MarginMatch: Improving Semi-Supervised Learning with Pseudo-Margins

Tiberiu Sosea
tsosea2@uic.edu

Cornelia Caragea
cornelia@uic.edu

ID: WED-PM-326

UIC COMPUTER SCIENCE

The University of Illinois Chicago

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Summary

Semi-Supervised Learning mitigates the requirement for labeled by leveraging unlabeled data.

FixMatch

$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \tau) \times H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))$$

FlexMatch

$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \mathcal{T}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}) \times H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))$$

MarginMatch

c acts as “ground truth”

$$\mathrm{PM}^t_c(\hat{x}) = z_c - max_{c!=i}(z_i)$$

c is the argmax of the model at prediction t

$$\mathrm{APM}^t_c(\hat{x}) = \frac{1}{t}\sum_{j=1}^{t} \mathrm{PM}^j_c(\hat{x})$$

flexmatch

The model may have different argmax values at different iterations. This phenomenon is captured using averaged pseudo-margins

argmax

argmax

$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \mathbb{1}(\mathrm{APM}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}(\hat{x}_i) > \gamma^t) \times \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \mathcal{T}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}) \times H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))$$

Summary

Semi-Supervised Learning mitigates the requirement for labeled by leveraging unlabeled data.

FixMatch

Confidence threshold

$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \boxed{\tau}) \times H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))$$

FlexMatch

Confidence threshold

$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \boxed{\mathcal{T}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}}) \times H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))$$

MarginMatch

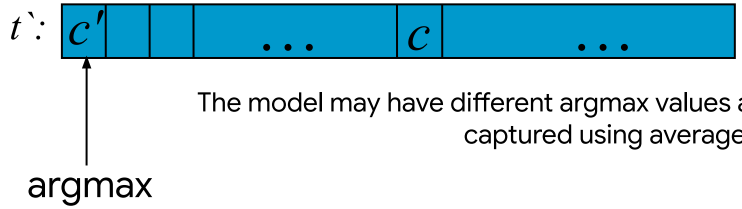
c acts as "ground truth"

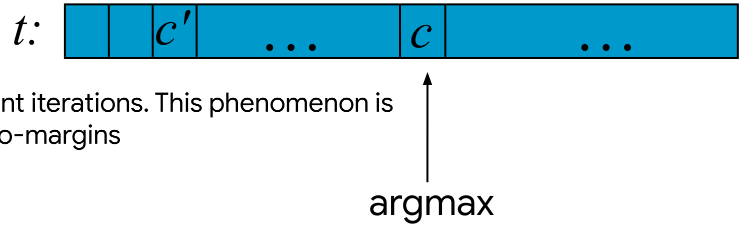
c is the argmax of the model at prediction t

$$\mathrm{PM}_c^t(\hat{x}) = z_c - max_{c!=i}(z_i)$$

$$\mathrm{APM}_c^t(\hat{x}) = \frac{1}{t} \sum_{j=1}^{t} \mathrm{PM}_c^j(\hat{x})$$

flexmatch





The model may have different argmax values at different iterations. This phenomenon is captured using averaged pseudo-margins

$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \mathbb{1}(\mathrm{APM}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}(\hat{x}_i) > \gamma^t) \times \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \mathcal{T}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}) \times H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))$$

Summary

Semi-Supervised Learning mitigates the requirement for labeled by leveraging unlabeled data.

FixMatch

$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \tau) \times H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))$$

FlexMatch

$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \mathcal{T}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}) \times H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))$$

MarginMatch

$$\mathrm{PM}^t_c(\hat{x}) = z_c - max_{c!=i}(z_i)$$

$$\mathrm{APM}^t_c(\hat{x}) = \frac{1}{t}\sum_{j=1}^{t} \mathrm{PM}^j_c(\hat{x})$$

c acts as “ground truth”

c is the argmax of the model at prediction t

flexmatch

$t`$: | c' | | | ... | c | ... |

t: | | | c' | ... | c | ... |

The model may have different argmax values at different iterations. This phenomenon is captured using averaged pseudo-margins

argmax

argmax

$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \mathbb{1}(\mathrm{APM}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}(\hat{x}_i) > \gamma^t) \times \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \mathcal{T}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}) \times H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))$$

Summary

Semi-Supervised Learning mitigates the requirement for labeled by leveraging unlabeled data.

