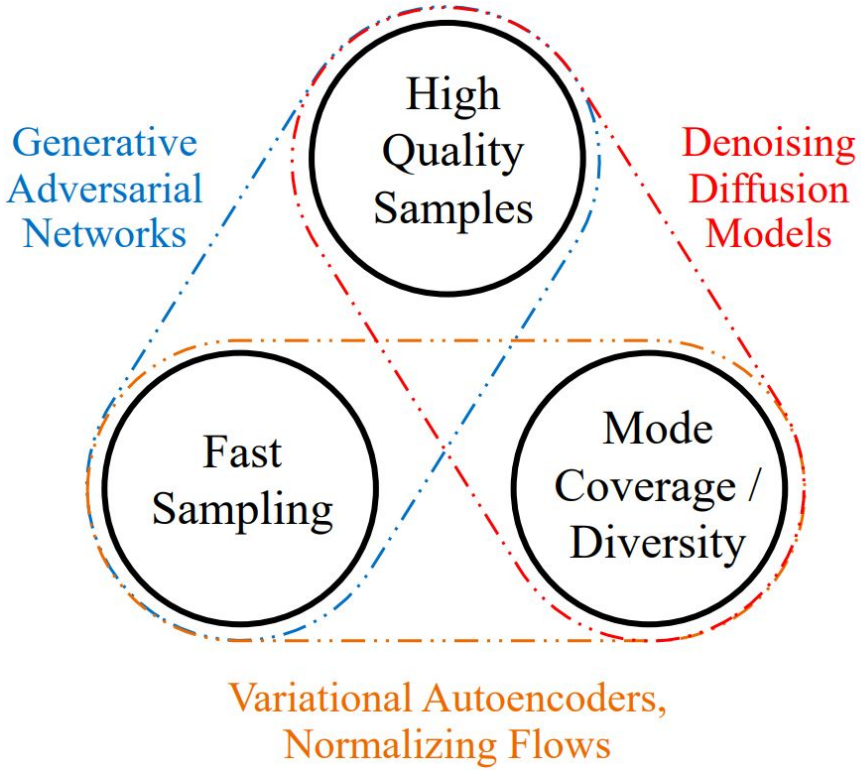
Tackling the Generative Learning Trilemma with Denoising Diffusion GANs

