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.

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:

$$L_p = \frac{1}{B|W|} \sum_{b,w} \|w \odot (F_b(x_0(b)))_{\text{per}} - F_b(x_0(b))\|_1$$

$$L_\mathcal{I} = \frac{F_b(x_0) - F_b(s_b(\hat{\theta}))}{\|F_b(x_0) - F_b(s_b(\hat{\theta}))\|_1}$$

$$L = \lambda L_\mathcal{I}(x_0, s_b(\hat{\theta})) + (1 - \lambda) L_p(x_0, s_b(\hat{\theta}))$$

4 Experiments

<table>
<thead>
<tr>
<th>Method</th>
<th>FID</th>
<th>IS</th>
<th>Prec.</th>
<th>Recall</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>USPS</td>
<td>48.73</td>
<td>23.34</td>
<td>0.42</td>
<td>0.38</td>
<td>26.45</td>
</tr>
<tr>
<td>MUNIT</td>
<td>56.50</td>
<td>28.99</td>
<td>0.34</td>
<td>0.51</td>
<td>20.25</td>
</tr>
<tr>
<td>Sketch-YOG</td>
<td>19.94</td>
<td>45.84</td>
<td>0.78</td>
<td>0.53</td>
<td>38.85</td>
</tr>
<tr>
<td>Ours</td>
<td>6.46</td>
<td>89.91</td>
<td>0.68</td>
<td>0.56</td>
<td>54.47</td>
</tr>
</tbody>
</table>

Qualitative experiment

Test on real human sketches

Image editing and interpolation

References