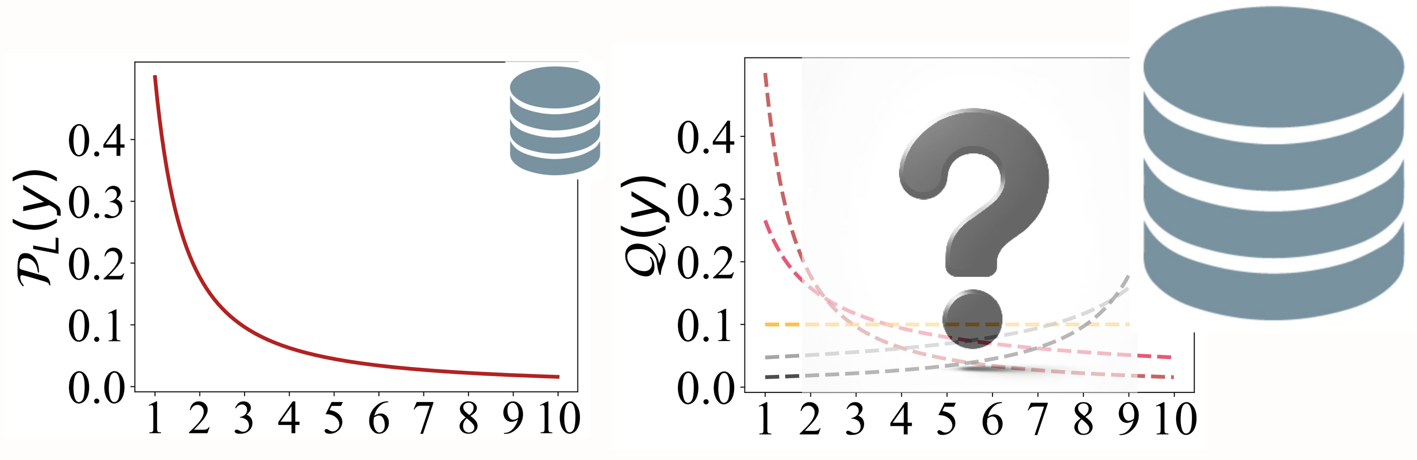
Flexible Distribution Alignment: Towards Long-tailed Semi-supervised Learning with Proper Calibration

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

Linköping University, Sweden

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 traditional 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, in a nutshell

