(ICLR 25') Complementary Label Learning with Positive Label Guessing and Negative Label Enhancement

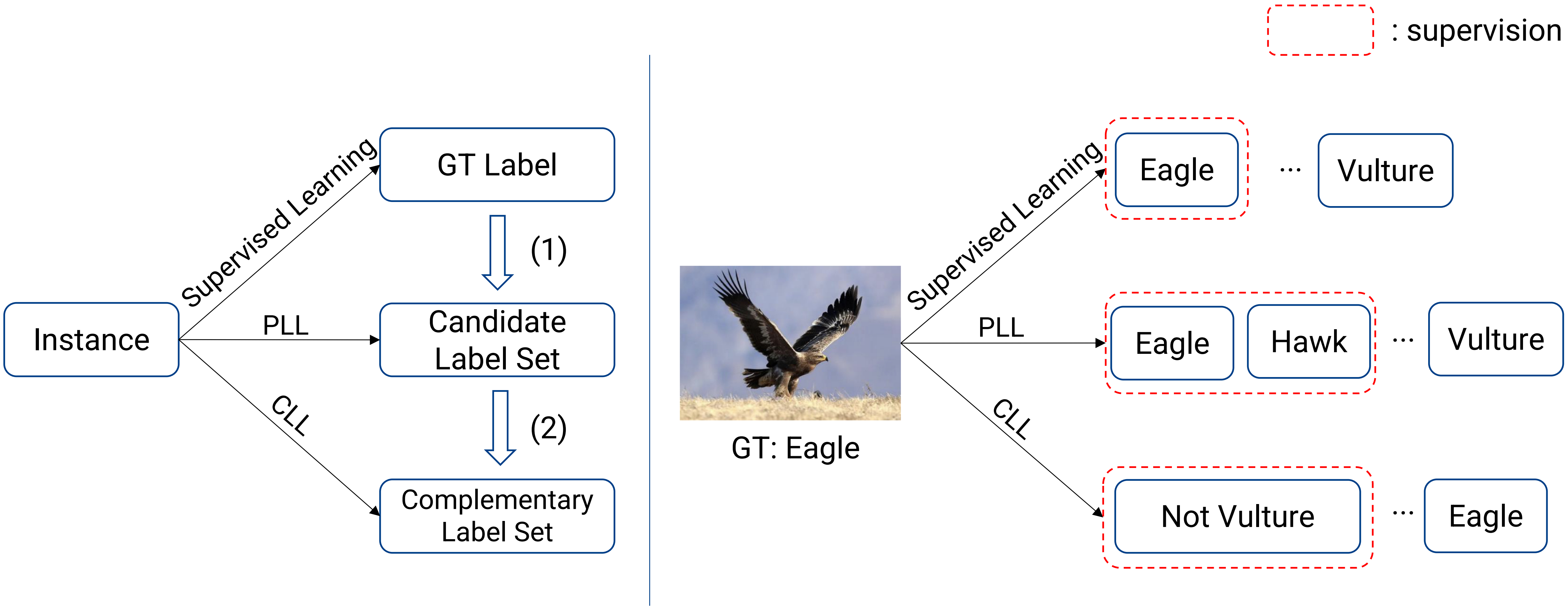
Yuhang Li (李宇航)
Southeast University
Nanjing, China

CONTENTS

01

Problem Setup

➢ Imperfect Information: Partial Label Learning & Complementary Label Learning (PLL & CLL)



(1): Reduce the need of expertise;

(2): Further reduce the need of expertise; Privacy need.

➢ Imperfect Information: Partial Label Learning & Complementary Label Learning (PLL & CLL)

- **Goal: Learn a classifier that can minimize the estimated risk on training set.**

Multi-Class Classification

$$\widehat{R}(f) = \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}(f(\boldsymbol{x}), y_i).$$

Complementary Label Learning

$$\widehat{R}(f) = \frac{1}{N}\sum_{i=1}^{N}\bar{\mathcal{L}}_{CLL}(f(\boldsymbol{x}_i), \bar{Y}_i),$$

- **Previous Methods**

1. Design more suitable and solid **Loss functions (risk minimizer)**; Weakness: **Static**
2. Better **Representation Learning**. Weakness: **Training-oriented**

- **Our Method – PLNL (PLG and NLE)**

1. **Status-aware**
2. **Data-oriented**
3. **Theoretical-guaranteed**

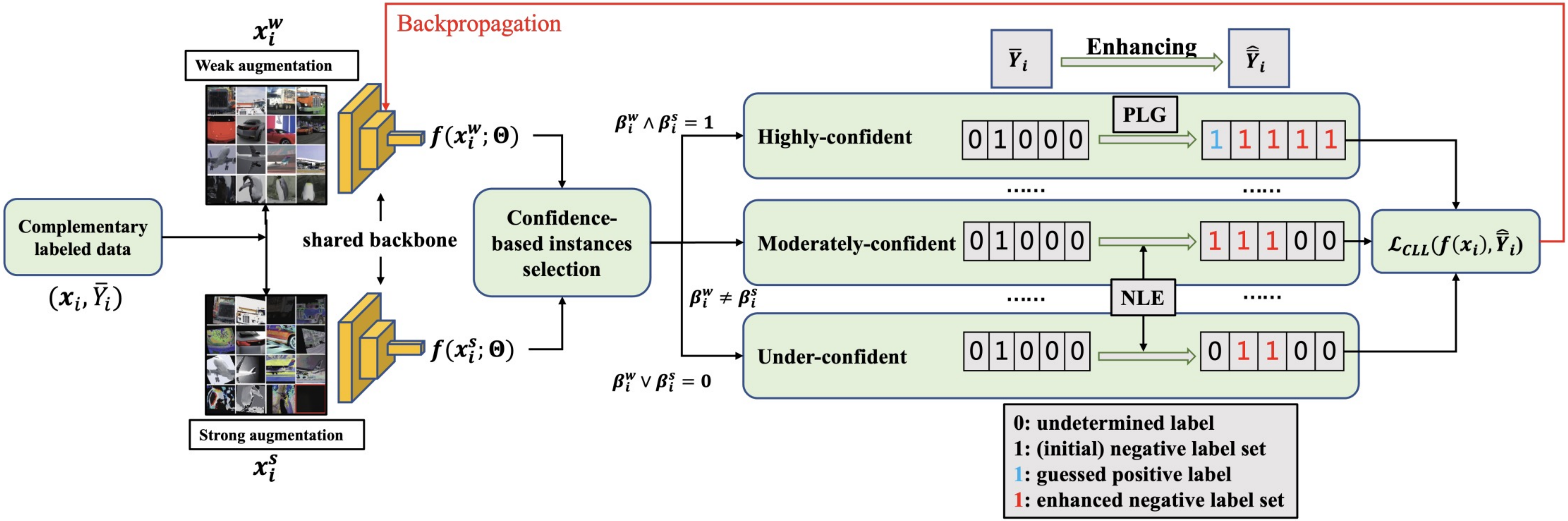
02

Methodology

➢ (ICLR 25') Complementary Label Learning with Positive Label Guessing and Negative Label Enhancement

1. How **Status-aware?**

Motivation/Finding: **Divide and Conquer (PLG and NLE)** based on **confidence of the model output.**

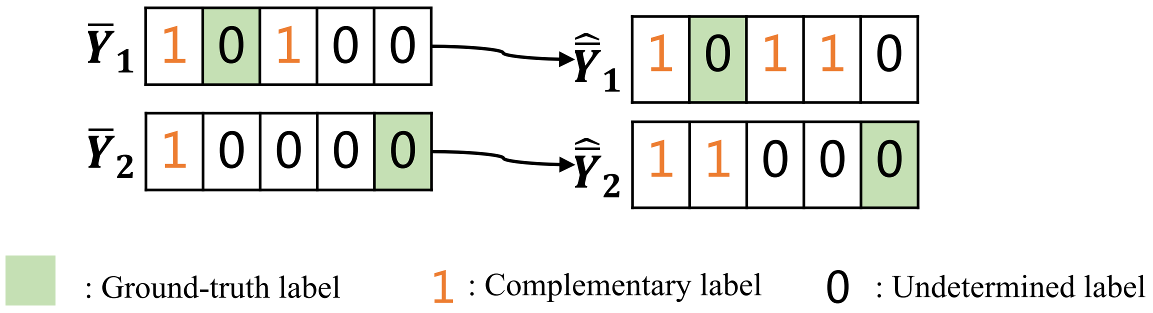


➢ (ICLR 25') Complementary Label Learning with Positive Label Guessing and Negative Label Enhancement

2. How **Data-oriented?**

Motivation/Finding: **More supervision** will bring performance gain.

