Consistency is All You Need

Kaiwen Zheng

2023.12.08

Content

- Consistency Models (ICML 2023)

Technical Improvement

- Improved Techniques for Training CM (ICLR 8866)
- Consistency Trajectory Models (ICLR 8666)

Applications to Text-to-Image Model

- Latent Consistency Models (ICLR 6555)
- LCM-LoRA

Diffusion Probabilistic Models (DPMs)

forward SDE $\mathrm{d}\boldsymbol{x}_t = \boldsymbol{f}(\boldsymbol{x}_t, t)\mathrm{d}t + g(t)\mathrm{d}\boldsymbol{w}_t$

$q_0(\boldsymbol{x}_0)$

$q_T(\boldsymbol{x}_T)$

reverse SDE $\mathrm{d}\boldsymbol{x}_t = [\boldsymbol{f}(\boldsymbol{x}_t, t) - g(t)^2 \nabla_{\boldsymbol{x}} \log q_t(\boldsymbol{x}_t)]\mathrm{d}t + g(t)\mathrm{d}\bar{\boldsymbol{w}}_t$

probability flow ODE $\frac{\mathrm{d}\boldsymbol{x}_t}{\mathrm{d}t} = \boldsymbol{f}(\boldsymbol{x}_t, t) - \frac{1}{2} g(t)^2 \nabla_{\boldsymbol{x}} \log q_t(\boldsymbol{x}_t)$

score, unknown

$\mathbf{s}_\theta(\mathbf{x}, t)$ ⟷ $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$

score matching

Song Y, Sohl-Dickstein J, Kingma DP, et al. Score-Based Generative Modeling through Stochastic Differential Equations. ICLR 2021.

The Forward Process

forward SDE

$$d\boldsymbol{x}_t = f(t)\boldsymbol{x}_t dt + g(t) d\boldsymbol{w}_t$$

$$\alpha_t = e^{\int_0^t f(\tau)\mathrm{d}\tau}, \quad \sigma_t^2 = \alpha_t^2 \int_0^t \frac{g^2(\tau)}{\alpha_\tau^2}\mathrm{d}\tau$$

forward transition kernel

$$q_{0t}(\boldsymbol{x}_t|\boldsymbol{x}_0) = N(\alpha_t \boldsymbol{x}_0, \sigma_t^2 \boldsymbol{I})$$

"noise schedule"

$$\boldsymbol{x}_t = \alpha_t \boldsymbol{x}_0 + \sigma_t \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{I})$$

Parameterizations in DPMs

score matching $\quad E_{\boldsymbol{x}_0,\epsilon}[\lambda_1(t)\|\boldsymbol{s}_\theta(\boldsymbol{x}_t,t) - \nabla \log q_t(\boldsymbol{x}_t)\|_2^2]$

intractable

denoising score matching $\quad E_{\boldsymbol{x}_0,\epsilon}[\lambda_1(t)\|\boldsymbol{s}_\theta(\boldsymbol{x}_t,t) - \nabla \log q_{0t}(\boldsymbol{x}_t|\boldsymbol{x}_0)\|_2^2]$

$$= -\frac{\boldsymbol{x}_t - \alpha_t \boldsymbol{x}_0}{\sigma_t^2} = -\frac{\boldsymbol{\epsilon}}{\sigma_t}$$

Define $\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t,t) = -\sigma_t \boldsymbol{s}_\theta(\boldsymbol{x}_t,t)$

$$\boldsymbol{x}_t = \alpha_t \boldsymbol{x}_0 + \sigma_t \boldsymbol{\epsilon}$$

noise prediction $\quad E_{\boldsymbol{x}_0,\epsilon}[\lambda_2(t)\|\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t,t) - \boldsymbol{\epsilon}\|_2^2]$

Parameterizations in DPMs

noise prediction $E_{x_0,\epsilon}[\lambda_2(t)\|\epsilon_\theta(x_t,t)-\epsilon\|_2^2]$

$$x_t=\alpha_t x_0+\sigma_t\epsilon \;\Rightarrow\; x_0=\frac{x_t-\sigma_t\epsilon}{\alpha_t} \;\Rightarrow\; x_\theta(x_t,t)=\frac{x_t-\sigma_t\epsilon_\theta(x_t,t)}{\alpha_t}$$

data prediction $E_{x_0,\epsilon}[\lambda_3(t)\|x_\theta(x_t,t)-x_0\|_2^2]$

Model Type	Training Objective	Example Paper
"noise": noise prediction model ϵ_θ	$E_{x_0,\epsilon,t}\left[\omega_1(t)\|\epsilon_\theta(x_t,t)-\epsilon\|_2^2\right]$	DDPM, Stable-Diffusion
"x_start": data prediction model x_θ	$E_{x_0,\epsilon,t}\left[\omega_2(t)\|x_\theta(x_t,t)-x_0\|_2^2\right]$	DALL·E 2
"v": velocity prediction model v_θ	$E_{x_0,\epsilon,t}\left[\omega_3(t)\|v_\theta(x_t,t)-(\alpha_t\epsilon-\sigma_t x_0)\|_2^2\right]$	Imagen Video
"score": marginal score function s_θ	$E_{x_0,\epsilon,t}\left[\omega_4(t)\|\sigma_t s_\theta(x_t,t)+\epsilon\|_2^2\right]$	ScoreSDE

Content

- **Consistency Models (ICML 2023)**

Technical Improvement

- Improved Techniques for Training CM (ICLR 8866)
- Consistency Trajectory Models (ICLR 8666)

Applications

- Latent Consistency Models (ICLR 6555)
- LCM-LoRA

Consistency from Diffusion ODEs

probability flow ODE

$$\frac{\mathrm{d}\boldsymbol{x}_t}{\mathrm{d}t} = \boldsymbol{f}(\boldsymbol{x}_t, t) - \frac{1}{2}g(t)^2\nabla_{\boldsymbol{x}} \log q_t(\boldsymbol{x}_t)$$

diffusion ODE

$$\frac{\mathrm{d}\boldsymbol{x}_t}{\mathrm{d}t} = f(t)\boldsymbol{x}_t + \frac{g^2(t)}{2\sigma_t}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t),$$

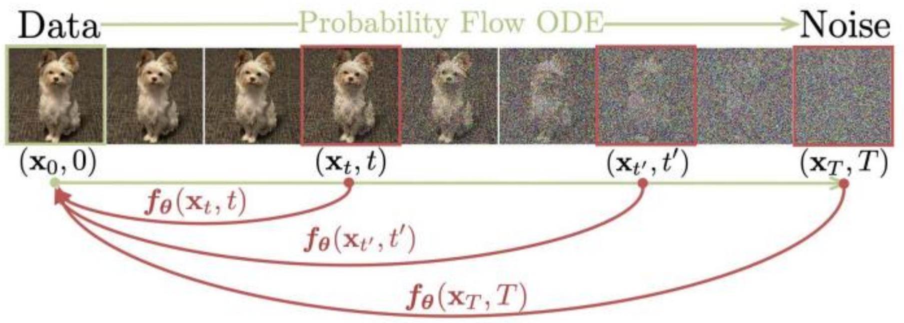


Figure 1: Given a Probability Flow (PF) ODE that smoothly converts data to noise, we learn to map any point (*e.g.*, $\mathbf{x}_t$, $\mathbf{x}_{t'}$, and $\mathbf{x}_T$) on the ODE trajectory to its origin (*e.g.*, $\mathbf{x}_0$) for generative modeling. Models of these mappings are called consistency models, as their outputs are trained to be consistent for points on the same trajectory.

