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

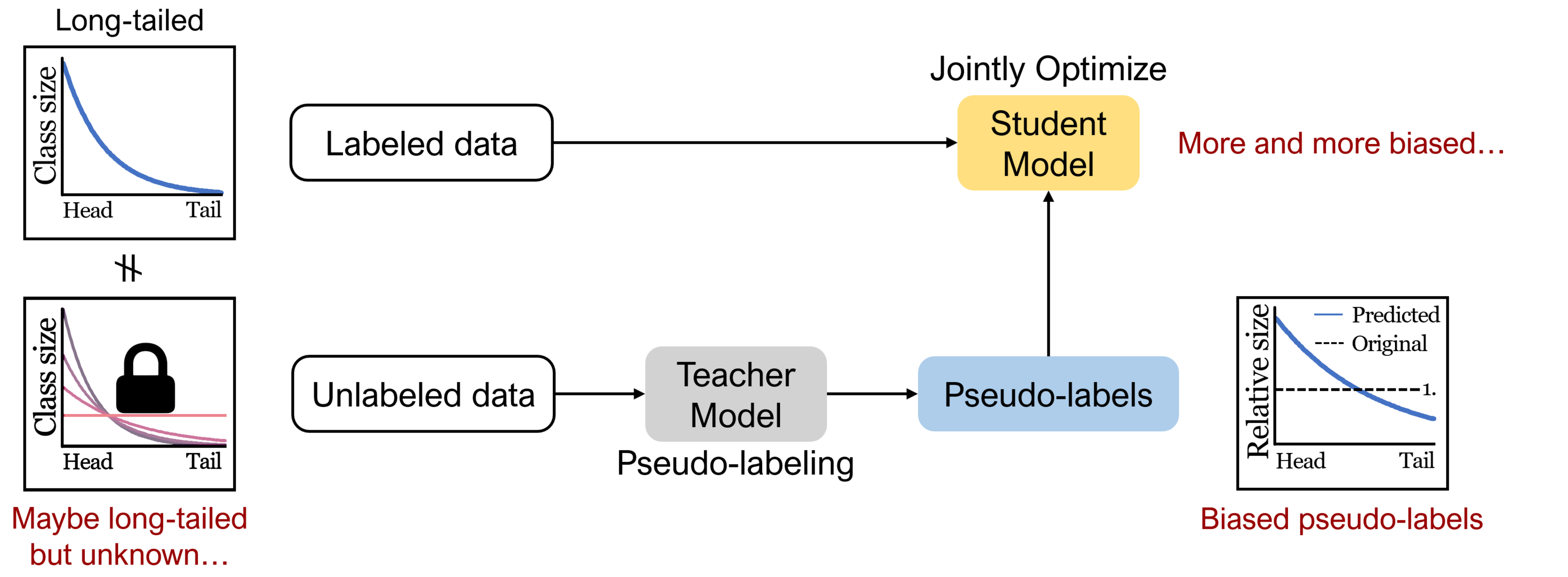
Youngtaek Oh[1], Dong-Jin Kim[2], In So Kweon[1]
[1]KAIST. [2]UC Berkeley / ICSI.

Introduction

Challenges in Imbalanced Semi-supervised Learning (SSL)

