Breaking Class Barriers: Efficient Dataset Distillation via Inter-class Feature Compensator

Xin Zhang[1,2], Jiawei Du[1,2], Ping Liu[3], Joey Tianyi Zhou[1,2*]

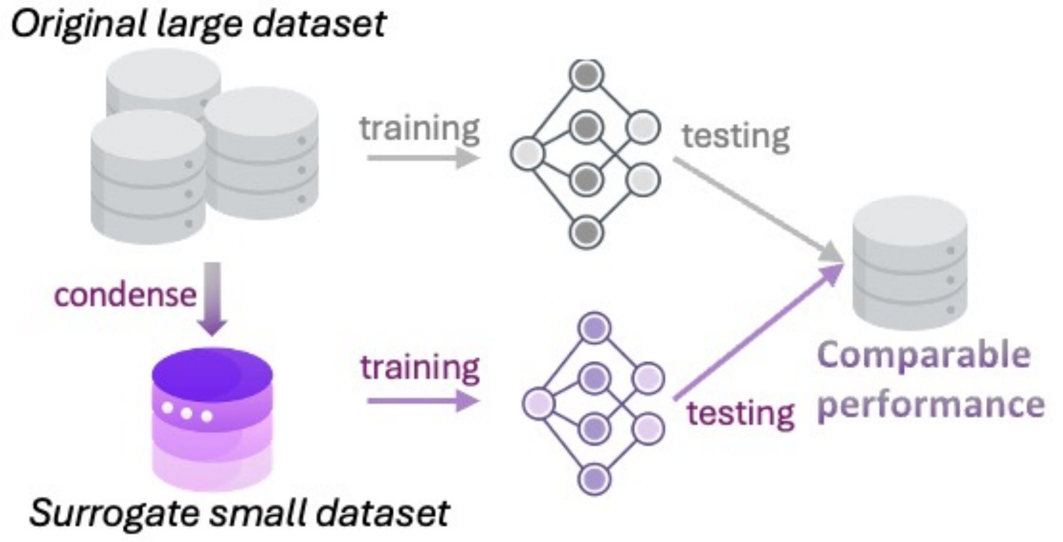
*[1]Centre for Frontier AI Research (CFAR), A*STAR, Singapore,*
*[2]Institute of High Performance Computing (IHPC), A*STAR, Singapore, [3]University of Nevada, Reno*

