DASO: Distribution-Aware Semantics-Oriented Pseudo-label for Imbalanced Semi-Supervised Learning

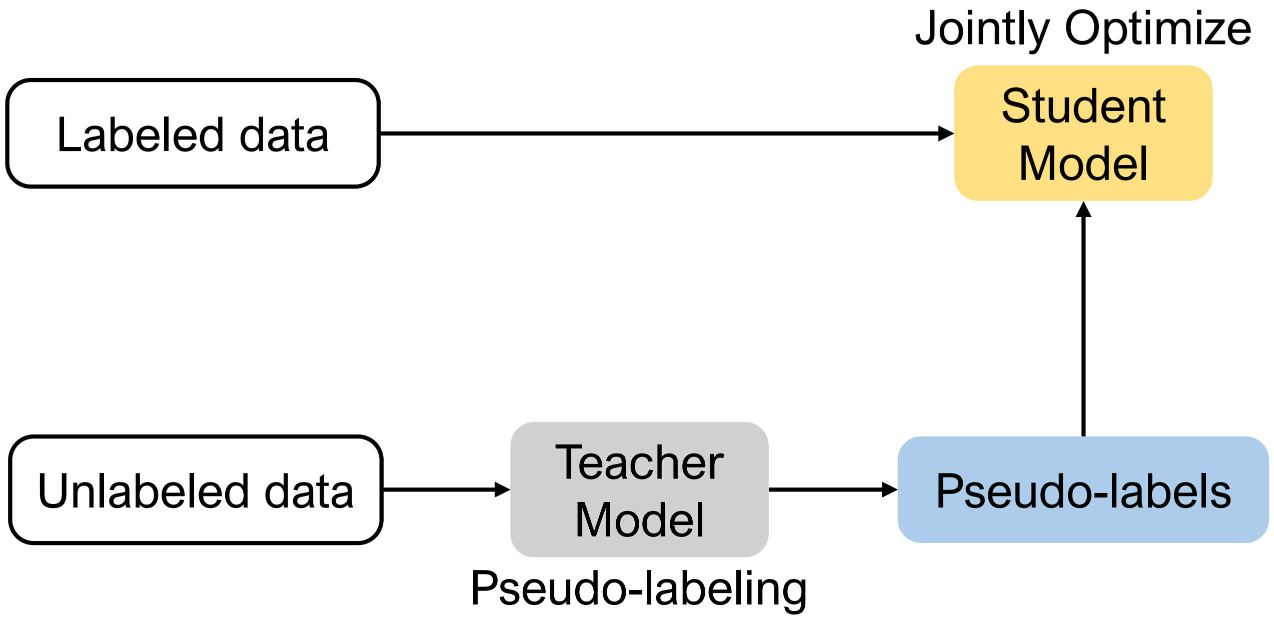
Youngtaek Oh[1] Dong-Jin Kim[2] In So Kweon[1]

