sification tasks using the style embedding. Specifically, we train a 3-layers MLP with the style embedding we get in different methods, including our SA-VAE, vanilla VAE and VAE-GAN [16], in which the encoder network and decoder network of Vanilla VAE and VAE-GAN have the same setting as the SA-VAE. To simplify the training, we only use 100 unique characters per style as the training data to do style classification and use 100 styles per character to do character content classification, then test the prediction accuracy over the rest of samples. The results are shown in Table 1.

Table 1: Classification accuracy using the style embedding

Methods	Style Accuracy	Content Accuracy
Vanilla VAE	43.44%	28.55 %
VAE-GAN	38.76%	28.24%
SA-VAE	**47.42**%	**1.28** %

We can see that compared with two other models, the style embedding extracted by our SA-VAE contains very little content information but more style information, which means that our SA-VAE can mainly extract the style features from characters and lead to a better result on one-shot generation. Fig. 11 provides a more powerful proof that our intercross pair-wise optimization method provides a powerful ability of disentanglement. Except our SA-VAE, the generation results produced by other two methods are not meaningful characters at all, since they carry too much content information from the style specified characters.

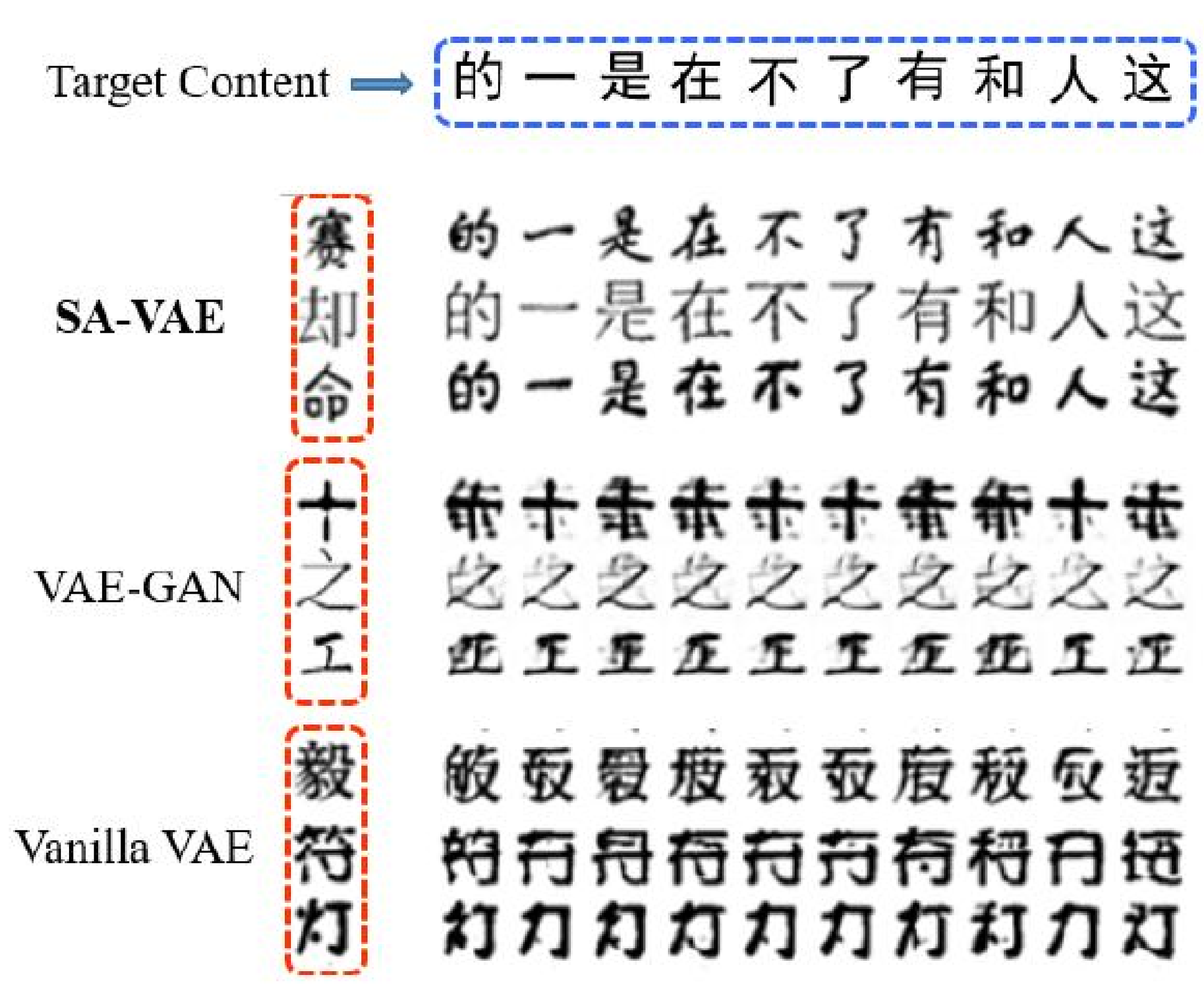

Figure 11: We show the one-shot generation results produced by three models. The style specified characters are in the left red dotted frame and the content specified characters are in the up blue dotted frame.

