Bidirectional Consistency Models

Liangchen Li[*, 1] Jiajun He[*, 2]
*equal contribution [1]Independent Researcher [2]University of Cambridge

Check out our paper!

Code and weights released!

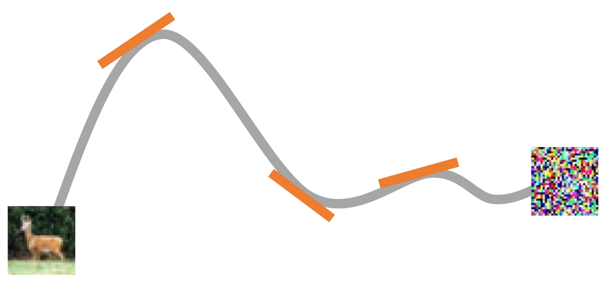
TL;DR: We extend consistency models to Bidirectional Consistency Models for fast sampling and its inversion.

Motivation

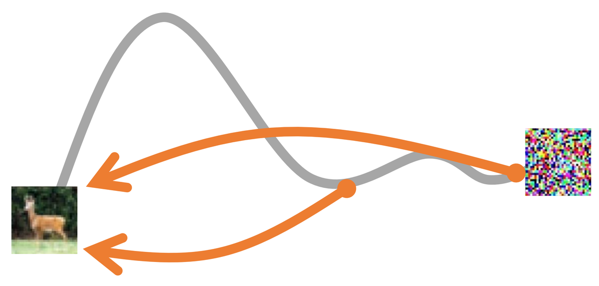
- Diffusion models requires **hundreds of NFEs** for high-quality samples; consistency models (CMs) only requires **1-2 NFE**;
- (ODE-based) diffusion models can map noise ⟷ image
- Consistency models only support noise ⟶ image

Motivation

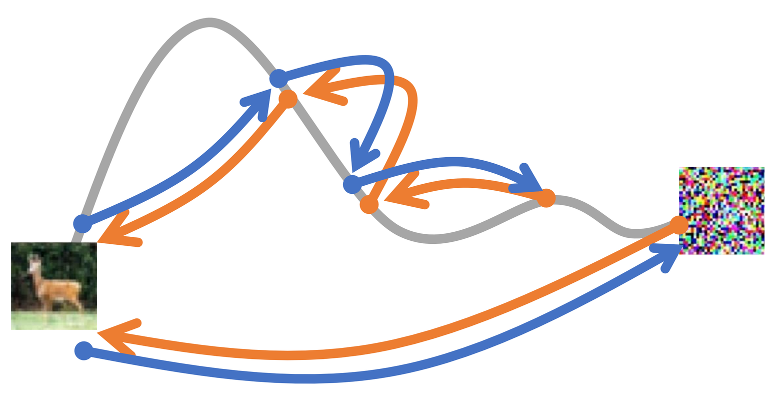
Diffusion Models estimate **scores** along the PF ODE:



Consistency Models estimates **starting points** of the PF ODE:



Bidirectional Consistency Models estimates **the points on the entire PF ODE** towards both denoising and noising directions:



Methods

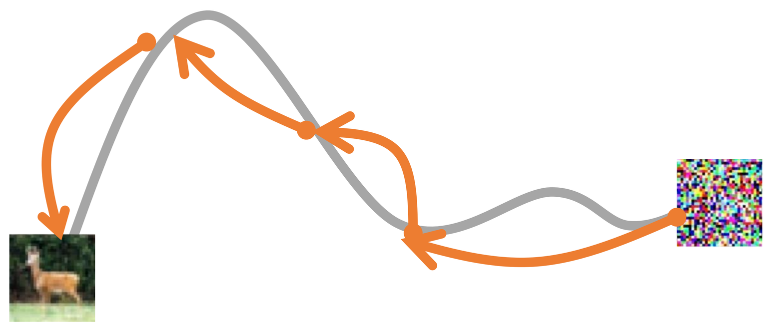
- We train a network $f_\theta(x, t_1, t_2)$ mapping x from time step t_1 to t_2;
- Given training image x, Gaussian noise z, and random time steps t, t', we calculate:
  1. Target image:
  $$x_0 \leftarrow f_{\mathrm{sg}(\theta)}(x + tz, t, 0)$$
  2. Estimator of x_0:
  $$x_0{}' \leftarrow f_\theta(x + (t+\delta)z, t+\delta, 0)$$
  3. Estimator of $x_{t'}$:
  $$x_{t'} \leftarrow f_\theta(x + tz, t, t')$$
  4. New estimator of x_0:
  $$x_0{}'' \leftarrow f_{\mathrm{sg}(\theta)}(x_{t'}, t', 0)$$
- We minimize $d(x_0, x_0')$ and $d(x_0, x_0'')$ together:

$$\ell = \frac{1}{\delta} d(x_0, x_0') + \frac{1}{|t - t'|} d(x_0, x_0'')$$

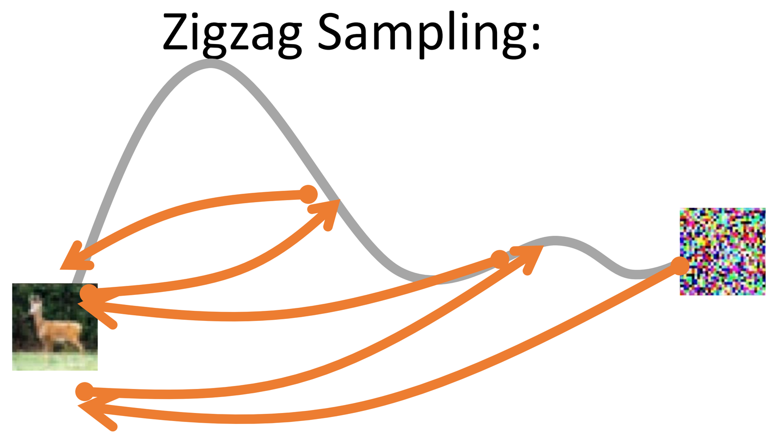
Consistency training loss — 'soft' trajectory constraint

