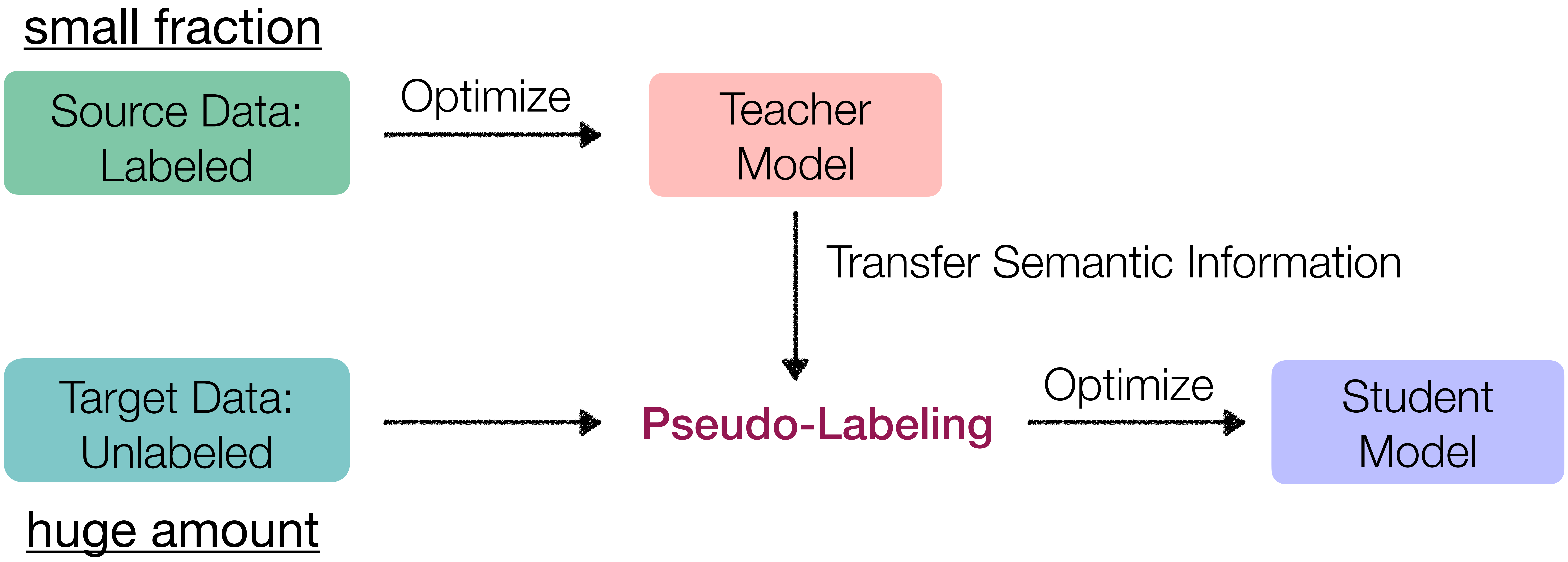
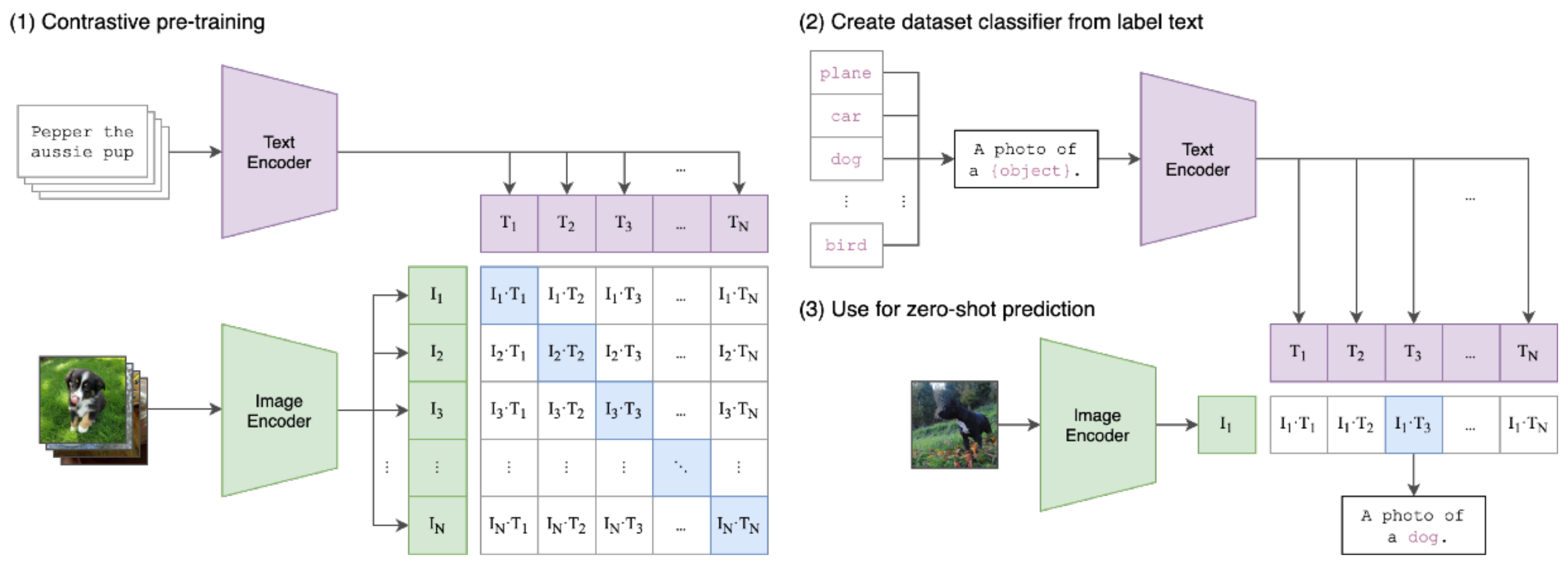
Debiased Learning from Naturally Imbalanced Pseudo-Labels

Xudong Wang, Zhirong Wu, Long Lian, Stella X. Yu
UC Berkeley / ICSI, Microsoft Research
CVPR 2022

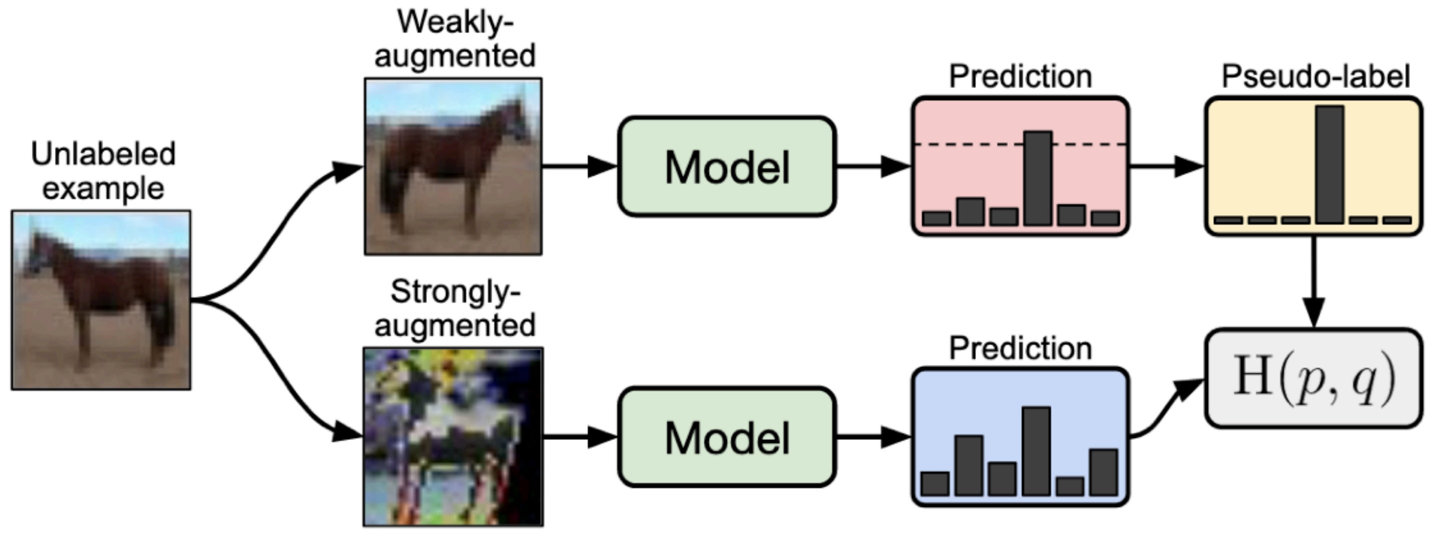
What is Pseudo-Labeling?



