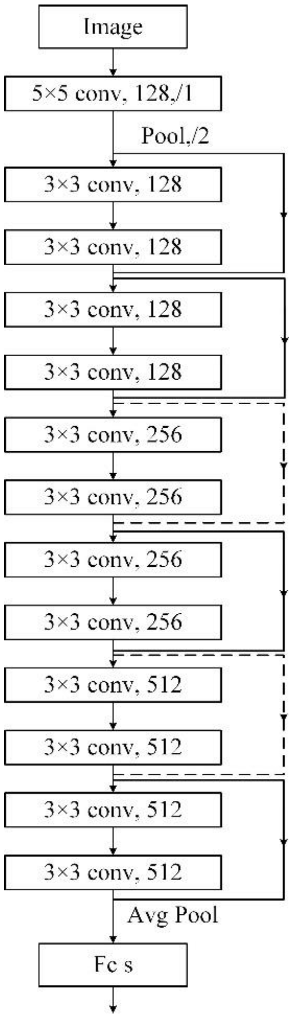

Figure 2. The structure of Wide-ResNet-14

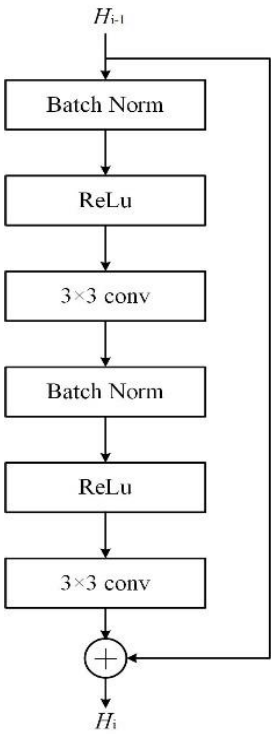

Figure 3. The structure of ResNet block

ISSN: 2414-1895 DOI: 10.6919/ICJE.202105_7(5).0068

First of all, the first layer of convolution uses 5x5 convolution kernels and the size of step is 1, and padding size is 1 so that the input and output of convolution operation have the same feature maps, and then through the average pooling, the size of the filter is the 3x3 and the size of step is 2 so as to reduce the figure size, simplifying network computing complexity. Since then, feature maps are put in the first residual block.

The Wide-ResNet-14 contains 6 residual blocks and each of their structure is shown in figure 3.

First, the batchnormalization[9]algorithm is used to normalize the output data of each cycle to achieve the purpose of stablizing and accelerating Network Convergence and preventing the overfitting. Then, the data processed by the BN algorithm is activated through the ReLU activation function to increase the nonlinear relationship between the layers of the neural network and complete the task of extracting complex features that needed to be completed by the neural network. At last, the activated data is fed into a convolution layer in which convolutional kernal is set to 3x3, step size is set to 1 and padding is set to 1 so that the input H_{i-1} and the output H_i of each residual block have the same size. The second convolution layer is input through BN algorithm and ReLU[10] activation function, and the relevant parameters are the same as the first convolution layer. Finally, the output of the second convolution and the input of the residual block are used as the input of the next residual block through matrix addition operation, and the feature mapping that needs to be learned is finally transformed into $F(x) + x$ in the residual idea. In the first residual element, BN algorithm and ReLU activation function are not used before the convolution layer.

The average pooling layer is set after 6 residual blocks, and a full connected layer is set after the average pooling layer. The parameters of layers are as shown as table 1. Among them, Basis Block is the parameter of the network that test in Cifar-10 and the Improved Block is the parameter of the improved network that proposed in this paper.

Table 1. Structure of parameters

Group name	Output size	Basis Block	Improved Block
Conv1	64×64	[3×3,16]	[5×5,128]
Conv2	32×32	5×[3×3,16]	2×[3×3,128]
Conv3	16×16	5×[3×3,32]	2×[3×3,256]
Conv4	8×8	5×[3×3,64]	2×[3×3,512]
Avg-pooling	1×1	8×8	8×8

Through increasing the number of some convolution output channels in the original residual model to widen the model and reducing the number of residual units makes the original deep residual model shallower, the improved residual network reduces the number of network layers from 32 to 14.

