Robust Semi-Supervised Learning when Not All Classes have Labels

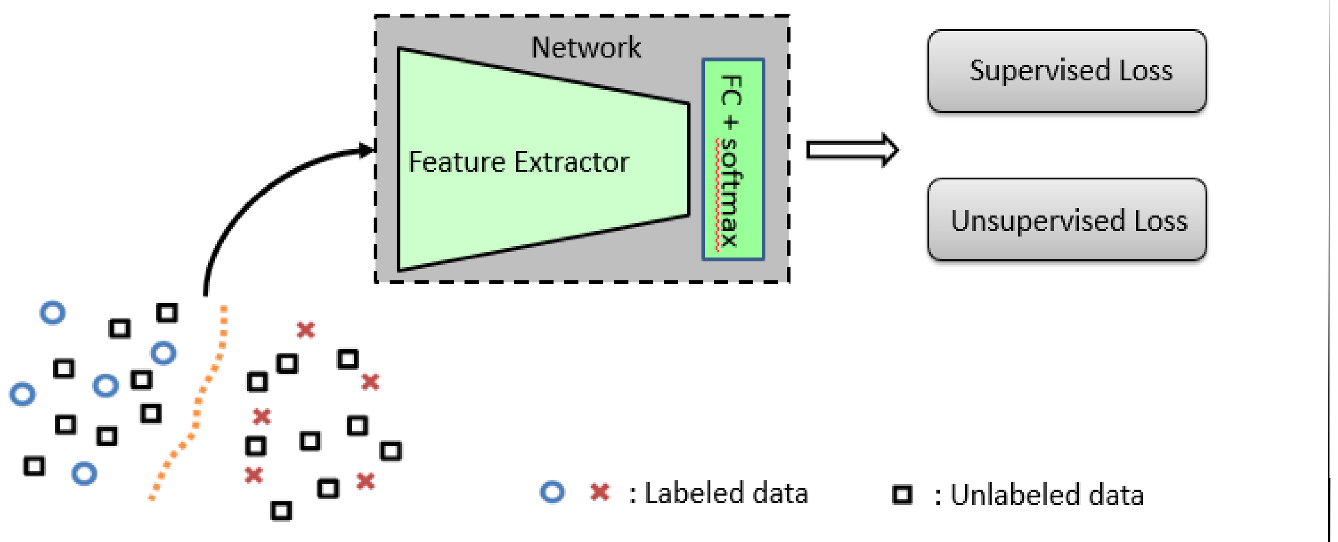
Lan-Zhe Guo*, Yi-Ge Zhang*, Zhi-Fan Wu, Jie-Jing Shao, Yu-Feng Li

National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China

Email: {guolz, zhangyg, wuzf, shaojj, liyf}@lamda.nju.edu.cn

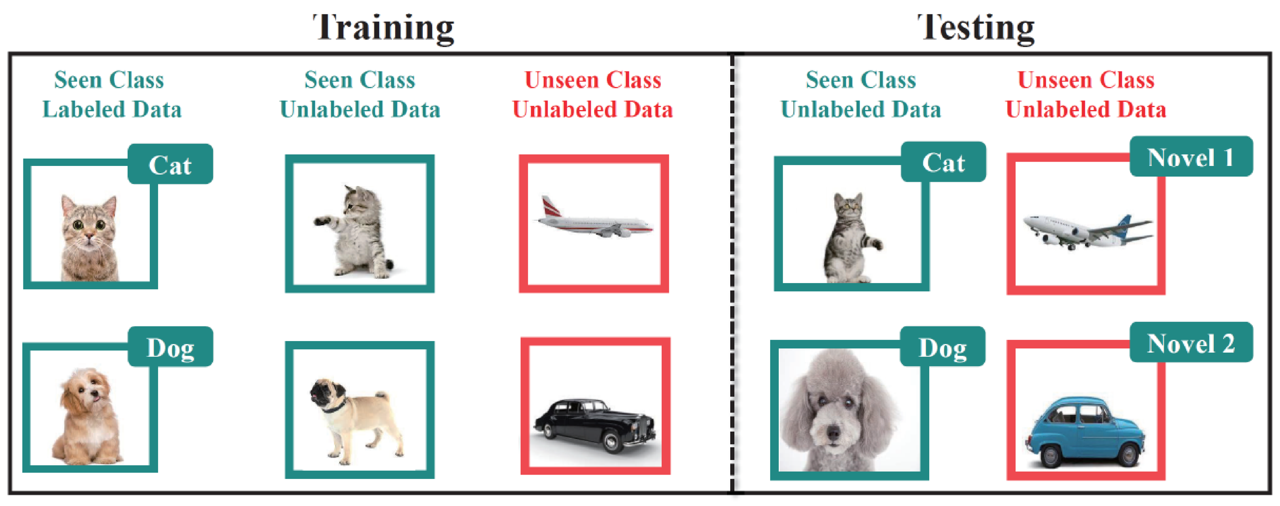
Motivation

SSL: learning from labeled and unlabeled data with a learning model



Traditional SSL assumes: "All classes have labeled samples! "

Real-world datasets often contain classes without labels.



