Image generation with Shortest-path Diffusion

Ayan Das*, Stathi Fotiadis*, Anil Batra, Farhang Nabiei, FengTing Liao, Sattar Vakili,
Da-Shan Shiu, Alberto Bernacchia
(* Equal Contributions)

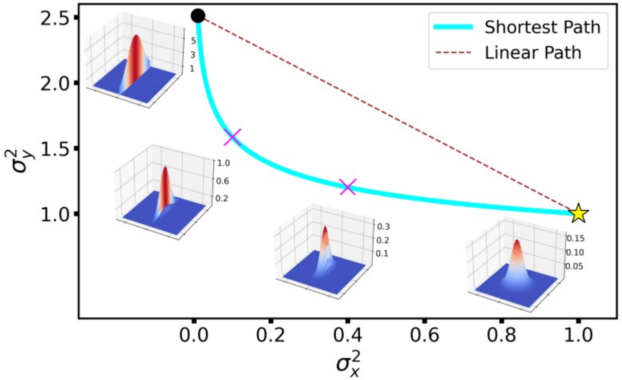
MediaTek Research
https://www.mtkresearch.com/

Introduction to Diffusion Models

- **Increasingly popular class of Generative Model**
- **Two primary components:**
  - **Reverse/Generative process**
    Going from $\mathcal{N}(\mathbf{0}, \boldsymbol{I})$ to data in *distribution space*
  - **Forward/Noising process**
    Specifies the exact "path" of travel
- **Forward specification**
  - By far, dominantly hand designed
  - Requires trial-and-error to find optimal path

Shortest path between distributions

- **Shortest path between two Gaussians**



- **Fisher metric** $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_0) \rightarrow \mathcal{N}(\mathbf{0}, \boldsymbol{I})$
  - Keeps maximum overlap between subsequent distributions

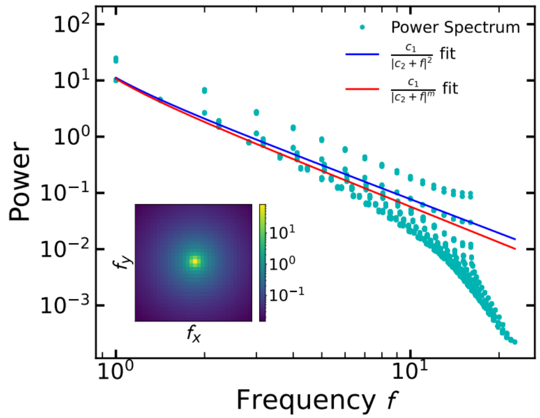
$$\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_0^{1-t}, t \in [0,1]$$

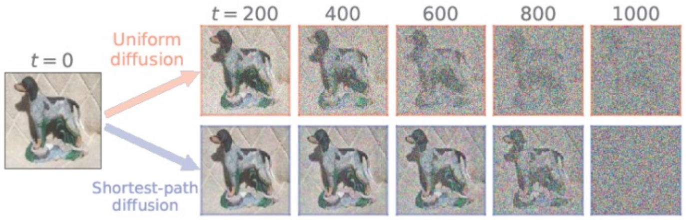
Modelling the covariance of natural images

- Translation invariant $\Sigma_0 = \boldsymbol{FDF}^H$
- Power spectrum[1] $\sim \frac{1}{\boldsymbol{f}^2}$
- We model D as $D_{ii} = \frac{c_1}{|c_2 + f_i|^m}$
- Implementation in Fourier space

$$\boldsymbol{u}_t = \boldsymbol{\Psi}_t^{1/2}\boldsymbol{u}_0 + (\boldsymbol{I} - \boldsymbol{\Psi}_t)^{1/2}\boldsymbol{\xi}_t$$

$$\boldsymbol{\Psi}_t = \left(\boldsymbol{I} - \boldsymbol{D}^{1-t/T}\right)(\boldsymbol{I} - \boldsymbol{D})^{-1}$$





[1] Hyvarinen, Huri & Hoyer, Natural Image Statistics (2009)

