SoFlow: Solution Flow Models for One-Step Generative Modeling

Tianze Luo, Haotian Yuan, Zhuang Liu

1-NFE generation results on the ImageNet 256x256 dataset
133 Github stars: https://github.com/luotianze666/SoFlow

Background & Motivation

- Flow Matching allows us to generate data by solving ODE
- $x_t = \alpha_t x_0 + \beta_t x_1,\ x_0 \sim p_{data},\ x_1 \sim N(0, I), t \in [0,1]$
- $\frac{\mathrm{d}X(t)}{dt} = v(X(t), t), v(x_t, t) = E[\alpha_t' x_0 + \beta_t' x_1 | x_t], X(1) = x_1$
- Iteratively solving the ODE is **time-consuming**
- Directly solve it with a network $f_\theta(x_t, t, s)$ for **one-step generation**
- It approximates $f(x_t, t, s)$, which is the solution function of the velocity ODE

Goals & Mathematical Principles

- Build a **strong** model that is **compatible with Classifier-free guidance (CFG)** and **doesn't require Jacobian-vector product (JVP) computation**
- Our model only needs to satisfy this PDE with a boundary condition

$$\partial_1 f_\theta(x_t, t, s) v(x_t, t) + \partial_2 f_\theta(x_t, t, s) = 0$$

$$f_\theta(x_t, t, t) = x_t$$

- Where $\partial_1 f_\theta \in R^{D \times D}$ is the Jacobian matrix function, $\partial_2 f_\theta \in R^D$ is the partial derivative function with respect to t, and D is the data dimension.

Model Parameterization

- Parameterization (satisfying the boundary condition):

$$\partial_1 f_\theta(x_t, t, s) v(x_t, t) + \partial_2 f_\theta(x_t, t, s) = 0$$

$$\boldsymbol{f_\theta(x_t, t, t) = x_t}$$

- We consider $f_\theta(x_t, t, s) = a(t, s)x_t + b(t, s)NN_\theta(x_t, t, s), a(t, t) = 1, b(t, t) = 0$
- $f_\theta(x_t, t, s) = x_t + (s - t)NN_\theta(x_t, t, s)$ Euler
- $f_\theta(x_t, t, s) = \cos\frac{\pi}{2}(s - t)x_t + \sin\frac{\pi}{2}(s - t)NN_\theta(x_t, t, s)$ Trignometric

Flow Matching Loss

- Flow matching loss for $t = s$ situation (satisfying the PDE):

$$\boldsymbol{\partial_1 f_\theta(x_t, t, s) v(x_t, t) + \partial_2 f_\theta(x_t, t, s) = 0}$$

$$f_\theta(x_t, t, t) = x_t$$

- They imply $\partial_3 f_\theta(x_t, t, t) = v(x_t, t)$, where $\partial_3 f_\theta \in R^D$ is the partial derivative of s

$$E_{x_t,t}[w_v(t, MSE)\frac{1}{D}\left|\left|\partial_3 f_\theta(x_t, t, t) - v(x_t, t)\right|\right|_2^2]$$

- $f_\theta(x_t, t, s) = a(t, s)x_t + b(t, s)NN_\theta(x_t, t, s), a(t, t) = 1, b(t, t) = 0$
- $\partial_3 f_\theta(x_t, t, t) = \partial_2 a(t, t)x_t + \partial_2 b(t, t)NN_\theta(x_t, t, t)$

Solution Consistency Loss

- Solution Consistency Loss for $s < t$ (satisfying the PDE):

$$\boldsymbol{\partial_1 f_\theta(x_t, t, s) v(x_t, t) + \partial_2 f_\theta(x_t, t, s) = 0}$$

