Diffusion Models Distillation

Alexander Nikolaev @alehandreus

SUMMER OF MACHINE LEARNING AT SKOLTECH

Problem

Diffusion models use complex multi-step process to generate samples. Can we use their knowledge to train one-step generators with the same quality?

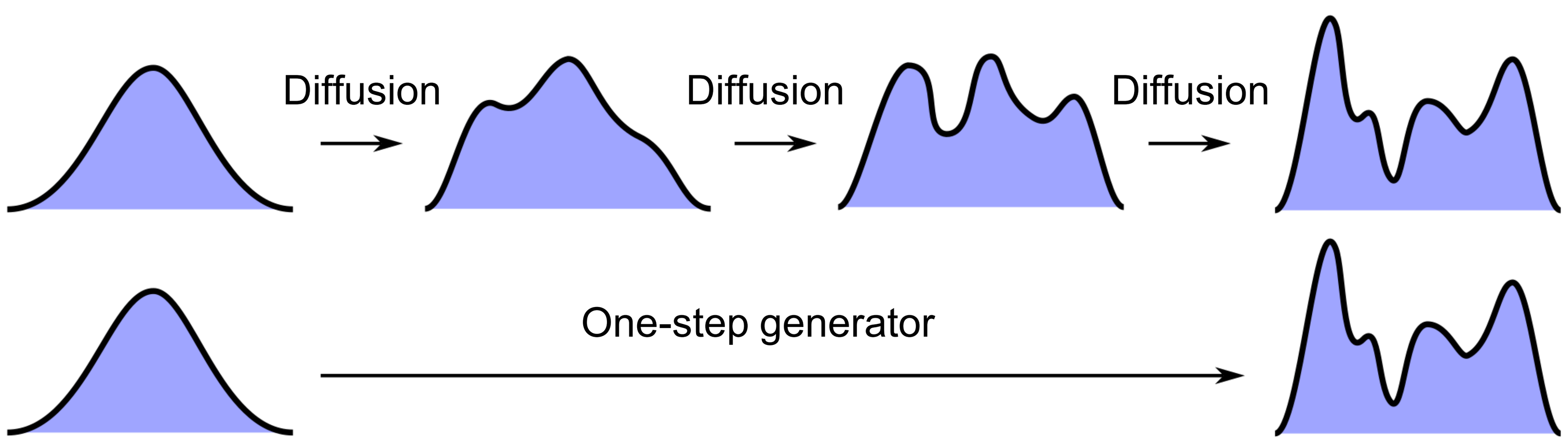


Figure 1. One-step vs multi-step generation

Recent approaches to this problem are:
- Probability Density Distillation
- Rectified Flows
- Consistency Models

Probability Density Distillation

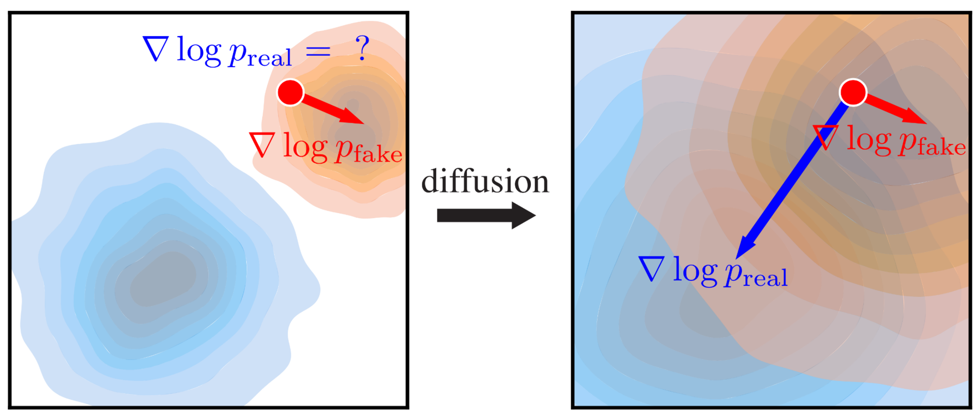
Samples from diffusion model have density p_{real} and objects from distilled generator have density p_{fake}.