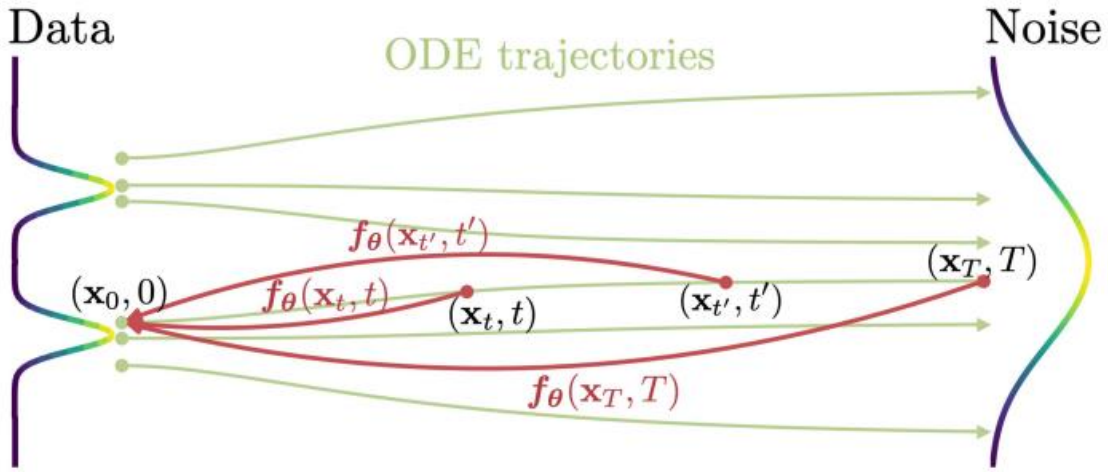


Figure 2: Consistency models are trained to map points on any trajectory of the PF ODE to the trajectory's origin.

consistency function $\boldsymbol{f}_{\boldsymbol{\theta}}(\mathbf{x}, t)$

How to parameterize f_θ?

consistency function $\boldsymbol{f}_{\boldsymbol{\theta}}(\mathbf{x}, t)$

s.t. $(\mathbf{x}_t, t) \mapsto \mathbf{x}_\epsilon \qquad t \in [\epsilon, T]$

boundary condition $\boldsymbol{f}(\mathbf{x}_\epsilon, \epsilon) = \mathbf{x}_\epsilon$

$$\boldsymbol{f}_{\boldsymbol{\theta}}(\mathbf{x}, t) = c_{\text{skip}}(t)\mathbf{x} + c_{\text{out}}(t)\boxed{F_{\boldsymbol{\theta}}(\mathbf{x}, t)}$$

Free-form NN

$$c_{\text{skip}}(\epsilon) = 1 \qquad c_{\text{out}}(\epsilon) = 0$$

Noise Schedule and Parameterization

- Following EDM, CM applied the *VE* schedule

$$\alpha_t = 1, \qquad \sigma_t = t$$

- The diffusion ODE is simply

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = -t\boldsymbol{s}_\phi(\mathbf{x}_t, t).$$

- The parameterizations:

$$\boxed{\boldsymbol{f}_{\boldsymbol{\theta}}(\mathbf{x}, t)} = c_{\text{skip}}(t)\mathbf{x} + c_{\text{out}}(t)F_{\boldsymbol{\theta}}(\mathbf{x}, t)$$

consistency function (CM) or data predictor (EDM)

CM

$$c_{\text{skip}}(t) = \frac{\sigma_{\text{data}}^2}{(t-\epsilon)^2 + \sigma_{\text{data}}^2}, \quad c_{\text{out}}(t) = \frac{\sigma_{\text{data}}(t-\epsilon)}{\sqrt{\sigma_{\text{data}}^2 + t^2}},$$

model transferable

EDM

$$c_{\text{skip}}(\sigma) = \sigma_{\text{data}}^2 / (\sigma^2 + \sigma_{\text{data}}^2)$$
$$c_{\text{out}}(\sigma) = \sigma \cdot \sigma_{\text{data}} / \sqrt{\sigma^2 + \sigma_{\text{data}}^2}$$

Types of CM

Consistency Distillation (CD)

Distill ODE trajectories of a teacher EDM model ϕ

$$\mathcal{L}_{CD}^{N}(\boldsymbol{\theta}, \boldsymbol{\theta}^{-}; \phi) := \mathbb{E}[\lambda(t_n) d(\boldsymbol{f}_{\boldsymbol{\theta}}(\mathbf{x}_{t_{n+1}}, t_{n+1}), \boldsymbol{f}_{\boldsymbol{\theta}^{-}}(\hat{\mathbf{x}}_{t_n}^{\phi}, t_n))]$$

$$\hat{\mathbf{x}}_{t_n}^{\phi} = \mathbf{x}_{t_{n+1}} - (t_n - t_{n+1}) t_{n+1} \boldsymbol{s}_{\phi}(\mathbf{x}_{t_{n+1}}, t_{n+1})$$

one-step ODE update

Consistency Training (CT)

Learn consistent ODE trajectories from data

$$\mathbb{E}[\lambda(t_n) d(\boldsymbol{f}_{\boldsymbol{\theta}}(\mathbf{x} + t_{n+1}\mathbf{z}, t_{n+1}), \boldsymbol{f}_{\boldsymbol{\theta}^{-}}(\mathbf{x} + t_n \mathbf{z}, t_n))]$$

$$\boldsymbol{\theta}^{-} \leftarrow \operatorname{stopgrad}(\mu \boldsymbol{\theta}^{-} + (1 - \mu)\boldsymbol{\theta})$$

"EMA self-teacher"

Sampling with CM

Algorithm 1 Multistep Consistency Sampling

Input: Consistency model $\boldsymbol{f_\theta}(\cdot, \cdot)$, sequence of time points $\tau_1 > \tau_2 > \cdots > \tau_{N-1}$, initial noise $\hat{\mathbf{x}}_T$

$\mathbf{x} \leftarrow \boldsymbol{f_\theta}(\hat{\mathbf{x}}_T, T)$

for $n = 1$ **to** $N - 1$ **do**

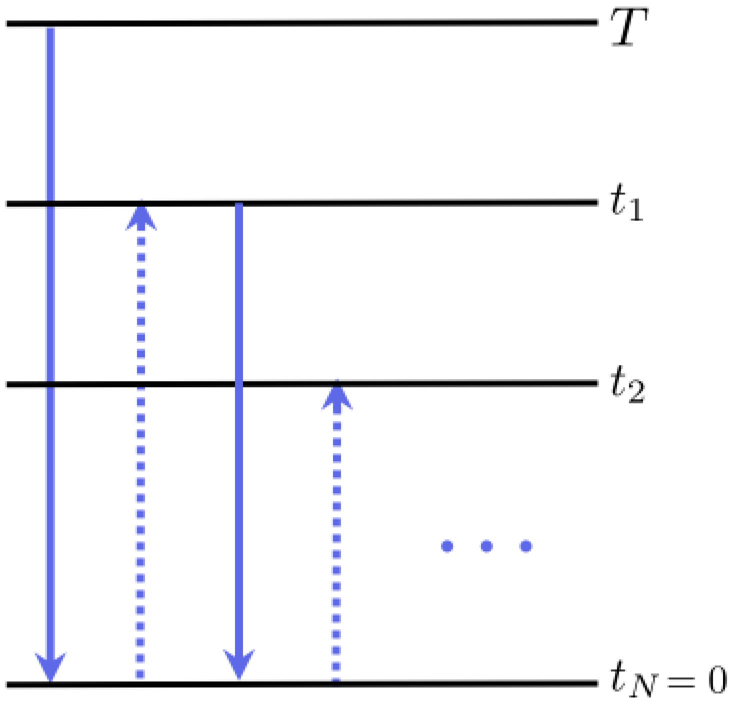
  Sample $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$

  $\hat{\mathbf{x}}_{\tau_n} \leftarrow \mathbf{x} + \sqrt{\tau_n^2 - \epsilon^2}\mathbf{z}$

