Mitigating Bias in Dataset Distillation

Justin Cui[1], Ruochen Wang[1], Yuanhao Xiong[1], Cho-Jui Hsieh[1]

[1]Department of Computer Science, University of California, Los Angeles

Please check out our paper!

Background

Dataset Distillation

Synthesizing large datasets into smaller ones while achieving similar performances

Applications

Continual Learning, Membership Inferences, Neural Architecture Search(NAS), etc

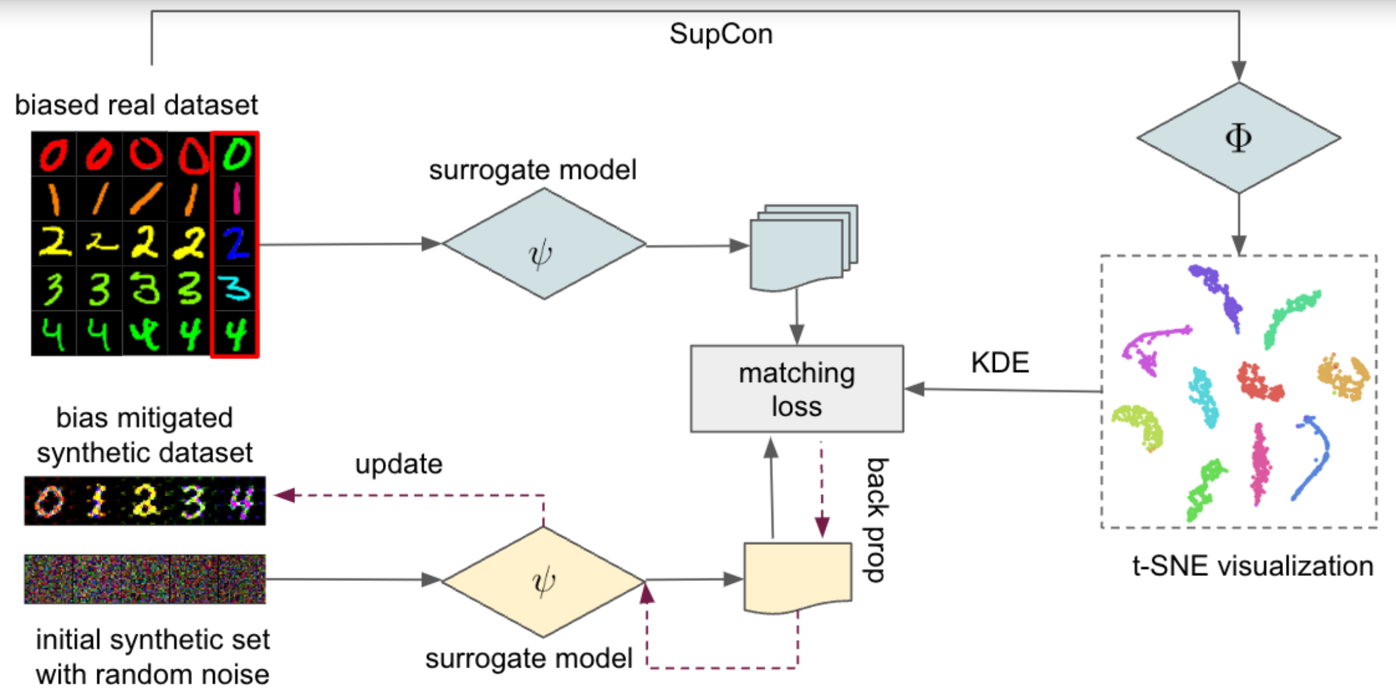
Motivation

DD methods often exhibit different behaviors on biased dataset than balanced datasets.

Contributions

- **Analysis** Dataset Distillation methods exhibit different behaviors on biased datasets, causing biases to be amplified on some datasets and suppressed on some others.
- **Method** We propose to use Kernel Density Estimation (KDE) with Supervised Contrastive (SupCon) Learning to re-weight the samples during distillation.
- **Results** Our method effectively boost various DD methods on bias-amplified dataset.