New Sampling Schemes

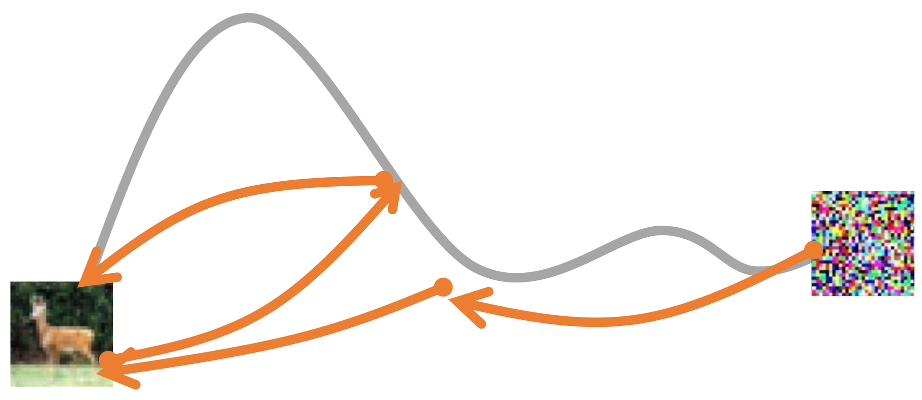
Ancestral Sampling:



Zigzag Sampling:



Combination of both can yield better performance:



Results

- In terms of sampling, BCM achieves competitive FID compared to CMs:

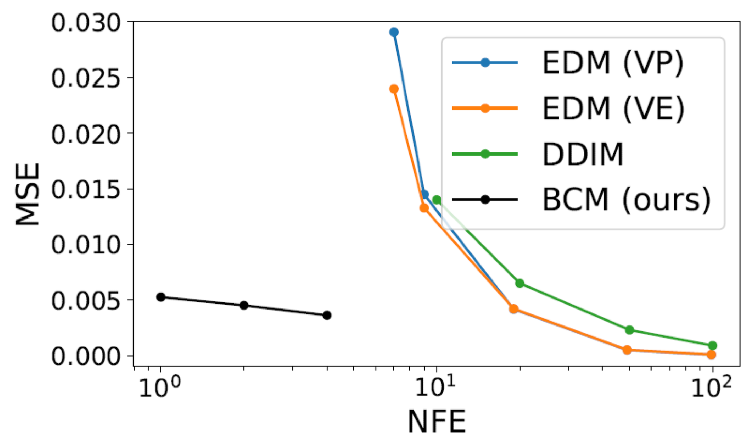
CIFAR-10

Methods	NFE	FID
iCT	1	2.83
	2	2.46
iCT-deep	1	2.51
	2	2.24
BCM	**1**	**3.10**
	2	**2.39**
	3	**2.50**
	4	**2.29**
BCM-deep	**1**	**2.64**
	2	**2.36**
	3	**2.19**
	4	**2.07**

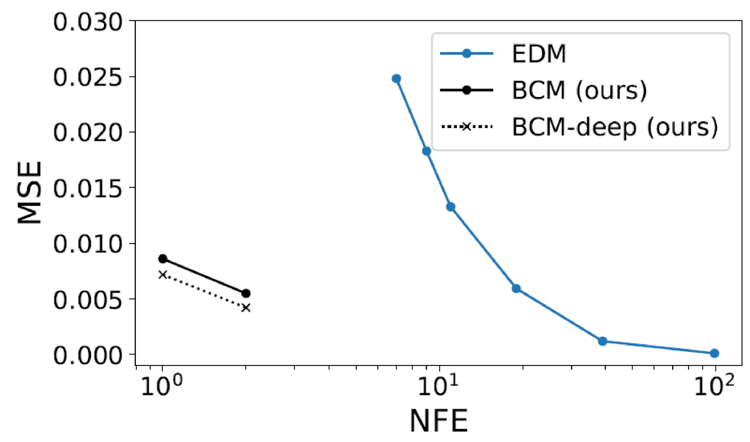
ImageNet-64

Methods	NFE	FID
iCT	1	4.02
	2	3.20
iCT-deep	1	3.25
	2	2.77
BCM	**1**	**4.18**
	2	**2.88**
	3	**2.78**
	4	**2.68**
BCM-deep	**1**	**3.14**
	2	**2.45**
	3	**2.61**
	4	**2.35**

- In terms of inversion, BCM achieves lower reconstruction error with fewer NFE:



(a) CIFAR-10.



(b) ImageNet-64.

- Interpolate between two real images and blind restoration of JPEG Images: