TUE-AM-329

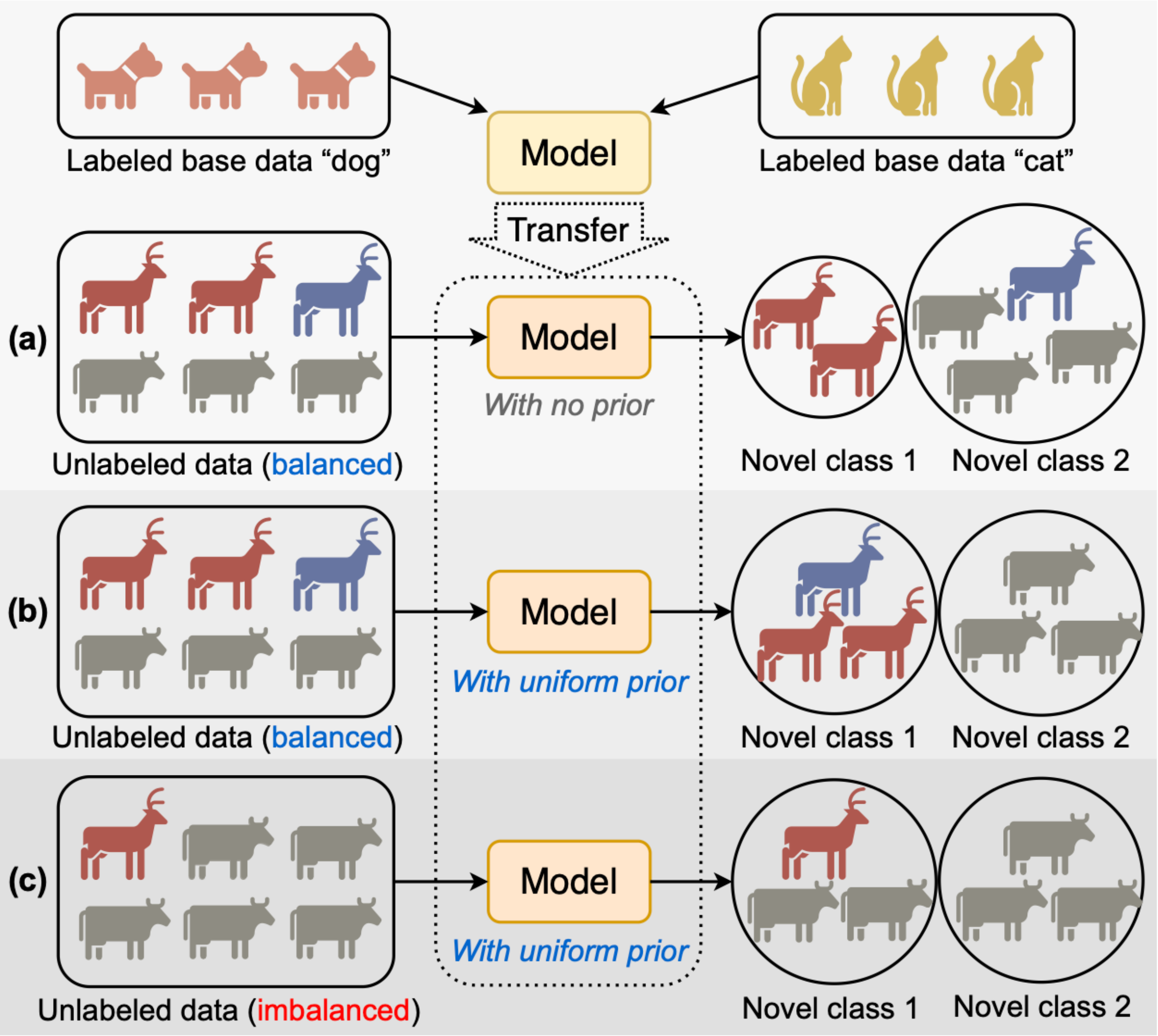
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Bootstrap Your Own Prior: Towards Distribution-Agnostic Novel Class Discovery

Muli Yang[1,2], Liancheng Wang[1], Cheng Deng[1], and Hanwang Zhang[2]

1: School of Electronic Engineering, Xidian University, Xi'an, China
2: School of Computer Science and Engineering, Nanyang Technological University, Singapore

Introduction



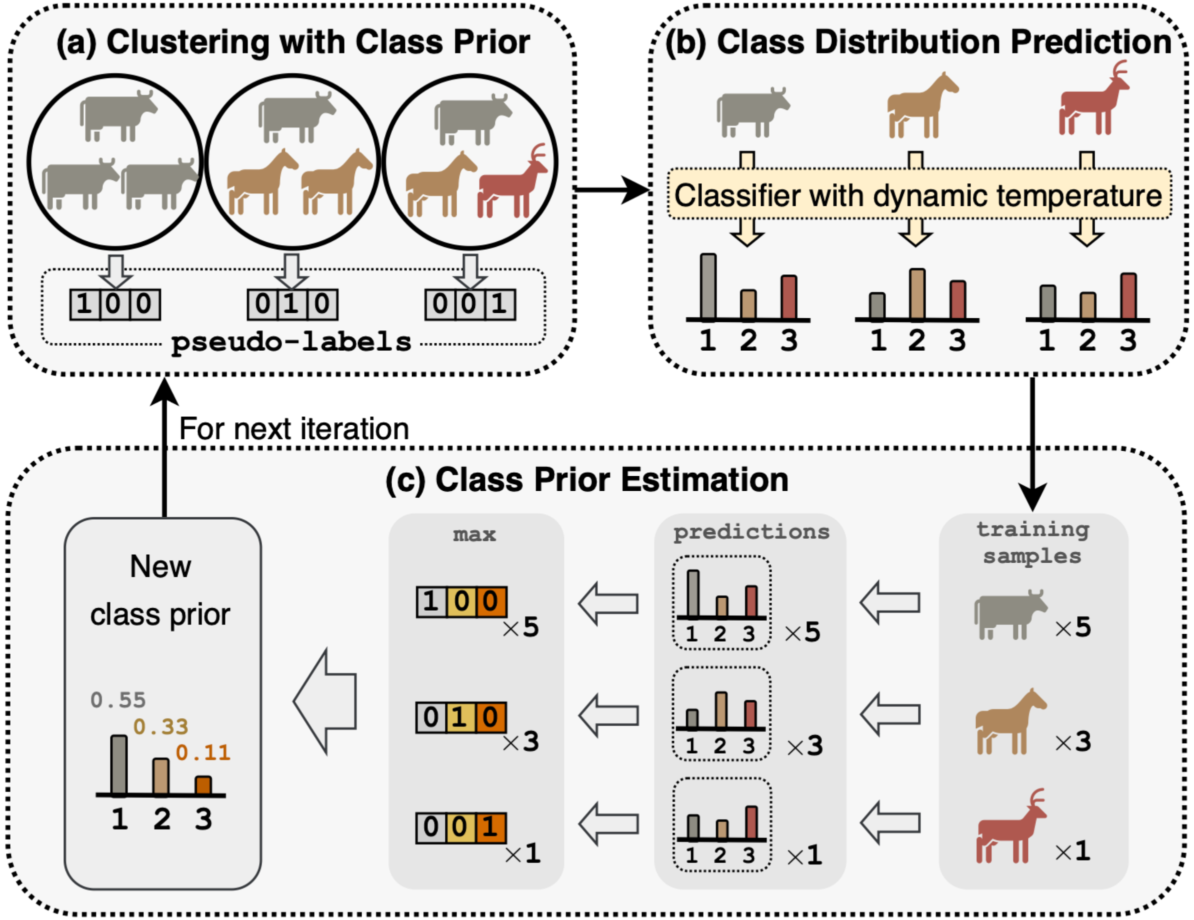
- An **impractical assumption** in existing works: *novel data is class-balanced.*

