Breaking the Synthetic-Real Domain Shortcut for Training-Free Generative Replay-based Class Incremental Learning

Tao Zhang[1,2], QiXuan Fan[1], Yiyuan Liang[1,2], Yanjie Wang[1,2], Song Yan[3], Tian Tian[1,2], Jiahuan Zhou[4], Luxin Yan[1,2], Sheng Zhong[1,2], Xu Zou*[1,2]

[1] Huazhong University of Science and Technology
[2] State Key Laboratory of Multispectral Information Intelligent Processing Technology
[3] University of Science and Technology of China
[4] Wangxuan Institute of Computer Technology, Peking University

Contact

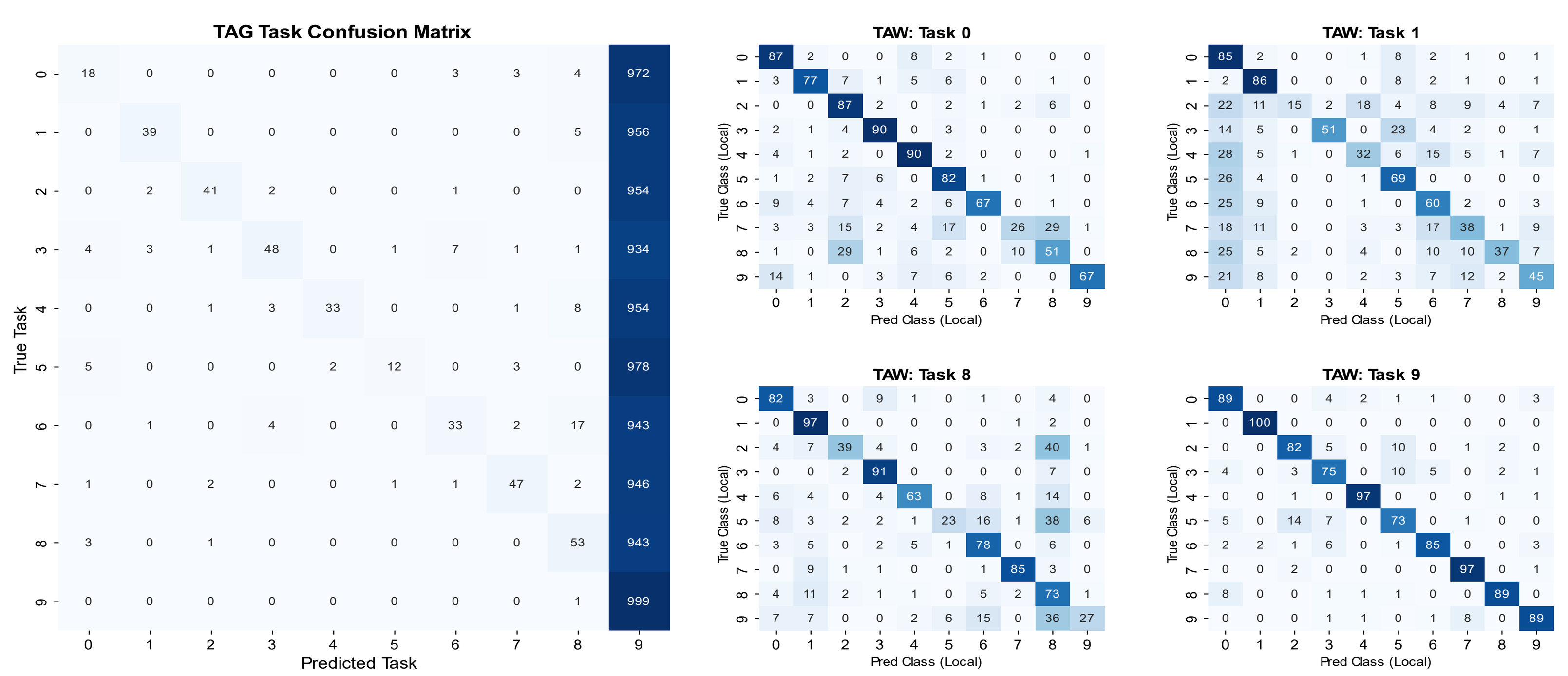
zhangtao_2023@hust.edu.cn

Motivation

- **Class Incremental Learning (CIL) requires models to learn new classes without catastrophically forgetting old ones. Training-free generative replay appears ideal for EFCIL: frozen T2I models synthesize old-class data without storing real exemplars or retraining generators.**

