Few-shot Font Style Transfer between Different Languages Supplementary Material

Chenhao Li, Yuta Taniguchi, Min Lu, Shin'ichi Konomi
HDI Lab, Kyushu University
{li.chenhao.995@s, taniguchi@ait, lu@artsci, konomi@artsci}.kyushu-u.ac.jp

Abstract

In this supplementary material, we first introduce the detailed network structure and some hyperparameter settings. We also provide additional visual comparison results between our model and EMD [4], and DFS [5]. Finally, we illustrate some examples of the font dataset that we constructed.

1. Network Structure

In this section, we introduce detailed hyperparameter settings and some network structures that are not explained before. Our basic setup follows Pix2Pix [2]. Both Generator G, Content Discriminator D_{content}, and Style Discriminator D_{style} are initialized with Normal Initialization. We train our model 20 epochs by using Adam optimizer [3] with β_1 = 0.5, β_2 = 0.999, and learning rate lr = 0.0002 in the first 10 epochs and an linear decay in the remaining 10 epochs. Note that for both G, D_{content}, and D_{style}, we use the same settings. Empirically, we set batch size to 256. Besides, unlike previous work [1] employed dropout to their generator to obtain randomness. Here, we don't use dropout because we have observed in experiments that this behavior will reduce the generative ability of the model. Instead, we add some slight random noise to the style code z_s.

In addition, we illustrate the detailed network structure of content encoder, decoder, and discriminators in Figure 1.

2. Visual Comparison

We illustrate the additional results in Figure 2. For each font, we randomly select 6 generated images. Experimental results show our method outperforms the other two methods [4, 5] on both printing fonts and handwritten fonts.

3. Font Dataset

In this section, we introduce the font dataset that we constructed for experiments in detail. As mentioned before, the dataset includes 847 gray-scale fonts each with approximately 1000 commonly used Chinese characters and 52 English letters in the same style. In addition, we use a ordinary font Microsoft YaHei as the content image. All content images are binary images with the white pixels equal to 1 and black pixels equal to 0, and it is only used for indexing the category of the character. We process the dataset by finding a bounding box around each glyph and resize it so that the larger dimension reaches 64 pixels, then pad to create 64×64 glyph images. Before inputting to the model, we will normalize the image so that all pixel values are in the range of -1 to 1. Figure 3 shows additional examples of the multi-language dataset. We randomly select 30 fonts and illustrate them. For each font, we choose 6 English letters and 20 Chinese characters as reference.

References

[1] Samaneh Azadi, Matthew Fisher, Vladimir G Kim, Zhaowen Wang, Eli Shechtman, and Trevor Darrell. Multi-content gan for few-shot font style transfer. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7564–7573, 2018.

[2] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[3] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[4] Yexun Zhang, Ya Zhang, and Wenbin Cai. Separating style and content for generalized style transfer. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8447–8455, 2018.

[5] Anna Zhu, Xiongbo Lu, Xiang Bai, Seiichi Uchida, Brian Kenji Iwana, and Shengwu Xiong. Few-shot text style transfer via deep feature similarity. *IEEE Transactions on Image Processing*, 2020.

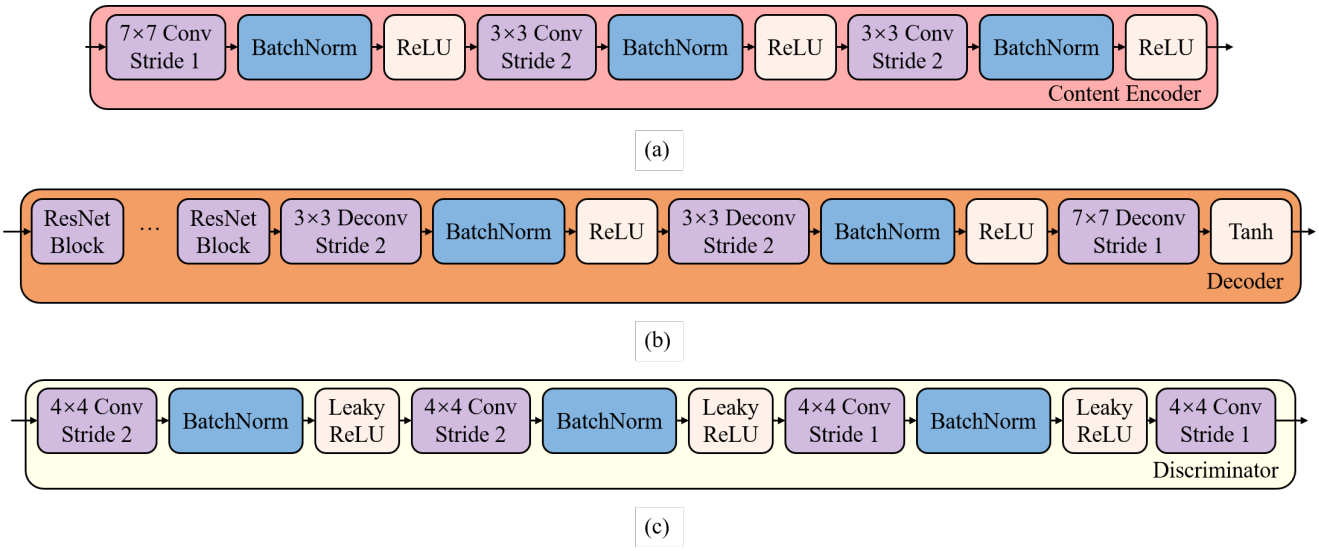

Figure 1. (a) Architecture layout of the proposed content encoder. (b) Architecture of the proposed decoder. (c) Architecture of the proposed discriminators

