



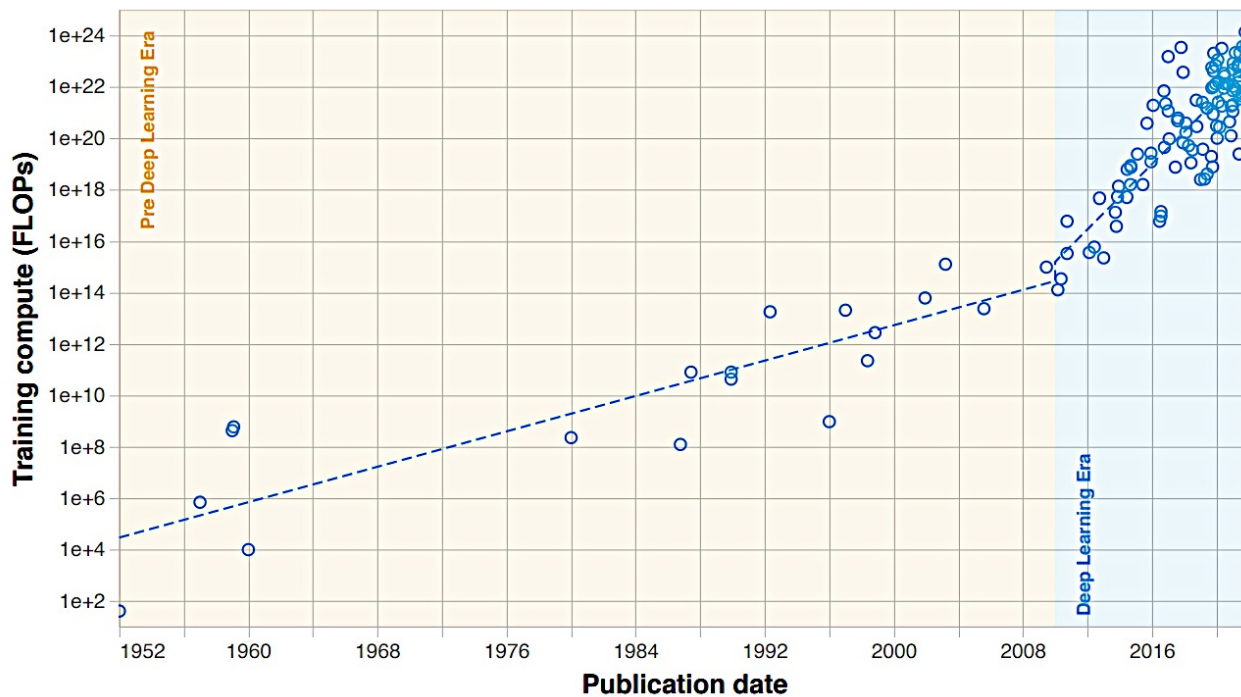
Learnability-guided Diffusion for Dataset Distillation

Jeffrey A. Chan-Santiago¹, Mubarak Shah¹

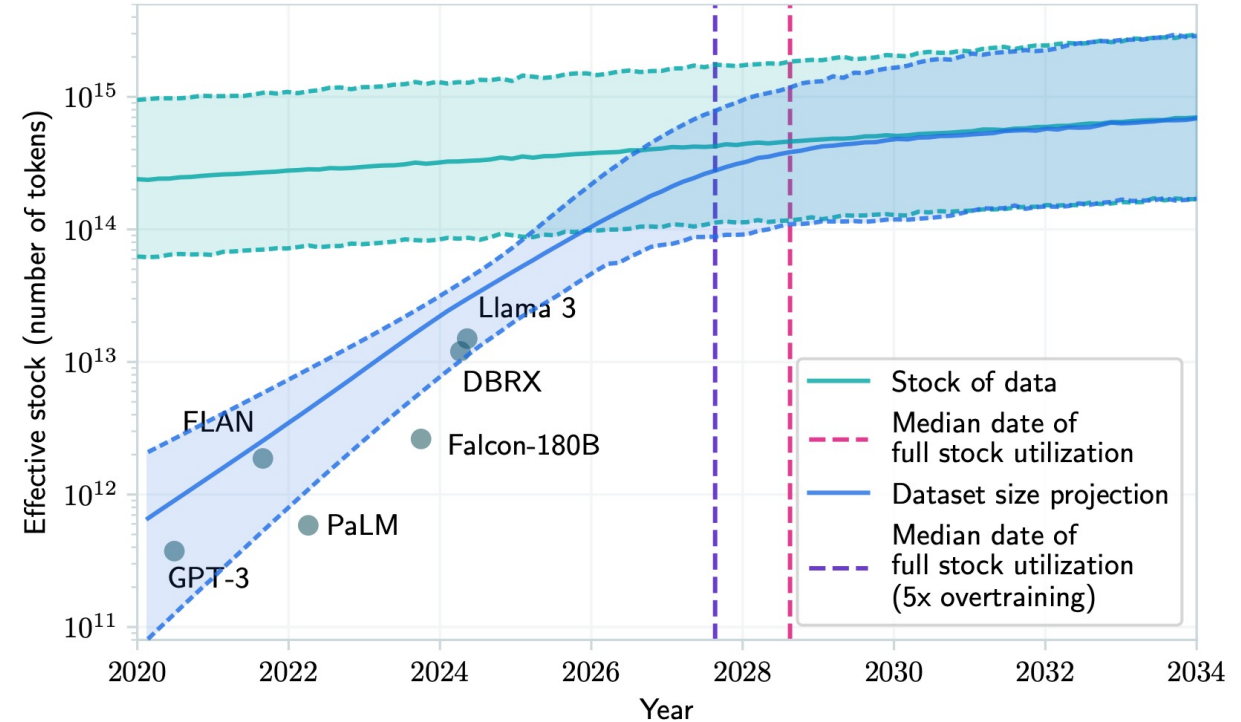
¹ Institute of Artificial Intelligence, University of Central Florida

Motivation: Scaling in Deep Learning Training

Scaling Trend: Need for larger datasets and models to achieve better performance.



