Noise Separation guided Candidate Label Reconstruction for Noisy Partial Label Learning

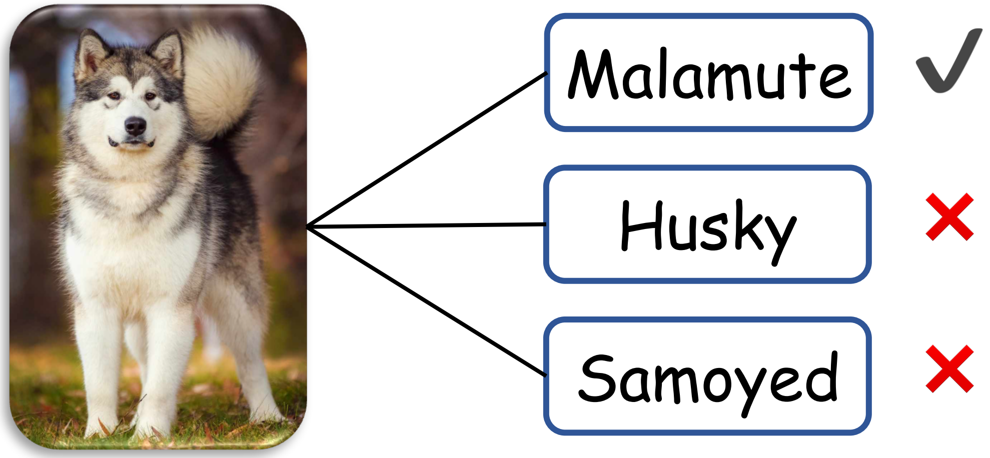
Xiaorui Peng[1], Yuheng Jia[2,3], FuChao Yang[1], Ran Wang[4,5], Min-Ling Zhang[2,6]

[1] College of Software Engineering, Southeast University
[2] School of Computer Science and Engineering, Southeast University
[3] Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University)
[4] Shenzhen Key Laboratory of Advanced Machine Learning and Applications, School of Mathematical Sciences, Shenzhen University
[5] Guangdong Provincial Key Laboratory of Intelligent Information Processing, Shenzhen University
[6] Key Laboratory of Computer Network and Information Integration (Southeast University)

Noisy Partial Label Learning (NPLL)

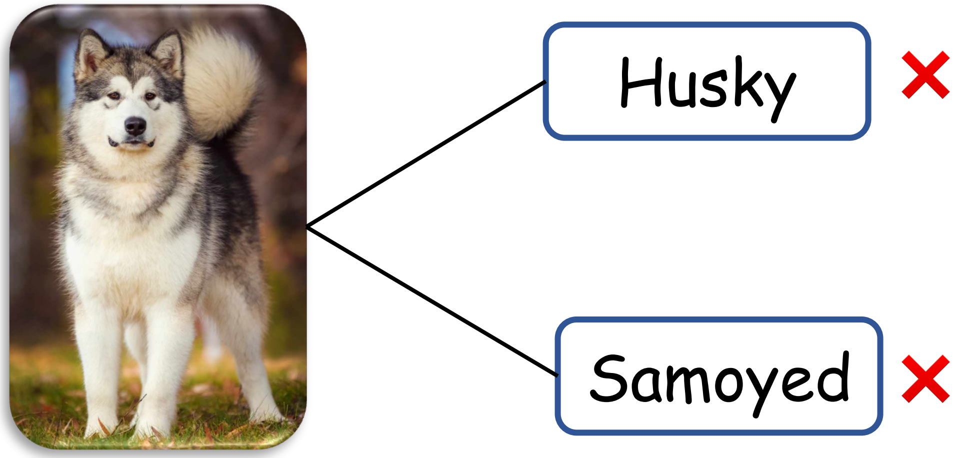
Partial Label Learning (PLL)

- a set of candidate labels
- only one valid (unknown)



Noisy Partial Label Learning (NPLL)

- a set of candidate labels
- only one valid or no one valid (unknown)



Motivation

Theorem 1. *Assume the loss function $\mathcal{L}(f(\boldsymbol{x}), y)$ is ρ-Lipschitz with respect to $f(\boldsymbol{x})$ for all $y \in \mathcal{Y}$ and uppper-bounded by M. For noise rate $0 < \epsilon < 1$ and mean CLS size for normal samples $1 < \alpha < C$, for any $\delta > 0$, with probability at least $1 - \delta$, we have*

$$R(\hat{f}) - R(f^*) \leq 2(1 - \frac{1-\epsilon}{\alpha})M + 4\sqrt{2}\rho \sum_{y=1}^{C} \mathfrak{R}_n(\mathcal{H}_y) + 2M\sqrt{\frac{log\frac{2}{\delta}}{2n}}.$$