paper codes

Centre for Frontier AI Research
A*STAR CFAR

Background

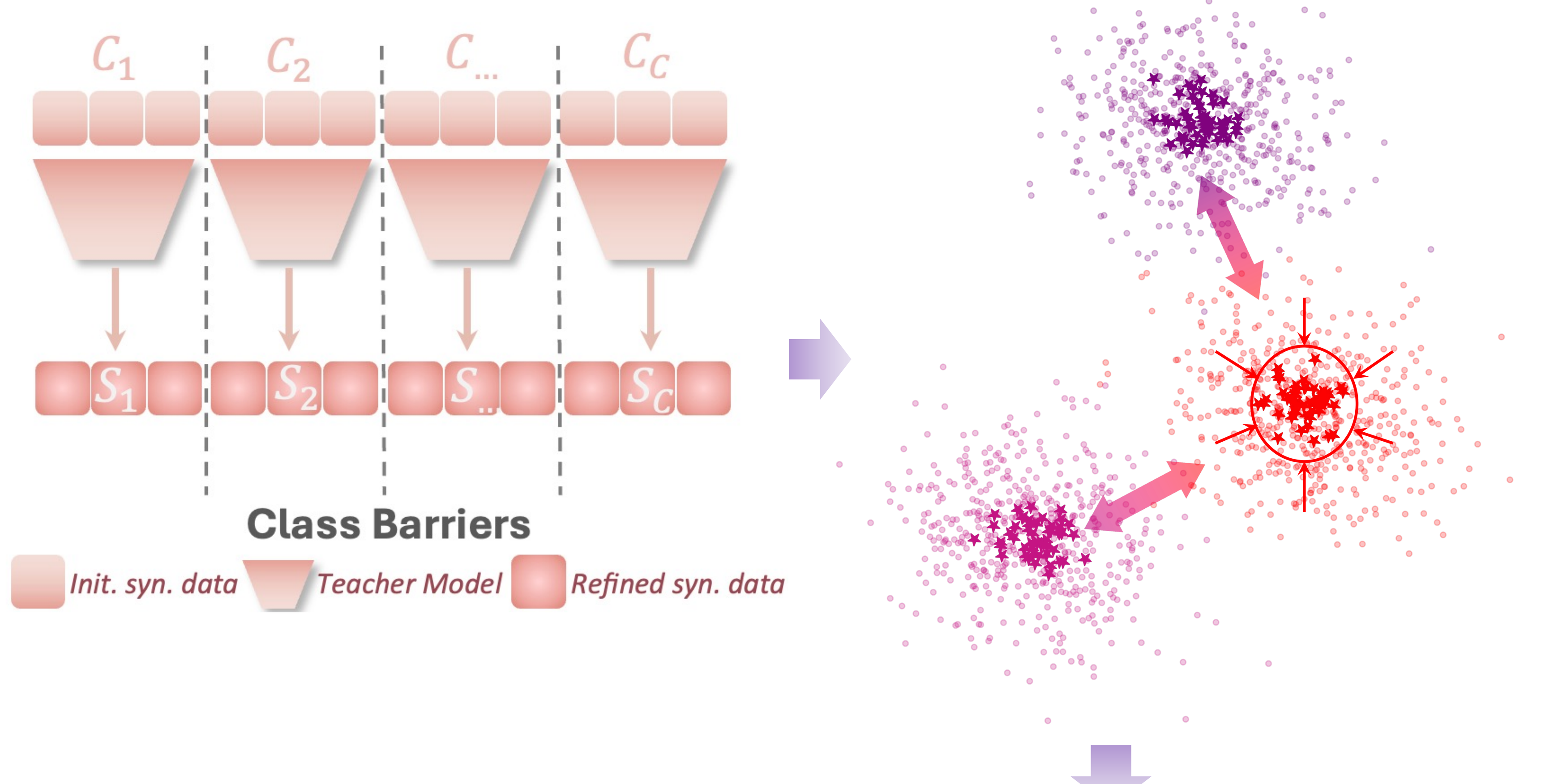
What is Dataset Distillation (DD) ?



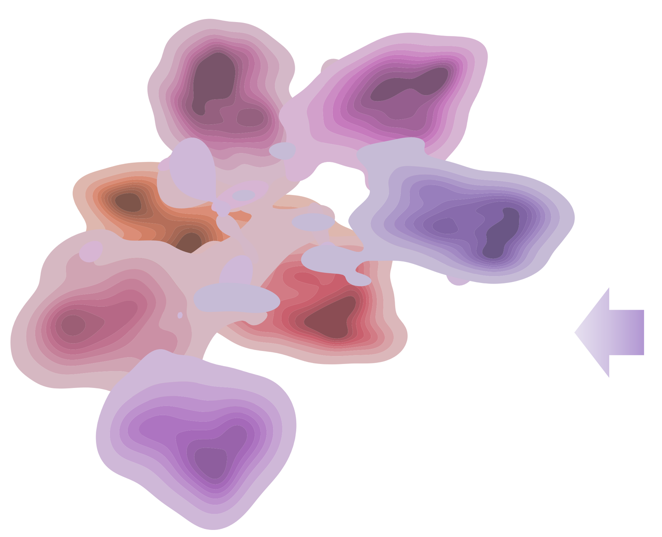
Dataset distillation *is to* ***synthesize*** *a* ***tiny and compact dataset*** *from a given* ***real and large dataset****, such that the former can yield a* ***comparable performance*** *as the latter.*

How do existing methods achieve DD?

