Paper Digest Meeting

DreamSampler: Unifying Diffusion Sampling and Score Distillation for Image Manipulation

Jeongsol Kim[1,*], Geon Yeong Park[1,*], and Jong Chul Ye[2]

[1] Dept. of Bio & Brain Engineering, KAIST
[2] Kim Jae Chul AI graduate school, KAIST

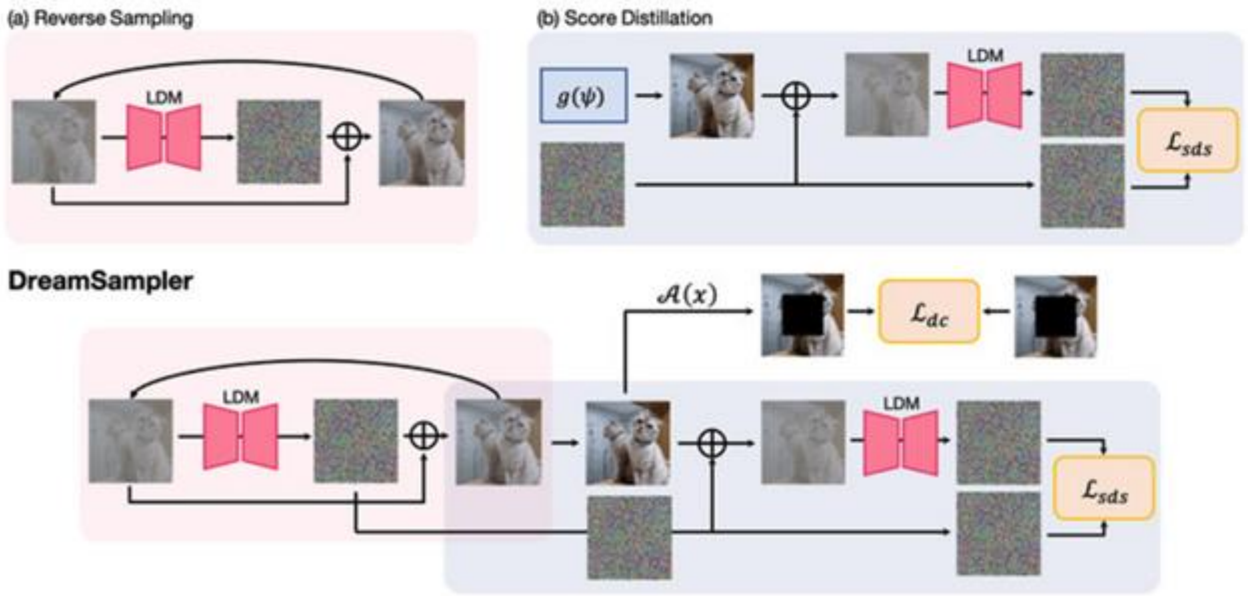
2024.08.04
Jaihoon Kim

KAIST Visual AI Group

Problem Definition

Motivation**:** Reverse process sampling lacks flexibility, and SDS suffers from mode collapse.

Goal: Combine the benefits of the two sampling methods for flexible and high-quality generation.



Key Ideas

DDIM Sampling from Optimization Perspective.

$$\boldsymbol{z}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\bar{\boldsymbol{z}} + \sqrt{1-\bar{\alpha}_{t-1}}\tilde{\boldsymbol{\epsilon}}, \quad \text{where} \quad \bar{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}} \|\boldsymbol{z} - \hat{\boldsymbol{z}}_{0|t}\|^2$$

1. Connection to Score Distillation Sampling (SDS):

$$\begin{aligned} \|\boldsymbol{z} - \hat{\boldsymbol{z}}_{0|t}\|^2 &= \left\| \frac{\boldsymbol{z}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}}{\sqrt{\bar{\alpha}_t}} - \frac{\boldsymbol{z}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_\theta(\boldsymbol{z}_t, t)}{\sqrt{\bar{\alpha}_t}} \right\|^2 \\ &= \frac{\sqrt{1-\bar{\alpha}_t}}{\sqrt{\bar{\alpha}_t}} \|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\boldsymbol{z}_t, t)\|^2, \end{aligned}$$

