Project ID: 32

2017-18

Xi'an Jiaotong-Liverpool University
西交利物浦大学

Investigation of GAN for Character Synthesis

WANG Chenhui

Department of Electrical and Electronic Engineering

Supervisor: Kaizhu Huang

Abstract— In this project, Generative Adversarial Network (GAN) is used for Chinese character generation, particularly with ConvNets. The novel architecture could learn handwriting style and generate more characters that resemble the inputs. The introduction of category and character embedding allow the model to learn from rich latent space, and some tricks intend to speed up model convergence are used. However, the required training set is too large for handling engineering issues.

Introduction

Generative Adversarial Network (GAN), based on the game theory, is comprised of generator and discriminator. **G** is trained to generate fake images that fool discriminator, while the task for **D** is to maximize the probability of images are from true data distribution. In this project, a modified GAN architecture is introduced which could learn handwriting style of Chinese characters. The generator is comprised of classical **encoder-decoder** architecture, which has 8 layers of **ConvNets** respectively.

On the deepest convolutional layer, character and category embedding are concatenated to handle multiple fonts. One character may appears in different font, so a non-trainable Gaussian noise is concatenated as style embedding to the character embedding. The generated characters are very similar to the real characters.

Methodology

In generator, the deepest layer contains the core features of source image (SIMSUN), therefore the style of handwriting could be embedded for character synthesis. Then the fake image (generated) is reconstructed by transpose convolution. The generator is classical encoder-decoder architecture:

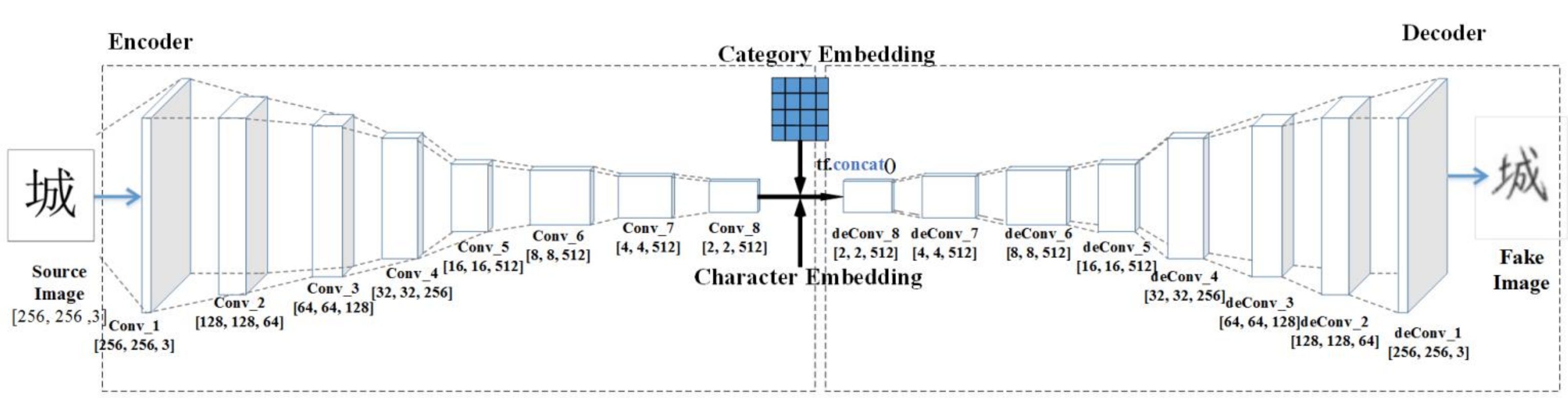

Figure 1. Generator architecture of GAN

The generated (fake) image is then fed into discriminator for adversarial training. The architecture of whole GAN is

