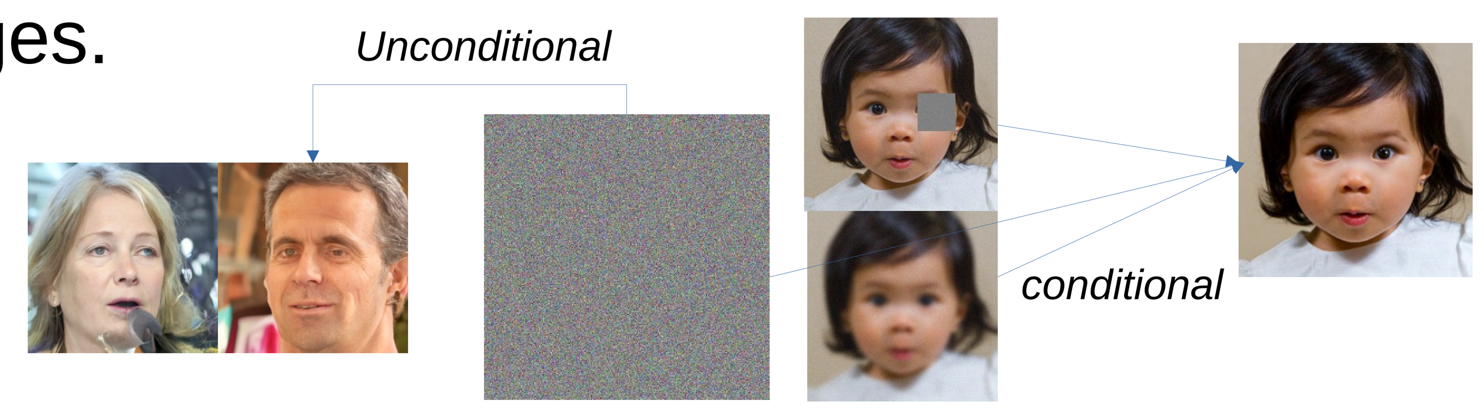
Diffusion Models for Solving Inverse Imaging Problems

Aditya Vikram Gupta

EE367 - Stanford University

Motivation

To solve ill-posed inverse problems such as inpainting, denoising, image reconstruction with generative priors. Diffusion models show great promise in unconditional generation. They model gradient of the log probability density of a data distribution, which guides the inverse process by acting as powerful score based prior that constraints the solution to the manifold of natural images.



Methods

3 methods build on top of a) pre-trained u-net based diffusion model pre-trained on FFHQ256 dataset b) VP formulation of forward and reverse diffusion process [7]

SDEdit [4]

$x_t = \sqrt{\bar{\alpha}_t} y + \sqrt{1-\bar{\alpha}_t} z, z \sim N(0, I)$

for t = T, …, 1:

$z \sim N(0, I)$ if $t > 1$ else $z = 0$

$\hat{x_0} = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t + (1-\bar{\alpha}_t) s_\theta(x_t, t))$

$x_{t-1} = \frac{\alpha_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t} \hat{x_0} + \sigma z$

end for

return x_0

Main idea: x_t partially noised image input + DDPM [3]

ScoreALD [5]

Problem: $\nabla_x \log p(\mathbf{b}|\mathbf{x_0}) \neq \nabla_x \log p_t(\mathbf{b}|\mathbf{x_t})$

Approach: $\nabla_x \log p(\mathbf{b}|\mathbf{x_0}) \approx \nabla_x \log p_t(\mathbf{b}|\mathbf{x_t})$

$x_t \sim N(0, I)$

for t = T. …. 1:

$z \sim N(0, I)$ if $t > 1$ else $z = 0$

$\hat{x_0} = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t + (1-\bar{\alpha}_t) s_\theta(x_t, t))$

$x_{t-1} = \frac{\alpha_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t} \hat{x_0} + \sigma z$

$x_{t-1} = x_{t-1} - \frac{1}{2(\sigma^2+\gamma^2)} \nabla_{x_t} \|\mathscr{A}(x_t) - y\|^2$

end for

return x_0

annealing factor

Main idea: x_t random noise + DDPM [3] + sampling with guidance using current noisy estimate i.e. x_t

DPS [6]

Problem: $\nabla_x \log p(\mathbf{b}|\mathbf{x_0}) \neq \nabla_x \log p_t(\mathbf{b}|\mathbf{x_t})$

Approach: $\nabla_x \log p(\mathbf{b}|\mathbf{x_0}) \approx \nabla_x \log p_t(\mathbf{b}|\mathbf{x_0} = \mathbb{E}[\mathbf{x_0}|\mathbf{x_t}])$