Exponential Increase in Compute

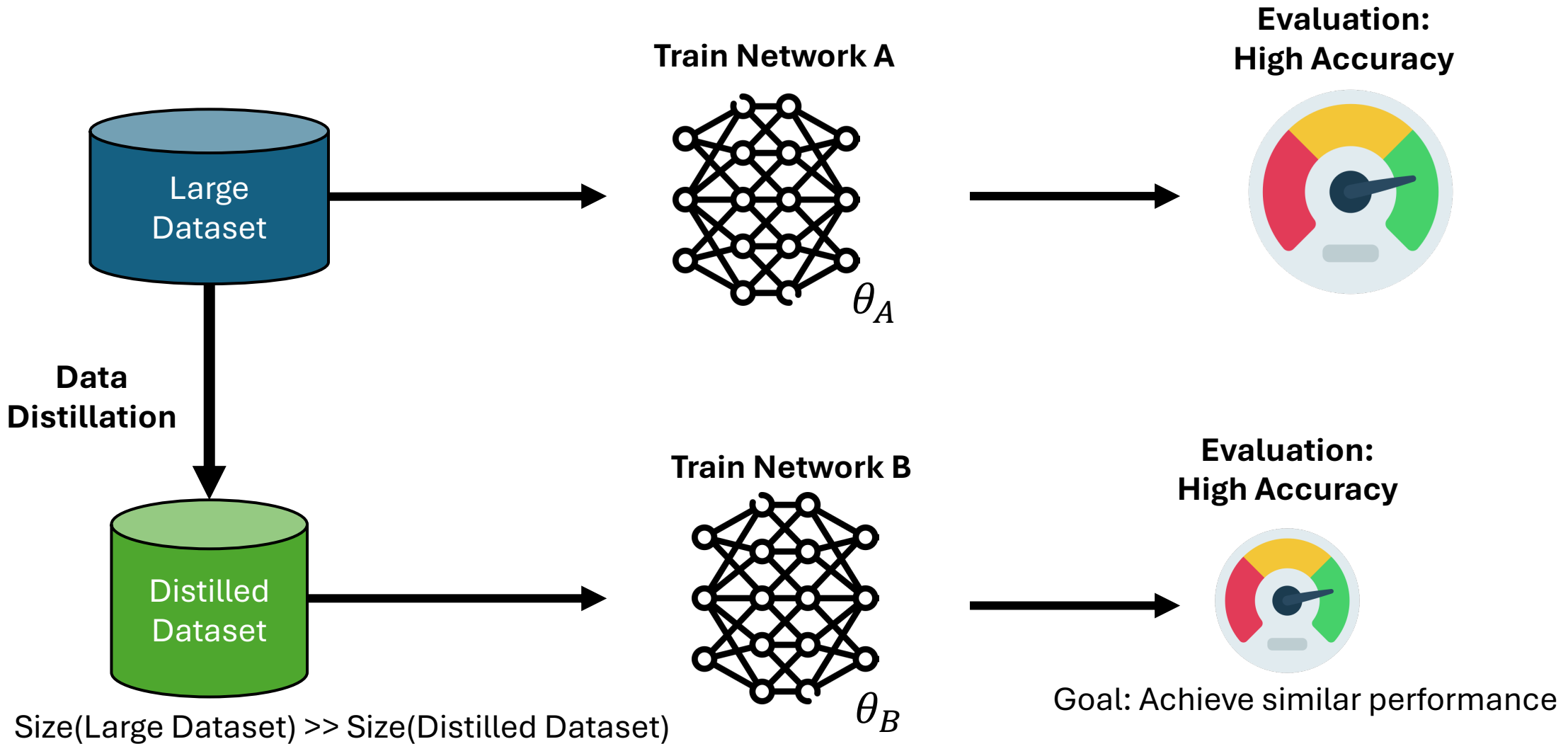


Saturation of Training Data

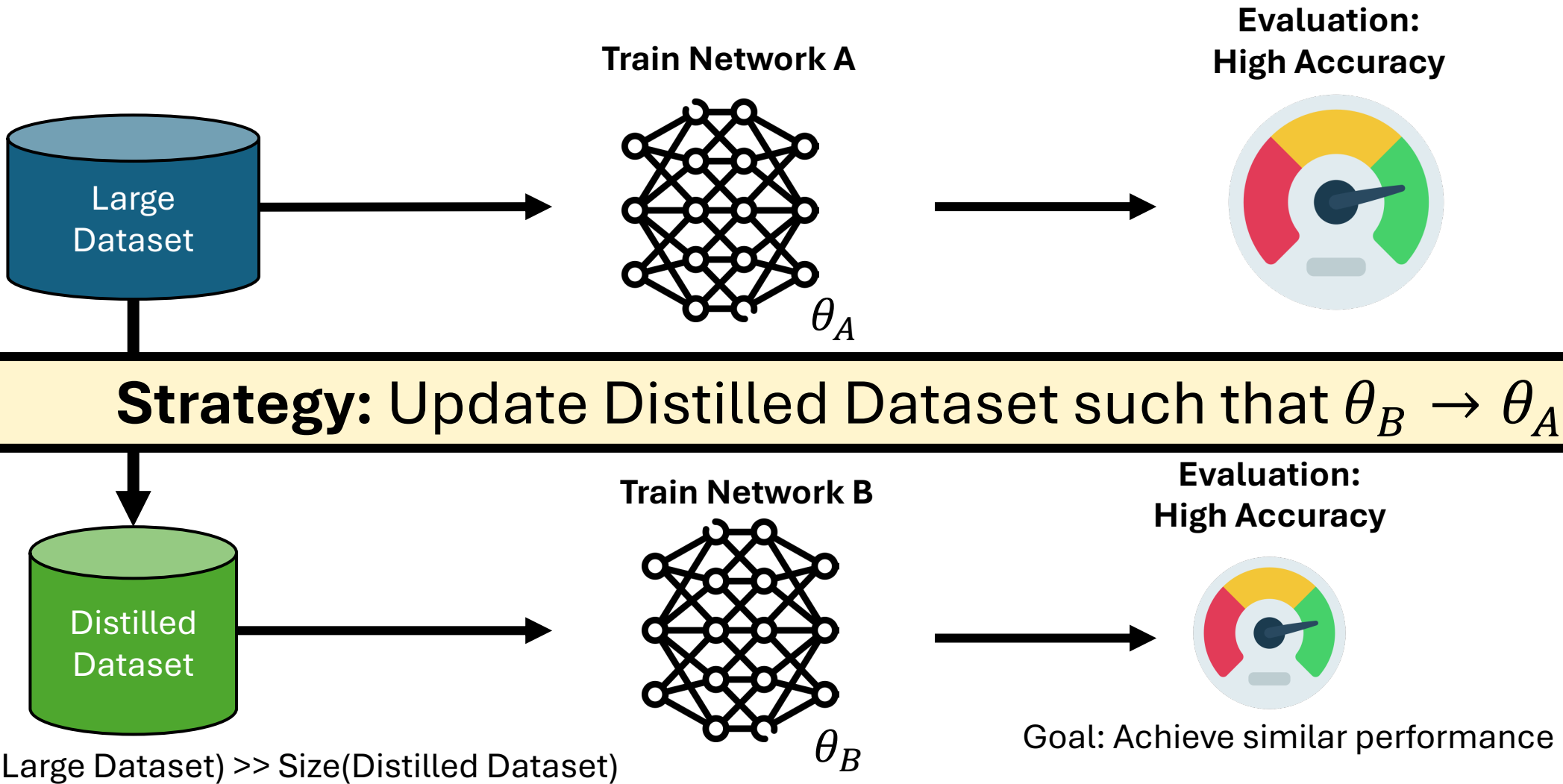
Resource Constraints: Challenging for researchers with limited computation and storage.

Goal: Achieve similar performance using only a fraction of the training data.

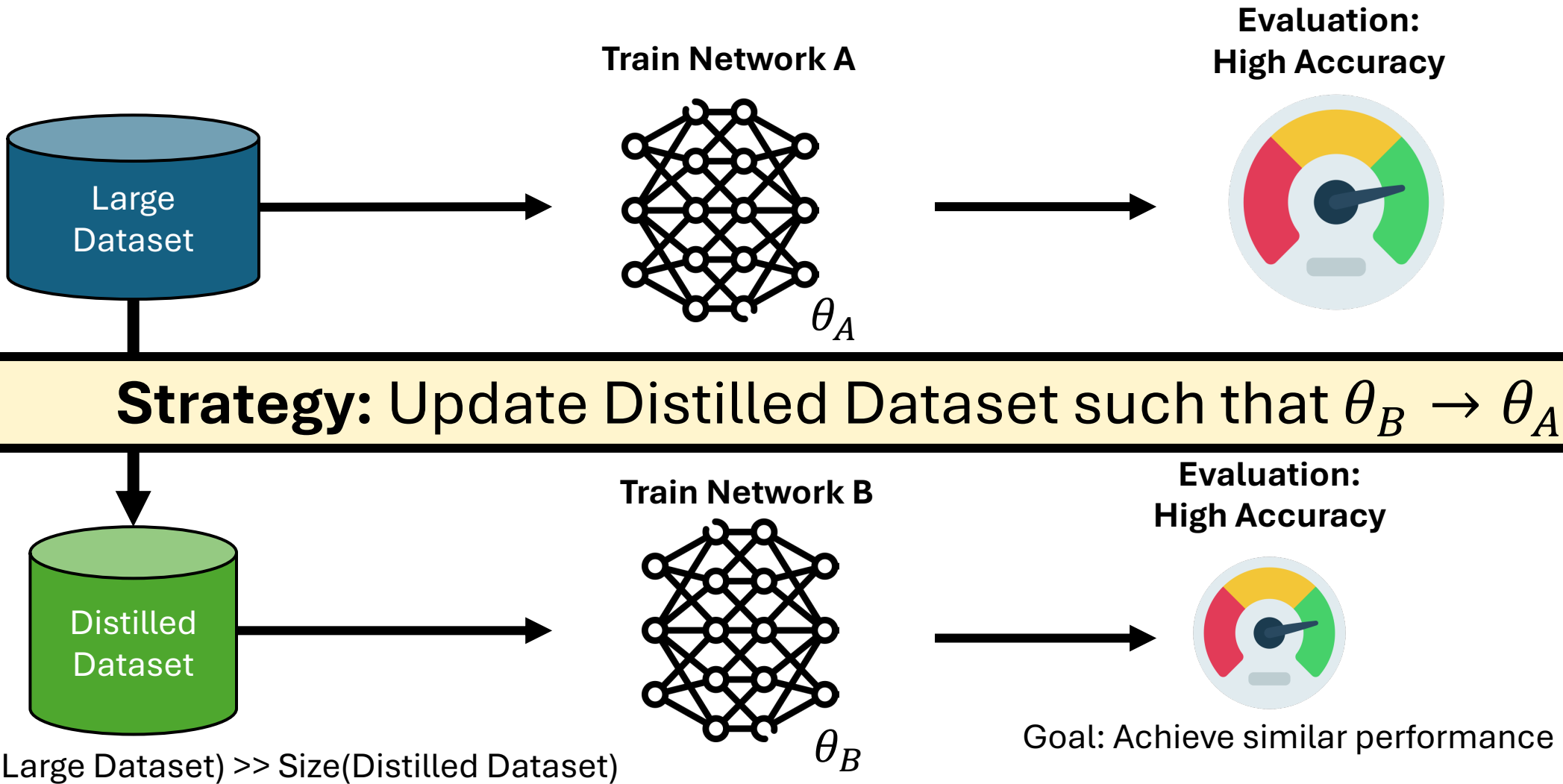
Dataset Distillation



Dataset Distillation



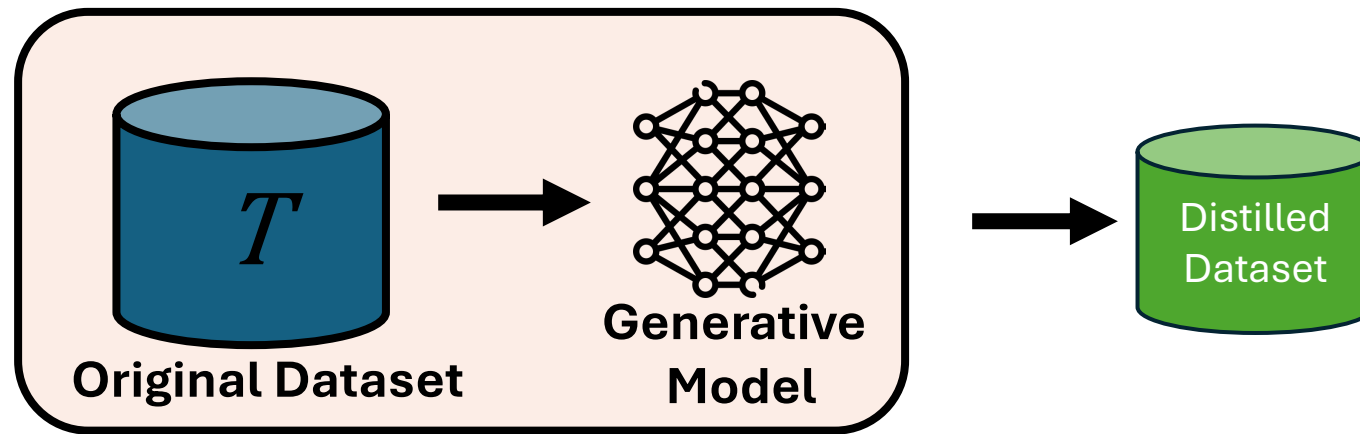
Dataset Distillation



Impact: Small dataset = less training cost

Related Work: Generative Dataset Distillation

Compresses the distribution of a large dataset into a generative model, which is then used to sample a distilled dataset.



Diffusion models effectively capture the underlying data distribution¹.

Samples that are both **diverse** and **representative** of the original data.

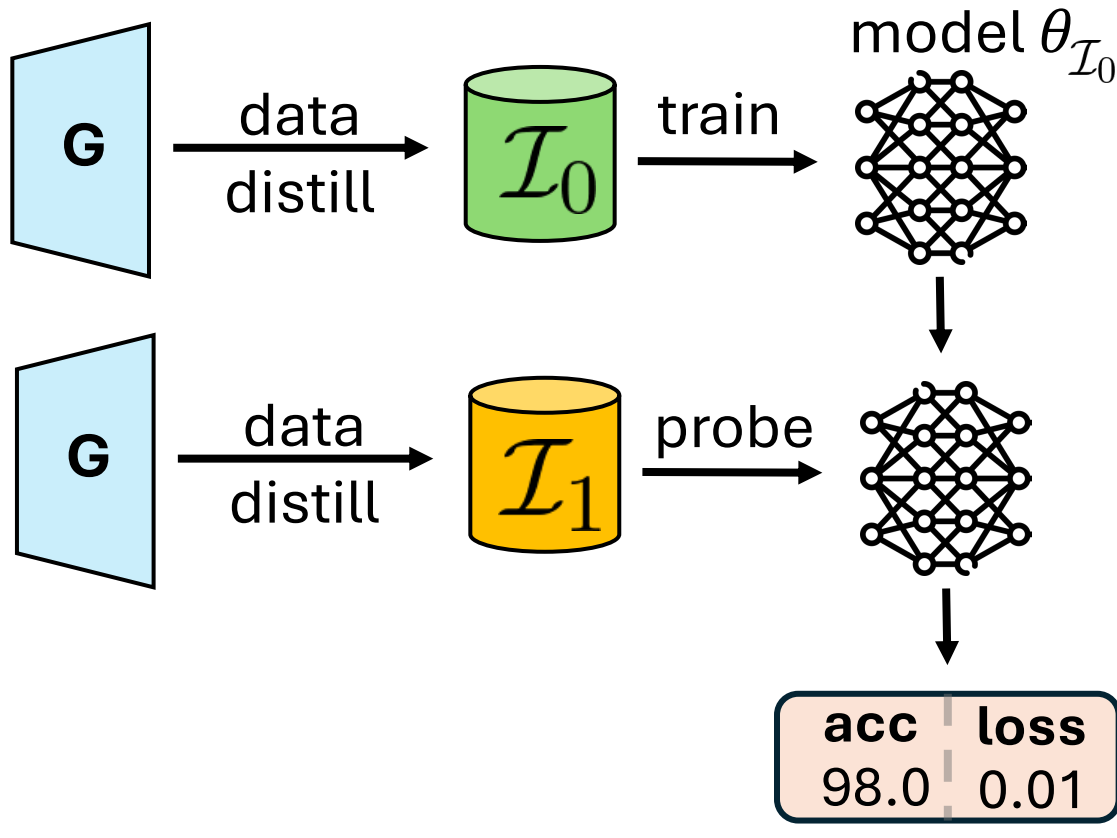
¹Dhariwal, Prafulla, and Alexander Nichol. "Diffusion models beat GANs on image synthesis." *NeurIPS* (2021)

Incremental Dataset Distillation

$$\mathcal{D}_s = \mathcal{I}_0 \cup \mathcal{I}_1$$

IPC 10 IPC 10

Previous Methods



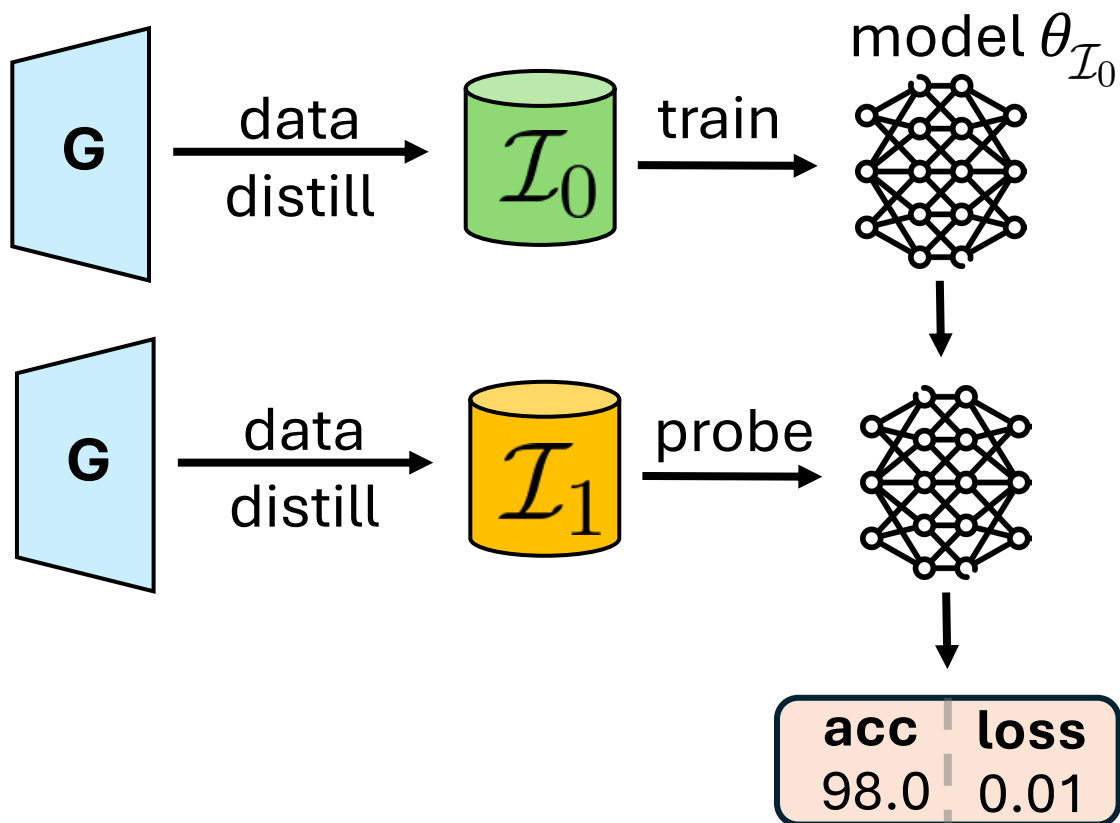
\mathcal{I}_1 is redundant (low loss)

