CReST: A Class-Rebalancing Self-Training Framework for Imbalanced Semi-Supervised Learning



Chen Wei¹ Kihyuk Sohn² Clayton Mellina² Alan Yuille¹ Fan Yang²

¹Johns Hopkins University ²Google Cloud AI

Code: https://github.com/google-research/crest



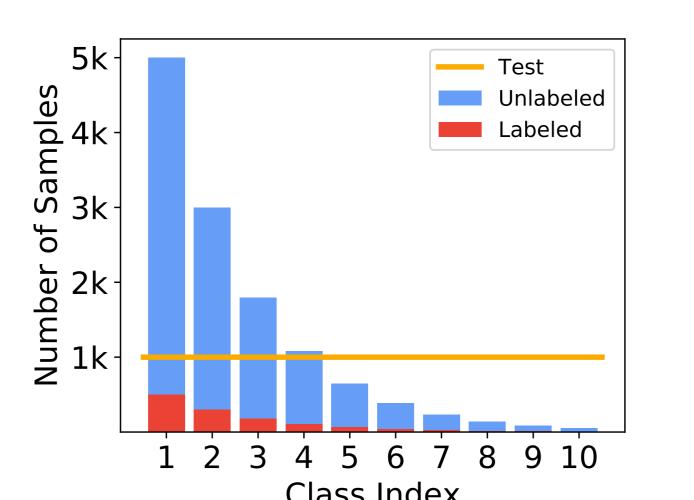
MOTIVATION

Challenge Description

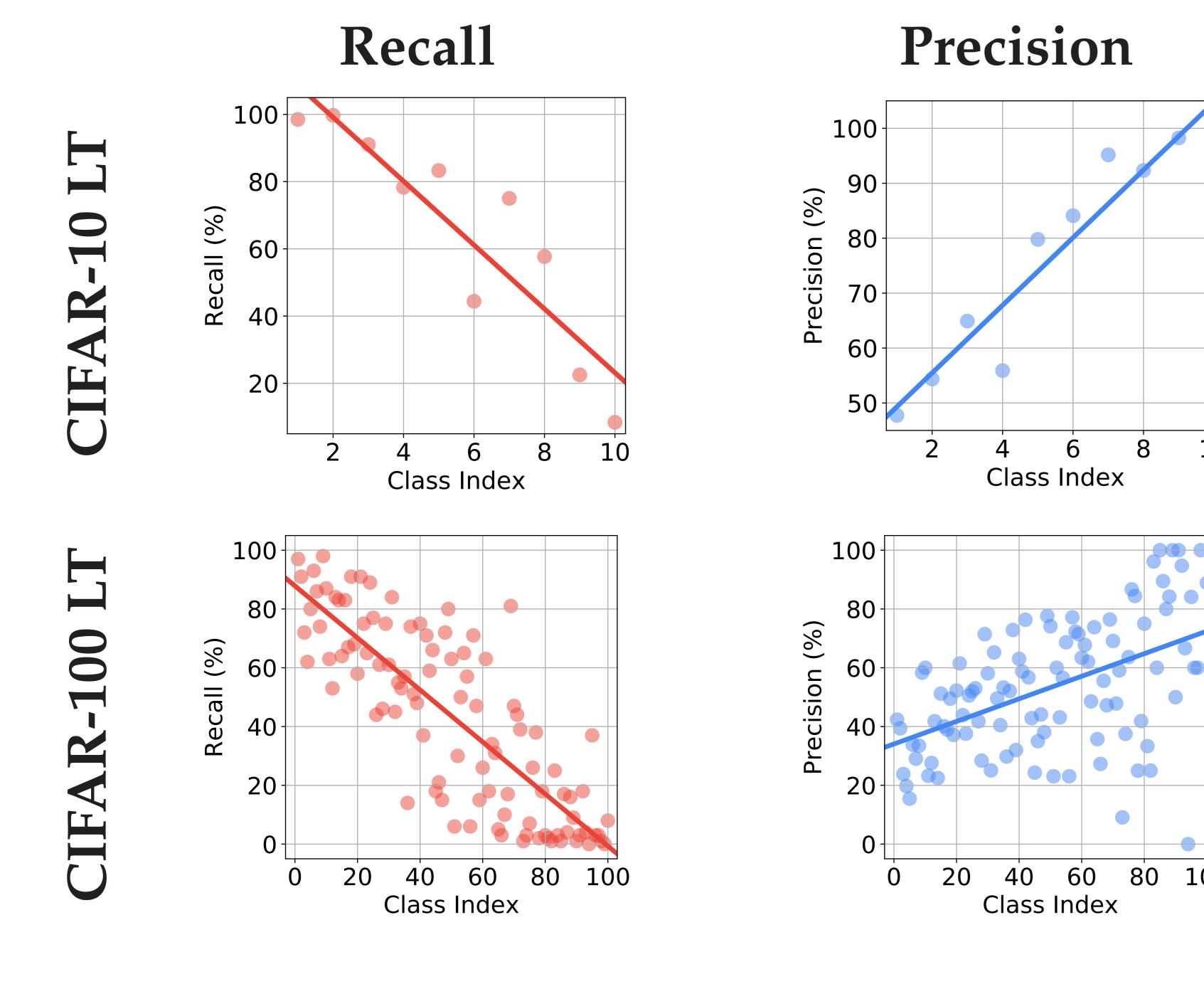
- Semi-supervised learning (SSL) on class-imbalanced data
- Imbalance makes SSL harder: Pseudo-labels are now biased
- SSL makes imbalance harder: Missing label information precludes rebalancing the unlabeled set