  $\mathbf{x} \leftarrow \boldsymbol{f_\theta}(\hat{\mathbf{x}}_{\tau_n}, \tau_n)$

end for

Output: $\mathbf{x}$

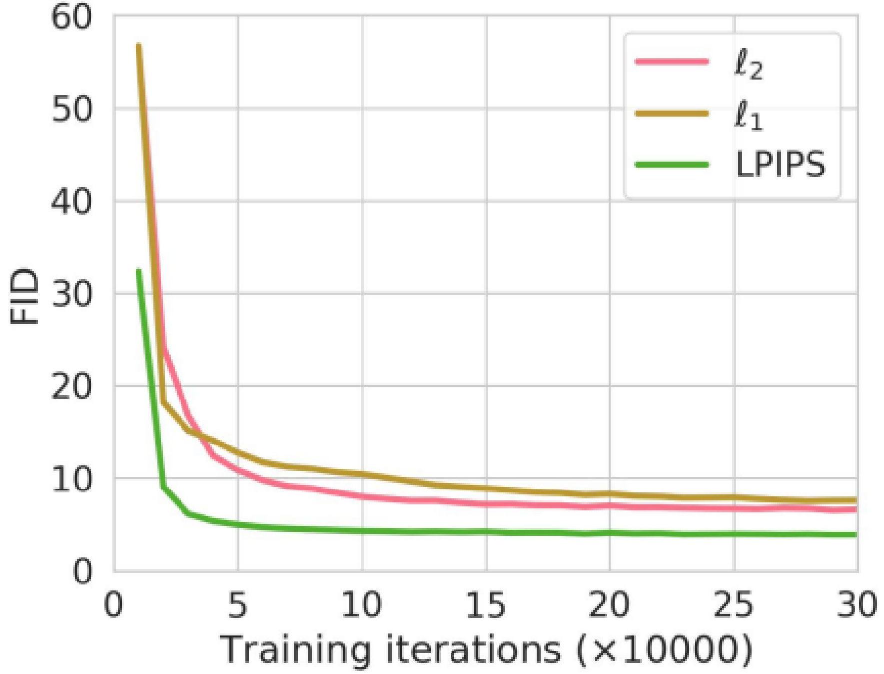


"Two-step generation often enhances the quality of one-step generation considerably, though increasing the number of sampling steps further provides diminishing benefits."

Choose the Distance Metric

$$\mathcal{L}_{CD}^{N}(\boldsymbol{\theta}, \boldsymbol{\theta}^{-}; \phi) := \mathbb{E}[\lambda(t_n) d(\boldsymbol{f}_{\boldsymbol{\theta}}(\mathbf{x}_{t_{n+1}}, t_{n+1}), \boldsymbol{f}_{\boldsymbol{\theta}^{-}}(\hat{\mathbf{x}}_{t_n}^{\phi}, t_n))]$$

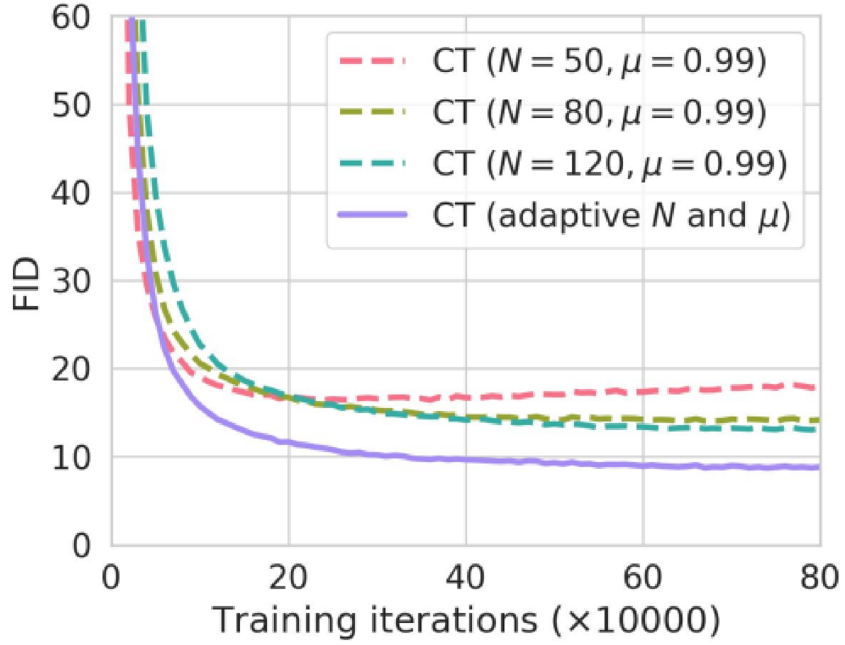
$$\mathbb{E}[\lambda(t_n) d(\boldsymbol{f}_{\boldsymbol{\theta}}(\mathbf{x} + t_{n+1}\mathbf{z}, t_{n+1}), \boldsymbol{f}_{\boldsymbol{\theta}^{-}}(\mathbf{x} + t_n\mathbf{z}, t_n))]$$



(a) Metric functions in CD.

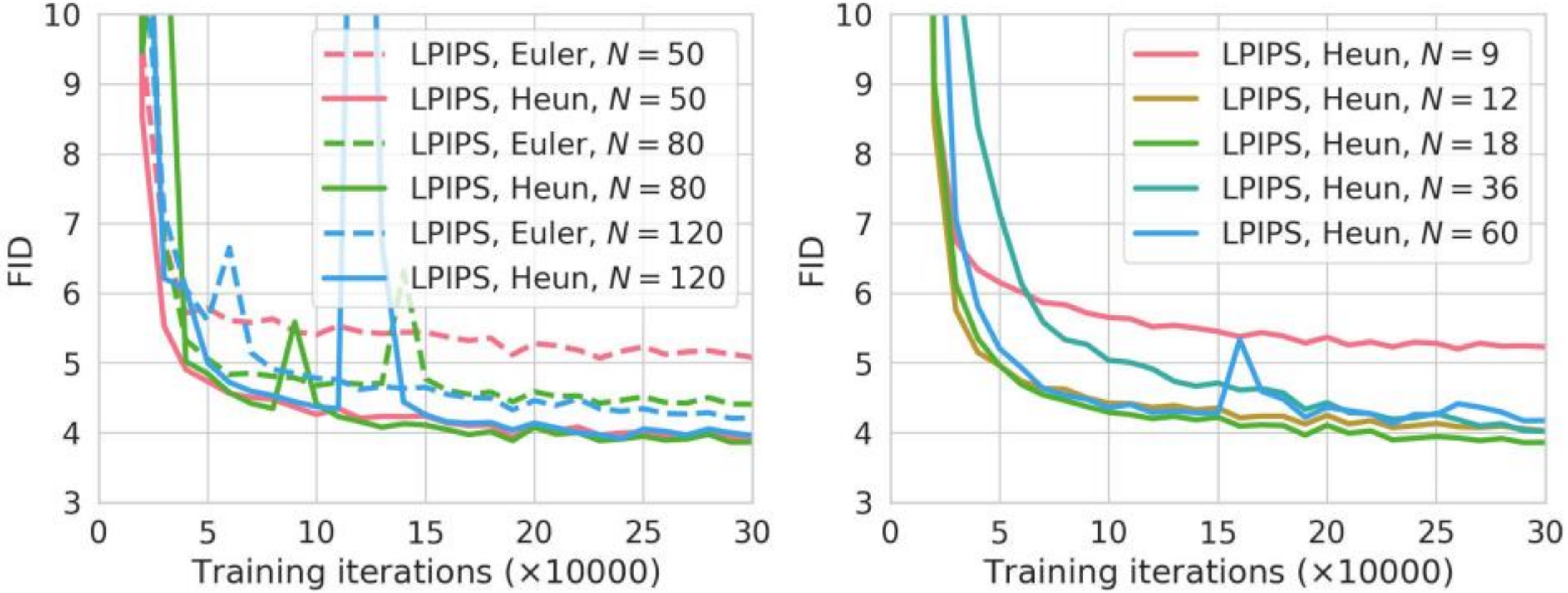
Choose the Number of Timesteps and EMA

- The number of timesteps: schedule $N(\cdot)$
- EMA rate: schedule $\mu(\cdot)$



(d) Adaptive N and μ in CT.

The One-Step ODE Solver



(b) Solvers and N in CD. (c) N with Heun solver in CD.

Results

Table 1: Sample quality on CIFAR-10. *Methods that require synthetic data construction for distillation.

