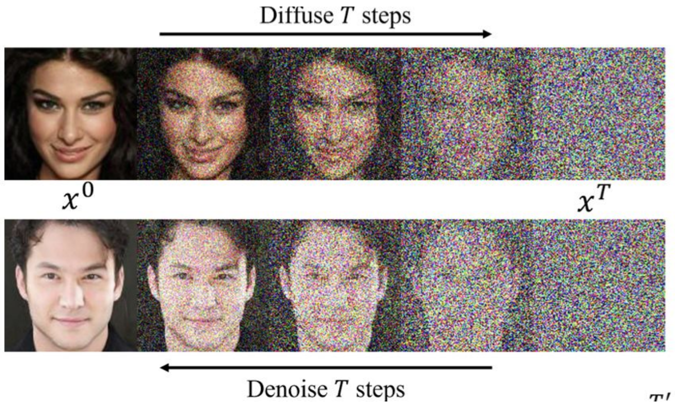
(Consistency Models)

Fast sampling

Elnikov Vladislav

Problem

Sampling process is slow because it requires hundreds to thousands of network evaluations to emulate a continuous process defined by differential equations.



Напомним наши результаты

ImageNet 64 × 64				
PD† (Salimans & Ho, 2022)	1	15.39	0.59	0.62
DFNO† (Zheng et al., 2022)	1	8.35		
CD†	1	6.20	0.68	0.63
PD† (Salimans & Ho, 2022)	2	8.95	0.63	**0.65**
CD†	2	**4.70**	**0.69**	0.64
ADM (Dhariwal & Nichol, 2021)	250	**2.07**	0.74	0.63
EDM (Karras et al., 2022)	79	2.44	0.71	**0.67**
BigGAN-deep (Brock et al., 2019)	1	4.06	**0.79**	0.48
CT	1	13.0	0.71	0.47
CT	2	11.1	0.69	0.56