Experimental setup

- **Datasets:**
  CIFAR10 (32x32) & ImageNet (64x64)
  - Representative of “Natural images”
  - Roughly holds the *translation invariant* assumption
- **Setup (for fair comparison)**
  - Same UNet architecture as iDDPM [1]
  - Same optimizer and learning rate as [1]
  - Analogous reverse process variance for sampling
- **Evaluation**
  - Computes FID with 50K samples

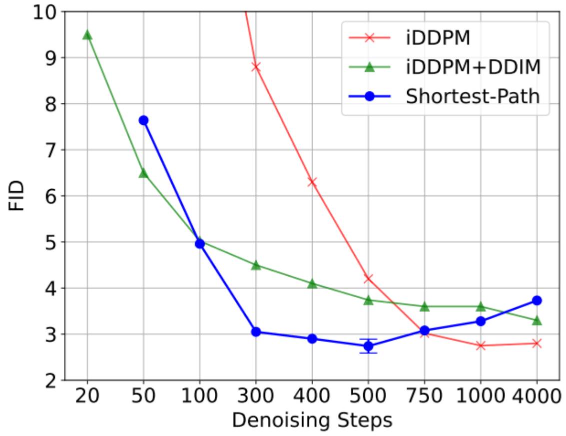
Only difference:
Our estimated non-uniform
forward noising schedule $\mathbf{\Psi}_t$

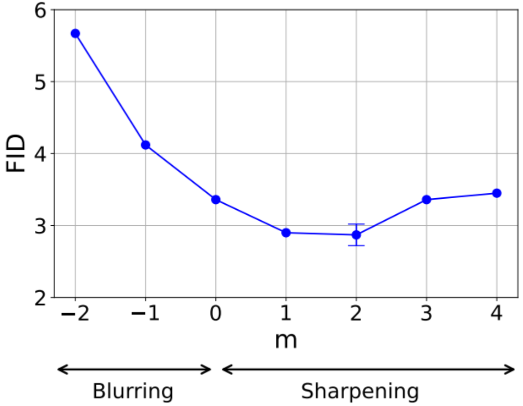
[1] Nichol, A. Q. and Dhariwal, P. “Improved denoising diffusion probabilistic model”, ICML 2021

CIFAR10 results

- **FID is lowest on the Shortest path**
  - Lowest point is at T = 500
  - Surpasses vanilla iDDPM
- **Our power spectrum model**
  - Found m = 2 to be optimal
  - Corresponds to "sharpening" rather than "blurring" .. .. as suggested by [1] & [2]

Methods	FID
Soft Diffusion	4.64
Blurring Diffusion	3.17
SPD (Ours)	**2.74**

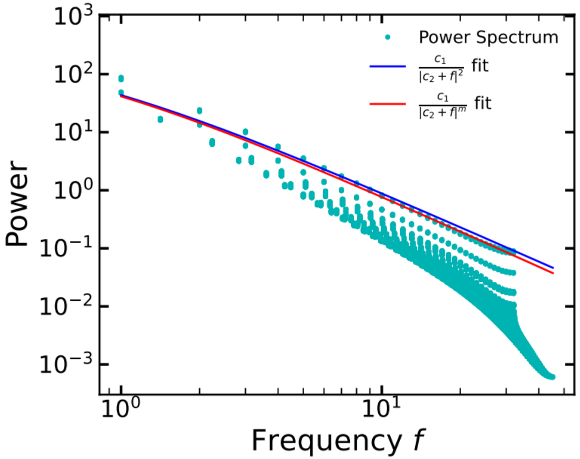




[1] Daras, G., Delbracio, M., Talebi, H., Dimakis, A. G., and Milanfar, P. "Soft diffusion: Score matching for general corruptions", 2022.
[2] Hoogeboom, E. and Salimans, T., "Blurring diffusion models", ICLR 2023

ImageNet64 results

- **Preliminary experiments are promising**
  - Unconditional model trained (and samples) with T = 1000
  - Better FID than iDDPM with less T and training iterations



Methods	Diffusion steps	Training steps	FID
iDDPM	4000	1.5M	19.2
SPD (Ours)	1000	1M	**13.7**

Quantitative results

Generated samples from **SPD (Ours)**

Thank you

Read the paper, or checkout our code →

mtkresearch/shortest-path-diffusion

SCAN ME