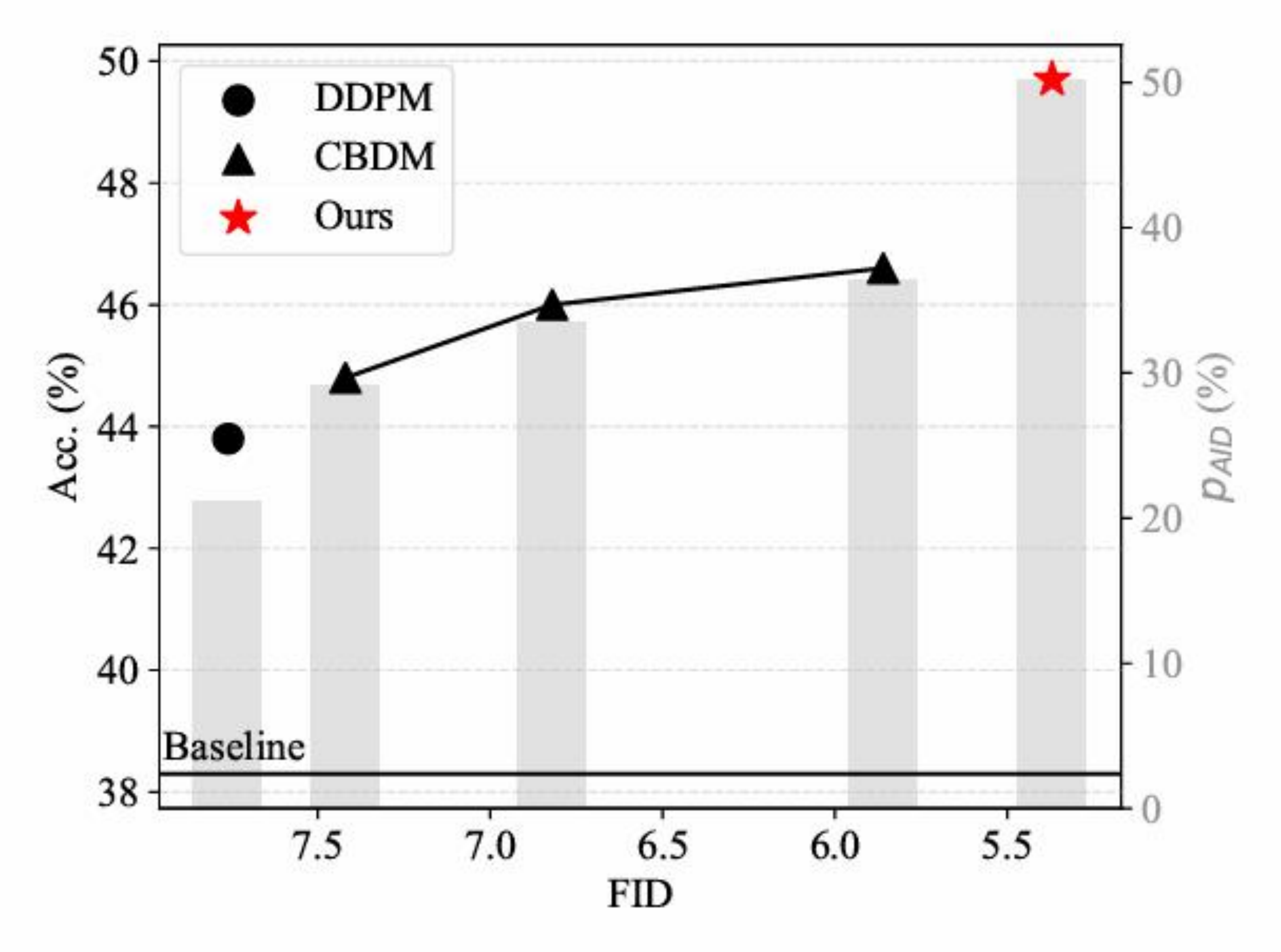
DiffuLT: Diffusion for Long-tail Recognition Without External Knowledge

