MGD^3: Mode-Guided Dataset Distillation using Diffusion Models

Jeffrey A. Chan-Santiago[1], Praveen Tirupattur[1], Gaurav Kumar Nayak[2], Gaowen Liu[3], Mubarak Shah[1]

[1] Center for Research in Computer Vision, University of Central Florida
[2] Mehta Family School of DS & AI, Indian Institute of Technology Roorkee, India
[3] Cisco Research

Project Page

ORAL

ICML International Conference On Machine Learning

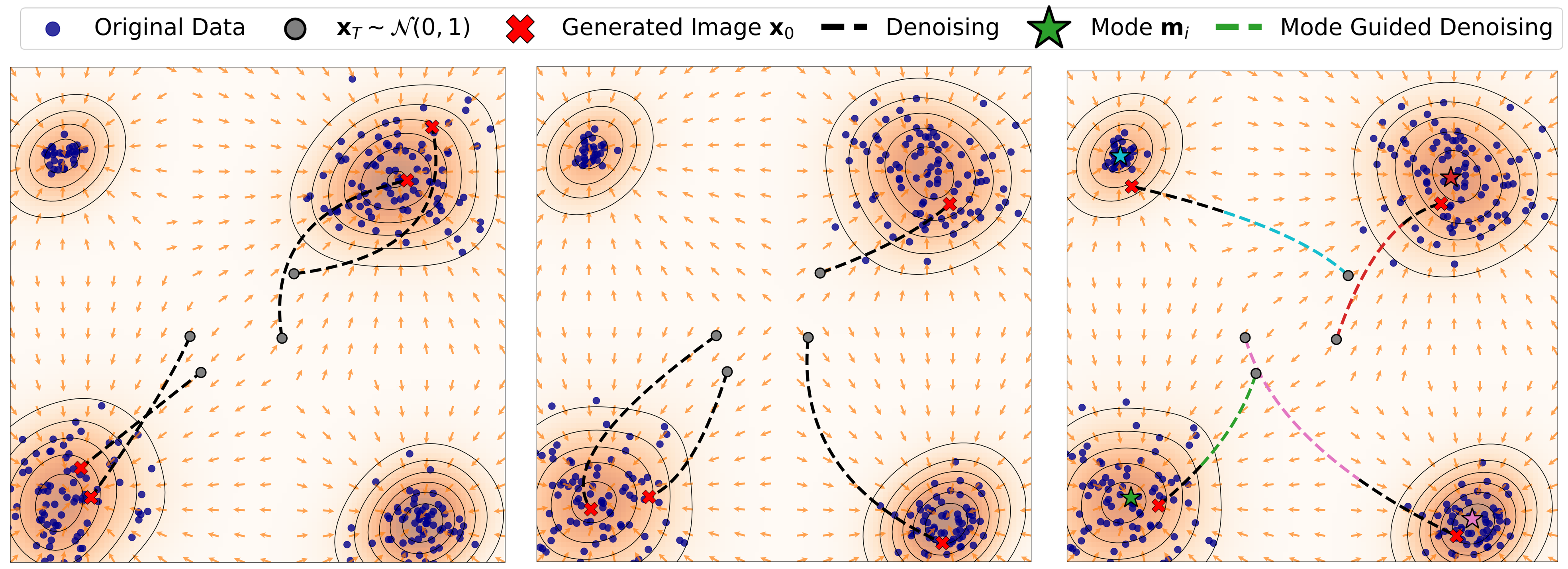
GENERATIVE DATASET DISTILLATION

- **Dataset distillation** synthesize a small dataset such that training a model on it achieves similar performance to training on the full dataset.
- Optimization-based methods were the first to emerge but are slow and hard to scale to large datasets and high resolutions.
- Generative dataset distillation compresses the full dataset into a generative model that samples the distilled data.
- The challenge is to generate a small, diverse, and representative dataset that closely approximates the full dataset with only a few samples.

CONTRIBUTIONS

- We propose a **training-free** method that to synthesize distilled datasets.
- Our approach achieves **diverse and representative** distilled datasets.
- Achieves **state-of-the-art performance** across multiple benchmarks.
- Works with text-to-image models, **even when not trained on target dataset**.