Some classes are difficult to label or newly occurred classes that can not be labeled in time.

Previous Close Set SSL methods suffer performance degradation problem with unseen unlabeled data.

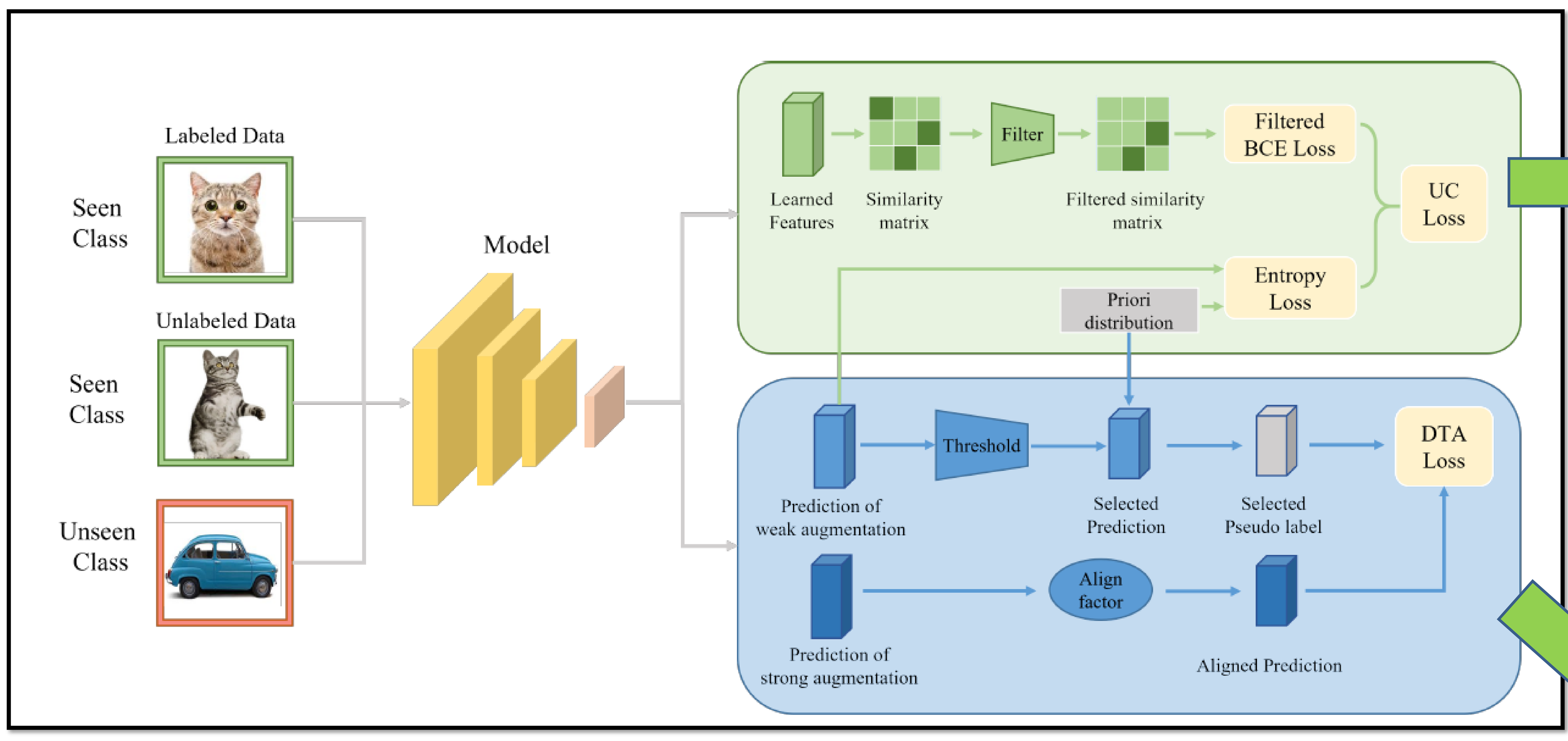
Previous Open Set SSL methods maintain performance on seen classes classification, but can not classify unseen classes.

Previous NCD methods maintain performance on detecting unseen classes, but can not classify seen classes accurately .

Can we design an robust SSL algorithm that can classify both classes with labeled samples and classes without them?

