SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

THE UNIVERSITY OF MELBOURNE

ICML International Conference On Machine Learning

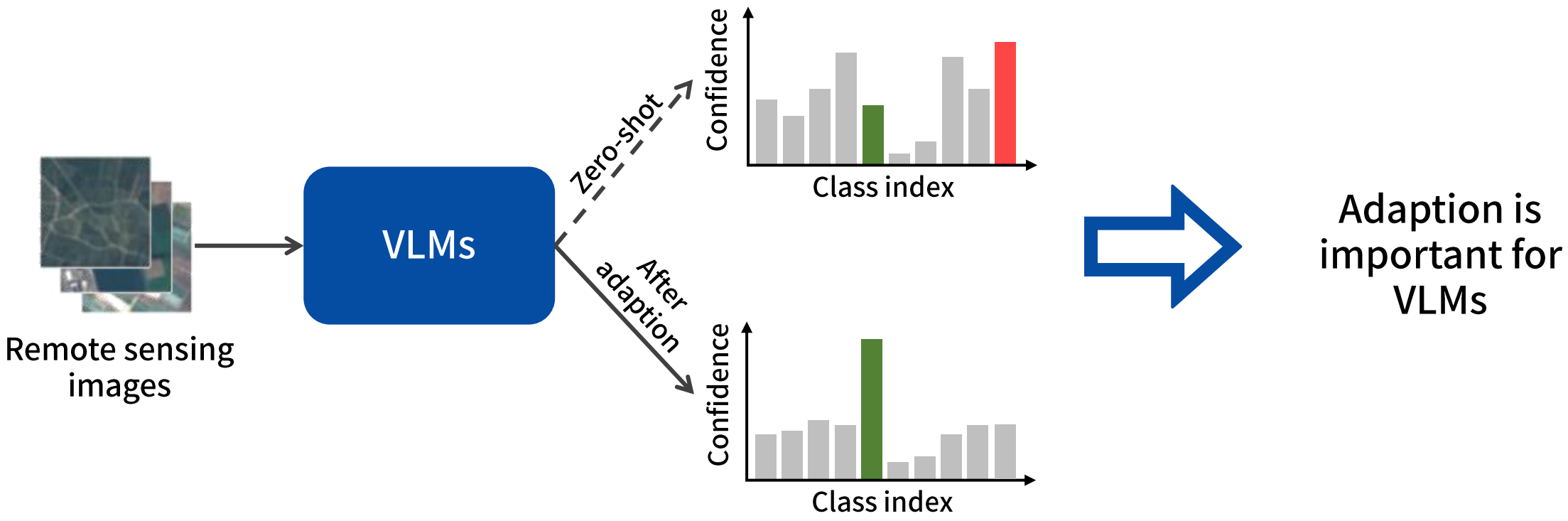
Candidate Pseudolabel Learning: Enhancing Vision-Language Models by Prompt Tuning with Unlabeled Data

Authors: Jiahan Zhang, Qi Wei, Feng Liu, Lei Feng

Date: 07/24/2024

Research Background

❖ Image classification task for VLMs with specialized domain



Research Background

❖ Image classification task for VLMs with specialized domain

Adaption is important for VLMs

But it still needs a lot of well-labeled data for adaption

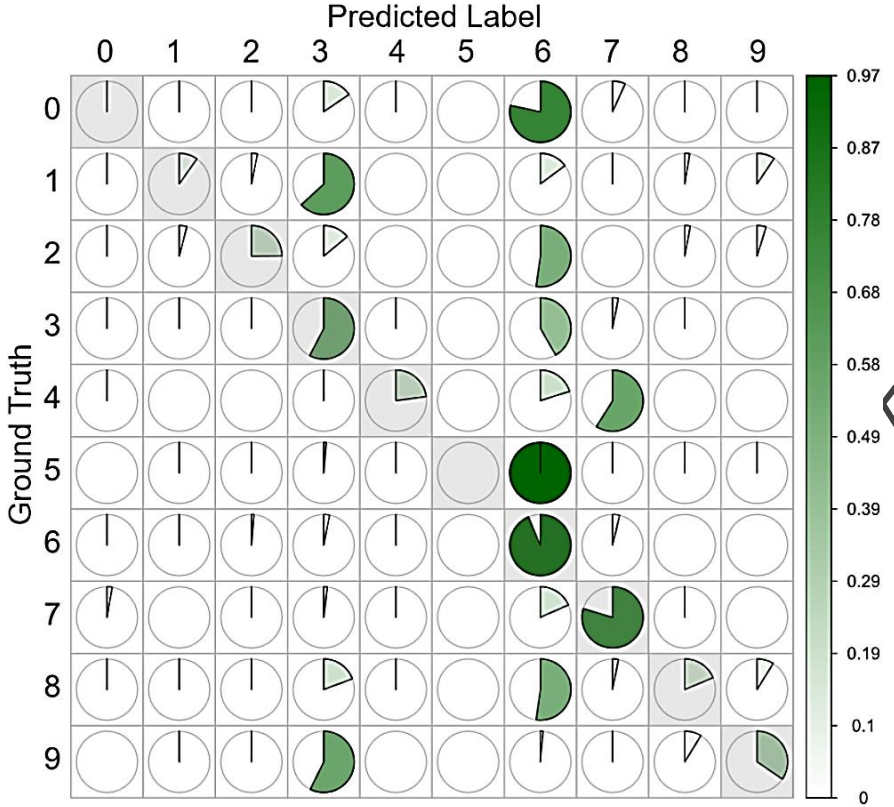
When a large amount of unlabeled data is available

Can we use the zero-shot capabilities of VLMs to exploit unlabeled data for adaption?