[1]KAIST [2]UC Berkeley / ICSI

CVPR 2022

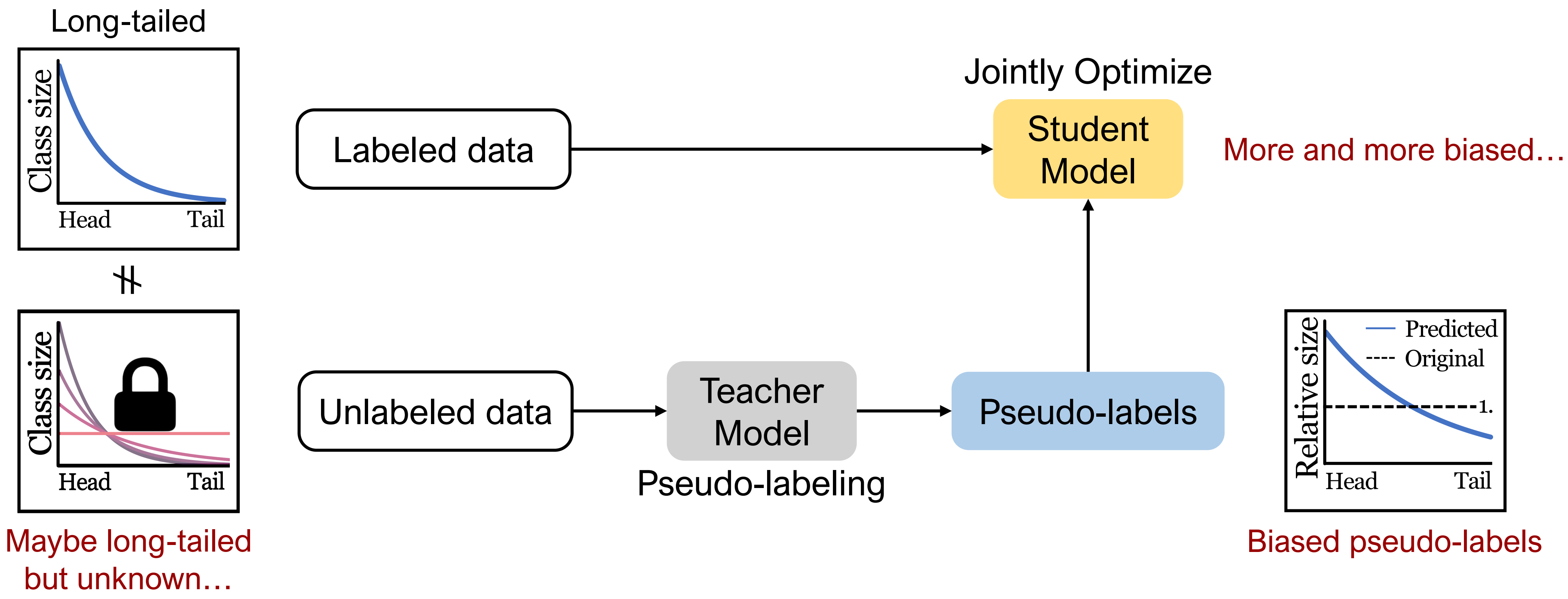
Background: Semi-supervised Learning (SSL)

- SSL utilizes both labeled data and unlabeled data for training a model.
  - ➔ For labeled data, the student model is directly optimized with labels.
  - ➔ *Pseudo-labels* based on the *teacher's predictions* are used to optimize the student with unlabeled data.



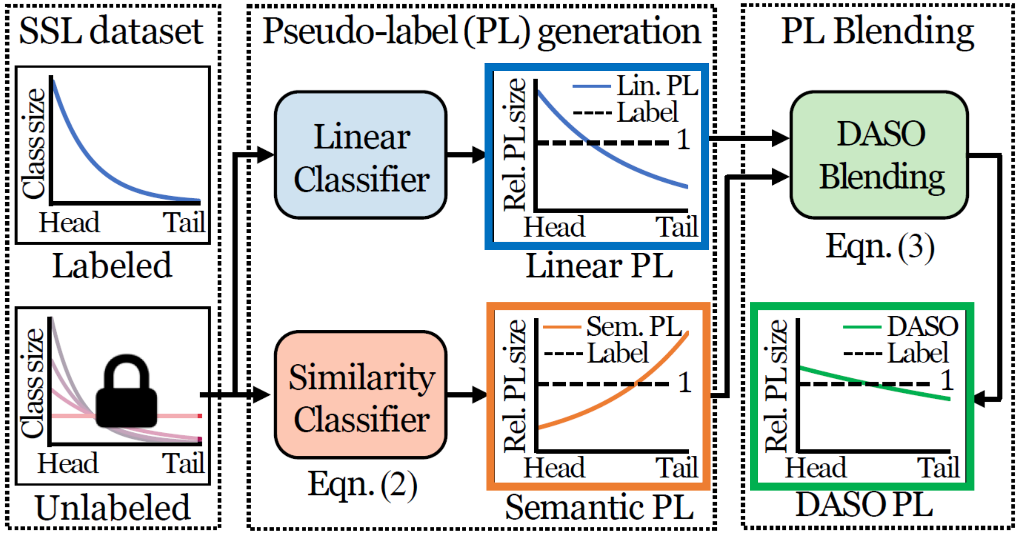
Background: Challenges in Imbalanced Semi-supervised Learning (SSL)

- Pseudo-labels from unlabeled data are biased, when learning on *long-tailed data*.
- Class distributions of unlabeled data are *unknown*, in practice.



Background: Glimpse of the proposed DASO framework

- DASO class-adaptively blends two complementarily biased pseudo-labels (PLs) to generate unbiased PL.
  - Linear PL from linear classifier (e.g., fc layer) and Semantic PL from similarity-classifier (e.g., [1-2]).



