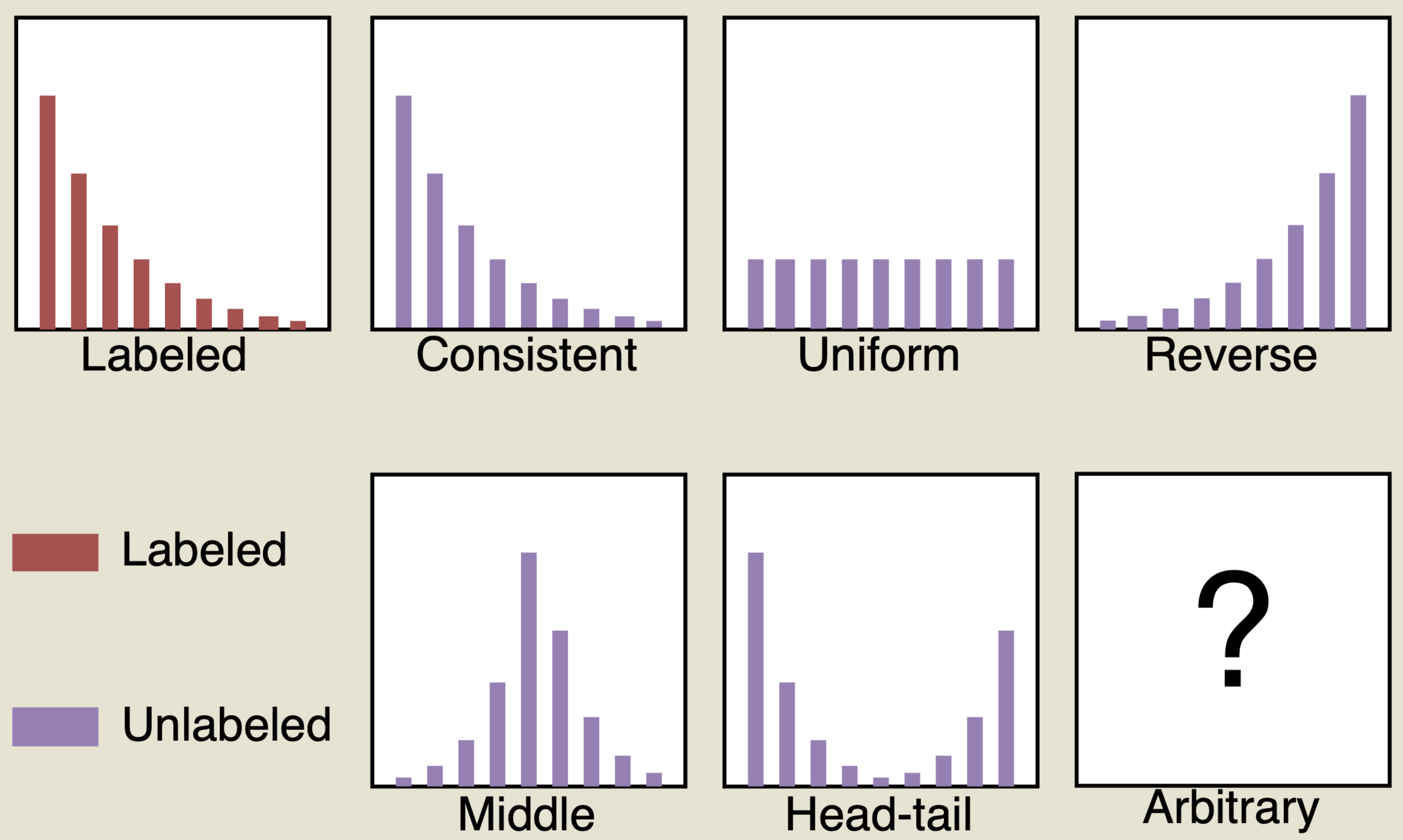
# SimPro: A Simple Probabilistic Framework Towards Realistic Long-Tailed Semi-Supervised Learning

Chaoqun Du*, Yizeng Han*, Gao Huang
Tsinghua University. *Equal contribution. Corresponding author.