Incremental Dataset Distillation

$$\mathcal{D}_s = \mathcal{I}_0 \cup \mathcal{I}_1$$

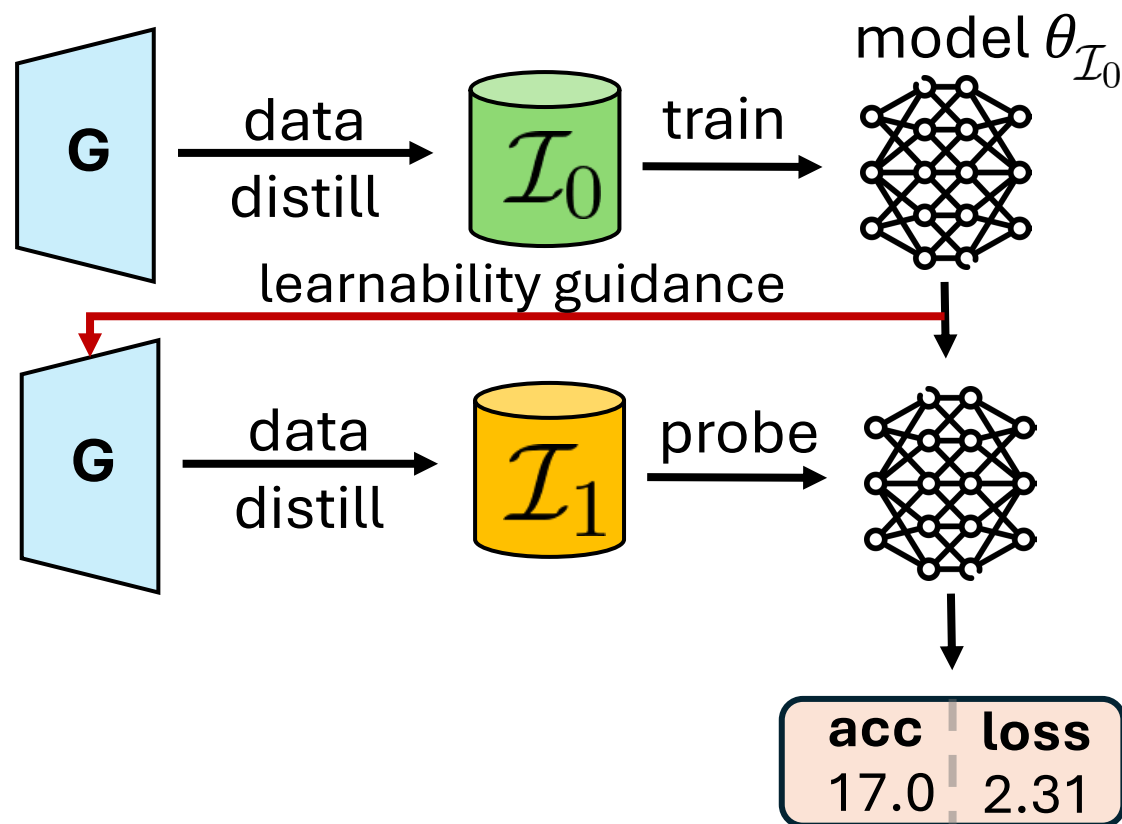
IPC 10 IPC 10

Previous Methods



\mathcal{I}_1 is redundant (low loss)

LGD (Ours)



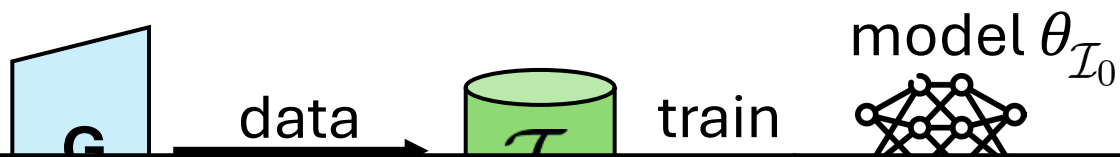
\mathcal{I}_1 is complementary (high loss)

Incremental Dataset Distillation

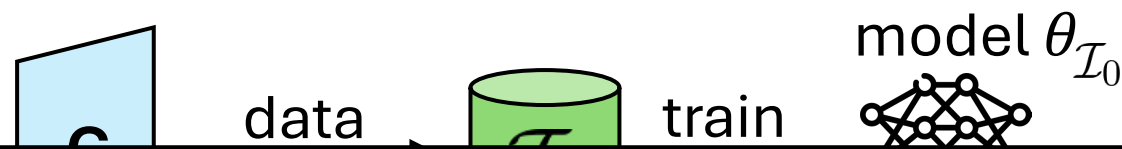
$$\mathcal{D}_s = \mathcal{I}_0 \cup \mathcal{I}_1$$

IPC 10 IPC 10

Previous Methods



LGD (Ours)



Why?:

- 1) Avoid the training signal redundancy
- 2) Each Increment Targets Current Model Knowledge Gap

distill



acc	loss
98.0	0.01

\mathcal{I}_1 is redundant (low loss)