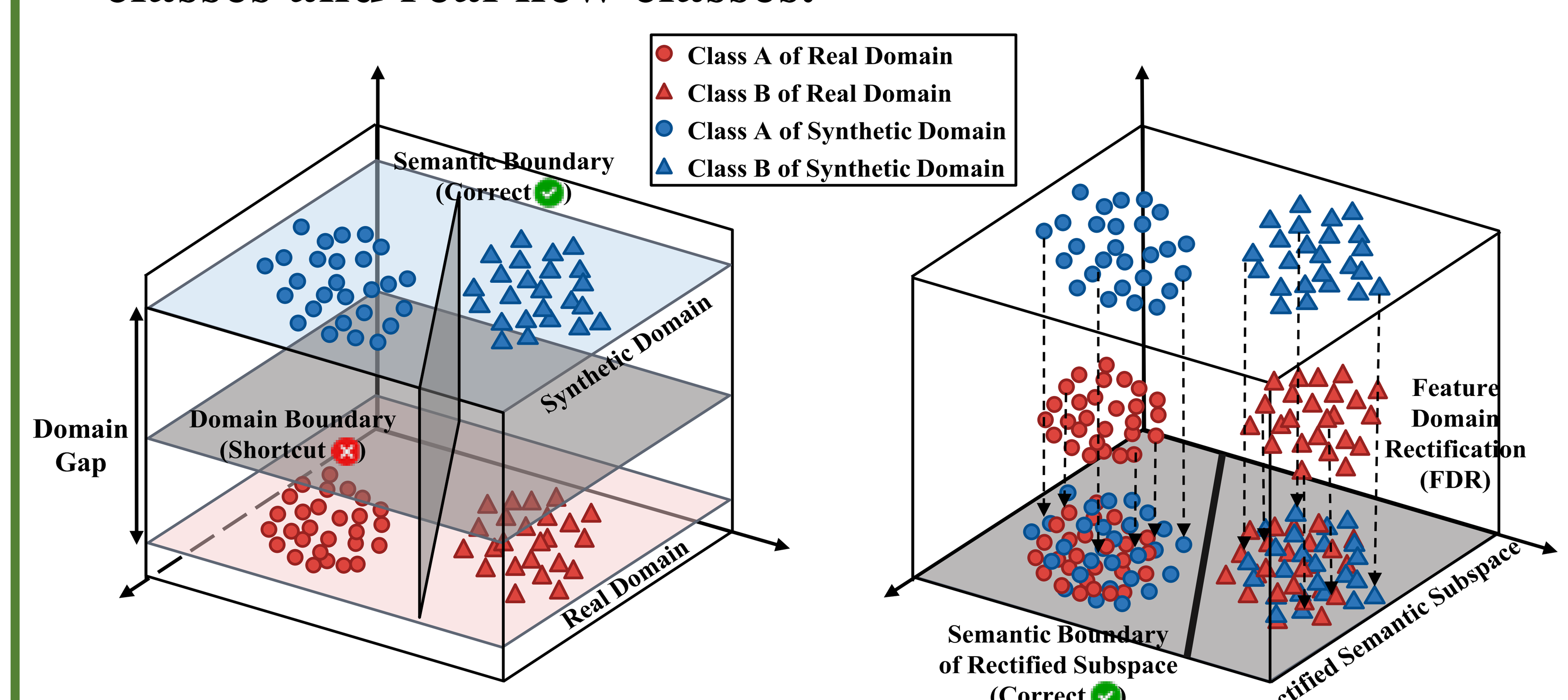
Key Observation:

- **TAW remains high: synthetic data preserves class semantics**
- **TAG collapses: real old-class data are misclassified into the current task.**