Our Setting

- Both labeled and unlabeled sets have **the same** imbalanced class distributions
- Evaluation on a balanced test set



OBSERVATION: RECALL VS. PRECISION



Setting

- FixMatch model on class-imbalanced CIFAR
- Per-class recall vs. precision
- Left: majority classes; Right: minority classes

Observation

- As one would guess, minority classes obtain poor recall
- Surprisingly, minority classes maintain high precision
- Minority predictions are more trustworthy

CLASS-REBALANCING SELF-TRAINING

Self-Training

- Multiple generations to train models iteratively
- A teacher model obtained at each generation
- **Pseudo-labels** $\hat{\mathcal{U}}$: The teacher's predictions on unlabeled data \mathcal{U}
- **Update** the labeled set \mathcal{X} with all the pseudo-labeled data $\hat{\mathcal{U}}$

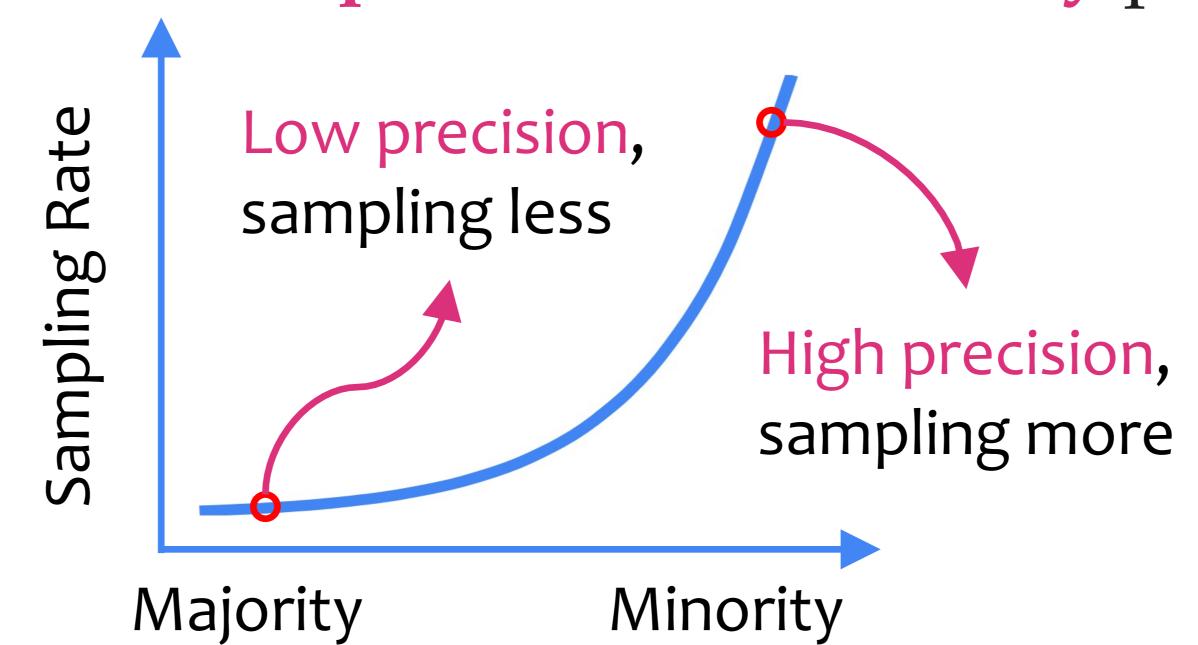
$$\mathcal{X}' = \mathcal{X} \cup \hat{\mathcal{U}} \tag{1}$$

