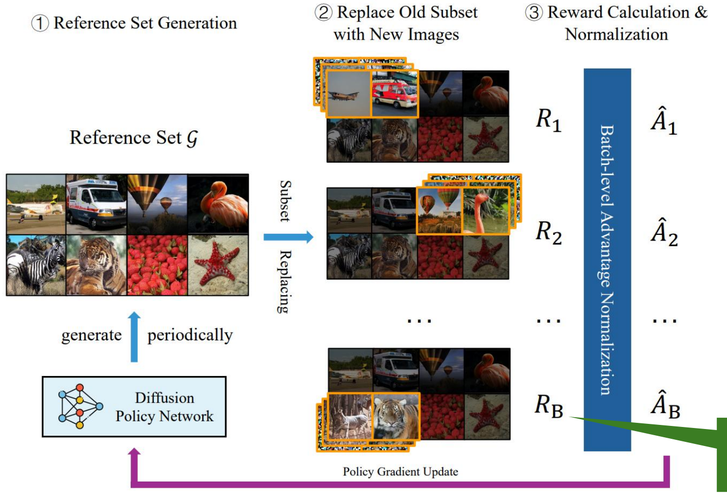
Optimizing Visual Generative Models via Distribution-wise Rewards

Ruihang Li[1,2,3], Mengde Xu[3], Shuyang Gu[3], Leigang Qu[4†], Fuli Feng[1], Han Hu[3], Wenjie Wang[1†] († Corresponding Authors)

[1] University of Science and Technology of China, [2] Shanghai Innovation Institute, [3] Hunyuan Frontier Lab, Tencent, [4] National University of Singapore

Core Idea

Method



Optimize the distribution, not isolated samples

Sample-wise RL rewards optimize each image independently, which can lead to reward hacking, visual artifacts, and diversity collapse. We instead optimize a **distribution-wise reward** that evaluates how well generated samples match the real data distribution, encouraging both high fidelity and broad sample diversity.

Turn distribution-wise FID into sample-wise Rewards

Directly optimizing FID is costly and sparse because it evaluates an entire generated distribution (~50K). We make it trainable by **subset replacement**: for each rollout, a small same-class subset (~50) of the generated reference set (~5K) is replaced by new samples. The updated set is compared with the real distribution using FID, and **−FID** becomes the reward assigned to those newly generated samples.

In Short...

Generate a *Refenence Set* with 5K images, replace 50 of them, recalculate the FID-5K, use **−FID** as rewards for the newly generated 50 images.

Experiments

How effective is distribution-wise RL?

SiT-XL/2 improves from FID 8.30 → 5.77 **(30% better!)**, with FD-DINOv2 reduced from 230.39 → 164.88.

Is the gain just FID overfitting?

No. KID, MMD, FD-DINOv2, Precision, Density, and Coverage all improve, showing broader distributional gains.

What training recipe works best?

Ablations suggest using a 5K reference set, replacing 50 images, and refreshing the reference set every 10 steps.

Model	Training Steps	FID ↓	FD_{DINOv2} ↓
ADM	1.98M	10.94	-
ADM-U	1.98M	7.49	-
LDM-8	4.8M	15.51	-
LDM-4	178K	10.56	-
DiT-XL/2	400K	19.50	-
DiT-XL/2	7M	9.60	-
SiT-XL/2	400K	17.20	-
SiT-XL/2	7M	8.30	230.39
+ Ours (RS)	+ 120	6.98	183.75
+ Ours (RL)	+ 450	**5.77**	**164.88**

Metric	SiT Original	+ Ours (RL)	Change
FID↓	8.30	5.77	↓30.5%
KID↓	0.0043	0.0020	↓53.5%
MMD↓	0.0029	0.0015	↓48.3%
FD_{DINOv2} ↓	230.39	164.88	↓28.5%
Precision↑	0.6983	0.7286	↑4.3%
Recall↑	0.7527	0.7262	−3.5%
Density↑	0.7673	0.8594	↑12.0%
Coverage↑	0.8698	0.8950	↑2.9%

Ablation Studies

- Selecting the global top 25% of samples for training is optimal. Just drop the samples with worse FIDs.
- RL training after Rejection Sampling fine-tuning provided no performance gain, likely due to overfitting from the RS phase. We therefore adopted a pure RL approach.