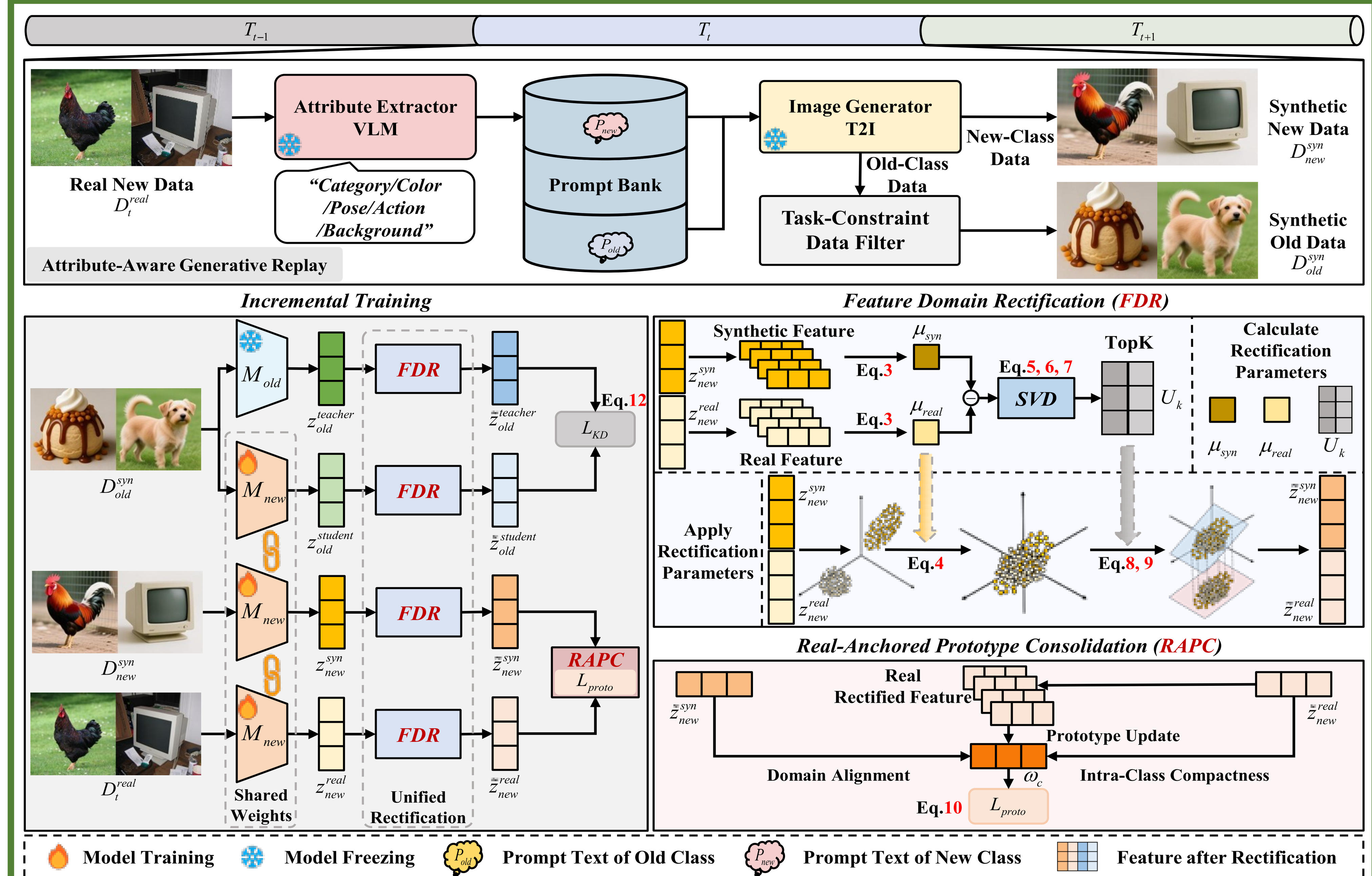
Core Problem: Synthetic–Real Domain Shortcut

- **The bottleneck is not synthetic data quality, but the domain-induced feature bias between synthetic old classes and real new classes.**