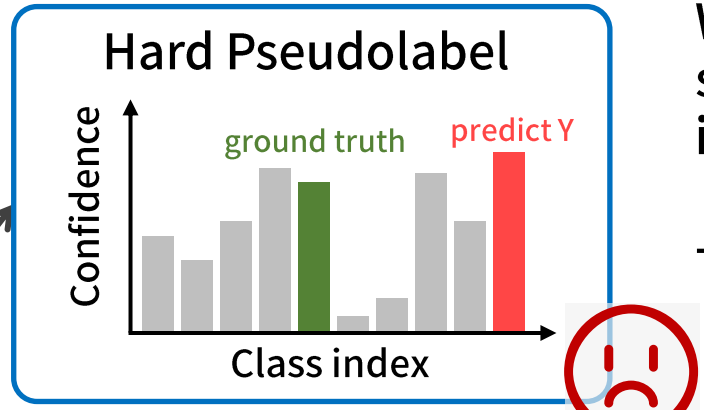
INTRODUCTION

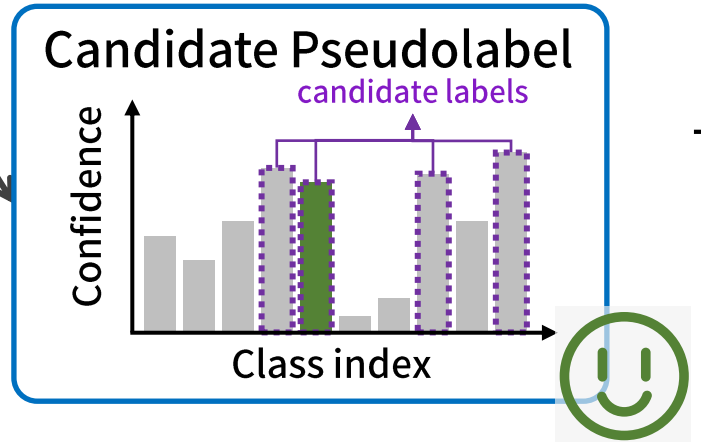
Comparison and Motivation

❖ How can we better use the zero-shot capabilities of VLMs to exploit unlabeled data?



Confusion matrix between true labels and zero-shot predicted label of EuroSAT





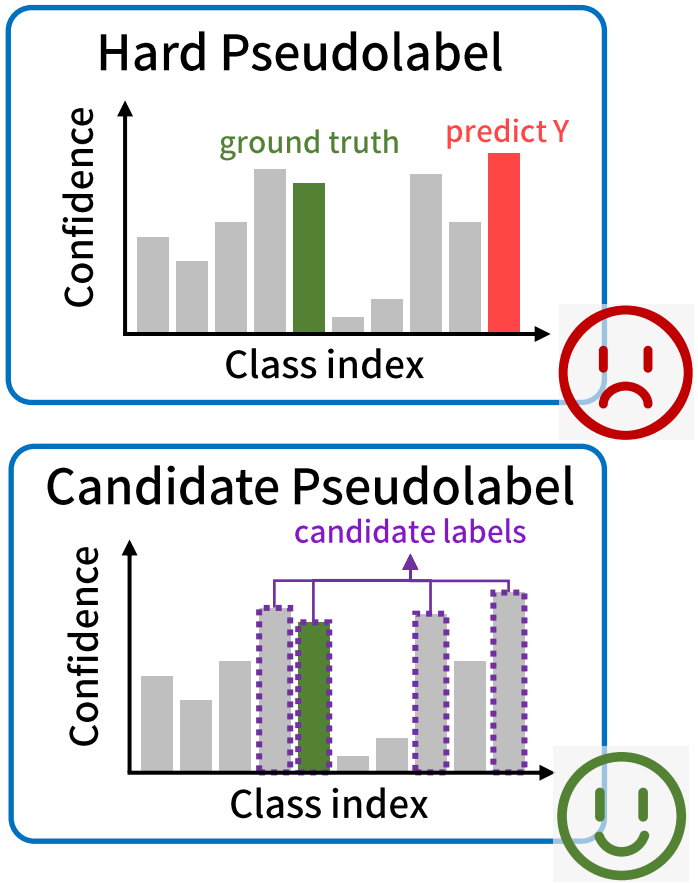
We observed that lots of zero-shot predicted labels are **incorrect** and **imbalanced:**

→ Directly applying them as hard pseudolabels will inherit these shortcomings

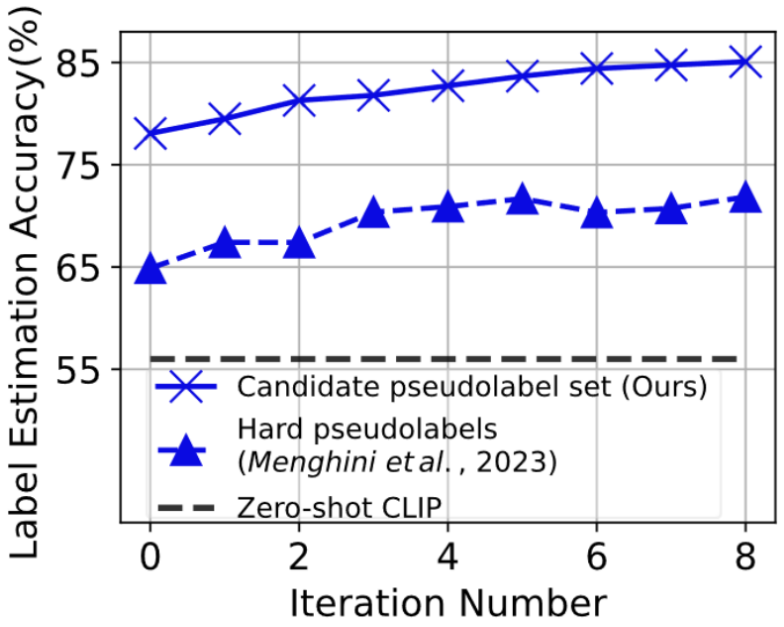
→ Generating a set of potential pseudolabels from the predictions will alleviate these issues

