Flexible Distribution Alignment: Towards Long-tailed Semi-supervised Learning with Proper Calibration

LINKÖPING UNIVERSITY

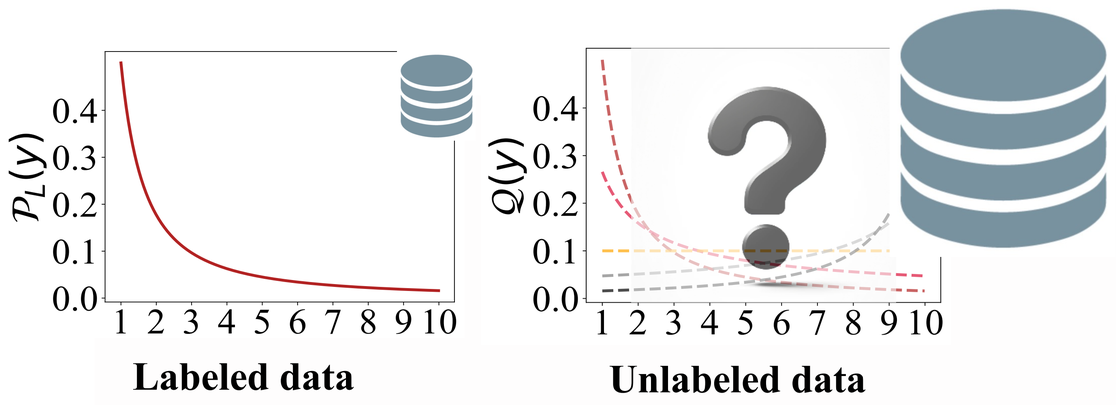
Emanuel Sánchez Aimar, Nathaniel Helgesen, Yonghao Xu, Marco Kuhlmann, Michael Felsberg

ECCV

Paper Code

Motivation

Long-tailed Semi-supervised Learning (LTSSL)



Classification under **class imbalance** and **limited labeled data**

Unlabeled data distribution is **unknown** and can be **different** from labeled data

Limitations of Previous Approaches

Consistency Regularization + Confidence thresholding + Pseudo-labeling

Ignores tail/underconfident samples (**unfair**)

Distribution Alignment

Aligns pseudo-labels with fixed/predefined priors (**unrealistic/wrong assumptions**)

Inefficient use of unlabeled data
Biased classifier during inference
Poorly-calibrated probabilities

Our Contributions

Flexible Distribution Alignment

Adaptive! (no need of predefined priors)

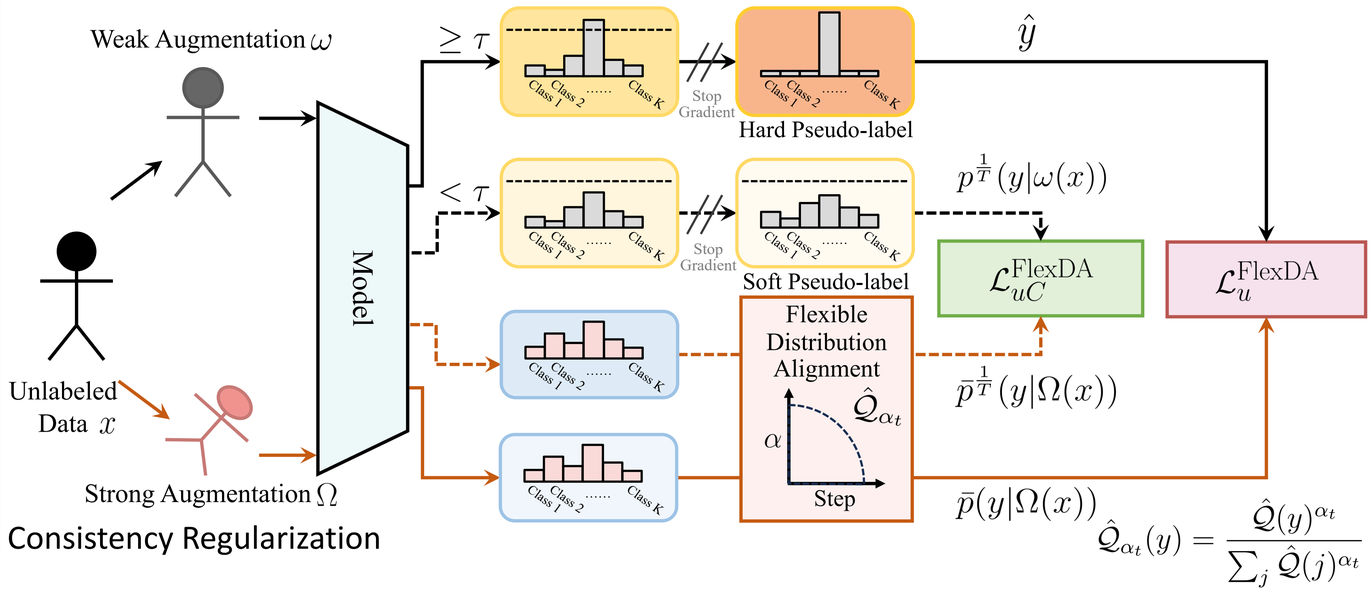
Full data utilization!

Calibrated model!