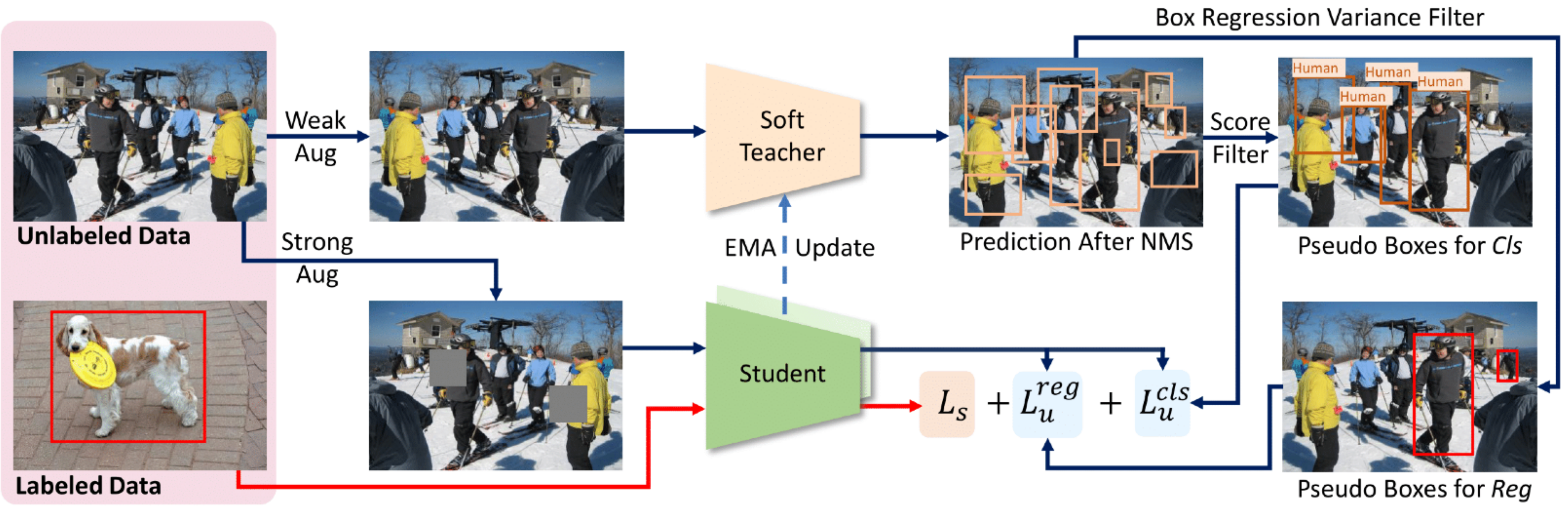
Pseudo-Labels are Widely Used



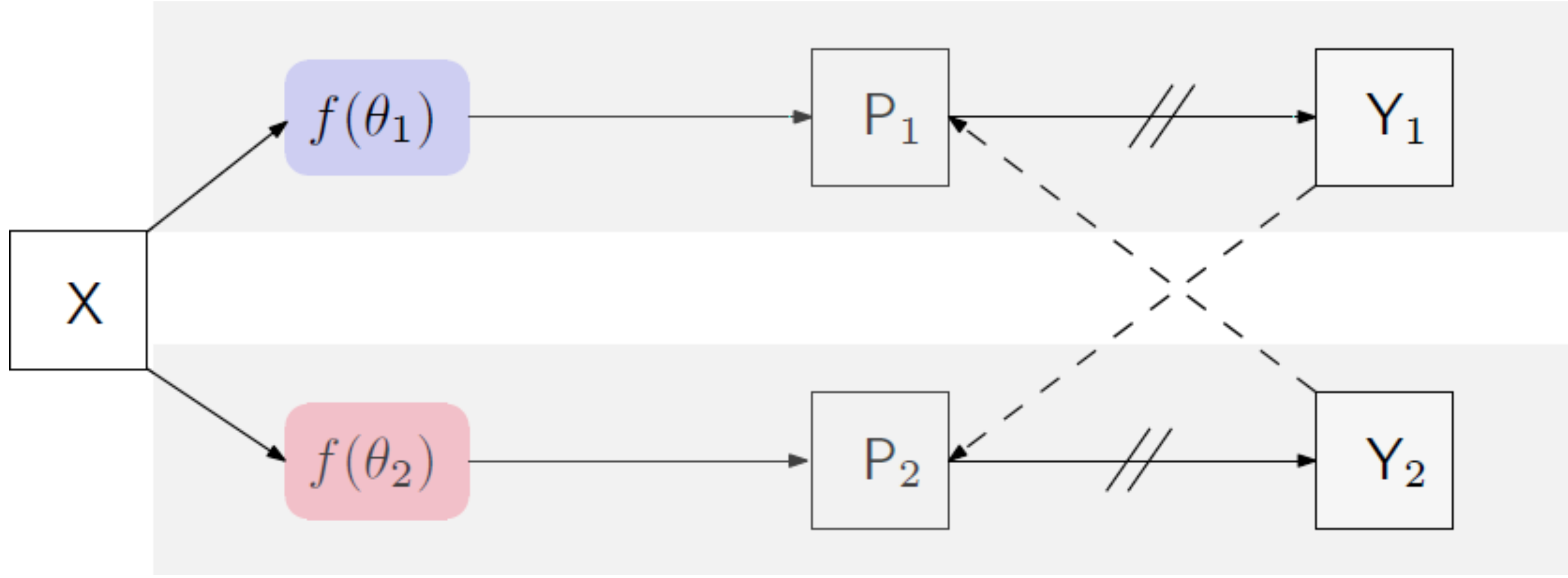
[Zero-shot Classification]
CLIP (*ICML 2021*)



[Semi-supervised Classification]
FixMatch *(NeurIPS 2020);* ***Noisy Student*** *(CVPR 2020)*



[Semi-supervised Object Detection]
Unbiased Teacher *(ICLR 2021);* ***Soft Teacher*** *(ICCV 2021)*