Proposed method

The Framework



Framework of the proposed algorithm. The unsupervised loss can be decomposed into UC loss (lies in the green box) and DTA loss (lies in the blue box).

Two issues need to be addressed

The first one is how to automatically classify unseen classes during model training.

$$\mathcal{L}_{UC}\begin{cases}\mathcal{L}_{FBCE} = -\frac{1}{B}\sum_{b=1}^{B}\log\left(p(\mathbf{x}_b^l)^\top p(\widetilde{\mathbf{x}}_b^l)\right) - \frac{1}{\mu B}\sum_{b=1}^{\mu B}\mathbb{I}(\cos(g(\mathbf{x}_b^u), g(\widetilde{\mathbf{x}}_b^u)) \geq d_k)\log\left(p(\mathbf{x}_b^u)^\top p(\widetilde{\mathbf{x}}_b^u)\right) \\ \mathcal{L}_{ENT} = \mathrm{KL}\Big(\frac{1}{B}\sum_{b=1}^{B}p(\mathbf{x}_i^l) + \frac{1}{\mu B}\sum_{b=1}^{\mu B}p(\mathbf{x}_b^u)\Big\|\mathcal{P}(\mathbf{y})\Big)\end{cases}$$

FBCE loss implements the clustering task by a pairwise similarity ground form while minimizes many error pairings on the basis of BCE loss.

Entropy loss avoids all samples being assigned to the same class.

The second one is how to synchronize the different learning paces caused by the different learning styles between seen and unseen classes.

First we observed the uncertainty between seen classes and unseen classes:

$$\mathrm{U} = \Big(\frac{1}{N_{seen}}\sum_{\mathbf{x}_i\in\mathcal{X}_{seen}}\widehat{p}_i\Big) - \Big(\frac{1}{N_{unseen}}\sum_{\mathbf{x}_j\in\mathcal{X}_{unseen}}\widehat{p}_j\Big)$$

We perform threshold adjustment using U, where threshold is :

$$\begin{cases}\tau & \text{for } x_i \text{ belongs to } X_{seen} \\ \tau-\beta U & \text{for } x_i \text{ belongs to } X_{unseen}\end{cases}$$

A priori knowledge is also used to adjust the logits:

$$F_{ali} = \log\mathcal{P}(X_{select})/\mathcal{P}(y)$$

$$\boxed{\mathcal{L}_{DTA} = \sum_{\mathbf{x}_i\in\mathcal{X}_{seen}\cup\mathcal{X}_{unseen}}\mathbb{I}\left(\widehat{p}_i \geq \tau_i\right)H\left(\widehat{\mathbf{y}}_i, p(\mathcal{A}(\mathbf{x}_i)) + F_{ali}\right)}$$

Main Results

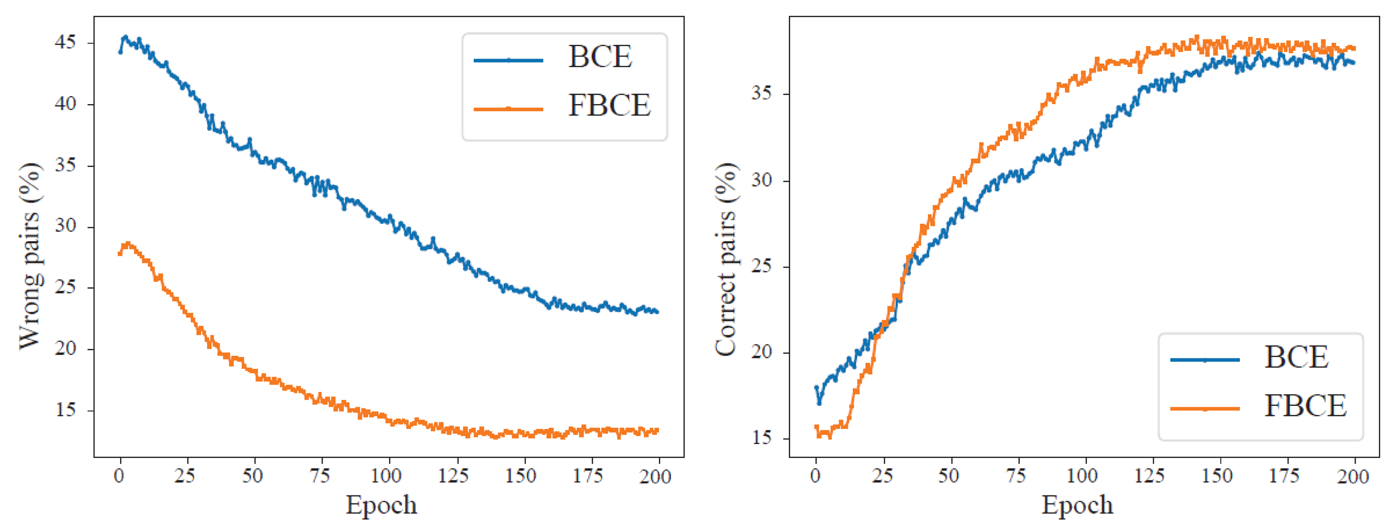
Classification accuracy of compared methods on seen, unseen and all classes.

