Hyperbolic Dataset Distillation

Wenyuan Li, Guang Li, Keisuke Maeda, Takahiro Ogawa, Miki Haseyama
Hokkaido University

Introduction

Background:

Dataset distillation (DD) aims to compress a large training set into a tiny synthetic core set that preserves the performance of training on the full data. Existing optimization-based DD methods are accurate but computationally expensive, while distribution-matching (DM) methods are more efficient but often less accurate. Our goal is to improve DM-based DD without changing the basic training pipeline.

Current Method Drawback:

- Existing dataset distillation (DM) methods are confined to Euclidean space, treating samples as i.i.d. points and failing to capture complex geometric and hierarchical relationships in data.

Our Method:

- We propose Hyperbolic Dataset Distillation (HDD), which performs distillation in hyperbolic space to naturally model tree-like and hierarchical structures.

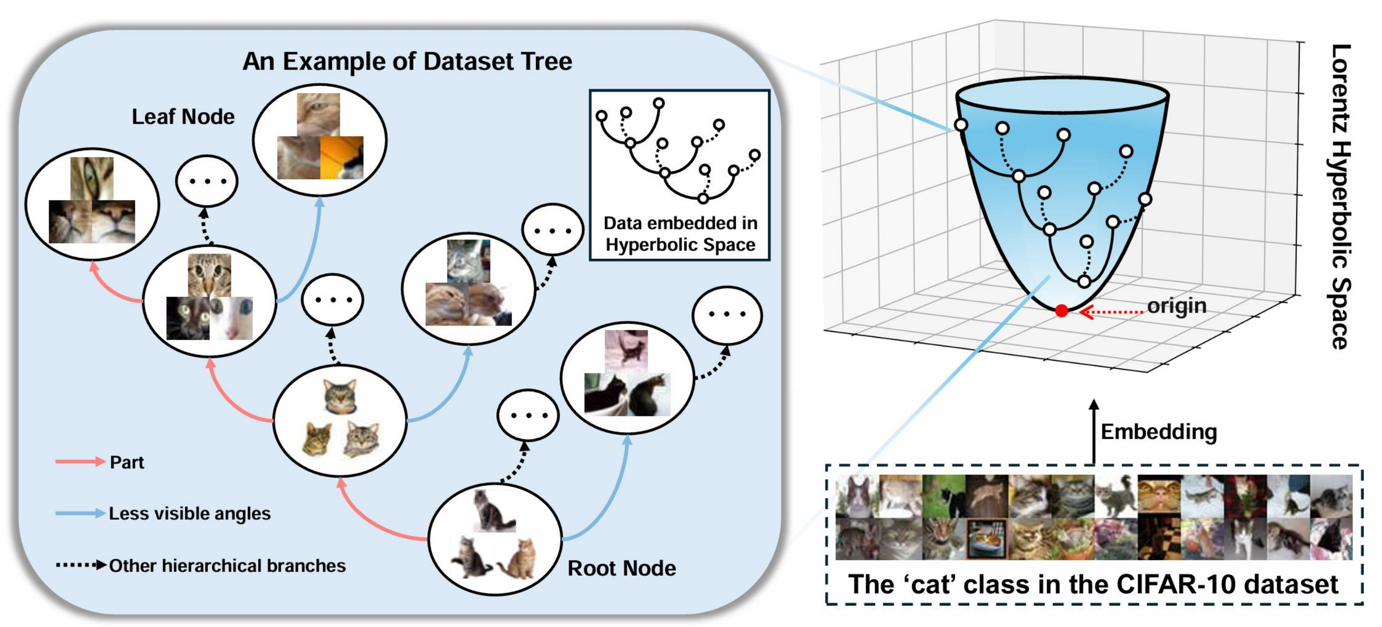


Figure 1: An example of hierarchical representation in hyperbolic space using the 'Cat' class from the CIFAR-10 Dataset.

Method

Method Overview

- Embed the dataset into hyperbolic space via exponential mapping, yielding a hierarchical representation.
- Assign level-dependent weights to samples to reflect their importance in the global geometry.
- Compute hyperbolic centroids of the original and synthetic datasets and measure their geodesic distance.
- Use this distance as a loss to update the synthetic data, aligning it with the class-specific prototypes of the original data.

Method

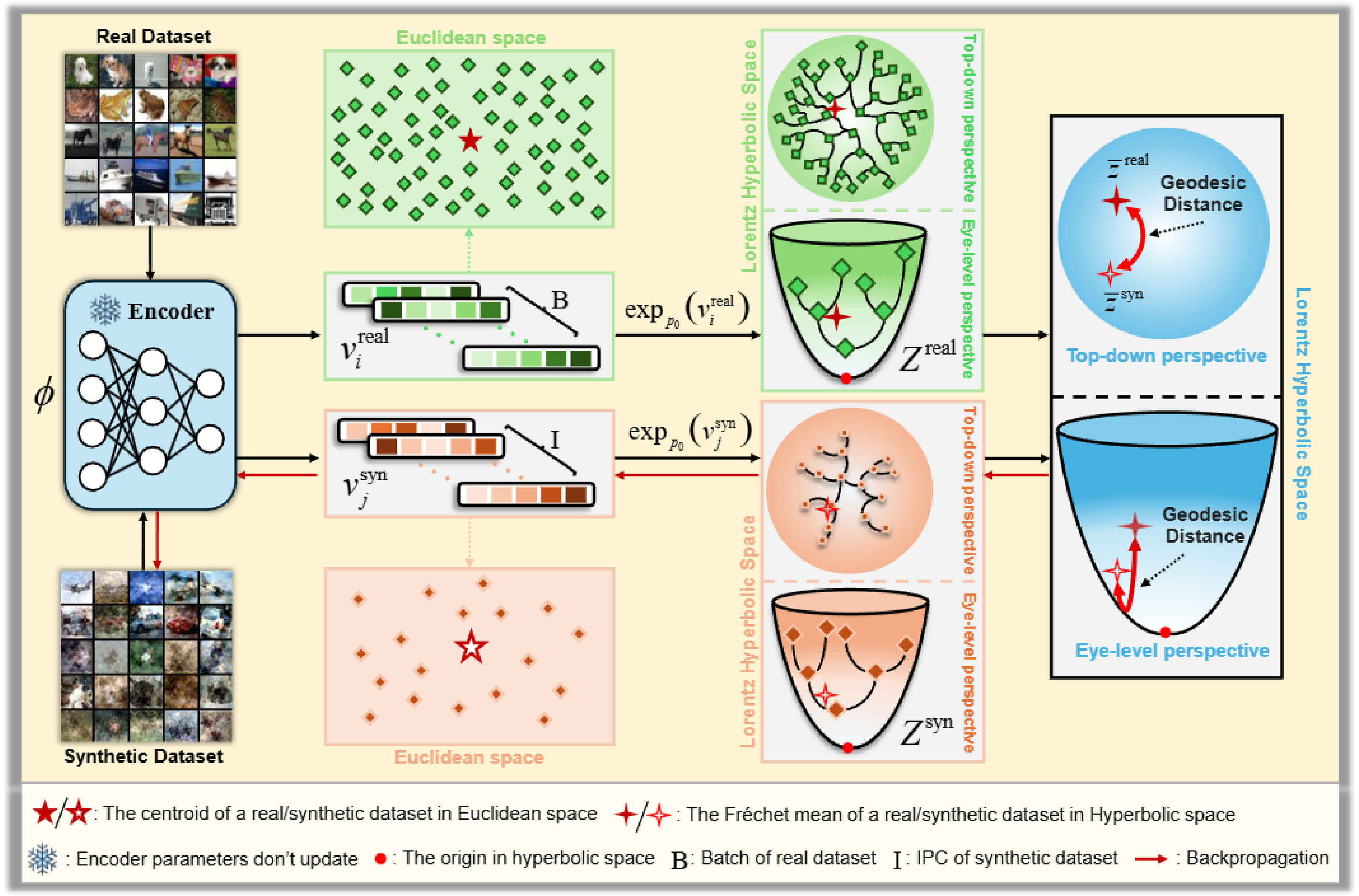


Figure 2: The framework of hyperbolic dataset distillation.

