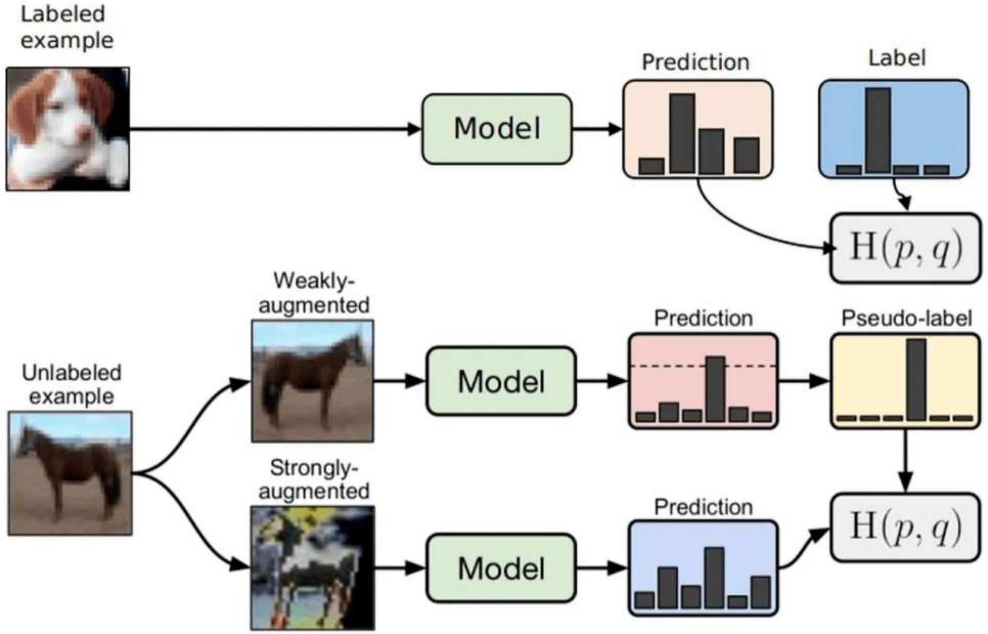
Boosting Semi-Supervised Learning by Exploiting All Unlabeled Data

Yuhao Chen[1] Xin Tan[2] Borui Zhao[1] Zhaowei Chen[1] Renjie Song[1] Jiajun Liang[1] Xuequan Lu[3]
[1]MEGVII Technology [2]East China Normal University [3]Deakin University
{yhao.chen0617, zhaoborui.gm, chaowechan}@gmail.com, xtan@cs.ecnu.edu.cn
{songrenjie, liangjiajun}@megvii.com, xuequan.lu@deakin.edu.au

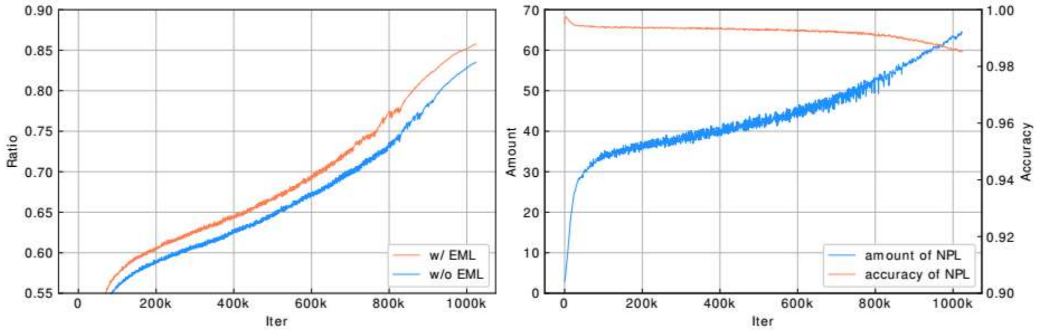
CVPR 2023

Background

How to utilize all unlabeled examples?

