Comparison of CNN based Illustration Drawing-Style Classification Systems

宋 航宇† 渡辺 裕†
Hangyu SONG † Hiroshi WATANABE †
†早稲田大学基幹理工学研究科
†Graduate School of Fundamental Science and Engineering, Waseda University

Abstract A series of 6-class CNN based drawing-style illustration classification systems with different structures were built and trained with same dataset. It was proved that CNN can be used for recognizing vague features such as drawing-style and for simple classification task with small dataset, the scale of network needs to be controlled.

1. Introduction

In recent years, CNN was used in classification systems in computer vision fields. In LSVRC-2010, AlexNet improved that CNN worked better than the previous method in image classification [1]. Later, VGG [2]and Inception (GoogleNet) [3] were introduced followed by ResNet [4] and other structures. With the development of the CNN, the accuracy of image classification grows rapidly.

To a large extent, drawing-style is a feature that depends on the habits of the author of the works. For example, the curls of hair, the way to draw the shadow in eyes and the shape of face. For different authors, their works may look different which suggests that illustrations may be classified by authors with different drawing-style.

2. Previous Research

Although the CNN had already been used for image classification with database Cifer10, ImageNet and Coco and got obviously satisfying results. The classification task was hardly done with CNN with the so called “Nijigen” (日本語：二次元) works (e.g. manga and illustrations).

A series of manga facial detection research proved that CNN can be used in the feature extracting of Nijigen works [5]. This suggests that it may be possible that CNN can be used to recognize some important feature of Nijigen works.

AlexNet that introduced in LSVRC 2012 was a milestone of CNN classification [1]. After AlexNet, by adding layers to the network to make it deeper, the VGG reached a higher accuracy in 2014 [2]. Google also introduced Inception (GoogleNet) which made the network wider by using more convolutional kernels with smaller size. This also made the performance of CNN better [3].

Mathematically, training a network is to make the loss of network reduce by adjusting the parameter of the network. In order to take use of the residual loss, ResNet was introduced. By combining ResNet and inception, Inception-ResNet was created [6].

3. Method

Although the CNN based classification systems reached the new height of accuracy when every time a new method was introduced. For a certain task, the best choice of the network structures varies for different conditions. Therefore, different networks were used to do the classification task. Then, the accuracy and the training process was compared for different method to decide the best choice for this task.

4. Experiments and Results

In the first stage, after 30000 steps’ training (with batch size of 32), the simple CNN model reached the accuracy at 78.53%, which was obviously higher than the mathematical expected random choices result within 6 classes (which should be about 16.67%). This proved that CNN was a valid method in the drawing style classification of illustrations.

In the second stage, 4 sets (from Set 0 to Set 3) of 10-layer CNN were built with different structure. In the previous study, one way to accelerate the training and improving the accuracy is batch normalization, which was used to overcome overfitting [7] instead of dropout. The other way to accelerate the training was to use Adam optimizer instead of SGD [8]. However, Adam was also reported to make the training unstable in some cases [9]. Also, “full convolution” (use a convolution layer with kernels with the same size of one channel of inputting tensor) worked well in some cases [10]. In this study, we also used this method in the experiment.

The result of experiment shows that, for all the sets, the accuracy obviously increased (lowest accuracy 83.03%, compared with 78.53% of simple CNN). Set 2 reached the highest accuracy at 91.87%.

After 30 hours’ training on 2 Nvidia P100, the accuracy of inception still wandered around 85% and the final accuracy was 87.90%. Compared with inception V4, the performance of inception-ResNet-V2 was better. After 30 hours’ training, accuracy reached 90.47%. However, this value was still lower than that of 10-Layer VGG.

5. Conclusion

As a conclusion, CNN can be used to recognize the drawing-style features of illustrations.

However, the fact that the accuracy of inception and inception-ResNet showed that, for small dataset (a dataset with 750 images was small compared with dataset such as ImageNet and CoCo), the scale of network needs to be controlled properly.

References

[1] A. Krizhevsky, I. Sutskever, G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25*, 2012.

[2] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*, abs/1409.1556, 2014

[3] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich. Going Deeper with Convolutions. *CoRR*, abs/1409.4842, 2014

[4] K. He, X. Zhang, S. Ren, J. Sun. Deep Residual Learning for Image Recognition. *CoRR*, abs/1512.03385, 2015

[5] H. Yanagisawa, T. Yamashita, H. Watanabe. A Study on Object Detection Method from Manga Images using CNN, in *International Workshop on Advanced Image Technology (IWAIT2018)*, Chiang Mai, No.16, pp.1-4, Jan. 2018

[6] C. Szegedy, S. Ioffe, V. Vanhoucke, A. Alemi. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning, *CoRR*, abs/1602.07261, 2016

[7] S. Ioffe and C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, *CoRR*, abs/1502.03167, 2015

[8] D. P. Kingma, J. Ba. Adam: A Method for Stochastic Optimization, In 3rd *International Conference for Learning Representations*, San Diego, 2015

[9] M. Arjovsky, S. Chintala and L. Bottou. Wasserstein GAN, *CoRR*, abs/1701.07875, 2017

[10] J. Long, E. Shelhamer, T. Darrell. Fully Convolutional Networks for Semantic Segmentation, *CoRR*, abs/1411.4038, 2014