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**

Early Stop of the Diffusion Process

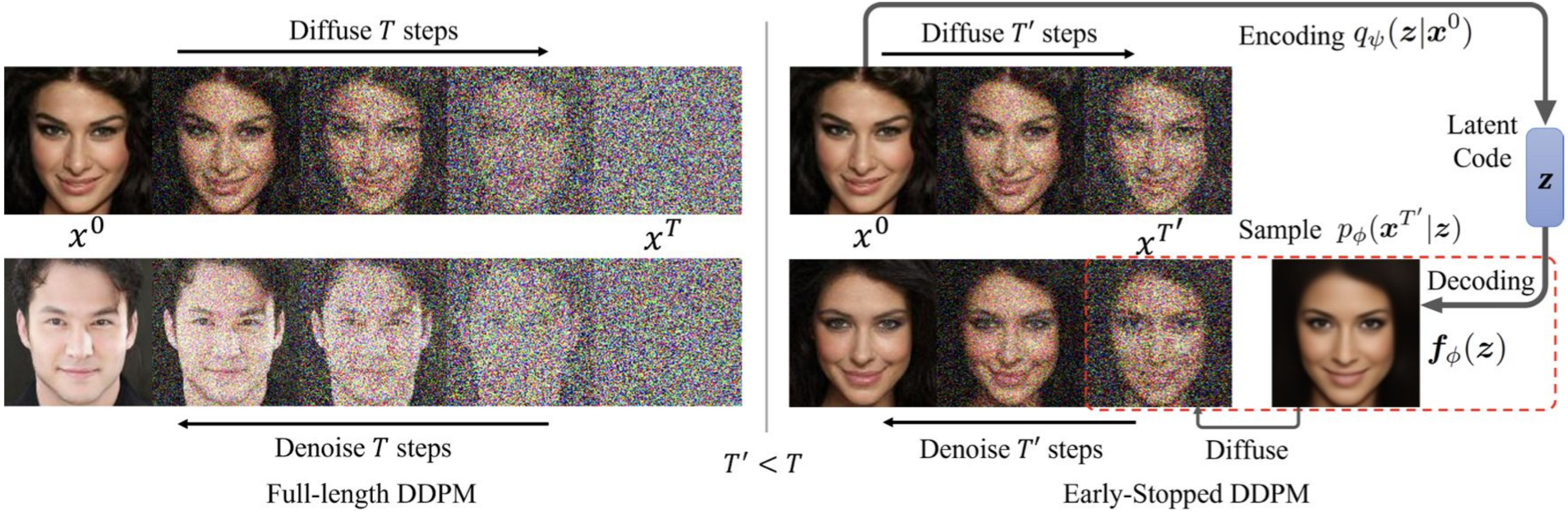


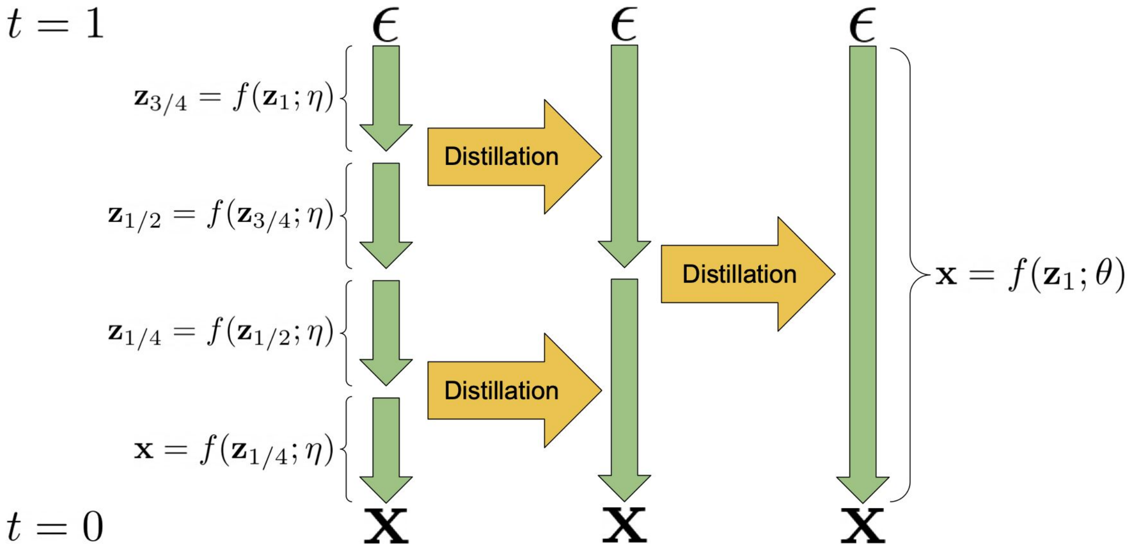
Figure 1: Compare Early-Stopped DDPM (ES-DDPM) with the full-length DDPM.

CelebA-64	Method	FID
Hybrid Models	StyleGAN2+100-step ES-DDPM (**Ours**)	3.01
	StyleGAN2+200-step ES-DDPM (**Ours**)	**2.55**
	1000-step DiffuseVAE (Pandey et al., 2022)	4.76
Score-based Methods	1000-step DDPM* (Ho et al., 2020)	3.26
	250-step PNDM (Liu et al., 2022)	2.71
	NCSN (Song & Ermon, 2019)	25.30
	NCSNv2 (Song & Ermon, 2020)	10.23
GAN-based Methods	COCO-GAN (Lin et al., 2019)	4.00
	StyleGAN2* (Karras et al., 2020)	4.55
	QA-GAN (Parimala & Channappayya, 2019)	6.42
VAE-based Methods	NCP-VAE (Aneja et al., 2020)	5.25

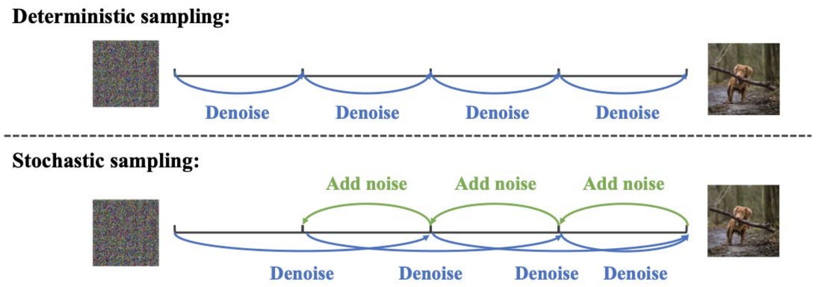
CelebA-128	Method	FID
Hybrid Models	StyleGAN2+100-step ES-DDPM (**Ours**)	**1.76**
	StyleGAN2+200-step ES-DDPM (**Ours**)	1.79
Score-based Methods	1000-step DDPM* (Ho et al., 2020)	5.65
GAN-based Methods	COCO-GAN (Lin et al., 2019)	5.74
	StyleGAN2* (Karras et al., 2020)	2.13
	PresGAN (Dieng et al., 2019)	29.12

Denoising Steps T'	0	100	200	300	400	500	600	700	800	900	1000
VAE+ES-DDPM	158.61	49.62	20.88	11.03	5.96	3.69	3.17	3.15	**3.12**	3.18	
DCGAN+ES-DDPM	33.31	15.54	9.97	7.21	5.49	3.98	3.37	**3.13**	3.14	3.24	3.20
StyleGAN2+ES-DDPM	7.18	5.52	5.02	4.60	4.03	3.51	3.26	**3.11**	3.16	3.17	