$$f_\theta(x_t, t, t) = x_t$$

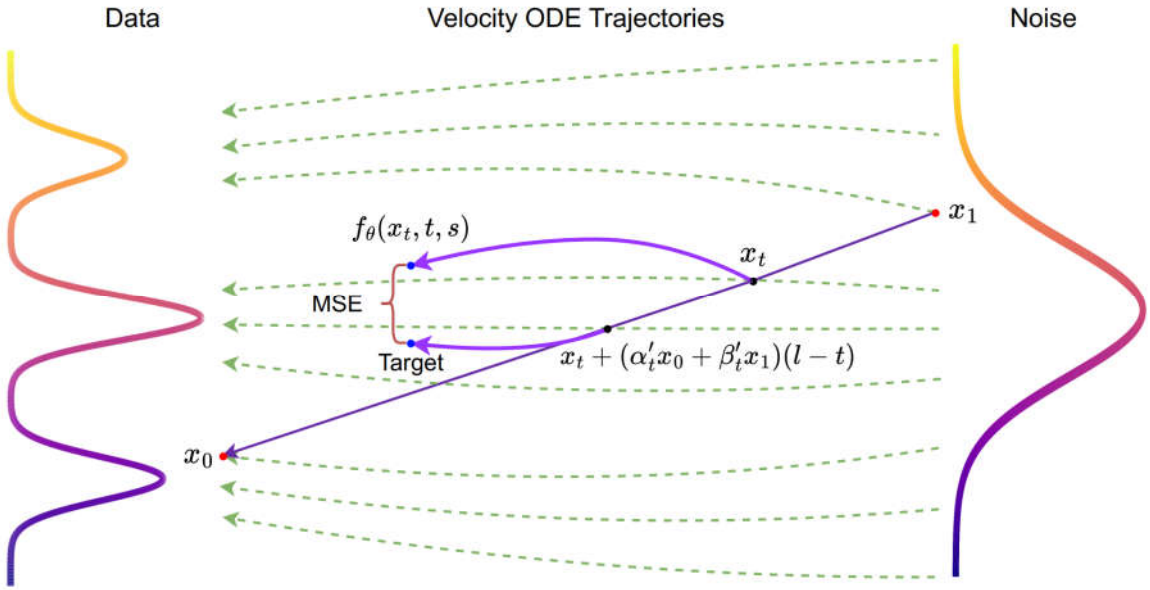
- An Equation with network's derivatives is not desirable
- $0 = \partial_1 f_\theta(x_t, t, s) v(x_t, t) + \partial_2 f_\theta(x_t, t, s) = \frac{f_\theta(x_t,t,s) - f_\theta(x_t + v(x_t,t)(l-t), l, s)}{t-l} + O(t-l)$
- Insight from Taylor expansion : $f_\theta(x_t, t, s) \to f_\theta(x_t + v(x_t, t)(l-t), l, s)$

An Illustrative Figure

- Insight from Taylor expansion : $f_\theta(x_t, t, s) \rightarrow f_\theta(x_t + v(x_t, t)(l-t), l, s)$

$$E_{x_t,t,l,s}[w_c(t,l,s,MSE)\frac{1}{D}\left|\left|f_\theta(x_t,t,s) - f_{\theta^-}(x_t + v(x_t,t)(l-t),l,s)\right|\right|_2^2]$$

- θ^- means stop-gradient operation, D is the data dimension.



Classifier-Free Guidance

- Flow matching loss with CFG:

$$E_{x_t,t}[w_v(t, MSE)\frac{1}{D}\left|\left|\partial_3 f_\theta(x_t, t, t, c) - v_{mix}\right|\right|_2^2]$$

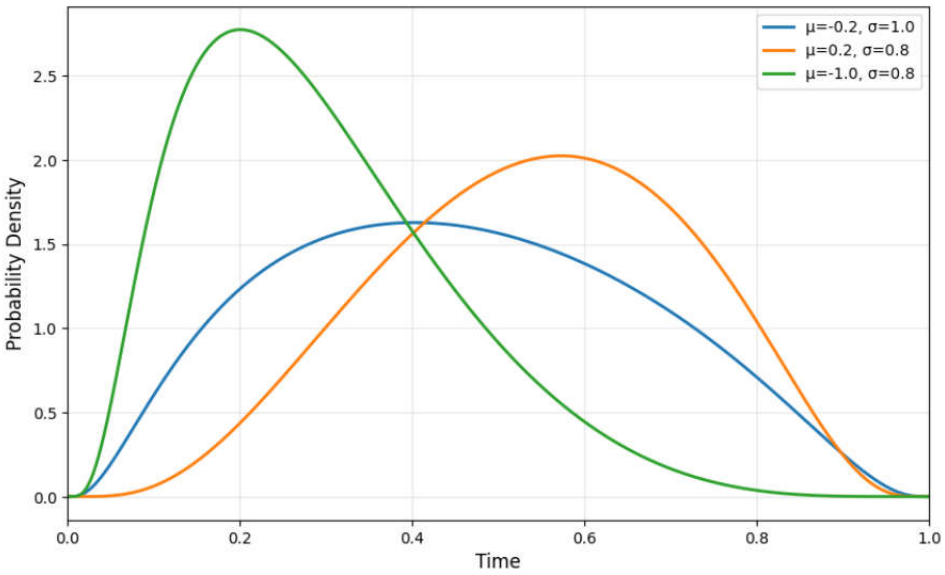
- Solution Consistency Loss with CFG:

$$E_{x_t,t,l,s}[w_c(t, l, s, MSE)\frac{1}{D}\left|\left|f_\theta(x_t, t, s, c) - f_{\theta^-}(x_t + v_{mix}(l-t), l, s)\right|\right|_2^2]$$