ICML International Conference On Machine Learning

## Background



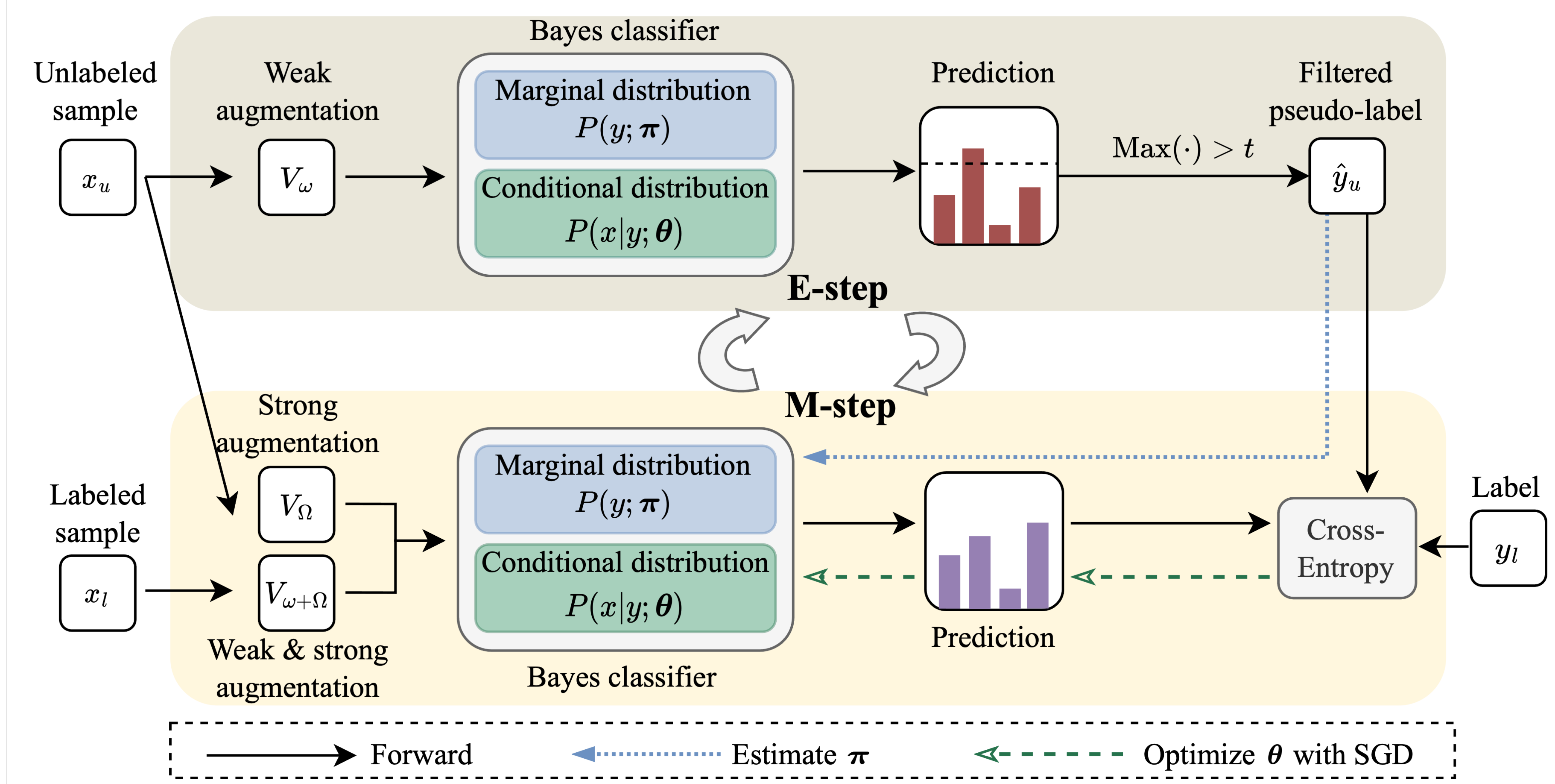

**Realistic Long-tailed semi-supervised learning (ReaLTSSL)**
Unlabeled data distribution is ***inconsistent*** with labeled data and ***unknown***