We propose **Distribution-agnostic NCD**, allowing novel data drawn from arbitrary unknown class distributions

- The **dilemma**: class distribution prior is necessary for accurate pseudo-label generation yet is unknown in our setting.

Approach: overview

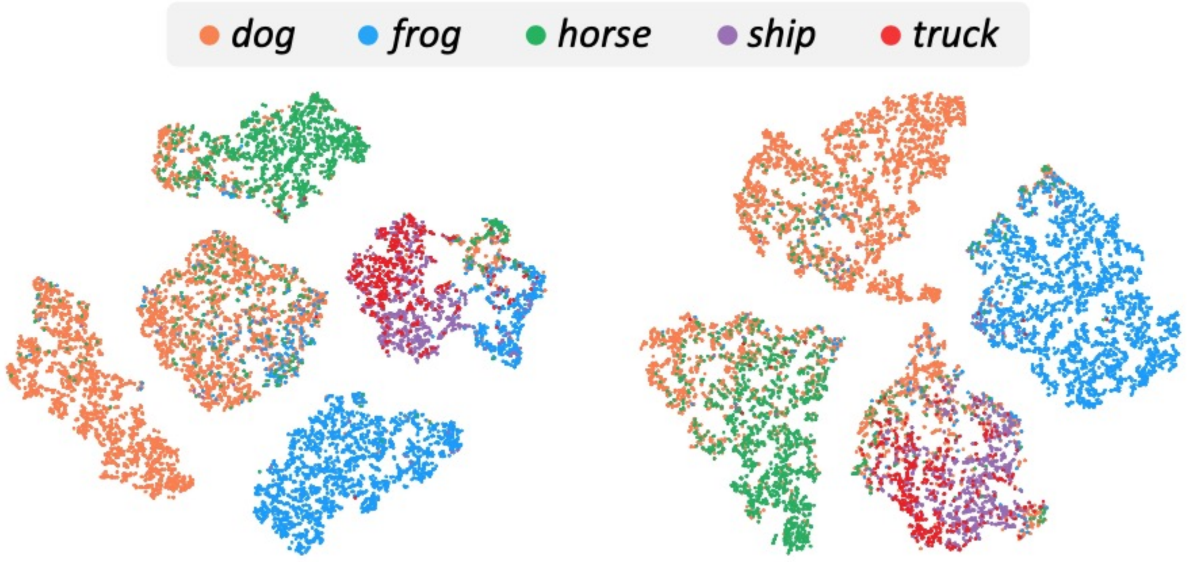
- We propose *Bootstrap Your Own Prior (BYOP)* for distribution-agnostic NCD, *i.e.*, to iteratively estimate the class distribution prior using the model prediction itself.



Experiments: visualizations

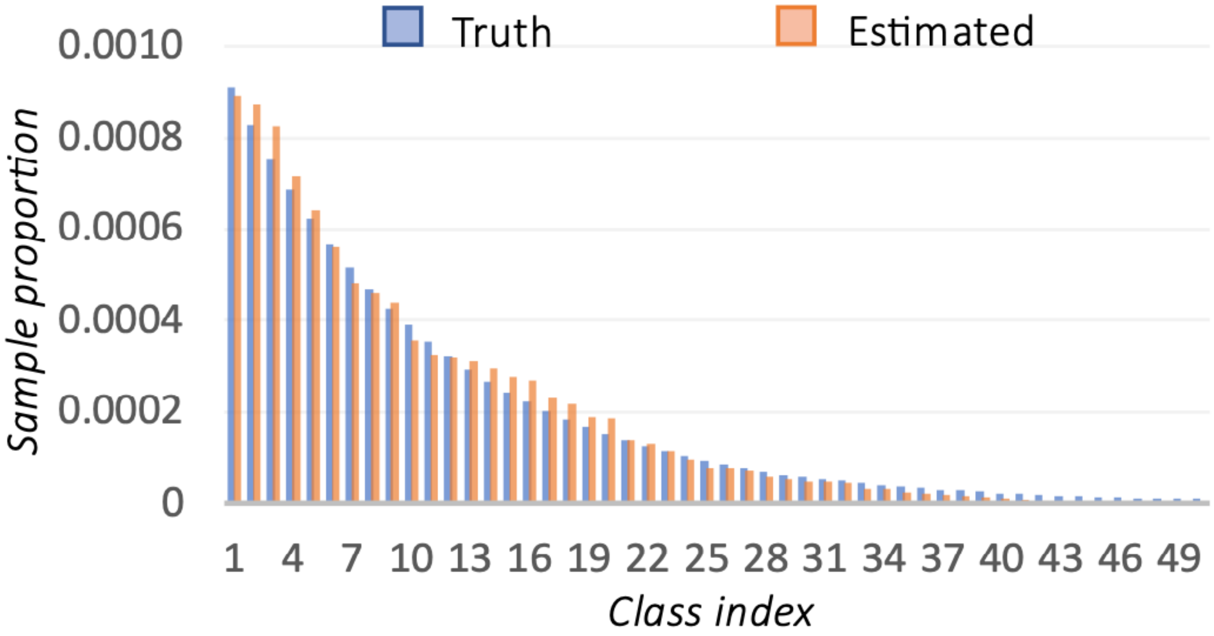
❑ **Left**: t-SNE visualizations; **right**: estimated prior *vs.* true prior

CIFAR10 w/ imbalance ratio 10



(a) UNO (b) UNO + BYOP

CIFAR100-50 w/ imbalance ratio 100



Novel Class Discovery (NCD)