Method Steps：

1. Hyperbolic Embedding via Exponential Map

Feature vectors v extracted from the encoder are mapped from Euclidean space into the Lorentz hyperbolic manifold L_K^n using the exponential map at the origin p_0. This enables the model to capture latent hierarchies:

$$z_i = \exp_{p_0}(v_i) = \cosh(\sqrt{-K}\|v_i\|)p_0 + \sinh(\sqrt{-K}\|v_i\|)\frac{v_i}{\sqrt{-K}\|v_i\|} \quad ,$$

where K is the negative curvature.

2. Riemannian Centroid Calculation

Instead of a simple Euclidean average, HDD computes the Riemannian mean (centroid) for both the real dataset batch (Z^{real}) and the synthetic dataset (Z^{syn}). Due to hyperbolic geometry properties, this centroid is naturally biased towards samples closer to the origin (prototypes):

$$\mathbf{c} = \sqrt{-K} \cdot \frac{\bar{\mathbf{z}}}{\sqrt{|\langle \bar{\mathbf{z}}, \bar{\mathbf{z}} \rangle_{\mathcal{L}}| + \epsilon}}, \quad \text{where} \quad \bar{\mathbf{z}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{z}_i.$$

3. Geodesic Distance Minimization (Loss Function)

The distillation objective is to minimize the Lorentzian hyperbolic distance (geodesic distance) between the centroid of the real data $\overline{z^{real}}$ and the centroid of the synthetic data ($\overline{z^{syn}}$). This aligns the global geometric distributions of the two datasets:

$$\mathcal{L}_{\mathrm{Lhd}} = \lambda\, d_L(\bar{z}^{\mathrm{real}}, \bar{z}^{\mathrm{syn}}) = \frac{\lambda}{\sqrt{-K}} \operatorname{acosh}\Big(-K \,\langle \bar{z}^{\mathrm{real}}, \bar{z}^{\mathrm{syn}} \rangle_{\mathcal{L}}\Big),$$

where d_L is the hyperbolic distance, and $\langle\cdot,\cdot\rangle_{\mathcal{L}}$ denotes the Lorentzian inner product.

Setting

- **Datasets**
We evaluate HDD on Fashion-MNIST, SVHN, CIFAR-10, CIFAR-100, TinyImageNet, and the higher-resolution ImageWoof subset of ImageNet (for hybrid architectures).
- **Network Architectures**
We use the standard Depth-3 ConvNet from DC/DM/IDM as our main backbone.
- **Implementation Details**
We follow the hyperparameter settings of DM/IDM/Dance, use DSA for augmentation, train synthetic data with SGD, and tune curvature K, scaling factor λ, and learning rate r per experiment.

Results

Table 1: Comparison of different methods on the FashionMNIST, SVHN, CIFAR10, and CIFAR100 datasets with IPC = 1, 10, and 50.