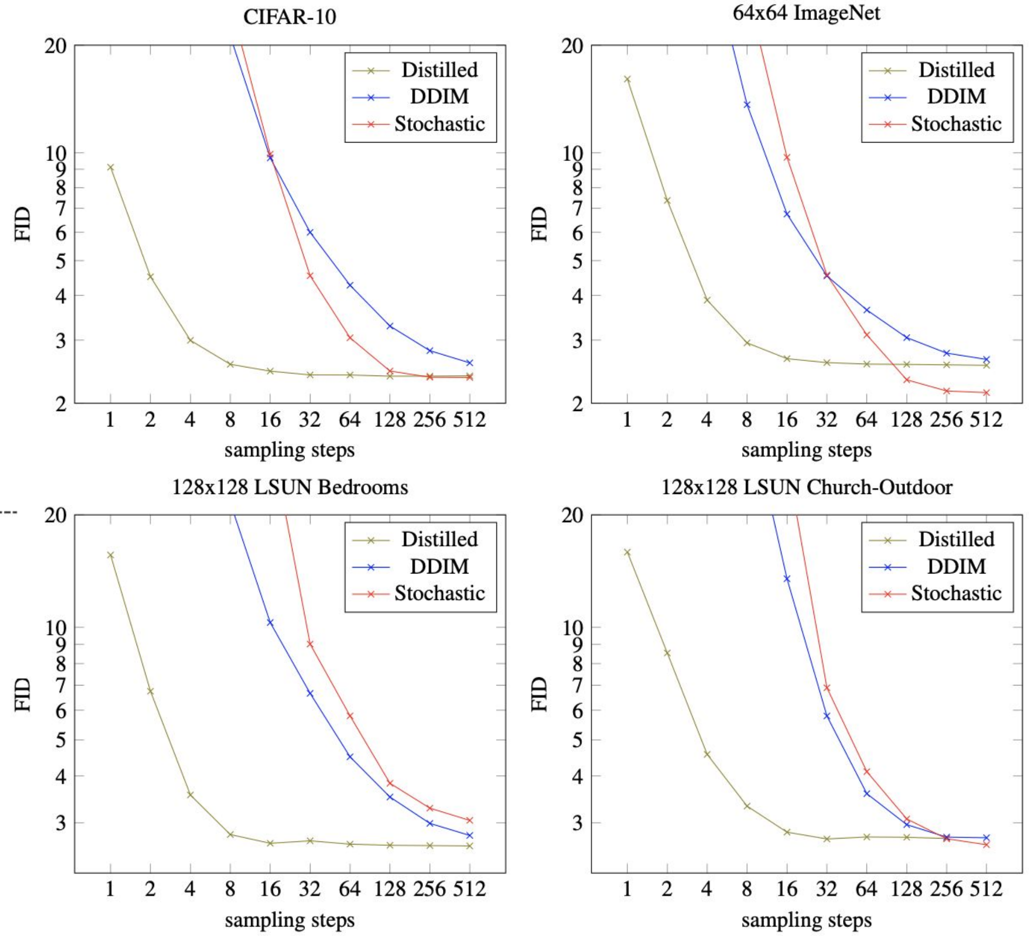
Дистилляция



Результаты

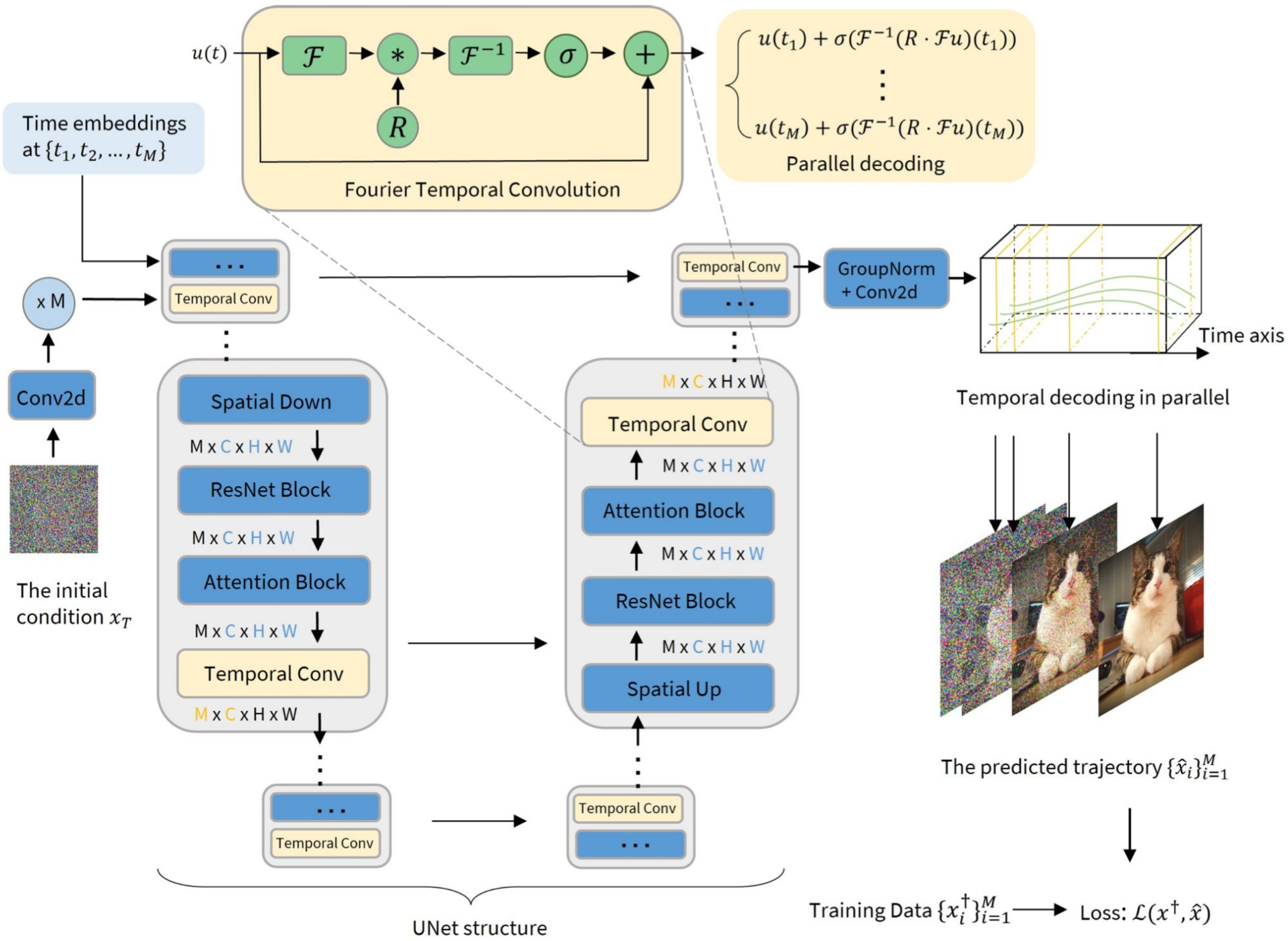


Deterministic sampling:
Denoise
Denoise
Denoise
Denoise
Stochastic sampling:
Add noise
Add noise
Add noise
Denoise
Denoise
Denoise
Denoise



CIFAR-10
64x64 ImageNet
128x128 LSUN Bedrooms
128x128 LSUN Church-Outdoor
FID
sampling steps
Distilled
DDIM
Stochastic
1 2 4 8 16 32 64 128 256 512
2 3 4 5 6 7 8 9 10 20

Diffusion Models via Operator Learning



Как выглядит процесс

$$(\mathcal{T}u)(t) = u(t) + \sigma\left(\left(\mathcal{K}u\right)(t)\right)$$

$$(\mathcal{K}u)(t) = \int_D (\mathcal{F}^{-1}R)(\tau)u(t-\tau)\mathrm{d}\tau, \forall t \in D.$$

$$R \cdot (\mathcal{F}u)_{j,k} = \sum_{l=1}^{d} R_{j,k,l}(\mathcal{F}u)_{j,l},$$

Результат

Method	NFE	FID
Ours	1	**3.78**
Knowledge distillation (Luhman & Luhman, 2021)	1	9.36
Progressive distillation (Salimans & Ho, 2021)	1	9.12
	2	4.51
	4	3.00

CIFAR-10

Method	Model evaluations	FID score
Ours	1	**7.83**
Progressive distillation	1	15.99
	2	7.11
	4	3.84

ImageNet-64

Проблема данных работ по моему мнению

Многие результаты тестируются только на CIFAR-10 и ImageNet-64. Много выводов делается на основе результатов в CIFAR-10, что не может не расстраивать.

Проблема данных работ по моему мнению

Нет понимания сколько точно работает каждая из моделей, насколько энергозатратны. Нет понимания, что будет лучше сходиться и выдавать лучшую картинку при большем количестве эволюций картинки (NFE - model evolutions).

Вывод

- Данная тема является очень новой, многие статьи совсем недавние: Diffusion Models via Operator Learning (2023), Consistency Models(2023), PROGRESSIVE DISTILLATION FOR FAST SAMPLING OF DIFFUSION MODELS (2022)
- Нужно более честные сравнения и на большем числе датасетов

\+ Данная тема очень перспективна и востребована