Smoothed Adaptive Weighting for Imbalanced Semi-Supervised Learning: Improve Reliability Against Unknown Distribution Data

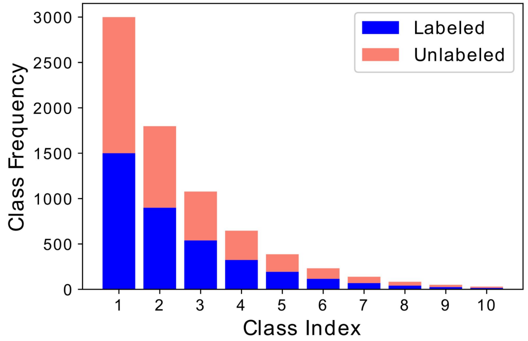
Zhengfeng Lai[1], Chao Wang[1], Henrry Gunawan[1],
Sen-Ching Cheung[2], Chen-Nee Chuah[1]

University of California, Davis[1]
University of Kentucky[2]

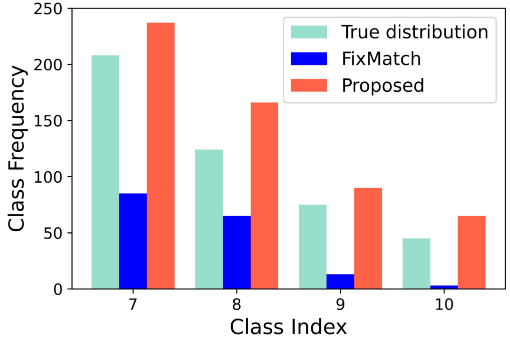
Imbalanced Semi-supervised Learning (SSL)

- SSL faces performance degradation when the unlabeled dataset is imbalanced
    - Designed with the assumption that both labeled set (L) and unlabeled set (U) are balanced
    - Pseudo labels during the self-training process can be ***biased*** towards the majority classes

- Recent class-imbalanced SSL[1,2]
    - Explicitly assume that U share similar distributions to L
    - In real-world scenarios, <u>U may have different distributions from L</u>
    - ***Can we relieve such assumptions?***



An example of imbalance dataset based on CIFAR-10



FixMatch's performance on the minority classes

[1] Kim, Jaehyung, et al. "Distribution aligning refinery of pseudo-label for imbalanced semi-supervised learning." NeurIPS 2020.
[2] Wei, Chen, et al. "Crest: A class-rebalancing self-training framework for imbalanced semi-supervised learning." CVPR 2021.

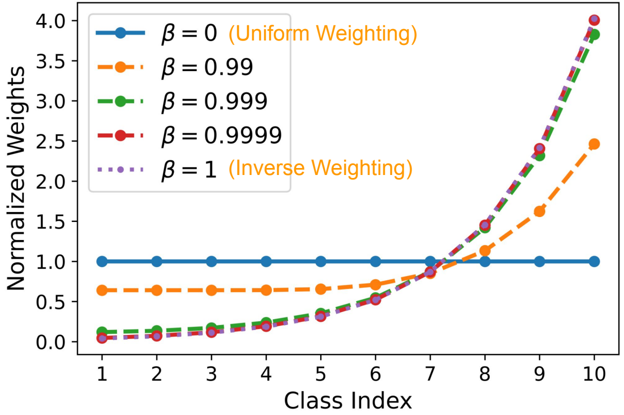
Main Contributions

- Verify the necessity and benefits of smoothed weights in the consistency loss
  - Uniform weighting? Inverse class-frequency weighting?
  - Smoothing weighting?
- SAW: Smoothed Adaptive Weighting
  - Estimate the distribution: does not assume that ***U*** has the same distribution as ***L***
    - *Effective number of samples*[2] is estimated based on pseudo labels
  - Calculate the smoothed weights by smoothing weighting schemes
- Evaluate the proposed methods under various scenarios
  - Hold-out tests are of various distributions besides balanced distribution reported in prior works

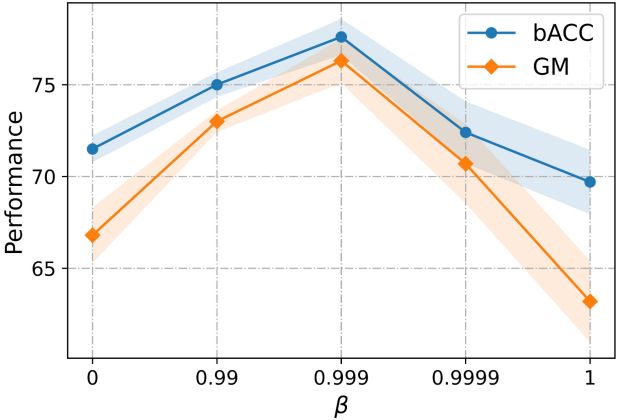
[1] Zhang, Bowen, et al. "Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling." NeurIPS 2021.
[2] Cui, Yin, et al. "Class-balanced loss based on effective number of samples." CVPR 2019.