Overall Workflow



Normal Training Objective

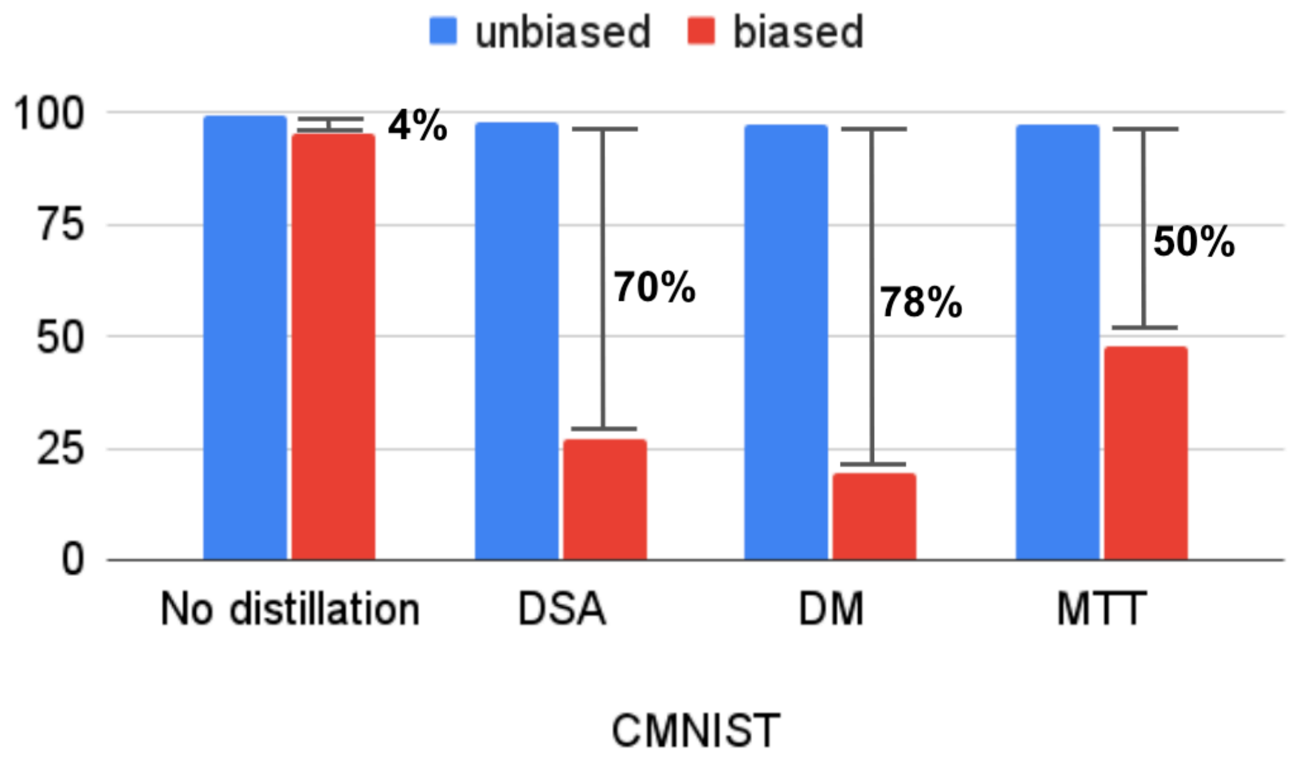
$$\min_{\mathcal{S}} \mathop{\mathbb{E}}_{\substack{v\sim P_v\\ \omega\sim\Omega}} \| \frac{1}{|\mathcal{T}|}\sum_{i=1}^{|\mathcal{T}|}\psi_v(\mathcal{A}(x_i,\omega)) - \frac{1}{|\mathcal{S}|}\sum_{j=1}^{|\mathcal{S}|}\psi_v(\mathcal{A}(s_j,\omega)) \|^2,$$

$$\min_{\mathcal{S}} D(\nabla_\theta\mathcal{L}_c^{\mathcal{T}}(\mathcal{A}(\mathcal{T},\omega^{\mathcal{T}}),\theta_t), \nabla_\theta\mathcal{L}_c^{\mathcal{S}}(\mathcal{A}(\mathcal{S},\omega^{\mathcal{S}}),\theta_t)),$$