3.5. Structure Knowledge of Chinese Characters

Our encoding method with human knowledge instead of the one-hot embedding provides a solution for the large dictionary issue of Chinese characters. To show the effectiveness of knowledge guidance, we compare the converge speed of three encoding methods – our content code, the one-hot embedding and the binary embedding. The results are shown in Fig. 12.

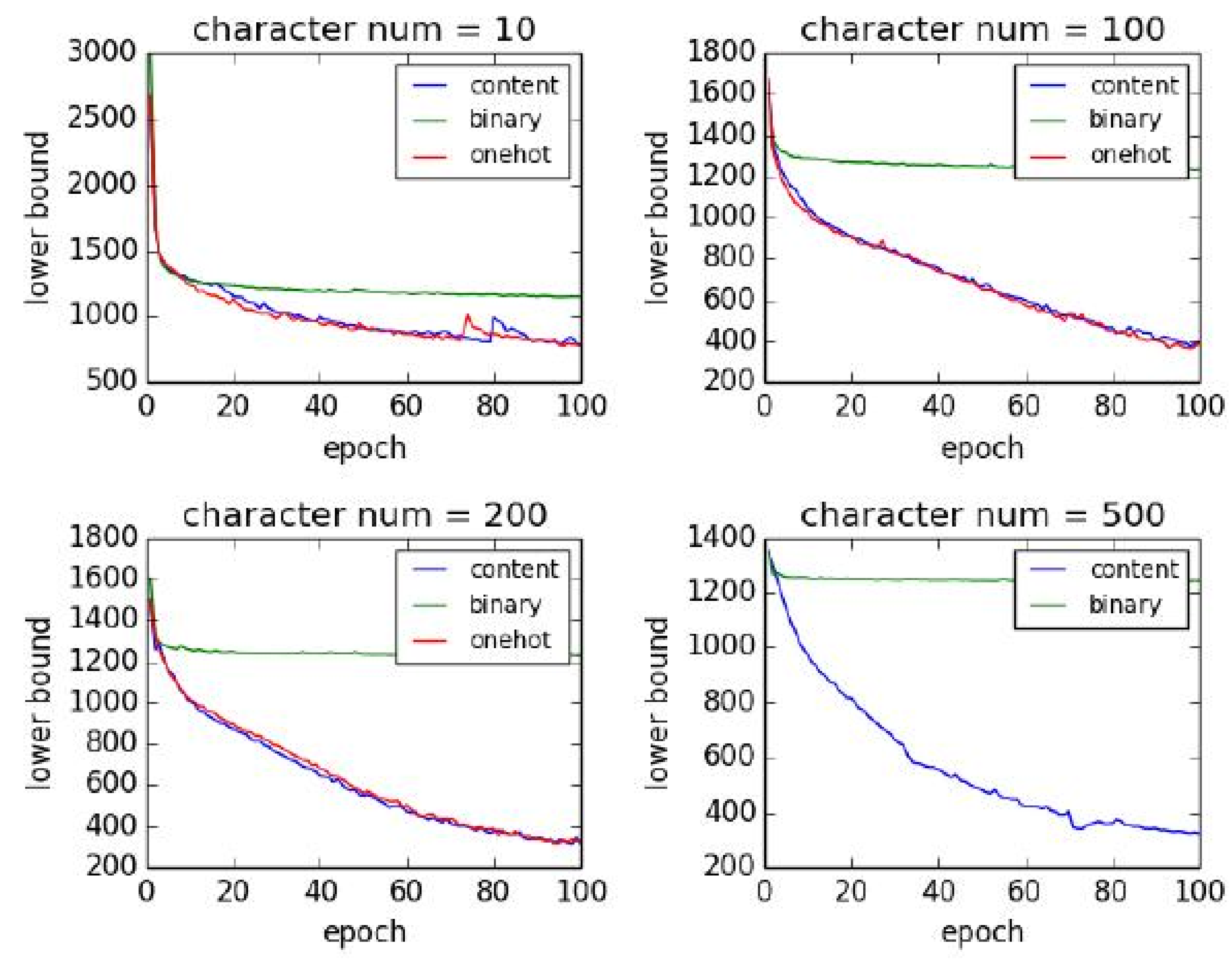

Figure 12: The training curves of three encoding methods using different numbers of unique characters, where the y-axis shows the negative lower bound and the x-axis shows the number of epochs during training. Notice that when we add the character numbers to 500, the model parameters using one-hot embedding will explode and we cannot show the corresponding curve.

Depending on the cost of expanding model capacity, the one-hot embedding may provide a comparable converge speed with our knowledgeable content code. However, this embedding method will make the model parameters exploded in a larger Chinese characters dictionary. The binary embedding consumes shorter length, which is only $\lceil \log_2 N \rceil$ bits, but it hardly converges no matter there are numerous characters or only a few characters.

4. Conclusions and Future Work

We have presented a novel Style-Aware VAE framework, which first achieves the stylized Chinese character generation by reading only one or a few characters. Our model disentangles the style information and the content information with the intercross pair-wise optimization method and shows a powerful one-shot or few-shot generalization ability with unseen styles. Finally, we present the impressive generation results both with printing styles and handwritten styles.

For the future work, we may collect more styles to our style-bank and deploy the method to real applications.

References

[1] H. Attias. A variational baysian framework for graphical models. In *Advances in neural information processing sys-*