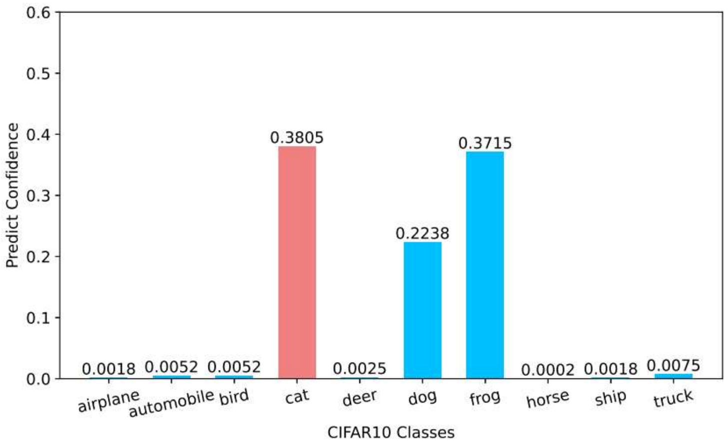


FixMatch



(a)

(b)



FullMatch = EML + ANL

Method

Entropy Meaning Loss(EML)

weakly-augmented: $Q^{(i)} = [q_1^{(i)}, ..., q_C^{(i)}]$

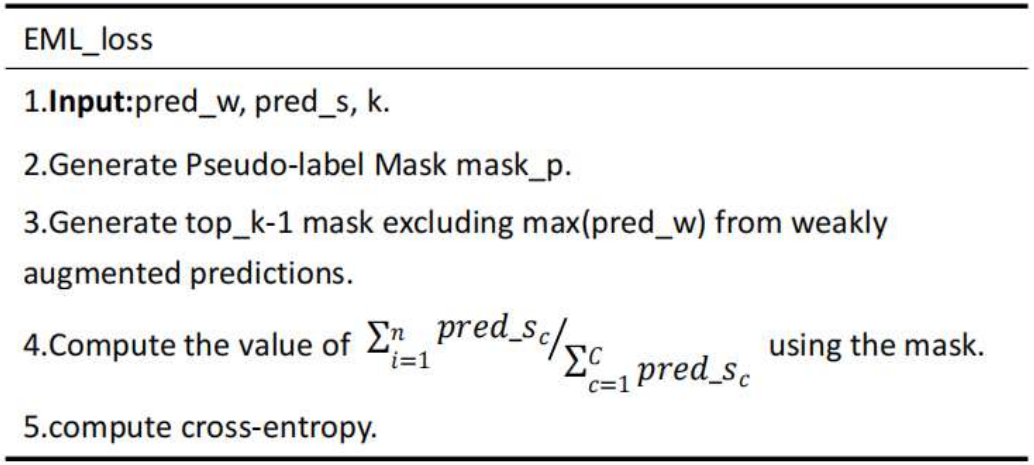
strongly-augmented: $P^{(i)} = [p_1^{(i)}, ..., p_C^{(i)}]$

mask:

$$s_c^{(i)} = \mathbb{1}(q_c^{(i)} \geq \tau)$$

$$u_c^{(i)} = \mathbb{1}(max(Q^{(i)}) \geq \tau) \cdot \mathbb{1}(s_c^{(i)} = 0)$$

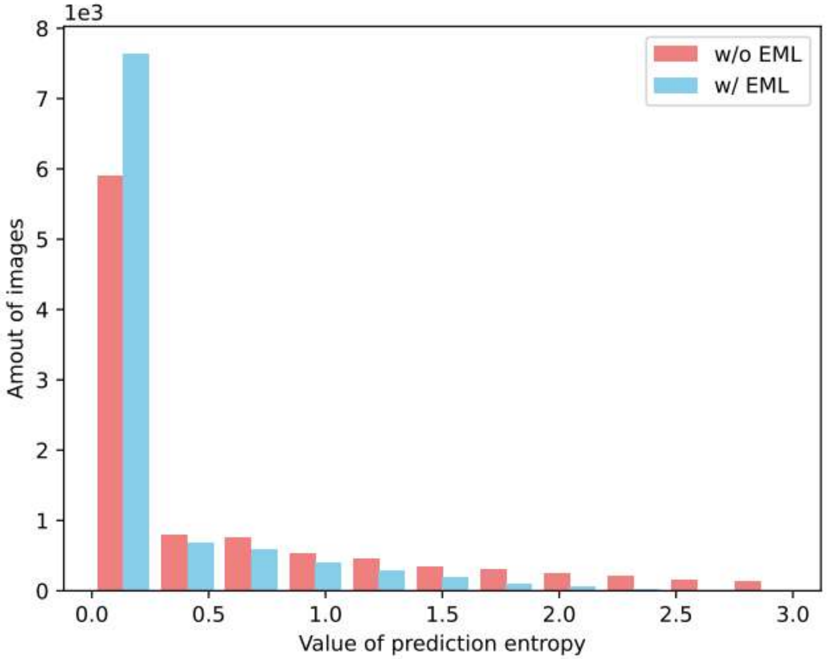
soft label: $$y_c^{(i)} = \frac{1 - \mathbb{1}(u_c^{(i)} = 0) \cdot p_c^{(i)}}{\sum_c \mathbb{1}(u_c^{(i)} = 1)}$$