The proof of Theorem 1 is provided in Appendix A.1. It can be observed that the generalization performance of $\hat{f}$ is primarily influenced by three factors: the noise rate ϵ, the mean CLS size α of normal samples, and the sample size n. As $n \to \infty$, $\epsilon \to 0$ and $\alpha \to 1$, Theorem 1 shows that the generalized error bound will be reduced, and the empirical risk minimizer $\hat{f}$ will get closer to the true risk minimizer f^*. **Obviously, a smaller noise rate ϵ and a smaller CLS size α will bring better generalization performance.**

The Proposed Method

Target : $\mathcal{D} = \{(\boldsymbol{x_i}, Y_i)\}_{i=1}^{n} \Rightarrow \mathcal{D} = \{(\boldsymbol{x_i}, \widehat{Y}_i)\}_{i=1}^{n} \Rightarrow \min_{\theta} \frac{1}{n}\sum_{i=1}^{n} \mathcal{L}_{PLL}(f(\boldsymbol{x_i};\theta), \widehat{Y}_i),$

Progressive Sample Separation

KNN-based pseudo label:

$$\boldsymbol{q_i} = Normalize\left(\sum_{j \in N_i} s_{ij}\boldsymbol{p_j}\right)$$

CLS-based pseudo label:

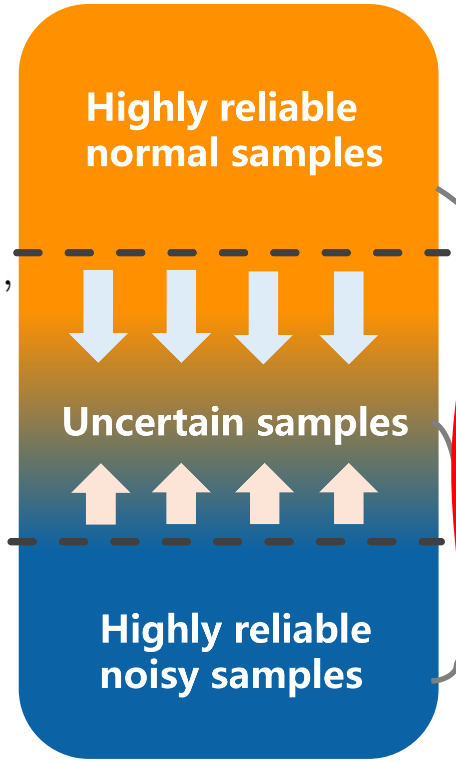
$$\widetilde{\boldsymbol{q}}_{\boldsymbol{i}} = Normalize\left(\boldsymbol{p_i} \odot \boldsymbol{S}(Y_i)\right),$$

consistency error:

$$E_i = -\sum_{j=1}^{C} \widetilde{q}_{ij} \log q_{ij},$$

Sample Separation with moving threshold

$$v_i = \begin{cases} 1, & E_i \le l, \\ 0, & l < E_i \le u, \\ -1, & E_i > u, \end{cases}$$