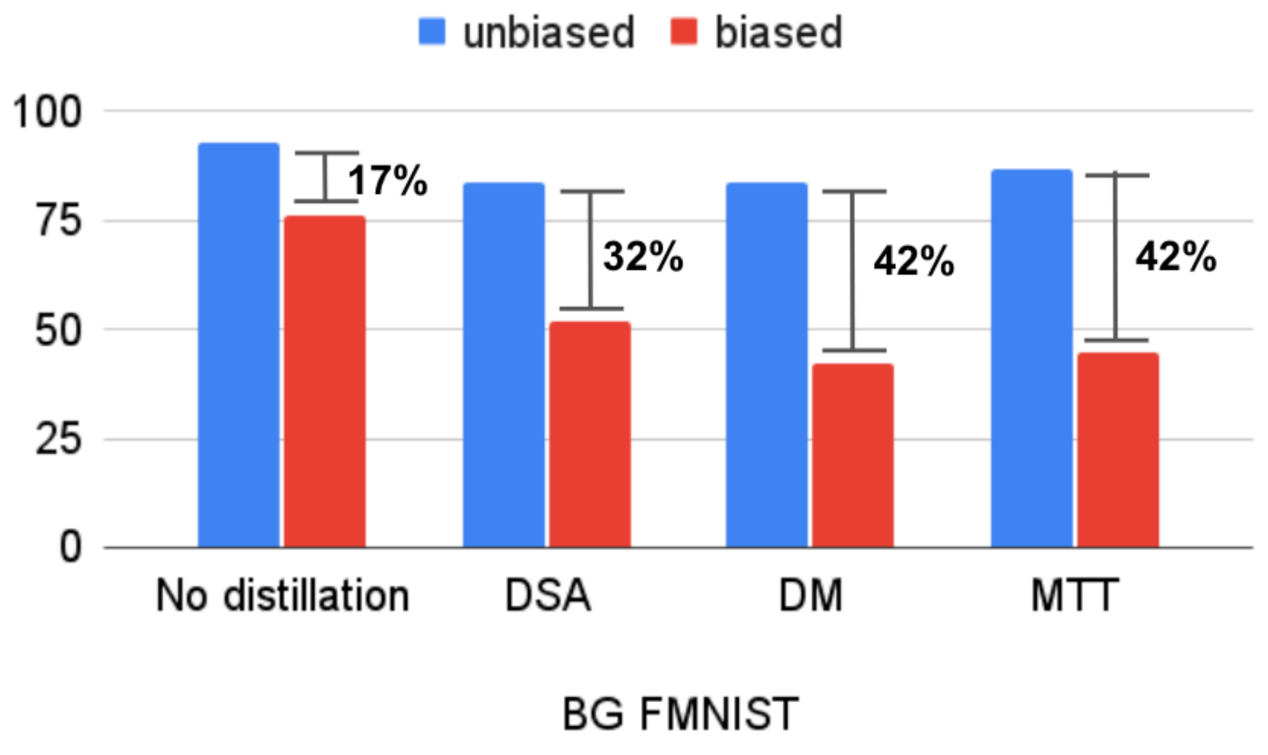
Training with KDE and SupCon

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(\|\Phi(x) - \Phi(x_i)\|\right)$$