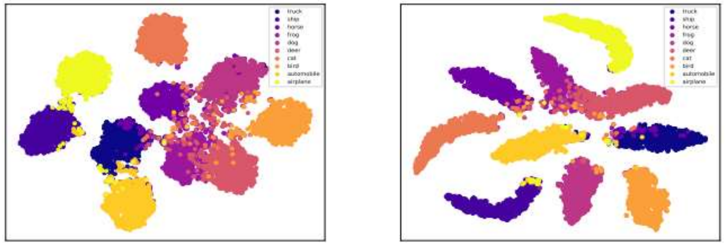


```
EML_loss
1.Input:pred_w, pred_s, k.
2.Generate Pseudo-label Mask mask_p.
3.Generate top_k-1 mask excluding max(pred_w) from weakly augmented predictions.
4.Compute the value of $\sum_{i=1}^{n} pred\_s_c / \sum_{c=1}^{C} pred\_s_c$ using the mask.
5.compute cross-entropy.
```

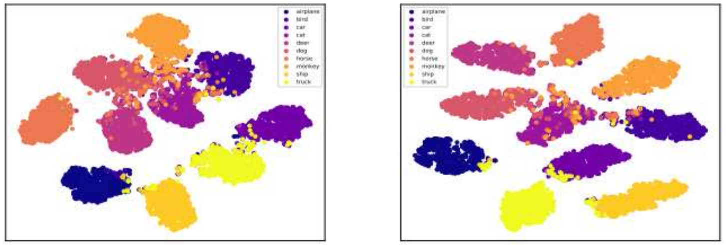
Method

Entropy Meaning Loss(EML)