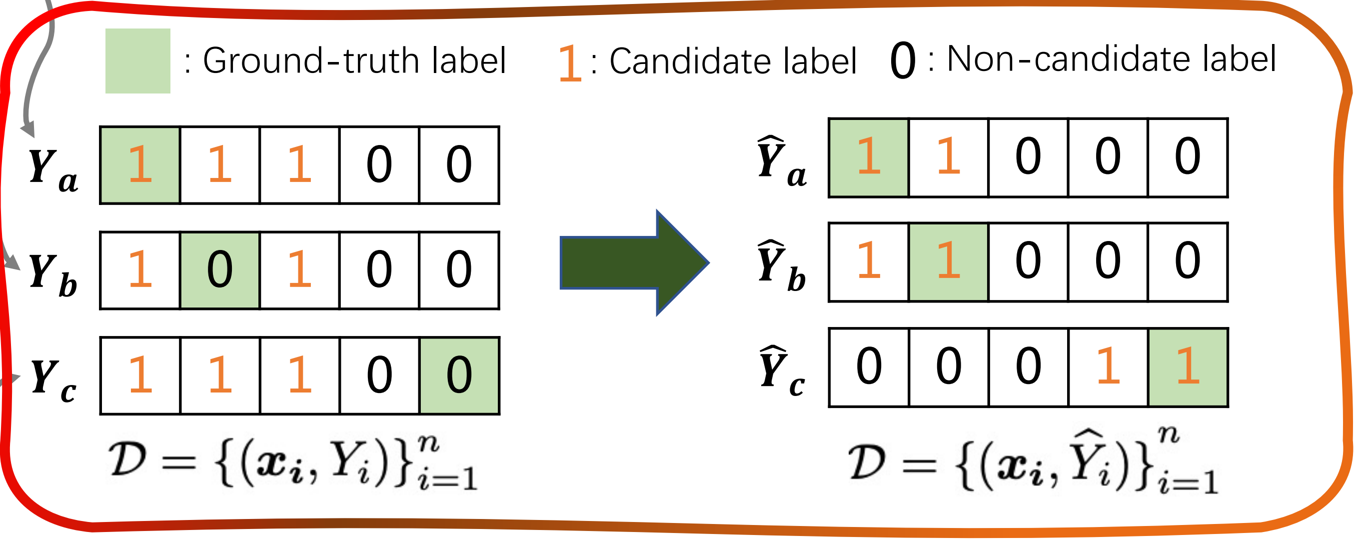
Reconstruction of Candidate Label Set

$$\min_{\widehat{Y}} \sum_{i=1}^{n} I(v_i \neq 0)\left(|\widehat{Y}_i| - \beta\langle \boldsymbol{q_i}, \boldsymbol{S}(\widehat{Y}_i)\rangle\right),$$
$$\text{s.t.} \quad \forall i, \text{if } v_i = 1, \widehat{Y}_i \neq \varnothing, \widehat{Y}_i \subseteq Y_i,$$
$$\forall i, \text{if } v_i = -1, \widehat{Y}_i \neq \varnothing, \widehat{Y}_i \subseteq \mathcal{Y} - Y_i,$$

$$\beta_i = \tau(\boldsymbol{q_i}),$$
$$\widehat{Y}_i = \{j | j \in Y_i, \ q_{ij} > \frac{1}{\beta_i}\}, \quad \text{if } v_i = 1,$$
$$\widehat{Y}_i = \{j | j \in \mathcal{Y} - Y_i, \ q_{ij} > \frac{1}{\beta_i}\}, \quad \text{if } v_i = -1.$$

$$c = \arg\max_{j \in \mathcal{Y} - Y_i}$$
$$c' = \arg\min_{j \in Y_i} q_{ij}$$

$$\widehat{Y}_i = Y_i \cup \{c\} \setminus \{c'\}, \quad \text{if } v_i = 0,$$



The Proposed Method

Target: $\mathcal{D}=\{(\boldsymbol{x_i}, Y_i)\}_{i=1}^{n} \rightarrow \mathcal{D}=\{(\boldsymbol{x_i}, \widehat{Y}_i)\}_{i=1}^{n} \rightarrow \min_{\theta} \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}_{PLL}(f(\boldsymbol{x_i};\theta), \widehat{Y}_i),$

TOP

Progressive Sample Separation

KNN-based pseudo label:

$$\boldsymbol{q_i} = Normalize\left(\sum_{j\in N_i} s_{ij}\boldsymbol{p_j}\right)$$

CLS-based pseudo label:

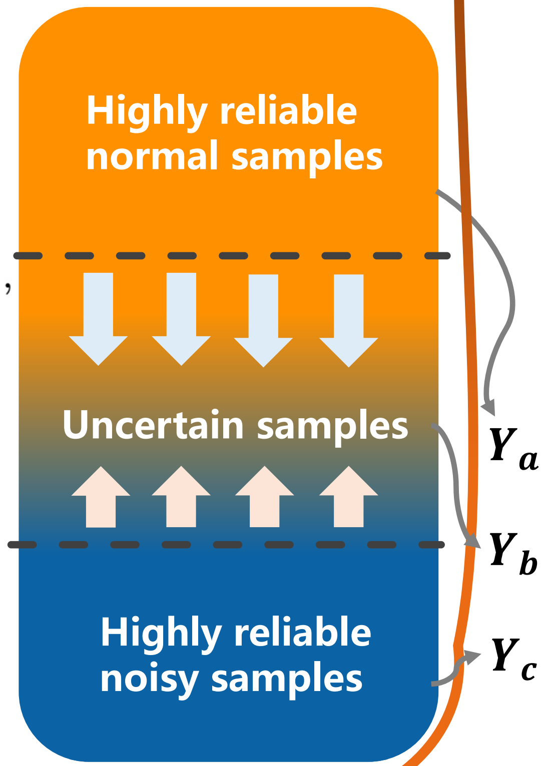
$$\widetilde{\boldsymbol{q}}_{\boldsymbol{i}} = Normalize\left(\boldsymbol{p_i}\odot \boldsymbol{S}(Y_i)\right),$$

consistency error:

$$E_i = -\sum_{j=1}^{C}\widetilde{q}_{ij}\log q_{ij},$$

Sample Separation with moving threshold

$$v_i = \begin{cases} 1, & E_i \le l, \\ 0, & l < E_i \le u, \\ -1, & E_i > u, \end{cases}$$



Reconstruction of Candidate Label Set

$$\min_{\widehat{Y}} \sum_{i=1}^{n} I(v_i \neq 0)\left(|\widehat{Y}_i| - \beta\langle \boldsymbol{q_i}, \boldsymbol{S}(\widehat{Y}_i)\rangle\right),$$

s.t. $\forall i$, if $v_i = 1, \widehat{Y}_i \neq \varnothing, \widehat{Y}_i \subseteq Y_i,$

$\forall i$, if $v_i = -1, \widehat{Y}_i \neq \varnothing, \widehat{Y}_i \subseteq \mathcal{Y} - Y_i,$

