Stochastic Forward–Backward Deconvolution
Training Diffusion Models with Finite Noisy Datasets

Haoye Lu Qifan Wu Yaoliang Yu

UNIVERSITY OF WATERLOO

VECTOR INSTITUTE | INSTITUT VECTEUR

ICML 2025

Introduction

Deep learning-based generative models:

- text generation, text-to-image generation, protein structure prediction

Training powerful generative models requires web-scale data ...

- ChatGPT: large collections of text data, such as books, articles, and web pages
- Stable diffusion: LAION, more than 12 million text-image pairs

Scaling Laws:

$$\text{bigger model} + \text{more data} \implies \text{better performance}$$

Jared Kaplan et al. Scaling Laws for Neural Language Models. arXiv:2001.08361, 2020.

Issues

Copyright: Training Datasets of this scale always have copyrighted contents

Privacy: They may contain sensitive personal information

Models could reproduce these samples during the sampling.

Example: Stable Diffusion

Daras et al. Consistent Diffusion Meets Tweedie: Training Exact Ambient Diffusion Models with Noisy Data. ICML 2024.

Consider an alternative method:

training generative models using data corrupted by noises.

Focus on: diffusion-based models and Gaussian noises.

Problem Setting

Given clean samples $\mathcal{D} = \{\mathbf{x}^{(k)}\}_{k=1}^{N}$ with $\mathbf{x}^{(k)} \sim p_{\text{data}}$, training dataset

$$\mathcal{D}_{\text{noisy}} = \{\mathbf{x}^{(k)} + \boldsymbol{\epsilon}^{(k)}\}_{k=1}^{N}, \qquad \boldsymbol{\epsilon}^{(k)} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\tau^2 \mathbf{I}).$$

Diffusion models will be trained using $\mathcal{D}_{\text{noisy}}$.

Note: all clean samples are corrupted only once.

Averaging multiple corrupted version $\Rightarrow$ clean sample

This problem is hard to solve

Very pessimistic sample complexity $\Theta\left((\log N)^{-2}\right)$

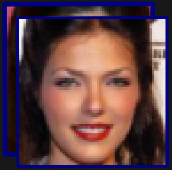
N : Number of noisy samples.

Takeaway: training models solely on noisy samples is *theoretically possible* but practically infeasible $\Longrightarrow$ pretraining on (small) copyright-free datasets is necessary.

Method: Stochastic Forward-Backward Deconvolution (SFBD)

Given:

Copyright-free
clean samples ($< 1\%$)



$\mathcal{D}_{\text{clean}}$

$\cdots$

Personal/copyrighted
corrupted samples ($> 99\%$)

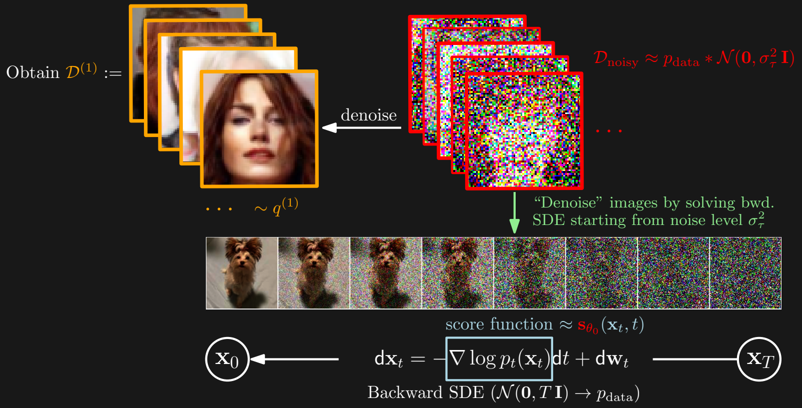
$\mathcal{D}_{\text{noisy}} \approx p_{\text{data}} * \mathcal{N}(\mathbf{0}, \sigma_\tau^2 \mathbf{I})$

$\cdots$

SFDB - Iteration 0

Step 1. Pretrain on the copyright-free clean samples to obtain diffusion model $\mathbf{s}_{\boldsymbol{\theta}_0}$.