- $v_{mix} = m\big(w(\alpha'_t x_0 + \beta'_t x_1) + (1-w)\partial_3 f_{\theta^-}(x_t, t, t, \phi)\big) + (1-m)\partial_3 f_{\theta^-}(x_t, t, t, c)$
- w is CFG strength, m is the velocity mix ratio to reduce the variance of the random term
- Simply replacing the velocity term with v_{mix} to enable CFG

Time Sampling Method

- We sample t, s from **logit-normal** distributions , i.e. $\text{sigmoid}(N(\mu, \sigma^2))$
- Then, l is determined by t, s and a schedule function $r(k, K)$, where k, K are current and total training steps
- $l = t + (s - t)r(k, K)$
- $r(k, K)$ decreases exponentially during training to gradually move l to t



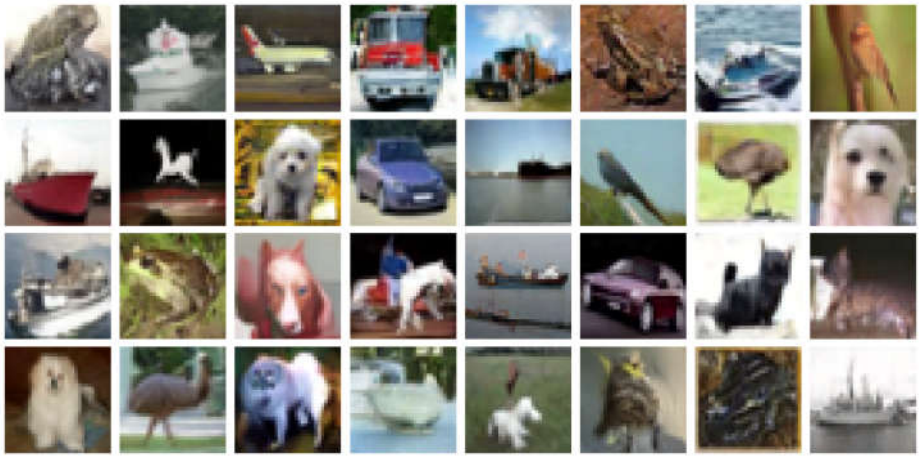
Main Results

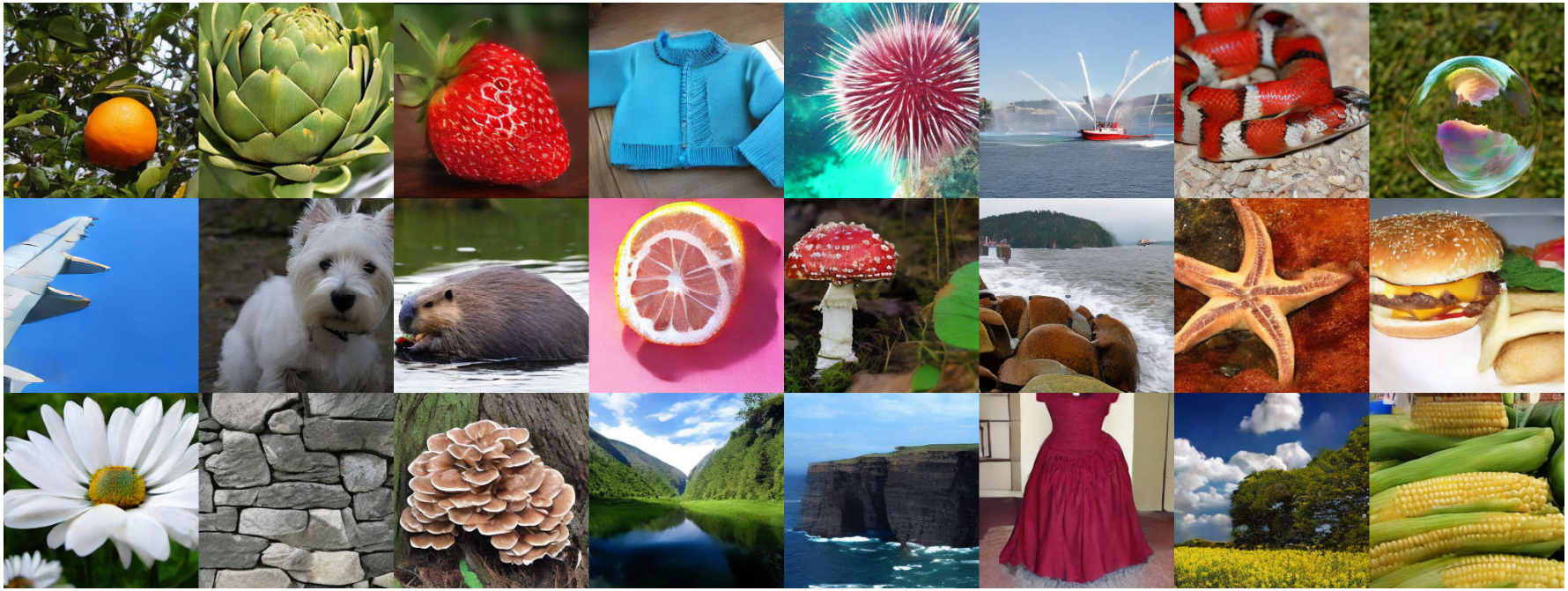
- 20 ablation studies to determine the hyperparameters (see our paper)
- ImageNet 256x256, 240 Epochs
- Strong 1-NFE, 2-NFE FID-50K performance compared with MeanFlow models (same DiT architecture and training steps)