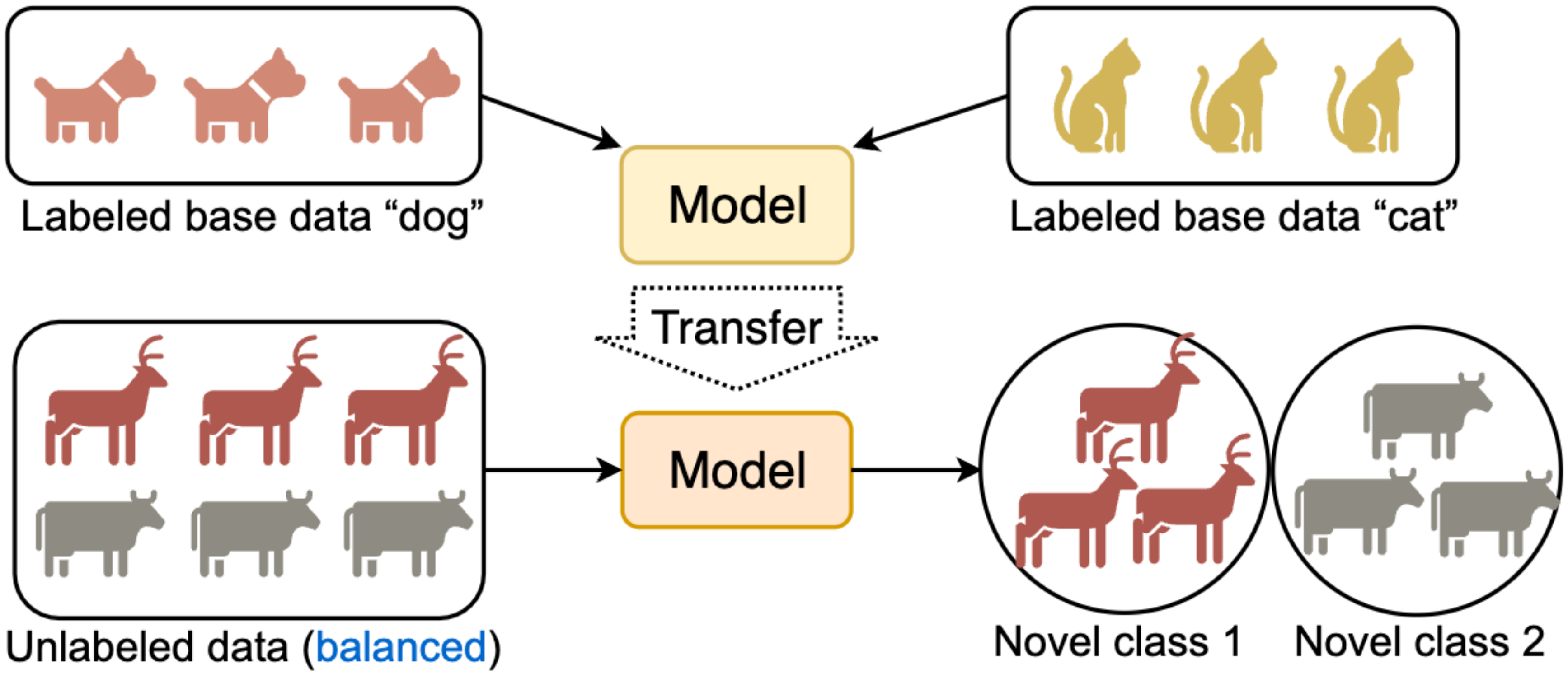
- NCD: Clustering unlabeled data according to semantic classes using the knowledge learned from labeled data

A naive solution that could be found with unsupervised clustering [1]

[1] Novel Class Discovery: an Introduction and Key Concepts, 2023

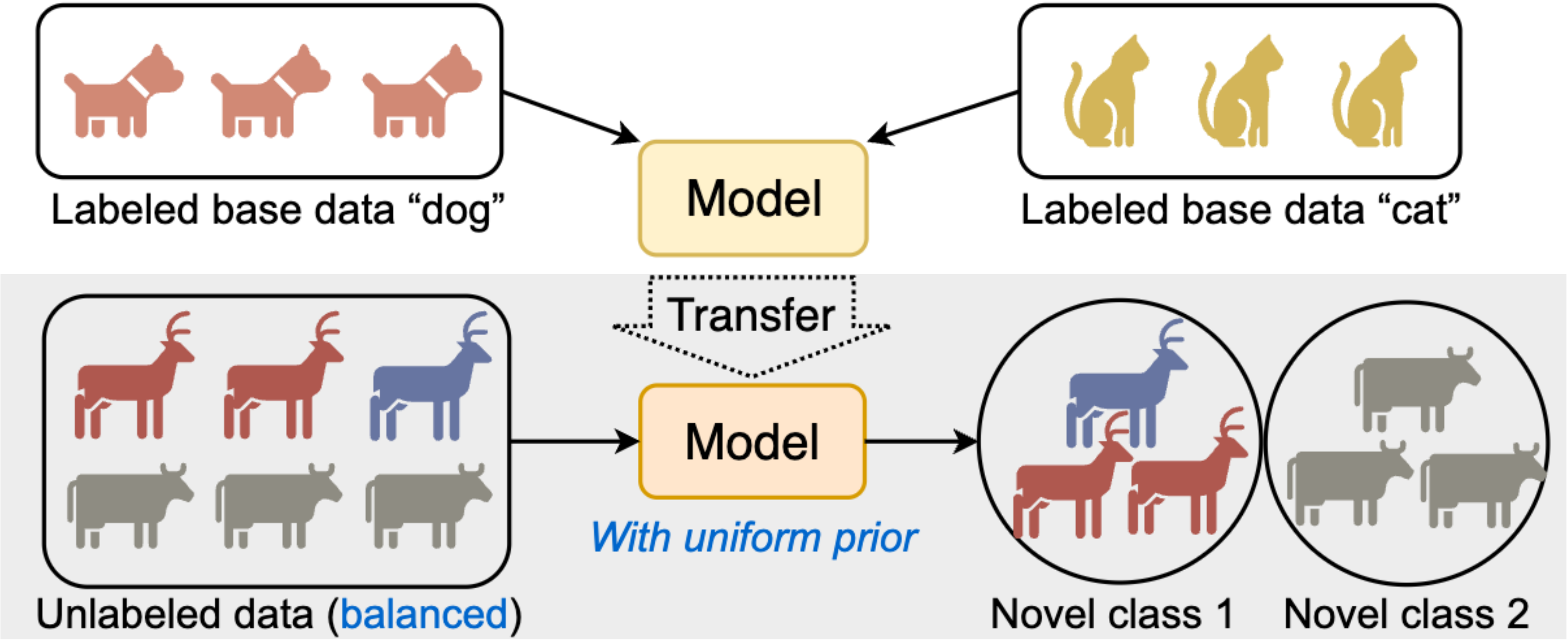
Introduction

- ❑ Existing works often hold a common assumption: unlabeled data has a uniform class distribution.



Introduction

- Existing works often hold a common assumption: unlabeled data has a uniform class distribution. But why?