(a) CIFAR-10

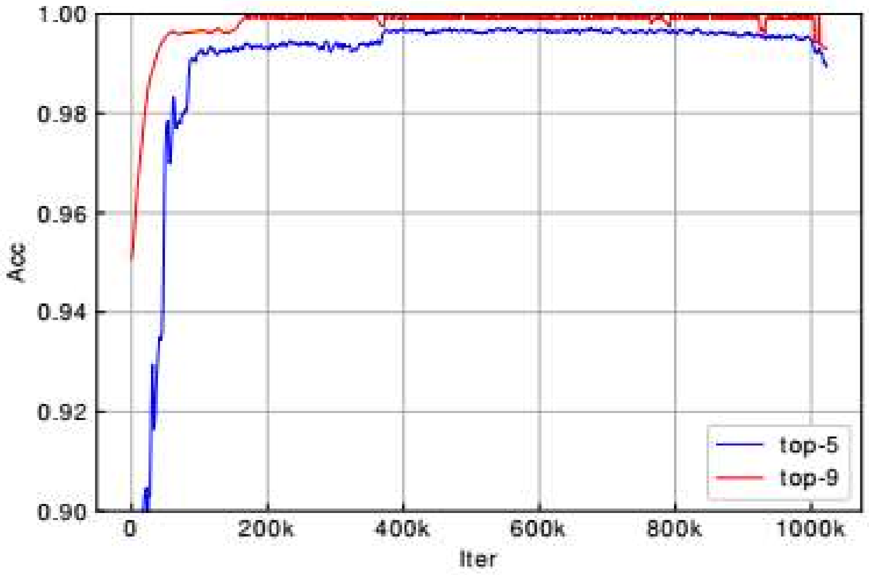


(b) STL-10

Method

top_k & ANL

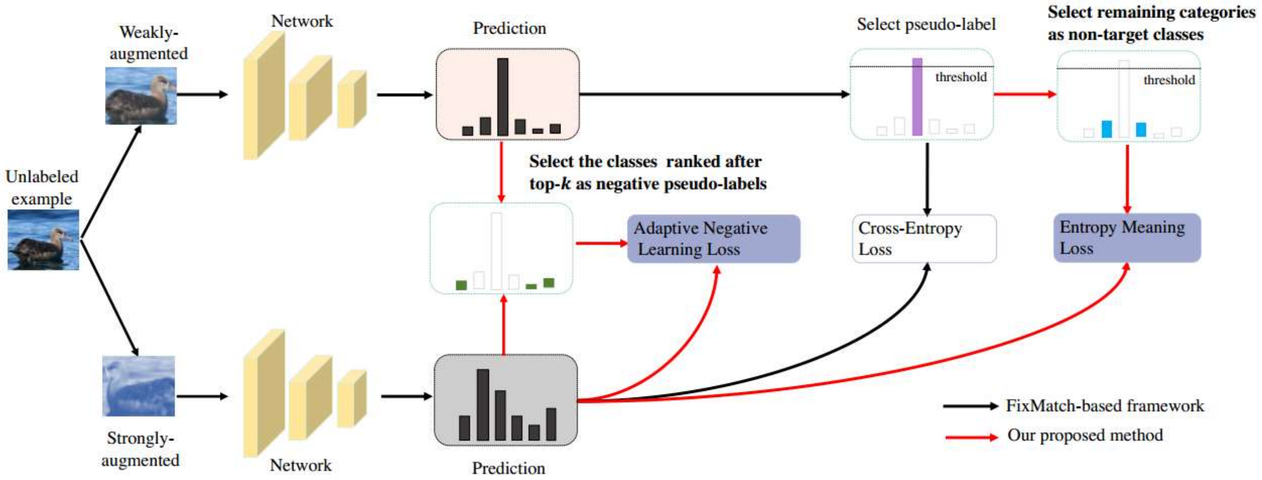
$$k = \arg\min_{\theta \in [2, C]} (Acc(P_t, \hat{Q}_t, \theta) = 100\%)$$

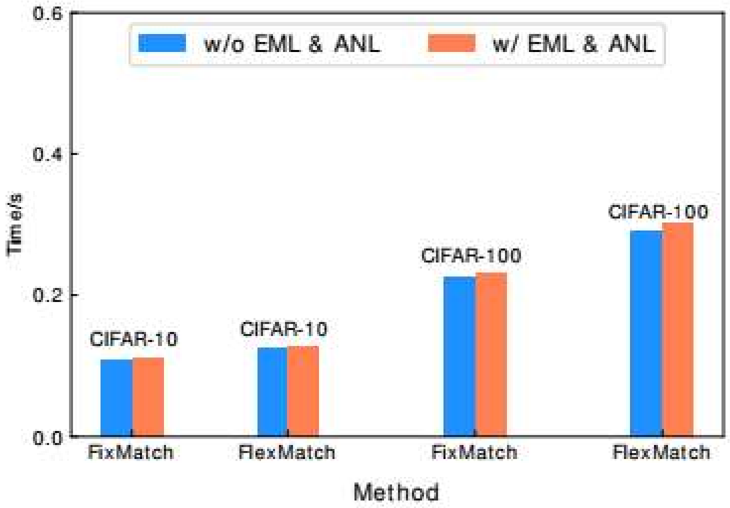


$$-\frac{1}{B}\sum_{i=1}^{B}\sum_{c=1}^{C} \mathbb{1}[Rank(q_c^{(i)}) > k]\; log(1 - p_c^{(i)})$$