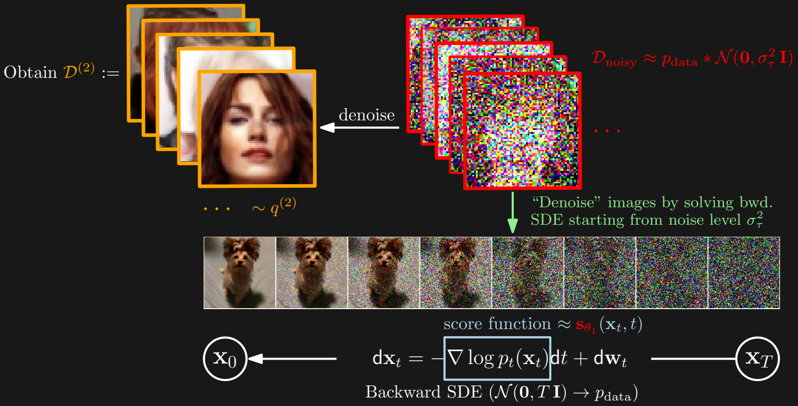
Step 2.



SFBD - Iteration 1

Step 1. Finetune $\mathbf{s}_{\boldsymbol{\theta}_0}$ on the denoised samples $\mathcal{D}^{(1)}$ to obtain diffusion model $\mathbf{s}_{\boldsymbol{\theta}_1}$.

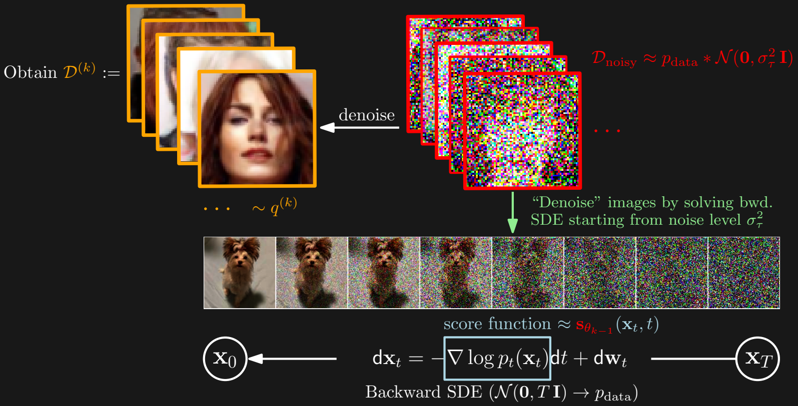
Step 2.



SFBD - Iteration k . . .

Step 1. Finetune $\mathbf{s}_{\boldsymbol{\theta}_{k-1}}$ on the denoised samples $\mathcal{D}^{(k)}$ to obtain diffusion model $\mathbf{s}_{\boldsymbol{\theta}_k}$.

Step 2.



We proved that

$$\boxed{q^{(k)} \to p_{\mathrm{data}} \quad \text{as } k \to \infty}$$

at a rate of $\mathcal{O}\left(M_0/\sqrt{k}\right)$, where M_0 depends on the pretrained model $\mathbf{s}_{\boldsymbol{\theta}_0}$:

- $M_0 \to 0$ if the pretrained model better estimates p_{data}.

Empirical Results

Pretraining was performed on 50 clean images.

Method	CIFAR10 (32 x 32)			CelebA (64 x 64)		
	σ_ζ	Pretrain	FID ↓	σ_ζ	Pretrain	FID ↓
DDPM (Ho et al., 2020)	0.0	No	4.04	0.0	No	3.26
DDIM (Song et al., 2021a)	0.0	No	4.16	0.0	No	6.53
EDM (Karras et al., 2022)	0.0	No	1.97	-	-	-
SURE-Score (Aali et al., 2023)	0.2	Yes	132.61	-	-	-
EMDiff (Bai et al., 2024)	0.2	Yes	86.47	-	-	-
TweedieDiff (Daras et al., 2024)	0.2	No	167.23	0.2	No	246.95
TweedieDiff (Daras et al., 2024)	0.2	Yes	65.21	0.2	Yes	58.52
SFBD (Ours)	0.2	Yes	13.53	0.2	Yes	6.49

SFBD also performs well for large σ_ζ.


H. Lu


Q. Wu


Y. Yu

Thank you!

Paper Link:

arXiv:2502.05446