Four eyes see more than two: Dataset Distillation with Mixture-of-Experts

BMVC 2025

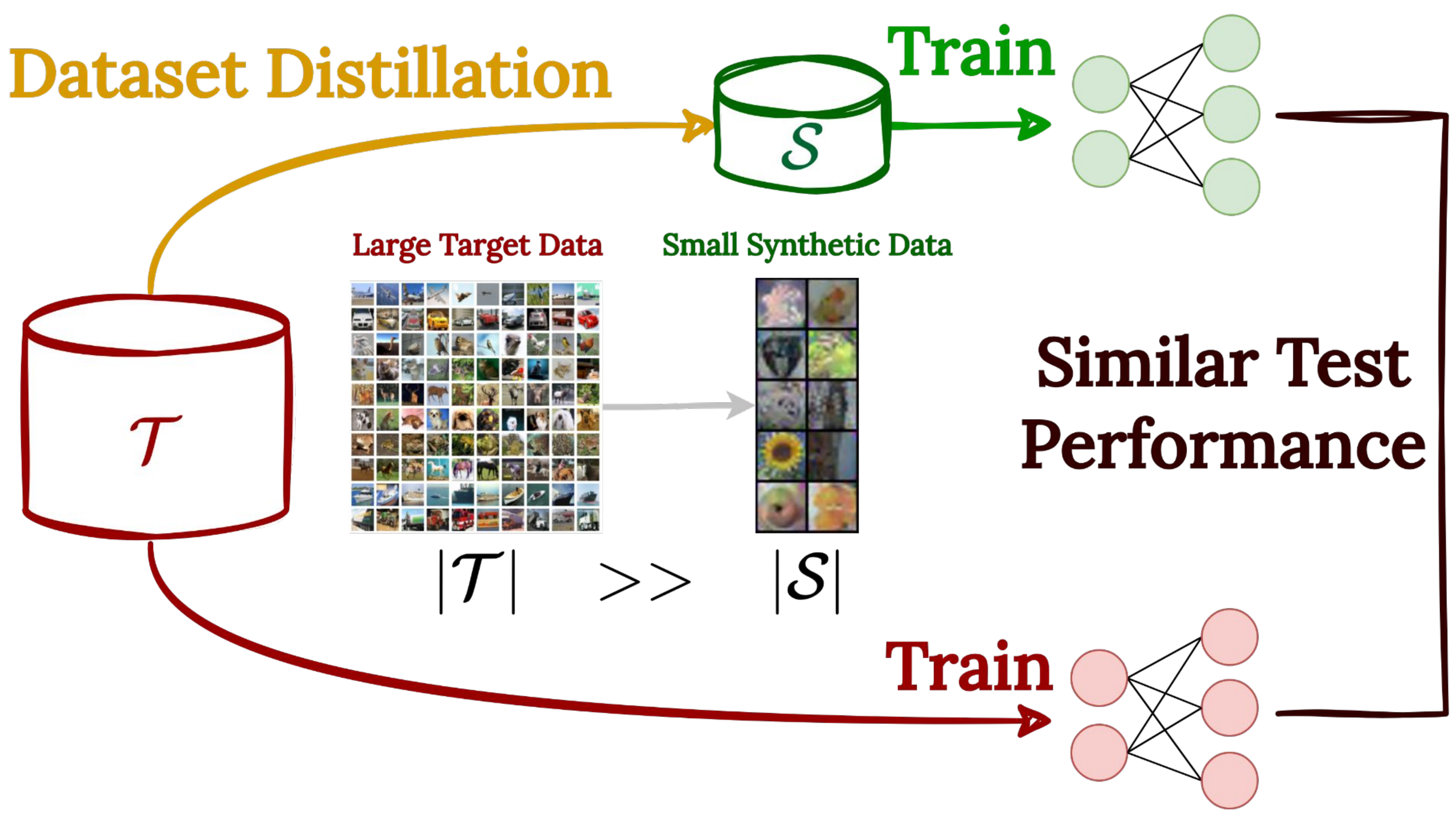
Jia-Jiun Yao[1], Sheng-Feng Yu[12], Wei-Chen Chiu[1]
[1]National Yang Ming Chiao Tung University [2]Macronix International Co., Ltd.

國立陽明交通大學
NATIONAL YANG MING CHIAO TUNG UNIVERSITY

MXIC
Macronix International Co., Ltd.