Align and Distill Everything All At Once

We propose ***ADELLO***, a simple flexible method for LTSSL:

- **Aligns** the model with the **correct prior**, **dynamically estimated** from pseudo-labels
- **Progressively debiases** the model during training
- **Leverages all data samples** with hard and soft pseudo-labels



Bias-adjusted Losses

Supervised:
$$\mathcal{L}_s^{\text{FlexDA}} = \frac{1}{B}\sum_{b=1}^{B} \mathcal{H}(y_b, \sigma(f(\omega(x_b)) + \log \frac{\mathcal{P}_L}{\hat{\mathcal{Q}}_{\alpha_t}}))$$

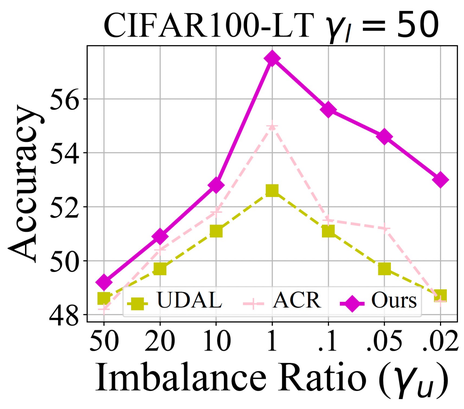
Consistency:
$$\mathcal{L}_u^{\text{FlexDA}} = \frac{1}{\mu B}\sum_{b=1}^{\mu B} \mathcal{M}(u_b) \cdot \mathcal{H}(\hat{y}_b, \sigma(f(\Omega(u_b)) + \log \frac{\hat{Q}}{\hat{\mathcal{Q}}_{\alpha_t}}))$$

Complementary Consistency:
$$\mathcal{L}_{uC}^{\text{FlexDA}} = \frac{1}{\mu B}\sum_{b=1}^{\mu B} \mathcal{M}^C(u_b) \cdot \mathcal{H}(p^{\frac{1}{T}}(y|\omega(u_b)), p^{\frac{1}{T}}(y|\Omega(u_b)))$$
where $\bar{p}^{\frac{1}{T}}(y|\Omega(u_b)) = \sigma(\frac{1}{T}(f(\Omega(u_b)) + \log \frac{\hat{Q}}{\hat{\mathcal{Q}}_{\alpha_t}}))$

Imbalance-aware temperature (after warmup):
$$T = \exp(\mathrm{KL}(\mathcal{P}_{\text{bal}} \| \hat{Q}))$$

Theoretically-sound: approximation of **bayes-optimal classifier**!

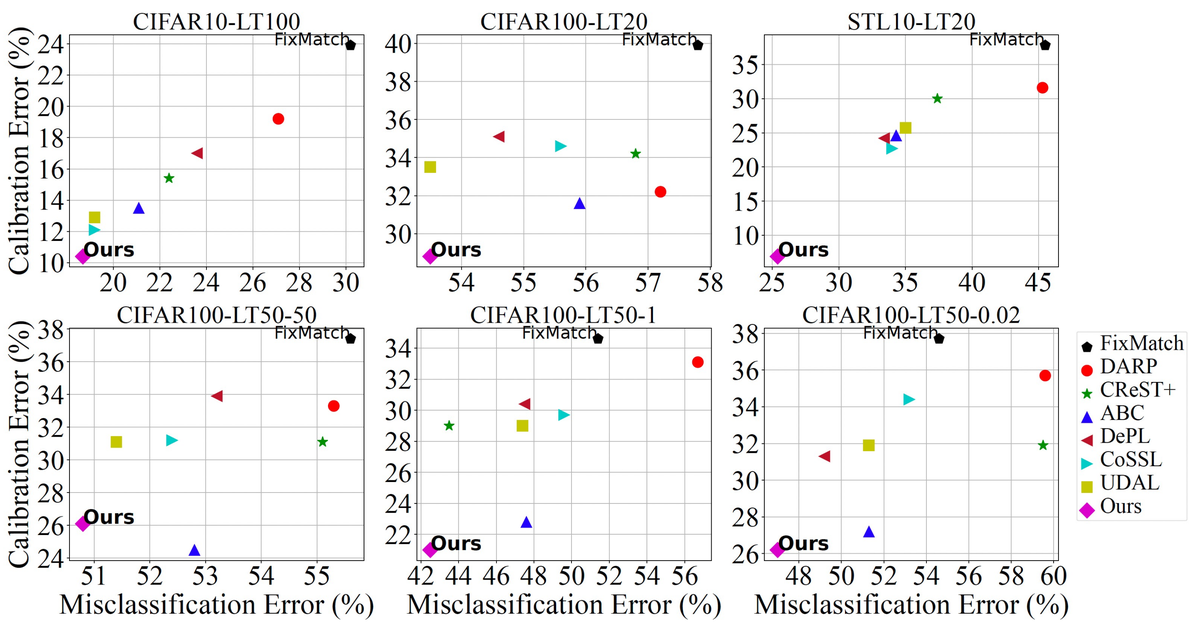
Results



Robust performance even under an increasing degree of distribution mismatch

ImageNet127 IR = 286	Balanced accuracy Resolution	
Method	32×32	64×64
FixMatch [58]†	29.7	42.3
+DARP [29]†	30.5	42.5
+DARP +cRT [29]†	39.7	51.0
+CReST+ [68]†	32.5	44.7
+CReST+ +LA [68]†	40.9	55.9
+CoSSL [16]†	43.7	53.8
+UDAL ($\alpha_{\min}$=0.55) [37]	40.2	49.4
+UDAL ($\alpha_{\min}$=0.1) [37]	44.1	52.3
+ADELLO (ours)	**47.5**	**58.0**

SOTA accuracy for challenging large-scale datasets under consistent case



ADELLO achieves strong confidence calibration across all datasets

Better-calibrated models tend to improve LTSSL performance!