A. Data Sampling

Following previous work, our tiny training set is sampled uniformly from the whole set. In Tab. 9, "1-way N-shot" denotes choosing one class randomly and then sampling N images from this class—the tiny set contains patterns of one specific class. It is indeed worse than uniform sampling, but the results are still acceptable, which means PRACTISE is robust with data from limited classes.

Data sampling	N=100	500	1000
1-way N-shot	$69.8_{\pm 0.47}$	$70.9_{\pm 0.42}$	$70.0_{\pm 1.84}$
uniform	$\mathbf{70.4}_{\pm 0.42}$	$\mathbf{71.8}_{\pm 0.07}$	$\mathbf{71.9}_{\pm 0.05}$

Table 9. Comparisons for different data sampling strategies.

B. Training Time

Evaluating the latency is efficient. The latency of raw ResNet-34 is 42ms, and testing it by 500 times costs only 21 seconds. Evaluating all 12 blocks requires about 5 minutes. Because of the tiny training set and limited training iterations, optimizing the model is also fast. Tab. 10 reports the costed time on one Titan Xp GPU. Computing one block's recoverability only takes about 6 minutes. The total training time of PRACTISE is only about 1.5 hours.

Latency	Recoverability	Finetuning	Total
0.4×12	5.8×12	11.3	85.7

Table 10. Training time (min).

C. Different Training Settings

Here, we conduct ablation studies for the learning rate γ and iterations. Tab. 11 presents results. We find that optimizing adaptors with 100 iterations is good enough while training all parameters requires more iterations.

Opt Eq.1	Opt Eq.4	$\gamma = 0.01$	0.02	0.04
100	1000	71.14	71.65	71.26
100	2000	71.52	**71.83**	71.46
1000	2000	71.61	71.82	71.37

Table 11. Ablative results of different hyperparameters.