❖ *"One instance for one class" paradigm*



Poor performance



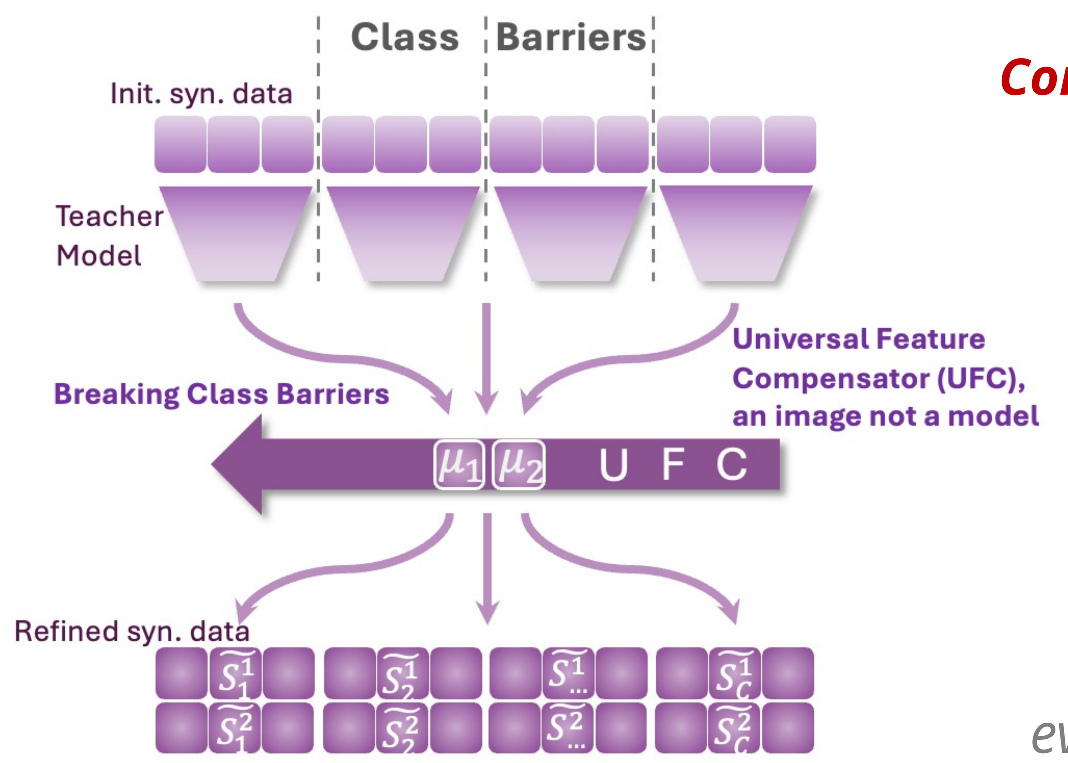
evaluated on validation set

❖ *Duplicated intra-class features*
❖ *Oversight of inter-class features*

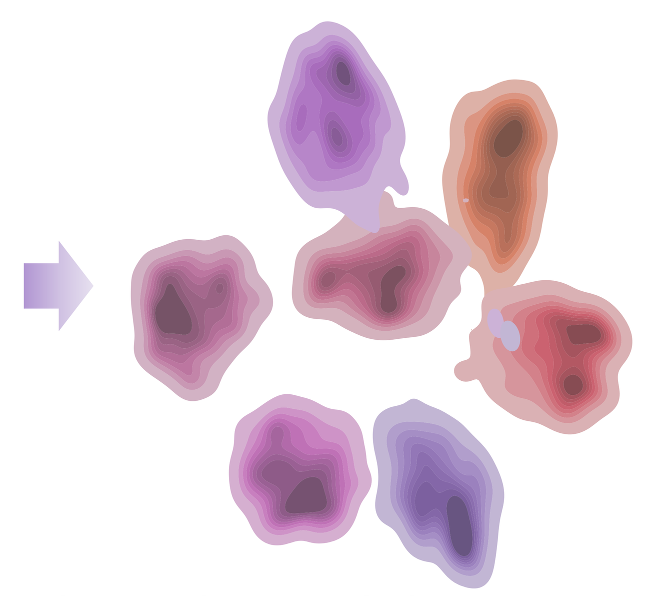
❖ *Inefficient Utilization of the Distillation Budget*
❖ *Collapsed feature diversity*