Dataset Distillation

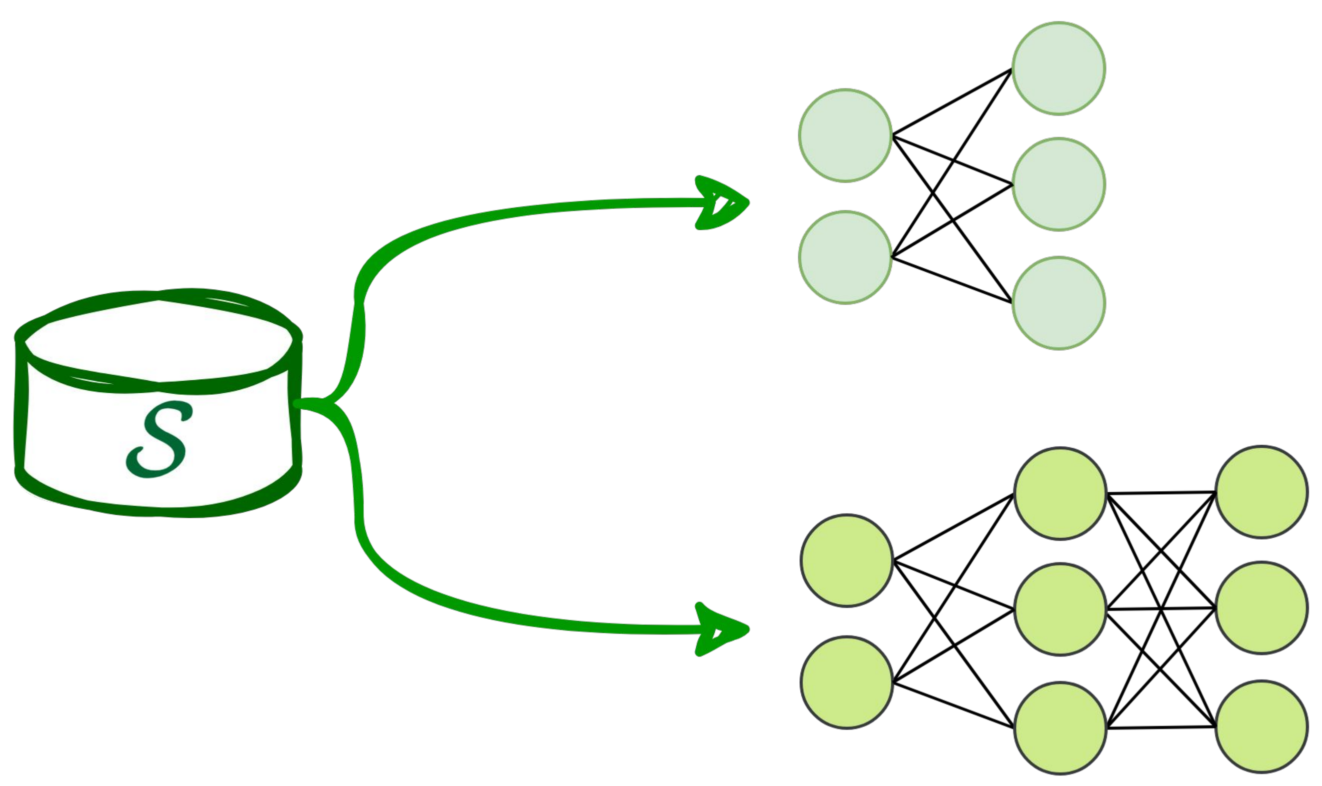
Dataset Distillation aims to generate a compact, synthetic training dataset that retains the training effectiveness of the target dataset.



A distillation expert is defined by an objective function and target dataset.

$$\mathcal{S} = \arg\min_{\mathcal{S}} \mathcal{L}(\mathcal{S}; \mathcal{T})$$

Once dataset distillation is complete, the synthetic dataset is available for training various networks.



