Supplemental Material for Aesthetic Text Logo Synthesis via Content-aware Layout Inferring

Yizhi Wang[1], Guo Pu[1], Wenhan Luo[2], Yexin Wang[2], Pengfei Xiong[2], Hongwen Kang[2], Zhouhui Lian[1*]
[1]Wangxuan Institute of Computer Technology, Peking University
[2]PCG, Tencent

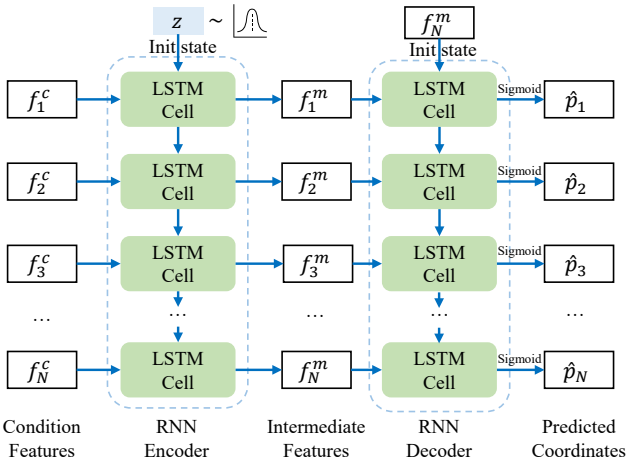

Figure 1. The detailed architecture of the coordinate generator G.

1. Architecture of the Coordinate Generator

In Figure 1, we give a detailed illustration of the coordinate generator G, which was not presented in our manuscript due to the limitation of paper length. In vanilla conditional GANs [1], z is fused with the condition by convolution layers or fully connected layers. In our task, we employ an RNN encoder whose initial state is set as z, to encode the condition input $\boldsymbol{f}^c$. By this means, the latent noise z is fused into each position of the intermediate features $\boldsymbol{f}^m$, which are further utilized to predict the coordinates $\hat{\boldsymbol{p}}$.

2. More Synthesis Results

In Figure 2, we show more synthesis results of our model, demonstrating that our synthesized results are visually pleasing and possess diverse styles at the same time.

*Corresponding author. E-mail: lianzhouhui@pku.edu.cn

3. More Ablation Results

In our manuscript, we mentioned that the synthesized layouts usually do not conform with the text semantics if we do not encode the text linguistics information as inputs, e.g., starting a new line that breaks a token. It can be somehow verified by Figure 8 in our manuscript, where the layout for one text is not suitable for another text with the same font. In Figure 3, we demonstrate more results to prove this claim. When the text information is not exploited, the tokens "强大" and "脚尖" are split into two lines, which does not conform to the design rules. Besides, the sizes of the characters "如" and "果" in a token are dramatically different, which affects the readability of synthesized results.

4. Visualization of Latent Space

We visualize the latent space of $z \in \mathbb{R}^{128}$ by PCA (Principal Component Analysis), and the results are shown in Figure 4. Specifically, for each case in the testing dataset (about 300 examples), we randomly sample the latent code 10 times to generate different layouts. Afterwards, we perform PCA on the 300×10 vectors to reduce the dimension of them and annotate some of them on a 2D space. We can find that similar layouts are located closer while dissimilar layouts are located farther away. For example, (1) vertical layouts (B2, C2, H2, E3) tend to locate in the left parts; (2) horizontal layouts (A1-E1, H1, G2) tend to locate in the centre and upper parts; (3) multi-line layouts (A2, D2, E2, F2) tend to locate in the bottom-right parts; (4) irregular layouts (F1, G1) tend to locate in the borders of distribution. The latent noise z is orthogonal to the length of input, For example, the lengths of B2, C2 and H2 are different but they present the same layout style (vertical), which verifies the effectiveness of our model design in Figure 1. Through the visualization method, we can guide the designers to explore the latent space for selecting their favorite layouts.

