DiffSketching: Sketch Control Image Synthesis with Diffusion Models

Qiang Wang, Di kong, Fengyin Lin, Yonggang Qi✉

{wanqqiang,dikong,fylin,qiyg}@bupt.edu.cn

Beijing University of Posts and Telecommunications, Beijing, China

1 Introduction

We propose DiffSketching, the first cross-domain sketch-to-image synthesis method utilizing diffusion models. Our method can be self-supervised when matching inputs, overcoming the large domain gap between sketch and generator's parameter space.

(a) Human sketch   (b) Synthesis images by DiffSketching
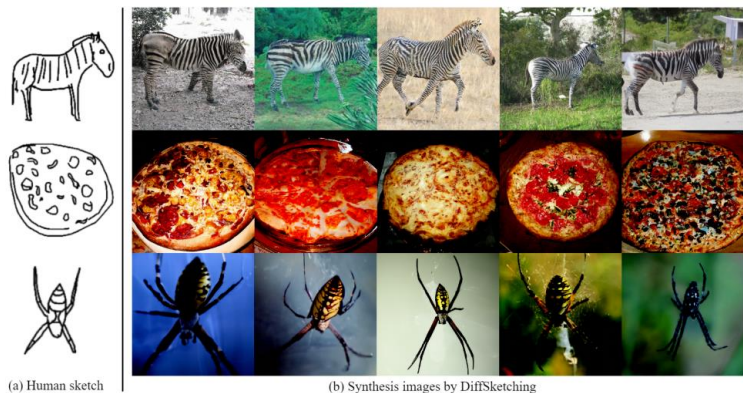
2 Contributions

1. Our method performs better on benchmarks than GAN-based models.

2. We can guide the generation process more finely and eliminate the singularity and uncertainty of input sketches.

3. Our method is capable of editing images and conducting image interpolation.
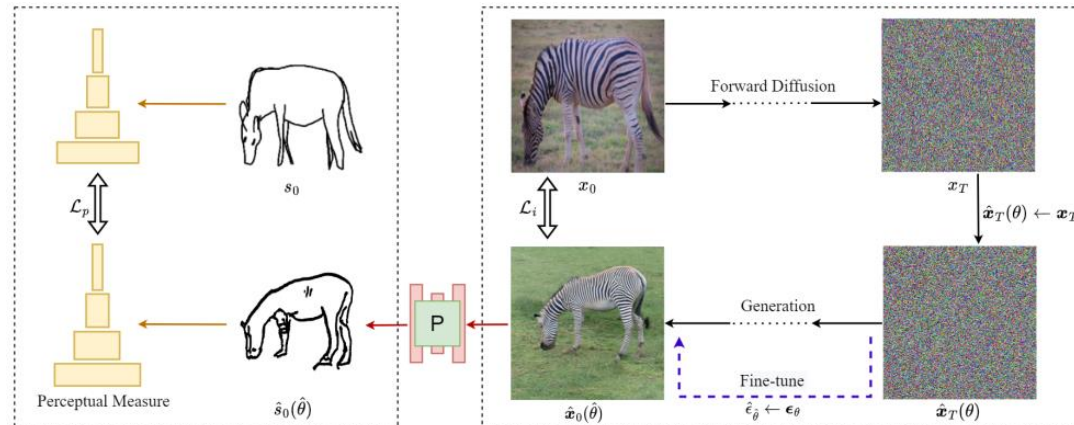
3 Method

We use perceptual diversity learning and image constraint identity learning to fine-tune the generation of diffusion models:

$$\mathcal{L}_p = \sum_{l=1}^{L} \frac{1}{H_l W_l} \sum_{h,w} \| w_l \odot (F_s(\hat{s}_0(\hat{\theta}))_{hw}^l - F_s(s_0)_{hw}^l) \|_2^2$$

$$\mathcal{L}_i = \frac{\mathbf{F_i(x_0)} \cdot \mathbf{F_i(\hat{x}_0(\hat{\theta}))}}{\|\mathbf{F_i(x_0)}\| \|\mathbf{F_i(\hat{x}_0(\hat{\theta}))}\|}$$

$$\mathcal{L} = \lambda \mathcal{L}_i(x_0, \hat{x}_0(\hat{\theta}))) + (1-\lambda) \mathcal{L}_p(s_0, \hat{s}_0(\hat{\theta}))$$

4 Experiments

Quantitative experiment

Method	FID ↓	IS ↑	Precision ↑	Recall ↑	Human ↑
USPS	48.73	23.74	0.42	0.38	26.45%
MUNIT	56.50	28.99	0.34	0.51	20.23%
Sketch-YOG	19.94	48.94	**0.70**	0.53	18.85%
Ours	**6.46**	**89.91**	0.68	**0.56**	**34.47%**
Ours(w/o $\mathcal{L}_p$)	7.22	83.43	0.33	0.39	N/A
Ours(w/o $\mathcal{L}_i$)	11.78	63.09	0.40	0.44	N/A
Ours(Quickdraw)	6.65	87.42	0.67	0.49	N/A

Qualitative experiment

Test on real human sketches

Input sketch   Synthesis images
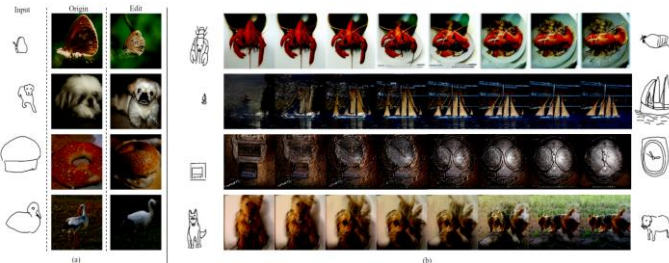
Fine-grained sketch controlling

Image editing and interpolation

References

[1] Sheng-Yu Wang, David Bau, and Jun-Yan Zhu. Sketch your own gan. In *Proceedings of ICCV*, pages 14050–14060, 2021.
[2] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proceedings of ECCV*, pages 172–189, 2018.
[3] Runtao Liu, Qian Yu, and Stella X Yu. Unsupervised sketch to photo synthesis. In *ECCV*, pages 36–52. Springer, 2020.