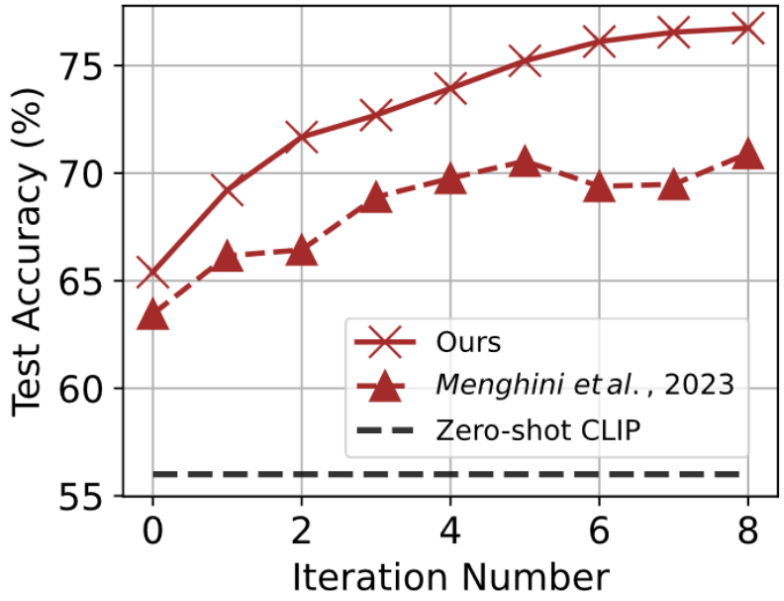
Comparison and Motivation

❖ Advantages of candidate pseudolabels over hard pseudolabels



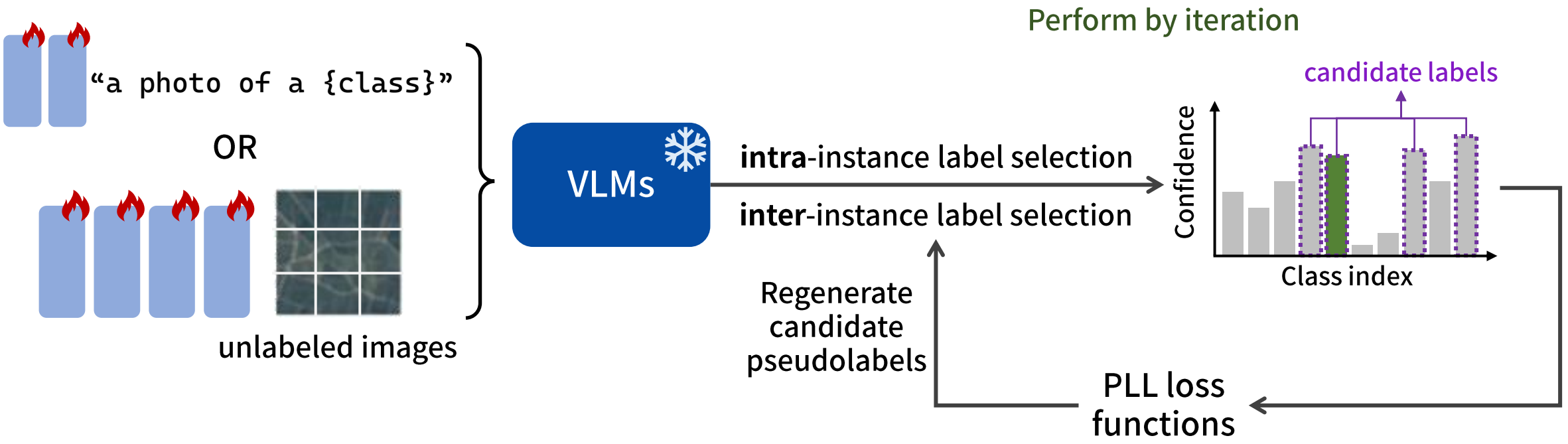
- Higher label estimation accuracy (the rate at which the true label is included in the pseudolabels)
- Fully leverage the zero-shot capabilities/information from VLMs
- Result in improved performance on test accuracy





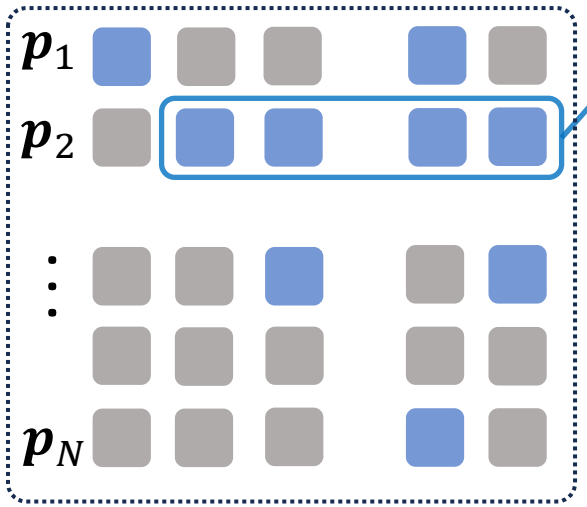
Overview of Workflow

❖ Overall workflow for prompt tuning with unlabeled data in **CPL** (**C**andidate **P**seudolabel **L**earning)



Scheme for Generating Candidate Pseudolabels

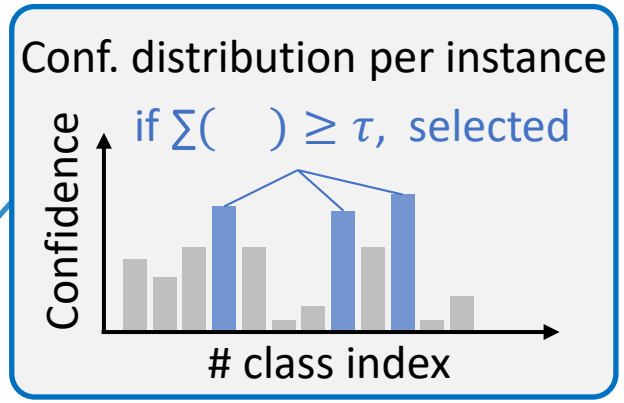
1. The strategy of intra-instance label selection