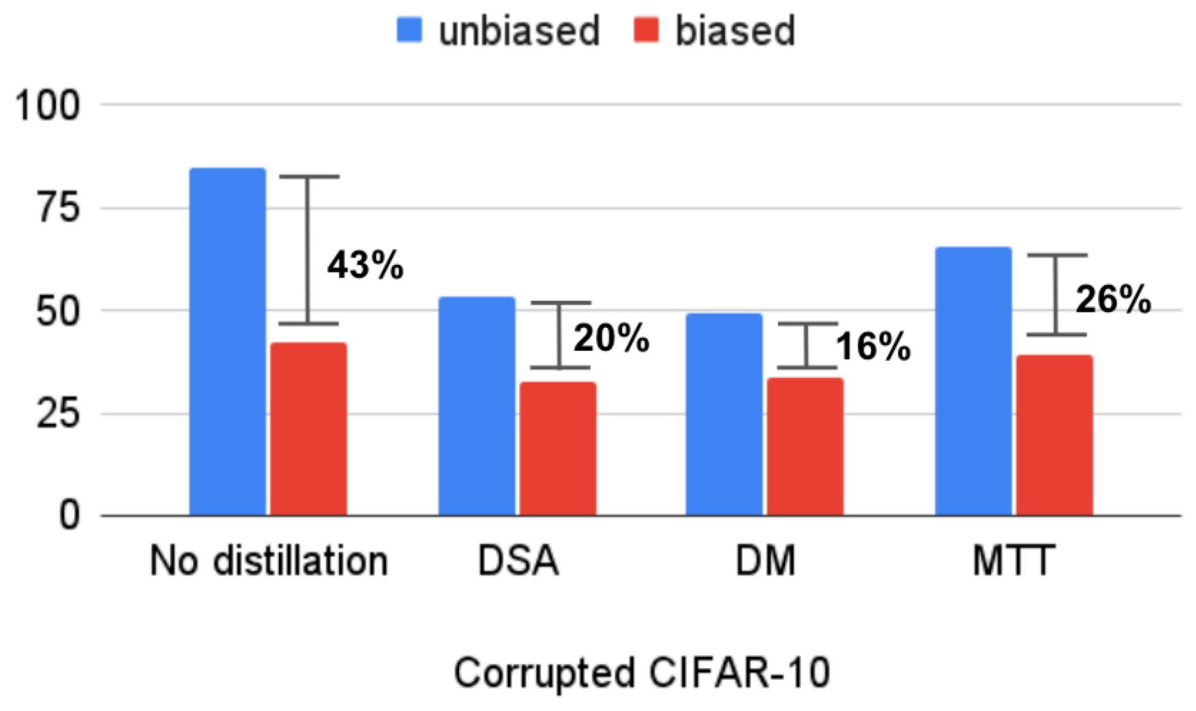
Overall Performance on Balanced Dataset



CMNIST



BG FMNIST

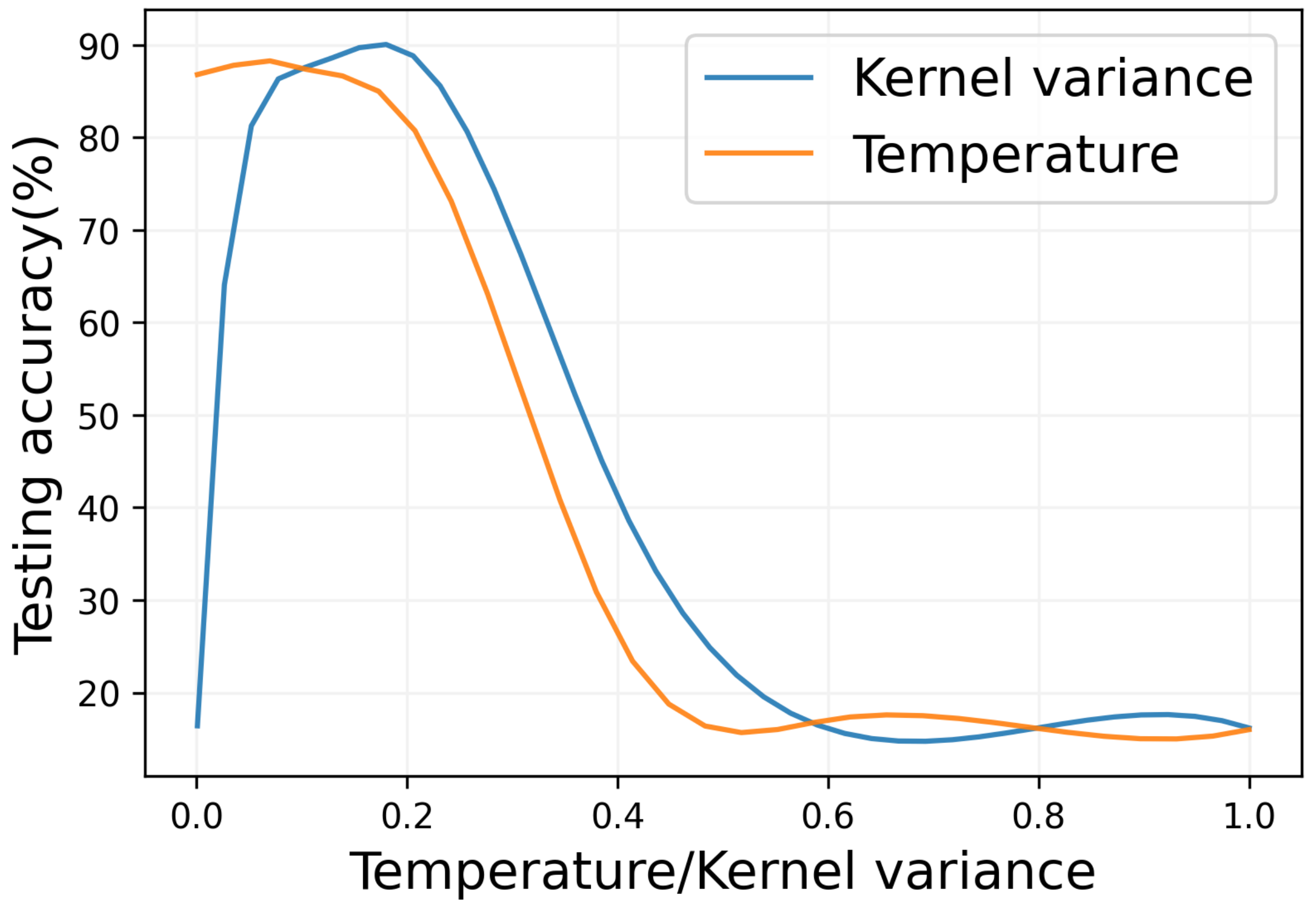


Corrupted CIFAR-10

Key Results

Dataset	Method	Bias-conflict Ratio (1.0%) 1	10	50	Bias-conflict Ratio (2.0%) 1	10	50	Bias-conflict Ratio (5.0%) 1	10	50
CMNIST	Random	22.6±1.0	13.5±0.3	16.0±0.1	22.8±1.0	16.2±0.1	20.7±0.1	19.9±0.5	17.4±0.4	27.2±0.2
	MTT	24.2±0.3	27.6±0.4	18.1±0.6	**28.3±0.3**	42.0±0.4	26.0±0.5	29.2±0.9	47.7±0.8	33.9±1.2
	IDM	20.1±0.9	18.8±1.2	17.6±1.6	20.2±1.0	20.1±0.5	19.1±1.3	20.8±0.9	22.3±1.1	26.5±1.5
	DREAM	21.3±0.4	15.9±1.2	17.9±0.8	26.2±0.5	17.1±1.3	17.5±1.5	23.7±1.2	30.8±1.5	50.4±1.0
	DM	25.4±0.1	18.6±0.2	22.6±0.5	24.8±0.4	18.5±0.6	23.6±0.8	25.3±0.3	19.6±0.9	23.8±1.3
	DM+Ours	**28.0±0.5**	64.9±0.3	75.4±1.1	26.4±0.7	50.6±1.2	75.7±1.0	32.2±1.0	86.5±1.2	91.5±0.9
	DSA	26.1±0.3	16.5±0.2	14.5±0.2	25.2±0.3	16.8±0.3	30.9±0.4	25.9±0.5	27.3±0.4	68.5±1.2
	DSA+Ours	27.9±0.4	**76.7±1.1**	**81.4±0.8**	26.4±0.2	**75.3±0.3**	**83.0±1.2**	**32.6±0.1**	**91.9±0.7**	**94.0±0.8**
BG FMNIST	Random	40.0±0.2	40.4±1.6	35.2±0.4	30.2±0.6	43.2±0.2	33.5±1.0	36.2±1.3	44.6±1.2	41.2±1.1