Why do we need smoothed weighting?

- Uniform weighting -> Ignore the class imbalance problem
- Inverse class-frequency weighting -> May erroneously overemphasize the weaker classes (overfitting)[2]



The effect of weighting scheme[1]



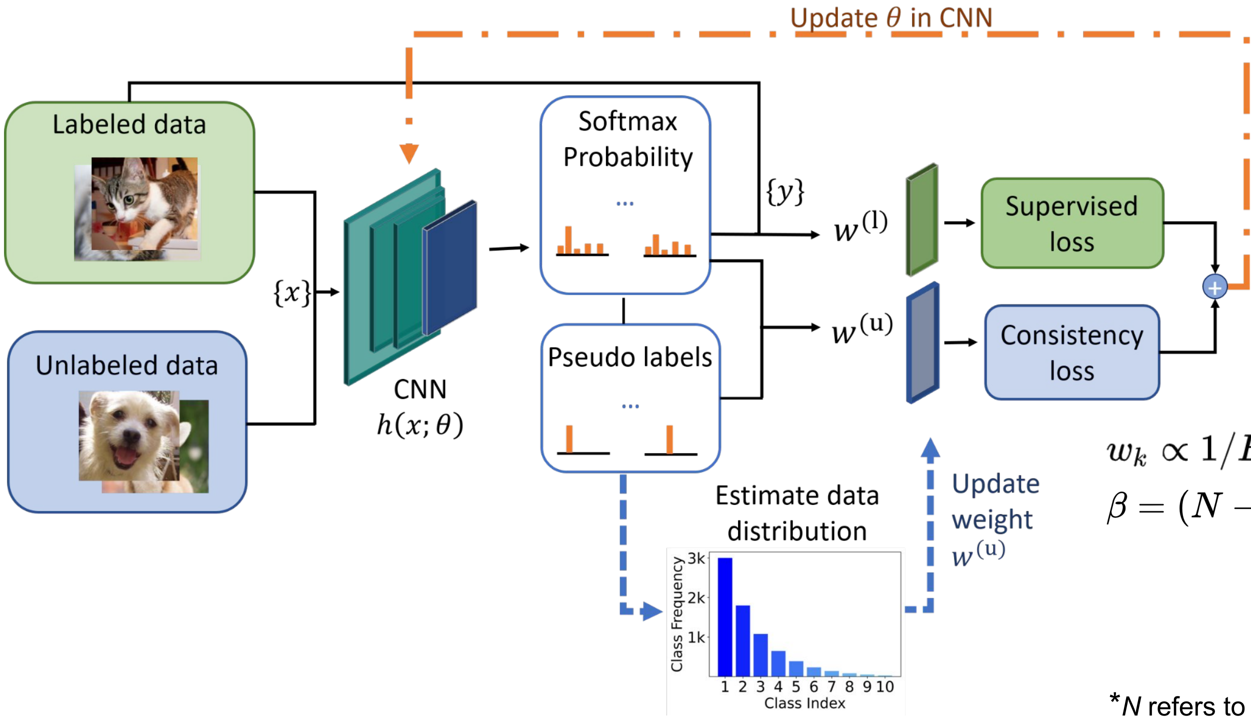
FixMatch on CIFAR-10 with the weighting scheme[1] when imbalance ratio is 100

$$w_k \propto 1/E_k, \text{ where } E_k = (1-\beta^{n_k})/(1-\beta)$$
$$\beta = (N-1)/N$$

[1] Cui, Yin, et al. "Class-balanced loss based on effective number of samples." CVPR 2019.
[2] Tang, Kaihua, et al. "Long-tailed classification by keeping the good and removing the bad momentum causal effect." NeurIPS 2020.

SAW: Smoothed Adaptive Weighting Scheme

Weighted consistency loss: $\mathcal{L}_{cw}(x; w, \theta) := \sum_{k=1}^{C} w_k \cdot p(x;\theta)_k \cdot \log(h(\texttt{pertub}(x);\theta)_k)$



We use *effective samples*[1] as one smoothing example.

$w_k \propto 1/E_k$, where $E_k = (1-\beta^{n_k})/(1-\beta)$

$\beta = (N-1)/N$

*N refers to the estimated average number of each class.

[1] Cui, Yin, et al. "Class-balanced loss based on effective number of samples." CVPR 2019.

Experimental Design

- Training sets scenarios
  - Both $\boldsymbol{L}$ and $\boldsymbol{U}$: long-tailed distributions
    - S1) $\boldsymbol{U}$ has the same distributions as $\boldsymbol{L}$ for varying imbalanced ratios
    - S2) $\boldsymbol{U}$ has different distributions as $\boldsymbol{L}$ for varying imbalanced ratios
- Hold-out test sets scenarios
  - a) Balanced distributions
  - b) Reversed distributions
  - c) Same distributions as $\boldsymbol{L}$ in the training set
- A real-world medical imaging application
  - Grey/white matter segmentation in gigapixel images[1]

[1] Lai, Zhengfeng, et al. "Joint Semi-supervised and Active Learning for Segmentation of Gigapixel Pathology Images with Cost-Effective Labeling." ICCV Workshop 2021.