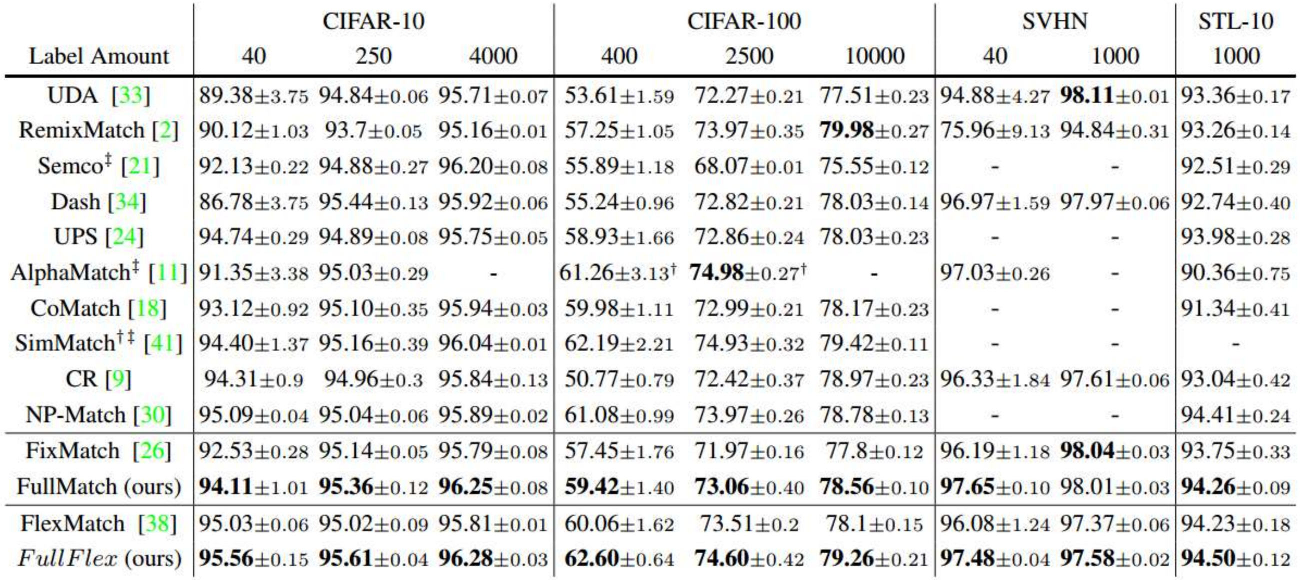
Method

FullMatch



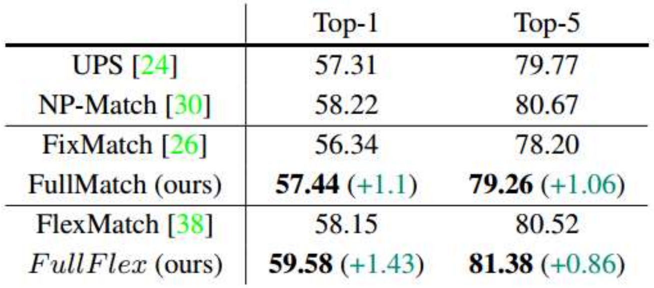


Experiment



Label Amount	CIFAR-10 40	CIFAR-10 250	CIFAR-10 4000	CIFAR-100 400	CIFAR-100 2500	CIFAR-100 10000	SVHN 40	SVHN 1000	STL-10 1000
UDA [33]	89.38±3.75	94.84±0.06	95.71±0.07	53.61±1.59	72.27±0.21	77.51±0.23	94.88±4.27	**98.11**±0.01	93.36±0.17
RemixMatch [2]	90.12±1.03	93.7±0.05	95.16±0.01	57.25±1.05	73.97±0.35	**79.98**±0.27	75.96±9.13	94.84±0.31	93.26±0.14
Semco‡ [21]	92.13±0.22	94.88±0.27	96.20±0.08	55.89±1.18	68.07±0.01	75.55±0.12	-	-	92.51±0.29
Dash [34]	86.78±3.75	95.44±0.13	95.92±0.06	55.24±0.96	72.82±0.21	78.03±0.14	96.97±1.59	97.97±0.06	92.74±0.40
UPS [24]	94.74±0.29	94.89±0.08	95.75±0.05	58.93±1.66	72.86±0.24	78.03±0.23	-	-	93.98±0.28
AlphaMatch‡ [11]	91.35±3.38	95.03±0.29	-	61.26±3.13†	**74.98**±0.27†	-	97.03±0.26	-	90.36±0.75
CoMatch [18]	93.12±0.92	95.10±0.35	95.94±0.03	59.98±1.11	72.99±0.21	78.17±0.23	-	-	91.34±0.41
SimMatch†‡ [41]	94.40±1.37	95.16±0.39	96.04±0.01	62.19±2.21	74.93±0.32	79.42±0.11	-	-	-
CR [9]	94.31±0.9	94.96±0.3	95.84±0.13	50.77±0.79	72.42±0.37	78.97±0.23	96.33±1.84	97.61±0.06	93.04±0.42
NP-Match [30]	95.09±0.04	95.04±0.06	95.89±0.02	61.08±0.99	73.97±0.26	78.78±0.13	-	-	94.41±0.24
FixMatch [26]	92.53±0.28	95.14±0.05	95.79±0.08	57.45±1.76	71.97±0.16	77.8±0.12	96.19±1.18	**98.04**±0.03	93.75±0.33
FullMatch (ours)	**94.11**±1.01	**95.36**±0.12	**96.25**±0.08	**59.42**±1.40	**73.06**±0.40	**78.56**±0.10	**97.65**±0.10	98.01±0.03	**94.26**±0.09
FlexMatch [38]	95.03±0.06	95.02±0.09	95.81±0.01	60.06±1.62	73.51±0.2	78.1±0.15	96.08±1.24	97.37±0.06	94.23±0.18
FullFlex (ours)	**95.56**±0.15	**95.61**±0.04	**96.28**±0.03	**62.60**±0.64	**74.60**±0.42	**79.26**±0.21	**97.48**±0.04	**97.58**±0.02	**94.50**±0.12