Confidence Score Matrix

$(p_1, p_2, \ldots, p_N)^\top$

1) Intra-inst. Label Selection



- Similar to top-K confident labels selection
- And further consider varying levels of identification difficulty to determine different K values for each instance

- The vector of conf. scores for instance i:

$$\boldsymbol{p_i} = \boldsymbol{g}(\boldsymbol{f}_\theta(\boldsymbol{x})) = (p_{i1}, p_{i2}, \ldots, p_{iC})^\top$$

- The intra-instance candidate set:

$$S_i^{\text{intra}} = \text{MinSize}(\{c \mid \sum_{c=1}^{C} p_{ic} \geqslant \tau\})$$

to get the set with the minimal size that the cumulative conf. score just surpasses a threshold

- Determine the threshold:

$$\tau = \text{Quantile}(\text{Sort}(\hat{\boldsymbol{p}}), \alpha)$$

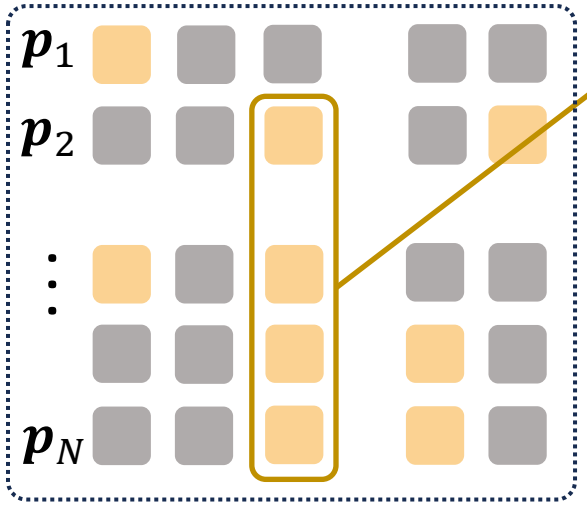
a function that returns the value at the given quantile α

Scheme for Generating Candidate Pseudolabels

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

ICML International Conference On Machine Learning

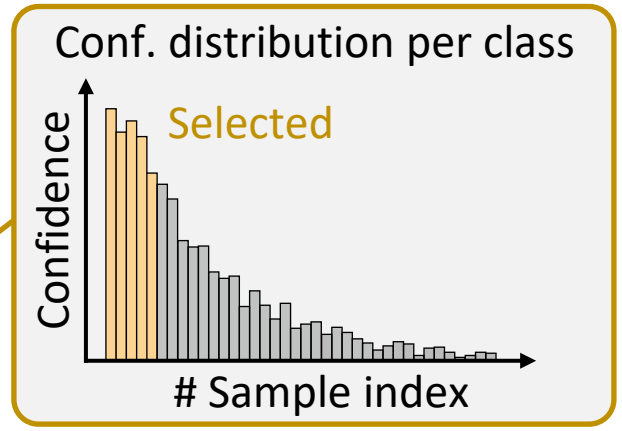
2. The strategy of inter-instance label selection



Confidence Score Matrix

$(p_1, p_2, \ldots, p_N)^\top$

2) Inter-inst. Label Selection



- To balance the ratio of each category in the candidate pseudolabel set
- To further refine the candidate pseudolabels

- Conf. scores across all unlabeled instances for class c:

$$\boldsymbol{q}_c = (p_{1c}, p_{2c}, \ldots, p_{Nc})$$

- The inter-instance candidate set:

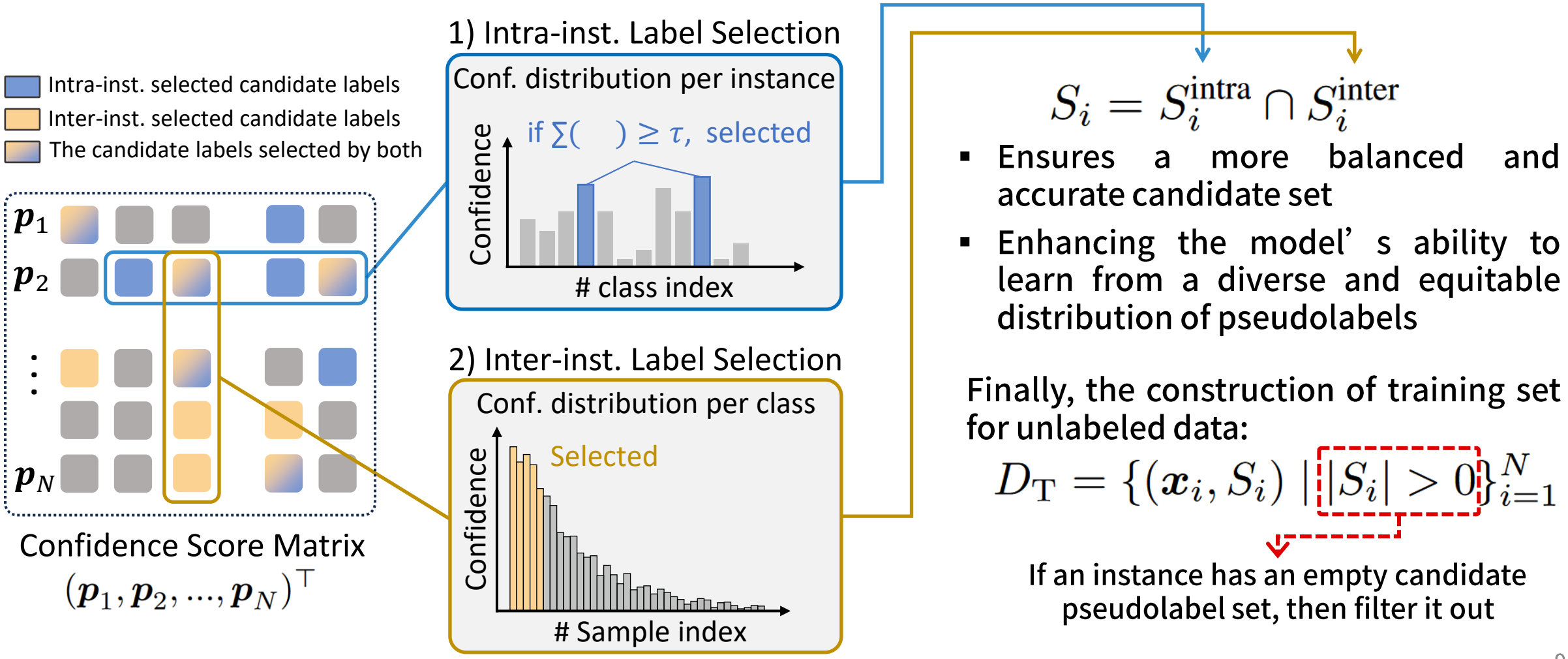
$$S_i^{\text{inter}} = \{c \mid p_{ic} > \text{Quantile}(\text{Sort}(\boldsymbol{q}_c), \beta)\}_{c=1}^{C}$$