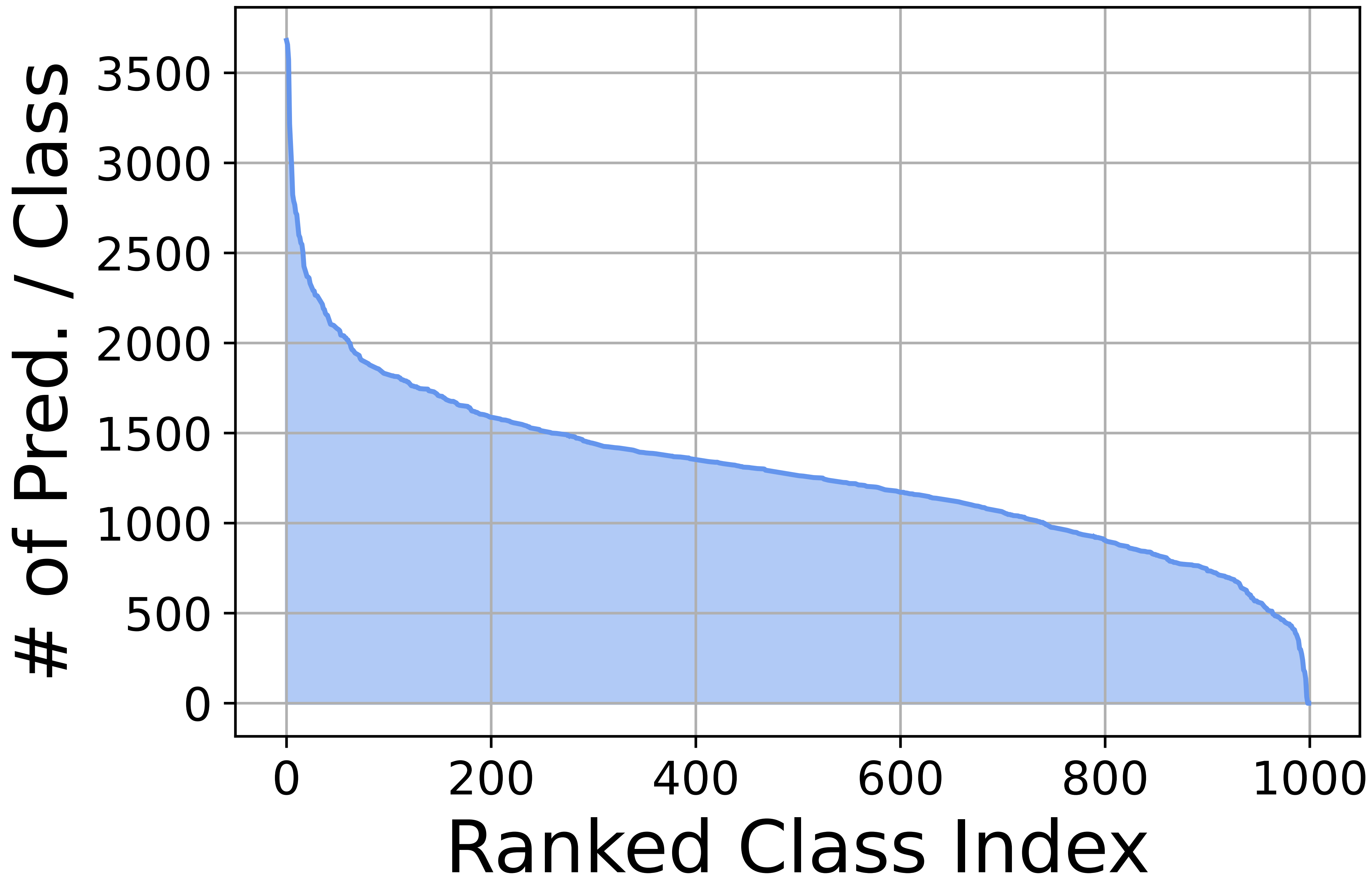


[Semi-supervised Semantic Segmentation]
CPS *(CVPR 2021)*

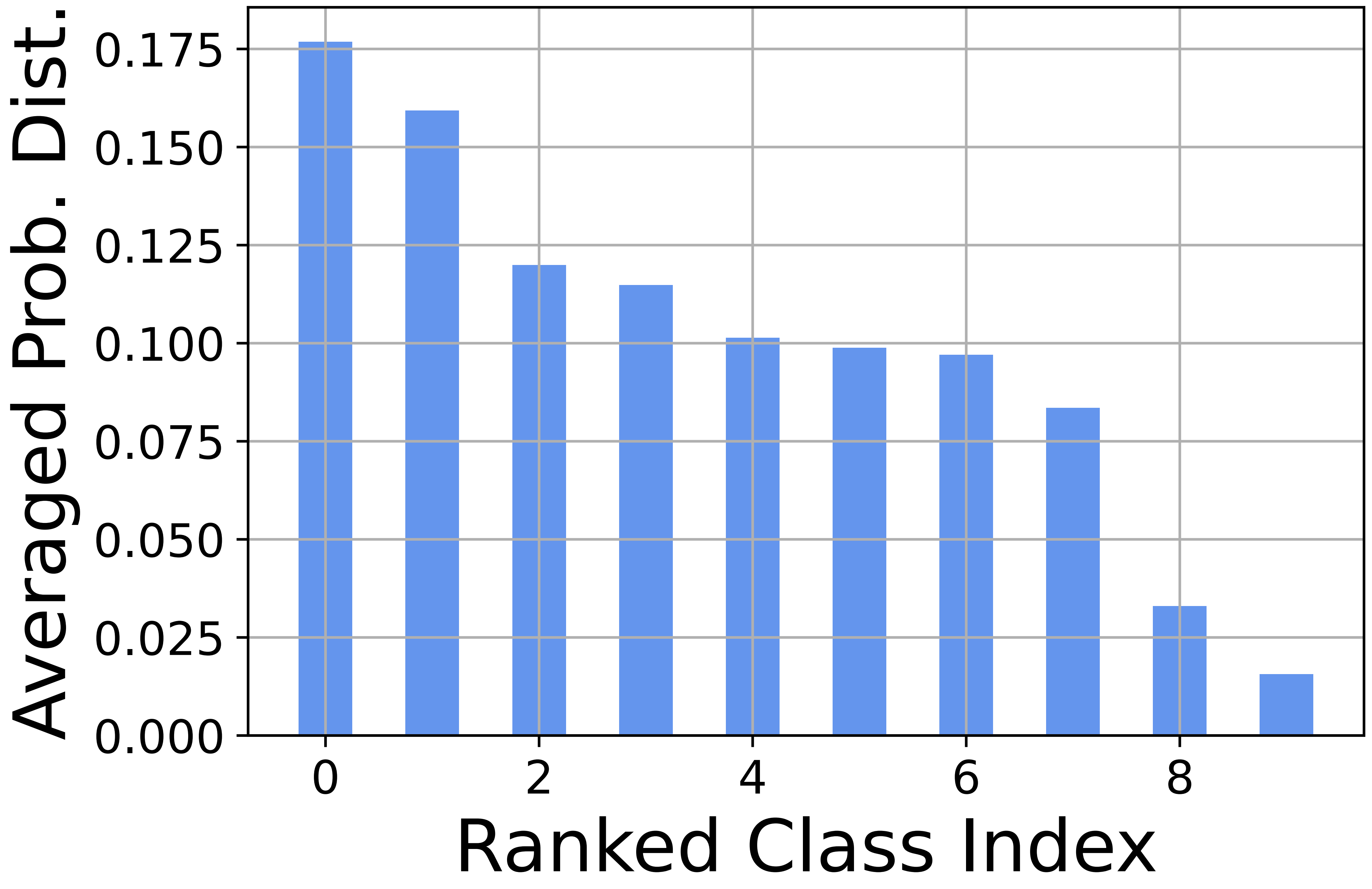
Pseudo-Labels are Naturally Imbalanced!