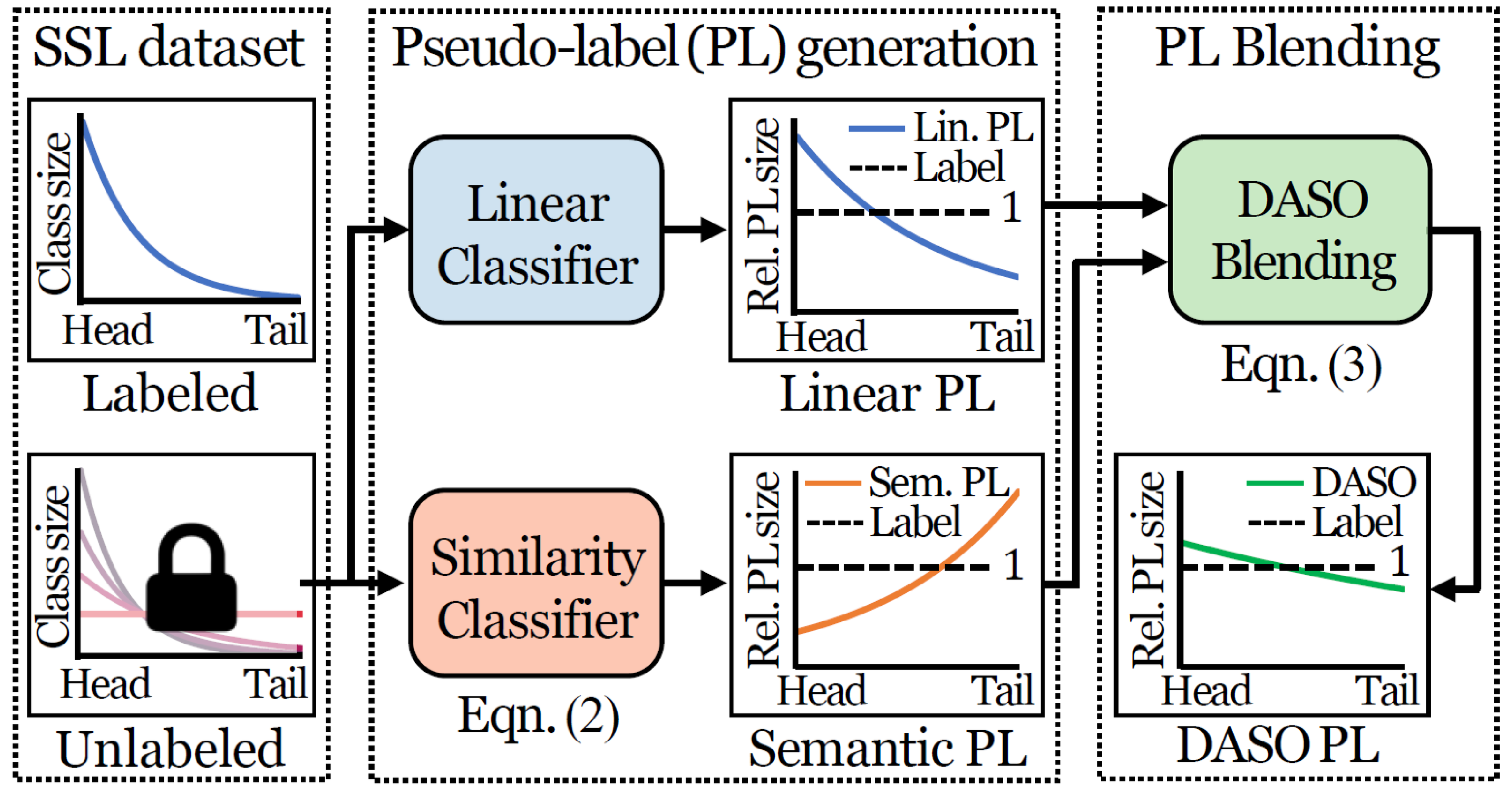
- ❑ Biased pseudo-labels (PLs) when learning with long-tailed data.
- ❑ Unknown class distribution of unlabeled data, in practice.
- ➔ Goal: (1) *unbiased pseudo-labels* (2) *without relying on any class prior*.

Practical Imbalanced SSL scenarios



Glimpse of the DASO framework

DASO class-adaptively blends two complementarily biased PLs from different classifiers to generate unbiased PL.

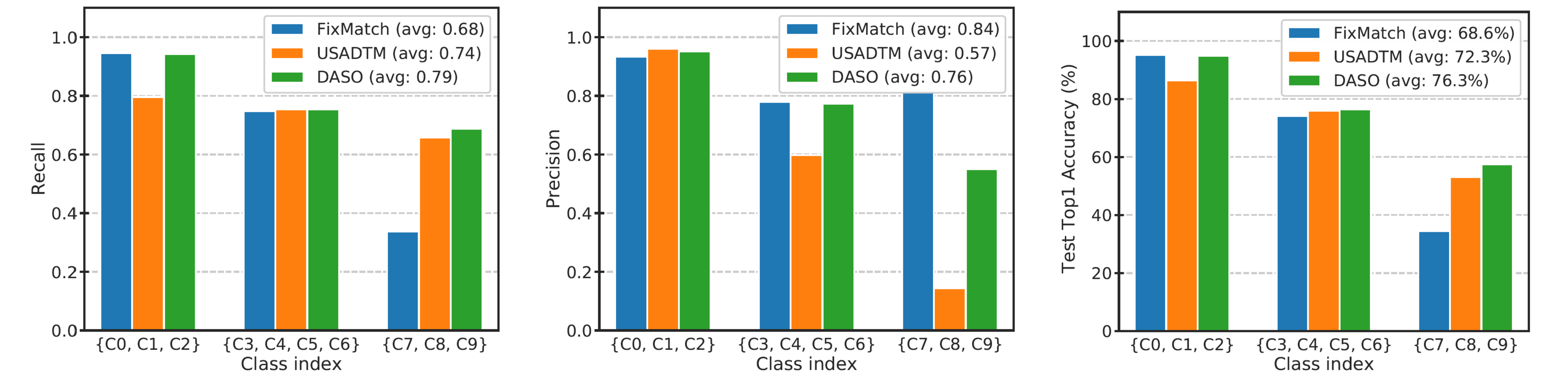


Proposed Method

Motivation: Trade-offs between linear and semantic pseudo-label

- ❑ Properties on linear PL w/ FixMatch and semantic PL w/ USADTM.
- ☹ FixMatch: Biased towards majority classes
- ☹ USADTM: Biased towards minority classes
- ☺ DASO: More semantic PL to the minorities mis-predicted to the head.