$\rightarrow$

$$\beta_i = \tau(\boldsymbol{q_i}),$$

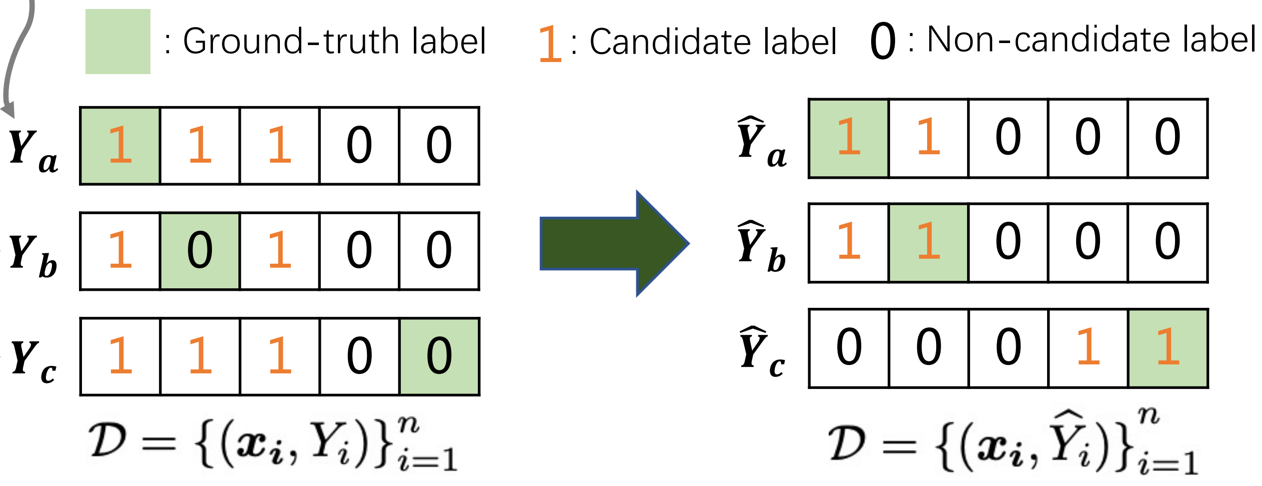
$$\widehat{Y}_i = \{j | j \in Y_i,\ q_{ij} > \frac{1}{\beta_i}\}, \quad \text{if } v_i = 1,$$

$$\widehat{Y}_i = \{j | j \in \mathcal{Y} - Y_i,\ q_{ij} > \frac{1}{\beta_i}\}, \quad \text{if } v_i = -1.$$

$$c = \arg\max_{j\in\mathcal{Y}-Y_i}$$
$$c' = \arg\min_{j\in Y_i} q_{ij}$$

$\rightarrow$

$$\widehat{Y}_i = Y_i \cup \{c\} \setminus \{c'\}, \quad \text{if } v_i = 0,$$



The Proposed Method

Target : $\mathcal{D}=\{(\boldsymbol{x_i},Y_i)\}_{i=1}^{n} \Rightarrow \mathcal{D}=\{(\boldsymbol{x_i},\widehat{Y}_i)\}_{i=1}^{n} \Rightarrow \min_{\theta} \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}_{PLL}(f(\boldsymbol{x_i};\theta),\widehat{Y}_i),$

Progressive Sample Separation

KNN-based pseudo label:

$$\boldsymbol{q_i} = Normalize\left(\sum_{j\in N_i} s_{ij}\boldsymbol{p_j}\right)$$

CLS-based pseudo label:

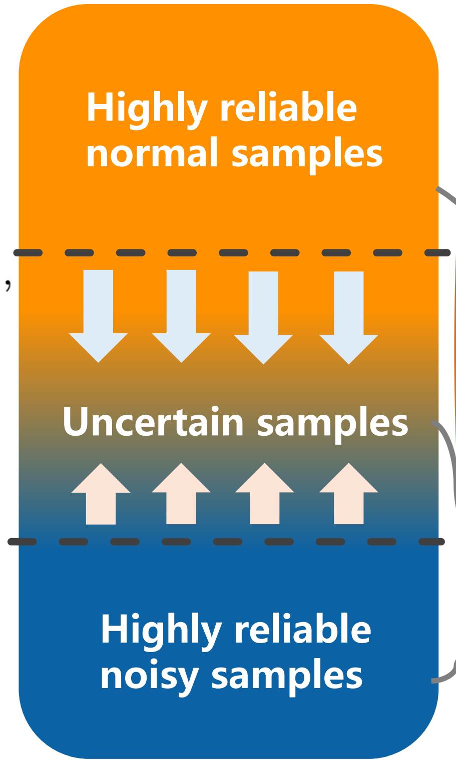
$$\widetilde{\boldsymbol{q}}_{\boldsymbol{i}} = Normalize\left(\boldsymbol{p_i}\odot \boldsymbol{S}(Y_i)\right),$$

consistency error:

$$E_i = -\sum_{j=1}^{C}\widetilde{q}_{ij}\log q_{ij},$$

Sample Separation with moving threshold

$$v_i = \begin{cases} 1, & E_i \le l, \\ 0, & l < E_i \le u, \\ -1, & E_i > u, \end{cases}$$