(a) Baseline: Domain Shortcut in Naive Replay **(b) Ours: Subspace Rectification via DREAM**

Method



Proposition 3.1 (Domain Shortcut Phenomenon)**.** *Define the domain bias vector as $\Delta_{dom} = \mathbb{E}[z_{dom}(real)] - \mathbb{E}[z_{dom}(syn)]$. When the domain variance significantly outweighs semantic variance ($|\Delta_{dom}|^2 \gg Var(z_{sem})$, the model prioritizes the Δ_{dom} direction, misclassifying all real data as a new task. Consequently, real test old-class data ($d = real$) are wrongly classified as the new tasks.*

Corollary 3.2 (Shared Domain Difference Subspace)**.** *Since the generator is **training-free**, the statistical characteristics of the synthetic domain remain time-invariant across all classes. Similarly, the real images share a consistent natural domain distribution. Consequently, the domain gap $\mathcal{S}_{dom}$, arising from the discrepancy between these two stable distributions, is structurally consistent and shared. Formally:*

$$\forall c \in \mathcal{C}_{1:T}, \quad (\mathbb{E}[z|c, real] - \mathbb{E}[z|c, syn]) \in \mathcal{S}_{dom}.$$

Corollary 3.3 (Structured Domain Shift vs. Isotropic Intra-class Variance)**.** *Attributable to the fixed nature of the training-free generator, the synthesized data exhibits consistent stylistic tendencies. Consequently, the domain bias is not stochastic noise but comprises inherent systematic artifacts. This bias defines a **structured component** Σ_{dom} that dominates specific directions in the feature space. Conversely, intra-class semantic variations (e.g., color, pose, background) are characterized as **statistically isotropic components** $\sigma_{sem}^2 I$. Therefore, the covariance of class-conditional residuals decomposes as:*

$$Cov(z|c) \approx \Sigma_{dom} + \sigma_{sem}^2 I, \quad s.t.\ \lambda_1(\Sigma_{dom}) \gg \sigma_{sem}^2.$$

- **DREAM: Domain-Regularized Exemplar-free Alignment Model**
- **DREAM introduces no extra trainable parameters. It breaks the Domain Shortcut by purely geometric feature operations: hierarchical SVD to estimating the domain-sensitive subspace with paired new-class real-synthetic data and Soft Orthogonal Projection to suppress domain bias while preserving semantic information.**
- **Attribute-Aware Generative Replay**
- ***FDR*: Feature Domain Rectification**
- ***RAPC*: Real-Anchored Prototype Consolidation**

Result

- **Comparative experiments on four datasets. Across five settings, our method achieves the highest overall TAG Accuracy, demonstrating consistent superiority.**