	MTT	39.0±1.2	48.0±1.5	45.3±0.9	38.9±1.4	59.2±1.1	59.3±0.8	48.1±1.4	45.2±1.3	62.3±0.8
	IDM	40.7±1.0	42.3±0.9	38.4±1.2	41.1±0.8	37.2±1.3	40.5±0.9	43.4±0.4	46.6±0.8	42.0±1.2
	DREAM	39.7±0.9	46.4±1.2	46.1±1.5	45.0±1.0	46.0±0.8	44.5±0.8	43.5±0.9	52.4±1.5	53.2±1.2
	DM	41.0±0.3	42.2±0.8	43.9±0.4	40.1±0.6	40.1±0.9	44.4±0.5	41.7±0.5	42.0±1.2	44.6±0.9
	DM+Ours	**44.6±0.5**	50.6±0.2	57.2±0.6	**51.4±0.7**	62.3±0.4	63.0±1.0	**49.4±0.2**	61.8±0.6	65.0±0.8
	DSA	43.4±0.4	45.8±0.5	40.7±0.9	43.7±0.5	47.6±0.3	48.4±0.8	44.7±0.6	52.8±0.5	59.3±0.6
	DSA+Ours	44.4±0.6	**57.0±1.0**	**58.3±0.8**	48.5±1.2	**64.4±0.9**	**65.1±0.8**	46.2±0.6	**66.4±0.6**	**71.2±1.1**
Corrupted CIFAR-10	Random	16.4±0.6	26.9±0.2	32.7±0.3	19.1±0.1	23.2±0.2	33.5±0.1	11.8±0.1	26.4±0.3	34.2±0.4
	MTT	23.5±0.4	25.4±1.5	33.3±0.5	24.1±0.3	**36.3±0.4**	35.7±0.2	24.2±0.8	**39.0±0.3**	39.5±0.4
	IDM	**26.3±1.0**	30.2±0.4	36.6±0.8	**26.2±0.8**	29.5±0.2	35.0±0.7	26.0±0.8	31.4±1.1	36.5±0.9
	DREAM	23.7±0.9	25.0±1.2	33.6±0.9	25.9±0.8	25.2±0.6	32.1±1.3	25.6±0.9	24.5±0.8	33.7±1.2
	DM	25.1±0.4	32.9±0.3	37.6±0.8	25.0±0.1	32.9±0.1	37.7±0.2	24.6±0.4	33.5±0.8	38.7±0.4