to select the candidate labels of relatively higher confidence levels within the class

Scheme for Generating Candidate Pseudolabels

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN
ICML International Conference On Machine Learning

❖ Final candidate sets and training set for unlabeled instances



$$S_i = S_i^{\text{intra}} \cap S_i^{\text{inter}}$$

- Ensures a more balanced and accurate candidate set
- Enhancing the model' s ability to learn from a diverse and equitable distribution of pseudolabels

Finally, the construction of training set for unlabeled data:

$$D_{\mathrm{T}} = \{(\boldsymbol{x}_i, S_i) \mid |S_i| > 0\}_{i=1}^{N}$$

If an instance has an empty candidate pseudolabel set, then filter it out

Learning with Candidate Pseudolabels

❖ The construction of training target for unlabeled data

The training set for unlabeled data:

$$D_{\mathrm{T}} = \{(\boldsymbol{x}_i, S_i) \mid |S_i| > 0\}_{i=1}^{N}$$

Transform the candidate set into training target:

$$\begin{cases} s_{ic} = 1 \text{ if } c \in S_i \\ s_{ic} = 0 \text{ if } c \notin S_i \text{ for } c \in [C] \end{cases} \longrightarrow (\boldsymbol{x}_i, \boldsymbol{s}_i)$$

❖ The training objectives:

If a labeled set D_{L} and an unlabeled set D_{UL} are provided

$$D_{\mathrm{L}} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{O}$$

$$\mathcal{L} = \mathcal{L}_{\mathrm{L}} + \lambda \mathcal{L}_{\mathrm{UL}} = \frac{1}{b_1} \sum_{i=1}^{b_1} L_{\mathrm{ce}}(\boldsymbol{x}_i, \boldsymbol{y}_i) + \lambda \frac{1}{b_2} \sum_{i=1}^{b_2} L_{\mathrm{p}}(\boldsymbol{x}_i, \boldsymbol{s}_i)$$

partial-label learning loss func.

If only an unlabeled data D_{UL} is provided

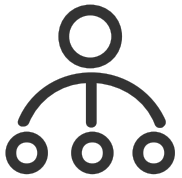
$$\mathcal{L} = \mathcal{L}_{\mathrm{UL}} = \frac{1}{b_2} \sum_{i=1}^{b_2} L_{\mathrm{p}}(\boldsymbol{x}_i, \boldsymbol{s}_i)$$

Experimental Settings

Design extensive experiments across 3 dimensions

- **Learning paradigm variety:**

  semi-supervised learning (SSL), unsupervised learning (UL), and transductive zero-shot learning (TRZSL). All involving the access of unlabeled data



- **Prompt tuning variety:**

  visual prompt tuning (VPT) or text prompt tuning (TPT)

- **Task variety:**

  FGVC-Aircraft, EuroSAT, CUB, Flowers102, RESISC45, DTD, CALTECH-101, UCF-101, and CIFAR-100 (9 classification tasks)

EXPERIMENTS

Experimental Results

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

ICML International Conference On Machine Learning

❖ The pros of CPL compared with previous hard pseudolabel on prompt tuning:

Table 1: Comparison results of top-1 test accuracy (%) on six benchmarks when applying **Textual prompts** as tuning strategy. Note that "✓" and "✗" denote whether full unlabeled data are utilized for fine-tuning or not, respectively.

Methods		Flowers102			RESISC45			DTD		
		SSL	UL	TRZSL	SSL	UL	TRZSL	SSL	UL	TRZSL
Zero-shot CLIP		$63.67_{0.00}$		$63.40_{0.00}$	$54.48_{0.00}$		$54.46_{0.00}$	$43.24_{0.00}$		$43.45_{0.00}$
FPL	✗	$75.96_{0.74}$	$65.67_{0.23}$	$80.97_{0.00}$	$68.13_{0.55}$	$63.07_{0.38}$	$72.11_{0.00}$	$37.10_{5.45}$	$44.96_{0.55}$	$46.30_{0.03}$
CPL (Ours)	✗	$\mathbf{77.36}_{0.24}$	$\mathbf{70.01}_{0.21}$	$\mathbf{84.60}_{0.10}$	$\mathbf{71.73}_{0.57}$	$\mathbf{68.47}_{0.34}$	$\mathbf{72.16}_{0.26}$	$\mathbf{54.63}_{0.79}$	$\mathbf{48.92}_{0.17}$	$\mathbf{59.79}_{1.32}$
GRIP	✓	$83.60_{0.48}$	$69.84_{1.06}$	$86.26_{0.00}$	$74.11_{0.68}$	$70.55_{0.88}$	$81.07_{0.00}$	$56.07_{0.85}$	$46.09_{1.06}$	$65.30_{0.01}$
CPL (Ours)	✓	$\mathbf{89.66}_{0.36}$	$\mathbf{72.90}_{0.78}$	$\mathbf{87.35}_{0.76}$	$\mathbf{80.98}_{0.11}$	$\mathbf{77.39}_{0.44}$	$\mathbf{85.85}_{0.49}$	$\mathbf{61.21}_{0.56}$	$\mathbf{51.91}_{0.71}$	$\mathbf{68.00}_{0.34}$