MOTIVATION



Pretrained Diffusion
No Fine-tuning
Low Diversity

MinMax Diffusion
Requires Fine-tuning
High Diversity

Ours
No Fine-tuning
High Diversity

METHOD

1 Mode Discovery:

Estimate the N modes per Class

$X_c = \{x_i \mid y_i = c\}$ samples from a class c

$M(X_c, N) \rightarrow \{m_1, \dots, m_n\}$

2 Mode Guidance:

Sample one image per mode m_i

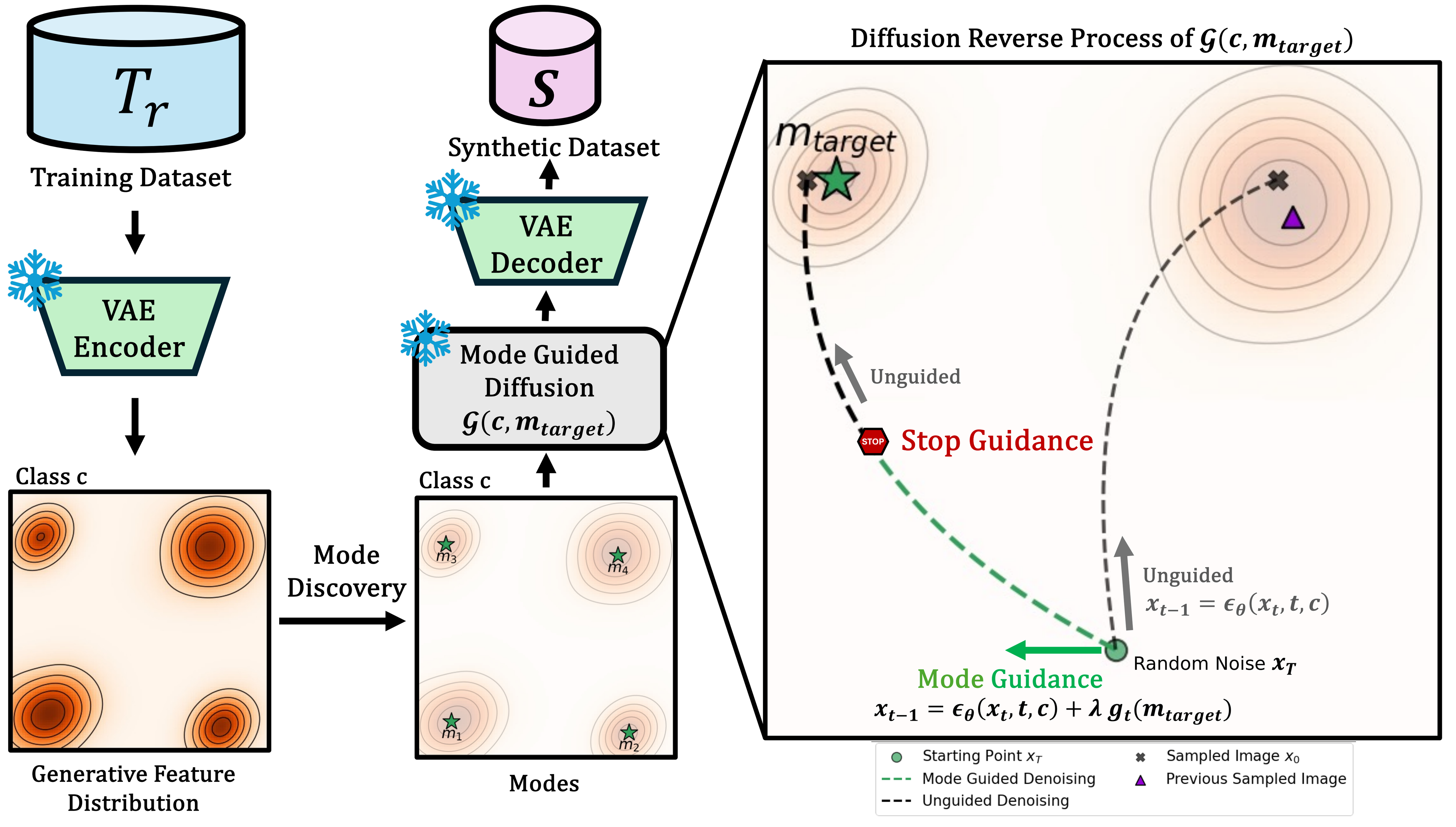
$$\widehat{\epsilon_\theta}(x_t, t, c) = \epsilon_\theta(x_t, t, c) + \lambda\, \boldsymbol{g(x_t, t, m_i)}\sigma_t$$

$$g(x_t, t, m_i) = m_i - \frac{1 - \sqrt{1-\bar{\alpha}_t}\,\epsilon}{\sqrt{\bar{\alpha}_t}} = m_i - x_0^t$$

