SPLAM: Accelerating Image Generation with **S**ub-**P**ath **L**inear **A**pproximation **M**odel

Chen Xu[1,2*], Tianhui Song[1,2*], Weixin Feng[2], Xubin Li[2],

Tiezheng Ge[2], Bo Zheng[2], Limin Wang[1,3]

[1]Nanjing University [2]Alibaba Group [3]Shanghai AI Lab

* Equal contribution

Oral Presentation

Preliminaries： Effectiveness of Consistency models

- Training for one ideal denioser $\boldsymbol{D}_{\boldsymbol{\theta}}$ for $\boldsymbol{x}_t$:

$$Minimize\ \mathcal{L} = \mathbb{E}[|\boldsymbol{D}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) - \alpha(t)\boldsymbol{x}_0|]$$

- The approximation in Consistency Models :

One well-learned generator $\boldsymbol{f}_{\boldsymbol{\theta}}$: $\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}, t-1) \approx \boldsymbol{x}_0$

- and the approximated learning objective ($\boldsymbol{D}_{\boldsymbol{\theta}} = \alpha(t)\boldsymbol{f}_{\boldsymbol{\theta}}$):

$$Minimize\ \mathcal{L} = \mathbb{E}[|D_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) - \alpha(t)\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}, t-1)|]$$

- also as

$$\mathcal{L}_{Approx}(\boldsymbol{\theta}) = \mathbb{E}[|\boldsymbol{x}_t - \frac{\alpha(t)}{\alpha(t-1)}x_{t-1} + \frac{\alpha(t)}{\alpha(t-1)}\sigma(t-1)\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}, t-1) - \sigma(t)\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)\ |]$$

Preliminaries： Effectiveness of Consistency models

- The approximated learning objective for Consistency Models:

$$\mathcal{L}_{Approx}(\boldsymbol{\theta}) = \mathbb{E}[|\boldsymbol{x}_t - \frac{\alpha(t)}{\alpha(t-1)} x_{t-1} + \frac{\alpha(t)}{\alpha(t-1)} \sigma(t-1) \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}, t-1) - \sigma(t) \boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)\ |]$$

- The error estimation with the accumulative approximation $\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}, t-1) \approx \boldsymbol{x}_0$

$$|\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) - \boldsymbol{x}_0| \le \sum_{t' \in [1,2,\dots,t]} |\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_{t'}, t') - \boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_{t'-1}, t'-1)|$$

Motivation ： Optimization on the upper bound

$$|\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) - \boldsymbol{x}_0| \leq \sum_{t' \in [1,2,\ldots,t]} |\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_{t'}, t') - \boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_{t'-1}, t'-1)|$$

- Converge slowly when T is large

Motivation：Optimization on the upper bound

$$|\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) - \boldsymbol{x}_0| \le \sum_{t' \in [1,2,\dots,t]} |\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_{t'}, t') - \boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_{t'-1}, t'-1)|$$

↓

$$|\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) - \boldsymbol{x}_0| \le \sum_{t' \in [k,2k,\dots,t]} |\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_{t'}, t') - \boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_{t'-k}, t'-k)|$$