Number of complementary labels			
2	3	4	5
82.73 ± 1.36	84.16 ± 0.76	86.78 ± 0.87	89.73 ± 0.24
79.30 ± 0.81	82.55 ± 0.87	85.29 ± 1.25	86.20 ± 0.96
86.91 ± 0.35	88.60 ± 0.31	89.58 ± 0.21	90.37 ± 0.19
85.61 ± 0.27	87.03 ± 0.36	89.39 ± 0.21	90.29 ± 0.17
88.90 ± 0.60	**92.30 ± 0.78**	**93.63 ± 0.75**	**94.53 ± 0.47**



➢ **(ICLR 25') Complementary Label Learning with Positive Label Guessing and Negative Label Enhancement**

2. How **Data-oriented?**

Method: **Enhancing the negative supervision (NLE)** via instances' nearest neighbors in feature space.

- **Complementary Label Sharing Mechanism**

Assumption 1. $\forall(\boldsymbol{x}_i, \bar{Y}_i) \in \mathcal{D}$ *and its k-NN instances* $(\boldsymbol{x}_i^{(j)}, \bar{Y}_i^{(j)})$, *the positive label y_i exists in its k-NN instances' complementary label set $\bar{Y}_i^{(j)}$ with probability no more than α_k, any negative label $y_i' \neq y_i$ exist in its k-NN instances' complementary label set $\bar{Y}_i^{(j)}$ with probability no less than β_k.*

Not "dog"

⇩

Not "dog, bird, snake"

Not "bird"

Not "snake"

...

➢ (ICLR 25') Complementary Label Learning with Positive Label Guessing and Negative Label Enhancement

2. How **Data-oriented?**

Method: **Enhancing the negative supervision (NLE)** via instances' nearest neighbors in feature space.

- **Complementary Label Sharing Mechanism**

Assumption 1. $\forall(\boldsymbol{x}_i, \bar{Y}_i) \in \mathcal{D}$ *and its k-NN instances* $(\boldsymbol{x}_i^{(j)}, \bar{Y}_i^{(j)})$, *the positive label* y_i *exists in its k-NN instances' complementary label set* $\bar{Y}_i^{(j)}$ *with probability no more than* α_k, *any negative label* $y_i' \neq y_i$ *exist in its k-NN instances' complementary label set* $\bar{Y}_i^{(j)}$ *with probability no less than* β_k.

- **k-NN Label Frequency**

$$\boldsymbol{F}_{ij} = \sum_{v=1}^{k} \mathbb{I}(j \in \bar{Y}_i^{(v)}),$$

- **Enhanced Complementary Label Set**

$$\widehat{\bar{Y}}_i = \{c | c \in \bar{Y}_i \vee c \in \text{top-}\tau_i\text{-max}_{j \in Y_i}(\boldsymbol{F}_{ij})\}.$$

➢ **(ICLR 25') Complementary Label Learning with Positive Label Guessing and Negative Label Enhancement**

2. How **Data-oriented?**

Method: **Enhancing the negative supervision** via instances' nearest neighbors in feature space.

- k-NN Label Frequency $$\boldsymbol{F}_{ij} = \sum_{v=1}^{k} \mathbb{I}(j \in \bar{Y}_i^{(v)}),$$
- Enhanced Complementary Label Set $$\widehat{\bar{Y}}_i = \{c | c \in \bar{Y}_i \vee c \in \text{top-}\tau_i\text{-max}_{j \in Y_i}(\boldsymbol{F}_{ij})\}.$$

- **The cooperation with Vision Language Models**

Feature Extractor	CIFAR-10, SCLL Accuracy	CIFAR-10, SCLL $1-\epsilon_2$	CIFAR-100, MCLL Accuracy	CIFAR-100, MCLL $1-\epsilon_2$
MoCo	93.12%	95.64%	61.82%	75.34%
BLIP-2	**95.84%**	**99.91%**	**69.85%**	**93.34%**
PreActResNet-18	94.78%	97.79%	64.33%	80.84%

➢ **(ICLR 25') Complementary Label Learning with Positive Label Guessing and Negative Label Enhancement**

3. How **Theoretical-guaranteed?**

Motivation/Finding: Bounded PLG/NLE errors, bounded generalization error **under mild assumption.**