Class-rebalancing Self-Training

• Update the labeled set \mathcal{X} with a selective subset $\hat{\mathcal{S}} \in \hat{\mathcal{U}}$

$$\mathcal{X}' = \mathcal{X} \cup \hat{\mathcal{S}} \tag{2}$$

- Rules to sample \hat{S} from \hat{U} : Class-rebalancing sampling
 - Favor the less frequent but trustworthy predictions



PROGRESSIVE DISTRIBUTION ALIGNMENT

Distribution Alignment (DA)

• Goal: Rebalance biased predictions q so that the aggregate of refined predictions matches a less biased target distribution

$$\tilde{q} = \text{Normalize}(q \frac{p(y)}{\tilde{p}(y)})$$
 (3)

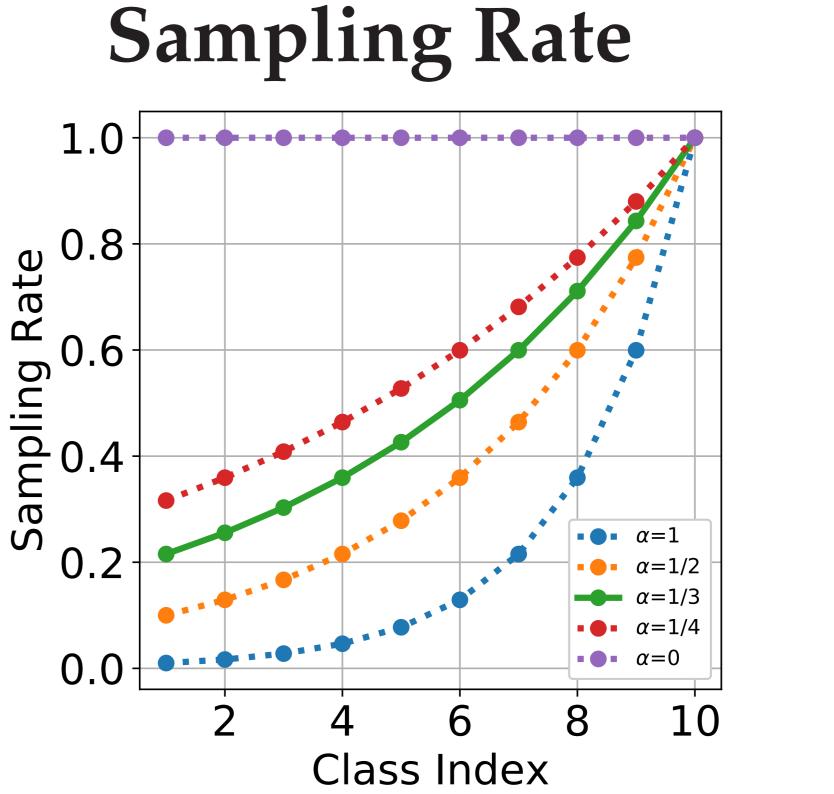
- Target distribution: p(y) the class distribution of the <u>dataset</u>
- Scaling factor: $\frac{p(y)}{\tilde{p}(y)}$ where $\tilde{p}(y)$ is the distribution of the <u>predictions</u>