Motivation：Optimization on the upper bound

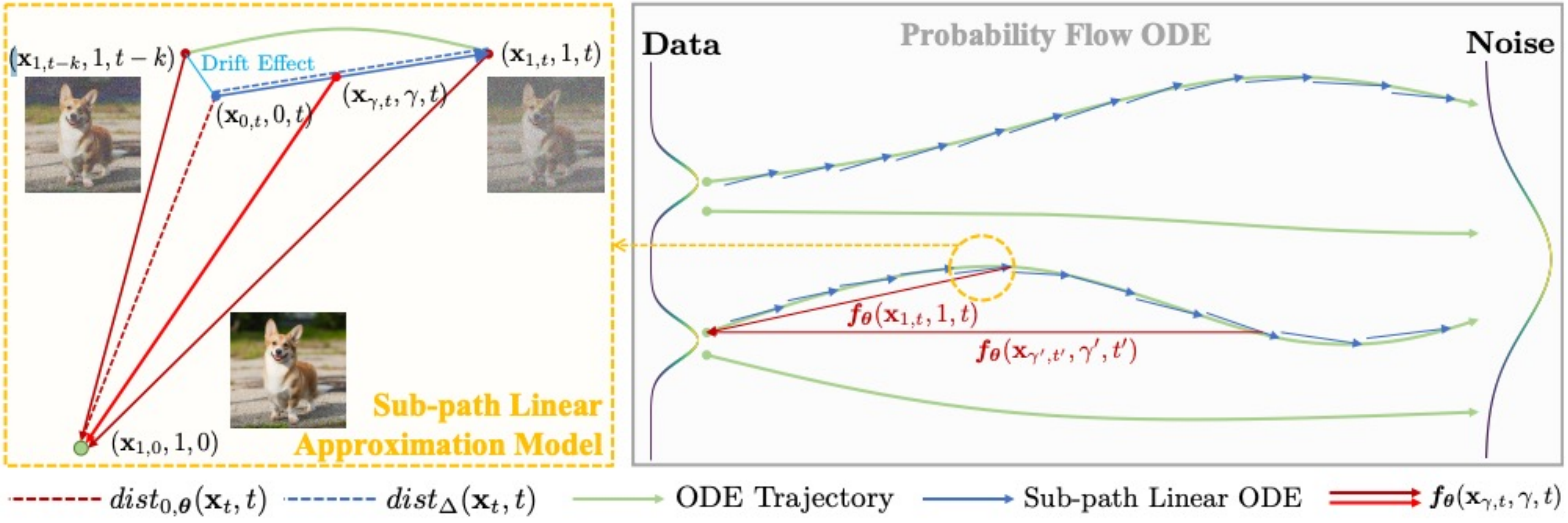
- The changed learning objective

$$\mathcal{L}_{Approx}(\boldsymbol{\theta}) = \mathbb{E}[|\underbrace{\boldsymbol{x}_t - \frac{\alpha(t)}{\alpha(t-k)}x_{t-1}}_{dist_{\Delta}} + \underbrace{\frac{\alpha(t)}{\alpha(t-k)}\sigma(t-k)\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-k}, t-k)}_{dist_{0,\boldsymbol{\theta}}} - \sigma(t)\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)\ |]$$

to be learned

Our SPLAM

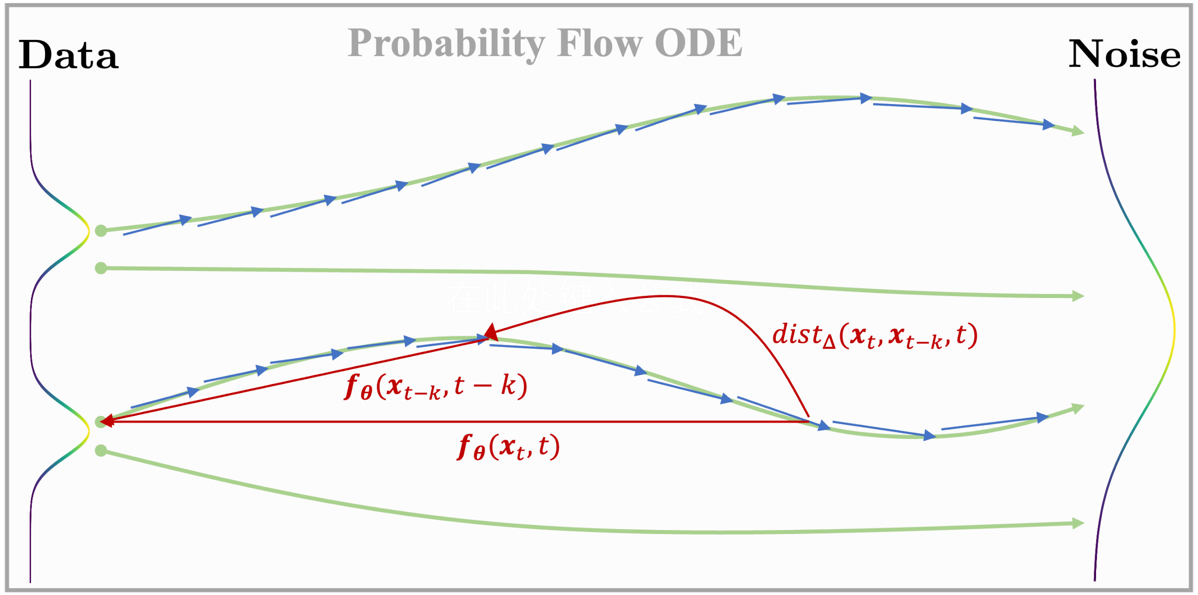
- Treat the Sampled PF-ODE trajectory as a series of connected sub-paths
- Build a better estimation for each sub-path from its start point $\boldsymbol{x}_t$ to the end $\boldsymbol{x}_{t-k}$ especially for optimizing $dist_{\Delta}$