Meshchaninov Viacheslav

Centre of Deep Learning and Bayesian Methods
HSE University

Generative learning trilemma



Reverse process approximation

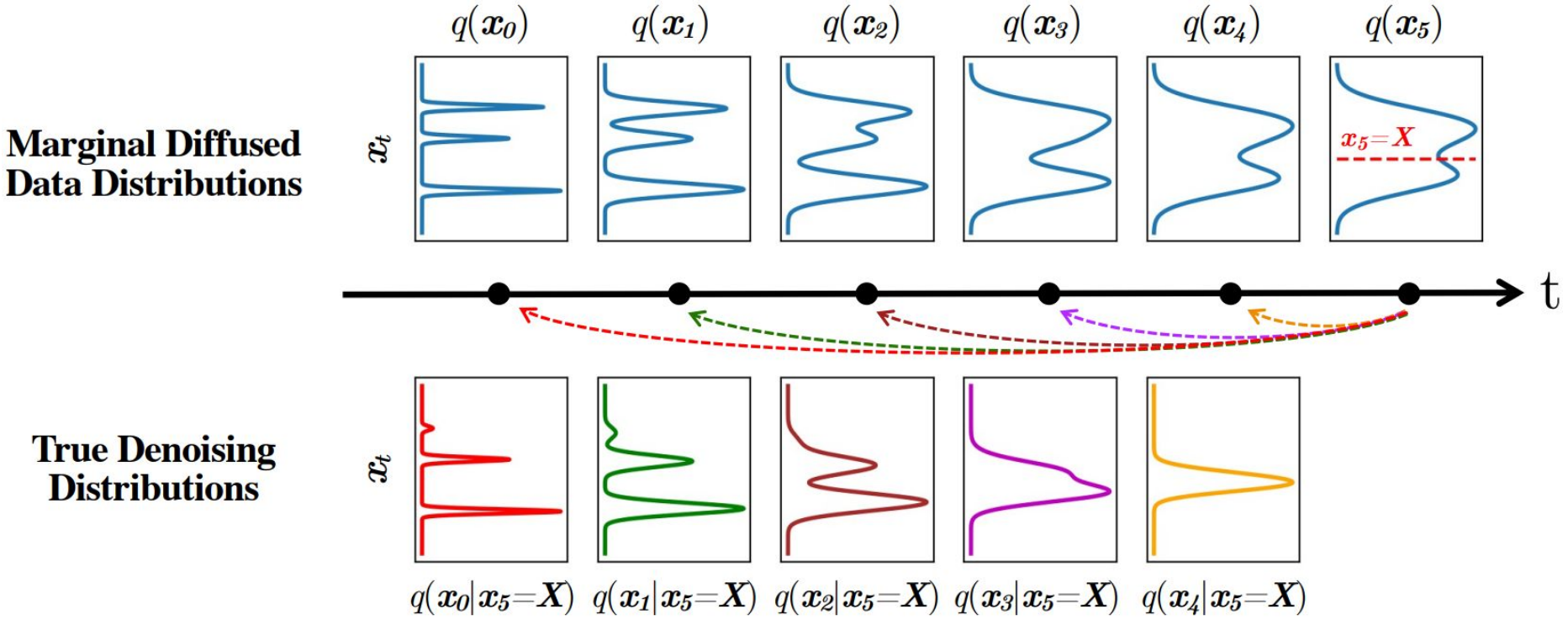
$$q(x_{t-1}|x_t) \propto q(x_t|x_{t-1}) * q(x_{t-1})$$

$$q(x_t|x_{t-1}) \sim \mathcal{N}(\sqrt{1-\beta_t}x_{t-1}, \beta_t I)$$

$q(x_{t-1}|x_t)$ — Gaussian in two situations:

- If β_t is infinitesimal, then the product is dominated by $q(x_t|x_{t-1})$
- If data marginal is Gaussian

Non-Gaussian denoising distribution



Parametrizing the implicit denoising model

DDPM: $p_\theta(x_{t-1}|x_t) = E_{x_0 \sim p_\theta(x_0|x_t)} q(x_{t-1}|x_t, x_0)$