Method	FashionMNIST			SVHN			CIFAR10			CIFAR100		
IPC Ratio (%)	1 0.017	10 0.17	50 0.83	1 0.014	10 0.14	50 0.7	1 0.02	10 0.2	50 1	1 0.2	10 2	50 10
Random	51.4±3.8	73.8±0.7	82.5±0.7	14.6±1.6	35.1±4.1	70.9±0.9	14.4±2.0	26.0±1.2	43.4±1.0	4.2±0.3	14.6±0.5	30.0±0.4
Herding	67.0±1.9	71.1±0.7	71.9±0.8	20.9±1.3	50.5±3.3	72.6±0.8	21.5±1.2	31.6±0.7	40.4±0.6	8.4±0.3	17.3±0.3	33.7±0.5
K-Center	66.9±1.8	54.7±1.5	68.3±0.8	21.0±1.5	14.0±1.3	20.1±1.4	21.5±1.3	14.7±0.7	27.0±1.4	8.3±0.3	7.1±0.2	30.5±0.3
Forgetting	-	-	-	12.1±5.6	16.8±1.2	27.2±1.5	13.5±1.5	23.3±1.0	23.3±1.1	4.5±0.3	15.1±0.2	30.5±0.4
DC [69]	70.5±0.6	82.3±0.4	83.6±0.4	31.2±1.4	76.1±0.6	82.3±0.3	28.3±0.5	44.9±0.5	53.9±0.5	12.8±0.3	25.2±0.3	-
DSA [67]	70.6±0.6	**84.6±0.3**	88.7±0.3	27.5±1.4	79.2±0.5	84.4±0.4	28.8±0.7	52.1±0.5	60.6±0.5	13.9±0.3	32.3±0.3	42.8±0.4
CAFE [55]	77.1±0.9	83.0±0.4	84.8±0.4	42.6±3.3	75.9±0.6	81.3±0.3	30.3±1.1	46.3±0.6	55.5±0.6	12.9±0.3	27.8±0.3	37.9±0.3
CAFE+DSA [55]	73.7±0.7	83.0±0.3	88.2±0.3	42.9±3.0	77.9±0.6	82.3±0.4	31.6±0.8	50.9±0.5	62.3±0.4	14.0±0.3	31.5±0.2	42.9±0.2
DCC [28]	-	-	-	34.3±1.6	76.2±0.8	83.3±0.2	34.0±0.7	54.4±0.5	64.2±0.4	14.6±0.3	33.5±0.3	39.4±0.4
G-VBSM [49]	-	-	-	-	-	-	-	46.5±0.7	54.3±0.3	16.4±0.7	38.7±0.2	45.7±0.4
DataDAM [47]	-	-	-	-	-	-	32.0±1.2	54.2±0.8	67.0±0.4	14.5±0.5	34.8±0.5	**49.4±0.3**
DM [68]	70.7±0.6	83.4±0.1	88.1±0.6	21.9±0.4	72.8±0.3	82.6±0.3	26.4±0.3	48.5±0.6	62.2±0.5	11.4±0.3	29.7±0.3	43.0±0.4
DM with HDD	72.1±0.2	84.0±0.1	**88.8±0.4**	25.0±0.2	75.1±0.2	83.0±0.3	28.7±0.2	50.3±0.3	63.2±0.4	13.3±0.2	30.1±0.1	43.8±0.2
IDM [70]	77.4±0.3	82.4±0.2	84.5±0.1	65.3±0.3	81.0±0.1	85.2±0.3	45.2±0.5	57.3±0.3	67.2±0.1	22.1±0.2	44.7±0.3	46.5±0.4
IDM with HDD	**78.5±0.2**	83.8±0.2	86.4±0.3	**67.8±0.2**	**84.0±0.2**	**87.6±0.1**	**47.0±0.1**	**61.3±0.1**	**69.7±0.2**	**25.3±0.2**	**45.4±0.1**	48.9±0.3
Whole Dataset		93.5±0.1			95.4±0.1			84.8±0.1			56.2±0.3	

Our method achieves consistent performance improvements over the mainstream distribution matching methods, DM and IDM, across all datasets, demonstrating the effectiveness of the HDD approach.

Table 3: Comparison of different methods on the CIFAR10, CIFAR100, and ImageWoof datasets.

Method	CIFAR10			CIFAR100			ImageWoof	
IPC Ratio (%)	1 0.02	10 0.2	50 1	1 0.2	10 2	50 10	1 0.11	10 1.10
DATM [21]	46.9±0.5	66.8±0.2	76.1±0.3	27.9±0.2	47.2±0.4	55.0±0.2	-	-
RDED [53]	23.5±0.3	50.2±0.3	68.4±0.1	19.6±0.3	48.1±0.3	57.0±0.1	18.5±0.9	40.6±2.0
D^4M [50]	-	56.2±0.4	72.8±0.5	-	45.0±0.1	48.8±0.3	-	-
IID (IDM) [11]	47.1±0.1	59.9±0.2	69.0±0.3	24.6±0.1	45.7±0.4	51.3±0.4	-	-
DSDM [35]	45.0±0.4	66.5±0.3	75.8±0.3	19.5±0.2	46.2±0.3	**54.0±0.2**	-	-
M3D [65]	45.3±0.3	63.5±0.2	69.9±0.5	26.2±0.3	42.4±0.2	50.9±0.7	-	-
Dance [64]	**47.2±0.3**	70.2±0.2	76.3±0.1	26.2±0.2	49.7±0.1	52.8±0.1	27.1±0.2	46.2±0.2
Dance with HDD	46.8±0.3	**70.8±0.2**	**77.1±0.2**	**27.7±0.3**	**50.2±0.2**	53.9±0.1	**27.6±0.2**	**46.6±0.1**
Whole Dataset		84.8±0.1			56.2±0.3		67.0±1.3	

On the distribution-matching-based hybrid architecture, our method can still benefit the base method and achieve performance gains.