FixMatch

$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \tau) \times H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))$$

FlexMatch

$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \mathcal{T}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}) \times H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))$$

MarginMatch

c acts as "ground truth"

$$\mathrm{PM}_c^t(\hat{x}) = z_c - max_{c!=i}(z_i)$$

c is the argmax of the model at prediction t

$$\mathrm{APM}_c^t(\hat{x}) = \frac{1}{t}\sum_{j=1}^{t} \mathrm{PM}_c^j(\hat{x})$$

flexmatch

$t`$: | c' | | | ... | c | ...

t: | | | c' | ... | c | ...

The model may have different argmax values at different iterations. This phenomenon is captured using averaged pseudo-margins

argmax

argmax

We employ an additional thresholding based on the APM values

$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \mathbb{1}(\mathrm{APM}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}\boxed{(\hat{x}_i) > \gamma^t}) \times \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \mathcal{T}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}) \times H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))$$

Motivation

- Semi-Supervised Learning is a powerful approach for training models on a large amount of unlabeled data without requiring a large amount of labeled data.

- Since unlabeled data can often be obtained with minimal human labor, any performance boost conferred by SSL often comes with low cost.

FixMatch and FlexMatch

$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \mathcal{T}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}) \quad \times$$
$$H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))$$

FixMatch and FlexMatch

Batch of Unlabeled Data

$$\mathcal{L}_u = \sum_{i=1}^{\boxed{\nu B}} \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \mathcal{T}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}) \times H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))$$

FixMatch and FlexMatch

$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \mathcal{T}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}) \quad \times$$

$$H(\boxed{\hat{p}_\theta(y|\pi(\hat{x}_i))}, p_\theta(y|\Pi(\hat{x}_i)))$$

First generate a pseudo-label by making predictions on weakly-augmented unlabeled examples.

FixMatch and FlexMatch

To ensure high quality pseudo-labels, discard this example if the confidence of the prediction is below a confidence threshold.

$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \boxed{\mathcal{T}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}}) \times H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))$$

FixMatch and FlexMatch

$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \mathcal{T}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}) \quad \times$$

$$\boxed{H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))}$$

Examples that pass the threshold are used during training.

FixMatch and FlexMatch

$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \mathcal{T}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}) \quad \times$$
$$H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))$$

Even with a very high confidence threshold, these methods still introduce errors.

Pseudo-label Data Quality Issues

Examples added to the training set with a wrong pseudo-label for FixMatch and FlexMatch.

These incorrect pseudo-labels are particularly harmful for deep neural networks, which can attain zero training error on any dataset, even on randomly assigned labels [Zhang et al. 2016], resulting in poor generalization capabilities.

How can we improve pseudo-labeled data quality?

- Previous models use confidence solely from the current iteration to enforce quality of pseudo-labels.
  - This provides only a myopic view of the model's behavior (i.e., its confidence) on unlabeled data (at a single iteration) and may result in wrong pseudo-labels even when the confidence threshold is high enough (e.g., if the model is mis-calibrated or overly-confident).

MarginMatch

- We propose MarginMatch, an SSL approach to improve pseudo-labeled data quality by monitoring the model's training dynamics on unlabeled data.

- MarginMatch leverages consistency regularization with weak and strong augmentations and pseudo-labeling.

Margins of Labeled Examples

Margin M at epoch t for $(\mathbf{x}, y) \in \mathcal{D}_{\text{train}}$ [Pleiss et al., 2020]

$$M^{(t)}(\mathbf{x}, y) = \overbrace{z_y^{(t)}(\mathbf{x})}^{\text{assigned logit}} - \overbrace{\max_{i \neq y} z_i^{(t)}(\mathbf{x})}^{\text{largest other logit}}.$$

Averaging the margins across the entire training yields the average margin of a labeled example.

$$\mathrm{AUM}(\mathbf{x}, y) = \frac{1}{T} \sum_{t=1}^{T} M^{(t)}(\mathbf{x}, y)$$

Pseudo-Margins of Unlabeled Examples

Average Pseudo-Margin **(APM)**

c acts as “ground truth”.

$$\mathrm{PM}_c^t(\hat{x}) = z_c - max_{c!=i}(z_i)$$

c is the argmax of the model at iteration t.

$$\mathrm{APM}_c^t(\hat{x}) = \frac{1}{t}\sum_{j=1}^{t}\mathrm{PM}_c^j(\hat{x})$$

t`: c' | | | $\dots$ | c | $\dots$

t: | | c' | $\dots$ | c | $\dots$

The model may have different argmax values at different iterations. This phenomenon is captured using averaged pseudo-margins.

argmax

argmax

- Examples with low APM are potentially mislabeled.