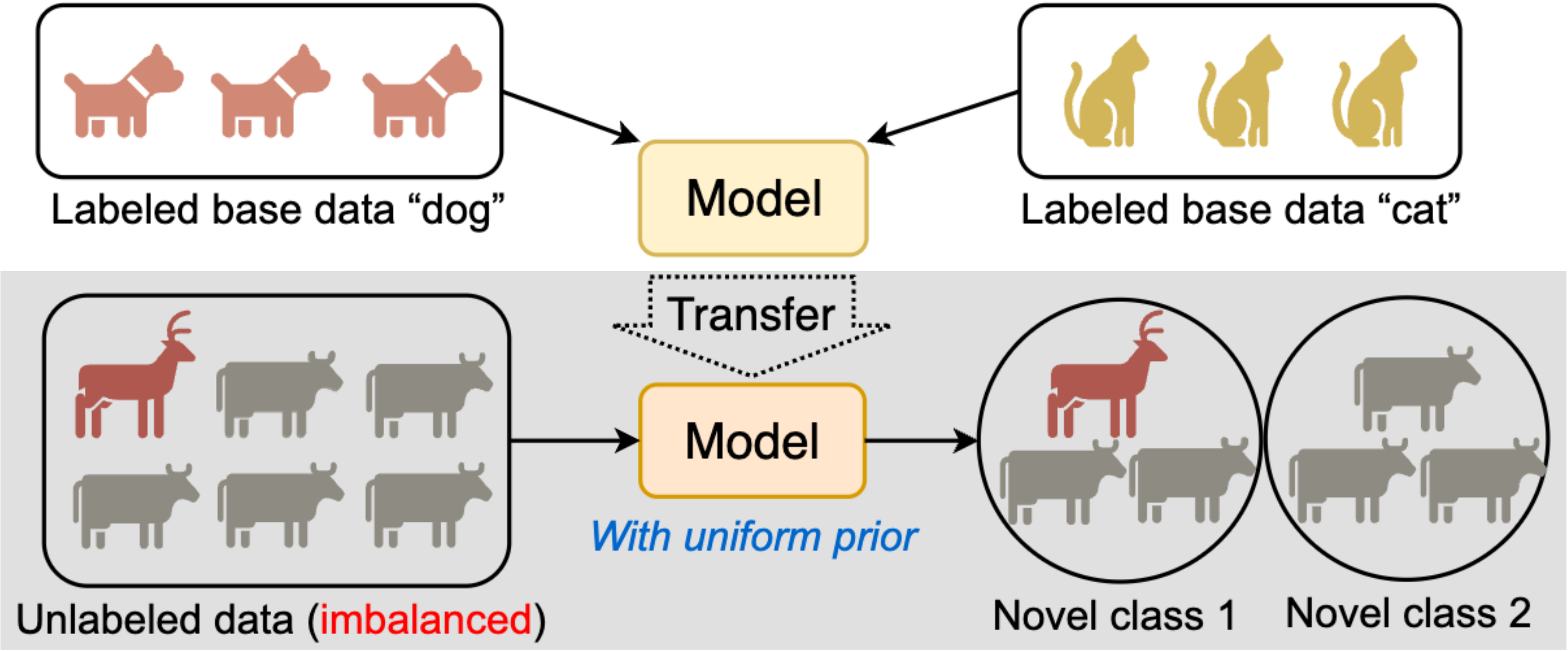


✓ When ambiguity presents, one can rely on the uniform clustering prior according to the uniform class distribution assumption

Introduction

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

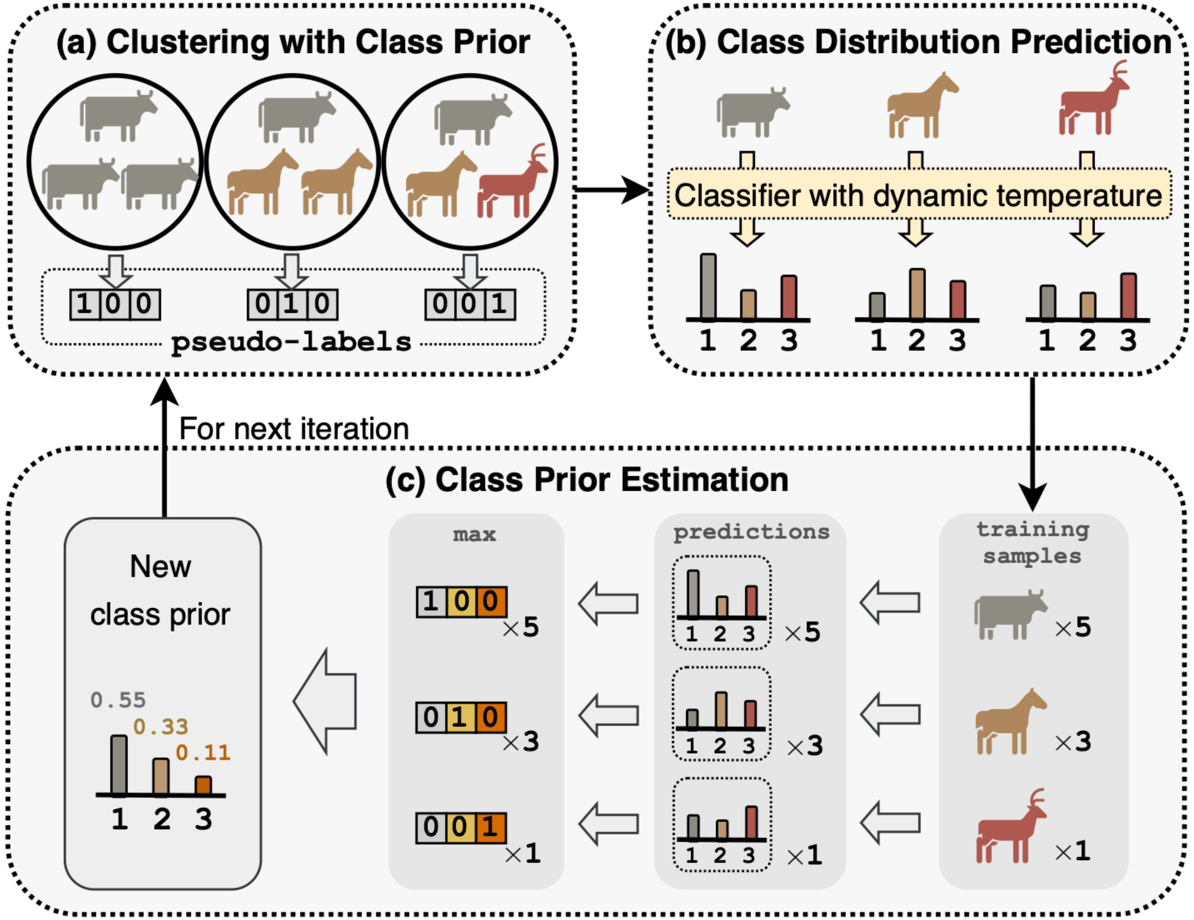
- ❑ However, unlabeled data can hardly be balanced in the real world.