Methods

- PF-ODE trajectory as a series of connected sub-paths
  - error estimation for two sample points can be defined as

$$\mathcal{L}_{Sub-p}(\boldsymbol{\theta}, k) = \mathbb{E}[|dist_{\Delta}(\boldsymbol{x}_t, \boldsymbol{x}_{t-k}, t) + dist_{0,\boldsymbol{\theta}}(\boldsymbol{x}_{t-k}, t-k, t) - \sigma(t)\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)|]$$

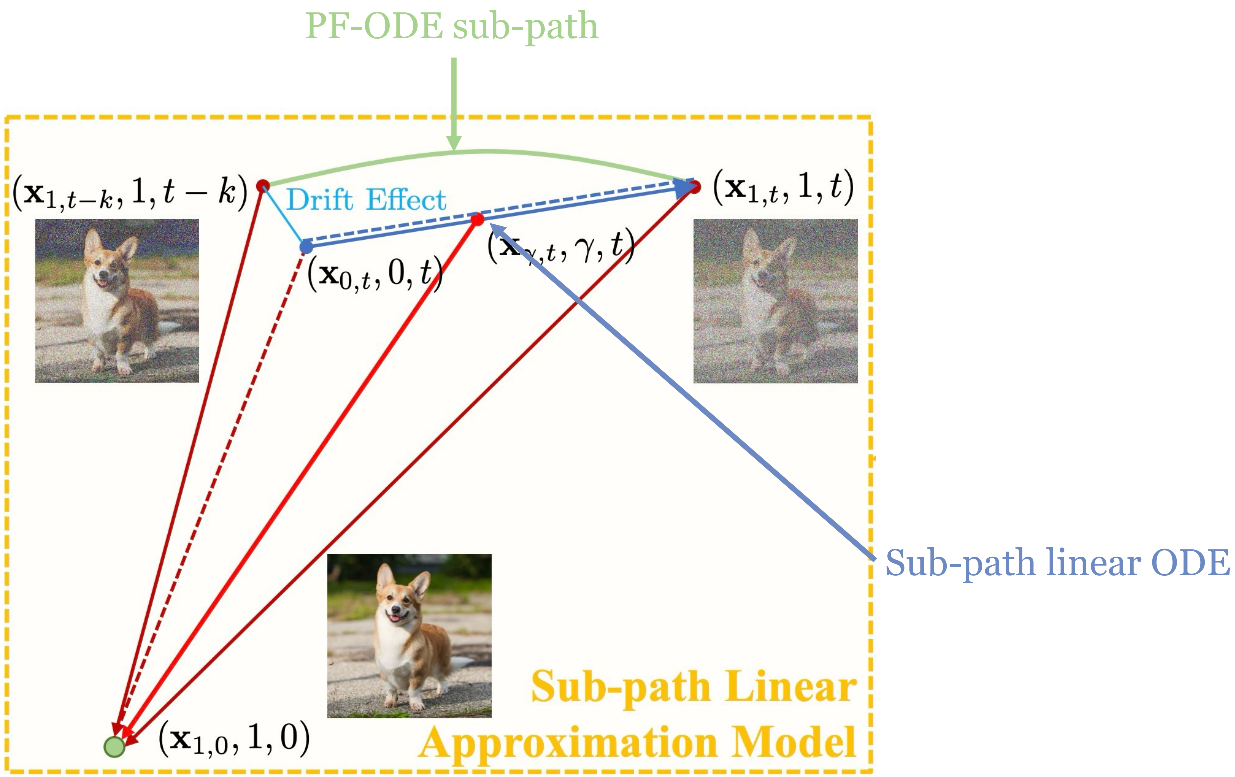


Methods

- We introduce Sub-path Linear ODE $\{x_{\gamma,t}\}_{\gamma\in[0,1]}$

$$\boldsymbol{x}_{\gamma,t} = \frac{\alpha(t)}{\alpha(t-k)}\boldsymbol{x}_{t-k} + \gamma * dist_{\Delta}(\boldsymbol{x}_t, \boldsymbol{x}_{t-k}, t)$$
$$= (1-\gamma)\frac{\alpha(t)}{\alpha(t-k)}\boldsymbol{x}_{t-k} + \gamma\boldsymbol{x}_t$$