As a result, the bottom of the network could learn from the original pixel to depict the local edge and texture features. Through a combination of various edge filters, intermediate filter layer could depict various features of objects and the top layer depict the whole feature.

3. Experiment results

3.1 Dataset

In this experimental dataset, 1900 pictures of water hyacinth and 4 other floating objects were collected by CCD camera, including 526 pictures of water hyacinth, 340 pictures of duckweeds, 280 pictures of plastic bags, 300 pictures of plastic bottles and 454 pictures of metal cans. In order to make the generalization ability of the network stronger, the data set was expanded. Through mirroring, rotation (±30° and ±60°), random clipping, adjusting brightness and adding noise, the dataset of 1900 images was expanded to 20000 images[11]. As is shown in figure 4.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202105_7(5).0068

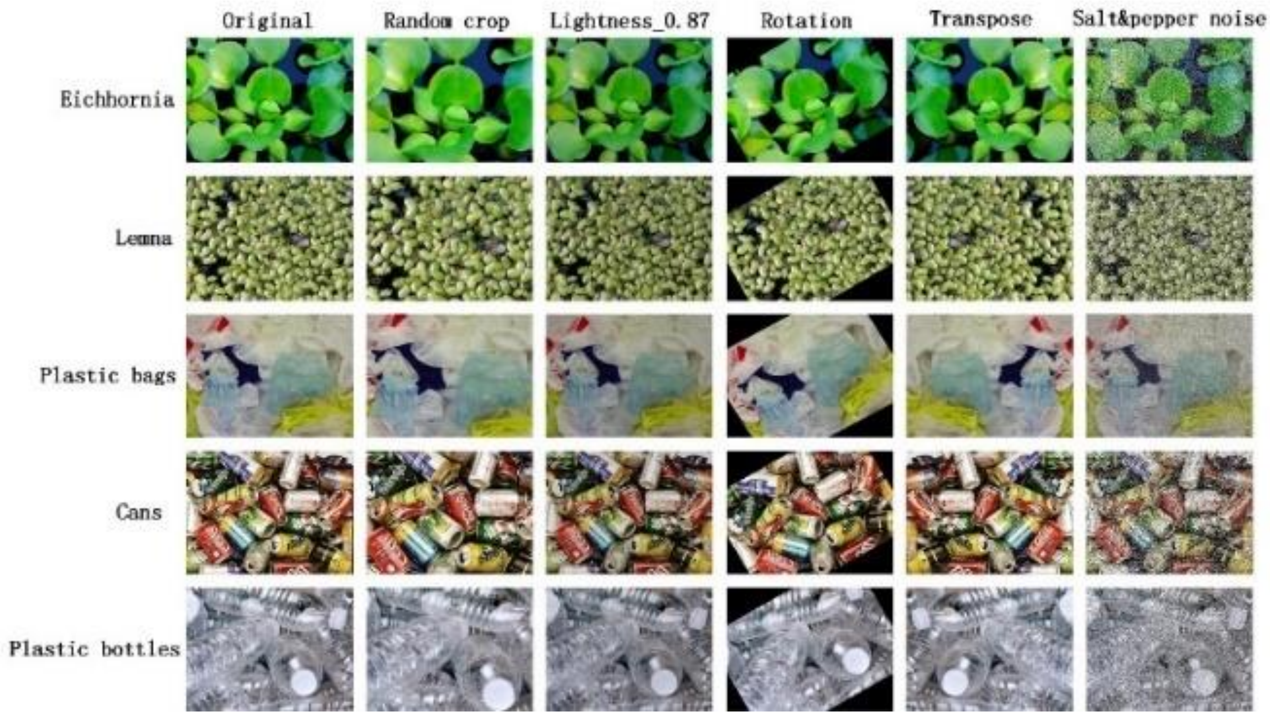

Figure 4. The dataset of MyFloats

And 80% of the dataset is classified as a training dataset and 20% as a test dataset.