- even when both source and target data are balanced

CLIP on ImageNet

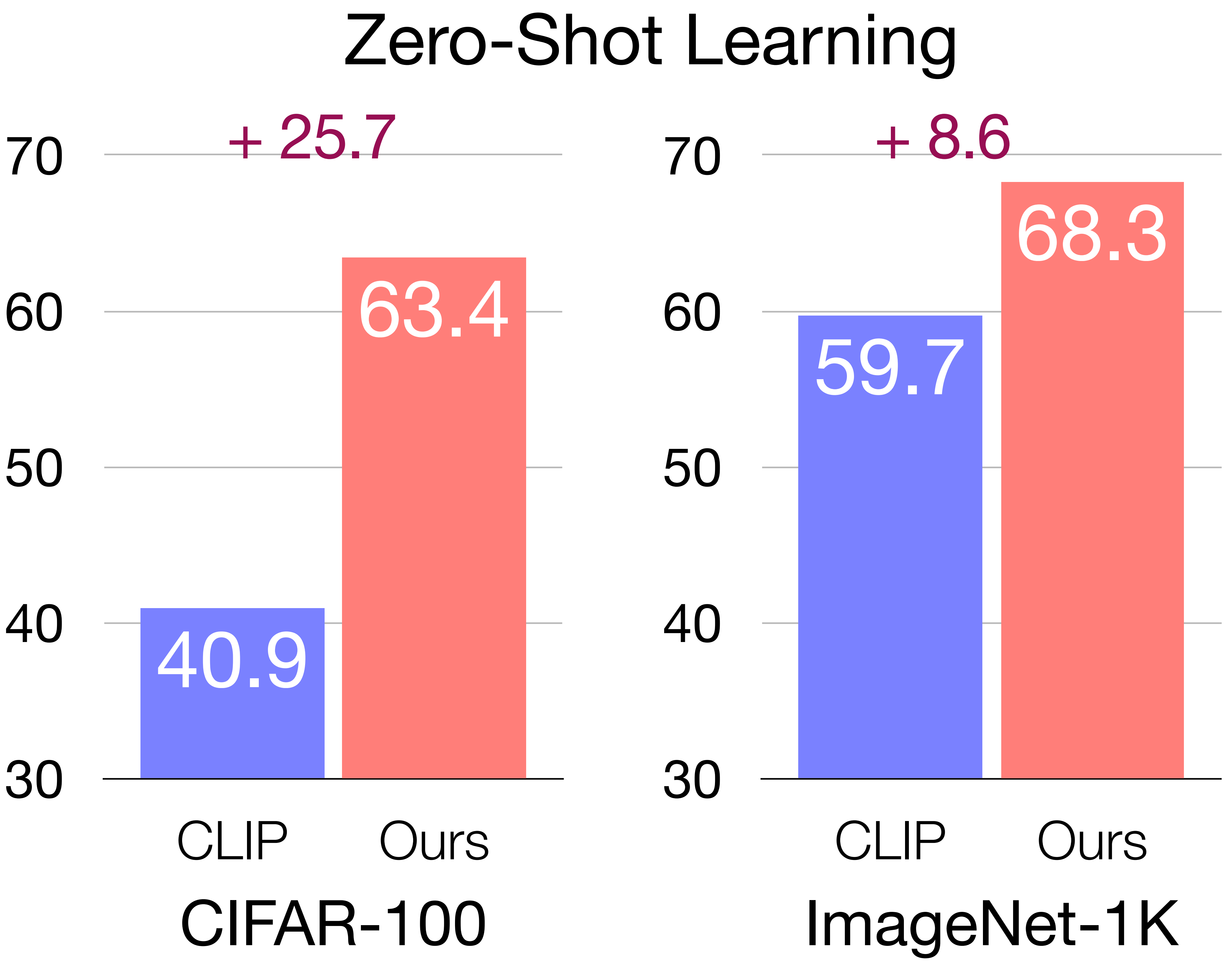


FixMatch on CIFAR



[1] CLIP: Radford, Alec, et al. "Learning transferable visual models from natural language supervision." ICML 2021
[2] FixMatch: Sohn, K., et al. "FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence." NeurIPS 2020.

Debiased Learning Delivers Significant Gains

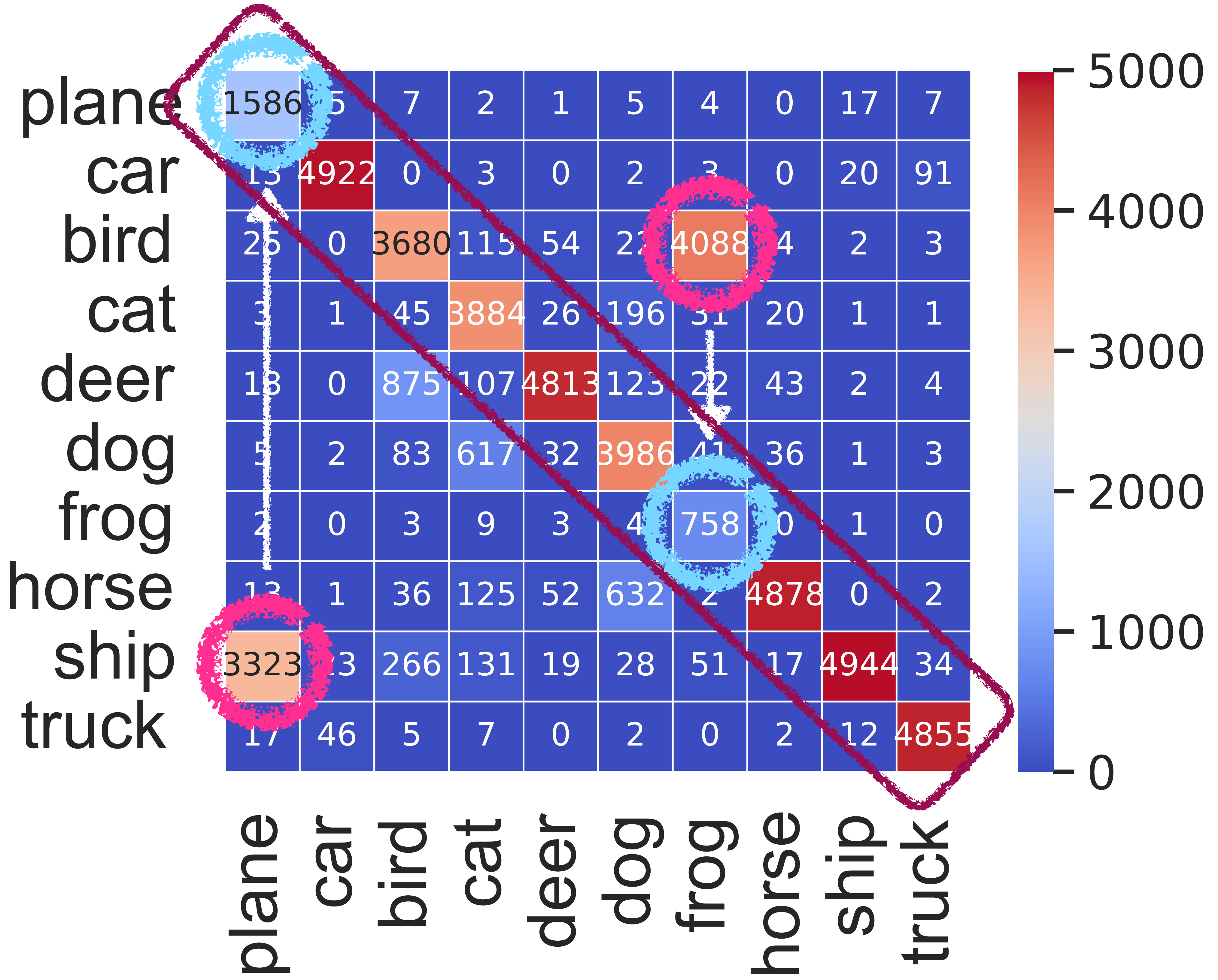


[1] CLIP: Radford, Alec, et al. "Learning transferable visual models from natural language supervision." ICML 2021
[2] FixMatch: Sohn, K., et al. "FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence." NeurIPS 2020.

$$\left.\begin{array}{l}\underline{\text{balanced}}\text{ source data}\\ \underline{\text{balanced}}\text{ target data}\end{array}\right\} \not\Rightarrow \underline{\text{balanced}}\text{ pseudo-labels}$$

Why FixMatch is Highly Biased?

