CroSel: Cross Selection of Confident Pseudo Labels for Partial-Label Learning

Shiyu Tian, Hongxin Wei, Yiqun Wang, Lei Feng

CVPR SEATTLE, WA JUNE 17-21, 2024

Intro and Motivation

The success of deep learning heavily relies on a massive amount of **fully labeled** data. However, it is **challenging** to obtain a large-scale dataset with completely accurate annotations in the real world.

What is PLL?

Each training example with a candidate label set that includes the true label.

A dog image *x* with candidate label set = {Toy-poodle, Dhole, <u>Bulldog</u>}

Challenge?

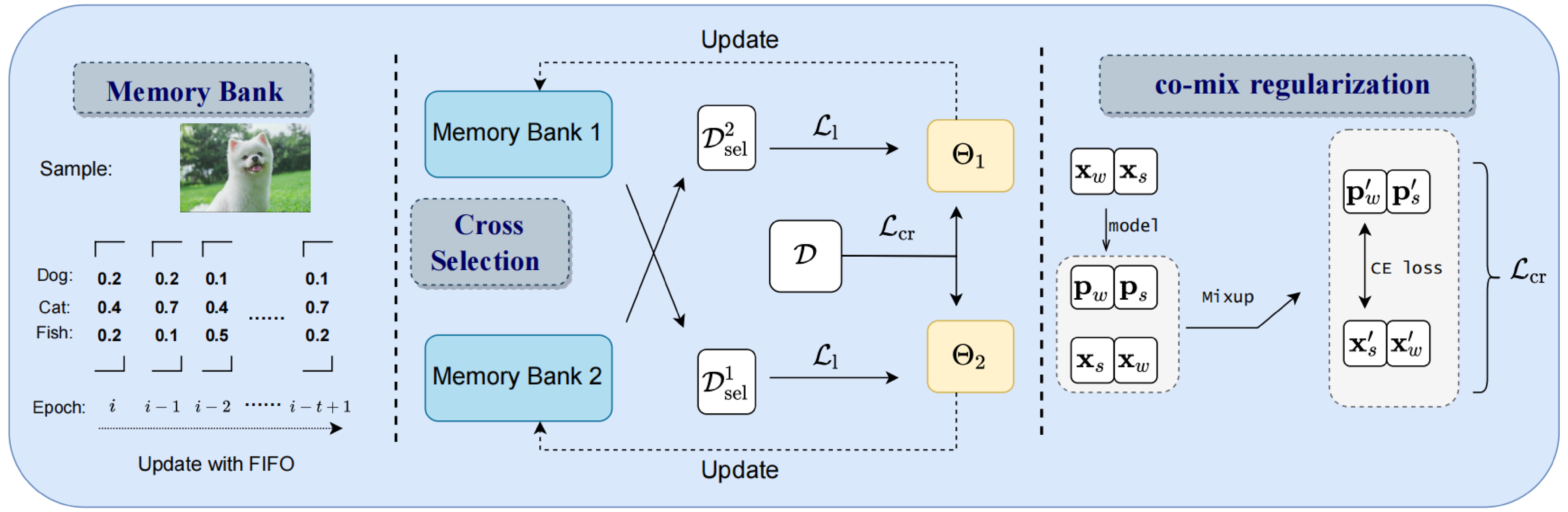
There exists the challenge of **label ambiguity** in PLL. As only one ground-truth label for each training example, and other labels in the candidate label set are actually wrong (**false positive**) labels.

Identify **more** 'true' labels → Train a **better** model

How to select the 'true' label in the candidate label set?

if a model can predict the same label with **high confidence** and **low volatility**, then there is a high probability that the label is the true label of this example.

Methodology



Cross Selection strategy

Our selection criteria consists of three parts：

Partial-label setting

Low volatility level

High confidence level

$$\beta_1 = \mathbb{I}(\operatorname{argmax}(\boldsymbol{q}^i) \in S),$$
$$\beta_2 = \mathbb{I}(\operatorname{argmax}(\boldsymbol{q}^i) = \operatorname{argmax}(\boldsymbol{q}^{i+1})),$$
$$\beta_3 = \mathbb{I}(\frac{1}{t}\sum_{i=1}^{t}\max(\boldsymbol{q}^i) > \gamma),$$

Training Objective for $\mathcal{L}_\text{l}$

$$\mathcal{L}_\text{l} = \frac{1}{|\mathcal{D}_\text{sel}|}\sum_{\boldsymbol{x}\in\mathcal{D}_\text{sel}}\mathcal{L}_\text{CE}(f(\boldsymbol{x}_\text{w}), \hat{y})$$

Co-mix consistency regularization term

based on the idea of **consistency regularization**

use **sharpen** and **mixup** to further augment the data

$$\boldsymbol{p}_i = \begin{cases} \frac{\exp(f_i(\boldsymbol{x})^{\frac{1}{T}})}{\sum_{i\in S}\exp(f_i(\boldsymbol{x})^{\frac{1}{T}})}, & i \in S, \\ 0, & i \notin S, \end{cases}$$

$$\lambda \sim \operatorname{Beta}(\alpha, \alpha),$$
$$\lambda' = \max(\lambda, 1-\lambda),$$
$$\boldsymbol{x}' = \lambda'\boldsymbol{x}_1 + (1-\lambda')\boldsymbol{x}_2,$$
$$\boldsymbol{p}' = \lambda'\boldsymbol{p}_1 + (1-\lambda')\boldsymbol{p}_2,$$

Training Objective for $\mathcal{L}_\text{cr}$

$$\mathcal{L}_\text{cr} = \frac{1}{2n}\sum_{i=1}^{2n}\mathcal{L}_\text{CE}(f(\boldsymbol{x}'_i), \boldsymbol{p}'_i)$$

Overall Training Objective

$$\mathcal{L}_\text{all} = \mathcal{L}_\text{l} + \lambda_\text{d} * \mathcal{L}_\text{cr} \qquad \lambda_\text{d} = (1 - r_\text{s}) * \lambda_\text{cr}$$

This hyperparameter decreases gradually as the picking ratio rises

Experimental results

Whether CroSel can outperform prior PLL methods?

Dataset	q	Ours	PoP	CRDPLL	PiCO	PRODEN	LWS	CC
CIFAR-10	0.1	97.31±.04%	97.17±.01%	**97.41±.06%**	96.10±.06%	95.66±.08%	91.20±.07%	90.73±.10%
	0.3	**97.50±.05%**	97.08±.01%	97.38±.04%	95.74±.10%	95.21±.07%	89.20±.09%	88.04±.06%
	0.5	**97.34±.05%**	96.66±.03%	96.76±.05%	95.32±.12%	94.55±.13%	80.23±.21%	81.01±.38%
SVHN	0.1	**97.71±.05%**	97.55±.06%	97.63±.06%	96.58±.04%	96.20±.07%	96.42±.09%	96.99±.17%
	0.3	**97.96±.05%**	97.50±.03%	97.65±.07%	96.32±.09%	96.11±.05%	96.15±.08%	96.67±.20%
	0.5	**97.86±.06%**	97.31±.01%	97.70±.05%	95.78±.05%	95.97±.03%	95.79±.05%	95.83±.23%
CIFAR-100	0.01	**84.24±.09%**	83.03±.04%	82.95±.10%	74.89±.11%	72.24±.12%	62.03±.21%	66.91±.24%
	0.05	**83.92±.24%**	82.79±.02%	82.38±.09%	73.26±.09%	70.03±.18%	57.10±.17%	64.51±.37%
	0.10	**84.07±.16%**	82.39±.04%	82.15±.20%	70.03±.10%	69.82±.11%	52.60±.54%	61.50±.36%

Datasets	Setting	Index	Performance
CIFAR-10	$q = 0.1$	S-ratio	99.09±.07%
		S-acc	99.79±.05%
	$q = 0.3$	S-ratio	98.10±.10%
		S-acc	99.55±.03%
	$q = 0.5$	S-ratio	96.25±.12%
		S-acc	99.44±.06%
SVHN	$q = 0.1$	S-ratio	97.25±.14%
		S-acc	99.84±.06%
	$q = 0.3$	S-ratio	76.42±.21%
		S-acc	99.77±.06%
	$q = 0.5$	S-ratio	73.21±.15%
		S-acc	99.34±.02%
CIFAR-100	$q = 0.01$	S-ratio	96.58±.13%
		S-acc	99.71±.06%
	$q = 0.05$	S-ratio	95.45±.21%
		S-acc	98.29±.15%
	$q = 0.10$	S-ratio	93.61±.12%
		S-acc	97.93±.11%

CroSel achieves **the best results** on different settings and shows great selection accuracy and selection ratio of pseudo labels.

Does each CroSel component work effectively?

Table 4. Results of thorough ablation experiments.

cr1	cr2	cr3	$\mathcal{L}_\text{cr}$	Acc	S-acc	S-ratio
✓				73.12%	95.15%	90.32%
✓	✓			72.00%	93.51%	85.09%
✓	✓	✓		70.68%	96.22%	78.65%
✓			✓	77.04%	91.02%	97.27%
✓	✓		✓	79.90%	94.66%	95.51%
✓	✓	✓	✓	**84.07%**	**97.93%**	**93.61%**

Table 5. Results for ablation studies on the scope of regularization.

Setting	Scope	Index	Performance
CIFAR-10 $q = 0.5$	All data	Acc	**97.34%**
		S-ratio	**96.25%**
		S-acc	**99.44%**
	Unselected data	Acc	90.32%
		S-ratio	93.27%
		S-acc	95.72%
	None	Acc	81.01%
		S-ratio	90.23%
		S-acc	89.72%
CIFAR-100 $q = 0.1$	All data	Acc	**84.07%**
		S-ratio	**93.61%**
		S-acc	**97.93%**
	Unselected data	Acc	77.61%
		S-ratio	90.12%
		S-acc	97.63%
	None	Acc	70.68%
		S-ratio	78.65%
		S-acc	96.22%

The selection strategy and regularization term **reciprocally promote each other.**

Code is available here https://github.com/jokersio-tsy/CroSel