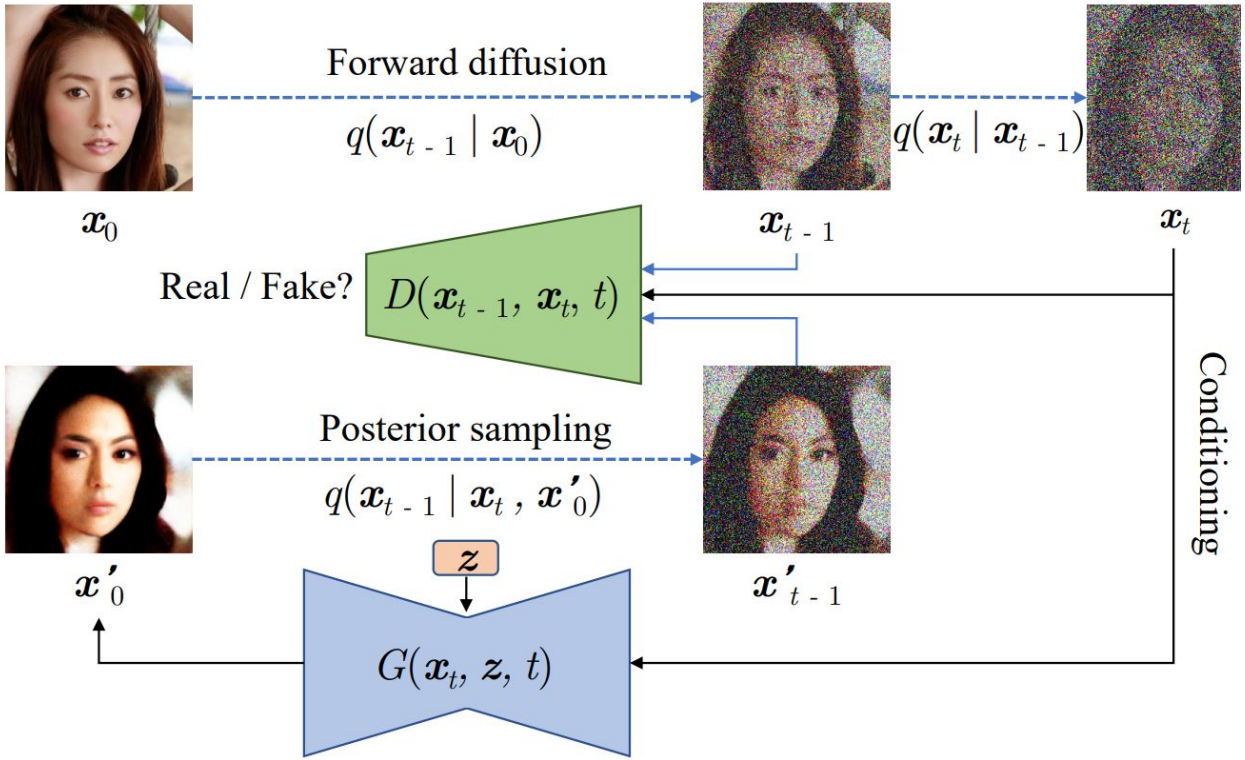
$p_\theta(x_0|x_t)$ — Gaussian

DDGAN: $p_\theta(x_{t-1}|x_t) = E_{x_0 \sim p_\theta(x_0|x_t)} q(x_t|x_{t-1}, x_0) =$

$= E_{z \sim \mathcal{N}(0,1),\, x_0 = GAN(x_t, t, z)} q(x_t|x_{t-1}, x_0)$

$p_\theta(x_0|x_t)$ — non-Gaussian

Training process



Objectives

Discriminator objective:

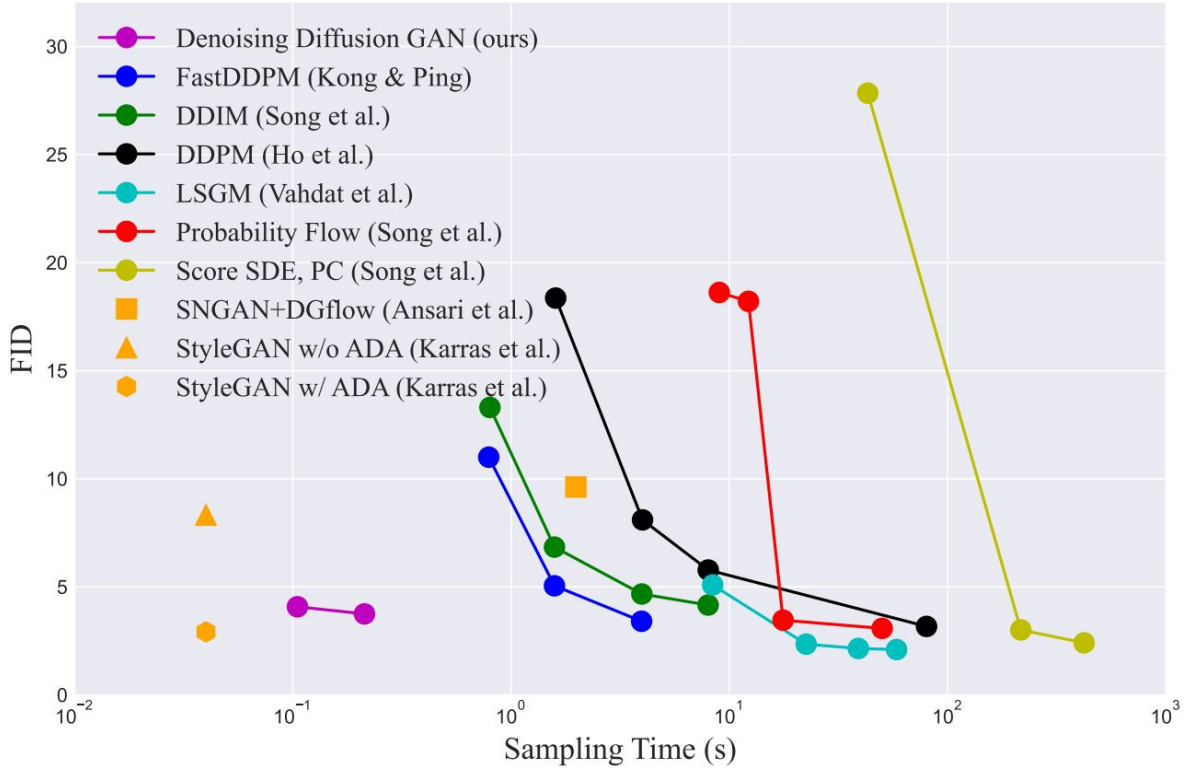
$$\min_{\phi} \sum_{t \geq 1} \mathbb{E}_{q(\mathbf{x}_t)} \left[\mathbb{E}_{q(\mathbf{x}_{t-1}|\mathbf{x}_t)} [-\log(D_\phi(\mathbf{x}_{t-1}, \mathbf{x}_t, t)] + \right.$$
$$\left. + \mathbb{E}_{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} [-\log(1 - D_\phi(\mathbf{x}_{t-1}, \mathbf{x}_t, t))] \right]$$