Distribution Matching Distillation (DMD): minimize KL-divergence between p_{real} and p_{fake}:

$$D_{KL}(p_{fake} \| p_{real}) = \mathop{\mathbb{E}}_{\substack{z \sim N(0;I),\\ x = G_\theta(z)}} -(\log p_{real}(x) - \log p_{fake}(x))$$

Figure 2. Intuition behind DMD

(a) Samples from G_θ have low p_{real} causing instability
(b) After adding noise, distributions overlap



Problem: low p_{real} values cause exploding gradients.
Solution: make distributions smoother by adding Gaussian noise.

Gradient update: $\nabla_\theta D_{KL} \simeq \mathop{\mathbb{E}}_{z,t,x,x_t} (\alpha_t(s_{fake}(x_t,t) - s_{real}(x_t,t))\nabla_\theta G_\theta(z))$

$s_{real}(x_t,t)$ and $s_{fake}(x_t,t)$ are score functions of p_{real} and p_{fake}.
Additional diffusion model is trained to predict $s_{fake}(x_t,t)$.

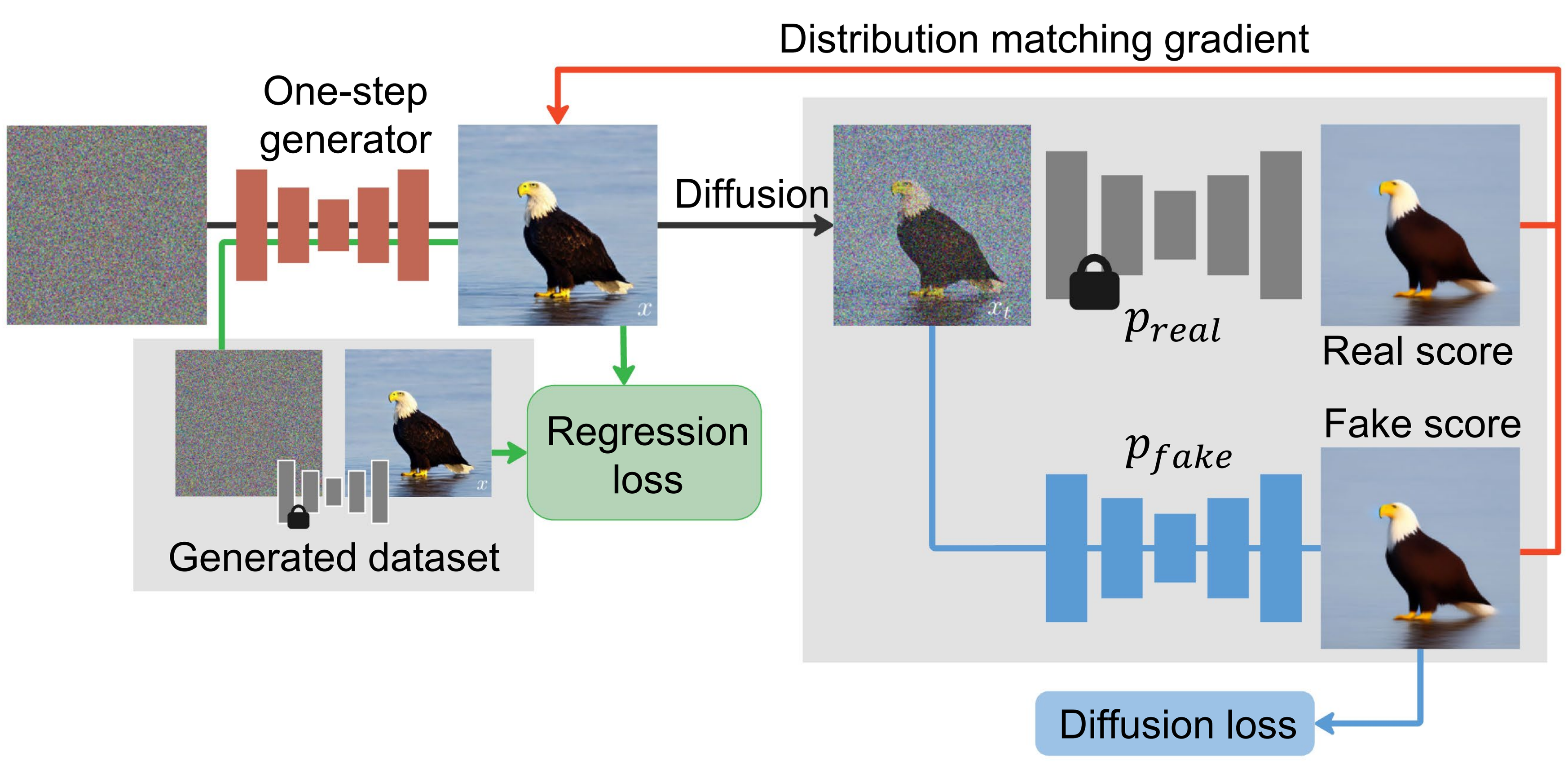


Figure 3. Distribution Matching Distillation

Rectified Flows

Trajectories between noise and image distributions are long and curved. Can we make them short and straight?

Transform noise Z_t to image Z_0:

$$dZ_t = v(Z_t,t)dt$$

Training objective:

$$\min_v \int_0^1 \underbrace{\mathbb{E}(|(X_1 - X_0) - v(X_t,t)|^2)}_{\text{Make predicted direction lie on a straight line}} dx$$

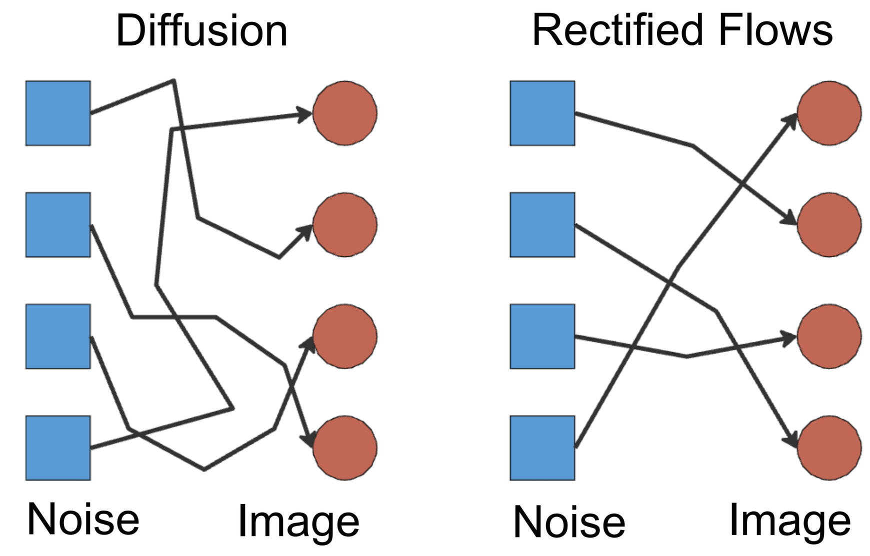


Figure 4. Rectified Flows trajectories

Consistency Models

Given a deterministic diffusion model, train $f_\theta(x_t,t)$ to predict x_0 from x_t in one step. Dataset of pairs (x_t,t) is sampled from pretrained deterministic diffusion model.

