The 38th Annual AAAI Conference on Artificial Intelligence
FEBRUARY 20-27, 2024 | VANCOUVER, CANADA

Roll With the Punches: Expansion and Shrinkage of Soft Label Selection for Semi-supervised Fine-Grained Learning

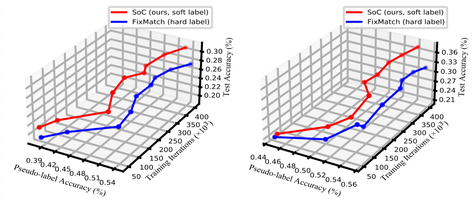
Yue Duan[1], Zhen Zhao[2], Lei Qi[3], Luping Zhou[2], Lei Wang[4], and Yinghuan Shi[1]

[1] Nanjing University, China [2] University of Sydney, Australia [3] Southeast University, China [4] University of Wollongong, Australia

Introduction

Semi-supervised learning (SSL) aims to leverage a pool of unlabeled data to alleviate the dependence of deep models on labeled data. However, *current SSL approaches achieve promising performance with clean and ordinary data, but become unstuck against the indiscernible unlabeled data.* A typical and worth discussing example is the **semi-supervised fine-grained visual classification (SS-FGVC)** [1], where the unlabeled data faced by the SSL model is no longer like the "house" and "bird" that are easy to distinguish, but like "*Streptopelia chinensi*" and "*Streptopelia orientalis*", which are difficult to distinguish accurately even for ornithologists. The mentioned scenario also hints at the practicality of SS-FGVC, i.e., it is resource-consuming to label the fine-grained data for supervised learning.

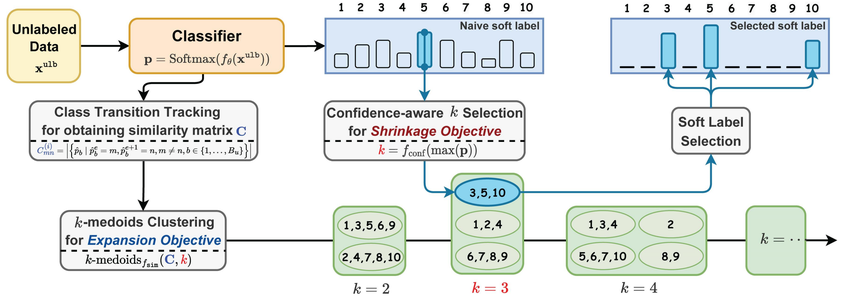
Motivation



(a) Semi-Aves (b) Semi-Fungi

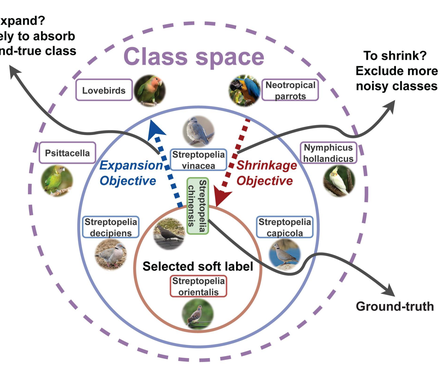
- **Fine-grained data may severely affect the quality of pseudo-labels and consequently pull down the model performance.**
- **The bad effects of incorrect hard label on the model have overridden its low entropy advantage while soft label can benefit the model because it could still provide useful information although it is wrong (as shown on the left).**
- **Using the traditional soft label may robbing Peter to pay Paul, because it contains all the classes, which obviously introduces too much noise into the learning.**

Overview of Method



We select a subset of the class space to serve as the selection range of candidate classes, encouraging the attendance of ground-truth class as much as possible (**Expansion Objective**) while rejecting noisy class as much as possible (**Shrinkage Objective**). Given the unlabeled samples, we perform **Class Transition Tracking** (see Sec. 3.2) to count the transitions of class predictions to obtain the similarity between classes. With obtained similarity, we perform k-medoids clustering [2] on the class space to obtain the clusters of candidate classes, which is used to select the soft pseudo-labels. For different samples, we decide the selection of k based on the confidence scores of their class predictions, where higher confidence corresponds to a larger k, i.e., a smaller selection range of candidate classes.