- ➢ A chicken-egg problem: prior is necessary, but is unknown (Distribution-agnostic NCD)

Approach: overview

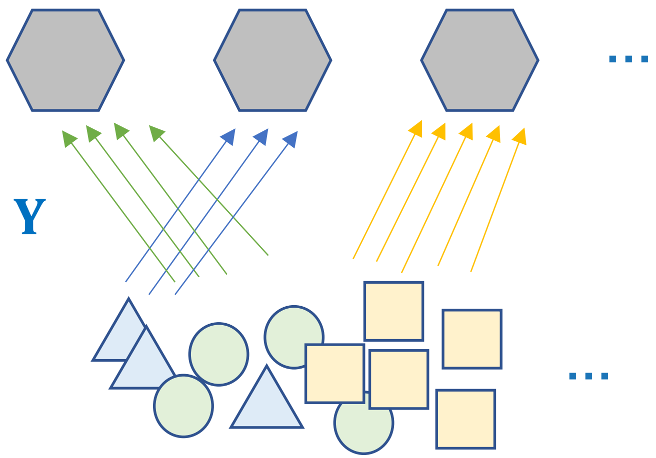
- We propose *Bootstrap Your Own Prior (BYOP)* for distribution-agnostic NCD, *i.e.*, to iteratively estimate the class distribution prior using the model prediction itself.



Approach: (1) clustering with class prior

❑ Clustering the unlabeled data using the current class prior

Clustering: assigning image features $\mathbf{X}$ to classfier weights $\mathbf{W}$ → An optimal transport (OT) problem



↓ C^n: # of novel classes; B: # of novel samples in a batch

$\mathbf{Y}$ is optimized by

$$\max_{\mathbf{Y}\in\mathcal{T}} \operatorname{tr}(\mathbf{Y}^\top \mathbf{W}^\top \mathbf{X})$$

$$\mathcal{T} = \left\{ \mathbf{Y} \in \mathbb{R}_+^{C^n \times B} \mid \mathbf{Y}\mathbf{1}_B = \boxed{\boldsymbol{p},} \mathbf{Y}^\top \mathbf{1}_{C^n} = \frac{1}{B}\mathbf{1}_B \right\}$$

Current class prior (updated in step (3))

Solved by the Sinkhorn-Knopp algorithm

Approach: (2) class distribution prediction

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

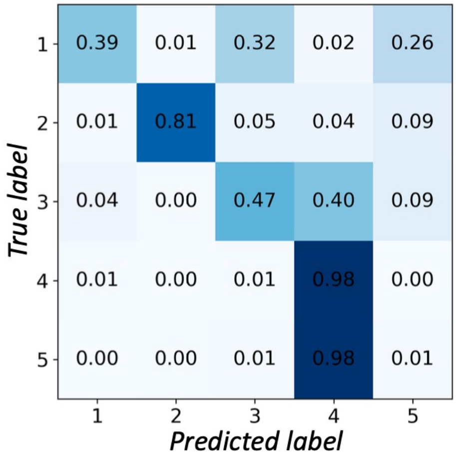
- ❑ Training a classifier using (pseudo-)labels for base/novel data

Ambiguity in the pseudo-labels → ambiguous prediction for novel samples → inaccurate prior estimation → …

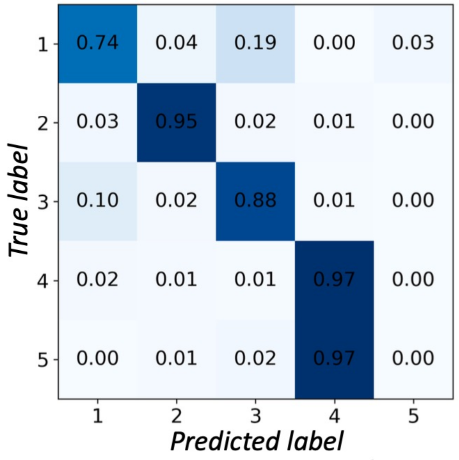
CE loss: $\mathcal{L}(\boldsymbol{x},\boldsymbol{y}) = -\sum_{c=1}^{C} y_c \log(\hat{y}_c)\,, \quad \hat{\boldsymbol{y}} = \sigma(\boldsymbol{q}/\tau)$

softmax

dynamic temperature: $\tau' = \tau/\rho\,, \quad \rho = \max\left(\sigma(\boldsymbol{q}/\tau)\right)$



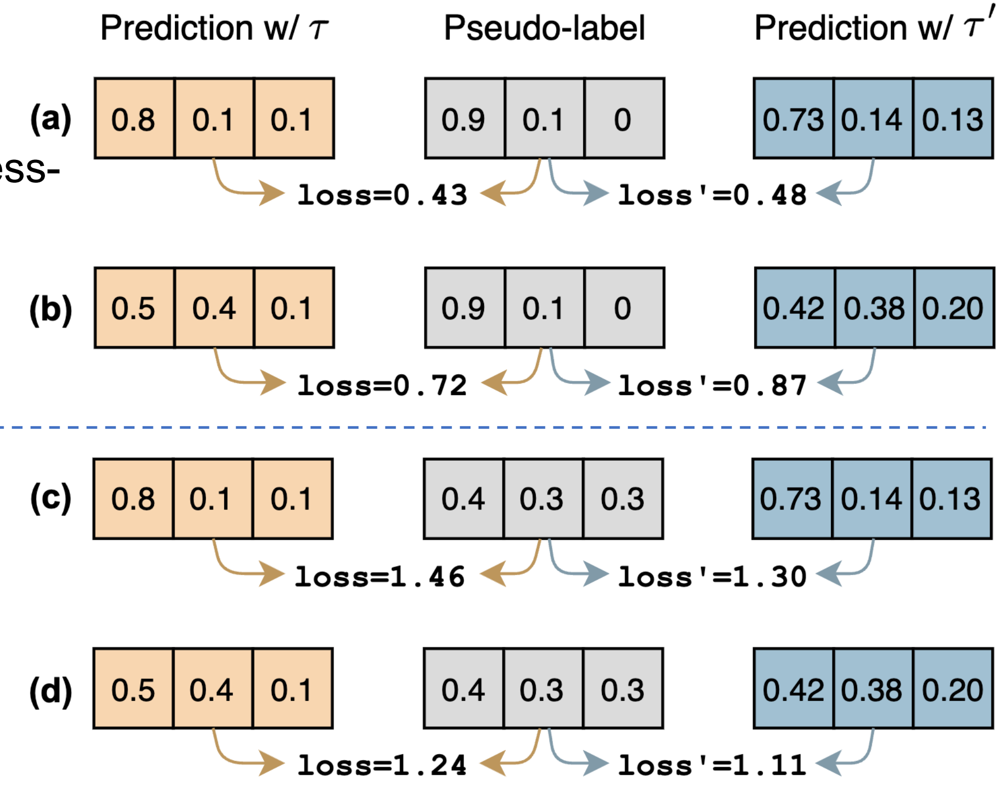
(a) BYOP w/ τ



(b) BYOP w/ τ'