	DM+Ours	24.2±1.2	**33.4±0.9**	**39.4±0.8**	25.3±0.5	34.2±0.5	**39.7±0.4**	**26.6±0.5**	33.5±0.6	**40.2±0.4**
	DSA	25.5±0.3	31.9±0.8	34.1±0.5	25.1±0.2	32.0±0.1	34.2±0.3	25.7±0.5	32.8±0.6	35.6±0.5
	DSA+Ours	26.0±0.1	32.6±0.8	35.0±0.6	25.2±0.8	33.2±0.2	35.8±0.6	26.0±0.3	32.5±0.7	36.6±0.3

Kernel Variance & Temperature



Ablation Study on Training Time Debiasing

Table 3. Ablation study test accuracy (%) on applying de-biasing method to train with synthetic datasets from DM, assessed under 5% bias-conflicting samples and IPC 10 and 50.

IPC	CMNIST 10	CMNIST 50	BG FMNIST 10	BG FMNIST 50
DM	19.6±0.9	23.8±1.3	42.0±1.2	44.6±0.9
DFA	25.8±1.0	31.2±1.3	11.0±2.1	17.6±1.9
SelecMix*	43.3±1.3	53.7±1.5	57.2±1.1	58.7±0.9
DM+Ours	**86.5±1.2**	**91.5±0.9**	**61.8±0.6**	**65.0±0.8**

Ablation Study on Surrogate Model

Table 4. Ablation study test accuracy (%) on applying de-biasing methods to surrogate models on CMNIST with 5% bias-conflict samples and IPC 1, 10 and 50.

Method	IPC 1	IPC 10	IPC 50
DM	25.3±0.3	19.6±0.9	23.8±1.3
DM+DFA	26.1±0.3	20.5±0.5	25.4±0.4
MTT	29.2±0.9	47.7±0.8	33.9±1.2
MTT+SelecMix	18.1±0.5	29.9±0.8	52.1±0.3
DM+Ours	**32.2±1.0**	**86.5±1.2**	**91.5±0.9**