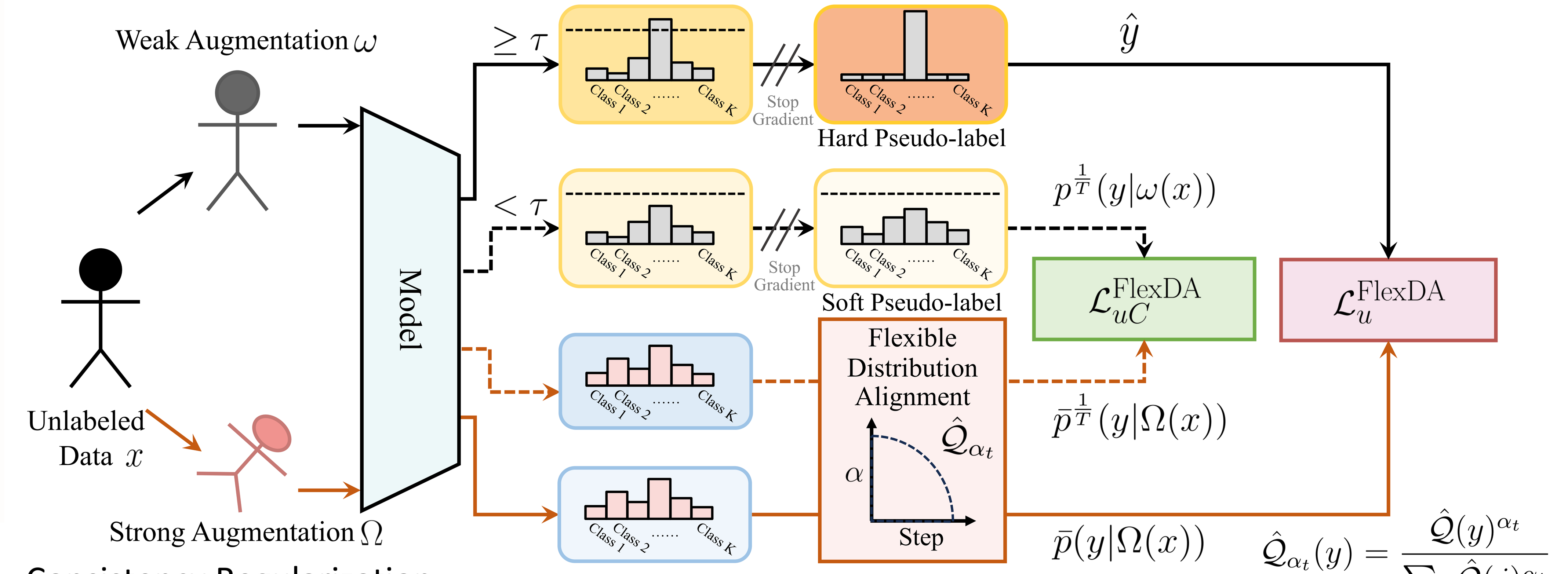
Flexible Distribution Alignment

- Adaptive! (no need of predefined priors)
- Full data utilization!
- Calibrated model!

Method: Align and Distill Everything All At Once (ADELLO)

We propose 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



Implementation: Bias-adjusted Losses

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

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

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

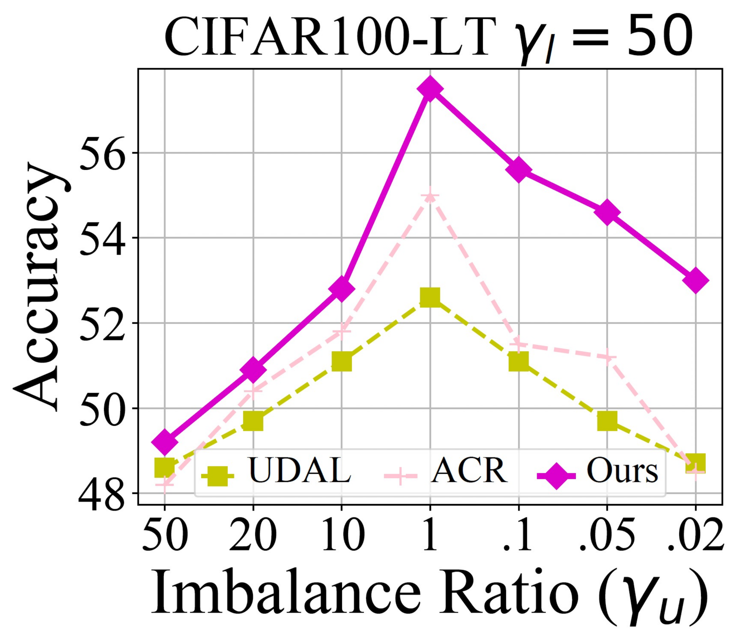
where $\bar{p}^{\frac{1}{T}}(y|\Omega(u_b)) = \sigma(\frac{1}{T}(f(\Omega(u_b)) + \log\frac{\hat{Q}}{\hat{Q}_{\alpha_t}}))$