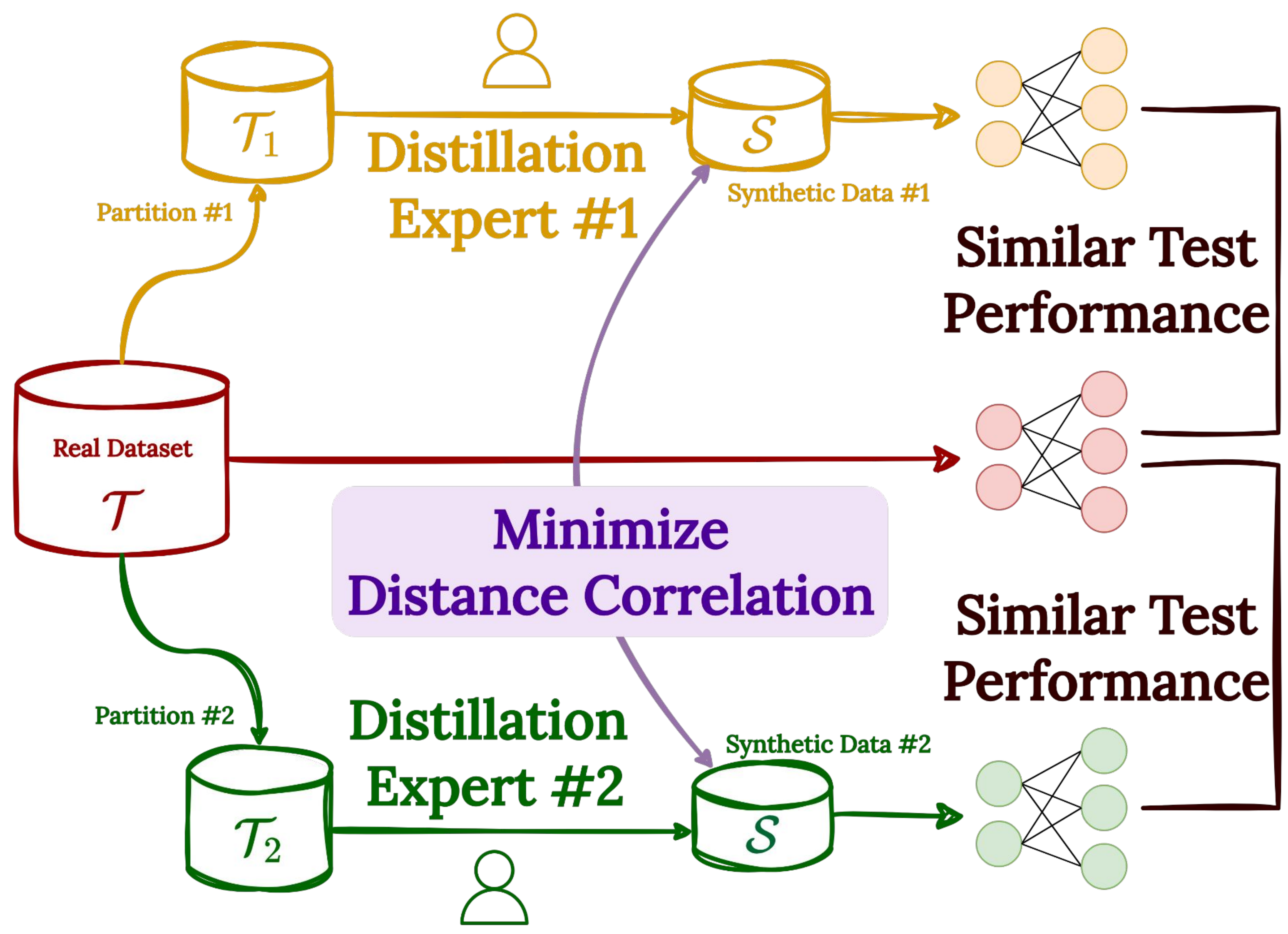
Cross-Architecture Performance

Existing dataset distillation methods often suffer from performance degradation when transferring distilled datasets across network architectures.