METHOD	NFE (↓)	FID (↓)	IS (↑)
Diffusion + Samplers			
DDIM (Song et al., 2020)	50	4.67	
DDIM (Song et al., 2020)	20	6.84	
DDIM (Song et al., 2020)	10	8.23	
DPM-solver-2 (Lu et al., 2022)	10	5.94	
DPM-solver-fast (Lu et al., 2022)	10	4.70	
3-DEIS (Zhang & Chen, 2022)	10	**4.17**	
Diffusion + Distillation			
Knowledge Distillation* (Luhman & Luhman, 2021)	1	9.36	
DFNO* (Zheng et al., 2022)	1	4.12	
1-Rectified Flow (+distill)* (Liu et al., 2022)	1	6.18	9.08
2-Rectified Flow (+distill)* (Liu et al., 2022)	1	4.85	9.01
3-Rectified Flow (+distill)* (Liu et al., 2022)	1	5.21	8.79
PD (Salimans & Ho, 2022)	1	8.34	8.69
CD	1	**3.55**	**9.48**
PD (Salimans & Ho, 2022)	2	5.58	9.05
CD	2	**2.93**	**9.75**
Direct Generation			
BigGAN (Brock et al., 2019)	1	14.7	9.22
Diffusion GAN (Xiao et al., 2022)	1	14.6	8.93
AutoGAN (Gong et al., 2019)	1	12.4	8.55
E2GAN (Tian et al., 2020)	1	11.3	8.51
ViTGAN (Lee et al., 2021)	1	6.66	9.30
TransGAN (Jiang et al., 2021)	1	9.26	9.05
StyleGAN2-ADA (Karras et al., 2020)	1	2.92	**9.83**
StyleGAN-XL (Sauer et al., 2022)	1	**1.85**	
Score SDE (Song et al., 2021)	2000	2.20	**9.89**
DDPM (Ho et al., 2020)	1000	3.17	9.46
LSGM (Vahdat et al., 2021)	147	2.10	
PFGM (Xu et al., 2022)	110	2.35	9.68
EDM (Karras et al., 2022)	35	**2.04**	9.84
1-Rectified Flow (Liu et al., 2022)	1	378	1.13
Glow (Kingma & Dhariwal, 2018)	1	48.9	3.92
Residual Flow (Chen et al., 2019)	1	46.4	
GLFlow (Xiao et al., 2019)	1	44.6	
DenseFlow (Grcić et al., 2021)	1	34.9	
DC-VAE (Parmar et al., 2021)	1	17.9	8.20
CT	1	**8.70**	**8.49**
CT	2	**5.83**	**8.85**

Content

- Consistency Models (ICML 2023)

Technical Improvement

- **Improved Techniques for Training CM (ICLR 8866)**
- Consistency Trajectory Models (ICLR 8666)

Applications to Text-to-Image Model

- Latent Consistency Models (ICLR 6555)
- LCM-LoRA

Motivation

- CD requires an additional model and has limited performance
- CD and CT relies on LPIPS, which may leak ImageNet features and inflate FID

Goal:

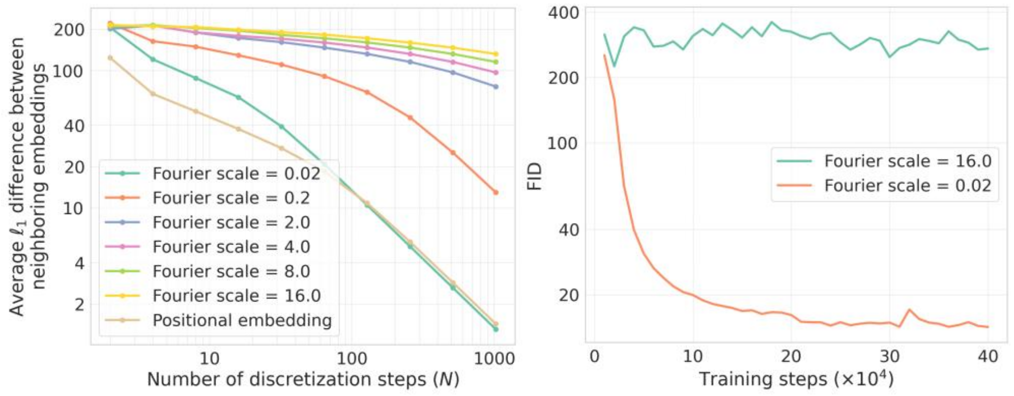
- Improve the two-step generation of CT to 100-step generation of diffusion models

Improved Techniques (1): weighting, Fourier scale and dropout

- Weighting: *larger weight at lower noise levels*

Weighting function	$\lambda(\sigma_i) = 1$	$\lambda(\sigma_i) = \frac{1}{\sigma_{i+1} - \sigma_i}$

- Fourier scale: *less sensitive noise embedding layer*



(a) Sensitivity of noise embeddings. (b) Continuous-time CT.

- Dropout: larger rate

Improved Techniques (2): remove EMA for teacher

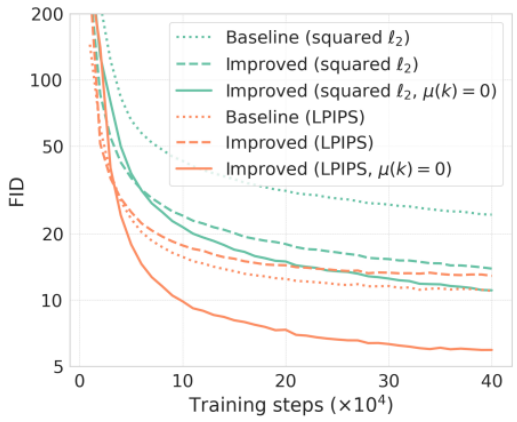
- EMA causes inconsistency for CT *even when the data is a single point ξ*

Proposition 1. *Given the notations introduced earlier, and using the uniform weighting function $\lambda(\sigma) = 1$ along with the squared ℓ_2 metric, we have*

no signals of ξ

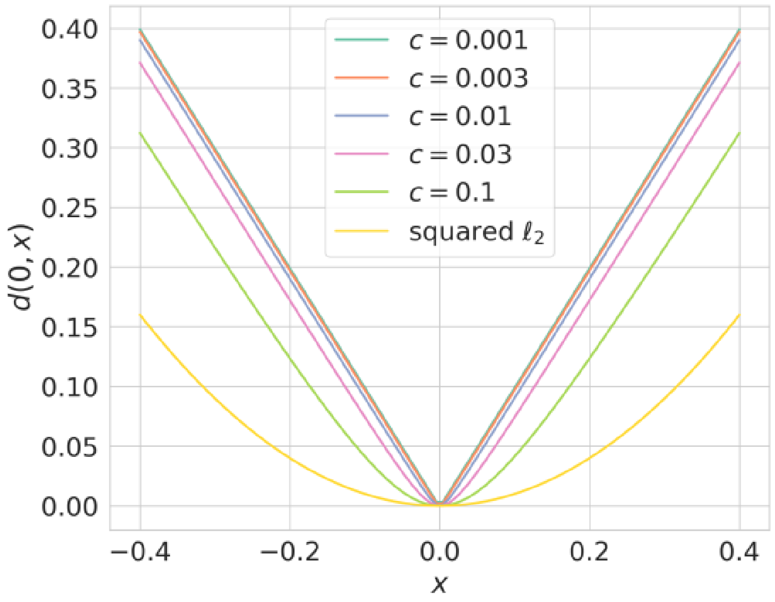
$$\lim_{N\to\infty} \mathcal{L}^N(\theta,\theta^-) = \lim_{N\to\infty} \mathcal{L}^N_{CT}(\theta,\theta^-) = \mathbb{E}\left[\left(1-\frac{\sigma_{min}}{\sigma_i}\right)^2(\theta-\theta^-)^2\right] \quad \text{if } \theta^- \neq \theta \tag{6}$$

$$\lim_{N\to\infty} \frac{1}{\Delta\sigma}\frac{\mathrm{d}\mathcal{L}^N(\theta,\theta^-)}{\mathrm{d}\theta} = \begin{cases} \frac{\mathrm{d}}{\mathrm{d}\theta}\mathbb{E}\left[\frac{\sigma_{min}}{\sigma_i^2}\left(1-\frac{\sigma_{min}}{\sigma_i}\right)(\theta-\xi)^2\right], & \theta^- = \theta \\ +\infty, & \theta^- < \theta \\ -\infty, & \theta^- > \theta \end{cases} \tag{7}$$



