Distilling Diffusion Models into Conditional GANs

Minguk Kang[1,2], Richard Zhang[2], Connelly Barnes[2], Sylvain Paris[2], Suha Kwak[1], Jaesik Park[3], Eli Shechtman[2], Jun-Yan Zhu[4], Taesung Park[2]

[1]Pohang University of Science and Technology, [2]Adobe Research, [3]Seoul National University, [4]Carnegie Mellon University

Motivation: "Diffusion + GAN for one-step generation"

Denoising inference is time-consuming and expansive.

Model distillation

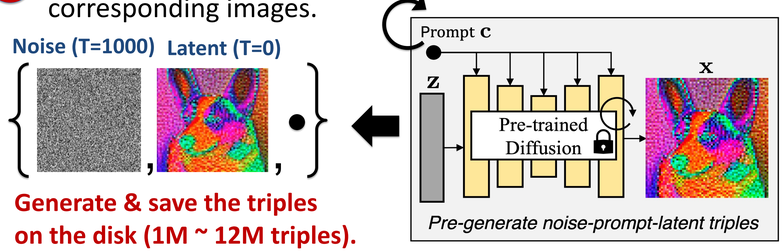
One-step image generation

We propose one-step Diffusion2GAN generator!

a) 1024px image in 0.16s
b) High-quality
c) ODE preserving distillation
d) Diverse image generation

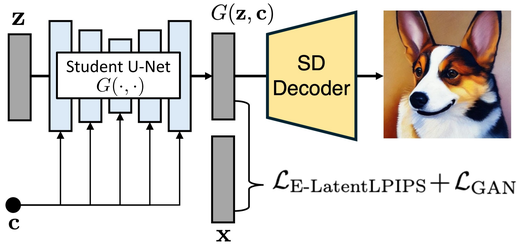
Distillation procedure: "Training a conditional GAN"

1. Simulate randomly sampled Gaussian noises and get their corresponding images.



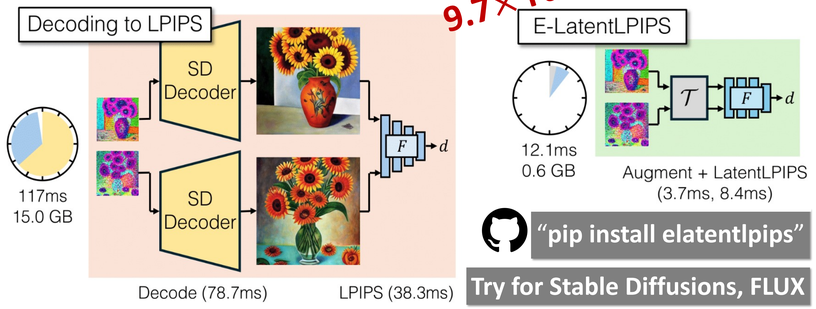
Generate & save the triples on the disk (1M ~ 12M triples).

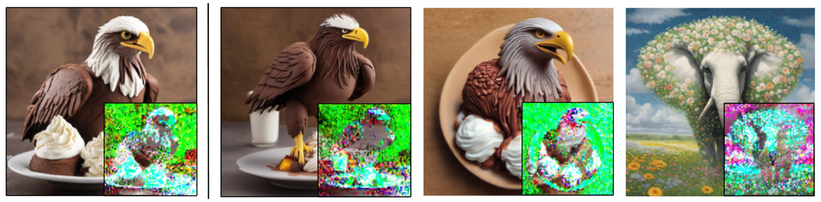
2. Train a conditional GAN where the inputs are noise and prompts, and the targets are their ODE solutions.



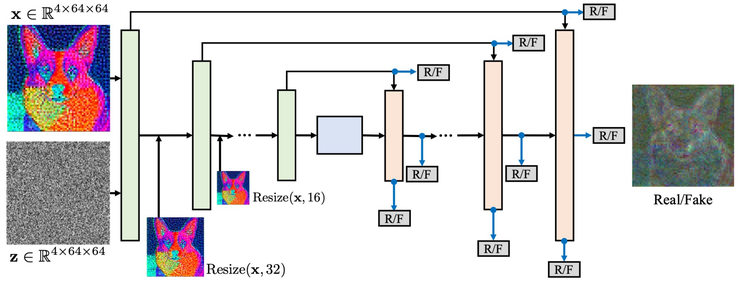
Latent space training: "Everything in latent space"

G Perceptual loss for LDM





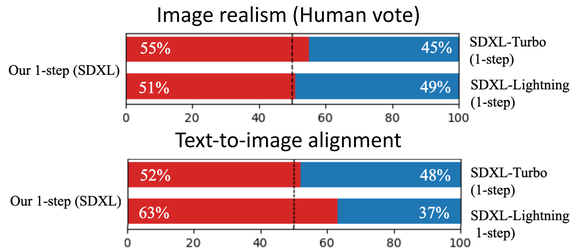
D Multi-scale I/O U-Net discriminator
(Initialized with a pre-trained diffusion model)



Evaluation: "Comparable to the latest distillation work"

"A cinematic shot of a little pig priest wearing sunglasses."

ODE preserving distillation: "Mimic teacher diffusion."



Our project page:

SCAN ME