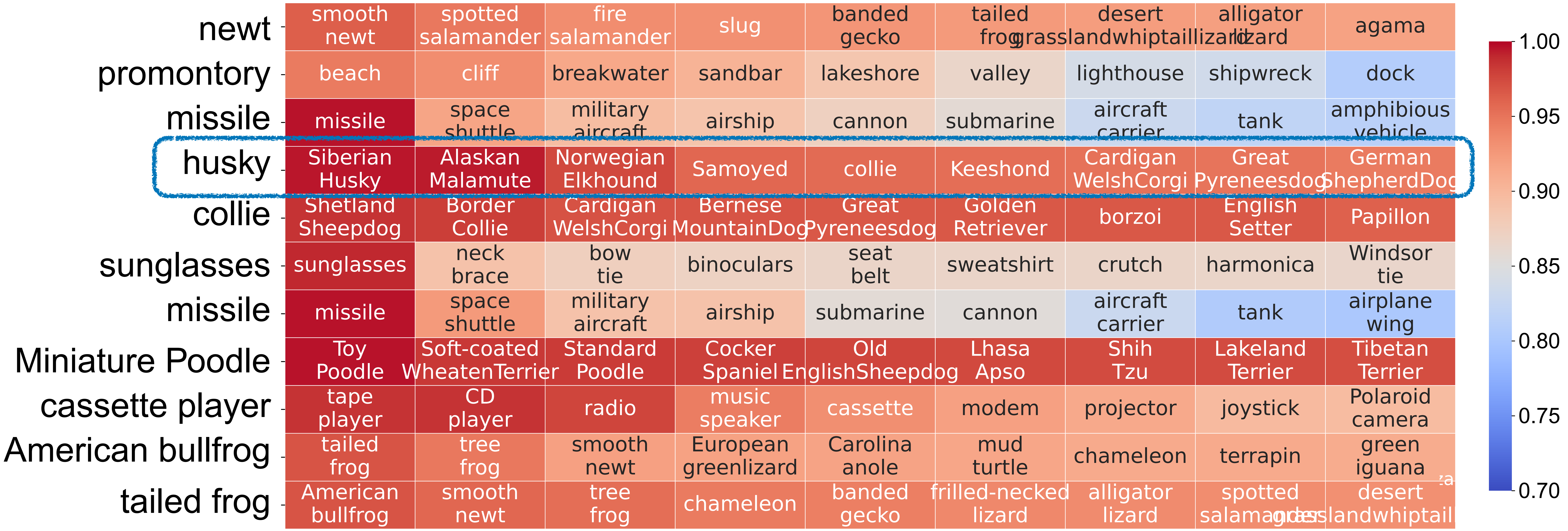
- High inter-class correlation ⇒ more missed pseudo-labels



Confusion Matrix on CIFAR-10

Why CLIP is Highly Biased?

- High inter-class correlation ⇒ more missed pseudo-labels



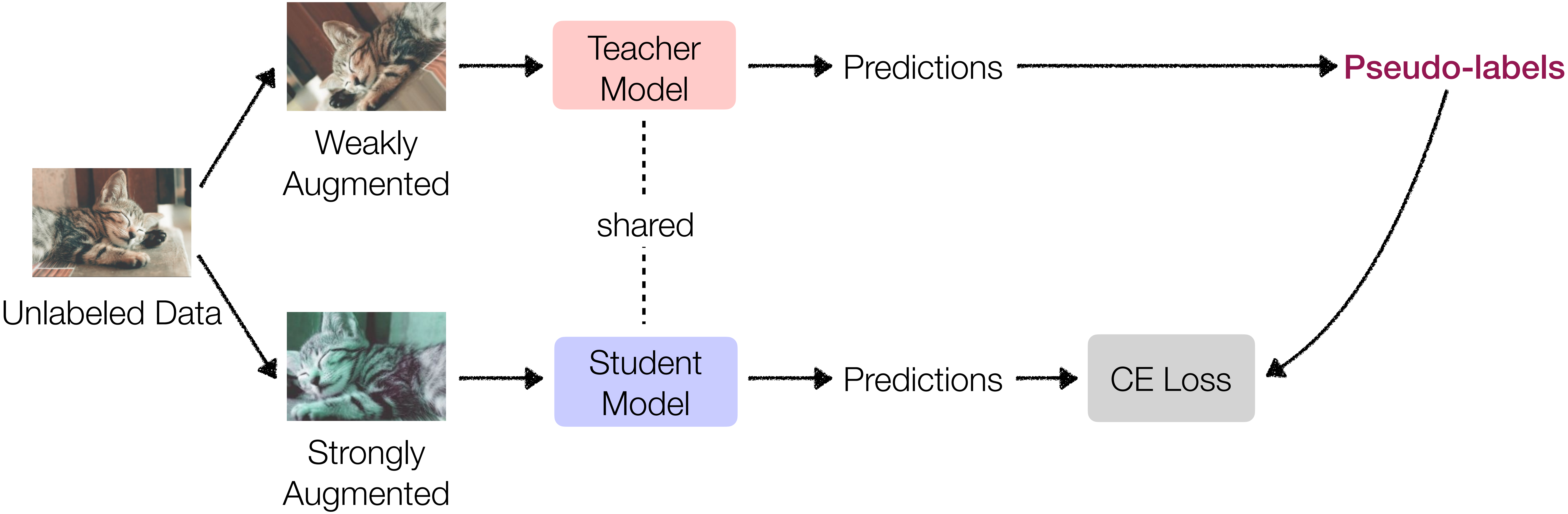
Correlation Matrix for Least-10 Classes and Nearest Neighbors (ImageNet)

What Should Be Addressed to Debias the Model?

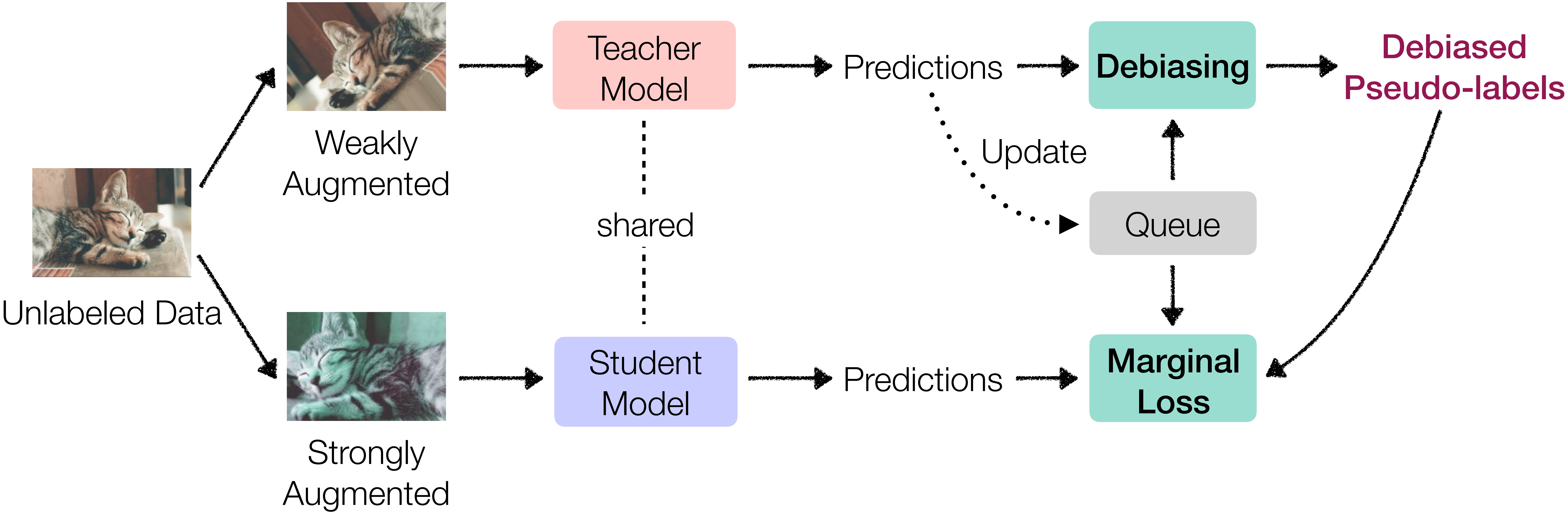
- Imbalanced class-distribution of pseudo-labels

- High inter-class confusion

Pseudo-Labeling with FixMatch



Debiased Pseudo-Labeling (w. FixMatch)