- PLG error bound

$$\epsilon_1 = \mathbb{P}(y_i \in \widehat{\bar{Y}}_i) \leq \psi(\boldsymbol{X}, \bar{\boldsymbol{Y}})^{(K-1-s_i)},$$

- NLE error bound

$$\epsilon_2 = \mathbb{P}(y_i \in \widehat{\bar{Y}}_i) \leq \sum_{j=1}^{k} \binom{|Y_i|-1}{|Y_i|-\tau_i} I_{\beta_k}(k-j+1, j)^{(|Y_i|-\tau_i)} b_{\alpha_k}(k, j),$$

- Generalization bound

$$R(\hat{f}) - R(f^*) \leq 2(1 - \frac{1-\epsilon}{K-\bar{s}})B + 4\rho K \mathfrak{R}_N(\mathcal{F}) + 2KB\sqrt{\frac{log\frac{2}{\delta}}{2N}}.$$

The first to prove the upper bound for Complementary Label Learning with Pseudo Label Noise.

➢ (ICLR 25') Complementary Label Learning with Positive Label Guessing and Negative Label Enhancement

Main Results

Table 1: Comparison of classification accuracies between different methods on four datasets with a single complementary label per instance. The results (mean ± std) are reported over 3 random trials. The best results are highlighted in bold (The same applies hereinafter).

Method	STL-10	SVHN	FMNIST	CIFAR-10
UB-EXP	28.84±0.54%	88.93±0.17%	87.96±0.08%	62.90±0.06%
UB-LOG	20.41±0.46%	89.59±0.08%	87.59±0.14%	70.28±0.12%
SCL-EXP	31.03±0.61%	88.66±0.20%	88.31±0.09%	72.35±0.10%
SCL-LOG	30.74±0.72%	89.26±0.24%	88.03±0.10%	79.87±0.14%
POCR	34.96±0.32%	96.65±0.14%	92.29±0.07%	94.15±0.09%
SELF-CL	30.87±0.72%	90.13±0.23%	84.86±0.10%	88.95±0.22%
ComCo	32.43±0.28%	91.41±0.35%	85.42±0.40%	89.36±0.76%
Ours	**55.25±0.36%**	**97.58±0.18%**	**93.38±0.06%**	**94.78±0.12%**

Table 2: Comparison of classification accuracies between different methods on five datasets with multiple complementary labels per instance. The results (mean ± std) are reported over 3 random trials.

Method	STL-10	SVHN	FMNIST	CIFAR-10	CIFAR-100
UB-EXP	60.85±0.12%	95.23±0.09%	92.34±0.28%	91.13±0.23%	34.43±0.08%
UB-LOG	62.84±0.17%	94.76±0.07%	91.84±0.29%	92.01±0.21%	52.76±0.15%
SCL-EXP	62.96±0.10%	95.28±0.14%	92.20±0.27%	91.85±0.25%	47.81±0.09%
SCL-LOG	61.60±0.14%	94.88±0.16%	91.51±0.25%	92.67±0.18%	49.40±0.19%
POCR	74.51±0.29%	97.14±0.09%	94.76±0.26%	96.09±0.27%	53.16±0.11%
SELF-CL	69.85±0.20%	91.58±0.30%	94.92±0.21%	92.23±0.16%	57.65±0.25%
ComCo	73.28±0.19%	95.41±0.23%	92.01±0.16%	91.38±0.73%	57.88±0.95%
Ours	**77.11±0.14%**	**98.13±0.11%**	**95.16±0.13%**	**96.80±0.28%**	**64.33±0.43%**

Ablation Study

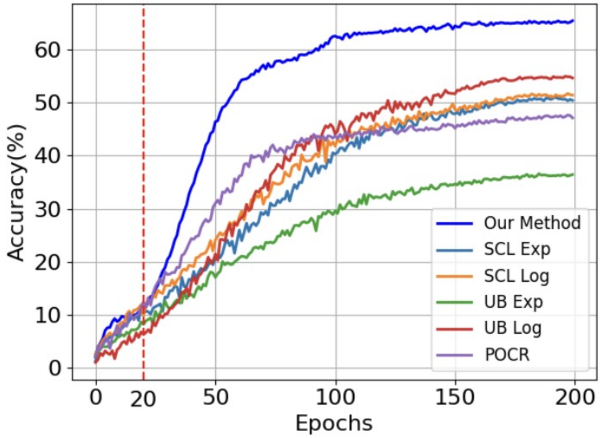
Table 4: Classification accuracy of degenerated methods on three settings.

