

## ABSTRACT

- We introduce **consistency** models, a new family of generative models optimized for producing high-quality samples efficiently.
- Consistency models support fast one-step generation, offer quality enhancement via multistep generation, and allow flexible zero-shot image editing without model re-training.
- Training can be performed through distillation from pretrained diffusion models or directly from data as standalone generative models.
- In the context of diffusion distillation, consistency models produce state-of-the-art onestep samples.
- As standalone generative models, consistency models outperform other single-step, non-adversarial generative models as well as many GANs.

Code



# **Score-Based Diffusion Models**





### **Estimating the score function:**

## **Sample generation:**

# **Consistency** Models Yang Song, Prafulla Dhariwal, Mark Chen, Ilya Sutskever OpenAI

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

Numerical SDE solvers Numerical ODE solvers Score-based MCMC (predictor-corrector)

> All sampling methods listed above demand repeated score network evaluations





• Consistency Training (CT), valid when  $|t_{n+1} - t_n| \rightarrow 0$ 

 $\mathbb{E}[\lambda(t_n)\|_{L^2}$ 

Continuous-Time Consistency Training

 $\mathbb{E} \left[ \lambda(t) \boldsymbol{f}_{\boldsymbol{\theta}}(t) \right]$ 

## **Properties:**

# **Consistency Models**

$$\forall t \in [0,T] : \boldsymbol{f}_{\boldsymbol{\theta}}(\mathbf{x}_t,t) = \mathbf{x}_0$$

$$f_{\theta}(\mathbf{x}, t) = c_{\text{skip}}(t)\mathbf{x} + c_{\text{out}}(t)F_{\theta}(\mathbf{x}, t)$$
$$c_{\text{skip}}(0) = 1 \qquad c_{\text{out}}(0) = 0$$

One ODE solver step using  
pretrained diffusion model  
$$f_{\theta}(\mathbf{x}_{t_{n+1}}, t_{n+1}) - f_{\theta}(\hat{\mathbf{x}}_{t_{n}}, t_{n}) \parallel ]$$
  
ing Student  
n model Teacher  
model  $\theta^{-} = \text{EMA}(\theta)$ 

$$f_{\theta}(\mathbf{x}+t_{n+1}\mathbf{z},t_{n+1}) - f_{\theta}(\mathbf{x}+t_n\mathbf{z},t_n)$$

$$\mathbf{x}_{t}, t)^{\mathsf{T}} \left( \frac{\partial \boldsymbol{f}_{\boldsymbol{\theta}^{-}}(\mathbf{x}_{t}, t)}{\partial t} + \frac{\partial \boldsymbol{f}_{\boldsymbol{\theta}^{-}}(\mathbf{x}_{t}, t)}{\partial \mathbf{x}_{t}} \cdot \frac{\mathbf{x}_{t} - \mathbf{x}}{t} \right) \right]$$

• These loss functions only provide gradients; their values are not useful for model comparison.

 Analogous to temporal difference learning in RL and bootstrap your own latents in unsupervised learning.



CT TESUILS				
METHOD	NFE $(\downarrow)$	$FID(\downarrow)$	IS (†)	ſ
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	
СТ	2	5.83	8.85	-





(c) Left: A stroke input provided by users. Right: Stroke-guided image generation.