Table 1. **Top-1 accuracy** (%) **for CIFAR-10/100, SVHN and STL-10 datasets on 3 different folds.** *FullFlex* indicates applying our method to FlexMatch. † indicates introducing an additional technique named DA (Distribution Alignment) [2]. ‡ represents the result comes from the original paper.



	Top-1	Top-5
UPS [24]	57.31	79.77
NP-Match [30]	58.22	80.67
FixMatch [26]	56.34	78.20
FullMatch (ours)	**57.44** (+1.1)	**79.26** (+1.06)
FlexMatch [38]	58.15	80.52
FullFlex (ours)	**59.58** (+1.43)	**81.38** (+0.86)

Table 2. **Top-1 and Top-5 accuracy (%) on ImageNet.** In green are the values of performance improvement over the baselines.

Experiment

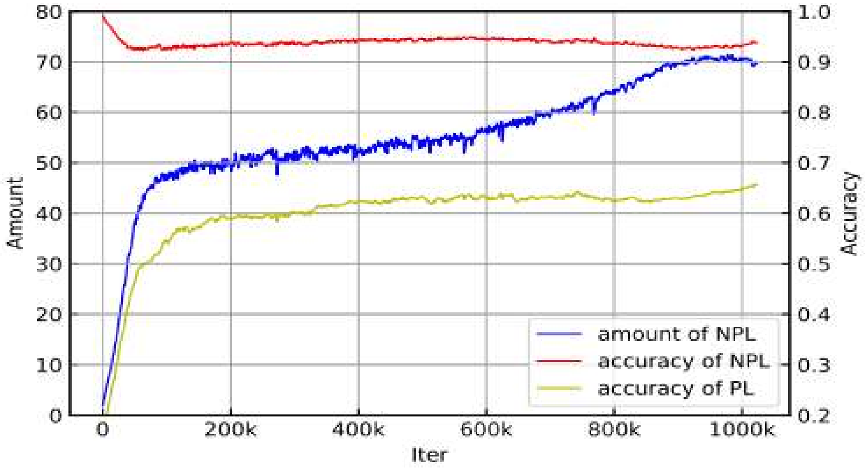


Figure 1. Visualize the experimental results on CIFAR-100 with 400 label samples.