Results

- Measuring metric: bACC (balanced accuracy) and GM (geometric mean)
- Imbalanced ratios (γ): for the labeled set, it is set as 100.
- Two recent state-of-the-art imbalanced SSL algorithms
  - CReST (assume ***U*** and ***L*** have the same distributions)
  - DARP (use the confusion matrix on ***L*** to estimate the distribution of ***U***)

S2a): U has a ***different*** distribution from L and the test set is balanced. (CIFAR-10)

Algorithm	$\gamma_u = 1$	$\gamma_u = 50$	$\gamma_u = 150$
ReMixMatch (Berthelot et al., 2020)	48.3±0.14 / 19.5±0.85	75.1±0.43 / 71.9±0.77	72.5±0.10 / 68.2±0.32
ReMixMatch* (Berthelot et al., 2020)	85.0±1.35 / 84.3±1.55	77.0±0.12 / 74.7±0.04	72.8±0.10 / 68.8±0.21
ReMixMatch* + DARP (Kim et al., 2020)	**89.7**±0.15 / **89.4**±0.17	77.4±0.22 / 75.0±0.25	73.2±0.11 / 69.2±0.31
ReMixMatch* + CReST (Wei et al., 2021)	45.9±1.27 / 20.1±1.99	70.2±0.45 / 65.8±0.71	65.4±0.34 / 62.9±0.15
ReMixMatch* + SAW	88.3±0.15 / 88.9±0.10	**80.3**±0.36 / **79.6**±0.40	**74.0**±0.94 / **72.4**±0.94
FixMatch (Sohn et al., 2020)	68.9±1.95 / 42.8±8.11	73.9±0.25 / 70.5±0.52	69.6±0.60 / 62.6±1.11
FixMatch + DARP (Kim et al., 2020)	**85.4**±0.55 / **85.0**±0.65	77.3±0.17 / 75.5±0.21	72.9±0.24 / 69.5±0.18
FixMatch + CReST (Wei et al., 2021)	60.2±1.34 / 35.9±2.50	65.8±0.78 / 67.1±0.84	60.1±1.44 / 51.4±1.68
FixMatch + SAW	83.9±0.44 / 83.3±0.47	**81.5**±2.25 / **80.9**±2.30	**76.8**±0.31 / **75.4**±0.37

S2b): U has a ***different*** distribution from L and the test set is ***imbalanced*** and of ***reversed*** distributions. (CIFAR-10)

Algorithm	$\gamma = 50$	$\gamma = 100$	$\gamma = 150$
ReMixMatch (Berthelot et al., 2020)	71.0±0.55 / 83.5±0.29	54.7±0.51 / 74.4±0.47	41.5±1.69 / 66.4±1.22
ReMixMatch + DARP (Kim et al., 2020)	66.9±0.75 / 80.5±0.46	49.7±1.55 / 70.5±0.90	35.8±1.81 / 60.9±2.42
ReMixMatch + CReST (Wei et al., 2021)	64.3±0.25 / 75.7±0.34	51.2±0.92 / 72.1±0.85	39.2±1.46 / 65.8±1.88
ReMixMatch + SAW	**86.3**±0.61 / **86.1**±0.64	**77.0**±0.59 / **76.0**±0.42	**71.5**±0.30 / **68.9**±0.26
FixMatch (Sohn et al., 2020)	70.5±0.26 / 82.2±0.31	51.0±1.65 / 71.5±1.24	38.5±1.15 / 63.4±0.31
FixMatch + DARP (Kim et al., 2020)	72.2±0.62 / 82.8±0.17	57.6±0.36 / 74.8±0.48	46.5±1.26 / 68.1±0.10
FixMatch + CReST (Wei et al., 2021)	69.4±0.35 / 80.1±0.41	52.4±0.32 / 70.3±0.28	42.9±1.45 / 67.4±1.07
FixMatch + SAW	**78.7**±0.77 / **84.2**±0.36	**64.3**±1.96 / **76.4**±0.88	**57.5**±2.83 / **70.5**±1.50

**More results on other scenarios can be found in the main paper.*

Discussion & Future Work

- SAW can complement consistency-based SSL algorithms
  - We verified the feasibility of adding weights to the consistency loss
  - We investigated the necessity and benefits of smoothed weights
  - SAW does not require the unlabeled data to have similar distributions as the labeled data
- Limitation & Future Work
  - Still assume U and L contain the same classes
  - Study more various scenarios in the imbalanced setting
  - Investigate sophisticated ways of estimating the distribution in the unlabeled data