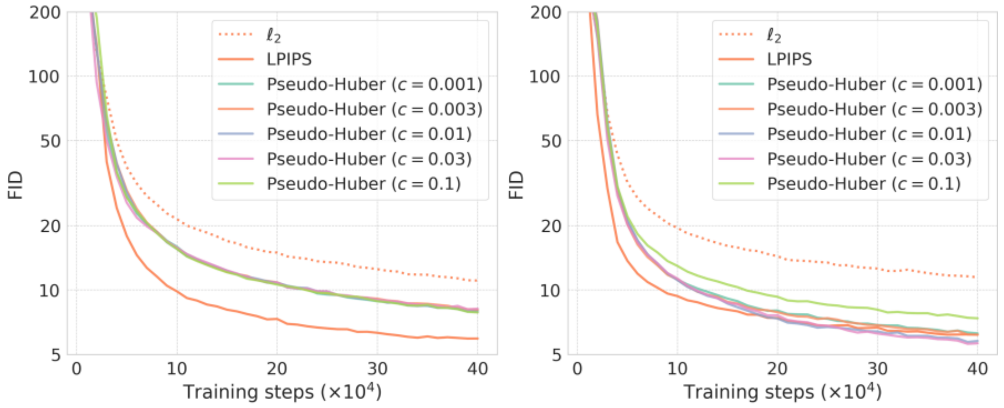
Improved Techniques (3): Pseudo-Huber loss

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\|\boldsymbol{x} - \boldsymbol{y}\|_2^2 + c^2} - c$$



(a) $d(\mathbf{0}, \boldsymbol{x})$ as a function of $\boldsymbol{x}$.

$c = 0.00054\sqrt{d}$, d is data dimensionality



(b) $s_0 = 2, s_1 = 150$ (c) $s_0 = 10, s_1 = 1280$

Improved Techniques (4): discretization/noise schedule

Discretization curriculum	$N(k) = \left\lceil \sqrt{\frac{k}{K}((s_1+1)^2 - s_0^2) + s_0^2} - 1 \right\rceil + 1$	$N(k) = \min(s_0 2^{\lfloor \frac{k}{K'} \rfloor}, s_1) + 1$, where $K' = \left\lfloor \frac{K}{\log_2 \lfloor s_1/s_0 \rfloor + 1} \right\rfloor$
	$s_0 = 2, s_1 = 150, \mu_0 = 0.9$ on CIFAR-10 $s_0 = 2, s_1 = 200, \mu_0 = 0.95$ on ImageNet 64×64	$s_0 = 10, s_1 = 1280$ $c = 0.00054\sqrt{d}$, d is data dimensionality
Noise schedule	σ_i, where $i \sim \mathcal{U}[\![1, N(k) - 1]\!]$	σ_i, where $i \sim p(i)$, and $p(i) \propto \operatorname{erf}\left(\frac{\log(\sigma_{i+1}) - P_{\text{mean}}}{\sqrt{2} P_{\text{std}}}\right) - \operatorname{erf}\left(\frac{\log(\sigma_i) - P_{\text{mean}}}{\sqrt{2} P_{\text{std}}}\right)$

Results

Direct Generation			
Score SDE (Song et al., 2021)	2000	2.38	9.83
Score SDE (deep) (Song et al., 2021)	2000	2.20	9.89
DDPM (Ho et al., 2020)	1000	3.17	9.46
LSGM (Vahdat et al., 2021)	147	2.10	
PFGM (Xu et al., 2022)	110	2.35	9.68
EDM* (Karras et al., 2022)	35	2.04	9.84
EDM-G++ (Kim et al., 2023)	35	1.77	
IGEBM (Du & Mordatch, 2019)	60	40.6	6.02
NVAE (Vahdat & Kautz, 2020)	1	23.5	7.18
Glow (Kingma & Dhariwal, 2018)	1	48.9	3.92
Residual Flow (Chen et al., 2019)	1	46.4	
BigGAN (Brock et al., 2019)	1	14.7	9.22
StyleGAN2 (Karras et al., 2020b)	1	8.32	9.21
StyleGAN2-ADA (Karras et al., 2020a)	1	2.92	9.83
CT (LPIPS) (Song et al., 2023)	1	8.70	8.49
	2	5.83	8.85
iCT (ours)	1	2.83	9.54
	2	2.46	9.80
iCT-deep (ours)	1	2.51	9.76
	2	2.24	9.89

Content

- Consistency Models (ICML 2023)

Technical Improvement

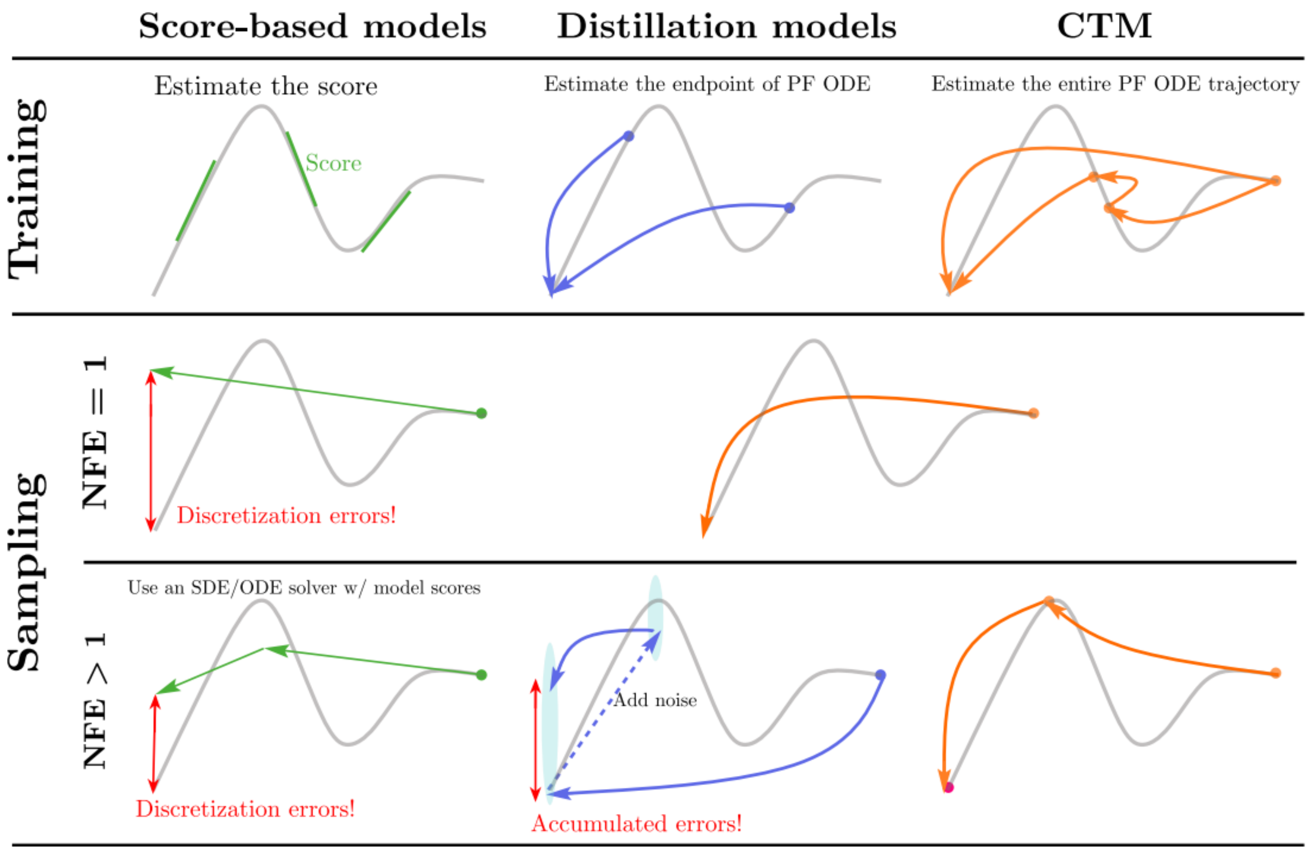
- Improved Techniques for Training CM (ICLR 8866)
- **Consistency Trajectory Models (ICLR 8666)**

Applications to Text-to-Image Model

- Latent Consistency Models (ICLR 6555)
- LCM-LoRA

Motivation

- Is it reasonable to always predict the clean data at time 0 ?