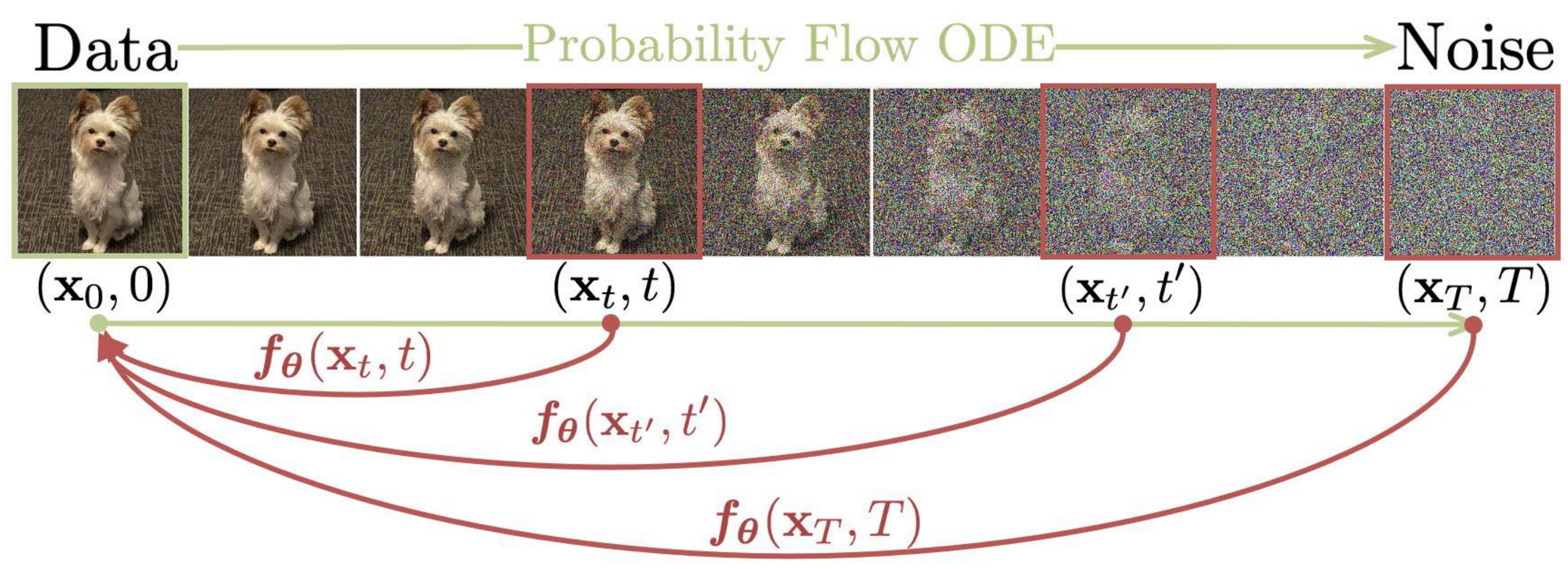


Figure 5. Consistency Model training scheme.

Do we really need pretrained diffusion model?
We do. Consistency model is trained to follow deterministic probability flows. These flows are not available without diffusion model.

What else?
- Consistency trajectory models — train $f_\theta(x_r, r, l)$ to predict x_l to allow multistep sampling.
- Latent consistency models (LCM) — open project supporting LORAs, real-time generation and more.

Results

Family	Model	Latency	FID (↓)
1. Probability Density Distillation	Distribution Matching	0.09s	**11.49**
2. Rectified Flows	InstaFlow	0.09s	13.10
3. Consistency models	LCM + LORA	0.19s	23.62
Teacher	Stable Diffusion 1.5	2.59s	8.78

Table 1. Sample quality comparison on text-to image generation on *MS COCO-30k*. Taken from [3]

References

1 Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Fredo Durand, William T. Freeman, and Taesung Park. **One-step diffusion with distribution matching distillation**. In *CVPR*, 2024

2 Xingchao Liu, Chengyue Gong, and Qiang Liu. **Flow straight and fast: Learning to generate and transfer data with rectified flow**. In *ICLR*, 2023

3 Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. **Consistency models**. In *ICML*, 2023

PDF version