ECE371 Neural Networks and Deep Learning Assignment 1

Luo Shuxian 22308135
School of Electronics and Communication Engineering
Sun Yat-sen University, Shenzhen Campus
luoshx7@mail2.sysu.edu.cn

Abstract: By making adjustments to pre-trained ResNet152 model, the classification accuracy of this model on the Flower dataset has improved. By proposing an extensive training strategy such as the tuning learning rate, the model's accuracy reached 80 percent.

Keywords:

Neural NetworkResnet152,Image Recognition

1 Introduction

This research explores the application of deep learning models and training techniques on a specific dataset to achieve high accuracy in image recognition.In this task, we focused on fine-tuning a pre-trained ResNet152 model for the VGG11 model that did not perform well. By implementing a comprehensive training strategy within a single main function, the accuracy rate improved a lot. This was made possible through advanced data augmentation techniques, such as random cropping and random erasing, which helped improve the model's robustness. Collectively, these strategies contributed to the improved performance and stability of the model throughout the training process.

2 Related Work

Early models such as R-CNN[1]relied on complex multi-stage processing pipelines, including region proposal and feature extraction. However, with the development of technology, the introduction of YOLO[2]has unified the task of object detection into a single neural network, achieving real-time object detection and significantly improving detection speed and efficiency.
The emergence of Fully Convolutional Networks[3] has brought a breakthrough to semantic segmentation tasks. By using fully convolutional layers to process input images of arbitrary sizes, FCN has significantly improved segmentation accuracy and flexibility.
To enhance model performance and flexibility, researchers have continuously optimized network architectures. The introduction of Spatial Pyramid Pooling[4] has addressed the issue of fixed-size input in convolutional neural networks, enabling models to better handle images of varying sizes.
Modern models are increasingly focusing on balancing real-time performance and accuracy. YOLO[2]has achieved real-time object detection while maintaining high accuracy by simplifying the network structure, which is of great significance for practical applications.

3 Method

In this study, we selected the pre-trained ResNet152 model and fine-tuned it to improve it's classification accuracy on the dataset.
First, all images in the base dataset were mixed and processed then divided into two parts–training data and test data in a ration of 8: 2, making them an independent data set to test model performance

during training and to prevent overfitting. We process data with attempts at random augmentation, including random cropping and random erasing, making the training results stable and improving it's robustness.
Then we choose two models for comparison,ResNet152 and VGG11.
We selected Adam optimizer for its adaptive learning rate characteristics which is suitable for optimizing non-stationary objectives and sparse gradients. The learning rate change strategy we adapt raises the learning rate at the beginning of training and then gradually lowers it later to tune model parameters. The model can converge quickly in the begining of training and avoids divergence.
By doing so we are supposed to gain a model that can study quickly with stability.

4 Experiments

Model	**ePoch**	**Training Accuracy (%)**	**Validation Accuracy (%)**
Resnet152	0	79.75	80.14
	1	79.96	80.14
	2	79.96	80.14
	3	79.96	80.14
VGG11	0	69.96	65.77
	1	67.80	64.00
	2	65.40	64.34
	3	63.00	62.00

The experimental results indicate that the recognition accuracy of the model is superior to that of another model, and it also converges more rapidly. The adjusted model demonstrates better performance in all aspects. The training loss of the model decreased from 0.5257 to 0.5037, while the

Epoch	Training Accuracy	train Loss
1	0.7825	0.5257
2	0.7996	0.5081
⋮	⋮	⋮
10	0.7996	0.5092

Table 1: Training and Validation Accuracy

validation accuracy stabilized at 0.7996. This indicates that the model is capable of generalizing well without showing signs of overfitting. Under the conditions of this experiment, it can be concluded that the ResNet152 model is effective. and the model also has further potential for improvement through methods such as adjusting the learning rate.

References

[1] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587. IEEE, 2014.

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 779–788. IEEE, 2016.

[3] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3431–3440. IEEE, 2015.

[4] K. He, X. Zhang, S. Ren, and J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. In *Proceedings of the European Conference on Computer Vision*, pages 99–115. Springer, 2014.