3.2 Experiment environment

The experimental environment is Tensorflow, a deep learning framework developed by Google, whose system structure consists of front-end and back-end. The front-end architecture provides support libraries in a variety of languages like C++, python and Java and triggers the back-end program to run based on the C API. Programming language is Python 3.6. CPU is Intel core i7-8750h 2.21GHz with 16-GB random access memory, GPU is configured to NVIDIA GeForce GTX 1060 with 6-GB memory. In the experiment, the data flow graph of convolutional neural network was built in TensorFlow environment with Python, and the training data was input into the graph. After the data flow graph was started, the weight updating module was used to update the weight automatically. After the network training, the test data was input into the data flow graph for diagnosis.

3.3 Result analysis

Gradient descent algorithm is used to optimize the network[12] by using the improved model proposed by the platform for accelerated training, the hyperparameters of the network is shown as table 2.

Table 2. The hyperparameters of the network

Parameters	Value
batch size	128
init_lr /ir_decay_factor	0.1/0.1
lr_decay_step0/lr_decay_step1	40 000/60 000
train_steps	80 000
Momentum	0.9
Weight_decay	0.0002

In the training process, in order to get the best experimental results on the existing hardware equipment, the size of the training batch was selected as 128, and the learning rate attenuated with the number of steps was used to train the network parameters, and the initial learning rate was 0.1.

The current step number is attenuated by 0.1 times in 40 000 and 60 000 steps respectively, so as to approach the optimal solution as much as possible and improve the accuracy. And the training cruve of MaFloating dataset is shown as figure 5.

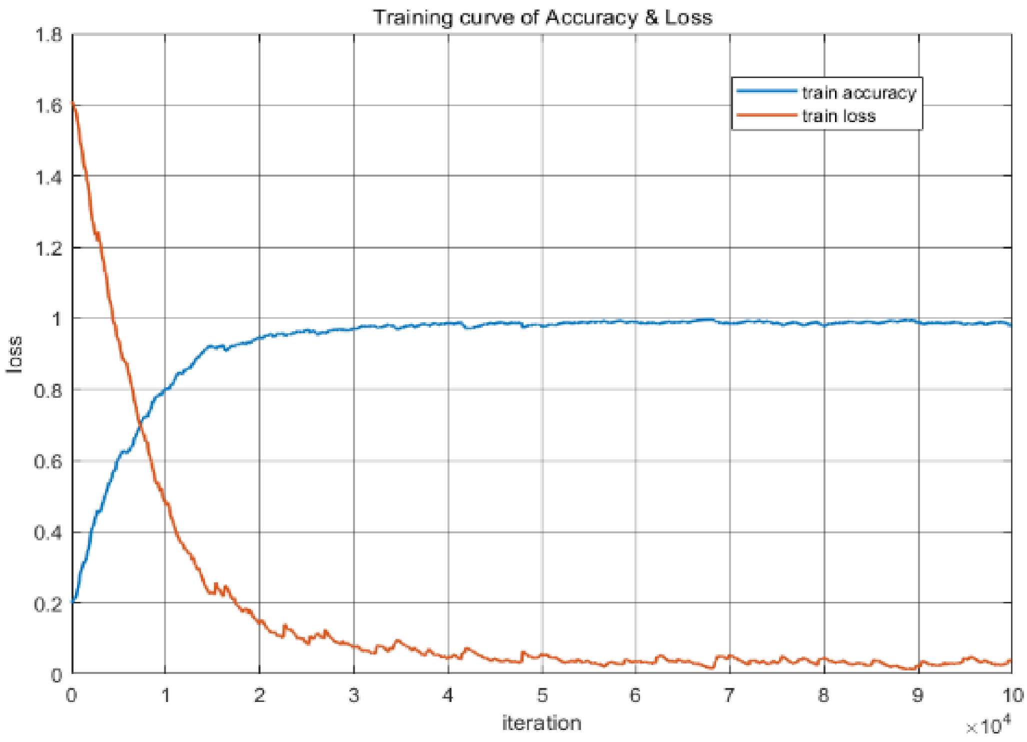

Figure 5. The training cruve of MyFloats

The training accuracy and loss converge stably after the 40000 steps and it refers to a good performace with the parameters setting.