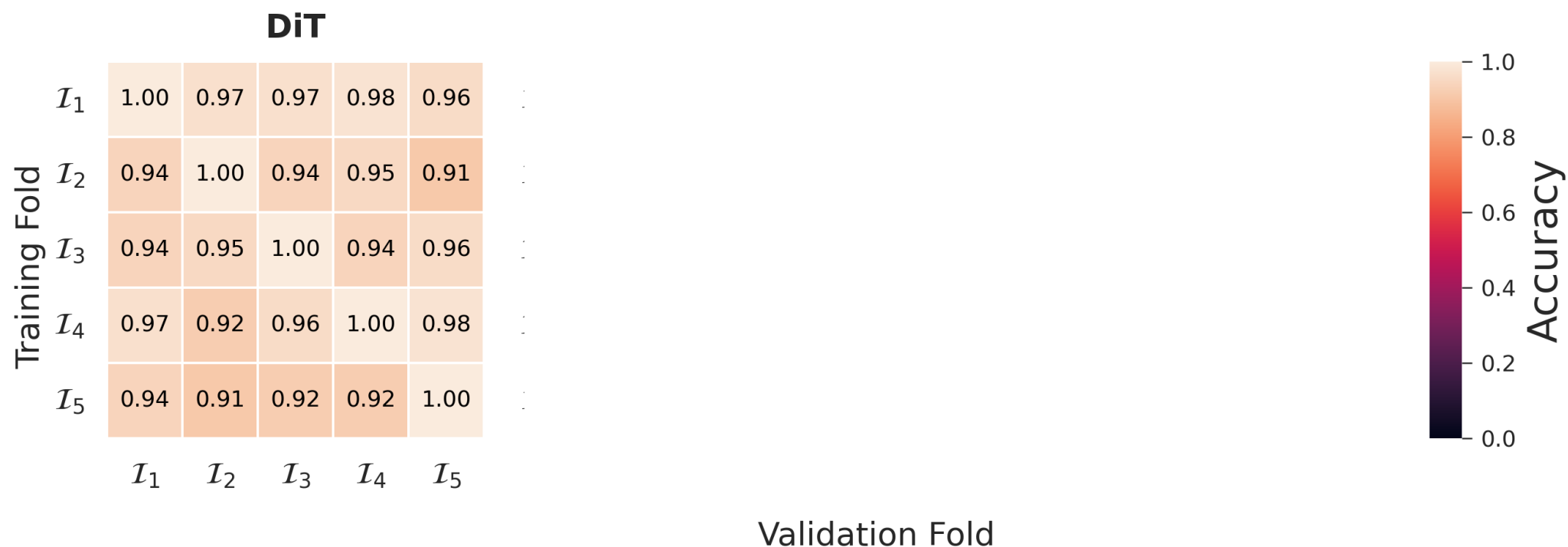
distill



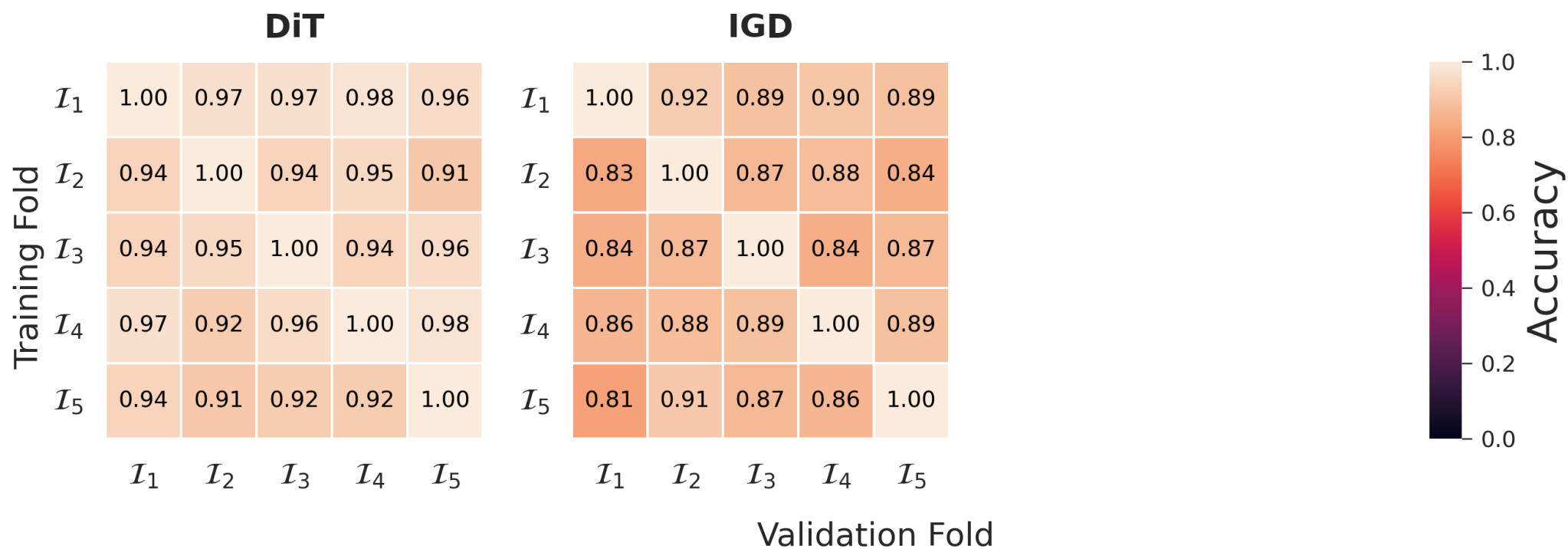
acc	loss
17.0	2.31

\mathcal{I}_1 is complementary (high loss)

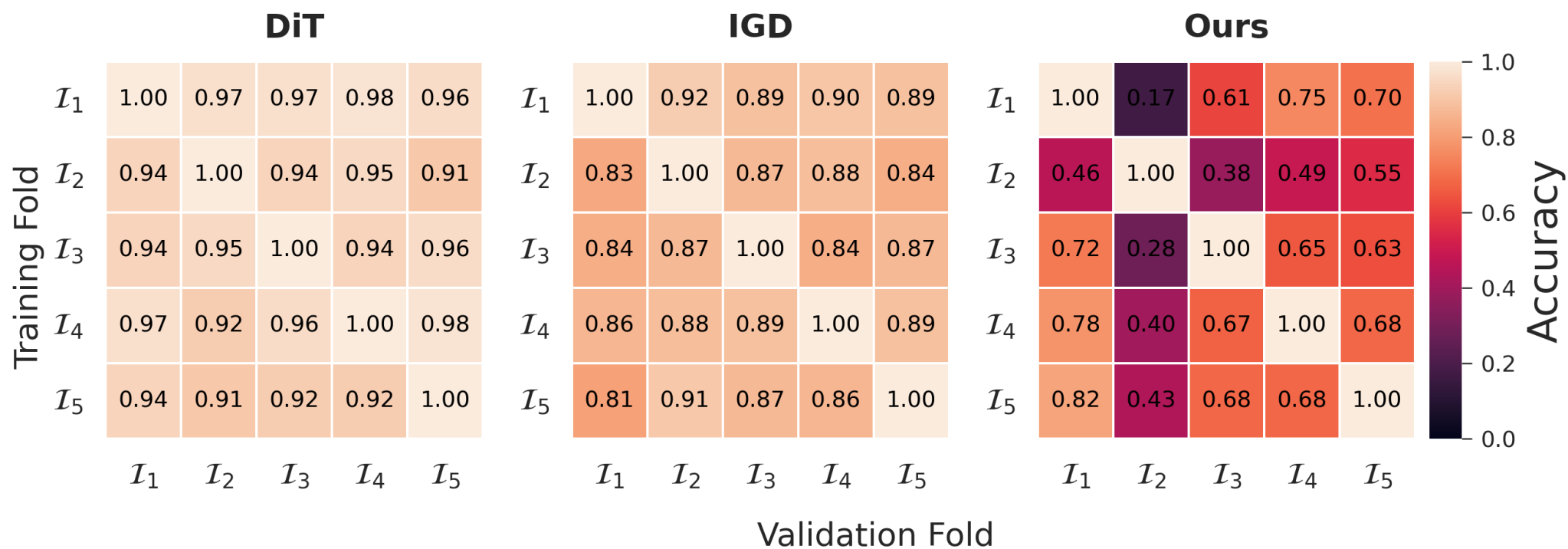
Redundant Training Signal



Redundant Training Signal



Redundant Training Signal



Problem Formulation

Instead of distilling all samples at once, partition the surrogate into disjoint increments built stage by stage.

- 1 Partition:** Split the distill dataset into K disjoint increments, each adding new IPC samples.

$$D_S = I_0 \cup I_1 \cup \dots \cup I_k$$

- 2 Cumulative Training:** At stage i , train model θ_i on the accumulated dataset D_i .

$$D_i = \cup_{n=1}^i I_n \quad \theta_i = \text{train}(\theta_{i-1}, D_i)$$

- 3 Maximize Training Signal:** Find the increment that maximizes the current model's ability to learn.

$$I_i^* = \arg \max_I \mathcal{L}(\theta_{i-1}, I)$$

- 4 Learnability:** Use a reference model to restrict to learnable samples.

$$I_i^* = \arg \max_I [\mathcal{L}(\theta_{i-1}, I) - \mathcal{L}(\theta^*, I)]$$

Learnability Guidance

Learnability Score: High S indicates the current model struggles while the reference model does not.

$$S(x, y) = \mathcal{L}(\theta_i, x, y) - \omega \cdot \mathcal{L}(\theta^*, x, y)$$

Learnability Guidance: Push current generation toward learnable samples

$$\widetilde{\epsilon}_\phi(x_t, t, y) = \widehat{\epsilon}_\phi(x_t, t, y) + \lambda \cdot \nabla_{x_t} S(x_t, y)$$

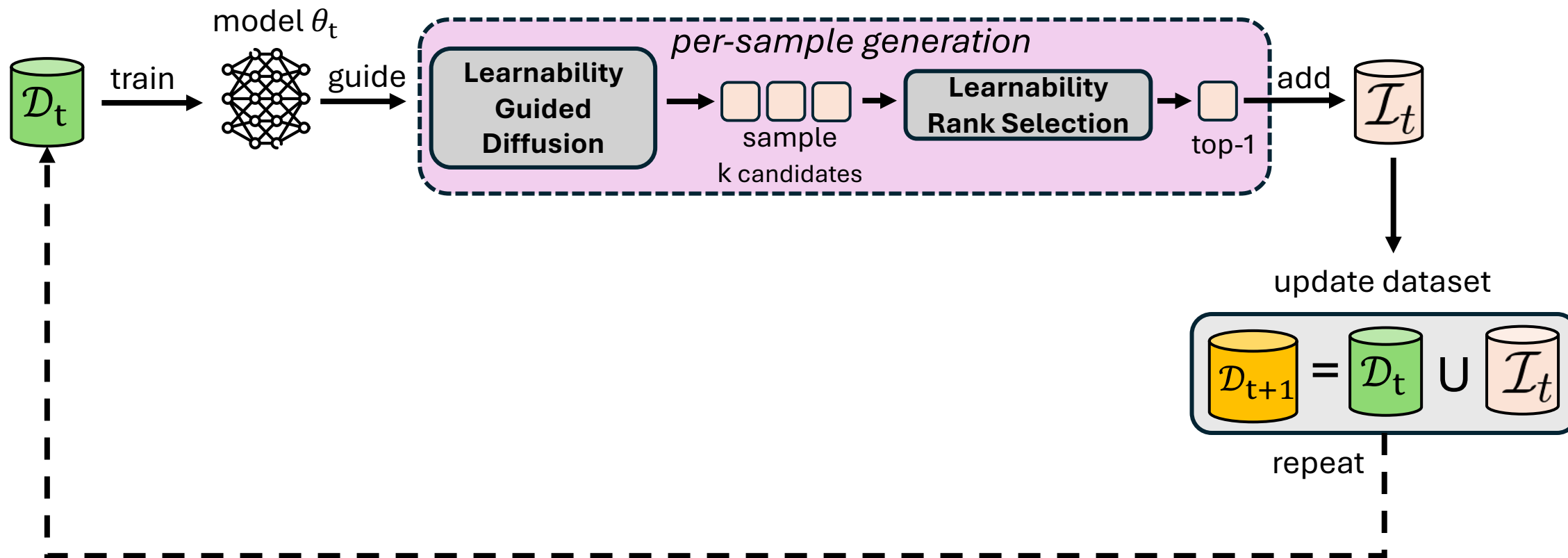
Learnability Rank Selection

Learnability Score: High S indicates the current model struggles while the reference model does not.