- from a sub-path start $\boldsymbol{x}_t$ to a drifted end $\frac{\alpha(t)}{\alpha(t-k)}\boldsymbol{x}_{t-k}$

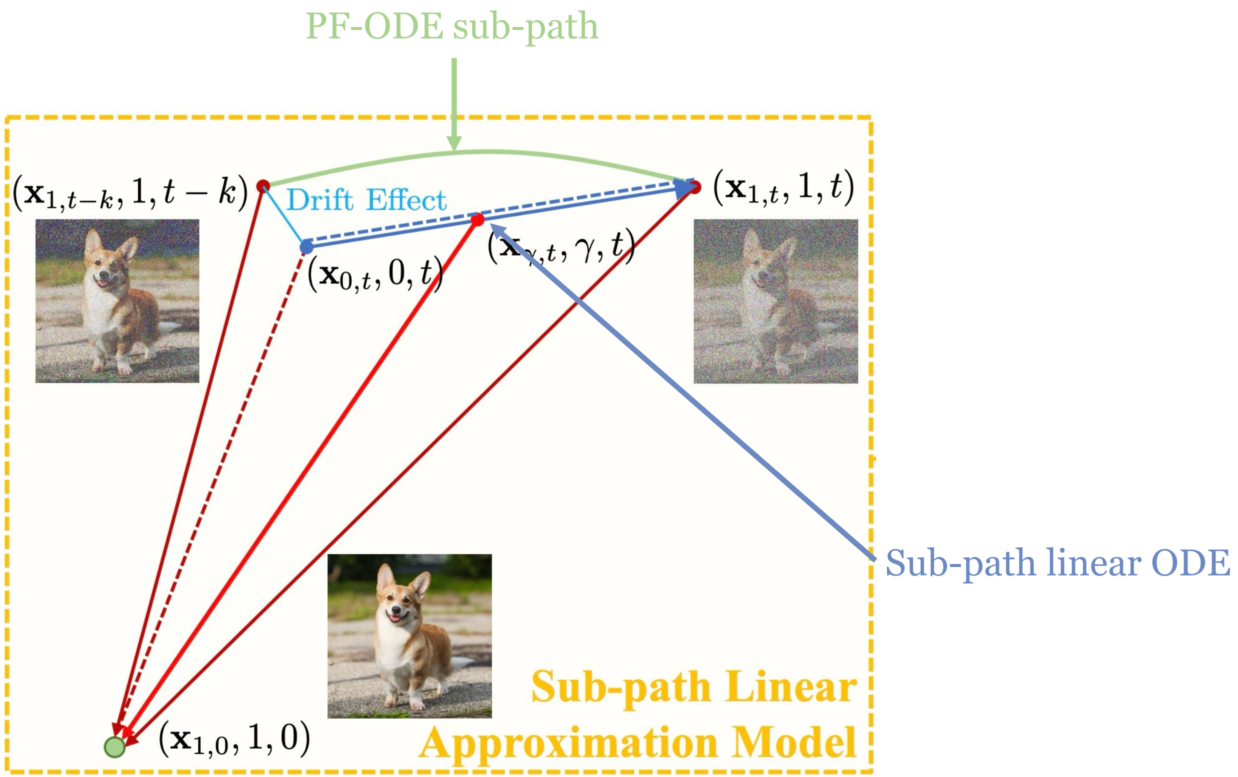


$$x_{\gamma,t} = (1-\gamma)\frac{\alpha(t)}{\alpha(t-k)}x_{t-k} + \gamma x_t$$

Methods

- Important property for our SL-ODE $\{x_{\gamma,t}\}$

$$d\boldsymbol{x}_{\gamma,t} = \gamma * dist_{\Delta}(\boldsymbol{x}_t, \boldsymbol{x}_{t-k}, t)d\gamma$$



$$x_{\gamma,t} = (1-\gamma)\frac{\alpha(t)}{\alpha(t-k)}x_{t-k} + \gamma x_t$$

Methods

- Important property for our SL-ODE $\{x_{\gamma,t}\}$

$$d\boldsymbol{x}_{\gamma,t} = \gamma * dist_{\Delta}(\boldsymbol{x}_t, \boldsymbol{x}_{t-k}, t)d\gamma$$

