CalliPaint: Chinese Calligraphy Inpainting with Diffusion Model

Qisheng Liao[1], **Zhinuo Wang**[2], **Gus Xia**[1], **Muhammad Abdul-Mageed**[1,3]

[1]MBZUAI, [2]Brown University, [3]The University of British Columbia
[1]{Qisheng.Liao,Gus.Xia,Muhammad.Mageed}@mbzuai.ac.ae
[2]zhinuo_wang@brown.edu

Abstract

Chinese calligraphy can be viewed as a unique form of visual art. Recent advancements in computer vision hold significant potential for the future development of generative models in the realm of Chinese calligraphy. Nevertheless, methods of Chinese calligraphy inpainting, which can be effectively used in the art and education fields, remain relatively unexplored. In this paper, we introduce a new model that harnesses recent advancements in both Chinese calligraphy generation and image inpainting. We demonstrate that our proposed model CalliPaint can produce convincing Chinese calligraphy.

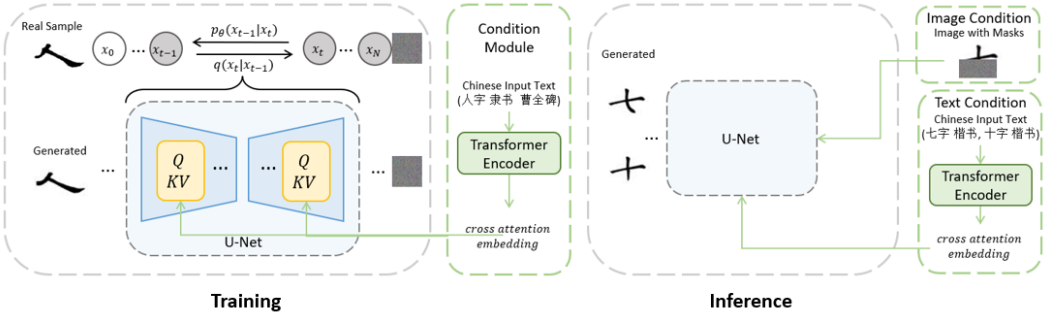


Figure 1: An illustration of our model. The left side displays the structure of our training procedure. During training, every image includes annotated text that showcases the character, script, and style of that specific image. The U-Net models are trained to acquire the ability to produce such images starting from Gaussian noise. The right side illustrates the process of inference. In this instance, we employ two distinct textual conditions while maintaining the same image condition. The resulting outcomes exhibit identical unmasked sections, but the masked portions are generated differently depending on the text conditions.

1 Introduction

Chinese calligraphy is a highly regarded and traditional art form that involves writing Chinese characters with a brush and ink. It is not just a means of communication but a visual art that has deep cultural and historical significance in China and other East Asian countries. Lately, there has been a noticeable trend in the use of machine learning for the creation of Chinese calligraphy. Examples include Zi2zi [1], CalliGAN [2], and ZiGAN [3], which typically employ a Generative Adversarial Network (GAN) architecture. While, Calliffusion [4], a new proposed work for Chinese calligraphy generation, uses Denoising Diffusion Probabilistic Models(DDPMs) [5] for generations.

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

	Regular		Semi-cursive		Cursive		Clerical		Seal		Total	
	Script	Character	Script	Character	Script	Character	Script	Character	Script	Character	Script	Character
Real Samples‡	0.91	0.93	0.83	0.81	0.88	0.68	0.96	0.83	0.97	0.81	0.88	0.78
Calliffusion‡	0.96	0.94	0.86	0.95	0.88	0.64	0.97	0.91	0.99	0.79	0.93	0.85
CalliPaint	0.98	0.96	0.95	0.99	0.99	0.93	0.99	0.98	0.99	0.90	0.98	0.95

Table 1: The performance of our generated data in different scripts in accuracy. ‡ indicate the results from Calliffusion.

	People know Calligraphy		People do not know Calligraphy		Total	
	Acc	P-Value	Acc	P-Value	Acc	P-Value
Type 1	0.28	0.94	0.22	0.94	0.24	0.99
Type 2	0.04	0.63	0.08	0.69	0.06	0.66

Table 2: The accuracy and p-value of each type of question in our survey.

