NeurIPS
#6348

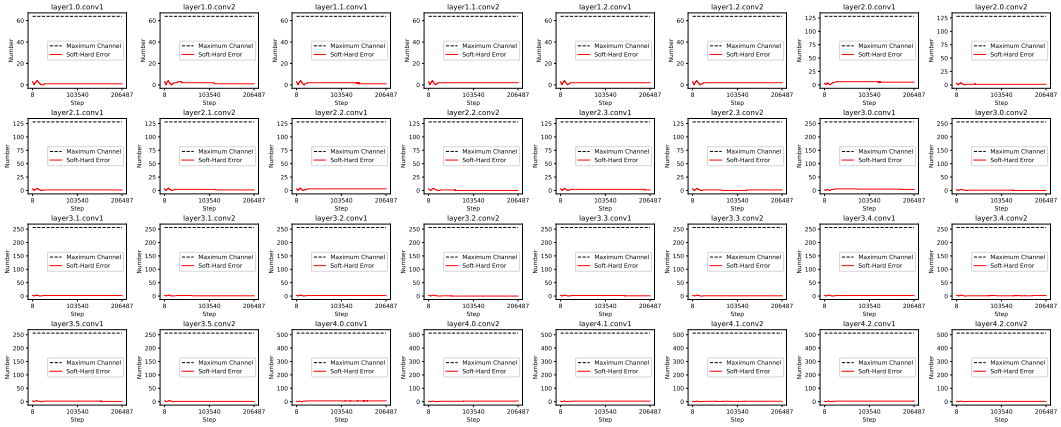

Figure t1. The soft and hard channel difference w.r.t. the maximum channel number of ResNet-50 layers during the training on CIFAR-100.

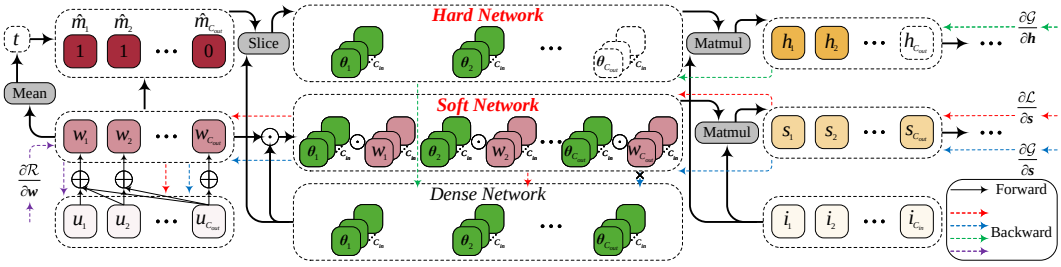

Figure t2. The revised framework, which emphasizes the concept of soft and hard networks.

Table t1. Supplementary results of ResNet-50 on Imagenet. CHEX* means implementing CHEX with our training schedule.

Method	Unpruned top-1 (%)	Pruned top-1 (%)	Top-1 drop (%)	FLOPs (%)	Pretrain epochs	Prune epochs
CHEX	77.80	**76.00**	1.80	25.00	0	250
CHEX*	76.15	75.33	0.82	30.00	0	200
Ours	76.15	75.39	**0.76**	**24.21**	0	**200**
SCOP	76.15	75.26	0.89	45.40	90	**140**
Ours	76.15	**75.68**	**0.47**	**35.00**	0	200
SCOP	76.15	75.95	0.20	54.70	90	**140**
CHEX	77.80	**77.40**	0.40	**50.00**	0	250
Ours	76.15	77.13	**-0.98**	53.49	0	200

Table t2. Comparison with the simple combination of pruning and knowledge distillation.

	Prune	Prune+Distillation	**Ours**
Epochs	250	500	500
Top-1 Acc (%)	76.93	77.64	**79.77**

Table t3. ResNet-50 on CIFAR-100 using different gap measures.

Metrics	L1	L2	**Ours (KL)**
Top-1 Acc (%)	77.09	76.95	**79.77**

Table t4. Training efficiency and GPU RAM comparison with different methods.

	RST-S	Depgraph	OTO v2	IMP-Refill	Ours
Top-1 Acc (%) (1x training schedule)	75.02	49.07	77.04	75.12	79.77
Top-1 Acc (%) (2x training schedule)	75.54	50.83	77.21	75.66	-
GPU time per epoch (s)	44.50	70.97	79.36	74.12	50.13
Peak GPU memory (MB) (training)	4329	4319	4221	4261	4710
Peak GPU memory (MB) (inference)	1351	1365	1262	1329	1279