When PL is less-ambiguous: encourage confidence

When PL is ambiguous: preventing learning



Approach: (3) class prior estimation

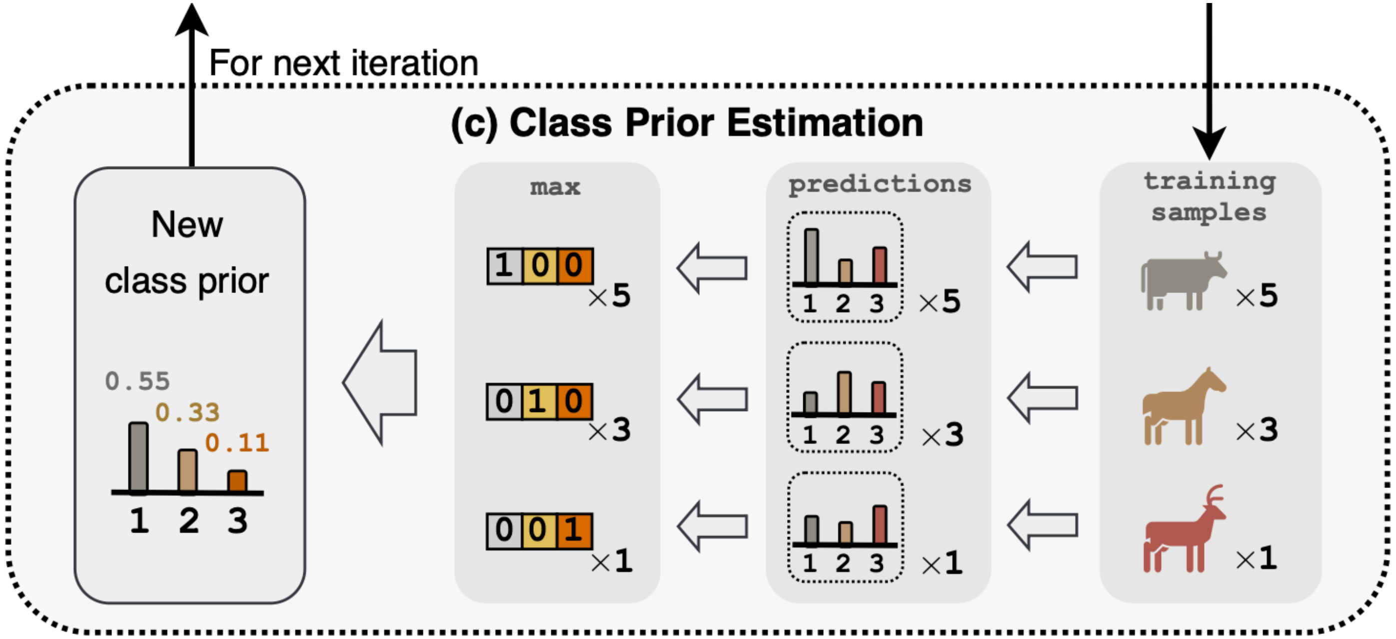
- Estimating the class prior based on the model prediction

$$r_c = \frac{1}{K}\sum_{k=1}^{K} \mathbb{1}\left(c = \arg\max_{c' \in S} q^u_{k,c'}\right)$$
$$S = \{1, \dots, C^u\}$$

$\boldsymbol{p} = [\frac{1}{C^u}, \dots, \frac{1}{C^u}]$ (iteration = 0)

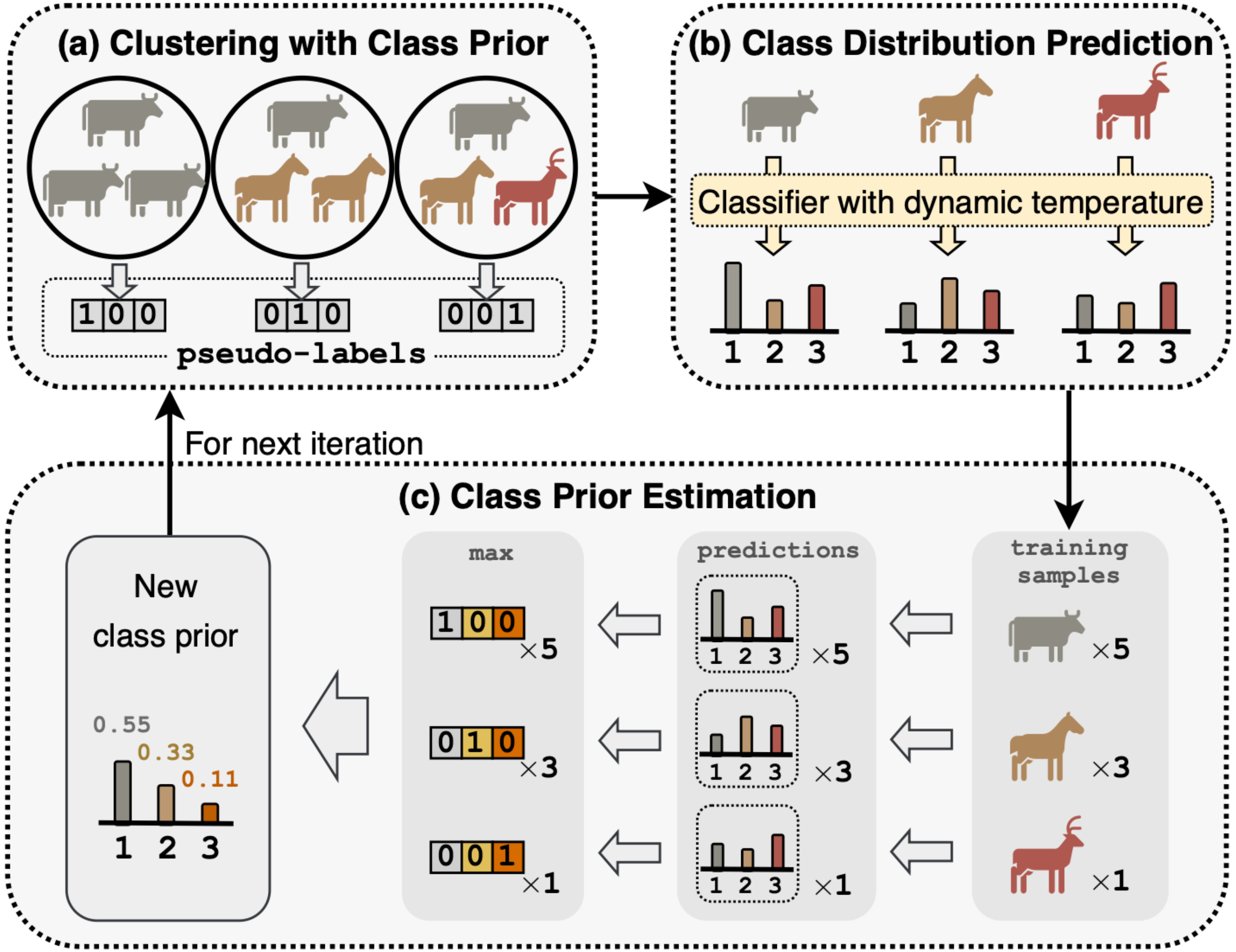
$\boldsymbol{p} \leftarrow \mu \boldsymbol{p} + (1-\mu)\boldsymbol{r}$ (afterwards)