DDPMs generate samples that match the data after a certain amount of time. In the forward diffusion process, a small amount of Gaussian noise is added to a data point sampled from a real data distribution in multiple steps, resulting in a sequence of noisy samples. While in the reverse diffusion process, the images are rebuilt based on the Gaussian distribution. Using the foundation of DDPMs, RePaint [6] is designed for the task of image inpainting. In RePaint, a novel scheduler is introduced, enabling image inpainting to be performed during the inference stage without the need for additional training.

In this work, we employ the Calliffusion model and RePaint scheduler to design a new framework that can do Chinese calligraphy inpainting with text conditions. This framework could be used not only in the art field but also in education.

2 Model Architecture

In training, we follow the settings of Calliffusion. Namely, we use a short description of Chinese text input to control the generations including the characters, scripts, and styles. In the inference phase, we need to provide an extra image with masks to guide the model to fulfill the masked parts. The details of our model architecture are introduced in Figure 1.

3 Evaluation

We assess our outcomes using a dual approach, considering both objective and subjective criteria.

In terms of objective evaluation, we employ the same pre-trained classifier that Calliffusion uses to identify the inpainted images. We choose the same conditions tested in Calliffusion but provide the gold images with random masks as image conditions. The findings presented in Table 1 demonstrate that, following the process of inpainting, the image quality from our model maintains a high standard. The classifier also successfully achieves accurate character predictions. Specifically, the average accuracy for inpainted images stands at 0.95, in contrast to the reported accuracy of 0.85 in Calliffusion. This difference in performance can be attributed to the fact that, in Calliffusion, the images under evaluation are entirely generated by the model. Meanwhile, in the case of inpainting, real images are extra conditions and the model only generates the masked parts, which is an easier task.

For subjective evaluation, we design a survey and invite 30 native Chinese speakers to answer. Twelve of them knew Chinese calligraphy before. The results in Table 2 indicate that it is hard for native Chinese speakers to distinguish the calligraphy inpainted by our model from real calligraphy. The examples of questions in our survey are shown in Figure 4 in Appendix B.

4 Conclusion

This paper represents a stride in the direction of employing diffusion models for Chinese calligraphy inpainting. The utility of inpainting extends across diverse domains. Nevertheless, we are currently confronted with certain drawbacks, notably the relatively slow inference time. We anticipate the resolution of this issue in forthcoming developments.

Ethical Implications

The ethical considerations of this project involve the intention to offer users an efficient tool for restoring missing elements in calligraphy images, which has broad applications such as correcting the inappropriate part of new learners' calligraphy and making it become an artwork written by famous calligraphers. The training data exclusively originate from publicly available websites, and the calligraphy art itself is written by ancient Chinese calligraphers.

References

[1] Yuchen Tian. zi2zi: Master chinese calligraphy with conditional adversarial networks, 2017.

[2] Shan-Jean Wu, Chih-Yuan Yang, and Jane Yung-jen Hsu. Calligan: Style and structure-aware chinese calligraphy character generator. *arXiv preprint arXiv:2005.12500*, 2020.

[3] Qi Wen, Shuang Li, Bingfeng Han, and Yi Yuan. Zigan: Fine-grained chinese calligraphy font generation via a few-shot style transfer approach. In *Proceedings of the 29th ACM International Conference on Multimedia*, pages 621–629, 2021.

[4] Qisheng Liao, Gus Xia, and Zhinuo Wang. Calliffusion: Chinese calligraphy generation and style transfer with diffusion modeling. *arXiv preprint arXiv:2305.19124*, 2023.

[5] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.

[6] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. Repaint: Inpainting using denoising diffusion probabilistic models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11461–11471, 2022.

A Qualitative Results

The qualitative results are shown in Figure 2. In Figure 2(a) and Figure 2(b), we show the generation of the same character in different scripts.In 2(c), we show the generation with some flaws. In 2(d), we show the generation of non-existing characters.

The flaw of Figure 2(c) is shown in Figure 3(a). There is an extra stroke left in the generation that comes from the unmasked part of the original image. Based on this observation, we intentionally manipulate the generation to force the model to generate some non-existing character shown in Figure 2(d). In Figure 3(b), we use red circles and a green circle to compare the non-existing character and the correct one. More specifically, the left part of the character in our example shoule be " 礻 " but we intentionally use " 礻 " as the image condition to inpainting. Based on this setting, the model generate the image that does not exist.