Direction toward mode m_i

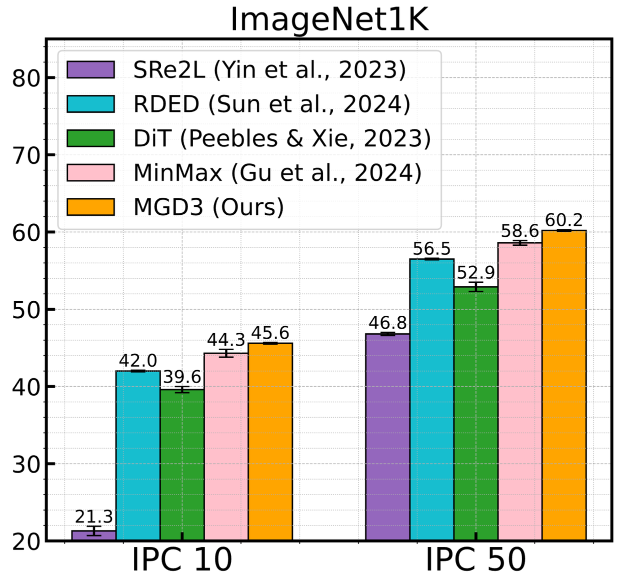
3 Stop Guidance:

Stop Mode Guidance when $t < t_{stop}$



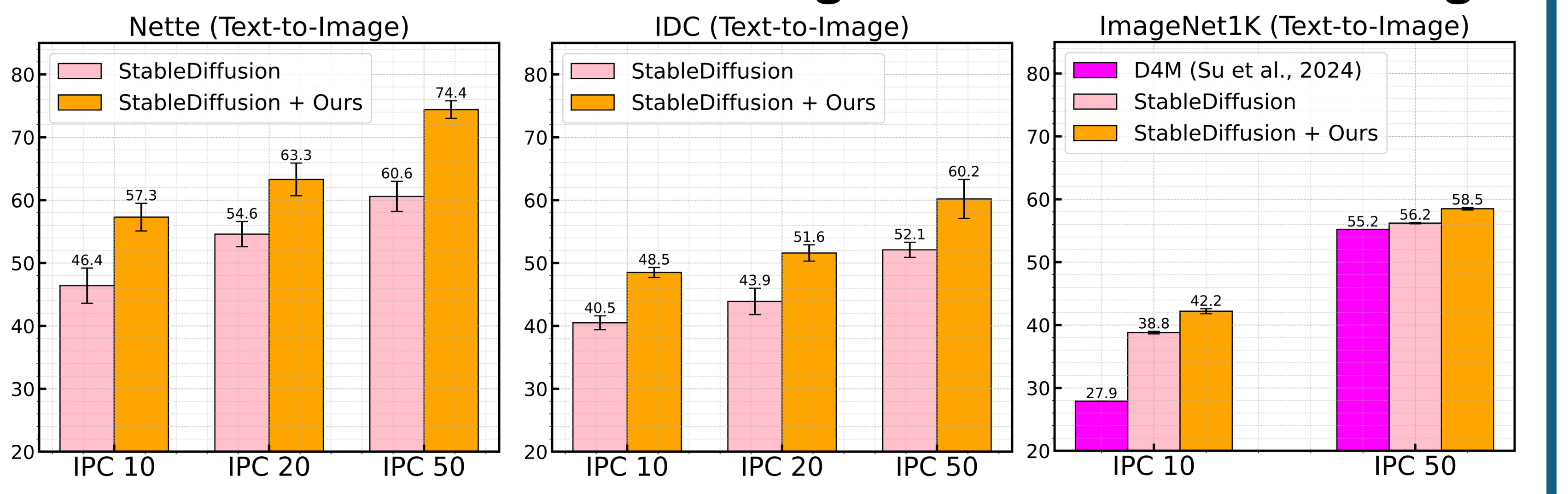
COMPARISON WITH PREVIOUS WORK

	ImageNette			ImageIDC		
Image per Class	10	20	50	10	20	50
Random	$54.2_{\pm1.6}$	$63.5_{\pm0.5}$	$76.1_{\pm1.1}$	$48.1_{\pm0.8}$	$52.5_{\pm0.9}$	$68.1_{\pm0.7}$
DM (Zhao & Bilen, 2023)	$60.8_{\pm0.6}$	$66.5_{\pm1.1}$	$76.2_{\pm0.4}$	$52.8_{\pm0.5}$	$58.5_{\pm0.4}$	$69.1_{\pm0.8}$
MinMaxDiff (Gu et al., 2024)	$62.0_{\pm0.2}$	$66.8_{\pm0.4}$	$76.6_{\pm0.2}$	$53.1_{\pm0.2}$	$59.0_{\pm0.4}$	$69.6_{\pm0.2}$
LDM (Rombach et al., 2022)	$60.3_{\pm3.6}$	$62.0_{\pm2.6}$	$71.0_{\pm1.4}$	$50.8_{\pm1.2}$	$55.1_{\pm2.0}$	$63.8_{\pm0.4}$
