Learning to Self-Train for Semi-Supervised Few-Shot Classification

Xinzhe Li[1] Qianru Sun[2,6] Yaoyao Liu[3] Shibao Zheng[1] Qin Zhou[4] Tat-Seng Chua[5] Bernt Schiele[6]

[1]Shanghai Jiao Tong University [2]Singapore Management University [3]Tianjin University [4]Alibaba [5]National University of Singapore [6]MPI for Informatics

max planck institut informatik

SHANGHAI JIAO TONG UNIVERSITY

Code is on GitHub

Task & Motivation & Contributions

- Few-shot classification (FSC) is challenging due to the scarcity of labeled training data, *e.g.* only one labeled image per class.
- One solution is meta-learning that transfers experiences learned from similar tasks to the target task [1].
- Another solution is semi-supervised learning that additionally use unlabeled data in training [4].
- In our work, we combine these two solutions and achieve the top performance, *e.g.* 70.1% on miniImageNet 5-way 1-shot setting.

Semi-supervised few-shot classification (SSFSC)

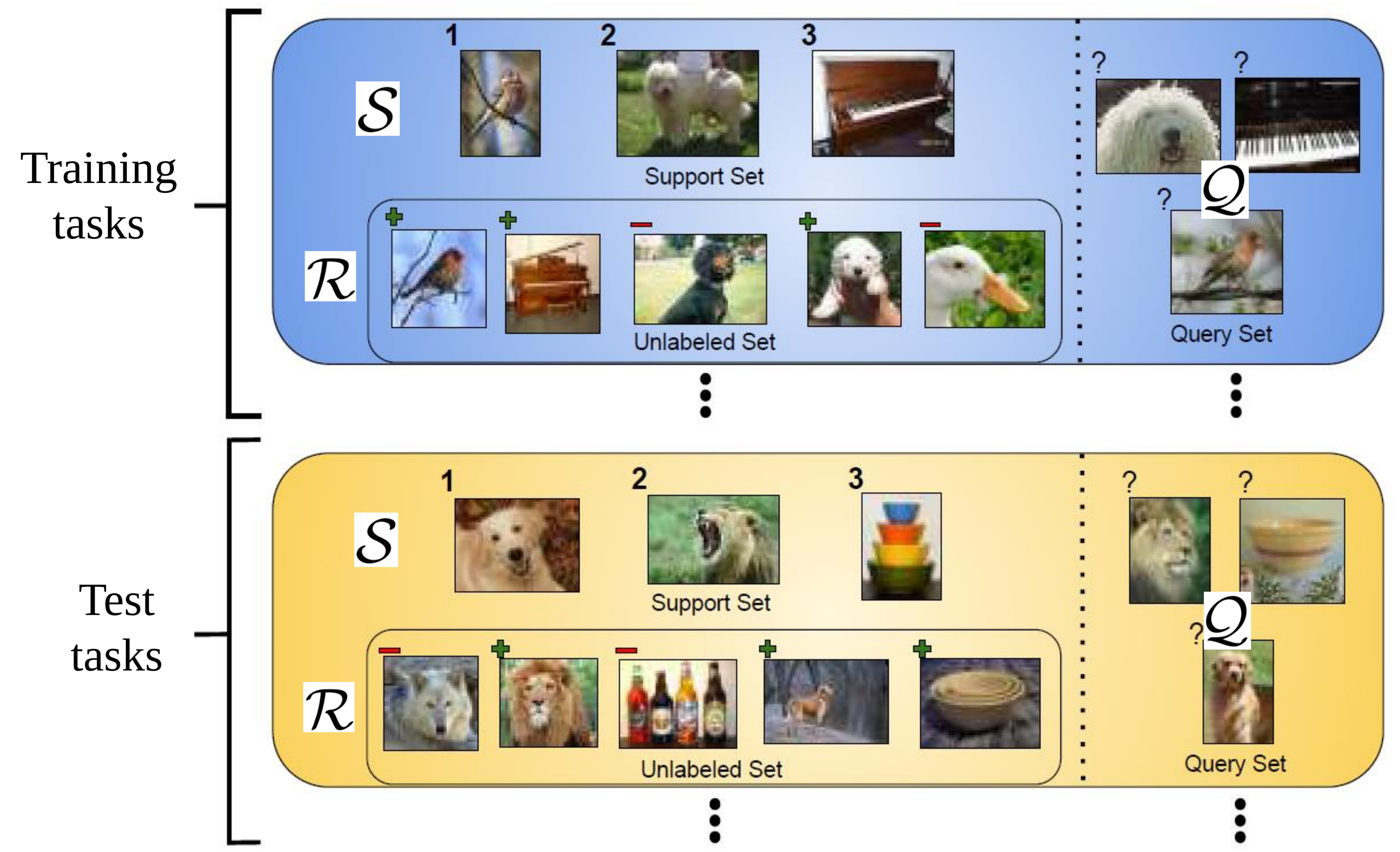


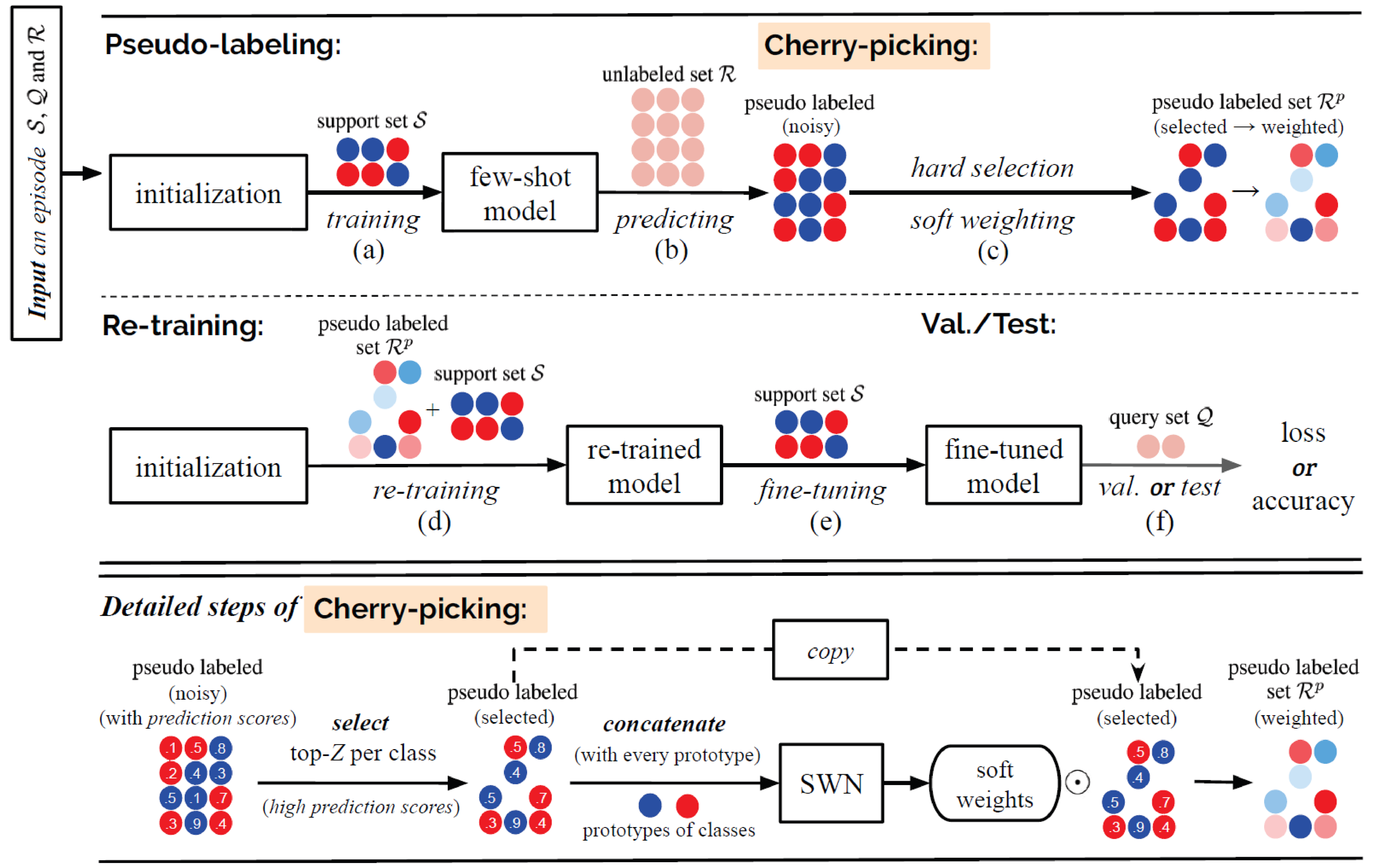
Image from [2]

- **A novel self-training strategy** that prevents the model from drifting due to label noise and enables robust recursive training.
- **A novel meta-learned cherry-picking method** that optimizes the weights of pseudo labels particularly for fast and efficient self-training.
- **Extensive experiments on two benchmarks** --- miniImageNet and tieredImageNet, on which our method achieves the top performance.