Biased towards head classes

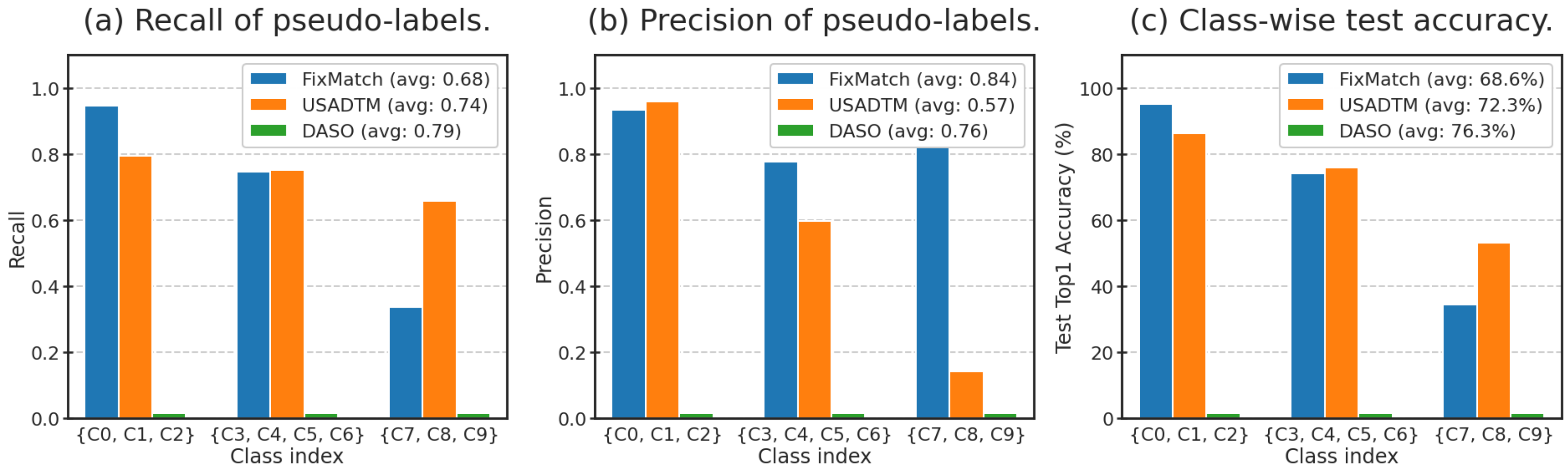
Biased towards tail classes

Less biased!

[1] Prototypical Networks for Few-shot Learning, NIPS'17.
[2] Unsupervised Semantic Aggregation and Deformable Template Matching for Semi-Supervised Learning, NeurIPS'20.

Motivation: A closer look at the bias of 'Linear' and 'Semantic' pseudo-label

- Properties on linear PL from FixMatch [1] and semantic PL from USADTM [2]:



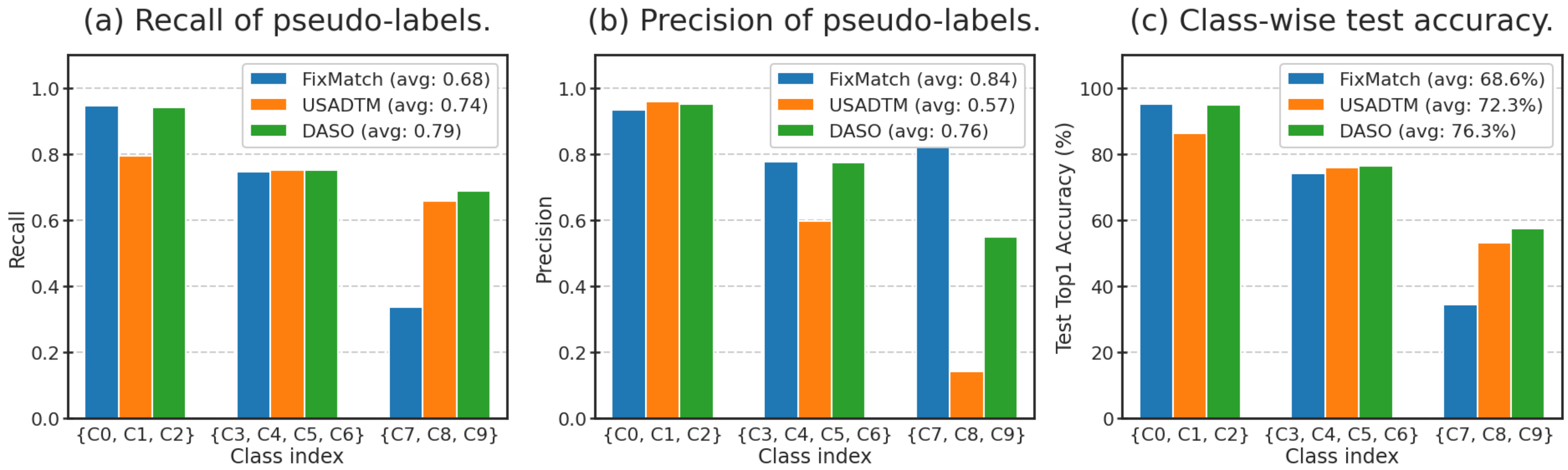
Analysis on (a) recall and (b) precision of pseudo-labels (PLs), and (c) test accuracy on CIFAR10-LT