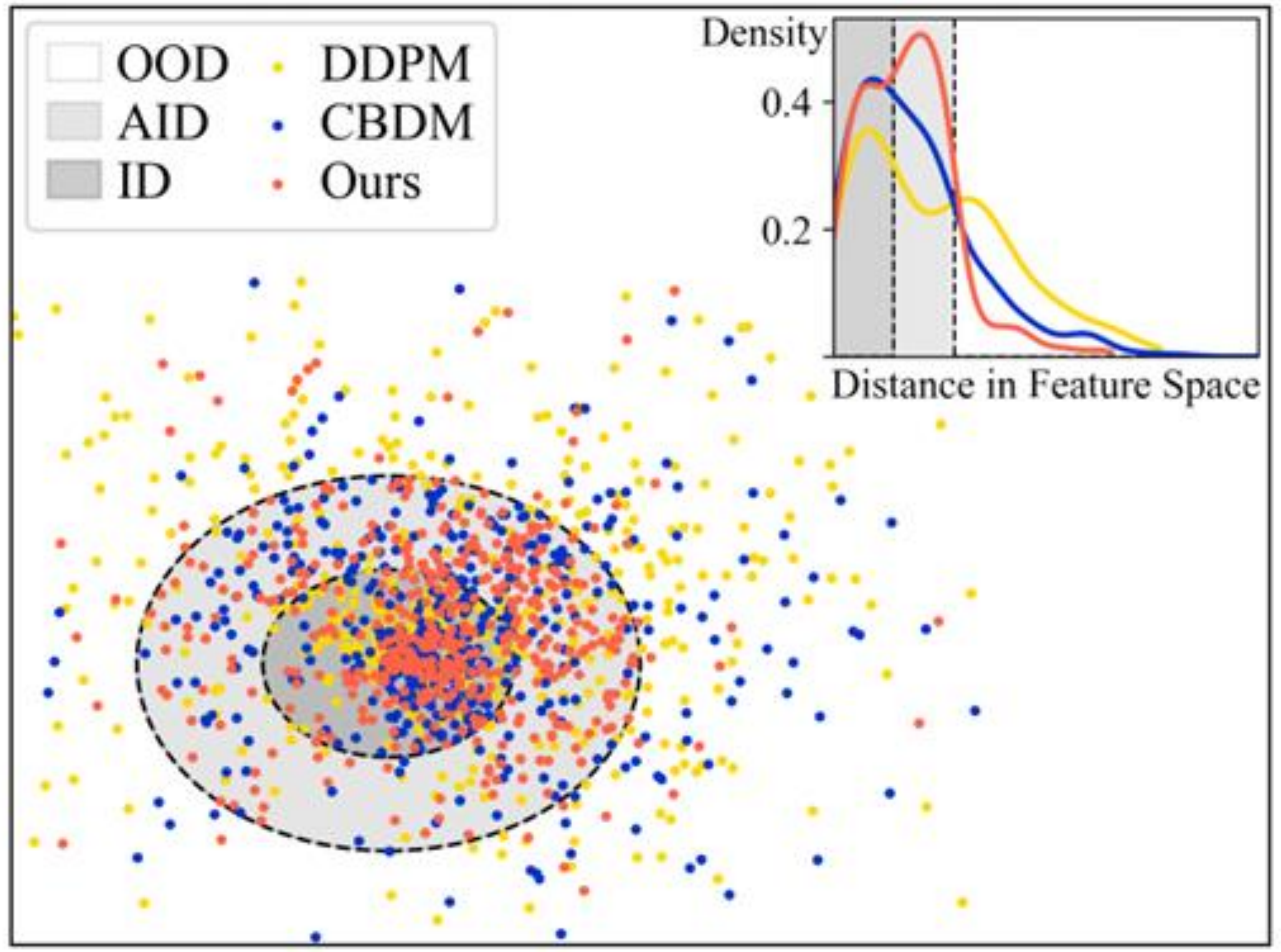
Jie Shao Ke Zhu Hanxiao Zhang Jianxin Wu*
National Key Laboratory for Novel Software Technology, Nanjing University, China
School of Artificial Intelligence, Nanjing University, China

Introduction

- **Objective:** Balance long-tailed datasets without external resources.
- **Method:** Use a diffusion model trained solely on the long-tail dataset to generate samples, improving long-tail classification.
- **Key Findings:**
  - Improving generative model performance boosts classifier accuracy.
  - Most impactful generated samples: AID (Approximately In-Distribution).
- **Proposed Pipeline:** DiffuLT (Diffusion model for Long-Tail recognition):
  - Step 1: Initial training of feature extractor and diffusion model with supervision for AID generation.
  - Step 2: Generate samples to balance the dataset.
  - Step 3: Retrain classifier on enriched dataset with reduced synthetic sample impact.

Introduction





Contributions

- **DiffuLT**
  - first to tackle long-tail recognition by synthesizing images with diffusion models
- **AID samples**
  - Blend information from head and tail classes
  - play a key role in boosting classifier accuracy
- Loss function that steers the diffusion model to generate more AID samples
- Experiments on CIFAR10-LT, CIFAR100-LT, and ImageNet-LT

Method

Table 1: FID of different generation models and their corresponding classifiers' accuracy.