A Coupled Optimization Goal for SS-FGVC



We divide the class space of the SS-FGVC scenario into clusters with different granularity, where each cluster (e.g., the **blue** circle and **red** circle) contains classes that are more similar to each other. We encourage the soft pseudo-label to select the classes in a cluster with smaller granularity and higher probability of containing ground-truth (i.e., "Streptopelia chinensis"), by optimizing **Expansion Objective** (to absorb more candidate classes) and **Shrinkage Objective** (to shrink the cluster for rejecting noisy classes).

Objective 1 (Expansion Objective). Encourage the pseudo-label to contain the ground-truth class as much as possible ($y_i^\star \in \mathcal{Y}$ is the ground-truth label of x_i^{ulb}):

$$\max_\theta \mathbb{E}_{\mathbf{x}_i^{\mathrm{ulb}} \in \mathcal{D}^{\mathrm{ulb}}} \left[\mathbb{1}(y_i^\star \in \{c \mid g_{i,(c)} = 1\}) p_{i,(y_i^\star)} \right]$$

Objective 2 (Shrinkage Objective). Encourage the pseudo-label to contain as few classes as possible:

$$\min_\theta \mathbb{E}_{\mathbf{x}_i^{\mathrm{ulb}} \in \mathcal{D}^{\mathrm{ulb}}} \sum_{c=1}^{K} g_{i,(c)}$$

Experiments

Dataset	Pseudo-label	Method	Year	from scratch Top1	from scratch Top5	from ImageNet Top1	from ImageNet Top5	from iNat Top1	from iNat Top5
Semi-Aves	—	Supervised oracle	—	57.4±0.3	79.2±0.1	68.5±1.4	88.5±0.4	69.9±0.5	89.8±0.7
		MoCo (He et al. 2020)	CVPR' 20	28.2±0.3	53.0±0.1	52.7±0.1	78.7±0.2	68.6±0.1	87.7±0.1
	Hard label	Pseudo-Label (Lee et al. 2013)	ICML' 13	16.7±0.2	36.5±0.8	54.4±0.3	78.8±0.3	65.8±0.2	86.5±0.2
		Curriculum Pseudo-Label (Cascante-Bonilla et al. 2021)	AAAI' 21	20.5±0.5	41.7±0.5	53.4±0.8	78.3±0.5	69.1±0.3	87.8±0.1
		FixMatch (Sohn et al. 2020)	NIPS' 20	28.1±0.1	51.8±0.6	*57.4±0.8*	78.5±0.5	70.2±0.6	87.0±0.1
		FlexMatch (Zhang et al. 2021)*	NIPS' 21	27.3±0.5	49.7±0.8	53.4±0.2	77.9±0.3	67.6±0.5	87.0±0.2
		MoCo + FlexMatch*	NIPS' 21	35.0±1.2	58.5±1.0	53.4±0.4	77.0±0.2	68.9±0.3	87.7±0.2
	Soft label	KD-Self-Training (Su, Cheng, and Maji 2021)	CVPR' 21	22.4±0.4	44.1±0.1	55.5±0.1	79.8±0.1	67.7±0.2	87.5±0.2
		MoCo + KD-Self-Training (Su, Cheng, and Maji 2021)	CVPR' 21	31.9±0.1	56.8±0.1	55.9±0.2	80.3±0.1	70.1±0.2	88.1±0.1
		SimMatch (Zheng et al. 2022)*	CVPR' 22	24.8±0.5	48.1±0.6	53.3±0.5	77.9±0.8	65.4±0.2	86.9±0.3
		MoCo + SimMatch*	CVPR' 22	32.9±0.4	57.9±0.3	53.7±0.2	78.8±0.5	65.7±0.3	87.1±0.2
		SoC	Ours	*31.3±0.8 (↑11.4%)*	*55.3±0.7 (↑6.8%)*	57.8±0.5 (↑0.7%)	*80.8±0.5 (↑1.3%)*	*71.3±0.3 (↑1.6%)*	*88.8±0.2 (↑1.1%)*
		MoCo + SoC	Ours	**39.3±0.2 (↑12.3%)**	**62.4±0.4 (↑6.7%)**	**58.0±0.4 (↑3.8%)**	**81.7±0.4 (↑1.7%)**	**70.8±0.4 (↑1.0%)**	**88.9±0.5 (↑0.9%)**
Semi-Fungi	—	Supervised oracle	—	60.2±0.8	83.3±0.9	73.3±0.1	92.5±0.3	73.8±0.3	92.4±0.3
		MoCo (He et al. 2020)	CVPR' 20	33.6±0.2	59.4±0.3	55.2±0.2	82.9±0.2	52.5±0.4	79.5±0.2
	Hard label	Pseudo-Label (Lee et al. 2013)	ICML' 13	19.4±0.4	43.2±1.5	51.5±1.2	81.2±0.2	49.5±0.4	78.5±0.2
		Curriculum Pseudo-Label (Cascante-Bonilla et al. 2021)	AAAI' 21	31.4±0.6	55.0±0.6	53.7±0.2	80.2±0.1	53.3±0.5	80.0±0.5
		FixMatch (Sohn et al. 2020)	NIPS' 20	32.2±1.0	57.0±1.2	56.3±0.5	80.4±0.5	58.7±0.7	81.7±0.2
		FlexMatch (Zhang et al. 2021)*	NIPS' 21	36.0±0.9	59.9±1.1	59.6±0.5	82.4±0.5	60.1±0.6	82.2±0.5
		MoCo + FlexMatch*	NIPS' 21	44.2±0.6	67.0±0.8	59.9±0.8	82.8±0.7	61.4±0.6	83.2±0.4
	Soft label	KD-Self-Training (Su, Cheng, and Maji 2021)	CVPR' 21	32.7±0.2	56.9±0.2	56.9±0.3	81.7±0.2	55.7±0.3	82.3±0.2
		MoCo + KD-Self-Training (Su, Cheng, and Maji 2021)	CVPR' 21	39.4±0.3	64.4±0.5	58.2±0.5	84.4±0.2	55.2±0.5	82.9±0.2
		SimMatch (Zheng et al. 2022)*	CVPR' 22	36.5±0.9	61.7±1.0	56.6±0.4	81.8±0.6	56.7±0.3	80.9±0.4
		MoCo + SimMatch*	CVPR' 22	42.2±0.5	67.0±0.4	56.5±0.2	82.5±0.3	57.4±0.2	81.3±0.4
		SoC	Ours	*39.4±2.3 (↑7.9%)*	*62.5±1.1 (↑1.3%)*	*61.4±0.4 (↑3.0%)*	*83.9±0.6 (↑1.8%)*	*62.4±0.2 (↑3.8%)*	*85.1±0.2 (↑3.5%)*
		MoCo + SoC	Ours	**47.2±0.5 (↑6.8%)**	**71.3±0.2 (↑6.4%)**	**61.9±0.3 (↑3.3%)**	**85.8±0.2 (↑3.6%)**	**62.5±0.4 (↑1.8%)**	**84.7±0.2 (↑1.8%)**

We provide comparisons with multiple baseline methods reported in [1] and the state-of-the-art SSL methods based on our re-implementation (marked as $*$). The models are trained from scratch, or ImageNet/iNat pre-trained or the model initialized with MoCo [3] learning on the unlabeled data. Our results are averaged on 3 runs while the standard deviations ±Std. are reported. Meanwhile, we mark out the *best SSL results* and the **best MoCo results**.

References

[1] Su, J.-C.; Cheng, Z.; and Maji, S. 2021. A realistic evaluation of semi-supervised learning for fine-grained classification. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

[2] Kaufman, L.; and Rousseeuw, P. J. 2009. *Finding groups in data: an introduction to cluster analysis*. John Wiley & Sons.

[3] He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. *In IEEE/CVF Conference on Computer Vision and Pattern Recognition*.