LDM + Disentangled Diffusion (D^{4}M (Su et al., 2024))	$59.1_{\pm0.7}$	$64.3_{\pm0.5}$	$70.2_{\pm1.0}$	$52.3_{\pm2.3}$	$55.5_{\pm1.2}$	$62.7_{\pm0.8}$
LDM + **MGD^3 (Ours)**	$61.9_{\pm4.1}$	$65.3_{\pm1.3}$	$74.2_{\pm0.9}$	$53.2_{\pm0.2}$	$58.3_{\pm1.7}$	$67.2_{\pm1.3}$
DiT (Peebles & Xie, 2023)	$59.1_{\pm0.7}$	$64.8_{\pm1.2}$	$73.3_{\pm0.9}$	$54.1_{\pm0.4}$	$58.9_{\pm0.2}$	$64.3_{\pm0.6}$
DiT + Disentangled Diffusion (D^{4}M (Su et al., 2024))	$60.4_{\pm3.4}$	$65.5_{\pm1.2}$	$73.8_{\pm1.7}$	$51.1_{\pm2.4}$	$58.0_{\pm1.4}$	$64.1_{\pm2.5}$
DiT + **MGD^3 (Ours)**	**$66.4_{\pm2.4}$**	**$71.2_{\pm0.5}$**	**$79.5_{\pm1.3}$**	**$55.9_{\pm2.1}$**	**$61.9_{\pm0.9}$**	**$72.1_{\pm0.8}$**

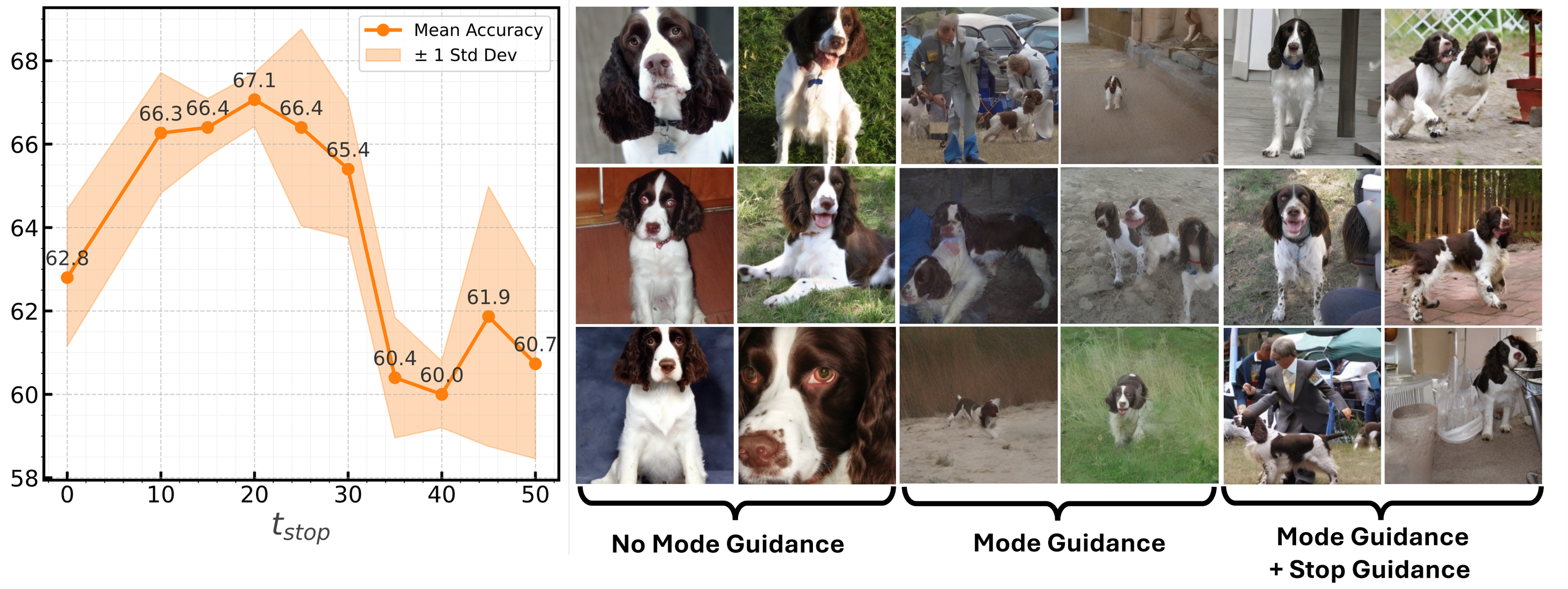


RESULTS

Dataset Distillation Without Target Distribution Pretraining



When should mode guidance stop?



