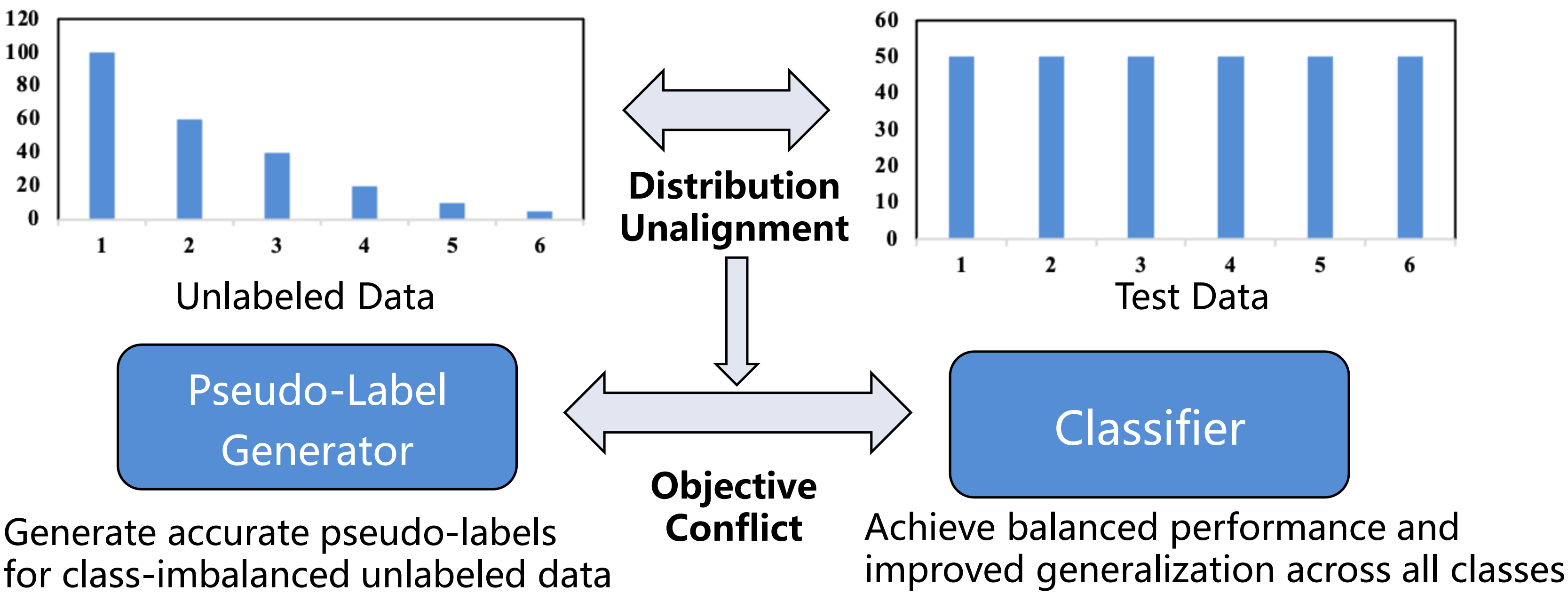
Debiased Label Enhancement for Class-Imbalanced Semi-Supervised Learning

Yifan WANG, Biao LIU, Xin GENG, Ning XU

Frontiers of Computer Science, DOI: 10.1007/s11704-026-51784-8

Problems & Ideas

- Problems of conventional class-imbalanced SSL approaches:
  - Using the classifier itself to generate pseudo-labels creates an **objective conflict**.
  - Classifiers seek uniform accuracy while pseudo-labeling follows skewed data, causing severe **confirmation bias**.
- Ideas: A disentangled debiased label enhancement framework (DALE) that explicitly disentangles pseudo-label generation from the classification task.

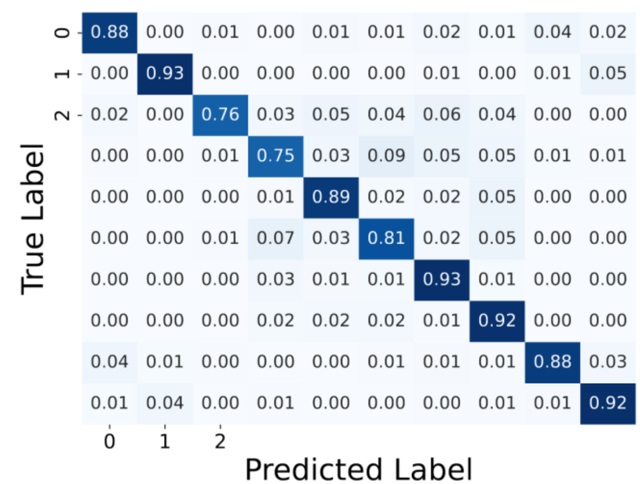


The motivation of DALE. The distribution unalignment between skewed unlabeled data (left) and balanced test data (right) leads to an inherent objective conflict between the pseudo-label generator and the final classifier.

Main Contributions

- Contributions:
  - We proposed a disentangled design helps resolve the conflicting goals of pseudo-label generator and classifier in imbalanced semi-supervised learning. We introduce a novel pseudo-labeler model that is specifically designed for the task of generating debiased pseudo-labels on the unlabeled data.
  - We investigate the effectiveness of using corrected loss to obtain high-quality pseudo-labels and provide a theoretical proof that the performance of the predictive model is guaranteed to improve as the quality of the pseudo-labels increases.

Algorithm	CIFAR-10-LT	
	$\gamma_l = \gamma_u$	$\gamma_l \neq \gamma_u$
DALE-	$82.16_{\pm 0.15}$	$83.57_{\pm 0.14}$
DALE	$\mathbf{84.23_{\pm 0.39}}$	$\mathbf{86.89_{\pm 0.45}}$



Left: Decoupling the pseudo-label generator (DALE) yields higher accuracy than the coupled version (DALE-). Right: The confusion matrix shows DALE effectively mitigates class-level bias, ensuring balanced performance across all classes.