Proposed Method

Breaking Class Barriers:
Efficient DD via Inter-class Feature Compensator

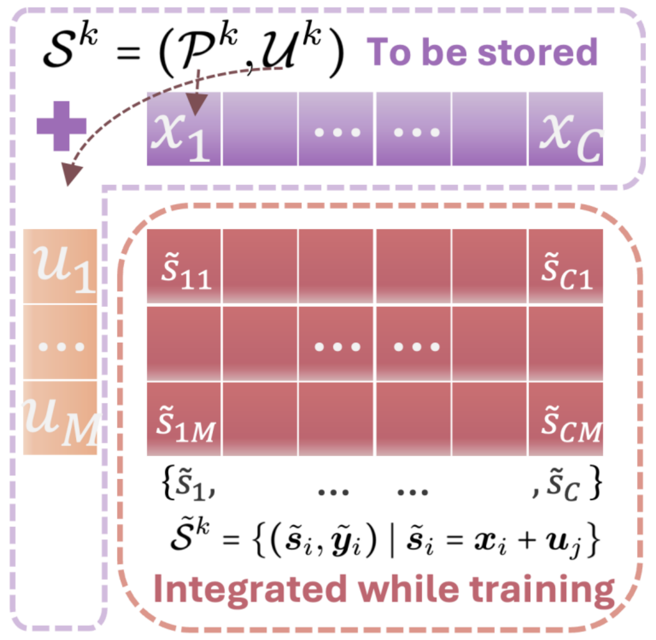


Considerable performance boost



evaluated on validation set

❖ *Implementation*



○ *Design:*

$$\tilde{\mathcal{S}}^k = \{(\tilde{\boldsymbol{s}}_i, \tilde{\boldsymbol{y}}_i) \mid \tilde{\boldsymbol{s}}_i = \boldsymbol{x}_i + \boldsymbol{u}_j, \text{ for each } \boldsymbol{x}_i \in \mathcal{P}^k \text{ and each } \boldsymbol{u}_j \in \mathcal{U}^k\}$$

○ *Optimization:*

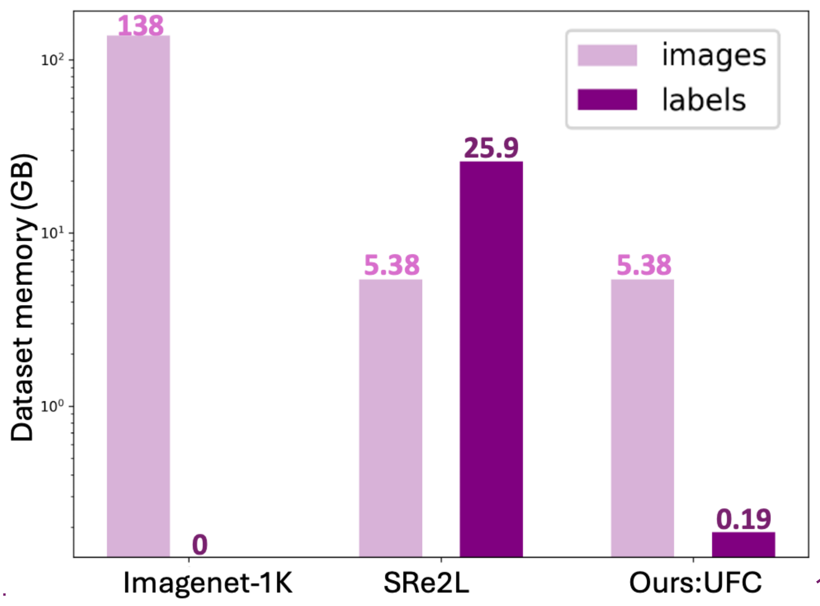
$$\arg\min_{\boldsymbol{u}_j \in \mathbb{R}^d} \sum_{(\boldsymbol{x}_i, \boldsymbol{y}_i) \in \mathcal{P}^k} \left[\ell\left(f_{\theta_\mathcal{T}}, \boldsymbol{x}_i + \boldsymbol{u}_j, \boldsymbol{y}_i\right) + \alpha \mathcal{L}_{\mathrm{BN}}\left(f_{\theta_\mathcal{T}}, \boldsymbol{x}_i + \boldsymbol{u}_j\right)\right],$$

$$\text{where} \quad \mathcal{L}_{\mathrm{BN}}\left(f_{\theta_\mathcal{T}}, \boldsymbol{x}_i + \boldsymbol{u}_j\right) = \sum_l \|\mu_l(\tilde{\mathcal{S}}_j^k) - \mu_l(\mathcal{T})\|_2 + \sum_l \|\sigma_l^2(\tilde{\mathcal{S}}_j^k) - \sigma_l^2(\mathcal{T})\|_2$$