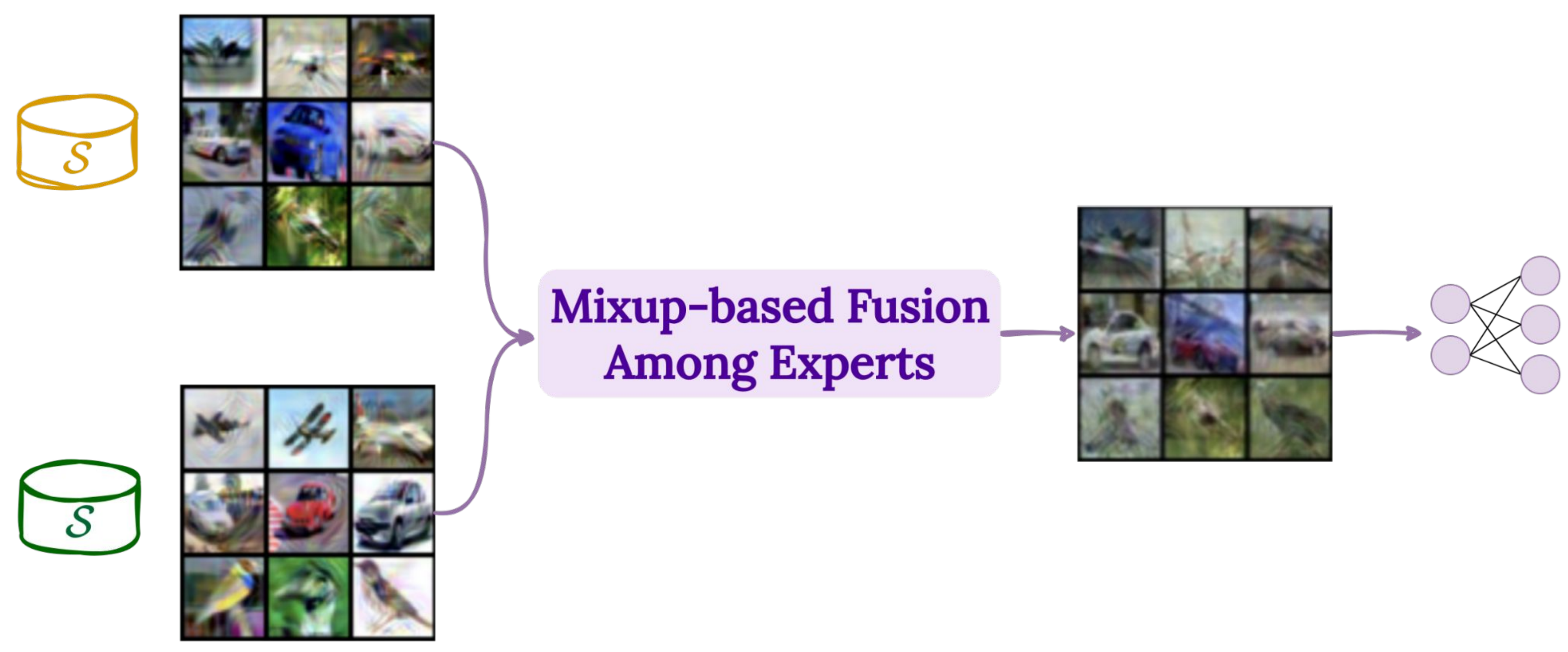
The proposed novel mixture-of-experts framework for dataset distillation enhances cross-architecture performance through increased diversity of distilled dataset.

Dataset Distillation with Mixture-of-Experts

Distillation Stage: Fostering Diversity Through Partitioned Data and Multiple Experts



Application Stage: Enhance distilled dataset transferability and performance via mixup.



Experimental Results

Target Budget IPCxNoE	CIFAR-10 10 5x2	CIFAR-10 10 10x1	CIFAR-10 20 10x2	CIFAR-10 20 20x1	CIFAR-100 10 5x2	CIFAR-100 10 10x1	CIFAR-100 20 10x2	CIFAR-100 20 20x1
Gradient Matching (IDC) [1]								
ConvNet-3	**53.32**	52.73	**55.52**	54.05	**31.44**	31.03	**36.03**	35.33
VGG11	**51.41**	49.70	**53.37**	51.74	**29.62**	27.38	**34.53**	31.25
ResNet18	**51.56**	47.51	**54.06**	50.06	**30.74**	28.59	**36.08**	34.09
AlexNet	**46.34**	40.69	**52.01**	44.04	**26.45**	20.02	**33.84**	25.76
Distribution Matching (IDM) [2]								
ConvNet-3	**49.04**	47.61	**52.31**	50.48	26.87	**27.45**	**32.38**	31.67
VGG11	44.53	**44.65**	**48.31**	47.70	**23.06**	22.90	28.57	**28.84**
ResNet18	**45.05**	42.84	**49.52**	46.02	**24.29**	25.35	**32.08**	32.04
AlexNet	**37.19**	34.96	**44.32**	41.71	**19.49**	17.07	**27.16**	22.59
Trajectory Matching (MTT) [3]								
ConvNet-3	53.82	**56.17**	60.14	**60.82**	**30.00**	29.53	28.50	**29.80**
VGG11	**49.86**	46.59	**52.04**	47.27	**29.23**	27.99	**27.80**	27.72
ResNet18	**52.14**	46.17	**58.13**	53.94	**30.19**	27.81	**27.55**	26.05
AlexNet	**44.16**	32.97	**52.11**	50.01	**25.27**	23.88	**28.10**	27.67

※ IPC: Image Per Classes NoE: Number of Expert

[1] Jang-Hyun Kim et al. Dataset condensation via efficient synthetic-data parameterization. In International Conference on Machine Learning (ICML), 2022
[2] Ganlong Zhao, Guanbin Li, Yipeng Qin, and Yizhou Yu. Improved distribution matching for dataset condensation. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2023
[3] George Cazenavette et al. Dataset distillation by matching training trajectories. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022