Type	Methods	CIFAR-10		CIFAR-100			ImageNet-Subset			TinyImageNet		
		T=2	T=5	T=5	T=10	T=20	T=5	T=10	T=20	T=5	T=10	T=20
Baselines	Joint	93.27	91.05	73.61	70.45	71.45	80.46	77.13	76.41	60.84	58.55	54.34
	Fine-Tuning	70.28	42.84	35.12	24.72	15.89	38.44	26.29	17.18	28.21	20.06	12.82
Exemplar-Free	EWC (Kirkpatrick et al., 2017)	72.10	42.24	39.95	25.29	17.05	41.36	27.74	17.38	29.61	20.47	14.10
	LwF (Li & Hoiem, 2017)	79.94	56.28	41.71	30.69	19.87	52.55	39.27	26.87	32.02	22.47	15.26
	PASS (Zhu et al., 2021b)	63.72	59.75	63.31	52.01	41.84	55.75	33.75	27.30	39.55	30.03	18.64
	IL2A (Zhu et al., 2021a)	63.42	51.66	58.67	43.28	40.54	62.66	43.46	35.59	36.56	32.38	16.70
	SSRE (Zhu et al., 2022)	62.84	51.52	56.96	43.41	31.07	52.25	46.00	34.96	33.23	28.82	16.22
	FeTrIL (Petit et al., 2023)	83.47	66.59	58.68	47.14	37.25	58.40	46.44	37.64	40.46	30.77	23.68
	SEED (Rypesc et al., 2024)	<u>87.89</u>	74.03	63.05	62.04	57.42	69.08	67.55	62.26	<u>51.16</u>	<u>48.77</u>	<u>39.68</u>
Exemplar-Based	T-CIL (Hwang et al., 2025)	-	65.86	-	56.25	-	-	-	-	-	31.48	-
	T-CIL+DER (Hwang et al., 2025)	-	74.93	-	69.98	-	-	-	-	-	47.79	-
Exemplar-Free Generative Replay	ABD (Smith et al., 2021)	84.07	72.12	60.78	54.00	43.32	67.12	57.06	45.75	45.80	41.59	35.65
	R-DFCIL (Gao et al., 2022)	85.90	74.98	64.67	59.18	49.76	68.42	59.36	49.99	49.36	44.54	39.52
	DiffClass (Meng et al., 2024)	-	-	<u>69.77</u>	<u>68.05</u>	<u>67.10</u>	<u>74.85</u>	<u>73.87</u>	<u>72.51</u>	-	-	-
	AHR (Nori et al., 2025)	-	<u>77.12</u>	-	54.43	-	-	-	-	-	-	-
	DREAM (Ours)	**91.08**	**88.98**	**71.55**	**69.73**	**68.48**	**78.37**	**76.57**	**74.43**	**55.54**	**52.92**	**51.00**

- **Ablation Study**

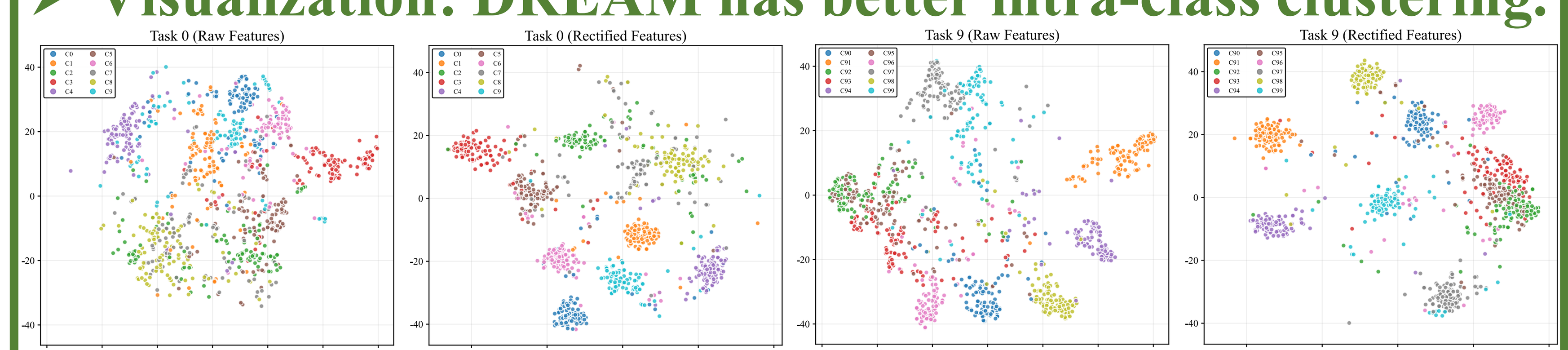
Component Effectiveness

Centering	HSR	RAPC	Acc (%)
Baseline (Raw Generative Replay)			33.27
✓			41.02
	✓		58.46
✓	✓		67.69
✓	✓	✓	**69.73**

Robustness to Generators

Type	Generator Source	w/o DREAM	DREAM
Multi-source	50% SD1.5 + 50% Qwen	37.92	68.53
	50% Qwen + 50% SD1.5	**40.37**	67.87
	Random mix	39.33	67.77
	Avg	39.21	68.06
Single-source	SD1.5	32.65	68.84
	Qwen-Image	33.27	**69.73**
	Avg	32.96	69.29

- **Visualization: DREAM has better intra-class clustering.**



- **DREAM eliminates domain bias and effectively breaks the "Domain Shortcut".**