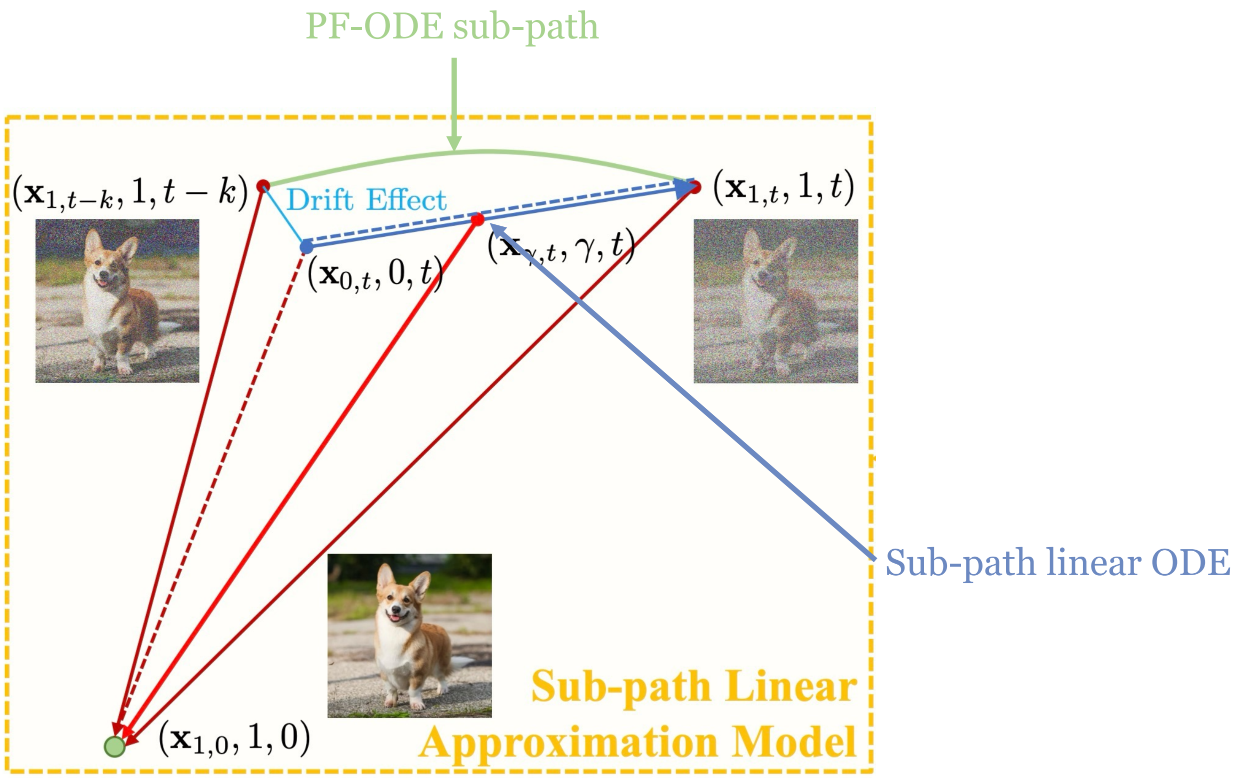
- The error estimation when $x_{\gamma,t}$ is involved:

$$\boldsymbol{f_\theta}(\boldsymbol{x}_{\gamma,t}, \gamma, t) = \frac{\boldsymbol{x}_{\gamma,t} - \sigma(\gamma, t)\boldsymbol{\epsilon_\theta}(\boldsymbol{x}_{\gamma,t}, \gamma, t)}{\alpha(t)}$$

- Using the approximation strategy in CM:

$$\mathcal{L}_{SL-ODE} = |\boldsymbol{f_\theta}(\boldsymbol{x}_{\gamma,t}, \gamma, t) - \boldsymbol{f_\theta}(\boldsymbol{x}_{1,t-k}, 1, t-k)|$$

$$and\ \boldsymbol{x}_{1,t-k} \equiv x_{t-k}$$



$$x_{\gamma,t} = (1-\gamma)\frac{\alpha(t)}{\alpha(t-k)}x_{t-k} + \gamma x_t$$

Methods

- Important property for our SL-ODE $\{x_{\gamma,t}\}$

$$\mathcal{L}_{SL-ODE} = |\boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_{\gamma,t}, \gamma, t) - \boldsymbol{f}_{\boldsymbol{\theta}}(\boldsymbol{x}_{1,t-k}, 1, t-k)|$$

- Final learning objective for SPLAM:

$$\mathcal{L}_{\text{SPLAM}}(\boldsymbol{\theta}, k) = \mathbb{E}|\gamma * dist_{\Delta}(\boldsymbol{x}_t, \boldsymbol{x}_{t-k}, t) + dist_{0,\boldsymbol{\theta}}\,(\boldsymbol{x}_{t-k}, t-k, t) - \sigma(\gamma, t)\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_{\gamma,t}, \gamma, t)\ |]$$