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



(a) Recall of pseudo-labels. (b) Precision of pseudo-labels. (c) Class-wise test accuracy.

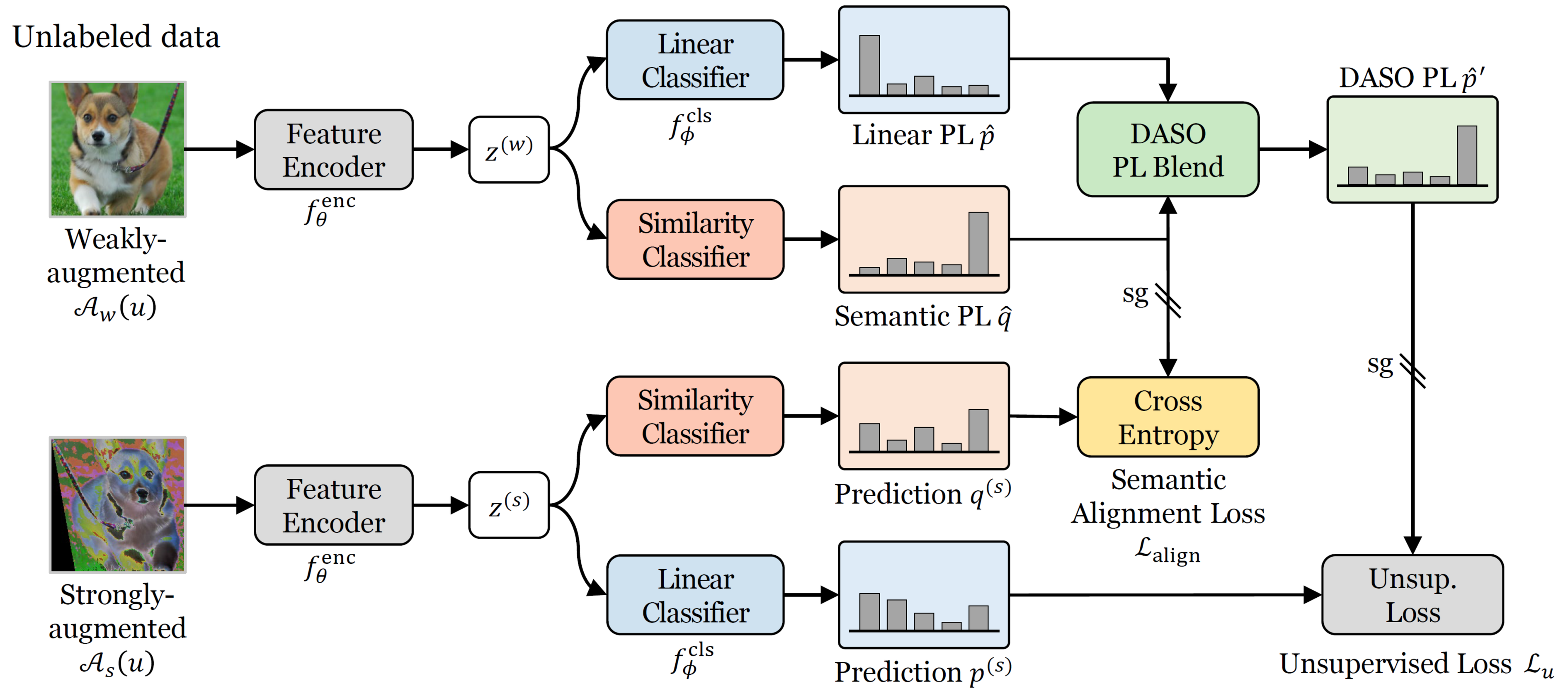
Distribution-Aware Semantics-Oriented (DASO) Pseudo-label Framework

- ❑ **Blending PLs**: class-adaptive blending of semantic PL into linear PL.

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

- ❑ **Semantic alignment loss**: balanced feature space for unbiased predictions.

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



Experiments

❑ Comparisons under identical imbalance: $\gamma_l = \gamma_u$.