CTM consistency function:

$$G(\mathbf{x}_t, t, s) := \mathbf{x}_t + \int_t^s \frac{\mathbf{x}_u - \mathbb{E}[\mathbf{x}|\mathbf{x}_u]}{u} \, \mathrm{d}u$$

Jump from time t to s
(√) deterministic sampling
(√) likelihood computation

How to Parameterize G?

consistency function

$$G(\mathbf{x}_t, t, s) := \mathbf{x}_t + \int_t^s \frac{\mathbf{x}_u - \mathbb{E}[\mathbf{x}|\mathbf{x}_u]}{u} \, \mathrm{d}u$$

boundary condition

$$G(\boldsymbol{x}_t, t, t) = \boldsymbol{x}_t, \qquad G(\boldsymbol{x}_t, t, 0) = f(\boldsymbol{x}_t, t)$$

$$G(\mathbf{x}_t, t, s) = \frac{s}{t}\mathbf{x}_t + \left(1 - \frac{s}{t}\right) \boxed{g(\mathbf{x}_t, t, s)}$$

The Property of g

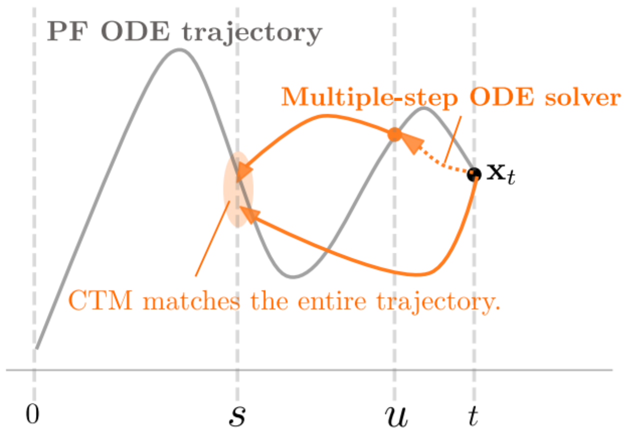
$$\lim_{s \to t} g(\mathbf{x}_t, t, s) = \mathbf{x}_t + t \lim_{s \to t} \frac{1}{t-s} \int_t^s \frac{\mathbf{x}_u - \mathbb{E}[\mathbf{x}|\mathbf{x}_u]}{u} \mathrm{d}u = \mathbb{E}[\mathbf{x}|\mathbf{x}_t]$$

$g(\boldsymbol{x}_t, t, t)$ is the data predictor!

$$\boldsymbol{x}_\theta(\boldsymbol{x}_t, t) = \frac{\boldsymbol{x}_t - \sigma_t \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)}{\alpha_t}$$

We can add extra score matching loss

CTM training loss



$$\mathcal{L}_{\text{CTM}}(\boldsymbol{\theta};\boldsymbol{\phi}) := \mathbb{E}_{t\in[0,T]}\mathbb{E}_{s\in[0,t]}\mathbb{E}_{u\in[s,t)}\mathbb{E}_{\mathbf{x}_0,p_{0t}(\mathbf{x}|\mathbf{x}_0)}\Big[d\big(\mathbf{x}_{\text{target}}(\mathbf{x},t,u,s),\mathbf{x}_{\text{est}}(\mathbf{x},t,s)\big)\Big]$$

$$\mathbf{x}_{\text{target}}: \quad G_{\text{sg}(\boldsymbol{\theta})}\big(G_{\text{sg}(\boldsymbol{\theta})}\big(\texttt{Solver}(\mathbf{x}_t,t,u;\boldsymbol{\phi}),u,s\big),s,0\big)$$

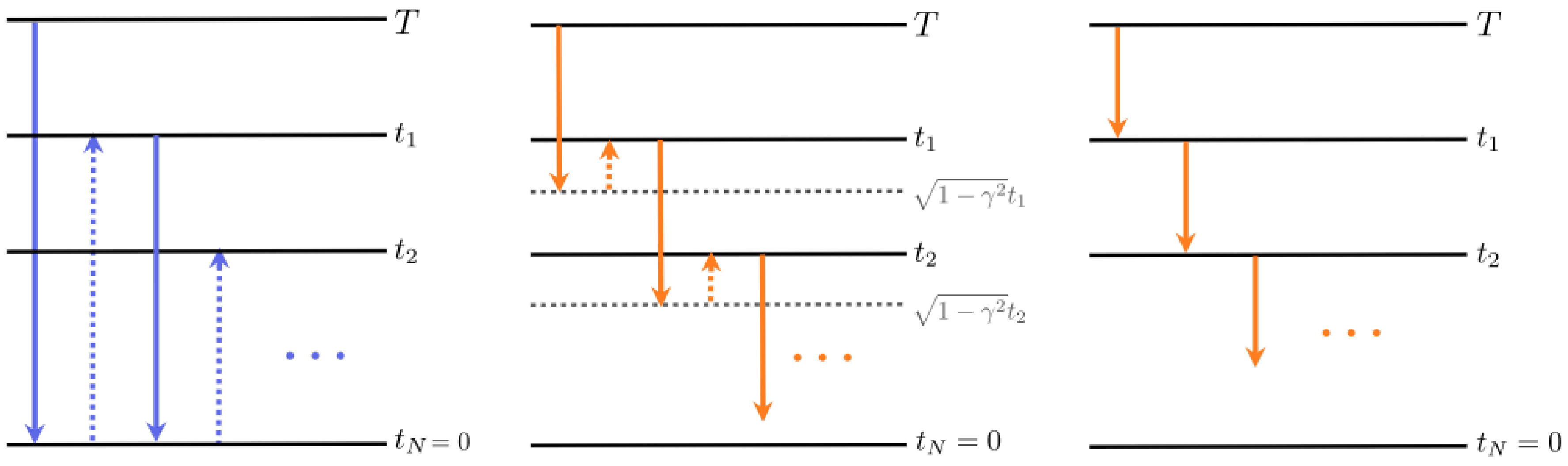
$$\mathbf{x}_{\text{est}}: \quad G_{\text{sg}(\boldsymbol{\theta})}\big(G_{\boldsymbol{\theta}}(\mathbf{x}_t,t,s),s,0\big)$$

A combination of CD and CT!

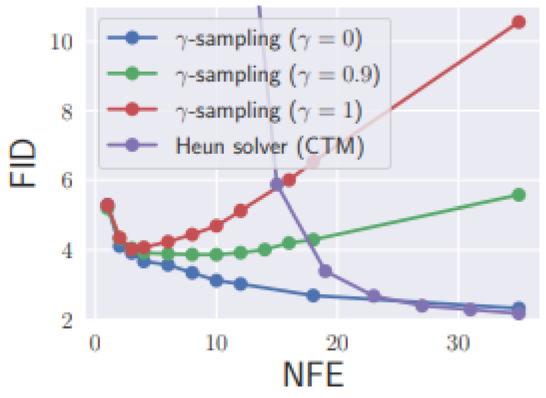
$$\mathcal{L}_{\text{DSM}}(\boldsymbol{\theta}) = \mathbb{E}_{t,\mathbf{x}_0,\mathbf{x}_t|\mathbf{x}_0}\big[\|\mathbf{x}_0 - g_{\boldsymbol{\theta}}(\mathbf{x}_t,t,t)\|_2^2\big]$$

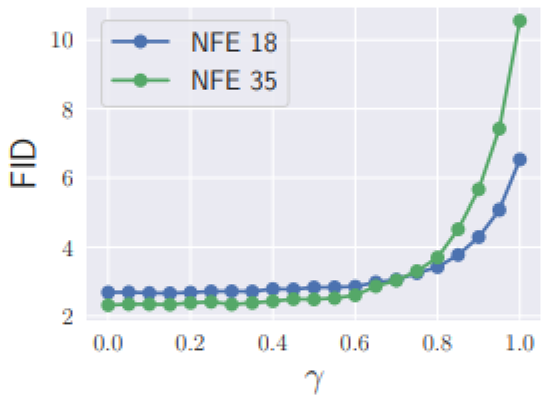
$$\mathcal{L}_{\text{GAN}}(\boldsymbol{\theta},\boldsymbol{\eta}) = \mathbb{E}_{p_{\text{data}}(\mathbf{x}_0)}[\log d_{\boldsymbol{\eta}}(\mathbf{x}_0)] + \mathbb{E}_{t,\mathbf{x}_t}\big[\log\big(1-d_{\boldsymbol{\eta}}(\mathbf{x}_{\text{est}})\big)\big]$$

γ-Sampling



(a) $\gamma = 1$ (Fully stochastic) (b) $1 > \gamma > 0$ (c) $\gamma = 0$ (Deterministic)





Results

Table 2: Performance comparisons on CIFAR-10.