- We use an APM threshold to eliminate erroneous examples:

APM Thresholding

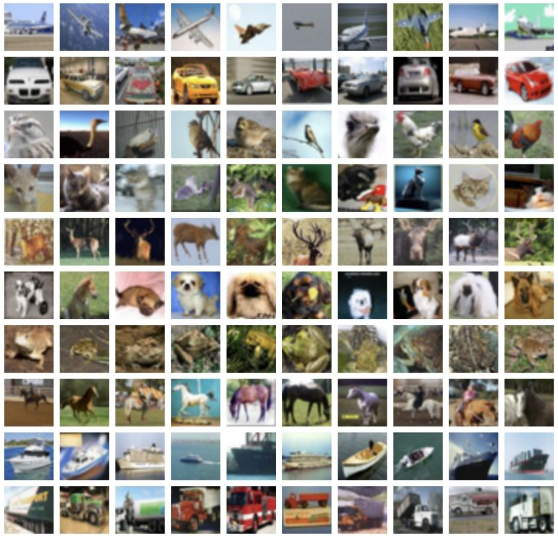
$$\mathcal{L}_u = \sum_{i=1}^{\nu B} \boxed{\mathbb{1}(\mathrm{APM}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))}(\hat{x}_i) > \gamma^t)} \times$$
$$\boxed{\mathbb{1}(\max(p_\theta(y|\pi(\hat{x}_i))) > \mathcal{T}^t_{\hat{p}_\theta(y|\pi(\hat{x}_i))})} \times$$
$$H(\hat{p}_\theta(y|\pi(\hat{x}_i)), p_\theta(y|\Pi(\hat{x}_i)))$$