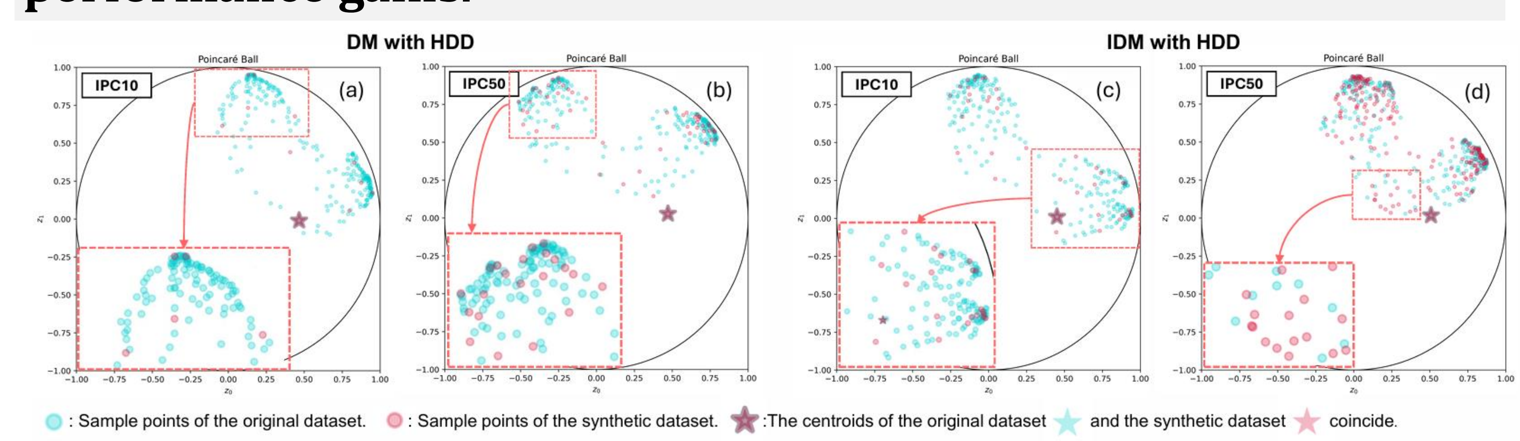


Figure 3: The distributions of the original and synthetic datasets in the Poincaré hyperbolic space are visualized.

The original CIFAR-10 data and the HDD-distilled synthetic set have almost perfectly aligned centroids on the Poincaré ball, and the sparse synthetic distribution still roughly follows the original (denser where the original is dense, sparser where it is sparse). Compared to the original data, which shows strong edge accumulation, the synthetic samples are more concentrated toward the root (center), revealing a clear root-centric bias.

This research was supported in part by JSPS KAKENHI Grant Numbers JP23K11211, JP23K21676, JP24K02942, JP24K23849, and JP25K21218.