Methods		CUB			EuroSAT			FGVCAircraft		
		SSL	UL	TRZSL	SSL	UL	TRZSL	SSL	UL	TRZSL
Zero-shot CLIP		$51.82_{0.00}$		$51.57_{0.00}$	$32.88_{0.00}$		$30.54_{0.00}$	$17.58_{0.00}$		$17.86_{0.00}$
FPL	✗	$55.29_{0.59}$	$53.04_{0.53}$	$55.44_{0.20}$	$62.05_{1.64}$	$48.96_{1.49}$	$53.70_{26.87}$	$20.02_{0.77}$	$16.62_{0.67}$	$17.55_{0.37}$
CPL (Ours)	✗	$\mathbf{56.37}_{0.45}$	$\mathbf{54.18}_{0.05}$	$\mathbf{64.01}_{0.17}$	$\mathbf{64.84}_{2.15}$	$\mathbf{51.45}_{1.97}$	$\mathbf{54.03}_{2.27}$	$\mathbf{22.37}_{0.66}$	$\mathbf{18.90}_{0.20}$	$\mathbf{28.47}_{0.43}$
GRIP	✓	$56.65_{0.33}$	$51.42_{0.21}$	$59.48_{0.38}$	$58.66_{2.64}$	$57.21_{1.77}$	$92.33_{0.69}$	$16.98_{0.82}$	$15.22_{0.71}$	$26.08_{0.25}$
CPL (Ours)	✓	$\mathbf{58.53}_{0.24}$	$\mathbf{53.47}_{0.36}$	$\mathbf{66.20}_{0.50}$	$\mathbf{77.51}_{0.80}$	$\mathbf{67.26}_{0.47}$	$\mathbf{93.78}_{0.12}$	$\mathbf{22.48}_{0.63}$	$\mathbf{18.35}_{0.27}$	$\mathbf{30.86}_{0.70}$

Table 2: Comparison results of top-1 test accuracy (%) on six benchmarks when applying **Visual prompts** as tuning strategy. Note that "✓" and "✗" denote whether full unlabeled data are utilized for fine-tuning or not, respectively.

Methods		Flowers102			RESISC45			DTD		
		SSL	UL	TRZSL	SSL	UL	TRZSL	SSL	UL	TRZSL
Zero-shot CLIP		$63.67_{0.00}$		$63.40_{0.00}$	$54.48_{0.00}$		$54.46_{0.00}$	$43.24_{0.00}$		$43.45_{0.00}$
FPL	✗	$67.03_{0.65}$	$65.50_{0.41}$	$71.94_{0.00}$	$65.14_{0.25}$	$62.24_{0.22}$	$67.85_{0.00}$	$47.60_{1.09}$	$47.69_{0.48}$	$52.43_{0.00}$
CPL (Ours)	✗	$\mathbf{70.58}_{0.13}$	$\mathbf{68.94}_{0.16}$	$\mathbf{78.13}_{0.31}$	$\mathbf{68.85}_{0.13}$	$\mathbf{67.97}_{0.17}$	$\mathbf{72.18}_{0.27}$	$\mathbf{52.64}_{0.68}$	$\mathbf{50.37}_{0.46}$	$\mathbf{55.90}_{0.69}$
GRIP	✓	$67.95_{1.2}$	$63.09_{0.56}$	$77.18_{0.00}$	$71.22_{0.77}$	$68.43_{0.61}$	$82.19_{0.00}$	$54.57_{4.86}$	$50.51_{0.99}$	$62.78_{0.00}$
CPL (Ours)	✓	$\mathbf{73.52}_{0.37}$	$\mathbf{67.25}_{0.41}$	$\mathbf{80.14}_{0.73}$	$\mathbf{78.46}_{0.74}$	$\mathbf{72.97}_{0.58}$	$\mathbf{86.67}_{0.33}$	$\mathbf{58.74}_{0.81}$	$\mathbf{53.42}_{0.56}$	$\mathbf{65.31}_{0.78}$

Methods		CUB			EuroSAT			FGVCAircraft		
		SSL	UL	TRZSL	SSL	UL	TRZSL	SSL	UL	TRZSL
Zero-shot CLIP		$51.82_{0.00}$		$51.57_{0.00}$	$32.88_{0.00}$		$30.54_{0.00}$	$17.58_{0.00}$		$17.86_{0.00}$
FPL	✗	$52.86_{0.42}$	$53.17_{0.06}$	$54.17_{0.16}$	$52.47_{2.53}$	$48.79_{3.69}$	$68.68_{14.74}$	$20.14_{0.26}$	$18.28_{0.33}$	$16.28_{0.45}$
CPL (Ours)	✗	$\mathbf{53.37}_{0.55}$	$\mathbf{53.28}_{0.31}$	$\mathbf{56.43}_{0.21}$	$\mathbf{66.37}_{2.10}$	$\mathbf{52.83}_{2.10}$	$\mathbf{74.02}_{1.34}$	$\mathbf{21.52}_{0.68}$	$\mathbf{20.10}_{0.51}$	$\mathbf{26.73}_{0.08}$
GRIP	✓	$\mathbf{53.83}_{0.11}$	$\mathbf{52.91}_{0.26}$	$54.92_{0.17}$	$63.48_{3.09}$	$63.68_{3.42}$	$96.97_{0.77}$	$19.43_{0.50}$	$17.51_{0.61}$	$26.42_{0.30}$
CPL (Ours)	✓	$49.50_{0.42}$	$52.11_{0.24}$	$\mathbf{56.37}_{0.06}$	$\mathbf{72.03}_{1.24}$	$\mathbf{68.93}_{1.15}$	$\mathbf{98.31}_{0.18}$	$\mathbf{20.51}_{0.68}$	$\mathbf{18.26}_{0.38}$	$\mathbf{30.26}_{0.46}$