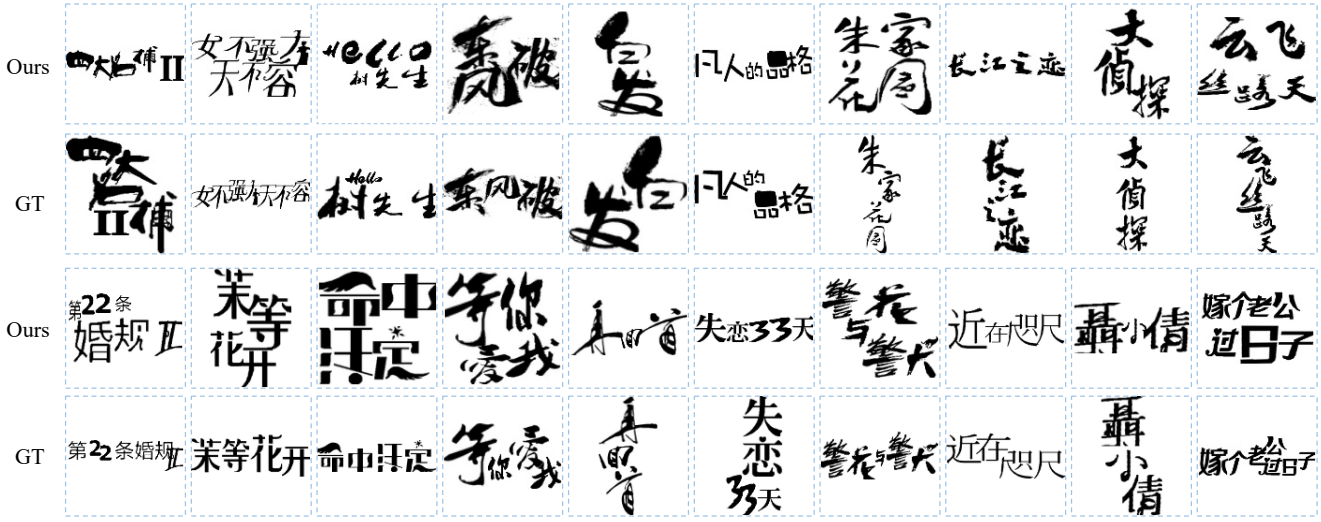

Figure 2. More synthesis results of our network. "GT" denotes ground truth (human-designed).

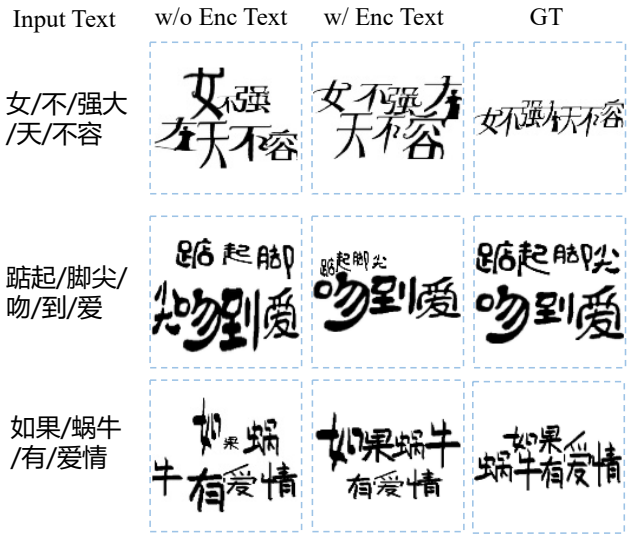

Figure 3. Ablation results of encoding text linguistic information ("Enc Text"). '/' is the symbol for splitting tokens.

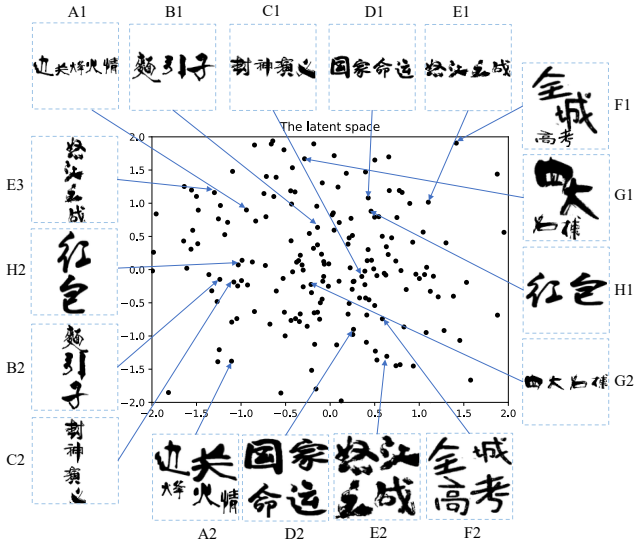

Figure 4. The visualization results of the latent noise z. The boxes annotated with the same starting letter (e.g., A1, A2 or E1, E2, E3) denote the different synthesized layouts for the same input content.

5. Font Generation

We adapt the network proposed by [2] to synthesize new fonts and predicting their attributes simultaneously. We collect 400 fonts and their attribute annotations (one-hot) from Internet. The network architecture of font generation is shown in Figure 5, where a font style latent space is learnt and we can sample from it to generate new fonts and derive the corresponding font attributes.

References

[1] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. 1

[2] 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. 2

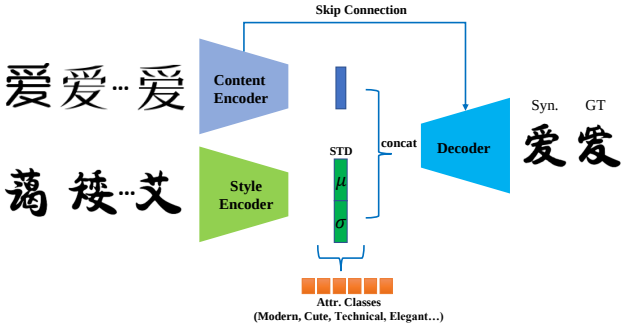

Figure 5. Synthesizing Chinese fonts with attribute predictions.