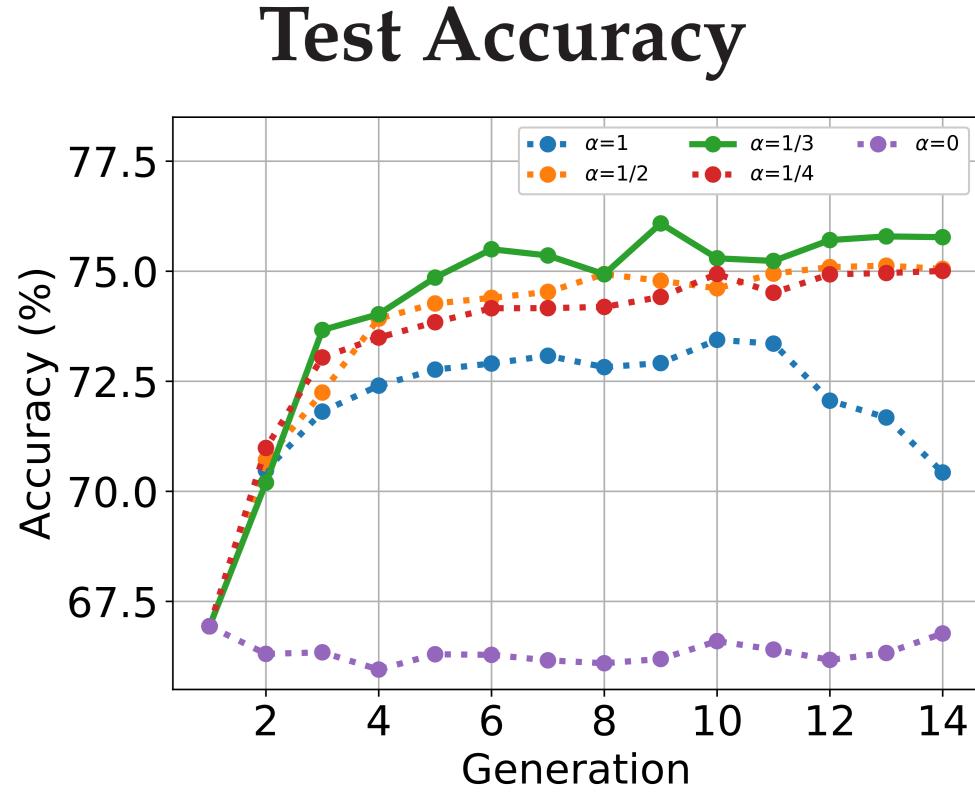
Progressive Distribution Alignment

- Target distribution: Normalize $(p(y)^t)$ where temperature $t \in [0, 1]$ t < 1 enhances the rebalancing strength
- Progressively rebalancing: Gradually decrease t from 1 to t_{\min}
 - Pseudo-labeling balance vs. precision