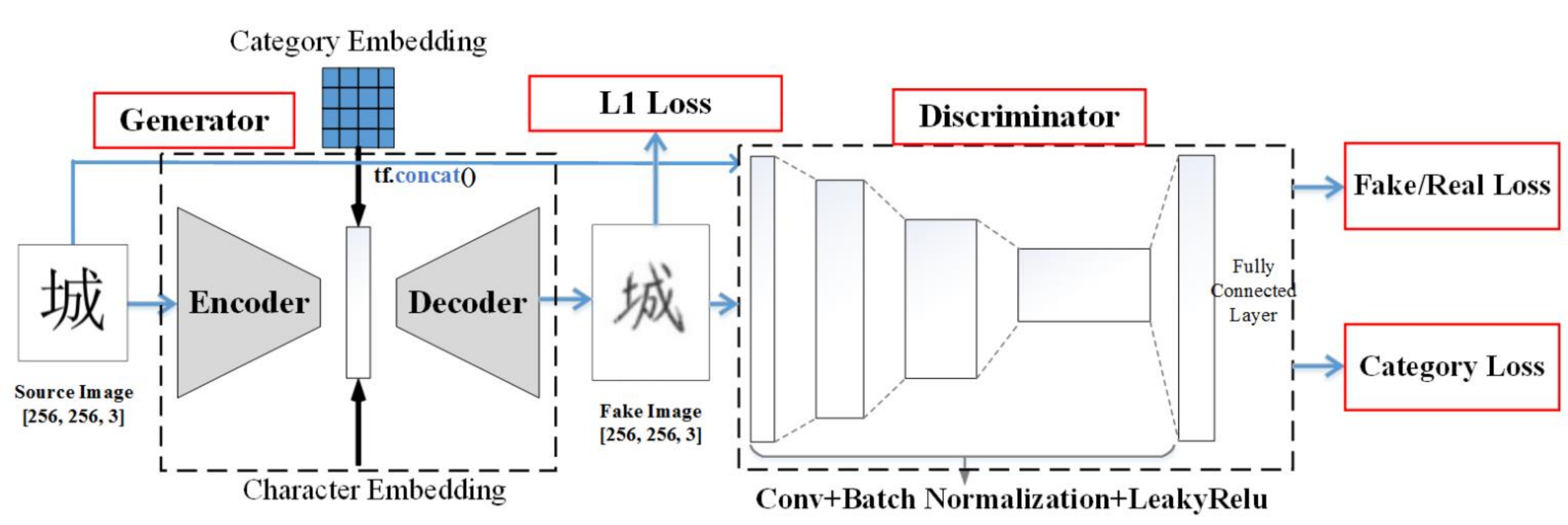

Figure 2. Modified Generative Adversarial Network

During training, the generator and discriminator losses are set as:

$$L_{GAN}(G,D) = E_{x,y\sim p_{data}(x,y)}[logD(x,y)] + E_{x\sim p_{data}(x),z\sim p_z(z)}\log[1 - D(x,G(x,z))]$$

To improve the convergence speed, L1 loss is introduced:

$$L_{L1}(G) = E_{x,y\sim p_{data}(x,y),z+p_z(z)}[\|y - G(x,z)\|]$$

Total Loss:

$$G^* = arg\min_G \max_D L_{GAN}(G,D) + \lambda L_{L1}(G)$$

Results and Discussion

After 1200 epochs training on single Titan X, the network is able to generate characters that similar to the real ones.