Method	STL-10 SCLL	CIFAR-10 SCLL	CIFAR-100 MCLL
PLNL	**55.25**	**94.78**	**64.33**
PLNL v1	49.25	93.75	63.09
PLNL v2	49.82	92.01	58.94
PLNL v3	53.22	94.28	63.14
POCR	34.96	94.15	53.16

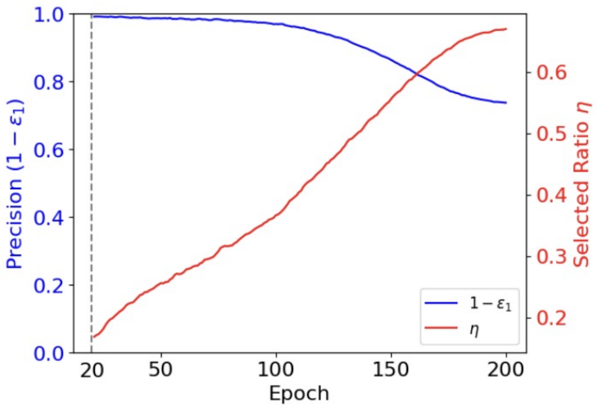
All components are indispensible.

➢ (ICLR 25') Complementary Label Learning with Positive Label Guessing and Negative Label Enhancement