○ *Efficient static labeling:*

$$f_{\theta_\mathcal{T}}[\lambda \tilde{\boldsymbol{s}}_i + (1-\lambda)\tilde{\boldsymbol{s}}_j] \approx \lambda f_{\theta_\mathcal{T}}(\tilde{\boldsymbol{s}}_i) + (1-\lambda) f_{\theta_\mathcal{T}}(\tilde{\boldsymbol{s}}_j), \forall \tilde{\boldsymbol{s}}_i, \tilde{\boldsymbol{s}}_j \in \tilde{\mathcal{S}}$$

Reduces the size of soft labels by up to 99%

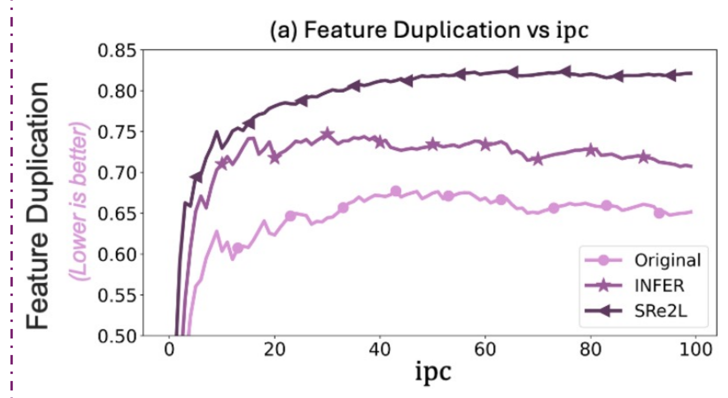


Experiments

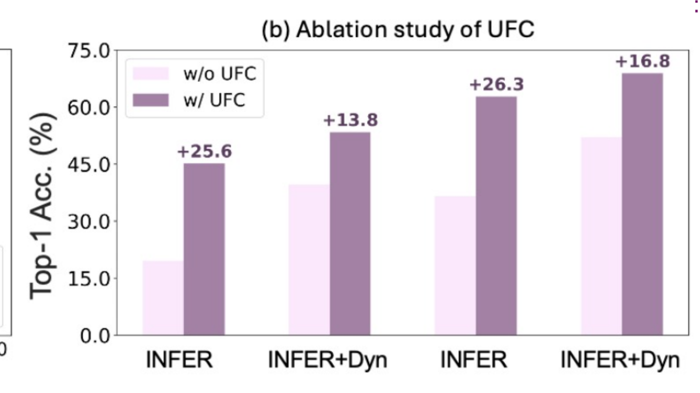
Results:

ipc	Compression Ratio 10	Compression Ratio 50	ResNet18 10	ResNet18 50	ResNet-50 10	ResNet-50 50	ResNet-101 10	ResNet-101 50
Random	0.78%	3.90%	10.5 ±0.4	31.4 ±0.3	9.3 ±0.3	31.5 ±0.2	10.0 ±0.4	33.1 ±0.1
SRe2L* (Yin et al., 2024)	0.81%	4.04%	9.8 ±0.1	17.3 ±0.5	8.7 ±0.3	17.2 ±0.4	8.8 ±0.2	15.8 ±0.2
G-VBSM* (Shao et al., 2024)	0.81%	4.04%	11.9 ±0.2	32.9 ±0.1	14.5 ±0.2	38.1 ±0.2	13.9 ±0.1	38.9 ±0.4
INFER	0.81%	4.04%	**28.7 ±0.2**	**51.8 ±0.2**	**26.9 ±0.3**	**53.3 ±0.3**	**26.5 ±0.1**	**52.2 ±0.3**
SRe2L (Yin et al., 2024)	4.53%	22.67%	21.3 ±0.6	46.8 ±0.2	28.4 ±0.1	55.6 ±0.3	30.9 ±0.1	60.8 ±0.5
G-VBSM (Shao et al., 2024)	4.53%	22.67%	31.4 ±0.5	51.8 ±0.4	35.4 ±0.8	58.7 ±0.3	38.2 ±0.4	**61.0 ±0.4**
INFER+Dyn	4.53%	22.67%	**36.3 ±0.3**	**55.6 ±0.2**	**38.3 ±0.5**	**63.4 ±0.3**	**38.9 ±0.5**	60.7 ±0.1