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

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

Classification results

Robust performance even under an increasing degree of distribution mismatch



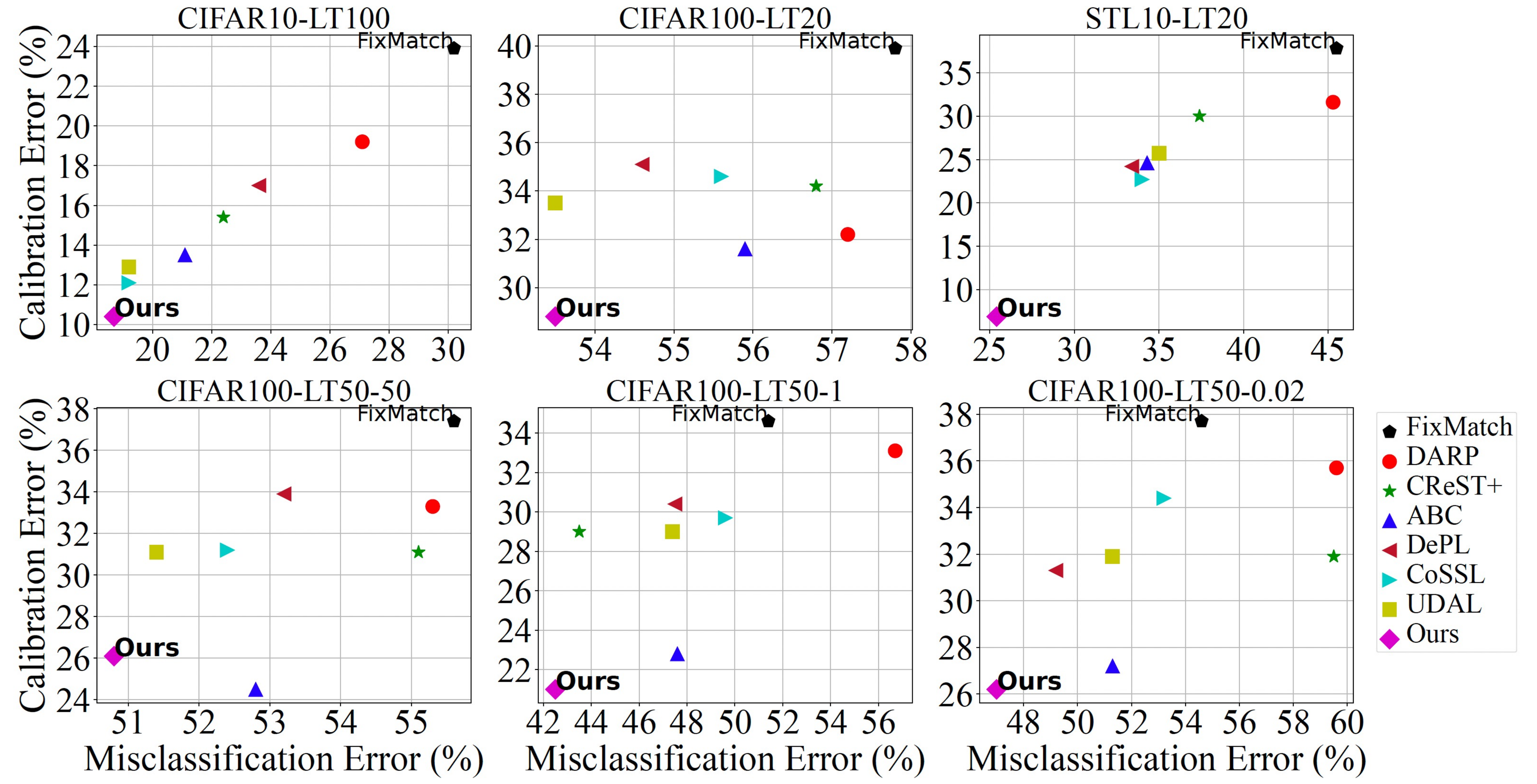
SOTA accuracy for challenging large-scale datasets under consistent case

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**

Calibration study

ADELLO achieves strong confidence calibration across all datasets

Better-calibrated models tend to improve LTSSL performance!



Ablations

Consistent alignment using all available data helps mitigate distribution shifts and achieve proper calibration!

Table 5: Influence of ADELLO components on model generalization (Test accuracy).

Components	CIFAR100-LT50 ↑			STL10-LT20 ↑
$\gamma_u \rightarrow$	50	1	0.02	N/A
FixMatch	44.4±0.6	48.6±1.0	45.4±1.6	54.5±4.3
+FlexDA	48.6±0.7	53.6±1.0	51.2±0.9	67.1±1.6
+CCR	44.7±0.7	51.6±1.6	47.2±2.1	61.1±2.9
+FlexDA+CCR	**49.2**±0.6	57.5±1.3	**53.0**±0.9	**74.6**±0.4
+FlexDA+KD	49.1±0.6	**58.2**±1.1	52.8±1.1	74.4±0.5

Table 6: Influence of ADELLO components on model calibration (ECE/MCE).

Components	CIFAR100-LT50 ↓	STL10-LT20 ↓
$\gamma_u \rightarrow$	50	N/A
FixMatch	37.4±0.4 / 57.3±1.1	37.8±4.5 / 55.1±4.9
+FlexDA	31.4±0.4 / 52.0±2.4	23.6±1.4 / 49.5±4.5
+CCR	36.3±0.7 / 56.8±1.8	22.2±2.3 / 38.5±4.2
+FlexDA+CCR	**26.1**±0.9 / **46.2**±0.6	**6.9**±0.3 / **25.9**±1.0
+FlexDA+KD	33.4±0.6 / 57.5±2.0	10.0±0.5 / 31.3±7.5

Paper

Code