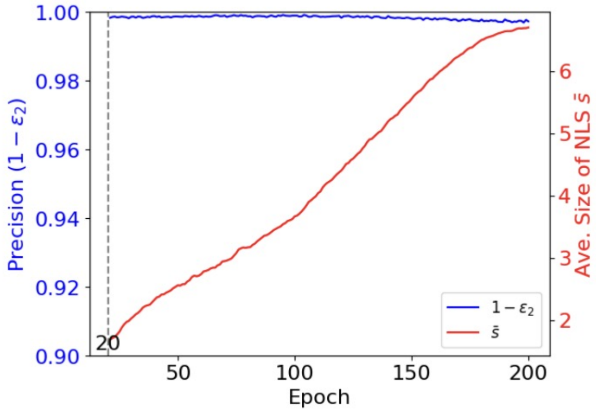
Further Analysis



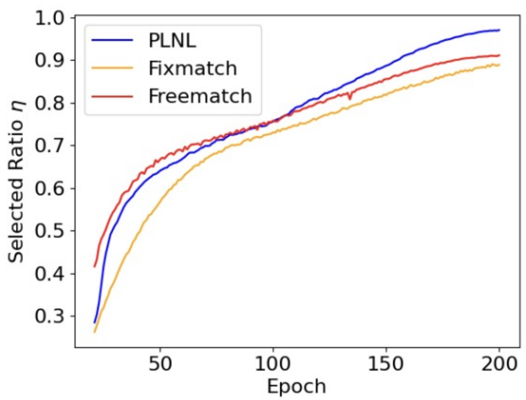
(a) Accuracy Comparison of CLL methods



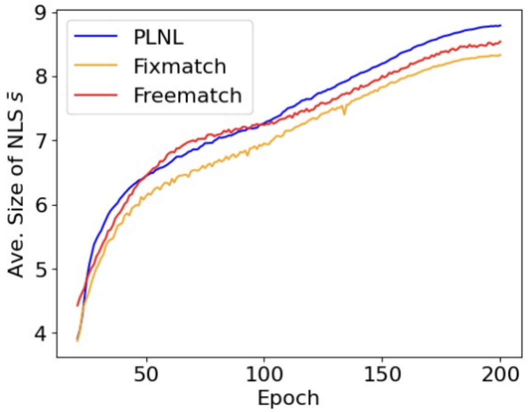
(b) PLG Precision vs. Selected Ratio



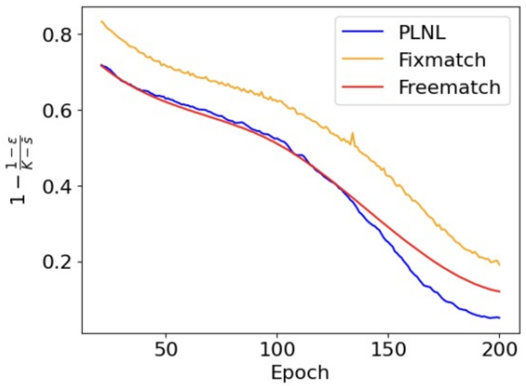
(c) NLE Precision vs. Ave. Num. of NLS



(a) Selected Ratio vs. Epoch



(b) Ave. Num. of NLS vs. Epoch



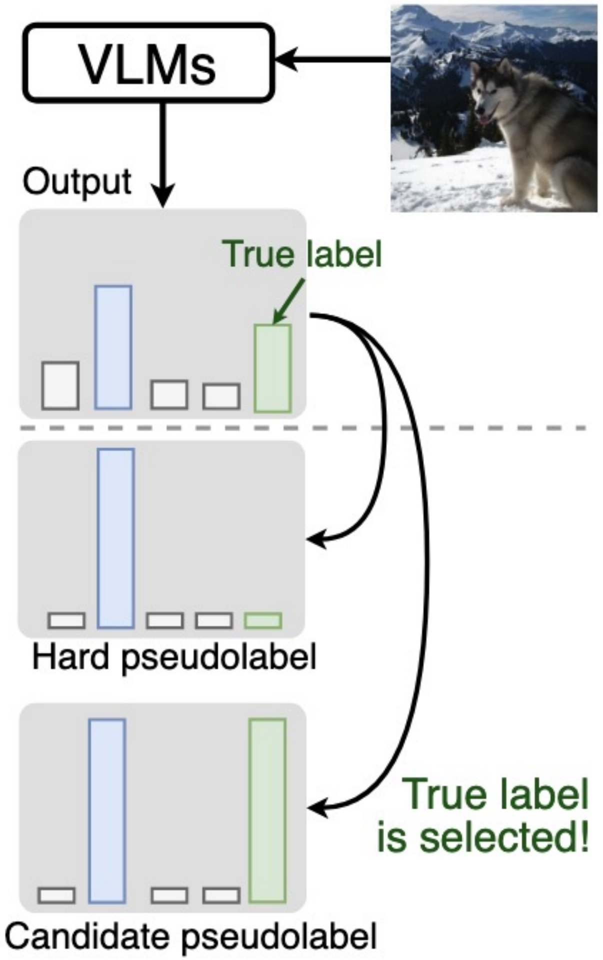
(c) Convergence of $(1-\frac{1-\epsilon}{K-\bar{s}})$