	CIFAR10-LT				CIFAR100-LT			
	$\gamma = \gamma_l = \gamma_u = 100$		$\gamma = \gamma_l = \gamma_u = 150$		$\gamma = \gamma_l = \gamma_u = 10$		$\gamma = \gamma_l = \gamma_u = 20$	
	$N_1 = 500$	$N_1 = 1500$	$N_1 = 500$	$N_1 = 1500$	$N_1 = 50$	$N_1 = 150$	$N_1 = 50$	$N_1 = 150$
FixMatch [1]	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 [2]	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+ [3]	**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	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

❑ Comparisons under diverse imbalances on unlabeled data: $\gamma_l \neq \gamma_u$.

	CIFAR10-LT ($\gamma_l \neq \gamma_u$)				STL10-LT (γ_u: *unknown*)			
	$\gamma_u = 1$ (*uniform*)		$\gamma_u = 1/100$ (*reversed*)		$\gamma_l = 10$		$\gamma_l = 20$	
	$N_1 = 500$	$N_1 = 1500$	$N_1 = 500$	$N_1 = 1500$	$N_1 = 150$	$N_1 = 450$	$N_1 = 150$	$N_1 = 450$
FixMatch [1]	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 [2]	82.5±0.75	84.6±0.34	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 [3]	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+ [3]	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	**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

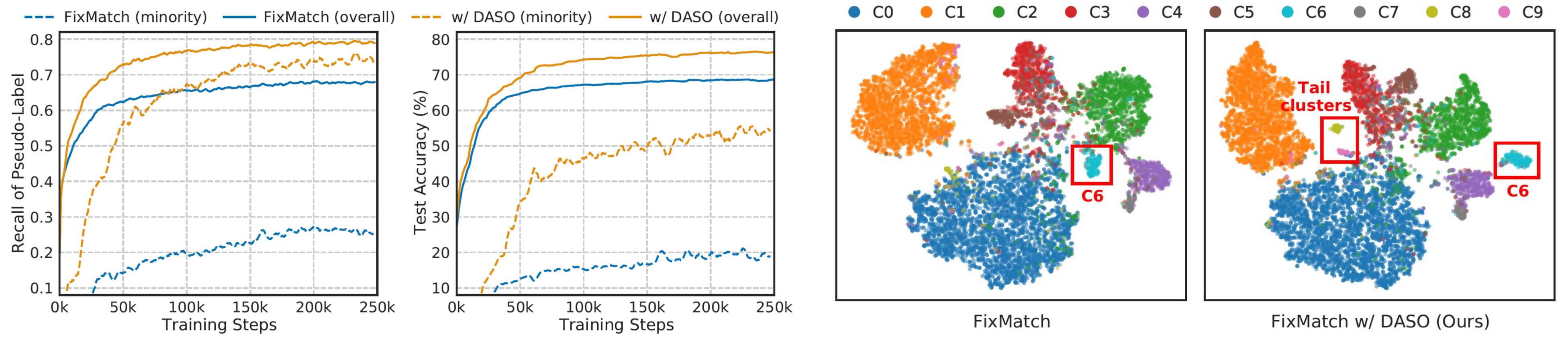
❑ Large-scale experiments with realistic scenarios.

Semi-Aves benchmark

- ❑ Long-tailed distributions $\gamma_l \neq \gamma_u$
- ❑ Large open-set class examples $\mathcal{U}_{\text{out}}$
- ❑ Total unlabeled data $\mathcal{U} = \mathcal{U}_{\text{in}} + \mathcal{U}_{\text{out}}$

Method	Semi-Aves			
	$\mathcal{U} = \mathcal{U}_{in}$		$\mathcal{U} = \mathcal{U}_{in} + \mathcal{U}_{out}$	
	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 [1]	53.8±0.17	53.8±0.13	45.7±0.89	46.1±0.50
w/ DARP [2]	52.3±0.48	52.1±0.48	46.3±0.70	46.4±0.61
w/ CReST [3]	52.1±0.36	52.2±0.27	43.6±0.69	43.6±0.68
w/ CReST+ [3]	53.9±0.38	53.8±0.38	45.1±1.09	45.2±1.00
w/ DASO	**54.5**±0.08	**54.6**±0.12	**47.9**±0.41	**47.9**±0.38

❑ Understanding the effects of DASO on debiasing pseudo-labels.



References

[1] K. Sohn et al., FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence, NIPS 2020.
[2] J. Kim et al., DARP: Distribution Aligning Refinery of Pseudo-label for Imbalanced Semi-supervised Learning, NIPS 2020.
[3] C. Wei et al., CReST: A Class-Rebalancing Self-Training Framework for Imbalanced Semi-Supervised Learning, CVPR 2021.