- Linear PL from FixMatch [1]: biased towards head classes.
- Semantic PL from USADTM [2]: reversely biased towards tail classes.
- Complementary each other ➡ useful cue for reducing the overall bias!

[1] FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence, NeurIPS'20.
[2] Unsupervised Semantic Aggregation and Deformable Template Matching for Semi-Supervised Learning, NeurIPS'20.

Motivation: A closer look at the bias of 'Linear' and 'Semantic' pseudo-label

- DASO → Blending more semantic PL on the *minorities mis-predicted to the head classes*!



Analysis on (a) recall and (b) precision of pseudo-labels (PLs), and (c) test accuracy on CIFAR10-LT

- DASO preserves the recall values on the majority classes.
- DASO increases the recall values on the minority classes, while maintaining the precisions.

[1] FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence, NeurIPS'20.

[2] Unsupervised Semantic Aggregation and Deformable Template Matching for Semi-Supervised Learning, NeurIPS'20.

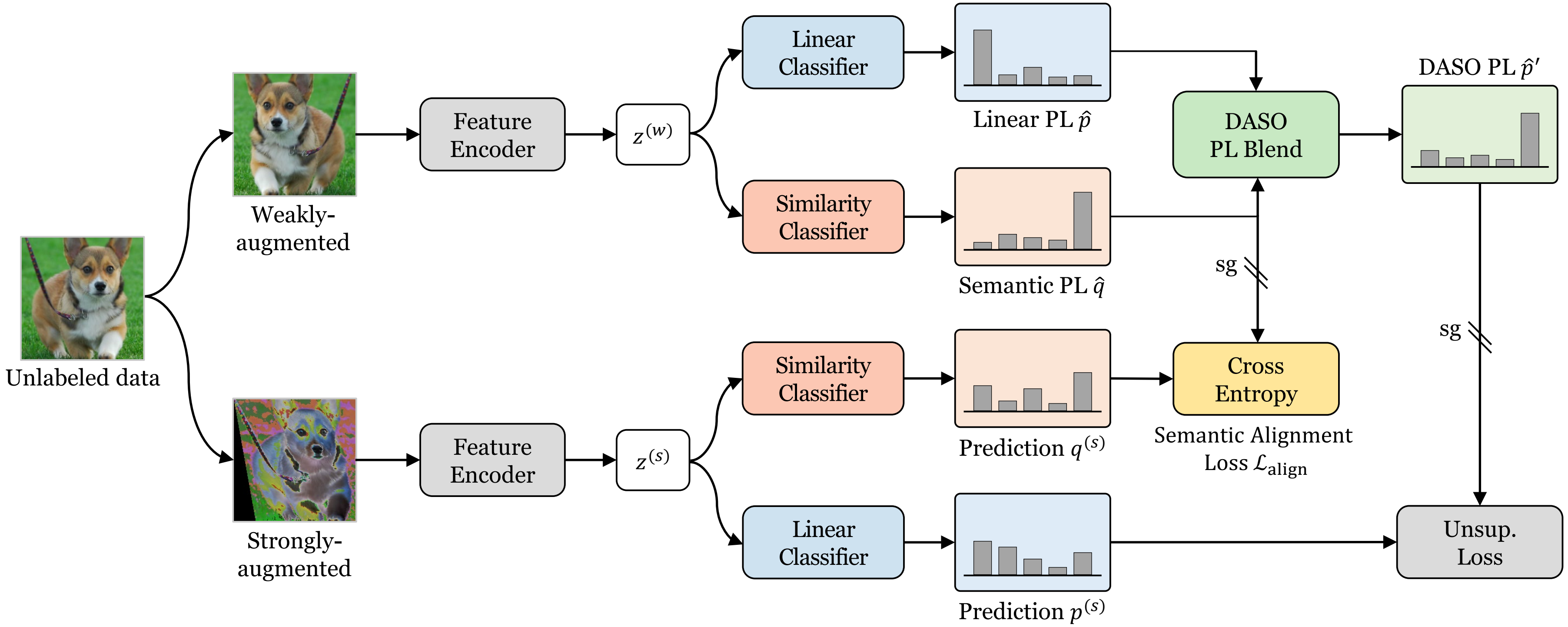
Method: Overall DASO framework

- Full DASO framework introduces two core components, *building upon existing SSL learner*.
  - **Blending pseudo-labels**: *adaptively blends* semantic PL into linear PL along with different classes for debiasing.

$$\hat{p}' = (1 - v) \cdot \hat{p} + v \cdot \hat{q},$$

  - **Semantic Alignment Loss**: constructs *balanced feature representations* for unbiased classifier predictions.