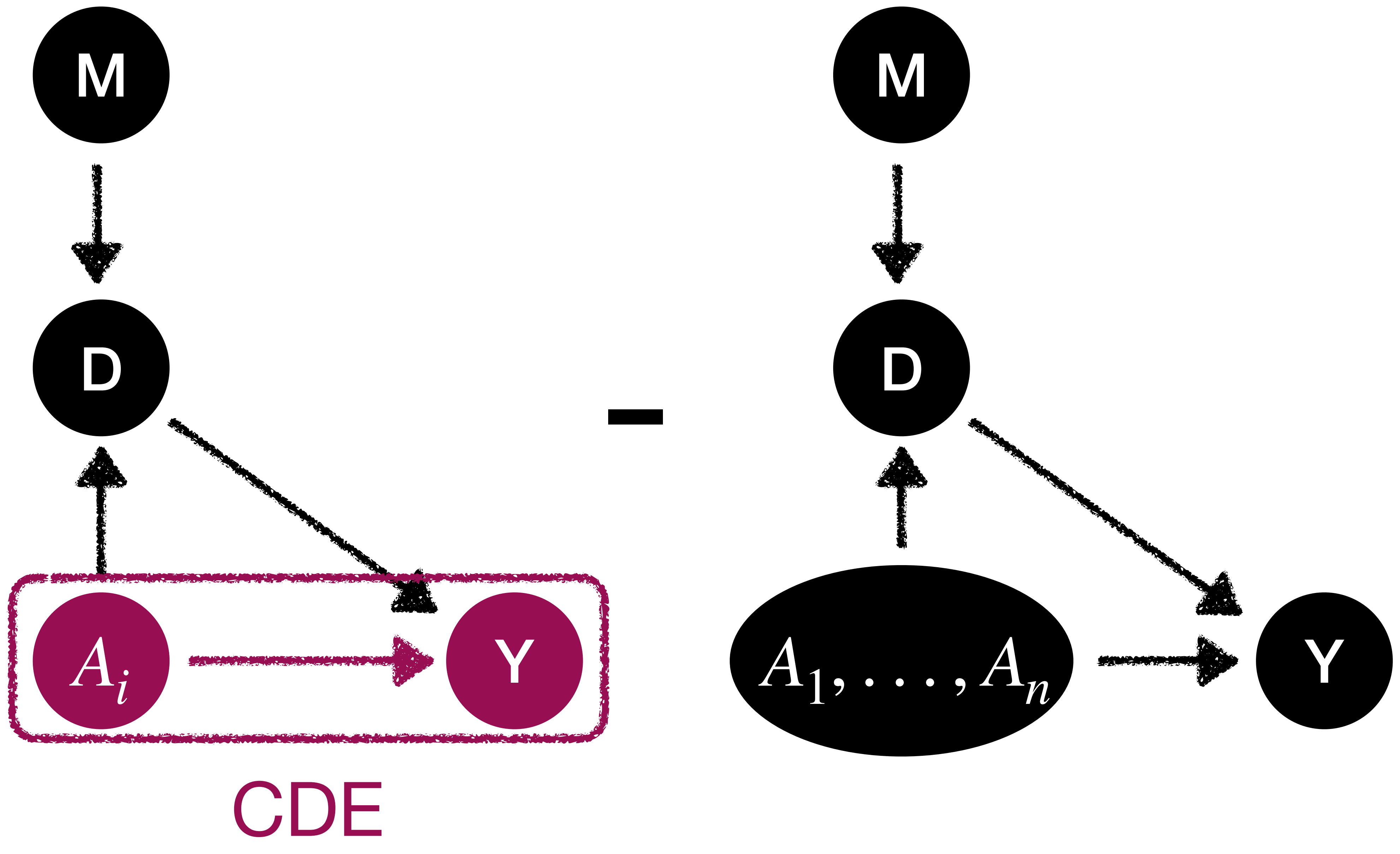


<u>Our method can be integrated into various methods!</u>

Debias with Adaptive Causal Inference

- Target: generate unbiased predictions via pursuing Controlled Direct Effect

$$\text{CDE}(Y_i) = [Y_i | do(A_i), do(D)] - [Y_i | do(\hat{A}), do(D)]$$



A_i: input data

Y: prediction

D: mediator

M: model bias

n: instance number

Debias with Adaptive Causal Inference

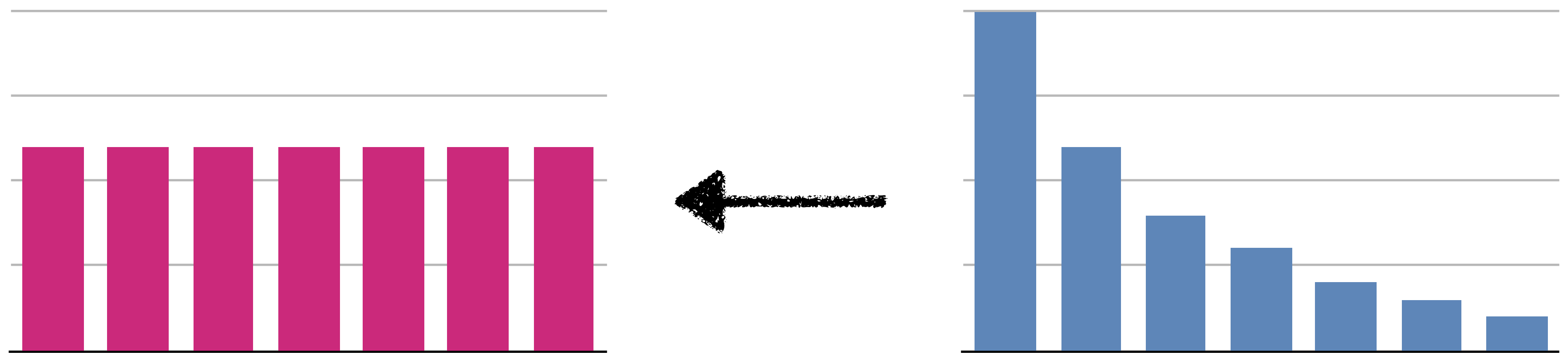
- Target: purse Controlled Direct Effect (CDE)

biased logit

$$\boxed{\tilde{f}_i} = \boxed{f(\alpha(x_i))} - \lambda \boxed{\log \hat{p}}, \ \textbf{where } \hat{p} = \frac{1}{n}\sum_{k=1}^{n} p_k$$

debiased logit

indirect effect
(dynamically changed)