Classes	Dataset	SSL	Open-Set SSL		NCD			
		Fixmatch	DS3L	CGDL	DTC	RankStats	ORCA	OURS
Seen	CIFAR-10	71.5	77.6	72.3	53.9	86.6	88.2	**89.5**
	CIFAR-100	39.6	55.1	49.3	31.3	36.4	66.9	**68.7**
	ImageNet-100	65.8	71.2	67.3	25.6	47.3	89.1	**91.0**
	Average	59.0	68.0	63.0	36.9	56.8	81.4	**83.1**
Unseen	CIFAR-10	50.4	45.3	44.6	39.5	81.0	90.4	**92.2**
	CIFAR-100	23.5	23.7	22.5	22.9	28.4	43.0	**47.0**
	ImageNet-100	36.7	32.5	33.8	20.8	28.7	72.1	**75.5**
	Average	36.9	33.9	33.6	27.7	46.0	68.5	**71.6**
All	CIFAR-10	49.5	40.2	39.7	38.3	82.9	89.7	**91.3**
	CIFAR-100	20.3	24.0	23.5	18.3	23.1	48.1	**52.1**
	ImageNet-100	34.9	30.8	31.9	21.3	40.3	77.8	**79.6**
	Average	34.9	31.7	31.7	26.0	48.8	71.9	**74.3**

Our proposal achieves significant performance gain compared with NCD method and sota Open-Set SSL methods.

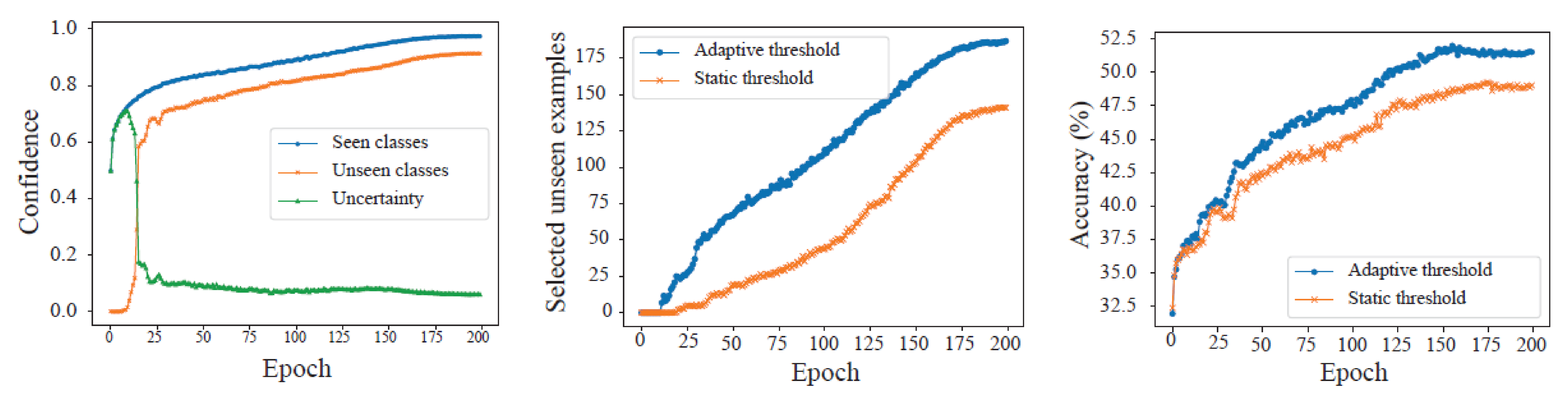
Ablation Study

Our proposal FBCE Loss can reduce the proportion of wrong pairs while increase the proportion of right pairs.



(a) Wrongly selected seen-unseen pairs (b) Correctly selected unseen-unseen pairs

Our proposal Adaptive Threshold achieves significant performance gain compared with UC Model.



(a) Learning difference between seen and unseen classes (b) Selected pseudo-labels for unseen classes (c) Accuracy of pseudo-labels for unseen classes

Conclusion

- ➢We figure out that some classes are difficult to label or newly occurred classes that can not be labeled in time in SSL.
- ➢We proposed NACH algorithm to classify both classes with labeled samples and classes without labeled samples.
- ➢We further aim at giving generalization risk analysis on classes without labeled samples.

More details