1. Add arbitrary regularization function:

$$\min_{\boldsymbol{z}} \|\boldsymbol{z} - \hat{\boldsymbol{z}}_{0|t}\|_2^2 + \lambda_{reg}\mathcal{R}(\boldsymbol{z})$$

Method

DreamSampler Framework:

Algorithm 1 Score Distillation

Require: T, ζ, g, ψ, $\mathcal{E}_\phi$, $\{\bar{\alpha}_t\}_{t=1}^{T}$

$\boldsymbol{z}_0 \leftarrow \mathcal{E}_\phi(\boldsymbol{x}_0)$
for $i = T$ **to** 1 **do**
  $t \sim U[0, T]$
  $\tilde{\boldsymbol{\epsilon}} \sim \mathcal{N}(0, \boldsymbol{I})$
  $\boldsymbol{z}_t \leftarrow \sqrt{\bar{\alpha}_t}\boldsymbol{z}_0 + \sqrt{1-\bar{\alpha}_t}\tilde{\boldsymbol{\epsilon}}$
  $\hat{\boldsymbol{\epsilon}}_\theta \leftarrow \boldsymbol{\epsilon}_\theta^\omega(\boldsymbol{z}_t, t, c)$
  $\nabla_\psi \mathcal{L}_{ds} \leftarrow \tilde{\boldsymbol{\epsilon}} - \hat{\boldsymbol{\epsilon}}_\theta$
  $\psi \leftarrow \psi - \zeta \nabla_\psi \mathcal{L}_{ds}$
  $\boldsymbol{z}_0 \leftarrow \mathcal{E}_\phi(g(\psi))$
end for
return ψ

Algorithm 2 DreamSampler

Require: T, ζ, g, ψ, $\mathcal{E}_\phi$, $\{\bar{\alpha}_t\}_{t=1}^{T}$

$\boldsymbol{z}_0 \leftarrow \mathcal{E}_\phi(\boldsymbol{x}_0), \boldsymbol{\epsilon}_\theta(\boldsymbol{z}_{T+1}) := \boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{I})$
for $i = T$ **to** 1 **do**
  $t \leftarrow i,\ \boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{I})$
  $\tilde{\boldsymbol{\epsilon}} \leftarrow \frac{\sqrt{1-\bar{\alpha}_{t-1}-\eta^2\beta_t^2}\,\hat{\boldsymbol{\epsilon}}_\theta + \eta\beta_t\boldsymbol{\epsilon}}{\sqrt{1-\bar{\alpha}_t}}$
  $\boldsymbol{z}_t \leftarrow \sqrt{\bar{\alpha}_t}\boldsymbol{z}_0 + \sqrt{1-\bar{\alpha}_t}\tilde{\boldsymbol{\epsilon}}$
  $\hat{\boldsymbol{\epsilon}}_\theta \leftarrow \boldsymbol{\epsilon}_\theta^\omega(\boldsymbol{z}_t, t, c)$
  $\nabla_\psi \mathcal{L}_{ds} \leftarrow \tilde{\boldsymbol{\epsilon}} - \hat{\boldsymbol{\epsilon}}_\theta$
  $\psi \leftarrow \psi - \zeta[\nabla_\psi \mathcal{L}_{ds} + \lambda_{reg}\nabla_\psi \mathcal{R}]$
  $\boldsymbol{z}_0 \leftarrow \mathcal{E}_\phi(g(\psi))$
end for
return ψ

Application

Real Image Editing

$$\min_{\boldsymbol{z}} \|\boldsymbol{z} - \hat{\boldsymbol{z}}_{0|t}(c_\phi)\|^2 + \gamma R(\boldsymbol{z}), \quad \textit{where} \quad R(\boldsymbol{z}) := \frac{\|\boldsymbol{z} - \hat{\boldsymbol{z}}_{0|t}(c_{tgt})\|^2}{(1-\gamma)}$$

Algorithm 3 DreamSampler for Image Editing

Require: source image $\boldsymbol{x}$, image encoder $\mathcal{E}_\phi$, latent diffusion model $\boldsymbol{\epsilon}_\theta$, null-text embedding c_ϕ, conditioning text embedding c_{tgt}.