Hyperbolic Dataset Distillation

Supplementary material

Baseline

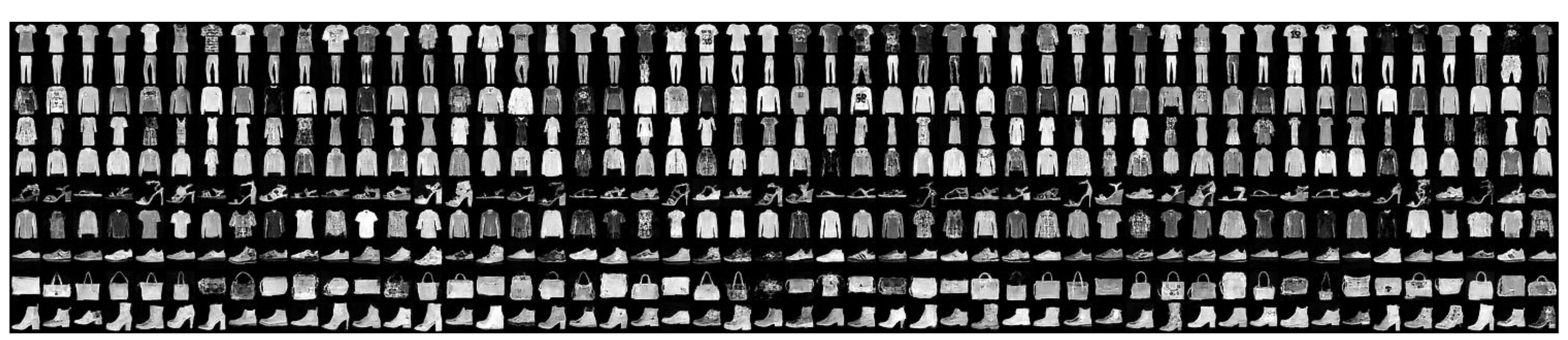
Distribution Matching (DM) [68] is the first to use maximum mean discrepancy to optimize synthetic data to match the distribution of the original data.

Improved Distribution Matching (IDM) [70] enhances DM by addressing feature imbalance through Partitioning and Expansion augmentation, and correcting invalid MMD estimation using enriched semi-trained model embeddings and class-aware distribution regularization, resulting in more accurate feature alignment and improved performance.

Dual-view distribution AligNment for dataset CondEnsation (DANCE) [64] introduces a dual-view approach to dataset condensation by leveraging expert models: it performs pseudo long-term distribution alignment via a convex combination of initialized and trained models to align inner-class distributions without persistent training, and applies distribution calibration using expert models to mitigate inter-class distribution shift and preserve class boundaries.

Distilled Images

(a) (b)



Hyperparameter Details

Table 7: Hyperparameter details of DM with HDD and IDM with HDD.

Dataset	IPC	DM with HDD			IDM with HDD		
		$-1/K$	λ	r	$-1/K$	λ	r
FashionMNIST	1	1	20	1	2	40	0.5
	10	1	40	1	2	60	1
	50	1	60	1	2	80	0.2
SVHN	1	1	10	1	2	120	0.5
	10	1	50	1	2	120	1
	50	1	100	1	2	120	0.2
CIFAR 10	1	1	1	1	3	80	0.5
	10	1	20	1	3	100	1
	50	1	80	1	3	120	0.2
CIFAR 100	1	1	10	1	2	60	0.5
	10	2	100	1	2	80	0.2
	50	2	120	1	2	100	0.6
TinyImageNet	1	-	-	-	2	80	0.5
	10	-	-	-	2	100	0.5
	50	-	-	-	2	120	0.6

Table 8: Hyperparameter details of Dance with HDD.

Dataset	IPC	DM with HDD		
		$-1/K$	λ	r
CIFAR-10	1	1.8	20	0.02
	10	0.2	40	0.2
	50	2	60	0.5
CIFAR-100	1	2	40	0.02
	10	1.5	80	0.1
	50	2	120	0.5
ImageWoof	1	0.6	100	0.1
	10	0.5	120	0.1

References

[64] Hansong Zhang, Shikun Li, Fanzhao Lin, Weiping Wang, Zhenxing Qian, and Shiming Ge. Dance: Dual-view distribution alignment for dataset condensation. *arXiv preprint arXiv:2406.01063*, 2024.

[68] Bo Zhao and Hakan Bilen. Dataset condensation with distribution matching. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 6514–6523, 2023.

[70] Ganlong Zhao, Guanbin Li, Yipeng Qin, and Yizhou Yu. Improved distribution matching for dataset condensation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7856–7865, 2023.

Contact information and related resources

Wenyuan Li： wenyuan@lmd.ist.hokudai.ac.jp

Guang Li： guang@lmd.ist.hokudai.ac.jp

GitHub

Webpage