Figure 2. Additional visual comparison of our FTransGAN (**4th rows**) with EMD [4] (**2nd rows**) and DFS [5] (**3rd rows**), the observed style images are illustrated in the **1st rows** and the ground truth images are in the **5th rows**. For each font, we randomly select 6 generated images as reference.

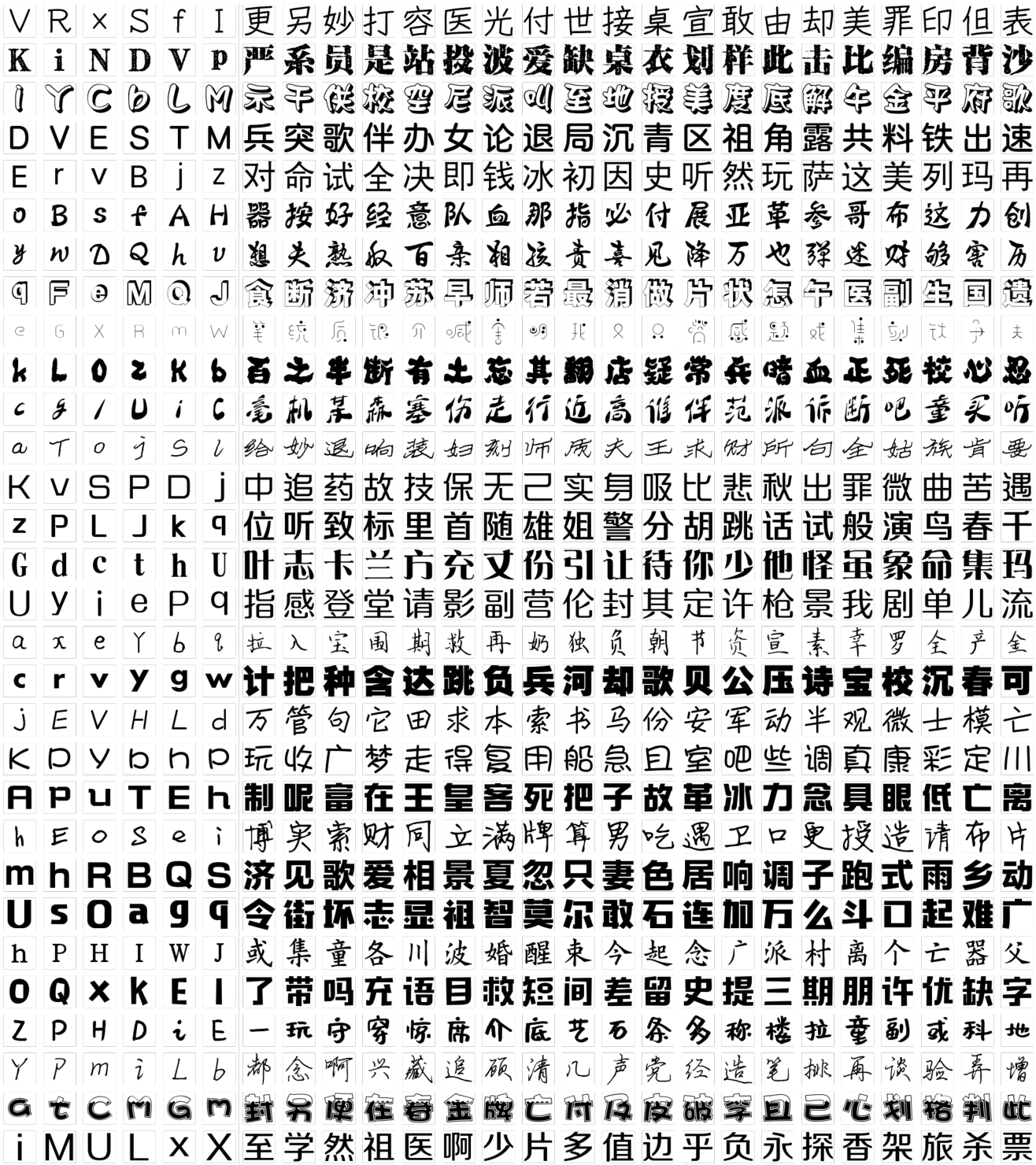

Figure 3. Additional examples of the font dataset that we constructed for our experiments. Each row represents a font, we randomly select 6 English letters and 20 Chinese characters as reference.