Table 1 evaluation results of experiments on DigitAnatomy dataset

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **NUM** | **AUC(%)** | **Acc(%)** | **F1(%)** | **NUM** | **AUC(%)** | **Acc(%)** | **F1(%)** |
| Digit\_exp1 | 66.97 | 63.30 | 59.89 | Digit\_exp7 | 72.96 | 68.80 | 61.10 |
| Digit\_exp2 | 63.96 | 61.40 | 47.84 | Digit\_exp8 | 68.17 | 65.20 | 58.17 |
| Digit\_exp3 | 72.15 | 68.50 | 63.16 | Digit\_exp9 | 67.69 | 64.60 | 64.10 |
| Digit\_exp4 | 66.15 | 62.80 | 60.00 | Digit\_exp10 | 61.08 | 58.00 | 54.35 |
| Digit\_exp5 | 60.27 | 59.40 | 50.73 | Digit\_exp11 | 62.97 | 60.50 | 53.36 |
| Digit\_exp6 | - | - | - |  |  |  |  |

Digit\_exp1 is regarded as an original experiment with the best parameter values which have been set by the author.

For a missing initial value in\_channels in the code, three experiments are implemented (exp1, exp2, exp3). With the value of in\_channels set to 3 in exp3, the nearly best evaluation result on DigitAnatomy in my experiments is obtained.

Exp4 shows an experiment where the model is trained for 1000 epochs. The increase in epochs to 1000 is a suggestion from the author, which can help make the results more stable.

Experiments Exp5, Exp6, Exp7, and Exp8 were implemented with different learning rates: 0.001, 0.1, 0.00001, and 0.000001, respectively. In exp6, the training process was interrupted, so the evaluation process could not be completed. The reason is that an initial learning rate was set too large, which leads to numerical overflow and a very unstable training process. Gradient explosion makes the value of loss function very large, which causes the model fails to converge.

Experiments Exp9, Exp10 and Exp11 were implemented with different batch size: 32(with test batch set to 64), 8, 32(with test batch set to 2), respectively. A larger batch size makes the training process more stable. The stability can be observed in loss curves. From the generated images in exp10, we can observe that the image quality cannot be improved with a relatively small batch size. In exp9, the image quality is better.

Table 2 evaluation results of experiments on ZhangLab dataset

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **NUM** | **AUC(%)** | **Acc(%)** | **F1(%)** | **NUM** | **AUC(%)** | **Acc(%)** | **F1(%)** |
| Zhang\_exp1 | 85.23 | 79.33 | 84.44 | Zhang\_exp9 | 83.64 | 79.17 | 84.30 |
| Zhang\_exp2 | 87.01 | 81.09 | 85.61 | Zhang\_exp10 | 82.74 | 76.76 | 82.25 |
| Zhang\_exp3 | 86.49 | 81.57 | 85.68 | Zhang\_best2 | 81.66 | 74.52 | 80.05 |
| Zhang\_exp4 | 84.61 | 76.92 | 80.17 | Zhang\_exp11 | 77.54 | 75.32 | 81.84 |
| Zhang\_exp5 | 81.78 | 77.72 | 83.19 | Zhang\_exp12 | 80.58 | 76.28 | 82.79 |
| Zhang\_exp6 | 83.25 | 76.76 | 82.42 | Zhang\_exp13 | 82.15 | 77.88 | 82.79 |
| Zhang\_exp7 | 82.99 | 75.80 | 80.16 | Zhang\_exp14 | 71.81 | 71.96 | 78.79 |
| Zhang\_exp8 | 72.78 | 73.24 | 79.81 |  |  |  |  |

The experiments on this ZhangLab dataset generally performed better than those on DigitAnatomy dataset. Hence, the generated image is better for observing. Zhang\_exp1 is regarded as an original experiment with the best parameter values which have been set by the author. Experiments Exp1, Exp2, Exp3 were implemented with different training epochs: 200, 500, 1000, respectively. After 500 and one 1000 epochs, the evaluation results look like similar. But the quality of generated image is better in exp3 than in exp2. In addition, the experiments for five loss weights are implemented to explore how these loss weights affect the evaluation results and quality of generated images. In exp1, exp4, exp5 and best2, the distillation weight is respectively set to 0.001, 0.1, 0.0001 and 0.01. With the decrease in distillation weight, the quantity of generated images is becoming worse. In exp1, exp6 and exp7, the teacher weight is respectively set to 0.01, 0.1 and 0.001. The experiment exp1 obtain the best evaluation result and best generated image. In exp1, exp8, exp9 and exp10, the generator weight and discriminator weight are respectively set to 0.005, 0.05, 0.0005 and 0.001. The reason why the two weights are always set equal is to ensure that the updates of the generator and the discriminator are coordinated at each step of the training process so that the training process is stable and effective. Keeping the weights equal helps balance this adversarial relationship. The experiment exp1 obtain the best evaluation result and best generated image as well. In the experiments from 11 to 14, the reconstruction weight is changed. The high quality of generated images and the reconstruction weight are particularly relevant. In the weight range of 5-20, we might get good results. With the weight set to 10, we can obtain the best result so far.

Code Extension

I wrote some code for visualizing the changes in losses and standard metrics during training process. By plotting these curves, we can directly monitor these values and then decide which parameters should be changed in the next experiment.

图形用户界面, 文本

描述已自动生成 文本

描述已自动生成