$$S(x, y) = \mathcal{L}(\theta_i, x, y) - \omega \cdot \mathcal{L}(\theta^*, x, y)$$

Learnability Rank Selection: Sample k samples and choose the Top-1 based on learnability score.

Incremental Distillation Loop



Qualitative Results

DIT



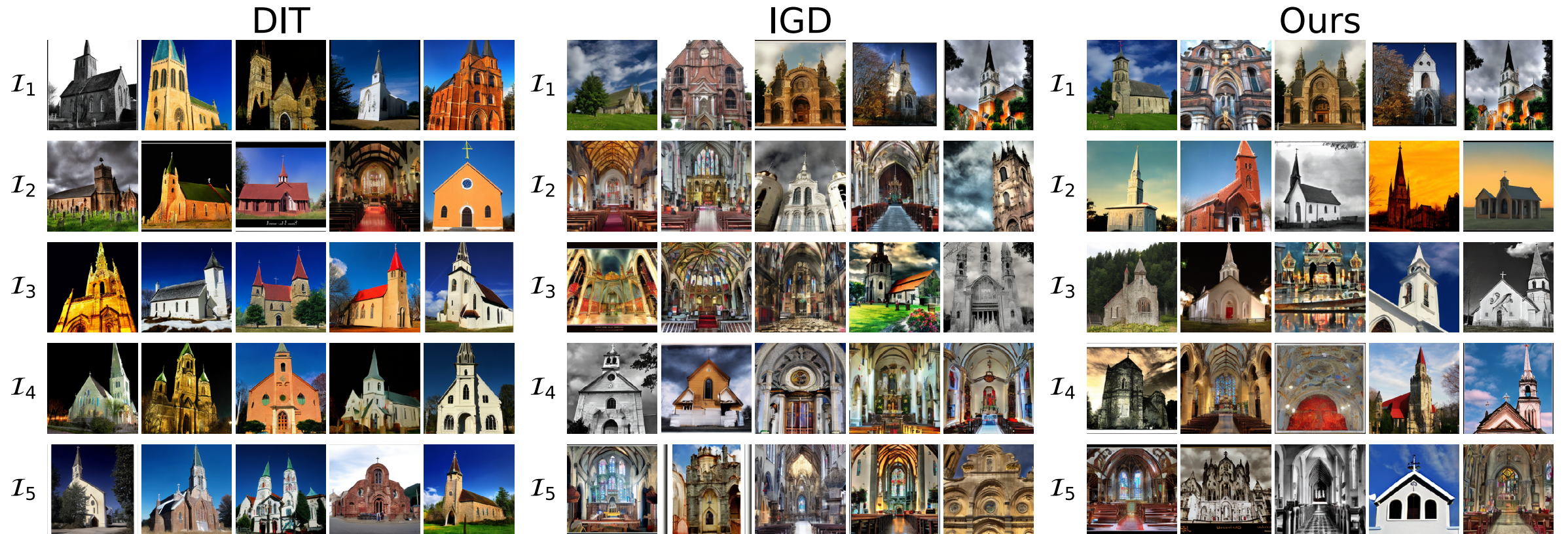
Qualitative Results

DIT

IGD



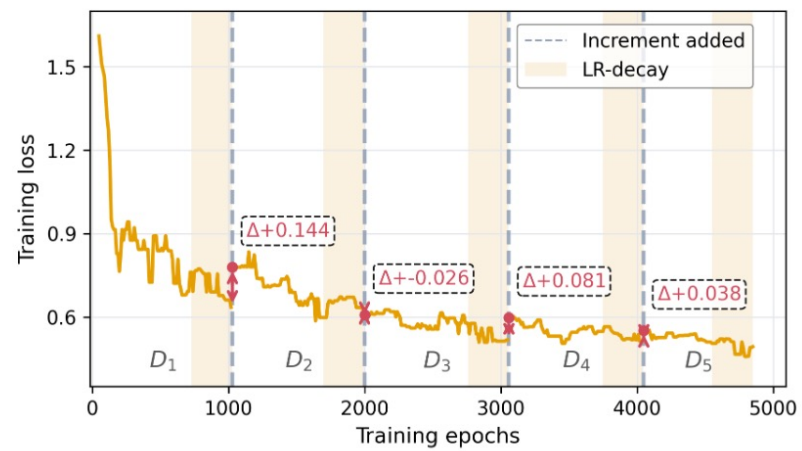
Qualitative Results



Results

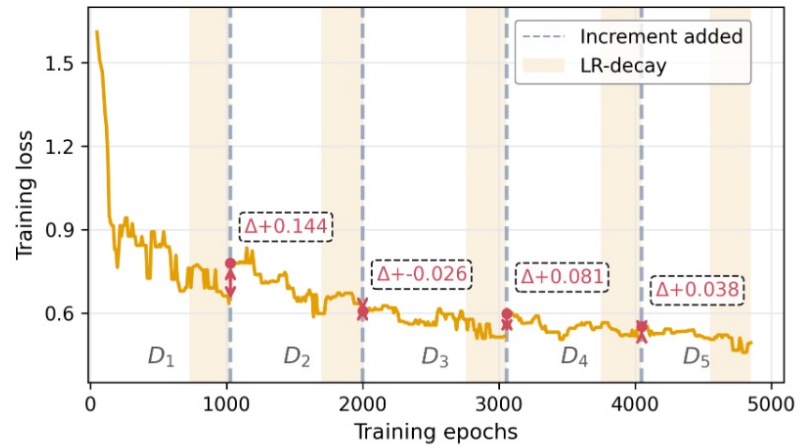
Dataset	Model	IPC	Random	DM [37]	IDC-1 [17]	DiT [23]	Minimax [10]	IGD [5]	MGD ³ [4]	LGD (Ours)	Full
Nette	ConvNet-6	50	71.8±1.2	70.3±0.8	72.4±0.7	74.1±0.6	76.9±0.9	80.9±0.9	80.9±2.3	82.6±0.7	94.3±0.5
		100	79.9±0.8	78.5±0.8	80.6±1.1	78.2±0.3	81.1±0.3	84.5±0.7	86.5±0.9	87.2±0.7	
	ResNetAP-10	50	77.3±1.0	76.7±1.0	77.4±0.7	76.9±0.5	78.2±0.7	81.0±1.2	81.2±1.0	84.3±0.5	94.6±0.5
		100	81.1±0.6	80.9±0.7	81.5±1.2	80.1±1.1	81.3±0.9	85.2±0.5	85.5±1.0	87.2±0.9	
	ResNet-18	50	75.8±1.1	75.0±1.0	77.8±0.7	75.2±0.9	78.1±0.6	81.0±0.7	81.5±3.4	85.0±0.9	95.3±0.6
		100	82.0±0.4	81.5±0.4	81.7±0.8	77.8±0.6	81.3±0.7	84.4±0.8	85.6±0.2	86.9±0.6	
Woof	ConvNet-6	50	41.9±1.4	43.8±1.1	42.6±0.9	48.5±1.3	50.7±1.8	54.2±0.7	53.4±0.4	53.9±2.2	85.9±0.4
		100	52.3±1.5	50.1±0.9	51.0±1.1	54.2±1.5	57.1±1.9	61.1±1.0	59.2±0.9	61.9±2.0	
	ResNetAP-10	50	50.1±1.6	47.8±1.2	49.3±0.9	55.8±1.1	59.8±0.8	62.7±1.2	59.4±0.2	62.5±0.7	87.2±0.6
		100	59.2±0.9	59.8±1.3	56.4±0.5	62.5±0.9	66.8±1.2	69.7±0.9	66.1±0.8	71.1±0.8	
	ResNet-18	50	54.0±0.8	53.9±0.7	54.5±1.0	57.4±0.7	60.5±0.5	62.0±1.1	63.9±0.3	65.1±0.7	89.0±0.6
		100	63.6±0.5	64.9±0.7	57.7±0.8	62.3±0.5	67.4±0.7	70.6±1.8	71.3±0.5	72.9±0.6	

