Table 1: Extended downstream task transfer learning with the pre-trained model

Method	Encoder	CIFAR-10	CIFAR-100	Aircraft	Flowers	Food-101	Cars	Pets
Supervised (from scratch)	Mob-V3 (1×)	92.97	73.69	65.37	79.89	60.30	68.18	70.97
Supervised (fine-tune)	Mob-V3 (1×)	94.53	78.86	68.29	89.94	75.84	82.43	85.87
XD (Ours)	Mob-V3 (1×)	**94.80**	**79.00**	**71.39**	**90.05**	75.71	**82.77**	**89.42**
SACL + XD (Ours)	Mob-V1 (1.5×-1×)	**94.92**	**79.64**	**72.21**	**90.48**	**76.12**	**83.14**	**90.24**

Table 2: ImageNet-100 test accuracy with linear evaluation protocol based on ViT-Tiny [12] encoder.

Methods	Encoder	Training Epochs	Linear Eval Acc. (%)
Barlow Twins [31]	ViT-Tiny [12] (# of Param = 5.5 Million)	400	62.56
*DINO [8]		400	63.04
XD (Ours)		400	**63.92 (+0.88)**
SACL+XD (Ours)	1.25× - 1×	400	**64.97 (+1.93)**

*: Reported DINO results from [11].

Table 3: Activation count comparison between the proposed method and the distillation-based CL

Method	Encoder	Teacher	Act. Count (E+07)	ImageNet-1K Accuracy (%)
SACL+XD (Ours)	Eff-B0 (1.5×-1×)	N/A	1.54	**65.32**
XD Only (Ours)	Mob-V3 (1×)	N/A	**0.90**	**59.34**
SSL-Small	Mob-V3	N/A	0.90	48.70
SSL-Small	Eff-B0	N/A	0.68	55.90
ReKD	Mob-V3	ResNet-101	2.03	59.60
ReKD	Mob-V3	ResNet-50	1.53	56.70
ReKD	Eff-B0	ResNet-50	1.75	63.40
SEED	Mob-V3	ResNet-50	1.52	55.20
SEED	Eff-B0	ResNet-101	2.26	61.30

Table 4: Training time comparison between the proposed method and the distillation-based CL

Model	Training Method	Teacher	Training time / epoch	GPU Type	Batch Size
SEED	MobileNet-V3	ResNet-50	35 min 20 sec	A100 (80G)	256
SACL-XD (Ours)	MobileNet-V3 1.5×-1×	N/A	26 min 02 sec	A100 (80G)	256
XD Only (Ours)	MobileNet-V3	N/A	16 min 15 sec	A100 (80G)	256

Table 5: Comparison between the proposed method and other supervised high-water marks

Model	Training Method	CIFAR-10 Acc (%)	CIFAR-100 Acc (%)	# of (remained) Param. (M)
ResNet-50	Supervised Learning	94.75	78.23	25.6
ResNet-50	Supervised + GraNet [Ref-1]	94.64	77.89	2.6 (90% sparsity)
ResNet-50	Supervised + RigL [Ref-2]	94.45	76.50	2.6 (90% sparsity)
Mob-V1	**SACL+XD (Ours) + Finetune**	**94.92**	**79.64**	3.2
Mob-V3	**XD (Ours) + Finetune**	**94.80**	**79.00**	3.0

Table 6: ImageNet-1K test accuracy with linear evaluation protocal based on MobileNet-V3 and EfficientNet-B0 trained by different contrastive learning/distillation methods.

Method	Encoder	Linear Eval. (%)	Epochs	Pre-train	Teacher	Training FLOPs (e+17)
SACL-XD (Ours)	Eff-B0 (1.5×-1×)	**65.32 (+2.12)**	200	✗	-	**24 (2.9× ↓)**
SACL-XD (Ours)	Mob-V3 (1.5×-1×)	**61.69 (+1.79)**	200	✗	-	**15 (64.7× ↓)**
SACL-XD (Ours)	Mob-V1 (1.5×-1×)	**59.34**	200	✗	-	**19**
XD only (Ours)	Mob-V3 (1×)	**59.42**	200	✗	-	**7.2**
XD only (Ours)	Mob-V3 (1×)	**57.16**	**100**	✗	-	**3.6**
XD only (Ours)	Mob-V1 (1×)	**55.84**	**100**	✗	-	**9.0**
SSL-Small [24]	Mob-V3 (1×)	47.90	800	2 epochs	-	19
SSL-Small [24]	Eff-B0 (1×)	55.90	800	2 epochs	-	34

Table 7: CIFAR-10 linear evaluation test accuracy based on ResNet-20 trained by SACL+XD with different asymmetrical architectures.

Method	Encoder	Linear Eval Acc. (1× model)	Teacher	Teacher Pre-trained by	Training FLOPs (e+16)
SACL+XD (Ours)	ResNet-20 (6×-1×)	**86.81 (+7.18)**	-	-	8.60
SACL+XD (Ours)	ResNet-20 (4×-1×)	**84.04 (+4.41)**	-	-	3.90
SACL+XD (Ours)	ResNet-20 (2×-1×)	**82.31 (+2.68)**	-	-	0.98
*SEED [14]	ResNet-20 (1 ×)	82.86	ResNet-20 (6×)	MoCo [7]	180
*SEED [14]	ResNet-20 (1 ×)	81.36	ResNet-20 (6×)	Barlow Twins [31]	180
Barlow Twins [31]	ResNet-20 (1 ×)	79.63	-	-	0.25
VICReg [3]	ResNet-20 (1 ×)	79.13	-	-	0.25