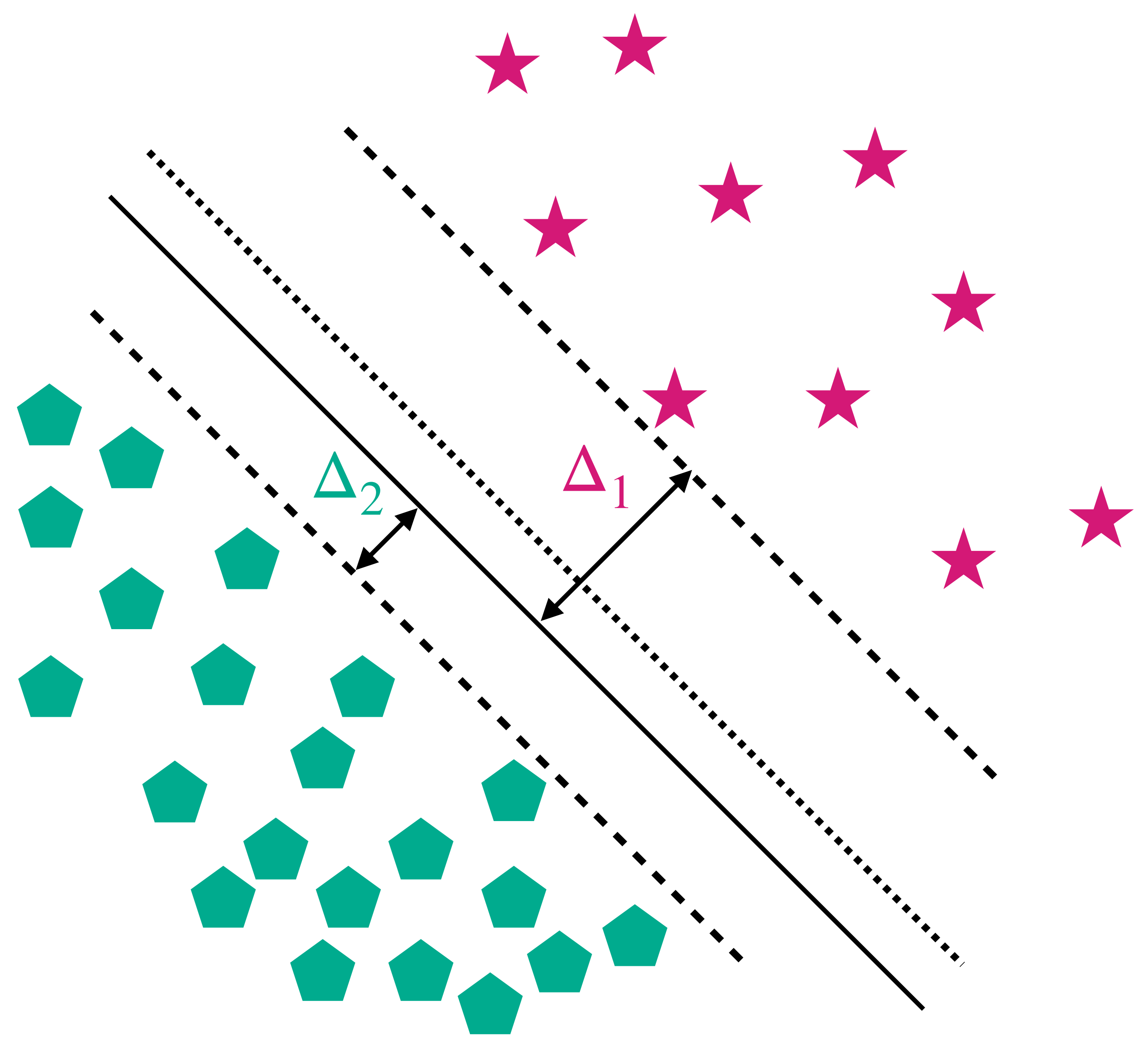


Debias with Adaptive Marginal Loss

- Target: counteract inter-class confusion.

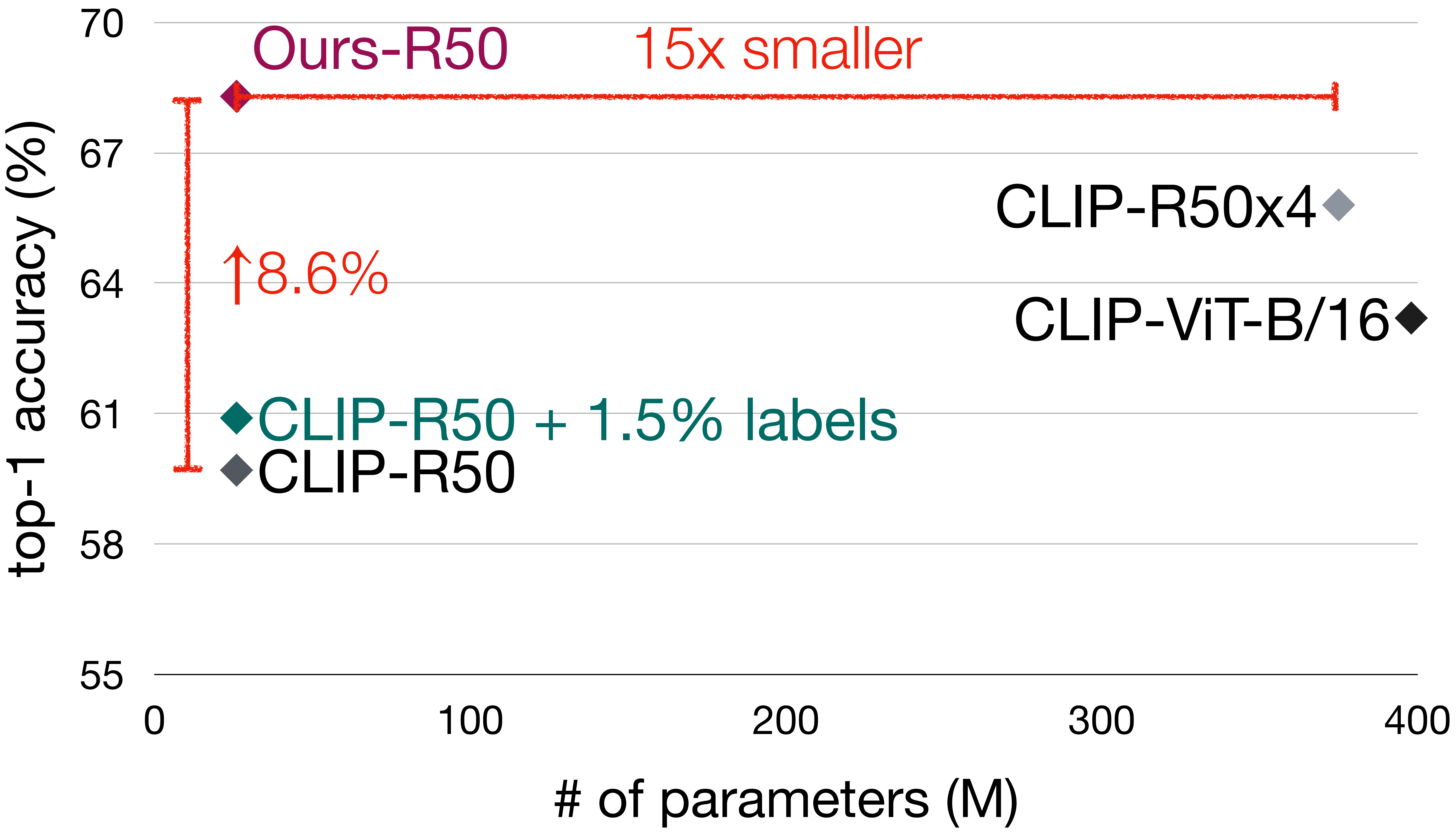
$$\mathcal{L}_{\text{AML}} = -\log \frac{e^{(z_{\hat{y}_i} - \Delta_{\hat{y}_i})}}{e^{(z_{\hat{y}_i} - \Delta_{\hat{y}_i})} + \sum_{k \neq \hat{y}_i}^{C} e^{(z_k - \Delta_k)}}$$

$$\Delta_j = \lambda \log\left(\frac{1}{\hat{p}_j}\right) \text{ for } j \in \{1, \ldots, C\}$$