$$\mathcal{L}_{\text{align}} = \mathcal{H}\left(\hat{q}, q^{(s)}\right),$$



Experiments: Experimental setups

- Benchmark datasets
  - (Small-scale) Synthetically long-tailed CIFAR-10/100 and STL-10.
    - Denote γ_l and γ_u as the imbalance ratio / N_1 and M_1 as the head-class size for the labeled data and unlabeled data, respectively.
    - *Test various imbalance scenario by changing those parameters above, including $\gamma_l \neq \gamma_u$ as well as $\gamma_l = \gamma_u$ scenarios.*
  - (Large-scale) Semi-Aves benchmark [1]
    - Both labeled data ($\mathcal{X}$) and unlabeled data ($\mathcal{U}$) have *long-tailed distributions, but mismatch between them (e.g., $\gamma_l \neq \gamma_u$)*.
    - The whole unlabeled data ($\mathcal{U}$) include *large portions of open-set class examples ($\mathcal{U}_{out}$)*; $\mathcal{U} = \mathcal{U}_{\text{in}} + \mathcal{U}_{\text{out}}$.
- Primary baseline methods: DARP [2] and CReST [3].
  - Both baseline methods require *true or estimated class distributions of unlabeled data*.
  - *DASO does not rely on such ideal assumption.*

[1] The Semi-Supervised iNaturalist-Aves Challenge at FGVC7 Workshop, arXiv preprint:2103.06937.
[2] Distribution Aligning Refinery of Pseudo-label for Imbalanced Semi-supervised Learning, NeurIPS 2020.
[3] CReST: A Class-Rebalancing Self-Training Framework for Imbalanced Semi-Supervised Learning, CVPR 2021.

Experiments: Equal class imbalance

- Class distributions of labeled data and unlabeled data are the same ($\gamma_l = \gamma_u$)

