Image Synthesis and Editing with Stochastic Differential Equations

Chenlin Meng[1] Yang Song[1] Jiaming Song[1]
Jiajun Wu[1] Jun-Yan Zhu[2] Stefano Ermon[1]
[1]Stanford University [2]Carnegie Mellon University

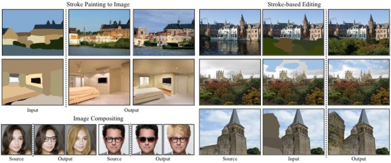

Figure 1: Stochastic Differential Editing (SDEdit) is a unified image synthesis and editing framework based on stochastic differential equations. SDEdit allows stroke painting to image, image compositing, and stroke-based editing without task-specific model training and loss functions.

Abstract

*We introduce a new image editing and synthesis framework, Stochastic Differential Editing (SDEdit), based on a recent generative model using stochastic differential equations (SDEs). Given an input image with user edits (*e.g.*, hand-drawn color strokes), we first add noise to the input according to an SDE, and subsequently denoise it by simulating the reverse SDE to gradually increase its likelihood under the prior. Our method does not require task-specific loss function designs, which are critical components for recent image editing methods based on GAN inversion. Compared to conditional GANs, we do not need to collect new datasets of original and edited images for new applications. Therefore, our method can quickly adapt to various editing tasks at test time without re-training models. Our approach achieves strong performance on a wide range of applications, including image synthesis and editing guided by stroke paintings and image compositing.*

1. Introduction

Denoising diffusion probabilistic models [23, 50] and score-based generative models [52, 55] are a new class of generative models that have found great success in image generation [52, 23, 53, 51, 55, 28, 13], audio synthesis [33] and graph generation [37]. One of the latest development in this direction, generative modeling with stochastic differential equations (SDEs [55]), has demonstrated comparable or better sample quality than generative adversarial networks (GANs) [17], with more stable training and better mode coverage. Specifically, Song et al. [55] proposed to perturb a data distribution according to the trajectory of an SDE by injecting Gaussian noise, which smoothly transforms any data distribution (*e.g.* images) to a tractable Gaussian prior distribution. To perform sampling, a neural network is trained to estimate the gradient of the data distribution, and subsequently uses it to solve the reverse stochastic process—a process that converts any Gaussian noise vector back to a data sample. Despite recent progress on uncondi-