Results

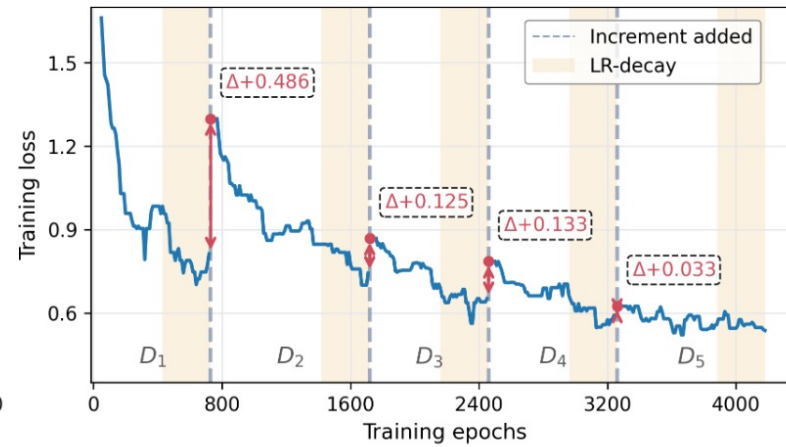


(a) DiT Incremental Training

Results

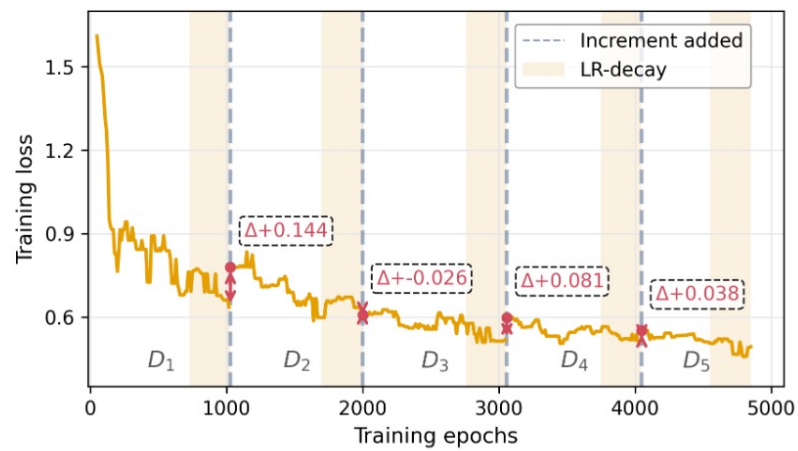


(a) DiT Incremental Training

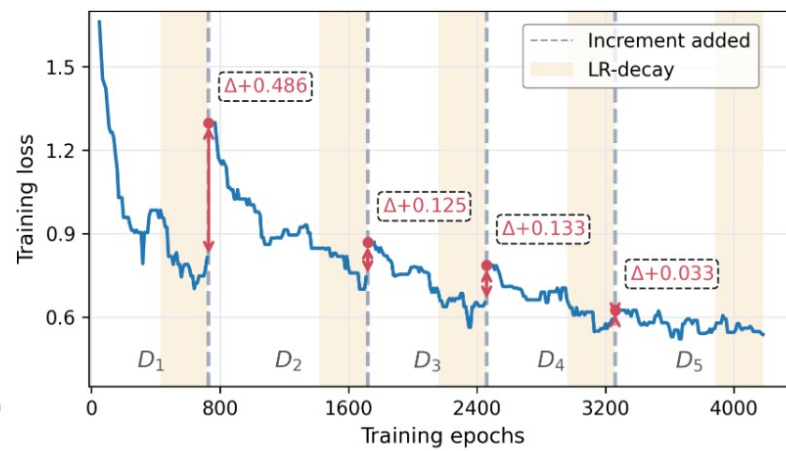


(b) Ours Incremental Training

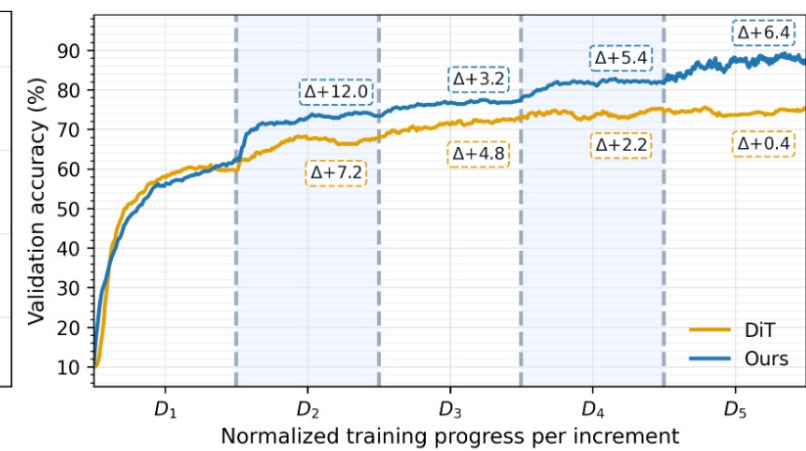
Results



(a) DiT Incremental Training

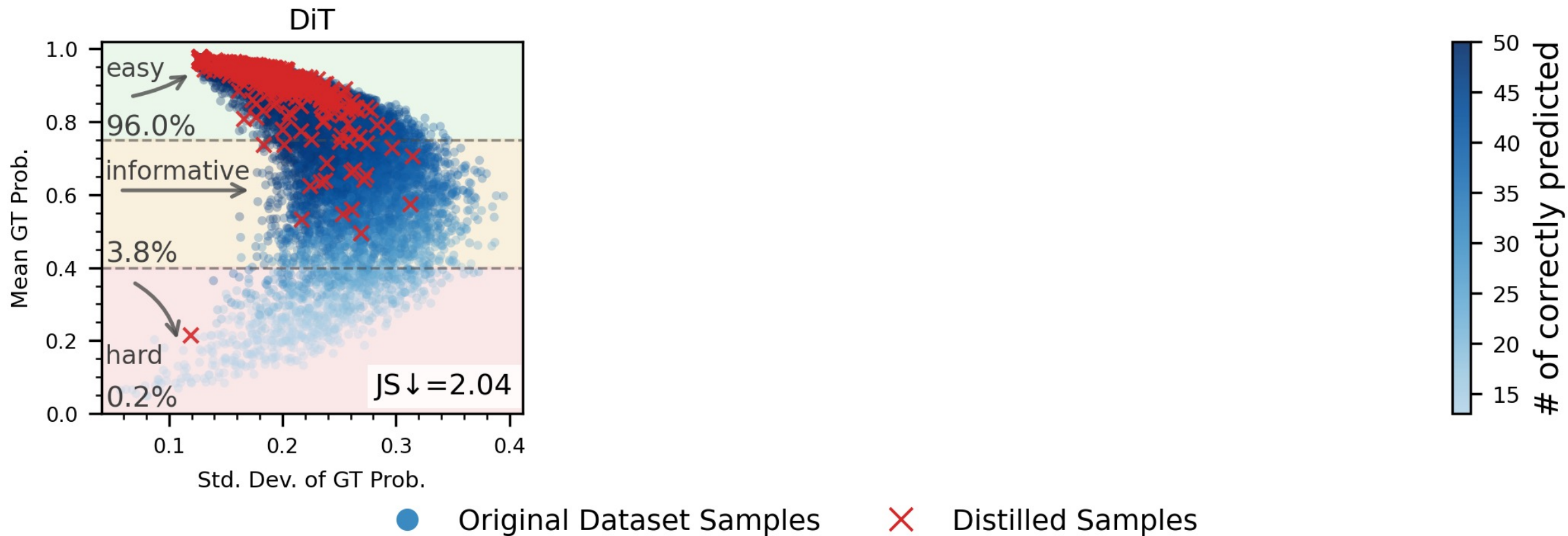


(b) Ours Incremental Training

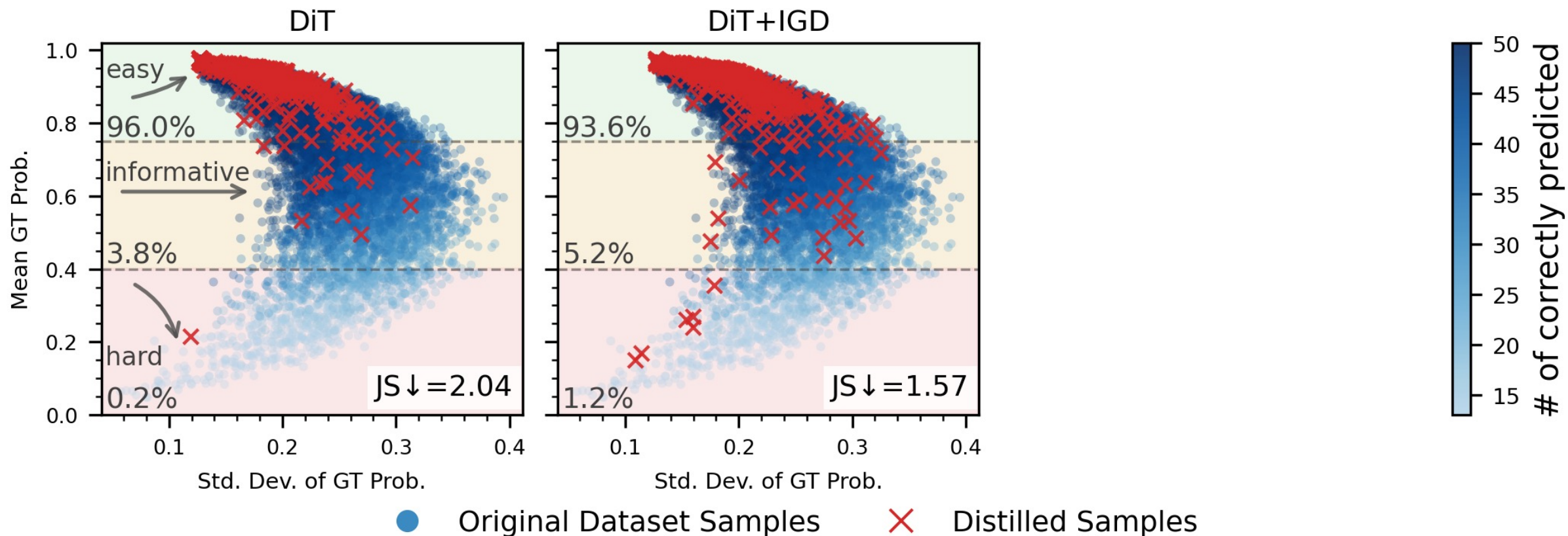


(c) Validation Accuracy Progression Comparison

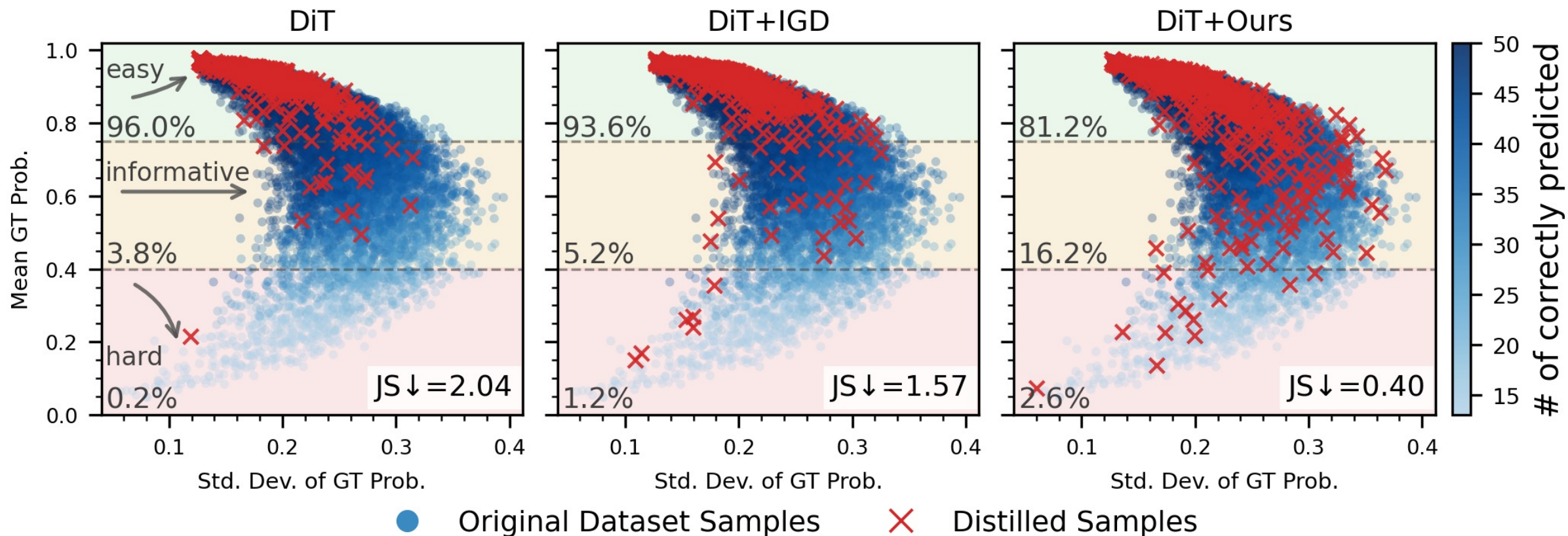