Generator objective:

$$\max_\theta \sum_{t \geq 1} \mathbb{E}_{q(\mathbf{x}_t)} \mathbb{E}_{p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)} [\log(D_\phi(\mathbf{x}_{t-1}, \mathbf{x}_t, t))]$$

Generative results on CelebA-HQ-256

Sample quality vs sampling time trade-off



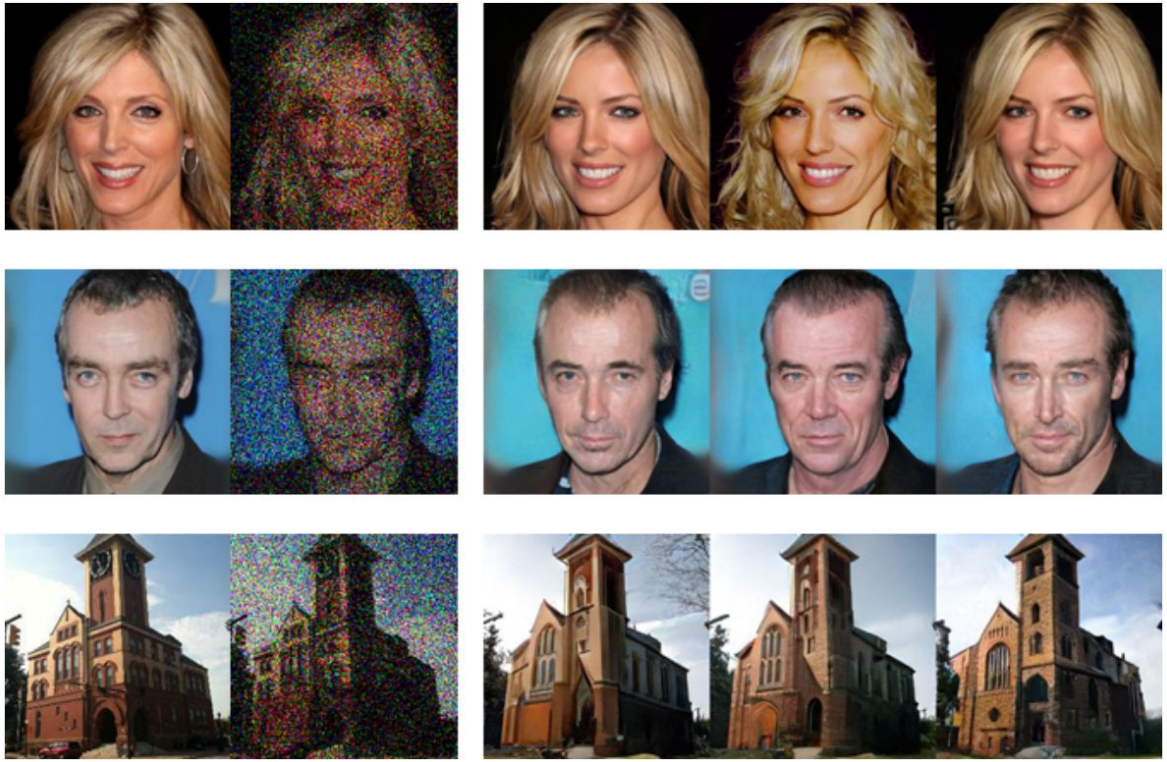
Results for unconditional generation on CIFAR-10