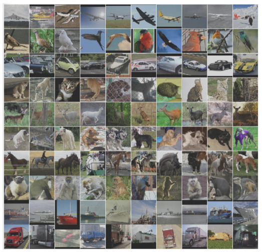
Confidence Thresholding

Evaluation

- Datasets:
  - CIFAR-10
  - CIFAR-100
  - SVHN
  - STL-10

  - ImageNet
  - WebVision





- Performance measures: error rate/accuracy

Results

Dataset	CIFAR-10			CIFAR-100			SVHN			STL-10		
#Labels/Class	4	25	400	4	25	100	4	25	100	4	25	100
Pseuso-Labeling	$74.61_{0.26}$	$46.49_{2.20}$	$15.08_{0.19}$	$87.45_{0.85}$	$57.74_{0.28}$	$36.55_{0.24}$	$64.61_{5.60}$	$25.21_{2.03}$	$9.40_{0.32}$	$74.68_{0.99}$	$55.45_{2.43}$	$32.64_{0.71}$
UDA	$10.79_{3.75}$	$5.32_{0.06}$	$4.41_{0.07}$	$48.95_{1.59}$	$29.43_{0.21}$	$23.87_{0.23}$	$5.34_{4.27}$	$4.26_{0.39}$	$1.95_{0.01}$	$37.82_{8.44}$	$9.81_{1.15}$	$6.81_{0.17}$
MixMatch	$45.24_{2.15}$	$12.76_{1.14}$	$7.13_{0.34}$	$62.15_{2.17}$	$41.51_{1.19}$	$28.16_{0.24}$	$46.18_{1.78}$	$3.98_{0.17}$	$3.5_{0.13}$	$34.15_{1.54}$	$8.95_{0.32}$	$10.41_{0.73}$
ReMixMatch	$5.27_{0.19}$	$4.85_{0.13}$	$4.04_{0.12}$	$47.15_{0.76}$	$27.14_{0.23}$	$23.78_{0.12}$	$4.23_{0.31}$	$3.18_{0.04}$	$1.94_{0.06}$	$31.51_{0.75}$	$8.54_{0.48}$	$6.19_{0.24}$
FixMatch	$7.8_{0.28}$	$4.91_{0.05}$	$4.25_{0.08}$	$48.21_{0.82}$	$29.45_{0.16}$	$22.89_{0.12}$	$3.97_{1.18}$	$\mathbf{3.13_{1.03}}$	$1.97_{0.03}$	$38.43_{4.14}$	$10.45_{1.04}$	$6.43_{0.33}$
FlexMatch	$5.04_{0.06}$	$5.04_{0.09}$	$4.19_{0.01}$	$39.99_{1.62}$	$26.96_{0.08}$	$22.44_{0.15}$	$8.19_{3.20}$	$7.78_{2.55}$	$6.72_{0.30}$	$29.15_{1.32}$	$8.23_{0.13}$	$5.77_{0.12}$
MarginMatch	$\mathbf{4.91_{0.07}}$	$\mathbf{4.73_{0.12}}$	$\mathbf{3.98_{0.02}}$	$\mathbf{36.97_{1.32}}$	$\mathbf{23.71_{0.13}}$	$\mathbf{21.39_{0.12}}$	$\mathbf{3.75_{1.20}}$	$3.14_{1.17}$	$\mathbf{1.93_{0.01}}$	$\mathbf{25.37_{3.58}}$	$\mathbf{7.31_{0.35}}$	$\mathbf{5.52_{0.15}}$

Dataset	ImageNet		WebVision	
	TOP-1	TOP-5	TOP-1	TOP-5
Supervised	48.39	25.49	49.58	26.78
FixMatch	43.66	21.80	44.76	22.65
FlexMatch	42.02	19.49	43.87	22.07
MarginMatch	**41.05**	**18.28**	**43.08**	**21.13**

Results

Dataset	CIFAR-10			CIFAR-100			SVHN			STL-10		
#Labels/Class	4	25	400	4	25	100	4	25	100	4	25	100
Pseuso-Labeling	$74.61_{0.26}$	$46.49_{2.20}$	$15.08_{0.19}$	$87.45_{0.85}$	$57.74_{0.28}$	$36.55_{0.24}$	$64.61_{5.60}$	$25.21_{2.03}$	$9.40_{0.32}$	$74.68_{0.99}$	$55.45_{2.43}$	$32.64_{0.71}$
UDA	$10.79_{3.75}$	$5.32_{0.06}$	$4.41_{0.07}$	$48.95_{1.59}$	$29.43_{0.21}$	$23.87_{0.23}$	$5.34_{4.27}$	$4.26_{0.39}$	$1.95_{0.01}$	$37.82_{8.44}$	$9.81_{1.15}$	$6.81_{0.17}$
MixMatch	$45.24_{2.15}$	$12.76_{1.14}$	$7.13_{0.34}$	$62.15_{2.17}$	$41.51_{1.19}$	$28.16_{0.24}$	$46.18_{1.78}$	$3.98_{0.17}$	$3.5_{0.13}$	$34.15_{1.54}$	$8.95_{0.32}$	$10.41_{0.73}$
ReMixMatch	$5.27_{0.19}$	$4.85_{0.13}$	$4.04_{0.12}$	$47.15_{0.76}$	$27.14_{0.23}$	$23.78_{0.12}$	$4.23_{0.31}$	$3.18_{0.04}$	$1.94_{0.06}$	$31.51_{0.75}$	$8.54_{0.48}$	$6.19_{0.24}$
FixMatch	$7.8_{0.28}$	$4.91_{0.05}$	$4.25_{0.08}$	$48.21_{0.82}$	$29.45_{0.16}$	$22.89_{0.12}$	$3.97_{1.18}$	$\mathbf{3.13_{1.03}}$	$1.97_{0.03}$	$38.43_{4.14}$	$10.45_{1.04}$	$6.43_{0.33}$
FlexMatch	$5.04_{0.06}$	$5.04_{0.09}$	$4.19_{0.01}$	$39.99_{1.62}$	$26.96_{0.08}$	$22.44_{0.15}$	$8.19_{3.20}$	$7.78_{2.55}$	$6.72_{0.30}$	$29.15_{1.32}$	$8.23_{0.13}$	$5.77_{0.12}$
MarginMatch	$\mathbf{4.91_{0.07}}$	$\mathbf{4.73_{0.12}}$	$\mathbf{3.98_{0.02}}$	$\mathbf{36.97_{1.32}}$	$\mathbf{23.71_{0.13}}$	$\mathbf{21.39_{0.12}}$	$\mathbf{3.75_{1.20}}$	$3.14_{1.17}$	$\mathbf{1.93_{0.01}}$	$\mathbf{25.37_{3.58}}$	$\mathbf{7.31_{0.35}}$	$\mathbf{5.52_{0.15}}$