$\boldsymbol{z}_0 \leftarrow \mathcal{E}_\phi(\boldsymbol{x})$

$\boldsymbol{z}_T \leftarrow \text{Inversion}(\boldsymbol{z}_0)$

for $t \in [T, 0]$ **do**

$\hat{\boldsymbol{\epsilon}}_\theta \leftarrow \boldsymbol{\epsilon}_\theta(\boldsymbol{z}_t, t, c_\phi) + \gamma[\boldsymbol{\epsilon}_\theta(\boldsymbol{z}_t, t, c_{tgt}) - \boldsymbol{\epsilon}_\theta(\boldsymbol{z}_t, t, c_\phi)]$

$\bar{\boldsymbol{z}} \leftarrow (\boldsymbol{z}_t - \sqrt{1-\bar{\alpha}_t}\hat{\boldsymbol{\epsilon}}_\theta)/\sqrt{\bar{\alpha}_t}$

$\tilde{\boldsymbol{\epsilon}} \leftarrow (\sqrt{1-\bar{\alpha}_{t-1} - \eta^2\beta_t^2}\hat{\boldsymbol{\epsilon}}_\theta + \eta\beta_t\boldsymbol{\epsilon})/\sqrt{1-\bar{\alpha}_{t-1}}$

$\boldsymbol{z}_t \leftarrow \sqrt{\bar{\alpha}_{t-1}}\bar{\boldsymbol{z}} + \sqrt{1-\bar{\alpha}_{t-1}}\tilde{\boldsymbol{\epsilon}}$

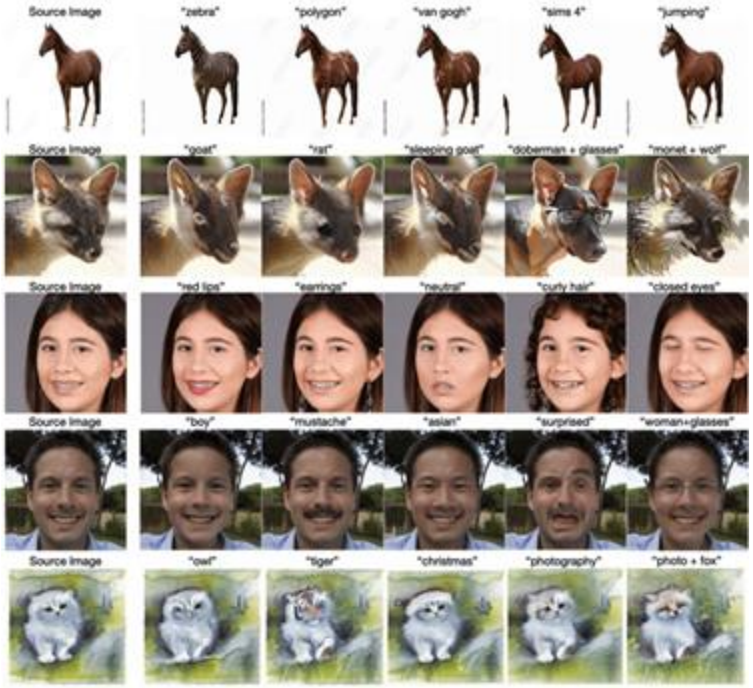
end for

Inverse Problems

$$\min_{\psi} (1-\gamma)\lambda_{SDS}\|\mathcal{E}(g(\psi)) - \hat{\boldsymbol{z}}_{0|t}(c_{\boldsymbol{y}})\|^2 + \gamma\lambda_{DC}\|\boldsymbol{y} - \mathcal{A}g(\psi)\|^2,$$