Algorithm	CIFAR10-LT $\gamma=\gamma_l=\gamma_u=100$ $N_1=500$ $M_1=4000$	CIFAR10-LT $\gamma=\gamma_l=\gamma_u=100$ $N_1=1500$ $M_1=3000$	CIFAR10-LT $\gamma=\gamma_l=\gamma_u=150$ $N_1=500$ $M_1=4000$	CIFAR10-LT $\gamma=\gamma_l=\gamma_u=150$ $N_1=1500$ $M_1=3000$	CIFAR100-LT $\gamma=\gamma_l=\gamma_u=10$ $N_1=50$ $M_1=400$	CIFAR100-LT $\gamma=\gamma_l=\gamma_u=10$ $N_1=150$ $M_1=300$	CIFAR100-LT $\gamma=\gamma_l=\gamma_u=20$ $N_1=50$ $M_1=400$	CIFAR100-LT $\gamma=\gamma_l=\gamma_u=20$ $N_1=150$ $M_1=300$
Supervised	47.3 ±0.95	61.9 ±0.41	44.2 ±0.33	58.2 ±0.29	29.6 ±0.57	46.9 ±0.22	25.1 ±1.14	41.2 ±0.15
w/ LA [36]	53.3 ±0.44	70.6 ±0.21	49.5 ±0.40	67.1 ±0.78	30.2 ±0.44	48.7 ±0.89	26.5 ±1.31	44.1 ±0.42
FixMatch [48]	67.8 ±1.13	77.5 ±1.32	62.9 ±0.36	72.4 ±1.03	45.2 ±0.55	56.5 ±0.06	40.0 ±0.96	50.7 ±0.25
w/ DARP [26]	74.5 ±0.78	77.8 ±0.63	67.2 ±0.32	73.6 ±0.73	49.4 ±0.20	58.1 ±0.44	43.4 ±0.87	52.2 ±0.66
w/ CReST+ [54]	**76.3** ±0.86	78.1 ±0.42	67.5 ±0.45	73.7 ±0.34	44.5 ±0.94	57.4 ±0.18	40.1 ±1.28	52.1 ±0.21
w/ DASO (Ours)	76.0 ±0.37	**79.1** ±0.75	**70.1** ±1.81	**75.1** ±0.77	**49.8** ±0.24	**59.2** ±0.35	**43.6** ±0.09	**52.9** ±0.42
FixMatch + LA [36]	75.3 ±2.45	82.0 ±0.36	67.0 ±2.49	78.0 ±0.91	47.3 ±0.42	58.6 ±0.36	41.4 ±0.93	53.4 ±0.32
w/ DARP [26]	76.6 ±0.92	80.8 ±0.62	68.2 ±0.94	76.7 ±1.13	50.5 ±0.78	59.9 ±0.32	**44.4** ±0.65	53.8 ±0.43
w/ CReST+ [54]	76.7 ±1.13	81.1 ±0.57	**70.9** ±1.18	77.9 ±0.71	44.0 ±0.21	57.1 ±0.55	40.6 ±0.55	52.3 ±0.20
w/ DASO (Ours)	**77.9** ±0.88	**82.5** ±0.08	70.1 ±1.68	**79.0** ±2.23	**50.7** ±0.51	**60.6** ±0.71	44.1 ±0.61	**55.1** ±0.72
FixMatch + ABC [32]	78.9 ±0.82	**83.8** ±0.36	66.5 ±0.78	80.1 ±0.45	47.5 ±0.18	59.1 ±0 .21	41.6 ±0.83	53.7 ±0.55
w/ DASO (Ours)	**80.1** ±1.16	83.4 ±0.31	**70.6** ±0.80	**80.4** ±0.56	**50.2** ±0.62	**60.0** ±0.32	**44.5** ±0.25	**55.3** ±0.53

- DASO shows better performance gains compared to baseline DARP and CReST+.
- DASO can improve various supervised / semi-supervised frameworks under imbalance.

[1] Long-tail learning via logit adjustment, ICLR 2021.
[2] ABC: Auxiliary Balanced Classifier for Class-imbalanced Semi-supervised Learning, NeurIPS 2021.

Experiments: Various class distributions for unlabeled data

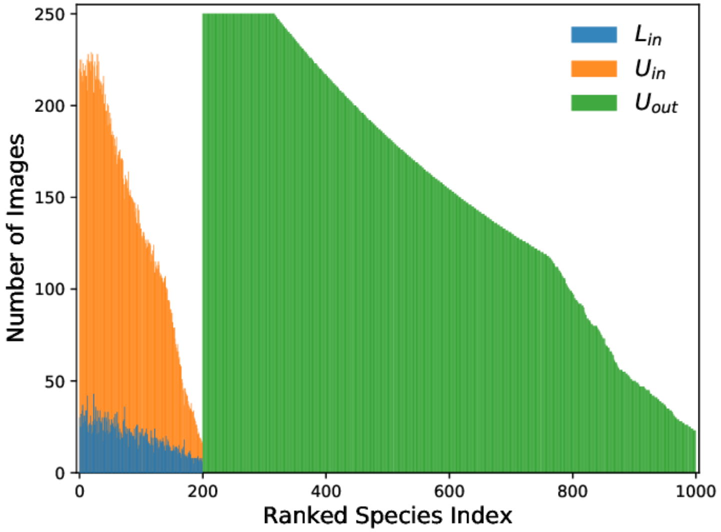
- Various class distributions of unlabeled data, different from that of labeled data ($\gamma_l \neq \gamma_u$)