- The original learning objective

$$\mathcal{L}_{Sub-p}(\boldsymbol{\theta}, k) = \mathbb{E}[|dist_{\Delta}(\boldsymbol{x}_t, \boldsymbol{x}_{t-k}, t) + dist_{0,\boldsymbol{\theta}}\,(\boldsymbol{x}_{t-k}, t-k, t) - \sigma(t)\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)\ |]$$

Methods : Distillation from Latent Diffusion Models

Algorithm 1 Sub-Path Linear Approximation Distillation (SPLAD)

Input: dataset $\mathcal{D}$, initial model parameter $\boldsymbol{\theta}$, learning rate η, EMA decay rate μ, ODE solver $\Phi(\cdot,\cdot;\phi)$, distance estimation $|\cdot|$, a fixed guidance scale w, step size k, VAE encoder $\mathcal{E}(\cdot)$, noise schedule $\alpha(t),\sigma(t)$

$\boldsymbol{\theta}^- \leftarrow \boldsymbol{\theta}$

repeat

  sample $(x,c)\sim\mathcal{D}, t\sim\mathcal{U}[k,T]$ and $\gamma\sim\mathcal{U}[0,1]$ ← Unform sampling for γ

  convert x into latent space: $z=\mathcal{E}(x)$

  sample $\boldsymbol{z}_t\sim\mathcal{N}(\alpha(t)z,\sigma(z)^2I)$

  $\hat{\boldsymbol{z}}^{\Phi}_{t_{\phi,0}} \leftarrow \boldsymbol{z}_t$, $i\leftarrow 0$ ← Multple Estimation for ODE solvers

  repeat

    $\hat{\boldsymbol{z}}^{\Phi}_{t_{\phi,i+1}} \leftarrow \hat{\boldsymbol{z}}^{\Phi}_{t_{\phi,i}} + w\Phi(\hat{\boldsymbol{z}}^{\Phi}_{t_{\phi,i}}, t_{\phi,i}, t_{\phi,i+1}, c;\phi) + (1-w)\Phi(\hat{\boldsymbol{z}}^{\Phi}_{t_{\phi,i}}, t_{\phi,i}, t_{\phi,i+1}, \emptyset;\phi)$

    $i\leftarrow i+1$

  until $k = i * k_\phi$

  $\boldsymbol{z}_{\gamma,t} \leftarrow (1-\gamma)*\frac{\alpha(t)}{\alpha(t-k)}\hat{\boldsymbol{z}}^{\Phi}_{i-k} + \gamma * \boldsymbol{z}_t$ ▷ Sample a point on the SL-ODE.

  $\mathcal{L}(\boldsymbol{\theta},\boldsymbol{\theta}^-;\phi) \leftarrow |(\boldsymbol{F}_{\boldsymbol{\theta}}(\boldsymbol{z}_{\gamma,t},c,\gamma,t) - \boldsymbol{F}_{\boldsymbol{\theta}^-}(\hat{\boldsymbol{z}}^{\Phi}_{1,t-k},c,1,t-k))|$

  $\boldsymbol{\theta}\leftarrow\boldsymbol{\theta}-\eta\nabla_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta},\boldsymbol{\theta}^-;\phi)$

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

until convergence

Experiment Results: better FIDs with faster convergence

Table 2: Quantitative results for SDv1.5. Baseline numbers are cited from [50] and [49]. All the results of LCM are our reproduction whose performance is aligned as stated in the paper. † Results are evaluated by us using the released models.

(a) Results on MSCOCO2014-30k, $w = 3$.