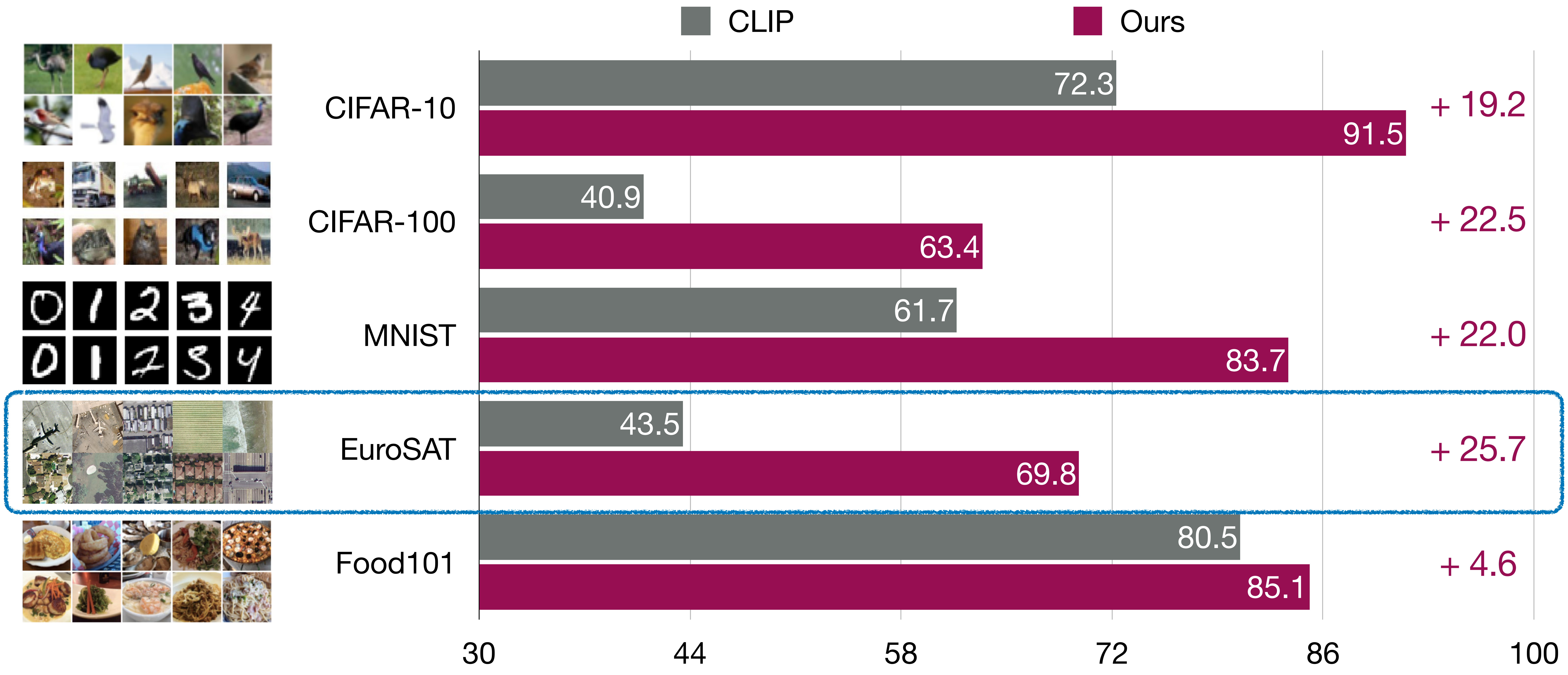


Zero-shot Learning: New SOTA on ImageNet

- outperforms CLIP fine-tuned with labels or CLIP using a 15x larger model

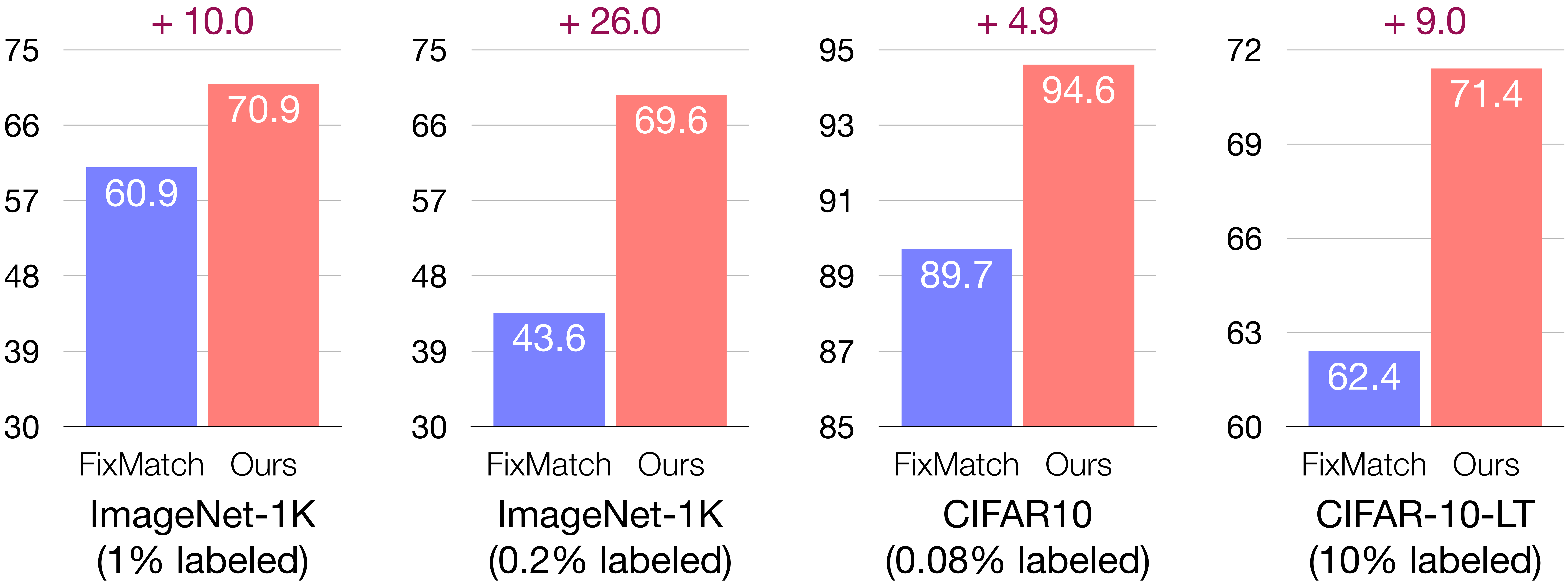


Zero-shot Learning: Stronger Robustness to Domain Shift



Semi-supervised Learning: New SOTA

- on both balanced and long-tailed data



Debiased Learning from Naturally Imbalanced Pseudo-Labels

Xudong Wang Zhirong Wu Long Lian Stella X. Yu

Code

Contributions

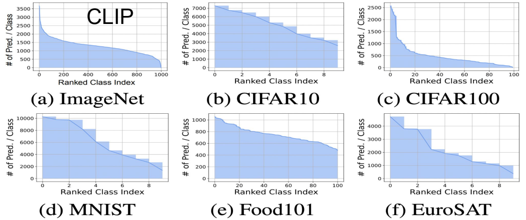
First Insight: Pseudo-labels by machines are naturally imbalanced, just like ground-truth labels by humans.

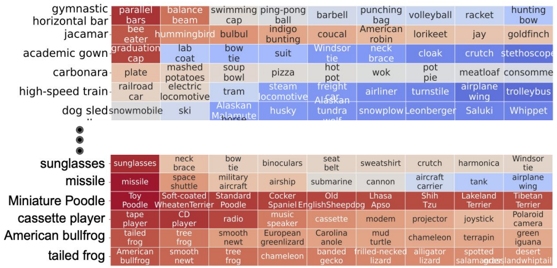
First debiased learning algorithm for pseudo-labels w/o knowing actual classification margins.

New SOTA on ImageNet: +9% on zero-shot, +26% on 0.2% semi-supervised learning; a universal add-on.

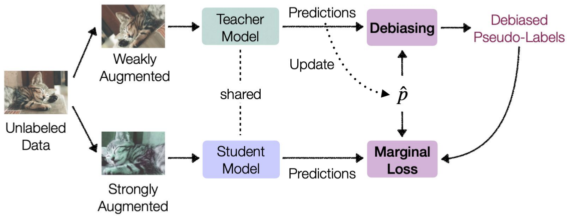
Pseudo-Labels Are Naturally Imbalanced

due to intrinsic data similarity, even when the model is trained and tested on balanced data; pseudo-labeled tail classes have stronger inter-class confusion.



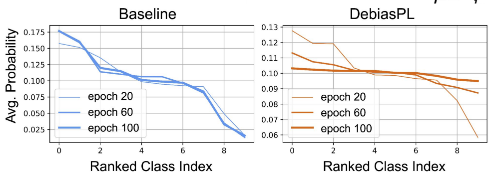


Our Debiased Pseudo-Labeling



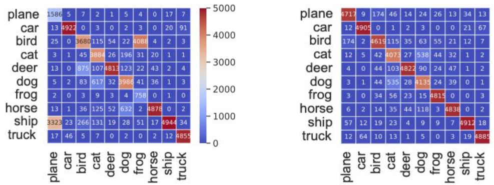
Adaptive debiasing on weak augmentations: Offset to reduce the class bias, more on pseudo-labeled **head** classes

$$\tilde{f}_i = f(\alpha(x_i)) - \lambda \log \hat{p}, \quad \hat{p} \leftarrow m\hat{p} + (1-m)\frac{1}{\mu B}\sum_{k=1}^{\mu B} p_k$$



Adaptive margin on strong augmentations: Offset to reduce inter-class confusion, more on pseudo-labeled **tail** classes

$$\mathcal{L}_{\mathrm{AML}} = -\log \frac{e^{(z_{\hat{y}_i} - \Delta_{\hat{y}_i})}}{e^{(z_{\hat{y}_i} - \Delta_{\hat{y}_i})} + \sum_{k \neq \hat{y}_i}^{C} e^{(z_k - \Delta_k)}} \qquad \Delta_j = \lambda \log\left(\frac{1}{\hat{p}_j}\right)$$



New SOTA: Large Gains, Robust, Simple & Lean