method	epochs	params	NFE	FID↓
Generative Adversarial Networks				
BigGAN (Brock et al., 2018)	-	112M	1	6.95
StyleGAN-XL (Sauer et al., 2022)	-	166M	1	**2.30**
GigaGAN (Kang et al., 2023)	-	569M	1	3.45
Masked and Autoregressive Models				
Mask-GIT (Chang et al., 2022)	555	227M	8	6.18
MagViT-v2 (Yu et al., 2023)	1080	307M	64	1.78
LlamaGen-XL (Sun et al., 2024)	300	775M	576	2.62
VAR (Tian et al., 2024)	350	2.0B	10	1.80
MAR (Li et al., 2024)	800	943M	64	**1.55**
RandAR-XL (Pang et al., 2025)	300	775M	256	2.22
Multi-step Diffusion Models				
LDM-4-G (Rombach et al., 2022)	170	395M	250×2	3.60
MDTv2 (Gao et al., 2023)	700	676M	250×2	1.63
DiT-XL/2 (Peebles and Xie, 2023)	1400	675M	250×2	2.27
SiT-XL/2 (Ma et al., 2024)	1400	675M	250×2	2.06
FlowDCN-XL/2 (Wang et al., 2024)	400	675M	250×2	2.00
SiT-REPA-XL/2 (Yu et al., 2024)	800	675M	250×2	**1.42**
Few-step Diffusion Models				
iCT-XL/2† (Song and Dhariwal, 2023)	-	675M	1 / 2	34.24 / 20.30
Shortcut-XL/2 (Frans et al., 2024)	250	675M	1 / 4	10.60 / 7.80
IMM-XL/2 (Zhou et al., 2025)	3840	675M	1×2 / 2×2	7.77 / 3.99
MeanFlow-B/2 (Geng et al., 2025)	240	131M	1	6.17
MeanFlow-M/2 (Geng et al., 2025)	240	308M	1	5.01
MeanFlow-L/2 (Geng et al., 2025)	240	459M	1	3.84
MeanFlow-XL/2 (Geng et al., 2025)	240	676M	1 / 2	3.43 / 2.93
SoFlow-B/2	240	131M	1 / 2	4.85 / 4.24
SoFlow-M/2	240	308M	1 / 2	3.73 / 3.42
SoFlow-L/2	240	459M	1 / 2	3.20 / 2.90
SoFlow-XL/2	240	676M	1 / 2	**2.96** / **2.66**

Table 2 FID-50K results for class-conditional generation on ImageNet 256×256. ×2 denotes an NFE of 2 per sampling step incurred by CFG. Entries in the format "1 / 2" indicate that the corresponding FID scores are reported for 1-NFE and 2-NFE sampling, respectively. † Results as reported in Zhou et al. (2025).

Visual Samples





Computational Costs

- SoFlow models eliminate the need of JVP computation
- More compatible with accelerated attention implementations
- **Faster** (23%) and **less GPU memory** consumption (31%)

	MeanFlow (Math Attn.)	Soflow (Math Attn.)	Soflow (Efficient Attn.)
GPU Memory	51.45GB	38.95 GB	35.44GB
Training Speed	2.39 iters/sec	2.84 iters/sec	2.94 iters/sec