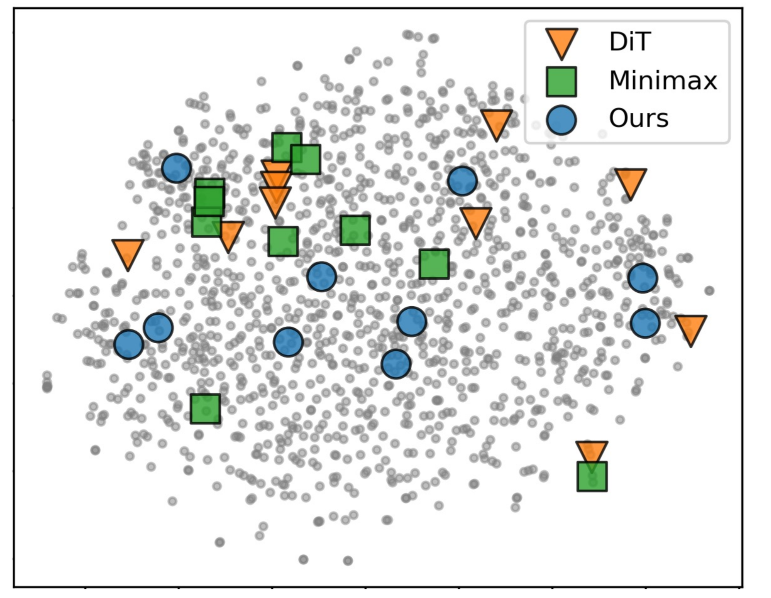
Effect of Mode Discovering

Mode Discovery	Accuracy
Random	$59.6_{\pm1.8}$
DBSCAN	$61.3_{\pm1.9}$
Spectral Clustering	$64.5_{\pm2.1}$
k-Means (centroid)	$66.4_{\pm2.4}$
k-Means (closest sample)	$64.6_{\pm0.4}$
GMM	**$66.9_{\pm0.4}$**

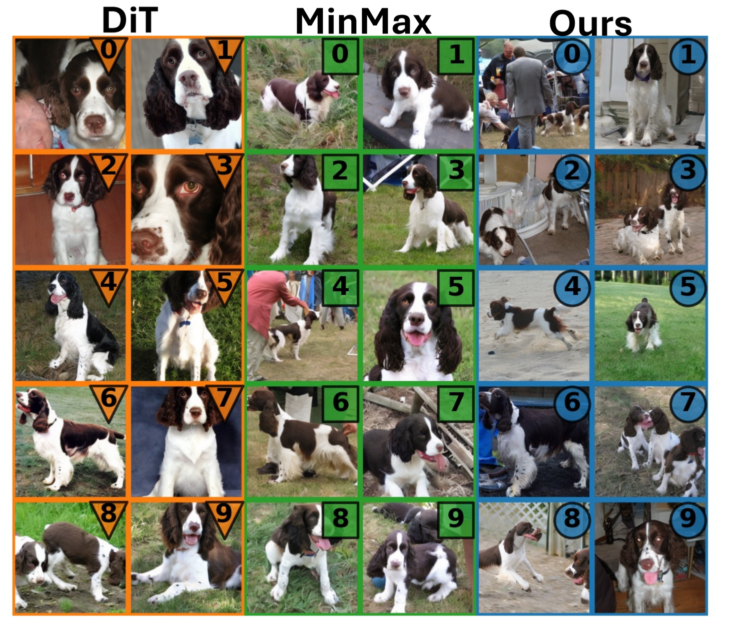
Ablation with Diffusion Samplers

Method	DDPM	DDIM
LDM	$60.3_{\pm3.6}$	$60.4_{\pm3.1}$
LDM + **Ours**	$61.9_{\pm4.1}$	$62.3_{\pm1.1}$
DiT	$58.8_{\pm2.1}$	$61.4_{\pm2.4}$
DiT + **Ours**	**$66.4_{\pm2.4}$**	**$66.6_{\pm0.6}$**

Qualitative Results



Wider t-SNE spread shows our method generates a more diverse dataset.



Our method yields greater diversity in backgrounds, poses, and viewpoints.