$x_t \sim N(0, I)$

for t = T, …, 1:

$\hat{x}_0 = \mathbb{E}[x_0|x_t]$

$z \sim N(0, I)$ if $t > 1$ else $z = 0$

$\hat{x_0} = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t + (1-\bar{\alpha}_t) s_\theta(x_t, t))$

$x_{t-1} = \frac{\alpha_t(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t} \hat{x_0} + \sigma z$

$x_{t-1} = x_{t-1} - \zeta_t \nabla_{x_t} \|\mathscr{A}(\hat{x_0}) - y\|^2$

end for

return x_0

Practical hyper parameter $\zeta_t = \frac{\zeta}{\left\|\nabla_{x_t} \|\mathscr{A}(\hat{x}_0) - \mathbf{y}\|^2\right\|}$

Main idea: x_t random noise + DDPM [3] + conditioning on current denoised estimate $\hat{x_0}$

Related Work

- BM3D [1] Conventional approach gold standard: Collapse on complex textures, over smoothing, artifacts on high noise level
- [2],[7] introduce concept of forward and reverse diffusion process & [3] demonstrates it for high quality image generation:
  main problem: sub-optimal log likelihood

References

[1] Sarjanoja, Boutellier, Hannuksela, "BM3D image denoising using heterogeneous computing platforms," DASIP, 2015

[2] Sohl-Dickstein, Weiss, Maheswaranathan, Ganguli, Deep Unsupervised Learning using Nonequilibrium Thermodynamics, ICML, 2015

[3] Ho, Jain, Abeel, Denoising Diffusion Probabilistic Models, NeurIPS, 2020

[4]Meng et al., "SDEdit: Guided Image Synthesis and Editing with StochasticDifferential Equations" ICLR 2022.

[5]Jalal et al., "Robust Compressed Sensing MRI with Deep Generative Priors,"NeurIps, 2021.

[6]Chung et al., "Diffusion Posterior Sampling for General Noisy Inverse Problems,"ICLR 2023.

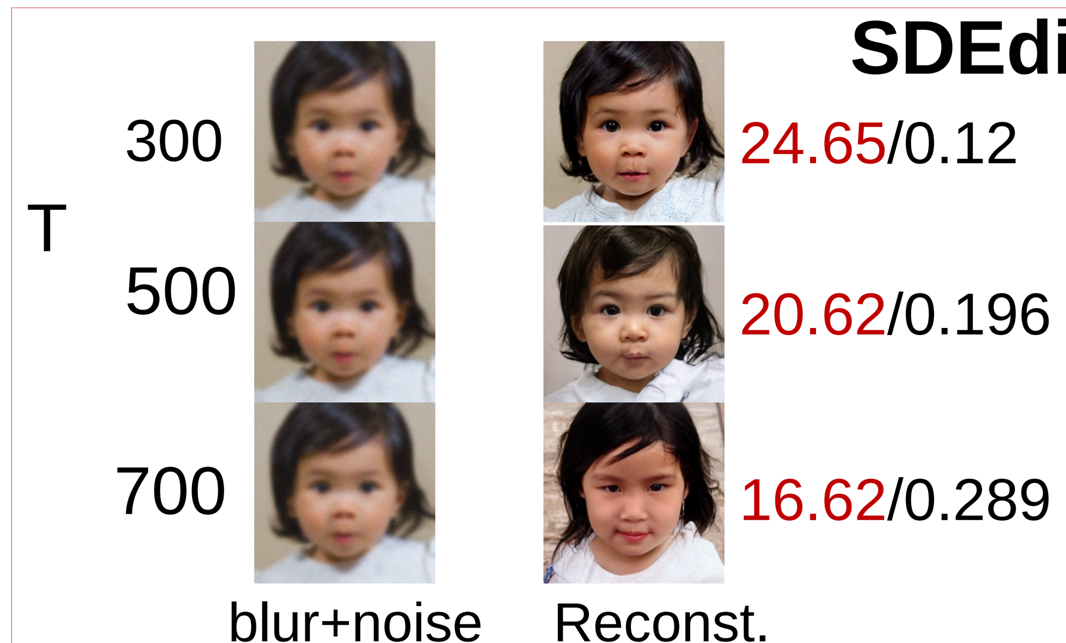
[7]Y. Song, S. Ermon, "Generative modeling by estimating gradients of the data distribution", NeurIPS 2019

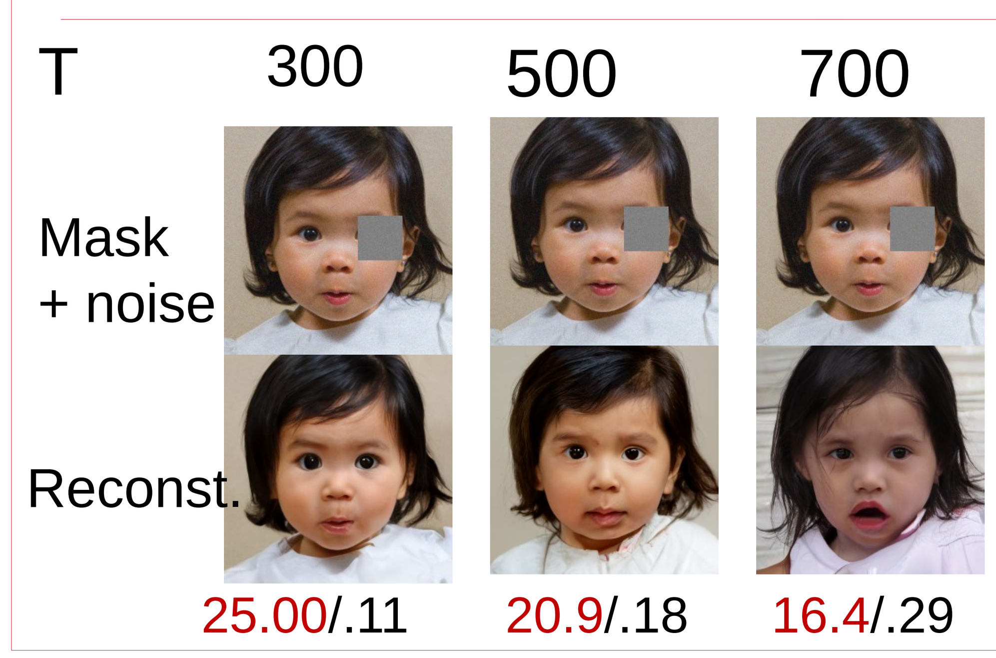
Experimental Results

Metric 1: PSNR (higher is better)

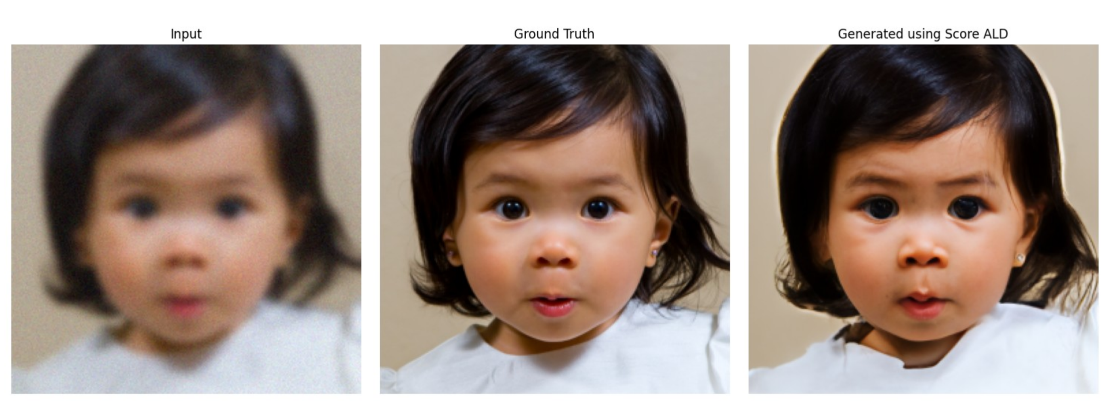
Metric 2: LPIPS (lower is better)