Model	NFE	Unconditional		Conditional
		FID↓	NLL↓	FID↓
GAN Models				
BigGAN (Brock et al., 2018)	1	8.51	✗	-
StyleGAN-Ada (Karras et al., 2020)	1	2.92	✗	2.42
StyleGAN-D2D (Kang et al., 2021)	1	-	✗	2.26
StyleGAN-XL (Sauer et al., 2022)	1	-	✗	1.85
Diffusion Models – Score-based Sampling				
DDPM (Ho et al., 2020)	1000	3.17	3.75	-
DDIM (Song et al., 2020a)	100	4.16	-	-
	10	13.36	-	-
Score SDE (Song et al., 2020a)	2000	2.20	3.45	-
VDM (Kingma et al., 2021)	1000	7.41	2.49	-
LSGM (Vahdat et al., 2021)	138	2.10	3.43	-
EDM (Karras et al., 2022)	35	2.01	2.56	1.82
Diffusion Models – Distillation Sampling				
KD (Luhman & Luhman, 2021)	1	9.36	✗	-
DFNO (Zheng et al., 2023)	1	3.78	✗	-
2-Rectified Flow (Liu et al., 2022)	1	4.85	✗	-
PD (Salimans & Ho, 2021)	1	9.12	✗	-
CD (official report) (Song et al., 2023)	1	3.55	✗	-
CD (retrained)	1	10.53	✗	-
CD + GAN (Lu et al., 2023)	1	2.65	✗	-
CTM (ours)	1	1.98	**2.43**	1.73
PD (Salimans & Ho, 2021)	2	4.51	-	-
CD (Song et al., 2023)	2	2.93	-	-
CTM (ours)	2	**1.87**	**2.43**	**1.63**
Models without Pre-trained DM – Direct Generation				
CT	1	8.70	✗	-
CTM (ours)	1	2.39	-	-

The GAN loss is tricky in improving FID.

Content

- Consistency Models (ICML 2023)

Technical Improvement

- Improved Techniques for Training CM (ICLR 8866)
- Consistency Trajectory Models (ICLR 8666)

Applications to Text-to-Image Model

- Latent Consistency Models (ICLR 6555)
- LCM-LoRA

Latent Consistency Models (LCM) w.r.t. CM

- Parameterization for more general noise schedule

$$\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{z}, \boldsymbol{c}, t) = c_{\text{skip}}(t)\boldsymbol{z} + c_{\text{out}}(t)\left(\frac{\boldsymbol{z} - \sigma_t \hat{\boldsymbol{\epsilon}}_\theta(\boldsymbol{z}, \boldsymbol{c}, t)}{\alpha_t}\right)$$

- To cope with classifier-free guidance:

$$\mathcal{L}_{\mathcal{CD}}\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{-}; \Psi\right) = \mathbb{E}_{\boldsymbol{z}, \boldsymbol{c}, \omega, n}\left[d\left(\boxed{\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{z}_{t_{n+1}}, \omega, \boldsymbol{c}, t_{n+1})}, \boldsymbol{f}_{\boldsymbol{\theta}^{-}}(\hat{\boldsymbol{z}}_{t_n}^{\Psi,\omega}, \omega, \boldsymbol{c}, t_n)\right)\right]$$

augmented consistency function with scale ω

- *Skipping timesteps* for accelerated training

$$\mathcal{L}_{\mathcal{CD}}\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{-}; \Psi\right) = \mathbb{E}_{\boldsymbol{z}, \boldsymbol{c}, \omega, n}\left[d\left(\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{z}_{t_{n+k}}, \omega, \boldsymbol{c}, \boxed{t_{n+k}}), \boldsymbol{f}_{\boldsymbol{\theta}^{-}}(\hat{\boldsymbol{z}}_{t_n}^{\Psi,\omega}, \omega, \boldsymbol{c}, \boxed{t_n})\right)\right]$$

Skipping timesteps for accelerated training

$$\mathcal{L}_{\mathcal{CD}}\left(\boldsymbol{\theta}, \boldsymbol{\theta}^{-}; \Psi\right) = \mathbb{E}_{\boldsymbol{z}, \boldsymbol{c}, \omega, n}\left[d\left(\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{z}_{t_{n+k}}, \omega, \boldsymbol{c}, t_{n+k}), \boldsymbol{f}_{\boldsymbol{\theta}^{-}}(\hat{\boldsymbol{z}}_{t_n}^{\Psi,\omega}, \omega, \boldsymbol{c}, t_n)\right)\right]$$

k too small: slow convergence

k too large: large discretization error

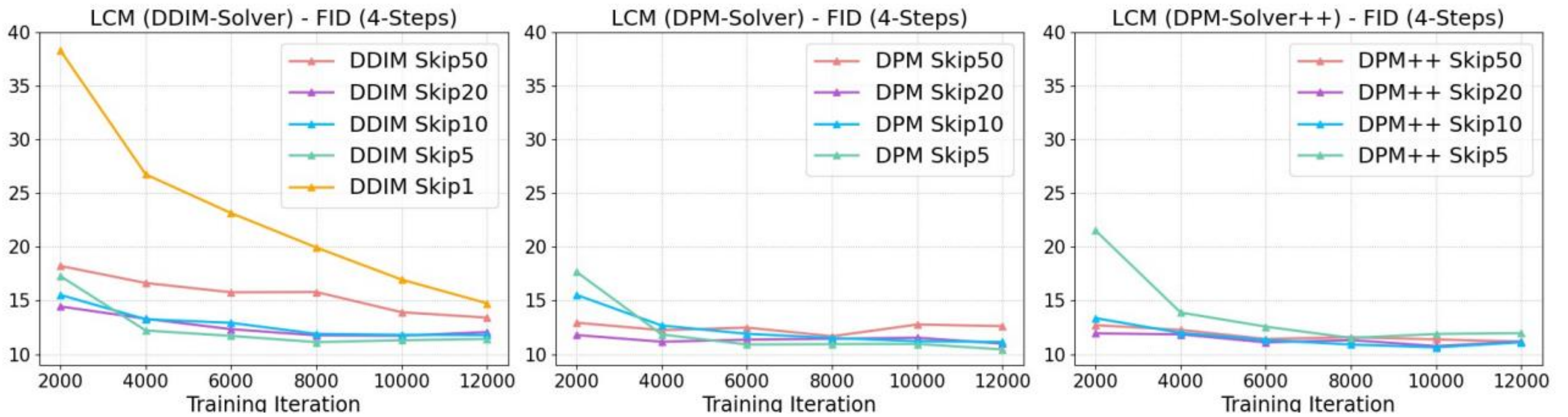


Figure 3: Ablation study on different ODE solvers and skipping step k. Appropriate skipping step k can significantly accelerate convergence and lead to better FID within the same number of training steps.

LCM: Results

Model (512 × 512) Reso	FID ↓				CLIP Score ↑			
	1 Step	2 Steps	4 Steps	8 Steps	1 Steps	2 Steps	4 Steps	8 Steps
DDIM (Song et al., 2020a)	183.29	81.05	22.38	13.83	6.03	14.13	25.89	29.29
DPM (Lu et al., 2022a)	185.78	72.81	18.53	12.24	6.35	15.10	26.64	29.54
DPM++ (Lu et al., 2022b)	185.78	72.81	18.43	12.20	6.35	15.10	26.64	**29.55**
Guided-Distill (Meng et al., 2023)	108.21	33.25	15.12	13.89	12.08	22.71	27.25	28.17
LCM (Ours)	**35.36**	**13.31**	**11.10**	**11.84**	**24.14**	**27.83**	**28.69**	28.84

Table 1: Quantitative results with $\omega = 8$ at 512×512 resolution. LCM significantly surpasses baselines in the 1-4 step region on LAION-Aesthetic-6+ dataset. For LCM, DDIM-Solver is used with a skipping step of $k = 20$.