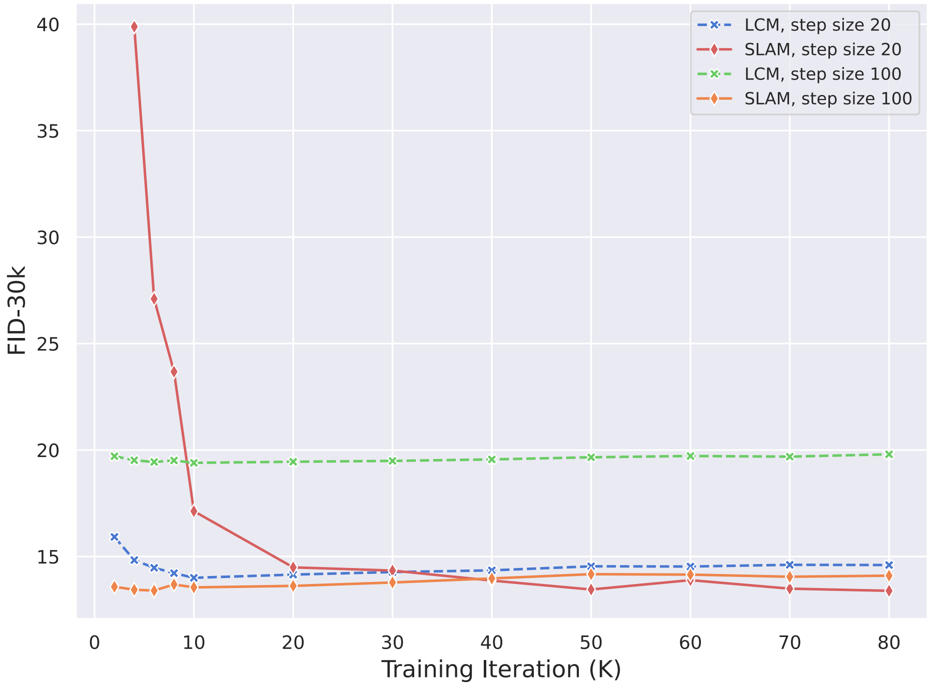
Family	Methods	Latency(↓)	FID(↓)
Unaccelerated	DALL-E [33]	-	27.5
	DALL-E2 [32]	-	10.39
	Parti-750M [51]	-	10.71
	Parti-3B [51]	6.4s	8.10
	Parti-20B [51]	-	7.23
	Make-A-Scene [5]	25.0s	11.84
	Muse-3B [4]	1.3	7.88
	GLIDE [29]	15.0s	12.24
	LDM [34]	3.7s	12.63
	Imagen [35]	9.1s	7.27
	eDiff-I [1]	32.0s	6.95
GANs	LAFITE [54]	0.02s	26.94
	StyleGAN-T [38]	0.10s	13.90
	GigaGAN [13]	0.13s	9.09
Accelerated Diffusion	DPM++ (4step) [24]	0.26s	22.36
	UniPC (4step) [52]	0.26s	19.57
	LCM-LoRA (4step) [27]	0.19s	23.62
	InstaFlow-0.9B [21]	0.09s	13.10
	InstaFlow-1.7B [21]	0.12s	11.83
	UFOGen [49]	0.09s	12.78
	DMD [50]	0.09s	11.49
	LCM (2step) [26]	0.12s	14.29
	SPLAM (2step)	0.12s	12.31
	LCM (4step) [26]	0.19s	10.68
	SPLAM (4step)	0.19s	**10.06**
Teacher	SDv1.5 [34]†	2.59s	8.03

(b) Results on MSCOCO2017-5k, $w = 3$.

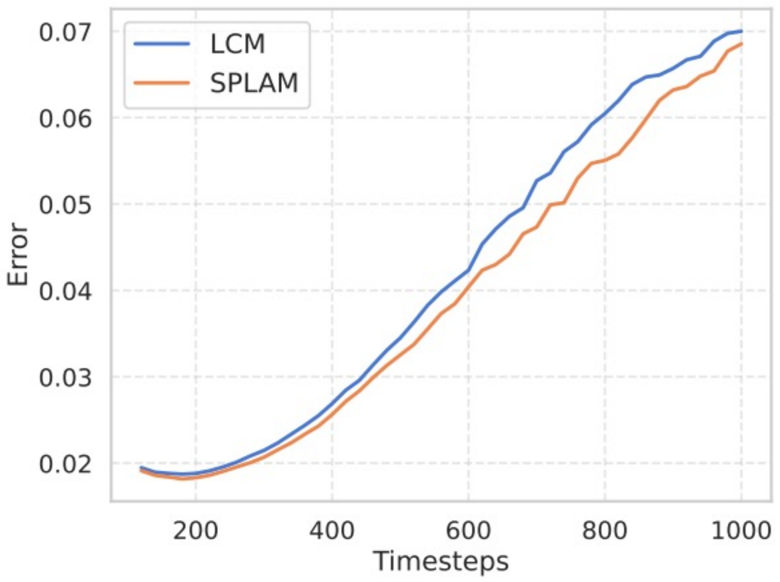
Methods	#Step	Latency(↓)	FID(↓)
DPM Solver++ [24]†	4	0.21s	35.0
	8	0.34s	21.0
Progressive Distillation [36]	1	0.09s	37.2
	2	0.13s	26.0
	4	0.21s	26.4
CFG-Aware Distillation [16]	8	0.34s	24.2
InstaFlow-0.9B [21]	1	0.09s	23.4
InstaFlow-1.7B [21]	1	0.12s	22.4
UFOGen [49]	1	0.09s	22.5
LCM [26]	2	0.12s	25.22
	4	0.19s	21.41
SPLAM	2	0.12s	23.07
	4	0.19s	**20.77**