Model	IS↑	FID↓	Recall↑	NFE ↓	Time (s) ↓
Denoising Diffusion GAN (ours), T=4	9.63	3.75	0.57	4	0.21
DDPM (Ho et al., 2020)	9.46	3.21	0.57	1000	80.5
NCSN (Song & Ermon, 2019)	8.87	25.3	-	1000	107.9
Adversarial DSM (Jolicoeur-Martineau et al., 2021b)	-	6.10	-	1000	-
Likelihood SDE (Song et al., 2021b)	-	2.87	-	-	-
Score SDE (VE) (Song et al., 2021c)	9.89	2.20	0.59	2000	423.2
Score SDE (VP) (Song et al., 2021c)	9.68	2.41	0.59	2000	421.5
Probability Flow (VP) (Song et al., 2021c)	9.83	3.08	0.57	140	50.9
LSGM (Vahdat et al., 2021)	9.87	2.10	0.61	147	44.5
DDIM, T=50 (Song et al., 2021a)	8.78	4.67	0.53	50	4.01
FastDDPM, T=50 (Kong & Ping, 2021)	8.98	3.41	0.56	50	4.01
Recovery EBM (Gao et al., 2021)	8.30	9.58	-	180	-
Improved DDPM (Nichol & Dhariwal, 2021)	-	2.90	-	4000	-
VDM (Kingma et al., 2021)	-	4.00	-	1000	-
UDM (Kim et al., 2021)	10.1	2.33	-	2000	-
D3PMs (Austin et al., 2021)	8.56	7.34	-	1000	-
Gotta Go Fast (Jolicoeur-Martineau et al., 2021a)	-	2.44	-	180	-
DDPM Distillation (Luhman & Luhman, 2021)	8.36	9.36	0.51	1	-
SNGAN (Miyato et al., 2018)	8.22	21.7	0.44	1	-
SNGAN+DGflow (Ansari et al., 2021)	9.35	9.62	0.48	25	1.98
AutoGAN (Gong et al., 2019)	8.60	12.4	0.46	1	-
TransGAN (Jiang et al., 2021)	9.02	9.26	-	1	-
StyleGAN2 w/o ADA (Karras et al., 2020a)	9.18	8.32	0.41	1	0.04
StyleGAN2 w/ ADA (Karras et al., 2020a)	9.83	2.92	0.49	1	0.04
StyleGAN2 w/ Diffaug (Zhao et al., 2020)	9.40	5.79	0.42	1	0.04

Ablation studies on CIFAR-10

Model Variants	IS↑	FID↓	Recall↑
T = 1	8.93	14.6	0.19
T = 2	**9.80**	4.08	0.54
T = 4	9.63	**3.75**	**0.57**
T = 8	9.43	4.36	0.56
One-shot w/ aug	8.96	13.2	0.25
Direct denoising	9.10	6.03	0.53
Noise generation	8.79	8.04	0.52
No latent variable	8.37	20.6	0.42

Multi-modality of denoising distribution



Conclusion

Advantages:

1. GAN architecture allows denoising distribution to become multimodal and complex in contrast to DDPM
2. The diffusion process smoothens the data distribution making the discriminator less likely overfit