### Key idea

**Class distribution** is a **powerful statistic** for mastering long-tailed learning.

**Estimate & exploit it!**

## Method




### Theoretical analysis

**Probabilistic model.**

$$P(\boldsymbol{y},\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{\pi}) = P(\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{\pi})P(\boldsymbol{x}).$$

**E-step** (expected complete data log-likelihood).

$$\begin{aligned}\mathcal{Q}(\boldsymbol{\theta},\boldsymbol{\pi};\boldsymbol{\theta}',\boldsymbol{\pi}') &= \mathbb{E}_{\boldsymbol{y}|\boldsymbol{x};\boldsymbol{\theta}',\boldsymbol{\pi}'}\left[\log P(\boldsymbol{y},\boldsymbol{x};\boldsymbol{\theta},\boldsymbol{\pi})\right] \\ &= \sum_i \log P(y_i|x_i;\boldsymbol{\theta},\boldsymbol{\pi}_l) \\ &\quad + \sum_{j,y} P(y|x_j;\boldsymbol{\theta}',\boldsymbol{\pi}')\log P(y|x_j;\boldsymbol{\theta},\boldsymbol{\pi}_u).\end{aligned}$$

**M-step** (optimizing $\theta$ and $\pi$).

Optimal $\pi$ that maximizes $Q(\theta, \pi; \theta', \pi')$.

$$\hat{\boldsymbol{\pi}}_l = \frac{1}{N}\sum_{i=1}^{N} y_i, \quad \hat{\boldsymbol{\pi}}_u = \frac{1}{M}\sum_{j=1}^{M} P(y|x_j;\boldsymbol{\theta}',\boldsymbol{\pi}').$$

Objectives for optimizing $\theta$.

$$\mathcal{L}_l = -\frac{1}{B}\sum_{i=1}^{B}\log\frac{\exp(f_{\boldsymbol{\theta}}(x_i,y_i))}{\sum_{y'}\phi_{y'}^{\tau}\exp(f_{\boldsymbol{\theta}}(x_i,y'))}.$$

$$\mathcal{L}_u = -\frac{1}{\mu B}\sum_{j=1}^{\mu B}\mathbb{I}(\max_y(q_y)) \geq t)\sum_y q_y \log p_y.$$

Optimal $\Phi$ for learning a Bayes classifier.

$$\begin{aligned}\hat{\boldsymbol{\phi}} &= [\hat{\phi}_1,\hat{\phi}_2,\cdots,\hat{\phi}_K] \\ &= \frac{1}{N+M}\Big(\sum_i y_i + \sum_j P(y|x_j;\boldsymbol{\theta}',\boldsymbol{\pi}')\Big).\end{aligned}$$

The Bayes classifier for unlabeled or test data.