Model (768 × 768) Reso	FID ↓				CLIP Score ↑			
	1 Step	2 Steps	4 Steps	8 Steps	1 Steps	2 Steps	4 Steps	8 Steps
DDIM (Song et al., 2020a)	186.83	77.26	24.28	15.66	6.93	16.32	26.48	29.49
DPM (Lu et al., 2022a)	188.92	67.14	20.11	**14.08**	7.40	17.11	27.25	29.80
DPM++ (Lu et al., 2022b)	188.91	67.14	20.08	14.11	7.41	17.11	27.26	**29.84**
Guided-Distill (Meng et al., 2023)	120.28	30.70	16.70	14.12	12.88	24.88	28.45	29.16
LCM (Ours)	**34.22**	**16.32**	**13.53**	14.97	**25.32**	**27.92**	**28.60**	28.49

Table 2: Quantitative results with $\omega = 8$ at 768×768 resolution. LCM significantly surpasses the baselines in the 1-4 step region on LAION-Aesthetic-6.5+ dataset. For LCM, DDIM-Solver is used with a skipping step of $k = 20$.

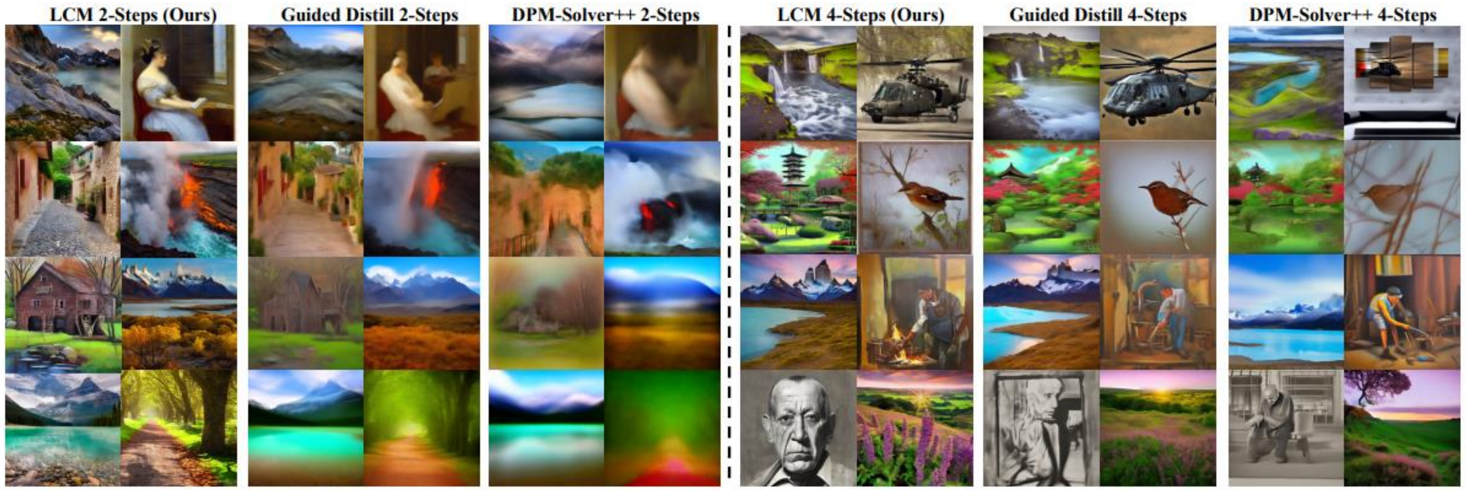
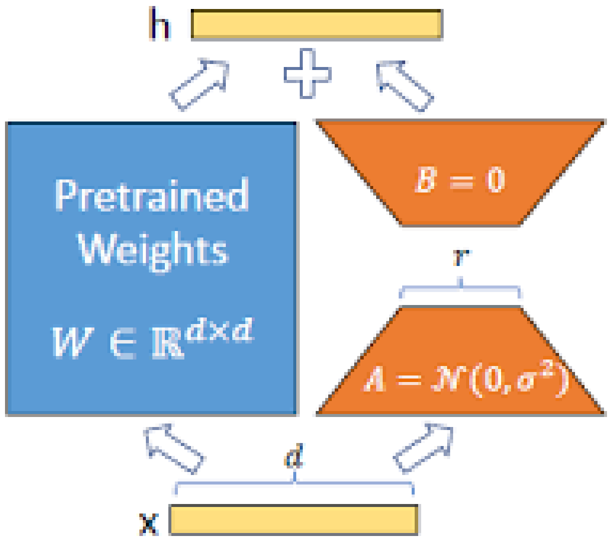


Figure 2: Text-to-Image generation results on LAION-Aesthetic-6.5+ with 2-, 4-step inference. Images generated by LCM exhibit superior detail and quality, outperforming other baselines by a large margin.

LoRA

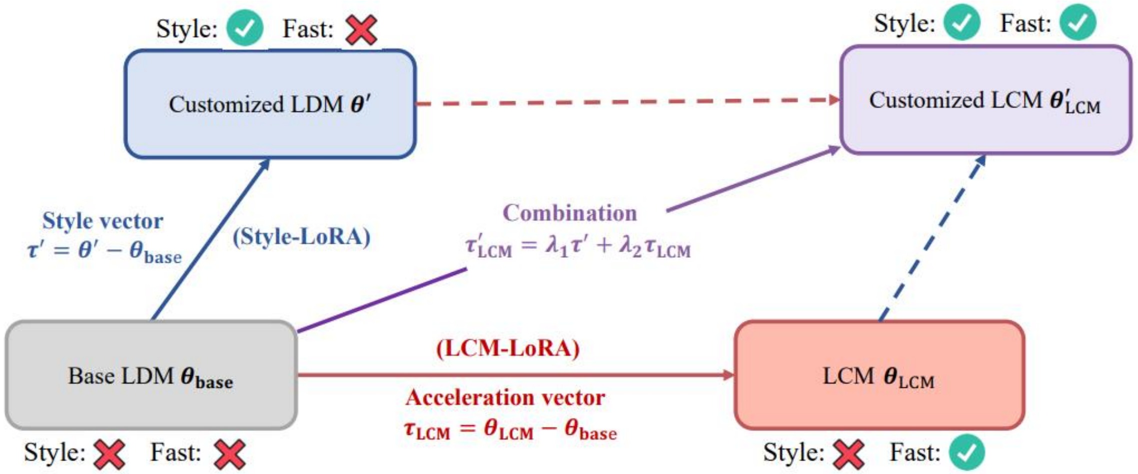


$$W' = W + \Delta W$$
$$\Delta W = AB^T,$$

Merge LoRA with LoRA

$$\Delta W = (\alpha_1 A_1 + \alpha_2 A_2)(\alpha_1 B_1 + \alpha_2 B_2)^T$$

LCM-LoRA



Model	SD-V1.5	SSD-1B	SDXL
# Full Parameters	0.98B	1.3B	3.5B
# LoRA Trainable Parameters	67.5M	105M	197M

References

[1] Yang Song, Prafulla Dhariwal, Mark Chen, Ilya Sutskever. Consistency models. ICML 2023

[2] Yang Song, Prafulla Dhariwal. Improved Techniques for Training Consistency Models. arXiv:2310.14189 (ICLR 8866)

[3] Dongjun Kim, Chieh-Hsin Lai, Wei-Hsiang Liao, Naoki Murata, Yuhta Takida, Toshimitsu Uesaka, Yutong He, Yuki Mitsufuji, Stefano Ermon. Consistency Trajectory Models: Learning Probability Flow ODE Trajectory of Diffusion. arXiv:2310.02279 (ICLR 8666)

[4] Simian Luo, Yiqin Tan, Longbo Huang, Jian Li, Hang Zhao. Latent Consistency Models: Synthesizing High-Resolution Images with Few-Step Inference. arXiv:2310.04378 (ICLR 6555)

[5] Simian Luo, Yiqin Tan, Suraj Patil, Daniel Gu, Patrick von Platen, Apolinário Passos, Longbo Huang, Jian Li, Hang Zhao. LCM-LoRA: A Universal Stable-Diffusion Acceleration Module. arXiv:2311.05556