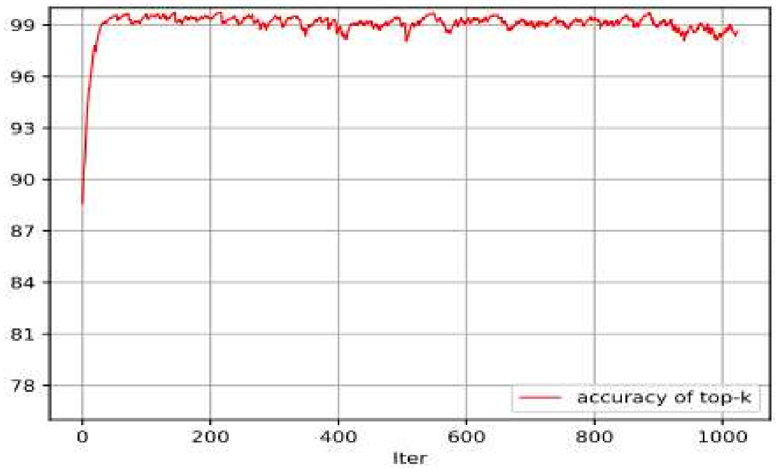
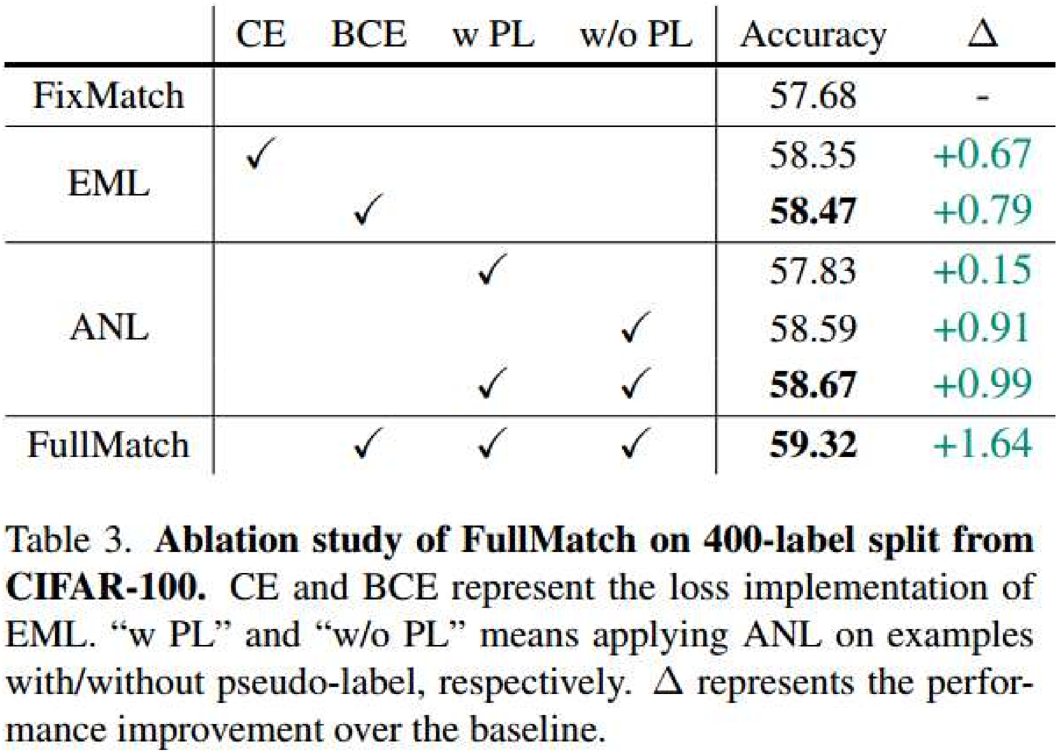


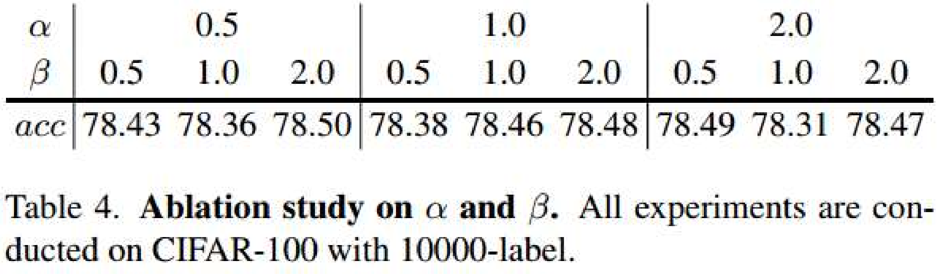
Figure 2. Visualizations the accuracy of top-k in the first 1K iterations on CIFAR-100 with 400 label samples.

Ablation Study



	CE	BCE	w PL	w/o PL	Accuracy	Δ
FixMatch					57.68	-
EML	✓				58.35	+0.67
		✓			**58.47**	+0.79
ANL			✓		57.83	+0.15
				✓	58.59	+0.91
			✓	✓	**58.67**	+0.99
FullMatch		✓	✓	✓	**59.32**	+1.64

Table 3. **Ablation study of FullMatch on 400-label split from CIFAR-100.** CE and BCE represent the loss implementation of EML. “w PL” and “w/o PL” means applying ANL on examples with/without pseudo-label, respectively. Δ represents the performance improvement over the baseline.



α	0.5			1.0			2.0		
β	0.5	1.0	2.0	0.5	1.0	2.0	0.5	1.0	2.0
acc	78.43	78.36	78.50	78.38	78.46	78.48	78.49	78.31	78.47

Table 4. **Ablation study on α and β.** All experiments are conducted on CIFAR-100 with 10000-label.

Thanks