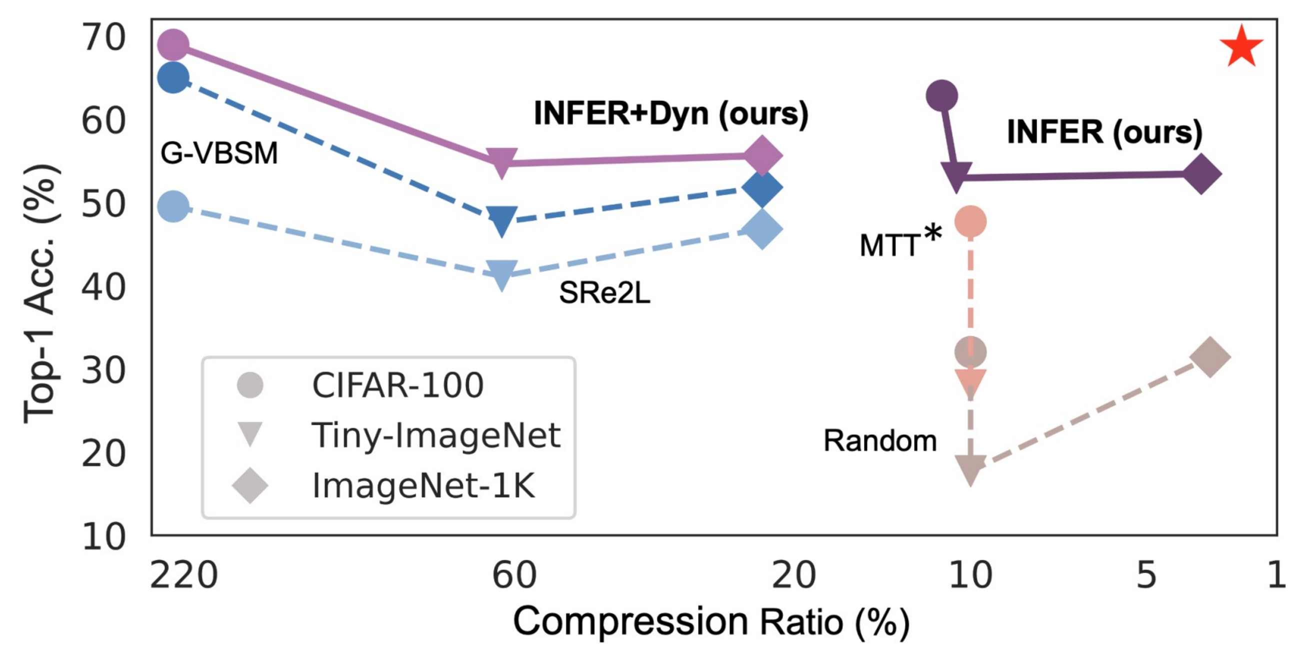
Comparison with SOTAs on ImageNet-1K



The change in feature duplication with the increase of ipc



The ablation study of UFC



Better performance & higher compression ratio

Visualization

References:
X. Zhang, J. Du, P. Liu, and J. T. Zhou, "Breaking Class Barriers: Efficient Dataset Distillation via Inter-Class Feature Compensator," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2025.
Z. Yin, E. Xing, and Z. Shen, "Squeeze, Recover and Relabel: Dataset Condensation at ImageNet Scale From A New Perspective," in *Proc. 37th Conf. Neural Information Processing Systems (NeurIPS)*, 2023.