Model	FID	Acc. (%)
Baseline	-	38.3
DDPM	7.76	43.8
CBDM ($\tau = 3$)	7.42	44.8
CBDM ($\tau = 2$)	6.82	46.0
CBDM ($\tau = 1$)	5.86	46.6

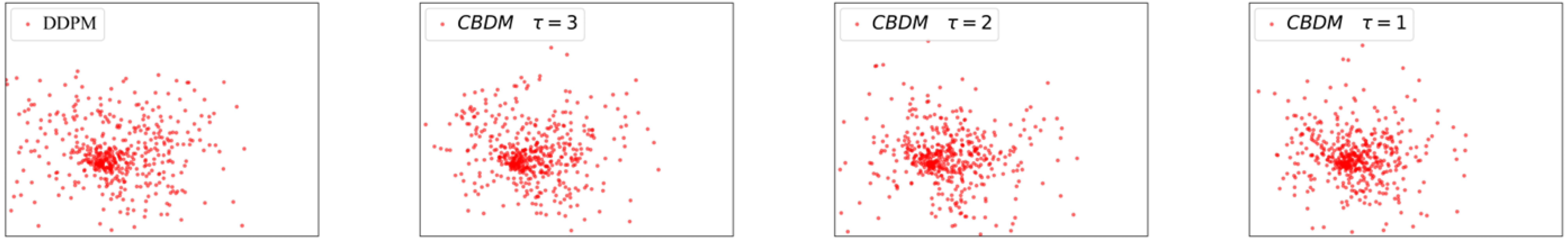


Figure 2: Visualization of generated samples for class 90 in feature space using t-SNE. The associated model is indicated in the upper-left corner.

Method

$$d_i = \|f_i - f_o\|_2 : \quad \begin{cases} d_i \le d_f, & \text{ID} \\ d_f < d_i \le 2d_f, & \text{AID} \\ d_i > 2d_f, & \text{OOD} \end{cases}$$

Table 3: Quantities, overall classifier enhancement, and average improvement per sample for different groups of data generated by diffusion model.

Group	$\|\mathcal{D}_{\text{gen}}\|$	Acc. (%)	$\Delta\text{Acc}/\|\mathcal{D}_{\text{gen}}\|$
Baseline	-	38.3	
ID	21,511	44.2	2.75×10^{-4}
AID	11,886	45.2	5.78×10^{-4}
OOD	5,756	36.2	-3.61×10^{-4}

Table 4: Diffusion trained with varying proportions of head class data and the corresponding results for tail classes.

p_h	p_{AID}	Acc_t (%)
-	-	25.0
0	25.8	26.0
40	33.2	29.7
80	35.7	32.5
100	39.1	32.8

Figure 3: Examples of three groups of generated samples.

Method

- AID loss:

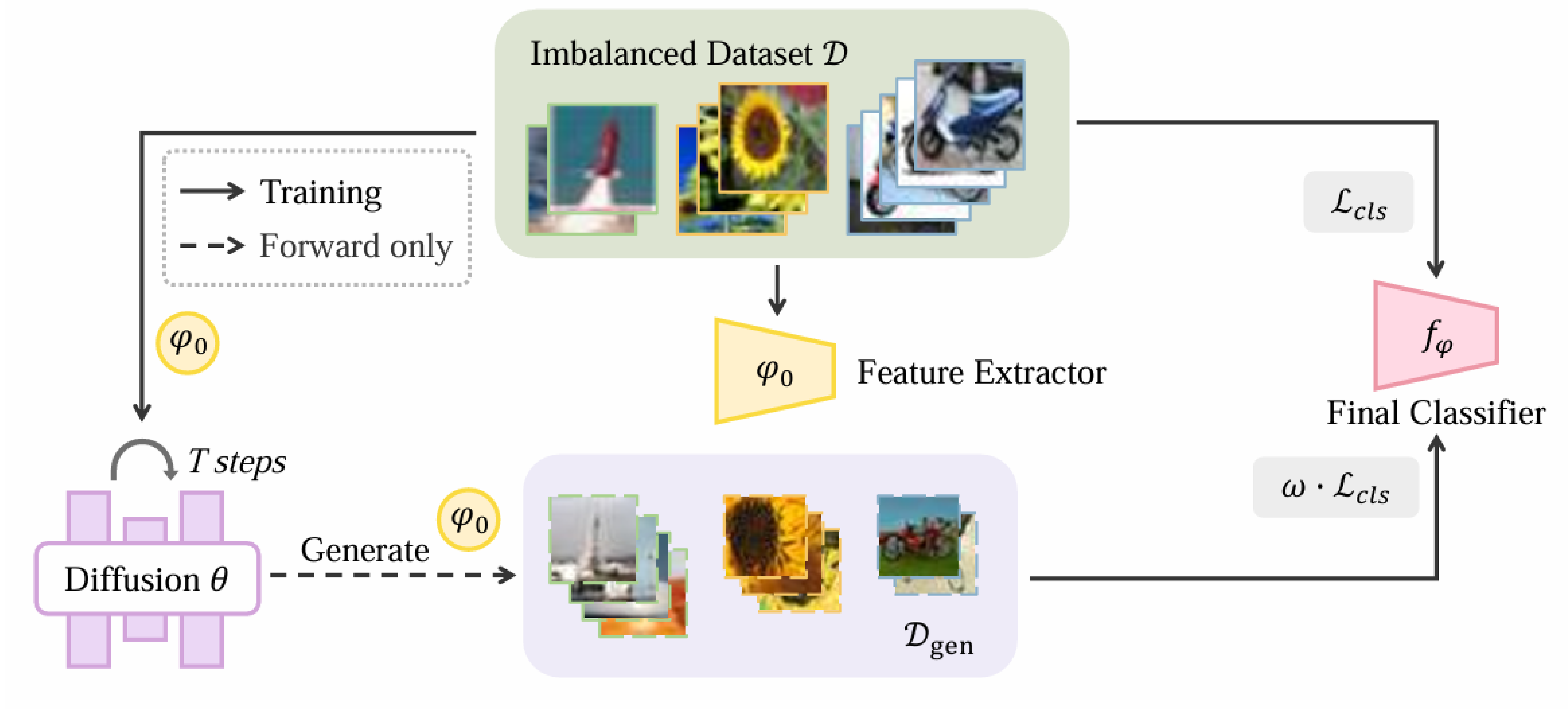
$$d_t = \frac{\sqrt{1-\bar{\alpha}_T}}{\sqrt{1-\bar{\alpha}_t}} \|\varphi_0(\boldsymbol{x}_0) - \varphi_0(\boldsymbol{x}_0 + \boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t, y))\|_2,$$

$$L_{\mathrm{AID}} = \alpha \mathbb{E}_{t\sim[1,T],\boldsymbol{x}_0,\epsilon_t} \|d_t - \frac{3}{2} d_f\|^2 .$$

- Weighted cross entropy:

$$L_{\mathrm{cls}} = -\sum_{(x,y,y_g)\in\mathcal{D}\cup\mathcal{D}_{\mathrm{gen}}} (\omega y_g + (1-y_g)) \log \frac{\exp(f_{\varphi,y}(x))}{\sum_{i=1}^{M} \exp(f_{\varphi,c_i}(x))} ,$$

Method – Stage 3



Experiments

Table 5: FID of diffusion model, proportion of samples, and corresponding classifier accuracy

Method	FID	p_{ID}	p_{AID}	p_{OOD}	Acc. (%)
DDPM	7.76	39.1	21.2	39.7	43.8
CBDM	5.86	44.8	36.3	18.9	46.6
Ours	5.37	40.7	50.1	9.2	49.7

Table 6: Methods and types of retained samples, pre-filtering counts, and classification accuracy.

Method	Kept	G-Num	Acc. (%)
CBDM	All	39,153	46.6
CBDM	AID	108,684	48.1
CBDM	ID & AID	48,414	47.1
Ours	All	39,153	49.7

Experiments

Table 7: Results on CIFAR100-LT and CIFAR10-LT datasets. The imbalance ratio r is set to 100, 50 and 10. The highest-performing results are in bold, with the second-best in underline. Additionally, we present the results for different groups (many, medium, and few) in CIFAR100-LT with $r = 100$.