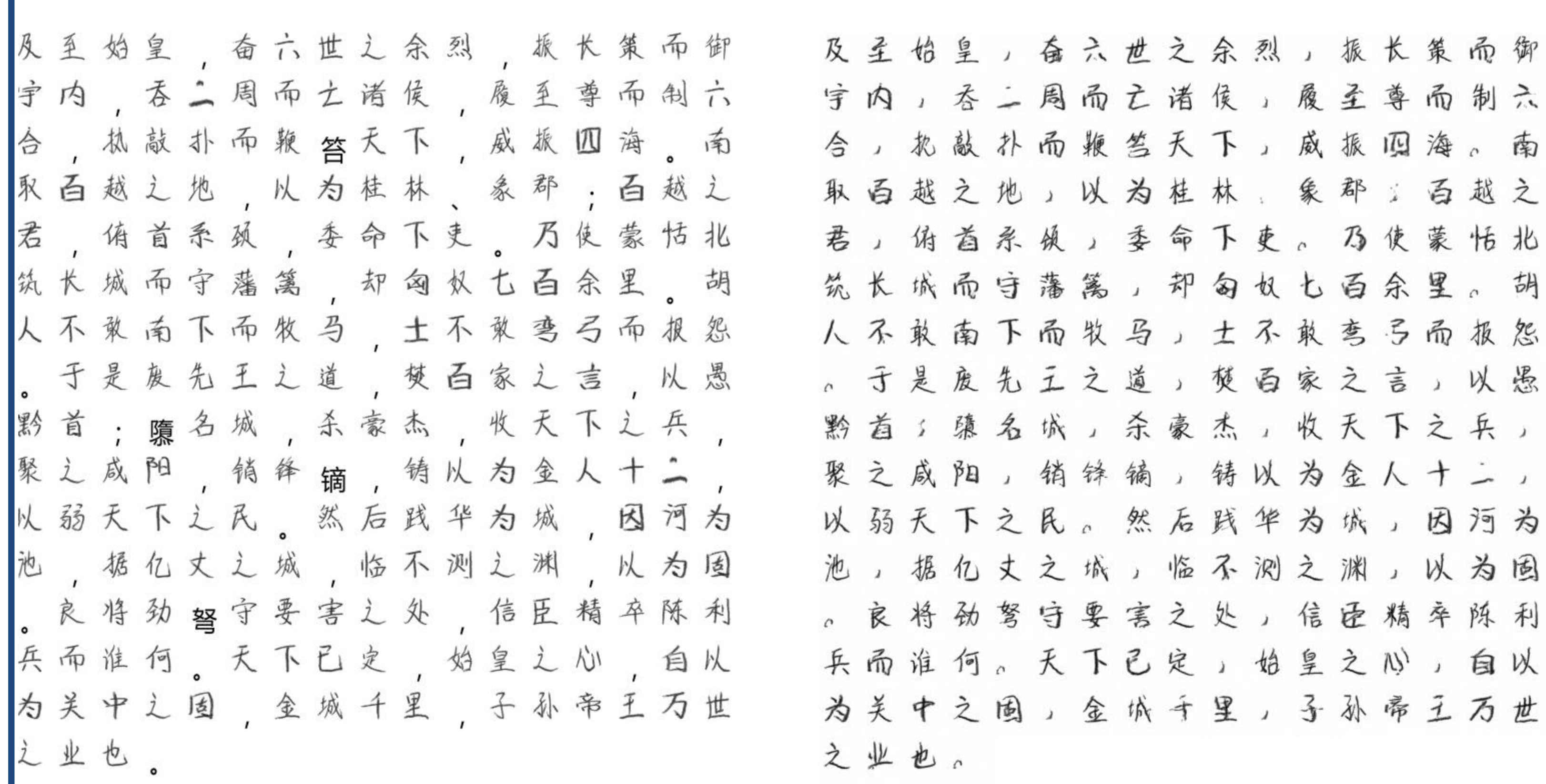

Figure 3. Generated characters (left) and real characters (right)

The network has good performance, however, some characters are hard to learn so they keep the same as source images. The loss functions of generator and discriminator are visualized:

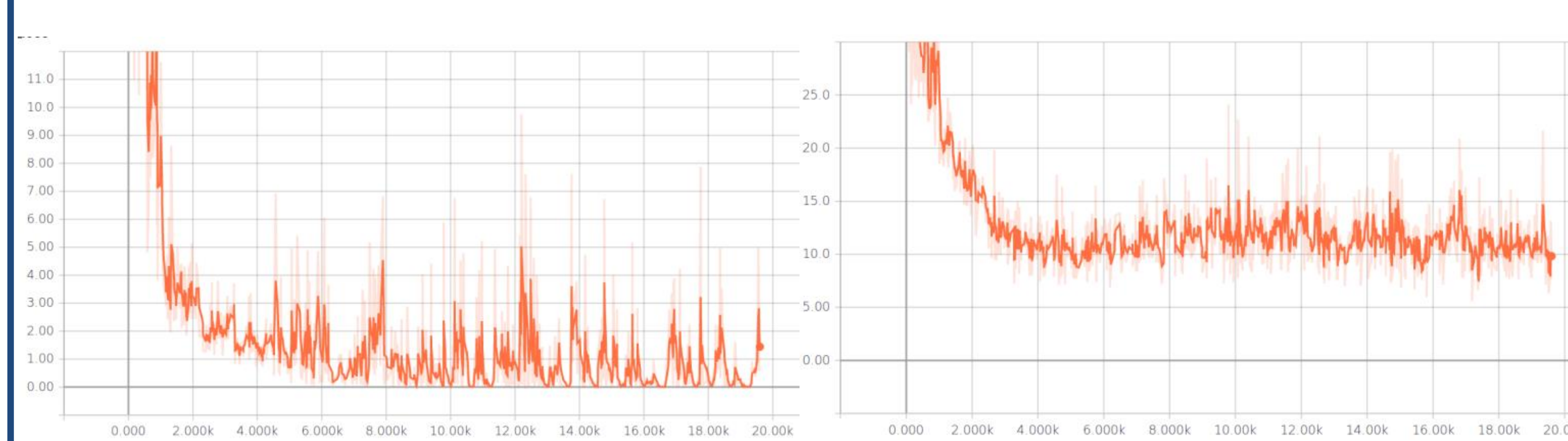

Figure 4. Loss functions of discriminator (left) and generator (right)

The discriminator converges quickly and approaching to 0, and the generator oscillates around some positive value. This may due to insufficient training or model instability. To avoid sparse gradients, ReLU activation function is replaced with LeakyReLU. Batch normalization is applied during reconstruction of images.

Conclusion and Future Work

The generated characters are very similar to ground truth, however, with some tiny defects. These unsuccessful samples may caused by the unstable discriminator and generator. Ideally, the discriminator and generator should converge to zero with no oscillation. From tensorboard visualization, the losses are unstable. To overcome this issue, loss functions are required to be reconsidered.

Although the model performance is good, it requires large training set. In this project, 2000+ samples obtained from open source database are used for training. It is impractical for an individual who wants to generate characters that similar to his handwriting. The next step will be minimize the training set, which requires a novel design of Generative Adversarial Network.

Selected References

[1]P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. *Image-to-image translation with conditional adversarial networks In CVPR, 2017.*