Reconstruction of Candidate Label Set

$$\min_{\widehat{Y}} \sum_{i=1}^{n} I(v_i\neq 0)\left(|\widehat{Y}_i| - \beta\langle \boldsymbol{q_i}, \boldsymbol{S}(\widehat{Y}_i)\rangle\right),$$
$$\text{s.t.}\quad \forall i, \text{if } v_i=1, \widehat{Y}_i\neq\varnothing, \widehat{Y}_i\subseteq Y_i,$$
$$\forall i, \text{if } v_i=-1, \widehat{Y}_i\neq\varnothing, \widehat{Y}_i\subseteq \mathcal{Y}-Y_i,$$

$\Rightarrow$

$$\beta_i = \tau(\boldsymbol{q_i}),$$
$$\widehat{Y}_i = \{j \mid j\in Y_i,\ q_{ij} > \frac{1}{\beta_i}\},\quad \text{if } v_i = 1,$$
$$\widehat{Y}_i = \{j \mid j\in \mathcal{Y}-Y_i,\ q_{ij} > \frac{1}{\beta_i}\},\quad \text{if } v_i = -1.$$

$$c = \arg\max_{j\in\mathcal{Y}-Y_i}$$
$$c' = \arg\min_{j\in Y_i} q_{ij}$$

$\Rightarrow$

$$\widehat{Y}_i = Y_i \cup \{c\}\setminus\{c'\},\quad \text{if } v_i = 0,$$

: Ground-truth label　1 : Candidate label　0 : Non-candidate label

Y_a	1 (ground-truth)	1	1	0	0
Y_b	1	0 (ground-truth)	1	0	0
Y_c	1	1	1	0	0 (ground-truth)

$\mathcal{D}=\{(\boldsymbol{x_i},Y_i)\}_{i=1}^{n}$

$\Rightarrow$

$\widehat{Y}_a$	1 (ground-truth)	1	0	0	0
$\widehat{Y}_b$	1	1 (ground-truth)	0	0	0
$\widehat{Y}_c$	0	0	0	1	1 (ground-truth)

$\mathcal{D}=\{(\boldsymbol{x_i},\widehat{Y}_i)\}_{i=1}^{n}$

Experimental results

Table 1: Accuracy comparisons on CIFAR10 and CIFAR100 under various ambiguity levels η and noise levels γ. Bold indicates the best result. Accuracies are presented in percentage (%) form. All experiments were conducted three times under the same three distinct random seeds.

Method	CIFAR10								
	$\eta = 0.3$			$\eta = 0.4$			$\eta = 0.5$		
	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$
PRODEN [ICML'20]	77.74 ± 0.53	67.20 ± 0.99	57.74 ± 0.56	71.43 ± 0.54	59.28 ± 0.82	46.87 ± 1.40	63.94 ± 0.75	49.38 ± 1.13	32.03 ± 1.33
CC [NeurIPS'20]	75.09 ± 0.37	63.48 ± 1.72	54.42 ± 0.34	68.08 ± 0.94	54.46 ± 0.36	42.24 ± 1.31	58.22 ± 0.24	44.38 ± 1.60	28.57 ± 1.67
CRDPLL [IMCL'23]	84.61 ± 0.19	80.12 ± 0.46	71.43 ± 0.93	81.60 ± 0.46	72.79 ± 0.39	53.24 ± 2.30	76.92 ± 1.04	56.78 ± 0.76	32.60 ± 1.04
PaPi [CVPR'23]	89.80 ± 0.36	86.36 ± 1.06	78.45 ± 0.61	86.71 ± 0.65	81.78 ± 0.52	59.02 ± 1.67	86.34 ± 0.67	73.06 ± 1.16	47.16 ± 1.35
FREDIS [ICML'23]	92.09 ± 0.29	87.91 ± 1.74	84.15 ± 0.19	89.25 ± 2.18	84.78 ± 2.50	77.74 ± 0.70	88.10 ± 0.59	79.73 ± 2.70	52.68 ± 1.22
PiCO+ [TPAMI'24]	94.12 ± 0.35	94.22 ± 1.19	89.56 ± 0.52	93.84 ± 0.96	92.96 ± 0.92	85.94 ± 1.48	92.21 ± 0.66	89.63 ± 1.47	75.59 ± 1.32
ALIM-Scale [NeurIPS'23]	94.97 ± 0.27	94.10 ± 0.16	93.31 ± 0.27	94.38 ± 0.24	93.41 ± 0.25	89.62 ± 0.66	93.92 ± 0.09	90.10 ± 0.56	69.78 ± 1.07
ALIM-Onehot [NeurIPS'23]	95.13 ± 0.10	94.39 ± 0.23	93.68 ± 0.14	94.64 ± 0.08	94.17 ± 0.04	88.88 ± 0.30	94.07 ± 0.15	90.71 ± 0.73	65.57 ± 2.04
Ours	$\mathbf{96.91 \pm 0.17}$	$\mathbf{96.80 \pm 0.14}$	$\mathbf{96.47 \pm 0.19}$	$\mathbf{96.78 \pm 0.10}$	$\mathbf{96.23 \pm 0.66}$	$\mathbf{96.03 \pm 0.56}$	$\mathbf{96.55 \pm 0.02}$	$\mathbf{94.54 \pm 1.84}$	$\mathbf{82.63 \pm 1.70}$