Method	CIFAR100-LT			CIFAR10-LT			Statistics		
	100	50	10	100	50	10	Many	Med.	Few
CE	38.3	43.9	55.7	70.4	74.8	86.4	65.2	37.1	9.1
Focal Loss Lin et al. [2017]	38.4	44.3	55.8	70.4	76.7	86.7	65.3	38.4	8.1
LDAM-DRW Cao et al. [2019a]	42.0	46.6	58.7	77.0	81.0	88.2	61.5	41.7	20.2
cRT Kang et al. [2019]	42.3	46.8	58.1	75.7	80.4	88.3	64.0	44.8	18.1
BBN Zhou et al. [2020a]	42.6	47.0	59.1	79.8	82.2	88.3	-	-	-
RIDE (3 experts) Wang et al. [2020]	48.0	-	-	-	-	-	68.1	49.2	23.9
CAM-BS Zhang et al. [2021a]	41.7	46.0	-	75.4	81.4	-	-	-	-
MisLAS Zhong et al. [2021b]	47.0	52.3	63.2	82.1	85.7	90.0	-	-	-
DiVE He et al. [2021]	45.4	51.1	62.0	-	-	-	-	-	-
CMO Park et al. [2022]	47.2	51.7	58.4	-	-	-	**70.4**	42.5	14.4
SAM Rangwani et al. [2022]	45.4	-	-	81.9	-	-	64.4	46.2	20.8
CUDA Ahn et al. [2023]	47.6	51.1	58.4	-	-	-	67.3	50.4	21.4
CSA Shi et al. [2023b]	46.6	51.9	62.6	82.5	86.0	90.8	64.3	49.7	18.2
ADRW Wang et al. [2024b]	46.4	-	61.9	83.6	-	90.3	-	-	-
H2T Li et al. [2023]	48.9	53.8	-	-	-	-	-	-	-
DiffuLT	51.5	56.3	63.8	84.7	86.9	90.7	69.0	51.6	29.7
DiffuLT + BBN	<u>51.9</u>	<u>56.7</u>	<u>64.0</u>	<u>85.0</u>	<u>87.2</u>	**90.9**	69.5	<u>51.9</u>	<u>30.2</u>
DiffuLT + RIDE (3 experts)	**52.4**	**56.9**	**64.2**	**85.3**	**87.3**	<u>90.9</u>	<u>70.3</u>	**52.1**	**30.7**

Experiments

Table 8: Results on ImageNet-LT. We deploy ResNet-10 and ResNet-50 as classifier backbones. Top-performing results are highlighted in bold, with second-best outcomes underlined.

	ResNet-10	ResNet-50			
	All	All	Many	Med.	Few
CE	34.8	41.6	64.0	33.8	5.8
Focal Loss Lin et al. [2017]	30.5	-	-	-	-
OLTR Liu et al. [2019b]	35.6	-	-	-	-
cRT Kang et al. [2019]	41.8	47.3	58.8	44.0	26.1
RIDE (3 experts) Wang et al. [2020]	45.9	54.9	<u>66.2</u>	51.7	34.9
MisLAS Zhong et al. [2021b]	-	52.7	-	-	-
CMO Park et al. [2022]	-	49.1	**67.0**	42.3	20.5
SAM Rangwani et al. [2022]		53.1	62.0	52.1	34.8
CUDA Ahn et al. [2023]	-	51.4	63.1	48.0	31.1
CSA Shi et al. [2023b]	42.7	49.1	62.5	46.6	24.1
ADRW Wang et al. [2024b]	-	54.1	62.9	52.6	37.1
DiffuLT	<u>50.4</u>	<u>56.4</u>	63.3	<u>55.6</u>	<u>39.4</u>
DiffuLT + RIDE (3 experts)	**51.1**	**56.9**	64.1	**55.8**	**39.9**

Learning And Mining from DatA

NEURAL INFORMATION PROCESSING SYSTEMS

Experiments

Table 9: Ablation experiments to verify the effect of each module.

Gen.	L_{AID}	Filt.	Weight	Acc. (%)
				38.3
✓				46.6
✓	✓			49.7
✓	✓	✓		50.3
✓	✓	✓	✓	**51.5**

Table 10: Performance with different weights ω and hyper-parameter α.

ω	Acc. (%)	α	Acc. (%)
0	38.3	0	38.3
0.1	49.2	0.1	49.7
0.3	51.5	0.5	49.5
0.5	50.1	1.0	48.3
0.7	50.3	2.0	45.1
1.0	50.3	4.0	43.3

NANJING UNIVERSITY 1902

Learning And Mining from DatA

NEURAL INFORMATION PROCESSING SYSTEMS

Conclusion

- **Novel Approach:** Proposed a data-centric method using AID samples for long-tail classification.
- **Key Contributions:**
  - Developed an AID-focused diffusion model to enrich datasets.
  - Demonstrated the critical role of AID samples in boosting classifier accuracy.
- **Impact:**
  - Provides a robust framework without needing external data.
  - Adaptable to various performance-critical applications.
- **Limitations & Future Work:**
  - Current method is time-consuming; future work will focus on optimizing training and generation speeds.