Dataset	ImageNet		WebVision	
	TOP-1	TOP-5	TOP-1	TOP-5
Supervised	48.39	25.49	49.58	26.78
FixMatch	43.66	21.80	44.76	22.65
FlexMatch	42.02	19.49	43.87	22.07
MarginMatch	**41.05**	**18.28**	**43.08**	**21.13**

Results

Dataset	CIFAR-10			CIFAR-100			SVHN			STL-10		
#Labels/Class	4	25	400	4	25	100	4	25	100	4	25	100
Pseuso-Labeling	$74.61_{0.26}$	$46.49_{2.20}$	$15.08_{0.19}$	$87.45_{0.85}$	$57.74_{0.28}$	$36.55_{0.24}$	$64.61_{5.60}$	$25.21_{2.03}$	$9.40_{0.32}$	$74.68_{0.99}$	$55.45_{2.43}$	$32.64_{0.71}$
UDA	$10.79_{3.75}$	$5.32_{0.06}$	$4.41_{0.07}$	$48.95_{1.59}$	$29.43_{0.21}$	$23.87_{0.23}$	$5.34_{4.27}$	$4.26_{0.39}$	$1.95_{0.01}$	$37.82_{8.44}$	$9.81_{1.15}$	$6.81_{0.17}$
MixMatch	$45.24_{2.15}$	$12.76_{1.14}$	$7.13_{0.34}$	$62.15_{2.17}$	$41.51_{1.19}$	$28.16_{0.24}$	$46.18_{1.78}$	$3.98_{0.17}$	$3.5_{0.13}$	$34.15_{1.54}$	$8.95_{0.32}$	$10.41_{0.73}$
ReMixMatch	$5.27_{0.19}$	$4.85_{0.13}$	$4.04_{0.12}$	$47.15_{0.76}$	$27.14_{0.23}$	$23.78_{0.12}$	$4.23_{0.31}$	$3.18_{0.04}$	$1.94_{0.06}$	$31.51_{0.75}$	$8.54_{0.48}$	$6.19_{0.24}$
FixMatch	$7.8_{0.28}$	$4.91_{0.05}$	$4.25_{0.08}$	$48.21_{0.82}$	$29.45_{0.16}$	$22.89_{0.12}$	$3.97_{1.18}$	$\mathbf{3.13_{1.03}}$	$1.97_{0.03}$	$38.43_{4.14}$	$10.45_{1.04}$	$6.43_{0.33}$
FlexMatch	$5.04_{0.06}$	$5.04_{0.09}$	$4.19_{0.01}$	$39.99_{1.62}$	$26.96_{0.08}$	$22.44_{0.15}$	$8.19_{3.20}$	$7.78_{2.55}$	$6.72_{0.30}$	$29.15_{1.32}$	$8.23_{0.13}$	$5.77_{0.12}$
MarginMatch	$\mathbf{4.91_{0.07}}$	$\mathbf{4.73_{0.12}}$	$\mathbf{3.98_{0.02}}$	$\mathbf{36.97_{1.32}}$	$\mathbf{23.71_{0.13}}$	$\mathbf{21.39_{0.12}}$	$\mathbf{3.75_{1.20}}$	$3.14_{1.17}$	$\mathbf{1.93_{0.01}}$	$\mathbf{25.37_{3.58}}$	$\mathbf{7.31_{0.35}}$	$\mathbf{5.52_{0.15}}$