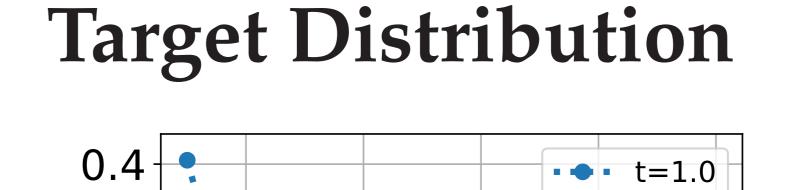
RESULTS

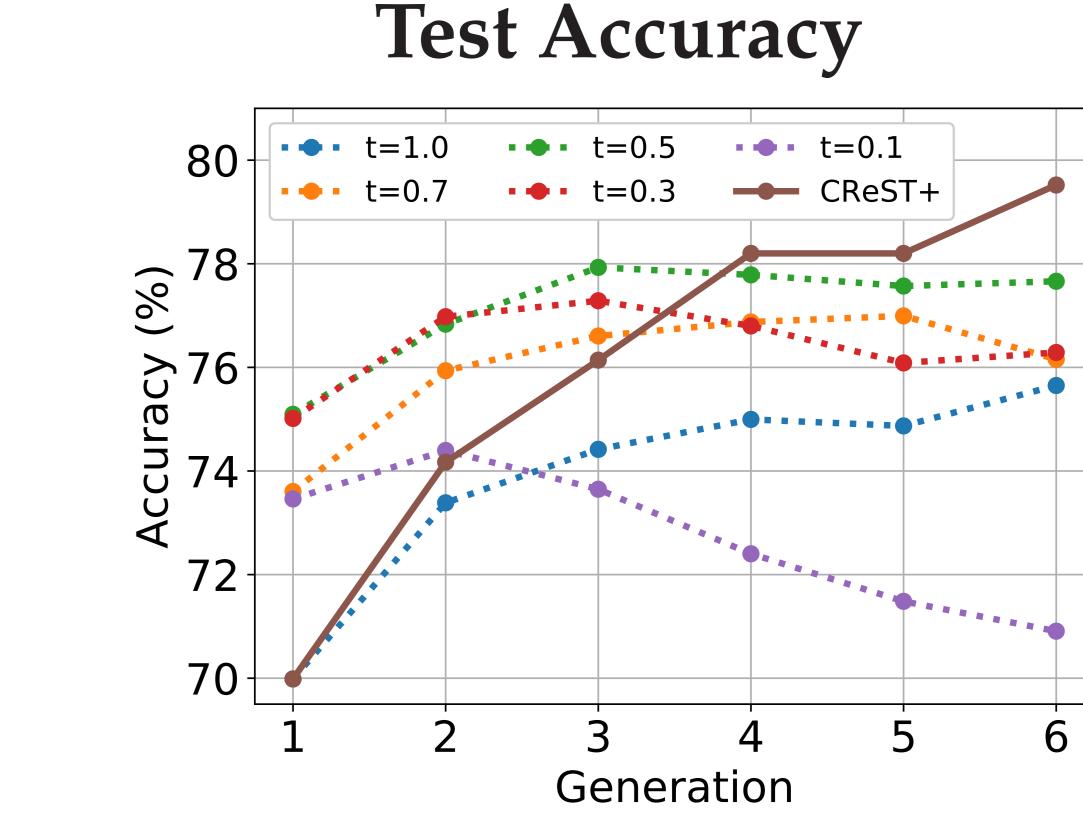
• Class-Rebalancing Self-Training (CIFAR10-LT)





• Progressive Distribution Alignment (CIFAR10-LT)





• Varying Label Fraction β and Imbalance Ratio γ

	CIFAR10-LT				CIFAR100-LT			
	$\beta = 10\%$		$\beta = 30\%$		$\beta = 10\%$		$\beta = 30\%$	
Method	$\gamma = 50$	$\gamma = 100$						
FixMatch	79.4	66.3	81.9	73.1	33.7	28.3	43.1	38.6
w/ CReST	83.8	75.9	84.2	77.6	37.4	32.1	45.6	40.2
w/ CReST+	84.2	78.1	84.9	79.2	38.8	34.6	46.7	42.0

Comparison with DARP

Method	$\gamma = 50$	$\gamma = 100$	$\gamma = 150$
FixMatch	79.2	71.5	68.4
w/DARP	81.8	75.5	70.4
w/ CReST	83.0	75.7	70.8
w/ CReST+	83.9	77.4	72.8

CReST+: CReST w/ progressive DA.

• Comparison on ImageNet127

Method	Gen ₁	Gen ₂	Gen
Supervised (100% labels)	75.8	_	_
Supervised (10% labels)	46.0	_	_
FixMatch (10% labels)	65.8	_	_
w/DA (t = 0.5)	69.1	_	-
w/ CReST	65.8	67.6	67.
w/ CReST+	68.3	70.7	73.