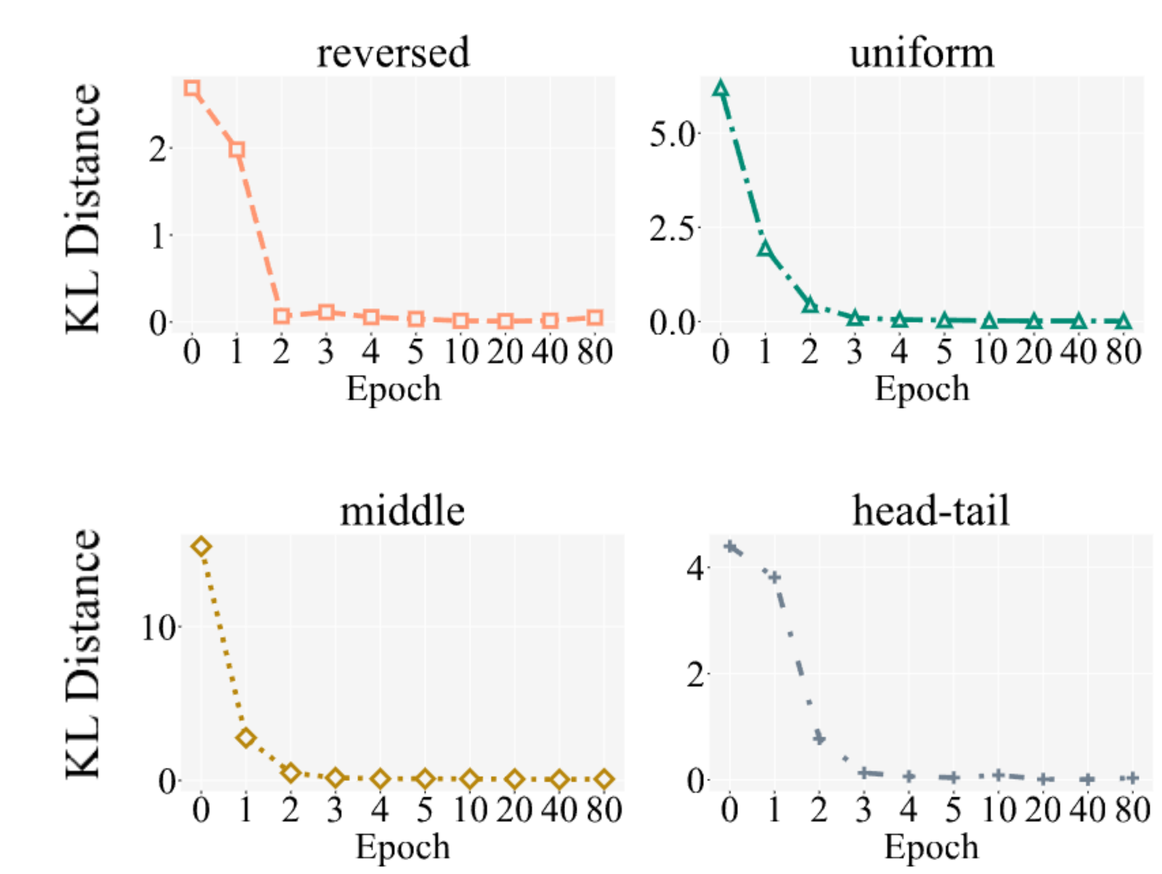
$$P(y|x;\boldsymbol{\theta};\hat{\boldsymbol{\pi}}) = \frac{P(y;\hat{\boldsymbol{\pi}})\exp(f_{\boldsymbol{\theta}}(x,y))}{\sum_{y'}P(y';\hat{\boldsymbol{\pi}})\exp(f_{\boldsymbol{\theta}}(x,y'))},$$

$$\text{or}\quad P(y|x;\boldsymbol{\theta}) = \frac{\exp(f_{\boldsymbol{\theta}}(x,y))}{\sum_{y'}\exp(f_{\boldsymbol{\theta}}(x,y'))}.$$

## Experiments

| | $\gamma_t \approx 286$ | |
|---|---|---|
| *ImageNet-127* | $32\times 32$ | $64\times 64$ |
| FixMatch (Sohn et al., 2020) | 29.7 | 42.3 |
| w/ DARP (Kim et al., 2020) | 30.5 | 42.5 |
| w/ CReST+ (Wei et al., 2021) | 32.5 | 44.7 |
| w/ CoSSL (Fan et al., 2022) | 43.7 | 53.9 |
| w/ ACR (Wei & Gan, 2023) | 57.2 | 63.6 |
| w/ SimPro | **59.1** | **67.0** |
| *ImageNet-127* | $\gamma_t = 1$ | |
| FixMatch (Sohn et al., 2020) | 38.7 | 46.7 |
| w/ ACR† (Wei & Gan, 2023) | 49.5 | 56.1 |
| w/ ACR (Wei & Gan, 2023) | 50.6 | 57.3 |
| w/ SimPro | **55.7** | **63.8** |
| *ImageNet-1k* | $\gamma_t = 1$ | |
| FixMatch (Sohn et al., 2020) | – | – |
| w/ ACR† (Wei & Gan, 2023) | 13.2 | 23.4 |
| w/ ACR (Wei & Gan, 2023) | 13.8 | 23.3 |
| w/ SimPro | **19.7** | **25.0** |

### Quality of estimation




Paper

Code