Dataset	ImageNet		WebVision	
	TOP-1	TOP-5	TOP-1	TOP-5
Supervised	48.39	25.49	49.58	26.78
FixMatch	43.66	21.80	44.76	22.65
FlexMatch	42.02	19.49	43.87	22.07
MarginMatch	**41.05**	**18.28**	**43.08**	**21.13**

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

UIC COMPUTER SCIENCE

Results

Dataset	CIFAR-10			CIFAR-100			SVHN			STL-10		
#Labels/Class	4	25	400	4	25	100	4	25	100	4	25	100
Pseuso-Labeling	$74.61_{0.26}$	$46.49_{2.20}$	$15.08_{0.19}$	$87.45_{0.85}$	$57.74_{0.28}$	$36.55_{0.24}$	$64.61_{5.60}$	$25.21_{2.03}$	$9.40_{0.32}$	$74.68_{0.99}$	$55.45_{2.43}$	$32.64_{0.71}$
UDA	$10.79_{3.75}$	$5.32_{0.06}$	$4.41_{0.07}$	$48.95_{1.59}$	$29.43_{0.21}$	$23.87_{0.23}$	$5.34_{4.27}$	$4.26_{0.39}$	$1.95_{0.01}$	$37.82_{8.44}$	$9.81_{1.15}$	$6.81_{0.17}$
MixMatch	$45.24_{2.15}$	$12.76_{1.14}$	$7.13_{0.34}$	$62.15_{2.17}$	$41.51_{1.19}$	$28.16_{0.24}$	$46.18_{1.78}$	$3.98_{0.17}$	$3.5_{0.13}$	$34.15_{1.54}$	$8.95_{0.32}$	$10.41_{0.73}$
ReMixMatch	$5.27_{0.19}$	$4.85_{0.13}$	$4.04_{0.12}$	$47.15_{0.76}$	$27.14_{0.23}$	$23.78_{0.12}$	$4.23_{0.31}$	$3.18_{0.04}$	$1.94_{0.06}$	$31.51_{0.75}$	$8.54_{0.48}$	$6.19_{0.24}$
FixMatch	$7.8_{0.28}$	$4.91_{0.05}$	$4.25_{0.08}$	$48.21_{0.82}$	$29.45_{0.16}$	$22.89_{0.12}$	$3.97_{1.18}$	$\mathbf{3.13_{1.03}}$	$1.97_{0.03}$	$38.43_{4.14}$	$10.45_{1.04}$	$6.43_{0.33}$
FlexMatch	$5.04_{0.06}$	$5.04_{0.09}$	$4.19_{0.01}$	$39.99_{1.62}$	$26.96_{0.08}$	$22.44_{0.15}$	$8.19_{3.20}$	$7.78_{2.55}$	$6.72_{0.30}$	$29.15_{1.32}$	$8.23_{0.13}$	$5.77_{0.12}$
MarginMatch	$\mathbf{4.91_{0.07}}$	$\mathbf{4.73_{0.12}}$	$\mathbf{3.98_{0.02}}$	$\mathbf{36.97_{1.32}}$	$\mathbf{23.71_{0.13}}$	$\mathbf{21.39_{0.12}}$	$\mathbf{3.75_{1.20}}$	$3.14_{1.17}$	$\mathbf{1.93_{0.01}}$	$\mathbf{25.37_{3.58}}$	$\mathbf{7.31_{0.35}}$	$\mathbf{5.52_{0.15}}$

Dataset	ImageNet		WebVision	
	TOP-1	TOP-5	TOP-1	TOP-5
Supervised	48.39	25.49	49.58	26.78
FixMatch	43.66	21.80	44.76	22.65
FlexMatch	42.02	19.49	43.87	22.07
MarginMatch	**41.05**	**18.28**	**43.08**	**21.13**

Conclusion and Future work

- We proposed a novel semi-supervised learning method that improves the pseudo-label quality using training dynamics.

- Future work:
  - We aim to further explore our method in settings when there is a mismatch between the labeled and unlabeled data distributions (i.e., making use of out-of-domain unlabeled data).

Thank you!

Contact: *tsosea2@uic.edu* *cornelia@uic.edu*

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

NSF

UIC COMPUTER SCIENCE