❖ The pros of CPL when combined with existing label-free CLIP fine-tuning pipeline:

	Flowers-102	UCF-101	CIFAR-100	EuroSAT	DTD	CALTECH-101
CLIP	66.6	61.0	64.2	45.1	42.9	90.5
CLIP-PR	57.7	57.9	63.2	44.2	40.1	84.8
UPL	71.5	63.9	65.8	62.2	48.0	90.6
LaFTer	71.0	68.2	74.6	73.9	46.1	93.3
LaFTer + Ours	**76.7**	**71.0**	**77.3**	**82.2**	**56.3**	**93.4**

EXPERIMENTS

More Experimental Results

SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

ICML International Conference On Machine Learning

❖ CPL performance when employing various PLL loss functions:

	Methods		SSL	UL	TRZSL
	Zero-shot CLIP		$43.24_{0.00}$		$43.45_{0.00}$
Use a fixed amount of unlabeled data	FPL	✗	$37.10_{5.45}$	$44.96_{0.55}$	$46.30_{0.03}$
	CPL $_{\text{Soft CE}}$	✗	$51.83_{0.62}$	$47.02_{0.37}$	$59.69_{0.59}$
	CPL $_{\text{CC}}$	✗	$54.63_{0.79}$	$48.92_{0.17}$	$59.79_{1.32}$
	CPL $_{\text{RC}}$	✗	$54.98_{0.49}$	$49.96_{0.15}$	$59.42_{0.44}$
	CPL $_{\text{CAV}}$	✗	$55.50_{0.29}$	$48.69_{0.66}$	$59.44_{0.13}$
	CPL $_{\text{LW}}$	✗	$55.21_{0.74}$	$49.82_{0.91}$	$59.24_{0.72}$
Use complete unlabeled data	GRIP	✓	$56.07_{0.85}$	$46.09_{1.06}$	$65.30_{0.01}$
	CPL $_{\text{Soft CE}}$	✓	$60.83_{0.66}$	$49.13_{0.10}$	$66.26_{0.77}$
	CPL $_{\text{CC}}$	✓	$61.21_{0.56}$	$51.91_{0.71}$	$68.00_{0.34}$
	CPL $_{\text{RC}}$	✓	$60.21_{0.46}$	$51.58_{0.11}$	$67.95_{0.31}$
	CPL $_{\text{CAV}}$	✓	$61.06_{0.50}$	$49.31_{0.19}$	$67.76_{0.53}$
	CPL $_{\text{LW}}$	✓	$60.20_{0.69}$	$52.23_{0.84}$	$68.29_{0.99}$

❖ CPL performance on imbalance dataset:

Methods	Balanced	Imbalanced δ=100	Imbalanced δ=50
LaFTer	74.64	65.63	66.59
CPL (w/o inter)	76.07	66.68	67.85
CPL	**77.32**	**67.70**	**69.65**

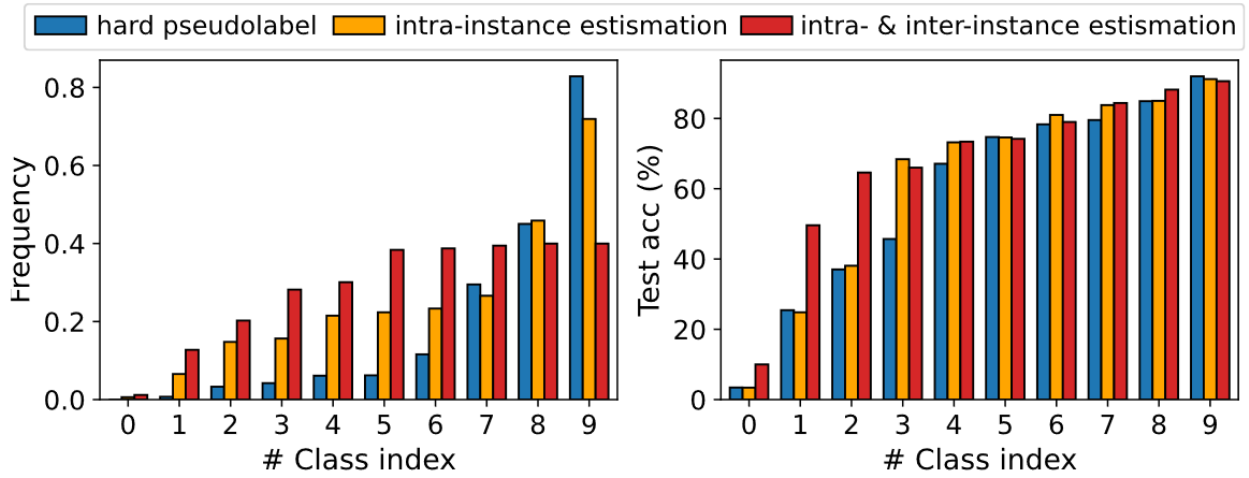
❖ CPL performance with different VLM encoder:

	Methods		SSL	UL	TRZSL
DTD	Zero-shot CLIP		$52.45_{0.00}$		$51.61_{0.00}$
	FPL	✗	$60.61_{1.56}$	$52.99_{0.43}$	$60.77_{0.54}$
	CPL	✗	$\mathbf{62.78}_{0.17}$	$\mathbf{57.23}_{0.19}$	$\mathbf{62.52}_{1.38}$
	GRIP	✓	$60.91_{0.00}$	$54.40_{0.00}$	$64.92_{0.00}$
	CPL	✓	$\mathbf{69.82}_{0.32}$	$\mathbf{57.20}_{0.45}$	$\mathbf{71.97}_{0.46}$
RESISC45	Zero-shot CLIP		$62.67_{0.00}$		$62.13_{0.00}$
	FPL	✗	$79.01_{0.55}$	$70.85_{0.66}$	$77.69_{0.83}$
	CPL	✗	$\mathbf{80.38}_{0.37}$	$\mathbf{76.01}_{0.19}$	$\mathbf{79.97}_{0.77}$
	GRIP	✓	$81.53_{0.00}$	$76.86_{0.00}$	$86.88_{0.00}$
	CPL	✓	$\mathbf{87.75}_{0.29}$	$\mathbf{80.88}_{0.86}$	$\mathbf{89.73}_{1.73}$
Flowers102	Zero-shot CLIP		$73.98_{0.00}$		$73.05_{0.00}$
	FPL	✗	$\mathbf{89.07}_{0.94}$	$77.81_{0.30}$	$91.84_{0.73}$
	CPL	✗	$88.37_{0.39}$	$\mathbf{82.98}_{0.14}$	$\mathbf{96.65}_{0.08}$
	GRIP	✓	$94.21_{0.00}$	$82.33_{0.00}$	$96.18_{0.00}$
	CPL	✓	$\mathbf{96.80}_{0.63}$	$\mathbf{83.94}_{0.69}$	$\mathbf{97.34}_{0.74}$

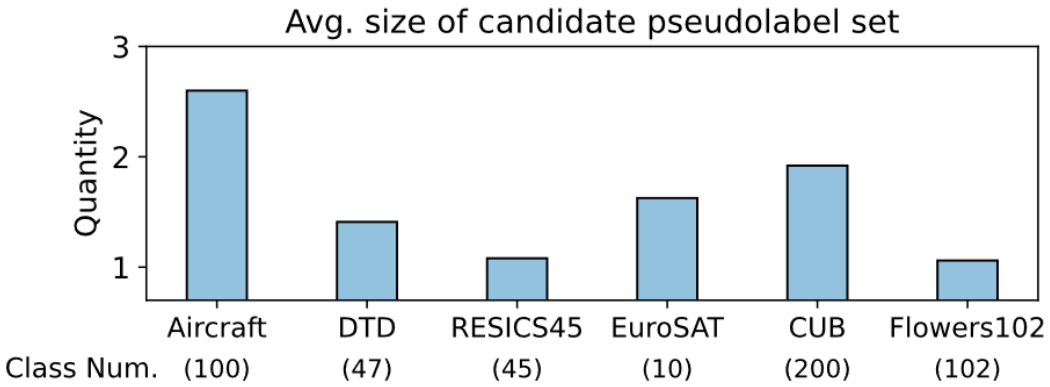
EXPERIMENTS

Ablation and Visualization

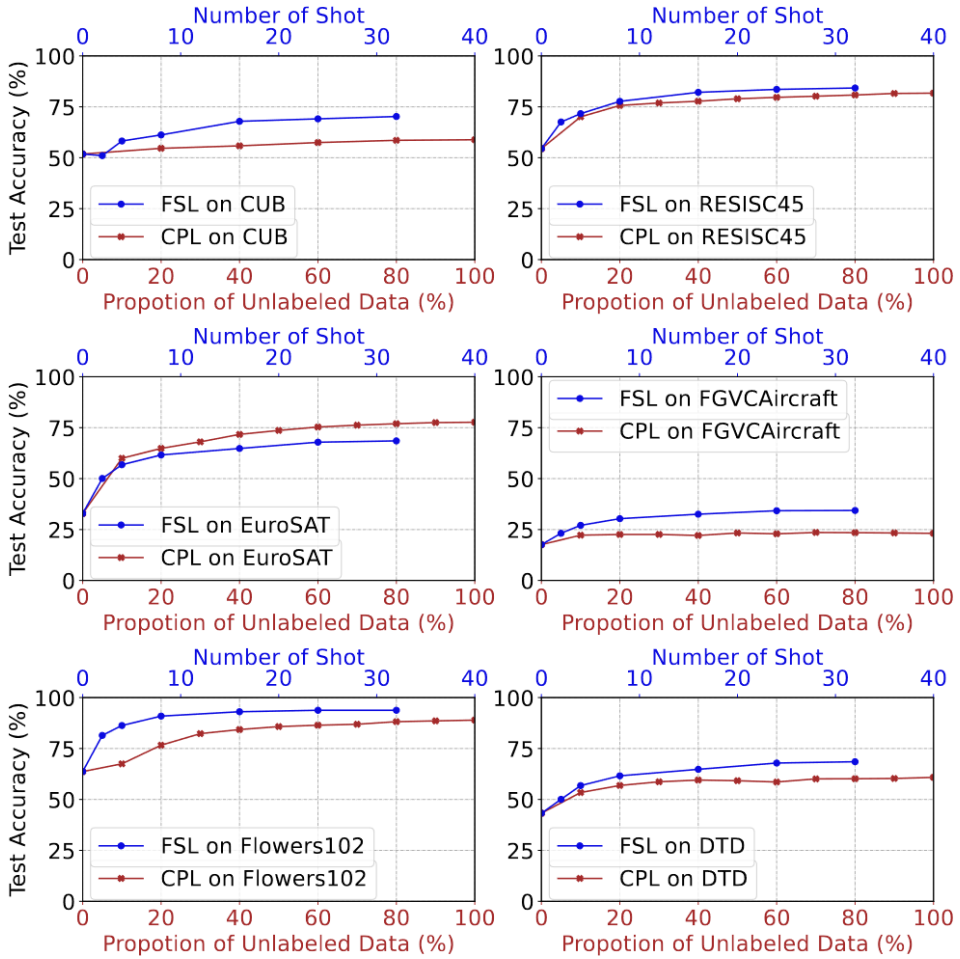
- The class-level distribution of the generated pseudolabels:



- The average size of the candidate set before the last iteration:



- Compare fully supervised few-shot learning and CPL



SINGAPORE UNIVERSITY OF TECHNOLOGY AND DESIGN

THE UNIVERSITY OF MELBOURNE

ICML International Conference On Machine Learning

Thanks For Your Listening!

Project Page: