Destylization of text with decorative elements-Supplementary Materials

Yuting Ma

NLPR, Institute of Automation, CAS
School of Artificial Intelligence, UCAS
mayuting2018@ia.ac.cn

Weiming Dong*

NLPR, Institute of Automation, CAS
CASIA-LLVision Joint Lab
weiming.dong@ia.ac.cn

Fan Tang*

School of Artificial Intelligence, Jilin University
tangfan@jlu.edu.com

Changsheng Xu

NLPR, Institute of Automation, CAS
CASIA-LLVision Joint Lab
csxu@nlpr.ia.ac.cn

1 U-NET AND RESNET

U-net is mainly used to segment tasks, ResNet is mainly used in areas such as target classification and as part of classical neural network, the backbone of computer vision tasks. We used U-NET and ResNet as network generators respectively, and it can be seen from the experimental results that ResNet is more suitable for destylized tasks. Figure 1 and Figure 2 compare the results in Chinese and English, respectively.

Figure 1: The picture shows the contrast between the effects of Chinese characters. The first line is a style image with decorative elements. The second line is the result of processing U-Net as a generator. The third line is the result of processing ResNet as a generator.

2 AUXILIARY CLASSIFIER

In order to improve the effect of the model, we added an auxiliary classifier structure to the model to help the model better process the pixel details of text edges. We use convolutional layer and multilayer perceptron as auxiliary classifier structures and use cross entropy loss function to train the classifier. Based on the content of the text as the basis for classification, the same text content, different font and style as one. Figure 3 and Figure 4 are the comparison results of Chinese and English with and without auxiliary classifiers.

*Co-corresponding authors

Figure 2: This is a demonstration of the English effect. The first line is the original style image. The second line is the result of processing U-NET as a generator. The third line is the result of processing ResNet as a generator.

Figure 3: The first line is the original style image. The second line is the result of having no auxiliary classifier. The third line is the result of adding an auxiliary classifier.

3 THE NUMBER OF RESIDUAL BLOCKS

We also considered the different depths of the ResNet network and selected the appropriate number of residual blocks through experiments. For the generator with ResNet as the basic model, 4

Figure 6: The first line is the original style image. The second line is the result of 4 residual blocks. The third line is the result of 8 residual blocks. The fourth line is the result of our model (6 residual blocks).

Figure 7: The first line is the original style image. The second line is the result of single scale training. The third line is the result of multi-scale training.

Figure 8: The first line is the original style image. The second line is the result of single scale training. The third line is the result of multi-scale training.

Figure 4: The first line is the original style image. The second line is the result of having no auxiliary classifier. The third line is the result of adding an auxiliary classifier.

residual blocks, 6 residual blocks and 8 residual blocks were used for ablation experiments, and the experimental results were shown in Figure 5 and Figure 6. Through experiments, we find that the ResNet network with 6 residual blocks is most suitable for the task in this paper.

Figure 5: The first line is the original style image. The second line is the result of 4 residual blocks. The third line is the result of 8 residual blocks. The fourth line is the result of our model (6 residual blocks).

4 SINGLE SCALE TRAINING AND MULTIPLE SCALE TRAINING

During training, we used two different methods to train the network to prove that the training strategy we used was more effective. First,

a patch 256 pixels in size is used to train the network. Then, three different scales of patch training network of 64,128,256 are used. Figure 7 and Figure 8 are the comparison results of Chinese and English of the two different methods.