B Examples of Subjective Surveys

In Figure 4, we show two example questions in our subjective survey. Figure 4(a) is the question that asks respondents to find genuine calligraphy and Figure 4(b) asks respondents to find the fake one. The correct answer for both questions is Option A.

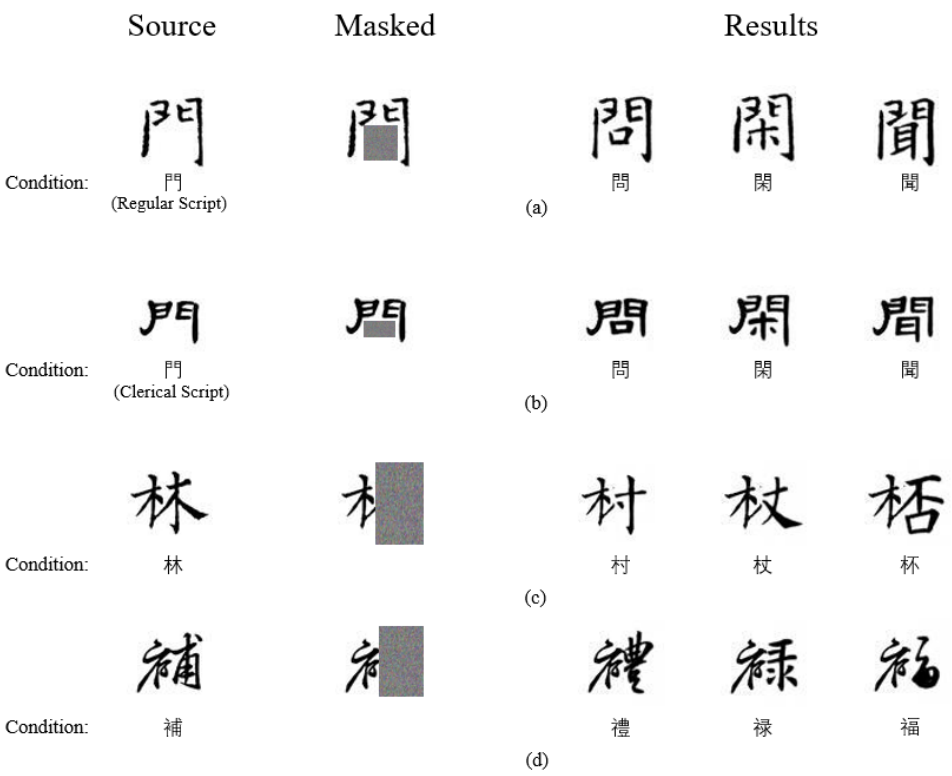
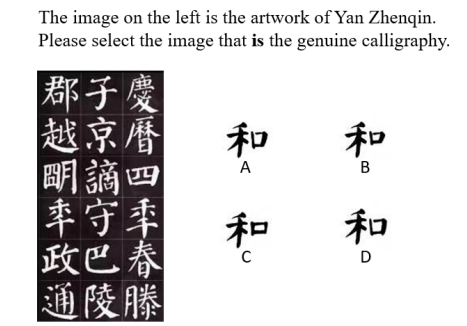


Figure 2: The qualitative results of our model.

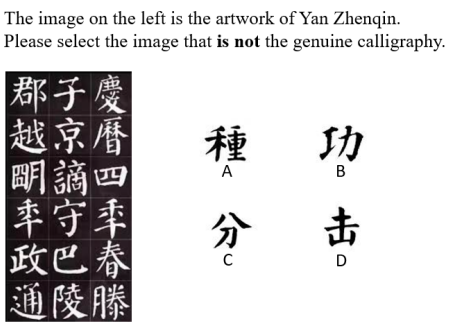
(a) The generation with flaws. The red circles denote the flaw, an extra stroke that comes from the unmasked part of the original image.

(b) The generation of non-existing characters. The red circles denote the flaw. The image with the green circle is the correct existing character.

Figure 3: The generation with flaws and non-existing characters.



(a) Question that asks the respondent to find the genuine calligraphy.



(b) Question that asks the respondent to find the fake calligraphy.

Figure 4: Two example questions in our subjective surveys.