First-in-first-out queue: $\mathcal{K} = \{\boldsymbol{q}^u_1, \dots, \boldsymbol{q}^u_K\}$



Approach: overview

- Bootstrap Your Own Prior (BYOP) for distribution-agnostic NCD



Experiments: datasets

❑ Each dataset is split into base subset and novel subset; each subset is again split into training data and testing data

Subset →	Base		Novel	
Dataset ↓	Images	Classes	Images	Classes
CIFAR10	25K	5	25K	5
CIFAR100-20	40K	80	10K	20
CIFAR100-50	25K	50	25K	50
Tiny-ImageNet	50K	100	50K	100

❑ For distribution-agnostic NCD, training data is imbalanced (controlled by imbalance ratio); testing data is balanced

Experiments: evaluation metrics

- For distribution-agnostic NCD, training data is imbalanced (controlled by imbalance ratio); testing data is balanced
- Metric 1: traditional protocol
  - *Clustering acc on training data of novel subset*
- Metric 2: task-aware protocol
  - *Classification acc on testing data of base subset*
  - *Clustering acc on testing data of novel subset*
- Metric 3: task-agnostic protocol (Generalized NCD)
  - *Classification/clustering acc on {testing data of base subset, testing data of novel subset}*
- Clustering acc: $$ACC = \frac{1}{M}\sum_{i=1}^{M}\mathbb{1}[y_i^u = \mathrm{map}(\hat{y}_i^u)]$$ (Using Hungarian algorithm)

Experiments: comparison with SOTAs

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Dataset →	CIFAR10 *(imbalance ratio: 100)*							CIFAR10 *(imbalance ratio: 10)*						
Protocol →	Trad.	Task-aware			Task-agnostic			Trad.	Task-aware			Task-agnostic		
Method ↓	Nov.	Base	Nov.	All	Base	Nov.	All	Nov.	Base	Nov.	All	Base	Nov.	All
RS [12]	46.3	**71.8**	43.2	57.5	–	–	–	69.7	87.4	63.6	75.5	–	–	–
RS+ [12]	45.3	64.4	50.1	57.3	**64.4**	55.5	**60.0**	66.5	77.3	63.3	70.3	77.3	62.3	69.8
NCL [56]	47.2	71.6	43.1	57.4	–	–	–	62.6	86.9	56.9	71.9	–	–	–
UNO [11]	43.9	69.6	52.2	60.9	56.0	55.6	55.8	59.6	88.1	59.1	73.6	78.2	58.8	68.5
UNO + BYOP	**59.3**	70.1	53.3	61.7	56.6	**56.6**	56.6	63.2	88.5	61.7	75.1	78.4	61.0	69.7
ComEx [47]	44.6	70.0	53.8	61.9	57.8	55.1	56.5	68.1	87.9	63.5	75.7	81.3	63.3	72.3
ComEx + BYOP	57.0	71.4	**54.5**	**63.0**	59.3	56.0	57.7	**72.1**	**88.7**	**65.5**	**77.1**	**82.2**	**65.4**	**73.8**

Dataset →	CIFAR100-50 *(imbalance ratio: 100)*							CIFAR100-50 *(imbalance ratio: 10)*						
Protocol →	Trad.	Task-aware			Task-agnostic			Trad.	Task-aware			Task-agnostic		
Method ↓	Nov.	Base	Nov.	All	Base	Nov.	All	Nov.	Base	Nov.	All	Base	Nov.	All
RS [12]	30.7	40.8	23.3	32.1	–	–	–	27.4	48.4	23.7	36.1	–	–	–
RS+ [12]	29.6	35.5	23.0	29.3	35.5	22.5	29.0	26.0	38.5	23.4	31.0	38.5	21.9	30.2
NCL [56]	30.4	39.9	21.8	30.9	–	–	–	27.8	47.0	23.4	35.2	–	–	–
UNO [11]	25.7	44.8	21.3	33.1	36.9	22.6	29.8	33.7	63.9	32.5	48.2	53.6	30.8	42.2
UNO + BYOP	**35.5**	45.4	23.4	34.4	37.3	24.0	30.7	**38.3**	64.6	**34.9**	**49.8**	54.1	**33.0**	43.6
ComEx [47]	27.1	45.9	22.6	34.3	39.5	23.3	31.4	34.3	64.6	32.6	48.6	57.7	31.9	44.8
ComEx + BYOP	33.1	**46.9**	**24.1**	**35.5**	**40.9**	**24.4**	**32.7**	37.4	**65.3**	33.5	49.4	**58.5**	32.8	**45.7**

Experiments: comparison with SOTAs

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

- + Logit Adjustment* (post-hoc adjustment using estimated prior)

CIFAR100-50 with imbalance ratio 100 (task-aware)

Subset →	Base (*test*)				Novel (*test*)			
Method ↓	Many	Med.	Few	All	Many	Med.	Few	All
UNO [11]	**78.5**	41.9	14.1	44.8	23.8	24.7	**15.4**	21.3
UNO + BYOP	76.6	44.2	15.4	45.4	**33.6**	24.1	12.4	23.4
UNO + BYOP†	76.3	**47.9**	**18.9**	**47.7**	28.6	**30.2**	12.9	**23.8**