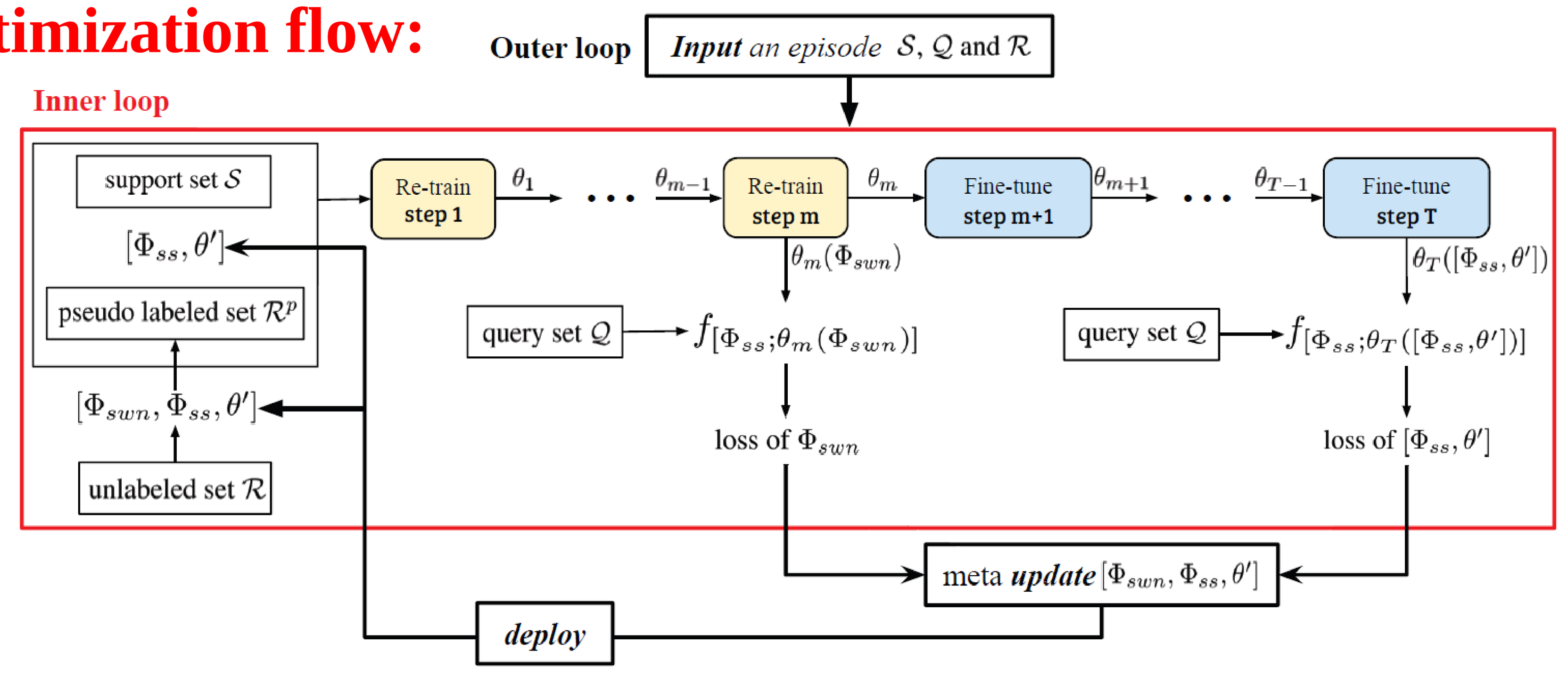
Framework & Optimization flow

Self-Training (inner-loop; base-learning):

- Pseudo-labeling the unlabeled data
- Cherry-picking the better pseudo-labeled data
- Re-training the base-learner with cherry-picked data



Optimization flow:



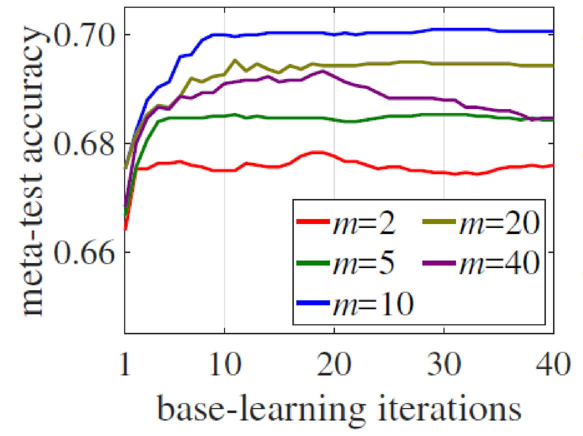
Learning to Self-Train (outer-loop; meta-learning): meta updates!

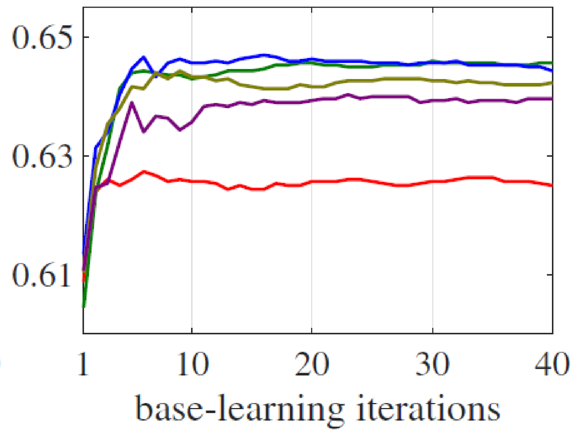
Experiment results on ImageNet-based benchmarks

Classification accuracies (%) in ablative settings (middle blocks), compared to the related SSFSC works (bottom block) with same backbone --- MTL [3]. "fully supervised": using the labels of unlabeled data. "w/*D*": adding unlabeled data from three distracting classes that are excluded in the support set [2, 5].

		miniImageNet		tieredImageNet		mini w/*D*		tiered w/*D*	
		1(shot)	5	1	5	1	5	1	5
fully supervised (upper bound)		80.4	83.3	86.5	88.7	-	-	-	-
no meta	*no selection*	59.7	75.2	67.4	81.1	54.4	73.3	66.1	79.4
	hard	63.0	76.3	69.8	81.5	61.6	75.3	68.8	81.1
	recursive,hard	64.6	77.2	72.1	82.4	61.2	75.7	68.3	81.1
meta	*hard* (Φ_{ss}, θ')	64.1	76.9	74.7	83.2	62.9	75.4	73.4	82.5
	soft	62.8	75.9	73.1	82.8	61.1	74.6	72.1	81.7
	hard,soft	65.0	77.8	75.4	83.4	63.7	76.2	**74.1**	82.9
	recursive,hard,soft	**70.1**	**78.7**	**77.7**	**85.2**	64.1	**77.4**	73.5	83.4
	mixing,hard,soft	66.2	77.9	75.6	84.6	**64.5**	76.5	73.6	**83.8**
Masked Soft k-Means *with* MTL		62.1	73.6	68.6	81.0	61.0	72.0	66.9	80.2
TPN *with* MTL		62.7	74.2	72.1	83.3	61.3	72.4	71.5	82.7
Masked Soft k-Means [2]		50.4	64.4	52.4	69.9	49.0	63.0	51.4	69.1
TPN [5]		52.8	66.4	55.7	71.0	50.4	64.9	53.5	69.9

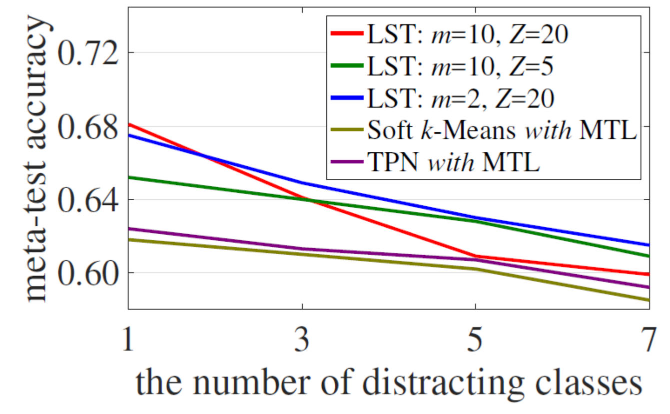
Are the meta-learned *soft* weights of pseudo labels useful?





LST: *recursive,hard,soft* *recursive, hard* (no meta)

How about more distracting classes?



Curves from: miniImageNet, 5-way, 1-shot. (See more settings in the paper)

References

[1] C. Finn *et al*. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, 2017.
[2] M. Ren *et al*. Meta-learning for semi-supervised few-shot classification. In *ICLR*, 2018.
[3] Q. Sun *et al*. Meta-transfer learning for few-shot learning. In *CVPR*, 2019.
[4] C. Olivier *et al*. Semi-supervised learning. Cambridge, Mass.: *MIT Press*, 2006.
[5] Y. Liu *et al*. Transductive propagation network for few-shot learning. In *ICLR*, 2019.

Code is on GitHub