Algorithm	CIFAR10-LT ($\gamma_l \neq \gamma_u$) $\gamma_u = 1$ (uniform)		CIFAR10-LT ($\gamma_l \neq \gamma_u$) $\gamma_u = 1/100$ (reversed)		STL10-LT ($\gamma_u = N/A$) $\gamma_l = 10$		STL10-LT ($\gamma_u = N/A$) $\gamma_l = 20$	
	$N_1 = 500$ $M_1 = 4000$	$N_1 = 1500$ $M_1 = 3000$	$N_1 = 500$ $M_1 = 4000$	$N_1 = 1500$ $M_1 = 3000$	$N_1 = 150$ $M = 100k$	$N_1 = 450$ $M = 100k$	$N_1 = 150$ $M = 100k$	$N_1 = 450$ $M = 100k$
FixMatch [48]	73.0 ±3.81	81.5 ±1.15	62.5 ±0.94	71.8 ±1.70	56.1 ±2.32	72.4 ±0.71	47.6 ±4.87	64.0 ±2.27
w/ DARP [26]	79.6 ±1.80	85.8 ±0.28	70.1 ±0.22	80.0 ±0.93	66.9 ±1.66	75.6 ±0.45	59.9 ±2.17	72.3 ±0.60
w/ CReST [54]	83.2 ±1.67	87.1 ±0.28	70.7 ±2.02	**80.8** ±0.39	61.7 ±2.51	71.6 ±1.17	57.1 ±3.67	68.6 ±0.88
w/ CReST+ [54]	82.2 ±1.53	86.4 ±0.42	62.9 ±1.39	72.9 ±2.00	61.2 ±1.27	71.5 ±0.96	56.0 ±3.19	68.5 ±1.88
w/ DASO (Ours)	**86.6** ±0.84	**88.8** ±0.59	**71.0** ±0.95	80.3 ±0.65	**70.0** ±1.19	**78.4** ±0.80	**65.7** ±1.78	**75.3** ±0.44

- DASO is far robust to the changes in imbalance of unlabeled data.
  - ➔ Limited gains from DARP and CReST due to improper assumption ($\gamma_l = \gamma_u$) under $\gamma_l \neq \gamma_u$ scenarios.

Experiments: Realistic scenarios

- Semi-Aves benchmark with *long-tailed distribution* and large *open-set class unlabeled data*



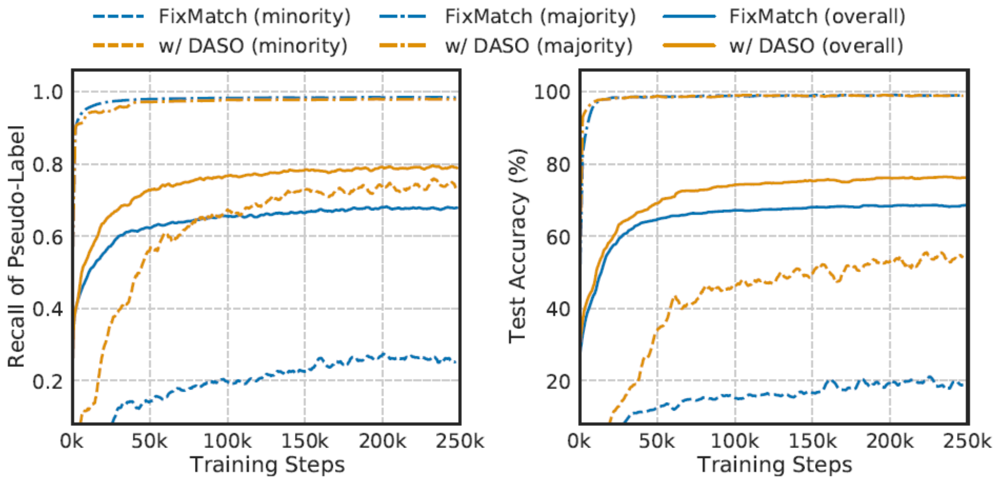
Class distributions of each data split in Semi-Aves.
Image reference: [1]