Experiments

Identity Renderer (g)

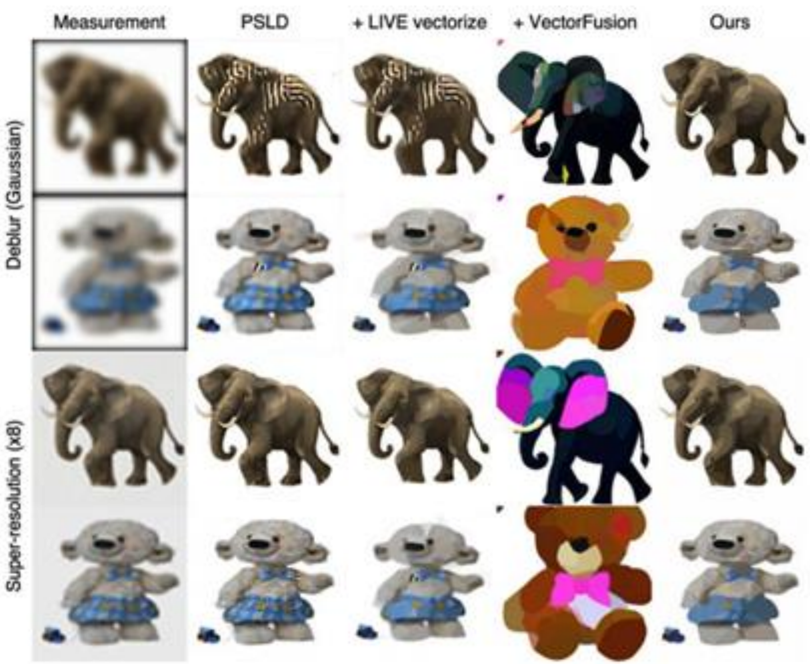


Real Image Editing

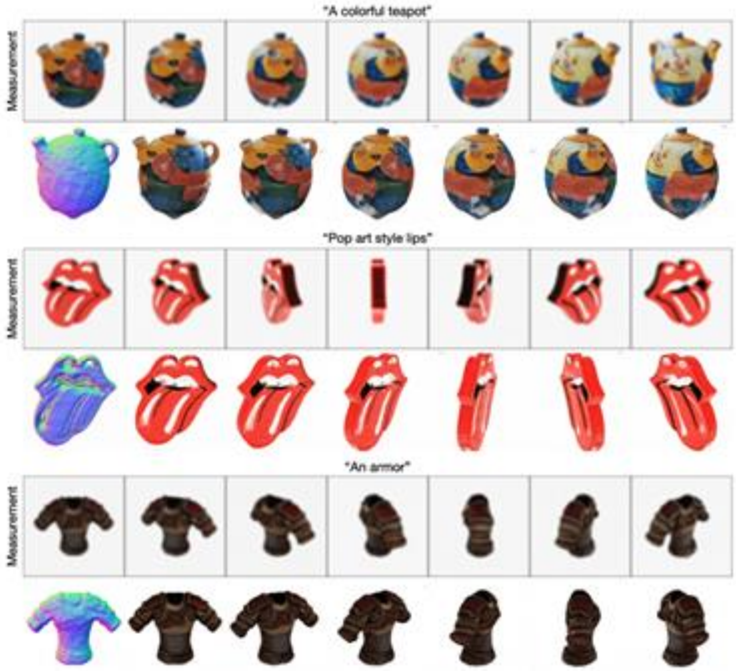
Image Inpainting

Experiments

Non-Identity Renderer (g)



SVG Inverse Problem



Novel View Synthesis from Blurry Inputs