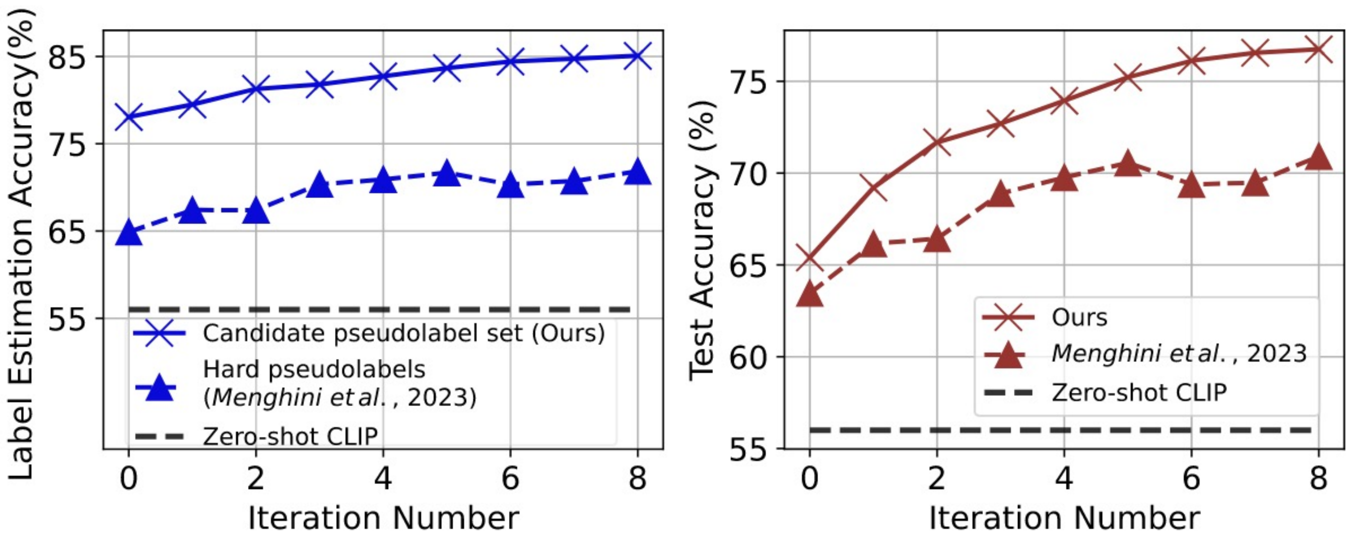
04

Future Work

➢ The cooperation with Language Models : Weakly-Supervised Learning

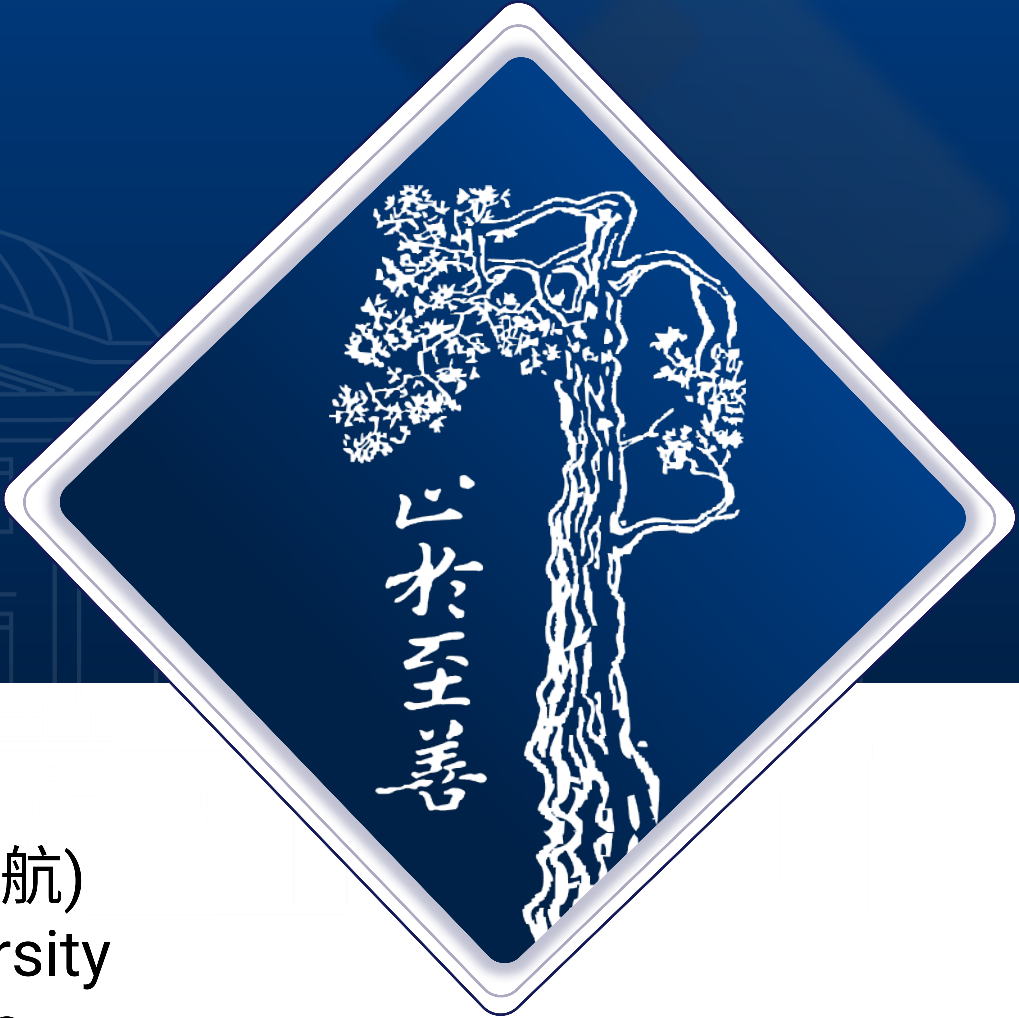
- A New Perspective to Leverage Unlabeled Data





東南大學
SOUTHEAST UNIVERSITY

Thanks!



Yuhang Li (李宇航)
Southeast University
Nanjing, China