(c) Results on MSCOCO2014-30k, $w = 8$.

Family	Methods	Latency(↓)	FID(↓)
Accelerated Diffusion	DPM++ (4step)	0.26s	22.44
	UniPC (4step) [52]	0.26s	23.30
	LCM-LoRA (4step) [27]	0.19s	23.62
	DMD [50]	0.09s	14.93
	LCM (2step) [26] [26]	0.12s	15.56
	SPLAM (2step)	0.12s	14.50
	LCM (4step) [26] [26]	0.19s	14.53
	SPLAM (4step)	0.19s	**13.39**
Teacher	SDv1.5 [34]†	2.59s	13.05

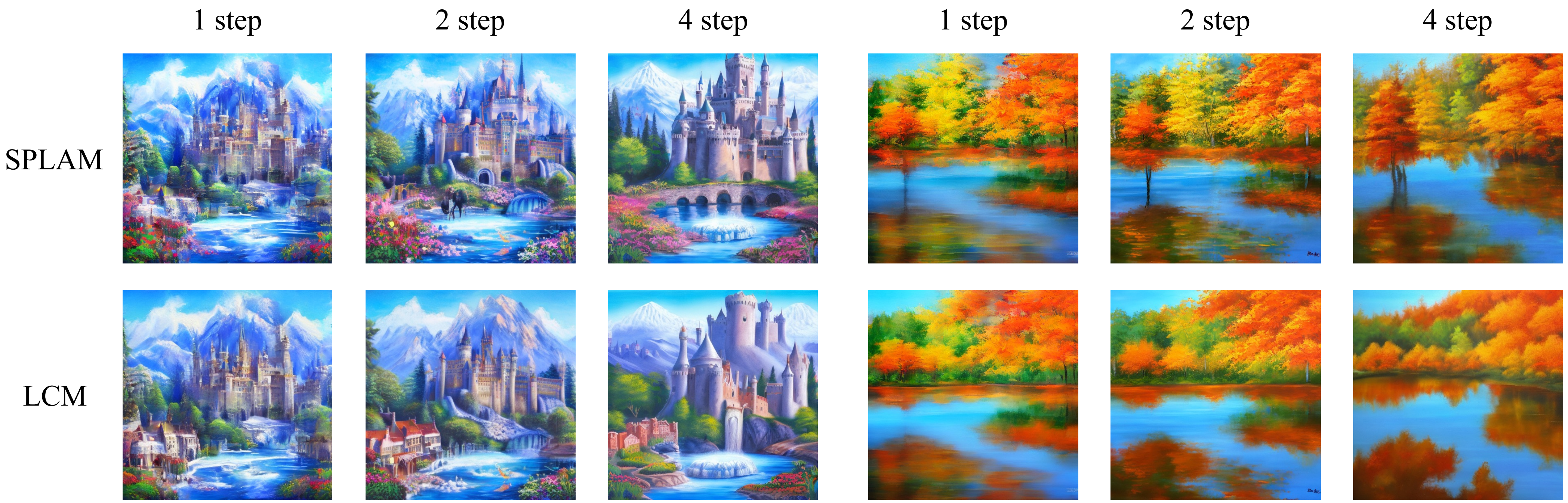


Experiment Results: Directly estimate the errrors



$$|x_0 - f_\theta(x_t, t)| \leq \sum_{t' \in [k, 2k, \ldots t]} |f_\theta(x_{t'}, t') - f_\theta(x_{t'-k}, t' - k)|$$

Experiment Results: Generation of images



Experiment Results: Generation of images

Thanks

媒体计算研究组
Multimedia Computing Group

https://hypnosxc.github.io
chenxu24568@gmail.com

https://mcg.nju.edu.cn/index.html
https://github.com/MCG-NJU

Paper, code, and models are available