$$\boldsymbol{q}^\dagger = \boldsymbol{q} - \tau \log(\boldsymbol{p})$$

*Long-Tail Learning via Logit Adjustment, ICLR'21

Experiments: ablation study

- ❑ Estimated prior $\boldsymbol{p}$, dynamic temperature τ

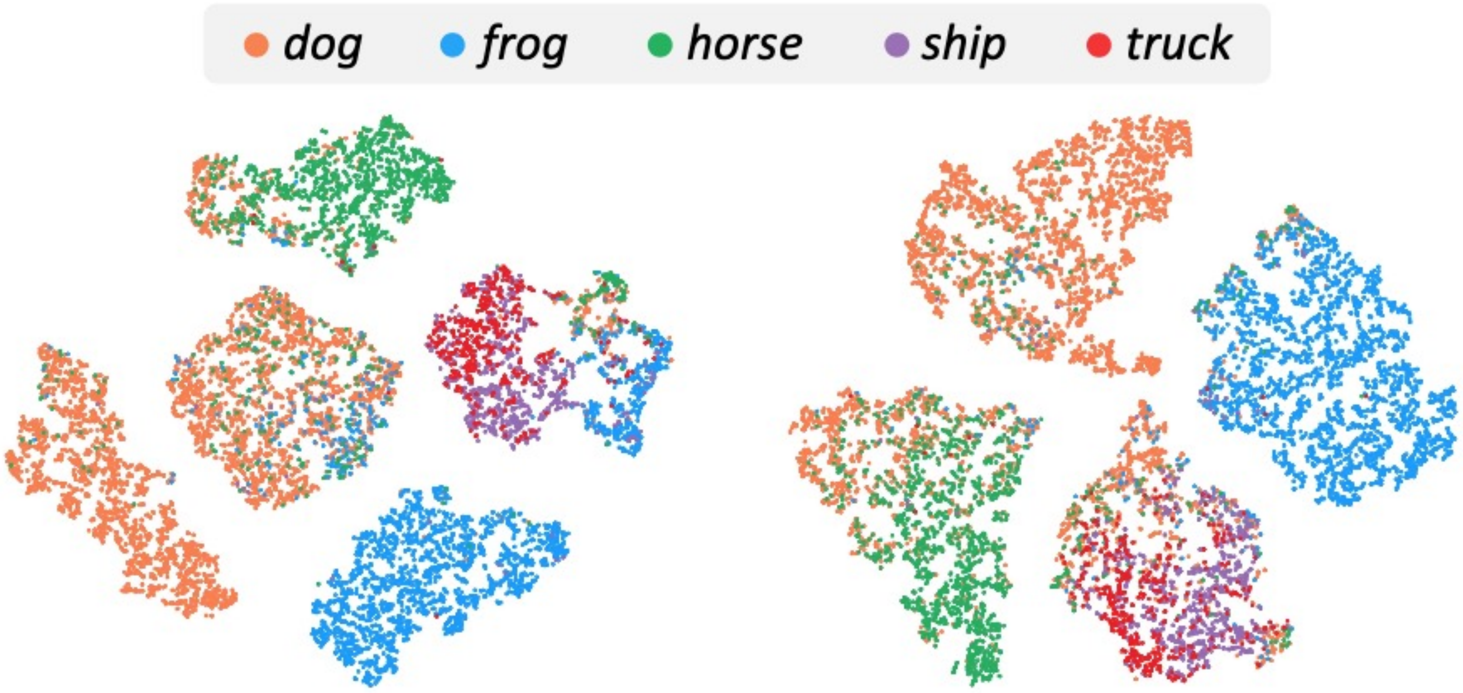
CIFAR100-50 with imbalance ratio 100 (traditional, task-aware)

Subset *(split)* →	Novel *(training)*				Base *(test)*				Novel *(test)*			
Method ↓	Many	Med.	Few	All	Many	Med.	Few	All	Many	Med.	Few	All
Uniform $\boldsymbol{p}$	24.6	**33.9**	**16.1**	25.7	**78.5**	41.9	14.1	44.8	23.8	24.7	15.4	21.3
Oracle $\boldsymbol{p}$	32.1	28.1	12.0	30.8	77.4	43.4	15.1	45.3	27.2	**28.3**	12.9	22.8
Estimated $\boldsymbol{p}$	29.7	31.0	14.6	29.4	76.1	43.4	15.2	44.9	28.7	23.4	**15.6**	22.6
Uniform $\boldsymbol{p}$ + dynamic τ	26.8	33.6	15.7	27.5	76.5	43.7	15.4	45.2	27.0	27.8	12.5	22.4
Oracle $\boldsymbol{p}$ + dynamic τ	**39.1**	28.7	11.0	**36.5**	76.6	43.6	**16.6**	**45.6**	**36.0**	22.7	11.0	23.2
Estimated $\boldsymbol{p}$ + dynamic τ	37.7	28.6	13.9	35.5	76.6	**44.2**	15.4	45.4	33.6	24.1	12.4	**23.4**

Experiments: visualization

- t-SNE visualizations of the 5 novel classes (training data of CIFAR10)

imbalance ratio: 10



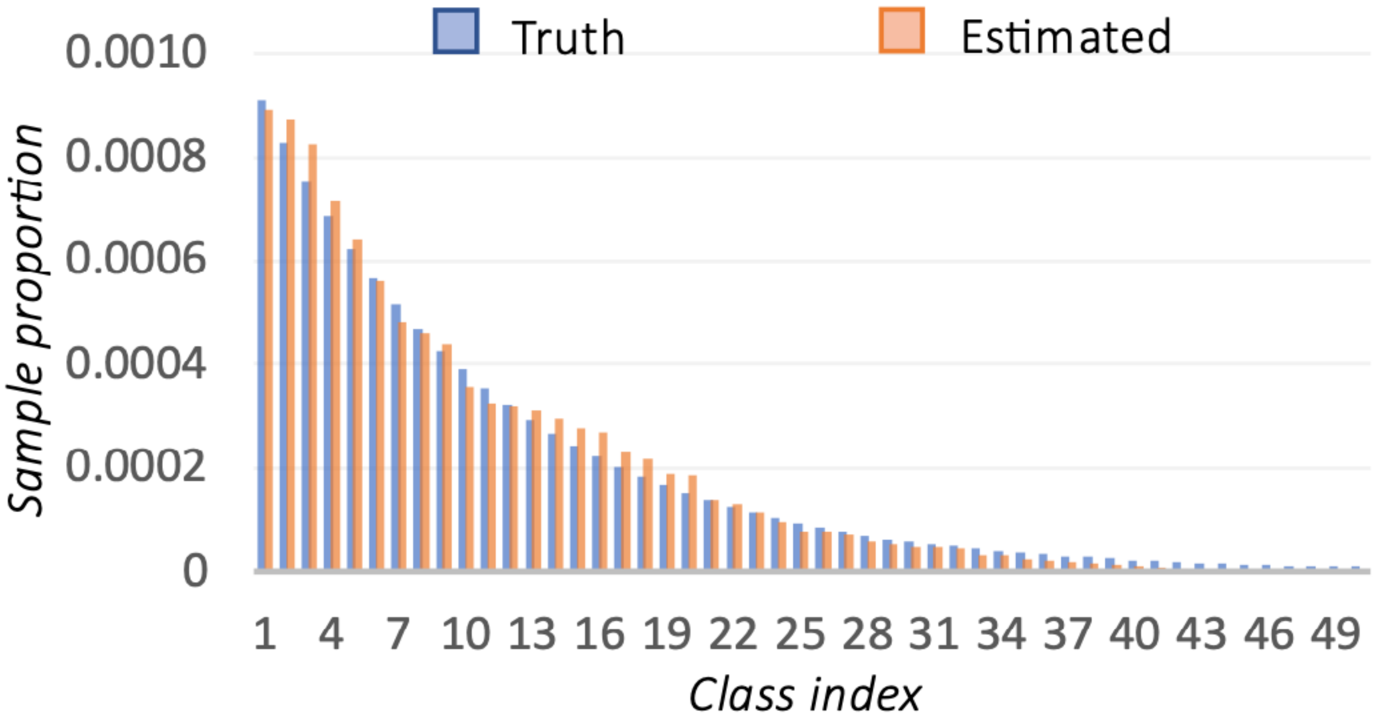
(a) UNO

(b) UNO + BYOP

Experiments: visualization

- Estimated prior vs. true prior

CIFAR100-50 with imbalance ratio 100



- Code: https://github.com/muliyangm/BYOP
- Thank you for listening!