Method	CIFAR100								
	$\eta = 0.03$			$\eta = 0.05$			$\eta = 0.1$		
	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$
PRODEN [ICML'20]	57.83 ± 0.49	48.66 ± 0.31	40.10 ± 0.37	55.39 ± 0.61	45.36 ± 1.16	36.11 ± 0.40	50.88 ± 1.12	40.02 ± 1.40	28.81 ± 0.89
CC [NeurIPS'20]	57.73 ± 0.70	48.66 ± 0.28	38.26 ± 1.31	55.93 ± 0.70	45.41 ± 1.23	35.31 ± 0.07	51.81 ± 0.36	40.69 ± 0.65	28.56 ± 0.29
CRDPLL [IMCL'23]	63.91 ± 0.53	59.16 ± 0.14	55.16 ± 0.36	63.02 ± 0.52	57.77 ± 0.48	53.64 ± 0.29	61.43 ± 0.21	54.77 ± 0.05	48.50 ± 0.36
PaPi [CVPR'23]	69.83 ± 0.57	61.99 ± 0.24	59.71 ± 0.68	68.64 ± 0.61	62.72 ± 0.95	58.63 ± 0.25	67.64 ± 0.56	61.98 ± 0.70	55.60 ± 0.51
FREDIS [ICML'23]	66.94 ± 0.10	61.85 ± 0.41	57.99 ± 0.35	67.48 ± 0.57	62.72 ± 0.77	57.19 ± 0.68	66.09 ± 0.42	57.60 ± 0.64	45.09 ± 0.72
PiCO+ [TPAMI'24]	74.32 ± 0.43	72.68 ± 0.28	67.31 ± 0.58	73.33 ± 0.48	70.17 ± 0.62	65.01 ± 0.48	62.67 ± 0.46	56.25 ± 0.84	47.75 ± 1.08
ALIM-Scale [NeurIPS'23]	76.39 ± 0.71	75.40 ± 0.60	74.58 ± 0.25	76.02 ± 0.31	75.33 ± 0.14	74.49 ± 0.69	75.27 ± 0.22	71.06 ± 1.41	64.61 ± 2.37
ALIM-Onehot [NeurIPS'23]	76.29 ± 0.19	74.83 ± 0.12	73.39 ± 1.14	74.92 ± 0.48	74.40 ± 0.06	71.49 ± 1.02	61.24 ± 0.57	58.01 ± 1.03	47.27 ± 1.82
Ours	$\mathbf{81.74 \pm 0.16}$	$\mathbf{80.73 \pm 0.16}$	$\mathbf{79.95 \pm 0.20}$	$\mathbf{80.76 \pm 0.08}$	$\mathbf{80.17 \pm 0.20}$	$\mathbf{78.89 \pm 0.41}$	$\mathbf{79.98 \pm 0.23}$	$\mathbf{79.21 \pm 0.88}$	$\mathbf{76.18 \pm 1.67}$

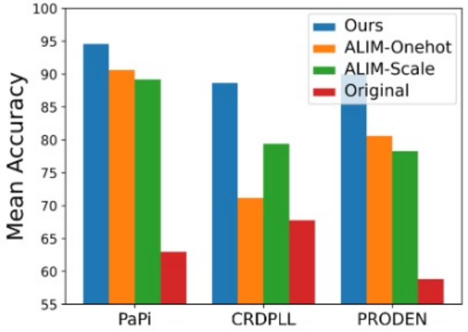
Experimental results

Table 2: Accuracy comparisons when the methods are used as a plug-in on CIFAR10 and CIFAR100 under various ambiguity levels η and noise levels γ. Bold indicates the best result. Accuracies are presented in percentage (%) form.

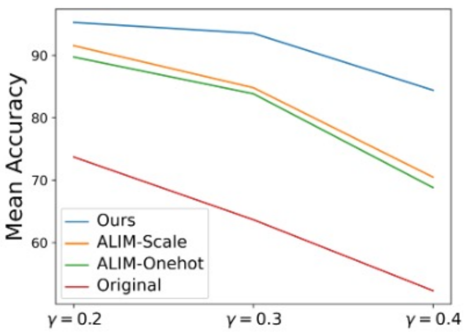
Method	CIFAR10								
	$\eta = 0.3$			$\eta = 0.4$			$\eta = 0.5$		
	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$
PRODEN	78.00	67.57	57.75	71.31	60.22	48.38	64.62	49.95	31.93
PRODEN + ALIM-Onehot	90.83	88.64	84.87	89.15	84.95	77.71	86.63	79.89	42.83
PRODEN + ALIM-Scale	92.05	89.83	83.22	90.58	85.78	71.27	87.10	66.34	38.14
PRODEN + Ours	**94.35**	**94.10**	**93.30**	**94.21**	**93.80**	**90.48**	**94.00**	**93.27**	**62.29**
CRDPLL	84.40	79.61	71.46	81.97	72.43	55.06	76.93	56.40	31.96
CRDPLL + ALIM-Onehot	88.30	83.64	74.21	86.12	77.04	56.58	80.75	60.72	32.95
CRDPLL + ALIM-Scale	92.06	90.42	85.86	90.81	85.36	73.85	87.06	68.57	40.50
CRDPLL + Ours	**95.29**	**95.25**	**94.86**	**95.17**	**94.44**	**90.43**	**94.28**	**83.40**	**54.66**
PaPi	69.83	61.99	59.71	68.64	62.72	58.63	67.64	61.98	55.60
PaPi + ALIM-Onehot	95.64	94.88	92.57	95.67	94.26	91.01	94.45	90.50	66.50
PaPi + ALIM-Scale	95.11	94.40	92.32	95.94	94.15	90.25	93.24	88.41	58.92
PaPi + Ours	**96.74**	**96.67**	**96.41**	**96.84**	**95.47**	**95.39**	**96.53**	**95.41**	**81.85**

Method	CIFAR100								
	$\eta = 0.03$			$\eta = 0.05$			$\eta = 0.1$		
	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$
PRODEN	58.10	48.98	40.30	54.89	46.60	35.85	51.87	41.63	29.84
PRODEN + ALIM-Onehot	74.07	71.18	68.26	72.47	69.76	66.36	68.17	62.33	53.44
PRODEN + ALIM-Scale	74.98	73.50	69.05	74.49	72.14	66.88	70.64	64.17	55.10
PRODEN + Ours	**75.66**	**74.59**	**72.29**	**75.08**	**73.68**	**70.19**	**74.74**	**69.93**	**61.47**
CRDPLL	64.36	59.01	55.20	62.45	58.26	53.37	61.26	54.72	48.69
CRDPLL + ALIM-Onehot	68.31	63.70	57.20	67.21	64.24	55.37	67.09	62.28	50.19
CRDPLL + ALIM-Scale	70.93	67.99	58.60	70.09	67.30	57.43	68.24	63.03	53.74
CRDPLL + Ours	**74.46**	**72.93**	**69.96**	**73.58**	**72.39**	**68.19**	**72.27**	**69.55**	**66.16**
PaPi	69.83	61.99	59.71	68.64	62.72	58.63	67.64	61.98	55.60
PaPi + ALIM-Onehot	80.37	79.20	77.58	79.64	78.14	76.08	77.77	74.04	58.63
PaPi + ALIM-Scale	81.50	80.23	78.77	80.51	78.79	77.11	79.16	75.86	63.02
PaPi + Ours	**81.70**	**80.69**	**80.08**	**80.85**	**79.88**	**79.29**	**79.78**	**78.23**	**78.10**

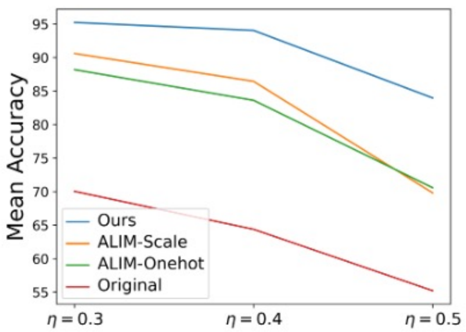
Experimental results



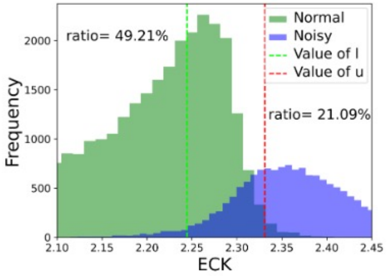
(a) varying PLL methods



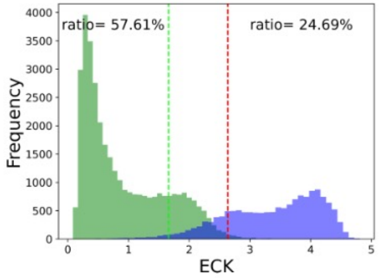
(b) varying noise level



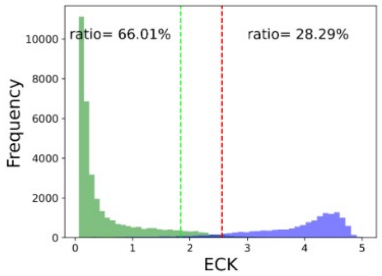
(c) varying ambiguity level



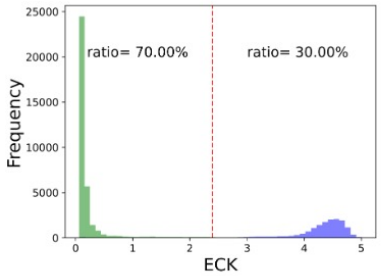
(a) 100 epoch



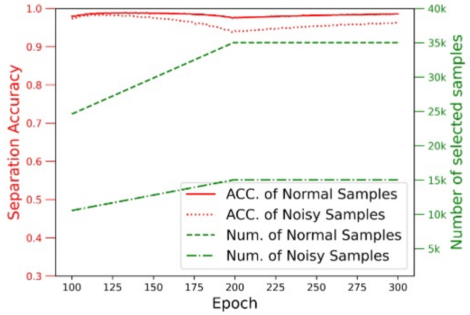
(b) 140 epoch



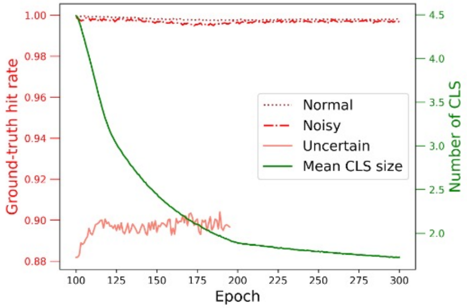
(c) 180 epoch



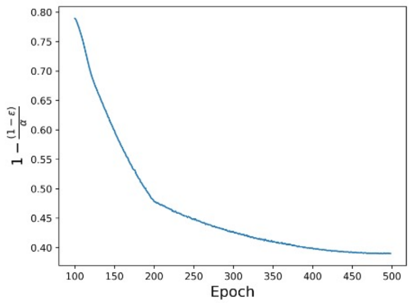
(d) 500 epoch



(a) Separation quantity vs Separation quality



(b) Ground-truth hit rate vs Mean CLS size



(c) The first term of Theorem 1

Experimental results

Table 3: Accuracies (%) on fine-grained datasets.

Method	CIFAR100H $\eta = 0.5$ $\gamma = 0.2$	CUB-200 $\eta = 0.03$ $\gamma = 0.3$	Flower $\eta = 0.05$ $\gamma = 0.2$
PaPi	63.94	43.56	74.95
PaPi + ALIM-Onehot	69.29	48.58	76.07
PaPi + ALIM-Scale	74.34	51.44	78.47
PaPi + Ours	**76.93**	**52.78**	**81.72**

Table 4: Accuracies (%) on real-world datasets.

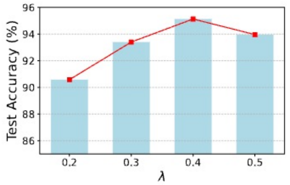
Method\Dataset	Treeversity2	Treeversity3	Benthic2#
Papi	81.07	82.55	80.90
PaPi + ALIM-Scale	82.72	83.47	81.46
PaPi + ALIM-Onehot	84.54	86.01	82.24
PaPi + Ours	**86.41**	**86.67**	**83.47**

Table 7: The separation accuracy of **normal/noisy samples** for each NPLL method.

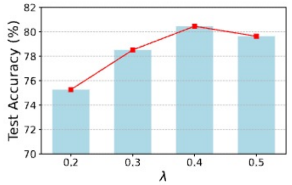
Method\Dataset	CIFAR10 (γ=0.2, η=0.5)	CIFAR10 (γ=0.3, η=0.5)	CIFAR10 (γ=0.4, η=0.5)
PiCO+	99.21%/48.90%	96.25%/69.14%	83.92%/76.11%
ALIM-Onehot	98.53%/94.44%	97.31%/93.36%	86.80%/79.33%
Ours	**99.35%/97.85%**	**98.91%/97.08%**	**92.68%/88.05%**

Table 5: Ablation study of our method (%).

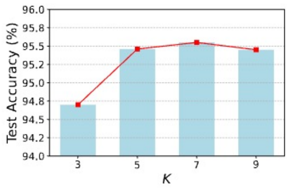
Method	CIFAR10 $\eta = 0.5$ $\gamma = 0.3$	CIFAR100 $\eta = 0.05$ $\gamma = 0.3$
Ours	**95.41**	**79.88**
Ours *v1*	94.09	78.68
Ours *v2*	95.39	75.16
Ours *v3*	81.32	74.62
Papi	61.98	62.72



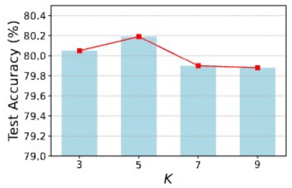
(c) λ on CIFAR10



(d) λ on CIFAR100



(g) K on CIFAR10



(h) K on CIFAR100

THANKS

Email: xiaorui@seu.edu.cn