

A MORE EXPERIMENT RESULT

The result of closed-set SelKD tasks on the CIFAR-10 dataset are presented in Table 5, and it is obvious that SelKD outperforms both KD and SKD in terms of top-1 accuracy.

Table 5: Top-1 accuracy (%) on CIFAR-10 for closed-set SelKD. We compare the performance of vanilla knowledge distillation (KD), self-knowledge distillation (SKD) and our selective knowledge distillation (SelKD) in the closed-set SelKD classification tasks.

Student Networks	Methods	SubTask 0	SubTask 1	Mean of SubTasks	
ResNet-8	Without KD	91.28	94.76	93.02	
	FitNet-KD	91.56	95.72	93.64	
	FitNet-SKD	91.42	95.34	93.12	
	FitNet-SelKD	91.88	95.88	93.88	
	FT-KD	91.98	96.22	94.10	
	FT-SKD	91.53	95.99	93.25	
	FT-SelKD	93.68	96.50	95.09	
	AT-KD	91.66	95.60	93.63	
	AT-SKD	91.68	95.09	93.65	
	AT-SelKD	92.90	96.12	94.51	
	ResNet-14	Without KD	94.42	96.54	95.48
		FitNet-KD	94.86	96.40	95.63
FitNet-SKD		91.56	95.72	93.64	
FitNet-SelKD		95.04	97.72	96.38	
FT-KD		94.56	97.04	95.80	
FT-SKD		91.56	95.72	93.64	
FT-SelKD		95.66	97.92	96.79	
AT-KD		94.02	96.76	95.39	
AT-SKD		91.56	95.72	93.64	
AT-SelKD	94.84	97.64	96.24		