Benchmark	Semi-Aves			
	$\mathcal{U} = \mathcal{U}_{\text{in}}$		$\mathcal{U} = \mathcal{U}_{\text{in}} + \mathcal{U}_{\text{out}}$	
Method	Last Top1	Med20 Top1	Last Top1	Med20 Top1
Supervised	41.7 ±0.32	41.7 ±0.32	41.7 ±0.32	41.7 ±0.32
FixMatch [48]	53.8 ±0.17	53.8 ±0.13	45.7 ±0.89	46.1 ±0.50
w/ DARP [26]	52.3 ±0.48	52.1 ±0.48	46.3 ±0.70	46.4 ±0.61
w/ CReST [54]	52.1 ±0.36	52.2 ±0.27	43.6 ±0.69	43.6 ±0.68
w/ CReST+ [54]	53.9 ±0.38	53.8 ±0.38	45.1 ±1.09	45.2 ±1.00
w/ DASO (Ours)	**54.5** ±0.08	**54.6** ±0.12	**47.9** ±0.41	**47.9** ±0.38

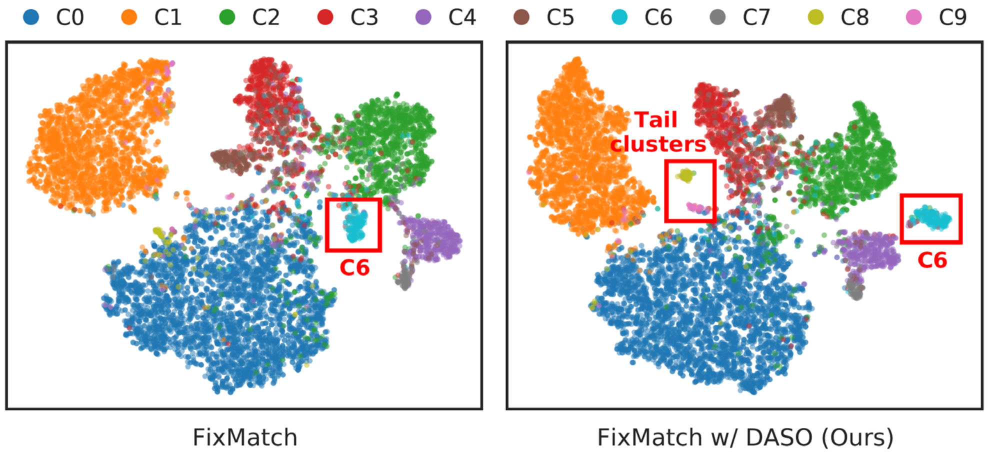
- DASO outperforms previous methods in both $\mathcal{U} = \mathcal{U}_{\text{in}}$ and $\mathcal{U} = \mathcal{U}_{\text{in}} + \mathcal{U}_{\text{out}}$ cases.
  - ➔ large gains even when open-set examples are dominant in unlabeled data.

[1] The Semi-Supervised iNaturalist-Aves Challenge at FGVC7 Workshop, arXiv preprint:2103.06937.

Analysis: Understanding the effects of DASO on debiasing pseudo-labels



Train curves for the recall of pseudo-labels and test accuracy.



t-SNE visualizations of feature representation from unlabeled data.

- Unbiased pseudo-label improves test accuracy.
  - ➔ improves performance on the minority classes, while preserving those from the majority classes.
- Tail-class clusters are better identified.
  - ➔ helps construct tail-class clusters, further reducing the biased predictions from the classifier.

Summary

Poster Session 2.2
22 June, 2:30PM-5:00PM
Poster ID: 180b

Webpage: ytaek-oh.github.io/daso
Code: github.com/ytaek-oh/daso
Contact: youngtaek.oh@kaist.ac.kr

TL;DR: DASO for debiasing pseudo-labels for imbalanced semi-supervised learning.

- We propose a new pseudo-labeling framework for SSL under *imbalanced data*, termed **Distribution-Aware Semantics-Oriented (DASO) Pseudo-label**, with two core components:
  - Class-adaptive blending of two complementary pseudo-labels.
  - Semantic alignment loss to construct balanced feature representations.
- We show DASO is versatile and applicable to various practical scenarios.
  - Easily built upon existing SSL algorithms and coupled with re-balancing frameworks.
  - Effective under various class distributions of unlabeled data.
- More experimental results and visualizations can be found in the paper!