When training and testing have been separately carry on the self-made MyFloats dataset and cifar-10 dataset, for objects were not aligned in the experiment, Top - 1 accuracy rate is obtained as is shown in table 3.

Table 3. Top - 1 accuracy rate(%) comparison

Dataset	ResNet-32	Widen-ResNet-14
MyFloats	91.23	96.78
CIFAR-10	63.42	71.26

It is obvious that, compared with the original residual neural network model, the improved residual neural network Widen-ResNet-14 has a better performance on MyFloats and CIFAR-10. And it is the reason that improved residual blocks could improve the utilization of the residual unit, leading to Widen-ResNet-14 more effective than the original ResNet model which is helpful to improve the competence of extracting features.

4. Conclusion

The Widen-ResNet-14 proposed in this paper have a higher accuracy more 5% than the original network both in MyFloats and CIFAR-10, especially have a better performance in CIFAR-10 which suggests that the improved residual network has a strong ability to extract the features of objects. However, all the experimental data weren’t preprocessed well such alignment, background information reduction and so on and it is possible that the model proposed in this paper could have some defects. As a result, the next research emphasis should be focused on how to find a balance

ISSN: 2414-1895 DOI: 10.6919/ICJE.202105_7(5).0068

between excellent data preprocessing and model adjustment in depth and width to give full play to residual networks' ability.

Acknowledgments

Natural Science Foundation.

References

[1] P. He, L. Shen, Y. Zhang, et al. Study on the solution of water hyacinth problem based on Kunshan area, JiangSu Water Resources, 6(2018), 42-45,49.

[2] Y. Q. Wang, L. Ma, Y. Tian, State-of-the-art of Ship Detection and Recognition in Optical Remotely Sensed Imagery, IEEE/ CAA Journal of Automatica Sinica (JAS). 2011,37(9),1029-1039.

[3] L. Deng, L. Yan, S. H. Zhang, et al. Intelligent Recognition and Judgment System of Floating Objects on Water Surface Based on Machine Vision, Electronic Test, 2019(17),133-134.

[4] N. Li, Y. Wang, S. Xu, et al. Recognition of floating objects on water surface with small sample based on Alexnet, Computer Applications and Software, 2019,36(2),245-251.

[5] L. Lei, J. Ai, J. Peng, et al. Evaluation of deep neural networks for floating objects detection, Inner Mongolia Environmental Sciences, 2019,31(6),117-120,123.

[6] K. M. He, X. Zhang, S. Ren, et al. Deep Residual Learning for Image Recognition, Computer vision and pattern recognition, (2016): 770-778.

[7] S. Zagoruyko, N. Komodakis, Wide residual networks, The 27th British Machine Vision Conference, BMVC 2016, NewYork, UK: BMVA, 2016: 87.1-87.12.

[8] K. M. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2011 33(12): 2341-2353.

[9] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, The 32nd International Conference on Machine Learning, ICML 2015,Lille,France, 2015: 448-456.

[10] A. Esteva, B. Kuprel, R. Novoa, et al. Dermatologist-level classification of skin cancer with deep neural networks, Nature,2017,542(Feb.2 TN.7):115-118.

[11] C. Szegedy, W. Liu, Y. Jia, et al. Going Deeper with Convolutions, 2015,1-9.

[12] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, et al. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? , IEEE Transactions on Medical Imaging, 2016, 35(5): 1299-1312.