SDEdit



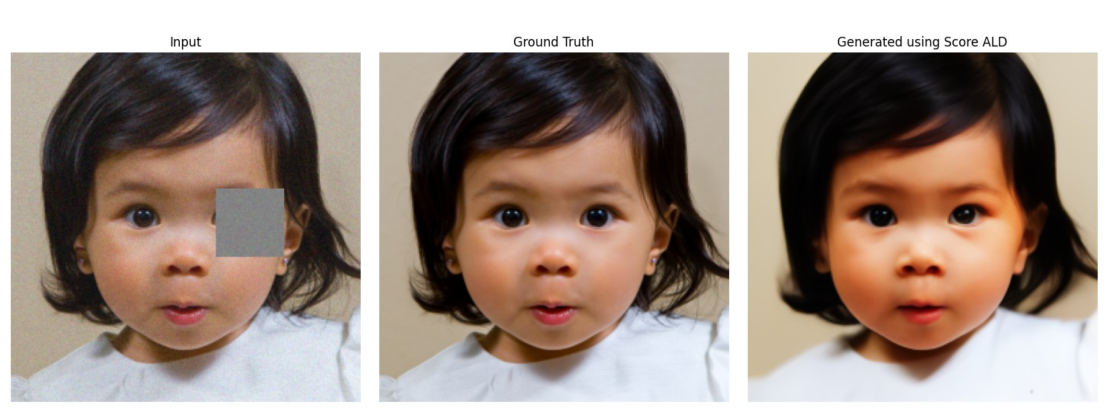


ScoreALD



Anneal [10,15] Deconv. 22.54/0.15

T=1000 for all experiments

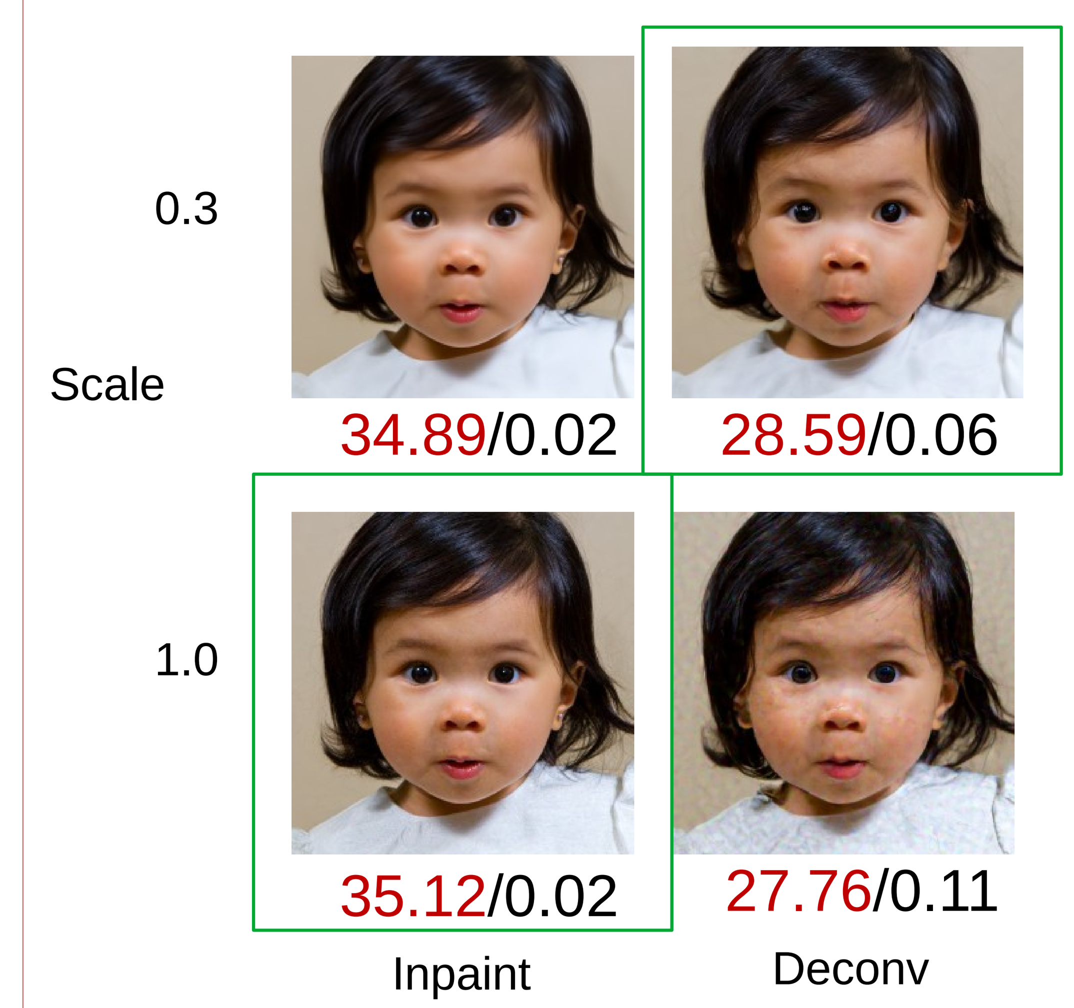


Anneal [15,20] Inpainting 21.95/0.14

*best results shown here, more in the full paper for full effect of hyperparameter

DPS

T=1000 for all exp



*here we can see which hyperparameter is the best for which task