Measuring Training Dynamics of Distilled Dataset



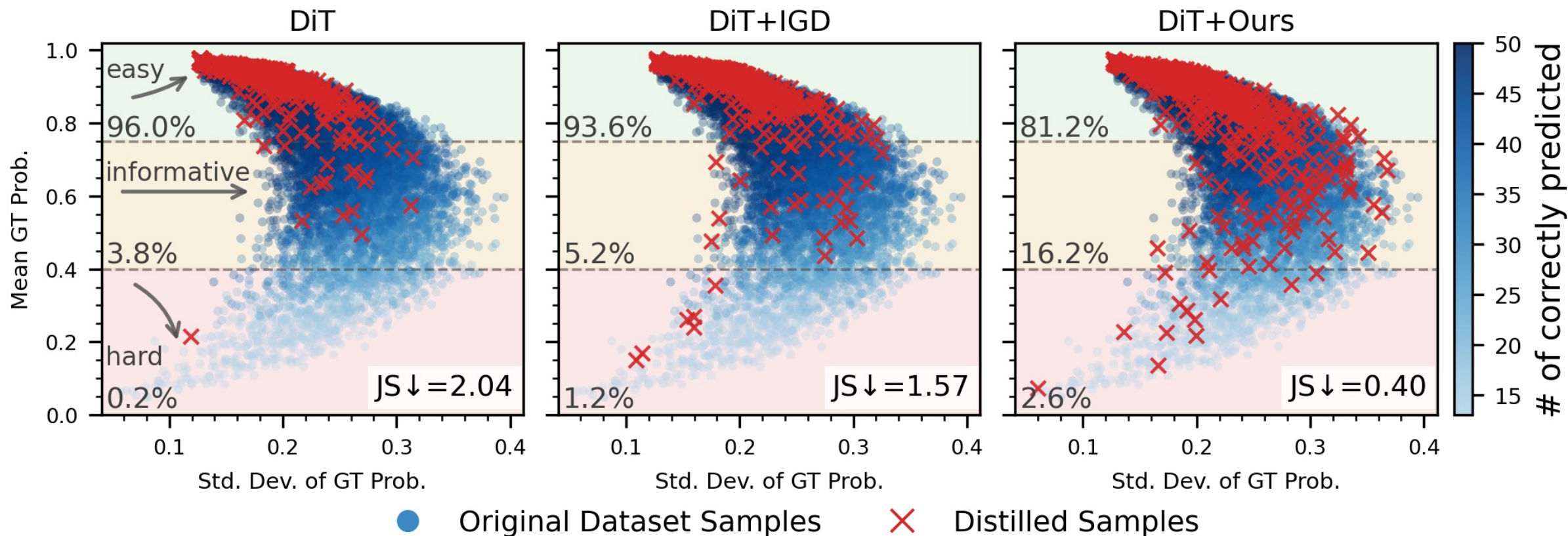
Measuring Training Dynamics of Distilled Dataset



Measuring Training Dynamics of Distilled Dataset



Measuring Training Dynamics of Distilled Dataset



Impact: Ours produce 5x informative samples

Conclusion

- ❖ Learnability-driven incremental Framework
- ❖ Learnability Guided Diffusion
- ❖ Training Signal Redundancy Analysis across distilled Increments
- ❖ State-of-the-art or competitive performance with improved sample efficiency